The Role of Sialic Acid Acetylerase in the Maintenance of B Cell Self Tolerance

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>McQuay, Amy Brook. 2015. The Role of Sialic Acid Acetylerase in the Maintenance of B Cell Self Tolerance. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:17463957">http://nrs.harvard.edu/urn-3:HUL.InstRepos:17463957</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
The Role of Sialic Acid Acetylemesterase in the Maintenance of B Cell Self Tolerance

A dissertation presented

by

Amy Brook McQuay

to

The Division of Medical Sciences

in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the subject of Immunology

Harvard University
Cambridge, Massachusetts
January 2015
The Role of Sialic Acid Acetylersterase in the Maintenance of B Cell Self Tolerance

Abstract

Sialic acid acetylersterase (SIAE) removes 9-O-acetyl moieties from acetylated sialic acids. The B cell receptor (BCR) inhibitory receptor CD22 cannot bind 9-O-acetylated α2-6-linked sialic acid-containing ligands. Therefore, the removal of these moieties by SIAE is important for inhibition of signaling through the BCR by CD22. Previous studies on Siae-deficient mice revealed a role for SIAE in the maintenance of B cell tolerance.

Deep sequencing of the SIAE exons in patients from several autoimmune cohorts revealed numerous single nucleotide polymorphisms (SNPs). Fluorometric enzymatic assays revealed that about half of these encode catalytically dead proteins while others have reduced activity. Coimmunoprecipitation studies indicated that mutant SIAE associates with wildtype SIAE in a multimer and acts in a dominant-interfering manner to decrease activity of the wildtype protein. We used monoclonal antibodies against different human SIAE epitopes in quantitative western blotting assays to assess whether variant-encoded SIAE proteins are misfolded and if activity and folding correlate. We found that most catalytically dead, disease-associated variants of SIAE are partially misfolded. Circular Dichroism studies were used to further investigate misfolding of mutant SIAE. Preliminary CD data indicate that mutant SIAE proteins are partially misfolded but retain significant structural integrity. These data are
consistent with the finding that mutant SIAE proteins are able to multimerize with wildtype SIAE. We used FPLC to investigate the oligomeric structure of SIAE and found that SIAE exists as a dimer.

We compared the results of the enzymatic assays of SIAE variants to the predictions generated by three commonly used algorithms; Polyphen-2, SIFT, and Provean. We found that the predictions of the algorithms were erroneous for between 11% (PolyPhen-2) and 28% (SIFT) of SIAE variants erroneous predictions for a given variant were often made by more than one algorithm, pointing to a need for non-computational predictive methods for the investigation of the effects of SNPs.
Table of Contents

Abstract ........................................................................................................................................... iii

Acknowledgements ...................................................................................................................... viii

Attributions ................................................................................................................................... xii

List of Abbreviations ................................................................................................................... xiv

Chapter 1

General Introduction ...................................................................................................................... 1

B Cell Receptor Structure ........................................................................................................... 2

B Cell Receptor Signaling ............................................................................................................ 3

Inhibition of B Cell Receptor Signaling ..................................................................................... 11

Differential B Cell Receptor Signaling ....................................................................................... 13

Siglecs ........................................................................................................................................... 13

CD22 ............................................................................................................................................ 18

Sialic Acid Acetylesterase ............................................................................................................ 25

B Cell Tolerance .......................................................................................................................... 30

Genome-Wide Association Studies of Autoimmune Disease ...................................................... 41

Circular Dichroism Spectroscopy for the Examination of Rare Variant ..................................... 46
Algorithms for the Examination of Rare Variants .........................................................52

Chapter 2

Comparison of catalytic activity of rare genetic variants of Siae with predictive algorithms for mutant protein function .................................................................58

Summary .........................................................................................................................59

Introduction ....................................................................................................................60

Methods .........................................................................................................................61

Results ............................................................................................................................63

Conclusions ....................................................................................................................80

Chapter 3

Oligomeric Structure and Structural Analysis of SIAE and Variant-Encoded Mutants....82

Summary .........................................................................................................................83

Introduction ....................................................................................................................84

Methods .........................................................................................................................86

Results ............................................................................................................................92

Conclusions ....................................................................................................................108

Contributions ...............................................................................................................109
Chapter 4

Studies on Siae-deficient BCR knockin mice ................................................................. 110

Summary ....................................................................................................................... 111

Introduction .................................................................................................................. 112

Methods ...................................................................................................................... 116

Results ......................................................................................................................... 120

Conclusions ............................................................................................................... 131

Contributions ............................................................................................................. 131

Chapter 5

Conclusions and Future Directions ............................................................................. 132

References ................................................................................................................... 145
Acknowledgements

To Shiv, thank you for allowing me to join your lab and for not giving up on me. I would not have been able to complete this degree without your patience and support. You’ve been a wonderful mentor, both in guiding my experiments and in helping me to persevere when I wanted to give up. You’ve been a great support through some difficult things that have occurred with my health and in my personal life. Your patience, acceptance, and encouragement during those difficult times were crucial to my ability to finish. You always encouraged me to put my health and my family life first. I appreciate your support of work-life balance. Thank you for advocating for me on multiple occasions. I am so grateful for everything you’ve done to help me through the process of completing my PhD. Your lab is a pleasant and fun place to work, largely because you genuinely care about the people who work there and support their needs and goals both in research and in life outside of the lab. Thank you!

To members of the Pillai lab, past and present, it was a pleasure to work with you all. Thank you all for friendship, advice, and collaboration. To Khaleda, thank you for your kindness and support and for all of your help with everything in lab. To Hamid, thank you for providing so much help with the AM14 mouse project. I could not have done those experiments without all of your help and guidance. Thank you for reading and editing this dissertation. Thank you for being a friend in lab and out. To Yurie, thank you for help with single cell sequencing. To Ian, thank you for providing so much advice on the purification of SIAE for FPLC and CD studies. To Vinay, thank you for your help with FPLC. To Katie, thank you for help with the monoclonal antibody experiments and protein purification for CD studies. To Vasant, thank you for so much help throughout, when I worked on the PKK project, for help expressing SIAE mutants in 293T, immunoprecipitations, and enzyme assays. To Cariappa, thank you for so much...
good advice, both scientific and personal. To Ilka, thank you for being so kind and supportive. You’ve been a great friend! To Kai Xin, thank you for always having a smile or a hug ready. It was great sharing a bay with you!

To the members of my committee – Shannon Turley, Kai Wucherpfennig, and Diane Mathis – thank you for sound counsel, helpful insights, and encouragement. I appreciate your time and your guidance.

To David Cardozo, thank you for being a great sounding board during some tough points. Thank you for guiding me out of a lab and situation that was not working for me and into a better situation. Thank you for great advise, applicable not just to grad school, but to life in general. Thank you for pushing me to take a dance class and spend time taking care of myself when I’d forgotten to do so.

To Susan Perkins, thank you for all that you do. I know how busy you must be keeping everything running smoothly in the program. Yet you somehow are always willing to help out in a pinch and to do so with a smile and words of encouragement.

To my family, every single one of you is wonderful and each of you has helped to grow me into the person that I am. Knowing that such a large group of people was rooting for me has kept me going through some of the harder times. To my mom, everything I am and everything I have ever accomplished I owe to your unwavering love and support. You’ve always been my soft landing place. More than anyone else, you have always believed in me. You are courageous and strong in ways I can only hope to emulate someday. Thank you for teaching me how to love. Thank you for giving me the courage and freedom to be my authentic self. I love you. To my dad, thank you for the left-brain. Thank you for teaching me how to reason and to
troubleshoot. Thank you for talking me down by always reminding me that even the worst-case scenario is usually not so bad. You set me up for a bright future. I love you. To my brothers, Matt and Scott, I am so lucky to have such loving, caring siblings. Scotty, I’m glad for the closeness we have from growing up together, and Matt, I’m glad for the closeness we’ve gained as adults. You are two of the best men I know. I admire you both for living so honestly and openly. I just really fucking love you both. Thank you for having my back.

I have made some of my closest friends during graduate school. To Jana, you’ve listened to me, hugged me, advised me, laughed with me and cried with me. Thank you for being my best friend. Muah! To Rose, John, and Ignacio, you took Hadleigh and I into your home when we needed a safe place to land and a home in Boston for a time. For that I will be forever grateful. Rose you’ve been a wonderful friend throughout these grad school years. You’re kind, generous, and one of the sweetest people I’ve ever known. To Keturah, thank you for the laughs and support during grad school. Thank you for helping Hadleigh and I begin to rebuild our lives. You have been a true friend. To Beckie, thank you for being such a sweetheart. I’ve always enjoyed our conversations and it was so nice to have a good friend in the Navy Yard. To Beth, you’re awesome! Thanks for many laughs and some good, wholesome fun. To Tingting, you were a breath of fresh air, almost always optimistic, funny and sweet. We both know I would not have made it through the earlier years without you. Thank you. You all are like sisters to me and I love you tremendously.

To Farron, thank you for being a wonderful friend during the most difficult time in my life and as I finished up the writing of this dissertation. You’ve helped to rebuild my confidence and renew my hope for a future full of love, laughter, and all of that other good stuff. I’m so fortunate and glad to have met you when I did. I love you.
To Alan, thank you for the support and encouragement you’ve offered during these grad school years. Thank you for your patience. Thank you for Hadleigh.

To Hadleigh Elise, my beloved Peach, you came into existence at the beginning of the writing process and you’ve filled my life with joy, with purpose, with inspiration, and with more love than I thought possible. Oh how you’ve grown my heart! You are, quite simply, my everything. Being your Mommy is a privilege and a pleasure. I love you madly with every cell in my body, Little Girl, always and forever.

To Patrick, you’ve been my almost constant writing companion. You’ve snuggled me and kept my feet warm. You’ve reminded me to take too frequent breaks for stretching, water, snacks, and walks. You’re the best dog and buddy a girl could have and I love you.
Attributes

Chapter 2

Erin Wei generated the U2OS-hSIAE stably-transfected cell line used to purify the hSIAE-FLAG that was used for the generation of anti-hSIAE monoclonal antibodies. Mutant hSIAE-expression vectors were generated by Vasant Chellappa and Ira Surolia. Kathryn Pedrick helped with the transfections, immunoprecipitations, and quantitative western blots shown in Figures 9 and 10. Dalya Ataca and Vasant Chellappa performed the experiments shown in Figure 11.

Chapter 3

Ian Rosenberg advised on the purification of SIAE protein used for FPLC and circular dichroism studies shown in Figures 12-16. Ian Rosenberg and Vinay Mahajan assisted me with the generation of the FPLC data shown in Figures 12 and 13. Vinay Mahajan generated the SIAE-Fc fusion protein used for FPLC sizing shown in Figure 13. Bunsyo Shiotani advised and provided a protocol for transient transfections used to purify protein for circular dichroism studies. Circular dichroism spectroscopy was performed in the laboratory of Dr. Gregory Verdine.

Chapter 4

AM14 HC mice were generously provided Mark Shlomchik. Hamid Mattoo and Cariappa Annaiah helped with mouse breeding and genotyping. Yurie Sekagami assisted with mouse genotyping. Kerstin Kiefer provided assistance with the experiments shown in Figure 18. These experiments were conducted in the laboratory of Ann Marshak-Rothstein, using their AM14Vκ8 mice, protocols, and reagents. Experiments shown in Figures 19 and 20 were conducted with
Hamid Mattoo. Yurie Sekagami assisted with PCR and sequencing of cDNA generated from single cell sorting.
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AID</td>
<td>Activation-induced cytidine deaminase</td>
</tr>
<tr>
<td>APC</td>
<td>Antigen presenting cell</td>
</tr>
<tr>
<td>BCR</td>
<td>B cell receptor</td>
</tr>
<tr>
<td>BM</td>
<td>Bone marrow</td>
</tr>
<tr>
<td>CASD1</td>
<td>Capsule structure domain containing 1</td>
</tr>
<tr>
<td>CDR</td>
<td>Complementarity determining region</td>
</tr>
<tr>
<td>CD</td>
<td>Crohn’s disease</td>
</tr>
<tr>
<td>CD</td>
<td>Circular Dichroism</td>
</tr>
<tr>
<td>cDC</td>
<td>Conventional dendritic cell</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>CFSE</td>
<td>Carboxyfluorescein succinimidyl ester</td>
</tr>
<tr>
<td>CMAH</td>
<td>Cytidine monophosphate-N-acetylleucaminic acid hydroxylase</td>
</tr>
<tr>
<td>cRSS</td>
<td>Cryptic recombination signal sequence</td>
</tr>
<tr>
<td>CSR</td>
<td>Class switch recombination</td>
</tr>
<tr>
<td>DAMP</td>
<td>Damage-associated molecular pattern</td>
</tr>
<tr>
<td>DC</td>
<td>Dendritic cell</td>
</tr>
<tr>
<td>pDC</td>
<td>Plasmacytoid dendritic cell</td>
</tr>
<tr>
<td>DEL</td>
<td>Duck egg lysozyme</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast protein liquid chromatography</td>
</tr>
<tr>
<td>FPR</td>
<td>False positive rate</td>
</tr>
<tr>
<td>GC</td>
<td>Germinal center</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-Wide Association Study</td>
</tr>
<tr>
<td>HC</td>
<td>Heavy chain (of immunoglobulin)</td>
</tr>
<tr>
<td>HEL</td>
<td>Hen egg lysozyme</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IGF2R</td>
<td>Insulin-like growth factor 2 receptor</td>
</tr>
<tr>
<td>IgH</td>
<td>Immunoglobulin Heavy Chain</td>
</tr>
<tr>
<td>ITAM</td>
<td>Immunoreceptor tyrosine-based activation motif</td>
</tr>
<tr>
<td>ITIM</td>
<td>Immunoreceptor tyrosine-based activation motif</td>
</tr>
<tr>
<td>Kdn</td>
<td>2-keto-3-deoxynononic acid</td>
</tr>
<tr>
<td>L-CPL</td>
<td>Left-handed circularly polarized light</td>
</tr>
<tr>
<td>LSE</td>
<td>Luminal sialic acid O-acetylesterase</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>miRNA</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>MPL</td>
<td>Monophosphoryl lipid A</td>
</tr>
<tr>
<td>MS</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>MZ</td>
<td>Marginal zone</td>
</tr>
<tr>
<td>Neu5Gc</td>
<td>N-glycolyl neuraminic acid</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>nsSNP</td>
<td>Non-synonymous single nucleotide polymorphism</td>
</tr>
<tr>
<td>OVA</td>
<td>Ovalbumin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PEI</td>
<td>Polyethyleneimine</td>
</tr>
<tr>
<td>PSIC</td>
<td>Position-Specific Independent Counts</td>
</tr>
<tr>
<td>PTK</td>
<td>Protein tyrosine kinase</td>
</tr>
<tr>
<td>RA</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>RAG</td>
<td>Recombination activation gene</td>
</tr>
<tr>
<td>R-CPL</td>
<td>Right-handed circularly polarized light</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RS</td>
<td>Recombination sequence</td>
</tr>
<tr>
<td>RSS</td>
<td>Recombination signal sequence</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SHM</td>
<td>Somatic hypermutation</td>
</tr>
<tr>
<td>SIAE</td>
<td>Sialic acid acetylesterase</td>
</tr>
<tr>
<td>SIAT</td>
<td>Sialic acid 9-O-acetyl transferase</td>
</tr>
<tr>
<td>Siglec</td>
<td>Sialic acid-binding immunoglobulin-type lectin</td>
</tr>
<tr>
<td>SLE</td>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>ST6GAL1</td>
<td>Beta-galactoside alpha-2,6-sialyltransferase 1</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris-buffered saline</td>
</tr>
<tr>
<td>TCR</td>
<td>T cell receptor</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptor</td>
</tr>
<tr>
<td>TPR</td>
<td>True positive rate</td>
</tr>
</tbody>
</table>
Chapter One

General Introduction
In this introduction, the following topics will be discussed: the structure of the B cell receptor, signaling through the B cell receptor, inhibition of B cell receptor signaling, Siglecs, Sialic Acid Acetylerase, B cell tolerance, Genome-Wide Association Studies for autoimmune diseases, Circular Dichroism Spectroscopy, and algorithms used for the prediction of effects of rare variants.

**B Cell Receptor Structure**

The B cell receptor is a multiprotein complex (Figure 1). It consists of two identical membrane-bound immunoglobulin (Ig) heavy chain polypeptides and two identical Ig light chain polypeptides, which together make up the antigen-binding part of the receptor, and associated chains described below which contribute to signaling. The heavy chain polypeptides are held together by two disulfide bonds and each heavy chain is bound to a light chain by a single disulfide bond. Each heavy and light chain contains an amino-terminal variable region, which combined together make up the antigen recognition portion of the receptor, and a C-terminal constant region. The constant region of the Ig heavy chain contributes to the effector function of the B cell receptor, in the membrane-bound form, as well as the effector function of secreted immunoglobulins. Secreted immunoglobulins are also referred to as antibodies. The membrane-bound form of Ig heavy chains also contains transmembrane regions and short cytoplasmic tails but these cytoplasmic tails lack the ability to transduce signals. Signaling through the B cell receptor is conveyed through its associated Igα/Igβ heterodimer. The Igα/Igβ heterodimer is held together by a disulfide bond but associates with the rest of the receptor complex non-covalently. The longer cytoplasmic tails of the Igα and Igβ chains each contain one
immunoreceptor tyrosine-based activation motif (ITAM). The ITAM motif is an 14-16-amino acid sequence (YxxL/Ix_{(6-8)}YxxL/I) that contains tyrosine residues that can be phosphorylated by intracellular Src-family tyrosine kinases. Phosphorylated ITAMs can recruit and activate Syk family tyrosine kinases, as discussed below.

**B Cell Receptor Signaling**

Signaling through the B cell receptor is critical for the survival, development, maturation, selection, and proliferation of B cells. A mouse deficient in B cells, due to deletion of the Ig µ heavy chain, demonstrated an arrest in B cell development at the pre-B cell stage [1]. In mice in which the Ig variable region could be inducibly deleted, it was shown that when surface BCR was deleted from the surface of mature B cells, these cells underwent rapid cell death [2]. Further, several studies that demonstrate alteration of signaling strength or disruption of normal BCR signaling reveal that such changes impact normal B cell developmental fates [3-6].

B cell antigen receptor signaling is initiated following recognition of multivalent antigen, resulting in BCR aggregation. This aggregation results in the BCR being translocated into lipid rafts, which are rich in members of the Src-family of protein tyrosine kinases (PTKs), including Lyn, Fyn, Blk, or Lck [7]. This close proximity to Src kinases allows for phosphorylation of ITAMs on the Igα/Igβ heterodimer [8]. Lyn is believed to be the primary Src-family kinase member involved in phosphorylation of the Igα and Igβ ITAMs following BCR aggregation. Interestingly, in addition to its positive role in BCR signaling, Lyn is also involved in the attenuation of BCR signaling, downstream of CD22 [9]. This negative role of Lyn will be discussed in a later section.
Figure 1. Structure of the B cell Receptor (BCR). This antigen-binding receptor is composed of membrane-bound immunoglobulin (heavy and light chains) and one disulfide-linked heterodimer Igα/Igβ.
The initial threshold of signaling through the BCR is regulated by two non-receptor-associated molecules, Csk and CD45. C-terminal Src tyrosine kinase (Csk) functions to repress the Src-family PTKs by phosphorylation of a tyrosine in the carboxy terminus which results in a conformational change that blocks the active site [10]. Conversely, CD45 is a transmembrane tyrosine phosphatase that functions primarily to de-repress Src-family PTKs, like Lyn, by phosphorylation of an amino-terminal kinase domain tyrosine (Figure 2) [11].

Following phosphorylation of Igα/Igβ ITAMs, Src-family kinases bind via their Src-homology 2 (SH2) domains, resulting in amplification of ITAM phosphorylation and recruitment of additional effector proteins [12]. While Igα/Igβ ITAMs are initially monophosphorylated, double phosphorylation of a minor proportion of these ITAM tyrosines allows for binding by the SH2 domains of the tyrosine kinase Syk and thereby Syk activation [13]. Syk then couples the BCR to distal signaling by phosphorylating the B cell linker protein, BLNK (also known as SLP-65), which serves as a molecular scaffold for multiple downstream signaling pathways and plays an important role in propagating the signal from the cell surface to the nucleus. In the absence of BLNK, B cells have highly impaired BCR signaling and calcium influx and fail to recruit PLCγ2 to the surface. BLNK contains five phosphorylation sites used to recruit downstream effector molecules PLCγ2, Vav and Btk and additional downstream linker proteins, Grb2 and Nck (Figure 3) [14-17].
Figure 2. Regulation of Src family kinases by Csk and CD45. Src-family kinases contain two tyrosine residues (represented by red bars). Phosphorylation of the carboxy-terminal tyrosine by Csk results in a conformational change, as a result of an interaction between the inhibitory phosphotyrosine and the SH2 domain, and repression of the Src-family kinase. Phosphorylation of the amino-terminal tyrosine by CD45 de-represses the Src-family kinase.
Figure 3. B cell receptor signaling. See text for details. Taken from reference [9].
Early on, BCR signaling is enhanced in mature B cells by the B cell co-receptor complex of CD19, CD21, and CD81. This complex can be co-ligated with the BCR through antigen binding. CD21, also known as complement receptor 2, is a receptor for the C3d fragment of complement. Antigens that have activated complement can crosslink the BCR with CD21 and the co-receptor complex [18]. This cross-linking induces phosphorylation of CD19 by tyrosine kinases associated with the BCR. CD19 is then phosphorylated by Src-family kinases creating binding sites for the SH2 domains of the p85 subunit of phosphatidylinositol 3-OH kinase (PI3K), Vav, and Lyn (Figure 3). This localizes PI3K to its lipid substrates in the membrane [19]. The role of the tetraspanin CD81 is thought to be as a chaperon for the co-receptor complex but tetraspanins including CD81 also participate in integrin binding [20]. CD19 is expressed on B cells early in development, before CD21 and CD81 expression, and can contribute to signaling in immature B cells in the absence of co-ligation of CD21 to the BCR [21]. B cells from mice lacking CD19 do not proliferate and do not have fully activated intracellular signaling in response to BCR cross-linking [22].

Activation of Bruton’s tyrosine kinase (Btk), a TEC family PTK is also critical for proper BCR signaling [23]. Activation of Btk is dependent upon the action of PI3K. PI3K phosphorylates phosphatidylinositol 4,5-bisphosphate (PIP$_2$), which is ubiquitous in the plasma membrane, to create phosphatidylinositol 3,4,5-triphosphate (PIP$_3$). The PH domain of Btk then binds to PIP$_3$. This binding is required both for the recruitment of Btk to the membrane and for its activation [24]. The activation of Btk after its recruitment to the membrane depends on its phosphorylation by Src family kinases (Figure 3).

Downstream of activation of Lyn, Syk, and Btk and the phosphorylation of BLNK, pathways often converge and interact, resulting in some redundancy in the pathways leading to
the activation of many transcription factors. These transcription factors, including NFκB, NFAT, MAX, AFT-2, Jun, c-Myc, and Elk-1, effect the necessary changes in gene expression which control the survival, proliferation, development, and differentiation of B cells. While some of the precise mechanisms and interactions remain unresolved, a simplified discussion of these downstream pathways follows (see Figure 3).

The phosphodiesterase Phospholipase C Gamma 2 (PLCγ2) plays an important role in BCR signal transduction through several important pathways resulting in activation of all of the previously mentioned transcription factors. While PLCγ2 contains a PH domain and could be recruited by PIP3, recruitment by BLNK through its SH2 domain appears to be more important [16]. Although Btk recruitment requires binding through the PH domain by PIP3, BLNK binding through the Btk SH2 domain is important as dual phosphorylation of PLCγ2 by Syk and Btk is required for its maximum activation [16, 17]. Upon activation, PLCγ2 cleaves phosphoinositide PI(4,5)P2 into the second messengers inositol triphosphate (IP3) and diacylglycerol (DAG). Production of IP3 opens calcium channels and results in the release of Ca2+ intracellular stores, primarily from the endoplasmic reticulum, and influx of extracellular Ca2+. Increased calcium results in activation of calmodulin, which in turn binds and activates the phosphatase calcineurin. Calcineurin then activates the nuclear factor of activated T cells, NFAT [25, 26]. Along with increases in intracellular Ca2+ levels, DAG, which remains associated with the membrane, activates members of the protein kinase C (PKC) family, a family of serine/threonine kinases. Upstream of NF-κB activation, DAG activates PKCβ, which phosphorylates the adaptor protein CARMA1, allowing it to associate with Bcl-10 and MALT1. This complex then recruits the IκB kinase (IKK) complex, which consists of IKKα, IKKβ, and IKKγ, and TAK1, a MAP3K family member. TAK1 then phosphorylates serine residues on IKKα and IKKβ [27]. The activated
IKK complex then phosphorylates IκB, targeting it for degradation. NF-κB components are held in the cytoplasm by IκB so this degradation allows NF-κB to move to the nucleus [28, 29].

