1

Characterization of Protein Molecules Prepared by Total Chemical Synthesis

Stephen B. H. Kent

Department of Chemistry, The University of Chicago, Chicago, IL 60637, USA

> *"Nevertheless, the chemical enigma of Life will not be solved until organic chemistry has mastered another, even more difficult subject, the proteins, in the same way as it has mastered the carbohydrates."*
>
> Source: Emil Fischer (Nobel Lecture 1902)

1.1 Introduction

Proteins are the "natural prodicts" of the twenty-first century. Protein molecules are ubiquitous in the biological world, with numerous diverse functions that range from acting as structural materials such as the keratins, to integral membrane proteins that serve as ion channels or as active molecular transporters in cells, to proteins that act as hormonal messengers in higher animal species, and to proteins that regulate gene expression [1]. The most important function of protein molecules is as enzymes, the potent and specific catalysts of the chemical reactions of biological metabolism, without which life would be impossible [2]. Thanks to modern DNA sequencing methods applied to genome [3] and metagenome [4] sequencing, vast numbers of proteins are being discovered at the nucleic acid level as open reading frames that code for a protein's polypeptide chain.

The central dogma of protein science is that the amino acid sequence of the polypeptide chain encodes the folded structure of the protein molecule in its natural environment, and that it is the folded structure of the protein molecule that gives rise to its biological function(s) [5]. Proteins range in mass from less than 5 kDa to more than 100 kDa. The median size of globular protein molecules is ~35–45 kDa, comprising a polypeptide chain of ~300–400 amino acid residues. Proteins are typically made up of two or more domains, autonomous units of folding, each of ~120–160 amino acid residues [6].

Protein molecules are not simply really big peptides. In its native environment, each globular protein molecule has a defined folded structure that gives rise to the

Total Chemical Synthesis of Proteins, First Edition. Edited by Ashraf Brik, Philip Dawson, and Lei Liu.

functional properties of that protein, including biochemical and biological activities. Synthetic proteins are organic molecules of high molecular mass, comprised of linear polypeptide chains (typically of 50–300 or more amino acid residues) that fold to form complex, dynamic structures. The large size of proteins together with the intricacy of their covalent and folded structures creates special challenges in the characterization of these synthetic molecules. For this reason, it is important for researchers to rigorously characterize protein molecules prepared by total chemical synthesis to meet standards similar to those used in synthetic organic chemistry. Analytical methods and criteria for the rigorous characterization of synthetic proteins have recently been enunciated [7].

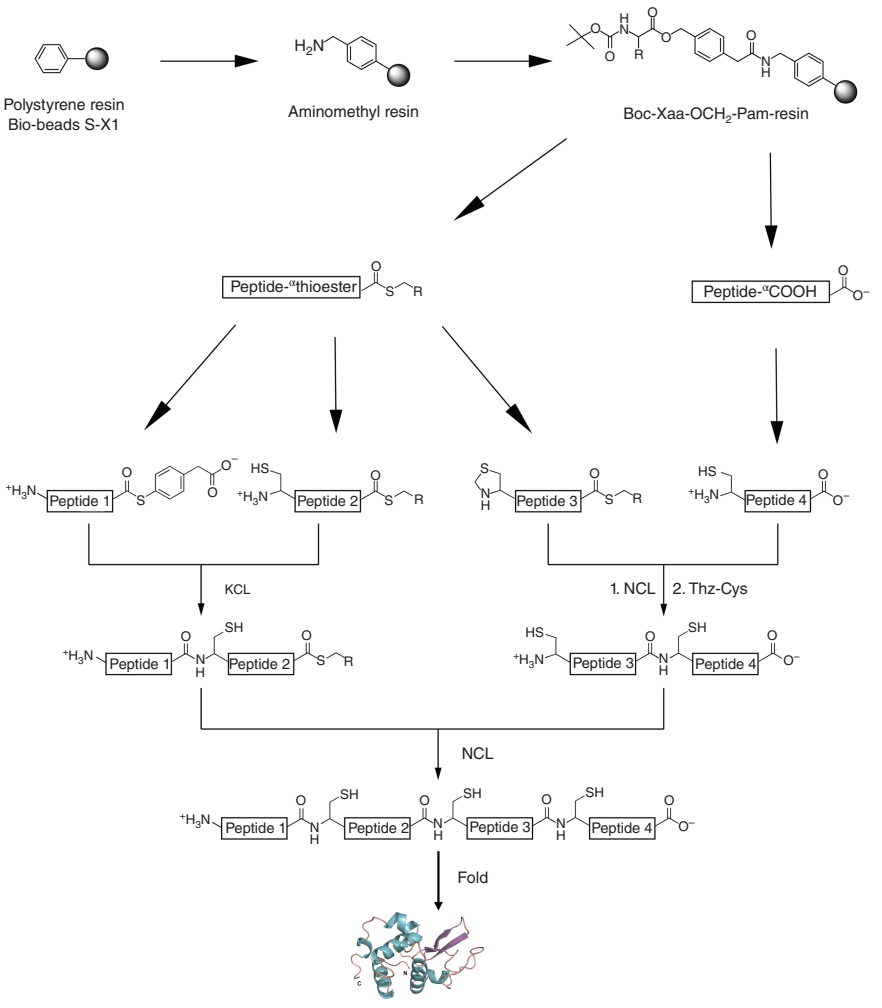
1.2 Chemical Protein Synthesis

> *"The chemical ligation approach .... breaks the conceptual shackles imposed by the peptide bond, frees us from the linear paradigm of the genetic code, and opens the world of proteins to the entire repertoire of chemistry."*
>
> Stephen Kent (23rd European Peptide Symposium 1994)

Chemistry, enabled by total synthesis, has an essential role to play in developing a fundamental understanding of the principles that give rise to the structures and biological functions of protein molecules. Thanks to modern chemical ligation methods [8–10], the total synthesis of protein molecules in the research laboratory is both practical and increasingly robust.

