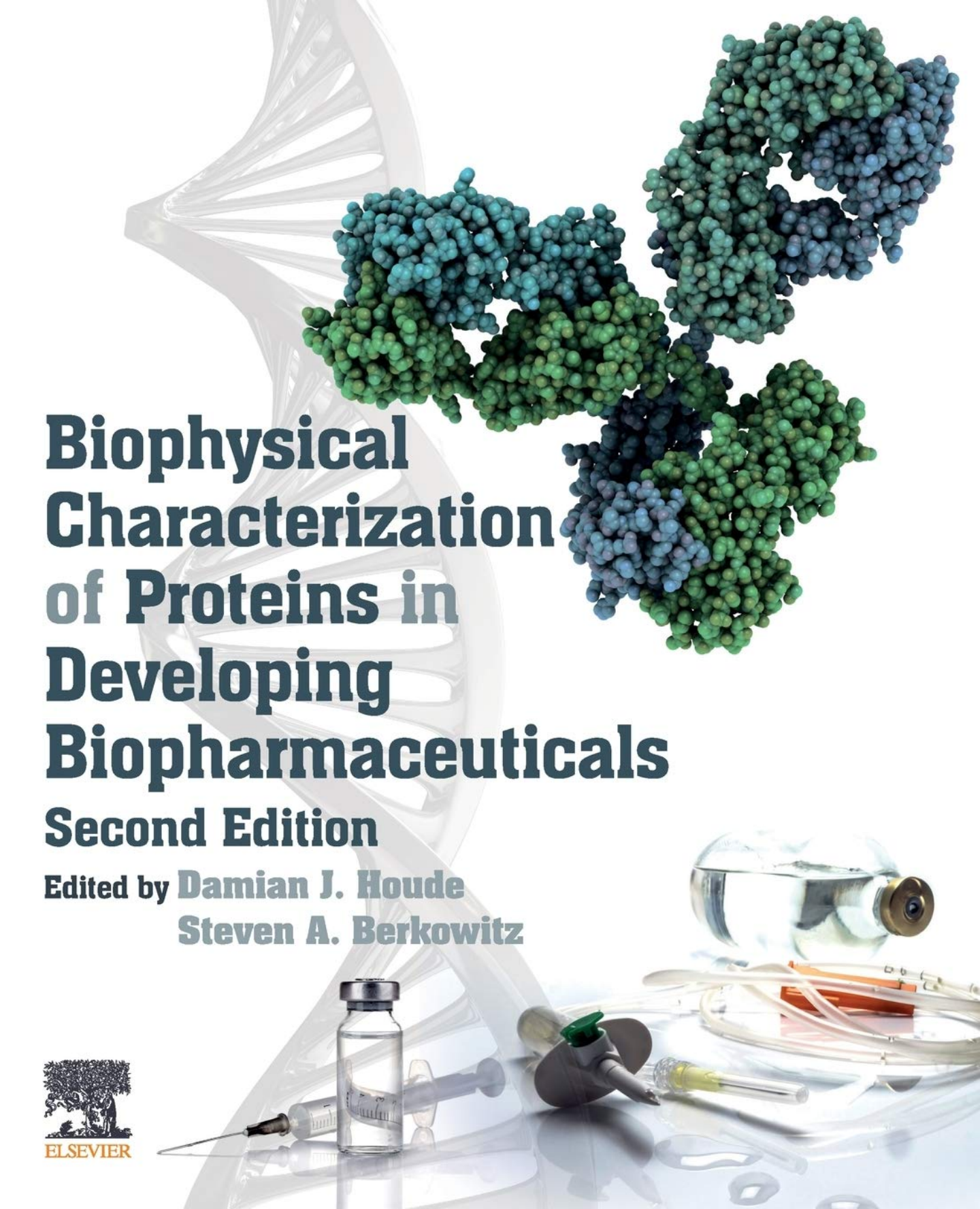

Biophysical Characterization of Proteins in Developing Biopharmaceuticals

Second Edition

Edited by Damian J. Houde
Steven A. Berkowitz

ELSEVIER

BIOPHYSICAL CHARACTERIZATION OF PROTEINS IN DEVELOPING BIOPHARMACEUTICALS

SECOND EDITION

Edited by

DAMIAN J. HOUDE
Biomolecular Discovery, Relay Therapeutics, Cambridge, MA, United States

STEVEN A. BERKOWITZ
Consultant, Sudbury, MA, United States

ELSEVIER

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-444-64173-1

Publisher: Susan Dennis
Acquisition Editor: Anita Koch
Editorial Project Manager: Sara Pianavilla
Production Project Manager: Surya Narayanan Jayachandran
Cover Designer: Matthew Limbert

Typeset by TNQ Technologies

Contributors

Yves Aubin Centre for Biologics, Regulatory Research Division, Evaluation, Biologics and Genetic Therapies Directorate, Health Products and Food Branch, Health Canada, Ottawa, ON, Canada

Steven A. Berkowitz Consultant, Sudbury, MA, United States

George M. Bou-Assaf Analytical Development, Biogen, Cambridge, MA, United States

Mark L. Brader Drug Product Analytical Development, Moderna, Cambridge, MA, United States

Richard K. Burdick Burdick Statistical Consulting, LLC, Colorado Springs, CO, United States

John F. Carpenter Department of Pharmaceutical Sciences, Center for Pharmaceutical Biotechnology, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

Stephen J. Demarest Eli Lilly Biotechnology Center, San Diego, CA, United States

Ertan Eryilmaz Takeda, Cambridge, Massachusetts, United States

Verna Frasca Field Applications Manager, Malvern Panalytical, Northampton, MA, United States

Darron L. Freedberg Structural Biology Section, Laboratory of Bacterial Polysaccharides, Silver Spring, MD, United States

John P. Gabrielson Elion Labs, A Division of KBI Biopharma, Inc., Louisville, CO, United States

Andrea Hawe Coriolis Pharma, Munich, Germany

Damian J. Houde Biomolecular Discovery, Relay Therapeutics, Cambridge, MA, United States

David A. Keire Food and Drug Administration, St. Louis, MO, United States

Francis Kinderman Amgen, Thousand Oaks, California, United States

Lee Makowski Department of Bioengineering and Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, United States

John P. Marino National Institute of Standards and Technology, Institute for Bioscience and Biotechnology Research, Rockville, MD, United States

Alan G. Marshall Ion Cyclotron Resonance Program, National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL, United States; Department of Chemistry & Biochemistry, Florida State University, Tallahassee, FL, United States

A.J. Miles Institute of Structural and Molecular Biology, Birkbeck College, University of London, London, United Kingdom

John S. Philo Alliance Protein Laboratories, San Diego, CA, United States

Angelika Reichel Coriolis Pharma, Munich, Germany

Deniz B. Temel Bristol-Myers Squibb, Devens, Massachusetts, United States

B.A. Wallace Institute of Structural and Molecular Biology, Birkbeck College, University of London, London, United Kingdom

Daniel Weinbuch Coriolis Pharma, Munich, Germany

William F. Weiss IV Bioproduct Research and Development, Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN, United States

Sarah Zölls Coriolis Pharma, Munich, Germany

Prefaces for the second edition

Critical to the development of any successful therapeutic drug is our ability to identify and manufacture the drug such that its beneficial therapeutic effect can be safely delivered to the patient. In the case of protein biopharmaceuticals, these large, heterogeneous (complex) and marginally stable molecules are often very sensitive to their micro-environment. This makes the process of developing and manufacturing a protein therapeutic extremely challenging. Throughout this entire process a protein biopharmaceutical must maintain its complex and delicate structure (or conformation) to realize its beneficial therapeutic attributes, while avoiding the potential harmful effects in failing to achieve this goal.

When we set out to write the first edition of this book, our goal was to provide a general resource that specifically dealt with the many challenges associated with the testing and characterization of the higher order structure and biophysical properties of protein biopharmaceuticals from a practical point of view to support its safety and beneficial therapeutic activity. As stated in the book's first preface we wanted to keep the reader focused on obtaining a pragmatic understanding and knowledge of the utility of biophysical tools and how they are used to meet these challenges by understanding what information can realistically be extracted from these tools. While we felt we had initially achieved our goal, the progression of time inevitably led to better and improved scientific developments and to the realization that there was room for improvements.

As a result, in writing this second edition we have undertaken the job of updating old information, correcting mistakes, improving clarity and the introduction of new topics that were not covered in the first edition. Therefore, we gathered our co-authors once again, invited a few new ones, and tasked ourselves with the goal to achieve these objectives. In so doing, all original chapters have been updated, corrected and enhanced, while new chapters have been added.

Globally, the format of the book has remained the same, consisting of three sections. Section I, which deals with the complexity of proteins and the relevance of biophysical methods in the biopharmaceutical industry. It has for the most part been altered to remove errors and achieve clarity. Section II, which discusses the biophysical tools and techniques most commonly used in the biopharmaceutical industry to characterize protein therapeutic molecules has similarly been altered, but has also been enhanced by the addition of a new chapter (Chapter 14) dedicated to the area of chromatography and electrophoresis. The tools in this chapter, which we did not cover in the first edition of the book (with the exception of size-exclusion chromatography), are typically not thought of or classified as biophysical tools. Nevertheless, an important objective in adding this chapter is to bring more attention

to their unrealized linkage as effective biophysical characterization tools without getting too deep into the details of their inner workings (which are extensively covered in many excellent books and review articles that are solely dedicated to these two enormously important techniques).

Overall, however, Section III of the book has experienced the most significant change and expansion via the addition of four new chapters that cover the following:

- Chapter 15, which deals with the biophysical characterization of complex biopharmaceuticals;
- Chapter 16, which deals with the rigor of statistical analysis;
- Chapter 17, which deals with biopharmaceutical developability;
- Chapter 18, which deals with technical decision making.

Finally, we would like to point out that in writing this second edition we have made a particular effort, wherever possible, to better link and cross-reference information in each chapter to bring more cohesion to the book as appose to just providing the reader with a collection of isolated chapters. We think his cohesion is in particular made apparent by the four additional chapters in Section III (described above).

In the end, we and our coauthors hope we have further enhanced the initial objective of the first edition of the book, of enlightening the reader to the challenges, tools and inner workings of the task associated with the biophysical characterization of protein biopharmaceuticals. An integral part of today's modern and challenging world of developing lifesaving drugs.

List of abbreviations and symbols

(T)RPS	(Tunable) resistive pulse sensing
3D	Three dimensional
A_{22} or B_2	Second viral coefficient
AAV	Adeno-associated virus
AC	Alternating current
AC-SINS	Affinity-capture self-interaction nanoparticle spectroscopy
AFFF-MALLS	Asymmetric field flow fractionation with multi-angle laser light scattering
ACN	Acetonitrile
ACS	Ammonium camphor sulfonate
ADC or ADCs	Analog to digital converter or Antibody drug conjugate(s)
ADCC	Antibody dependent cell-mediated cytotoxicity
AF4	Asymmetric flow field flow fractionation
AFM	Atomic force microscopy
API	Active pharmaceutical ingredient
APR	Aggregation prone regions
AQL	Acceptable quality level
ASTM	American society for testing and materials
ATP	Analytical target profile
ATR	Attenuated total reflectance
AUC	Analytical ultracentrifugation
BiAb or bsAb	Bispecific antibody
BLA	Biological license application
BMI	Backgrounded membrane imaging
BSA	Bovine serum albumin
CA	Capsid protein
CAD	Collision-activated dissociation
CCD	Charge-coupled device
CD	Circular dichroism
CDER	Center for drug evaluation and research
CDR	Complementarity-determining region
CEX	Cation-exchange chromatography
cGMP	Current good manufacturing practices
C_H	Immunoglobulin gamma heavy chain constant domain
CH1 or C_H1	Immunoglobulin gamma heavy chain constant domain 1
CH2 or C_H2	Immunoglobulin gamma heavy chain constant domain 2
CH3 or C_H3	Immunoglobulin gamma heavy chain constant domain 3
CHO	Chinese hamster ovary
CIC	Cross-interaction chromatography
CID	Collision induced dissociation

