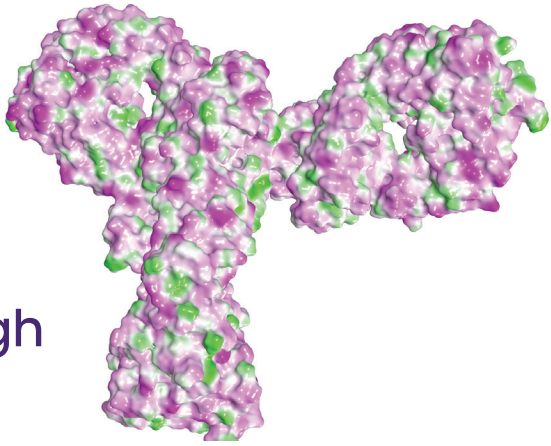
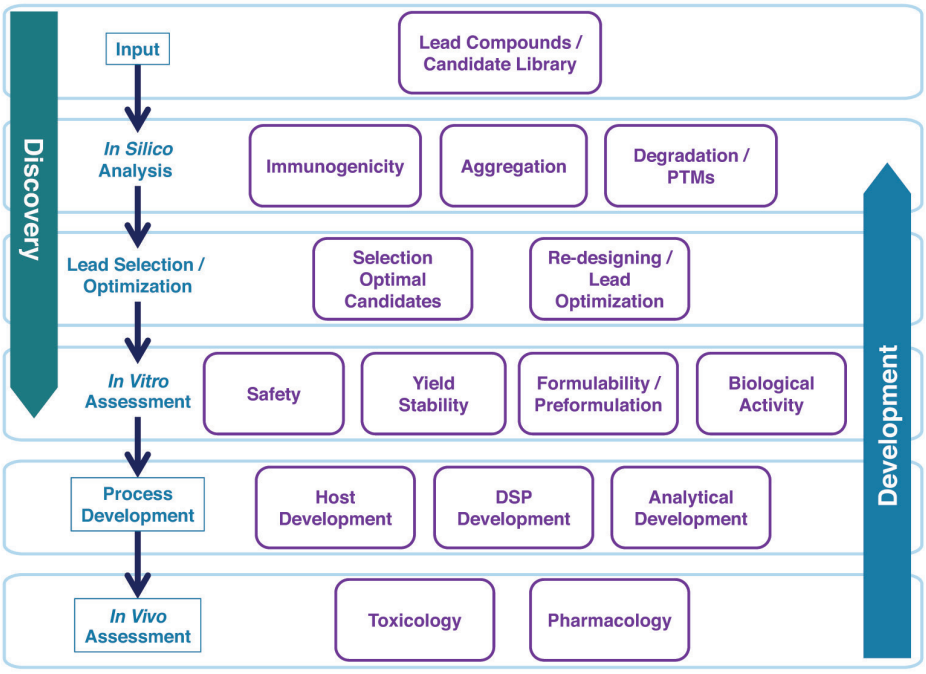
DEVELOPABILITY OF BIOTHERAPEUTICS

Computational Approaches



Edited by
Sandeep Kumar
Satish Kumar Singh



CRC Press
Taylor & Francis Group

DEVELOPABILITY OF BIOTHERAPEUTICS

Computational Approaches

DEVELOPABILITY OF BIOTHERAPEUTICS

Computational Approaches

Edited by
Sandeep Kumar
Biotherapeutics Pharmaceutical Sciences
Pfizer, Inc.

Satish Kumar Singh
Biotherapeutics Pharmaceutical Sciences
Pfizer, Inc.

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an informa business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20151120

International Standard Book Number-13: 978-1-4822-4616-5 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

SECTION II Developability Practices in the Biopharmaceutical Industry

Foreword

This book is about developability. But what a funny term *developability* is. A neologism, it can only be a transitional term, for, like its cousin, *manufacturability*, it implies that issues of biopharmaceutical drug development (and by analogy issues of manufacturing) should be taken into account during the discovery phase. It implies the need for achieving the "seamless interdependent whole" discussed in the first chapter, whereas at the same time acknowledging that we are still so far from it that we need to force ourselves to take into account parts of the process of creating new biopharmaceuticals while we pursue other parts without fully accounting for their interdependence.

Why are we in this situation? The answer was given by Francis Bacon more than 400 years ago. Before we give the answer, a little background is in order. Bacon is the thinker who transformed human aspirations from "abstract meditation" to "fruit and works." In doing so, he invented the experimental method with the major purpose of improving human health and extending life. Shortly thereafter, René Descartes built upon Bacon's method so "that we could be spared an infinity of diseases, of the body as well as of the mind." In order to accomplish this, he introduced the wholesale application of mathematics into science. It is only in the past few decades, however, that we have obtained the tools for applying the Baconian–Cartesian method to the creation of new pharmaceuticals. Even with these tools, we have been slow to apply them, to bring about holistic pharmaceutical creation.

Bacon tells us why. In discussing his famous idols of the mind, which run deep and hold us back from progress, he writes, "The human intellect, from its own character, easily supposes that there is more order and regularity in things than it finds." Do we really think that focusing on the properties that make a molecule bind to its target will absolve us from designing that molecule to do the other things that we wish it to do? Do we really think that our focusing on design of binding to a target will lead to the molecule interacting as we wish it to with other parts of the body, or having the stability to be made into a pharmaceutical product, or having suitable properties for manufacturing? Clearly, we do not think there is such regularity, but our minds lead us to believe that these issues will sort themselves out, as if there were such regularity.

This volume addresses the disparity between what we know is true, on the one hand, and how we act, on the other. It shows us how, in the field of biopharmaceutical creation, our mind's idols can be overcome. The idea is that if we learn about the wide-ranging and effective technical solutions available today, we will be motivated to apply them, and therein bring about the regularity that will not happen of its own accord. The editors and contributors of this volume discuss solutions that run the gamut, from chemical and physical stability to pharmacokinetics/pharmacodynamics (PK/PD) analysis and prediction to supply chain issues, immunogenicity, formulation, modality selection, polydispersity, effects of post-translational modifications, upstream and downstream production, and even epitope predictions. They show us that the Baconian–Cartesian tools that have formed the basis of the tremendously successful mathematical physics over the last several hundred years have now been

extended to the realm of complex biological systems and, as such, have the potential to enhance human health in a hitherto unprecedented way.