Characterization of a protein prepared by total chemical synthesis should begin with the synthetic process itself. Analytical control of the bond-forming steps, purification and analysis of each synthetic intermediate, and folding of the full-length synthetic polypeptide chain to form a defined tertiary structure are integral to verifying the structure of the final synthetic protein product. The steps involved in a typical total chemical synthesis of a protein molecule are listed below. Key aspects pertinent to **characterization** of the synthetic protein are in **bold**:

- **Establish a verified amino acid sequence of the polypeptide chain of the target protein molecule.**
  - Use a database such as UniProt (www.uniprot.org); include posttranslational processing where that is known to give the mature protein molecule.
  - Resolving data base/literature ambiguities in the reported amino acid sequence of a protein can be problematic. It is easy to make a mistake and to end up making an incorrect target sequence.
- Design a (convergent) synthesis of the protein's polypeptide chain, starting from peptide segments containing fewer than 40–50 amino acid residues (Scheme 1.1).
- Stepwise solid-phase synthesis (SPPS) of peptide segments equipped with suitable functionalities for chemical ligation.
- Boc chemistry [11] or Fmoc chemistry SPPS [12], with **documentation of the amino acid sequence actually made.**



Scheme 1.1 Convergent chemical synthesis of a protein molecule from four synthetic peptide segments prepared by stepwise SPPS. Key: NCL, native chemical ligation; KCL, kinetically controlled ligation; Thz-Cys, conversion of N-terminal thiazolidine-CO– to Cys–. Source: Adapted from Durek et al. [11].

- Purification by preparative high-performance liquid chromatography (HPLC), which should be performed under *displacement mode* [13].
- Combined fractions should be checked by a high resolution technique based on a ***separation principle distinct from that used in the purification***, such as capillary isoelectric focusing (CIEF) [14]. Because HPLC is used for the purification of the synthetic peptide segment, it is NOT sufficient to use analytical HPLC or LCMS run under similar chromatographic conditions to verify its homogeneity.
- *Direct infusio*n electrospray ionization mass spectrometry (ESI-MS) for checking purified peptide segments for impurities that have different masses.

- **Confirmation of the covalent structure of each synthetic peptide segment.**

- **Precise mass measurement**
  - from analytical LCMS, the mass spectrometric data *must* be collected across the *entire UV absorbing peak corresponding to the purified peptide*.
  - a correct mass is a necessary but not sufficient analytical criterion.
- **Amino acid sequence**
  - by MALDI-TOF MS 'ladder sequencing' [15] of terminated byproducts that are invariably present at low-levels in crude synthetic peptides made by SPPS.
  - by MS-MS of the purified peptide. This process is rendered more straightforward because the target sequence of the peptide segment is known.
- Convergent covalent condensation of the peptide segments
  - Chemical conversion of cryptic functional groups to the reactive form suitable for chemical ligation (e.g. Thz–peptide to Cys–peptide; [16] peptide–$CONHNH_2$ to peptide–thioester [17]).
  - Native chemical ligation (NCL); [18] pH 7.0; aqueous 6 M Gu.HCL as a near-universal solvent; ambient temperature; arylthiol catalyst [19].
- **Analytical control of synthetic steps**
  - "Hands-on" real-time analytical LCMS during ligation reactions is an essential tool for following the course of each bond-forming step.
- **Purification and characterization of each intermediate product**
  - Using the same protocols as those used for the synthetic peptide segments.
- **Purification to homogeneity and rigorous characterization of the full-length synthetic polypeptide chain.**
  - Purification by preparative HPLC, which should be performed under *displacement mode* [13].
  - Combined fractions should be checked by a high-resolution technique based on a ***separation principle distinct from that used in the purification***, such as CIEF [14]. Because HPLC has been used for the purification of the synthetic polypeptide, it is NOT sufficient to use analytical HPLC or LCMS run under similar conditions to verify its homogeneity.
  - *Direct infusio*n ESI-MS is also a good technique for checking the full-length synthetic polypeptide chain for impurities that have different masses.
- Folding the synthetic polypeptide chain to form the functional protein molecule
  - For disulfide-containing proteins, standard thiol redox couple conditions are used along with moderate amounts of solubility-enhancing agents ("denaturants") to keep misfolded polypeptide chains in solution so that they too are eventually able to fold correctly [20].
  - For proteins that do not contain disulfide bonds, it is frequently sufficient to simply dilute the polypeptide chain into native buffer or to slowly dialyze from 6 M Gu.HCl into native buffer conditions.
- Analytical control of protein folding
  - In the case of disulfide-containing proteins, the folded protein molecule will usually elute *earlier* than the unfolded polypeptide chain on analytical reverse phase HPLC, because hydrophobic side chains are less exposed in the folded, disulfide cross-linked protein molecule [21].

- For proteins that do not contain disulfide bonds, other techniques such as CD-ORD[1] or multidimensional nuclear magnetic resonance (NMR) must be used to monitor the progress of folding.

