CHAPTER 7

Recombinant and Semisynthesis of Peptides

P. FONTE*, A. MALCANGI, M. STUCCHI AND A. OPPEDISANO

Olon SpA, via Solferino, 10036 Settimo Torinese (TO), Italy
*E-mail: pfonte@olonspa.it

7.1 Introduction

7.1.1 Manufacture of Peptide Therapeutics

Peptide therapeutics are emerging as a potent new class of active pharmaceutical ingredients (APIs), even though historically their original usage dates back to the 1920s, when insulin was first used in humans to treat diabetes, and in the last decade more and more peptide drugs have entered the market. This is a rich field of research, as illustrated by the fact that both academic groups and pharma industries are actively researching and investing money. This is a clear sign of a field that is not yet mature, with many possibilities for further improvements and discoveries.

The preparation of peptides is a very deceptive area of work, where the apparently simple chemistry involved can lead to underestimation of the complexity of the system. Indeed, most of the synthetic work revolves around the formation of an amide bond to forge the peptide carbon backbone. However, when applied to real systems, this simple operation can meet unexpected difficulties: the unique chemical and physical properties of peptides,

Green Chemistry Series No. 85
Sustainability in TIDES Chemistry: Green Approaches to Oligonucleotides and Oligopeptides Synthesis
Edited by Alessandra Tolomelli, Lucia Ferrazzano and Walter Cabri
© The Royal Society of Chemistry 2024
Published by the Royal Society of Chemistry, www.rsc.org

the formation of secondary structures, the low solubilities, the unusual and surprising reactivity of certain amino acid sequences, and so on. All of these properties can lead to many struggles along the path to the final target. In light of the issues, most syntheses rely on the leverage of solid-phase peptide synthesis (SPPS), which is a very well-established technique, with a deep background of knowledge and a vast database of solutions. SPPS is a widely known methodology that is generally economically advantageous and easy to develop in order to prepare a target molecule; moreover, it has some clear strong points on its side: excess reagents/reactants are easy to remove, scale-up is very straightforward, there is ample opportunity to implement automation in the process, and lastly the insertion of non-natural amino acids or non-canonical structures is facile. The other most common methodology employed in the preparation of peptides is liquid-phase peptide synthesis (LPPS), which is very similar to SPPS but performed in solution.

On the other hand, both techniques have clear disadvantages: often peptides and amino acids show very limited solubility in the solvent or system used in the synthesis, which can be a cause, with other factors, of the low yields that are sometimes observed; in addition, both methodologies use hazardous solvents and/or reagents, which makes them not very appealing from a sustainability point of view. Finally, the cost of these techniques increases with each amino acid added to the sequence, and the cost impact for peptides with sequences longer that 20–25 amino acids can be very high, with a negative impact on the cost of the final drug substance and on the profit for the manufacturer.

While peptides with longer sequences (40 or even 50 amino acids long) are making their way to the market, recombinant technologies are now becoming of commercial interest; more specifically, in the field of pharmaceutical peptides there is a growing requirement for more complex structures, which are not simple to access *via* SPPS or LPPS. Generally, the chemical synthesis of peptides longer than 20 amino acids is cumbersome and expensive. As more complex structures and longer amino acid chains are coming through the pipeline, the classical methodologies of SPPS and LPPS for the preparation of APIs are less competitive, and at this point recombinant deoxyribonucleic acid (rDNA) technologies are gaining the edge (see Figure 7.1).

Currently, about 80 peptides are approved for human use, more than 150 are in clinical development, and many more are in preclinical studies.[1] Every year new therapeutic peptides are approved by the US Food and Drug Administration (FDA) and regulatory agencies around the world. Peptide therapeutics have distinct features compared with small molecules: they are generally more potent and more specific and can establish interactions with the target receptors that are just not feasible for small molecules. In addition, peptides can target more types of receptors and, thanks to their nature, peptides are less harmful to the human body. On the other hand, peptides are easily degraded *in vivo*, making them difficult to administer orally, and they do not easily cross the cell membrane.

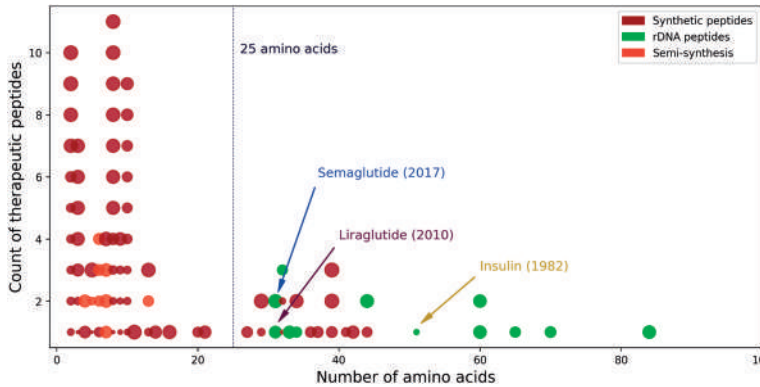


Figure 7.1 Number of therapeutic peptides as a function of the number of amino acids. Dark brown dots represent APIs manufactured *via* synthesis, light brown dots represent semisynthetic APIs and green dots represent APIs manufactured *via* rDNA. The size of the dots is proportional to the year of production (larger = more recent).

