1

Introduction to Biotherapeutics

Jennie R. Lill

Department of Microchemistry, Proteomics & Lipidomics, Genentech Inc., South San Francisco, CA, USA

Abbreviations

ADAs antidrug antibodies
ADC antibody–drug conjugate
ADCC antibody-dependent cell-mediated cytotoxicity
CDR complementary-determining region
Fab antigen binding fragment
Fc crystallizable fragment
NMR nuclear magnetic resonance
PEG polyethyleneglycol
PTM posttranslational modification

1.1 Introduction

Biotherapeutics, also known as biologics, include protein-based and nucleic acid-based drugs that are commonly derived by recombinant expression in living organisms although a few are made by chemical synthesis. This book focuses on the characterization of protein-based biotherapeutics, exploring the various analytical technologies that have enabled in-depth molecular characterization while discussing current triumphs and limitations.

The first human protein therapeutic derived from recombinant DNA technology was human insulin (Humulin®) created at Genentech, developed by Eli Lilly, and approved by the US Food and Drug Administration (FDA) in 1982. Since that time, major advancements in both recombinant DNA technology
and recombinant protein production have contributed to the development of several hundred biotherapeutics [1] including relatively simple molecules such as interferons, insulin, and the human growth hormone to more complexly engineered moieties including ADCs such as trastuzumab emtansine [2] and brentuximab vedotin [3].

Unlike conventional small molecule (chemical) drugs such as aspirin, antibiotics, and various chemo-therapeutics, the manufacturing process for biotherapeutics is typically far more cumbersome as they are larger compounds with more complex structures and their production can be extremely sensitive to changes in fermentation and environmental conditions. In addition, biotherapeutics are often less stable than many small molecules and can be prone to aggregation [4] or deamidation, oxidation, and other modifications [5]. Since the manufacturing of biotherapeutics is often dependent upon the host cells of living organisms, complex process development is required to ensure reproducible fermentations, isolation, and characterization [6].

### 1.2 Types of Biotherapeutics and Manufacturing Systems

There are several different types of marketed biotherapeutics including antibody-based drugs, anticoagulants, blood factors, bone morphogenetic proteins, engineered protein scaffolds, enzymes, Fc (cystallizable fragment) fusion proteins, growth factors, hormones, interferons, interleukins, and thrombolytics (Figure 1.1).

![Figure 1.1 Various categories of the main types of biotherapeutics currently marketed. Source: Carter [7]. Reproduced with permission of Elsevier.](image-url)
Antibody-based drugs represent the largest and most rapidly expanding class of biotherapeutics [1]. Figure 1.2 shows the diverse mechanisms by which the antibody structure can be modified to increase its biotherapeutic potential.

Humanized and other chimeric versions of these antibodies now dominate the market [11] and in the past 5 years have accounted for nearly 30% of all approvals. Various antibody isotypes are now being explored to provide a wealth of functional diversity that is present through the various IgG subclasses that can be exploited to improve clinical safety and performance by increasing stability, reducing adverse events, modulating effector functions, and by the engagement of two antigens by a single antibody [8]. Several variants that have been Fc engineered for reduced effector function have entered the clinic, for example, Eculizumab, a novel engineered IgG isotype, IgG2m4, with reduced Fc functionality. IgG2m4 is engineered based on the IgG2 isotype with four key amino acid residue changes derived from IgG4 (H268Q, V309L, A330S, and P331S). This antibody was demonstrated to have an overall reduction in complement and Fc gamma receptor binding in in vitro binding analyses while maintaining the normal in vivo serum half-life in rhesus [12].

Biosimilars (biologically identical antibodies, for example) and so-called biobetters (moieties with improved properties such as pharmacodynamic (PD) and pharmacokinetic (PK) readouts, higher potency, longer half-lives, and less immunogenicity, for example) are also starting to emerge, which presents new challenges in terms of testing for the presence of liabilities such as degradative properties, changes in immunogenicity through addition of novel contaminant proteins from new manufacturing processes, and so on. New formats such as glucagon-like peptide 1GLP fused proteins, for example, Eperzan (albiglutide) [13], and PEGylated proteins such as Plegridy (e.g., peginterferon beta-1a) [14] offer improved PK or PD properties but also increased analytical challenges due to their larger masses and increased heterogeneity.