- **Purification of the folded synthetic protein to give a single defined molecular species.**
  - By one or more of the techniques of reverse-phase HPLC, ion exchange chromatography, size exclusion chromatography.
- **Homogeneity of the purified synthetic protein must be verified by a high-resolution technique based on a separation principle *distinct* from that used for purification.**
  - CIEF [14] is the preferred technique for rigorously establishing the homogeneity of the purified synthetic protein molecule.
- **Precise measurement of the mass of the synthetic protein molecule by *direct infusion* ESI-MS (direct infusion ESI-MS).**
  - *Direct infusion* ESI-MS provides data representing *all* the molecular species present in the final synthetic product. It thus provides both a precise experimental measurement of the mass of the synthetic protein and at the same time can reveal the presence of any other products that have a mass distinct from that of the target protein molecule. All measured mass data must include an experimental uncertainty (Figure 1.1).
  - HPLC with on-line ESMS can be used. However, impurity protein species present in the synthetic product may not be eluted from the reverse phase support. Mass spectrometric data from LCMS *must* **be acquired over the entire UV-absorbing peak** corresponding to the purified synthetic protein product.
    - ***It is NOT acceptable to report a mass determined at a single time point of the HPLC chromatogram, while ignoring molecular species with different masses that are present at other time points under the same UV absorbing peak.***
  - Wherever possible, the **monoisotopic mass** of a synthetic protein should be reported (together with experimental uncertainty), along with the calculated monoisiopic mass. An example is shown below (Figure 1.2).
- **Verification of the cysteine pairing for disulfide containing proteins.**
  - Proteolytic digestion followed by LCMS peptide mapping [24].
  - Cysteine pairing can also be verified by high resolution X-ray crystallography (see below).
- **Conformational homogeneity of the synthetic protein molecule.**
  - Multidimensional NMR "fingerprinting" can be used to establish that the synthetic protein has a single folded structure [11, 25]. Notes: (i) there are examples of natural protein molecules that form more than one defined folded

1 While a low-resolution technique such as CD-ORD is useful for monitoring the progress of folding, it must be supplemented with high-resolution characterization of the folded structure of the synthetic protein by techniques such as NMR and X-ray crystallography.

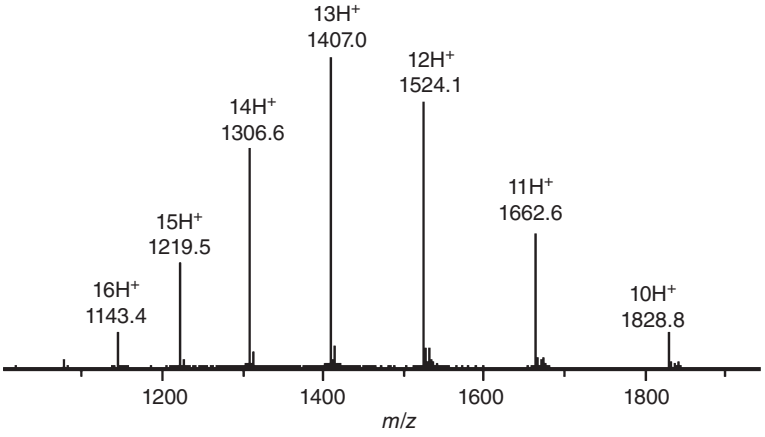


Figure 1.1 Direct infusion ESI-MS data for [Lys[24,38,83]]erythropoietin aglycone prepared by total chemical synthesis. Impurities of mass less than the target protein molecule are essentially absent, as shown by the lack of peaks on the low *m/z* side of each charge state. The peaks on the high *m/z* side of each charge state are Na and Ca ion adducts. Source: Adapted from Liu et al. [22].

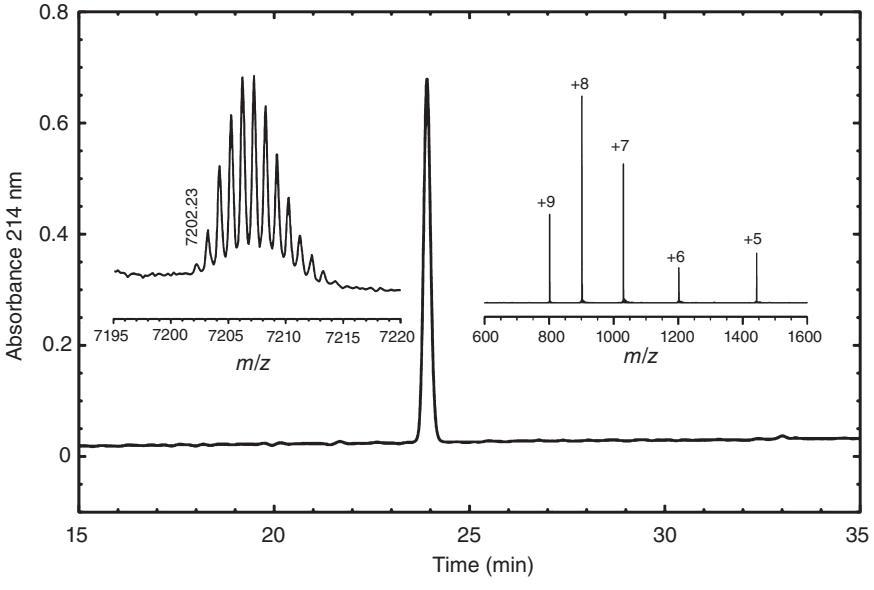


Figure 1.2 Analytical HPLC and MS characterization of a synthetic protein. With modern mass spectrometric instrumentation, the monoisotopic mass can be experimentally determined and compared with the theoretical mass. Source: Adapted from Durek et al. [23].