In light of these considerations, rDNA routes for the preparation of new peptides are increasing. With this methodology, it is possible to engineer bacteria and yeasts, among others, to naturally express the desired target peptide. As one can imagine, the costs associated with such production methods are minimally dependent on the peptide length, hence the preparation of long peptides is economically advantageous at this point. Overall, rDNA techniques have many appealing aspects: natural organisms are exploited and engineered to deliver therapeutic peptides, carried out in aqueous systems, employing non-hazardous reagents such as molds and syrups. The sustainability of the process in this case goes hand in hand with the economic advantages of the methodology. In addition, the building of the molecule is performed by the natural organism, and this comes with the clear advantage of the natural precision in codifying the correct structure, with the correct amino acid sequence and the required stereo form. Nevertheless, the industrialization of such processes is more difficult to develop and more time is required in the R&D phase to scale up a productive process. Additionally, some restriction is mandatory on the fermented peptides, as natural organisms cannot express non-natural structures or non-canonical amino acids. Finally, when comparing and exploring the various possibilities for producing peptides, one must also consider the differences in the quality of the products: the quality profiles of the same peptides prepared *via* biotechnologies or *via* SPPS are radically different: in one case the most common impurities includes deletion and duplication of amino acids in the sequence, with many process-related impurities being formed during the many synthetic steps; in the other situation, most of the impurities are product related, where degradation pathways and bio-derived residues are the most critical issues.

Taken together, the rDNA and SPPS processes for manufacturing therapeutic peptides are two possibilities that do not overlap but on the contrary cover different areas of needs in the industry panorama. The ever-changing scenario of API manufacturing, with multiple players, such as regulatory agencies, sustainability targets, and market requirements themselves, will naturally set the emphasis on the most suitable technologies from process to process.

Currently, the high impact of APIs prepared *via* biotechnologies is exemplified by the striking success of semaglutide. This is a peptide produced by Novo Nordisk, originally prescribed as an injectable antidiabetic (Ozempic) and subsequently marketed as an orally disposable formulation (Rybelsus), with a large impact on the market. Thereafter, the drug was shown to have an impressive weight-loss effect, and sales skyrocketed. Many countries, such as the USA, saw a huge increase in medical prescriptions of the drug and Novo Nordisk, the company holding the patents for semaglutide, in 2023 registered a market value higher than the gross domestic product (GDP) of Denmark (the country of location of the company headquarters), and it is the second most valuable company in Europe.

7.1.2 Manufacturing Sustainability

The concept of industrial sustainability and the idea of realizing clean industrial processes have gained importance over time. There is a need to reshape industrial process technologies to prevent pollution at the source. Biotechnology represents an opportunity to move from fossil-based to bio-based industry, reducing global warming. Fermentation, as a water-based process, has clear ecological advantages. The raw materials used are renewable resources and wastes may be biodegraded in wastewater treatment plants or recycled to produce clean fuel. Nevertheless, an assessment of the environmental impact of novel developed processes is important to have a clear understanding of the product life cycle, from raw materials sourcing, processing, manufacture, and transport to end-of-life disposal, and to identify more sustainable process options.

The use of specific analytical tools such as life-cycle assessment (LCA) can provide decision support on how to design future technologies. LCA provides a holistic view of a process. The application of this methodology may support the identification of the types of microorganisms to be utilized, can drive the selection of optimal substrates, and can allow the choice of process model and bioreactor technology to minimize energy consumption.[2–4]

The result of the scale-up of a process to the full commercial scale can also be predicted by LCA and negative environmental impact effects can be reduced.[4] The identification of hotspots can, in fact, be used for process optimization. The environmental impact of a bioprocess is primarily due to energy and water consumption and the improvement of the process yield through strain optimization is the most effective strategy to reduce it.[5] Continuing technical innovation and the development of recombinant DNA

technology increased the number of industrial biotechnology applications. This technology provides a powerful method for combining diverse genetic capabilities, thus permitting the engineering of organisms with new specific capabilities. The tailor-made synthesis of products by modified microorganisms opens up new possibilities for the production of chemicals for different purposes.

This approach can be useful every time there is a problem in the supply chain of the product, for example if the product is present in low concentration in the natural source or if it is difficult to extract and purify it from the natural source. rDNA technology can be an opportunity in the case of a product that is difficult to produce by chemical synthesis or may be used to design chemicals that would not occur by natural pathways.

An example of an important application of rDNA is represented by human insulin. This hormone cannot be obtained directly from the natural source and extraction from animal tissues has several drawbacks. The possibility of producing recombinant human insulin in microorganisms represented a major success and paved the way to the production of recombinant human hormones and their therapeutic use.

7.2 Human Insulin: The First Hormone Produced Using Recombinant DNA Technology

7.2.1 History of Insulin

Human insulin is comprised of 51 amino acids and has a molecular weight of 5808 Da. It is produced by beta cells of the pancreas and plays a key role in regulating carbohydrate and fat metabolism in the body. Insulin is synthesized as a single polypeptide known as preproinsulin (110 amino acids). The mature form consists of an A-chain with 21 amino acids and a B-chain containing 30 amino acids and the two polypeptides are linked together by two disulfide bonds. Further, the A-chain has an intrachain disulfide bond.[6,7]

Insulin is secreted by beta cells of islets of Langerhans in a glucose-dependent manner and acts to stimulate glucose uptake in the liver, muscle, and adipose tissues. Like other protein hormones, insulin binds with specific receptors on the outer membrane of target cells, which causes metabolic processes to take place within those cells. Insulin was the first human hormone produced using rDNA technology.