cIEF	Capillary isoelectric focusing
CIU	Collision-induced unfolding
C_L	Immunoglobulin gamma light chain constant domain
CMC	Manufacturing and Control
COSY	Correlation spectroscopy
cP	Centipose
CPL	Circularly polarized light
CQA or CQAs	Critical quality attribute(s)
CSA	Camphor sulfonic acid
CZE	Capillary (free) zone electrophoresis
D	Deuterium or translational diffusion coefficient or electric dipole
DAC	Deutscher arzneimittel-codex
DC	Direct current
DI	Developability index
DLS	Dynamic light scattering
DoE	Design of experiment
DNA	Deoxyribonucleic acid
DOSY	Diffusion ordered spectroscopy
DP	Drug product
dPLIMSTEX	Dilution PLIMSTEX
DRI	Differential refractive index detector
DS	Drug substance
DSA	4,4-dimethyl-4-silapentane-1-ammonium trifluoroacetate
DSC	Differential scanning calorimetry
DSF	Differential scanning fluorimetry
DSLS	Differential static light scattering
DSS	4,4-dimethyl-4-silpentane-1-sulfonic acid
DTT	Dithiothreitol
ECD	Electron capture dissociation or equivalent circular diameter
ECHOS	Easy comparability of higher order structure
EDTA	Ethylene diamine tetra-acetic acid
EM	Electromagnetic radiation or electron microscopy
EMEA	European Medicines Agency
ESD	Equivalent sphere diameter
ESI	Electrospray ionization
ESZ	Electrical sensing zone
ET	Electron tomography
ETD	Electron transfer dissociation
EX1	H/D exchange mechanism in which the rate constant for protein folding/unfolding is much slower than the rate constant for H/D exchange
EX2	H/D exchange mechanism in which the rate constant for protein folding/unfolding is much faster than the rate constant for H/D exchange
Fab	Immunoglobulin gamma fragment antigen binding
Fc	Immunoglobulin gamma fragment crystallizable (constant region)
FcγRIIIa	Immunoglobulin gamma Fc receptor RIIIa
FcRn	Neonatal Fc receptor
FDA	Food and Drug Administration
FFF	Field flow fractionation
FID	Free induction decay

FIX	Blood clotting factor IX
FL	Fluorescence
FT-ICR	Fourier transform ion cyclotron resonance
FTIR or FT-IR	Fourier transform infrared spectroscopy
fuc	Fucose
FUV-CD	Far ultraviolet circular dichroism
FVIII	Blood clotting factor VIII
gal	Galactose
GlcNAc	N-acetylglucosamine
GLP	Good laboratory practices
GMP	Good manufacturing practices
H/DX-MS or HDX-MS	Hydrogen/deuterium exchange mass spectrometry
HSA	Human serum albumin
HCH	Human growth hormone
HCl	Hydrochloric acid
HCLF	High concentration liquid formulation
HDC	Hydrodynamic chromatography
HDX	Hydrogen/deuterium exchange
HF5	Hollow fiber flow field flow fractionation
HGH	Human growth hormone
hI	Hydrophobicity index
HIC	Hydrophobic interaction chromatography
HILIC	Hydrophilic interaction chromatography
HMQC	Heteronuclear multiple quantum coherence spectroscopy
HMW	High molecular weight
HOS	Higher-order structure
HPLC	High performance liquid chromatography
HRR-DSC	High ramp rate differential scanning calorimetry
HSQC	Heteronuclear single quantum coherence spectroscopy
HT	High tension
HX	Hydrogen exchange
ICH	International conference on harmonization of technical requirements for registration of pharmaceuticals for human use
icIEF	Imaging capillary isoelectric focusing
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IEC	Ion-exchange chromatography
IEF	Capillary isoelectric focusing
IF	Intrinsic fluorescence
IFN IFNβ or IFNβ1a	Interferon-β-1a
IgG1	Immunoglobulin gamma 1 or immunoglobulin G1
ILP	Integer linear programming
IM	Ion mobility
ITC	Isothermal titration calorimetry
IUP	Intrinsically unstructured protein
IUR	Intrinsically unstructured region
IV	Intravenous injection
k_D	Diffusion Interaction Parameter
LC/MS	Liquid chromatography/mass spectrometry

LMW	Low molecular weight
LNPs	lipid nanoparticles
LO	Light obscuration
LOQ	Limit of quantitation
LS	Light scattering
mAb	Monoclonal antibody
MALDI	Matrix-assisted laser desorption/ionization
MALLS or MALS	Multiangle laser light scattering
man	Mannose
MD	Molecular dynamics
MEM	Maximum entropy method
MEMS	Micro-Electro-Mechanical Systems
MFI	Micro-flow imaging
MHz	Megahertz
MRE or [M.R.E]	Mean residue ellipticity
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS/MS	Mass spectrometry/mass spectrometry or tandem mass spectrometry
MW	Molecular weight
NEM	N-ethylmalemide
NIBS	Noninvasive back scattering technique
nIEF	Native isolelectric focusing
NIST	National Institute of Science and Technology
NMR	Nuclear magnetic resonance or nuclear magnetic resonance spectroscopy
NNLS	Nonnegative least squares
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect spectroscopy
NTA	Nanoparticle tracking analysis
OCD	Oriented circular dichroism
OD	Optical density
OQ	Operation qualification
OS-GAGs	Oversulfated glycosaminoglyclans
PBS	Phosphate buffered saline
PCA	Principal component analysis
PDA	Photodiode-array
PDB	Protein data bank
PDI	Polydispersity index
PEG	Polyethylene glycol
PEM	Photoelastic modulator
PFG	Pulsed field gradient
PFGE	Pulsed field gradient echo
Phe	Phenylalanine
pI	Isoelectric point
PK	Pharmacokinetic
PL	Path length
PLIMSTEX	Protein–ligand interactions by mass spectrometry, titration, and H/D exchange
PMT	Photomultiplier tube
PPC	Procedure Performance Criterion
PPI	Protein-protein interactions

PPQ	Procedure Performance Qualification
PQ	Performance qualification
Pr	Probability
PTM or PTMs	Posttranslational modification(s)
QA	Quality Analysis
QbD	Quality by design
QToF	Quadrupole time of flight
QTPP	Quality target product profile
RDCs	Residual dipolar couplings
RF	Radio-frequency
rhGM-CSF	Recombinant human granulocyte colony stimulating factor
rhuEPO	Recombinant human erythropoietin
RI	Refractive index
rmAb	Recombinant monoclonal antibody
RMM	Resonant mass measurement
RP-HPLC, RPLC or rpLC	Reversed-phase high performance liquid chromatography or reversed-phase liquid chromatography
RPS	Resistive pulse sensing
RS	Reference standard
RT	Room temperature or retention time
S or σ	Standard deviation
S/N	Signal to noise
SANS	Small angle neutron scattering
SAP	Spatial Aggregation Propensity
SAXS	Small angle X-ray scattering
SC or SubQ	Subcutaneous injection
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE-AUC	Sedimentation equilibrium analytical ultracentrifugation
SEC	Size-exclusion chromatography
SE-HPLC or HP-SEC	Size-exclusion high performance liquid chromatography or high-performance size-exclusion chromatography
SEM	Scanning electron microscopy
SFC	Supercritical fluid chromatography
SIC	Self-interaction chromatography
SIMCA	Soft independent modeling of class analogy
SIMSTEX	Self-association interactions by mass spectrometry, self-titration, and H/DX
SLS	Static light scattering
SMP	Submicron particles
SMR	Suspended microchannel resonator
SPE	Solid-phase extraction
SRCD	Synchrotron radiation circular dichroism
STEM	Scanning transmission electron microscopy
SUPREX	Stability of unpurified proteins from rates of H/D exchange
SV-AUC	Sedimentation velocity analytical ultracentrifugation
SVD	Singular value decomposition
SVP	Subvisible particles
T1	Longitudinal relaxation time constant
T2	Transverse relaxation time constant
TCEP-HCl	Tris(2-carboxyethyl)phosphine hydrochloride

TDA	Taylor dispersion analysis
TEM	Transmission electron microscopy
TIC	Total ion current
TM-DSC or MT-DSC	Temperature-modulated differential scanning calorimetry
TMU	Target measurement uncertainty
TPP	Target product profile
Try	Tyrosine
TSP	Trimethylsilyl propionate
Tyr	Tryptophan
UPLC or UHPLC	Ultrahigh performance liquid chromatography or ultra-performance liquid chromatography
USP	United States Pharmacopeia
UV	Ultraviolet light
UV–VIS	Ultraviolet–visible spectroscopy
V_H	Immunoglobulin gamma heavy chain variable domain
VIS	Visible light
V_L	Immunoglobulin gamma light chain variable domain
VLP or VLPs	Virus-like particle(s)
WAXS	Wide angle X-ray scattering
WCX	Weak-cation exchange chromatography
WSD	Weighted spectral difference
XIC	Extracted ion chromatogram
Z	Net charge

CHAPTER

1

The complexity of protein structure and the challenges it poses in developing biopharmaceuticals

Steven A. Berkowitz[a], *Damian J. Houde*[b]

[a]Consultant, Sudbury, MA, United States; [b]Biomolecular Discovery, Relay Therapeutics, Cambridge, MA, United States

1.1 The basics of protein higher order structure (HOS)

Proteins are an important class of large biological molecules that are classified more generally as macromolecules or polymers. However, given their biological origin, these unique molecules are often referred to as biomacromolecules or biopolymers. They are truly complex, particularly when compared to synthetic (man-made) polymers and even other types of biopolymers, e.g., DNA. One of the main reasons for this complexity arises from their basic building blocks, which in synthetic polymer chemistry are referred to as monomer units. In the case of most synthetic polymers, the chemical composition consists typically of only one type of monomer (although some synthetic polymers called copolymers or block-copolymers are composed of two or possibly more different monomer units). Proteins made in nature via a process called translation utilizing the genetic code are composed of not one, two, or even three different monomer units, but rather are composed of as many as 20 different "naturally" occurring monomer units called *amino acids*. These 20 amino acids (or proteinogenic amino acids, which does not include the other know, but rare proteinogenic amino acids selenocystine or pyrrolysine) are referred to as the standard amino acids. Although not all proteins contain all 20 amino acids, most do. The presence of such a large diversity in chemical composition, in virtually every protein, is a key element for their structural complexity, which in turn gives rise to their diverse functionality. Indeed, this chemical complexity, coupled with the large number of amino acid units or *residues* (*N*) present in proteins (that can number in the thousands), and the uniqueness of the amino acids linear sequential arrangement (which in protein chemistry is called the *primary* (1°) structure, see Fig. 1.1A), enables a

Biophysical Characterization of Proteins in Developing Biopharmaceuticals, Second Edition
https://doi.org/10.1016/B978-0-444-64173-1.00001-9

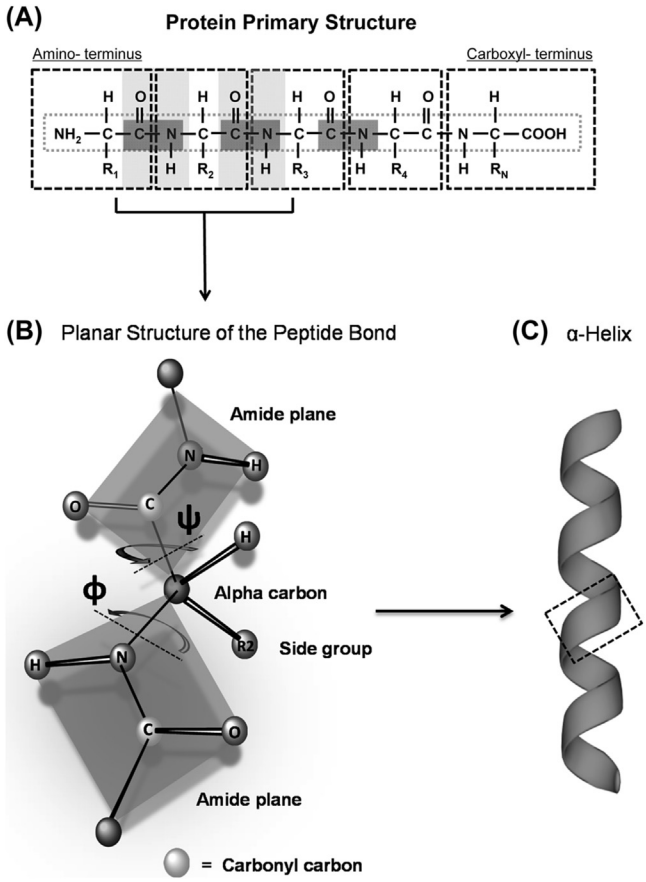

FIG. 1.1 (A) The linear sequential ordering of amino acids (represented by the rectangular black dashed boxes) in a protein is referred to as its primary structure. The extreme left amino acid corresponds to the amino-terminus, while the extreme right amino acid corresponds to the carboxyl-terminus end of the protein chain. The gray shaded area corresponds to the peptide bonds that link all the amino acid units in a protein, yielding the polypeptide backbone (or chain) indicated by the red (gray in print version) dotted rectangle. (B) An illustration of the planar structure of two adjacent amide planes (each resulting from the double bond character, due to resonance, of the peptide bond shown as black dashes), corresponding to the light blue (light gray in print version) shaded areas in (A), where the bottom amide plane is formed from the peptide bond between the carboxyl group of amino acid 1 (containing R_1) and the amino group of amino acid 2 (containing R_2) and the top amide plane is formed from the peptide bond formed between the carboxyl group of amino acid 2 and the amino group of amino acid 3 (containing R_3). Due to steric issues, angular rotation around CαN (expressed by Φ, phi) and CCα (expressed by Ψ, psi) bonds are limited. (C) A representation of a common secondary structure, the α-helix. The small rectangle outlined in black dashes corresponds to a small section of the helical arrangements of the amide planes, shown in (B).

staggering number of different possible proteins, 20^N, to be made. Given the enormous array of different proteins that can be made, the cell has exploited this diversity in protein structure to create proteins to perform nearly every functional and structural role needed for its existence.