Typically, expression of non-mAb biotherapeutics has been performed in Escherichia coli or a noneukaryotic system. This has many advantages for biotherapeutics that are not reliant on PTMs for their optimal activity. Over the years, however, there has been a gradual increase in the prevalence of mammalian expression systems. Of the mammalian expression systems, the Chinese hamster ovary (CHO) cell-based model (reviewed by Krawitz and Sandoval in Ref. [11]) remains the most employed expression system with a smaller percentage of therapeutics manufactured in other mammalian cell lines such as the murine myeloma line, NSO, and baby hamster kidney cells [15, 16]. Nonmammalian eukaryotic expression systems such as yeast [17] are also utilized, each again presenting their own challenges with regard to the correct PTM of the protein, occasionally adding to adverse properties [18, 19].

More recently transgenic animal production systems (e.g., expression of recombinant products in the animals’ milk [19, 20], rabbits, and goats) have been explored as a means of biopharmaceutical production although to date
Figure 1.2 Monoclonal antibody (mAb) structure can be modified on the basis of the desired mechanism of action. Immunoglobulin G1 (IgG1) is the most effective naturally occurring human IgG isotype at mediating antibody-dependent cell-mediated cytotoxicity (ADCC). Glycomodified afucosylated mAbs (part a) (such as Obinutuzumab) demonstrate enhanced binding to IgG Fc receptors (FcγRs) and enhanced ADCC. In addition antibody-dependent cellular phagocytosis, a process mediated by macrophages, can also occur [8]. Afucosylated mAbs are produced using cell lines that lack the enzymes responsible for fucosylation. Modifying the amino acid sequence of mAb Fc (part b), as was done to produce ocaratuzumab [9], can also result in enhanced binding to FcγRs and enhanced ADCC. For mechanisms of action in which ADCC is not desirable, IgG4 may be a more appropriate isotype, as IgG4 mAbs do not mediate ADCC to the same degree as IgG1 (part c) although this isotype can still engage macrophage effector function via nanomolar affinity binding to FcγRI. Nivolumab, an IgG4 mAb that blocks programmed cell death protein 1 (PD1) on T cells, is one such example. Producing radioimmunoconjugates involves linking the radioisotope to the mAb. A stable linker is most desirable (part d) to limit the leakage of the free radioactive isotope. Conversely, optimal antibody–drug conjugates (ADCs) use a cleavable linker (part e). To avoid nonspecific toxicity, it is desirable for drugs used in ADCs to be cytotoxic once inside the target cell but nontoxic when bound to the mAb in the circulation. Linkers that are pH-sensitive or enzymatically cleaved are now a standard component of ADCs. Chimeric antigen receptor (CAR) T cells get their specificity from mAb variable regions but are a form of gene, not protein, therapy. They are produced by inserting DNA coding for the mAb variable region fused to DNA coding for signaling peptides into T cells (part f). Some bispecific antibodies lack a functional constant region so that they do not nonspecifically crosslink activating receptors and activate T cells (part g). The lack of a constant region on such constructs results in a short half-life, thus requiring continuous infusion to achieve the desired exposure. Source: Weiner [10]. Reproduced with permission of Nature Publishing Group.
there are many challenges associated with this type of biotherapeutic production with few benefits. Throughout this book the challenges of characterizing both the biotherapeutic moiety itself and the contaminant proteins such as CHO-derived proteins are discussed.