Diabetes mellitus was first described over 2000 years ago and represented a fatal medical condition until the early 1920s.[8] The discovery of insulin was a milestone in medical science. The work leading to the discovery of insulin was started in 1921 by the Canadian surgeon Frederick Grant Banting and his assistant Charles Best, who worked at the University of Toronto under the supervision of John Macleod. They isolated insulin from pancreatic animal tissues and carried out the initial studies administering insulin preparations to diabetic patients.[9,10] Banting and Best started a collaboration with Eli Lilly that led in 1922 to the first commercially available insulin.[11] In Europe,

two Danish companies, which in the late 1980s merged into Novo Nordisk, started to produce insulin.

Insulin from cows and pigs was used to treat diabetes for many years and saved millions of lives. Nevertheless, the use of animal insulin was problematic because its composition was not identical to that of human insulin and stimulated the production of antidrug antibodies, preventing the effective treatment of diabetic patients or even leading to worsening of their diabetes.[12] In addition, there was the problem of contamination of insulin of animal origin with proteases and the risk of hydrolysis of the product.[11] Major efforts were made in subsequent years to improve purification to achieve safety and stability of the product.

The bovine insulin amino acid sequence was elucidated by the British biochemist Frederick Sanger in 1955. For this discovery, Sanger was awarded the Nobel Prize for Chemistry in 1958.[13] In 1960, the human insulin sequence was elucidated and some comparative studies with bovine and porcine insulin evidenced a three and a single amino acid difference, respectively.[11] In order to obtain a therapeutic product identical in sequence with native human insulin, the total chemical synthesis was considered, but was found to be economically unaffordable.

The enzymatic conversion of porcine into human insulin was developed in the 1970s and found to be commercially feasible.[11,14] This result was achieved by considering that protein semisynthesis could be carried out very conveniently if protein fragments obtained by the action of proteinases could be recombined again by the action of the same proteinases. In order to achieve such an objective, techniques were needed to shift the conditions in such a manner that the enzymes remained catalytically active, and under one set of conditions peptide bond hydrolysis was favored and under the other peptide bond synthesis was favored. The addition of organic cosolvents was used to shift the equilibria in enzyme-catalyzed reactions towards synthesis.[14] The product obtained was termed "semisynthetic human insulin" or "human insulin (emp)" (emp = enzyme-modified porcine).

In the 1970s, rDNA methods for cloning and expressing genes in *Escherichia coli* were under intense development. The invention of DNA cloning by Stanley Cohen and Herbert Boyer led to the creation of various recombinant proteins with medicinal uses. In 1978 and in 1979, the successful expression in *E. coli* of two mammalian hormones, first somatostatin and then human insulin, was achieved by Genentech. Somatostatin represented only a proof of concept to test the feasibility of the method on a 14-amino acid mammalian hormone. The method used for insulin was essentially the same as that for somatostatin.[15] Human insulin, which is comprised of a total of 51 amino acids (21 in the A-chain and 30 in the B-chain), was the first therapeutically useful protein product resulting from rDNA technology.

The first recombinant insulin was approved by the FDA in 1982 for the treatment of diabetic patients. The successful production of human insulin in bacteria provided, for the first time, a practical, scalable source of human insulin.

7.2.2 Insulin Production Process

The use of cows and pigs to obtain the pancreatic hormone had a negative footprint in terms of sustainability and could not support the increasing world demand for insulin. The change from extraction to recombinant insulin production process led to several advantages:

- an improvement in quality;
- a reduction of the risk of immunogenic reaction that could compromise the efficacy of the therapy;
- the possibility of producing insulin on a large scale at an affordable cost;
- an improvement in the sustainability of the production process.

The first production process developed by Genentech used an approach that required the expression of chemically synthesized complementary DNA (cDNA) encoding for the insulin A and B chains separately in *E. coli*. Each chain was fused to a carrier protein and produced independently. After expressing the fusion proteins and releasing the two chains by chemical cleavage, the two free chains were purified and co-incubated under optimum reaction conditions that promoted the generation of intact and bioactive insulin by disulfide bond formation. The first recombinant insulin available on the market was produced using this method.

Another approach consisted in the expression of a single chemically synthesized cDNA encoding for native human proinsulin in *E. coli* as inclusion bodies followed by solubilization and a refolding procedure. After purification, excision of C-peptide was performed with a bioconversion step (proteolytic digestion). This approach was more efficient and convenient than the two-chains combination approach for the large-scale production of therapeutic insulin. The method required a single fermentation and purification process and was economically affordable. Eli Lilly adopted this technology to produce Humulin, which entered the market in 1986.