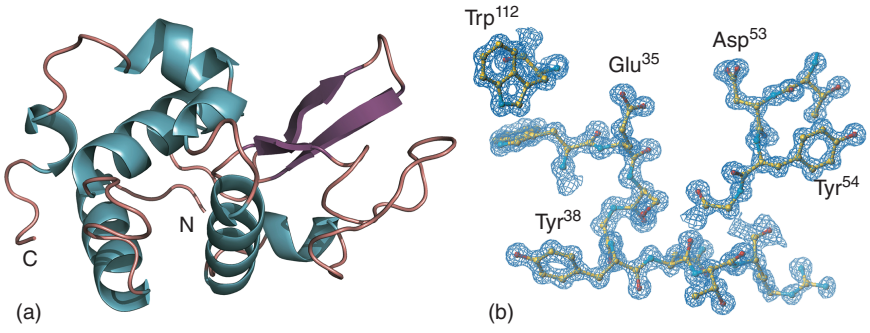


Figure 1.3 X-ray diffraction structure of crystalline human lysozyme prepared by total chemical synthesis. (a) Cartoon representation of the secondary structure of the folded human lysozyme synthetic protein molecule; (b) $2F_o - F_c$ electron density map , showing the quality of the data acquired at a resolution of 1.04 Å. Source: Durek et al. [11]. © 2007 National Academy of Sciences.

structure [26]; (ii) intrinsically disordered proteins only fold, if at all, in the presence of their target molecules [27].

- **Determination of the atomic structure of the folded protein molecule.**
  - X-ray crystallography, including racemic protein crystallography, [28] can be used to determine the structure of the folded, homogeneous synthetic protein molecule [29, 30].
  - X-ray structural data will also confirm the connectivity of any disulfides that are present, and at sufficiently high resolution can reveal the presence of aberrant chemical modifications of the synthetic protein (Figure 1.3).
  - Note that because fractional crystallization from protein mixtures is common, determination of the crystal structure of a synthetic protein by X-ray diffraction does not in itself establish that all of the synthetic protein molecules have that structure and are undamaged. Homogeneity of the synthetic protein molecule must be established *before* determination of the crystal structure.
  - For small proteins, the folded structure can be determined using natural abundance protein NMR techniques [31].
- Biological/biochemical assays.
  - Quantitative assays, especially measurements of the catalytic activity of a synthetic enzyme molecule, can be very informative but are inadequate as proof of homogeneity or correct structure of the synthetic protein.

The total synthesis of the chemokine CCL2 (MCP-1) by Grygiel et al. [32] is a near-perfect example of a properly documented total chemical synthesis of a protein molecule, illustrating essentially all of the synthetic steps described above together with meticulous and complete characterization of the synthetic protein as a single molecular species of defined structure (Figure 1.4).

Selected further examples of well-characterized synthetic proteins can be found in Refs [24, 33–42].

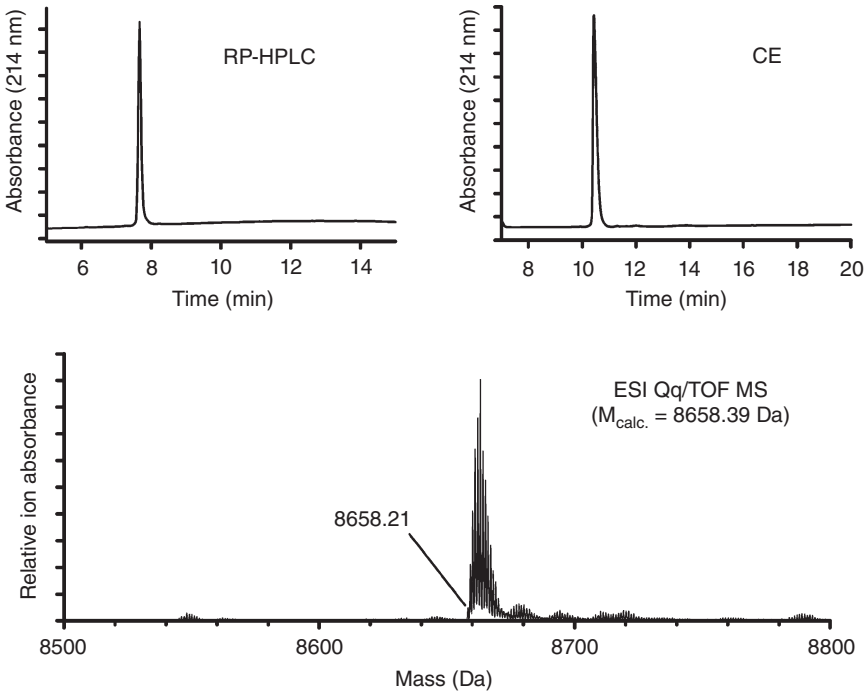


Figure 1.4 Analytical characterization of the protein CCL2 (MCP-1) prepared by total chemical synthesis. [32] Purity of the synthetic protein was verified by reverse phase HPLC and by capillary electrophoresis, two high resolution analytical methods based on different separation principles (hydrophobicity; charge). The monoisotopic mass was measured by LC-ESI Qq/TOF mass spectrometry. In addition to the data shown here, the disulfide bonds are verified by enzymatic digestion and LC-(MS-MS) of the resulting peptide fragments, and the crystal structure of the synthetic protein is determined at a resolution of 1.9 Å. Source: From Grygiel et al. [32].