In proteins, the amino acid units are linked together through a unique chemical bond called the *peptide bond*, which is also referred to as the amide link, see Fig. 1.1A. The collection of these peptide bonds in a given protein form a common element found in all proteins called the *polypeptide backbone* or *chain*, see Fig. 1.1A. A unique feature of the peptide bond is the planar structure that it forms between the carbonyl oxygen, carbon and the α-carbons (Cα or alpha carbon) of one amino acid and the amide nitrogen, hydrogen and α-carbons of an adjacent amino acid. The resulting planar feature of these linked atoms arises as a result of the partial double bond character that exists between the carbonyl carbon (C) and the amide nitrogen (*N*) atoms due to the presences of resonance structures, see Fig. 1.1B. This planar structure and its attributes play an important role in a protein's structure, as its presences confines the polypeptide backbone to only certain configurations, via steric effects, which restricts the angular range of bond rotation around the Cα−*N* (expressed by Φ, phi) and C-Cα (expressed by Ψ, psi) bonds. These restrictions have been summarized in a 2-dimensional graphical plot called the Ramachandran plot, developed by Ramachandran and others in 1963 [1]. Such a plot graphically shows how certain structural features of proteins can only exist within a limited range of angles characterized by Ψ and Φ, e.g., α-helix, see Fig. 1.1C. These restrictions play an important role in the development of protein's spatial structure or *higher order structure* (*HOS*).

1.1.1 The levels of protein HOS

In developing protein biopharmaceuticals and in studying proteins in general, the most important concept is "structure". In the previous section, we briefly discussed the most basic component of a protein's structure, its linear sequence of amino acids, or primary structure. However, the focus of this book is concerned with a protein's *three-dimensional* (3D) or spatial *structure*, also referred to as its *conformation* or HOS. Ultimately, when considering the structures of proteins, it is the HOS in concert with its primary structure (which also includes all the primary chemical bond modifications that occur to its amino acid units, see Section 1.1.4) that enables a protein to properly function or, as we will also discuss in latter sections, malfunction.

In terms of protein HOS, there are three different levels that have been defined. These three levels include: *secondary* (2°), *tertiary* (3°), and *quaternary* (4°) structure, see Fig. 1.2. The first two structural levels are concerned with a single polypeptide chain, while the latter is associated with protein structures that involve the interaction of two or more polypeptide chains. A protein's 2° structure refers to the local folding patterns of a protein's polypeptide chain, in which the α-helix (see Fig. 1.2A), the β-sheet, turns, and random coils are the most prominent resulting structural elements that are formed. These local folded elements can further participate in higher levels of folding that involve an array of secondary structural elements that give rise to the final 3D structure of a protein referred to as 3° structure of a protein; see Fig. 1.2B. The summation of 2°, 3° and (if present) 4° structure, along with its entire 1° structure, is what gives a protein its unique structure, chemical and physical properties and therefore its unique function. Indeed, it is this relationship between structure and function that is the genesis of the protein "structure-function" concept, which states that a protein's structure determines its function.

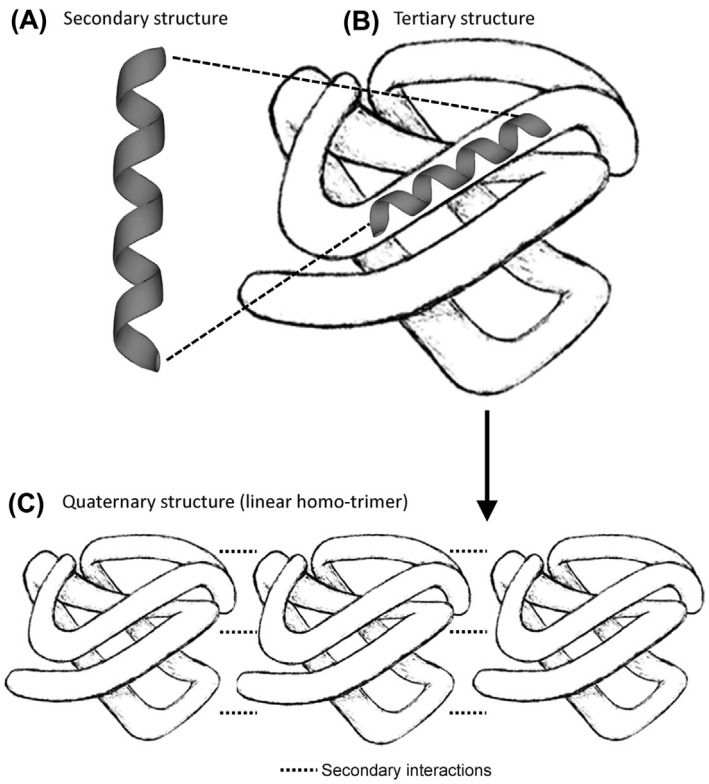

FIG. 1.2 Illustration of the three levels of a protein's HOS. (A) Representative secondary structural element, as illustrated by a ribbon representative structure of an α-helix. (B) A cartoon representation of the folding of all the secondary structural elements in a polypeptide chain, which gives rise to the polypeptide's tertiary structure. (C) A cartoon representation of the quaternary structure of a protein, which arises when the final protein structure involves the association of more than one polypeptide chain to form the final folded protein structure (also see Fig. 1.3).

Although the folding and interactions of the secondary structural elements can give rise to an enormous array of different protein tertiary structures, each with unique properties and functions, it's not uncommon to find that the 3° structure of a protein often consists of one or more commonly folded patterns called *motifs*, *super-secondary structures*, or *complex folds* [2–4]. These commonly folded structures contain several folded secondary elements involving only a portion of the entire polypeptide chain of a protein, which can blur some of the distinction between a protein's 2° and 3° structure. Hence, one might look at motifs, super-secondary structures, or complex folds as "local 3°structure", while referring to the 3° structure of the entire protein molecule as its "global 3°structure".

Another structural element that further subclassifies the structural level of a protein between what we call a protein's 2° and 3° structure is the concept of *domain* [5,6]. Domains are typically a much larger collection of folded structural elements than motifs, supersecondary structures, or complex folds. In terms of the global structure of a protein, domains correspond to one or more independent compact region of a protein's polypeptide chain, as

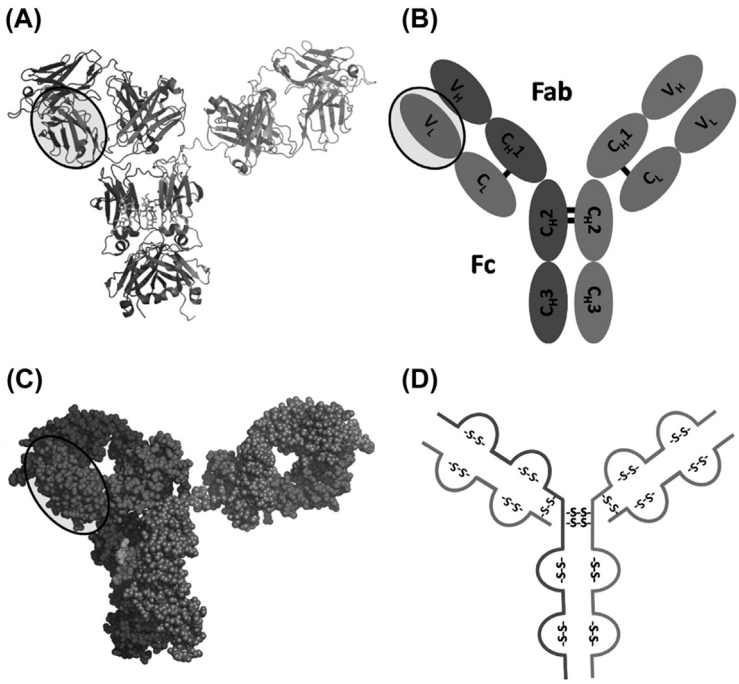

FIG. 1.3 Different representations of the HOS of a monomeric IgG1 antibody. The two heavy chains are color-coded in blue (light gray in print version) and gray, while the two light chains are both color-coded in red (dark gray in print version). (A) A ribbon model of an IgG1 antibody (PDB: 1HZH). The black circle corresponds to the variable domain on one of the IgG1 light chain (V_L). (B) A simplified cartoon of the monomeric IgG1 antibody indicating the various sections of individual domains present. The black lines linking the various interchain domains correspond to areas where covalent linkages exist (disulfide bonds) between different polypeptide chains in the IgG1 molecule. The black circle corresponds to the same V_L domain in the IgG1 molecule as shown in (A). (C) A space-filling structural model of the monomeric IgG1 antibody. The black circled region again corresponds to the same V_L domain in the IgG1 antibody as shown in (A). (D) A linear depiction of a monomeric IgG1 structure showing all the various covalent linkages (disulfide bonds) present in the IgG1 antibody. Those disulfide bonds present within the same polypeptide chain are referred to as *intrachain* disulfide bonds, while those disulfide bonds that link two different polypeptide chains are referred to as *interchain* disulfide bonds.

indicated by the black circles shown in Fig. 1.3A–C. Proteins containing two or more domains are frequently referred to as multidomain proteins. In these proteins, the domains are chemically linked by short sections of the polypeptide chain that are typically highly flexible, called a "linker", but nevertheless exist as stable and independent folded units. In certain cases, common domain structures can also be found in proteins much like that observed with motifs, super-secondary structures, or complex folds.

What is interesting about these folded elements is that there is a certain amount of change in the 1° structure that can be tolerated while still arriving at, effectively, the same folded structure. This observation explains the common presence of similar secondary, super-secondary, and even domain structures seen in different proteins with different sequences. Hence, the formation of these basic folding elements can display some level of discrepancy in terms of the required or allowable amino acid sequence variations and still give rise to

the same functioning protein. This feature plays an important role in biological evolution, in generating HOS building blocks, and in controlling and regulating groups of proteins that perform very similar functions in different biochemical pathways [7–9]. Nevertheless, it is important to mention that in proteins, there exist many sequence regions where even a slight change, i.e., one amino acid change or a minor chemical modification (e.g., oxidation, deamidation), can significantly alter a protein's structure and therefore its function [10,11].

For many proteins, however, the unique folded state of its polypeptide chain is not the last step in attaining a final overall 3D structure. Many proteins are composed of more than one polypeptide chains, which may be identical or nonidentical, giving these proteins an added level of structural complexity, 4° structure; see Fig. 1.2C.