Two key concepts form the twofold theme: informational and knowledge based. We need accurate information for our models to be able to describe reality, and we have reached the point where, while we can never have enough information, we have enough to make much faster progress. Moreover, we need to base our models on knowledge, not merely facts or data, but mechanistic understanding, as close to first principles as possible. We may not be at the point of having the ultimate mechanistic understanding of these complex systems, but we are much further along than what our current approach to creating biopharmaceuticals presumes. The editors go even deeper still in identifying the key aspects that need to be addressed: the knowledge-based approach is not just about mechanistic understanding but is "comprised of human intentions and culture as much as it is of tools and technologies." Bacon also said: "The human intellect swells and cannot stay still or rest, but aspires to go further, in vain." If we would only embrace the approach described here, not just the tools, but also the mindset, what we could accomplish may indeed be limitless.

In the Preface, the editors discuss the aim of remaining in Well Country as opposed to being forced to visit Sick City. Such metaphors emphasize the universalism of what we are trying to accomplish, for after all, biopharmaceuticals that can help anyone who needs them potentially help everyone. They also remind us of what we are trying to accomplish by creating new biopharmaceuticals. The Greek word for happy is *eudaimon*—possessing a "well" spirit. A necessary condition to being *eudaimon* is possessing a well body, in order words, being healthy. Applying the knowledge in this volume will make us healthier and therein happier.

Bernhardt L. Trout
Raymond F. Baddour, ScD (1949)
Professor of Chemical Engineering, MIT

Preface

Dear Reader,

If you consider that access to modern innovative medicines is a right of all patients and are interested in understanding how recent advances in computational sciences can help in this effort, then this is the book for you! A major portion of the developmental cost of new drugs is incurred during clinical trials, whose outcomes depend, in part, on the choices made during the discovery and selection (design) of the drug molecule, formulation, manufacturing, dosing, clinical trial design, and patient population selection. As you flip through the pages of this book, the use of computation in novel ways to improve the overall process of drug discovery and development will become evident. This book is focused mainly on developability of monoclonal antibody candidates and is organized in two sections. The first section describes applications of computational approaches toward discovery and development of biopharmaceutical drugs; the second section presents the best practices in developability assessments of early-stage drug candidates being followed by leading companies in this business.

Although we prefer to live in Well Country forever, forced visits to Sick City do happen every now and then. Advances in medicines have significantly improved the health and well-being of millions of people, particularly in the last century. The explosive growth of knowledge in biology and genetics has been driving this effort. However, the success rate for novel therapeutic entities is falling, raising their cost, especially for innovative biotherapeutics and vaccines. The consequent effects on the organizations involved in drug discovery and development; the medical systems, such as clinics, hospitals, insurance companies, and government-sponsored healthcare; and ultimately the individual patients and society at large are enormous.

Each year the research labs in pharmaceutical companies discover several promising novel drug candidates, but only a few of these newly discovered compounds reach the clinic after several years. Projects that succeed must recover costs for their own development as well as of the failed ones, and make reasonable profits from sales over the remaining duration of their exclusivity and after the loss of exclusivity, so that the business of bringing new medicines to patients can be sustained. This already low success rate, however, continues to fall and points to the risks and inefficiencies inherent to drug discovery, development, testing, and approval processes. While biological activity, rightly, is a major focus during the discovery of biopharmaceutical candidates, the macromolecular sequence–structural properties of these candidates can also inform us about their cell line expression levels, potential degradation routes, interactions with extractables and leachables, behaviors of highly concentrated solutions, immunogenicity, pharmacokinetics/dynamics, and so forth. Such insights, when they come via computation at early stages of lead candidate design or selection, can help make discovery and development of biologic drugs more efficient by removing empiricism and reducing developmental costs and attrition rates.

We are thankful to Hilary LaFoe for inviting us to edit this book, to Kari Budyk for coordinating the effort, and to Prof. Bernhardt Trout of MIT for providing the

Foreword. The enthusiastic contributors to the book chapters represent the vanguards of biopharmaceutical informatics performing cutting-edge research on developability issues in biopharmaceuticals. Without the generous time and effort put in by these very busy and outstanding scientists, this book would not have been possible. Discussions with numerous colleagues spread over nearly a decade are also gratefully acknowledged. It goes without saying that we could not have undertaken this journey without unwavering support, affection, and encouragement from our families and friends.

Most pharmaceutical industry executives, professionals, postdocs, students, and enthusiasts can appreciate the potential of computational approaches toward biopharmaceutical discovery and development, but awareness of the many different contributions that computation can make is generally lacking. This book provides examples and focuses on filling this void. This is still a nascent field and you, dear reader, are encouraged to explore it on your own.

S.K. and S.K.S.

Contributors

Anson K. Abraham
Pharmacokinetics, Pharmacodynamics & Drug Metabolism (PPDM)—Early Stage QP2
Merck Research Laboratories
West Point, Pennsylvania

Neeraj J. Agrawal
Amgen, Inc.
Thousand Oaks, California

Juan C. Almagro
Centers for Therapeutic Innovation (CTI)
Pfizer, Inc.
Boston, Massachusetts

Nicolas Angell
Product Attribute Sciences
Amgen, Inc.
Thousand Oaks, California

Andreas Arnell
Applied Protein Services
Lonza Biologics plc.
Great Abington, Cambridge, United Kingdom

Matthew P. Baker
Antitope Ltd. (an Abzena company)
Babraham Research Campus
Babraham, Cambridge, United Kingdom

Patrick M. Buck
Pharmaceutical Research and Development
Biotherapeutics Pharmaceutical Sciences
Pfizer, Inc.
Chesterfield, Missouri

Anuj Chaudhri
Computational Research Division
Lawrence Berkeley National Laboratory
Berkeley, California

Naresh Chennamsetty
Bristol-Myers Squibb Company
New Brunswick, New Jersey

Kristine Daris
Drug Substance Development
Amgen, Inc.
Thousand Oaks, California

Surjit B. Dixit
Zymeworks, Inc.
Vancouver, British Columbia, Canada

Thomas R. A. Gallagher
Applied Protein Services
Lonza Biologics plc.
Great Abington, Cambridge, United Kingdom

Ramón Gómez de la Cuesta
Research and Technology
Lonza Biologics plc.
Great Abington, Cambridge, United Kingdom