NF-κB activation can occur through another pathway downstream of PIP₃ generation by PI3K. PIP₃ recruits a serine/threonine kinase called Akt through its N-terminal PH domain, resulting in a conformational change that allows phosphorylation and activation of Akt [30-32]. Following its activation, Akt can then translocate to the cytoplasm and nucleus where it activates IKK [9]. Akt also has other important cytoplasmic substrates that affect B cell survival and proliferation, including Bad, a pro-apoptotic member of the Bcl-2 family [33] and glycogen synthase kinase-3 (GSK3), a kinase involved cell cycle regulation [34, 35], among others.

The mitogen activated protein kinase (MAPK) family, including the p38 kinase, c-Jun NH2-terminal kinase (JNK), and extracellular signal-regulated kinase (ERK) play important roles in activation of several transcription factors downstream of BCR signaling. Each of the MAPKs phosphorylate and activate different sets of transcription factors. MAX and ATF2 are activated by p38, ATF2 and Jun are activated by JNK, and c-Myc and Elk1 are activated by ERK [9]. There are integrated mechanisms for the activation of each of the MAPK proteins and their activation is differentially dependent on the PTKs Lyn, Syk, and Btk. Activation of ERK is completely dependent upon Syk activation while sustained ERK activity is dependent on Btk as well and Lyn activity seems to be unnecessary [36]. ERK activation occurs through the synergistic activities of Grb2 activation of Ras signaling through its recruitment of the guanine nucleotide exchange factor (GEF), SOS, and DAG activation of PKC, with the requirement for PKC activation having a greater effect, while there is no apparent requirement for increased Ca²⁺[37]. JNK activity is completely abolished in the absence of Syk or Btk activity and thus requires both, while Lyn activity is unnecessary [36]. JNK activation is completely dependent
upon Rac1 activation by the Rac1 GEF, Vav, and both PKC activation and intracellular calcium [36, 37]. Finally, p38 activity is not absolutely dependent upon Lyn, Btk, or Syk. However, it is completely abolished in Lyn/Syk double deficient cells, indicating that either Lyn or Syk are sufficient for p38 activation [36]. The activity of p38, like JNK, is completely dependent on Rac1 activation by Vav and PLCγ2-dependent signaling but is unaffected in the absence of IP3, suggesting that the requirement for PLCγ2 signaling is through the PKC pathway [37].

**Inhibition of B Cell Receptor Signaling**

Signaling through the BCR is modulated by several receptor-associated proteins and coreceptors. A positive modulator, CD19 was discussed previously. Two well-characterized negative regulators of BCR signaling are CD22 and FcγRIIB. CD22 will be discussed in greater detail in a later section. Briefly, CD22 is a member of the sialic acid-binding immunoglobulin-like lectin (SIGLEC) superfamily of lectins. CD22 binds to α2,6-linked sialic acid residues on N-glycans. However, specific physiologic ligands that induce inhibitory signaling through CD22 have not been identified. Upon binding of ligand, CD22 is phosphorylated by the tyrosine kinase Lyn on its three intracellular ITIM motifs [38]. This phosphorylation then results in the recruitment of the SH2-containing tyrosine phosphatase, SHP-1, which then inactivates factors downstream of the BCR [39].

FcγRIIB is a low-affinity receptor for IgG that, upon co-ligation with the BCR, is phosphorylated on its ITIM region by Lyn, like CD22. Upon phosphorylation, FcγRIIB primarily recruits the SH2-containing inositol 5’-phosphatase (SHIP) [9]. SHIP then prematurey terminates phosphoinositide hydrolysis and Ca^{2+} mobilization through the
hydrolysis of PI(3,4,5)P₃ to PI(3,4)P₂ [40] and by dephosphorylating CD19, which abrogates PI3K activation [41]. In addition, SHIP can inhibit Erk activation through Ras signaling by recruitment of the adaptor molecule downstream of kinase (DOK) to the plasma membrane, near the BCR. Upon recruitment, DOK is phosphorylated and then can recruit the Ras GTPase activating protein (RasGAP), which catalyzes the intrinsic GTPase activity of Ras [42].

Two other regulators of BCR signaling are the paired immunoglobulin-like receptors (PIRs) and the myeloid-associated immunoglobulin-like receptors (MAIRs, CD300 family in humans). These each exist in both activating and inhibitory isoforms. The activating PIR-A lacks ITIM motifs that are present in the cytoplasmic tail of PIR-B but extracellular regions of the two isoforms are homologous. PIR-A appears to function independently of the BCR but was shown to associate with the ITAM-containing FCRγ chain in mast cells [43]. PIR-B contains four ITIM motifs that are phosphorylated by Lyn upon co-ligation of PIR-B with the BCR. PIR-B then recruits SHP-1 and SHP-2 through their SH2 domains and these inhibit BCR signaling by dephosphorylating Sky and Btk, preventing PLCγ2 activation [44-46].

The MAIR family of paired activating and inhibitory receptors is the mouse counterpart of the human CD300 family. This family remains somewhat uncharacterized; however, MAIR-I and MAIR-II appear to be paired inhibitory and activating receptors, respectively, while MAIR-IV and MAIR-V appear to be paired activating and inhibitory receptors, respectively. These pairs, like PIR-A and PIR-B have highly homologous extracellular domains. MAIR-I and MAIR-V contain ITIM motifs in their cytoplasmic tails. MAIR-II and MAIR-IV have short cytoplasmic tails with no signaling motifs and both contain charged amino acids in the transmembrane domains, which allow them to associate with ITAM-bearing adaptor proteins. MAIR-I and MAIR-II are each expressed on subsets of B cells but MAIR-IV and MAIR-V are
not expressed on B cells. MAIR-I can has been shown to recruit SHP-1 and SHIP while MAIR-II has been shown to associate with DAP-12, which when phosphorylated can recruit Syk. The MAIR ligands and precise mechanisms of MAIR activating and inhibitory activities are not well defined [9, 47-49].

**Differential BCR Signaling**

As mentioned, BCR signaling plays a complex role in the development and function of B cells, controlling survival and death, anergy, receptor editing, differentiation, antigen processing, and proliferation. The many varied responses to signaling through the BCR are determined by several factors including signal strength, signal duration, co-receptor binding and activation, and environmental factors. The developmental stage of a B cell will also determine the relative expression levels of various co-receptors, signal-transducing proteins and regulatory proteins. The complex nature of, and regulation of, BCR signaling are necessary for the production of the large diversity of surface BCR and secreted antibodies necessary to protect the host from a limitless variety of pathogens while simultaneously allowing for the inactivation or deletion of self-reactive B cells and cells that do not express functional receptors.

**Siglecs**

As described previously, signaling through the B cell receptor (BCR) is attenuated by the sialic acid binding immunoglobulin-like lectin (Siglec) family member CD22 (Siglec-2). Siglecs, as defined in 1998 [50], are a family of type-I transmembrane proteins, and a subset of
Ig-type lectins, characterized by an N-terminal V-set domain, which binds to sialic acids on cell surface glycoconjugates, a variable number of C2-set Ig domains, the most N-terminal of which may also play a role in binding of sialic acids [51], a transmembrane region and a cytosolic tail. Siglecs have similar gene structures and are thought to have arisen by gene duplication [52, 53].

Siglecs have traditionally been divided into two groups based on sequence similarity and evolutionary conservation. In this categorization scheme, Sialoadhesin (Siglec-1), CD22 (Siglec-2), myelin-associated glycoprotein (MAG, or Siglec-4), and Siglec-15 are characterized by being conserved across mammals, but have lower sequence homology to each other. The other group, the CD33-related Siglecs, have greater homology to each other (50-99% sequence identity), are evolving rapidly, and do not have orthologues across mammalian species [51, 54]. The CD33-related Siglecs include CD33 (Siglec-3), and Siglecs-5 through -12, -14, and -16 in humans and CD33 (Siglec-3), Siglec-F (a functional paralog of human Siglec-8), Siglec-E (paralog of human Siglec-9), Siglec-G (paralog of human Siglec-10) and Siglec-H in mice [51, 54].

Siglecs can also be categorized based on features in their cytoplasmic tails and transmembrane domains which are relevant to their functions [51]. In this case, the first group consists of Siglec-1 and Siglec-4, which both lack signaling motifs in their cytoplasmic tails and have neutral transmembrane domains. These Siglecs are thought to mediate adhesion events and Siglec-1 has an extended extracellular region with 16 C2-set Ig domains. The second group is characterized by the presence of ITIM and ITIM-like motifs on their cytoplasmic tails and inhibitory biological functions. This group includes human Siglecs -2, -3, and -5 through -12 and mouse Siglecs-2, -E, -F, and –G. In this group, ligand recognition is thought to make these inhibitory motifs accessible to phosphorylation by Src-family kinases, which then results in
recruitment of tyrosine phosphotases such as SHP-1 and SHP-2. A third group can be characterized by the presence of a positively-charged lysine residue in their transmembrane domains and association with homodimers of DAP-12 through the transmembrane aspartate residues. This association allows these Siglecs to act as positive regulators of cell signaling [55]. It is believed that ligation of these Siglecs leads to conformational changes that enhance accessibility of the cytosolic ITAM motifs of DAP12 (or DAP-10, which may associate with Siglec-15) to tyrosine kinases, facilitating Syk family tyrosine kinase recruitment to DAP12 or PI3K recruitment to DAP-10. This group includes human Siglecs -14, -15, and -16 and mouse Siglecs -3, -15, and -H. Murine Siglec-3 does not contain the ITIM motif seen in the human ortholog but does contain an ITIM-like motif in its cytoplasmic tail. It also contains a positively charged residue in the transmembrane region. However, association with DAP-12 has not been demonstrated [51].

Siglecs have been found to be expressed primarily in immune cells [54] with the exceptions of MAG (Siglec-4), which is expressed at high levels in oligodendrocytes and Schwann cells [51], and Siglec-6, which is expressed in trophoblasts [56]. Individual Siglecs are expressed in a cell type-specific manner and as mentioned, have unique functions [54]. Briefly, a few specific examples of expression and function of Siglecs: Sialoadhesin (Siglec-1) is expressed on macrophages and has been shown to have roles in the phagocytic uptake of sialylated pathogens and in regulating autoimmune responses and inflammation [57]. Siglec-8 (murine Siglec-F) is expressed on eosinophils and plays a role in eosinophil apoptosis [58, 59]. Siglec-10 (murine Siglec-G) is expressed primarily on B cells as well as monocytes and dendritic cells and plays roles in maintenance of tolerance in B cells, the modulation of B1a B cell activation, and attenuating signaling in response to damage-associated molecular patterns (DAMPs) [60-63].
As the name implies, most Siglecs are thought to bind sialic acid containing ligands. Sialic acids are a family of N- or O- substituted derivatives of neuraminic acid (Figure 4). They are negatively-charged nine-carbon sugars that decorate the ends of glycoconjugates. The first carbon is the carboxylic carbon, carbons 2-6 make up a ring, and carbons 7-9 are an exocyclic side chain. Neuraminic acid (Neu) has an amino group at the C5 position, which can be acetylated to form N-acetyl neuraminic acid (Neu5Ac). The 5-N-acetyl group can then be hydroxylated by the cytidine monophosphate-N-acetylneuraminic acid hydroxylase (CMAH), yielding N-glycolyl neuraminic acid (Neu5Gc). The desamino form of neuraminic acid containing a hydroxyl group at the C5 position is 2-keto-3-deoxynononic acid (Kdn). These four core sialic acids can also be substituted on the hydroxyl groups of carbons 4, 7, 8, or 9 [64]. The many possible modifications generate a large variety of sialic acids.
Figure 4. Structure of sialic acids. Sialic acids are a family of negatively-charged nine-carbon sugars, $N$- or $O$- substituted derivatives of neuraminic acid, which decorate the ends of glycoconjugates. Taken from reference [51].
Generally, Siglecs have been shown to recognize, with low affinity, Neu5Ac sialic acids with α2-3 and α2-6 linkages to galactose. Exceptions include Siglec-7 and Siglec-11, which preferentially bind Neu5Ac α2-8-linked to another Neu5Ac (Neu5Ac(α2–8)Neu5Ac), which is then α2-3-linked to galactose. But, individual Siglecs have different characteristic sialyl specificity profiles [54]. For example, human CD22 preferentially binds Neu5Acα2–6Gal and Neu5Gcα2–6Gal in vitro. However, humans lack Neu5Gc owing to a mutation in the CMAH gene. Murine CD22 binds Neu5Gcα2–6Gal preferentially and Neu5Acα2–6Gal only with much lower affinity [54]. Defining the specific physiologically relevant sialylated ligands for each Siglec has remained difficult.

**CD22**

CD22, and its role in inhibitory signaling in B cells, are of particular interest and importance to our research. CD22 plays a critical role in establishing a threshold for B cell activation. It is expressed on low levels on pre-B cells and immature B cells and at high levels on mature B cells [65]. Expression of CD22 is down-regulated on plasma cells [66]. CD22 is one of the four Siglecs conserved between humans and mice. It has one V-set domain and six C2-set Ig domains in its extracellular region and three ITIM motifs, one ITIM-like motif, an SH2-binding domain, and two ITAM-like motifs in the cytosolic region [51, 67, 68]. The presence of ITAM-like motifs suggests that CD22 may be involved in positive signaling as well [69].

Upon BCR cross-linking, CD22 is rapidly tyrosine phosphorylated [70], a process which has been shown to require the Src kinase, Lyn [38, 71]. This phosphorylation forms docking
sites for the tyrosine phosphatase, SHP-1 [72]. SHP-1 then acts to dephosphorylate several components of the BCR signaling pathway, including Vav, CD19, BLNK [73-75], and the plasma membrane calcium-ATPase 4 (PMCA4), which attenuates BCR signaling through the promotion of Ca^{2+} efflux [76]. CD22 also recruits SHIP in a complex with She and Grb2, which may further inhibit calcium flux initiated through BCR activation (Figure 5) [68, 77, 78].

Other SH2-containing factors recruited by CD22 include Syk, PLCγ2 and PI3K [80-82]. It is unclear what role these factors, typically associated with positive signaling through the BCR, might be playing in the context of CD22 recruitment. However, coupled with the presence of ITAM-like motifs, in vitro studies demonstrating that CD22 cross-linking with anti-CD22 monoclonal antibodies promotes proliferation directly and enhances anti-Ig-stimulated proliferation [72, 83, 84], in vitro demonstration of enhanced intracellular calcium flux in cells treated with anti-Ig following anti-CD22 treatment [85], and a study showing JNK activation following CD22 cross-linking [86], there is indication that CD22 may also function as a positive regulator of BCR signaling in some circumstances [69].

CD22 has been shown to be specific for α2-6-linked sialic acid groups found on N-glycans [100]. The specific glycoprotein ligands containing these moieties relevant to CD22 function in vivo have not been definitively identified. Surface IgM and CD45 have been identified as possible cis ligands for CD22 through immunoprecipitation experiments [70, 101, 102]. However, mutation of a key arginine on CD22 required for sialic acid binding did not affect the immunoprecipitation of these molecules [103]. In addition, IgM and CD45 did not appear to be important ligands in in situ photoaffinity protein-glycan cross-linking studies, which instead showed CD22 associating in homomultimers within clathrin-coated pits [104]. These
studies indicate that interactions between CD22 and CD45 or IgM occur through non-glycan-mediated protein-protein interactions.
Figure 5. Inhibition of B-cell receptor (BCR) signaling by CD22. See text for details. Taken from reference [79].
Yet, cell lines expressing ligand-binding site mutants of CD22 had increased calcium responses to BCR signaling and reduced tyrosine phosphorylation of CD22 and reduced SHP-1 association with CD22 [105]. Similar results were observed when synthetic sialosides were used to abrogate ligand binding to CD22 [106]. These results indicate that sialic acid ligand binding is important for CD22 inhibitory function.

However, *in vivo* studies in CD22 knockout mice indicate that CD22 probably functions predominantly as an inhibitory receptor. There have been several studies using different CD22 knockout mice [65, 87-89]. All of the knockout mice exhibit increased intracellular calcium release in response to BCR signaling, consistent with a role for CD22 in inhibition. However, B cells in these mice have diminished proliferation in response to BCR signaling. While this could be consistent with loss of positive signaling, it likely reflects increased apoptosis induced by hyperstimulation of these CD22-deficient cells *in vitro* and increased turnover *in vivo* [65, 88, 90]. CD22-deficient mice also have decreased numbers of marginal zone (MZ) B cells and impaired T-independent antigen responses [91]. It is possible that the loss of marginal zone B cells is a result of increased BCR signaling strength [92]. Some CD22-deficient mice develop increased levels of multi-clonal, high-affinity antibodies against double-stranded DNA at around 8 months of age, supporting the model that CD22 regulates B cell activation thresholds to prevent autoimmunity in mature B cells [93]. Lyn-deficient [94-97] mice and SHP-1 B cell conditional knockout mice [98] have more striking autoimmune phenotypes. Lyn knockout mice have enhanced calcium flux in response to BCR stimulation, reduced numbers of recirculating B cells and high B cell turnover rates, IgM hyperglobulinemia, circulating autoantibodies, and IgG immune complex deposition in kidneys reminiscent of systemic lupus erythematosus (SLE). SHP-1-deficient mice have increased numbers of B-1a cells, which have been implicated in
autoimmune disease pathogenesis [99], increased serum immunoglobulin, and spontaneous SLE-like increases in anti-DNA antibodies and immune complex glomerulonephritis. Increased numbers of B-1a cells have also been seen in two of the CD22 knockout mouse lines [87, 89].

B cells from knockin mice that express CD22 that lacks the two N-terminal ligand binding domains (CD22ΔI-2) or that has the ligand-binding arginines, R130 and R137, replaced by alanine (CD22AA) have reduced surface CD22, reduced surface IgM expression, and elevated MHC Class II expression [107]. The authors of this study suggest that CD22 surface expression is likely stabilized by ligand binding and that reduced IgM and elevated MHC II surface expression reflect a chronically activated state due to the absence of CD22 inhibition. B cells from these mice also show reduced proliferation in response to BCR stimulation and increased cell turnover, similar to the phenotype observed in CD22-deficient mice. In addition, these mice had a defect in the marginal zone B cell population similar to CD22-deficient mice, indicating that ligand binding is important for the regulation of this population. But, calcium mobilization in B cells from these mice was comparable to wildtype. Similarly, CD22 phosphorylation and SHP-1 recruitment are not reduced in the B cells from these mice. The reason for this difference between cells from these mice and B cell lines expressing ligand-binding defective CD22 is not entirely clear but may reflect the decreased IgM expression or saturating levels of anti-IgM may have been used in these in vitro studies, masking differences that may have otherwise been observed between the mutant and wildtype B cells [105, 107, 108].

Somewhat conflicting results about the importance of sialic acid ligand binding in CD22 function are observed in ST6Gal1-deficient mice. ST6Gal1 is the transferase that adds α2-6-linked sialic acid to the N-acetyl termini of N-glycans to generate the CD22-binding epitope. Unlike CD22-deficient mice, ST6Gal1-deficient mice have reduced calcium mobilization in
response to BCR stimulation [109]. Like mice with defects in ligand binding described above, ST6Gal1-deficient mice have reduced B cell surface expression of CD22 and IgM. However, B cells from these mice do not have elevated levels of MHC Class II surface expression, indicating that reduced IgM and CD22 is not due to chronic activation. Another study showed that ST6Gal1-deficiency results in colocalization of CD22 and IgM, which results in CD22-dependent enhanced endocytosis of IgM [110]. This may account for the reduced calcium mobilization and in fact, B cell calcium flux levels and proliferation in response to BCR signaling are enhanced to CD22-deficient levels in ST6GAL1/CD22 double knockout mice [111, 112]. Like CD22-deficient mice, ST6GAL1-deficient mice have reduced marginal zone B cells and impaired B cell proliferation in response to BCR stimulation. These mice also have attenuated responses to both T-dependent and T-independent antigens.

Ligand-binding defective mice and ligand-deficient mice have some phenotypes which conflict, both with each other and with CD22-null mice, resulting in some difficulty in establishing an exact mechanism by which CD22 predominantly functions as an inhibitory coreceptor of the BCR. In particular, it is difficult to establish the precise role that α2-6-linked sialic acid binding plays in this inhibitory function. It is clear, however, that CD22 plays a key role in regulating BCR signaling and establishing a threshold for B cell activation. Therefore, it seems that defects in CD22 might result in autoimmunity but CD22-deficient mice do not develop overt autoimmunity. Only one line of CD22-deficient mice has been seen to develop isotype-switched, high-affinity anti-DNA antibodies by eight months of age [93]. This difference between CD22-deficient strains may simply reflect different background strains used in the generation of these mice. The mice that express high-affinity autoantibodies were generated using 129/Sv embryonic stem cells and significant portions of the 129/Sv chromosome.
7 still surround the CD22 gene following backcrossing. The other CD22-deficient mice were generated using B6 embryonic cells [113]. Predisposition to autoimmunity, and SLE in particular, has been demonstrated to correlate with the expression of particular portions of the 129/Sv genome on the B6 background [114, 115]. While CD22 polymorphisms have been described which are correlated with strains of autoimmune-prone mice [116], no such polymorphisms have been linked to autoimmunity in humans. It is likely that defects in CD22 are compensated by other levels of regulation of BCR signaling and that effects of CD22 defects are only likely to result in autoimmunity when there are simultaneous defects in other levels of control of tolerance.

**Sialic Acid Acetylerase**

Sialic acid acetylerase (SIAE) is an enzyme that removes the 9-O-acetyl moieties from 9-O-acetylated sialic acids (Figure 6) [117, 118]. This enzyme is relevant to signaling through CD22, and the maintenance of peripheral B cell tolerance, because CD22 is unable to bind 9-O-acetylated α2-6-linked sialic acid-containing ligands [119]. Therefore, we suspected that SIAE may play a role in regulating CD22 signaling and B cell receptor signaling by extension. SIAE was originally purified from rat liver membrane compartments and called luminal sialic acid O-acetylerase (LSE) to distinguish it from a cytosolic enzyme with similar activity [117, 120, 121]. The cDNA was independently purified by two groups from a pluripotent hematopoietic cell line [117] and through a differential expression study in a late pre-B cell line but not in an early pro-B cell line [118]. LSE was shown by immunoelectron microscopy to localize with lysosomes, despite its lack of activity at acidic pH, and was subsequently called lysosomal sialic
acid O-acetylesterase [122]. The same gene encoding LSE was also found to encode a splice variant (CSE) lacking the signal peptide sequence. CSE is targeted to the cytosol and maintains esterase activity [123]. The SIAE protein was later shown to be secreted from overexpressing stably-transfected cells and could access glycosylated proteins of the membrane either in a post-Golgi vesicle or on the cell surface [124]. However, SIAE has not been found to be present in even trace amounts in detailed analyses and annotation of all proteins making up the plasma proteome [125]. Furthermore, follow-up pulse-chase studies to investigate the relevance of secretion as a measure of SIAE function showed that, even in over-expressing transfected cells, SIAE is only minimally secreted, indicating that secretion of SIAE is most likely an artifact of over-expression and not physiologically relevant [126]. Flow cytometric analysis showed that SIAE was expressed intracellularly but not on the cell surface [126].
Figure 6. Modulation of CD22 inhibition of BCR signaling by SIAE and the putative SIAT.

Signaling through CD22 occurs upon binding of CD22 to a ligand containing α2-6-linked sialic acid, but cannot bind to this ligand if the sialic acid is acetylated. Sialic acid 9-O-acetyl transferase (SIAT) adds acetyl groups to sialylated N-glycans. SIAE removes the 9-O-linked acetyl group from alpha-2-6-linked sialic acid, allowing CD22 to bind its ligand. SIAT and SIAE likely access ligands in the
Knockout mice deficient for SIAE have a similar phenotype as CD22-deficient mice [124]. B cells in these mice have enhanced 9-O-acetylation of surface sialic acid. They also exhibit increased signaling through the BCR as demonstrated by accelerated and enhanced calcium release from intracellular stores upon BCR cross-linking. CD22 tyrosine phosphorylation and SHP-1 recruitment by CD22 were decreased in SIAE-deficient mice. These mice had a deficiency of MZ B cells and recirculating bone marrow perisinusoidal B cells. In addition, these mice have alterations in B cell proliferation that are dependent upon BCR signal strength and duration. SIAE-deficient mice spontaneously develop class-switched anti-DNA and anti-chromatin antibodies, as well as IgG immune complex deposits in the kidneys as early as 20 weeks of age whereas CD22-null mice do not develop high titers of anti-self antibodies until 9 months of age and do not develop glomerular immune complex deposition. This stronger autoimmune phenotype in SIAE-deficient than in CD22-deficient mice may indicate that SIAE is important for the function other inhibitory Siglec(s). Siglec G/10 has been proposed as another candidate that might be regulated by SIAE [127]. Rag knockouts reconstituted with hematopoietic stem cells from SIAE-deficient mice demonstrate that these phenotypes are lymphocyte intrinsic and not a result of lack of SIAE expression in other cell types.

