Evaluation of Predictive Computational Modelling in Biologic Formulation Development

by

Jayanthi Jayakumar

Bachelor of Science in Biological Engineering, Massachusetts Institute of Technology (2010)

Submitted to the MIT Sloan School of Management and the Department of Mechanical Engineering in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration

and

Master of Science in Mechanical Engineering

In conjunction with the Leaders for Global Operations (LGO) Program at the

MASSACHUSETTS INTITUTE OF TECHNOLOGY

June 2017

© Jayanthi Jayakumar, MMXVII. All rights reserved.
The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Author Signature redacted

Department of Mechanical Engineering & MIT Sloan School of Management

May 12, 2017

Certified by Signature redacted

Collin Stultz, Thesis Advisor

Professor of Electrical Engineering and Computer Science

Certified by Signature redacted

Roger Kamm, Thesis Advisor

Cecil and Ida Green Distinguished Professor of Biological and Mechanical Engineering

Certified by Signature redacted

Thomas Roemer, Thesis Advisor

Senior Lecturer in Operations Management and Executive Director of LGO

Accepted by Signature redacted

Rohan Abeyaratne

Chairman, Committee on Graduate Students, MIT Department of Mechanical Engineering

Accepted by Signature redacted

Maura Herson

Director of MIT Sloan MBA Program, MIT Sloan School of Management

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 20 2017
LIBRARIES
ARCHIVES

Evaluation of Predictive Computational Modelling in Biologic Formulation Development

by

Jayanthi Jayakumar

Submitted to the MIT Sloan School of Management and the Department of Mechanical Engineering on May 12, 2017 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Mechanical Engineering

ABSTRACT

Computational modelling has completely redefined the experimentation process in many industries, allowing large sets of design concepts to be tested quickly and cheaply very early in the innovation process. Harnessing the power of computational modelling for protein drug formulation has numerous, currently unrealized, benefits. This project aims to be the first step in the development of a high throughput predictive computational model to screen for excipients that would decrease protein aggregation in solution and thus increase its stability and enable clinical effectiveness.

Protein drug formulation currently relies heavily on empirical evidence from wet-lab experiments and personal experience. During the biologic drug development process, proteins that target specific disease pathways are identified, developed, isolated, and purified. Scientists then conduct a series of wet-lab experiments to identify the optimal formulation that will allow the protein to be used as a drug therapy. A critical part of formulation development is the identification of inactive ingredients called excipients that perform various important functions including prevention of protein aggregation. Despite their critical role in enabling proteins to be effective therapies, very little is understood about excipient-protein interaction. Furthermore, often a limited set of compounds are tested for their use as excipients since wet-lab experiments are expensive and time consuming.

This project accomplishes the following goals:

- Identification of databases of compounds that could be used as excipients in biologic formulation
- Development of a high throughput method to computationally model a target protein and 247 potential excipients
- Evaluation of potential relationship between computational output and wet-lab results based on experimentation with 32 of the 247 excipients
- Recommendations on next steps that include feedback on types of proteins and excipients to be tested for the validation of the method developed in this project

Thesis Advisor: Collin Stultz
Title: Professor of Electrical Engineering and Computer Science

Thesis Advisor: Roger Kamm
Title: Cecil and Ida Green Distinguished Professor of Biological and Mechanical Engineering

Thesis Advisor: Thomas Roemer
Title: Senior Lecturer in Operations Management and Executive Director of LGO

This page is intentionally left blank.

ACKNOWLEDGEMENTS

Thank you to the MIT Leaders for Global Operations (LGO) program and Amgen Inc. for their support of this thesis.

In particular, I would like to express my deepest gratitude to Pavan Ghatty, my mentor at Amgen. He not only played a critical role in guiding me on this project, but continuously inspired me about the incredible possibilities of using computational modeling in understanding atomic level interactions. I would also like to thank Sekhar Kanapuram, for his support and mentorship throughout the project. I am grateful to the numerous people at Amgen who played an important role in the successful execution of this project including Gerd Kleemann, Sam Guhan, Ping Yeh, Arnold McAuley, Larysa Alekseychyk, Vladimir Razinkov, Bill Callahan III, Twinkle Christian, Hao Lou, Han Ke, Derek Miller, Dave Stadolnik, Dollie Grajczak, and the Pre-Pivotal Drug Product Technology Team. I would also like to thank the LGO alumni and numerous other people throughout Amgen who always made time to meet with me, provided me with greater insight about the company, and inspired me. I am grateful to have shared this experience with Kevin Peng, my LGO friend and fellow intern at Amgen.

MIT will always be my home and its mission "Established for advancement and development of science, its application to industry, the arts, agriculture, and commerce" is one that I deeply believe. There are numerous people at MIT that are critical to the success of this project and my growth and development over the last couple years. I would like to express my sincere gratitude to my advisors Collin Stultz, Roger Kamm, and Thomas Roemer. Professor Stultz helped advanced my approach to research through his mentorship. Professor Kamm's and Professor Roemer's support and feedback throughout the project provided me with the confidence to pursue a relative new area of research for me. I am grateful to have had the opportunity to learn from them and hope to continue to do so in the future. I would also like to thank the LGO faculty for always offering their support. My LGO classmates have kept me grounded throughout the last two years and I look forward to the growth of our friendship in the coming decades.

My parents and my sister are everything I am and everything I aspire to be and their support throughout this whole experience has meant the world to me. Thank you to my mom, Vasanthi Jayakumar, for the unconditional love and providing me with the courage to conquer anything. To my dad, Jayakumar Thirumalai, for being my role model and providing me with an insatiable appetite to learn. To my sister, Madhuvanthi Jayakumar, for her supreme chillness and for being the source of my strength. I need to express my heartfelt thank you to Ravi Deedwania, my best friend, whose unwavering faith in me inspires me to be the best possible version of myself. I would also like to thank Sowmiya Sivanantham, Cathy Deedwania, Prakash Deedwania, Anne Deedwania, Jack Crane, David Crane, David Deedwania, Amit Agrawal, Anju Nattanmai, and Divya Murugesan. They made this Bostonian feel at home in southern California and helped me fall in love with it.

This page is intentionally left blank.

TABLE OF CONTENTS

This page is intentionally left blank.

LIST OF FIGURES

This page is intentionally left blank.

LIST OF TABLES

This page is intentionally left blank.

1. Background on Protein Therapeutics

1.1 Industry Overview

The treatment options available to patients can be broadly divided in two major categories: surgical intervention and pharmacological intervention. Historically, pharmacological intervention has relied on small molecules, which typically have a low molecular weight (< 900 Daltons) and can usually be delivered orally as a pill. Furthermore, these molecules are derived through chemical reactions and can be manufactured at a commercial scale without reliance on living systems. Small molecules like statins and NSAIDs can have high clinical efficacy and are still used as the first line of treatment in many diseases.

However, advancements in science and engineering over the last century have led to the development of more targeted and effective therapies called large molecules or biologics. Unlike small molecules, biologics have a high molecular weight and are often delivered through intravenous or subcutaneous methods. Biologics are typically not designed for oral delivery both because of their size (e.g., IgG antibodies are in the order of 150 kilo Daltons) and because they will be rendered ineffective after they are metabolized in the digestive system. Another important difference is that they are produced by living systems that range from bacteria and yeast cells to mammalian cells. The type of living system that is used for manufacturing depends on the protein itself (e.g., if the protein has a specific glycosylation pattern that can only be produced by certain types of cells) and the cost-effectiveness of the living system at a commercial scale. There are many different categories of biologics including protein therapy, cell therapy, gene therapy, vaccines, and tissue engineering. Of these, protein therapeutics are the most prevalent, with over 100 FDA approved drugs in the market (Leader, 2008).

Protein therapeutics have changed the treatment paradigm for many serious diseases since the approval of the first biologic, recombinant human insulin, in 1982 (Altman, 1982). Deeper understanding of disease pathophysiology has led to discoveries about the involvement of various proteins in the disease pathways. Often either mutations or variations in concentrations of key proteins lead to chronic, debilitating, and degenerative disorders. This realization has led scientists to identify and develop proteins to target those disease pathways. Although some protein therapeutics are endogenously produced and isolated from humans or other mammals, many protein therapeutics depend on recombinant DNA technology.

Once scientists know the amino acid sequence of a protein that plays a role in a disease pathway, they can use recombinant DNA technology to modify the protein and insert the DNA sequence in host cells. These host cells grow and replicate in large bioreactors where they act as miniature "factories" that produce the target protein. In a series of complex steps, these proteins are then isolated from the cells, purified, and formulated to be used as an injectable or infusible drug.

Protein therapeutics offer numerous clinical benefits that were previously unrealized with small molecules. Because they are targeted to particular disease pathways (e.g., a specific receptor on a cancerous cell), often they are more efficacious and have fewer side effects. Furthermore, they are less likely to trigger an immune response since they are usually based on proteins found endogenously in the human body. Lastly, the use of recombinant DNA technology allows scientists to engineer the protein for enhanced functionality and stability.

However, there are still important drawbacks with protein therapeutics for patients. While many protein therapeutics are disease modifying agents, few actually lead to cure, leaving large gaps in care. The intravenous or subcutaneous delivery of protein therapeutics negatively impacts patients' quality of life since they often either have to drive to the physician's office or an outpatient setting to receive treatment or inject themselves at home. Protein therapeutics, are significantly more expensive than small molecules. One study indicates that despite only 2% of the population receiving biologics, over 40% of prescription drug spend in the US is allocated to it (Glover 2015). These concerns regarding drug pricing continue to grow in importance as people's average lifespan increases and many protein therapeutics are used as chronic treatments.

Biotechnology companies and biopharmaceutical companies also have some business incentives to focus on protein therapeutics. Although both small molecules and protein therapeutics have to follow the drug approval process shown in Figure 1, protein therapeutics typically gain approval faster (Leader, 2008). In addition, 13%-14% of protein therapeutics in phase 1 are likely to gain approval compared to only 7.6% of small molecules in phase 1 (Hay, 2014). Although they are in general more difficult to manufacture than small molecules, recombinant DNA technology allows for fast, scalable, and reliable commercial production that is not available with non-recombinant proteins (Leader, 2008). Biologics typically also have longer exclusivity rights and less competition from biosimilars, compared to branded small molecules and their generic counterparts, because of the relative complexity of developing and manufacturing them and the difficultly associated with developing identical replication for biosimilars. This, along with the clinical benefits that it offers, enables biotechnology companies to price biologics on average 22 times more than small molecules (Richardson, 2013).

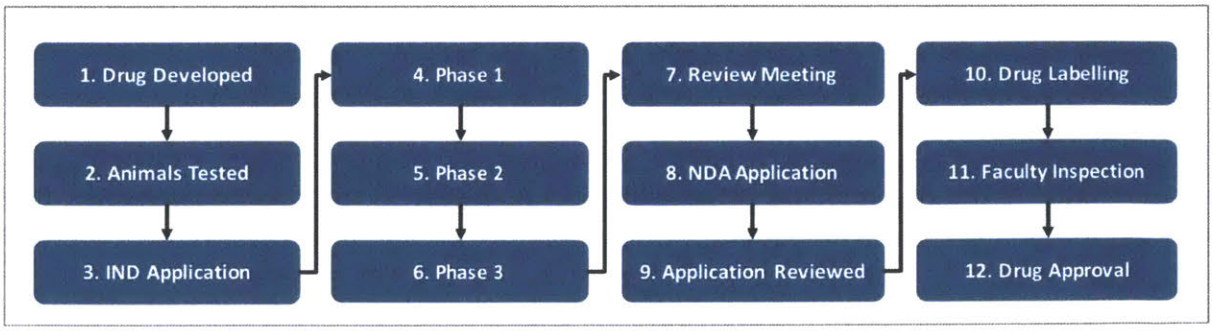


Figure 1: Overview of the FDA Drug Approval Process

("Drug Approval Process", 2016)

However, there are also important business risks that companies take by investing in protein therapeutics. Despite having a higher likelihood of getting approval compared to small molecules, there is also only a relatively low rate of successful progression to consecutive stages of development, as shown in Figure 2. Furthermore, a report by Tufts Center for the Study of Drug Development indicates that the cost to develop a new drug is approximately $2.6 billion, a 145% increase from approximately $1B in its 2003 report (Peters, 2014). While this is arguably offset by the high prices that biotechnology companies charge for protein therapeutics, there are increasing cost pressures and rising threat of biosimilars that are likely to affect their profit margin in the coming years.

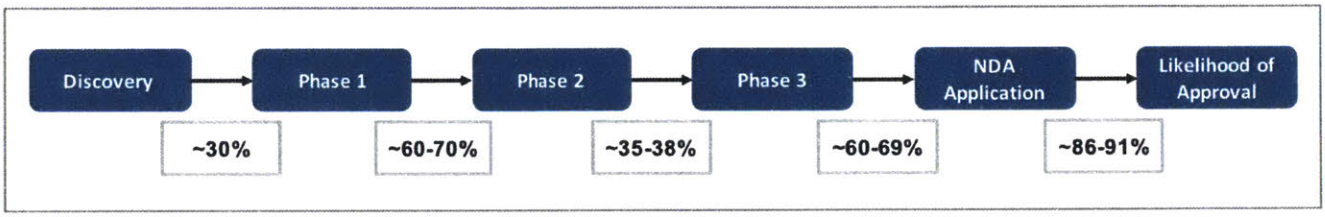


Figure 2: Rate of Success Between Each Stage of Development for Biologics (Hay, 2014; Bains, 2004)

Despite their business risks and clinical drawbacks, protein therapeutics will continue to play an important role in treating serious diseases given their ability to be targeted and efficacious, with fewer adverse events. Many current protein therapeutics are based on monoclonal antibodies. However, in the coming years, new modalities of protein therapeutics will continue to be introduced including bispecific T-cell engagers (BiTEs), protein-drug conjugates, and engineered zymogens (Tobin, 2014). Through the use of protein engineering, these new modalities are expected to offer better treatment options for patients.

1.2 Amgen's Role

Amgen is one of the world's leading independent biotechnology companies with a market cap of approximately $135 billion ("Amgen Inc.", 2017). It was founded in 1980 as Applied Molecule Genetics Inc., by venture capitalist, William Bowes, and a staff of three in Thousand Oaks, California ("The Amgen Story", 2015). It launched its first drug, Epogen, in 1989, just 7 years after the approval of recombinant human insulin in 1982, and its second drug, Neupogen, in 1992("The Amgen Story", 2015). Both Epogen and Neupogen are blockbuster biologics that are indicated for treatment of anemia in patients with Chronic Kidney Disease (CKD) and neutropenia in patients receiving chemotherapy, respectively ("The Amgen Story", 2015). Although sales for both biologics have been declining over the last couple years due to competition, Epogen has made over $40 billion in the last 28 years and Neupogen has made approximately $1 billion to $1.5 billion each year over the last few years (Pollack, 2012; "Amgen's 2015 Revenues", 2016).

Amgen currently has 20,000 employees and approximately $21.7 billion in total global revenues ("About Amgen", 2017). It has operations and/or affiliated subsidiaries in 52 countries and serves millions of patients in over 100 countries ("About Amgen", 2017). Starting with Epogen in 1989, Amgen currently has a total of 15 drugs in the market ("About Amgen", 2017). Of these, 12 are biologic and more specifically 10 of these are protein therapeutics. A more detailed portfolio of Amgen's currently marketed drugs is shown in Table 1.

Table 1: List of Amgen's Approved Drugs As of January 2017

Name	Initial U.S. Approval	Disease Area	Type
EPOGEN (EPOETIN ALFA) RECOMBINANT	1989	Nephrology	Biologic: Recombinant Protein
NEUPOGEN (FILGRASTIM)	1991	Hematology / Oncology	Biologic: Recombinant Protein
Enbrel etanercept	1998	Inflammation	Biologic: Recombinant Fusion Protein
Aranesp (darbepoetin alfa)	2001	Nephrology	Biologic: Recombinant Protein
Neulasta (pegfilgrastim)	2002	Hematology / Oncology	Biologic: Recombinant Protein with PEG
Sensipar (cinacalcet) Tablets 30mg · 60mg · 90mg	2004	Bone Health	Small Molecule
Vectibix (panitumumab) Injection for IV Infusion	2006	Hematology / Oncology	Biologic: Recombinant IgG2
Nplate® romiplostim	2008	Hematology / Oncology	Biologic: Peptide fusion protein
prolia (denosumab)	2010	Bone Health	Biologic: Recombinant IgG2
XGEVA (denosumab)	2010	Bone Health	Biologic: Recombinant IgG2
Kyprolis (carfilzomib)	2012	Hematology / Oncology	Small molecule
BLINCYTO (blinatumomab) for injection	2014	Hematology / Oncology	Biologic: BiTES
Corlanor (ivabradine) tablets	2015	Cardiovascular	Small molecule

IMLYGIC® (talimogene laherparepvec) SUSPENSION FOR INJECTION 10^6 PFU/mL and 10^8 PFU/mL single-use vials	2015	Hematology / Oncology	Biologic: Recombinant viral therapy
Repatha (evolocumab) injection 140 mg/mL	2015	Cardiovascular	Biologic: IgG2

("Products", 2017; "Amgen's 2015 Revenues", 2016; "Aranesp", 2016; "Blincyto", 2016; "Corlanor", 2017; "Enbrel", 2016; "Epogen", 2012; "Imlygic", 2015; "Krypolis", 2016; "Neulasta", 2016; "Neupogen", 2016; "NPLATE", 2016; "Products", 2017; "Prolia", 2017; "Repatha", 2016; "Sensipar", 2014; "Vectibix", 2015; "Xgeva", 2016.)

In addition to the 15 drugs that have already been launched, Amgen also has 43 drugs in various stages of development ("Pipeline" 2016). Of these, 15 are in Phase I, 7 are in Phase II, and 12 are in Phase III ("Pipeline" 2016). The remaining 9 molecules in development are biosimilars of blockbuster biologics currently in the market ("Pipeline" 2016). Approximately 10 of the drugs in various stages of development are already approved Amgen biologics that are being explored in other indications ("Pipeline" 2016). A majority of these drugs are being developed for hematology/oncology, followed by inflammatory diseases, cardiovascular diseases, and neurological disorders ("Pipeline" 2016). Figure 3 depicts the 11 different modalities that make up the drugs in the pipeline ("Pipeline" 2016). Approximately 73% of the pipeline drugs are biologics (including the 9 biosimilars), 18% are small molecules, and remaining are unknown to the public ("Pipeline" 2016). While Amgen pursues a modality-independent strategy towards drug development, the high proportion of its portfolio dedicated towards biologics emphasizes the importance of these molecules in the future of healthcare.

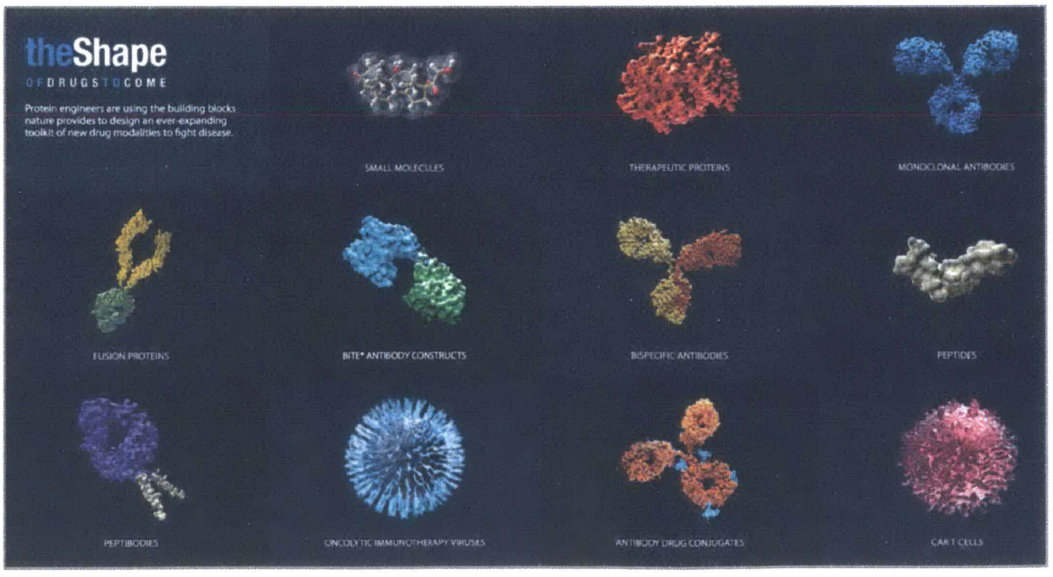


Figure 3: Modalities of Drugs Considered by Amgen

("The Shape of Drugs to Come", 2015)

Despite having a strong product portfolio and pipeline, Amgen faces numerous challenges in the coming years. Interestingly, many of these challenges require the development of better formulations. For example, some of its older drugs like Epogen, Neupogen, Enbrel, and Neulasta are already facing or are expected to face competition from biosimilars as their patents expire (Farooq 2016). In order to continue to maintain strong revenues from these older drugs, Amgen must explore opportunities to enhance their benefits and their ability to deliver better care to patients through careful lifecycle management. A strategy that is likely going to be very important with these drugs is to improve their delivery methods to increase ease of use for patients and doctors, improve quality of life for patients, and increase patient adherence and compliance. But utilization of new delivery devices will often require the drug to be reformulated to be compatible with the device. Another challenge Amgen faces is related to the development of its biosimilar portfolio. However, given the uncertain regulatory pathway in the US, Amgen faces numerous hurdles related to the launch of these drugs. One main concern is that the protein patent often expires earlier than the formulation patent for the originator drug. Thus Amgen may need a way to quickly develop new formulation for the biosimilar protein that is also similar to the originator drug's formulation. Last, but not least, Amgen also faces formulation challenges related to the new modalities of treatment (shown in Figure 3) it is developing. In the past, many of its drugs were made up of monoclonal antibodies or recombinant proteins that were often naturally stable and relatively easy to formulate. However, some of the new modalities of drugs pose interesting formulation challenges and as a result, Amgen may need better ways to determine optimal formulations faster.

1.3 Excipients in Protein Drug Development

1.3.1 Function of Excipients in Protein Therapies

Formulation development is an important step in the drug development process because it allows the drug candidate to become an effective therapeutic at a commercial scale. After the lead protein candidate has been identified through the process shown in Figure 4 below, it enters the formulation development step. Once isolated from the cells, proteins often become unstable since they no longer have the ecosystem of molecules present in the cells to stabilize them. Even a protein that is less sensitive to external conditions, may face issues over the course of the drug's shelf life. Stability issues include protein aggregation, chemical or thermal degradation, and denaturation. These stability issues have the potential to not only lower the bioavailability of the protein, but may render the protein ineffective or cause immunogenic reactions in patients.

Figure 4: Stages of Pre-clinical Discovery and Development

In addition to ensuring protein stability, scientists need to also consider other important factors during formulation development. Solubility and viscosity of proteins may require alterations so that clinically effective dosages of drugs can be delivered to patients. Formulation development can also help ensure that biocompatibility factors like isotonicity are maintained. From a business perspective, formulation development helps determine the shelf life for the drug. A protein that might otherwise become unstable in a month, could get an extension of life by two years, a common goal the industry aspires to, with the proper formulation. This allows biotech companies to take advantage of the economies of scale related to large scale manufacturing. Longer shelf life also results in cheaper drugs and more convenience for patients and physicians. Another growing trend in the industry is the focus on developing new delivery devices, which also require appropriate formulation changes. Biotech companies are motivated to develop new devices since it offers competitive differentiation and can result in increased patient adherence. The latter is a particularly important problem to tackle because patient adherence rates are expected to be approximately 20% - 50% for biologics (Degli Esposti, 2014). Delivery devices for Amgen's biologics vary from a multi-dose vial that can be used with a regular syringe to prefilled syringe, auto-injectors, and on-body injectors. Amgen also recently entered a licensing deal with Unilife for its wearable injectors ("Unilife and Amgen..." 2016). However new delivery devices often also require new formulations that are compatible with the specifications of the device. Formulation development is a critical part of the drug development process given its role in stabilizing the protein, ensuring appropriate solubility, viscosity, and isotonicity, helping lengthen the shelf life, and increasing compatibility with various delivery devices.

Although the exact process of formulation development varies depending on the biotech company, there are three primary stages to the process: pre-formulation, drug substance stabilization, and final formulation. Pre-formulation is used to understand the physical and biological properties of the protein candidate (Chang, 2002). These properties include understanding the protein's structure, solubility and stability parameters, viscosity, ionic strength, glycosylation, hydrophobicity, and self-association (Chang 2004). It is also helpful to know in vitro and in vivo activity of the protein either from studies done in early stage research or new studies. During pre-formulation, scientists also frequently establish the target pH and temperature range and the protein's response to shear stress and freeze/thaw cycles. At this stage, there is also a chance for formulation scientists to work with research scientists to identify opportunities to modify the protein itself for increased stability and/or solubility.

Once these parameters are better understood through the pre-formulation assessment, the next step is to identify a formulation that improves the drug substance stabilization. Drug substance is a purified solution containing only the protein candidate. At this stage, numerous possible tools can be used to optimize the stability including molecules called excipients, buffers, and pH adjustments. Table 2 below identifies some common stability related problems with the protein candidate, potential causes, and solutions.

*Table 2: Stability Issues with Protein Drug Candidates**

Problems	Potential Causes	Possible Solutions
Non-covalent aggregation	Solubility, structural changes, heat, shear, surface, denaturants, impurities	pH, ionic additives, amino acids, surfactants, protein concentration, raw material purity
Covalent aggregation	Disulfide scrambling, other unknown mechanisms	pH, inhibit non-covalent aggregation
Deamidation	pH < 5.0 or pH > 6.0	pH optimization
Cyclic imide	pH around 5	pH optimization
Cleavages	Protease impurity, other unknown mechanisms	pH, product purity, inhibitors
Oxidation	Active oxygen species, free radicals, metals, light, impurity	Excipient purity, free-radical scavenger, active oxygen scavengers, methionine
Surface denaturation, adsorption	Low protein concentration, specific affinity, protein hydrophobicity	Surfactants, protein concentration, pH

*(*Table from Chang, 2002)*

Once protein stability, reproducibility, and manufacturing capabilities have been established, the drug candidate is ready to be submitted for an IND application to gain approval for phase I and phase II trials. However, simultaneously it is important to begin the final stage of formulation development for phase III trials and commercial approval. While ensuring that safety, efficacy, and stability of the protein are maintained, developing a commercial formulation requires an understanding of the delivery device, clinical requirements (e.g., dosage), and expected transportation and storage conditions, especially given the patient- and physician-constraints of certain indications and costs. Excipients are heavily relied on at this point to change various properties of the protein solution including its solubility, viscosity, concentration, and compatibility with the delivery device. At this stage, it may also be important to revisit stability parameters to ensure that protein stability can be maintained over the expected shelf-life of the drug. Numerous studies are conducted with time point measurements over 1 – 2 years to ensure that the final formulation meets efficacy, safety, and delivery requirements. Regulatory approval for the final formulation is gained through the results of these experiments and typically the phase III trial results.

As seen above, excipients play a critical role in formulation development. The FDA defines excipients as "inactive ingredients" or "any component of a drug product other than an active ingredient." So these include macromolecule compounds like albumin to small molecules like amino acids, carbohydrates, lipids, salts, polyols, antioxidants and polymers. Excipients serve various functions in biologic formulation development. They can impact stability and solubility

by maintaining specific pH ranges as buffering agents or interacting directly with the proteins through non-covalent bonding (e.g., hydrogen bonds, van der Waals interactions, hydrophobic interactions). They may help provide isotonicity to help improve bio-compatibility during injection. They can also act as surfactants that reduce or prevent aggregation and surface adsorption. Excipients may also help to control protein oxidation, act as cofactors, help maintain protein conformation, and influence the viscosity. Ultimately they serve as the tool kit that allows for the formulation of a protein drug candidate into a stable drug that can typically be delivered through subcutaneous or intravenous methods at the right concentration and biocompatibility. They often serve to increase the bio-availability of the drug by helping to maintain the protein's stability and conformation.