Recombinant insulin represented a significant improvement in purity and quality consistency over semisynthetic insulin and permitted a worldwide supply.[16]

E. coli is not the only host utilized for insulin production.[17,18] In Europe, *Saccharomyces cerevisiae* was used by Novo Nordisk and currently several recombinant commercial insulin products are expressed in this strain.[7] The advantage of the yeast expression system is the extracellular secretion, which can simplify the purification of the protein. Although *S. cerevisiae* is still the predominant yeast system for insulin production, several alternative yeast hosts have become available in recent years. An example is *Pichia pastoris*, which is a very useful expression host with superior features. The use of integrative vectors allows the genetic stability of the strain to be improved. Its strong and tightly regulated methanol-inducible alcohol oxidase 1 (AOX1) promoter and its capacity to reach very high cell densities, by simple cultivation strategies, contribute to the high-level production of recombinant

proteins. *P. pastoris* has proven its applicability for insulin production, and a comparative study revealed a performance equal to or even better than that of *S. cerevisiae*. Biosimilar insulin analogs produced by Mylan (rh-insulin glargine and rh-insulin aspart) were obtained in *P. pastoris* and approved in 2018–2021. Also, Baxter received approval in 2022 for a biosimilar rh-insulin using *P. pastoris* as the expression system.

Except for Novo Nordisk, which is still using *S. cerevisiae*, all other players are using the higher performance expression system in *E. coli* or *P. pastoris* for the recombinant production of insulin and insulin analogs.

7.3 Glucagon and GLP-1 Analogs

The history of insulin is tightly bound to glucagon, another hormone peptide secreted by alpha cells in the pancreas, which has an opposite effect as it counter-regulates insulin and raises the levels of glucose in the blood.[19] Glucagon itself was isolated by Kimball and Murlin in 1922, while they were working on the isolation of insulin.[20] Glucagon is a 29-mer peptide, produced in the human body by proglucagon, which has a wide range of functions in health, and in particular is related to type I and type II diabetes. Like insulin, its amino acid sequence was clarified in the 1950s by Eli Lilly and its functions in the human body became more thoroughly understood in the 1970s. Its preeminent role in today's medicine is exemplified by its inclusion in the WHO's *Model List of Essential Medicines*. Its manufacture by rDNA technology was finally introduced in the early 1990s to supplant the isolation of the API from natural sources, both for ease of availability and to prevent contamination of the drug by infectious residues of animal origins. The first manufacturing route employed an engineered version of *S. cerevisiae*, providing a glucagon perfectly resembling that derived from animals.[21] The production process involves expression of the glucagon in a cultivation broth on agar, fed with yeast and glucose, followed by the recovery phase of the API from the fermentation broth: after completion of the productive phase, the pH of the both is increased to solubilize the target molecule, then the product is recovered by centrifugation to remove the yeast and other solid residues, and finally a purification step by reversed-phase high-performance liquid chromatography (RP-HPLC) is performed to achieve the required quality of the product. An alternative process was developed later, based on treatment on a cation-exchange resin of the supernatant obtained after the centrifugation step, followed by crystallization at the isoelectric point and a final RP-HPLC step to fix the quality of the API. The processes for the isolation of glucagon are complicated by its natural trend to aggregate and form fibrils; this event is apparently reversible in basic environments and therefore the manufacturing route involves steps of basic treatment of the stream to prevent any undesired phenomena.

GLP-1 (glucagon-like peptide 1) is a 37-amino acid peptide involved in glucose homeostasis and in the regulation of insulin and glucagon levels. Owing to its antidiabetic function, it is an important structure target for new APIs;

moreover, compared with other antidiabetes drugs, it has a much safer profile, with lower risks of hypoglycemia, and at the same time can have a body weight-lowering effect and a cardioprotective role.[22] However, the usefulness of GLP-1 is plagued by a very short half-life *in vivo* due to the proteolytic activity of the enzymes dipeptidyl peptidase-4 (DPP-4) and neutral endopeptidase (NEP 24.11); depending on the dosing method, its half-life ranges from 1.5 min to 1.5 h.[23] Interestingly, it seems that the cleavage by DPP-4 yields a GLP-1(9–36) product that acts as an antagonist of GLP-1 itself.

7.4 Liraglutide and Semaglutide

Native peptides are defined as peptides with the same amino acid sequence as in natural peptides, and they can either be produced by isolation from natural sources, prepared synthetically, or obtained by recombinant DNA techniques. They can exhibit distinct advantages over either small molecules or larger biologics: compared with small molecules, peptides exhibit higher target specificity and a better interaction with the receptors, and compared with larger biologics, such as proteins and antibodies, the cost of production is lower and they show less immunogenicity. On the other hand, peptides have weak points as drugs: they cannot easily permeate through the cell membranes and they are easily degraded *in vivo*, resulting in low stability when administered. These drawbacks result in the necessity to target mainly extracellular receptors and the need to adopt subcutaneous injection of the drug, because the oral bioavailability is very low.

This easy degradation of native peptides in the human body is a major hurdle and much research effort has been poured into the improvement of the pharmacokinetic properties of peptides. Selected engineering of the structure of the peptides, with insertion of non-natural amino acids and backbone modifications, is now widely employed to increase stability towards biodegradation and improve physical properties. The purpose is to open up new and novel interactions with target receptors. At the same time, developments in the formulation methodologies are aimed at increasing the solubility and bioavailability of peptides, so that oral and other routes of administration become feasible.[24] These modified peptides are defined as analogs, in contrast to native peptides.

Many efforts were then directed at modification of the backbone of GLP-1 and addition of side-chains to improve the physicochemical properties of peptides. Both liraglutide and semaglutide, among other APIs, were developed starting from GLP-1, with precise structural modification to reduce the metabolic degradation and improve pharmacokinetics.