1.3 Comments on Characterization of Synthetic Protein Molecules

> *"In the field of protein synthesis it is my confident hope that tomorrow's deeds will catch up with today's titles, and that we shall truly be able to obtain enzymatically active proteins as synthetic substances: as materials composed of a single molecular species."*
>
> Josef Rudinger (3rd American Peptide Symposium 1972)

1.3.1 Homogeneity

The single most important aspect of the characterization of a synthetic protein is rigorous verification of its homogeneity, i.e. that the synthetic product is a single

molecular species. In verifying the homogeneity of a synthetic protein it is essential to use analytical techniques based on separation principles that are distinct from the purification method(s) used in the synthesis of the protein molecule. Preparative HPLC is almost invariably used as the final purification step and closely related by-products from the synthesis may be co-purified with the desired protein product. For that reason, analytical HPLC is not a sufficient proof of molecular homogeneity of a synthetic protein product. *Direct infusion* ESI-MS is a good technique for checking for the presence of protein impurities that have masses that differ from the mass of the target protein molecule. However, impurities that have the same mass as the target protein will not be detected.

Characterization of a synthetic protein must also rule out the presence of inadvertent modifications of the covalent structure that may have been introduced in the course of chemical synthesis. For example, chemical manipulations of peptide chains can lead to the formation of isopeptide bonds (i.e. beta-amide links) at aspartic acid residues, formation of succinimides at Asn residues, deamidations of Asn and/or Gln side chains, *N*-to-*O* acyl shifts at Gly-Ser sequences, and any of a variety of other possible by-products. These abnormal modifications of the covalent structure are often not resolved by analytical HPLC and frequently result in small (1 Da) or zero differences in the observed mass of the synthetic protein molecule, and thus can be difficult or impossible to detect by mass spectrometry alone.

A state-of-the-art technique for detection of byproducts and for the rigorous verification of the molecular homogeneity of a synthetic protein is CIEF. CIEF is a charge-based, ultrahigh-resolution analytical method that can reveal the presence of impurities that do not directly involve a unit charge change, even subtle differences as benign as the replacement of an uncharged amino acid residue in the protein's polypeptide chain by another uncharged residue. For that reason, CIEF is to be preferred for documenting the absence of unexpected by-products and for the verification of the molecular homogeneity of synthetic peptides and proteins. Recent descriptions of state-of-the-art CIEF can be found here [43, 44].

> *The time has come for capillary isoelectric focusing data to be mandatory verification of homogeneity in the final characterization of a synthetic protein.*

1.3.2 Amino Acid Sequence

The amino acid sequence of a protein determines every aspect of its molecular structure. Ironically, the aspect of the structure of a chemically synthesized protein molecule that is most often NOT experimentally verified is the amino acid sequence of a protein's polypeptide chain. Because the target amino acid sequence of the synthetic protein's polypeptide chain is known, it is straightforward to verify the amino acid sequence of a synthetic polypeptide chain by enzymatic digestion and HPLC-(MS-MS) analysis of the peptide fragments [33].

The all-to-frequent failure to verify the amino acid sequence of a synthetic protein product should be rectified.

1.3.3 Chemical Analogues

An important objective of total protein synthesis is to make *uniquely chemical analogues* of the natural protein molecule, including non-coded amino acids [45], stable analogues of post-translational modifications [46], fixed elements of secondary structure [47], chemical engineering of the polypeptide chain backbone [48], site-specific labeling with isotopes [49] or fluorophores [50], glycoproteins [51, 52] and glycoprotein mimetics [53], novel topologies such as circular [54] or concatenated polypeptide chains [55], protein diastereomers [56–58], and mirror-image protein molecules [59]. Thus, in addition to characterizing features of protein molecular structure found in natural proteins, characterization of synthetic proteins must include *verification of the novel features introduced by chemical synthesis*.

1.3.4 Limitations of SPPS

Stepwise synthesis of polypeptide chains has its limitations. Because of the statistical accumulation of resin-bound by-products, even the most highly optimized stepwise SPPS methods should *never* be used for the preparation of peptide segments longer than ~50 amino acid residues. Depending on amino acid composition and the sequence of the target peptide, the practical limit for preparing homogeneous synthetic peptides free of closely related contaminant peptide co-products (e.g. deletions, terminations, deamidations, isopeptide bonds, *N*-to-*O* acyl shifts, covalent modifications) can be as short as 40 or even fewer amino acid residues [60]. Effective purification and careful characterization of each synthetic peptide segment used in total protein synthesis is essential.

1.3.5 Folding as a Purification Step

The reality is that no matter how meticulously they have been purified, 30–50 amino acid residue peptide segments prepared by stepwise SPPS will almost always contain some level of closely related peptide impurities. Even small amounts of impurities present in the peptide segments used in protein synthesis will carry over in each bond-forming step and frequently will be found in the full-length synthetic polypeptide chain after purification. Fortunately, particularly for disulfide-constrained globular proteins, the folding step can be an effective further purification in which closely related polypeptide chains that do not fold correctly are subsequently removed during purification of the folded synthetic protein. Examples of LCMS data before and direct infusion ESI-MS data after folding are shown in Figure 1.5.