M. Michael Gromiha
Department of Biotechnology
Indian Institute of Technology Madras
Chennai, India

Kannan Gunasekaran
Therapeutic Discovery
Amgen, Inc.
Thousand Oaks, California

Steffen Hartmann
Integrated Biologics Profiling Unit
Biologics Technical Development and Manufacturing
Novartis Pharma AG
Basel, Switzerland

Timothy P. Hickling
Pharmacokinetics, Dynamics and Metabolism—New Biological Entities
Immunogenicity Sciences
Worldwide Research and Development
Pfizer, Inc.
Andover, Massachusetts

Tim D. Jones
Antitope Ltd. (an Abzena company)
Babraham Research Campus
Babraham, Cambridge, United Kingdom

Anette C. Karle
Integrated Biologics Profiling Unit
Biologics Technical Development and Manufacturing
Novartis Pharma AG
Basel, Switzerland

Randal R. Ketchem
Therapeutic Discovery
Amgen, Inc.
Thousand Oaks, California

Hans P. Kocher
Integrated Biologics Profiling Unit
Biologics Technical Development and Manufacturing
Novartis Pharma AG
Basel, Switzerland

Sandeep Kumar
Pharmaceutical Research and Development
Biotherapeutics Pharmaceutical Sciences
Pfizer, Inc.
Chesterfield, Missouri

Li Li
Pharmaceutical Research and Development
Biotherapeutics Pharmaceutical Sciences
Pfizer, Inc.
Andover, Massachusetts

Ben Locwin
Lonza Biologics, Inc.
Portsmouth, New Hampshire

Alessandro Mascioni
Centers for Therapeutic Innovation (CTI)
Pfizer, Inc.
Boston, Massachusetts

Olga Obrezanova
Applied Protein Services
Lonza Biologics plc.
Great Abington, Cambridge, United Kingdom

Jason W. O'Neill
Therapeutic Discovery
Amgen, Inc.
Thousand Oaks, California

Russell H. Robins
Analytical Research and Development
Biotherapeutics Pharmaceutical Sciences
Pfizer, Inc.
Chesterfield, Missouri

Pratap Singh
Pharmacokinetics, Dynamics and Metabolism—New Biological Entities
Translational Modeling and Simulation (TMS)
Worldwide Research and Development
Pfizer, Inc.
Cambridge, Massachusetts

Satish K. Singh
Pharmaceutical Research and Development
Biotherapeutics Pharmaceutical Sciences
Pfizer, Inc.
Chesterfield, Missouri

Noel Smith
Applied Protein Services
Lonza Biologics plc.
Great Abington, Cambridge, United Kingdom

A. Mary Thangakani
Department of Crystallography and Biophysics
University of Madras
Chennai, India

Abhinav Tiwari
Pharmacokinetics, Dynamics and Metabolism—New Biological Entities
Translational Modeling and Simulation (TMS)
Worldwide Research and Development
Pfizer, Inc.
Cambridge, Massachusetts

Bernhardt L. Trout
Department of Chemical Engineering,
Massachusetts Institute of Technology
Cambridge, Massachusetts

Jesús Zurdo
Research and Technology
Lonza Biologics plc.
Great Abington, Cambridge, United Kingdom

Anup Zutshi
Pharmacokinetics, Dynamics and Metabolism—New Biological Entities
Translational Modeling and Simulation (TMS)
Worldwide Research and Development
Pfizer, Inc.
Cambridge, Massachusetts

Motivation for This Book

Biopharmaceuticals, particularly monoclonal antibodies and antibody-based biotherapeutics, have emerged as best-selling medicines in recent years, thereby delivering on promises from the early days of biotechnology. However, these innovative medicines are costly to develop and produce on a commercial scale. The biologics possess heterogeneous molecular structures and are vulnerable to physical and chemical degradation, such as aggregation, oxidation, deamidation, and fragmentation, because of stresses encountered by these macromolecules during manufacturing, shipping, and storage. At the same time, a drop in the number of novel small-molecule drugs being discovered and the failure of candidates during late-stage drug development (clinical trials) have led to unprecedented highs in the cost of bringing new medicines to use. These concomitant developments are among the major drivers for the rising costs of healthcare in the United States, Europe, emerging markets, and elsewhere in the world. At the same time, demand for innovative medicines is rapidly growing in both the developed and developing world. This is especially true for life-threatening diseases such as cancer, cardiac failure, and chronic diseases such as diabetes. Long considered to be the bane of the developed world, these diseases have now emerged as a major challenge to human health in the emerging markets and other developing countries as well. Therefore, several conflicting issues are being faced today by the pharmaceutical industry and by society at large in regard to continued access to new medicines. The most pressing challenges are sustainability of business via realization of costs associated with bringing new drugs to clinic and fulfillment of pharmaceutical companies' profit expectations, and the ability of payers, for example, governments, hospitals, insurance companies, individual patients, and their families, to afford these advanced medicines.

Innovative medicines do not have to be costly. Certainly, no patient should die just because the cost of modern life-saving medicines is beyond his or her reach. The high prices of innovative medicines, particularly anticancer biotherapeutics, have become a major issue in emerging markets and other developing as well as developed countries, and are proving to be a barrier to their widespread use, notwithstanding huge demand. Yet, much can be done to reduce the cost of biopharmaceutical drugs. *This is a winnable war.* Attention should be paid to details of protein sequence and structure during the development of biopharmaceuticals, and considerations regarding manufacturability, formulation development, flexible delivery options, and immunogenicity should be included, alongside potency and efficacy, at the early stages of lead candidate discovery and design. The development and use of appropriate computational biophysics techniques such as multiscale molecular modeling, dynamic simulations, and prediction can help de-risk the drug development pipelines at preclinical stages. Similarly, statistically robust planning and execution of clinical trials, including the design of appropriate bioassays and careful selection of patient groups that are most likely to positively respond to a candidate medicine, can go a long way toward preventing costly late-stage drug failures.