In addition to phenotypic similarity with CD22-null mice, SIAE-deficient mice also share phenotypic similarity with CMAH-null mice. In CMAH-null mice, N-glycans contain α2-6-linked Neu5Ac but do not have α2-6-linked Neu5Gc. Murine CD22 does not bind well to Neu5Ac-containing ligands [128]. B cells from CMAH-null mice have enhanced BCR signaling [129], higher levels of 9-O-acetylated sialic acid [124]. CMAH-null mice also had reduced numbers of MZ B cells and circulating perisinusoidal B cells [124].
The presence of, and phenotype in the absence of, SIAE indicates that there must be a sialic acid 9-O-acetyl transferase (SIAT) that adds the acetyl groups to sialylated N-glycans. This transferase would then also be regulating CD22 signaling upstream of SIAE (Figure 6). The CASD1 (capsule structure domain containing 1) protein has been identified as a putative candidate [130] but further studies are required to clarify the specificity of this enzyme.

Based on the SIAE-deficient phenotype of mice, indicating that SIAE does in fact play a role in regulating CD22 binding to ligand, we predicted that mutations in SIAE may predispose humans to autoimmune disease. Genome-wide association studies had not previously revealed any common variants in the SIAE gene among patients with autoimmune disease. Therefore, the Pillai lab sequenced every exon of the SIAE gene to determine if any loss-of-function variants would be found to be enriched among patients with autoimmune disease compared to healthy controls [131]. The original study sequenced the gene of 923 subjects from several autoimmune disease cohorts (including Crohn’s disease, juvenile idiopathic arthritis, multiple sclerosis, rheumatoid arthritis, Sjogren’s syndrome, systemic lupus erythematosus (SLE), type 1 diabetes, and ulcerative colitis) and 648 healthy controls. This study identified 14 previously-unidentified heterozygous non-synonymous single nucleotide polymorphisms (SNPs) in 19/923 subjects. A homozygous non-synonymous SNP (encoding the M89V mutant protein) was identified in 8 of the autoimmune subjects. Of the 648 healthy controls, 17 presented with one of 8 non-synonymous SNPs and none presented with the M89V-encoding gene in a homozygous state. However, in a study of a larger number of autoimmune subjects, Hunt et al. did find the M89V-encoding variant in the homozygous state in equal rates among subjects and healthy controls [132]. Only one SNP, encoding the T312M mutant protein, overlapped between autoimmune subjects and healthy controls. These SNPs were re-created in a FLAG-tagged human cDNA
expression vector in order to examine whether the mutant proteins were functional. Mutant proteins were assessed for catalytic activity and secretion from cells. Ten of the SNPs identified in subjects, and two of the SNPs identified in controls, encoded catalytically dead protein. The SNP encoding the catalytically dead T312M mutant was identified in two patients and one control. All of the catalytically dead mutants also failed to be secreted from transfected 293T cells except one, which was secreted to a reduced degree. The M89V mutant was catalytically normal but was not secreted. All of the defective variants functioned in a dominant-interfering manner in the transfection assay. The odds ratio for inherited defective SIAE alleles was 8.62 for all autoimmune disorders with a two-sided P-value of 0.0002, indicating that rare polymorphic variants in the SIAE gene result in a genetic link to susceptibility to common autoimmune diseases [131]. Additional sequencing of a larger number of subjects has brought the number of rare variants of SIAE up to more than 67 [126]. Additional sequencing has held that confirmed catalytically dead variants of SIAE are linked to autoimmunity with an Odds Ratio of 3.51 and a two-tailed p value by Fisher’s exact test of 0.0077 [126].

**B Cell Tolerance**

Historically, autoimmune disorders have been categorized as either T cell-mediated or B cell-mediated diseases based on the mechanisms of pathogenesis. Diseases were thought of as resulting from T cell-mediated inflammation, auto-antibody-induced tissue destruction or accumulation of immune complexes (macromolecular complexes of antibodies bound to antigen and sometimes including complement components) [133]. This dichotomous view of autoimmune disease pathogenesis has been replaced in recent years with the recognition that B
cells and T cells have cooperative roles in most autoimmune pathogenesis. Self-reactive B cells may capture and present self-antigen to inflammatory T cells in a disease driven by T cells. Alternatively, T cells may provide help to self-reactive B cell, driving positive selection and expansion of B cell clones which then produce pathogenic auto-antibodies [133].

The generation of the B cell specificity repertoire occurs in two broad steps, both of which include a significant degree of random variability, although neither are strictly stochastic. V(D)J recombination, which occurs in developing B cells in the bone marrow, involves the recombining and splicing of Variable, Diversity, and Joining gene segments to produce rearranged genes for the expression of immunoglobulin (Ig) heavy chains, and Variable and Joining segments to produce Ig light chains. The process is mediated by the Recombination Activating Genes (RAG1 and RAG2)-encoded Rag recombinase. The human heavy chain locus contains 51 Variable, 27 Diverse, and 6 Joining gene segments, but they are not all utilized in recombination with the same (strictly stochastic) frequencies. The same is true of usage of Variable and Joining segments for the two light chain loci.

Somatic Hypermutation (SHM) is the induced introduction of mutations, at a rate of $10^{3}$-to $10^{6}$-fold over the background mutation rate, in the variable regions of rearranged immunoglobulin genes. This process is mediated by Activation Induced Cytidine Deaminase (AID) in the Germinal Center (GC) within the spleen and lymph nodes. The mutation is significantly disproportionately targeted to the three Complementarity Determining Regions (CDRs), portions within the greater variable region that make the most contact with antigen.

The random variability introduced during V(D)J Recombination and SHM leads to the production of B cells with a vast number of different specificities thus allowing for B cell
responses to virtually any pathogen that might be encountered. The drawback of this is that the processes inevitably lead to the production of a large number of self-reactive B cells at each step. An estimated 55-75% of all B cells emerging from the bone marrow are self-reactive [134]. These cells must be removed from the repertoire or otherwise tolerated following each diversity-generating step. Because the maintenance of tolerance is of such critical importance, multiple mechanisms exist for educating and eliminating self-reactive B cells.

Central Tolerance

Central B cell tolerance mechanisms that occur in the bone marrow include VH replacement, pre-BCR censoring, receptor editing and, to lesser extents, clonal deletion and anergy induction. Receptor editing is a major mechanism of central tolerance induction. In early B cell development, some common lymphoid progenitors express key transcription factors that direct commitment to the B cell lineage, Kappa E2-Binding Factor (E2A), Early B Cell Factor (EBF), and Paired-Box Protein 5 (Pax5). These drive transcription of the Rag 1 and Rag 2 genes, \( \lambda 5 \) and vPreB surrogate light chain genes, and the \( Ig\alpha \) and \( Ig\beta \) genes [135]. Pro-B cells expressing all of these genes first undergo V(D)J rearrangement of one Ig heavy chain (IgH) locus, beginning with a \( D_H \) to \( J_H \) rearrangement then a \( V_H \) to \( DJ_H \) rearrangement. Nucleotides are added and removed at the junctions so approximately one-third of rearrangements will be successful (maintain triplet coding). The process of \( V_H \) replacement, in which a previously rearranged VDJ\(_{H}\) gene is recombined in a secondary rearrangement with another upstream \( V_{H} \) gene segment [136-141], may rescue cells that have either not initially made a productive rearrangement [142], or may also play a role in rescuing cells that have produced a self-reactive
IgH chain [143]. The mechanism of VH replacement is not well understood, but is known to be mediated by the Rag recombinase and utilizes a cryptic recombination signal sequence (cRSS) within the rearranged VDJ_H[144]. Additionally, the process was recently shown to be mediated by signaling through the BCR in immature B cells [145]. VH replacement seems to occur primarily in pro-B and pre-B cells [146-148] and it is unclear how important it is for tolerance induction.

Once a productive IgH chain V(D)J has been rearranged, it will be expressed and pair with the surrogate light chains and Igα/β dimer on the surface, forming the pre-BCR. Signaling through the pre-BCR allows the cell to survive and proliferate, and mediates allelic exclusion [149]. This positive selection of those cells which express a successfully rearranged light chain occurs in a ligand-independent fashion [150]. Allelic exclusion ensures that a single IgH chain is expressed on each cell, maintaining clonal specificity, and is accomplished through monoallelic activation of heavy chain rearrangement, so that only one allele is undergoing V(D)J recombination at a time, and through feedback inhibition as a result of pre-BCR signaling, which prevents the other allele from rearranging. The determination of which allele rearranges first, and the method by which an allele is prevented from rearranging once a productive arrangement has been made, are both mediated by epigenetic markings which make chromosomes inaccessible [135, 149]. Clonal specificity is one key in maintaining B cell tolerance, by ensuring that self-reactive BCRs are not positively selected along with non-self-reactive BCRs on the surface of a single cell. In addition to allelic exclusion, another mechanism of tolerance maintenance called pre-BCR censoring occurs at this pro-B cell to pre-B cell transition [133, 151, 152]. Surrogate light chain-deficient mice have elevated levels of anti-nuclear antibodies in the serum and mature autoantibody-secreting cells in the periphery [151]. A number of µ heavy
chain transcripts with unusual characteristics can be identified in pro-B cells and are not found in
pre-B cells. These heavy chains can be expressed on the surface in the absence of surrogate or
conventional light chains and have an increased number of positively charged amino acids in the
CDR3 region. When expressed in the Jurkat cell line with the Igα/β signal transduction
molecule, these µ heavy chains induce activation and apoptosis, suggesting that these µ heavy
chains are capable of inducing negative selection at the pro-B to pre-B cell transition [152]. The
specific mechanism by which the pre-BCR removes B cells expressing self-reactive IgH chains
from the repertoire is poorly understood but may involve self-antigen recognition coupled to
deletion [133]. Finally, signaling through the pre-BCR induces silencing of the surrogate light
chain λ5 and νPreB genes [153].

Still in the bone marrow, pre-BCR signaling induces rearrangement on one Igκ light
chain locus, which is the first light chain to rearrange. Light chain rearrangements involve the
joining of V to J gene segments. If a productive rearrangement does not occur on the first Igκ
allele, the second will rearrange. If neither Igκ allele rearranges successfully, the Igλ alleles will
rearrange in succession. When an immature B cell has made a productive Ig light chain
rearrangement, it will be co-expressed on the surface with IgH chain as the complete B Cell
Receptor (BCR). Tonic signaling through the BCR mediates cell survival and down regulates
expression of the Rag genes [135]. A large percentage of immature B cells express BCRs that
are self-reactive [135]. When these self-reactive cells are stimulated through the BCR by self
antigen, they are triggered to revert to the late pre-B cell stage, re-express the Rag proteins and
surrogate light chains, and undergo receptor editing. Receptor editing is the process by which an
auto-reactive light chain undergoes additional rearrangement of an upstream Vκ or Vλ segment
to a downstream Jκ or Jλ segment [154, 155]. This subsequent rearrangement is most likely first
attempted on the κ allele that was self-reactive [135]. The existing self-reactive rearranged VJκ exon is typically deleted. Further rearrangements can continue until a productive rearrangement occurs as long as upstream V and downstream J segments remain. Rearrangement of the λ light chains only occurs in cells undergoing receptor editing and in which the κ deleting elements have been removed from both κ chains [135, 156]. The κ deleting element is a non-coding sequence, similar to the recombination signal sequence recognized by Rag1/2 located downstream of the Constant κ region (Cκ), which, when recombined with a Vκ, deletes the Cκ, preventing expression of a κ light chain [157]. Receptor editing, rather than deletion, is now recognized as the predominant mechanism of central B cell tolerance [134, 158, 159]. Studies have suggested that up to 50% of mature B cells have undergone light chain receptor editing during development [160-163]. However, some cells that undergo receptor editing continue to express multiple different light chains, a phenomenon termed ‘allelic inclusion’. The resulting polyreactive B cells may escape into the mature B cell pool in the periphery, while remaining potentially autoreactive. This may be the case even when the original autoreactive BCR has high-avidity for self antigen [164-166]. Typically, the self-reactive BCR will be diluted on the surface of B cells by the non-autoreactive secondary receptor or the autoreactive receptor will not be expressed on the surface but may be secreted as IgM autoantibodies [167]. Therefore, although receptor editing is an important mechanism in self-tolerance, allelic inclusion may contribute to the escape of some self-reactive B cells to the periphery. While receptor editing is arguably the predominant mechanism of central self-tolerance, and can spare B cells that have already gone through heavy chain rearrangement, a considerable investment of energy, evidence also suggests that some central tolerance is mediated through clonal deletion and anergy.
Like receptor editing, deletion and anergy are dependent upon BCR signaling. Circumstances that determine whether cells are deleted, made anergic, or undergo receptor editing, is not fully predictable. However, avidity seems to play a significant role. Stronger signaling through the BCR seems to promote either deletion or editing, while weaker signaling seems to promote anergy [168-170]. Multivalent antigens, which cross-link the BCR, are more likely to induce receptor editing while soluble antigens are more likely to induce anergy [133, 159, 169, 171]. It is generally accepted that deletion occurs when cells have exhausted the possibility for further editing [168]. However, the relevance for deletion in central tolerance is also supported by studies of mice with deleted recombination sequences (RS), which are the mouse equivalent of the κ deleting elements. B cells in these mice cannot undergo receptor editing and have defective self-tolerance and increased autoantibody production. When RS-deficient mice were crossed to anti-apoptotic Bcl2-transgene expressing mice, the phenotype was even more profound, indicating that receptor editing is complemented by deletion [156]. Deletion can also occur in the periphery and anergy is initiated in the bone marrow and the process is completed in the periphery [135].

*Peripheral Tolerance*

Peripheral B cell tolerance is maintained by several mechanisms, including clonal deletion, anergy, clonal ignorance, antigen-specific inhibition by T regulatory cells, and elimination by the Fas/FasL pathway. Immature B cells migrate from the bone marrow to the spleen and become transitional T1 B cells. These cells mature into T2 transitional B cells and begin to express IgM and IgD. When T1 or T2 B cells encounter new multivalent self-antigen,
not present in the bone marrow, and bind with high avidity, they are most likely clonally deleted [133]. This deletion is rapid and mediated by pro-apoptotic Bcl2 family member, Bim [166, 172].

T1 or T2 cells which encounter antigen and bind with low avidity will be made anergic [133, 166]. Anergy is a condition in which cells are made unresponsive to antigen, but are not deleted. The anergic state is characterized by expression of markers consistent with an arrest at the transitional B cells stage. However, mature B cells which encounter self antigen in the absence of a second signal are made anergic and acquire this phenotype as well [173]. Anergic B cells are short-lived in the periphery, surviving only for about 5 days [174]. This decreased lifespan appears to be the result of competition with non-self-reactive B cells for localization within follicular niches and B Cell Activating Factor (BAFF), rather than due to any intrinsic survival defect within anergic B cells [135, 175-177]. Inability to compete for localization within follicular niches prevents self-reactive, anergic B cells from receiving T cell help, a phenomenon termed “follicular exclusion” [176, 178, 179]. Although anergic B cells continue to express BCR, there are no signals transduced upon binding. Cambier and Getahun propose the BCR as a three-position molecular switch in which unoccupied BCRs remain “off” because ITAMs remain unphosphorylated, acutely occupied BCR are “on” and signal is transduced following biphosphorylation of ITAMs resulting in recruitment of Syk and downstream activation pathway signaling, and chronic BCR occupancy results in ITAM monophosphorylation, engagement of Lyn, and stimulation of inhibitory pathway signaling [173, 175].

Very weakly self-reactive B cells could be innocuous. However, because they are not deleted or made unresponsive, if allowed to receive T cell help and enter into GCs where they
might undergo SHM, they have the potential to become self-reactive with a higher avidity. Therefore, those B cells that bind with very low avidity are made clonally ignorant [133, 166].

Clonal ignorance is established by inhibitory signaling pathways, including the CD22/Siglece pathway. This mechanism of B cell tolerance is not dependent upon BCR signaling, but instead works to establish a threshold for B cell activation. If permitted, activation could allow uptake of self-antigen complexed to foreign peptide, upregulation of CCR7, migration to the T cell zone, presentation of the foreign peptide to T cells, and T-B collaboration. Therefore, although defects in BCR signaling are associated with failed receptor editing, deletion, and anergy, enhanced BCR signaling is observed when defects in the CD22 pathway leading to ignorance are associated with autoimmunity [127, 133].

Another issue that presents in the GC is that non-self-reactive B cells may become self-reactive following SHM [133]. B cells in the GC have pro-apoptotic gene expression profiles, including downregulation of anti-apoptotic Bcl-2 and upregulation of Bim. Therefore, GC B cells are deleted if they do not receive secondary survival signals from T follicular helper (T\textsubscript{FH}) cells [166]. Typically, T\textsubscript{FH} cells would not be presented with self antigen-derived peptides in the GC and would therefore not provide the necessary signal. However, as described above, in rare instances, a B cell may mutate to strong self-reactivity and self-antigen may complex with foreign peptide and self-reactive T\textsubscript{FH} might provide the necessary help for the selection and expansion of self-reactive B cells. It is not entirely clear how this tolerance would be maintained in this very rare scenario. IgG cloning of single B cells from Fc\gammaRIIB-deficient mice revealed that in the absence of the IgG inhibitory receptor, there is an increase in highly poly- and autoreactive IgG+ GC B cells that had undergone SHM, indicating that Fc\gammaRIIB plays a role in eliminating self-reactive GC B cells. In addition, the frequency of autoreactive IgG+ plasma
cells in the spleen and bone marrow was much lower, revealing an FcγRIIB-independent tolerance checkpoint between the GC and plasma cell transition [133, 180].

T regulatory cells are believed to play roles in the maintenance of B cell self-tolerance, either through deletion, or induction of anergy, of self-reactive B cells in the periphery. Enzyme-linked immunosorbent assay (ELISA) studies of recombinant Ig isolated from single mature naïve B cells taken from CD40L-deficient patients revealed high proportions of autoreactive antibodies, indicating that CD40-CD40L interactions between B and T cells are responsible for regulating self-reactive B cells. These patients have a decreased frequency of MHC Class II-restricted CD4+ Treg cells. In addition, Ig isolated from single B cells from a patient deficient in MHC Class II were also analyzed by ELISA and showed that autoreactive B cells in this patient were not deleted [181]. Taken together, this study indicated that Tregs mediate peripheral B cell tolerance through CD40-CD40L and MHC Class II-TCR interaction. In another study, CD4+CD25+ regulatory T cells were pre-activate and then co-cultured with B cells in the presence of B cell activators. B cells underwent cell death in a cell-contact dependent manner, not mediated by Fas/FasL interactions, but relying on the upregulation of granzymes and perforin in the CD4+CD25+ T cells. In addition, the cells death was specific to activated, antigen presenting B cells, indicating that CD4+CD25+ Tregs can regulate B cell survival in an antigen-dependent manner [182].

Fas signaling following Fas ligand (FasL) binding leads to apoptosis. Under normal conditions, B cells activated through the BCR express c-FLIP (cellular Flice-like protein), an inhibitor of Fas activation, and are thus protected from the apoptosis mediated by FasL on activated T cells [135]. Mutations in Fas in the lpr (lymphoproliferation) mouse, and FasL in the gld (generalized lymphoproliferative disease) mouse result in a lupus-like phenotype [183, 184].
In humans, mutations in the genes encoding Fas and FasL are found in the rare autoimmune lymphoproliferative syndrome (ALPS) \cite{185, 186}. MRL-\textit{lpr} mice expressing Fas under the distal \textit{lck} (lymphocyte-specific protein tyrosine kinase) promoter in T and B cells are rescued from the \textit{lpr} phenotype \cite{187}. Fas is upregulated once B cells are activated and is expressed at high levels in germinal center B cells. Activated B cells do not express FasL so Fas-mediated apoptosis of autoreactive B cells must occur through interaction with FasL in \textit{trans} \cite{188}. This elimination is thought to be mediated primarily by CD4\(^+\) T cells \cite{189, 190}. In addition, \textit{lpr} mice lacking B cells due to a deletion of the J\(_{H}\) locus, had a reduced accumulation of CD4\(^+\) memory T cells, indicating that autoreactive B cells deleted through Fas/FasL interactions with CD4\(^+\) T cells would, if not deleted, further activate self-reactive T cells through antigen presentation \cite{191}. Conditional knockout mice in which Fas is deleted specifically in the B cell compartment had enlarged spleens and lymph nodes, high autoantibody titers, and accumulation of CD4\(^+\) T cells \cite{192, 193}. Conditional knockout of Fas specifically in GC B cells reproduces the phenotype, demonstrating that Fas/FasL interactions regulating B and T cell homeostasis initiates in the germinal center \cite{193}. Fas/FasL signaling eliminates low-affinity and bystander cells activated non-specifically by CD40L on activated T helper cells, but not those cells protected by BCR stimulation and c-FLIP \cite{190}. Chronically activated autoreactive B cells from anti-HEL (hen egg lysozyme) BCR- and TCR- transgenic mice are eliminated, whereas naïve HEL-specific B cells proliferated and produced antibody, in the presence HEL-specific CD4\(^+\) T cells. But, in the absence of Fas, these chronically activated B cells were not eliminated, indicating that Fas/FasL interactions are critical for the elimination of self-reactive B cells in the germinal center \cite{189}. 


Genome-Wide Association Studies of Autoimmune Disease

Autoimmunity is the result of any aberrant response of the immune system against self-antigens. Any disease resulting from such an aberrant immune response may be considered an autoimmune disease. There are a large number of clinically diverse autoimmune diseases that target a wide variety of organs, tissues, or cell types. Although aberrant immune responses are often clearly involved, the specific mechanisms behind autoimmune pathogenesis have remained elusive for most autoimmune diseases. Family and twin studies showed early on that there is a clear heritable component to autoimmune disease. The accepted paradigm has been that susceptibility is the result of genetic changes that cause a break in immunological tolerance, increase in autoantigens, and/or increased inflammation and that environmental insults could then lead to autoimmune pathogenesis. It is also generally believed that most autoimmune diseases have complex, non-Mendelian, genetic backgrounds. Some early gene candidate approaches focused on Major Histocompatibility Complex (MHC) genes based on a model in which a general break in tolerance or shift in B or T cell signaling thresholds would pre-dispose to autoimmunity and then polymorphisms in MHC would cause the tissue specificity of the resulting disease [194].

More recently, research has focused on the study of common variants through the use of genome-wide association studies (GWAS). A typical GWAS compares polymorphic patterns across the genome in controls and patients in order to identify variants that show associations with a common complex disease or trait [195]. In general, GWAS results identify common variants that have minimal effects, and low Odds Ratios, for common diseases. Information obtained from GWAS help to provide possible targets for further investigation and may provide some insight into the relative importance of particular targets. However, GWAS results are only
a starting point. Two common approaches for using GWAS data for elucidating biological causes of disease pathogenesis are described by Hu and Daly [196]. In the traditional bottom-up approach, a single variant or gene from GWAS results is chosen and the function of the protein product and the defect caused by the variant are studied to gain insight into the disease pathogenesis. In the top-down approach, many or all of the variants identified by a GWAS are investigated together in order to glean insight into the networks and pathways that are disrupted during disease pathogenesis.

In recent years, many important results for autoimmune disease research have come out of GWAS [197]. A current catalog, updated weekly, of all results from GWAS is available online at http://www.genome.gov/gwasstudies [198]. Over 200 loci have been found to be linked to autoimmune diseases through GWAS [199]. For most autoimmune diseases, the loci with the strongest single-gene effects are in the MHC. A recent GWAS and meta-analysis of two other prior GWAS found 41 genetic loci with linkage to Type I Diabetes, including regions containing the new candidate genes IL10, IL19, IL20, GLIS3, CD69 and IL27 [200]. A meta-analysis of six GWAS for Ulcerative Colitis identified 29 additional loci, in addition to confirming 18 previously identified loci, including regions containing genes for IL1R2, IL8RA-IL8RB, IL7R, IL12B, DAP, PRDM1, JAK2, IRF5, GNA12 and LSP1 [201]. There are fewer identified susceptibility genes for Rheumatoid Arthritis. A GWAS in 2007 identified variants in the MHC locus PTPN22, and a single nucleotide polymorphism (SNP) in linkage disequilibrium with TRAF1 (tumor necrosis factor receptor–associated factor 1) and C5 (complement component 5) [202]. It is not clear which gene allele is, or if both are, involved in RA pathogenesis and there are compelling reasons to believe either might be relevant. Interestingly, for autoimmune disease, many common variants identified through GWAS have been shown to have pleiotropic
effects [203]. Cotsapas et al. analyzed 107 variants identified across 7 autoimmune diseases and found that about half of the variants were commonly linked between multiple diseases. They found that the shared variants within disease clusters were in genes encoding interacting proteins, indicating that genetic variants conferring risk for multiple diseases influence common pathways and mechanisms of disease pathogenesis [204]. This supports long-known tendencies of autoimmune diseases to cluster within families. Another interesting trend is that GWAS are identifying targets relevant to the establishment of signaling thresholds that play important roles in establishing and maintaining self-tolerance in immune cells [194, 205, 206]. While each individual target may have a low effect on disease susceptibility, multiple subtle alterations in threshold maintenance may be additive.