1.3 Types of Analyses Performed

Throughout this book a variety of analytical procedures are described. Many of them have been implemented for characterizing biotherapeutic molecules for as long as these moieties have existed. Others have evolved as the need arises. One such example of developing such sets of tools to answer a newly arisen problem is for the de novo sequencing of antibodies [21, 22]. Occasionally antibodies are discovered that are of great interest for preclinical testing, for which the cDNA or any genetic information is not available. In these scenarios, researchers have to sequence antibodies at the protein level, one amino acid at a time, and then reverse engineer the antibodies to the nucleotide level. In Chapter 6 Castellana and Guthals provide technical details and review the innovative approaches employed to quickly gain sequence information through a de novo approach.

As well as sequence information at the amino acid level, PTM profiling is also an important element in characterizing biotherapeutics [23]. There are a plethora of cotranslational modifications and PTMs that play key roles in the folding of proteins, in their secretion, and in their ultimate stability and effector functionality in vivo.

Glycosylation is important both for antibody secretion by B-cells and for in vivo antibody effector function. Glyco-engineering is a rapidly growing field, whereby glycosylation sites and composites are engineered to produce antibodies with specific glycoforms which may have an effect on therapeutic efficacy. Obinutuzumab (Gazyva®) [24], for example, is a humanized therapeutic monoclonal antibody that binds to an epitope on the B cell antigen, CD20. This antibody is engineered in a platform that allows control of the proteins’ glycosylation, in this case the platform enforces the overexpression of two glycosylation enzymes MGAT3 and the golgi mannosidase 2. This results in the generation of antibodies with bisected nonfucosylated sugars, thereby increasing the antibodies’ ability to activate natural killer cells. This means that Obinutuzumab can induce cell death through a dual mechanism of action, both by the antibody directly binding to B cells and by antibody-mediated cytotoxicity by recruiting the immune system to attack B cells. Some types of glycosylation are sometimes not beneficial. For example, Cetuximab, a chimeric mouse–human IgG1 monoclonal antibody against the epidermal growth factor receptor (EGFR) approved for use in colorectal cancer and squamous-cell carcinoma of the head and neck has a high prevalence of hypersensitivity reactions which has been attributed to cross reactivity to a V domain
glycosylation site. In some patients severe adverse events have been observed including anaphylaxis, which was found to be due to the generation of patient-specific antibodies to the galactose-alpha-1,3-galactose modification [25].

Protein terminal modifications have the effect of modifying a protein’s function, half-life, or cellular localization. Pyroglutamate formation, for example, is a highly prevalent modification whereby glutamine and glutamate at the N termini of recombinant monoclonal antibodies can cyclize spontaneously to pyroglutamate (pE) \textit{in vitro} [26]. Proteolytic processing is also an irreversible modification that affects the vast majority of proteins, often with great functional consequences. Intracellular proteolytic processing has distinct effects on the functionality of proteins and can either abrogate or antagonize function, modify half-life, or also determine cellular localization. During protein synthesis, manufacturing, purification, and storage proteolysis events can occur, thereby changing a protein’s functionality or stability [27]. Either through direct mass spectrometric analysis as reviewed in Chapter 2, through a variety of historic analytical techniques such as gel electrophoresis or Edman degradation, or by employing a variety of new biochemical-based methodologies for determining the termini of recombinant proteins as reviewed in Chapter 3, the determination of proteolytic processing remains a key analytical need for the characterization of biotherapeutic moieties.

Beyond linear sequence determination, structural analyses are also instrumental in the overall characterization of biotherapeutics. The biomolecular architecture is a vital component in dictating the specificity and overall efficacy of therapeutic proteins. The higher order structure (HOS) of a protein includes the secondary, tertiary, and quaternary structures of a protein that are required for its function. There is a diverse range of biophysical methods including circular dichroism, isothermal calorimetry, which are available for the characterization of a protein HOS, each of them with associated benefits and limitations. Related to conformational analysis is structural analysis as it pertains to epitope and paratope mapping. Again, several well-established techniques such as nuclear magnetic resonance (NMR) [28] and X-ray crystallography as well as some newer techniques such as mass spectrometric-based structural tools [29] including hydrogen deuterium exchange are described in Chapters 4 and 5.