Over the past couple of years, an increasing number of pharmaceutical companies have begun to embrace the umbrella concept of developability. At its core,

developability is a risk assessment and mitigation exercise that seeks to improve the likelihood that a biotherapeutic drug candidate discovered to be efficacious against a target will be successfully developed into a medicine available in clinics. The major regulatory agencies, such as the U.S. Food and Drug Administration (FDA) and European Medicines Evaluation Agency (EMEA), are also emphasizing quality by design (QbD) approaches to improve biopharmaceuticals. Several computational methodologies and biophysical techniques are being adapted, and examples of successful designs are beginning to emerge. Simultaneously, there is a growing awareness of improving clinical trial designs using statistical and mathematical modeling to prevent unnecessary drug candidate failures. However, this is still a very nascent field, and few books dealing with this subject are currently available in the market. As far as we are aware, there is currently no book that details the applications of computational and molecular modeling techniques toward biopharmaceutical drug development at preclinical stages. Yet, the importance of this subject and the need to increase awareness of industry leaders, regulators, clinicians, and the general scientifically interested public of matters related to biopharmaceutical drug development cannot be overstated. Beside these, there is also a need to train next-generation pharmaceutical, biophysical, and medical scientists in emergent issues facing the drug industry and on how innovative uses of computation can lead to highly efficacious, affordable, and safer medicines that are also convenient to use.

PURPOSE OF THE BOOK

This book serves as a primary reference and textbook for computational applications addressing the issues in biopharmaceutical development. The targeted audience for this book is pharmacy, medicine, and life science students and educators at the tertiary level, industrial research and development mid- and senior-level management, regulatory agencies, and scientists concerned with public health issues. Biopharmaceutical drug discovery and formulation professionals, as well as scientists interested in bioinformatics and computational biophysics, may also find this book of interest.

It is expected that this book will be bought by libraries supporting the schools of pharmacy, medicine, and life sciences in the United States and international universities and by companies invested in biopharmaceutical research and development. It is also hoped that this book raises awareness about the promise of computational research among pharmaceutical scientists and becomes a catalyst for innovative applications of computational design to biopharmaceutical drug development and delivery.

DESCRIPTION OF CONTENTS

Section I: Principles of Biopharmaceutical Informatics

Chapter 1: Biopharmaceutical Informatics: Applications of Computation in Biologic Drug Development

Chapter 1 defines the term *biopharmaceutical informatics* and describes the applications of computational biophysics toward understanding the challenges encountered during biopharmaceutical drug development.

Chapter 2: Computational Methods in the Optimization of Biologic Modalities

Once a target has been validated as druggable, there are a number of small-molecule and biologic-based modalities that can potentially be used. The process by which a molecular modality is optimized has a bearing on the overall success of the program. Therefore, this chapter reviews the considerations involved in the optimization of biologic modalities at the early stages of drug discovery.

Chapter 3: Understanding, Predicting, and Mitigating the Impact of Post-Translational Physicochemical Modifications, including Aggregation, on the Stability of Biopharmaceutical Drug Products

Several physicochemical modifications, such as aggregation, oxidation, deamidation, glycation, glycosylation, and disulfide scrambling, can adversely impact the molecular integrity of the active ingredient in biopharmaceutical drug products. These instabilities can arise from several sources, including extractables and leachables from drug delivery components, such as glass/silica from vials, silicon oil on prefilled syringes, and metal ions from injection needles. This chapter describes the consequences of physicochemical degradation on the stability, efficacy, and pharmacokinetics/pharmacodynamics (PK/PD) of biopharmaceuticals and attempts to identify potential degradation sites in the sequence and structure of the biotherapeutic candidates. An important part of this chapter is the issue of aggregation encountered during commercial manufacturing, storage, and shipping of biotherapeutics and how it can be mitigated via rational protein design. This chapter describes the computational efforts to understand the aggregation mechanism and predict aggregation-prone regions in proteins. A distinction is made between aggregation due to colloidal properties of liquid biopharmaceutical formulations and the one due to inherent conformational (in)stability of the protein needed to withstand the insults faced by the protein molecule during manufacturing, shipping, and storage.

Chapter 4: Preclinical Immunogenicity Risk Assessement of Biotherapeutics

A great advantage of biotherapeutics over small-molecule drugs is highly specific target binding and nearly complete absence of non-mechanism toxicity. However, administration of biotherapeutics, including recombinant and plasma-derived human proteins, often leads to undesirable immune responses among patients. These responses can vary from transient non-significant injection site inflammations to life-threatening events in rare instances. Another common immune response is the development of anti-drug antibodies. This chapter describes the computational ability to predict B- and T-cell immune epitopes in biotherapeutics and how such tools can help in preclinical immunogenicity risk assessments and the design of deimmunized biologics.

Chapter 5: Application of Mechanistic Pharmacokinetic–Pharmacodynamic Modeling toward the Development of Biologics

The PK/PD and distribution of monoclonal antibodies and other biopharmaceuticals in human tissues depend on their sequence and structural properties, as well as route

of administration. This chapter describes the computational efforts aimed at mathematical modeling of PK/PD profiles of biotherapeutics.

Chapter 6: Challenges in High-Concentration Biopharmaceutical Drug Delivery: A Modeling Perspective

Subcutaneous delivery of high-concentration biopharmaceutical drugs is desirable from the perspective of patient compliance and convenience. However, successful development of such products requires overcoming several challenges related to colloidal behavior of the drug substance, such as viscosity and syringeability. This chapter focuses on the computational efforts to understand viscosity issues in biotherapeutics.

Section II: Developability Practices in the Biopharmaceutical Industry

Chapters 7–9: Best Practices

These three chapters describe the best practices for developability assessment being followed by three of the major biopharmaceutical companies: Novartis, Amgen, and Pfizer.

Chapter 10: Developability Assessment Workflows to De-Risk Biopharmaceutical Development

Several pharmaceutical companies and contact research organizations (CROs) are beginning to develop *in silico* methods tailored toward improving bioprocess yields, biophysical stability, and safety profiles of biopharmaceuticals. One of these companies was contacted to describe its methods and technologies.