New excipients, defined as "inactive ingredients that are intentionally added to therapeutic and diagnostic products, but that: (1) are not intended to exert therapeutic effects at the intended dosage, although they may act to improve product delivery (e.g., enhance absorption or control release of the drug substance; and (2) are not fully qualified by existing safety data with respect to the currently proposed level of exposure, duration of exposure, or route of administration" ("Guidance for Industry" 2005). They are labelled as inactive ingredients by the FDA because even though they serve to enhance the properties of the protein and/or the delivery device, they are expected to have minimal to no impact on patient safety and should produce no off-target effects. These strict regulations requiring extensive safety data for new excipients came about after a disaster in the early 20th century due to a new formulation of a drug called Elixir Sulfanilamide. The pharmaceutical company added an excipient called diethylene glycol to switch the drug from a pill to a liquid formulation (Ballentine, 1981). However, this excipient was found later to be an antifreeze agent that resulted in the death of over 100 patients before the FDA was able to recall it (Ballentine, 1981). Since then, the FDA has published the Federal Food, Drug, and Cosmetic Act of 1938, the Guidance for Industry Nonclinical Studies for the Safety Evaluation of Pharmaceutical Excipients (2005), and other documents to regulate and provide guidance to pharmaceutical and biotech companies on the extensive toxicology and safety data necessary for the approval of new excipients. Thus toxicology studies should be done for new excipients during early stage clinical trials and patient safety data should be collected by phase III trials. The FDA may require additional studies if the formulation is changed after phase III trials. The one exception to these guidelines is if the excipients are already approved by the FDA for a particular mode of delivery and listed in the FDA's Inactive Ingredient Database. These excipients may be used for new drug formulations with the same mode of delivery with potentially less extensive studies.

1.3.2 Current Methods for Selecting Excipients

Excipient selection occurs through two different methods. Some biotech companies and/or contract research organizations (CROs) begin with platform formulations for different molecule types. Platform formulations may specify a pH or a general pH range, the buffer, and a set of excipients that have been shown to support certain protein types like monoclonal antibodies. Some platform formulations may even specify development guidelines including types of experiments to perform and time points to collect. The purpose of platform formulation is to

reduce the time and cost required for formulation development, especially for a class of molecules that are believed to behave similarly. This strategy is also often used to develop formulations for first-in-human Phase I and Phase II trials and is refined for phase III and commercial use. Any refinement to platform formulations rely heavily on personal experience and expertise of the scientists developing the formulations and/or through the use of the trial-and-error method.

Another approach that is commonly used with excipient selection, particularly for formulation development of novel proteins, is design of experiments (DOE) (Hwang, 2005). If high-throughput instruments and assays are available, DOE provides a systematic approach to simultaneously test different combinations of excipients for formulation development. It also allows for statistical analysis of the results to create design space to choose the optimal combination of the variables tested. This is because DOE not only provides information on each of the variables and the experimental outcome measured, but also the relationship between the different variables. DOE also provides a method to identify the most critical variables for formulation development, providing scientists with guidance on factors that require more careful optimization. Although DOE is a promising approach, especially given the increasing use of automation to perform high throughput experiments, some scientists may still take an empirical approach based on past experiences and expertise.

Regardless of the approach taken to determine the excipients, there are some experiments that are commonly performed to test their impact. A summary of these experiments used to test protein stability are shown in Table 3 below. Although not all of these experiments are conducted for each protein's formulation development, it is recommended to do a variety – time and resources permitting – to gain more information regarding the excipient and protein interaction. In addition, other experiments may be performed depending on the stage of formulation development and the need of the protein. For example, especially during the third stage of formulation development, scientists may run experiments to optimize viscosity, solubility, and compatibility with the delivery device.

*Table 3: Analytical Approaches to Assess Impact of Excipients on Protein Stability Attributes**

Methods	Examples	Protein Stability
Column chromatography	HPLC, FPLC, low pressure LC; size-exclusion, reversed-phase, ion-exchange, hydrophobic, affinity columns, coupled with UV, fluorescence, RI, and other analytical instruments as detectors	Most physical and chemical degradations, excipient impurities, leacheates
Electrophoresis	SDS-PAGE, native PAGE, isoelectric focusing, capillary electrophoresis	Degradations with changes in size and/or charge
Spectroscopy	CD, fluorescence, FTIR, UV, Raman, NMR	Structural changes, chemical modifications of side groups
Thermal Analysis	Differential scanning calorimetry, thermogravimetric anlaysis, thermomechanical analysis	Protein structure, lyophilized cake structure, powder characterizations
Light scattering/ turbidity	Dynamic light scattering, other light scattering devices, turbidity, particle size determination, particle counter	Aggregation, precipitation, molecular weight determination
Other micro characterization methods	Peptide mapping, peptide sequencing, amino acid analysis, mass spectrometry, other specific analyses for individual reactive groups	Identification of impurities and chemical degradation, analysis of complex proteins, e.g., antibody and glycoprotein

*(*Table from Chang, 2002)*

1.3.3 Areas of Opportunities

Given its important role in translating a research molecule into a commercial drug, the formulation development process is fairly well defined in many biotech companies and CROs. Whether it is a platform approach, design of experiments (DOE), trial-and-error, or a combination, companies may have unique methods for formulation development. Typically, formulation scientists are also given limited timeline and a limited set of resources to produce formulations that meet certain thresholds for the different stages of the development process. These requirements, in addition to the FDA regulations, are often from the product management and commercial teams that set goals for development. It is not clear that the best formulation, or even a close to best formulation is identified for protein drugs since scientists are often working to meet the requirements with severe resources constraints. Thus there is significant room for improvement of this process.

First, an increased collaboration between the discovery research team team and the formulation development team may allow for optimization of the protein structure to meet formulation need. This could be a challenging process since scientists who are involved in selecting the optimal drug candidate during the pre-clinical stages will have three main considerations:

- Selecting a molecule that is likely to provide the most clinical benefits
- Ensuring that it can successfully and relatively easily be scaled-up for commercial scale manufacturing
- Identifying a structure that has high stability and may be easy to manufacture

Thus far, the first two considerations often take priority since it has a direct impact on the commercial success of the drug and scientists have been able to deliver formulations that meet the requirements. However, as biotech companies explore new modalities of protein therapies that provide new challenges for formulation, it may become even more critical to select protein candidates from the discovery research stage that have shown to be more stable.

Second is that the existing methods take up significant resources and do not provide space for continuous innovation with formulation development. The three approaches discussed in the previous section naturally provide a limitation on the number of formulations that can be tested and do not really provide an incentive to identify the optimal formulation. In particular, with the platform formulation approach, even the addition of new knowledge to the process is limited since the same platform is tested with potentially slight variations. Additionally, even with time and resources constraints, all three of these approaches can be a burdensome process because these are not set up to gain scientific insight into excipient-protein interaction.

The last is that improved communication and collaboration between the formulation department and the manufacturing department may help efficiently deliver final formulations that are also optimized for manufacturing capabilities. Although not a major issue at the moment given that many protein therapies are monoclonal antibodies with shared properties, this could become particularly important with new modalities. There are numerous benefits that could be realized from increased communication between these groups, from consideration of the cost of specific excipients at a commercial scale to understanding the flexibility and the doable strict requirements for manufacturing department.

2. Introduction to Project

2.1 Problem Statement

Formulation development is increasingly becoming a critical challenge in protein drug development. There are in fact three industry trends that will only further amplify this in the coming years:

1. New modalities of protein therapeutics that will likely require new formulation designs
2. Advancements in delivery devices that may require re-formulation of currently approved protein therapeutics
3. Biosimilars may require new formulation if the formulation patent life for the originator drug has not yet expired

It is important that advancements in formulation development match the advancements in protein engineering and the changing formulation needs. Although the increased utilization of high-throughput equipment has made it easier to experimentally test a larger set of formulation designs, the process to discover new excipients and to screen them have not changed notably over the last couple of decades. Scientists still primarily rely on platform formulations, DOE, trial-and-error, and personal experience. Furthermore, the cost, resources, and time required to develop new formulations, particularly for molecules for whom the platform formulations do not work, are still high. Lastly, it is not clear that optimal formulations are identified; rather it often seems like a formulation that sufficiently meets expectations is developed given rapid turn around deadlines.

There are numerous types of formulation challenges associated with the three main trends noted above. At the core is the ability to identify excipients that could perform various tasks like reducing viscosity, increasing stability, increasing solubility. In particular, the need for excipients that will reduce protein aggregation is likely going to increase in the coming years due to the increased preference for liquid formulations across all three trends. The level of protein aggregation is a marker of the protein's stability and is an important endpoint to track during the various stages of the drug development process. High levels of aggregation can result in a loss of clinical efficacy, changes in biodistribution of the drug, and potentially trigger immunogenicity.

Protein aggregation is defined as the irreversible association between two or more protein monomers to form dimers, trimers, or even macroscopic particles. There are numerous processes that may lead to protein aggregation. It has been shown that aggregation is often the result of association between protein monomers that are in non-native conformational state (Wang 2010, Gokarn 2006). Typically, a small proportion of proteins are mis-folded, unfolded, or denatured in equilibrium. This is related to the conformational stability of the protein and may increase throughout the different processing steps including during the expression of proteins in cells, purification, freeze/thaw, shaking and shearing, pressurization during filtration, and initial formulation development (Wang 2010). The exposure of its hydrophobic regions and the greater degree of flexibility it has enables the protein to aggregate more in its non-native conformational state (Wang 2010, Gokarn 2006). Proteins may also aggregate due

to changes in its structure and make-up after chemical degradation. Although less frequent, protein aggregation may occur in its native state (Wang 2010). This could either be due to its colloidal stability, resulting in association through electrostatic, hydrophobic, and van der Waals forces, or through direct chemical linkages, particularly due to intermolecular disulfide bond/exchange (Wang 2010). Different proteins may have varying levels of propensity to aggregate through these various methods.

There are no clear regulatory guidelines regarding acceptable levels of aggregation (Guidance for Industry 2014). The type of aggregation and quantity of aggregates that my illicit an immune response or lead to reduced clinical benefit is protein dependent (Guidance for Industry 2014). Through the formulation development process, scientists not only try to attain an understanding of the type of aggregation the protein is likely to go through but also need to understand the threshold at which aggregates affect clinical results for that particular protein. Based on this understanding, scientists try to identify excipients that help stabilize the protein and reduce aggregation. However, this could be a long, arduous path in which platform formulations, DOE, trial-and-error, and personal experience are necessary to identify the excipients and the concentrations in which they are effective. Furthermore, numerous types of experiments (e.g., size exclusion chromatography, dynamic light scattering) are necessary to identify aggregates and particle formations due to limitations of any one experimental method (Guidance for Industry 2014). Lastly, these experiments need to be conducted over various time points to monitor changes in aggregation and determine the shelf-life of the protein therapeutic. As industry trends lead to increased need for timely and cost-effective formulation development, the lack of process improvements in identifying excipients that could reduce aggregation will only further exacerbate the currently lengthy and expensive process.

2.2 Project Goals

The goal of this project is three-fold:

1. Identify new compounds to be tested computationally and experimentally for their functionality as excipients that would impact aggregation
2. Develop method to computationally model excipients against a target protein
3. Perform wet-lab experiments and compare results against computational outcomes

This project aims to be the first step in the development of a predictive computational model to help identify excipients that will reduce protein aggregation for commercial biologic formulation development at Amgen. The data and recommendations presented in this thesis will be used to develop a relationship between computational and experimental results through an iterative process. Ultimately a predictive computational model will be built and incorporated into Amgen's work flow for formulation development, particularly for proteins that have aggregation problems and require changes to the platform formulation.

2.3 Business Implications

There are five key foreseeable business implications associated with the development of a high-throughput predictive computational model to identify excipients for formulation development. However, the most important role of this project is to introduce and gain acceptance for the use of computational technologies in formulation development, a process that has not changed over the last few decades. There were numerous barriers to the introduction of computational modelling in formulation development including:

- The opportunity cost of investing money and people-resources towards a long-term goal like developing computational modelling compared to investing in the immediate needs for resources in drug development
- The high risk of investing in supporting research that does not seemingly directly impact current drugs in development and the bottom line
- Potential regulatory challenges associated with the utilization of computational data and new excipients
- Gaining buy-in from expert scientists who have developed their own knowledge base and processes and may feel discouraged about the benefits of computational modelling

This project helped to overcome these barriers by identifying immediate potential benefits of computational modelling to gain a deeper understanding of excipient-protein interaction. It also helped to demonstrate that advancement in atomic-level modelling capabilities and computational power will enable the development process to occur at a fast and cost-effective manner. Lastly, the incorporation of expertise from formulation scientists from the initial stages of development enables direct alignment with scientists' needs.

There are numerous unrealized benefits that the development of a predictive computational model may add to formulation development. (1) It allows for the identification of new excipients that could perform various functions like reducing aggregation, by screening through databases with thousands of compounds in a systematic way. (2) A key benefit of computational modelling is its ability to test compounds must faster than conducting wet-lab experiments. As an example, 50 compounds were tested in approximately two to three days computationally and over 2.5 weeks of work experimentally for this project. (3) Computational modelling also allows screening experiments to be more cost-effective. For this project, while there was $0 variable cost in computationally testing 50 compounds, it costs approximately $15,000 to purchase them from Sigma Aldrich for experimental evaluation. (4) Lastly, computational modelling will provide a greater understanding of the excipient-protein interaction at an atomistic level, enabling smarter formulation development process overtime.

This project may also have additional business implications since new compounds are tested for their role as excipients. It is particularly difficult to assess the cost-benefits of using new excipients given the relative uncertainty in their development process and the difficulty in directly quantifying its clinical benefits for the final protein therapeutic. As such, formulation development often revolves around excipients that have already been accepted by the FDA unless there is a clear need that is not fulfilled by those. Through this project, a new set of

compounds are introduced to the experimental design space to potentially play a role in future formulation development. The true value of these compounds will only be understood based on its specific use in protein therapeutic development.

2.4 Project Approach

This project has three distinct phases that are outlined in Figure 5 below. Each of the subsections in this chapter will explain the considerations and methodology for each of these stages.

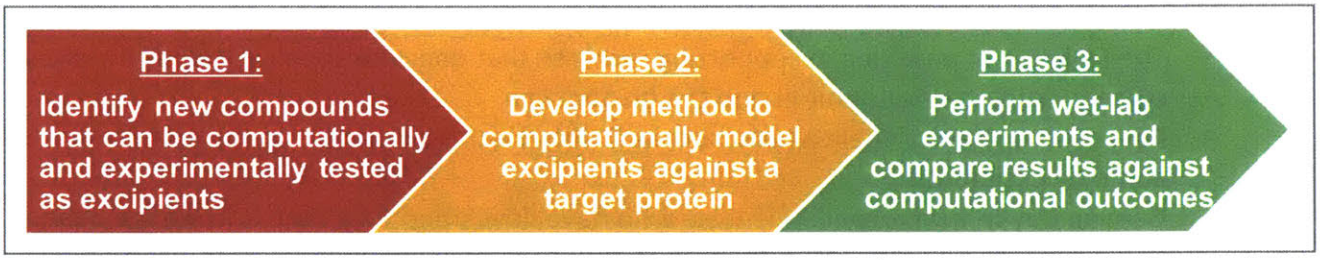


Figure 5: Overview of Project Approach

The primary hypothesis that will be tested is that excipients from Phase I that are computationally shown to have a non-specific interaction with the target antibody in Phase 2, are the ones that will likely reduce antibody aggregation. These excipients will likely interact with the different regions of the antibody, providing it a broad coverage and preventing the antibody from interacting with each other. The level of aggregation and protein stability will be the primary outcomes measured in Phase 3.

2.4.1 Methodology for Excipient Selection

This first phase of this project required the identification and selection of compounds that can be tested as excipients. In particular, since the goal of this project is to identify the impact of excipients on protein aggregation, compounds that could increase aggregation and compounds that could decrease aggregation were sought after. However, there are currently no clear attributes that help distinguish excipients based on their propensity to impact protein aggregation in solution. Furthermore, it is not obvious that excipients used in other liquid protein formulations to reduce aggregation would necessarily have the same properties with the target protein tested in this project.

So with that in mind, two different approaches were used to identify compounds. The first was to identify interesting compounds based on conversations with formulation scientists at Amgen and reviewing secondary literature for excipients that have demonstrated a reduction in aggregation. However, this resulted in a collection of molecules that had a variety of different attributes that gave the appearance of a random sample set. Thus, this set of molecules was not used because they did not constitute a systematic process and may potentially produce confounding results.

The second approach was based on identifying databases of compounds with specific properties or information. Table 4 below describes six major databases of compounds that were considered for this project and could be used in the future to identify new excipients for formulation. Compounds from Sigma Metabolites were computationally and experimentally studied for this project. Chapter 3 explains in more detail the process and rationale associated with excipient selection.

Table 4: Databases of Compounds Considered for Use as Excipients

Name	Number of Compounds (approximate)	Description
Chemical Entities of Biological Interest (ChEBI)	40,000	• Natural and synthetic "small" compounds • Molecules encoded by genomes (e.g., nucleic acids, proteins) are not included • Includes ontological classification • Provides SDF and SMILES format
FDA's Inactive Ingredients Database	3,200	• Excipients approved by the FDA • Information provided on route of administration, dosage form, and amount
PubChem	63,000,000	• Information on biological activities of small molecules • Provides 2D and 3D structures, SMILES, crystal structures • Vendor information available, if possible
Sigma Metabolites	250	• Metabolites and cofactors
The Human Metabolome Database	42,000	• Metabolites from the human body • Contains chemical, clinical, and biochemistry / molecular biology data • Provides SMILES structure
ZINC Database	35,000,000	• Commercially available • Only "biologically relevant" molecules and their representations • Built for virtual screening (especially docking) • Provides SMILES, mol2, SDF, pdbqt, and flexibase structure formats

(Sources: "PubChem Compounds", 2017; "Inactive Ingredients Database Download", 2017; "Metabolites and Cofactors on the Metabolic Pathways Chart", 2017; "Metabolites & Cofactors", 2017; Wishart, 2013; Hastings, 2013; Irwin, 2012; "Category: ZINC", 2015.)

2.4.2 Details on Target Antibody

It is important to note that although the focus of this project is on excipients, the target protein used to test these excipients significantly impacts the computational and experimental results. Anti-streptavidin antibody (ASA), shown in Figure 6, was used as the target protein based on its availability through Amgen for research and publication.

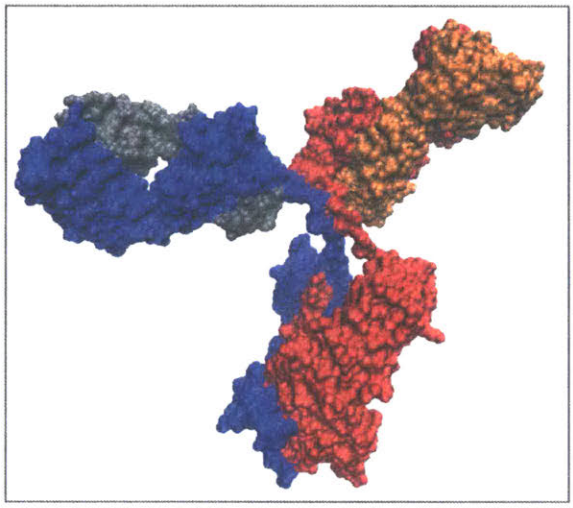


Figure 6: Anti-streptavidin Antibody

ASA is a monoclonal antibody that binds to streptavidin, a protein from the bacterium Streptomyces avidinii ("Anti-Streptavidin antibody [S3E11]", 2017). Streptavidin has one of the strongest non-covalent binding affinity to biotin, with a disassociation constant (Kd) of 10^{-14} moles/liter ("Anti-Streptavidin antibody [S3E11]", 2017). As such, it is commonly used in immunoassays against biotinylated proteins ("Anti-Streptavidin antibody [S3E11]", 2017). ASA binds with streptavidin without blocking its ability to bind with biotin ("Anti-Streptavidin antibody [S3E11]", 2017). ASA is an immunoglobulin G (IgG) molecule and commonly available as either IgG1 or IgG2 subclasses.

For this project mammalian anti-streptavidin IgG2 (referred to as ASA2) was used for computational modelling and wet-lab experiments. ASA2 has been shown to have higher aggregation compared to ASA1 under "physiological pH and mildly elevated temperatures" (Franey 2010). In particular, this study indicates that the IgG1 and IgG2 subclasses of ASA were incubated at 20 mg/mL concentration in 20mM sodium phosphate, 5% (w/v) sorbitol, at pH 7.0 and temperature of 45°C for up to 12 weeks (Franey 2010). The study identifies that the greater number of disulfide bonds in ASA2 leads to increased aggregation compared to ASA1 (Franey 2010). Based on the findings from the Franey study, ASA2 was used for this project. ASA2, given its higher propensity for aggregation, is more likely to show a greater range in aggregation across different excipients and controls than ASA1. Additional details regarding the computational model and experimental conditions under which ASA2 was studied are discussed in the sections below.

2.4.3 Methodology for Computational Modelling

2.4.3.1 AutoDock versus Molecular Dynamics Simulation

Phase 2 of this project was focused on developing a computational model to simulate the excipient-protein interaction. The goal of this phase is to not only gain a stronger scientific understanding through atomistic level modelling of excipients and ASA2, but also to use the results to hypothesize on outcomes from wet-lab experiments on aggregation. Currently there are no computational models that are widely accepted and used in formulation development for protein drug therapies. Thus there are two possible options that could be pursued. Under the first option, a completely new method to computationally model excipient-protein interaction is developed from scratch. However, this is not only out of the scope of this project due to resource and time constraints, but might not be the most effective starting point since understanding of excipient-protein interactions in solution is limited. The second option is to use existing molecular modelling tools to piece together a stronger understanding of excipient-protein interaction and through an iterative process, build a composite model that uniquely serves the needs of formulation development.

This project serves as the first step of the iterative process described above. AutoDock Vina and Molecular Dynamics (MD) Simulation were explored initially as options to model excipient protein-interaction. AutoDock Vina is a molecular docking tool that uses Amber force fields to calculate free energy of ligand-macromolecule binding at different sites of the macromolecule. It uses a scoring function that calculates the standard chemical potentials of the system using a machine learning approach and outputs a specified number of binding positions and binding affinities at those positions. Within a user-specified search space, AutoDock Vina attempts to find the global minimum free energy of binding. MD Simulation is a computational method to study the trajectory of atoms and molecules in a defined time period using Newtonian mechanics. Given a macromolecule and a specified concentration of ligands and buffer, MD simulations can model the trajectory of each of the molecules in the user provided time scale. Figure 7 below visually represents the output from AutoDock Vina and MD Simulation for lysozyme and a small molecule called niacinamide.

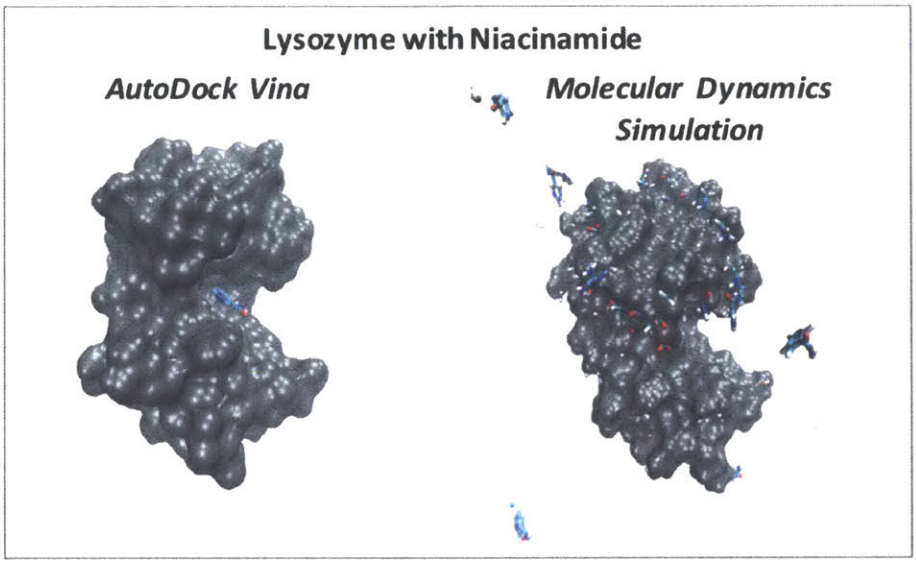


Figure 7: Visual Representation of Output from AutoDock Vina and Molecular Dynamics Simulation

Despite these interesting functionalities of both AutoDock Vina and MD Simulation, only one of these models could be pursued at this stage of computational development due to time constrains. Focusing on one model allows for effective use of resources and targeted computational and wet-lab experimental design. In order to select a tool to pursue further, both AutoDock Vina and MD Simulation were evaluated across five main criteria. These criteria are described below and in greater detail in Table 5.

1. **Input:** Files are quickly generated
2. **User Interface:** Easy to use
3. **Modelling Requirements**: Fast for high-throughput screening
4. **Modelling Complexity:** Provides additional insight into excipient-protein interaction
5. **Output**: Can be quickly analyzed to identify excipients for formulation

AutoDock Vina was pursued as the primary modelling tool in this project because it meets the initial criteria for a high-throughput screening tool to select excipients. Section 4.1 discusses in greater detail the reasoning for choosing AutoDock Vina over MD Simulation. Section 2.4.3.2 below goes through the detailed methods that were used to screen almost 250 excipients on AutoDock Vina.

Table 5: Overview of AutoDock Vina and MD Simulation

Key Categories	Category	AutoDock Vina	Molecular Dynamics Simulation
Input	Files required	PDBQT	Force fields: Non-bonded parameters and bonded parameters
	Ease of generation	Easy	Difficult (for high accuracy)
User Interface	Operating Systems	Windows, Linux, & Macs	Linux & Macs
	Coding Necessary	Yes	Yes
Modelling Requirements	Computational power	Low	High
	Time	Low	High
Modelling Complexity	pH	Change protonation	Change protonation
	Buffer	Implicit	Can add
	Protein surface conformation	No change	Changes
	Protein structural conformation	Manual iterations	Manual iterations
	Excipient concentration	One molecule	Match experimental values
Output	Binding affinity	Yes	No*
	Binding sites	Yes	No**
	Residence times	No	Yes
	Difficulty for analysis	Low	High
	*Although not a direct output, can be calculated **Not a direct output, can be calculated, but computationally very expensive		

2.4.3.2 AutoDock Input, Run Criteria, Output

AutoDock consists of a set of automated docking tools produced by The Molecular Graphics Laboratory at The Scripps Research Institute (Morris 2007). Five different versions of AutoDock

have been release since 1989/1990, with the latest versions being AutoDock 4 and AutoDock Vina (Morris 2007). Of these two, AutoDock Vina is used for this project since it improves accuracy of binding mode predictions compared to AutoDock 4, does not require the calculation of grid maps, and is approximately two orders of magnitude faster (see Figure 8). This enables a relatively easy and quick scan of the entire target antibody for potential binding spots.

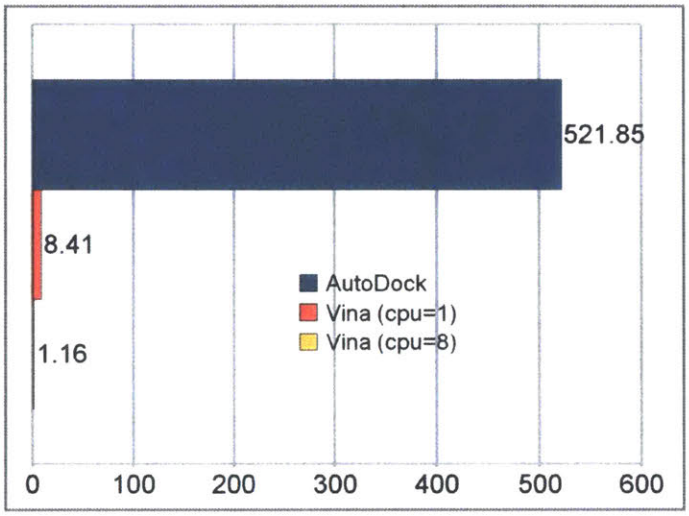


Figure 8: AutoDock versus AutoDock Vina (Average Time (Minutes) per Complex) (Trott, 2010)

There are three key inputs that are required to run AutoDock Vina. The first is a PDBQT (Protein Data Bank Partial Charge (Q) & Atom Type (T)) file of the macromolecule; the second is a set of PDBQT files for the excipients; and third is the input criteria to run the simulation on AutoDock Vina. As mentioned in Section 2.4.2, ASA2 was selected as the macromolecule for AutoDock Vina simulations. The structure of ASA2 was provided for this project by Clark et al. (2013). However, the structure did not include the sugars in-between the Fc regions and were added from the intact crystal structure of 1 HZH (Saphire 2001). Thus the entire structure of ASA2 contained approximately 29,194 atoms, 1,314 amino acid residues, and 16 sugars. ASA2's overall structural variability in different conformation was also considered for this project. As shown in Figure 9 below, ASA2 has the flexibility to take on many different structural conformations due to factors like the hinge region that connects the Fc segment to the Fab arms.