When referring to a protein's 4° structure, a lack of clarity or confusion can unfortunately arise. An example is illustrated in Fig. 1.3. In this figure, a monomeric intact IgG1 antibody is shown. However, this protein could be referred to as a protein dimer (made of two identical protein units) or a protein tetramer made of four separate polypeptide chains, which in this case are chemically cross linked via covalent bonds called disulfide bonds (which is the most common primary bond used in nature to cross-link parts of polypeptides). Such a choice of descriptive words unfortunately can lead to some confusion. As a result, some care should be taken when describing the basic structure of a protein. In the case of the 4° structure of IgG1 molecule, as shown in Fig. 1.3, the use of a tetramer in the context of its 4° structure would be correct. However, in the context of a complete functioning unit (in its lowest complete form) the molecule is a monomer.

1.1.2 Stabilizing the HOS of proteins

In all three levels of a protein's HOS (i.e., 2°, 3°, and 4°), various changes in the conformation of the polypeptide chain(s) occur as a protein folds to reach its final native structure. These changes are typically accompanied by an increase in overall structural order, which imparts a significant reduction in the protein's entropy that by itself is highly unfavorable, in terms of the overall free-energy change. However, as a protein folds, various weak noncovalent (secondary) bonds form via ionic, dipoles (hydrogen bonds), nonpolar (hydrophobic effect), and van der Waals interactions. These weak bonds involve the interactions of amino acid side chains, as well as elements of the polypeptide backbone, particularly the amide hydrogen. While individually these interactions are weak, during the folding process their large number and the cooperative way they form provide the necessary enthalpic and entropic driving forces (release of structured water via the hydrophobic effect) to override the large unfavorable decrease in entropy that occurs as a protein folds into its native (more ordered) conformation. The stabilization of the folded protein, however, is only marginal. Comparing the level of stabilization against the average thermal energy content of a protein molecule (which is equal to kT, where $k =$ Boltzmann constant and $T =$ temperature) and the distribution of this energy, in terms of the amount of thermal energy per molecule, a variety of these weak secondary bonds can be broken as a function of time. Such spatial and temporal rupturing of these weak secondary bonds enables a protein to display dynamic structural properties in its conformation (sometimes referred to as *protein breathing*). This dynamic property can play an important role in a protein's function [12–15] and stability [16,17]. This dynamic property, however, can also constitute a weakness for protein biopharmaceuticals, given the

wide range of stressful environments an average biopharmaceutical must endure during its biosynthesis, purification, formulation, packaging/storage, patient handling, and its administration. Hence, in searching for a good therapeutic biopharmaceutical, scientists look for molecules with high stability, such that the dynamic properties of the protein do not result in loss of activity or adverse structural changes. Proteins that have such attributes are said to have good developability properties.

In addition to weak secondary bonds, stabilization of the HOS of a protein can be achieved through primary bonds formed between folded elements within a protein. As already mentioned, the most common such bond is the disulfide bond, see Fig. 1.3D. Although the number of disulfide bonds found in a given protein typically amounts to only a few such bonds per protein molecule (and may not even exist within some proteins), they often play important roles in a protein's overall structure-function and stability [18]. Disulfide bonds can occur both within a single polypeptide chain (where they are referred to as *intrachain* disulfide bonds; see Fig. 1.3C) and between two different polypeptide chains in the same protein (where they are referred to as *interchai*n disulfide bonds; see Fig. 1.3C). Disulfide bonds also occur between two different protein molecules where they function to stabilize large complex multiprotein supramolecular structures [19]. Unfortunately, however, the formation of disulfide bonds can go astray leading to altered HOS structures or aggregates via disulfide scrambling or exchange between other disulfide bonds or free cysteine residues in the same protein or different proteins. These modes of protein degradation [20–27] are another reason why the biopharmaceutical scientist need to constantly scrutinize the structure of the biopharmaceutical during its development.

1.1.3 Dynamics properties of a Protein's HOS

The HOS of virtually all proteins is primarily held together by a large array of relatively weak bonds. In the context of a protein's thermal energy content, these bonds can break enabling various levels of fluctuations within a protein's HOS that can span an enormous time range, from 10^{-15} s to tens of seconds and even longer [12,28]. Again, the fluctuations in a protein's conformation essentially occur because of the opening or breaking of various weak secondary bonds. The extent of these fluctuations in terms of amplitude and location is very dependent on many factors, e.g., environmental conditions, the strength of each secondary bond, the distribution of these bonds within the protein, as well as the distribution of thermal energy within the protein. Variations in these (and other) factors will determine the location of which secondary bonds will break in a protein's HOS and therefore, the nature of the conformational change(s) and the population of protein molecules in a specific conformation as a function of time. While these changes are for the most part contained to a region where the secondary bond(s) break, changes might also extend to other areas of the protein, via allosteric effects. Due to the random nature of the thermal energy fluctuations within a protein, a range of different conformations and populations of different conformational states will exist at any one time. For the most part, the extent of change in a protein's HOS are typically not that large and are often reversible allowing the altered protein structure to return to its more stable conformations.

Consequently, in solution proteins exist as an *ensemble* of different conformations, rather than as a single fixed unique conformation. This ensemble is limited and controlled by the

interplay of the overall structure of the protein and its physicochemical environment. However, under appropriate conditions, involving some form of stress or subtle changes in a protein's chemical structure, changes in conformation may cause a protein to display different physicochemical properties. In the case of a protein biopharmaceutical, changes in its physicochemical properties could alter the drug's ability to bind with its therapeutic target or enable it to bind to different materials it encounters, e.g., various container closure surfaces [29–33]. Other possible adverse events include the formation of aggregates that are nonfunctional and/or even more concerning, immunogenic [34–36]. It should be noted that the formation of aggregates and their associated link to loss of protein function and/or immunogenicity corresponds to one of the most common forms of protein degradation that is closely monitored in the biopharmaceutical industry.

1.1.4 Finer structural alteration of proteins

Once a protein is synthesized, or as it is being synthesized, additional primary structural changes can occur in vivo. In most cases, these changes are due to additional enzymatic processing reactions involving a multitude of potential chemical modifications to various amino acids, as well as changes involving cleavage or cross-linking reactions. These reactions may or may not play an important role in the normal function/activity of a protein, but rather may represent alterations that play out to the determent of the cell or even the organism due to an immunogenic response. Generally, most modifications are confined to the protein's surface. However, modifications can also occur to the protein's interior due to the dynamic properties of its structure (which exposes these buried internal areas) or during its synthesis when these normally buried internal areas had not had a chance to properly fold. Such alterations can lead to changes in the local or global HOS of the protein. In general, these modifications are referred to as *posttranslational modifications* (PTMs). Principally, PTMs occur in vivo and the number of different PTMs that a protein can experience is quite large [37]. In eukaryotes, one of the more common (and biopharmaceutically relevant) PTMs is glycosylation. This modification involves the enzymatic addition of carbohydrate (also called glycan or sugar) units to a protein at specific asparagine (where they are called N-linked glycan) or serine or threonine (where they are called O-linked glycan) amino acid [38]. While most PTMs occur in vivo (inside the cell), PTMs can also occur in vitro (outside the cell). These latter PTMs, however, typically represent forms of protein degradation that occur due to direct physical or chemical interactions (e.g., oxidation, deamidation, glycation, etc.) and are also of great concern in the biopharmaceutical industry as they are often linked to instability leading to lose of drug activity and adverse effects [3,39–44].

1.2 The search for how proteins attain their correct HOS: the protein folding problem

In the 1950s and 1960s, biophysical research led scientists to the realization that a protein's HOS is effectively dictated by its primary sequence. Christian Anfinsen was the key scientist who formalized this idea, and in 1972 was awarded the Nobel Prize in chemistry for his contributions [45]. In the scientific literature, this idea has been frequently referred to as the

"Anfinsen dogma" or the "thermodynamic hypothesis". The folding path a protein takes to achieve its correct functional HOS is intrinsically dictated by its 1° structure (which may also include PTMs). How the folding process advances so efficiently, in combination with the way a protein is synthesized in vivo, in the specific physicochemical environment within the cell, has fascinated scientists for many years [46]. This fascination stems from the realization that proteins achieve their correct HOS within a matter of milliseconds to seconds!

In the 1960s, Cyrus Levinthal prosed the following interesting and simple problem concerning protein folding. For a protein consisting of 100 amino acids in an initially unfolded state, how long would it take this protein to find, through a completely random process, its appropriate native HOS given its physicochemical environment [28]? This problem is nicely restated in the words of Amit Kessel and Nir Ben-Tal in their book "Introduction to Proteins: Structure, Function and Motion" [47] as follows:

> Assuming that the protein folding process involves the free sampling of all possible conformations of the protein (i.e., of each residue independently), and that each residue has at least three states, then the folding of a 100-residue protein is excepted to sample $3^{100} = 5 \times 10^{47}$ conformations. Now if we assume that it takes a protein 1 picosecond to sample a single conformation, then the time it takes to sample all possible conformations in order to find the right one should be $3^{100} \times 10{-}^{12}$ s $= 5 \times 10^{35}$ s $= 1.6 \times 10^{28}$ years. This period of time is about 10^{18} times longer than the age of the universe!!

This simple problem proposed by Levinthal is called "Levinthal's Paradox" and was a significant driving force for the generating what is called "the protein folding problem". Clearly, the nature of protein folding is nowhere as simple as starting with the completely synthesized and unstructured (denatured or random coil) form of a protein, which is then allowed to undergo a completely random sampling process of conformational space. Protein folding must proceed via a process that is enormously more efficient, but how!!? Answers to this problem appear to lie within the idea of a "funnel-shaped folding energy landscape" [48–52], see Fig. 1.4, which might possibly take advantage of the way proteins are made in vivo along with a concept of "divide and conquer". In this process a protein proceeds to fold through a hierarchy of subassembly units called a "foldon" [53,54]. These units can fold somewhat independent of each other in parallel to form relatively local higher order structures that can eventually collapse into the final native HOS of the protein.

In general, the funneling process of protein folding is likely not as simple as that portrayed in Fig. 1.4A. Rather, it is expected to be more complex and treacherous, as indicated in Fig. 1.4B. In the latter scenario, a folding protein could encounter conformational states that are not as optimally folded as its native state and contain high activation energy barriers that inhibit its search to find the most stable conformation. Hence, the protein in these states would find itself trapped, due to the high energy of activation needed to transition the misfolded state back into a more unfolded state so it can find its more stable and native form. Although these misfolded protein forms may be encountered at very low levels under normal conditions, the situation could escalate under stressed conditions, such as forcing a cell to produce a large quantity of one protein in a very short period. For such a situation, a higher frequency of misfolded or metastable folded protein states could be encountered leaving the biopharmaceutical scientist with a more difficult purification process that results in a lower protein drug yield.

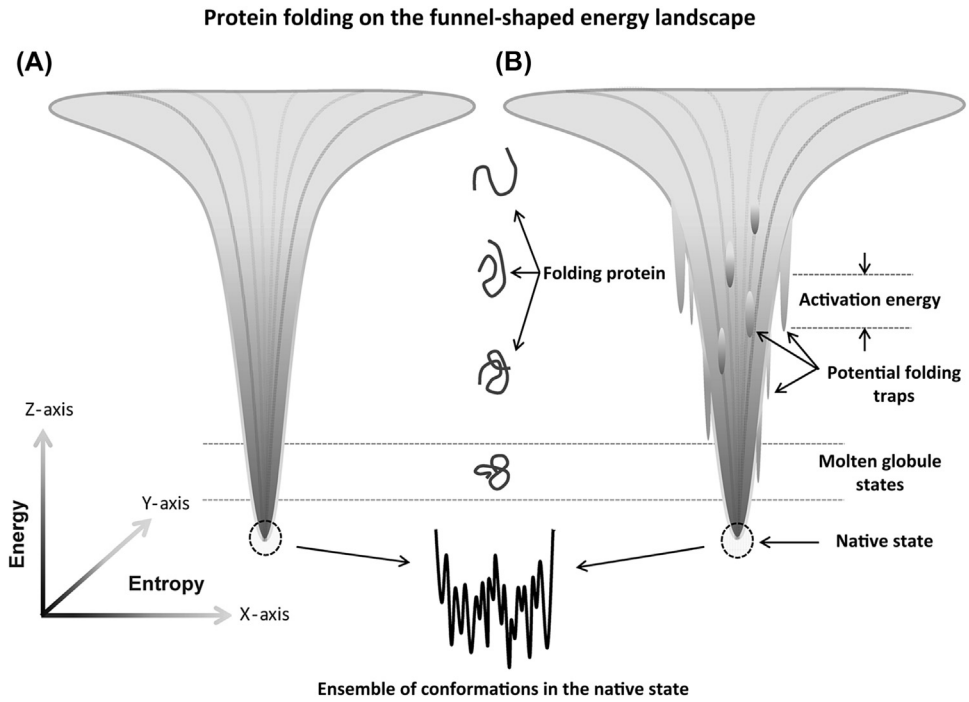

FIG. 1.4 A graphical view of the three-dimensional funnel-shaped energy landscape for protein folding. The top of each funnel corresponds to the completely unfolded protein. The bottom of each funnel plot corresponds to the fully folded protein molecule in its native state, which under closer scrutiny actually consists of a large array of slight different energetically folded states (conformations) that differ in most cases by a small amount of free energy thus enabling the native protein to exist in solution as an ensemble of different conformations. (A) A folding process free of situations where it can be trapped in incomplete or partial folded state. (B) A folding process that enables partially folded proteins to be potentially trapped due to the presence of smaller shaped folding funnels with relatively large energy of activation that must be overcome in order to escape and find its final native state.