In a recent candidate gene study drawing data from a prior GWAS, Lee et al. identified a non-coding variant in FOXO3A which is associated with the prognosis, but not diagnosis, of Crohn’s Disease (CD), Rheumatoid Arthritis (RA), and malaria, predicting a milder course of CD and RA and a more severe malaria [207, 208]. The group had previously shown that transcriptional differences in the IL-2 and IL-7 signaling pathways correlate with disease prognosis in CD and other diseases. They set out to determine whether genetic variation in other immune pathways, that is not associated with disease development, might correlate with disease prognosis by comparing data from the existing GWAS with allied phenotypic data that they used to identify patients with either aggressive or indolent CD. This led to the identification of the SNP in FOXO3A, which encodes FOXO3, a transcription factor with non-redundant functions in suppressing cytokine production. They show that the prognosis-associated SNP regulates reaccumulation of FOXO3 in the nucleus after it has been exported to the cytoplasm during inflammatory responses by controlling its transcription. Earlier recovery of nuclear FOXO3
occurs with the indolent CD-associated allele is present and triggers a TGFβ1-dependent pathway in monocytes that reduces production of proinflammatory TNFα and increases production of anti-inflammatory IL-10. They also showed that variation at this SNP is associated with prognosis in RA and malaria, both diseases in which TNFα and IL-10 have been implied [207]. This interesting study provides an example of the extension of the usefulness of data garnered from GWAS beyond the more common focus on identifying variants associated with disease susceptibility in order to predict prognosis and possible identify patient-specific targeted therapeutics [208].

Although many important findings have come from GWAS, they have several limitations. Early obstacles to their use, including the high cost of whole-genome sequencing and limited genome-wide marker coverage, have been largely overcome by high-throughput and array-based genotyping advances and the Human Genome, dbSNP, and HapMap projects [209]. However, GWAS still require large numbers of subjects and controls in order to obtain appropriate statistical power. Another issue is that there are few cases in which risk loci identified from GWAS harbor coding variants in genes believed to participate in immune processes. In the majority of cases, the loci harbor only genes with unknown function or non-coding regions [196]. Similarly, due to linkage disequilibrium, many variants within risk loci may have equivalent levels of disease association, making the identification of the causal variant difficult [195, 197, 210]. Candidate gene and family-based linkage studies generally identify alleles that have large effects on disease susceptibility. With the exception of MHC loci, GWAS have only led to the identification of loci with weak to modest disease risk [194]. This modest risk attributable to most identified loci have lead to concerns about “missing heritability” – the associated variants in total only account for a small proportion of total genetic risk [196, 197,
Several causes of this missing heritability have been proposed. Lee et al., propose that the causal variants may not be in sufficient linkage disequilibrium with the SNPs on commercial arrays used for detection or that the effect sizes at individual SNPs are too small to reach significance in GWAS [211]. Other possible explanations are that gene-gene (epistatic) or gene-environment interactions are not detectable by GWAS, that more causal common variants are unaccounted for by arrays or limited sample sizes, and that large-effect rare variants are unidentified by GWAS [196]. This final explanation is called the Rare Variant Hypothesis.

Although GWAS are well-suited for identifying common variants in complex, common disease, they miss genes for which there may be many rare variants that contribute significantly to disease susceptibility [206]. In other words, the Common Disease/Common Variant Hypothesis cannot account for the total genetic risk load [195]. Some estimate that validated common SNPs identified by GWAS may account for only 10-40% of the total inheritance of complex traits [196]. Common variants are likely to have arisen in populations thousands of years ago and have remained common due to low effect risks while rare variants have arisen more recently in evolutionary history, some possibly even arising de novo, and, owing to their greater effect risks, are unlikely to become common in the population [206]. Compensating for the effect of rare variants requires deep resequencing and functional studies, which can be both labor intensive and expensive. Rivas et al. recently undertook a large next-generation resequencing project of 56 genes in loci identified by GWAS to be associated with Crohn’s Disease, using 350 subjects and 350 controls. They then followed up by genotyping 70 rare protein-altering variants in a much larger set of subjects and controls. In this way, they identified additional risk variants in NOD2 and protective variants in IL23R and CARD9 and rare associations in several other genes [212].
No signal has been identified in any GWAS for SIAE. In addition to an ongoing deep-sequencing project in autoimmune subjects and controls, which has identified over 60 rare, non-synonymous variants [126, 131], our lab has undertaken functional studies to determine which of these variants have adverse effects on protein function, and we have compared the results of our functional studies to predictions from three algorithms in order to assess how accurately these algorithms predict the effects of coding changes on protein function. These studies are discussed in greater detail in other parts of this thesis.

**Circular Dichroism Spectroscopy**

Plane-polarized light is made up of two circularly polarized components of equal magnitude and opposite rotation, one left-handed (counter-clockwise) and one right-handed (clockwise). Circular Dichroism (CD) refers to the differential absorption of left-handed circularly polarized light (L-CPC) and right-handed circularly polarized light (R-CPL) by a molecule containing chiral chromophores. A chromophore may be intrinsically chiral by virtue of its structure, or made chiral by virtue of being linked to a chiral center or placed in an asymmetric environment by virtue of the 3-dimentional structure of the molecule [213]. Chiral chromophores are said to be “optically active” because they absorb one CPL state to a greater extent than the other, resulting in a non-zero CD signal. This differential absorption of CPL states varies as a function of wavelength according to the equation \( \Delta A(\lambda) = A(\lambda)_{L-CPL} - A(\lambda)_{R-CPL} \) [213, 214]. CD spectropolarimeters generally report the difference in absorbance of L-CPL and R-CPL in terms of ellipticity (\( \theta \)) in degrees where \( \theta = 32.98 \Delta A \).
CD spectropolarimeters can use three methods to measure the CD effect. In the modulation method, which is the most commonly used, incident radiation is continuously switched between L-CPL and R-CPL components. To accomplish this, plane-polarized light is split into the two circularly polarized components by passage through a modulator (usually piezoelectric quartz crystal tightly fused to a thin plate of isotropic material) subjected to an alternating electric field. The transmitted, switched polarized light is then detected by a photomultiplier [213]. In the direct subtraction method, the absorbances of the two components are measured separately and subtracted from each other. In the ellipsometric method, the ellipticity of the transmitted radiation is measured.

CD is a valuable technique for analyzing protein structures in solution. Proteins contain multiple chromophores of interest including the peptide bond (absorption below 240 nm), aromatic amino acid side chains (absorption between 260-320 nm), and disulphide bonds (weak, broad absorption centered around 260 nm). Non-protein cofactors also absorb at characteristic peaks over a wide spectral range. In addition, ligands with no intrinsic chirality may acquire chirality when bound to protein and may also contribute to the CD signal.

A number of types of information can be obtained from CD spectroscopy. Absorption below 240 nm (called the far UV) is primarily due to the peptide bond. Different types of common regular secondary structure, such as α-helices, β-sheets, and β-turns, have characteristic CD spectra in the far UV region (Figure 7) [215]. Several algorithms exist for estimating the secondary structure composition of a protein based on CD spectra that draw from datasets of CD spectra of proteins with known structures solved by X-ray crystallography [213, 216-218]. An online server (DICHROWEB, http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) has also been established which allows entered CD data to be analyzed by these various algorithms [218, 219].
Figure 7. Typical far-UV curves representing pure protein secondary structures. Solid line, \(\alpha\)-helix; long dashed line, antiparallel \(\beta\)-sheet; dotted line, type I \(\beta\)-turn; cross-dashed line, extended \(3_1\)-helix or poly (Pro) II helix; short-dashed line, irregular structure. Taken from reference [220].
The region between 260-320 nm (called the near UV) contains spectra that arise from the absorption by aromatic amino acids phenylalanine, tyrosine, and tryptophan. Each absorbs at a characteristic wavelength (Figure 8). The shape of a near UV spectrum and the sizes of peaks will depend on the number of each aromatic amino acid present, their mobility, their environment, and their spatial disposition in the protein [213]. Near UV spectra are not sufficiently characterized to reveal significant insights into structure but progress towards that end has been made by the sequential removal of aromatic amino acids by site-directed mutagenesis [221-223]. However, some valuable information about protein tertiary structure can be obtained from near UV spectra by comparing wildtype and mutant forms. Near-UV spectra can also provide valuable information about the “molten globule” [224] thermodynamic state of proteins [225].
Figure 8. Near UV Circular dichroism spectrum for type II dehydroquinase from *Streptomyces coelicolor*. The wavelength ranges corresponding to signals from Phe, Tyr and Trp side chains are indicated. Taken from reference [220].
CD has been used to assess the integrity of cofactor binding sites [226]. Generally, organic cofactors have little to no CD signal when in free solution and are only visible by CD when binding in a protein cofactor-binding site has conferred chirality. Therefore, CD signals in a characteristic spectral region for a cofactor can be used to make predictions about integrity of cofactor binding sites and the binding and interaction between cofactors and protein ligands [213, 214]. Cofactors may absorb within both the near- and far-UV regions but are often well-separated from the characteristic absorption regions of peptide bonds and aromatic amino acid side chains [214].

Similarly, CD can be used to detect conformational changes in proteins themselves, such as those that occur upon ligand or cofactor binding [226]. When used in this way, CD can provide valuable information about how much protein conformation changes upon ligand binding, conformational changes over varying concentrations of ligand, varying temperature. Using time-resolved CD, the rate at which conformational changes occur, can also be monitored [213]. Continuous- or stopped-flow CD methods are also helpful in detecting the special class of conformational changes that occur during the acquisition of the native structure of proteins during biosynthesis. These experiments usually involve studying the refolding of denatured proteins and detect the rates at which secondary and tertiary structure are acquired at the millisecond or sub-millisecond time scale. Interesting results from these studies indicate that smaller proteins (<100 amino acids) likely fold in a rapid two-state mechanism, with no detectable intermediates, while larger proteins fold in multi-stage pathways in which secondary structure is mostly acquired prior to the formation of tertiary structure contacts [213, 227-229].

CD spectroscopy is a relatively low-resolution technique for protein structural studies, compared to X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. The
benefits of CD spectroscopy, in addition to those described above, are that the technique is
significantly less demanding both in terms of time and quantity of protein required (1 milligram
or less). In addition, CD is done in solution and is not destructive to the sample, so the purified
protein can be re-collected and used in other studies. NMR required high concentrations of
purified protein. X-ray crystallography requires that protein be crystallized suitably for
diffraction and is only appropriate for smaller proteins or portions of proteins (≤40 kDa). CD
studies can be helpful prior to X-ray crystallographic studies for confirming the integrity of
purified recombinant protein or integrity of the structure of individual domains of multi-domain
proteins. Used prior to NMR studies, CD can verify that structure is maintained under
conditions, such as high concentration, acidic pH, the presence of certain denaturing agents,
elevated temperatures, or long acquisition times, used for NMR, by confirming that there is not
loss of CD signal in those conditions. Finally, CD studies are possible under conditions that may
be considered to be more physiologically relevant [230].

**Algorithms for the Examination of Rare Variants**

Single nucleotide polymorphisms (SNPs) are the most frequent type of DNA variation in
humans. Non-synonymous SNPs (nsSNPs), those SNPs within coding regions that result in
amino acid changes in the gene protein product, along with SNPs found in gene expression
regulatory regions, are thought to have the greatest impact on human phenotype. Perhaps the
most significant goal of SNP research is to understand the effect that these genetic variants have
on human phenotype differences, and particularly on susceptibility to complex human diseases
[231]. However, not every SNP, or even every nsSNP, will have a functional effect on a protein,
impact gene expression or result in a phenotype. Further complicating SNP research is the sheer number of SNPs present in the human genome, an estimated 10 million [232]. Even the number of SNPs in one candidate gene can be quite significant. Fortunately, only about 1-2% of all SNPs are non-synonymous, which does somewhat limit the number that need to be addressed when studying variations affecting protein sequence [232, 233].

There are many possible effects of amino acid substitutions caused by nsSNPs. Such a change may affect the catalytic site of an enzyme, a ligand-binding site, or result in abnormal folding, instability, or protein aggregation. Amino acid changes may also affect sites of post-translational modifications, such as glycosylation, phosphorylation, or cleavage sites. There are many computational methods available that predict the impact of amino acid substitutions on protein function and are therefore very useful for studying SNPs [232, 234]. Three of the most commonly used are the SIFT (Sorting Intolerant From Tolerant), PolyPhen (Polymorphism Phenotyping), and Provean (Protein Variation Effect Analyzer). We investigated and compared the use of these three algorithms to results obtained from some of our functional and structural studies.

The SIFT algorithm was initially introduced in 2001 and was most recently updated in 2012 [235, 236]. This algorithm is a sequence homology-based tool predicated on the idea that protein function is strongly correlated with protein evolution, so that mutations affecting protein sequence located in more highly conserved regions are expected to be more deleterious to protein function [235]. SIFT works by locating sequences within a database with homology to a query sequence, identifying those sequences for proteins most likely to have similar functions, using this subset of sequences to obtain multiple alignment scores and, finally, calculating the normalized probabilities of all possible substitutions at each position. The algorithm gives an
output in the form of a tolerance score (from 0 to 1) for each substitution with higher scores reflecting substitutions that are less likely to impact the function of the protein. Predictions of “tolerated” substitutions versus those likely to “affect protein function” occur at a cutoff value of 0.05 [235, 237, 238]. SIFT is available on a web server at http://sift.jcvi.org/ or as a standalone download through Amazon Cloud.

The PolyPhen algorithm was introduced as a web server in 2002 [231]. PolyPhen-2 is an updated version that was introduced in 2010 [239]. Like SIFT, PolyPhen-2 is based, in part, on evolutionary conservation of protein sequence (phylogenetics). PolyPhen-2 also incorporates sequence and structure-based predictive features and is much more involved and complex. The algorithm uses eight sequence-based and three structure-based features.

PolyPhen-2 initially tries to identify the query sequence in the UniProtKB/Swiss-Prot database and uses the feature table of the entry in order to determine whether the amino acid variant occurs at a site annotated in the feature table as an active site, binding site, transmembrane region, or other key structural site. It also memorizes all sites that are annotated for future comparison against 3D structures, if known. If a structure is known of homologous 3D structures can be identified, PolyPhen-2 checks to see whether the variant site is in spatial contact with any sites annotated to have functional or structural significance. Mapping the substitution site to 3D structures will also inform as to whether the substitution will affect the hydrophobic core, electrostatic interactions, ligand binding interactions, and other important features.

The alignment pipeline chooses a set of homologous sequences from BLAST+ [240], then constructs and refines a multiple sequence alignment. Reliably aligned sequences are
clustered and only compact clusters are taken into account in the calculation of the Position-Specific Independent Counts (PSIC) scores [241] of the conserved amino acid and variant. PSIC scores reflect the likelihood, based on conservation, that a given amino acid will occupy a specified position in the amino acid sequence. The greater the difference between the PSIC scores of the conserved and variant amino acid, the greater impact the variant is predicted to have on the structure and function of the protein. PSIC score differences of 1.5 or above are considered damaging.

PolyPhen-2 also calculates the effects of amino acid substitution on protein surface area, surface propensity (hydrophobic potential), side chain volume, and a number of other factors related to protein structure and function.

After PolyPhen-2 extracts the various sequence- and structure-based features for the site of a given amino acid substitution, it feeds the information into a Naïve Bayes probabilistic classifier. The classifier was trained using two pairs of data sets. The first, HumDiv, was compiled from all damaging alleles in the UniProtKB database known to effect molecular function to cause Mendelian diseases compared to differences between human proteins and closely related mammalian homologs assumed to be non-damaging. The second pair, HumVar, compares all disease-causing mutations in UniProtKB to all common human nsSNPs without known involvement in disease (and treated as non-damaging). Users can choose between results based on HumDiv or HumVar-trained PolyPhen-2 models. The HumVar results should be used to assess the effects of variants on Mendelian diseases while HumDiv results should be used for the analysis of effects of rare variants on more complex phenotypes [242].
Ultimately, PolyPhen-2 calculates a Naïve Bayes probability that a mutation is damaging along with false positive rate (FPR) and true positive rate (TPR) and assigns a qualitative prediction of benign, possibly damaging, or probably damaging based on pairs of FPR thresholds. PolyPhen-2 uses 5%/10% FPR for the HumDiv model and 10%/20% FPR for the HumVar model as the thresholds for these classifications. If a lack of available data for a given query does not allow a prediction, the output is reported as unknown. More detailed explanations of the PolyPhen-2 algorithm can be found in the most recent review (see especially Supplementary Methods) [239] and at the About tab from the PolyPhen-2 website (http://genetics.bwh.harvard.edu/pph2).

The Provean algorithm was introduced in 2012 [243, 244]. It is considerably simpler by comparison. Provean collects sequences related to the query sequence from the NCBI NR protein database using BLASTP. It then clusters the homologous sequences with global sequence homology of 80% or higher using CD-HIT [245] and chooses the top 30 clusters most closely related to the query to include in the supporting sequence set. A pair-wise delta alignment score is then computed for each sequence compared to the query sequence according to the equation

$$
\Delta(Q,v,S) = A(Q',S) - A(Q,S)
$$

where $Q'$ is the variant of the $Q$ query sequence caused by variant $v$ and $A(P_1, P_2)$ is the semi-global alignment score between two protein sequences $P_1$ and $P_2$, based on the given amino acid substitution matrix BLOSUM62 [246] and gap penalties. The delta scores are averaged within each cluster. Then, the average delta scores for each cluster are again averaged across all
clusters, with each cluster weighted equally, to compute the final unbiased Provean score. If the score is equal or below a pre-set threshold (default threshold is -2.5), it is considered deleterious. The pre-set threshold can be changed in order to change the stringency of detection. Lower thresholds give greater specificity and higher confidence, while higher thresholds give greater sensitivity (greater number of deleterious variants) but lower confidence.
Chapter Two

Comparison of catalytic activity of rare genetic variants of Sialic Acid Acetylerase with predictive algorithms for mutant protein function
Summary

Limited tools are available to examine the function of rare genetic variants linked to disease. Catalytically defective alleles of sialic acid acetyesterase are associated with human autoimmunity and are relatively easy to identify. The analysis of these variants for functionality is also quite straightforward given the existence of a fluorometric enzyme activity assay. We have analyzed 64 rare variant alleles of this gene and have established that approximately half are defective, with disease-associated alleles being exclusively heterozygous loss-of-function variants. A battery of monoclonal antibodies reveal that disease-associated variants are partly misfolded, consistent with both their loss of catalytic activity and their retained ability to associate with the wildtype protein. Functional analyses correlated only partially with the predictive assays Polyphen 2, SIFT, and Provean for variant protein function.
Introduction

With the advent of exome sequencing, a very large number of rare genetic variants are being discovered both in diseased and healthy individuals. One gene that harbors numerous rare genetic variants encodes sialic acid acetylersterase (SIAE), an enzyme that removes 9-\(O\)-acetyl groups from sialic acid and which permits \(\alpha2,6\)-linked sialic acid external to a lactosamine moiety on N-glycans to bind to CD22 [117-119]. CD22 is a member of the Sialic acid-binding immunoglobulin-like lectin family (Siglec), also known as Siglec 2, which inhibits B cell receptor signaling only when it can bind to its sialic acid-containing ligands [113]. Studies on an Siae mutant mouse have revealed an important role for this esterase in the maintenance of immunological tolerance [51, 124, 127, 133].

Rare genetic variants encoding catalytically defective alleles of SIAE are associated with human autoimmunity. All disease-associated variants of SIAE have been shown to be catalytically defective heterozygous alleles [126, 131]. A genotyping-based study failed to reveal an association between SIAE and disease primarily because of the inclusion of a catalytically normal common variant that has since been shown to be functionally competent [132]. Re-examination of the data focusing only on catalytically defective SIAE alleles in this study demonstrated an enrichment of defective SIAE alleles in disease subjects [126]. This gene represents a particularly strong candidate for the evaluation of broader approaches to the analysis of disease-associated rare genetic variants.
Methods

Immunoprecipitations. 293T cells were transfected at 50% confluency on 10-cm tissue culture dishes with 5µg of FLAG-tagged hSIAE expressing plasmid DNA and 33µl of polyethylenimine (Polysciences, Inc. Cat.# 23966) at 1mg/ml concentration. Transfected 293T cells were allowed to expand for 48-72 hours then each plate was lysed in 1 ml of 4°C lysis buffer (1% NP-40, 150mM NaCl, 50mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0). Cell lysates were pre-cleared by ultracentrifugation at 13,000 rpm for 15 minutes at 4°C two times. For immunoprecipitation, 30µl of 10% slurry of Protein G-Agarose (Sigma Cat.# P7700), 200µl of 3% slurry of Protein A-Sepharose (Sigma Cat.# P9424) and 10ug of anti-FLAG or anti-hSIAE monoclonal antibody was added to each 1ml lysate. Lysates were then incubated on a rotator at 4°C overnight. Protein A/G lysate mixture was then centrifuged at 13,000 rpm for 15 minutes. Lysate was removed and beads were washed with cold lysis buffer three times. After the final wash, beads were resuspended in 100µl of 2X Laemmli’s SDS Sample Buffer (Boston BioProducts Cat. #BP-111R).

Quantitative Western Blots. Immunoprecipitated protein samples were boiled for 10 minutes with frequent vortexing, to remove protein from Protein A and G beads, placed on ice for 2 minutes, then centrifuged at 13,000 rpm for 5 minutes to pellet the beads. 40µl of sample buffer was loaded into wells of 10% polyacrylamide mini-gels. Gels were then run at 130V until the 25kDa band of Dual Color Precision Plus Protein Standards (Bio-Rad Cat. #161-0374) ran out of gel (running buffer was Boston BioProducts Cat. #BP-150 at 1X). Gels were transferred to Immobilon-FL PVDF membrane (Millipore Cat. # IPFL00010) at 30V overnight at 4°C (transfer buffer was Boston BioProducts Cat. #BP-190 at 1X with 20% methanol). Membranes were then blocked in 5% milk buffer (in 0.1% Tween-PBS) for 2 hours at room temperature. Membranes
were then incubated with primary monoclonal ANTI-FLAG M2 antibody at 1:1000 dilution (Sigma Cat. #F1804) at 4°C overnight. Membranes were washed with 0.1% Tween-PBS for 15 minutes three times. Membranes were incubated with secondary IRDye 800CW Goat anti-Mouse IgG (H + L) antibody at 1:10,000 dilution (Li-Cor Cat. #926-32210) at room temperature for 1 hour. Membranes were again washed with 0.1% Tween-PBS for 15 minutes three times. Membranes were then analyzed on a Li-Cor Odyssey Imager.

Western blot data were normalized to account for differences in transfection efficiency. We determined the average intensities of bands in the blots using Li-Cor Odyssey software. The average intensities of the untransfected samples that were immunoprecipitated with anti-FLAG antibody were taken as the background and were subtracted from the intensities of the other bands. We then compared the amount of mutant protein immunoprecipitated by an anti-hSIAE antibody to the amount bound by the anti-FLAG antibody. Then this ratio was compared to the ratio of the amount of wildtype SIAE protein immunoprecipitated by the anti-hSIAE antibody to the amount of wildtype SIAE immunoprecipitated by the anti-FLAG antibody. This is represented by the following equation:

\[
\frac{\text{Mutant immunoprecipitated with anti-hSIAE}}{\text{Wildtype immunoprecipitated with anti-hSIAE}} \div \frac{\text{Mutant immunoprecipitated with anti-FLAG}}{\text{Wildtype immunoprecipitated with anti-FLAG}}
\]

SIAE variant functional analysis. Each SIAE variant was recreated by site-directed mutagenesis as a C-terminal Flag-tagged human SIAE cDNA in an expression vector and transfected into 293T cells. Protein was quantified in lysates by immunoprecipitation and quantitative western blotting. Esterase activity of immunoprecipitated SIAE variants was
analyzed by a fluorometric assay using 4-methylumbelliferyl acetate as a substrate (protocol modified from [247]) and normalized to protein levels.

**Use of SIFT.** SIFT results were obtained from the SIFT Sequence website ([http://sift.jcvi.org/www/SIFT_seq_submit2.html](http://sift.jcvi.org/www/SIFT_seq_submit2.html)). The UniRef 90 April 2011 database was selected. Median conservation of sequences was set to 3.00. Sequences with greater than 90 percent identity to the query sequence were removed. Data used was collected June 2013.

**Use of PolyPhen-2.** PolyPhen-2 results were obtained from the batch query option ([http://genetics.bwh.harvard.edu/pph2/bgi.shtml](http://genetics.bwh.harvard.edu/pph2/bgi.shtml)). Query sequence was NP_733746.1. Version v2.2.2r398 was used to access data in July, 2013. Sequence database was UniProtKB/UniRef100 Release 2011_12. Structures database was PDB/DSSP Snapshot 03-Jan-2012. Classifier model was HumDiv. Genome Assembly was GRCh37/hg19.

**Use of Provean.** Provean results were obtained from the batch query option ([http://provean.jcvi.org/human_protein_batch_submit.php](http://provean.jcvi.org/human_protein_batch_submit.php)). Query sequence was NP_733746.1. There were 30 clusters generated from 76 supporting sequences. Results were obtained in June 2013.

**Results**

Disease-associated SIAE alleles have so far proven to be exclusively heterozygous loss-of-function rare genetic variants and they have been shown to have the potential to function in a dominant interfering manner [131]. Although a small fraction of the SIAE expressed in transfected non-lymphoid cells is secreted, there is no evidence for the secretion of this protein *in vivo*. Further, there is no expression of SIAE on the surface of B cells (although it is expressed intracellularly) indicating that this enzyme functions in an intracellular post-Golgi vesicular
location [126]. Nevertheless, because overexpressed and epitope tagged secreted SIAE is relatively easy to purify, we generated FLAG-tagged SIAE and a SIAE-murine Fc fusion protein and purified these proteins from the supernatants of stably transfected cells.