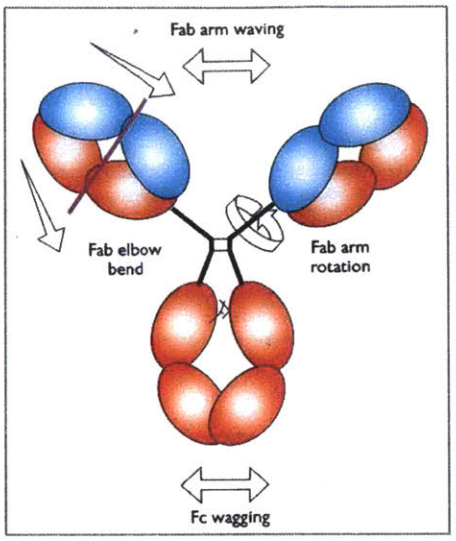


Figure 9: Areas of Flexibility for IgG

(Brekke, 1995)

Computational analysis of ASA2 using SASSIE produced 56,511 distinct conformations (Clark, 2013). The structural conformations taken on in solution depends highly on factors like the pH, temperature, and other compounds in solution. Thus, the part of the antibody that is exposed for interactions with excipients can be highly variable and can strongly influence the excipient-protein relationship. AutoDock Vina keeps the macromolecule, ASA2, static and thus does not allow for dynamic variations in structural conformations. Three different structures of ASA2 were used based on findings from Clark 2013 and analysis of probability of occurrence in pH 5.2 buffer. Figure 10 below shows the final three conformations that were used in this study. AutoDock Vina was run three times for each excipient-conformation combination, resulting in a total of 9 AutoDock Vina runs for every excipient.

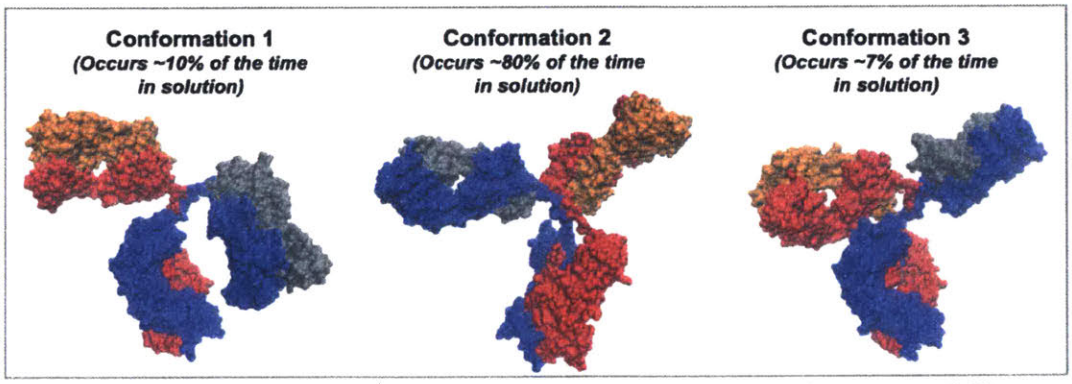


Figure 10: Final Conformations of ASA2 Used for AutoDock Vina Modelling

The structure for ASA2 was provided in the Protein Data Bank (PDB) format. At a high level, PDB files contains information on coordinates of each atom in a molecule and their connectivity. There are six different line items of information, or records, that can be provided for the PDB file. Table 6 below describes the information that is required for each record type.

Table 6: PDB Format: Information Necessary for Different Record Type

Selected Protein Data Bank Record Types	
Record Type	**Data Provided by Record**
ATOM	atomic coordinate record containing the X,Y,Z orthogonal Å coordinates for atoms in standard residues (amino acids and nucleic acids).
HETATM	atomic coordinate record containing the X,Y,Z orthogonal Å coordinates for atoms in nonstandard residues. Nonstandard residues include inhibitors, cofactors, ions, and solvent. The only functional difference from ATOM records is that HETATM residues are by default not connected to other residues. Note that water residues should be in HETATM records.
TER	indicates the end of a chain of residues. For example, a hemoglobin molecule consists of four subunit chains that are not connected. TER indicates the end of a chain and prevents the display of a connection to the next chain.
HELIX	indicates the location and type (right-handed alpha, *etc.*) of helices. One record per helix.
SHEET	indicates the location, sense (anti-parallel, *etc.*) and registration with respect to the previous strand in the sheet (if any) of each strand in the model. One record per strand.
SSBOND	defines disulfide bond linkages between cysteine residues.

("Introduction to Protein Data Bank Format", 2014)

Figure 11 below is an example of a PDB format and contains a sample of records for hemoglobin. The first item of the record indicates the type. The second item (column) indicates the atom serial number, which is typically the line number of the record. The third item indicates the atom and the fourth item indicates the residue in which the atom is found. The fifth item indicates the chain in which the residue is located. Hemoglobin contains two alpha chains and two beta chains. The entries shown in Figure 11 are all from one of the two alpha chains in hemoglobin. The sixth item is the residue number within the chain. Items seven through nine provides the xyz coordinates of the record. Item ten is the occupancy; item eleven is the temperature factor; and lastly item twelve is the element symbol.

Type	Atom Serial #	Atom	Residue	Chain	X	Y	Z	Occu-pancy	Temp-erature Factor	Element Symbol
ATOM	1058	N	ARG	A 141	-6.466	12.036	-10.348	7.00	19.11	N
ATOM	1059	CA	ARG	A 141	-7.922	12.248	-10.253	6.00	26.80	C
ATOM	1060	C	ARG	A 141	-8.119	13.499	-9.393	6.00	28.93	C
ATOM	1061	O	ARG	A 141	-7.112	13.967	-8.853	8.00	28.68	O
ATOM	1062	CB	ARG	A 141	-8.639	11.005	-9.687	6.00	24.11	C
ATOM	1063	CG	ARG	A 141	-8.153	10.551	-8.308	6.00	19.20	C
ATOM	1064	CD	ARG	A 141	-8.914	9.319	-7.796	6.00	21.53	C
ATOM	1065	NE	ARG	A 141	-8.517	9.076	-6.403	7.00	20.93	N
ATOM	1066	CZ	ARG	A 141	-9.142	8.234	-5.593	6.00	23.56	C
ATOM	1067	NH1	ARG	A 141	-10.150	7.487	-6.019	7.00	19.04	N
ATOM	1068	NH2	ARG	A 141	-8.725	8.129	-4.343	7.00	25.11	N
ATOM	1069	OXT	ARG	A 141	-9.233	14.024	-9.296	8.00	40.35	O
TER	1070		ARG	A 141						
HETATM	1071	FE	HEM	A1	8.128	7.371	-15.022	24.00	16.74	FE
HETATM	1072	CHA	HEM	A1	8.617	7.879	-18.361	6.00	17.74	C
HETATM	1073	CHB	HEM	A1	10.356	10.005	-14.319	6.00	18.92	C
HETATM	1074	CHC	HEM	A1	8.307	6.456	-11.669	6.00	11.00	C
HETATM	1075	CHD	HEM	A1	6.928	4.145	-15.725	6.00	13.25	C

Figure 11: Sample PDB File for Hemoglobin

("Introduction to Protein Data Bank Format", 2014)

The PDB files for each of three ASA2 conformations were converted to PDBQT format to be compatible with AutoDock Vina. In addition to the coordinates and connectivity information in the PDB files, PDBQT files contain partial charges and AutoDock atom types for each atom (Morris, 2007). PDBQT files also have united atom representation (Morris, 2007). This results in all non-polar hydrogens being combined with heavy atoms and only polar hydrogens are explicitly shown (Morris, 2007). PDB files can be converted to PDBQT files using the prepare_receptor4.py python script provided by AutoDock (Huey, 2010). The PDBQT files of ASA2 were directly fed into AutoDock Vina without any additional modification.

In total, 247 excipients were modelled against ASA2 using AutoDock Vina. These excipients are listed in Exhibit A of the Appendix. They range in molecular weight from 60 g/mol to 850 g/mol and vary in the type of compounds from amino acids to sugars, lipids, and nucleotides. Simplified molecular-input line-entry system (SMILES) structure for each excipient was

downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/), ChEBI (https://www.ebi.ac.uk/chebi/init.do), or ZINC (http://zinc.docking.org/). Open Babel was used to convert the SMILES structures to PDB files and python script, prepare_ligand4.py, provided by AutoDock was used to convert the PDB files to PDBQT files (O'Boyle 2011, Huey 2007).

Once prepared, the PDBQT files of ASA2 and excipients were fed into AutoDock Vina along with few other specifications on the run criteria. The pH of ASA2 and the excipients were included in the protonation states of these molecules at pH 7. In addition to this, the center and xyz dimensions of the search space was provided based on adding 10Å to each dimension of ASA2's size. The maximum number of binding modes that was specified was 30; however, AutoDock Vina only produced 20 binding modes per excipient-protein combination likely due to either finding only that many interesting binding modes or because it was limited by the energy range. Although no additional variables were specified, AutoDock Vina has options to modify other variables including varying the exhaustiveness of the search, providing a specific energy range, and specifying an explicit random seed "AutoDock Vina Manual" 2010".

AutoDock Vina produced two different output files. One of them was a log file that contained the binding affinities at each of the 20 different binding positions. The other was a PDBQT file with the coordinates of the binding positions. Thus, for each excipient-protein combination, it was possible to visualize the 20 different binding positions using Visual Molecular Dynamics (VMD) or other molecular graphics systems. Some of the excipients tested were visualized in this manner to gain a qualitative insight into the range of binding sites. However, in order to develop a quantitative and systematic way to analyze the results for 247 excipients, each of which was modelled 9 times, the following methods were used.

Binding Affinity: The standard error for calculating binding affinity in AutoDock Vina is around 2.85 kcal/mol (Trott 2010). On average, the range of difference in binding affinities across all of the runs for each excipient was around -2.07 kcal/mol. Only 35 excipients had a range greater than 2.85 kcal/mol. Thus it is assumed that the 20 different binding positions identified are more or less equivalently favorable. Therefore, at first the average binding affinity was calculated across the 3 runs for each of the 3 conformations (60 binding affinities / conformation). Then a weighted average was calculated across the three different conformations based on the probability of occurrence as described in Figure 10. This resulted in each excipient having one value for binding affinity that described its overall attraction to ASA2.

Footprint: Footprint was a new metric that was derived to understand the degree to which each excipient covered the surface of ASA2. All 60 binding positions for each conformation of ASA2 were superimposed for every excipient. The area of of ASA2 that is not interacting with the metabolite were subtracted away. The remaining area represented the excipient's coverage on that conformation. This value was divided by the total surface area of ASA2 in that conformation to understand the footprint or the proportion of the antibody that are covered by these 60 positions. Then a weighted average of the footprints for the three different conformations based on their probability of occurrence (as described in Figure 10) was calculated. Similar to the calculation for binding affinity, this resulted in one final value for each

excipient. This footprint value provides an idea of the proportion of ASA2 that is covered by each excipient.

Spread: One key hypothesis of this project, described in greater detail in Section 4.3, is that an excipient that provides more coverage of ASA2 is more likely to reduce aggregation because it prevents antibody-to-antibody interaction. Thus, in addition to footprint, another quantitative metric was necessary to understand if the excipient is only interacting with a certain component of ASA2 or if it is interacting with the different components. Spread was calculated by first calculating the center of mass for each binding position. The average center of mass (μ) for each conformation was calculated across the 60 different binding positions. The variance of the center of mass for each conformation was calculated using the following formula to derive the spread per conformation: $VAR = \Sigma(xi - \mu)^2$. Lastly the weighted average of the spreads for the three different conformations of the antibody were used to calculate one final spread value for each excipient. A larger spread indicates that the excipient is attracted to and covers different parts of the antibody. A smaller spread indicates that the excipient likely has more localized interactions with the antibody.

The average binding affinity, spread, and footprint were plotted on graphs and to the extent possible a variety of excipients were selected from different quadrants of the graph for wet-lab experiments. Excipient selection was constrained by two main factors: (1) excipient has to be safe to handle during wet-lab experiments and (2) excipients should have a molecular weight less than 500kD to avoid solubility issues. A more detailed analysis of the excipient selection process and a final list of excipients selected are described in Section 4.4.

2.4.4 Methodology for Wet-lab Experiments

2.4.4.1 ASA2 Preparation

Purified ASA2 from mammalian cell culture was provided by Amgen in 20mM sodium acetate solution at a concentration of 30mg/ml. Approximately 12ml of the ASA2 stock was dialyzed into a pH 5.2 10mM glutamic acid+10mM sodium phosphate buffer, using a Pierce Slide-A-Lyzer Dialysis Cassette at a million fold dilution, including three buffer changes, at 4°C. This process was repeated to get another 12ml of the ASA2 stock into a pH 7.5 10mM glutamic acid+10mM sodium phosphate buffer and another 12ml into a pH 3.5 10mM glutamic acid+10mM sodium phosphate buffer. The concentration of the dialyzes samples were measured using a spectrophotometer, with absorbance at 280nm and an extinction coefficient, ε, of 1.6 (mL/(mg*cm)) (Clark 2013). The concentrations of the ASA2 samples after dialysis and the final concentrations used to set up the first run and second run of the wet-lab experiments are listed in Table 7. The post-dialysis samples were diluted using their respective pH buffers to achieve a target concentration 10mg/mL for experimentation.

Table 7: Concentrations of ASA2 Samples

pH	Post-Dialysis Concentration	First Run Concentrations	Second Run Concentrations
3.5	20.84 mg/mL	10.43 mg/mL	9.89 mg/mL
5.2	18.76 mg/mL	9.81 mg/mL	10.33 mg/mL
7.5	21.1 mg/mL	10.13 mg/mL	10.43 mg/mL

2.4.4.2 Excipient Sample Set-up

The final list of excipients tested experimentally and the high and low concentrations used are shown in Table 8 below. Excipient preparation occurred through the following steps:

- Solubility of each of the 32 excipients was determined based on secondary sources
- High concentration stock solutions were 80% of maximum molarity based on solubility limits for each excipient
    - 125 μL of high concentration stock solution necessary per excipient for one run of all three experiments (DLS, DSF, and SEC)
    - Eight times this volume (1mL) was produced approximately
    - The grams of excipients necessary was based on the following formula: Molarity at maximum solubility * 80% * Molecular Weight*(1*10^-3L)
    - Each excipient was weighed out into 2.0mL Eppendorf tubes and dissolved in approximately 1mL of pH 5.2 10mM glutamic acid+10mM sodium phosphate buffer
    - Each sample was titrated to pH 5.2 approximately, if necessary, using 5N HCl, 1N HCl, 10N NaOH, and 1N NaOH
    - The approximate final high concentration used in experiments (and shown in Table 8) is half the molarity of the stock concentration since 125 μL of this sample was added to 125 μL of ~10 mg/mL ASA2
- Low concentration stock solutions are at 20% of the molarity of the high concentration stock solutions
    - 125 μL of low concentration stock solution necessary per excipient for one run of all three experiments (DLS, DSF, and SEC)
    - 25 μL of high concentration stock solution was added to 100 μL of pH 5.2 10mM glutamic acid+10mM sodium phosphate buffer
    - 125 μL of ASA2 sample was added to 125 μL of the low concentration excipient solution, effectively cutting in final concentration in half (as shown in Table 8)
- Stock solutions for pH3.5 and pH7.5 were created from the pH 5.2 high concentration stock solution for each excipient
    - Approximately 300μL of pH 5.2 stock solution was titrated pH 3.5 and another 300μL was titrated to pH 7.5
    - 125 μL of the titrated samples were used for the high concentration experiments for each pH and 25 μL was used for the low concentration and diluted by 100 μL buffer at each pH

Table 8: Concentrations of Excipients Used in Wet-lab Experiments

Excipients	High Concentration		Low Concentration	
	Molarity	Stoichio-metric Ratio	Molarity	Stoichio-metric Ratio
b-Alanine	2.3E+00	69279	4.6E-01	13856
DL-b-Aminoisobutyric acid	1.4E+00	40816	2.7E-01	8163
Oxaloacetic acid	2.8E-01	8492	5.7E-02	1698
L-(-)-Malic acid	2.9E-01	8734	5.8E-02	1747
D-(+)-Glucose	3.0E-01	8914	5.9E-02	1783
O-Phosphorylethanolamine	2.9E-01	8640	5.8E-02	1728
α-Ketoglutaric acid disodium salt hydrate	2.2E-01	6515	4.3E-02	1303
L-Serine	2.1E-01	6220	4.1E-02	1244
Betaine	1.9E-01	5695	3.8E-02	1139
L-Threonine	2.2E-01	6634	4.4E-02	1327
Sodium succinate dibasic hexahydrate	1.5E-01	4414	2.9E-02	883
D-(+)-Cellobiose	1.5E-01	4471	3.0E-02	894
Adonitol ≥99%	1.3E-01	3892	2.6E-02	778
Adenosine 5'-monophosphate disodium salt	1.1E-01	3320	2.2E-02	664
L-Arginine	1.1E-01	3368	2.2E-02	674
myo-Inositol ≥99%	1.2E-01	3459	2.3E-02	692
Thymidine 5′-monophosphate disodium salt hydrate	1.1E-01	3184	2.1E-02	637
L-Carnitine hydrochloride	1.0E-01	3086	2.1E-02	617
D-Glucuronic acid sodium salt monohydrate	8.2E-02	2472	1.6E-02	494
L-Leucine	6.7E-02	2010	1.3E-02	402
L-Glutathione	6.1E-02	1830	1.2E-02	366
Sodium phenylpyruvate powder	5.9E-02	1778	1.2E-02	356
Adenosine 5'-diphosphate sodium salt	4.8E-02	1450	9.7E-03	290
L-Tryptophan	1.8E-02	527	3.5E-03	105
Phospho(enol)pyruvic acid trisodium salt hydrate	2.2E-02	657	4.4E-03	131
Homogentisic acid	1.8E-02	538	3.6E-03	108
Glycocyamine	6.3E-03	189	1.3E-03	38
Orotic acid	5.1E-03	152	1.0E-03	30
Indole-3-acetic acid sodium salt	8.2E-03	246	1.6E-03	49
S-(5′-Adenosyl)-L-homocysteine crystalline	1.1E-03	32	2.2E-04	6

L-Tyrosine	2.2E-04	7	4.5E-05	1
Xanthine	4.9E-05	1	9.7E-06	0.3

2.4.4.3 Master Plate Set-up

"Master" plates were set up with all of the formulations once stock solutions were prepared for ASA2 and the excipients at the appropriate pHs. Specifically, three 96-well plates were created for each run, with each plate containing the high and low concentrations of the 32 excipient-protein samples in that pH and four controls of just ASA2 in 10mM glutamic acid+10mM sodium phosphate buffer at the relevant pH. Preparation of master plates involved pipetting 125 μL of high or low concentration of excipients into a well, adding 125 μL of ASA2 at the appropriate pH, and pipetting a few times to mix the sample. The 96-well plate design is shown in Figure 12 below. This design was replicated for each of the 3 pHs and across the first and second run. Samples for the three wet-lab experiments were taken out of this plate and replaced back, where relevant.

	1	2	3	4	5	6	7	8	9	10	11	12
A	b-Alanine	DL-b-Aminoisobutyric acid	Oxaloacetic acid	L-(-)-Malic acid	D-(+)-Glucose	O-Phosphorylethanolamine	α-Ketoglutaric acid disodium salt hydrate	L-Serine	Betaine	L-Threonine	Sodium succinate dibasic hexahydrate	D-(+)-Cellobiose
B	Adonitol		Adenosine 5'-monophosphate disodium salt	L-Arginine	myo-Inositol	Thymidine 5'-monophosphate disodium salt hydrate	L-Carnitine hydrochloride	D-Glucuronic acid sodium salt monohydrate	L-Leucine		L-Glutathione reduced	Sodium phenylpyruvate powder
C	Adenosine 5'-diphosphate sodium salt			L-Tryptophan		Phospho(enol)pyruvic acid trisodium salt hydrate		Homogentisic acid crystalline	Glycocyamine	Orotic acid		Indole-3-acetic acid sodium salt
D	ASA2 - Control	ASA2 - Control	S-(5'-Adenosyl)-L-homocysteine crystalline	L-Tyrosine		Xanthine						
E	b-Alanine	DL-b-Aminoisobutyric acid	Oxaloacetic acid	L-(-)-Malic acid	D-(+)-Glucose	O-Phosphorylethanolamine	α-Ketoglutaric acid disodium salt hydrate	L-Serine	Betaine	L-Threonine	Sodium succinate dibasic hexahydrate	D-(+)-Cellobiose
F	Adonitol		Adenosine 5'-monophosphate disodium salt	L-Arginine	myo-Inositol	Thymidine 5'-monophosphate disodium salt hydrate	L-Carnitine hydrochloride	D-Glucuronic acid sodium salt monohydrate	L-Leucine		L-Glutathione reduced	Sodium phenylpyruvate powder
G	Adenosine 5'-diphosphate sodium salt			L-Tryptophan		Phospho(enol)pyruvic acid trisodium salt hydrate		Homogentisic acid crystalline	Glycocyamine	Orotic acid		Indole-3-acetic acid sodium salt
H	ASA2 - Control	ASA2 - Control	S-(5'-Adenosyl)-L-homocysteine crystalline	L-Tyrosine		Xanthine						

High Concentration
Low Concentration

Figure 12: Wet-lab Experiments Plate Design

2.4.4.4 Size Exclusion Chromatography

High-throughput size exclusion chromatography (SEC) was used to determine the proportion of excipient-protein sample that is composed of soluble high molecular weight components (i.e., aggregates). Through this method, the larger aggregates elute from the column first, followed by monomeric protein, and lastly any fragments or excess metabolites. Waters Acquity H-class UPLC system at the Amgen facility was used to run this experiment in a high-throughput manner. Approximately 60 μL from master plates were pipetted into new 96-well plates for each sample, including the controls, across each concentration and pH, for each run. Each sample is automatically collected from the 96-well plate at a rate of 0.4 mL/min and a runtime of minutes. About 10 μg of each sample were run through a gel filtration column and eluted

with 100mM sodium phosphate buffer at pH 6.76. An UV diode array detector was used to capture the absorbance of the sample as it elutes off the column at 280nm. Empower, the software, was used to collect the results and identify the concentration of different components using Integration under the absorbance curve The graphs shown in Figure 13 below show the absorbance curve for ASA2 – Control sample, pH 3.5, at T0 and 1 week. The graphs have also been integrated and show the % of HMW (high molecular weight), % of Main (ASA2 monomers), and % of LMW (fragments and/or excipients) components of the solution. For this particular sample, an increase in aggregation is observed after 1 week. For the first run, T0, 1 week, and 1-month time points were collected. T0 and 1-month time points were collected for the second run. In between the runs, any remaining samples were collected and placed back into the master plates and incubated in 40°C.

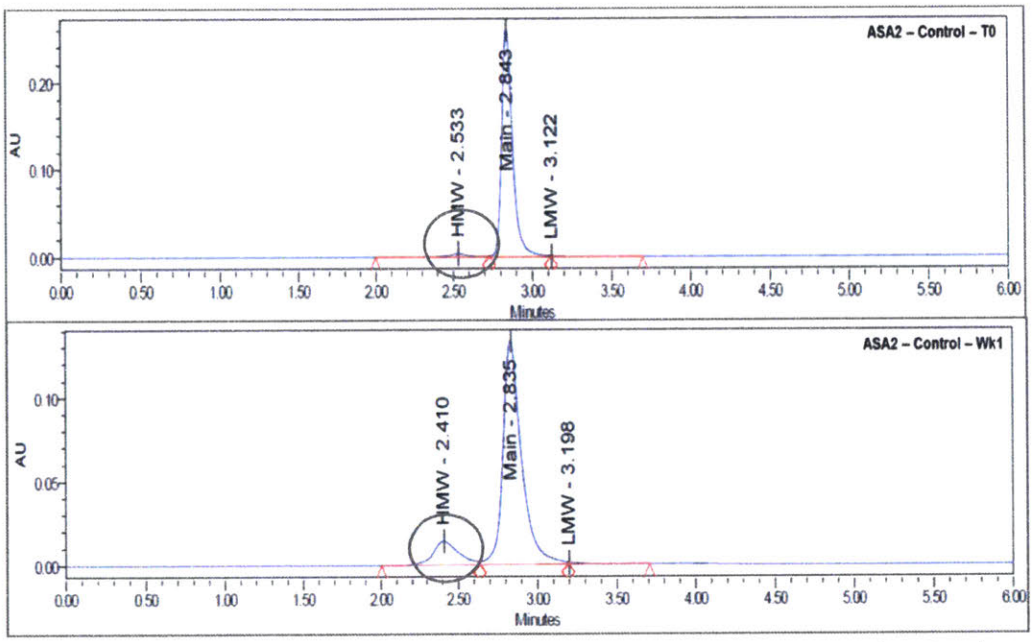


Figure 13: SEC Output and Analysis for ASA2 in pH 3.5 at T0 and 1 Week

2.4.4.5 Dynamic Light Scattering

Dynamic Light Scattering (DLS) is used to determine the distribution of particle size in solution. DynaPro Plate Reader II with Touch from Wyatt Technology was used to read excipient-protein samples in 96-well plates that replicated the design of the master plates. Approximately 60 μL of each formulation was transferred from the master plate to a black bottom imaging plate. The samples were covered and plates were centrifuged for 2 minutes at 1000 RPM to get rid of air bubbles. The 96-well plates were placed into the Plate Reader (one at a time) and data was collected after 15-minutes of equilibration. Light from lasers is shined through the particles in suspension and the resulting scattering of light is captured to understand various properties of the sample, including the mass distribution of particles in sample and the size distribution. While DLS results can depict the level of aggregation in a sample, it was used in this project to

identify any changes in the radius of ASA2 monomer. Figure 14, shown below, depicts changes in radius with serine and alanine compared to control. The changes in the radius of the monomer could be an interesting metric to analyze the stability of the monomer and potentially connect it with aggregation.

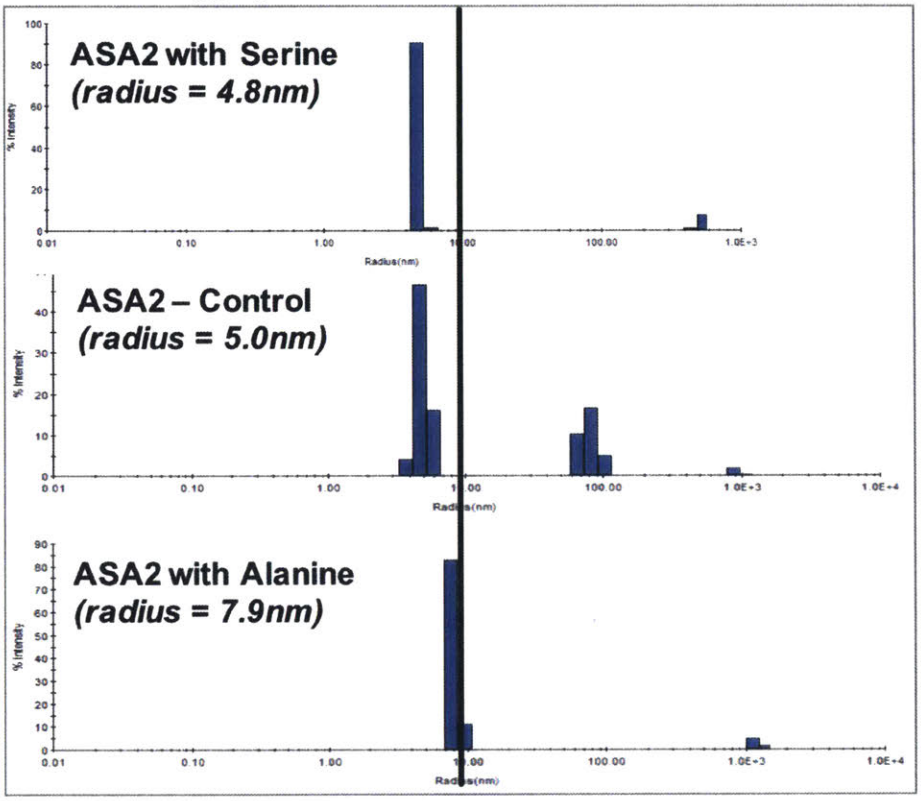


Figure 14: Sample DLS Results and Analysis

2.4.4.6 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) is a technique used to determine the melting temperature (Tm) of a protein. Approximately 19 μL of each excipient-protein sample was transferred from the master plate to a similarly designed 96-well plate. SYPRO Orange, a fluorescent dye that binds to the hydrophobic areas often in the core of the antibody, was added to the sample. About 2 μL of SYPRO Orange from stock was added to 10mL of 10mM glutamic acid buffer at pH 4.8 to dilute the stock solution. Approximately 1 μL from this diluted solution was added to each sample in the 96-well plate. The final plate was placed in a Bio-Rad CFX96 thermal cycler platform instrument. As the samples were heated slowly over time, the protein started to unfold at different rates and release the dye. The rate of release of the SYPRO Orange dye over time is used to determine Tm in the CFX Manager software.