Liraglutide closely resembles GLP-1, with 97% identity, so the modifications made to the peptide structure are minimal, which should guarantee retention of the activity of GLP-1. In more detail, liraglutide is the 7–37 sub-sequence of the original backbone of GLP-1; there is a single amino acid substitution compared with GLP-1(7–37). During the medicinal chemistry development, many amino acid substitutions and modifications to the

backbone were tested, and finally it was concluded that the most valuable ones, in terms of stability *versus* proteolytic degradation and activity *in vivo*, are the following:

- replacement of a lysine at position 34 with an arginine residue;
- acylation of lysine at position 26 with a fatty acid side-chain, linked to the ε-amino group through a glutamic acid spacer.[25]

These apparently simple, but crucial, modifications play a key role in the behavior of the API: the amino acid substitution prevents degradation *in vivo*, and the acylation increases the binding to the receptor. The outcome is that liraglutide has a half-life of about 8–10 h when administered *via* intravenous injection and up to 13–15 h for subcutaneous administration, making it viable for daily injection (see Figure 7.2).

The production of liraglutide has been patented with both SPPS[26] and fermentation processes, even though the manufacture of liraglutide for commercial purposes is carried out only by fermentation, for economic reasons. A caveat is necessary: the engineered organism used in the process can express the amino acid modification, as the arginine inserted is a natural amino acid, but cannot handle the insertion of an acylated lysine. The process therefore entails a hybrid approach: the peptide backbone is expressed by the natural organism, the peptide is isolated and purified as necessary, and at this point the peptide is modified chemically to introduce the modification of interest. In the production process for liraglutide, as patented by Novo Nordisk, this path is described as follows: the expression of the peptide backbone is carried out by properly engineered *S. cerevisiae*, then the peptide is purified and a chemical acylation takes place. More recent reports described the use of either *S. cerevisiae* or *E. coli*, both ending in a similar downstream process.[27]

The original report by Novo Nordisk very simply described a method for the preparation of acylated peptides: the fermentation broth is processed by initial purification by RP-HPLC, followed by precipitation of the compound

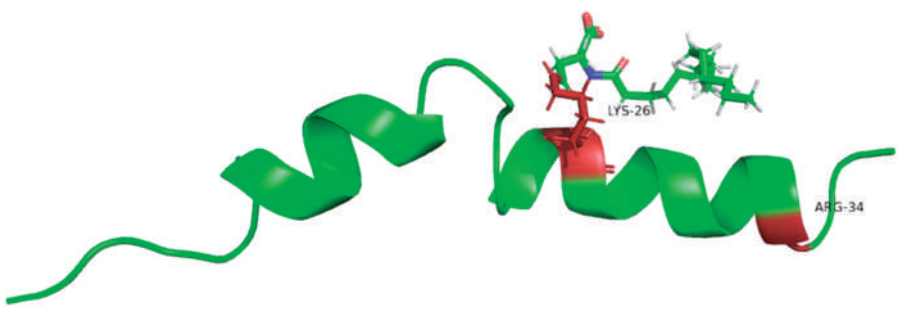


Figure 7.2 Secondary structure of liraglutide. Highlighted in red are the modified portions compared with GLP-1. 3D structure re-elaborated from Protein Data Bank, ref. 64.

at isoelectric pH (~5.4) and its isolation by centrifugation. Here the semisynthesis begins, with the necessity to acylate the lysine residue. Therefore, the isolated compound is at first treated with Et_3N to reach a basic pH of 11.6, then an active ester of the desired lipophilic side-chain carboxylic acid is added (palmityl–Glu–OH in the case of liraglutide preparation).[28] This acylation is the key step in the preparation: Lys26 has to be acylated with the required side-chain, while at the same time leaving unreacted the free amino group of the N-terminal histidine. Key to this selectivity is the pH of the reaction conditions: by operating in an aqueous environment, at a controlled pH of 10.5–11.5, there is a synthetically controlled chemoselectivity towards the primary amino group of the lysine. Indeed, under these conditions the free N-terminal amino group of histidine is less reactive, and the potentially interfering amidine group of arginine, which is protonated under these conditions, does not become acylated.

The final step of the semisynthesis is the removal of the protecting group used in the production process of the side-chain, needed to shield the carboxylic acid functional group on the glutamic acid spacer. Depending on the protecting group employed (usually a methoxy ester or a *tert*-butyl ester), a basic aqueous treatment or an acidic treatment with trifluoroacetic acid (TFA) is applied, and finally the target functionalized peptide is ready for a final purification. The final purification stage using RP-HPLC is crucial to attain a purity that meets the standard required by regulatory agencies; most often to achieve the target a C_{18} stationary phase is used in combination with acidic water and acetonitrile as the mobile phase. As alternatives, C_8- or even C_4-based-resins can be used, and ethanol and formic acid are common modifiers for eluting mixtures. To achieve a higher purity, in some processes even a double RP-HPLC purification has been applied. In cases such as liraglutide, where a semisynthetic approach is followed, the purification needs to address both process-related and product-related impurities. As one can imagine, very fine control of the conditions is required to achieve the target quality; therefore, during the R&D development phase, a crucial point is the determination of the structures of key impurities, whether they are critical from the purification point of view or not, and if they are critical then care must be taken to limit their formation by carefully tuning the process. All these evaluations can take up a considerable amount of the development time of the process and they are part of the control strategy of a process.