1.2.1 In vivo production of proteins: revisiting the protein folding problem

Another unique attribute of proteins is the complex manner with which they are made in vivo. Protein synthesis involves a complex array of cellular machinery, the main component of which is the ribosome. In vivo, proteins are synthesized from the N-terminus to the C-terminus in a sequential manner at a rate of 50–300 amino acids/min [55,56]. The specific ordering and chemical coupling of the amino acids for a given protein is achieved by a process called *translation*, which controls the protein synthesis process dictated by the genetic coding information stored in a specific messenger RNA (m-RNA). As the nascent protein chain is synthesized and exposed to the cell matrix, it can begin to fold. However, it should be noted that the first 50–60 amino acids in the growing polypeptide are initially limited to some extent in their ability to freely fold, due to the physical restrictions (steric hindrance) of the environment within the ribosome [57]. This idea of concurrent, in vivo, protein synthesis and folding are referred to as *cotranslational protein folding* [58] and likely plays an important role in the folding of newly synthesized polypeptide.

The importance of cotranslational protein folding most likely arises because only the growing polypeptide chain that has advanced beyond the ribosome tunnel will be able to fully participate in the folding process. This allows only a portion of the growing protein chain to fold without the interference from other parts of the protein that has either not been synthesized or is located in the ribosome tunnel. As a result, this should improve the efficiency of the sequential folding of the local higher order structural elements characterized as foldon units to proceed in a more orderly manner. Such foldon units most likely correspond to local HOS elements that are present in the final native protein. Nevertheless, as these local higher order structural elements are formed, they must search out and undergo higher levels of folding as the protein chain continues to grow. As a result, these various hierarchy of folded structural elements are probably not arranged or packed optimally (as they are in the protein's final native state) until the entire protein is fully synthesized and release from the ribosome. Once this happens, what remaining loose arrangement of folded (or partially folded) structural elements that still exist must collapse into the final native structure of the protein. This final consolation of folded or partially folded structural elements most likely proceeds through the interactions of key amino acid side chains to make the final functional protein (notwithstanding any additional changes in HOS resulting from PTMs and the formation of quaternary structures).

Consequently, cotranslational protein folding constrains and to some extent guides the overall protein folding process. By limiting the number of folding pathways (specifically bad folding pathways, which would significantly increase the amount of time required to find the correct native conformation) available to a protein, relative to the situation where folding only begins once the protein is fully synthetized and release from the ribosome, could make the role of cotranslational protein folding truly an important attribute in the successful folding of a protein.

1.2.2 In vivo production of proteins: avoiding and eliminating folding errors via the use of chaperones

In vivo, there are mechanisms involving other proteins, called *chaperones*, that help proteins that are folding avoid the situation of being misfolded. This task is achieved via the chaperone's ability to assist a folding protein to avoid folding traps by participating in the protein folding process through protein–protein interactions [59–63]. In addition to chaperones, there also exists in vivo cellular machinery whose function is to identify the presences of misfolded proteins and eliminate them via proteolytic hardware existing within the cell [64]. However, these systems are not perfect, and failure to remove or prevent these erroneously folded proteins from accumulating within the cell can alter the cell, causing adverse effects that could eventually lead to its death. In the case of producing a protein biopharmaceutical, once a misfolded protein is released into the cell culture media, it then becomes the problem for the process scientist to develop appropriate purification strategies to remove the misfolded protein from the final protein drug product. If these erroneously folded proteins are not removed, they could lead to adverse effects when the final drug product is administered to a patient. Hence biophysical analysis of the biopharmaceutical's HOS again becomes an important activity in developing protein biopharmaceuticals with minimal levels of these misfolded forms in the final drug product.

1.3 Surprises in the world of protein folding: intrinsically disordered or unstructured proteins (an apparent challenge to the protein Structure–Function paradigm)

Within the past two decades, it has been realized that many proteins, especially in eukaryotes or multicellular organism, exist within the cell with no defined HOS [28]. Rather, these proteins appear to be disordered or unstructured, approaching what might be called a random coil structure, anomalous to what is frequently seen with synthetic polymers or denatured proteins. However, when these proteins interact with their target molecule(s) they commonly appear to take on a level of organized HOS. Hence, this structural disorder is transient in many cases and a disorder-to-order transition occurs during their functioning (i.e., interacting with their binding target). Such behavior could play an important role in allowing these proteins to bind to an array of different partner molecules by taking advantage of the plasticity of their polypeptide chain's flexibility [28]. This process is liable to be modulated by other factors within the cell, which control and regulate the binding partners they interact with. Indeed, the level of disordered proteins is higher in eukaryotes or multicellular organism, in comparison to prokaryotes, where high levels of signaling and regulation is required. This unique class of proteins has been referred to as intrinsically disordered proteins (IDPs) or intrinsically unstructured proteins (IUPs) [65,66]. The existence of these IDPs would appear to present a challenge to the paradigm of structure–function discussed earlier in this chapter.

With the realization of the existence of IDPs, many of the large random coil-like regions of proteins consisting of 20–30 or more amino acids in length are now being referred to as intrinsically disordered or unstructured regions (IDRs or IURs) [66]. These structural elements are commonly seen as linkers between ordered protein regions such as domains where they are thought to also play important roles in providing protein flexibility, allowing proper folding or to facilitate domain–domain interactions or domain binding to functioning binding targets. At present, IDPs have not made their way into the biopharmaceutical industry, although it is probably only a matter of time until such a protein drug will appear.

1.4 Proteins and the biopharmaceutical industry: problems and challenges

Although proteins can be chemically synthesized external to the cell, their high cost (which is a function of protein size), as well as their overall complexity leads to poor economics for building a viable commercial drug business via this approach. Over the years, however, scientists have figured out how to get cells to produce significantly large amounts of a specific protein, by manipulating their DNA via recombinant DNA technology. The development of this capability was the key in enabling the biopharmaceutical drug industry to flourish. Overall this process of making protein biopharmaceutical differ greatly from the classical process used to make simple organic drug molecules called pharmaceuticals, see Fig. 1.5. Cellular and molecular biologist can now produce protein biopharmaceuticals at concentrations expressed in the culture media volume as great as 10 g/L [67]. Nevertheless, the challenges of doing this

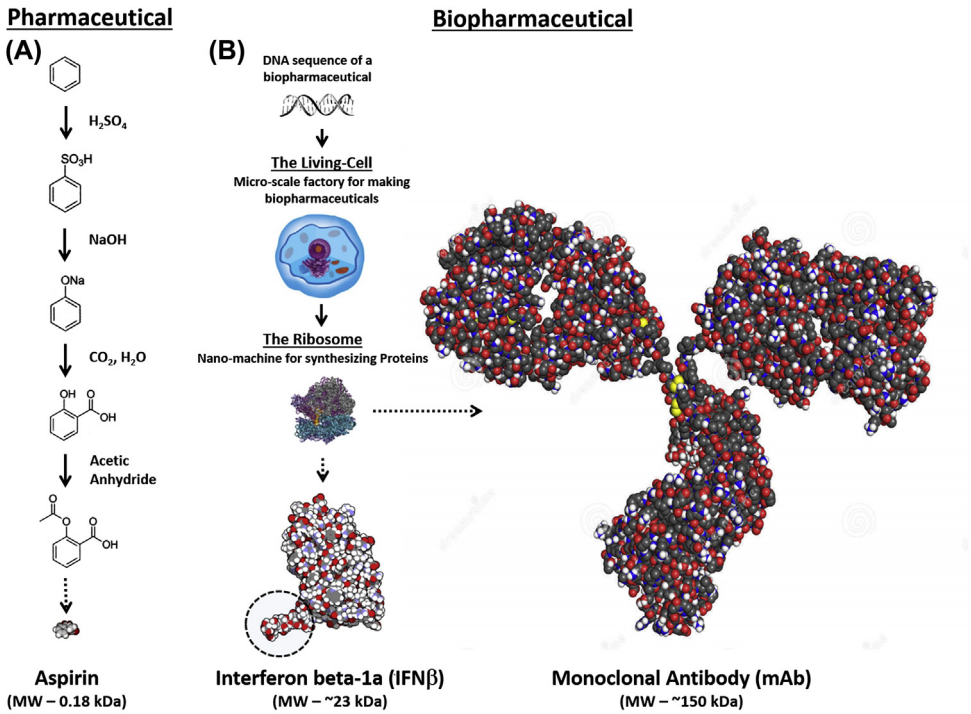

FIG. 1.5 A simple comparison illustrating differences in the process of making a pharmaceutical versus making a biopharmaceutical: (A) Coarse outline of the sequential chemical reactions for making a pharmaceutical, using aspirin as an example and (B) a coarse outline of the basic steps for making a biopharmaceutical, which consists of first synthesizing a piece of DNA containing the correct nucleotide sequence code for making the desired biopharmaceutical's polypeptide chain(s), the insertion of this DNA into an initial small collection of cells (the microscale factories for making the biopharmaceutical) using recombinant DNA technology, the large-scale growth of these cells during which the cell's internal protein synthesizing nano-machine (the ribosome, a complex cellular organelle composed of many proteins and several pieces of RNA) are directed to synthesize the target biopharmaceutical, illustrated here as either interferon beta-1a (IFNβ) or a monoclonal antibody (mAb). Note that the space-filling molecular models of aspirin, IFNβ and mAb have all been displayed roughly on the same arbitrary scale to help provide the reader with an approximate perspective on how they would relatively compare to each other on the basis of size. The dashed circle highlighting part of the structure of IFNβ corresponds to the carbohydrate-containing portion of this biopharmaceutical that plays a dominant role in giving rise to its microheterogeneity shown in chapter 2 in Figure 2.3 when coupled with other posttranslational modifications (PTMs). *Reproduced with permission from Berkowitz SA [116].*

successfully are significant. Forcing a cell to produce unusually large amounts of a single protein presents unique problems to the cell. Particularly in terms of making sure that all the protein molecules are properly folded and have consistent physical, chemical, and biological properties. Hence, to achieve this goal requires the constant and diligent monitoring and characterization of the protein biopharmaceutical's HOS.