More specifically, the release of the dye in each sample was captured as Relative Fluorescence Units (RFU) over temperature. The first derivative of the RFU versus temperature curve, -d(RFU)/dT, shows the rate of fluorescence release. The minimum in the –d(RFU)/dT curve shows the melting temperature. The graph in Figure 14 depicts the results from the CFX

Manager software for ASA2-Control and two different formulations. Excipient that stabilizes ASA2 led to an increase in Tm, like glycosamine shown in Figure 14. T0 was collected for the first run. The 1-month time point was collected for both the first and the second run. The samples were discarded after each run since ASA2 was completely denatured after the heating process.

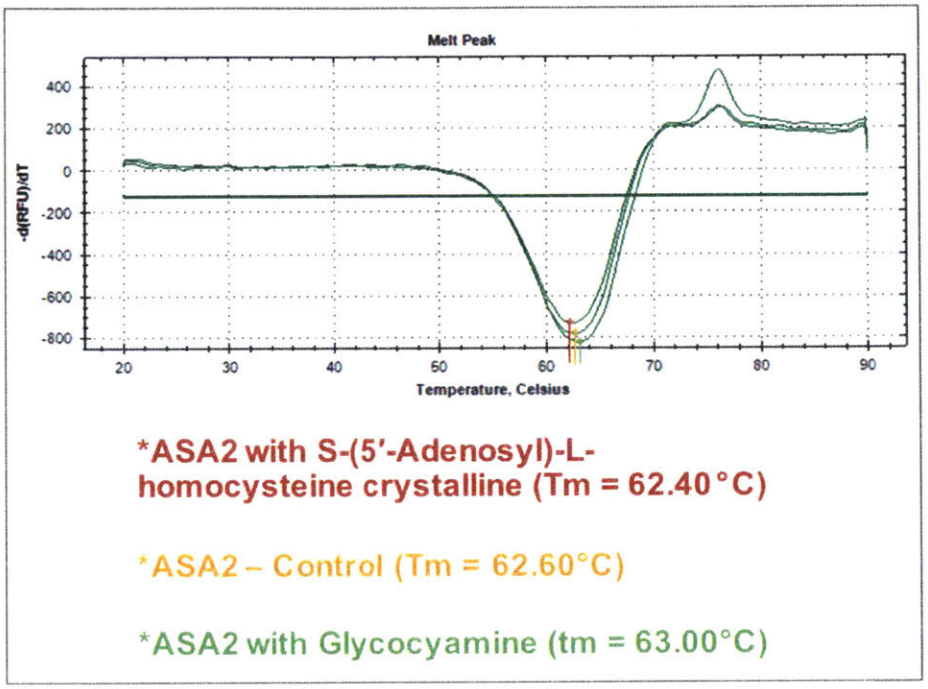


Figure 15: Sample DSF Results and Analysis

This page is intentionally left blank

3. Excipient Selection

As described in Section 2.4.1, six different databases of compounds were considered for their use as excipients in formulation development. These databases are: (1) Chemical Entities of Biological Interest (ChEBI), (2) FDA's Inactive Ingredients Database (IID), (3) PubChem, (4) Sigma Metabolites, (5) The Human Metabolome Database, and (6) ZINC Database. For the purpose of this project, the FDA's IID and Sigma Metabolites were the top two choices and ultimately compounds from Sigma Metabolites were tested computationally and experimentally. The sections below discuss in detail the advantages and disadvantages of each of these two databases and the rationale for choosing Sigma Metabolites.

While the other four databases contain a larger number of compounds and may present more interesting solutions for unique formulation challenges, they were out of scope for this project. Given that the primary goal of this project is to build a proof-of-concept model and through that process, achieve the secondary goal of identifying new excipients, the time, cost, and effort required to model these large datasets would not have a business justification to pursue at this time. However, once a robust model has been developed, it would be a worthwhile endeavor to computationally model the compounds in one or more of these databases so that they can be tested with new proteins.

3.1 Characteristics of Excipients Approved by FDA

FDA's Inactive Ingredients Database (IID) is one of the most promising databases to start with in any formulation development process for two main reasons:

1. Faster approval and fewer requirements by the FDA for excipients already in the database
2. Higher level of confidence that excipient will work, particularly for a similar formulation and protein modality

However, it is important to be cautious of the fact that biotech companies and specialized formulation development companies might have patents on some of the excipients in the database that restrict their usage to certain formulation and/or require heavy licensing fees. Furthermore, changes in modalities, concentration, and chemical structure of the excipient may warrant additional studies by the FDA. Thus it would still be important to be cautious about the use of compounds from this database. (FDA's IID: https://www.fda.gov/Drugs/InformationOnDrugs/ucm113978.htm)

This history and overview of the FDA's regulatory approach towards excipients is described in Section 1.3.1. FDA began to review excipients, identified primarily as inactive ingredients, in the regulatory approval pathway of new drugs after the Federal Food, Drug, and Cosmetic Act of 1938 (Davis 2006). Since then, the FDA has published Guidance Documents to provide more insight into its expectations for new excipients ("Search for FDA Guidance Documents" 2017). In addition to this, the FDA also publishes the Inactive Ingredients Database (IID) that contain most, if not all excipients that are present in approved drugs. If there are any excipients missing

from this database at all, it is likely because they are considered to be "active" ingredients that directly impact the therapeutic response of the drug.

An updated version of the IID is published by the tenth working day of every quarter (i.e., April, July, October, and January) ("Inactive Ingredients Database Download" 2017). The database can be downloaded and/or interactively used online to search for excipients. There are seven key variables defined for each excipient: (1) Name of the inactive ingredient, (2) Route, (3) Dosage Form, (4) CAS Number, (5) UNII, (6) Potency Amount, and (7) Potency Unit ("Inactive Ingredients Database Download" 2017). The names of the inactive ingredients are typically based on the submission by the first manufacturer; however, recently the names were changed to be consistent with the FDA's Substance Registration System (SRS) ("Inactive Ingredient Search..." 2017). Since the various synonyms of a particular compound are not present in the IID, the SRS can also be used to identify the preferred name ("Inactive Ingredient Search..." 2017). The route indicates the type of delivery for the drug (e.g., intravenous, oral) and the dosage form indicates its formulation (e.g., lyophilized powder, solution, tablet). The CAS Number is the Chemical Abstracts Service Number that is used by the American Chemical Society to register unique compounds and can be used to search the compounds in other database ("Inactive Ingredient Search..." 2017). UNII, Unique Ingredient Identifier, is similar to CAS Number in that it is unique alpha-numeric identifier provided to each new compound and used by the United States Pharmacopeia (USP) and the SRS to collect information on these compounds ("Inactive Ingredient Search..." 2017). The potency amount and unit describe the maximum amount excipient used in the approved drug product per dose. However, this does not indicate the maximum daily intake (MDI) of the excipient since patients may have to take multiple doses of a drug in a given day.

For this project, October 2015 IID was used for further analysis since it was the latest available update at the time of execution of the research in February 2016. A total of approximately 13,583 line-items were listed in this database. However, only 3,209 unique compounds were present based on ingredient names. The remaining 10,374 line items were repeats of the same compounds but in different route, dosage form, or potency amount. Of the 3,209 excipients on this list, 291 were used in intravenous or subcutaneous formulation. Since generally protein therapeutics are formulated for intravenous or subcutaneous delivery, this filter is the most time-effective way to identify excipients used in this formulation. Among these 291 compounds, 35 are already used in Amgen's commercialized study. Discussions with Amgen's senior scientists revealed that they have also had experience with an additional 173 compounds from the list. This leaves a total of 83 compounds (shown in Table 9) that are completely new to Amgen's formulation development team and should be considered in the future for computational and wet-lab experiments.

Table 9: Potential Compounds for Consideration as Formulation Excipients for Parenteral Administration

.beta.-cyclodextrin sulfobutyl ether sodium	gentisic acid	octreotide

bibapcitide	gentisic acid ethanolamide	oleic acid
brocrinat	gentisic acid ethanolamine	oxidronate disodium
calcium gluceptate	gluceptate sodium	pentasodium pentetate
calcobutrol	gluceptate sodium dihydrate	pentetate calcium trisodium
caldiamide sodium	gluconolactone	pentetic acid
caloxetate trisodium	glucuronic acid	perflutren
calteridol calcium	hetastarch	poloxamer 188
diacetylated monoglycerides	hexylresorcinol	polyglactin
diatrizoic acid	hydroxyethylpiperazine ethane sulfonic acid	povidone k12
dimethicone	hydroxypropyl cellulose (type h)	povidone k17
dimethyl sulfoxide	iobenguane	protamine sulfate
dipalmitoylphosphatidylglycerol, dl-	lactic acid, dl-	sodium ascorbate
disodium citrate sesquihydrate	lactic acid, l-	sodium benzoate
disodium sulfosalicylate	lactic acid, unspecified form	sodium bicarbonate
disofenin	lactobionic acid	sodium carbonate
distearoylphosphatidylcholine, dl-	lidofenin	sodium chlorate
egg phospholipids	magnesium stearate	sodium citrate, unspecified form
ethanolamine hydrochloride	maleic acid	sodium desoxycholate
ethylenediamine	mebrofenin	sodium gluconate
ethylene-vinyl acetate copolymer (15% vinyl acetate)	medronate disodium	sodium hypochlorite
ethylene-vinyl acetate copolymer (28% vinyl acetate)	medronic acid	sodium metabisulfite
ethylene-vinyl acetate copolymers	metaphosphoric acid	sodium trimetaphosphate
exametazime	methanesulfonic acid	triethyl citrate
fosveset	methylboronic acid	tromantadine
gadolinium oxide	n,n-dimethylacetamide	urethane
gamma cyclodextrin	niacinamide	versetamide
gelatin	nioxime	

While these compounds were strongly considered for computational modelling, they were not ultimately used in this project. These compounds are diverse and their impact on protein aggregation is not clear apriori. The only two unifying attributes of these compounds are that they are approved by the FDA as excipients in IV and SC formulations and are new to Amgen. Of the two options considered (compounds from the FDA's IID and metabolites from Sigma), the compounds from FDA's IID seem like it could produce results whose trends might be difficult to extrapolate.

3.2 Rationale for Selecting Metabolites

Metabolites from Sigma Aldrich were chosen for computational modelling and wet-lab experiments for this project. Metabolites present an interesting opportunity for formulation development. Endogenous metabolites in particular may have lower likelihood of triggering adverse events or immune reaction since these compounds are already found in the human body. This could also lead to fewer non-clinical and clinical studies to be required by the FDA for

excipient approval. Furthermore, metabolites in certain classes have already shown to have an impact on protein stability. For example, sugars like mannitol and fructose and amino acids like lysine and arginine have shown to reduce or prevent aggregation by lowering the thermodynamic activity of proteins (Basavaraj 2014). It is possible that exploration of metabolites in these classes or with similar structure could result in excipients that are even better at these function. However, it would be also important to be cautious about the impact of these metabolites on the body since variations in concentrations could result in adverse events.

As of May 2016, The Human Metabolome Database had 41,993 compounds (Wishart 2013). Of these, approximately 29,289 are endogenous and 3098 of these are actually detected (Wishart 2013). While this database serves as a repository for Amgen to explore in the future, metabolites from Sigma Aldrich were used for this project because it contained (1) a smaller sample set that was reasonable to model within the time constrains of the project, (2) metabolites from different classes including amino acids, carbohydrates, and lipids, (3) easy to purchase if selected for wet-lab experiments. Sigma Aldrich has approximately 301 unique metabolites that are available to be purchased. Of these approximately 80 are amino acid metabolites, 103 are carbohydrate metabolites, 76 are lipid / cholesterol metabolites, 40 are nucleotide metabolites, and 32 are metabolites from the TCA cycle or porphyrin metabolites (The sum of these add to greater than 301 since some metabolites are double counted in different classes) ("Metabolites & Cofactors", 2017). The complete list of metabolites available from Sigma for purchase (and modelled in this project) are provided in Exhibit A of the Appendix. Many of these compounds are already used as excipients in commercialized drugs and listed in the IID. (However, it is difficult to identify these very quickly because of small differences in naming conventions.) These compounds present a really interesting opportunity for this project because they provide diversity in chemical structure, size, and type of molecules that could potentially give a range of computational and experimental results. However, they are all metabolites, many of which have already been used as excipients, and can potentially be analyzed within classes to understand the nuances of excipient-protein interactions.

4. Computational Modelling

4.1 Rationale for Using AutoDock

As described in Section 3.2, metabolites from Sigma Aldrich was used for screening through computational modelling. There were three primary reasons for using these compounds: (1) they were readily available for purchase for wet-lab experiments; (2) metabolites, particularly ones that are endogenously found in the human body, are likely to be safer in clinical application; (3) many metabolites have shown to increase protein stability. The list of these compounds are in Exhibit A of the appendix. The next phase of this project involved developing a computational modelling technique to understand excipient-protein interaction and evaluating the compounds listed in Exhibit A to identify excipients for wet-lab experimentation. Computational modelling is currently not used for formulation development in the biotech industry. As such, there were no existing tools available for this particular need. This resulted in two options for the project – to either develop an innovative new method to model excipient-protein interaction or identify a novel application for existing modelling technologies. While an entirely new modelling methodology would provide numerous benefits, including having specific capabilities to address questions related to formulation development, it might not be the best first approach. It would not only require time and resources that are beyond the scope of this project but also would not effectively take advantage of the information derived from existing methods. As a result, two different currently available modelling methodologies were considered for this project – AutoDock Vina and molecular dynamics simulation. Section 2.4.3 describes in detail the inputs and outputs for each of these modelling systems.

The ultimate utility of a modelling tool to identify excipients for wet-lab experimentation comes from its ability to meet the following criteria:

1. Input files are quickly generated
2. Easy to use
3. Modelling is fast
4. Modelling provides additional insight into excipient-protein interaction
5. Output can be quickly analyzed to identify excipients for formulation

The most important advantage of molecular dynamics simulation is in its ability to provide atomistic level information on excipient-protein interaction. Residence time is the direct output of molecular dynamics simulation and provides information on the amount of time an excipient spends near the protein. But other variables can be indirectly calculated from this including identifying the regions of the protein the excipient most closely interacts with and the binding affinity at that region. In addition to this, concentrations of the excipient and the choice of buffer can be specified leading to a more accurate representation of experimental conditions. Lastly, although the structural conformation of the protein does not change in molecular dynamics simulation, its surface does dynamically change, leading to a more meaningful insight on excipient-protein interaction. However, despite these advantages related to point 4, it is important to note that MD simulations typically only cover short timescales due to the immense computational power they require. Thus it might be challenging to extrapolate the nature of excipient-protein interactions. Furthermore, the business justification to pursue MD

simulations at this point in development would be tough given its drawbacks in points 1-3 and 5. Developing accurate input files are extremely challenging as it requires developing force fields for both the excipients and the protein. Arguably once force fields are developed for excipients, they can be used in screening exercises for other proteins. However, the development of accurate force fields are difficult, extremely time consuming, and will likely require specialized resources.

AutoDock Vina was used in this project because of its ability to provide insight into an excipient's propensity to interact with different parts of a protein through fast modelling, relatively easy to develop input files, and easy to analyze output files. Although AutoDock Vina has been typically used for screening small molecule drug candidates against a biological target, it could potentially have a novel application in identifying excipients for formulation development. As described in Section 2.4.3, the rate limiting step with AutoDock Vina is likely the collection of SMILES structures for all of the excipients. However, once they are collected, they can quickly be converted into pdbqt files for AutoDock Vina and be used in the future for other proteins. It requires fairly minimal amount of other information, requires less computational power than MD simulations, and most importantly, it is an effective tool for high throughput screening. It also creates easy to analyze output files: binding affinities for the different locations of interactions and the coordinates for those interactions.

Theoretically AutoDock Vina met all of the five modelling needs listed above; however, a small computational test was performed to ensure that AutoDock Vina functions as expected. Lysozyme is an enzyme that breaks down bacterial cell walls by enabling hydrolysis between N-acetylglucosamine and N-acetylmuramic acid. The active pocket in lysozyme interacts more favorably with sugars compared to other small molecules. Knowing this, two different small molecules were tested against lysozyme. UDP-N-acetyl-alpha-D-muramate contains a sugar component and was hypothesized to primarily have docking positions in the active pocket. Methanesulfonic acid (MsOH) on the other hand was hypothesized to have non-specific interaction with lysozyme. The results form AutoDock Vina confirm the hypothesis on the expected behavior of MsOH and UDP-N-acetyl-alpha-D-muramuate. Based on this, AutoDock Vina was used to study ASA2 and approximately 250 Sigma metabolites.

4.2 AutoDock Output and Analysis

There are two main outputs from AutoDock: binding affinity and the coordinates of the binding site for 20 different positions. Section 2.4.3.2 describes the methods that were used to analyze binding affinity and to convert coordinates of binding sites into spread and footprint. Figure 17 below shows the results from AutoDock Vina for approximately 250 excipients against ASA2.

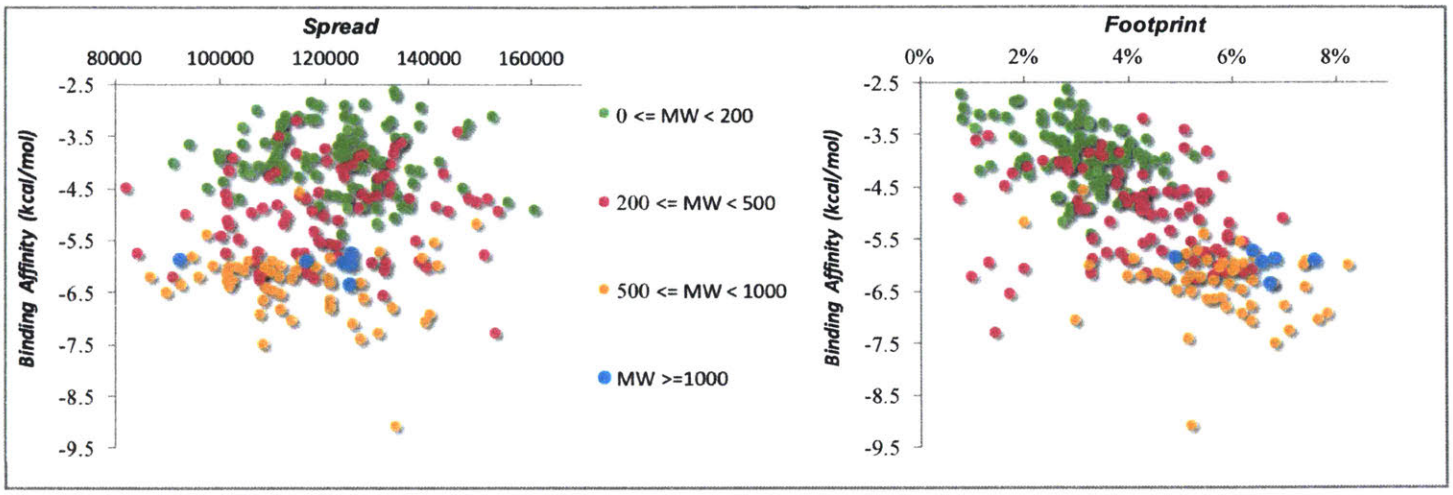


Figure 16: Analysis of AutoDock Vina Results for 247 Excipients

Generally, binding affinity and footprint increase with the molecular weight of the excipients. Molecular weight in this case can be used as a close approximation for the size of the molecule. The larger the excipient, the greater its binding affinity with ASA2. This is likely due to the greater number of non-covalent interactions it has with ASA2, including potentially a greater number of hydrogen bonds. The larger the excipient, the more likely it is to cover a larger surface area of ASA2, leading to a larger footprint. The molecular weight of the excipient does not seem to influence the spread, indicating that the size of the excipient does not influence the parts of ASA2 that interacts with.

4.3 Hypothesis on Impact of Selected Excipients on Protein

Using both quantitative analyses and visualization of ASA2's interaction with many of the excipients using VMD helped generate the primary hypothesis form AutoDock Vina modelling. Our hypothesis was that excipients with larger spread, high footprint, and high binding affinity are likely to reduce aggregation. These excipients are likely to reduce antibody-to-antibody interaction and therefore reduce aggregation.

While this hypothesis was used to guide excipient selection for wet-lab experiments, additional hypotheses may be generated from AutoDock Vina results. For example, if there are particular areas of the antibody that has a propensity for aggregation, then AutoDock Vina could be used to identify excipients that specifically interact with those areas.

4.4 Final Excipients Selected for Experiments

Excipients were chosen in order to experimentally test the validity of the hypothesis on the AutoDock Vina results. As such the excipients shown in Figure 19 were chosen to be from each of the four quadrants of the Spread versus Binding Affinity graph.

- Quadrant 1: High spread, low binding affinity
- Quadrant 2: Low spread, low binding affinity
- Quadrant 3: Low spread, high binding affinity
- Quadrant 4: High spread, high binding affinity

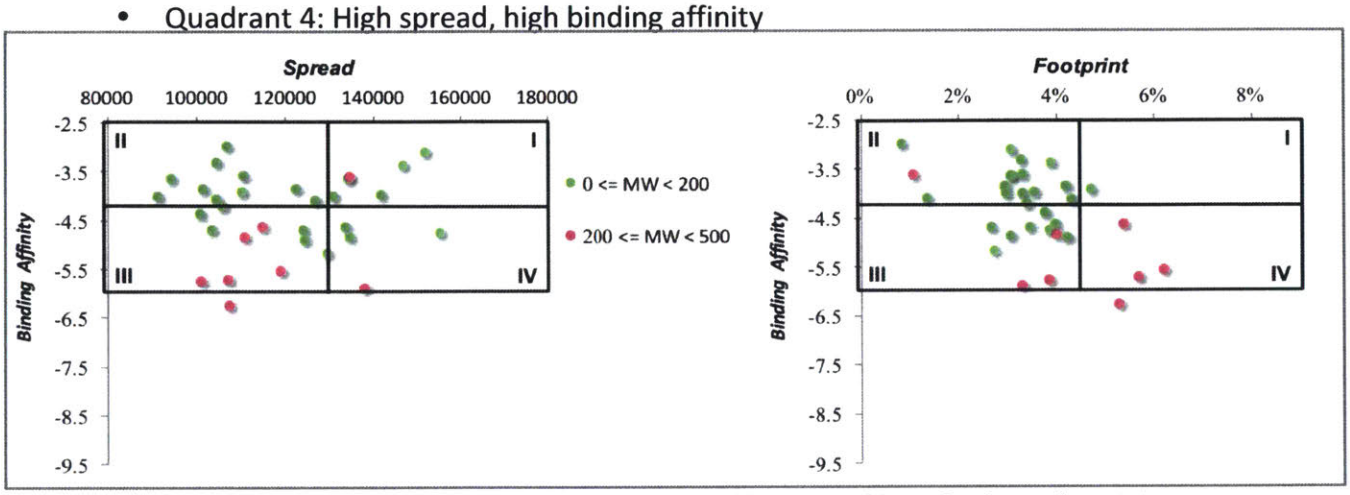


Figure 17: Spread, Footprint, and Binding Affinity Profiles of Selected Excipients

Excipients that were above 500 Daltons were eliminated from wet-lab experiments. This is roughly based on Lipinski's rule of five, of which one indicates that orally active small molecule drug should not be larger than 500 Daltons. But in general, the larger the excipient, the tougher it is to achieve necessary concentrations. It can also potentially impact the maximum possible concentration of antibody in solution. In addition to this, some excipients were eliminated because of the toxicity issues to conduct wet-lab experiments. Lastly, some excipients were chosen (e.g., arginine) because they have previously shown to have a propensity to reduce aggregation in proteins. A final list of excipients that were tested experimentally are shown in Table 8.

5. Wet-lab Experiments

5.1 Rationale for SEC, DLS, and DSF

Scientists currently rely on wet-lab experiments for formulation development. These experiments are used to evaluate various attributes that are important in clinical effectiveness and commercial viability of biologics. Some attributes like protein aggregation are important to evaluate across all protein therapeutics and experiments that test for aggregation are systematically performed for every drug. The degree to which some other wet-lab experiments are conducted is dependent on the needs of the protein. For example, experiments related to solubility or viscosity are more likely to be conducted for proteins that have issues related to it. Overall wet-lab experiments are conducted for two reasons:

1. To identify the right conditions that meet the goals at each stage of formulation development (i.e., pre-formulation, drug substance stabilization, and final formulation), including identifying the appropriate excipients
2. To gather data for clinical and commercial specifications (e.g., aggregation, stability, and shelf-life over lifecycle of the drug)

The goal of this project was to study the impact of 32 excipients on ASA2, particularly as it results to protein aggregation. There are numerous experiments that can be conducted to study protein aggregation. This project takes advantage of the high-throughput techniques available at Amgen to understand protein aggregation from multiple perspectives across thousands of design conditions. Section 2.4.4 explains in detail the methodologies used to conduct size exclusion chromatography (SEC), dynamic light scattering (DLS), and differential scanning fluorimetry (DSF). Table 10 below summarizes the design conditions for the experiments. This section will cover in more detail the rationale behind choosing these experiments and the key insights they add to understanding ASA2 aggregation. This section will also further explain some of rationale behind key formulation design choices, including the different concentrations of the excipients, incubation at 40C, and pH of solution.

Table 10: Overview of Design for Wet-lab Experiments

	Size Exclusion Chromatography (SEC)	Dynamic Light Scattering (DLS)	Differential Scanning Fluorimetry (DSF)
Measurement	Aggregation	Protein size	Melting Temperature
Total Number of Excipients	32	32	32
Temperature	40°C	40°C	40°C
Time Points	T0, 1 week, 1 month	T0, 1 month	T0, 1 month
pH	3.5, 5.2, 7.5	3.5, 5.2, 7.5	3.5, 5.2, 7.5
Concentrations	High, Low	High, Low	High, Low
Total Number of Experiments	576	384	384

Size exclusion chromatography (SEC) is a standard, reliable, and relatively easy-to-use experimental technique to identify the proportion of aggregates in a sample. A gel-based filtration column is used to pass samples through and separate its components by size through different flow rates. Smaller molecules (e.g., protein monomers, degraded protein components, or excess excipients) penetrate the pores of the beads in the gel and follow a longer path, leading to slower flow rates. Protein aggregates, on the other hand, are less likely to penetrate the pores and will likely flow directly through the column. High-throughput SEC was conducted at Amgen that enabled the screening of more than 1000 different formulation samples. In addition to separating the ASA2 aggregates from the monomers, a built-in UV detection also helped to quantitatively measure the proportion of the sample made up of these different components: high molecular weight, main (aka monomer), and low molecular weight. This output directly addressed the primary goal of the project – to understand the impact of each excipient on the level of ASA2 aggregation. In addition, once the samples were set up in 96-well plates, the high-throughput SEC set-up took approximately 1 hour to set-up and then it automatically ran the different samples (2 plates at a time) at a pace of approximately 6 minutes per plate. The output from SEC showed UV absorbance at 280nm at different time points. Area under the curve was integrated at the high molecular weight (HMW) peaks to identify the proportion of aggregates in the sample. However, there are two challenges related to SEC that are important to recognize. The first is that with a larger number of samples, the gel filtration column itself may be degraded over time and may impact the results. The second is that aggregates that are insoluble and settle to the bottom of the plate are not measured through this process.

Dynamic light scattering (DLS) is another commonly used technique in the analysis of protein aggregation. DLS provides information on the polydispersity of a sample and the hydrodynamic radius of its components. Similar to the SEC, DLS can be done in a high-throughput manner using the DynaPro Plate Reader by Wyatt Technology at Amgen. This allowed us to run almost 800 formulation samples at a pace of approximately 2 hours per 96-well plate (~2 minutes per sample). DLS captures intensity of scattered light from a sample over time and uses autocorrelation to provide outputs such as the hydrodynamic radius and polydispersity. While it is fast and easy to use, it is more sensitive to large particles and may produce less reliable results in a heterogeneous sample (Chaudhuri 2014). DLS can also be used to detect the hydrodynamic radius of the ASA2 monomer, giving insight into the level of compactness of the protein and by extension, to its stability. Excipients that enable the protein's native, more compact conformations, have been shown to reduce aggregation; while, excipients that lead to an expansion of the native state have been shown to increase aggregation (Kendrick 1998).