In order to examine whether disease-associated SIAE mutant proteins may be partially misfolded, we immunized mice with purified SIAE and generated a battery of monoclonal antibodies against the full-length protein. Different clones producing anti-SIAE antibodies were confirmed to be distinct by determining that they all had distinct light chains (data not shown). FLAG-tagged wildtype SIAE, S127A SIAE (an active site mutant not expected to be misfolded [248]) and disease-related SIAE mutant proteins were generated by transfecting 293T cells and each protein was immunoprecipitated separately with anti-FLAG antibodies and each of 5 different anti-SIAE monoclonal antibodies (7H2, 1G5, 2F2, 4A4 and 1D11). As seen in Figure 9A, when comparing wildtype SIAE, S127A SIAE and three disease-related SIAE proteins, C196F SIAE, G212R SIAE and F404S SIAE, the 4A4 monoclonal antibody recognized wildtype and S127A SIAE efficiently but did not recognize disease-related SIAE proteins effectively. A similar phenomenon was noted with other disease-related mutations as seen in Figure 9 B, C and D, and summarized in Figure 10 and Table I. In contrast, the 7H2 and 1G5 monoclonal antibodies bound similarly to wildtype and S127A SIAE protein but showed enhanced binding to most of the defective, disease-associated mutants (Figures 9, 10, and Table I). The other two monoclonal antibodies, 2F2 and 1D11 bound to some defective mutations with either enhanced or decreased affinity but bound most mutant SIAE proteins similarly to wildtype SIAE (Figures 9, 10, and Table I). Another interesting observation is that the R479C mutant is bound poorly by all but one of the monoclonal antibodies. This mutant has a comparatively high Odds Ratio for
disease-linkage and may be grossly misfolded while most other mutants are partially misfolded [126].
Figure 9. Anti-SIAE monoclonal antibodies bind SIAE mutants with differing affinities.

Immunoprecipitation of wildtype and mutant SIAE proteins by a battery of monoclonal antibodies followed by quantitative Western blotting with anti-FLAG antibodies. Wildtype SIAE (A-D) was compared with S127A SIAE, C196F SIAE, G212R SIAE, F404S SIAE (A), M89V SIAE, C266G SIAE, Q309P SIAE, R393H SIAE (B), R230W SIAE, R314H SIAE, R479C SIAE (C) and A3G SIAE, T312M SIAE, Y349C SIAE and K400N SIAE (D).
Figure 9 (Continued)

A

<table>
<thead>
<tr>
<th>Untransfected</th>
<th>Anti-FLAG</th>
<th>7H2 anti-hSIAE</th>
<th>Anti-FLAG</th>
<th>1G5 anti-hSIAE</th>
<th>4A4 anti-hSIAE</th>
<th>Anti-FLAG</th>
<th>2F2 anti-hSIAE</th>
<th>1D11 anti-hSIAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wildtype</td>
<td>S127A-hSIAE</td>
<td>Wildtype</td>
<td>G196F-hSIAE</td>
<td>G212R-hSIAE</td>
<td>Wildtype</td>
<td>F404S-hSIAE</td>
<td></td>
</tr>
</tbody>
</table>

Normalized Lysate Data

- anti-FLAG IPs
- 7H2 anti-hSIAE IPs
- 1G5 anti-hSIAE IPs
- 4A4 anti-hSIAE IPs
- 2F2 anti-hSIAE IPs
- 1D11 anti-hSIAE IPs
Figure 9 (Continued)

B

<table>
<thead>
<tr>
<th>hSIAE Mutant</th>
<th>Anti-FLAG</th>
<th>7H2 Anti-hSIAE</th>
<th>Anti-FLAG</th>
<th>1GS Anti-hSIAE</th>
<th>4AA Anti-hSIAE</th>
<th>Anti-FLAG</th>
<th>2F2 Anti-hSIAE</th>
<th>1D11 Anti-hSIAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Wild-type</td>
<td>M89V-hSIAE</td>
<td>C266G-hSIAE</td>
<td>Q309P</td>
<td>R393H</td>
<td>Wild-type</td>
<td>M89V-hSIAE</td>
<td>C266G-hSIAE</td>
</tr>
</tbody>
</table>

**Normalized Lysate Data**

The data shows the normalized lysate expression levels for different hSIAE mutants with and without the anti-FLAG antibodies. The x-axis represents the hSIAE mutants, while the y-axis shows the normalized lysate data.
Figure 9 (Continued)

C

<table>
<thead>
<tr>
<th>Anti-FLAG</th>
<th>7H2 Anti-hSIAE</th>
<th>1G5 Anti-hSIAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untransfected</td>
<td>Wildtype</td>
<td>Wildtype</td>
</tr>
<tr>
<td>R230W-hSIAE</td>
<td>R479C-hSIAE</td>
<td>R314H-hSIAE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anti-FLAG</th>
<th>4A4 Anti-hSIAE</th>
<th>2F2 Anti-hSIAE</th>
<th>1D11 Anti-hSIAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untransfected</td>
<td>Wildtype</td>
<td>Wildtype</td>
<td>Wildtype</td>
</tr>
<tr>
<td>R230W-hSIAE</td>
<td>R479C-hSIAE</td>
<td>R314H-hSIAE</td>
<td>R479C-hSIAE</td>
</tr>
</tbody>
</table>

Normalized Lysate Data

- anti-FLAG IPs
- 7H2 anti-hSIAE IPs
- 1G5 anti-hSIAE IPs
- 4A4 anti-hSIAE IPs
- 2F2 anti-hSIAE IPs
- 1D11 anti-hSIAE IPs

<table>
<thead>
<tr>
<th>hSIAE Mutant</th>
<th>WT</th>
<th>R230W</th>
<th>R314H</th>
<th>R479C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutant with hSIAE antibody</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutant without hSIAE antibody</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wildtype with hSIAE antibody</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wildtype without hSIAE antibody</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

69
Figure 9 (Continued)

D

<table>
<thead>
<tr>
<th></th>
<th>Anti-FLAG</th>
<th>7H2 Anti-hSIAE</th>
<th>1G5 Anti-hSIAE</th>
<th>4A4 Anti-hSIAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untransfected</td>
<td>Wildtype</td>
<td>A3G-hSIAE</td>
<td>T312M-hSIAE</td>
<td>Y349C-hSIAE</td>
</tr>
<tr>
<td></td>
<td>Wildtype</td>
<td>A3G-hSIAE</td>
<td>T312M-hSIAE</td>
<td>Y349C-hSIAE</td>
</tr>
<tr>
<td></td>
<td>Wildtype</td>
<td>A3G-hSIAE</td>
<td>T312M-hSIAE</td>
<td>Y349C-hSIAE</td>
</tr>
<tr>
<td></td>
<td>Wildtype</td>
<td>A3G-hSIAE</td>
<td>T312M-hSIAE</td>
<td>Y349C-hSIAE</td>
</tr>
</tbody>
</table>

Normalized Lysate Data

Mutant with anti-FLAG antibody

hSIAE Mutant

-1.000

0.000

1.000

2.000

3.000

4.000

5.000

6.000

anti-FLAG IPs

7H2 anti-hSIAE IPs

1G5 anti-hSIAE IPs

4A4 anti-hSIAE IPs

2F2 anti-hSIAE IPs

1D11 anti-hSIAE IPs
Figure 10. A comparison of immunoreactivities of various SIAE variant proteins with different monoclonal antibodies.
Table I. Summary of anti-SIAE monoclonal antibody binding to SIAE variant mutant proteins. Data summary represents averages of three repeats of the experiments. Antibody-variant protein binding is listed as either enhanced, reduced, or similar (compared to antibody binding to wildtype SIAE).

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Secretion</th>
<th>Activity</th>
<th>7H2</th>
<th>1G5</th>
<th>4A4</th>
<th>2F2</th>
<th>1D11</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3G</td>
<td>Normal</td>
<td>Normal</td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
</tr>
<tr>
<td>K400N</td>
<td>Normal</td>
<td>Normal</td>
<td><strong>enhanced</strong></td>
<td><strong>enhanced</strong></td>
<td>similar</td>
<td>enhanced</td>
<td></td>
</tr>
<tr>
<td>C196F</td>
<td>Defective</td>
<td>Defective</td>
<td>enhanced</td>
<td>enhanced</td>
<td>reduced</td>
<td>similar</td>
<td>similar</td>
</tr>
<tr>
<td>G212R</td>
<td>Defective</td>
<td>Defective</td>
<td>enhanced</td>
<td>enhanced</td>
<td>reduced</td>
<td>reduced</td>
<td>similar</td>
</tr>
<tr>
<td>R230W</td>
<td>Defective</td>
<td>Defective</td>
<td>enhanced</td>
<td>enhanced</td>
<td>reduced</td>
<td>reduced</td>
<td>similar</td>
</tr>
<tr>
<td>T312M</td>
<td>Defective</td>
<td>Defective</td>
<td>enhanced</td>
<td>enhanced</td>
<td>reduced</td>
<td>similar</td>
<td>similar</td>
</tr>
<tr>
<td>Q309P</td>
<td>Defective</td>
<td>Defective</td>
<td>enhanced</td>
<td>enhanced</td>
<td>reduced</td>
<td>similar</td>
<td>similar</td>
</tr>
<tr>
<td>R314H</td>
<td>Defective</td>
<td>Defective</td>
<td>enhanced</td>
<td>enhanced</td>
<td>reduced</td>
<td>reduced</td>
<td>similar</td>
</tr>
<tr>
<td>R393H</td>
<td>Defective</td>
<td>Defective</td>
<td>enhanced</td>
<td>enhanced</td>
<td>reduced</td>
<td>enhanced</td>
<td>similar</td>
</tr>
<tr>
<td>R479C</td>
<td>Defective</td>
<td>Defective</td>
<td>reduced</td>
<td>reduced</td>
<td>reduced</td>
<td>similar</td>
<td>reduced</td>
</tr>
<tr>
<td>Y349C</td>
<td>Reduced</td>
<td>Reduced</td>
<td><strong>enhanced</strong></td>
<td><strong>enhanced</strong></td>
<td>similar</td>
<td><strong>enhanced</strong></td>
<td>similar</td>
</tr>
<tr>
<td>F404S</td>
<td>Reduced</td>
<td>Reduced</td>
<td>enhanced</td>
<td>similar</td>
<td>reduced</td>
<td>enhanced</td>
<td>similar</td>
</tr>
<tr>
<td>C266G</td>
<td>Defective</td>
<td>Reduced</td>
<td>enhanced</td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
</tr>
<tr>
<td>M89V</td>
<td>Defective</td>
<td>Normal</td>
<td><strong>enhanced</strong></td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
</tr>
<tr>
<td>S127A</td>
<td>Normal</td>
<td>Defective</td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
<td>similar</td>
</tr>
</tbody>
</table>
We have now identified over 70 non-synonymous rare variants of the SIAE gene. Some were identified through deep sequencing of all SIAE exons in patients and healthy controls in our lab, as reported previously [131], others were identified from the Exome Variant Server [126, 249], the 1000 Genomes Project [250], and dbSNP [251]. Functional studies of additional variants identified since our previous publication have been conducted. The variants were recreated in plasmids for transfection and overexpression, and were subsequently immunoprecipitated and analyzed by quantitative western blot and enzyme activity assays (Figure 11). The S127A mutant is an active site variant used as a control. Variants encoding SIAE(E96K), SIAE(Q101K), SIAE(S228C), SIAE(S255A), SIAE(L323F), SIAE(R340H), SIAE(N357S), SIAE(M360K), SIAE(S371L), SIAE(A385T), SIAE(R387H), and SIAE(C443R) are severely catalytically defective. More modest catalytic defects were observed in the variants encoding SIAE(R27H), SIAE(Y31C), SIAE(A171V), SIAE(R210C), SIAE(L366V), SIAE(R393C), SIAE(A394T) and SIAE(M456T). Variants encoding, SIAE(R62H), SIAE(N107K), SIAE(A147V), SIAE(S158F), SIAE(T206I), SIAE(D246N), SIAE(P251L), SIAE(T286M), SIAE(S325P), SIAE(D334N), SIAE(R369K), SIAE(D370H), SIAE(Q428L), SIAE(K434T), SIAE(V459I), SIAE(H472Q), and SIAE(G514A) are catalytically normal or nearly normal.
Figure 11. Analysis of SIAE variants identified in patients with autoimmunity. A and B)

Each SIAE variant found in subjects with autoimmunity was re-created by site-directed mutagenesis in a human SIAE cDNA. Each SIAE variant cDNA was transfected into 293T cells. Variant-encoded proteins were purified from each lysate by immunoprecipitation with anti-Flag antibodies and examined for esterase activity by 4-methylumbelliferyl acetate fluorometric assay. Data in the graphs is presented following normalization for lysate SIAE protein content. (The data in this figure represents the work of Dalya Ataca and Vasant Chellappa.)
Figure 11 (Continued)

A
Figure 11 (Continued)

B
There are a few computational approaches available that can predict the impact of amino acid substitutions on protein function. Polyphen 2 (Polymorphism Phenotyping v2) is a Web server and software that estimates the probability that a mutation in a human protein will affect protein structure or function. The algorithm calculates the Position-Specific Independent Counts (PSIC) scores of the conserved amino acid and a variant — based on characterization of the substitution site and amino acid contacts, multiple alignment of homologous sequences, and mapping of the substitution site to known 3-dimensional structures — then calculates the difference between the two PSIC scores, higher differences predict greater impact of the variant to the structure and function of the protein [233, 234, 239, 242]. The SIFT (Sorting Intolerant From Tolerant) algorithm searches for homologous protein sequences from different species and obtains multiple alignments then calculates the tolerance index (from 0 to 1) for each possible substitution at the site. Amino acid substitutions with higher indexes are less likely to impact the structure of the protein [233, 234, 238]. Provean (Protein Variation Effect Analyzer) collects sequences related to the query sequence from BLAST. It then clusters sequences with homology of 75% or higher and chooses the top 30 clusters most closely related to the query to include in the supporting sequence set. A delta alignment score is then computed for each sequence and the scores are averaged within and across clusters to compute the final Provean score. If the score is equal or below a pre-set threshold (default threshold is -2.5), it is considered deleterious. Lower pre-set thresholds give higher confidence, while higher pre-set thresholds give greater number of deleterious variants but lower confidence [243, 244].

We were interested in comparing our enzymatic activity functional data ([131] and Figure 11) with predictions from PolyPhen-2, SIFT, and Provean algorithms. We used default settings for all three algorithms. PolyPhen-2 predictions are benign, possibly damaging, or probably
damaging. As shown in Table II, PolyPhen-2 predictions are erroneous 11.3% of the time (6 out of 53 variants). The same number of variants (6 out of 53 variants) are predicted to be possibly damaging. PolyPhen-2 predictions were more frequently incorrectly predicted to be benign, when the SIAE variant was defective, than probably damaging, when the SIAE variant was normal. SIFT predictions are erroneous 28.3% of the time (15 out of 53 variants). Like PolyPhen-2, the SIFT algorithm is more likely to predict that an amino acid substitution will be tolerated, when it is defective, than to predict that a substitution will affect protein function when it is normal. Provean predictions are incorrect 18.9% of the time (10 out of 53 variants). Provean in also more likely to incorrectly predict that a substitution will be neutral than to incorrectly predict that a substitution will be deleterious. Most of the variant impacts that are incorrectly predicted are common between at least two of the algorithms and most of the variants predicted to be possibly damaging by PolyPhen-2 are incorrectly predicted by the SIFT and Provean algorithms.
Table II. Comparison of PolyPhen-2, Provean and SIFT predictions with results of functional enzyme assay. Erroneous predictions are highlighted in yellow or green. PolyPhen-2 “possibly damaging” predictions are highlighted in blue or orange.

<table>
<thead>
<tr>
<th></th>
<th>PolyPhen 2 HumDiv</th>
<th>Provean Prediction</th>
<th>SIFT results</th>
<th>Assay</th>
<th>Estimated Activity % of WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3G</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>87</td>
</tr>
<tr>
<td>R37H</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>30</td>
</tr>
<tr>
<td>Y31C</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>25</td>
</tr>
<tr>
<td>R62H</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>M89V</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>72</td>
</tr>
<tr>
<td>E96K</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Defective</td>
<td>15</td>
</tr>
<tr>
<td>Q101K</td>
<td>Possibly Damaging</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>N107K</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>85</td>
</tr>
<tr>
<td>S127A</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>A171V</td>
<td>Possibly Damaging</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>A175F</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>75</td>
</tr>
<tr>
<td>A177M</td>
<td>Possibly Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>C196F</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>0</td>
</tr>
<tr>
<td>R201C</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>30</td>
</tr>
<tr>
<td>T206I</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>G212R</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>S228C</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>3</td>
</tr>
<tr>
<td>R230W</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>D246N</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>P251L</td>
<td>Possibly Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>S255A</td>
<td>Possibly Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>20</td>
</tr>
<tr>
<td>C266G</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Defective</td>
<td>45</td>
</tr>
<tr>
<td>T286M</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>85</td>
</tr>
<tr>
<td>L296F</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Normal</td>
<td>80</td>
</tr>
<tr>
<td>Q309P</td>
<td>Probably Damaging</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>T312M</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>6</td>
</tr>
<tr>
<td>R314H</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>L323F</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>0</td>
</tr>
<tr>
<td>S325P</td>
<td>Benign</td>
<td>Neutral</td>
<td>Affect Protein Function</td>
<td>Normal</td>
<td>65</td>
</tr>
<tr>
<td>D334N</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>95</td>
</tr>
<tr>
<td>R340H</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>Y349C</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>37</td>
</tr>
<tr>
<td>N357S</td>
<td>Benign</td>
<td>Deleterious</td>
<td>Tolerated</td>
<td>Normal</td>
<td>20</td>
</tr>
<tr>
<td>M360K</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>5</td>
</tr>
<tr>
<td>L366V</td>
<td>Possibly Damaging</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>50</td>
</tr>
<tr>
<td>R369K</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>D370H</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>S371L</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>3</td>
</tr>
<tr>
<td>A385T</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>10</td>
</tr>
<tr>
<td>R387H</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>0</td>
</tr>
<tr>
<td>R393C</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>30</td>
</tr>
<tr>
<td>R393H</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>A394T</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>40</td>
</tr>
<tr>
<td>K400N</td>
<td>Possibly Damaging</td>
<td>Neutral</td>
<td>Affect Protein Function</td>
<td>Normal</td>
<td>80</td>
</tr>
<tr>
<td>F404S</td>
<td>Benign</td>
<td>Deleterious</td>
<td>Tolerated</td>
<td>Defective</td>
<td>50</td>
</tr>
<tr>
<td>Q428L</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>K434T</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>C443R</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td></td>
</tr>
<tr>
<td>M456F</td>
<td>Benign</td>
<td>Neutral</td>
<td>Affect Protein Function</td>
<td>Normal</td>
<td>70</td>
</tr>
<tr>
<td>V459I</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>H472Q</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
<tr>
<td>R479C</td>
<td>Probably Damaging</td>
<td>Deleterious</td>
<td>Affect Protein Function</td>
<td>Defective</td>
<td>0</td>
</tr>
<tr>
<td>G514A</td>
<td>Benign</td>
<td>Neutral</td>
<td>Tolerated</td>
<td>Normal</td>
<td>&gt;100</td>
</tr>
</tbody>
</table>
Conclusions

Rare disease-associated variants can be identified through the deep sequencing of candidate genes or through genome-wide association studies. Once identified, it is useful to study the effects of variants on protein structure and function. In the case of the SIAE protein, a functional assay of enzymatic activity exists. However, for many proteins, such functional assays are not available and other methods of predicting variant effects are helpful.

We have investigated the possibility that a battery of monoclonal antibodies against SIAE with different epitope specificities might be useful in predicting protein misfolding for variants we have identified as defective and have compared binding of the antibodies to both defective and functionally normal variants. We did not find a direct correlation between SIAE variant function and diminished antibody binding. Two antibodies (7H2 and 1G5) bind to most defective variants more strongly than to wildtype SIAE. Another antibody (4A4) did have diminished binding to most of the defective variants while two others (2F2 and 1D11) have similar binding to most variants and wildtype protein but enhanced or reduced binding to some variants. One defective variant (R479C) was bound by 4 of the 5 antibodies less effectively than wildtype, possibly indicating that this variant causes significant disruption to protein structure. The battery of anti-SIAE monoclonal antibodies was not useful for making any specific conclusions about protein structure and function. However, some of the monoclonal antibodies did seem to distinguish between variants with defective and normal function with either enhanced or reduced binding compared to binding to wildtype SIAE. Therefore, we believe that such a set of monoclonal antibodies could prove to be one useful tool in investigating possible structural changes caused by amino acid substitutions in proteins. Such a set might be
particularly useful if the specific epitopes (or regions or domains) recognized by each monoclonal antibody could be determined.

We also compared the results of our enzymatic activity assay to the prediction of three popular algorithms for the prediction of damaging amino acid substitutions. Overall, we found that PolyPhen-2, the most elaborate of the algorithms tested, made the fewest erroneous predictions (approximately 11% of the time) while SIFT had the highest erroneous prediction rate (approximately 28% of the time) for the SIAE variants. Another interesting finding was that for variants erroneously predicted to be non-damaging, two or all three of the algorithms made the same erroneous prediction. This is also true for variants erroneously predicted to be damaging, indicating that use of predictive algorithms should include multiple algorithms and predictions of damaging substitutions should be further investigated even when they are not substantiated by all algorithms. We consider these error rates to be considerable and believe these results point to the need for other, non-computational, predictive methods that can be used in the laboratory in conjunction with these algorithms.
Chapter Three

Oligomeric Structure and Structural Analysis of SIAE and Variant-Encoded Mutants
Summary

Disease-associated variants of SIAE, previously identified through deep exon sequencing, are exclusively found in the heterozygous state. Previously published studies showed that SIAE exists as a dimer or higher-order multimer and that defective mutants of the SIAE protein act in a dominant-interfering manner [131]. Because mutant SIAE proteins are able to co-immunoprecipitate with wildtype SIAE, and diminish the activity of the wildtype in co-transfection experiments, we sought to determine the oligomeric structure of SIAE. Fast Protein Liquid Chromatography (FPLC) shows that the protein is dimeric. Because some mutant SIAE proteins, though defective, are still able to form multimers, we were also interested in investigating the extent to which SIAE mutants are misfolded, both compared to wildtype and compared to each other. Preliminary circular dichroism spectroscopy results indicate that disease-associated variants likely encode partially misfolded proteins but likely retain significant structural integrity, consistent with the retained ability to form multimers.
**Introduction**

In previously published work, we reported that many variants of the SIAE protein are defective when studied in *in vitro* experiments. We co-transfected expression vectors encoding V5-tagged wildtype murine Siae and FLAG-tagged Siae mutant proteins into 293T cells. The proteins were then co-immunoprecipitated from the cell lysates and analyzed by quantitative western blot and enzyme activity assays. These experiments showed that wildtype and mutant proteins did co-immunoprecipitate, indicating that the Siae protein forms dimers or higher-order multimers. In addition, comparing Siae enzyme activity levels of the wildtype protein to the expression levels observed via quantitative western blot with anti-V5 antibody showed that defective mutants acted in a dominant-interfering manner on wildtype protein activity, further suggesting the formation of multimers [131]. We elected to investigate the size and oligomeric structure of the SIAE protein by FPLC. Because over-expressed and epitope-tagged wildtype SIAE is secreted from transfected cells and is relatively easy to purify, we generated FLAG-tagged SIAE and a SIAE-murine Fc fusion protein and purified these proteins from the supernatants of stably transfected cells for use in these studies.

Although a fraction of the wildtype SIAE expressed in transfected non-lymphoid cells is secreted, there is no evidence for the secretion of this protein *in vivo*. Further, there is no expression of SIAE on the surface of B cells (although it is expressed intracellularly) indicating that this enzyme functions in an intracellular post-Golgi vesicular location [126]. Still, we hypothesized that misfolding might explain why defective mutants of SIAE are not secreted from transfected cells while wildtype SIAE is. Furthermore, we hypothesized that mutants that are nearly or completely enzymatically dead and not secreted at all might be significantly more misfolded than mutants that have reduced enzymatic activity or secretion. We used circular
dichroism spectroscopy in order to determine if we could detect relative structural changes among mutants. Because defective mutants of SIAE are not secreted, this required optimizing a protocol for purification of wildtype and mutant SIAE from lysates of transiently-transfected cells.