1.4 Future perspectives

Nature has provided us with various types of protein scaffolds to explore as frameworks for building new types of biotherapeutics and there is a growing field of using these scaffolds as alternatives to antibodies [30]. Each of these types of engineered molecular structures offers new advantages in terms of stability and specificity. One example of this is the cystine knot mini proteins/peptides (knottins); these are peptide-based alternative molecules to
monoclonal antibodies which are raised/designe against tumor-associated receptors and other antigens of interest. Knottins contains a disulfide-bonded core that exhibits a high level of resistance to proteolysis and increased thermal stability. Knottins emerged as an attractive molecular candidate for drug development as they fill the niche between small molecule drug design and protein biologics. Knottins have the potential to bind clinical targets with both high selectivity and affinity [31]. There are several naturally occurring knottins that have been approved as biotherapeutics for the treatment of pain [32] and irritable bowel syndrome and for tumor imaging purposes [33].

Elucidating disulfide bonding patterns of any biomolecules, but in particular a structure which relies on disulfide bonding patterns for their folding, stability, and activity, is an important part of molecular characterization. A variety of techniques can be employed for doing this from simple intact molecular weight measurement to more complex top-down proteomic protocols [34] and these methodologies continue to mature as more of these types of molecules emerge on the market.

Another growth area for biotherapeutics is increasing the molecules’ *in vivo* half-life. For therapeutics that involves frequent or uncomfortable delivery, for example, injectable ocular therapeutics, or to make drugs that have poor PD properties more tolerable, there is a strong drive to create molecules that have increased *in vivo* stability (and potentially decreased immunogenicity). Some common mechanisms of molecular half-life extension include the generation of Fc fusion proteins, the formulation of biomolecules into various different nanoparticle systems, or the addition of stabilizing peptides. For a comprehensive list of these efforts please refer to Table 1.1.

The addition of albumin to stabilize the half-life of proteins has also been explored. Albumin is the most abundant plasma protein in humans and mice and is highly soluble, extremely stable, and has a circulatory half-life of ~20 days in man [36]. By fusing albumin to therapeutic proteins, these molecules become less susceptible to renal filtration and circulatory clearance and one of the main reasons albumin has a long half-life is due to its ability to bind to FcRn. Association, conjugation, or fusion of therapeutic drugs to albumin has been shown to correlate with superior PK.

In addition to albumin addition, PEGylation, the process of adding polyethylene glycol (PEG) chains to a molecule by incubating a reactive derivative of PEG with the biomolecule of interest, also appears to work universally for improving therapeutic protein *in vivo* half-life [37]. PEGylation also increases the molecule’s hydrodynamic size, thereby prolonging its time in circulation by reducing renal clearance. In addition, PEGylation can also make hydrophobic drugs with poor PK properties more water-soluble. There are a number of PEGylated molecules on the market and many companies are exploring this concept on a wider range of biotherapeutics [38, 39]. Characterizing these modified proteins, such as albumin-conjugated or PEGylated species, brings along increased analytical challenges, some of which are covered by Bakalarski *et al.* in Chapter 2 and by Ellerman *et al.* in Chapter 4.
Table 1.1 The circulatory half-life of a therapeutic protein can be extended by several strategies depending upon the endogenous clearance mechanism of the drug.