Differential Scanning Fluorimetry (DSF) provides information on the protein's thermostability (Chaudhuri 2014). SYPRO Orange, a fluorescent dye that preferentially binds to hydrophobic regions of a protein, was added to each ASA2 formulation, including the controls. Since the hydrophobic regions of ASA2 are typically located in the interior, like most other antibodies, the dye was not able to bind to these regions and its fluorescent signals were instead quenched by the aqueous buffer (Chaudhuri 2014). However, as the temperature of the sample rises over time, the protein begins to unravel and lose its secondary, tertiary, and quaternary structures. This exposes the hydrophobic regions and results in higher fluorescent readings as SYPRO Orange binds to these molecules. The curve produced by the change in fluorescence over increasing temperature can be used to find melting temperature (Tm) of the protein. Typically, more stable proteins are shown to have higher melting temperature and smaller aggregation rates. Likewise, structures of less stable molecules begin to unravel at lower temperatures and expose hydrophobic regions that have a propensity for aggregation in aqueous solution (Chaudhuri 2014). Since all of the ASA2 formulation samples were incubated at 40°C, DSF provides insight into excipients that reduce aggregation by increasing the thermostability of the protein.

In addition to determining the type of experiments to conduct, decisions were also made regarding experimental design factors based on feedback from Amgen's formulation scientists, MIT advisors, and secondary research. One of the first decisions that was made was on pH. This is an important decision point in traditional formulation development since pH plays a critical role in the stability of the protein. Changes in pH can change the protonation states of the amino acid residues in the protein, leading to the potential disruption of some of the hydrogen bonds and electrostatic interactions. This can then result in changes to the secondary, tertiary, and quaternary structures of ASA2, destabilization of the native conformation, and increase in aggregation. While in the case of formulation development, a pH that leads to increase in stability is identified, the goal for this project was to test pHs that destabilize the protein. ASA2 is a fairly stable antibody at pH 5.2 and one of the concerns of this project was that it might be difficult to observe the impact of excipients if the antibody is stable regardless. Therefore, ASA2

was also tested in pH 3.5 and pH 7.5 buffers. These two pHs were theoretically selected to help destabilize ASA2 and understand the impact of excipients in continuing to prevent aggregation.

As a result of testing each excipient-ASA2 formulation sample in three different pH conditions (pH 3.5, 5.2, 7.5), a decision also had to be made regarding the buffers for each pH. The goal of this experiment was to focus on the impact of only the excipients on the protein, to the extent possible. So it was necessary to have the same buffer across all of three pH samples. Based on feedback from an Amgen scientist, a 10mM glutamic acid+10mM sodium phosphate buffer was used. The benefit of using glutamic acid and sodium phosphate was that the pH of the buffer could be adjusted for the most part without using other acids and bases for titration. For the most part, pH was adjusted by varying the volume ratio of glutamic acid to sodium phosphate to make up the three buffers.

The results from pH 3.5 experiments were not used for final analysis in this project. ASA2 samples were dialyzed directly into the respective pH 3.5, pH 5.2, and pH 7.5 buffers. However, the excipient stock solutions had to be titrated to the appropriate pH because of the strong influence of some of the excipients on the pH of the buffer. Titration was done with 5N HCl, 1N HCl, 10N NaOH, and 1N NaOH. While ASA2-Control samples exhibited an increase in aggregation at 1 week in pH 3.5, ASA2 samples with excipients were for the most part completely denatured within 1 week. While this maybe due to lack of favorable excipient-protein interaction at pH 3.5, it is likely due to the addition of HCl during the titration steps. The hypothesis that the results from the pH 3.5 were confounded because of HCl was further validated using a small experiment. The experimental set-up and results are shown in Exhibit B of the Appendix. Thus the analysis of the results from wet-lab experiments are focused on pH 5.2 and pH 7.5 samples.

The concentration of excipients typically plays an important role in the excipients' ability to affect aggregation of the protein. As such, a high concentration and a low concentration were tested for every excipient with ASA2. The motivation behind the high concentration was to understand the impact of excipients on ASA2 when they are in far excess by moles compared to ASA2. However, the concentration of each excipient was constrained by its maximum solubility. For example, Xanthine, with the lowest solubility among the 32 excipients tested, was at most on a 1:1 stoichiometric ratio with ASA2. Alanine, on the other hand, had the highest solubility and at that concentration had a stoichiometric ratio of ~70,000:1 with ASA2. Given this large difference and the difficulties with arbitrarily choosing one concentration, the high concentration of the stock solution was determined to be at 80% of the maximum solubility of each excipient. The final excipient-ASA2 formulations contained half of this concentration since 125 μL of the excipient stock was added to 125 μL of ASA2. The low concentration stocks were chosen to be approximately at 20% of the molarity of high concentration stocks. The significant difference between the high and low concentrations of excipients were tested to understand the influence of excess concentration of excipients on ASA2's stability and aggregation.

In addition to pH, temperature of the sample also plays a critical role in protein aggregation and thermostability. All ASA2 was incubated in 40°C for one month. Since ASA2 is naturally a fairly

stable molecule with a low propensity for aggregation, the goal of incubating it at 40°C was to potentially disrupt that stability and understand the impact of the excipients. In addition, incubation at 40°C could also provide greater insight into excipients that specifically improve the thermostability of ASA2, but may not necessarily have an impact on colloidal stability.

The goal behind testing a large sample of excipients and a variety of experimental design factors was not only to better understand each excipient's relationship with ASA2 but also to understand trends across excipients compared to the computational data. Section 5.2 provides the results from the wet-lab experiments and Chapter 6 discusses the comparison between wet-lab results and computational results. The ability to run high-throughput wet-lab experiments played a critical role in being able to explore so many formulation designs efficiently.

5.2 Results and Analysis

The results from SEC, DLS, and DSF for all of the formulation samples at pH 5.2 and pH 7.5 are shown in Figures 20 - 27 below. Unfortunately, since ASA2 is a fairly stable molecule, there are no notable trends across all of the excipients. However, a few interesting results do emerge from these experiments. There are more excipients that show higher aggregation of ASA2 in the high concentration samples compared to the low concentration samples across both PH. Interestingly, many of the excipients that lead to higher aggregation are also acids and show higher aggregation at pH 5.2 compared to pH 7.5. In light of the results from the HCl experiments shown in Exhibit B of the Appendix, it is possible that a higher concentration of protonated acids interacts with ASA2 differently and leads to an increase in aggregation.

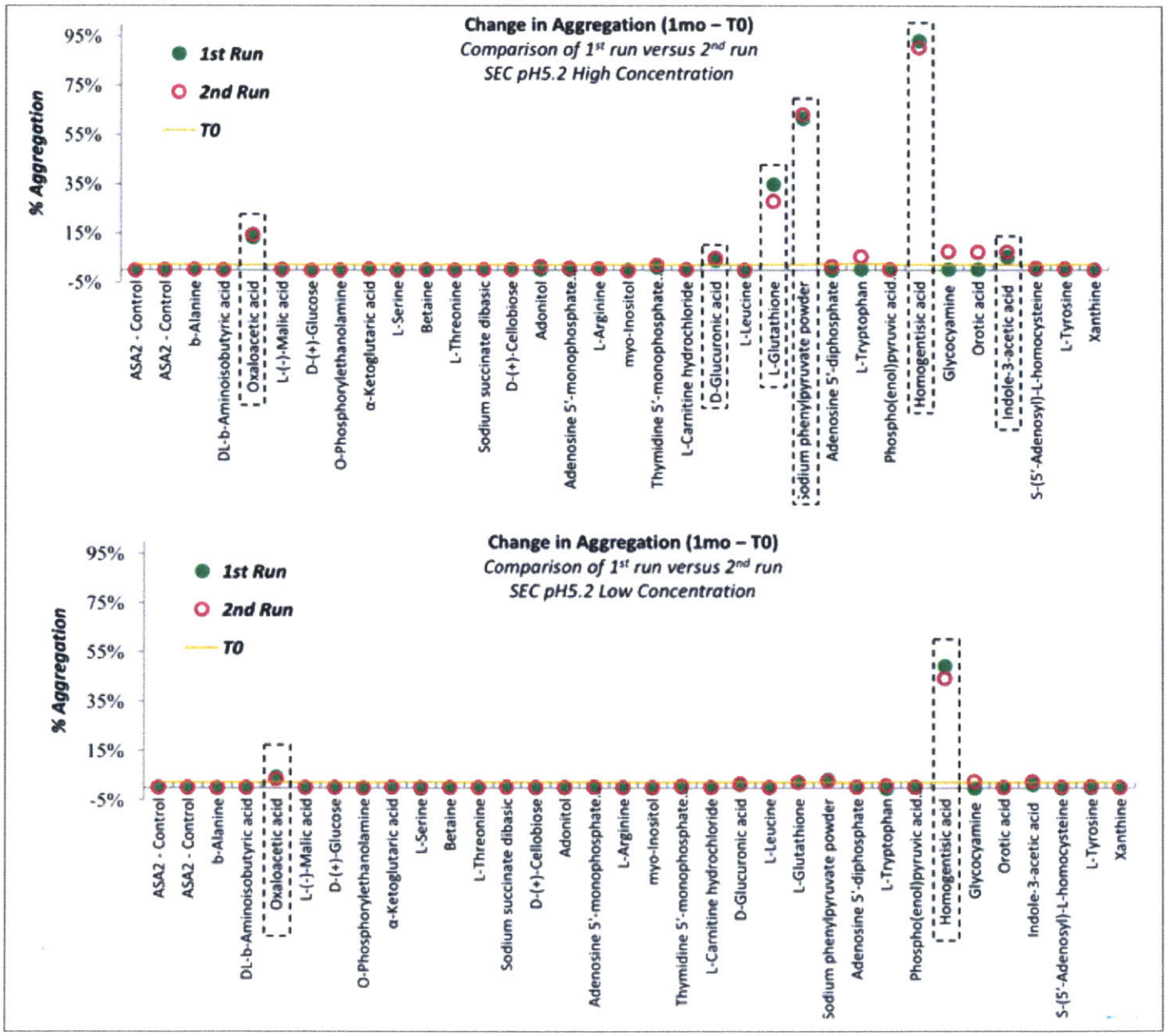


Figure 18: Change in ASA2 Aggregation Over One Month at pH

The figures show a change in aggregation from baseline for ASA2 with high and low concentrations of excipients at pH 5.2. The yellow line indicates the level of aggregation at T0. The green dots and pink circles indicate change in aggregation from the baseline for two replicates at 1 month. The excipients highlighted in dotted lines are ones that showed the most notable level of change from baseline.

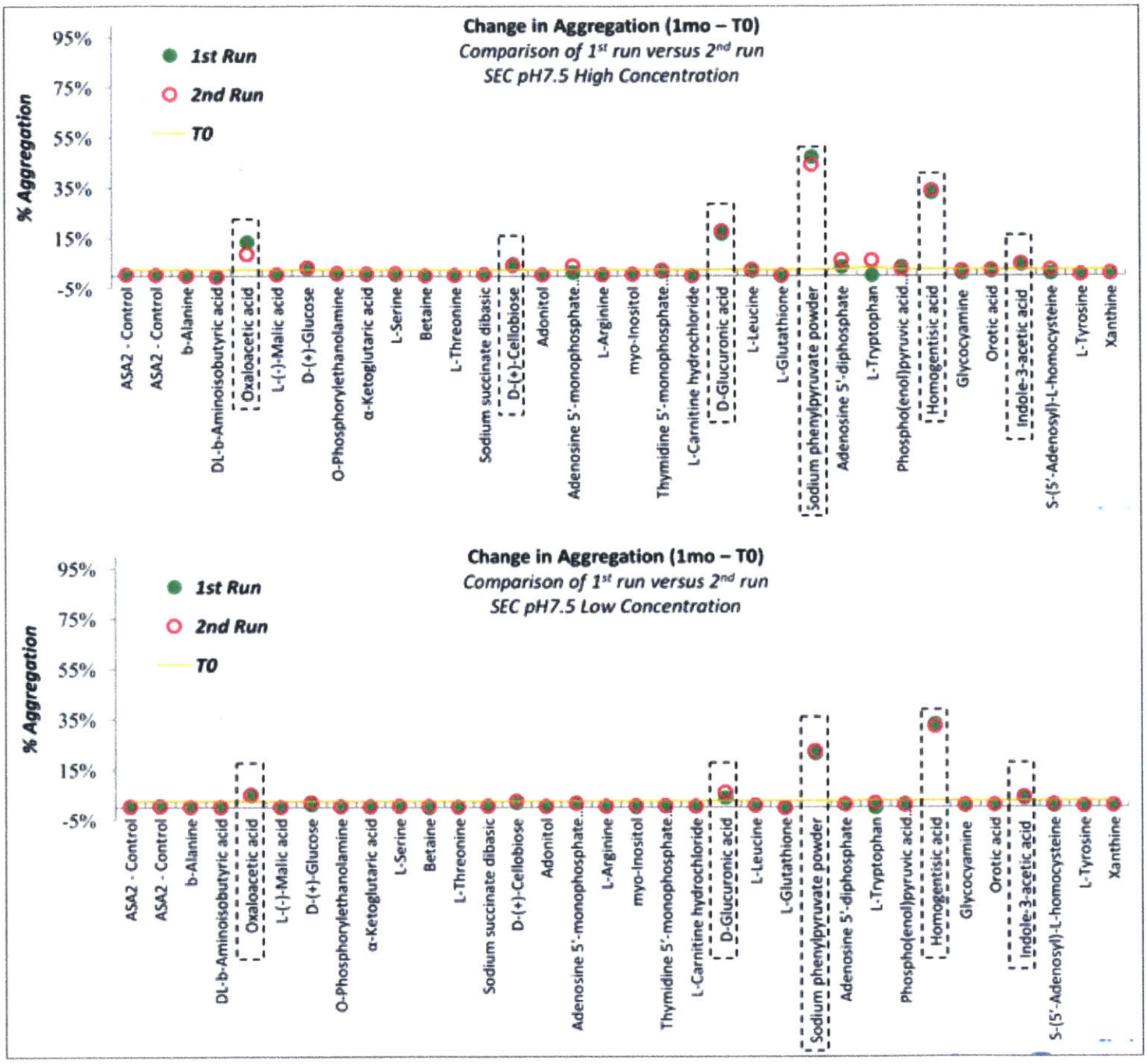


Figure 19: Change in ASA2 Aggregation Over One Month at pH

The figures show a change in aggregation from baseline for ASA2 with high and low concentrations of excipients at pH 7.5. The yellow line indicates the level of aggregation at T0. The green dots and pink circles indicate change in aggregation from the baseline for two replicates at 1 month. The excipients highlighted in dotted lines are ones that showed the most notable level of change from baseline.

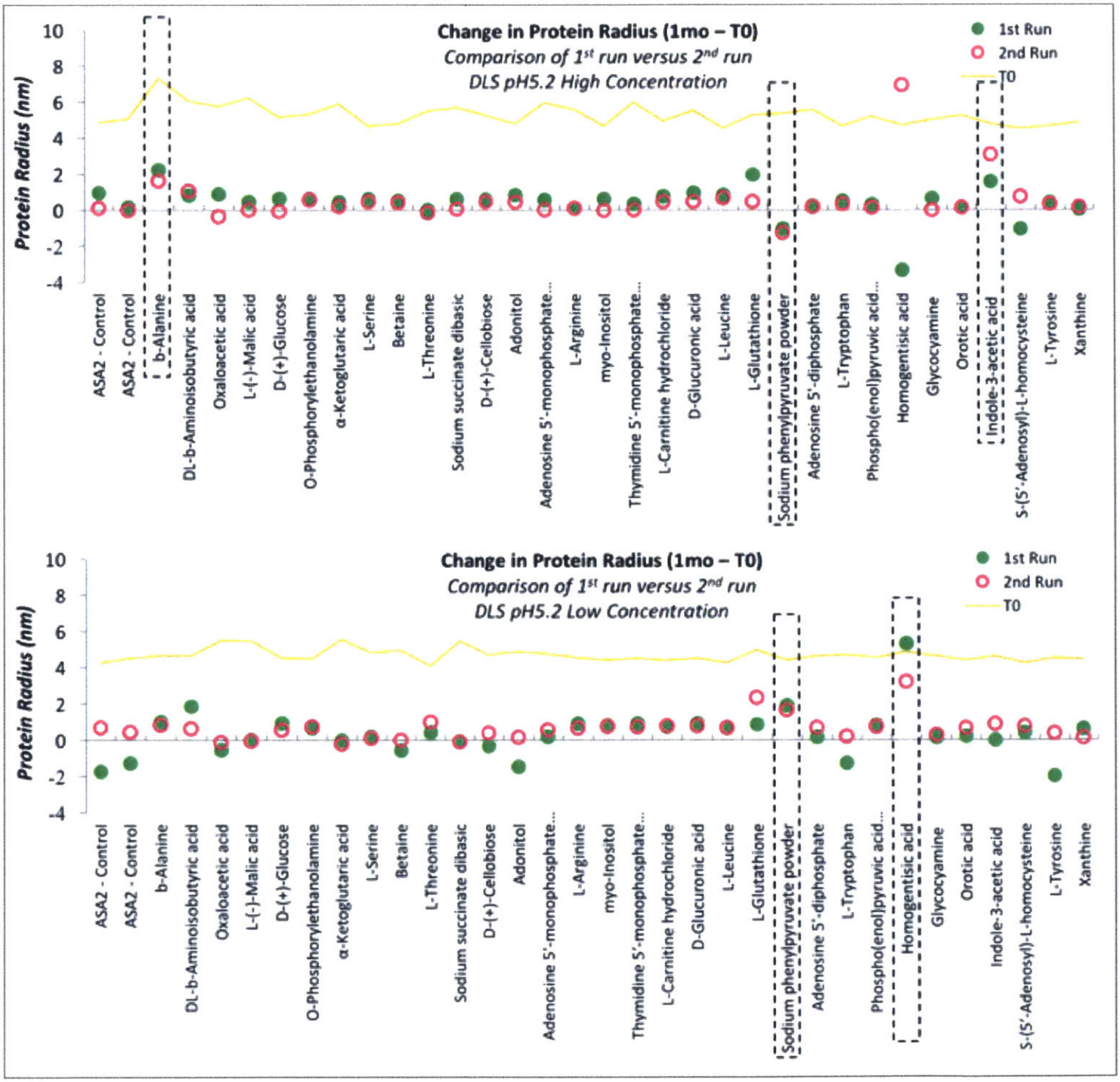


Figure 20: Change in Hydrodynamic Radius of ASA2 Over One Month at pH 5.2

The figures show a change in ASA2 radius from baseline for formulations with high (top graph) and low (bottom graph) concentrations of excipients at pH 5.2. The yellow line indicates the protein radius at T0. The green dots and pink circles indicate a change in radius at 1 month for two replicates. The excipients highlighted in dotted lines are ones that showed the most notable level of change from baseline for both replicates.

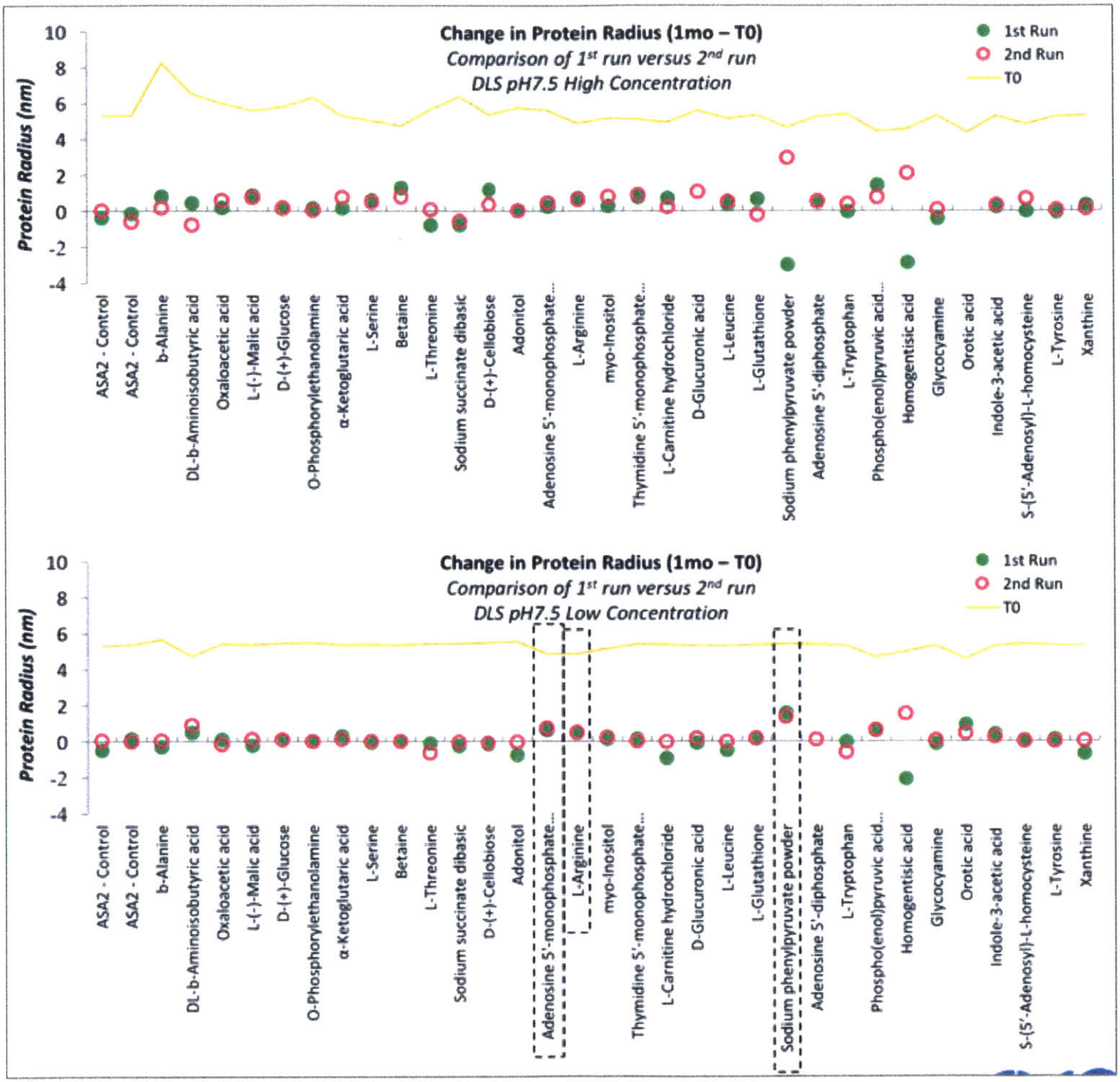


Figure 21: Change in Hydrodynamic Radius of ASA2 Over One Month at pH 7.5

The figures show a change in ASA2 radius from baseline for formulations with high (top graph) and low (bottom graph) concentrations of excipients at pH 7.5. The yellow line indicates the protein radius at T0. The green dots and pink circles indicate a change in radius at 1 month for two replicates. The excipients highlighted in dotted lines are ones that showed the most notable level of change from baseline for both replicates.

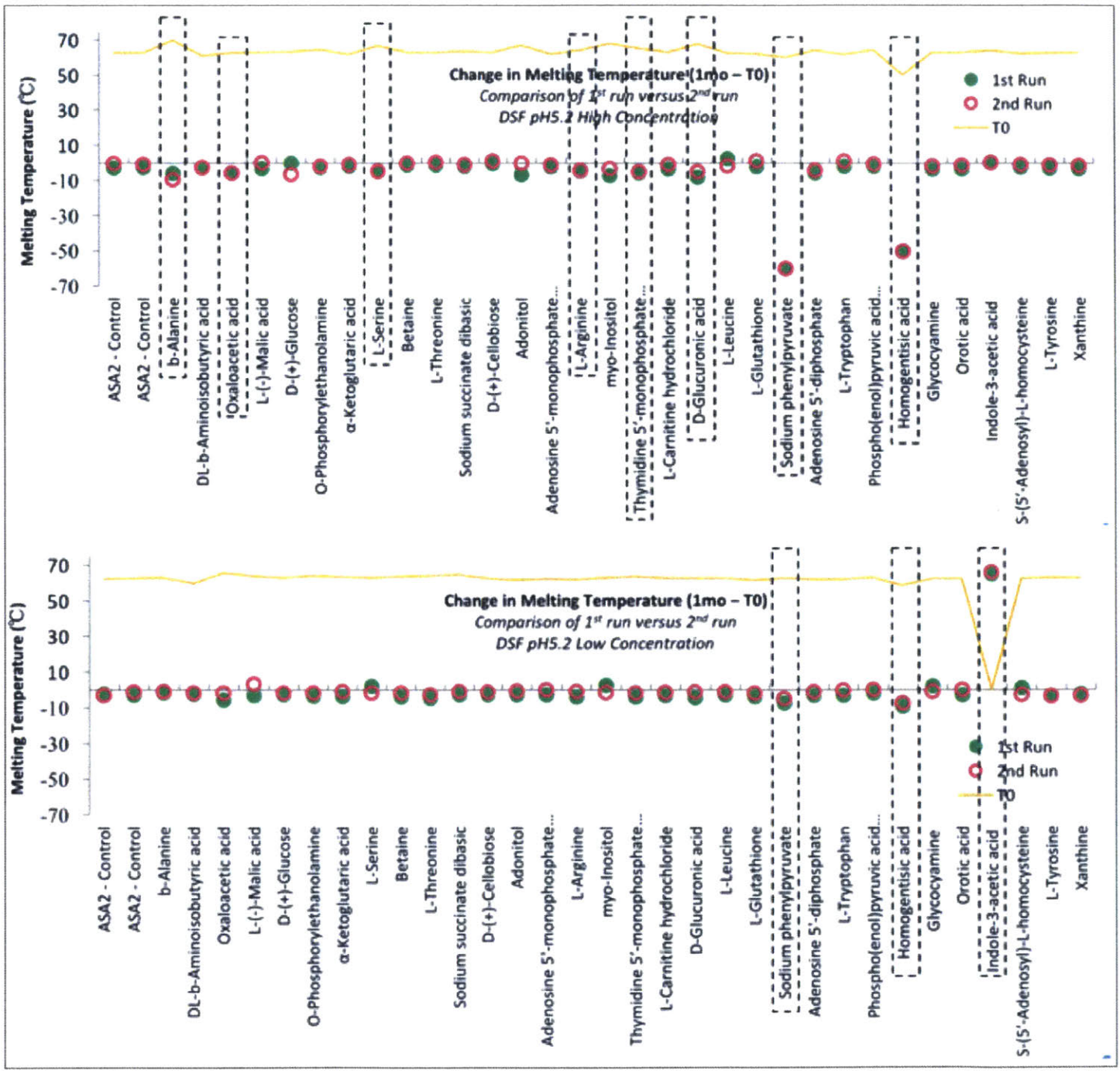


Figure 22: Change in Melting Temperature for ASA2 Over One Month at pH 5.2

The figures show a change in the melting temperature (Tm) for ASA2 from baseline for formulations with high (top graph) and low (bottom graph) concentrations of excipients at pH 5.2. The yellow line indicates the melting temperature at T0. The green dots and pink circles indicate a change in the melting temperature at 1 month for two replicates. The excipients highlighted in dotted lines are ones that showed the most notable level of change from baseline for both replicates.