Circular dichroism spectroscopy measures the difference in the absorption of left-handed circularly polarized light (L-CPL) and right-handed circularly polarized light (R-CPL) by a molecule containing chiral chromophores and the signal varies as a function of wavelength. It is useful for ascertaining some basic information about protein secondary structure [215, 218, 219] and has been used to analyze the effects of single amino acid changes on protein secondary structure. Cases in which amino acid changes have no effect, and others in which amino acid changes have dramatic effects, on protein secondary structure visualized by CD spectroscopy have been documented. For example, in studying the effect of the defective R23Q mutant type II dehydroquinase from *Streptomyces coelicolor*, it was found that this mutant did not result in significant changes in the secondary structure and that structural changes could not account for the loss of catalytic activity [213, 252]. In another example, mutants of the DJ-1 protein, identified as the cause for an autosomal recessive, early onset form of familial Parkinson’s disease, were examined by CD spectroscopy. An active site mutant identified from the crystal structure, C106A, did not result in any significant alteration in the primarily α-helical structure of DJ-1. The disease-associate L166P mutation, on the other hand, resulted in a dramatic alteration to the CD spectra of DJ-1, representing an unfolded, random coil structure of the mutant protein [253].
Methods

Protein Purification prior to Fast Protein Liquid Chromatography

Stably transfected U2OS-SIAE cells were grown in culture to confluency on large plates at which point supernatant media was collected. Supernatant media was then centrifuged in 250 ml flasks at 6000 rpm for 15 minutes at 4°C to remove dead cells and other debris. Supernatants were then stored at 4°C in the dark until use.

Ammonium Sulfate Precipitation

A total volume of 3000 ml of supernatant was used for purification of secreted SIAE-FLAG protein. The amount of ammonium sulfate used for precipitation was calculated using the website http://www.encorbio.com/protocols/AM-SO4.htm. The temperature entered was 10°C, the starting percentage of ammonium sulfate was 0%, and the desired final percentage of ammonium sulfate was 60%. These conditions were used to calculate an appropriate mass of ammonium sulfate of 1135.79 grams. Supernatant was divided into two 2 Liter beakers with magnetic stir bars on stir plates and covered with aluminum foil. The ammonium sulfate was added to the stirring supernatant in 100-gram increments. Once the ammonium sulfate had dissolved and the temperature of the supernatant had reached 10°C, the mixture was left to stir at room temperature for one additional hour. At that point, the ammonium sulfate was pelleted by centrifugation in 250 ml flasks at 8500 rpm for 30 minutes at 4°C in a G23 rotor. The majority of protein in supernatant was determined to precipitate with the ammonium sulfate during previous optimization of the protocol. Supernatant was removed and ammonium sulfate pellets were resuspended in a total volume of 100 ml of autoclaved phosphate-buffered saline (PBS). Dialysis bags were cut to 10-inch length and cleaned and opened by gentle rubbing under warm
water for 3-5 minutes and then rinsing under deionized water for an addition 1-2 minutes. Clips were added to one end of each bag to seal then 20ml of the resuspended ammonium sulfate in PBS solution was added to each bag and the other sides were sealed. Bags were placed into a 4ml beaker of fresh PBS with a stir bar, beaker was covered with aluminum foil and left to stir at 4° C for 3 days, during which time, PBS was replaced twice daily and dialysis bags were rotated to re-dissolve ammonium sulfate which had settled to the bottoms of the bags. After 3 days, solutions were removed from dialysis bags and placed in 50 ml Falcon tubes, then centrifuged at 5000 rpm for 10 minutes at 4° C to remove residual salt.

**SIAE-FLAG purification using EZview Red Anti-FLAG M2 affinity gel columns**

EZview Red Anti-FLAG M2 affinity gel beads (Sigma Cat. #F2426) were washed and equilibrated 3 times with 10-ml portions cold Tris-buffered saline (TBS 50mM Tris-HCl, 150mM NaCl, pH 7.4). Beads were then resuspended in the dialyzed protein sample solution from above (1 ml of original bead slurry per 50 ml of dialyzed protein sample solution) and placed into 50 ml Falcon tubes then left rocking gently at 4°C overnight to allow FLAG-tagged protein to bind to beads. Beads and solution were then poured into a column and solution was allowed to run through at 4°C. Collected solution was poured through the column two additional times to bind any more free SIAE-FLAG protein to the column beads then stored at 4°C. The column was then washed with 100 ml of cold TBS by gravity flow.

**Elution of SIAE-FLAG with free FLAG peptide**

The SIAE-FLAG protein was then eluted from the column with free FLAG peptide (Sigma Cat. #F3290) at a concentration of 100 µg/ml. Protein was eluted in a total of 8 1-ml
fractions of FLAG peptide. The first fraction was collected while the column was still at 4°C. The column was then placed at room temperature for 1 hour before the subsequent fractions were collected.

Fractions were tested for SIAE content and purity by running them on a gel by SDS-PAGE and Coomassie staining. Single bands were observed at the expected size around 53 kDa and the majority of total SIAE-FLAG protein was eluted in fractions 1-3, the largest portion in fraction 2. A significantly fainter band was observed in the lane for fraction 4 and no band was clearly observed in fraction 5. The presence of hSIAE-FLAG in fractions 1-3 was also confirmed by western blot with anti-FLAG antibody (Sigma Cat. #F3165).

**Enzyme Activity Assay**

An enzyme activity assay was performed on samples from fractions 2 and 5 to confirm the presence of active protein in fraction 2 and to assess whether active protein would be found in fraction 5.

For the assay, 10µl of SIAE-FLAG fractions 2 or 5, or 10µl of 0.1M sodium phosphate buffer as a negative control, were added to Eppendorf tubes containing 190µl of sodium phosphate buffer. In the dark, 10µl of 50µg/µl 4-methylumbelliferyl acetate substrate was added to each tube. Tubes were wrapped in aluminum foil, vortexed, and allowed to shake at 37°C for 15 minutes. Tubes were then centrifuged at 13,000 rpm for 10 minutes in the dark to pellet precipitated substrate. While still in the dark, 300µl of 50mM citric acid was added to each tube to stop the reaction. Tubes were centrifuged for an additional 10 minutes. Then, 150µl of each reaction was added to each of three wells in a black 98-well plate for triplicate readings. The plate was covered with aluminum foil for transport to the plate reader. Results were read on a
Fluoroscan II with exitation wavelength of 355nm and emission wavelength of 460nm and analyzed using Spectrosoft software. This assay confirmed that the SIAE-FLAG enzyme in fraction 2 was active and that no active protein was detected in fraction 5, so that all active protein was eluted in fractions 1-4.

**Fast Protein Liquid Chromatography (FPLC)**

Fraction 2 of SIAE-FLAG purified from U2OS-SIAE cell supernatants as described above was used for FPLC. A gel filtration Superdex 200 10/300 column (GE) was equilibrated with filtered, degassed TBS (50mM Tris-HCl, 150mM NaCl, pH 7.4). The 200 µl of purified, active SIAE-FLAG enzyme from fraction 2 was applied to the column at a flow rate of 15ml/hour. The protein was eluted with TBS in 400µl fractions. Two peaks were observed and eluted around 10.3 ml (fraction 26) and 13.8 ml (fractions 35-36). Enzyme activity assay confirmed that the peak of activity was eluted in fractions 35-36, which very little activity observed in fractions 34 and 37. There was no enzyme activity in the peak that eluted at 10.3 ml.

**Protein Purification for Circular Dichroism.**

*Large-scale transient transfections of 293T cells*

For circular dichroism studies, protein was purified from cell culture lysates. Large-scale transient transfections of 293T cells with wildtype or mutant SIAE-FLAG-expressing plasmid were performed according to a protocol provided by Dr. Bunsyo Shiotani (postdoctoral fellow in the lab of Dr. Lee Zou) and adapted from Durocher YS et al., 2002 and Pham PL et al., 2003 [254, 255]. For each wildtype or mutant SIAE-FLAG protein purification, 293T cells were plated in 40 large (150mm) tissue culture dishes and allowed to adhere and expand overnight.
The next day, cells were usually at approximately 40% confluency. Large transfection mixtures were prepared containing 12.5 µg of plasmid DNA in 100 µl of autoclaved MilliQ water and 100 µl of 1 mg/ml polyethyleneimine (PEI, Polysciences, Inc. Cat. #23966) per plate. Mixtures were gently shaken, incubated at room temperature for 5 minutes, and subsequently placed on ice for at least 10 minutes. Then, 200 µl of the mixture was added dropwise, care taken to avoid dislodging cells, to each plate. Plates were gently swirled to disperse the plasmid/PEI mixture into the culture media. The cells were then allowed to incubate for two to three days until cells were nearly confluent. Media was then removed from the plates and the cells were rinsed with 10 ml of cold, sterile PBS.

**Lysis of transfected 293T Cells**

Cells were scraped gently from the plates in the cold PBS and transferred to 50 ml Falcon tubes. Cells were centrifuged at 1200 rpm for 10 minutes then PBS was removed from the cell pellets. The cells were lysed in 2 ml of cold lysis buffer per plate – 80ml total for 40 plates of cells (Lysis buffer: 150 mM NaCl, 1% NP40, 50 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). Cells in lysis buffer were vortexed gently then incubated on ice for 30 minutes.

**Removal of nucleic acid contamination from lysates**

To pre-clear lysates, they were centrifuged at 13,000 rpm for 15 minutes at 4°C. Lysates were then transferred to fresh 50 ml Falcon tubes. To degrade nucleic acid in lysates, 20 µl of Universal Nuclease (Pierce Cat. #88701) was added to 80 ml of cell lysate and the lysates were then allowed to shake at room temperature for 30 minutes. To precipitate any remaining undigested nucleic acid contamination and some acidic proteins, 8.9 ml of 5% PEI was added to
80ml of nuclease-treated lysates (0.5% final PEI concentration). This was incubated on ice for 10 minutes and PEI was then pelleted by centrifugation at 13,000rpm for 15 minutes. Prior optimization showed that SIAE-FLAG remained in the lysate. Lysate was transferred again to fresh 50 ml Falcon tubes.

*Ion exchange to remove some protein from lysates*

To the 90 ml of lysate, 6ml of Q Sepharose Fast Flow ion exchange beads (GE Life Sciences, Cat. #17-0510-01) were added to remove more protein. This was incubated on a shaker at room temperature for 30 minutes then beads were pelleted by centrifugation at 13,000 rpm for 15 minutes. Again, prior optimization showed that very little SIAE-FLAG stuck to the Q Sepharose beads so lysates were removed from the pellet and transferred to fresh 50 ml Falcon tubes.

*SIAE-FLAG purification using EZview Red Anti-FLAG M2 affinity gel columns*

EZview Red Anti-FLAG M2 affinity gel beads that had been washed and equilibrated with TBS (Sigma Cat. #F2426) were then resuspended in the lysate (2ml of bead slurry per 90 ml of lysate) and the tubes were left to rock at 4°C overnight. Beads and solution were then poured into a column and solution was allowed to run through at 4°C. Collected solution was poured through the column two additional times to bind any more free SIAE-FLAG protein to the column beads then stored at 4°C. The column was then washed with 100 ml of cold TBS.
Elution of SIAE-FLAG with free FLAG peptide

The wildtype or mutant SIAE-FLAG protein was then eluted from the column with free FLAG peptide (Sigma Cat. #F3290) at a concentration of 100 µg/ml in TBS buffer. Protein was eluted in a total of 15 1-ml fractions of FLAG peptide. All fractions were eluted at room temperature after the column had been allowed to warm to room temperature for one hour. Fractions were pooled. Free FLAG peptide was removed by running pooled samples through a G25 Sephadex column (Sigma). Protein was concentrated to 300 µg/ml using Millipore Amicon Ultra Centrifugal Filters (10,000 MWCO, Cat. #UFC901024). Protein concentrations were verified by Bradford assay using bovine serum albumin standard curves and the Bio-Rad Protein Assay Dye Reagent Concentrate (Cat. #500-0006), according to the protocol given.

Circular Dichroism

Circular Dichroism experiments were conducted to compare wildtype and mutant SIAE protein structures. CD measurements were made using a Jasco J-715 Spectropolarimeter (Jasco Analytical Instruments) using a cuvette with a 1 mm path length. Far-UV spectra scans were carried out from 190-260 nm at constant 20°C temperature. Protein samples were at 300 µg/ml (5.66 mM) in TBS (50mM Tris-HCl, 150mM NaCl, pH 7.4) and values were obtained by subtracting the baseline recorded for buffer alone under the same circumstances.

Results

In order to confirm that SIAE exists as a multimer and to determine the oligomeric state, we undertook Fast Protein Liquid Chromatography studies. We obtained FLAG-tagged human SIAE (SIAE-FLAG) from the supernatants of stably-transfected U2OS cells via ammonium
sulfate precipitation and purification on EZview Red Anti-FLAG M2 affinity gel columns. We confirmed that the purified protein was pure via SDS-PAGE and Coomassie blue staining and by western blot with anti-FLAG antibody and that it was active via enzyme activity assay (data not shown). Preliminary FPLC demonstrated that SIAE-FLAG eluted in two peaks between fractions 18-22 and 28-32 (Figure 12A). The fractions containing the peaks of protein were analyzed by SDS-PAGE followed by Coomassie blue staining and western blot with anti-FLAG antibody Figure 12B). Coomassie blue staining showed that the majority of SIAE-FLAG eluted in the second peak, while a smaller amount of protein appeared to co-elute with another protein in the first peak. We identified this other large protein, by mass spectroscopy, to be insulin-like growth factor 2 receptor (IGF2R). However, our attempts to verify that the SIAE and IGF2R interact via co-immunoprecipitation and western blot analysis did not confirm an association (data not shown). Western blot analysis confirmed that SIAE-FLAG was present in both peaks
Figure 12. Initial FPLC shows that purified SIAE-FLAG elutes in two peaks. A) FPLC with purified SIAE-FLAG. B) Fractions were analyzed by SDS-PAGE followed by Coomassie Blue staining (left) and western blot with anti-FLAG antibody (right). The expected monomer size of SIAE-FLAG is approximately 55 kDa.
In order to determine the oligomeric structure of SIAE, we ran FPLC with protein standards blue dextran, catalase, aldolase, bovine serum albumin (BSA) and ovalbumin along with SIAE-FLAG and an SIAE-Fc fusion protein generated by Vinay Mahajan (Figure 13A). The protein standards were used to generate a calibration curve used to determine the molecular weights of SIAE-FLAG to be 91.9 kDa, and SIAE-Fc to be 356 kDa (Figure 13B). The monomeric form of SIAE has a molecular weight of 53 kDa and SIAE-FLAG has a molecular weight of 54 kDa. A dimer of SIAE-FLAG would therefore have an expected molecular weight of 108 kDa. A monomer of SIAE-Fc is expected to have a molecular weight of 83 kDa. We expected the SIAE-Fc fusion to form a dimer of dimers, with a molecular weight of 332 kDa, in which Fc regions interact and SIAE regions also interact (Figure 13C). We therefore conclude that SIAE most likely forms a dimer, which supports previous data indicating that mutant SIAE variants function in a dominant negative manner.
Figure 13. FPLC sizing and determination of oligomeric structure of SIAE. A) Fast protein Liquid Chromatography of SIAE-FLAG and SIAE-Fc compared with catalase, ovalbumin, bovine serum albumin (BSA), blue dextran, and aldolase standards. B) Size calibration curve generated from the log of the molecular weights of the standards and the partition coefficients ($K_{av}$) of the standards, where $K_{av} = (V_e - V_o)/(V_t - V_o)$, $V_e$ is the elution volume, $V_o$ is the void volume (a column constant, 7.77 ml in this case), $V_t$ is total bed volume (a column constant, 23.56 ml in this case). Elution volumes are 8.19 ml for blue dextran, 12.83 ml for catalase, 13.05 ml for aldolase, 14.26 ml for BSA, and 15.17 ml for ovalbumin. After generation of the calibration curve, elution volumes for SIAE-FLAG (13.96 ml) and SIAE-Fc (11.52 ml) were then used to calculate $K_{av}$ values and molecular weights of SIAE-FLAG and SIAE-Fc.
Figure 13 (Continued)

A

![Graph showing various molecular weight markers.](image)
Figure 13 (Continued)

### B

![Graph showing Kav versus Log MR with markers for various proteins and equation Kav = \(-0.2621 \times \log(MR) + 1.6927\) with \(R^2 = 0.98526\).]

### C

![Diagram illustrating SIAE (53kDa) and Fc (~30kDa) in blue and red respectively.]
Most of the mutant variants of human SIAE are not secreted from cells in culture [131]. Therefore, in order to purify protein for circular dichroism studies to compare relative misfolding of the mutant proteins, we worked out a method for purifying SIAE-FLAG from cell lysates. This was significantly more complex than purifying protein secreted from supernatants, as we did for FPLC studies. We used the same SIAE-FLAG plasmid constructs described in chapter 2 of this thesis for large-scale transient transfections of 293T cells, using polyethyleneimine (PEI) as the transfection reagent. Once transfected cells were confluent, they were washed and lysed in cold lysis buffer, on ice, without any protease inhibitor since protease inhibitor inactivates SIAE protein. To remove nucleic acid contamination, lysates were pre-cleared by centrifugation and then treated with Universal Nuclease (Pierce). To further precipitate any remaining nucleic acid, and some of the acidic proteins, PEI was also used directly after nuclease treatment, followed by centrifugation to pellet the PEI, nucleic acid, and acidic proteins. As shown, following nuclease treatment and PEI precipitation, the majority of SIAE-FLAG remains in the lysate (Figure 14). Ion exchange with Q Sepharose beads was then used to remove some additional proteins from the mix. Very little SIAE remains bound to the Q Sepharose beads, as can be seen by eluting the bound protein from the beads with NaCl (Figure 14). After all of this initial clarification, lysates were then run on EZview Red anti-FLAG affinity gel columns to bind the SIAE-FLAG protein. After washing the column, the SIAE-FLAG was then eluted with free FLAG peptide. All fractions were pooled. Free FLAG was removed on G25 Sephadex column. As shown, very virtually no SIAE-FLAG protein flows through the anti-FLAG column and the protein eluted from anti-FLAG and G25 columns is quite pure (Figure 14). The SIAE-FLAG protein was then concentrated to approximately 300 µg/ml using Millipore Amicon Ultra Centrifugal Filters and concentrations were determined by Bradford assay (not shown).
Figure 14. Analysis of protein purification steps for circular dichroism. See text and methods for detailed descriptions of the purification steps. SDS-PAGE followed by Coomassie blue staining (left) and western blot with anti-FLAG antibody (right). Lanes are as follows: lysate - total lysate following pre-clearing, universal nuclease – lysate following treatment with Universal Nuclease (Pierce), PEI precipitation – lysate following nucleic acid and acidic protein precipitation and centrifugation, ion exchange – lysate remaining after Q Sepharose ion exchange, NaCl elute from Q beads – protein eluted from the beads following ion exchange, FLAG column flow-through – total protein that does not bind to the EZview anti-FLAG column, FLAG/G25 elute - total protein eluted from the anti-FLAG then G25 Sephadex columns, final concentrated protein is SIAE-FLAG following concentration on Millipore Amicon Ultra Centrifugal Filters. Representative wildtype SIAE-FLAG purification shown.
In our initial attempts at circular Dichroism, to assess folding in SIAE and some mutants, prior to the optimized purification protocol described above, we saw curves consistent with random coiling (Figure 15). This could mean that the wildtype protein has no significant secondary structural features. However, the Chou-Fasman predictive method ([http://www.biogem.org/cgi-bin/cho-fas.pl](http://www.biogem.org/cgi-bin/cho-fas.pl), [256, 257]) led us to expect that the protein would contain significant α-helical (70.2%) and β sheet (51.1%) content. The obtained curves for both wildtype and mutant proteins could also indicate that the protein is simply denatured or misfolded. Another possibility is that the protein might not have been sufficiently pure and free of nucleic acid and other proteins.
Figure 15. Circular dichroism spectroscopy of inactive wildtype and several mutant SIAE-FLAG proteins. Wildtype and mutant SIAE-FLAG proteins were purified from transiently-transfected 293T cell lysates. Transfections and purifications were as described above and in methods, except that protease inhibitor was included in the lysis buffer and the Universal Nuclease, PEI precipitation, and ion exchange steps were not included. The M89V mutant is enzymatically normal but is not secreted from cells in vitro. The S127A mutant is an active site mutation, not found in patients, and is secreted normally in vitro. The C196F mutant is enzymatically dead and not secreted in vitro. The F404S mutant has both reduced activity and reduced secretion in vitro compared to wildtype.
Therefore, we attempted to verify that the wildtype fraction of purified protein used for CD still had enzymatic activity (mutants were expected to be enzymatically dead). For comparison, we also used wildtype SIAE-FLAG protein that had been purified previously from the supernatants of stably transfected U2OS (as described for use in FPLC experiments). This previously purified protein had been highly active and had remained so for nearly 15 months. We found that the sample purified from the U2OS-SIAE supernatants was still active while the protein isolated from transfected 293T cell lysates, and used for CD, had no activity (data not shown). In our initial protein purification protocol for CD, we had included protease inhibitors. Although we knew that protease inhibitors rendered SIAE protein inactive, we had not realized that they might cause misfolding of the protein.

We next used the active wildtype SIAE-FLAG purified from stably-transfected U2OS cell supernatants and the inactive wildtype SIAE-FLAG purified from transiently-transfected 293T cell lysates for comparison in CD studies to determine if the active wildtype protein would have a different spectra (Figure 16). Active SIAE-FLAG protein did, in fact, give rise to a spectra more consistent with spectra representing α-helical and β-sheet content, as predicted. This indicated to us that we needed to modify the protocol for purifying SIAE-FLAG protein from transiently-transfected 293T cell lysates in order to consistently obtain active wildtype protein and that the activity of wildtype SIAE-FLAG should be verified prior to CD studies.
Figure 16. Circular dichroism spectroscopy comparison between active and inactive wildtype SIAE-FLAG proteins. Active wildtype SIAE-FLAG was purified from stably-transfected U2OS cell lysates as described in methods (purification for FPLC). Inactive wildtype SIAE-FLAG is the same used for prior experiment shown in Figure 15. Activity was confirmed as described in methods (data not shown).
After modifying the purification protocol so that protease inhibitors were left out of the lysis buffer and purity was improved (as described above), reliably active wildtype SIAE-FLAG could be obtained. This new protocol was used to obtain wildtype and mutant SIAE-FLAG proteins for circular dichroism (Figure 17). We also re-analyzed wildtype SIAE-FLAG purified from U2OS supernatants for additional comparison. An enzyme activity assay confirmed that the wildtype and mutant proteins had expected relative levels of activity (not shown). This time, spectra were obtained for wildtype and all mutant SIAE-FLAG proteins that are consistent with the expected α-helical and β-sheet content of the protein. However, there are some potentially interesting differences between wildtype and some mutants.

The spectra for wildtype SIAE-FLAG purified from stably-transfected U2OS cells has larger peaks than any of the other SIAE-FLAG proteins purified from transiently-transfected 293T lysates. There are a couple of different possible reasons for this. The first would be differences in protein concentration. Protein concentration can impact the magnitude of peaks in CD spectra [218, 230]. We confirmed protein concentration by Bradford assay to all be approximately 300 µg/ml (~5.6 µM). Therefore, this is unlikely to be the difference, since the magnitude of peaks in the CD spectra from supernatant-purified protein is significantly larger than those of lysate-purified proteins. The other possibility would be a difference in solvent. However, all proteins were ultimately eluted from columns in TBS so there is no difference in that respect. It is therefore unclear what accounts for the difference in peak magnitude between wildtype SIAE-FLAG purified from supernatant and from lysates.

In comparing the wildtype and mutant SIAE-FLAG proteins purified from transiently-transfected 293T lysates, the spectra for wildtype protein and three of the analyzed mutants look fairly different, while the spectra for the Y349C mutant looks somewhat similar to wildtype. We
had expected spectra for mutants might be similar or dissimilar to wildtype based on activity levels, with enzymatically defective mutants having more dissimilar spectra than mutants with normal, or merely reduced, enzyme activity levels. However, the spectra indicate that it may be useful to compare *in vitro* secretion levels of mutant and wildtype SIAE-FLAG, when comparing CD spectra for these proteins. Although there is no evidence for secretion of SIAE from B cells *in vivo*, and secretion is not likely to be physiologically relevant, secretion might be indicative of protein folding or misfolding, since most defective mutants are not secreted, while wildtype protein is secreted from transfected cells. The Y349C mutant has defective enzyme activity. However, Y349C secretion from non-lymphoid cells *in vitro*, is merely reduced compared to secretion of wildtype proteins. The M89V mutant has normal enzyme activity levels but is not secreted from transfected cells. The C196F and R479C mutants are both enzymatically defective and are not secreted from transfected cells. When comparing the CD spectra of the mutants and wildtype proteins, the mutants with defective secretion (M89V, C196F, R479C) have spectra that are similar, while the wildtype and Y349C mutant, which is secreted at a somewhat reduced level, give similar spectra. Overall, the shapes of the spectra for wildtype and all examined mutants are fairly similar, with a narrow positive peak around between 190 nm and 200 nm, and a broad, negative double peak between 200 nm and 240 nm. This double peak is consistent with α-helical content. However, pure α-helical protein would give a second dip in the double peak with a greater magnitude than the first peak. The shift more toward 200 nm in the positive peak, and the lower magnitude of the second dip in the negative peak, of SIAE-FLAG protein spectra is consistent with β-sheet content in SIAE-FLAG protein.
Figure 17. Circular dichroism spectroscopy of wildtype and M89V, C196F, Y349C, and R479C mutant SIAE-FLAG proteins from lysates and wildtype SIAE-FLAG from supernatant.
Conclusions

The sizing of proteins by size-exclusion chromatography/FPLC is not precise and is actually more a measure of hydrodynamic volume, that is the space taken up by a molecule, than of its actual mass. Therefore, determination of the multimeric state of a protein based on the monomeric mass calculated from the protein sequence requires estimation. The calculated mass of the SIAE-FLAG monomer is 54 kDa. The estimated mass of SIAE-FLAG from our FPLC study is 91.9 kDa, significantly smaller than the calculated mass of SIAE-FLAG dimer of 108 kDa, yet significantly larger than the calculated mass of a monomer of SIAE-FLAG. The calculated mass of SIAE-Fc is 83 kDa. The estimated mass of SIAE-Fc from our FPLC study is 356 kDa, somewhat larger than the calculated mass of 332 kDa for the predicted dimer of dimers. For both SIAE fusion proteins, the estimated mass based on FPLC is closer in size to the calculated mass of a dimer of SIAE-FLAG or dimer of dimers of SIAE-Fc than to a monomer of SIAE-FLAG or dimer of SIAE-Fc (with dimerization of the Fc region only). Therefore, based on previous evidence from cotransfection studies that FLAG-tagged SIAE mutants coimmunoprecipitate with V5-tagged wildtype SIAE, and function in a dominant-negative manner to suppress the enzymatic activity of wildtype SIAE, and our FPLC data, we conclude that SIAE most likely forms dimers.