Further studies have been carried out to improve the manufacturing steps and achieve better processes, with the introduction of new methodologies. To improve expression and reduce the degradation by host enzymes, it was found that the inclusion in the expressed molecule of a tag or carrier sequence could help. New expression of fusion proteins, made up of the peptide backbone and a tailored tag or carrier, was developed and used in manufacturing processes. The most commonly used tags are polyhistidine tag (His-tag), ubiquitin or ubiquitin-like, and thioredoxin, and clearly the overall process needs to be redefined to account for their implementation. The peptide, now fused to the tag, is recovered from the broth in an appropriate manner,

depending on how the target is expressed (intra- or intercellular, related to the producing organism); then a bioconversion step is required to cleave off the tag and free the desired peptide. Subsequently, depending on the type of carrier or tag cleaved, a different purification step is required to upgrade the stream. Generally, this purification can be achieved using affinity chromatography and the tags used are tailored to be easy to purify. Even if more steps are added to the process, the advantages gained in terms of both ease of purification and product recovery offset the negative impact.[29] After purification of the peptide, the steps of the semisynthesis are performed on the new material. Once the complete structure of the API has been constructed, most often a final deprotection step is needed, as the free carboxylic acid of the glutamic acid contained in the side-chain is protected, by either a methyl or a *tert*-butyl ester. In the former case, basic aqueous treatment is sufficient to reveal the carboxylic moiety; in the latter case, an acidic treatment, usually with TFA, is required to reveal the final molecule.

Interesting new developments in the preparation of liraglutide include the following:

- the overexpression of liraglutide by concatamers by an appropriate gene in a host cell (*i.e. E. coli*), the expressed multimer then being digested by proteases and carboxypeptidases to provide the peptide;[30]
- the use of new fusion proteins, which permits faster, cheaper, and more environmentally sustainable processes.[31,32]

Notwithstanding the successful results obtained with liraglutide, new and more efficient APIs were investigated to improve the drug profiles further. The strategic pathway for the modification of the structure of GLP-1 was repurposed by Novo Nordisk. Research was carried out with the objective of increasing the affinity for the receptor, which is positively correlated with the extension of activity efficacy and half-life, while at the same time retaining the positive attributes developed with liraglutide.[33] After initial studies, it became clear that modification of the amino acid sequence, the side-chain, and the spacer of the side-chain itself from the peptidic backbone had the most important influences on the binding and potency of the newly developed compounds. Therefore, the minimum extent of modifications was carried out, so that the profile of the API would remain as similar as possible to that of GLP-1 and liraglutide and minimize the risk of an autoimmune response *in vivo*. What resulted was semaglutide, an antidiabetic peptide analog administered by injection; compared structurally with liraglutide it has a further amino acid substitution where Aib8, a non-natural amino acid, is inserted in the sequence in place of an Ala, a γ-Glu-2xOEG spacer is used in place of a simple glutamic acid, and lastly a C_{18} diacid is used rather than a shorter monoacid side-chain. With these modifications, the half-life of semaglutide in humans is increased to ~165 h, opening up the way for a weekly dose administration (see Figures 7.3 and 7.4).

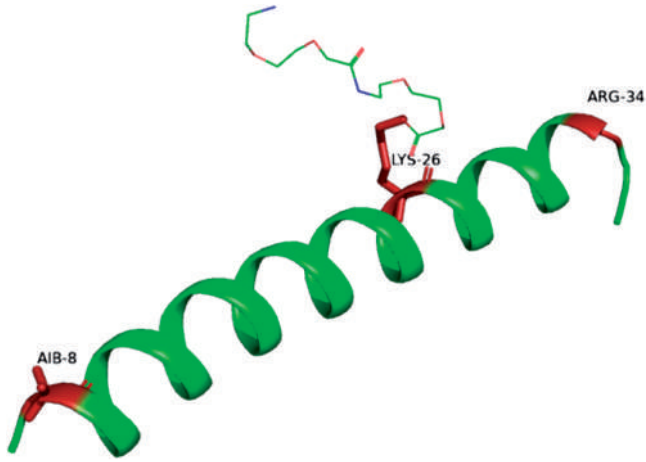


Figure 7.3 Secondary structure of semaglutide. Highlighted in red are the modified portions compared with GLP-1. 3D structure re-elaborated from Protein Data Bank, ref. 65 and 66.

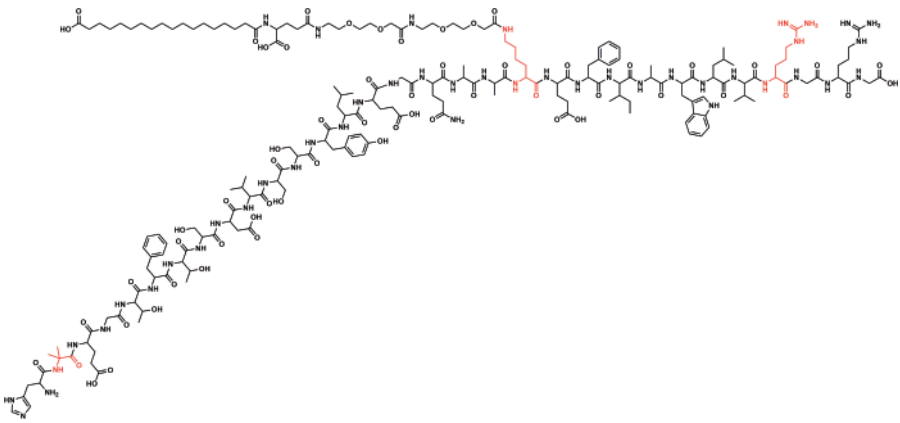


Figure 7.4 Primary structure of semaglutide. Highlighted in red are the modified portions compared with GLP-1.