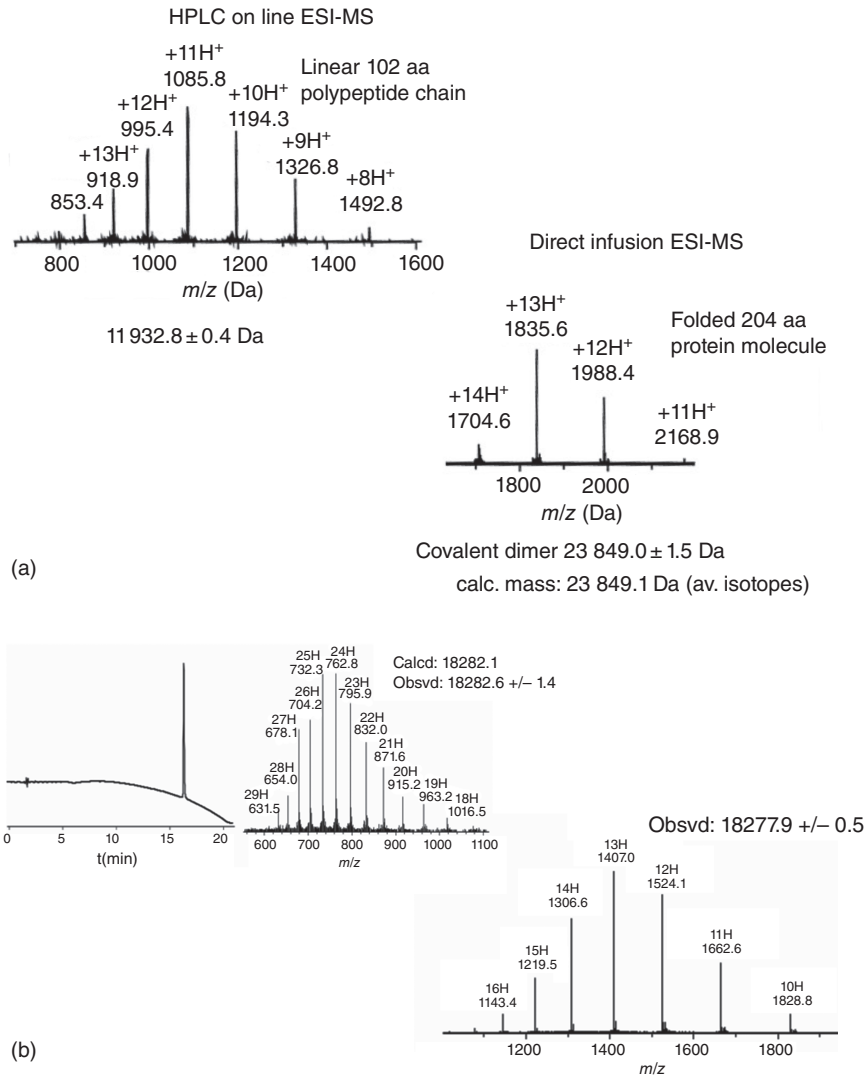


Figure 1.5 The folding step can significantly reduce the amounts of residual impurities present in synthetic polypeptide chains. Two examples are shown here. (a) VEGF A: (upper left) MS data from LCMS of purified linear VEGF A polypeptide chain – ESI-MS data were acquired across the entire UV absorbing peak in the LC chromatogram; (lower right) Direct infusion ESI-MS of purified folded covalent homodimer VEGF A synthetic protein molecule. Note the reduced amounts of impurities on the low *m/z* side of each charge state in the folded protein data. Source: Adapted from Mandal and Kent [21]. (b) [$Lys^{24,38,83}$]EPO: [upper left] MS data from LCMS of purified linear [$Lys^{24,38,83}$]EPO polypeptide chain – ESI-MS data were acquired across the entire UV absorbing peak shown in the LC chromatogram to the left; (lower right) Direct infusion ESI-MS of purified folded [$Lys^{24,38,83}$]EPO synthetic protein molecule. Note the essential absence of impurities on the low *m/z* side of each charge state in the folded protein data. Small peaks to the right of each charge state are sodium and calcium ion adducts. MS data representations have been digitally scaled to give the same vertical total ion count and horizontal *m/z* scales. Adapted from [22].

1.4 Summary

Total synthesis of protein molecules has been made straightforward by modern ligation chemistries, and is used in research laboratories throughout the world [61]. Optimally, chemical protein synthesis involves the integrated use of high-resolution analytical methods at multiple points in the synthetic process, combined with meticulous characterization of the final protein product. With the ready availability of powerful analytical techniques, it is no longer acceptable to only report a mass found at an ill-defined point in an LCMS chromatogram as "proof" of the purity and correct molecular structure of a synthetic protein.

The most important aspects of characterization of synthetic proteins as single molecular species of define structure are

- high-resolution charge-based analysis to verify molecular homogeneity
- measurement of the precise mass of the synthetic protein
- verification of the amino acid sequence of the protein's polypeptide chain
- determination of the structure of the synthetic protein by X-ray crystallography.

References

1 Burley, S.K., Berman, H.M., Christie, C. et al. (2018). RCSB protein data bank: sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci.* 27: 316–330.

2 Wolfenden, R. and Snider, M.J. (2001). The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.* 34: 938–945.

3 Mudge, J.M., Jungreis, I., Toby Hunt, T. et al. (2019). Discovery of high-confidence human protein-coding genes and exons by whole-genome PhyloCSF helps elucidate 118 GWAS loci. *Genome Res.* 29: 2073–2087.

4 Yooseph, S., Sutton, G., Rusch, D.B. et al. (2007). The sorcerer II global ocean sampling expedition: expanding the universe of protein families. *PLoS Biol.* 5 (3): e16.

5 Anfinsen, C.B. (1973). Principles that govern folding of protein chains. *Science* 181: 223–230.

6 Brocchieri, L. and Karlin, S. (2005). Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* 33: 3390–3400.

7 Kent, S.B.H. and Alewood, P.F. (2014). Synthetic biomolecules. *Curr. Opin. Chem. Biol.* 22: viii–xi.

8 Kent, S.B.H. (2009). Total chemical synthesis of proteins. *Chem. Soc. Rev.* 38: 338–351.

9 Bode, J.W. (2017). Chemical protein synthesis with the α-ketoacid–hydroxylamine ligation. *Acc. Chem. Res.* 50: 2104–2115.

10 Conibear, A.C., Watson, E.E., Payne, R.J., and Becker, C.F.W. (2018). Native chemical ligation in protein synthesis and semi-synthesis. *Chem. Soc. Rev.* 47: 9046–9068.