<table>
<thead>
<tr>
<th>Protease degradation</th>
<th>Novel/alternative delivery strategies avoiding need to extend half-life</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-terminal acetylation or C-terminal amidation</td>
<td>Controlled-release depot (subcutaneous, intramuscular, intravenous); e.g. Bydureon–poly(lactic-co-glycolic acid) (PLGA) microspheres (Alkermes)</td>
</tr>
<tr>
<td>Nonnatural amino acids at labile sites</td>
<td></td>
</tr>
<tr>
<td>Cyclization using disulfide bonds</td>
<td></td>
</tr>
<tr>
<td>Microspheres or nanoparticles</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Increase size/hydrodynamic volume to prevent clearance by kidneys</th>
<th>Increase size/hydrodynamic volume to prevent clearance by kidneys and attach to protein with a long half-life</th>
</tr>
</thead>
<tbody>
<tr>
<td>Di- or multimers</td>
<td>Attach to Fc (the natural antibody constant region)</td>
</tr>
<tr>
<td>Attach PEG—conjugation</td>
<td>Genetic fusion; e.g. Enbrel (Amgen), Mimetibody™ (Centocor), SynFusion (Biogen Idec/Syntonix)</td>
</tr>
<tr>
<td>Advanced PEGylation—modification of peptides (and prodrugs) by attaching with specific polymer chains</td>
<td>Covalent attachment (site-specific) (CovX/Pfizer)</td>
</tr>
<tr>
<td>Site-specific PEGylation (polytherics)—enables more selective PEGylation, reducing likelihood of protein deactivation upon conjugation and reducing immunogenicity</td>
<td>Attach to human serum albumin</td>
</tr>
<tr>
<td>Glyco-PEGylation (Neose)</td>
<td>Genetic fusion, albufuse® (Novozymes Biopharma/GSK/Teva/CSL)</td>
</tr>
<tr>
<td>Releasable PEGylation (Enzon)</td>
<td>Conjugation (Novozymes Biopharma, ConjuChem, Cardiovax)</td>
</tr>
<tr>
<td>Protein “rPEG”—genetic fusion</td>
<td>Albumin binding peptides/proteins/affinity tags (Genentech/Roche, Ablynx, Philochem, Affibody, Adnexus/BMS)</td>
</tr>
<tr>
<td>Poly-glycine</td>
<td>Albumin binding single domain antibodies fused to bioactive peptides (GSK/Domantis, Ablynx, Haptogen/Wyeth/Pfizer)</td>
</tr>
<tr>
<td>PASylation (XL-protein)</td>
<td>Albumin binding fatty acids (Novo Nordisk)</td>
</tr>
<tr>
<td>XTEN (Amunix)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Hyaluronic acid (Novozymes)</td>
<td></td>
</tr>
</tbody>
</table>
Future perspectives

PEG polymers themselves are often highly heterogeneous and the increase in molecular weight that occurs when biomolecules are PEGylated means that more specialized mass spectrometric techniques or other analytical methods need to be employed [40]. With an increase in these types of complex molecules being developed for improved drug stability, the analytical challenges posed with their characterization also increase in complexity.

In addition to analyzing the intact biomolecules both in vitro and in vivo for assessing molecular stability, PK and PD properties, additional types of analyses to assess traits such as immunogenicity are starting to emerge. All biotherapeutics, including monoclonal antibodies and their derivatives, are immunogenic to varying degrees in various patients, with chimeric antibodies representing more of an immunogenic risk than humanized antibodies. These mouse human chimeric antibodies can induce some patients to develop antidrug antibody (ADA) responses [41]. Certain factors are known to influence biotherapeutic immunogenicity including structural homology with respect to human amino acid sequences and various PTMs. Although frameworks can be designed to minimize the potential for immunogenicity, the complementarity-determining regions (CDRs) of antibodies and other variable domains of biotherapeutics can be highly sequence variable making it difficult to predict how immunogenic a regent maybe [42]. Several companies have adopted strategies to assess immunogenicity retrospectively for therapeutics that have shown an ADA response [43]. Prospectively, a T cell activation assay, whereby CD4+ T cells are monitored for activation by antigen-presenting cells (APCs) loaded

Table 1.1 (Continued)