The overall shapes of the CD spectra of the SIAE mutant proteins resemble the spectra of the wildtype SIAE. However, some shifts and differences in the magnitudes of the peaks may indicate that amino acid substitutions result in some slight, possibly localized, misfolding of the mutants examined. It does not appear that any of the defective mutants analyzed result in significant or complete loss of wildtype secondary structure. It is interesting that, of the mutants we analyzed, the one with the CD spectra most resembling wildtype, Y349C, is enzymatically
defective but still secreted from transfected cells at a reduced level, while the M89V mutant, which is not secreted but is enzymatically normal, gives a spectra that more closely resembles the C196F and R479C mutants, which are both enzymatically defective and not secreted from transfected cells. These results indicate that although secretion is not physiological, secretion from cells in vitro may be dependent on secondary structure and secretion might be a better indicator of structural effects of amino acid changes than enzymatic function.

**Contributions:** I purified SIAE proteins that were sent for monoclonal antibody production. I transfected wildtype SIAE and various variants of SIAE and analyzed the differential binding of monoclonal antibodies. I purified wildtype and SIAE mutant proteins and performed HPLC analyses and CD spectral studies. I compiled the data comparing the activity of SIAE proteins in various assays and compared the data with Polyphen 2 and SIFT data.

Mutagenesis of *SIAE* and the initial analysis of catalytic activity and secretion were conducted by Vasant Chellappa, Ira Surolia, and Dalya Ataca.
Chapter Four

Studies on Siae-deficient BCR knockin mice
Summary

We have hypothesized that Sialic Acid Acetylesterase (SIAE) mediates peripheral B cell tolerance by preventing weakly self-reactive B cells from receiving T cell help. Self-reactive B cells would essentially be maintained in a state of ignorance so that they will not be driven into a germinal center reaction or undergo somatic hypermutation to generate dangerous strongly self-reactive clones. The SIAE/CD22 pathway accomplishes this ignorance by establishing a signaling/activation threshold. However, in cases in which SIAE is defective, weakly self-reactive B cells might be activated by antigen, migrate to the T cell zone and into germinal centers where they may survive and undergo affinity maturation due to stronger signaling in the absence of CD22 inhibition. We also hypothesized that SIAE-deficient activated B cells may serve as better antigen presenting cells (APCs) to present self-epitope more strongly to T cells. Our laboratory has taken two approaches to investigate the role of the SIAE/CD22 pathway in mediating B cell tolerance. One involved transfer of B cells expressing a transgenic hen egg lysozyme-specific BCR into congenic recipients. The other, representing the work in this thesis, attempted to investigate the role of SIAE in mice with a knockin low affinity BCR.


**Introduction**

In Siae-deficient mice, B cell activation is enhanced. However, B cells from these mice exhibit defective proliferation and differentiation following BCR activation, possibly reflecting enhanced apoptosis [124]. We hypothesized that although Siae-deficient B cells do not exhibit enhanced proliferation as expected, they might contribute to autoimmunity by contributing to helper T cell activation. The contribution of B cells to the generation of T follicular helper (T<sub>FH</sub>) cells, following initial stimulation of helper T cells by dendritic cells has been previously demonstrated [258-262]. In order to investigate whether, in the absence of Siae, a weak antigen might facilitate T-B collaboration and the generation of T<sub>FH</sub> cells, another member of the laboratory, Hamid Mattoo, undertook transfer experiments with CD45.2-expressing B cells from wildtype or Siae-deficient MD4 mice (hen egg lysozyme-specific BCR transgenic mice) [171] along with CD45.2-expressing CD4<sup>+</sup> T cells from OT-II mice (ovalbumin-specific TCR transgenic mice) [263] into congenic CD45.1 recipients. Recipient mice were then immunized with a DEL-OVA (duck egg lysozyme-ovalbumin) conjugate, and analyzed for T<sub>FH</sub> phenotype cells. HEL-specific MD4 B cells recognize DEL with about a three log lower affinity than HEL itself. At day 3 post-immunization, there was no difference in the numbers of CD45.2 T<sub>FH</sub> cells between recipients that received wildtype and Siae-deficient MD4 B cells. This result was not unexpected, since T<sub>FH</sub> cell commitment at this time point is known to be B cell-independent and mediated primarily by dendritic cell priming [260]. However, at day 7, a time point at which B cells are known to be required for the generation or maintenance of a stable T<sub>FH</sub> phenotype, there was a significant increase in T<sub>FH</sub> cell numbers in recipient mice that received Siae-deficient MD4 B cells. This difference is physiologically relevant as even Siae-deficient mice in steady state, without any immunization or infection, showed increases in frequency and numbers of T<sub>FH</sub> cells.
in secondary lymphoid organs. We hypothesize that this is a result of altered B cell ignorance resulting in enhanced B-T interaction in response to self-antigens. To verify that the increase in T_{FH} cells is due to the presence of Siae-deficient B cells, wildtype and Siae-deficient mice were crossed onto a µMT (B cell-deficient) background. In that case, there is no longer an increase in T_{FH} cell numbers seen in Siae-deficient mice compared to wildtype, suggesting that the increase in T_{FH} cells in Siae-deficient mice is a B cell-dependent phenomenon.

Because T_{FH} cells play an important role in the induction of germinal center formation [264, 265], the acquisition of germinal center phenotype in wildtype and Siae-deficient MD4 B cells co-transferred with OT-II CD4^+ T cells was also investigated. Following DEL-OVA immunization, the majority of Siae-deficient MD4 B cells transferred acquired a germinal center phenotype, at a rate 6-8 fold higher than transferred wildtype MD4 B cells. Rates of somatic hypermutation (SHM) in wildtype and Siae-deficient MD4 B cells were also examined, following co-transfer and immunization, by single cell PCR and subsequent sequencing of the HEL-specific V_{H} and V_{κ} transgenes. The proportion of clones containing mutations and frequency of mutations per clone were elevated in Siae-deficient MD4 B cells.

While naïve follicular B cells recognizing specific antigen are not generally induced to proliferate by BCR signaling, such signaling may induce endocytosis of antigen and the CCR7 expression required for B cells to migrate to the T-B interface and interact with activated CD4+ T cells [266-268]. To investigate whether Siae deficiency lowers the threshold for induction of endocytosis and CCR7 expression, Mattoo et al. undertook adoptive transfer of CFSE-labeled follicular (FO) B cells from MD4 or siae^{−/−} MD4 mice into CD45.1 recipients. After transfer and immunization with DEL, siae^{−/−} MD4 FO B cells, but not wildtype MD4 B cells, were induced to express CCR7 and migrate toward the T cell zone. Immunization with HEL (a high
affinity ligand) induced CCR7 and migration equally well in both wildtype and siae^{A2/A2} MD4 FO B cells.

B cells also play a role in generation of CD4^{+} T cell memory [269]. Consistent with this role, DEL-OVA immunization over 75 days revealed a significant increase in CD4^{+} memory T cells in recipient mice that received OT-II CD4^{+} T cells and siae^{A2/A2} MD4 B cells compared with those that received the same T cells along with wildtype MD4 B cells. An increase in both central and effector memory CD4^{+} T cell compartments was also observed in unimmunized siae mutant mice compared to wildtype mice.

**B1-8 knockin mice fail to respond to immunization**

In B1-8 heavy chain gene knockin mice [270], in which the IgH locus has been knocked in with a rearranged heavy chain gene encoding a 4-hydroxy-3-nitrophenylacetyl (NP)-binding antibody heavy chain, pre-immunization with OVA to prime T cells, followed by immunization with NP-OVA, did not result in an increase in class-switched cells compared with non-immunized mice. However, when B1-8 B cells were transferred into wildtype mice that were primed with OVA and then immunized with NP-OVA, an appropriate T-dependent immune response was observed. In the course of these studies, also performed by Hamid Mattoo, we have assumed that the high levels of accumulated circulating anti-NP antibodies in the knockin animals makes immunization with an NP-protein conjugate ineffective since the bulk of the injected immunogen is potentially rapidly opsonized and cleared. In cell transfer recipients, the absence of circulating anti-NP antibodies allows the transferred NP-specific B cells and the OVA-specific T cells to respond to DEL-OVA and collaborate.
Since the function of Siae had been investigated using a rather artificial cell transfer approach using relatively high affinity transgenic MD4 B cells and OVA specific TCR transgenic OT-II T cells, we were interested in examining whether these studies could be performed directly in BCR knockin or transgenic mice – without transferring cells. Transferred cells may be trapped in certain tissues, go through homeostatic proliferation, and function in ways that are distinct from what might occur in a mouse not subjected to such manipulation. Since transgenic mice with high affinity BCRs secrete high affinity antibodies that can effectively sequester antigens used for immunization, we wished to explore whether a low affinity BCR knockin mouse might prove more suitable for studies on the consequences of loss of Siae function.

The AM14 Mouse Model

We wanted to investigate whether AM14 mice could be used to study the role of SIAE in the regulation of T cell-dependent clonal ignorance in the context of a low-affinity, self antigen-specific model without the requirement for transfer experiments. The AM14 mice, provided to us by Dr. Mark Shlomchik [271], are site-directed transgenics encoding a low-affinity IgM BCR specific for an anti-chromatin IgG2a (a rheumatoid factor) knocked in to the endogenous heavy chain locus. This means that, unlike the MD4 transgenic mice, B cells expressing the AM14 heavy chain can undergo class switching to another Ig isotype. In these mice, B cells are not made anergic and the BCR is not down-regulated. The B cells from these mice are clonally ignorant, making them useful as a model for studies of peripheral B cell self-tolerance. These mice do not progress to autoimmune disease on a normal IgG2a b background. On autoimmune-
prone backgrounds, such as the MRL/lpr background, AM14 cells are spontaneously activated to undergo class switch recombination (CSR) and somatic hypermutation (SHM) and to form extrafollicular plasmablasts [272, 273]. A similar response is seen when AM14 mice on a normal background are immunized with an anti-chromatin IgG2a [271]. The Shlomchik lab has shown that this formation of plasmablasts is T cell-independent but MyD88/TLR7/TLR9-dependent [271]. Given the low affinity of the BCR in the AM14 mouse model, we sought to investigate its use in a non-transfer context.

**Methods**

**Mice.** *Siae* exon 2 deleted mice have been previously described [124]. AM14 HC mice were obtained from Dr. Mark Shlomchik and were previously described [271]. AM14 HC mice were backcrossed into the C57BL/6 background. AM14 HC/SIAE\textsuperscript{A2/A2} were obtained by crossing heterozygous AM14 HC mice and homozygous Siae\textsuperscript{A2/A2} mice. Mice were genotyped using primers for Siae as described [124] and primers AM14F1 (5\' - GCTGTCTTGCAAGGCTTCTGGCTA - 3\') and AM14R1 (5\' - GGCCCCAGTAAGCAAACCCGTCGTA3\') for the AM14 heavy chain.

**IgG2a\textsuperscript{a}-specific B cell numbers assays.** These studies were conducted by me in the laboratory of Ann Marshak-Rothstein with Kerstin Kiefer according to their established protocols. Single cell suspensions of splenocytes were generated and stained with labeled B220 and 4-44 antibodies. Flow cytometric analysis was used to gate live cells then on B220\textsuperscript{+} cells (not shown). Histograms show the percentage of total B cells expressing the AM14 BCR (IgG2a\textsuperscript{a}-specific) as recognized by the idiotype specific 4-44 antibody.
**In vitro proliferation assays.** These studies were conducted in the laboratory of Ann Marshak-Rothstein with Kerstin Kiefer using their established protocols and reagents [274]. Briefly, mouse splenic B cells were purified using B220 magnetic beads (BD Biosciences) and were stimulated with the reagents specified for 24 hours then pulsed with 5 mCi/ml of [methyl 3H] thymidine for another 6 hours. Incorporation of [3H] thymidine was quantified using a liquid scintillation beta counter.

"**T cell-independent**" study, immunization and flow cytometry. These studies are assumed to be T-independent since unconjugated IgG2a was used as an antigen - likely providing a B cell epitope but not a T cell epitope. Three AM14 HC and three AM14 HC/siae<sup>A2/A2</sup> mice were injected intraperitoneally with 625 µg of Alum, and three additional AM14 HC and three AM14 HC/siae<sup>A2/A2</sup> mice were immunized intraperitoneally with 625 µg of Alum and 500 µg of IgG2a<sup>a</sup> (Sigma M7769) on day 0. All mice were boosted on day 42 with the same doses. On day 60, all mice were sacrificed and spleens were harvested. Splenocyte single cell suspensions were generated and treated with Hank’s buffered saline solution (HBSS). Red blood cells were lysed in Ammonium-Chloride-Potassium (ACK) lysis buffer. Cells were counted and divided. Cells were pre-treated with 2.4G2 (anti-CD16/CD32 [Fcγ III/II receptor], rat IgG2b, κ; BD) then surface stained with antibodies at appropriate dilutions for flow cytometry. All antibodies were purchased from BioLegend unless mentioned otherwise. Antibodies used for flow cytometry were IgG1-FITC, IgG3-FITC, SA-PE Txr (BD-Biosciences), IgM-APC, B220-PE, Annexin V-PacBlue, and 4-44-biotin (generated in the lab of Ann Marshak). Flow cytometric analysis was performed on a BD LSR II (BD Biosciences). Sample data was obtained using FACS Diva software (BD Biosciences) and analyzed using FlowJo2 software (v. 9.4.4, Tree Star, Inc.).
T cell-dependent studies, immunization, flow cytometry, single cell sorting, and AM14 HC sequencing. Six AM14 HC and five AM14 HC/\textit{siae}^{A2/A2} mice were injected with 625 μg of Alum, 95.2 μg of OVA, and 4.76 μg of monophosphoryl lipid A (MPL) intraperitoneally in order to prime T cells with OVA on day 0. IgG2a-OVA was conjugated using 2.5% gluteraldehyde for 15 minutes at room temperature with occasional mixing. The reaction was stopped with 1M Tris-HCl pH 7.5 and product was concentrated in Amicon Ultra tube. Efficiency of conjugation was confirmed by running on a 10% polyacrylamide Coomassie-stained gel and estimated to be about 25%. On day 22, three AM14 HC and three AM14 HC/\textit{siae}\textsuperscript{A2/A2} mice were injected intraperitoneally with 5 μg MPL, 625 μg of Alum, and approximately 100 μg of IgG2a-OVA while three AM14 HC mice and two AM14 HC/\textit{siae}^{A2/A2} mice received only MPL and Alum. On day 49, mice were boosted with the same quantities or antigen and adjuvant used on day 22. On day 60, all mice were sacrificed and spleens were harvested. Splenocyte single cells suspensions were generated, red blood cells lysed, and cells were washed and counted as described above. Cells were divided for flow cytometric analysis and for single cell sorting. Cells were pre-treated with 2.4G2 (anti-CD16/CD32 [Fc\textgreek{y} III/II receptor], rat IgG2b, κ; BD) then surface stained with antibodies at appropriate dilutions for flow cytometry. All antibodies were purchased from BioLegend unless otherwise noted. Antibodies used for flow cytometry and single cell sorting were IgG1-FITC, IgG3-FITC, SA-PETxR (BD Biosciences), IgM-APC, B220-PacBlue, GL7-APC (BD Biosciences), Annexin V-PacBlue, and 4-44-biotin (generated in the lab of Ann Marshak-Rothstein). Flow cytometric analysis was performed on a BD LSR II (BD Biosciences). Sample data was obtained using FACS Diva software (BD Biosciences) and analyzed using FlowJo2 software (v. 9.4.4, Tree Star, Inc.). Single cell sorting was performed with the assistance of Laura Prickett at the Harvard Stem Cell
Institute Flow Core on a FACS Aria II machine (BD Biosciences using FACS Diva software and data was analyzed using FlowJo2 software.

**Single Cell Sorting and AM14 heavy chain sequencing.** Following single cell sorting, sequencing of AM14 heavy chains between the CDR2 and CDR3 region was conducted. The method used was modified as needed from Tiller et al., 2009 and Tiller et al., 2010 using reagents and conditions described therein [180, 275]. Briefly, single B220+, 4-44+ (AM14), GL7hi, IgG1+ B cells from spleens of the indicated mice were isolated into 96-well PCR plates containing 0.4 µl of ice cold lysis buffer. Plates were then sealed with Microseal ‘F’ Film (BioRad) and immediately frozen on dry ice before being stored at -80°C. Total RNA from single sorted cells was reverse transcribed in the same sorting plate as described. Resulting cDNA was stored at -20°C. First round polymerase chain reaction (PCR) amplification was performed using outside primers AM14VHFR1 (5’-CTGCAGCAGCCTGGGAC-3’) and AM14R2 (5’-CCAGAGTCCCTTGGCCCCAGTAA-3’) using NEB Taq Polymerase. The first round of PCR was performed at 94°C for 2 minutes followed by 30 cycles of 94°C for 1 minute, 56.5°C for 1 minute, 72°C for 1 minute, and a final incubation at 72°C for 10 minutes. Second and third rounds of PCR were performed using semi-nested or nested primers AM14F1 (5’-GCTGTCCTGCAAGGCTTCTGGCTA-3’) and AM14R1 (5’-GGCCCCAGTAAGCAAACCGTGGCTGA-3’). Second round PCR was performed using NEB Taq Polymerase at 95°C for 2 minutes followed by 45 cycles of 95°C for 1 minute, 59°C for 1 minute, 72°C for 1 minute, and a final incubation at 72°C for 10 minutes. Third round of PCR was performed using Promega GoTaq Green Master Mix at 95°C for 2 minutes followed by 50 cycles of 95°C for 1 minute, 60.5°C for 1 minute, 72°C for 30 seconds, and a final incubation at 72°C for 10 minutes. PCR products were analyzed at each step on 3% agarose gels. Aliquots of
PCR products were sequenced with the AM14F1 primer. Sequencing data was analyzed using the DNAStar Lasergene 10 software suite.

**Results**

Our experiments were conducted using AM14 heavy chain knockin mice. Because B cells from the mice did not contain a light chain transgene, we wanted to verify that the mice would express some B cells specific for the IgG2a^a antigen. With the assistance of Kerstin Kiefer in the laboratory of Ann Marshak-Rothstein, we compared the number of IgG2a^a-specific B cells in the AM14 heavy chain (HC) transgenic mice and AM14 HC/SIAE^{A2/A2} mice to AM14 heavy chain and light chain double knockin mice (AM14V\kappa 8) [274] (Figure 18A). We did so using 4-44 and 4G7 idiootype specific antibodies for flow cytometry. As shown, approximately 2-10% of B cells from both the AM14 HC and AM14 HC/Siae^{A2/A2} mice express B cells specific for the IgG2a^a antigen prior to immunization. Although the percent of IgG2a^a-specific B cells was low, we expected the cells to proliferate upon immunization of the mice and therefore expected to access a sufficient number of antigen-specific cells to be able to conduct the experiments.

*In vitro* proliferation studies indicate that while purified B cells from the AM14 HC/SIAE^{A2/A2} mice proliferate comparably to those from AM14 HC and AM14V\kappa 8 mice in response to TLR9 (1826) and TLR4 (LPS) agonists, they did not proliferate as well in response to BCR stimulation (anti-IgM) or a specific anti-chromatin IgG2a^a (PL2-3) (Figure 18B). We also stimulated cells *in vitro* with immune complexes for preliminary studies. Pseudo PL2-3 is an Fab fragment-DNA complex that binds to the BCR and TLR9. Since there is no IgG2a^a involved, this stimulus is independent of AM14 and should activate all B cells. The 1D4-C11 complex is
an IgG2a antibody specific for biotin (1D4) complexed to CpG with biotin tags (C11). So, ID4-C11 could stimulate AM14 cells through the BCR and TLR-9. The 1D4-Sumo is ID4 mixed with CpG with Sumo tags so no immune complexes are formed but both the AM14 BCR and TLR-9 can be stimulated. After stimulation with PL2-3, anti-IgM, anti-chromatin IgG2a, and immune complexes, the B cells from the AM14 HC/siae^{\alpha2/\alpha2} mice proliferate significantly less than B cells from AM14 HC mice (Figure 18B).
**Figure 18. AM14 B cell proliferation.** A) Representative flow cytometric analysis of the percentage of AM14 BCR-expressing B cells. Total splenocytes from AM14Vκ8, AM14 HC, and AM14 HC/siae<sup>A2/A2</sup> mice were stained with B220 and one of either the 4-44 or 4G7 AM14 idiotypetype-specific antibodies and analyzed by flow cytometric analysis. Numbers shown indicate the percentage of total gates B220<sup>+</sup> cells that also express AM14 BCR (IgG2α-specific). B) B cell proliferation following stimulation with TLR9 (ODN 1826) or TLR4 (LPS) ligands, anti-chromatin IgG2α, anti-IgM, or immune complexes. B cells were purified from the spleens of AM14Vκ8, AM14 HC, and AM14 HC/siae<sup>A2/A2</sup> mice stimulated with the specified reagents for 24 hours, pulsed with 3H-thymidine for 6 hours.
Figure 18 (Continued)

A

```
<table>
<thead>
<tr>
<th>AM14Vk8</th>
<th>AM14 HC</th>
<th>AM14 HC/siae&lt;sup&gt;A2/A2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Cells</td>
<td># Cells</td>
<td># Cells</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10&lt;sup&gt;2&lt;/sup&gt;</td>
<td>10&lt;sup&gt;3&lt;/sup&gt;</td>
<td>10&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
<tr>
<td>10&lt;sup&gt;3&lt;/sup&gt;</td>
<td>10&lt;sup&gt;4&lt;/sup&gt;</td>
<td>10&lt;sup&gt;5&lt;/sup&gt;</td>
</tr>
<tr>
<td>10&lt;sup&gt;4&lt;/sup&gt;</td>
<td>10&lt;sup&gt;5&lt;/sup&gt;</td>
<td>10&lt;sup&gt;6&lt;/sup&gt;</td>
</tr>
<tr>
<td>10&lt;sup&gt;5&lt;/sup&gt;</td>
<td>10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>10&lt;sup&gt;7&lt;/sup&gt;</td>
</tr>
<tr>
<td>10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>10&lt;sup&gt;8&lt;/sup&gt;</td>
</tr>
<tr>
<td>10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>10&lt;sup&gt;8&lt;/sup&gt;</td>
<td>10&lt;sup&gt;9&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
```

B

```
<table>
<thead>
<tr>
<th>1826</th>
<th>LPS</th>
<th>PL2-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>cpm</td>
<td>cpm</td>
<td>cpm</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>1000</td>
<td>5000</td>
</tr>
<tr>
<td>1000</td>
<td>2000</td>
<td>10000</td>
</tr>
<tr>
<td>1500</td>
<td>3000</td>
<td>15000</td>
</tr>
<tr>
<td>2000</td>
<td>4000</td>
<td>20000</td>
</tr>
<tr>
<td>2500</td>
<td>5000</td>
<td>25000</td>
</tr>
<tr>
<td>3000</td>
<td>6000</td>
<td>30000</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>anti-IgM</th>
<th>Immune complexes</th>
</tr>
</thead>
<tbody>
<tr>
<td>cpm</td>
<td>cpm</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td>1000</td>
<td>2000</td>
</tr>
<tr>
<td>1500</td>
<td>3000</td>
</tr>
<tr>
<td>2000</td>
<td>4000</td>
</tr>
<tr>
<td>2500</td>
<td>5000</td>
</tr>
</tbody>
</table>
```

123
In order to explore the role of Siae in regulating peripheral B cell tolerance, we initially conducted an experiment designed to analyze T-cell independent responses. We crossed AM14 HC site-directed transgenic mice with siae^{Δ2/Δ2} mice. We then immunized AM14 HC/siae^{Δ2/Δ2} and AM14 HC/siae^{+/+} mice with IgG2a and alum or alum alone intraperitoneally. All mice were boosted on day 42. Mice were sacrificed on day 60 and splenocytes were used for flow cytometry to examine rates of class switching to IgG1 or IgG3 (Figure 19). There was no difference in class switching between AM14-expressing B cells from unimmunized and immunized mice, either in wildtype or Siae-deficient AM14 knockin mice.
Figure 19. T cell-independent immunization of AM14 HC and AM14 HC/siae^{A2/A2} mice. Mice were immunized with IgG2aa with alum or alum alone, boosted on day 42 and sacrificed on day 60. Flow cytometric analysis was conducted to assess class switching of AM14 BCR-expressing B cells to either A) IgG1 or B) IgG3.
We hypothesized that the IgG2a-specific B cells might require T cell help in order to be sufficiently activated and migrate into germinal centers in order to undergo CSR and SHM. To determine whether SIAE regulates T cell-dependent clonal ignorance in the AM14 model, we primed AM14/Siae\(^{−/−}\) and AM14/SIAE\(^{+/+}\) mice intraperitoneally with OVA. On day 22 post-priming, we immunized the mice with IgG2a\(^a\)-OVA conjugate with alum and Monophosphoryl lipid A (MPL). We boosted the mice with IgG2a\(^a\)-OVA on day 49. On day 61, mice were sacrificed and spleens were harvested for B cell isolation. We analyzed splenic B cells for class switching by flow cytometry and for hypermutation by single cell sorting, PCR and sequencing (Figure 20). We found that AM14 BCR-expressing B cells from Siae-deficient mice do not have an increased germinal center phenotype compared to those from wildtype mice. In addition, AM14 BCR-expressing B cells did not seem to undergo an increased entry into the germinal center upon immunization either in Siae-deficient or wildtype mice (Figure 20A). Consistent results were obtained when we looked for the numbers of AM14 BCR-expressing cells that had switched to IgG1. We did not see increased switching to IgG1 in Siae-deficient compared to wildtype AM14-expressing B cells. However, switching was also not increased in immunized versus unimmunized mice (Figure 20B).