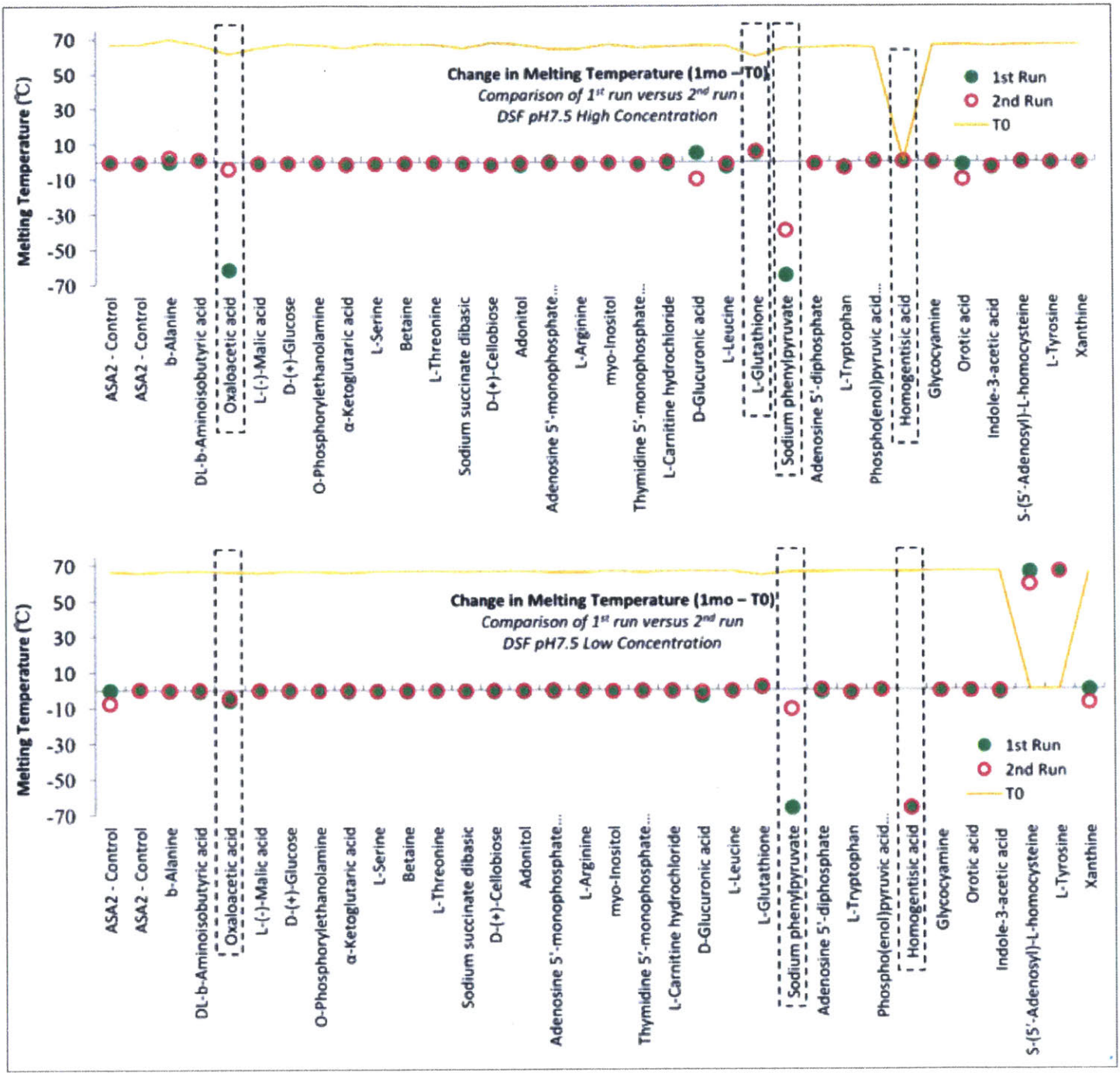


Figure 23: Change in Melting Temperature for ASA2 Over One Month at pH 7.5

The figures show a change in the melting temperature (Tm) for ASA2 from baseline for formulations with high (top graph) and low (bottom graph) concentrations of excipients at pH 7.5. The yellow line indicates the melting temperature at T0. The green dots and pink circles indicate a change in the melting temperature at 1 month for two replicates. The excipients highlighted in dotted lines are ones that showed the most notable level of change from baseline for both replicates.

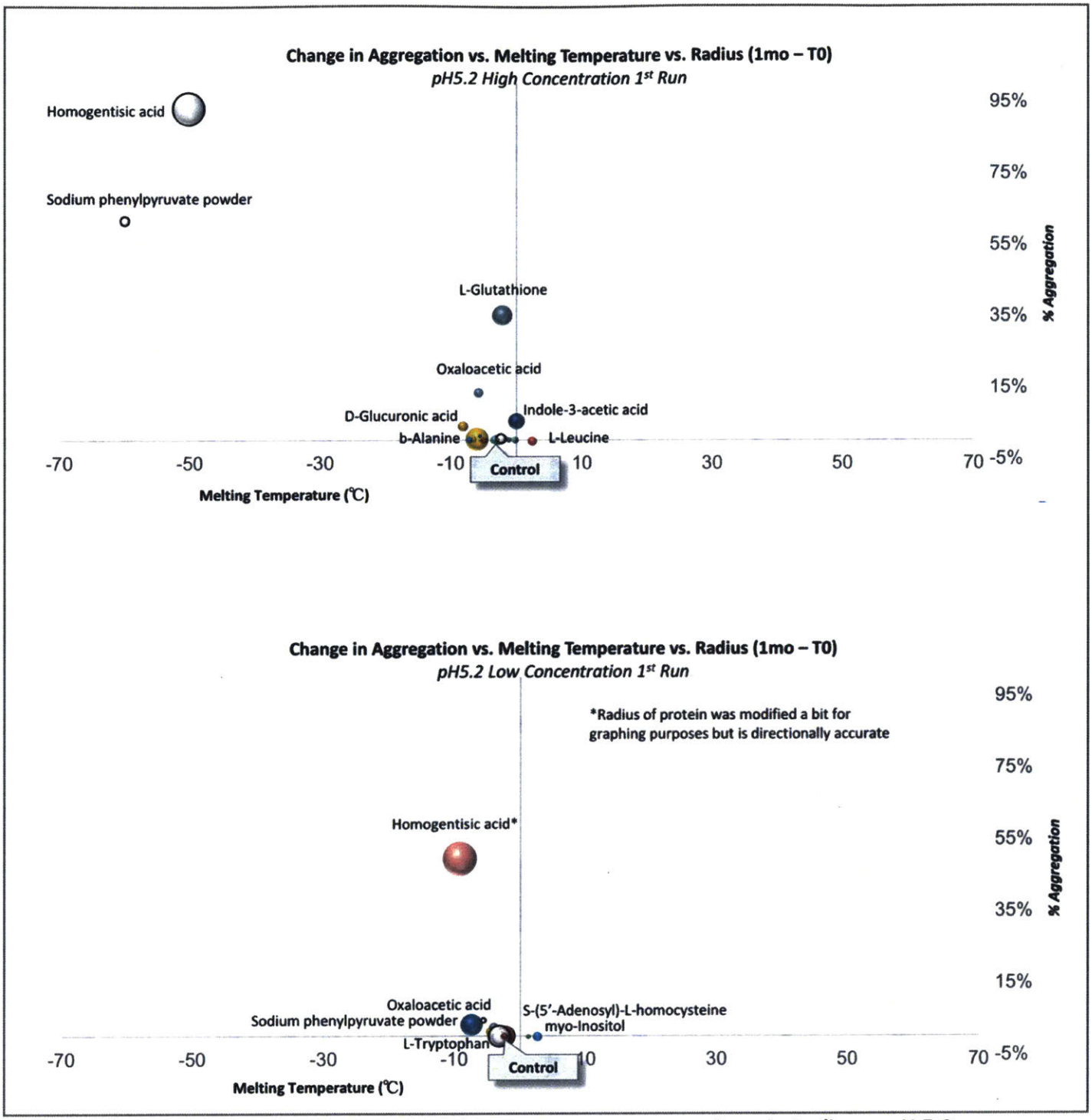


Figure 24: Aggregation versus Melting Temperature versus Protein Radius at pH 5.2

The figures show a change at 1 month compared to baseline for all ASA2 formulations for the first run (or replicate) at pH 5.2 for high (top graph) and low (bottom graph) concentrations of excipients. The change in aggregation from baseline is plotted in the y-axis, the change in melting temperature is plotted in the x-axis, and the change in protein radius is represented by the size of the circles. Each color represents a different ASA2 formulation.

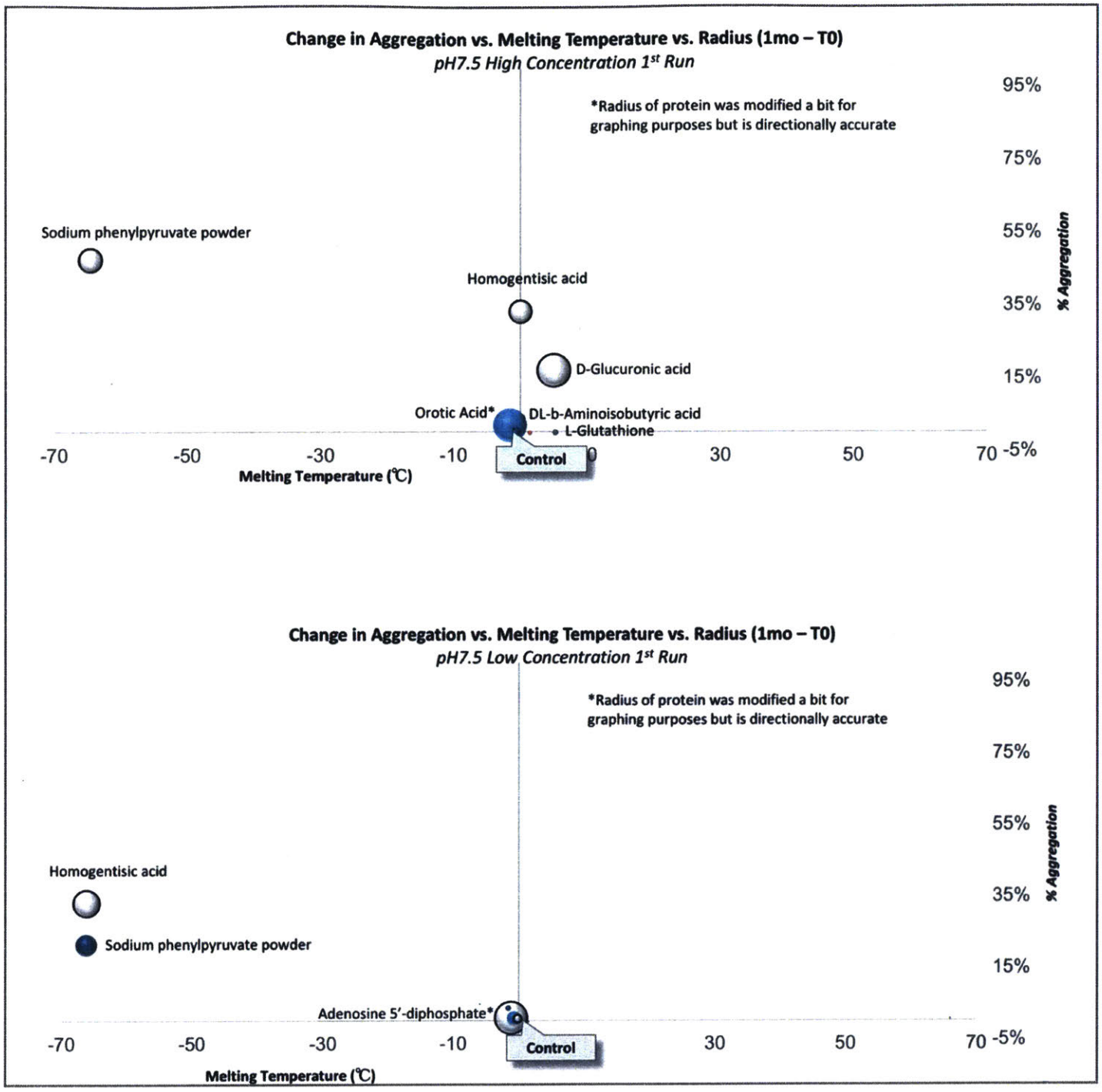


Figure 25: Aggregation versus Melting Temperature versus Protein Radius at pH 7.5

The figures show a change at 1 month compared to baseline for all ASA2 formulations for the first run (or replicate) at pH 7.5 for high (top graph) and low (bottom graph) concentrations of excipients. The change in aggregation from baseline is plotted in the y-axis, the change in melting temperature is plotted in the x-axis, and the change in protein radius is represented by the size of the circles. Each color represents a different ASA2 formulation.

This page is intentionally left blank.

6. In-silico and In-vitro Data Comparison & Analysis

6.1 Size Exclusion Chromatography Results versus Computational Results

The percentage of aggregates in each formulation sample was compared against the spread, footprint, and binding affinity. The hypothesis at the start of the project was that excipients that demonstrate a high spread, footprint, and binding affinity are the most likely to reduce aggregation. These excipients could cover the surface of ASA2 with favorable interactions, stabilizing the antibody and preventing the antibodies from interacting with each other. However, this trend was not observed. It was difficult to note any significant observations since most excipients did not seem to have an impact on ASA2 aggregation.

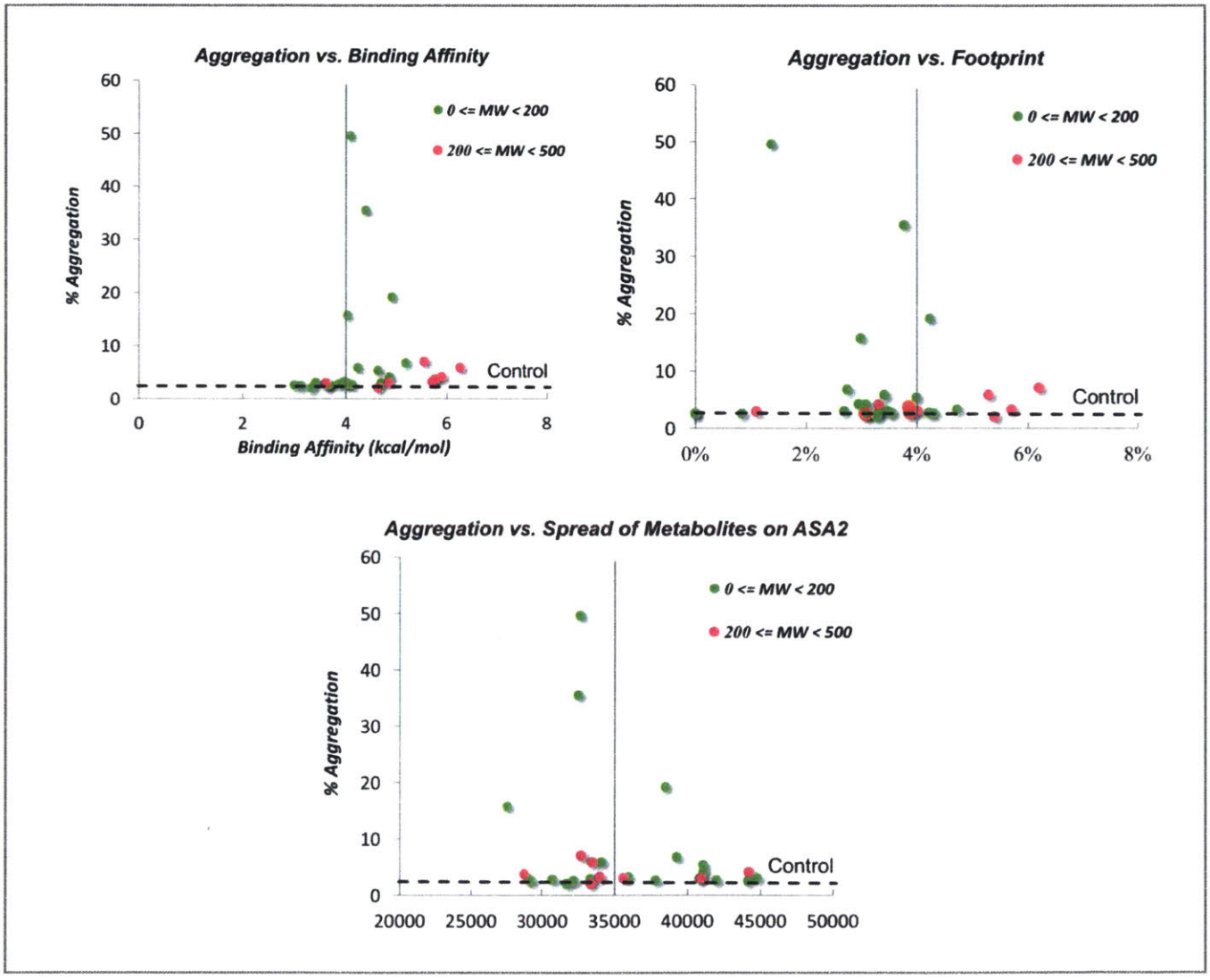


Figure 26: Aggregation versus Computational Results at pH 7.5, 1 month

However, there are trends that would be interesting to explore further. Figure 28 above, shows aggregation at pH 7.5, high concentration of excipients compared against binding affinity, footprint, and spread. Interestingly, the few excipients that have led to an increase in aggregation tend to have a higher binding affinity and low footprint. These are likely excipients that act more like small-molecule drugs than an excipient. Further analysis to identify the areas of interaction between ASA2 and these excipients using AutoDock Vina may offer insight into the reason for increased aggregation.

6.2 Dynamic Light Scattering Results versus Computational Results

Data from DLS was used to analyze the hydrodynamic radius of protein monomer with the hypothesis that a protein in a more compact, baseline-like state has more colloidal stability and that sample is likely to have lower aggregation. Based on the computational results, it was expected that excipients that have higher binding affinity, footprint, and spread are likely to stabilize ASA2 more by surrounding the antibody with favorable interactions. However, there are no correlations observed between the hydrodynamic radius of ASA2 from DLS and the computational results. Figure 29 shows results for pH 7.5, high concentration samples.

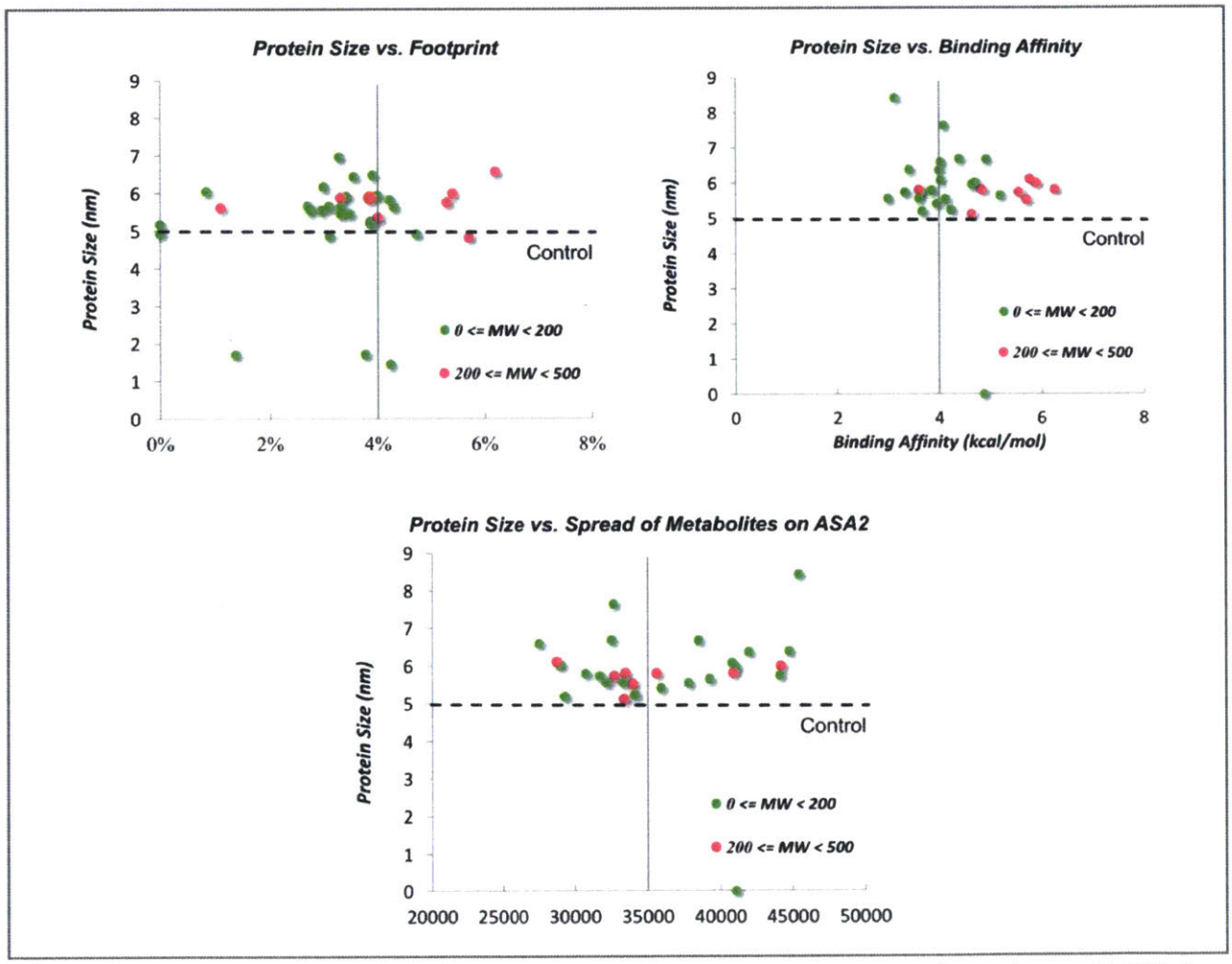


Figure 27: Hydrodynamic Radius of ASA2 ("Protein Size") versus Computational Results at pH 7.5, 1 month

6.3 Differential Scanning Fluorimetry Results versus Computational Results

The melting temperature of ASA2, derived from DSF, indicates its thermostability. Excipients that lead to higher melting temperature increase ASA2's stability and are likely to show less aggregation after incubation at 40°C. Excipients that lead to a lower melting temperature compared to control are likely destabilizing ASA2 and may result in increased aggregation after incubation at 40°C. Given the relatively high stability of ASA2-Control there were no excipients that demonstrated a significant increase in melting temperature. However, there are a few excipients that decreased the melting temperature of ASA2. Figure 30 shows the results for pH 7.5, high concentration samples. Similar to the trends observed with SEC, excipients that have resulted in a lower melting temperature have more drug-like properties – lower footprint and higher binding affinity. Further analysis on AutoDock Vina should be done to identify the specific areas of excipient-protein interaction for these formulations.

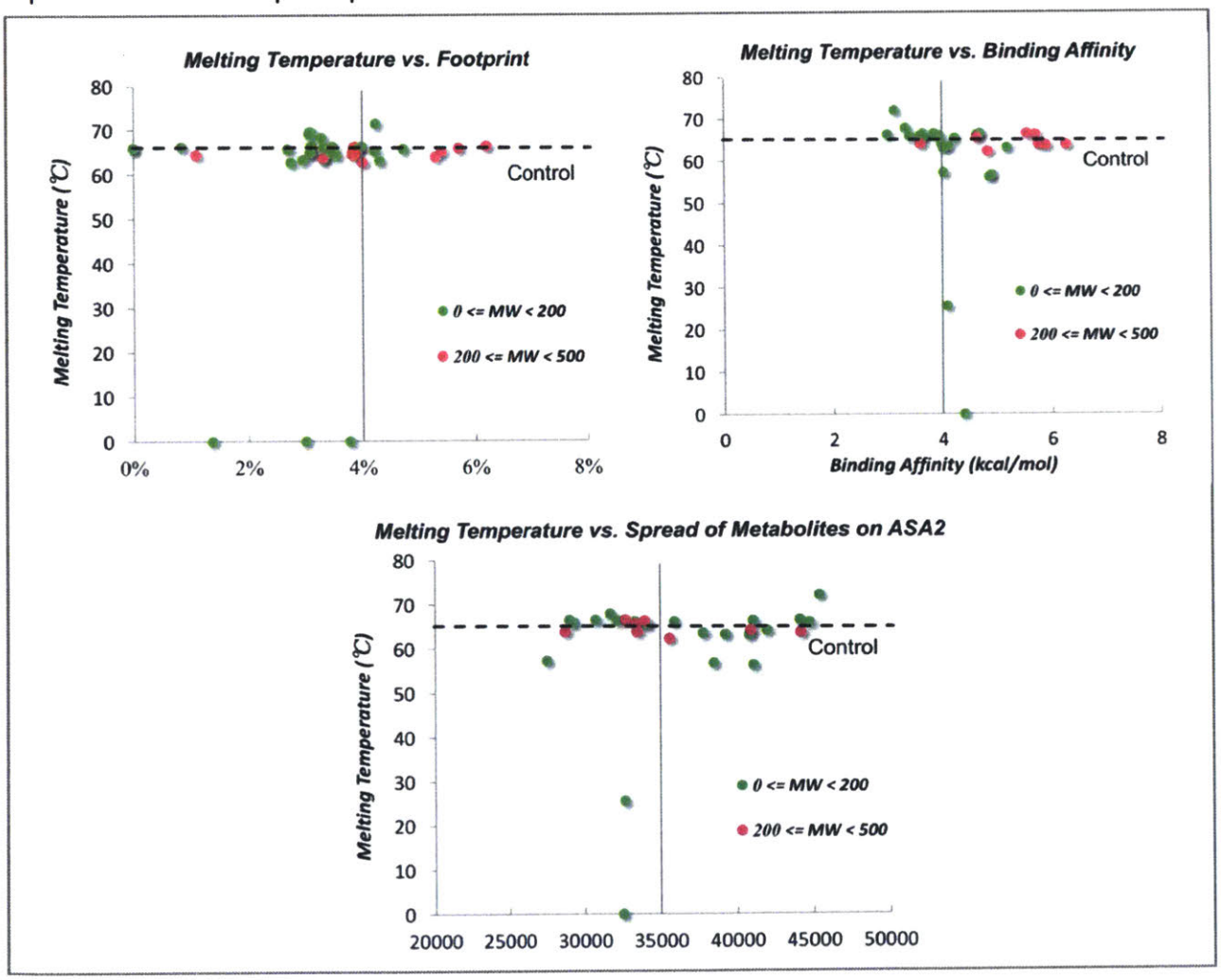


Figure 28: Melting Temperature versus Computational Results at pH 7.5, 1 month

6.4 Conclusions

6.4.1 Current Challenges and Areas for Improvement with AutoDock Vina

AutoDock Vina can potentially serve as an effective tool to model excipient-protein interaction because it is fast, easy to use, requires relatively low computational power, inputs are relatively easily generated, and outputs are relatively straight-forward to analyze. However, the biggest challenge with AutoDock Vina is understanding its capabilities and limitations in effectively modelling excipient-protein interaction. AutoDock is a molecular docking tool that is intended to predict non-covalent binding of small molecule ligand to macromolecules (Trott 2010). Its application has been most successful in screening small molecule drug candidates against a known protein target and target site, to identify leads in the drug development process (Trott 2010). It can also be used to predict the bound conformation of known ligand-macromolecule binders (Trott 2010). Thus its application in modelling excipient-protein interaction is novel and includes numerous challenges.

The search space provided for ASA2 in AutoDock Vina is larger than the typical search spaces for docking sites. In initial tests of 190 protein-ligand complexes by AutoDock Vina creators, the search space was 22.4Å in the x,y,z dimensions (Trott 2010). The size of this search base was based on the following logic:

- Initial sizes of each dimensions were based on minimal rectangular parallelepiped that covered the bound ligand structure (Trott 2010)
- The sizes were increased by 10Å in each of the three dimensions and then an additional 5Å was added to one of the directions of each dimension (Trott 2010)
    - The increase by 10Å ensured that the ligand has enough space to rotate in and the addition of 5Å in a randomly selected direction for each dimension prevented any bias from centering the search space on the binding location (Trott 2010)
- Ultimately each dimension was increased to 22.5Å to maintain compatibility with previous tests performed on AutoDock and to provide enough room for ligand movement (Trott 2010)

AutoDock Vina also recommends that the search space be no larger than 30Å in each dimension to make it easier for the program to search for and find the global minimum ("AutoDock Vina Manual" 2010). However, for ASA2, a search space with the size 144Å, 117Å, and 136Å was used in x,y,z dimensions, respectively. This accounts for the full size of the antibody and an addition of 10Å in each direction. The motivation behind using a large antibody is because the large antibody more closely simulates the type of protein therapeutics that are on the market and that Amgen works with. The reason to use its full size for AutoDock Vina is because it provided information on preferential binding sites across the whole antibody and could potentially be used to differentiate excipients that bound more strongly to certain parts of the antibody. However, since the search space was large, it is not clear that binding positions that represented the global minimum were actually found. Furthermore, the nondeterministic algorithm that guides the search process in AutoDock Vina led to large variations in each run,

that despite having numerous runs, the net results may be a wash. In order to address this issue, it would be helpful to consider the following steps for future simulations on AutoDock:

1. Identify hotspots on the antibody that are prone to aggregation and focus the search space to those areas
2. Increase exhaustiveness of AutoDock Vina search as this linearly increases the time spent on each search and increases the chances of finding the global minimum ("AutoDock Vina Manual" 2010)
    a. Although this was tried in this project with no noticeable difference in the results, it is possible that it might help in combination to the other recommendations provided here
3. User smaller proteins (e.g., lysozyme) for this proof of concept study
4. Results from three different runs for each of the three conformations (a total of 9 runs per excipient-ASA2 pair) were used in this project. However, it might be helpful to do more runs, particularly for larger search spaces

Another challenging aspect with using a large, complex antibody like ASA2 is determining the conformations that should be used in modelling. AutoDock Vina keeps the macromolecule static and rigid while the ligand has some conformational flexibility. Three different conformations of ASA2 was used in this project based on findings from Clark 2013. However, this project did not use the functionality in AutoDock Vina to identify certain side chains to be flexible during docking ("AutoDock Vina Manual" 2010). The combination of these two factors like significantly impacted the AutoDock Vina results. However, the conformations used for modelling may not actually represent the conformations ASA2 takes in the presence of those excipients at different pHs and 40C incubation, thus significantly altering the results. As a proof of concept study, it would be helpful to do the following:

1. Choose a smaller protein with fewer degrees of flexibility related to conformational changes; if possible identify target sites within the protein for the search space
2. Utilize AutoDock Vina's functionality to provide flexibility to side chains during docking

Another important factor that needs to be further reviewed is the protonation states of ASA2 and the excipients. pH plays an important role in determining the stability of the protein, its propensity for aggregation, and its interaction with excipients (Chaudhuri 2014). Currently, the only way to account for pH in AutoDock Vina is by altering the protonation states and charge distribution of molecules. For this project, protonation states of ASA2 and the excipients were all based on a pH of 7.0, even though experimentally pH 3.5, 5.2, and 7.5 were tested. Given the important role of pH, Autdock Vina should be tested again with excipient-protein combinations with protonation states that accurately matches the pH at which the solution will be maintained in.