<table>
<thead>
<tr>
<th>Increase size/hydrodynamic volume to prevent clearance by kidneys</th>
<th>Increase size/hydrodynamic volume to prevent clearance by kidneys and attach to protein with a long half-life</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydroxyethyl Starch (HESylation®—Fresenius Kabi)</td>
<td>Hydroxyethyl Starch (HESylation®—Fresenius Kabi)</td>
</tr>
<tr>
<td>Polysialic acid (PolyXen®—Xenetic Biosciences)</td>
<td>Polysialic acid (PolyXen®—Xenetic Biosciences)</td>
</tr>
<tr>
<td>Elastin-like polypeptide (ELP) technology (Phase Bio Pharmaceuticals Inc.)</td>
<td>Elastin-like polypeptide (ELP) technology (Phase Bio Pharmaceuticals Inc.)</td>
</tr>
</tbody>
</table>

Source: Sleep [35]. Reproduced with permission of Elsevier.

These include the reduction in the endogenous degradation of the drug; slow release/depot formulations; increasing the hydrodynamic volume of the drug by attachment of a large bulky polymer or extension of the therapeutic protein by addition of a linear but unstructured protein; or the addition of a large structured protein which additionally can take advantage of the FcRn-mediated recycling.
with biotherapeutics, can be employed alongside a major histocompatibility complex associated peptide proteomics (MAPPs) assay. The MAPPs assay involves the \textit{in silico} prediction of DR4 or other MHC class II peptide presentation using one of several programs such as SYPETHI, NETMHC, and so on. These algorithms predict potential T cell epitopes derived from the therapeutic protein. Additional analyses whereby MHC class II complexes are immune-precipitated from cells, peptides isolated and analyzed by mass spectrometry are also sometimes performed to complement the results from \textit{in silico} predictions. These types of analyses are starting to be incorporated into early molecular assessment workflows to minimize the potential risk of adverse immunogenicity of new molecules prior to ADAs being reported.

In addition to these types of analyses, additional complexities arise due to chemically modified/conjugated biotherapeutics. Antibody–drug conjugates and antibody–antibiotic conjugates are gaining momentum in the clinic and these hybrid molecules that are composed of an antibody, a cleavable linker, and a chemotherapeutic or potent antibiotic molecule raise their own set of analytical challenges. In Chapter 7, Liu describes the importance of measuring drug-to-antibody ratios (DARs) \cite{44} and the types of chromatographies and mass spectrometric techniques employed for characterizing these complex molecules. Liu describes the applications of increasingly diversified mass spectrometric techniques employed for characterizing the integrity of ADCs from the perspective of production to \textit{in vitro} and \textit{in vivo} testing as well as in terms of elucidating the mechanisms of ADC biotransformations.

So, now we have a variety of different biotherapeutic material in the research environment, manufacturing, and the clinic. How do we organize such materials? Prior to official lot validation that occurs once a biomolecule hits our pipeline, how do we ensure that the antigens we are using for immunization are from the same batch of cells, have been purified in the same manner and already characterized for correct protein sequence, folding, and structure? Biorepositories \cite{45, 46} are growing in popularity and necessity in biotechnology companies. Official curation, stable storage, and easy retrieval of various lots of cDNAs, antigens, antibodies, ADCs and all the associated modifications are now becoming commonplace to allow researchers more consistency in the quality and reproducibility of their assay and other results.

As analytical technologies increase in sensitivity and complexity, low-level heterogeneity of therapeutic biomolecules that were previously unknown becomes revealed. Many of these modifications will be irrelevant in terms of the protein’s stability or function; however, when reporting such heterogeneity to the FDA and other regulatory bodies one has to take care to not provide information that might be misconstrued. So how sensitive should the analysis be? At what percentile of the overall nonmodified protein should a mutation or modification be before it is reported or deemed important for functionality?
Some of the techniques involved in the molecular assessment of therapeutic biomolecules are described by Phung et al. in Chapter 11. These are all questions that analytical chemists and biochemists face as we embark on the design of new biotherapeutic moieties and see the emergence of biosimilars.

**Acknowledgments**

Thank you to Wendy Sandoval and Paul J. Carter for informative discussions and editing.

**References**