11 Durek, T., Torbeev, V.Y., and Kent, S.B.H. (2007). Convergent chemical synthesis and high resolution X-ray structure of human lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* 104: 4846–4851.

12 Dhayalan, B., Mandal, K., Rege, N. et al. (2017). Scope & limitations of Fmoc chemistry SPPS-based approaches to the total synthesis of Insulin Lispro via ester insulin. *Chem. Eur. J.* 23: 1709–1716.

13 Mehok, A.R., Mant, C.T., Gera, L. et al. (2002). Preparative reversed-phase liquid chromatography of peptides. Isocratic two-step elution system for high loads on analytical columns. *J. Chromatogr. A* 972: 87–99.

14 Creamer, J.S., Oborny, N.J., and Lunte, S.M. (2014). Recent advances in the analysis of therapeutic proteins by capillary and microchip electrophoresis. *Anal. Methods* 6 (15): 5427–5449.

15 Chait, B.T., Wang, R., Beavis, R., and Kent, S.B.H. (1993). Protein ladder sequencing. *Science* 262: 89–92.

16 Bang, D. and Kent, S.B.H. (2004). A one-pot chemical synthesis of Crambin. *Angew. Chem. Int. Ed. Engl.* 43: 2534–2538.

17 Fang, G.-M., Wang, J.-X., and Liu, L. (2012). Convergent chemical synthesis of proteins by ligation of peptide hydrazides. *Angew. Chem. Int. Ed.* 51: 10347–10350.

18 Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S.B.H. (1994). Synthesis of proteins by native chemical ligation. *Science* 266: 776–779.

19 Dawson, P.E., Churchill, M., Ghadiri, M.R., and Kent, S.B.H. (1997). Modulation of reactivity in native chemical ligation through the use of thiol additives. *J. Am. Chem. Soc.* 119: 4325–4329.

20 De Bernardez, C.E., Schwarz, E., and Rudolph, R. (1999). Inhibition of aggregation side reactions during in vitro protein folding. *Methods Enzymol.* 309: 217–236.

21 Mandal, M. and Kent, S.B.H. (2011). Total chemical synthesis of biologically active vascular endothelial growth factor. *Angew. Chem. Int. Ed.* 50: 8029–8033.

22 Liu, S., Pentelute, B.L., and Kent, S.B.H. (2012). Convergent chemical synthesis of [$Lys^{24,38,83}$]human erythropoietin. *Angew. Chem. Int. Ed.* 51: 993–999.

23 Durek, T., Vetter, I., Wang, C.-I.A. et al. (2013). Chemical engineering and structural and pharmacological characterization of the α-scorpion toxin OD1. *ACS Chem. Biol.* 8: 1215–1222.

24 Hsieh, Y.S.Y., Wijeyewickrema, L.C., Wilkinson, B.L. et al. (2014). Total synthesis of homogeneous variants of hirudin P6: a post-translationally modified anti-thrombotic leech-derived protein. *Angew. Chem. Int. Ed.* 53: 3947–3951.

25 Bang, D., Chopra, N., and Kent, S.B.H. (2004). Total chemical synthesis of Crambin. *J. Am. Chem. Soc.* 126: 1377–1383.

26 Sohma, Y., Hua, Q.-x., Liu, M. et al. (2010). Contribution of residue B5 to the folding and function of insulin and IGF-I. *J. Biol. Chem.* 285 (7): 5040–5055.

27 Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* 37 (12): 509–516.

28 Yeates, T.O. and Kent, S.B.H. (2012). Racemic protein crystallography. *Ann. Rev. Biophys.* 41: 41–61.

29 Torbeev, V.Y. and Kent, S.B.H. (2007). Convergent chemical synthesis and crystal structure of a 203 amino acid 'covalent dimer' HIV-1 protease enzyme molecule. *Angew. Chem. Int. Ed. Engl.* 46: 1667–1670.

30 Huang, Y. and Liu, L. (2015). Chemical synthesis of crystalline proteins. *Sci. China Chem.* 58 (12): 1779–1781.

31 Schroeder, I.S., Rash, L.D., Vila-Farrs, X. et al. (2014). Chemical Synthesis, 3D structure, and ASIC binding site of the toxin mambalgin-2. *Angew. Chem. Int. Ed.* 53: 1017–1020.

32 Grygiel, T.L.R., Teplyakov, A., Obmolova, G. et al. (2010). Synthesis by native chemical ligation and crystal structure of human CCL2. *Biopolymers – Pept. Sci.* 94: 350–359.

33 Zhang, Y., Hirota, T., Kuwata, K. et al. (2019). Chemical synthesis of atomically tailored SUMO E2 conjugating enzymes for the formation of covalently linked SUMO−E2−E3 ligase ternary complexes. *J. Am. Chem. Soc.* 141: 14742–14751.

34 Pan, M., Gao, S., Zheng, Y. et al. (2016). Quasi-racemic X-ray structures of K27lLinked ubiquitin chains prepared by total chemical synthesis. *J. Am. Chem. Soc.* 138: 7429–7435.

35 Huang, Y.-C., Chen, C.-C., Gao, S. et al. (2016). Synthesis of l- and d-ubiquitin by one-pot ligation and metal-free desulfurization. *Chem. Eur. J.* 22: 7623–7628.

36 Dang, B., Wu, H., Mulligan, V.K. et al. (2017). De novo design of covalently constrained mesosize protein scaffolds with unique tertiary structures. *Proc. Natl. Acad. Sci. U. S. A.* 114: 10852–10857.

37 Wilken, J., Hoover, D., Thompson, D.A. et al. (1999). Total chemical synthesis and crystal structure of the potent anti-HIV protein AOP-RANTES. *Chem. Biol.* 6: 43–51.