AutoDock Vina also does not take into account the impact of water or other buffers on the excipient-protein interaction. While this may not present as a challenge in the identification of small-molecule drug targets due to the need for leads that have a strong binding affinity to the protein, it could become problematic in finding excipients to prevent aggregation. The binding affinity for excipients is and likely should be less than that of a small molecule drug target to an

active site. However, the current modelling capabilities of AutoDock Vina do not capture the complexity of the interaction between water/buffer and ASA2 and water/buffer and excipients. The level of preferential interaction of water wit the excipients or the protein could have a considerable impact on the wet-lab results observed. One possible way to mitigate this risk is by using third-party add-ons like WaterDock to understand preferential binding of the protein to the ligand/excipient instead of water (Ross 2012).

But ultimately there are two important disadvantages to AutoDock Vina that might prevent its use in identifying excipients to reduce protein aggregation. The first is that AutoDock Vina models ligand-macromolecule interaction in a temperature-independent fashion (Trott 2010). However, temperature plays an important role in protein aggregation. Higher temperatures typically lead to greater protein instability resulting in multiple starting points for protein aggregation, including thermodynamic instability, protein denaturation, hydrophobic interaction, greater movement and physical interaction of proteins (Wang and Roberts 2010). AutoDock Vina currently does not provide tools to model the impact of temperature on the excipient and protein structures. The second is that AutoDock Vina ultimately models the interaction of a small molecule and a macromolecule and it might be too difficult to translate the findings from this interaction to their behavior in solution. It is possible that the AutoDock Vina results might be more helpful for formulations that are lyophilized (Tarar 2013). But in general since the industry is moving towards liquid formulation and thus it is important to have a computational modelling tool that would be effective in this setting.

6.4.2 Current Challenges and Areas for Improvement with Wet-lab Experiments

Unlike the methods for computational modelling, the methods for wet-lab experimentations to identify protein aggregation and stability are pretty well established. However, there are still challenges that emerged during the design and implementation phase of high-throughput experimentation conducted in this project. Addressing these issues may help provide more confidence and reliability to the results from the wet-lab experiments and its relationship to the computational results.

Approximately 1344 experiments were conducted in total, not including the study of control samples. Of all the different factors that are listed in the table above, there are three design choices that likely had the most direct impact on the outcome: concentrations and pH. The high concentrations of the excipients were determined based on 80% of the maximum solubility of each excipient in solution and for the low concentration, it was 20% of the high concentration. The actual molarity of each of the excipients tested is listed in Table 8. However, it is possible that at the high concentration, the stoichiometric ratio of excipients to protein is too high to see any differences in results for the most part and at the low end, it is too small, especially for some excipients. It is also challenging to compare the results across different excipients given the confounding factor of different concentrations for each excipient. Thus it would be helpful in future experiments to determine the concentration of excipients based on stoichiometric ratios that are the same across all the excipients.

The second design challenge is determining the pHs for experimentation. Since ASA2 is a natively stable antibody at pH 5.2, pH 3.5 and pH 7.5 we tried to destabilize it and test the additional benefit of the excipients. The other two experimental pHs (3.5 and 7.5) were determined based on feedback from Amgen scientists and some publications (Alekseychyk 2014). The motivation for selecting pH 3.5 and pH 7.5 was that it would provide conditions to destabilize ASA2 and promote aggregation without completely destroying the antibody. However, addition of certain excipients significantly changed these pHs immediately and individual samples needed to be titrated back to the appropriate pH. Furthermore, titration with HCl, especially in acidic conditions (i.e., pH 3.5) seems to have a strongly negative impact on ASA2. Due to the confounding results from HCl, the results from the entire set of experiments conducted at pH 3.5 could not be used for analysis in this project. Lastly, for ASA2, pH 7.5 does not seem to have significantly destabilizing effect compared to pH 5.2 in this project. The design process should take into account these challenges and perform empirical tests first to identify the appropriate pH conditions for experimentation.

The third design challenge is related to collecting data for replicates. There were no replicates designed in each of the 96-well plates (except for the controls) in the current methods for this project. This was because the goal of the project was to observe trends across excipients, the thought process that replicates can be independently set up in new plates, and because of time and resource constraints. A replication of all of the experiments were conducted, though only for some time points, across all of the samples, in new 96-well plates to verify the findings from the first run. However, there was too much variability introduced with setting up the second set of samples from the degree of pipetting to small differences in concentration of the protein and changes in the SEC column. Likely a better design approach for the next set up experiments is to build in replicates of three within the same run and plate for each sample.

In addition to these design challenges, there were also numerous challenges related to the set-up and execution of the wet-lab experiments. The pH of the samples not only presented a design challenge, but a setup challenge as well. Each of the 192 unique samples (32 excipients * 2 concentrations * 3 pHs), needed to be individually titrated to a pH that was close to the target pH. This not only increases the opportunity for error, but adds titration buffers at various concentration for each excipient, and changes the concentration of excipients. All of these factors could impact protein aggregation and stability and ultimately confound the results on the impact of the excipient itself. Each individual excipient solution was also developed by manually weighing out small quantities of the excipient and dissolving it in buffer. While to the extent possible, the amount weighed out was close to the target amount and was tracked, there is also room for error in this process given the small quantities. Lastly, given the large number of samples and high levels of aggregations in some samples (especially in pH 3.5), the SEC column's filtration property eroded and had to be replaced twice in the middle of the experiments, additional addition variability to the results.

Despite all of these challenges, likely the most important factor leading to inconclusive results is the fact that ASA2 is a fairly stable antibody that did not show a significant difference in aggregation at pH 5.2 and pH 7.5, even without any additional excipients, and after one month

of incubation in 40°C. Thus, with the exception of few excipients that have led to increases in aggregation, there are no clear differences from the baseline that is observed. Selecting a protein with a propensity for aggregation will likely produce more interesting results and insight into if there is a relationship between the computational and experimental analysis.

7. Next Steps & Recommendations

This project accomplishes two major steps in formulation development:

1. It serves as a starting to point to develop computational models that can do high throughput screening of excipients virtually
2. It identifies databases of compounds that could be used for both computational and wet-lab experiments

While the latter is the primary output from this project that could immediately add value to formulation scientists, a majority of the discussion in this section will be focused on recommendations to further test the role of computational modelling, specifically AutoDock Vina, in formulation development. Harnessing the power of computational modelling to test hundreds and thousands of formulation designs is a truly innovative way to develop novel formulations in a faster and more cost-effective manner. Furthermore, computational modelling has the power to provide insight into excipient-protein interaction at an atomic level, leading to a greater understanding of science and better drug development in the future. The benefits related to computational modelling is only going to become more important as the challenges related to formulation development become more urgent. There are three industry trends that are likely to further amplify the need for a robust high-throughput computational model to select excipients. The first is related to new formulation challenges that will likely come with the introduction of new modalities of protein therapeutics. The second is from re-formulation needs of currently approved protein therapeutics due to advancements in delivery devices and increased patient preference for convenient liquid or oral formulations. Lastly, biosimilars may sometimes require new formulations that match the functionality of the original drug due to delayed formulation patent expiry. The current reliance on DOE, trial-and-error, platform formulations, and empirical expertise may at best exacerbate problems related to the time and cost of drug development and at worst, may not provide a sufficient formulation at all. Developing a high-throughput computational model has significant potential to improve this process, that at its core has not changed over the last couple decades.

This project pursues AutoDock Vina as a way to computationally model excipients against a target protein, ASA2. A docking tool that has traditional been used to screen small molecule drug candidates against the active site in macromolecules, AutoDock Vina was used in this project to understand the degree of non-specific interaction between each excipient and ASA2. The ultimate goal of this endeavor was to screen for excipients that would reduce ASA2 aggregation or increase its stability, as measured through SEC, DLS, and DSF. However, results from this first stage of development do not show a significant correlation between computational results and experimental results. While it is possible that AutoDock Vina is not an appropriate tool to screen for excipients, its value in this process can not yet be missed. There are numerous confounding factors that are discussed in Section 6.4 that could be affecting the outcomes and should be explored further to definitively establish AutoDock Vina's value in this process.

Based on the findings and lessons learned through this project, Figure 31 summarizes the recommended next steps to do a final assessment on the role of AutoDock Vina in screening for excipients that help prevent aggregation and increase stability.

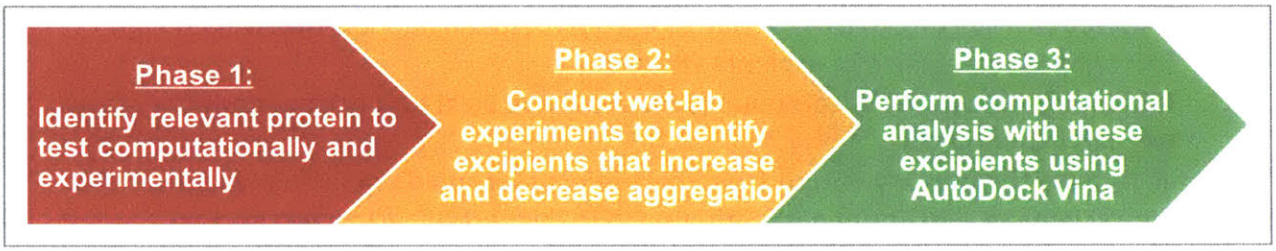


Figure 29: Overview of Recommended Project Approach

The first phase of this project should focused on identifying an appropriate protein to test computationally and experimentally. Computationally the protein should be small with only a few or no structural conformation changes. ASA2, tested in this project, is a large antibody with ~1314 amino acids, molecular weight of ~150 kD, and surface area of ~62,000Å. In addition to this, ASA2 can also take on various structural conformations because of the flexibility provided by the two hinge regions that serve as the attachment between each of the Fab arms and the Fc (Clark 2013). This is an especially important factor because the structural conformation of ASA2 strongly impacts the parts of the antibody that are exposed and can interact with the excipients. Experimentally, the protein should have a known propensity for aggregation at a defined pH and within a one-month time point at 40C or room temperature. One of the challenges with this project was that since ASA2 was a fairly stable molecule with very low aggregation, it was by nature difficult to identify excipients that reduced it even further. If a protein is selected that has a high propensity for aggregation, then it would also be helpful to collect information on any active sites that serve as the point of aggregation. Lastly, it would also be helpful to select a protein with tendencies for aggregation, that also has a known formulation to prevent it. This will help in setting up the experiments in the second phase of this project.

Lysozyme might potentially be good option to consider for this project. It is a relatively small protein, especially compared to ASA2. It has 129 amino acids, weighs about ~14 kD, and has a surface area of 6000Å ("Product Information: Lysozyme" 2017, Gregory 1995). Initial computational tests, as described in Section 4.1, shows that AutoDock Vina can differentiate between site-specific interactions and non-specific interaction for different excipients even when a large search space that covered the entire molecule was provided. Furthermore, lysozyme is available to be purchased commercially from many sources, including Sigma-Aldrich ("Product Information: Lysozyme" 2017). Lysozyme's propensity for aggregation under different experimental conditions needs to be further validated. However, initial journal reviews indicate that aggregation is a problem with lysozyme.

The second phase of this project should focus on wet-lab experiments. Similar to the experiments conducted in this project, SEC, DLS, DSF, and other potentially interesting

experiments should be conducted with clear design criteria. Data should be collected at T0, 1 week, 2 weeks, and 1 month across all of the different different experiments. Ensure that at least 3 replicates and controls are built in. Isolate lysozyme in a buffer at a target pH and eliminate other excipients. This will be helpful in understanding the effects of the excipients and minimize confounding variables. Also ensure that the final concentration of lysozyme used is high enough to meet the requirements for SEC, DLS, and DSF, and also high enough to enable aggregation. Lastly, with regards to the excipients, selection should be based on literature search and trial-and-error. Although this only replicates current processes for identifying excipients, hopefully it will ultimately help to evaluate AutoDock Vina. Ideally select three excipients that increase aggregation, three excipients that reduce aggregation, and three that have no effect. Evaluate excipients in three concentrations (high, medium, low), based on stoichiometric ratios and maintain these ratios across all of the excipients. The highest concentration should roughly cover the surface of the protein. Given the difficulty in preparation of the excipient samples described in Section 6.4, samples should be created in larger volumes and titrated to the target pH without HCl. Excipients should only be added to the lysozyme solution after this step is done. The list of excipients in Exhibit A of the Appendix from Sigma Aldrich that were evaluated for this project would be a good place to start.

Once the wet-lab experiments are completed or while waiting for the 1-month time point, computational analysis using AutoDock Vina should be done. Ensure that the protonation states and charges on lysozyme and the excipients match experimental conditions. The search space should be approximately 10Å larger in each dimension compared to the size of lysozyme. The exhaustiveness parameter in AutoDock Vina should also be explored further to determine if it has an effect on the outcome. Run AutoDock Vina at least 3 times for each excipient. If the results are very different, particularly as it relates to the spread, then more times may be needed. Water dock should be performed if necessary. If there are active sites that induces aggregation, then it would be helpful to do AutoDock runs with specifically those areas as the search space. The AutoDock Vina results should be analyzed based on the methods described in this project and visually using VMD or other tools.

If these results from AutoDock Vina correlate with wet-lab outcomes, then the model should be evaluated using two different lenses. The first is a more qualitative evaluation of the additional scientific insight offered by AutoDock Vina on the excipient-protein interaction. Often the business impact of such knowledge is difficult to quantify immediately, but may overtime be critical in advancing formulation development and drug development. The second is a quantitative evaluation of cost/benefits associated with continued development of AutoDock Vina in excipient selection. It is possible that even if strong trends are realized in this project, there might still not be a business justification to continue to pursue this model if there are too many criteria that need to be satisfied for the model to produce useful results. The project described in this section could also help conclusively rule out the value of AutoDock Vina in excipient selection to reduce aggregation and increase stability.

In addition to the project on evaluating the role for AutoDock Vina in excipient selection, there are several other projects that would be worth pursuing based on the results from this project.

It would be interesting to pursue additional studies to evaluate the role of HCl in protein degradation. There are some studies that already explore this area a little, but it would be especially valuable to pursue this topic of research in comparison to other acids that are used for titration. In addition to wet-lab experiments, computational modelling could also provide insight into this degradation process. Another interesting area that could be particularly helpful in systematic formulation development is building a curated database of excipients that are used or have been tested with different types of therapies, along with a summary of the function they have served in formulation, if they ae listed in the IID, etc. Currently this information is difficult to find and not collected or shared systematically. However, the process of formulation development could potentially be more efficient by the addition of easy-to-use database with regularly updated information on different excipients.

Formulation development is a critical area in the drug development process. It enables a therapeutic protein candidate to become an effective commercial drug by ensuring the follow attributes in liquid or lyophilized form:

- Maintaining stability
- Preventing aggregation
- Modifying the viscosity to meet dosage requirements
- Ensuring biocompatibility
- Enabling compatibility with delivery device
- Providing a long shelf-life

While there have been many advances overall in the experimentation process (e.g., high-throughput automated equipment, advanced computational algorithms for analysis of the data), the core process that is used for formulation development has not changed notably. After decades of experience, platform formulations were developed that serve as the starting point for certain types of protein. While this may still be an effective option for those proteins, introduction of new modalities and delivery devices will bring out new formulation challenges that might force scientists to go through extensive trial-and-error process. This project serves as a first step to highlight two potential areas for development that could help address impending formulation challenges. The first is to develop an easy-to-use comprehensive database that is systematically updated to include information on excipients, their function, and benefits in formulation, and shared with the entire formulation development for future use. The second is continuing the pursuit of developing computational modelling that can not only do high throughput screening for excipients to address formulation needs, but also reveal insight into excipient-protein interaction.

8. References

"About Amgen." Amgen Fact Sheet 9 Jan 2017: http://www.amgen.com/~/media/amgen/full/www-amgen-com/downloads/fact-sheets/fact_sheet_amgen.ashx.

Alekseychyk L, Su C, Becker GW, Treuheit MJ, Razinkov VI. "High-Throughput Screening and Stability Optimization of Anti-Streptavidin IgG1 and IgG2 Formulations." Journal of Biomolecular Screening. 2014; 19(9): 1290-1301.

Altman LK. "A New Insulin Given Approval for Use in U.S." *New York Times* 30 Oct. 1982: http://www.nytimes.com/1982/10/30/us/a-new-insulin-given-approval-for-use-in-us.html.

"Amgen Inc." Yahoo! Finance 10 March 2017: https://finance.yahoo.com/quote/AMGN?ltr=1.

"Amgen's 2015 Revenues Increased 8% To $21.7 Billion and Adjusted Earnings Per Share (EPS) Increased 19% To $10.38." *Amgen News Releases* 28 Jan. 2016: http://wwwext.amgen.com/media/news-releases/2016/01/amgens-2015-revenues-increased-8-percent-to-217-billion-and-adjusted-earnings-per-share-eps-increased-19-percent-to-1038/.

"Anti-Streptavidin antibody [S3E11]." Abcam. 2017: http://www.abcam.com/streptavidin-antibody-s3e11-ab10021.html.

"Aranesp." Highlights of Prescribing Information. April 2016: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/aranesp/ckd/aranesp_pi_hcp_english.ashx.

"AutoDock Vina Manual." AutoDock. 2010: http://vina.scripps.edu/manual.html.

Bains W. "Failure rates in drug discovery and development: will we ever get any better?" *Drug Discover World* Fall 2004: http://www.ddw-online.com/business/p148365-failure-rates-in-drug-discovery-and-development:-will-we-ever-get-any-better-fall-04.html.

Ballentine C. "Sulfanilamide Disaster." FDA Consumer Magazine June 1981: https://www.fda.gov/AboutFDA/WhatWeDo/History/ProductRegulation/SulfanilamideDisaster/ucm2007257.htm

Basavaraj S, Betageri GV. "Can formulation and drug delivery reduce attrition during drug discovery and development – review of feasibility, benefits, and challenges." Acta Pharmaceutica Sinica B; 4(1): Feb 2014: http://www.sciencedirect.com/science/article/pii/S2211383513001081.

"Blincyto." Highlights of Prescribing Information. Nov 2016: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/blincyto/blincyto_pi_hcp_english.ashx.

Brekke OH, Michaelsen TE, Sandlie I. "The structural requirements for complement activation by IgG: does it hinge on the hinge?" Immunology Today. February 1995; 16(2): 85-90.

"Category: ZINC." ZINC12. 2 June 2015: http://wiki.bkslab.org/index.php/ZINC_Database.

Chang BS, Hershenson S. "Practical approaches to protein formulation development." *Rationale Design of Stable Protein Formulations* 2002: pp. 1-25.

Chaudhuri R, Chen Y, Middaugh CR, Volkin DB. "High-Throughput Biophysical Analysis of Protein Therapeutics to Examine Interrelationships Between Aggregate Formation and Conformational Stability." AAPS J. January 2014; 16(1): 48-64.

Clark NJ, Zhang H, Krueger S, Lee HJ, Ketchem RR, Kerwin B, Kanapuram SR, Treuhet MJ, McAuley A, Curtis JE. "Small Angle Neutron Scattering Study of a Monoclonal Antibody Using Free-Energy Constraints." J. Phys. Chem. B. 2013; 117(45): 14029-14038.

"Corlanor." Highlights of Prescribing Information. Jan 2017: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/corlanor/corlanor_pi.ashx.

Degli Esposti L, Sangiorgi D, Perrone V, et al. Adherence and resource use among patients treated with biologic drugs: findings from BEETLE study. ClinicoEconomics and Outcomes Research: CEOR. 2014;6:401-407. doi:10.2147/CEOR.S66338. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4171999/.

Davis H. "Food and Drug Administration Perspective on Regulation of Pharmaceutical Excipients." Excipient Development for Pharmaceutical, Biotechnology, and Drug Delivery Systems. 2006: 3 – 13.

"Drug Approval Process." *U.S. Food and Drug Administration.* Accessed on 20 Jan. 2017: http://www.fda.gov/downloads/drugs/resourcesforyou/consumers/ucm284393.pdf.

"Epogen." Highlights of Prescribing Information. May 2012: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/epogen/epogen_pi_hcp_english.ashx.

Farooq R. "Amgen's Drug Pipeline Fortifying A Biotech Pioneer." Forbes 29 July 2016: http://www.forbes.com/sites/bidnessetc/2016/07/29/amgens-drug-pipeline-fortifying-a-biotech-pioneer/#d5a1e091d07b.

Franey H, Brych SR, Kolvenbach CG, Rajan RS. "Increased aggregation propensity of IgG2 subclass over IgG1: Role of conformational changes and volaent character in isolated aggregates." Protein Sci. 19 September 2010; 19(9): 1601-1615.

Glover L. "Why Are Biologic Drugs So Costly?" *U.S. News* 6 Feb. 2015: http://health.usnews.com/health-news/health-wellness/articles/2015/02/06/why-are-biologic-drugs-so-costly.

Gokarn YR, Kosky A, Kras E, McAuley A, Remmele RL. "Excipients for Protein Drugs." Excipient Development for Pharmaceutical, Biotechnology, and Drug Delivery Systems. 2006: 291 – 306.

Gregory R. "Protein-Solvent Interactions." New York: Marcel Dekker, 1995. https://books.google.com/books?id=zrpTLo14l7MC.

"Guidance for Industry: Nonclinical Studies for the Safety Evaluation of Pharmaceutical Excipients." *U.S. Department of Health and Human Services* May 2005: http://www.fda.gov/ohrms/dockets/98fr/2002d-0389-gdl0002.pdf.

"Guidance for Industry: Immunogenicity Assessment for Therapeutic Protein Products." *U.S. Department of Health and Human Services* August 2014: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM338856.pdf.

Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., and Steinbeck, C. (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res (2013) 41 (D1): D456-D463. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks1146.

Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. 2014; 32: 40-51. doi:10.1038/nbt.2786.

Huey R. "How to prepare a ligand file for AutoDock 4" AutoDock. 27 August 2007: http://autodock.scripps.edu/faqs-help/how-to/how-to-prepare-a-ligand-file-for-autodock4.

Huey R, Morris G. "How to prepare a receptor file for AutoDock 4" AutoDock. 19 November 2010: http://autodock.scripps.edu/faqs-help/how-to/how-to-prepare-a-receptor-file-for-autodock4.

Hwang R, et al. "Design of Experiments for Formulation Development." PharmaTech. Dec 2005: http://www.pharmtech.com/design-experiments-formulation-development-0.

"Imlygic." Highlights of Prescribing Information. Oct. 2015: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/imlygic/imlygic_pi.ashx.

"Inactive Ingredients Database Download." U.S. Food & Drug Administration. 18 January 2017: https://www.fda.gov/Drugs/InformationOnDrugs/ucm113978.htm.

"Inactive Ingredient Search for Approved Drug Products: Frequently Asked Questions." U.S. Food & Drug Administration. 25 January 2017: https://www.fda.gov/Drugs/InformationOnDrugs/ucm080123.htm.

"Introduction to Protein Data Bank Format." UCSF Computer Graphics Laboratory. May 2014: https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/tutorials/pdbintro.html.

Irwin, Sterling, Mysinger, Bolstad and Coleman, J. "ZINC12." Chem. Inf. Model. 2012 DOI: 10.1021/ci3001277. http://zinc.docking.org/.

Kendrick BS, Carpenter JF, Cleland JL, Randolph TW. "A transient expansion of the native state precedes aggregation of recombinant human interferon-Y." Proc Natl Acad Sci USA. 24 Nov 1998; 95(24): 14142-14146.

"Kyprolis." Highlights of Prescribing Information. Nov. 2016: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/kyprolis/kyprolis_pi.ashx.

Leader B, Baca QJ, Golan DE. Protein therapeutics: a summary and pharmacological classification. 2008; 7: 21-39. doi:10.1038/nrd2399.

"Metabolites & Cofactors." Sigma-Aldrich. 2017: https://www.sigmaaldrich.com/life-science/metabolomics/learning-center/metabolic-pathways/metabolites-and-cofactors.html.

"Metabolites and Cofactors on the Metabolic Pathways Chart." Sigma-Aldrich. 2017: https://www.sigmaaldrich.com/life-science/cell-biology/cell-biology-products.html?TablePage=14576263.

Morris, GM. "AutoDock..." AutoDock. 2007: http://autodock.scripps.edu/resources/science/AutoDock/.

Morris, GM. "What is the format of a PDBQT file?" AutoDock. 11 April 2007: http://autodock.scripps.edu/faqs-help/faq/what-is-the-format-of-a-pdbqt-file

"Neulasta." Highlights of Prescribing Information. Dec 2016: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/neulasta/neulasta_pi_hcp_english.ashx.

"Neupogen." Highlights of Prescribing Information. June 2016: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/neupogen/neupogen_pi_hcp_english.ashx.

"NPLATE." Highlights of Prescribing Information. April 2016: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/nplate/nplate_pi_hcp_english.ashx.

O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchinson GR. "Open Babel: An open chemical toolbox." Journal of Cheminformatics. 7 October 2011; 3: 33. DOI: 10.1186/1758-2946-3-33.

Peters S. "Cost to Develop and Win Marketing Approval for a New Drug is $2.6 Billion." *Tufts Center for the Study of Drug Development*. 18 Nov. 2014: http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study.

Philippidis A. "Top 25 Biotech Companies of 2016." *Genetic Engineering & Biotechnology News* 26 Sept. 2016: http://www.genengnews.com/the-lists/top-25-biotech-companies-of-2016/77900741.

"Pipeline." *Amgen* 6 Oct 2016: http://www.amgenpipeline.com/~/media/amgen/full/www-amgenpipeline-com/charts/amgenpipelinedownload.ashx.

Pollack A. "F.D.A. Approves New Anemia Drug." *New York Times* 27 Mar. 2012: http://www.nytimes.com/2012/03/28/health/policy/fda-approves-new-anemia-drug.html.

"Products." Amgen 2017: http://www.amgen.com/products/.

"Product Information: Lysozyme." Sigma-Aldrich. 2017: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/7/l7651dat.pdf.

"Prolia." Highlights of Prescribing Information. Jan. 2017: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/prolia/prolia_pi.ashx.

"PubChem Compound." PubChem. 2017: https://pubchem.ncbi.nlm.nih.gov/#.

"Repatha." Highlights of Prescribing Information. July. 2016: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/repatha/repatha_pi_hcp_english.ashx.

Richardson E. "Health Policy Brief: Biosimilars." *Health Affairs* 10 Oct. 2013: http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=100.

Ross GA, Morris GM, Biggin PC. Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. PLOS ONE. 2012; 7(3): e32036. doi: 10.1371/journal.pone.0032036

Saphire EO, Parren PW, Pantophlet R, Zwick MB, Morris GM, Rudd PM, Dwek RA, Stanfield RL, Burton, DR, Wilson IA. "Crystal structure of a neutralizing human IgG against HIV-1: a template for vaccine design." Science. 2001; 293: 1155-1159. http://www.rcsb.org/pdb/explore.do?structureId=1HZH.

"Search for FDA Guidance Documents." U.S. Food & Drug Administration. 18 February 2017: https://www.fda.gov/RegulatoryInformation/Guidances/.

"Sensipar." Highlights of Prescribing Information. 2014: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/sensipar/sensipar_pi_hcp_english.ashx.

Tarar HS. "Optimum protein-excipient interactions using molecular docking simulations." Journal of Undergraduate Research (Kansas University). 2013; 69-73: https://kuscholarworks.ku.edu/bitstream/handle/1808/15039/Tarar_jur14.pdf?sequence=1.

"The Amgen Story." Amgen. 2015: http://www.amgenhistory.com.

"The Shape of Drugs to Come." *Amgen Science* 27 July 2015: https://www.amgenscience.com/the-shape-of-drugs-to-come/.