Section I

Principles of Biopharmaceutical Informatics

1 Biopharmaceutical Informatics

Applications of Computation in Biologic Drug Development

Sandeep Kumar,[*] Russell H. Robins,[†] Patrick M. Buck,[*] Timothy P. Hickling,[‡] A. Mary Thangakani,[§] Li Li,[¶] Satish K. Singh,[*] and M. Michael Gromiha[**]

CONTENTS

[*] Pharmaceutical Research and Development, Biotherapeutics Pharmaceutical Sciences, Pfizer, Inc., Chesterfield, Missouri
[†] Analytical Research and Development, Biotherapeutics Pharmaceutical Sciences, Pfizer, Inc., Chesterfield, Missouri
[‡] Pharmacokinetics, Dynamics and Metabolism—New Biological Entities, Immunogenicity Sciences, Worldwide Research and Development, Pfizer, Inc., Andover, Massachusetts
[§] Department of Crystallography and Biophysics, University of Madras, Chennai, India
[¶] Pharmaceutical Research and Development, Biotherapeutics Pharmaceutical Sciences, Pfizer, Inc., Andover, Massachusetts
[**] Department of Biotechnology, Indian Institute of Technology Madras, Chennai, India

1.1 INTRODUCTION

Biologics, particularly monoclonal antibodies and antibody-based therapeutics (fragments of antigen binding [Fabs], antibody–drug conjugates [ADCs], and fragment crystallizable [Fc] fusion proteins), have emerged as an important class of therapeutics in the last couple of decades. Based on information available from the journal *mAbs*, more than 30 antibody-based therapeutics have been approved in the United States (US) and European Union (EU) as of February 2014 for the prevention or cure of diverse human diseases, including several cancers, respiratory syncytial virus (RSV) infection, rheumatoid arthritis, psoriasis, asthma, macular degeneration, multiple sclerosis, bone loss, and systemic lupus erythematosus. Besides this, another 10 antibody-based therapeutics are currently in review by US, EU, or Japanese regulatory agencies. According to a recent forecast,[1] four or five therapeutic monoclonal antibody (mAb) candidates currently in phase 3 clinical trials are expected to transition into regulatory review, and another three or four molecules are expected to be approved for marketing in the US or EU during 2014. Therefore, a robust growth in antibody-based therapeutics available to treat human diseases is anticipated in the near future. The success of antibody-based therapeutics in clinics is being enabled by several technological leaps in biology, antibody design, manufacturing, analytical characterization, formulation development, and delivery devices for such candidates. These biotechnological advances are occurring at an increasingly rapid pace in recent years and are facilitating development of both innovator and follow-on biologics. The success of biologics in the clinic is also fueling demand for these medicines throughout the world. However, these biotech successes have also been accompanied by sharp increases in drug development costs, regulatory hurdles, and attrition rates of therapeutic candidates at preclinical and clinical stages.[2] In addition, the rate of translation of therapeutic candidates discovered in laboratories into viable drugs available in clinics is declining.[3] Overall, high drug development costs and profit expectations of drug manufacturers are resulting in high pricing regimes that payers (governments, insurance companies, and individual patients) find increasingly difficult to afford.[2] Therefore, the current practices in biologic drug discovery and development need innovation for reducing costs of drug development and improving safety and flexible, patient-friendly delivery options. Availability of affordable, safe, and easily deliverable biologics can significantly expand the reach of biologics to all parts of the world and improve health for all humans.

The purpose of this chapter and also of this book is to highlight how computational tools and analyses can aid in biologic drug discovery and development, and describe a new discipline called biopharmaceutical informatics. Unlike small-molecule drugs, the discovery and development of biologics has been mainly a compartmentalized enterprise dominated by experimental processes of trial and error. However, currently available computational tools can be utilized at every stage of the drug discovery and development cycle, from target validation to design/selection of lead compounds to bioprocess optimization and formulation development to safety, efficacy, and pharmacokinetics to clinical trial design. For example, molecular design/selection and bioprocess development can significantly impact the biological activity and safety of a biopharmaceutical drug product.[2] Therefore, rigorous

molecular-level assessments of biologic candidates for physicochemical degradations, thermodynamic stability, immunogenicity, and other drug safety attributes at early stages of candidate design and selection can be very helpful in forecasting resources required to develop them and prevent late-stage drug failures. At the early stages, it may also be feasible to optimize the amino acid sequence of the lead candidate(s) for easier, cost-effective drug development and safety profiles than the candidates optimized only for target binding affinity.[4–8]

Biopharmaceutical informatics endeavors to use information technology, sequence- and structure-based bioinformatics analyses, molecular modeling and simulations, and statistical data analyses toward biologic drug development. In this chapter, we shall examine several applications of biopharmaceutical informatics toward biologic drug development. We focus on understanding potential molecular origins of physicochemical degradation of antibody-based biologics and how these may be related to product quality and safety. Development of databases containing experimental data on biophysical stability, safety, and preclinical/clinical immune observations, along with molecular sequence–structural analyses for several biologic candidates, can enable comparisons among molecular-level properties of well-behaved candidates with those of poorly behaved ones. Such comparisons will improve our understanding of how molecular-level properties of the biologic candidates impact their development as drug products. Here, we present case studies from our own work. Several aspects of this chapter shall be described in greater detail in subsequent chapters of this book. Therefore, this chapter does not review the literature comprehensively. Instead, our goal is to spark the reader's interest in different aspects of computational applications to biopharmaceutical drug development by highlighting interesting scientific advances and case studies. It is pertinent to mention here that computation is also being increasingly applied for structure-based design of biologic candidates during drug discovery to improve binding affinity and selectivity of these molecules toward their cognate targets. This is an important aspect of biopharmaceutical informatics. However, it is beyond the scope for this chapter. In yet another aspect of biopharmaceutical informatics, computational tools are also applied in drug development through systems pharmacology[9] and pharmacokinetic/pharmacodynamic modeling.[10] These areas are also not covered in this chapter.

1.2 APPLICATIONS OF INFORMATION TECHNOLOGY: INFORMATION SUPPLY CHAIN AND KNOWLEDGE-BASED DECISION MAKING

In essence, the development of biologic drugs is an information business. With the exception of physical clinical supplies, all development products are in fact informational, supplying the demands of primary (intended) and secondary (unintended) customers toward the development and licensure of high-value products for all stakeholders. Whether it is validating analytical methods, robust manufacturing process understanding, fit-for-purpose formulation design, quality assessments, clinical supply management, regulatory interactions, predictive simulations, or the like, it is imperative

to develop and maintain an effective information supply chain to meet the needs of internal and external customers within research and development. Therefore, there must be constant focus on enabling a knowledge-based approach in which high-quality decision making, development, and execution are facilitated and underwritten by facile access to high-quality data throughout the development value stream. To achieve this, it is essential to cost-effectively capture (produce), manage (curate), and make available (distribute) all pertinent information in a manner that sustains and leverages our collective intellectual legacy. Essentially, an effective information supply chain manifests as a knowledge-based development engine in which requisite information/knowledge is accessible to the right people, at the right time, and in the right format to enable value creation throughout the development cycle.