The process of finding, developing, and obtaining regulatory approval of a protein biopharmaceutical that is made using recombinant DNA technology proceeds through a sequence of key activities or basic phases of activity that is outlined in Fig. 1.6. Success of

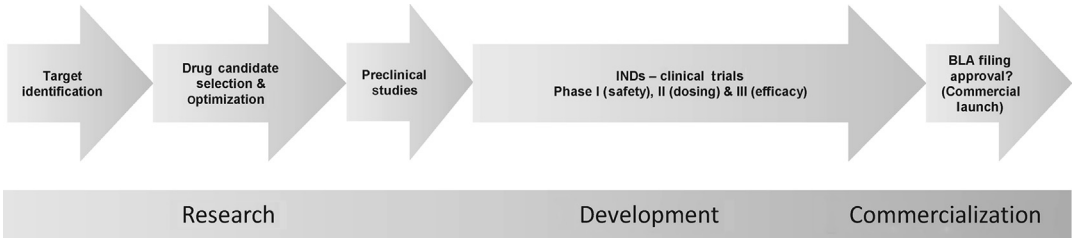

FIG. 1.6 A coarse overview of the basic areas and sequence of major activities involved in commercializing a protein biopharmaceutical. The relative length of each block arrow is roughly associated with the length of time typical spent at each stage, from research through commercialization. Overall, the cost in this process can easily be in excess of one or more billion dollars and can require more than an decade to develop [68]. These numbers can vary significantly from company to company and from drug to drug. As a result they are only approximate.

this process is highly dependent on biophysical characterization work associated with monitoring the consistency of the physicochemical properties of the protein drug, confirming the absence of changes in the drug's HOS (which might give rise to small unwanted subpopulations of altered molecules), and in assessing the potential impact that PTMs might have on a drug's structure.

During the first part of this chapter we have dealt with the very basic properties of protein structure. In the remaining sections, we will discuss how these properties are responsible for many of the potential problems that are of great concern to a large range of biopharmaceutical scientists. In addition, we will briefly look at some of the more novel types of protein biopharmaceuticals that have and are being developed that further challenge the task of biophysically characterizing these complex drugs.

1.4.1 Impact of PTMs on the HOS of protein biopharmaceuticals

The complex chemical composition of proteins, consisting of 20 chemically different naturally occurring building blocks (i.e., amino acids) effectively empower the cell with the needed components (chemistry set) to make the necessary array of proteins it needs to properly function. However, these amino acids also offer a range of chemically different targets that can undergo chemical changes, via direct chemical reactions or through the participation of various enzymatic reactions. The chemical changes that a protein biopharmaceutical can incur offer opportunities to alter the HOS of these molecules, impacting the consistency of manufacturing or worse, cause adverse events when administrated to a patient. As mentioned in section 1.1.4, these chemical changes, whether they occur in vivo or in vitro, are collectively referred to as PTMs. Many PTMs play roles in the biological function of a protein in vivo, while others are a result of normal degradation or aging. Hence, the idea that a given protein exists as a single defined unique chemical entity is misleading. In fact, nearly all protein biopharmaceuticals exist as a collection of highly similar variant forms. The range of these variants and their amounts in the final biopharmaceutical drug product is determined by the nature of the cell-line used, the cell culture conditions (e.g., raw materials and hold-times within the bioreactor), the resolution properties of the purification process, as well as our ability to detect and characterize them. The collection of highly similar proteins, *variant forms* or *proteoforms* [69] of effectively the same protein that characterize a biopharmaceutical

is referred to as *microheterogeneity* and is a unique property of these drugs. In making a protein biopharmaceutical, the attributes of microheterogeneity need to be carefully characterized, measured and controlled. In so doing, microheterogeneity becomes a fingerprint or a signature of the protein drug that is linked to its therapeutic behavior. However, it should be noted that manufacturers of biopharmaceuticals cannot "exactly" duplicate this fingerprint or signature microheterogeneity on a lot-to-lot basis. Nevertheless, within the concept of consistency of manufacturing, microheterogeneity needs to be contained to within an established and reasonable level of variation that is bounded by the biopharmaceutical's specifications. The task of establishing this level of variation permitted in a biopharmaceutical's microheterogeneity is a collaboration between the drug manufacturer and regulatory agencies (who will eventually review and approve the drug). Although, in the end it is the regulators who have the final say on what is acceptable, in terms of necessary specifications that defines this level of variation. These specifications are commonly associated with critical quality attributes (CQAs) and are attributes that are directly related to the structural characteristics that define the chemical, physical, and biological properties of the protein drug that impact the protein's therapeutic activity.

As mentioned, PTMs occur both inside (in vivo) and outside (in vitro) of the cell. Key factors that control in vivo PTMs include the following: cell line, culture media, and growth conditions. In the case of in vitro PTMs, once the protein biopharmaceutical is excreted into the culture media, the following are just a few key factors that can affect the protein drug product: temperature, pH, contact surfaces, light, metal contamination, released enzymes in the culture media or resulting from contamination, and sample handling (e.g., shaking, freezing, and thawing) [70–75]. From beginning to end, there is a host of environmental challenges that the protein biopharmaceutical must endure without altering its primary structure or HOS. In the case of the former, scientists in the biopharmaceutical industry have extensively used mass spectrometry, MS, as a key analytical tool to detect and characterize PTMs. This dependence on MS is primarily due to the change in mass that accompanies nearly all chemical reactions and the high mass accuracy and resolution of most commercially available MS instruments. However, it is worth noting that PTMs that involve isomerization reactions and or isobaric mass transitions can occur, yielding no obvious or little mass change and may require more sophisticated techniques to detect their presence [76–78]. In some cases, these isobaric mass modifications can be chromatographically separated or fragmented differentially via tandem MS [79]. In addition, although MS can detect, quantitate, and localize PTMs within a protein, the impact of a PTM on a protein's HOS is largely unknown. Therefore, the application of biophysical analysis and characterization, as discussed in this book, in combination with bioassays plays an important role in attempting to assess the impact of PTMs in the biopharmaceutical industry.

1.4.2 Impact of changes in noncovalent interactions (secondary bonds) on the HOS of protein biopharmaceuticals

Due to the high level with which a protein depends on an ensemble of weak secondary bonds or interactions to maintain its HOS and dynamic nature of its conformation, a protein can find itself potentially trapped in an altered conformation (metastable or intermediate state) without any change in its primary structure. A protein is particularly vulnerable to

these HOS changes when placed under stress conditions. Under such stress conditions the normally native protein can adopt a nonnative but energetically stable conformation (albeit not as stable as its native conformation). When this partially unfolded protein state is accompanied with a relatively large kinetic energy activation barrier, when the stress is removed the partially unfolded protein can find itself trapped; see Fig. 1.4B. In this situation, the protein would encounter significant difficulty in returning to its more stable native state. As a result, proteins can undergo an alteration in HOS "without" the need of requiring 1° structure change. The ability to detect these types of HOS changes by MS would be very difficult since the change in conformation would occur without any change in mass! Such changes in the HOS of a protein can in terms of mass spectroscopy be considered *silent* HOS changes. Thus, for the biopharmaceutical scientists to detect and quantify such changes a battery of biophysical tools are often required.

1.4.3 A more detail discussion concerning protein biopharmaceutical aggregations and its influence on HOS

Earlier in this chapter (Sections 1.1.2 and 1.1.3), it was mentioned that one of the most concerning properties linked to biopharmaceutical proteins is their ability to self-associate and form aggregates. While proteins can aggregate through many different mechanisms [80], in general, protein aggregation can be crudely classified to arise from two basic properties of a protein's stability, its colloidal, and conformational stability. Aggregation resulting from the attractive complex nature of the "normal" stable surface properties of a protein is a characteristic associated with the protein's colloidal properties and is related to its "colloidal stability". Aggregates formed via these properties are general referred to as "colloidal aggregates". However, due to the dynamic properties of proteins, these molecules can undergo a range of fluctuations in its HOS, especially under stress conditions. Thus, changes in a protein's conformational properties that expose buried (hydrophobic) chemical groups prone to self-associate with other similar or different chemical groups on another protein molecule are related to a protein's "conformational stability". Aggregates formed via these properties are generally referred to as "conformational aggregates". In some cases, aggregates that are formed may not neatly arise from one form of stability or the other. Rather they might arise through a hybrid combination of both [81,82].

In considering the unique self-associating properties of protein drugs, additional concerns surface due to the presence of "macromolecular crowding" [83]. These concerns play a in two areas which include: 1) the impact of the in vivo conditions of macromolecular crowding on the self-association process, and 2) the impact of in vitro conditions required in developing high protein biopharmaceutical concentration formulations [84,85]. In the former case, the direct administration of a protein drug into the blood stream of a patient is thought to lead to its rapid dilution, alleviating the adverse effect of a drug's own high concentration. However, the effect of macromolecular crowding resulting from the presence of other proteins in the blood and lymph system along with the changes in the environmental solution conditions, relative to the vialed-protein drug product can enhance self-association. As a result, better methods for assessing these situations are needed [86,87], see chapter 15. In the case of developing high protein drug formulations for subcutaneous (SC) injection (which enables patient

convenience, avoiding the costly and inconvenient process of administrating large amounts of protein drugs intravenously, IV), an additional question arises concerning the effects of macromolecular crowding on protein drug self-association. This is because upon SC injection a high protein concentration is deliberately created within the injected area that can remain at a high concentration for a relatively long time due to the slow passive ability of the drug protein to find its way into the patient's blood or lymph system in comparison to intravenous injection (IV). In addition, the formulated SC delivered protein drug required for these injections must also be stable and unaffected at these very high protein concentrations within its container closure, e.g., vial or prefilled syringe, for several years [88], see chapter 15.

In general, the characterization and assessment of protein self-association is by no means an easy task, especially at high protein concentrations. The bulk of the biophysical tools available to detect and characterize protein self-association have been developed for use on dilute protein solutions, often referred to as ideal solution conditions. Here, the details of macromolecular physical chemistry are simplified and much better understood. Tackling the solution behavior of proteins at concentrations as high as several 100 mg/mL forces the biopharmaceutical scientist into experimental space where their ability to interpret acquired data is extensively lacking, due to the poorly understood complexity of this situation and the absence of a completely well-developed theory. As a result, conducting useful biophysical characterization work is a real challenge resulting in biopharmaceutical scientists resorting to either very empirical methodologies [89,90], or to the extrapolation of data from very low concentrations to high concentrations [91–93] to make primitive and risky assessments, again see Chapter 15 for further discussion on this topic.

1.4.4 The novelty of different classes of protein biopharmaceuticals that create unique questions and challenges in characterizing and monitoring their HOS

Some protein biopharmaceuticals that have and are being developed contain unusual or "unnatural" constructions and/or properties. In these cases, unique questions, problems, and challenges arise concerning their physicochemical properties and HOS. Two of the main types of "unusual" or "unnatural" drug candidates involve: (1) the covalent coupling of a biopharmaceutical to another protein or chemical compound to create a fusion [78] or conjugated (e.g., pegylated proteins [94,95], antibody drug conjugate, ADC [96], and XTEN technology [97] protein biopharmaceutical); and (2) very large assembly of proteins such as virus or virus-like particles (VLPs) or nanoparticle that serve as drug delivery systems [98], see chapter 15. The following sections illustrate some of these novel biopharmaceuticals that give rise to unique questions that bring into play the importance of biophysical measurements concerning HOS.

1.4.4.1 Example 1: Fc fusion proteins

The fusion of an Fc (fragment crystallizable) part of an antibody (typically an IgG1 antibody) with that of another pharmaceutically relevant protein through recombinant technology, results in an Fc fusion protein. The reason for undertaking this fusion process is that in an antibody, the Fc portion has been shown to be responsible for increasing the antibody's circulation time [99–101]. Thus, by expressing a pharmaceutically relevant protein fused with

an Fc fragment, it is hoped a similar effect of increased circulation time will be achieved. As an example, fusing an Fc to the blood-clotting protein's Factor VIII (FVIII-Fc) and Factor IX (FIX-Fc) have been shown to reduce the clearance of these factors, while retaining the correct biological activity [102–105]. This should enable patients suffering from Hemophilia A/B to reduce the number of drug infusions. However, the fusion of two relatively large proteins (each $>$ 50 kDa) beckons the questions, "does the fusion of these two proteins cause any significant changes to either protein that would lead to an alteration in their corresponding HOS?" and, "would that fusion impact the functionality of either part of the molecule or potentially lead to an adverse effect?". While the answer to these questions will be protein-dependent, for FVIII-Fc and FIX-Fc, the answer is no, as revealed by a battery of biophysical studies [106–108].