Tobin PH, Richards DH, Callender RA, Wilson CJ. "Protein Engineering: A New Frontier for Biological Therapeutics." Curr Drug Metab. 2014; 15(7): 743-756.

Trott O, Olson AJ. "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading." J Comput Chem. 30 January 2010; 31(2): 455-461.

"Unilife and Amgen Enter Strategic Collaboration for Injectable Drug Delivery Systems." PR Newswire. 22 Feb 2016: http://www.prnewswire.com/news-releases/unilife-and-amgen-enter-strategic-collaboration-for-injectable-drug-delivery-systems-300223649.html.

"Vectibix." Highlights of Prescribing Information. 2015: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/vectibix/vectibix_pi.ashx.

Verma AS, Agrahari S, Rastogi S, Singh A. "Biotechnology in the Realm of History." *Journal of Pharmacy and Bioallied Sciences.* 2011;3(3):321-323. doi:10.4103/0975-7406.84430.

Wang W, Nema S, Teagarden D. "Aggregation – Pathways and influencing factors." International Journal of Pharmaceutics. 2010; 390: 89-99.

Wang W, Roberts CJ. "Aggregation of Therapeutic Proteins." Wiley. 2010.

Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, et al., HMDB 3.0 — The Human Metabolome Database in 2013. Nucleic Acids Res. 2013. Jan 1;41(D1):D801-7. 23161693. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531200/.

"Xgeva." Highlights of Prescribing Information. 2016: http://pi.amgen.com/~/media/amgen/repositorysites/pi-amgen-com/xgeva/xgeva_pi.ashx.

This page is intentionally left blank.

Appendix

Exhibit A: Metabolites from Sigma-Aldrich

(±)-Mevalonolactone ~97 % (titration)	D-Mannose 6-phosphate sodium salt ≥95% (enzymatic)	L-Tyrosine reagent grade ³98 % (TLC)
2-Butenoyl coenzyme A lithium salt ≥90% (HPLC)	D-Ribulose 1,5-bisphosphate sodium salt hydrate ~90 %	L-Valine reagent grade ³98 % (TLC)
2-Ketobutyric acid 99 %	D-Ribulose 5-phosphate sodium salt ~90 %	L-Xylulose ~95 % syrup
2,3-Diphospho-D-glyceric acid pentasodium salt	D-Sorbitol ≥98%	L(+)-Lactic acid ~98 %
2′-Deoxyadenosine 5′-di-phosphate sodium salt	D-Sphingosine from bovine brain ~99 % (TLC)	Lanosterol ≥93%, powder
2′-Deoxycytidine 5′-diphosphate sodium salt ≥96%	D-Xylulose	Lauroyl coenzyme A lithium salt 90-95 %
2′-Deoxycytidine 5′-monophosphate sodium salt Sigma Grade, ≥98%	Decanoyl coenzyme A monohydrate ³90 %	Leukotriene B4 ~100 µg/mL in ethanol, ≥97%
2′-Deoxycytidine 5′-triphosphate disodium salt ≥95%	Desmosterol ≥84% (GC)	Linoleic acid ≥99%
2′-Deoxyguanosine 5′-triphosphate sodium salt hydrate ≥96% (HPLC)	Dihydrouracil	Linoleoyl coenzyme A lithium salt ≥85%
2′-Deoxyuridine 5′-monophosphate disodium salt Sigma Grade	Dihydroxyacetone phosphate dilithium salt ≥93% (enzymatic)	Lithium acetoacetate ≥90% (enzymatic)
3-Hydroxy-DL-kynurenine	Dihydroxyacetone phosphate hemimagnesium salt hydrate ≥95% (TLC)	Lithium carbamoylphosphate dibasic hydrate ≥85%
3-Hydroxyanthranilic acid	Dimethylallyl pyrophosphate triammonium salt	Malonyl coenzyme A lithium salt ≥90% (HPLC)
3′-Dephosphocoenzyme A ≥90% (HPLC)	DL-3-Hydroxy-3-methylglutaryl coenzyme A sodium salt hydrate ≥90% (HPLC)	Methylmalonyl coenzyme A tetralithium salt hexahydrate 90-95 %
4-Hydroxyphenylpyruvic acid 98 %	DL-b-Aminoisobutyric acid	Methylmalonyl coenzyme A tetralithium salt hydrate ≥90% (HPLC)
4-Imidazoleacrylic acid ³99 %	DL-b-Hydroxybutyryl	myo-Inositol ≥99%

	coenzyme A lithium salt ≥90 %	
5-Aminoimidazole-4-carboxamide-1-b-D-ribofuranosyl 5'-monophosphate ~95 %	DL-Glyceraldehyde ≥90% (GC)	N-Acetyl-D-mannosamine
5-Aminolevulinic acid hydrochloride ~98 %	DL-Glyceraldehyde 3-phosphate diethyl acetal barium salt	N-Acetylneuraminic acid from Escherichia coli Type VI ≥98 %
5-Phospho-D-ribose 1-diphosphate pentasodium salt	DL-Homocysteine ≥95% (titration)	n-Heptadecanoyl coenzyme A lithium salt ≥90%
5-Pregnen-3b-ol-20-one ≥98 %	DL-Isocitric acid trisodium salt hydrate ≥93%	n-Propionyl coenzyme A lithium salt ≥85%
a-D-Galactose 1-phosphate dipotassium salt pentahydrate Type II ≥98 %	Ethanolamine ≥98%	N,N-Dimethylglycine ≥99%
a-D(+)Mannose 1-phosphate sodium salt Sigma Grade	Farnesyl pyrophosphate ammonium salt methanol:ammonia solution, ≥95% (TLC)	Ne,Ne,Ne-Trimethyllysine ≥97 % (TLC)
Acetoacetyl coenzyme A sodium salt hydrate 90-95 %	g-Aminobutyric acid ≥99 %	O-Acetyl-L-carnitine hydrochloride ≥99% (titration), powder
Acetyl coenzyme A sodium salt ~95 % powder	g-Linolenic acid ≥99 % liquid	O-Acetyl-L-serine hydrochloride
Acetyl coenzyme A sodium salt ≥93% (HPLC), powder	Geranyl pyrophosphate ammonium salt 1 mg/mL in methanol (:aqueous 10 mM NH4OH (7:3)), ≥95% (TLC)	O-Phospho-L-serine
Acetyl coenzyme A trilithium salt ~95 %	Glycerol ReagentPlus®, ≥99.0% (GC)	O-Phosphorylethanolamine
Acetylcholine chloride ≥99% (TLC)	Glycine ReagentPlus®, ≥99% (HPLC)	Octanoyl coenzyme A lithium salt monohydrate 90-95 %
Adenine hydrochloride hydrate ≥99%	Glycocyamine	Oleoyl coenzyme A lithium salt ≥90% (HPLC)
Adenosine 3',5'-cyclic monophosphate sodium salt monohydrate ~98 % (HPLC) powder	Glycolic acid BioXtra, ≥98.0% (titration)	Orotic acid ≥98% (titration), anhydrous
Adenosine 3'-phosphate 5'-phosphosulfate lithium salt hydrate ≥60%	Glyoxylic acid solution 50 % (w/w) in water	Orotidine 5'-monophosphate trisodium salt ≥99% (HPLC), powder

Adenosine 5'-diphosphate sodium salt bacterial 95-99 %	Guanine Sigma Grade crystalline ³99 % as anhydrous (HPLC)	Oxalic acid dihydrate ReagentPlus®, ≥99.0% (GC)
Adenosine 5'-triphosphate disodium salt Grade I ³99 %	Guanosine 5′-diphosphate sodium salt Type I, ≥96% (HPLC)	Oxaloacetic acid ~98 %
Adenosine 5′-diphosphate sodium salt bacterial, ≥95% (HPLC)	Guanosine 5′-diphospho-β-L-fucose sodium salt ≥85%	p-Coumaric acid
Adenosine 5′-monophosphate sodium salt from yeast, ≥99%	Guanosine 5′-diphosphoglucose sodium salt	Palmitoleoyl coenzyme A lithium salt ~90%
Adenosine 5′-phosphosulfate sodium salt ≥85%	Guanosine 5′-monophosphate disodium salt hydrate from yeast, ≥99%	Palmitoyl coenzyme A lithium salt ³90 %
Adenosine 5′-triphosphate disodium salt hydrate Grade I, ≥99%, from microbial	Guanosine 5′-triphosphate sodium salt hydrate ≥95% (HPLC), powder	Phospho(enol)pyruvic acid trisodium salt hydrate ≥97% (enzymatic)
Adenosine-5′-diphosphoglucose disodium salt ≥93%	Hemin BioXtra, from Porcine, ≥97.0% (HPLC)	Phosphocholine chloride calcium salt tetrahydrate Sigma Grade
Adonitol ≥99%	Hexanoyl coenzyme A trilithium salt trihydrate ³85 %	Phosphocreatine disodium salt hydrate ≥97%
Allantoin	Histamine dihydrochloride ≥99% (TLC), powder	Porphobilinogen powder
Anthranilic acid reagent grade ³98 %	Homogentisic acid crystalline	Progesterone ≥99%
Arachidonic acid from porcine liver ~99 % (capillary GC) oil	Hypotaurine ≥98% (TLC)	Prostaglandin E2 ≥93% (HPLC), synthetic
Arachidonoyl coenzyme A lithium salt ³85 %	Indole-3-acetic acid sodium salt BioReagent, plant cell culture tested, ≥98%	Psychosine from bovine brain lyophilized powder
Argininosuccinic acid disodium salt hydrate ~80%	Indole-3-pyruvic acid	Putrescine dihydrochloride ≥98% (TLC)
b-Alanine cell culture tested insect cell culture tested	Inosine ≥99% (HPLC)	Pyrocatechol ≥99%
b-Methylcrotonyl coenzyme A lithium salt ³90 %	Inosine 5′-monophosphate disodium salt hydrate from muscle, Sigma Grade, 99-100%	S-(5′-Adenosyl)-L-homocysteine crystalline
b-Nicotinamide adenine	Inulin from chicory	S-(5′-Adenosyl)-L-methionine

dinucleotide 2'-phosphate reduced tetrasodium salt ≥98 %		p-toluenesulfonate salt from yeast (L-methionine enriched), ≥80% (HPLC), ≥80% (spectrophotometric assay)
b-Nicotinamide adenine dinucleotide hydrate ≥99 %	Isobutyryl coenzyme A lithium salt ≥85%	Shikimic acid ≥99%
b-Nicotinamide adenine dinucleotide phosphate sodium salt ≥98 %	Isopentenyl pyrophosphate ammonium salt solution 1 mg/mL ≥95 % (TLC)	sn-Glycero-3-phosphocholine inner salt from egg yolk 5 mg/mL in methanol ~98 %
b-Nicotinamide adenine dinucleotide, reduced disodium salt hydrate ≥98 %	Isovaleryl coenzyme A lithium salt hydrate ≥90%	Sodium phenylpyruvate powder
Benzoyl coenzyme A lithium salt ≥90%	L-(-)-Malic acid 95-100 % (enzymatic)	Sodium pyruvate anhydrous, free-flowing, Redi-Dri™, ReagentPlus®, ≥99%
Betaine	L-(+)-Arabinose ≥99%	Sodium pyruvate ReagentPlus®, ≥99%
Betaine aldehyde chloride	L-(−)-Arabitol ≥98%	Sodium succinate dibasic hexahydrate ReagentPlus®, ≥99%
Butyryl coenzyme A dilithium salt hydrate 90-95 %	L-a-Lysophosphatidylcholine from bovine brain ~99 % Type V	Spermidine trihydrochloride ≥98% (TLC)
Cardiolipin solution bovine heart ~98 % (TLC) solution	L-Alanine ≥98% (TLC)	Spermine tetrahydrochloride
Chitin crab shells practical grade coarse flakes	L-Arginine reagent grade ≥98 % (TLC) powder	Sphingomyelin from bovine brain ≥98 % powder
Cholesterol Sigma Grade ≥99%	L-Asparagine ≥98% (HPLC)	Squalene ≥98%, liquid
Choline chloride ≥98%	L-Aspartic acid reagent grade ≥98 % (TLC)	Starch from wheat Unmodified
Choline chloride ≥99%	L-Carnitine hydrochloride from synthetic ~98 %	Succinyl coenzyme A sodium salt ≥85%
Chondroitin sulfate sodium salt bovine cartilage standard (for CPC (cetylpyridinium chloride) titration)	L-Carnosine ~99 % crystalline	Sucrose ≥99.5%
Citric acid monohydrate reagent grade ≥98 % (GC/titration)	L-Citrulline ≥98% (TLC)	Taurine ≥99%
Coenzyme A hydrate ≥85%	L-Cystathionine ≥98% (TLC)	Thymidine 5'-

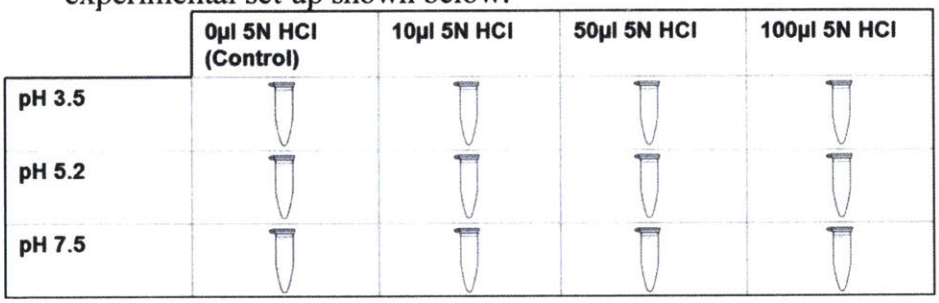
(UV, HPLC)		monophosphate disodium salt hydrate ≥99%
Coenzyme A sodium salt hydrate cofactor for acyl transfer	L-Cysteine from non-animal source, BioReagent, suitable for cell culture, ≥98%	Thymidine 5′-triphosphate sodium salt ≥96%
Coenzyme A trilithium salt ≥93%	L-Cysteinesulfinic acid monohydrate	Thymine ≥99%
Creatine anhydrous	L-Cystine ≥98% (TLC), crystalline	trans-4-Hydroxy-L-proline BioReagent, suitable for cell culture, ≥98.5%
Cytidine 5′-diphosphate sodium salt hydrate from yeast, ≥95%	L-Dihydroorotic acid ≥99%	trans-Cinnamic acid ³99 %
Cytidine 5′-diphosphocholine sodium salt dihydrate ~98%, from yeast, solid	L-Glutamic acid monosodium salt hydrate ≥99% (HPLC), powder	Tryptamine ³99 % crystalline
Cytidine 5′-triphosphate disodium salt ≥95%	L-Glutamine ReagentPlus®, ≥99% (HPLC)	Tyramine hydrochloride ≥98%
Cytidine-5′-monophospho-N-acetylneuraminic acid sodium salt ≥85% (HPLC)	L-Glutathione reduced ≥98.0%	Uracil ≥99.0%
D-(-)-3-Phosphoglyceric acid disodium salt ~95 % powder	L-Histidine ReagentPlus®, ≥99% (TLC)	Urea ReagentPlus®, ≥99.5%, pellets
D-(+)-Cellobiose	L-Histidinol dihydrochloride ≥98 (TLC)	Uric acid ≥99%, crystalline
D-(+)-Galactose ≥99%	L-Homoserine	Uridine 5′-(trihydrogen diphosphate) sodium salt from Saccharomyces cerevisiae 95-100%
D-(+)-Glucose ≥99.5% (GC)	L-Isoleucine reagent grade ³98 % (TLC)	Uridine 5′-diphospho-N-acetylgalactosamine disodium salt ≥98%
D-(+)-Mannose	L-Kynurenine ≥98% (HPLC)	Uridine 5′-diphospho-N-acetylglucosamine sodium salt ≥98%
D-(+)-Xylose ≥99%	L-Leucine reagent grade ³98 % (TLC)	Uridine 5′-diphosphogalactose disodium salt ≥97.0%
D-(−)-Arabinose ≥98%	L-Lysine ≥98% (TLC)	Uridine 5′-diphosphoglucose disodium salt hydrate from Saccharomyces cerevisiae ≥98%
D-(−)-Fructose ≥99%	L-Methionine reagent	Uridine 5′-

	grade ³98 % (TLC)	diphosphoglucuronic acid trisodium salt 98-100%
D-(−)-Ribose ≥99%	L-Ornithine monohydrochloride ~99 %	Uridine 5′-monophosphate disodium salt ≥99%
D-Fructose 1,6-bisphosphate sodium salt hydrate ³70 %	L-Phenylalanine reagent grade ³98 %	Uridine 5′-triphosphate trisodium salt hydrate from yeast, Type III, ≥96%
D-Fructose 6-phosphate disodium salt dihydrate ~98 % (enzymatic) amorphous powder	L-Proline ReagentPlus®, ≥99% (HPLC)	Xanthine ≥99.5% (HPLC), purified by recrystallization
D-Glucosamine 6-phosphate sodium salt ~98 %	L-Serine ReagentPlus®, ≥99% (HPLC)	Xylitol ≥99%
D-Glucose 6-phosphate sodium salt Sigma Grade crystalline	L-Threonine reagent grade ³98 % (TLC)	α-D-Glucose 1-phosphate disodium salt hydrate ≥97%
D-Glucuronic acid sodium salt monohydrate 97.5-102.5% (non-aqueous titration)	L-Tryptophan reagent grade ³98 % (TLC)	α-Ketoglutaric acid disodium salt hydrate analytical standard, ≥95%
D-Gulonic acid g-lactone 99 %		α-Lactose monohydrate

Exhibit B: Hydrochloric Acid (HCl) – ASA2 Test

Methods:

- Aliquot 1ml of buffer to four different Eppendorf tubes for each pH (10mM Glutamic Acid,10mM Sodium Phosphate buffer, pH 3.5, pH 5.2, and pH 7.5)
- For each pH buffer, add either 0µl, 10µl, 50µl, or 100µl of 5N HCl to result in the experimental set up shown below:



- Titrate the samples with HCl (with 5N HCl, 1N HCl, 10N NaOH, 1N NaOH, as needed) so that they are back to their target pH of 3.5, 5.2, or 7.5
- Populate a 96-well plate with 125µl of ASA2 at the respective pH and 125µl of each of the samples shown above
- Final plate design*:

	1	2	3	4	5	6	7	8	9	10	11	12
A	pH 3.5 - Control (0µl HCl)	pH 3.5 - 1.69µl HCl	pH 3.5 - 6.25µl HCl	pH 3.5 - 14µl HCl	pH 5.2 - Control (0µl HCl)	pH 5.2 - 1.25µl HCl	pH 5.2 - 6.88µl HCl	pH 5.2 - 12.50µl HCl	pH 7.5 - Control (0µl HCl)	pH 7.5 - 1.25µl HCl	pH 7.5 - 6.88µl HCl	pH 7.5 - 13.75µl HCl

*The volume of HCl shown in the plate for each sample accounts for any additional 5N HCl added during titration and dilution from ASA2

- Perform DSF, DLS, and SEC on samples at T0 and at week 1, after one week of incubation in 40C

Results:

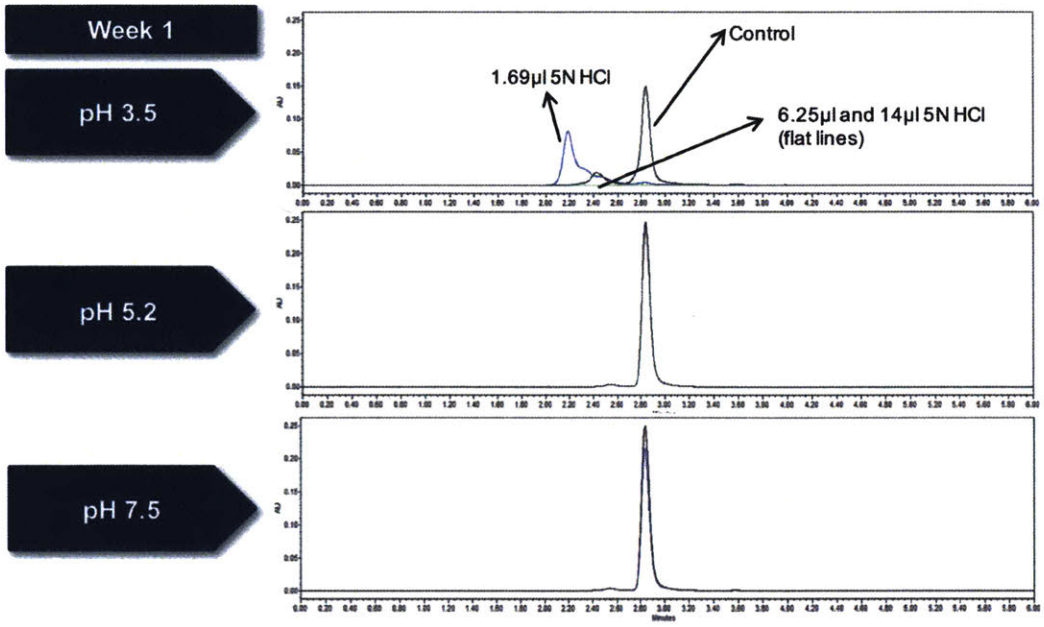
Overview

	DLS		DSF		SEC	
	Radius (nm)		Melting Temperature (°C)		% of Aggregates	
	T0	1wk	T0	1wk	T0*	1wk
pH 3.5 - Control (0µl HCl)	4.4	4.5	45.0	47.0	2.2	17.8
pH 3.5 - 1.69µl HCl	5.0	1.7	39.8	None	2.9	93.9
pH 3.5 - 6.25µl HCl	5.6	1.3	36.2	None	11.3	88.8**
pH 3.5 - 14µl HCl	6.5	179.4	35.8	None	20.9	79.1**
pH 5.2 - Control (0µl HCl)	5.1	4.9	62.8	62.0	1.9	2.3
pH 5.2 - 1.25µl HCl	5.2	5.2	62.8	62.2	2.0	2.3
pH 5.2 - 6.88µl HCl	5.1	5.3	60.2	60.6	2.0	2.3
pH 5.2 - 12.50µl HCl	5.1	4.9	59.8	60.0	2.0	2.3
pH 7.5 - Control (0µl HCl)	5.3	5.3	66.4	66.8	2.4	2.5
pH 7.5 - 1.25µl HCl	5.2	5.1	66.6	66.4	2.3	2.5
pH 7.5 - 6.88µl HCl	5.3	5.2	65.6	65.6	2.3	2.5
pH 7.5 - 13.75µl HCl	5.4	5.5	64.6	65.0	2.2	2.5

*T0 for SEC was actually measure on day 4; Initial T0 run was on a bad column so sample was refrigerated for 3 days until a new column arrived and was measured on day 4 on the new column

**ASA2 was completed destroyed in these samples; % shown is indicative of the high level of degradation and not exact for aggregation

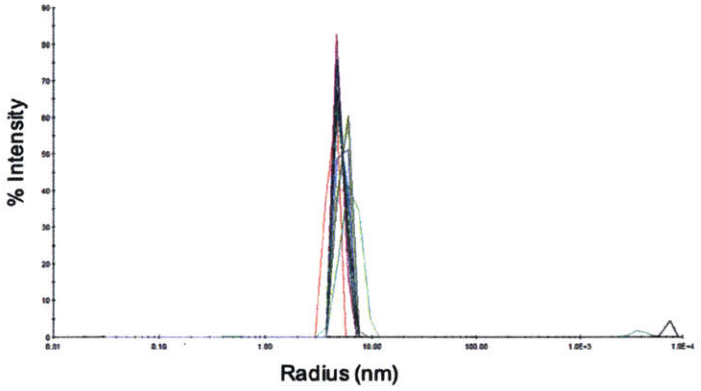
SEC Results



	% Aggregates
pH 3.5 - Control (0µl HCl)	17.8
pH 3.5 - 1.69µl HCl	93.9
pH 3.5 - 6.25µl HCl	--
pH 3.5 - 14µl HCl	--
pH 5.2 - Control (0µl HCl)	2.3
pH 5.2 - 1.25µl HCl	2.3
pH 5.2 - 6.88µl HCl	2.3
pH 5.2 - 12.50µl HCl	2.3
pH 7.5 - Control (0µl HCl)	2.5
pH 7.5 - 1.25µl HCl	2.5
pH 7.5 - 6.88µl HCl	2.5
pH 7.5 - 13.75µl HCl	2.5

DLS Results

Distribution of ASA2's Hydrodynamic Radius at T0



Distribution of ASA2's Hydrodynamic Radius at Week 1

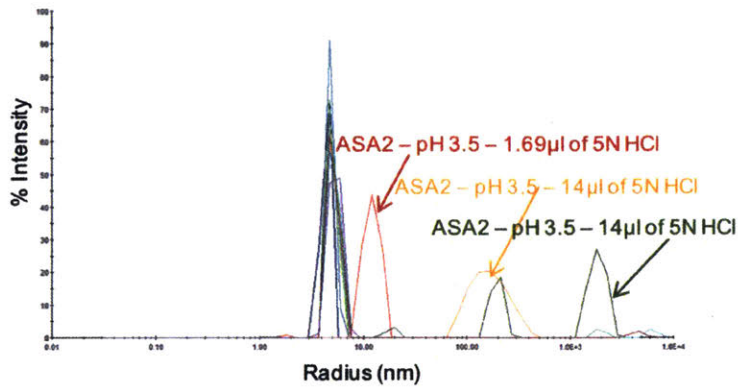


Plate Design

	1	2	3	4	5	6	7	8	9	10	11	12
A	pH 3.5 (Control) 0µl 5N HCl	pH 3.5 1.69µl 5N HCl	pH 3.5 6.25µl 5N HCl	pH 3.5 14µl 5N HCl	pH 5.2 (Control) 0µl 5N HCl	pH 5.2 1.25µl 5N HCl	pH 5.2 6.88µl 5N HCl	pH 5.2 12.50µl 5N HCl	pH 7.5 (Control) 0µl 5N HCl	pH 7.5 1.25µl 5N HCl	pH 7.5 6.88µl 5N HCl	pH 7.5 13.75µl 5N HCl

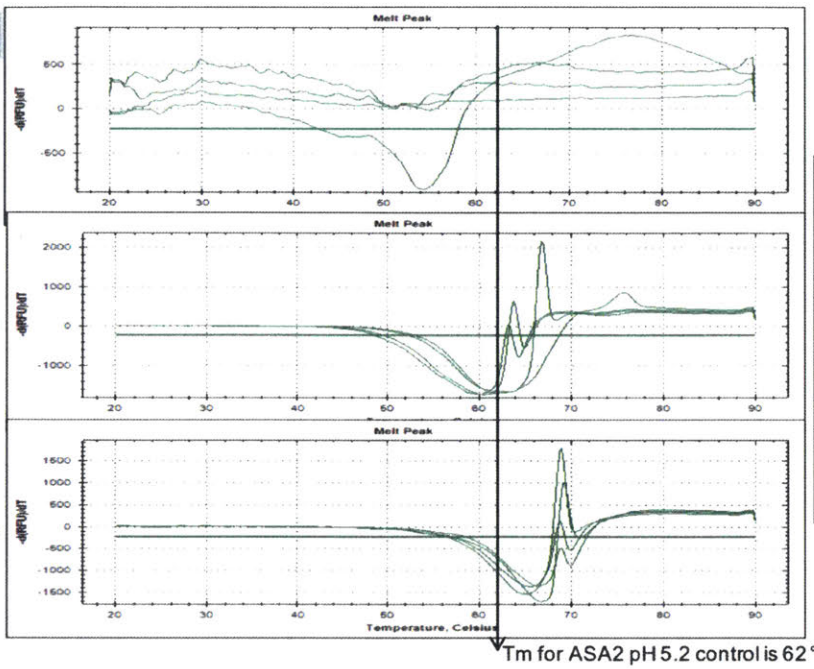
DSF Results

Week 1

pH 3.5

pH 5.2

pH 7.5



Tm for ASA2 pH 5.2 control is 62 °C

	Melting Temperature (°C) - Week 1
pH 3.5 - Control (0µl HCl)	47.0
pH 3.5 - 1.69µl HCl	None
pH 3.5 - 6.25µl HCl	None
pH 3.5 - 14µl HCl	None
pH 5.2 - Control (0µl HCl)	62.0
pH 5.2 - 1.25µl HCl	62.2
pH 5.2 - 6.88µl HCl	60.6
pH 5.2 - 12.50µl HCl	60.0
pH 7.5 - Control (0µl HCl)	66.8
pH 7.5 - 1.25µl HCl	66.4
pH 7.5 - 6.88µl HCl	65.6
pH 7.5 - 13.75µl HCl	65.0