The desire to take a knowledge-based approach is easy to state but remarkably difficult to describe in operational details, as it is comprised of human intentions and culture as much as it is of tools and technologies. The willingness to understand and exploit technology toward common goals is equally important, making this both a technical and a human enterprise. It is also important to note that knowledge management, as it is often called, is not something an organization builds or buys. Instead, it is merely the consequence of developing a fit-for-purpose information supply chain that embeds appropriate technologies within effective work processes operating within a culture of knowledge appreciation and awareness. All three are essential for success, though the tendency is to focus solely on technology.

Before dissecting the primary elements of an effective information supply chain (production, distribution, consumption), we must recognize two fundamental information markets, as they uniquely impact both design and culture. The primary data market is the most familiar one, in which the customer requesting that information be produced is also the customer in a functionally closed request–delivery loop exemplified by process development requesting sample analysis from analytical development using specific criteria and workflows. Such transactions are well described, repeatable, and readily embedded within both tools and processes. Data standards, quality, formats, and reporting are agreed upon up front. The hypotheses under testing, for example, product quality versus specifications, are also well defined. The secondary data market, on the other hand, is one in which the information consumer did not request that the information be created initially, but seeks to extract additional value from existing information created for other reasons, for example, correlating clinical event data with lot disposition/characterization data to establish key quality attributes. These information queries are unique in that the data were never collected or curated originally to test such hypotheses per se, and as a result, the requirements for data capture, curation, and reporting were not defined for such purpose when the information was created. Historically, the focus has been on the primary data market, with little cultural regard for the secondary data market. This cultural perspective is understandable, but it hinders our ability to extract full value from biologic drug development data. A broader learning perspective that enables hypothesis generation, not just hypothesis testing, is required for the growth of biopharmaceutical informatics.

In the simplest sense, the information supply chain consists of data capture (production), data curation (inventory management), and reporting (distribution channels). In most primary data markets, all three elements are contained within a platform tool

such as Laboratory Information Systems (LIMs), electronic Labortatory Notebook (eLN), and database structures. These include data authoring, data archiving, retrieval, and reporting all within a tool common to both the producer and consumer under well-defined criteria. In secondary data markets, producer and consumer often do not share expertise or access to a common tool or share criteria on data standards or definitions, making the exposure and retrieval of pertinent information difficult, if not impossible.

Therefore, an effective information supply chain that faithfully serves both our primary and secondary data markets across myriad biologic drug development interests and partners must have the follow capabilities:

- A culture that appreciates that all data have legacy as well as specific value and treats them as such
- A culture that is enabled with the capability to capture, curate, and use the information it produces
- Sufficiently common data standards and quality criteria to facilitate data retrieval and reutilization
- Sufficient tool interoperability to enable data discovery and distribution
- Work process alignment promoting the distribution of data in appropriate formats across end users
- Information consolidation layer capability enabling disparate systems to work as one when necessary or more practical
- Appropriate data analytics to reduce and extract meaning from our information

To acquire these capabilities, the following investments are usually required:

- Tools for the exposure and retrieval of desired information without reliance on social networks, local experts, or deep knowledge of multiple curation points
- Capture of appropriate metadata across work streams and partners with sufficient context to facilitate meaningful interpretation
- Development of intuitive end-user interfaces that facilitate compliance with business rules with minimal training or specific tool expertise
- Portfolio management systems that enable rapid and informed decision making, and preserve legacy learning across projects
- Implementation of single sourcing of key information to reduce redundancy, inefficiencies, and cross-verification burden
- Information utilization that is fully leveraged beyond the written word—audio, video, and imaging
- Tools to access information on demand agnostic of technology

The above discussion has presented an aspirational road map for utilizing information technology toward enabling goals of biopharmaceutical informatics. Besides this, there is also a need to develop scientific understandings, tools, and techniques to fully realize the potential of this field. The subsequent sections in this chapter describe our initial attempts.

1.3 DEVELOPABILITY ASSESSMENTS OF BIOLOGIC CANDIDATES: PREDICTING POTENTIAL PHYSICOCHEMICAL DEGRADATION SITES

Traditionally, biologic drug discovery and development has been compartmentalized into discovery and development organizations. During drug discovery, the main objective is to identify highly potent candidate molecules that are most likely to achieve a desired therapeutic effect. At this stage, high binding affinity and selectivity toward a particular receptor are the major drivers for design/selection of molecular candidates. Once selected, the biologic candidate proceeds into development for various stages of animal and human testing for safety and efficacy. Drug development scientists endeavor to stabilize the biologic molecule for commercially viable production, long shelf life, and delivery in a user-friendly format. This traditional approach implies that the amino acid sequence of the selected molecule is fixed and cannot be changed once it enters the development stages. Therefore, drug product development mainly involves use of external processes like formulation buffer, pH and excipient screening, lyophilization, and drug delivery devices to minimize physicochemical degradation of the drug molecule during storage, shipping, and administration. A limitation of this paradigm is that the drug product development fails or stalls for problematic molecules with poor stability or solubility if finding optimum formulation and delivery combinations proves difficult. Even if the optimum biopharmaceutical drug product is developed from a molecule with poor biophysical attributes, its delivery options may be limited. For example, it may be feasible to develop a biologic drug product for parenteral, but not for subcutaneous administration due to viscosity and syringeability related issues at high concentrations. Another limitation of this paradigm is that opportunities to optimize the selected candidate for desirable attributes, such as high cell line expression yields, safety (low immunogenicity), improved pharmacokinetics/pharmacodynamics (PK/PD), less frequent dosing, and flexible delivery options, are lost. Furthermore, lack of consideration of the discovery and development efforts as a seamless interdependent whole leads to expensive development, if not outright failures. All of these limitations contribute to the currently high costs of developing and manufacturing biologics. To overcome the above-mentioned limitations and capitalize on all opportunities available to a biologic drug product development program, it is essential to modify the above-described compartmentalized paradigm. This can be done by understanding sequence and structural features of biologic drug candidates and optimizing them, not only for potency, but also for developability, manufacturing costs, and safety. Below, we describe how computation can be used to assess biologic candidates for developability by taking into consideration potential chemical degradation sites, aggregation, immunogenicity, and high-concentration-solution behavior.