1.4.4.2 Example 2: PEGylated proteins and antibody drug conjugates (ADCs)

The chemical coupling of a polyethylene glycol polymer (PEG) to a protein biopharmaceutical yields a pegylated protein drug [94,95,109]. As observed with Fc fusion protein, pegylated proteins show a significant reduction in the clearance of the administrated modified protein drug from the body [109], significantly increasing their therapeutic value. However, just as in the case of Fc fusion proteins the conjugation of the PEG molecule to a protein beckons the same question concerning the impact of this modification on the HOS of the modified protein (as discussed in the previous section) and for the same reasons. Nevertheless, to date there are currently more than a dozen pegylated biopharmaceuticals approved and marketed [110] and more in development [111].

Similar situation exists for ADCs where a small generally toxic drug (called a payload) is covalently bound to a mAb that uniquely bind to a specific cell (typically a cancerous cell [96]). In this situation, the normally nonspecific toxic small molecule drug is turned into a highly specific targeting drug by taking advantage of the high specificity of the mAb. The mAb essentially guides (carries) the toxic pharmaceutical payload to its cellular targets (e.g., cancer cell) where it's internalized by the target cell and activated by its cleavage from the ADC causing highly specific cell (e.g. cancer cell) killing. In constructing these ADCs careful biophysical characterization studies are required to assess the impact of the small pharmaceutical and its load (drug to antibody ratio, DAR) on the mAb [112] to insure its overall HOS [113] (binding specificity) is not compromised.

1.4.4.3 Example 3: viruses, VLPs

The formation of very large (MDa) multisubunit protein complexes, such as a virus (e.g., used in gene therapy), VLP, nanoparticles (e.g., lipid nanoparticles, LNPs, liposomes, and exosomes) which are used as drug delivery systems present unique challenges when trying to biophysically characterize these complex biopharmaceuticals, which may be only partially composed of protein material (or even in some cases may not contain any proteins, e.g., non-protein biopharmaceuticals composed of specific nucleic acid molecules incased in LNPs or Liposomes). A critical challenge in developing these very large biopharmaceuticals is in assessing their homogeneity and overall structure. Due to their large size, the common workhorse tool for acquiring this information, size-exclusion chromatograph (SEC, see Chapter 7), may have limited utility because of limitations in its separation range (requiring SEC columns with pore sizes that are often larger than what is commercially commonly found to be

available, i.e., >1000 Å). In this case, alternative analytical tools such as asymmetric flow field flow fractionation (AF4, Chapters 10 and 15), analytical ultracentrifugation (AUC, Chapter 9), and nanoparticle tracking analysis (NTA, Chapter 10) and flow cytometry (Chapter 10) can be important biophysical tools capable of filling in this gap. Some of these tools, such as AUC can also provide additional characterization information (chemical composition), which is unique to these classes of very large complex biopharmaceuticals, e.g., the level of virus drug particles that are filled, partially filled empty in terms of nucleic acid material [114].

1.5 Conclusion

In this chapter, the authors have provided, in broad strokes, brief discussions on the fundamental structural properties of proteins. Since a protein's structure dictates its function, these properties empower proteins with important functional roles for maintaining the cascade of activities that characterize all living systems. However, these same structural properties also create a heavy but required characterization workload for the biophysical scientist working in the biopharmaceuticals industry. It is our hope that the reader is in a better position to understand the unique challenges the biopharmaceutical industry encounters in striving to bring these protein drugs to the market place. These challenges are significantly more daunting compared with those typically encountered in the pharmaceutical area, where the drug product corresponds to small organic molecules with much simpler, rigid and homogenous structure.

The developments that have occurred since the discovery of the structure of DNA [115], a little over a half century ago that ushered in the molecular biology era, has culminated in the last four decades with the successful commercial development of today's growing biopharmaceutical industry. Concurrent with this development has been the advancement of the bioanalytical sciences, which has led to the development of better instruments and methods. Not only are we able to better understand how these complex molecules work, but we are also capable of characterizing them to better assure their safety and consistency of manufacturing.

In the area of biophysical characterization, significant innovative developments in newer biophysical tools and techniques continue to occur along with the improvements in the older and traditional biophysical tools and techniques. This situation makes our ability to characterize the HOS of biopharmaceuticals on a routine level truly impressive. However, today knowing what we do know versus knowing what we don't know can be very sobering! It seems, to the authors, that the more we discover the less we realize we know. Although still lacking in our ability to characterize (biophysically) these fascinating protein drugs, the good news is we are moving forward, learning to do a better job in participating in the overall task of making more effective and safer drugs!

References

[1] Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol 1963;7:95–9.

[2] Creighton TE. The folded conformations of globular proteins. In: Proteins: structures and molecular properties. New York (NY): W.H. Freeman and Co.; 1992.

[3] Creighton TE. The biophysical chemistry of nucleic acids & proteins. Helvetian Press; 2010.
[4] Kessel A, Ben-Tal N. Primary structure [Chapter 2]. In: Introduction to proteins: structure function and motion. Boca Raton (FL): CRC Press; 2010.
[5] Kessel A, Ben-Tal N. Secondary structure. In: Introduction to proteins: structure function and motion. Boca Raton (FL): CRC Press; 2010.
[6] Kyte J. Structure in protein chemistry. New York (NY): Garland Publishing, Inc.; 2006.
[7] Kessel A, Ben-Tal N. Tertiary structure. In: Introduction to proteins: structure function and motion. Boca Raton (FL): CRC Press; 2010.
[8] Creighton TE. The physical and chemical basis of molecular biology. Helvetian Press; 2010.
[9] Creighton TE. The folded conformations of globular proteins. In: Proteins: structures and molecular properties. New York (NY): W.H. Freeman and Co.; 2010.
[10] Houde DJ, Bou-Assaf GM, Berkowitz SA. Deciphering the biophysical effects of oxidizing sulfur-containing amino acids in interferon-beta-1a using MS and HDX-MS. J Am Soc Mass Spectrom 2017;28(5):840−9.
[11] Bobst CE, Abzalimov RR, Houde D, Kloczewiak M, Mhatre R, Berkowitz SA, Kaltashov IA. Detection and characterization of altered conformations of protein pharmaceuticals using complementary mass spectrometry-based approaches. Anal Chem 2008;80(19):7473−81.
[12] Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature 2007;450:964−72.
[13] Smock RG, Gierasch LM. Sending signals dynamically. Science 2009;324:198−203.
[14] Teilum K, Olsen JG, Kragelund BB. Functional aspects of protein flexibility. Cell Mol Life Sci 2009;66:2231−47.
[15] Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science 2009;324:203−7.
[16] Kamerzell TJ, Middaugh CR. The complex inter-relationships between protein flexibility and stability. J Pharm Sci 2008;97:3494−517.
[17] Kamerzell TJ, Ramsey JD, Middaugh CR. Immunoglobulin dynamics, conformational fluctuations, and nonlinear elasticity and their effects on stability. J Phys Chem 2008;112:3240−50.
[18] Liu H, May K. Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. mAbs 2012;4:17−23.
[19] Chen PL, Wang M, Ou WC, Lii CK, Chen LS, Chang D. Disulfide bonds stabilize JC virus capsid-like structure by protecting calcium ions from chelation. FEBS Lett 2001;500:109−13.
[20] Toichi K, Yamanaka K, Furukawa Y. Disulfide scrambling describes the oligomer formation of superoxide dismutase (SOD1) proteins in the familial form of amyotrophic lateral sclerosis. J Biol Chem 2013;288:4970−80.
[21] Trivedi MV, Laurence JS, Siahaan TJ. The role of thiols and disulfides on protein stability. Curr Protein Pept Sci 2009;10:614−25.
[22] Ho SC, Wang T, Song Z, Yang Y. IgG aggregation mechanism for CHO cell lines expressing excess heavy chains. Mol Biotechnol 2015;57(7):625−34.
[23] Sakurai K, Nakahata R, Lee YH, Kardos J, Ikegami T, Goto Y. Effects of a reduced disulfide bond on aggregation properties of the human IgG1 CH3 domain. Biochim Biophys Acta 2015;854(10 Pt A).
[24] Nag M, Bera K, Basak S. Intermolecular disulfide bond formation promotes immunoglobulin aggregation: investigation by fluorescence correlation spectroscopy. Proteins 2015;83(1):169−77.
[25] Moritz B, Stracke JO. Assessment of disulfide and hinge modifications in monoclonal antibodies. Electrophoresis 2017;38(6):769−85.
[26] Morgan GJ, Usher GA, Kelly JW. Incomplete refolding of antibody light chains to non-native, protease-sensitive conformations leads to aggregation: a mechanism of amyloidogenesis in patients. Biochemistry 2017;56(50):6597−614.
[27] Serebryany E, Woodard JC, Adkar BV, Shabab M, King JA, Shakhnovich EI. An internal disulfide locks a misfolded aggregation-prone intermediate in cataract-linked mutants of human γD-crystallin. J Biol Chem 2016;291(36):19172−83.
[28] Kessel A, Ben-Tal N. Protein structural dynamics. Introduction to proteins: structure function and motion. Boca Raton (FL): CRC Press; 2010.
[29] Bee JS, Randolph TW, Carpenter JF, Bishop SM, Dimitrova MN. Effects of surfaces and leachables on the stability of biopharmaceuticals. J Pharm Sci 2011;100(10):4158−70.
[30] Majumdar S, Ford BM, Mar KD, Sullivan VJ, Ulrich RG, D'Souza AJ. Evaluation of the effect of syringe surfaces on protein formulations. J Pharm Sci 2011;100:2563−73.

[31] Sharma B. Immunogenicity of therapeutic proteins. Part 2: impact of container closures. Biotechnol Adv 2007;25:318–24.
[32] Sharma B. Immunogenicity of therapeutic proteins. Part 3: impact of manufacturing changes. Biotechnol Adv 2007;25:325–31.
[33] Sharma B. Immunogenicity of therapeutic proteins. Part 1: impact of product handling. Biotechnol Adv 2007;25:310–7.
[34] Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, et al. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 2002;416:507–11.
[35] Filipe V, Hawe A, Schellekens H, Jiskoot W. Aggregation and immunogenicity of therapeutic proteins. In: (editors: Wang W, Roberts CJ.) Aggregation of therapeutic proteins. John Wiley and Sons: Hoboken (NY), pp. 400–33.
[36] Rosenberg AS. Effects of protein aggregates: an immunologic perspective. AAPS J 2006;8:E501–7.
[37] Walsh CT. Posttranslational modification of proteins: expanding nature's inventory. Englewood , CO: Roberts and Co.; 2006. 490pp.
[38] Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al. Essentials of glycobiology. Cold Springs Harbor (NY): Cold Springs Harbor Laboratory Press; 2009.
[39] Kessel A, Ben-Tal N. Protein structure. Introduction to proteins: structure function and motion, CRC Press, Boca Raton (FL).
[40] Creighton TE. Biosynthesis of proteins. Proteins: structures and molecular properties, W.H. Freeman and Co., New York (NY).
[41] Walsh CT, Garneau-Tsodikova S, Gatto Jr GJ. Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed 2005;44:7342–72.
[42] Farley AR, Link AJ. Identification and quantification of protein posttranslational modifications. Methods Enzymol 2009;463:725–63.
[43] Walsh G, Jefferis R. Post-translational modifications in the context of therapeutic proteins. Nat Biotechnol 2006;24:1241–52.
[44] Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223–30.
[45] Dill KA, MacCallum JL. The protein-folding problem, 50 years on. Science 2012;338:1042–6.
[46] Levinthal C. How to fold graciously. Mossbauer Spectrosc Biol Syst 1969;67:22–6 [Proceedings].
[47] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167–95.
[48] Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol 1997;4:10–9.
[49] Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. Annu Rev Biophys 2008;37:289–316.
[50] Onuchic JN, Wolynes PG. Theory of protein folding. Curr Opin Struct Biol 2004;14:70–5.
[51] Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. Toward an outline of the topography of a realistic protein-folding funnel. Proc Natl Acad Sci USA 1995;92:3626–30.
[52] Panchenko AR, Luthey-Schulten Z, Wolynes PG. Foldons, protein structural modules, and exons. Proc Natl Acad Sci USA 1996;93:2008–13.
[53] Englander SW, Mayne L, Krishna MM. Protein folding and misfolding: mechanism and principles. Q Rev Biophys 2007;40:287–326.
[54] Fedorov AN, Baldwin TO. Protein folding and assembly in a cell-free expression system. Methods Enzymol 1998;290:1–17.
[55] Netzer WJ, Hartl FU. Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature 1997;388:343–9.
[56] Cabrita LD, Dobson CM, Christodoulou J. Protein folding on the ribosome. Curr Opin Struct Biol 2010;20:33–45.
[57] Fedorov AN, Baldwin TO. Cotranslational protein folding. J Biol Chem 1997;272:32715–8.
[58] Beissinger M, Buchner J. How chaperones fold proteins. Biol Chem 1998;379:245–59.
[59] Gottesman S, Wickner S, Maurizi MR. Protein quality control: triage by chaperones and proteases. Genes Dev 1997;11:815–23.
[60] Lee S, Tsai FT. Molecular chaperones in protein quality control. J Biochem Mol Biol 2005;38:259–65.
[61] Saibil H. Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev 2013;13:630–42.