The manufacturing process of semaglutide *via* semi-recombinant DNA preparation was patented by Novo Nordisk.[34] The process is similar to that for liraglutide, but in this case the insertion of a non-natural amino acid means that the full peptide backbone cannot be obtained *via* fermentation, but instead the semisynthetic approach this time needs to include a step of chain elongation to insert the missing amino acids. The manufacturing process begins with the upstream phase, with propagation of the expression plasmids in *E. coli*, then these plasmids are recovered and analyzed to confirm the correct inclusion of the proper sequence and finally transformed into

S. cerevisiae to produce the target peptide backbone. Two possible excisions of amino acids are reported: either two (His and Aib) or four (His, Aib, Glu, and Gly); the reasoning behind this choice is simple: activation of the C-terminal amino acids in these fragments harbors no risk of racemization of the amino acid involved (both Aib and Gly have no stereocenter in the α-position), thus preserving the optical purity of the target peptide. The expressed molecule is then recovered from the fermentation broth and purified as required; the final product of this phase is the selected shorter peptide, short by either two or four amino acids. At this point, a chemoselective acylation, on Lys26, in the presence of a free amino group at the N-terminus is applied, in a similar manner to that described for liraglutide. What is different with semaglutide is the nature of the side-chain that becomes attached to the lysine: in this case, a longer diacid is employed with a hydrophilic polyether amide spacer used to fuse the two parts of the molecule. The product can now be further processed to add the missing amino acids to the carbon backbone. The bis- or tetrapeptide (depending on the route chosen) is preactivated in the form of an active ester with standard coupling reagents and finally reacted with the acylated peptide. At this point, the synthesis is one step away from the target API; just a deprotection is required, as both the smaller peptide and the side-chain are added in the synthesis in a protected form, to avoid any formation of impurities due to cross-reactivity. Once the target structure has been revealed, a final RP-HPLC purification takes place with precise tuning of the conditions to achieve the required purity.

Further process developments now allow the use of a different culturing organism, either *E. coli* or *P. pastoris*, in addition to *S. cerevisiae*. This is obviously reflected in the requirement to adapt the upstream and downstream processes to isolate a peptide of adequate quality to be employed in the semisynthetic part of the process. The advantages of using fusion proteins or carriers to stabilize the molecule in the fermentation broth and simplify the purification were also introduced in the manufacturing of semaglutide.

Thanks to its very good pharmacokinetics and dynamics, more development work was carried out on the formulation of semaglutide: in addition to the classic injection, a different additive, SNAC {sodium 8-[(2-hydroxybenzoyl)amino]octanoate}, permits semaglutide to be formulated as tablets and be used in oral dosage form. More recently, even a route of administration *via* small rubber patches that adhere to the internal part of the cheek and release the drug has been found to be possible, opening up new frontiers for the delivery of peptides to patients.

As can be easily imagined, semaglutide is attracting considerable attention, and Novo Nordisk announced that, in addition to its antiobesity and diabetes regulation functions, semaglutide has been found to reduce the risk of cardiovascular diseases.[35] In addition, new limited reports have claimed the ability of semaglutide to act against drug and alcohol abuse, reducing dependence in patients under treatment. However, these reports have yet to find confirmation.

All this activity has accelerated the market demands and in turn the annual production of semaglutide is now over the metric ton level, and its manufacture must be able to withstand these rhythms without causing shortages and the production costs must be kept reasonable to allow patients to have access to the required medication. All these parameters need to be considered with the sustainability of the whole package in mind, from plasmid propagation to delivery to the patient, and the rDNA techniques are demonstrating themselves to be able to fulfill this role.

7.5 Recombinant Peptides and Protein Production: The Expression Systems

There is no universal expression platform for the production of recombinant peptides that is optimal for all of them. The choice of the host strain depends principally on protein structure, the presence of post-translational modifications, and the opportunity to have the product inside the cell or in the cultivation medium.

The fermentation yield is an important parameter in the evaluation of process costs and sustainability. The process mass intensity (PMI) depends on the production yield. The higher the productivity, the lower is the consumption of water, raw materials, and energy per kilogram of protein produced. The fermentation output depends on the cell density achieved and on the specific productivity of the strain.

Microorganisms with higher growth rates, such as *E. coli*, and that can reach a high cell density, such as *P. pastoris*, have clear economic and environmental advantages over less efficient expression systems.[36] A high cell density corresponds to high recombinant protein production.

The specific productivity depends on the genetic characteristics of the recombinant strain. Among the genetic characteristics, the presence of a strong inducible promoter in the expression cassette plays an important role. The expression pattern and efficiency rely on the selected promoter. If an inducible promoter is chosen, the growth and the production phase can be decoupled to reduce the metabolic load imposed by recombinant expression, while constitutive high-level expression frequently depletes the host cells.