Biologics comprise a variety of products, such as oligonucleotides, growth factors, cytokines, hormones, receptors, enzymes and clotting factors, prophylactic/therapeutic vaccines, monoclonal antibodies (mAbs), antibody components (Fabs), Fc fusion proteins, and antibody–drug conjugates (ADCs). These macromolecules possess highly complex heterogeneous three-dimensional (3D) molecular structures and are produced using recombinant DNA technologies in a variety of hosts or may

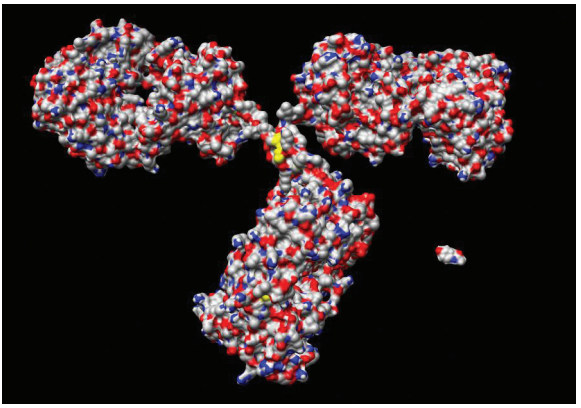


FIGURE 1.1 Molecular structure of a mAb versus that of a small-molecule drug is shown. The mAb shown in this case is a murine IgG2a mAb[18] whose crystal structure is available in the Protein Data Bank[19,20] entry 1IGT. The small-molecule drug shown here is acetaminophen.

be plasma derived. Figure 1.1 illustrates the complexity of biologic drug molecules by comparing the molecular structure of a mAb with a small-molecule drug. As stated earlier, mAbs are emerging as the most successful class of biopharmaceuticals, and it can be seen that their molecular structures are far bigger and much more complex than those of the small-molecule drugs. The mAbs and antibody-based drug candidates and products are the major focus of this chapter.

As a consequence of their size and structural complexities, biologics are vulnerable to numerous physicochemical stresses during manufacturing, shipping, storage, and administration. Degradation caused by these stresses can potentially compromise the potency and safety of these drug products. A number of physicochemical stresses potentially encountered by a biologic are shown in Figure 1.2. This figure illustrates that multiple stresses experienced by biologic molecules at various stages can result in common physical degradations, such as aggregation. Naturally occurring proteins found in organisms adapted to extreme environmental conditions, such as high and low temperatures and high acidity and salinity, also face similar stresses.[11] Therefore, strategies used by nature can potentially be applied to the molecular design and formulation of biologics. In particular, organisms adapted to high and low temperatures (thermophiles and psychrophiles) can teach us important lessons[12–17] toward improving protein stability and solubility without sacrificing potency or causing large-scale structural rearrangements. Striking a balance among protein activity, stability, solubility, and viscosity is consistent with the goal of biologic product development, which is to maintain molecular (physicochemical as well as structural) integrity of the protein coping with environmental stresses.

In a living cell, proteins can age due to several non-enzymatic covalent modifications that accrue gradually because of their oxygen-rich aqueous environment. Moreover, almost all amino acid residues found in natural proteins are vulnerable to one or another chemical degradation.[21] Likewise, it can be imagined that biologics can also age during storage. Therefore, biologic formulations must contain components

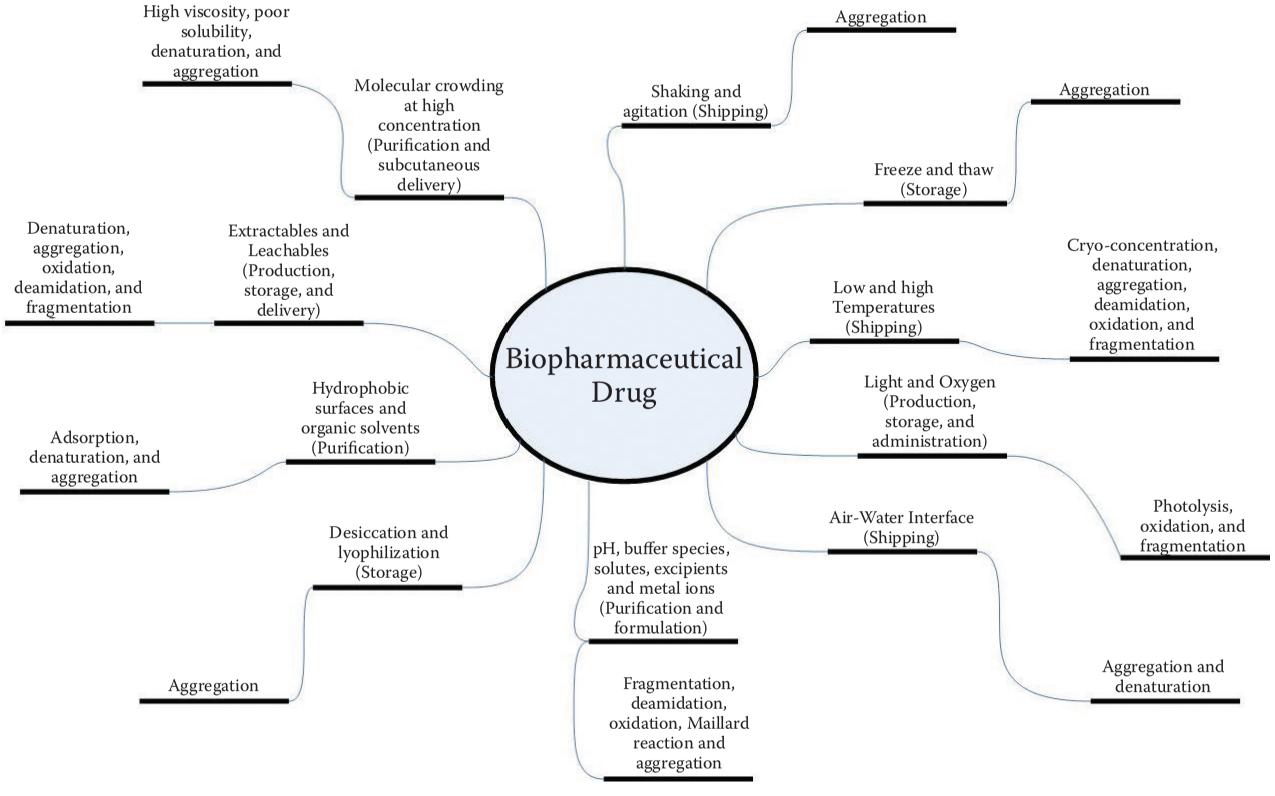


FIGURE 1.2 Physicochemical stresses (inner circle of bars) faced by a biopharmaceutical drug during various stages of production, shipping, storage, and administration and their consequences (outer circle of bars). Note that different stages of drug manufacturing, storage, and shipping may encounter common stresses.