[62] Quan S, Bardwell JC. Chaperone discovery. Bioessays 2012;34:973–81.
[63] Sorokin AV, Kim ER, Ovchinnikov LP. Proteasome system of protein degradation and processing. Biochemistry (Mosc) 2009;74:1411–42.
[64] Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. J Mol Graph Model 2001;19:26–59.
[65] Dunker AK, Babu MM, Barbar E, Blackledge M, Bondos SE, Dosztanyi Z, et al. What's in a name? Why these proteins are intrinsically disordered. Intrinsically Disord Proteins 2013;1:e24157.
[66] Huang Y-M, Hu W, Rustandi E, Chang K, Yusuf-Makagiansar H, Ryll T. Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. Biotechnol Prog 2010;26:1400–10.
[67] Britt KA, Schwartz DK, Wurth C, Mahler HC, Carpenter JF, Randolph TW. Excipient effects on humanized monoclonal antibody interactions with silicone oil emulsions. J Pharm Sci 2012;101:4419–32.
[68] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry new estimates of R&D. J Health Econ 2016;47:20–32.
[69] Smith LM, Kelleher NL. Proteoform: a single term describing protein complexity. Nat Methods 2013;10(3):186–7.
[70] Ha E, Wang W, Wang YJ. Peroxide formation in polysorbate 80 and protein stability. J Pharm Sci 2002;91:2252–64.
[71] Jiang Y, Nashed-Samuel Y, Li C, Liu W, Pollastrini J, Mallard D, et al. Tungsten-induced protein aggregation: solution behavior. J Pharm Sci 2009;98:4695–710.
[72] Kerwin BA, Remmele Jr RL. Protect from light: photodegradation and protein biologics. J Pharm Sci 2007;96:1468–79.
[73] Patel J, Kothari R, Tunga R, Ritter NM, Tunga BS. Stability considerations for biopharmaceuticals, Part I. Bioprocess Int 2011:20–31.
[74] Wang W. Instability, stabilization, and formulation of liquid protein pharmaceuticals. Int J Pharm 1999;185:129–88.
[75] Bossio RE, Marshall AG. Baseline resolution of isobaric phosphorylated and sulfated peptides and nucleotides by electrospray ionization FTICR ms: another step toward mass spectrometry-based proteomics. Anal Chem 2002;74:1674–9.
[76] Geiger T, Clarke S. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. J Biol Chem 1987;262:785–94.
[77] Hambly DM, Banks DD, Scavezze JL, Siska CC, Gadgil HS. Detection and quantitation of IgG 1 hinge aspartate isomerization: a rapid degradation in stressed stability studies. Anal Chem 2009;81:7454–9.
[78] Czajkowsky DM, Hu J, Shao Z, Pleass RJ. Fc-fusion proteins: new developments and future perspectives. EMBO Mol Med 2012;4(10):1015–28.
[79] Beck A, Reichert JM. Therapeutic Fc-fusion proteins and peptides as successful alternatives to antibodies. mAbs 2011;3:415–6.
[80] Philo JS, Arakawa T. Mechanisms of protein aggregation. Curr Pharmaceut Biotechnol 2009;10:348–51.
[81] Chi EY, Krishnan S, Kendrick BS, Chang BS, Carpenter JF, Randolph TW. Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. Protein Sci 2003;12:903–13.
[82] Mahler HC, Fischer S, Randolph TW, Carpenter JF. Protein aggregation and particle formation effects of formulation, interfaces, and drug product manufacturing operations. In: Wang W, Roberts CJ, editors. Aggregation of therapeutic proteins. John Wiley & Sons, Inc.; 2010. p. 301–5.
[83] Ellis RJ, Minton AP. Protein aggregation in crowed environments. Biol Chem 2006;387:485–97.
[84] Minton AP. Influence of macromolecular crowding upon the stability and state of association of proteins: predictions and observations. J Pharm Sci 2005;94:1668–75.
[85] Shire SJ, Shahrokh Z, Liu J. Challenges in the development of high protein concentration formulations. J Pharm Sci 2004;93:1390–402.
[86] Demeule B, Shire SJ, Liu J. A therapeutic antibody and its antigen form different complexes in serum than in phosphate-buffered saline: a study by analytical ultracentrifugation. Anal Biochem 2009;388:279–87.
[87] Wright RT, Hayes DB, Stafford WF, Sherwood PJ, Correia JJ. Characterization of therapeutic antibodies in the presence of human serum proteins by AU-FDS analytical ultracentrifugation. Anal Biochem 2018;550:72–83.

[88] Mathaes R, Koulov A, Joerg S, Mahler HC. Subcutaneous injection volume of biopharmaceuticals—pushing the boundaries. J Pharm Sci 2016;105(8):2255–9.
[89] Chaudhri A, Zarraga IE, Yadav S, Patapoff TW, Shire SJ, Voth GA. The role of amino acid sequence in the self-association of therapeutic monoclonal antibodies: insights from coarse-grained modeling. J Phys Chem 2013;117:1269–79.
[90] Kayser V, Chennamsetty N, Voynov V, Helk B, Forrer K, Trout BL. Evaluation of a non-Arrhenius model for therapeutic monoclonal antibody aggregation. J Pharm Sci 2011;100:2526–42.
[91] Li S, Xing D, Li J. Dynamic light scattering application to study protein interactions in electrolyte solutions. J Biol Phys 2004;30:313–24.
[92] Scherer TM, Liu J, Shire SJ, Minton AP. Intermolecular interactions of IgG1 monoclonal antibodies at high concentrations characterized by light scattering. J Phys Chem 2010;114:12948–57.
[93] Yadav S, Scherer TM, Shire SJ, Kalonia DS. Use of dynamic light scattering to determine second virial coefficient in a semidilute concentration regime. Anal Biochem 2011;411:292–6.
[94] Bailon P, Won CY. PEG-modified biopharmaceuticals. Expert Opin Drug Deliv 2009;6:1–16.
[95] Jevsevar S, Kunstelj M, Porekar VG. PEGylation of therapeutic proteins. Biotechnol J 2010;5:113–28.
[96] Lambert JM, Berkenblit A. Antibody-drug conjugates for cancer treatment. Annu Rev Med 2018;69:191–207.
[97] Podust VN, Balan S, Sim BC, Coyle MP, Ernst U, Peters RT, Schellenberger V. Extension of in vivo half-life of biologically active molecules by XTEN protein polymers. J Control Release 2016;240:52–66.
[98] Ma Y, Nolte RJM, Cornelissen JJLM. Virus-based nanocarriers for drug delivery. Adv Drug Deliv Rev 2012;64:811–25.
[99] Unverdorben F, Richter F, Hutt M, Seifert O, Malinge P, Fischer N, Kontermann RE. Pharmacokinetic properties of IgG and various Fc fusion proteins in mice. mAbs 2016;8(1):120–8.
[100] Roopenian DC, Akilesh S. FcRn: the neonatal Fc receptor comes of age. Nat Rev Immunol 2007;7:715–25.
[101] Roopenian DC, Christianson GJ, Sproule TJ, Brown AC, Akilesh S, Jung N, et al. The MHC class I-like IgG receptor controls perinatal IgG transport, IgG homeostasis, and fate of IgG-Fc-coupled drugs. J Immunol 2003;170:3528–33.
[102] Peters RT, Low SC, Kamphaus GD, Dumont JA, Amari JV, Lu Q, et al. Prolonged activity of factor IX as a monomeric Fc fusion protein. Blood 2010;115:2057–64.
[103] Shapiro AD, Ragni MV, Valentino LA, Key NS, Josephson NC, Powell JS, et al. Recombinant factor IX-Fc fusion protein (rFIXFc) demonstrates safety and prolonged activity in a phase 1/2a study in hemophilia B patients. Blood 2012;119:666–72.
[104] Dumont JA, Liu T, Low SC, Zhang X, Kamphaus G, Sakorafas P, et al. Prolonged activity of a recombinant factor VIII-Fc fusion protein in hemophilia A mice and dogs. Blood 2012;119:3024–30.
[105] Powell JS, Josephson NC, Quon D, Ragni MV, Cheng G, Li E, et al. Safety and prolonged activity of recombinant factor VIII Fc fusion protein in hemophilia A patients. Blood 2012;119:3031–7.
[106] Houde D, Berkowitz SA. Conformational comparability of factor IX-Fc fusion protein, factor IX, and purified Fc fragment in the absence and presence of calcium. J Pharm Sci 2012;101:1688–700.
[107] Li H, Bai S, Wei JY, Berkowitz SA, Brader ML. Calcium binding to a factor ix Fc fusion protein and effects on higher-order structure. J Pharm Sci 2011;100:4597–606.
[108] Kulman J, et al. Structural comparability between recombinant FVIII-Fc and its isolated FVIII and Fc constituents, Abstract 1135 presented at the 54th Annual meeting and exposition in Altlanta Ga 8–11 December 2012.
[109] Harris JM, Martin NE, Modi M. Pegylation: a novel process for modifying pharmacokinetics. Clin Pharmacokinet 2001;40:539–51.
[110] Turecek PL, Bossard MJ, Schoetens F, Ivens IA. PEGylation of biopharmaceuticals: a review of chemistry and nonclinical safety information of approved drugs. J Pharm Sci 2016;105(2):460–75.
[111] Ivens IA, Achanzar W, Baumann A, Brändli-Baiocco A, Cavagnaro J, et al. PEGylated biopharmaceuticals: current experience and considerations for nonclinical development. Toxicol Pathol 2015;43(7):959–83.
[112] Adem YT, Schwarz KA, Duenas E, Patapoff TW, Galush WJ, Esue O. Auristatin antibody drug conjugate physical instability and the role of drug payload. Bioconjug Chem 2014;25(4):656–64.
[113] Pan LY, Salas-Solano O, Valliere-Douglass JF. Conformation and dynamics of interchain cysteine-linked antibody-drug conjugates as revealed by hydrogen/deuterium exchange mass spectrometry. Anal Chem 2014;86(5):2657–64.

[114] Berkowitz SA, Philo JS. Monitoring the homogeneity of adenovirus preparations (a gene therapy delivery system) using analytical ultracentrifugation. Anal Biochem 2007;362:16–37.
[115] Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953;171:737–8.
[116] Berkowitz SA. Analytical characterization: structural assessment of biosimilarity. In: Endrenyi L, Declerck P, Chow S-C, editors. Biosimilar drug product development, vol. 216. Bca Raton, FL: CRC Press; 2015. p. 15–82.

Further reading

For a comprehensive review on many of the topics covered in this chapter, the authors highly recommend the following useful books:

[1] Creighton TE. Proteins: Structures and Molecular Properties. 2nd ed. W.H. Freeman and Co., NY.
[2] Creighton TE. The Biophysical Chemistry of Nucleic Acids & Proteins, Helvetian Press.
[3] Cantor CR, Schimmel PR. Biophysical Chemistry Part I: The Conformation of Biological Macromolecules, W.H. Freeman & Company, New York.
[4] Dickerson RE, Geis I. The Structure and Action of Proteins, Harper & Row, New York.
[5] Kaltashov IA, Eyles SJ. Mass Spectrometry in Biophysics, Wiley-Interscience, Hoboken (NJ).
[6] Kessel A, Ben-Tal N. Introduction to Proteins: Structure, Function and Motion, CRC Press, Boca Raton (FL).
[7] Kyte J. Structure in Protein Chemistry, Garland Publishing, Inc., NY.