38 Bacchi, M., Jullian, M., Sirigu, S. et al. (2016). Total chemical synthesis, refolding, and crystallographic structure of fully active immunophilin calstabin 2 (FKBP12.6). *Protein Sci.* 25: 2225–2242.

39 Ghassemian, A., Anderson Wang, C.I., Yau, M.-K. et al. (2013). Efficient chemical synthesis of human complement protein C3a. *Chem. Commun.* 49: 2356–2358.

40 Boerema, D.J., Tereshko, V.A., and Kent, S.B.H. (2008). Total Synthesis by modern chemical ligation methods & high resolution (1.1Å) X-ray structure of ribonuclease A. *Biopolymers – Pep. Sci.* 90: 278–286.

41 Avital-Shmilovici, M., Mandal, K., Gates, Z.P. et al. (2013). Convergent chemical synthesis of ester insulin: determination of the high resolution X-ray structure by racemic protein crystallography. *J. Am. Chem. Soc.* 135: 3173–3185.

42 Ho Yeung, H., Squire, C.J., Yosaatmadja, Y. et al. (2016). Radiation damage and racemic protein crystallography reveal the unique structure of the GASA/Snakin protein superfamily. *Angew. Chem. Int. Ed.* 55: 7930–7933.

43 Cologna, P.M.R. and S.M. (2017). Isoelectric point separations of peptides and proteins. *Proteomes* 5: 4. https://doi.org/10.3390/proteomes5010004.

44 Rozing, G. (2019). Imaged capillary iso-electric focusing: background, status, and perspectives. In: *Chromatography Today*, 8–14. International Labmate Ltd.

45 Wlodawer, A., Miller, M., Jaskolski, M. et al. (1989). Crystal structure of synthetic HIV-1 protease: conserved fold in retroviral proteases. *Science* 245: 616–621.

46 Kee, J.M., Villani, B., Carpenter, L.R., and Muir, T.W. (2010). Development of stable phosphohistidine analogues. *J. Am. Chem. Soc.* 132: 14327–14329.

47 Baca, M., Alewood, P., and Kent, S.B.H. (1993). Structural engineering of HIV-1 protease with a β-turn mimic of fixed geometry. *Protein Sci.* 2: 1085–1091.

48 Schnölzer, M. and Kent, S.B.H. (1992). Constructing proteins by dovetailing unprotected synthetic peptides: backbone engineered HIV protease. *Science* 256: 221–225.

49 Dhayalan, B., Fitzpatrick, A., Mandal, K. et al. (2016). Efficient total chemical synthesis of $^{13}C{=}^{18}O$ isotopomers of human insulin for isotope-edited FTIR. *ChemBioChem* 17: 415–420.

50 Dang, B., Kubota, T., Correa, A.M. et al. (2014). Total chemical synthesis of biologically active fluorescent dye-labeled Ts1 toxin. *Angew. Chem. Int. Ed.* 53: 8970–8974.

51 Yamamoto, N., Tanabe, Y., Okamoto, R. et al. (2008). Chemical synthesis of a glycoprotein having an intact human complex-type sialyloligosaccharide under the Boc and Fmoc synthetic strategies. *J. Am. Chem. Soc.* 130 (2): 501–510.

52 Murakami, M., Kiuchi, T., Nishihara, M. et al. (2016). Chemical synthesis of erythropoietin glycoforms for insights into the relationship between glycosylation pattern and bioactivity. *Sci. Adv.* 2 (e1500678): 1–12.

53 Kent, S.B.H. (2013). Bringing the science of proteins into the realm of organic chemistry: total chemical synthesis of SEP (synthetic erythropoiesis protein). *Angew. Chem. Int. Ed.* 52: 11988–11996.

54 Clark, R.J. and Craik, D.J. (2010). Native chemical ligation applied to the synthesis and bioengineering of circular peptides and proteins. *Biopolymers* 94 (4): 414–422.

55 Yan, L.Z. and Dawson, P.E. (2001). Design and synthesis of a protein catenane. *Angew. Chem. Int. Ed.* 40: 3645–3627.

56 Bang, D., Makhatadze, G.I., Tereshko, V. et al. (2005). Total chemical synthesis and X-ray crystal structure of a protein diastereomer: [D-Gln35]Ubiquitin. *Angew. Chem. Int. Ed. Engl.* 44: 3852–3856.

57 Bang, B., Gribenko, A.V., Tereshko, V. et al. (2006). Dissecting the energetics of protein α-helix C-cap termination through chemical protein synthesis. *Nat. Chem. Biol.* 2: 139–143.

58 Dang, B., Kubota, T., Shen, R. et al. (2017). Inversion of Thr and Ile side chain stereochemistry in a protein molecule: impact on the folding, stability, and structure of the ShK toxin protein molecule. *Angew. Chem. Int. Ed.* 56: 3324–3328.

59 Mandal, K., Uppalapati, M., Ault-Riché, D. et al. (2012). Chemical synthesis and X-ray structure of a heterochiral {D-protein antagonist plus VEGF-A} protein complex by racemic crystallography. *Proc. Natl. Acad. Sci. U. S. A.* 109: 14779–14784.

60 Behrendt, R., White, P., and Offer, J. (2016). Advances in Fmoc solid-phase peptide synthesis. *J. Pept. Sci.* 22: 4–27.

61 Agouridas, V., El Mahdi, O., Cargoët, M., and Melnyk, O. (2017). A statistical view of protein chemical synthesis using NCL and extended methodologies. *Bioorg. Med. Chem.* 25: 4938–4945.