that retard physicochemical degradation of the drug substances and maintain integrity of the biologic drug products over the duration of their shelf life. Several potential physicochemical degradations and their mechanisms have been described for biopharmaceuticals. Among these, deamidation, oxidation, isomerization, fragmentation, and aggregation are the common ones.[22] Most of the chemical degradation routes (i.e., deamidation, oxidation, etc.) arise from specific residues or residue pairs found in protein sequences.[23] Such degradation sites can be easily identified in the amino acid sequences of the biologic candidates. Aggregation-prone regions (APRs), susceptible to β-strand-mediated aggregation, are typically 5–10 residues long and can also be predicted using the amino acid sequence. However, APR prediction is complex and currently available methods are not 100% accurate.[6,24,25] Furthermore, β-strand-mediated aggregation can lead to several morphologies, ranging from amorphous β-aggregates to amyloid fibrils.[26,27] Using experimental data available on hexapeptides that form either amorphous β-aggregates or amyloid fibrils, Thangakani et al. have developed an algorithm, called Generalized Aggregation Proneness (GAP).[28,29] GAP scans a given amino acid sequence for amyloid fibril and amorphous β-aggregating hexapeptide segments based on propensities of amino acid residue pairs to occur together at the same or alternate faces of a β-strand that participates in aggregation.[29] Benchmarking studies using the available experimental data indicate that GAP performs at a significantly superior level than other APR prediction algorithms.[29] For example, Tsolis et al.[30] have recently compiled a set of 48 amyloid fibril-forming peptide sequences found in 33 amyloidogenic proteins. These sequences were used here to benchmark performances of several freely available APR prediction tools. A peptide sequence was considered to be amyloidogenic if it contained at least one APR (six or more consecutive residues identified as aggregating). The results are shown in Table 1.1. GAP is considerably more accurate than the other methods. In summary, the above discussion indicates that physicochemical degradation sites can be predicted from amino acid sequences of biologic drug candidates. Such predictions can be further refined using structural models of the biologic candidates.

To elaborate on the use of amino acid sequences and structural models toward predicting potential physicochemical degradation sites in biologic candidates, the human b12 monoclonal antibody,[35] whose full-length crystal structure is available in the Protein Data Bank (PDB)[19,20] entry 1HZH, is utilized. Consider that this is a biologic drug candidate at the initial stages of formulation development that needs to be assessed for potential physicochemical degradation routes. Figure 1.3 shows potential physicochemical degradation sites in the amino acid sequence of human b12 mAb, and Table 1.2 counts the number of such sites in variable and constant regions of the mAb. Figure 1.3 and Table 1.2 present the potential physicochemical degradation sites for a pair of heavy and light-chain sequences in the human b12 mAb. A full-length mAb contains two such pairs.

Figure 1.3 and Table 1.2 illustrate the complexity of macromolecules such as mAbs by pointing out that large portions of their amino acid sequences are inherently prone to one or another of the several physicochemical degradations. Chemical degradations often require that the involved residues be present on the protein surface so that they can interact with solvent, metal ions, redox agents, and so forth. Similarly, the APRs

TABLE 1.1
Performance of Different Aggregation Prediction Algorithms on 48 Sequences from 33 Amyloidogenic Proteins

Aggregation Prediction Algorithm	Number of Sequences Predicted to Contain at Least One APR	Number of Sequences Predicted to Contain No APRs	Total Number of Amyloidogenic Sequences	Accuracy[a] (%)
GAP[29]	40	8	48	83.3
Amylpred2[30]	30	18	48	60.4
AGGRESCAN[31]	32	16	48	66.7
TANGO[32]	14	34	48	29.2
WALTZ[33]	21	27	48	43.7
PASTA2[34]	14	34	48	29.2

[a] For each algorithm, accuracy was judged based on the number of correctly predicted sequences with at least one APR of six or more consecutive aggregation-prone residues. These benchmarks are for 48 amyloid fibril-forming sequences from 33 amyloidogenic proteins.[30] For a more comprehensive comparison, refer to Thangakani et al.[29]

need to lie at or near protein surfaces to be able to promote aggregation in response to a physical stress such as temperature. Therefore, building 3D structural models is essential to pinpoint which of the above marked sites are at greater risk of physicochemical degradations. For the human b12 antibody, a crystal structure of the full-length antibody is publicly available. 3D structures are often unavailable for most biopharmaceutical drug candidates. Moreover, crystallizing every candidate and its variants is costly and time-consuming. Computational techniques of protein structure prediction are commonly used during research and development of biopharmaceutical products. Homology-based models of biologic candidates can be rapidly derived if suitable templates are available. Homology-based protein structure prediction is a vast field with several applications to drug discovery and design,[38,39] and a review of this field is out of the scope for this chapter. Briefly, homology modeling relies on the premise that proteins with similar amino acid sequences have similar 3D structures. Therefore, it utilizes the similarity of a target protein's amino acid sequence with that of a template protein, whose experimental structure is available, to model the 3D structure of the target protein. Procedures for computational modeling of antibody-variable domains (Fvs and Fabs) have been developed in recent years, and these are proving helpful in structure-based design of antibody-based therapeutics.[40] Homology-based models of variable regions of antibodies and also of the full-length antibodies can prove useful in understanding physicochemical attributes of the candidates and for pinpointing potential physicochemical degradation sites.

Let us continue with our example of human b12 antibody and refine predictions made using the mAb sequence (Figure 1.3). As stated earlier, a 2.7 Å resolution crystal structure for this mAb is publicly available in the PDB entry 1HZH. This structure