Nevertheless, we still single cell sorted switched AM14-expressing cells from immunized mice in order to ascertain whether switched Siae-deficient cells would have undergone greater rates of SHM (Figure 20C). We obtained good sequencing data for 19 wildtype IgG1-switched AM14-expressing cells and 31 Siae-deficient IgG1-switched AM14-expressing cells. We sequenced the region of the AM14 transgene including complementarity determining regions CDR2 through CDR3. We found only two wildtype cells that contained only one base pair mutation each and only one Siae-deficient cell containing only a one base pair mutation.
Representative sequencing data, including the three sequences containing mutations, is shown in Figure 20D.
Figure 20. T cell-dependent immunization of AM14 HC and AM14 HC/siae^{A2/A2} mice. Mice were primed with OVA, alum, and MPL. On day 22 mice were immunized with IgG2a^{a}-OVA, alum, and MPL or alum and MPL alone. Mice were boosted on day 49 and sacrificed on day 60. Flow cytometric analysis was used to assess, A) entry to AM14 BCR-expressing B cells into germinal centers and B) switching of AM14 BCR-expressing cells to IgG1. C) Fluorescence activated single cell sorting was used to sort individual AM14 BCR-expressing IgG1-switched cells. D) Total RNA was purified from single sorted switched AM14 BCR-expressing B cells and used to generate cDNA, which was then amplified using AM14 heavy chain VDJ region primers between the CDR2 and CDR3 and sequenced. Shown are some representative sequence results including the two sequences from wildtype and one sequence from siae^{A2/A2} cells found to contain single base pair mutations.
Figure 20 (Continued)

A

<table>
<thead>
<tr>
<th></th>
<th>Wildtype Unimmunized</th>
<th>Wildtype IgG2a(^a)-OVA</th>
<th>Siae(^{Δ2/Δ2}) Unimmunized</th>
<th>Siae(^{Δ2/Δ2}) IgG2a(^a)-OVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annexin V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th></th>
<th>Wildtype Unimmunized</th>
<th>Wildtype IgG2a(^a)-OVA</th>
<th>Siae(^{Δ2/Δ2}) Unimmunized</th>
<th>Siae(^{Δ2/Δ2}) IgG2a(^a)-OVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annexin V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 20 (Continued)

C

Singlets

B cells

D

Majority

Am14 IgG (Continued)

Majority

Am14 IgG (Continued)
Conclusions

These studies indicate that immunization of AM14 HC and AM14 HC/siae^{A2/A2} mice with either IgG2a^a or with an IgG2a^a-OVA conjugate fails to drive meaningful levels of class switching or somatic hypermutation in IgG2a^a-specific B cells. In addition, immunization of AM14 HC/siae^{A2/A2} mice does not result in any consistently detectable increase in levels of CSR or SHM above those of AM14 HC mice. Therefore, we conclude that AM14 mouse model is unfortunately not useful for investigating the role of Siae or any other enzyme/protein in generation/maintenance of B cell ignorance in a non-transfer setting, despite the low affinity of the knocked-in BCR and the small number of antigen specific B cells present prior to immunization. It is possible that the levels of antigen-specific secreted antibody in the serum of the AM14 mice are still high enough to opsonize IgG2a^a or IgG2a^a-OVA antigen so that it is sequestered before it can stimulate naïve specific B cell to migrate to T cell zone, interact with T cells, or undergo affinity maturation in a germinal center reaction. However, it remains formally possible that there are other explanations for the lack of detectable activation of AM14 BCR-expressing B cells.

Contributions: The HEL BCR transgenic and B1-8 transgenic studies were conducted by Hamid Mattoo. I performed studies on AM14 cells working with Kristin Keifer in Ann Marshak-Rothstein's laboratory and animal studies on AM14 knockin mice working under the supervision of Hamid Mattoo.
Chapter Five

Conclusions and Future Directions
Sialic acid acetylesterase (SIAE) moderates signaling through the B cell receptor (BCR) by removing the 9-O-acetyl moieties from acetylated α2-6-linked sialic acids on CD22 ligands. This modification is critical since CD22 cannot bind to 9-O-acetylated α2-6-linked sialic acid-containing ligands [119]. Inhibitory signaling through the CD22 receptor occurs when CD22 binds ligand, is phosphorylated by Lyn on three intracellular ITIM motifs, resulting in the recruitment of SHP-1, which then inactivates several factors downstream of the BCR [38, 39]. This CD22-mediated inhibition of BCR signaling plays an important role in establishing a threshold for activation of B cells in peripheral lymphoid tissues.

Previous work demonstrated that SIAE-deficient mice have increased signaling through the BCR, decreased tyrosine phosphorylation and SHP-1 recruitment upon BCR stimulation, deficiencies in marginal zone (MZ) B cells and recirculating bone marrow (BM) perisinusoidal B cells, spontaneously develop class-switched anti-DNA and anti-chromatin antibodies and IgG immune complex deposits in the kidneys [124].

Deep exon sequencing of the SIAE gene in patients from multiple autoimmune disease cohorts has led to the identification of more than 67 rare disease-associated non-synonymous single nucleotide polymorphism (ns-SNP) variants [126, 131]. Many of these SNPs were recreated in FLAG-tagged human cDNA expression vectors and many of the encoded proteins were found to be catalytically dead in in vitro experiments. All but one of the catalytically dead variants failed to be secreted from transfected 293T cells and all were found to function in a dominant-interfering manner to repress activity of co-transfected wildtype protein [131]. Variants of SIAE confirmed to encode catalytically dead proteins are linked to autoimmunity with an Odds Ratio of 3.51 and a two-tailed p value by Fisher’s exact test of 0.0077 [126].
In the current work, we describe the follow-up studies to further understand the effect of rare, disease-associated variants on protein folding, structure, and function and to investigate the usefulness of algorithmic predictive assays in predicting these effects. We also describe studies to determine the oligomeric state of the SIAE protein. Finally, we describe the results of an AM14 mouse-based experiment to examine, *in vivo*, the role of SIAE in the maintenance of peripheral tolerance in a low-affinity, self antigen-specific model. In this chapter, we discuss some of the strengths and limitations of these studies and future directions for studies on the SIAE protein.

**Can the limitations of algorithmic predictive programs in predicting effects of rare genetic single amino acid variants be overcome by combining results with other methods?**

The functional analysis of rare, disease-associated genetic variants, whether they are identified through genome-wide association studies or deep sequencing of gene candidates, ultimately requires a gene-by-gene approach. For candidate genes like *SIAE*, which encode enzymatic proteins, functional analysis is aided by *in vitro* enzymatic assays. For genes with non-enzymatic protein products, functional analysis is made more complicated. Our initial *in vitro* functional analysis of *SIAE* variants identified through deep exon sequencing involved enzymatic assays to determine which variant encoded enzymatically defective protein, assessment of variant protein ability to co-immunoprecipitate with wildtype SIAE, and determination that defective variants function in a dominant-negative manner to inhibit enzymatic activity of wildtype protein, thus explaining the heterozygous state of most defective variants in autoimmune patients. In this thesis, we describe further work to examine the effect
the single amino acid substitutions have on protein structure for comparison to the effects they have on enzymatic function. We used several techniques for these investigations, including the generation of a battery of monoclonal antibodies used to assess relative misfolding of variants, the use of three popular predictive algorithms to compare the predictions about protein function to our enzymatic assay and structural results, and circular dichroism studies to investigate the effect of variant amino acid changes on protein structure.

Most of our anti-SIAE monoclonal antibodies, used in immunoprecipitation and quantitative western blots, did not distinguish between enzymatically defective and active SIAE variants. Two of the antibodies, 7H2 and 1G5, had enhanced binding to SIAE variants compared to wildtype SIAE, regardless of the enzymatic function of the variants. Two other antibodies, 2F2 and 1D11, showed differential (reduced or enhanced) binding to some variants and not others, but not in a manner that aligned with enzyme activity. One of the five antibodies used for our studies, 4A4, did seem to bind variant proteins with defective or significantly reduced enzyme activity less strongly than it bound to wildtype SIAE. We did find that one variant in particular, the R479C SIAE protein, was bound poorly by all but one of our monoclonal antibodies. We suspected that this variant might be more severely misfolded than some of the other defective variants, which would have been supported by the predictions from predictive algorithms as well, all three of which predicted R479C SIAE to be defective. However, our circular dichroism results did not reflect a more severe misfolding of this variant than that observed for a few other defective SIAE variants. Overall, although the group of monoclonal antibodies we generated was not useful for making specific conclusions about SIAE protein structure compared to function, one of the monoclonal antibodies did seem to distinguish between variants with defective and normal function with reduced binding compared to binding
to wildtype SIAE. Nevertheless, we believe that a set of monoclonal antibodies could prove to be a useful tool for investigating structural changes resulting from amino acid substitutions resulting from rare, disease-associated variants. In particular, such a set of monoclonal antibodies might be particularly useful if generated against specific portions of epitopes of a protein or if the epitopes recognized by the antibodies were determined. If the specific epitopes were known, the results of differential binding studies might be more informative about whether structural effects of single amino acid substitutions are global or more localized. In addition, we might be able to predict which antibodies would bind with altered strength to specific variants versus wildtype SIAE protein.

The three predictive algorithms that we used to compare to our enzymatic assay results and monoclonal antibody binding data - SIFT, PolyPhen-2, and Provean - had some significant limitations in our studies. Comparison of results of our SIAE enzymatic functional assay to the predictions generated by PolyPhen-2 reveal an 11.3% error rate for the 53 SIAE variants tested, with predictions more frequently erroneously predicted to be benign when the SIAE variant was catalytically defective. SIFT predictions were erroneous 28.3% of the time and is also more likely to erroneously predict that a substitution will be tolerated. Provean predictions were incorrect 18.9% of the time and were also more likely to erroneously predict a defective substitution to be neutral. These results indicate that a significant number of defective rare variants might be missed using each of the three predictive algorithms we tested in our studies. Furthermore, most of the defective variants erroneously predicted to be benign/tolerated/neutral were commonly erroneously predicted between at least two of the algorithms, indicating that any use of prediction algorithms should probably utilize multiple algorithms and any variant predicted to be damaging by any algorithm is worth investigating functionally. Most variants
erroneously predicted to be damaging were also mispredicted by two or three of the algorithms, indicating that use of these predictive algorithms might result in a high number of variants to be tested functionally which would then be found to be functionally normal. PolyPhen-2 is both the most elaborate, and least erroneous, of the three predictive algorithms we tested. We tested three of the most popular algorithms for predicting deleterious effects of rare variant-encoded single amino acid substitutions and have found that all three produce high rates of both false positive and false negative predictions. We conclude that multiple algorithms should be used to be truly predictive prior to use of functional assays and that other non-computational methods should be used in addition to verify results if efficient functional assays are not available.

The results of our circular dichroism studies are preliminary at best. The spectra seem to indicate that disease-associated variants likely encode partially misfolded proteins but likely retain significant structural integrity, consistent with their retained ability to form multimers and affect wildtype SIAE function in a dominant-negative manner. Briefly, the mutants analyzed that are not secreted from transfected cells – M89V, C196F, and R479C – appeared to be more structurally compromised than the Y349C mutant, which is catalytically defective but has only reduced secretion. The M89V mutant is catalytically active but not secreted, while the C196F and R479C mutants are catalytically dead and not secreted. We were somewhat surprised that the protein structure as observed through CD studies seems to be more similar to wildtype in a mutant that is catalytically dead with reduced secretion (Y349C) than in the mutant that is catalytically normal but not secreted (M89V). We had expected the M89V mutant to give a CD spectra that looked basically similar to wildtype. Still all of the mutant SIAE protein CD spectra vary considerably from the wildtype spectra.

We believe that CD spectroscopy could be a useful tool for investigating structure and,
possibly even predicting function, of disease-associated variant-encoded proteins. However, this approach does require optimizing a good protein purification protocol for obtaining sufficiently pure (at least 95% by SDS-PAGE/Coomassie Blue staining [213]) protein reliably. Fortunately, CD does not require large quantities of protein, unlike X-ray crystallography or nuclear magnetic resonance (NMR). In order to confirm our preliminary results, and obtain better CD spectra, it will be important to purify larger quantities of protein. We initially used SIAE protein concentrations that were somewhat on the low side of what is usable for obtaining good CD spectra. Although we used wildtype SIAE purified from supernatants of stably-transfected U2OS cells for comparison to the protein purified from lysates at the same concentration, when we utilized higher concentrations (not shown) the spectra were somewhat cleaner. So, scaling up the transfection and purification protocols to obtain a larger amount of SIAE protein from lysate would likely be beneficial. Although, at some point, scaling up protein concentration might make NMR practical and more informative.

Testing different solvents for final suspension of proteins may also be informative. We only attempted CD with SIAE proteins suspended in Tris-buffered saline (TBS). In our experience, SIAE protein is fairly stable when stored at 4°C in TBS, retaining high activity levels even after 15 months. However, it might also have been useful to lyophilize proteins prior to CD so that they could be stored more stably and purified sequentially in larger amounts per mutant. Solvent choice can affect conformational integrity of protein, and factors such as pH of the solvent (which should be considered with respect to the pKa of the protein) and ionic strength can affect surface charges and folding of the protein [213].

Finally, in order to expand and draw more conclusions from CD spectroscopy about the nature of SIAE variant protein structure and function, it would be beneficial to examine the
spectra of several more variant proteins in order to compare, in a more complete way, variants that are catalytically defective but secreted in some amounts, catalytically defective and not secreted, or catalytically normal and secreted, etc. If a definitive CD22 ligand could be identified, it might also be interesting to use CD spectroscopy to study how wildtype and variant SIAE proteins bind to that ligand.

**Do epigenetic mechanisms of SIAE regulation explain the “missing heritability” not accounted for by rare genetic variants identified through exon deep sequencing?**

Genetic studies involving both common and rare genetic variants have provided significant information about disease susceptibility. However, there is a significant amount of “missing heritability” or a “susceptibility gap”, meaning the risk attributable to all identified susceptibility loci and variants accounts for only a small proportion of the total genetic susceptibility [196, 197, 211]. Defective SIAE genetic variants are found in only 2-3% of any autoimmune disease cohort. What accounts for the rest of the apparent heritability of disease susceptibility? Obviously, other loci are likely to be involved. However, there may be other, perhaps epigenetic factors, at play. Epigenetics is the study of heritable changes in gene expression caused by DNA or RNA being repressed or expressed through chemical modifications, rather than changes in genetic sequence. Mechanisms that produce changes in gene expression include DNA methylation and chromatin remodeling through modification of histones (by acetylation, methylation, ubiquitylation, phosphorylation, sumoylation, ribosylation and citrullination).

Our laboratory has recently discovered that **SIAE translation is impaired in naïve B cells**
in over 50% of lupus and rheumatoid arthritis (RA) subjects but in less than 1% of controls (unpublished observations). Autoimmune subjects also have increased 9-O-acetylation of sialic acid. This impaired translation and increased 9-O-acetylation of sialic acid in autoimmune subjects appears to be mediated by selective expression of a specific miRNA. However, it is not yet clear if this selective miRNA expression is causal or secondary to disease. It is thought that approximately 50% of miRNAs are associated with CpG islands, which are highly-methylated regions of DNA and are typically poorly expressed [276]. In addition, some miRNA expression appears to be regulated through histone modification and others are regulated through combined DNA methylation and histone modifications that work in tandem [276]. Therefore, these findings in lupus and RA patients indicate that epigenetic regulation of a miRNA that mediates SIAE translation may account for a significant proportion of missing heritability for these diseases, not accounted for by SNPs, and the single base pair variants they encode, in the SIAE gene.

**Does SIAE regulate other Siglecs?**

The SIAE-deficient mice have a phenotype similar in many ways to that of CD22-deficient mice, including increased BCR signaling [124]. Interestingly, SIAE-deficient mice develop switched anti-DNA and anti-chromatin antibodies, as well as deposits of IgG immune complexes in the kidneys as early as 20 weeks of age. Mice deficient for CD22, on the other hand, do not spontaneously develop anti-DNA and anti-chromatin antibodies until about 9 months of age and immune complex kidney deposits are not seen in CD22-deficient mice. The stronger autoimmune phenotype in SIAE-deficient mice compared to CD22-deficient mice implies that SIAE might be regulating multiple Siglecs, not just CD22. Siglec G/10 may also be
regulated by SIAE function [127]. Siglec G appears to be somewhat restricted in its expression to cells of the B cell lineage, and is particularly expressed in B1a cells, where it acts as an inhibitory receptor and controls BCR-mediated calcium signaling [62]. Like CD22-deficient mice, Siglec G-deficient mice do not exhibit an autoimmune phenotype. However, CD22/Siglec G double deficient mice develop an autoimmune phenotype resembling that of SIAE-deficient mice with spontaneous accumulation of autoantibodies and glomerulonephritis [127, 277]. These results suggest that SIAE may play a role in regulating Siglec G. There is evidence that Siglec G binds sialic acid on certain CD24 glycoforms, but it is not yet known whether this recognition is limited to sialic acid that is non-9-O-acetylated [51]. Further investigation on a possible role for SIAE in the regulation of Siglec G, and possibly of other Siglecs, is needed.

**Is there a role for SIAE in other cell types?**

In addition to its role in B cells, SIAE may be functionally relevant in other immune cells. Data from the Immunological Genome Project (ImmGen) indicates that both red pulp macrophages (RPMs) and plasmacytoid dendritic cells (pDCs) express higher levels of Siae mRNA than B cells. In addition, both CD11b- and CD11b+ conventional dendritic cells (cDCs) express comparable levels of Siae mRNA to those expressed in B cells (not shown) [278]. Ilka Netravali has demonstrated that plasmacytoid dendritic cells (pDCs) and their progenitors express levels of 9-O-acetylated sialic acid similar to those levels seen on B cells (unpublished data). Therefore, she next investigated whether levels of 9-O-acetylated sialic acid would be increased in the absence of Siae in pDCs as it is in Siae-deficient B cells. However, staining with CHE-FcD, a chimeric Fc fusion protein that uses the influenza C hemagglutinin activity to
bind and detect 9-O-acetylated sialic acid, showed that these levels did not increase in Siae-deficient pDCs from either the bone marrow or spleen (unpublished data).

Netravali also showed that CHE-FcD staining is high in total splenic CD11b⁺ cDC cells but that different levels of CHE-FcD staining correlated with different cDC subsets. Specifically, CD8⁺ cDCs have lower-intensity CHE-FcD staining than CD4⁺ and CD8αCD4⁻ cDC subsets, which express similar and high levels of 9-O-acetylated sialic acid. However, in Siae-deficient animals, CHE-FcD staining levels on CD8⁺ cells increased to the levels seen in CD4⁺ and CD8αCD4⁻ cDC subsets, which were not altered compared to wildtype.

When she examined non-DC myeloid cells, Netravali found that, consistent with their high Siae mRNA expression levels, red pulp macrophages had increased levels of surface 9-O-acetylated sialic acid when Siae-deficient compared to wildtype. Surface 9-O-acetylated sialic acid on neutrophils, which have high CHE-FcD staining but low Siae mRNA expression levels, remained unchanged in Siae-deficient mice. Monocytes have modest levels of Siae mRNA expression but have reduced 9-O-acetylated sialic acid in Siae-deficient mice. The mechanism for this is not clear.

We had previously noted that Siae-deficient mice have a dramatic leukopenia characterized by loss of marginal zone B cells in the spleen [124]. Netravali further examined absolute numbers of myeloid cell types in the spleen and bone marrow of Siae-deficient mice. She found significant changes in absolute cell counts of almost all myeloid cell types in spleen. She noted 2-fold increase in granulocytes and 3-fold increase in red pulp macrophages. Most notably, she observed a 7-fold reduction in pDCs. To further explore this decrease in pDCs, Netravali looked at the numbers of pDCs and their progenitors in the bone marrow of Siae-
deficient mice. She observed a 2-fold increase in granulocytes, which may explain their increase in the spleen. However, pDCs and their progenitors were present in similar absolute numbers in the BM of wildtype and Siae-deficient mice, indicating that development of pDCs in the bone marrow is not affected in Siae-deficient mice and reduction in the periphery does not reflect defects in pDC development.

Taken together, these recent results indicate that Siae may play roles in regulating innate immune responses of cell types other than B cells, including red pulp macrophages, neutrophils, pDCs, and cDCs. However, further exploration of the role of Siae in these other cell types, and downstream effects on adaptive immune responses, including the interaction of cDCs and T cell responses, is required.

**CASD1 as a target for small molecule therapies for the treatment of autoimmune disease**

Many new small molecule therapies, including Syk inhibitors, JAK inhibitors, and BTK inhibitors are being investigated for use in the treatment of autoimmunity. The three most studied thus far are Fostamatinib, Tofacitinib, and Ibrutinib. Fostamatinib inhibits ATP binding by the catalytic domain of Syk and thereby inhibits Syk kinase activity [279]. It has been tested in Phase II clinical trials for lymphoma, rheumatoid arthritis, and autoimmune thrombocytopenia [279-282]. In addition, fostamatinib has been shown to inhibit disease progression in mouse models of lupus [281, 283, 284] and to block MHC class II-restricted presentation of B cells, priming of cytotoxic T cells, and disease onset in mouse models of type I diabetes [281, 285]. Tofacitinib is an inhibitor of JAK3. It has received FDA approval for the treatment of rheumatoid arthritis and is approved for RA treatment in several other countries [286-288].
Tofacitinib has also shown promise in Phase II trials for the treatment of ulcerative colitis [289], for which it has entered Phase III trials, Crohn’s Disease [290] and psoriasis [291-294]. In addition, one 25 year old male patient treated with tofacitinib for psoriasis, experienced a nearly complete remission of alopecia universalis in addition to decreased psoriasis [295]. Ibrutinib, a BTK inhibitor, selectively binds cysteine 481 and blocks BTK activation and kinase activity [296]. This inhibitor is both safe and efficacious in clinical trials and has been in Phase III trials for a number of B cell tumors and has been approved for treatment of mantle cell lymphoma [297, 298] and chronic lymphocytic leukemia [298] and is in development for diffuse large B-cell lymphoma and multiple myeloma [296, 299]. Ibrutinib and other BTK inhibitors have also shown promise in pre-clinical studies and mouse models of rheumatoid arthritis and systemic lupus erythematosus (SLE) [296, 300, 301].

Our lab recently knocked out CASD1 and confirmed it as the sialic acid 9-O-acetyl transferase (SIAT), which adds acetyl groups to sialylated N-glycans, in B cells and myeloid cells. When CASD1 is knocked out, surface 9-O-acetylation is decreased in most tested cell types, but is not decreased in CD4+ T cells, indicating that a second SIAT must be present in these cells (unpublished data). Based on the success of other small molecule inhibitors of enzymes involved in B cell signaling, we have hypothesized that small molecule inhibitors of CASD1 may be of therapeutic value for the treatment of autoimmune disease like lupus, rheumatoid arthritis, and others. By inhibiting CASD1, and preventing the addition of 9-O-acetyl groups to sialylated N-glycans, CD22 would be able to bind to ligand constitutively, allowing consistent prevention of B cell activation. The lab has therefore started a screen for small molecule inhibitors of CASD1.
References


126. Chellappa, V., et al., M89V Sialic acid Acetyl Esterase (SIAE) and all other non-synonymous common variants of this gene are catalytically normal. PLoS One, 2013. 8(1).