An increase in yield through strain optimization is probably the most effective action to obtain environmental improvements.

7.5.1 *Escherichia coli*

E. coli, first isolated by Theodor Escherich in 1885, was the first system to be used for recombinant protein production. For some *E. coli* strains the complete genome sequence has been published, making its genetic manipulation easy.

The *E. coli* expression system is the most extensively used bacterial expression system; its transformation is simple and has high efficiency. In addition

to its easy manipulation, *E. coli* has the advantage of a high growth rate. The fermentation time is short and the energy consumption is very low. Using *E. coli* it is possible to produce huge amounts of proteins with an economical and low environmental impact process.

An essential component of *E. coli*-based expression systems is the vector used to transfer the genetic information inside the bacterial cell.[37] The vector replicates independently inside the cell and is transferred to daughter cells. Generally, multiple copies of the vector are produced containing the rDNA. The copy number is controlled by the replicon (origin of replication) and varies from one vector to another. The use of a high copy number plasmid, such as the plasmid cloning vector pUC, does not always mean higher efficiency of recombinant protein production because the large amount of plasmid affects the bacterial growth rate. The high metabolic burden increases cellular stress level and can lead to plasmid rejection. The cells, in fact, eliminate the plasmid DNA to decrease the metabolic burden.[38,39] Over the course of cultivation, the number of bacterial cells bearing plasmids can decrease. The plasmid free cells have a higher growth rate because they do not have to replicate the plasmid DNA or produce the recombinant protein and consume the carbon source present in the cultivation medium, resulting in a limitation of carbon availability for the recombinant cells and consequently low recombinant protein production.

In order to avoid plasmid loss, a marker selection gene is usually present in the vector. The marker could be an antibiotic resistance gene or a metabolic selection marker. In the case of the presence on the plasmid of an antibiotic resistance gene, addition of antibiotic to the culture medium is required to maintain a high selectivity pressure. In this case, the recombinant protein purification process has to prove complete elimination of the antibiotic using sensitive analytical methods.

The metabolic selection marker represents a better choice. This system requires the use of a host strain that is lacking an essential gene, which was deleted from the genome and placed on the plasmid. Only the recombinant microorganisms will be able to grow.

The gene encoding the recombinant protein is cloned into the multiple cloning site (MCS) of the vector under the control of a promoter that will regulate the expression of the gene. The promoter controls the binding of RNA polymerase to DNA. The use of a strong promoter, with a high affinity for RNA polymerase, is essential to obtain an increase in the transcription of the heterologous gene. The T7 promoter system present in the pET plasmid cloning vector is the preferred one. This system requires the presence of the T7 RNA polymerase-encoding DE3 phage fragment, integrated in the bacterial chromosome, and under the control of lacUV promoter. The lacUV promoter is regulated by a chemical inducer, isopropyl-β-D-thiogalactopyranoside (IPTG), which activates transcription of T7 RNA polymerase when added to the culture. T7 RNA polymerase is faster than native *E. coli* enzyme and transcribes the recombinant gene under the control of T7 promoter that is present on the plasmid.[38–40]

The use of a strong promoter and a system with a high transcription efficiency is a key factor to achieve a high yield of the fermentation process. The recombinant protein production in the inducible expression system can be controlled and separated from the cell growth phase. Normally there is a high specific growth rate initially to support biomass formation and then there is a decrease in specific growth rate in the production stage to achieve maximum productivity.[41]

The direct expression of small recombinant peptides is very difficult, because they are easily degraded by the proteases present in the host cell. In addition, small peptides are not easily isolated and separated from the host proteins. In order to circumvent these technical difficulties, other essential elements are included in the expression cassette, namely the protein tags. There are different types of tags that have various functions. The protein tag can be formed by a short amino acids tail or by a protein. They can help the purification of the product, protect it from enzymatic degradation, or modify its solubility properties.

The affinity tags, generally constituted by a histidine tail, are used for the purification of the peptide drug by affinity chromatography. His-tagged proteins can be recovered by immobilized metal ion affinity chromatography using nickel or cobalt ion-loaded nitrilotriacetic acid–agarose resins. The host proteins are easily eliminated in the flow through. In other cases, the tag is a carrier protein that can increase or decrease the solubility of the recombinant product. The variation in solubility can give an advantage in the purification of the peptide or protect it from the protease enzymes present in the cytoplasm.

Heterologous proteins generally accumulate in the cytoplasm as inclusion bodies, which comprise insoluble misfolded aggregates of proteins. The use of carrier proteins that enhance the solubility of the recombinant protein can avoid the formation of inclusion bodies. In the case of peptides that are sensitive to the proteolytic action of the cytoplasmic enzymes, the formation of inclusion bodies may represent a form of protection, and the use of a different carrier protein, that reduces solubility, can induce recombinant protein precipitation as inclusion bodies.

If co-expression of a protein tag is introduced, a single chimeric polypeptide chain will be produced from the cell and an enzymatic or chemical cleavage will be necessary to release the protein drug. In the case of enzymatic removal, a protease cleavage site will be introduced in the expression cassette between the protein tag and recombinant drug. The correct removal of the protein tag is a very important step and high-fidelity endoprotease enzymes are used to guarantee the safety and efficacy of the protein therapy. The effect of an imprecise enzymatic cut is a product having one or more additional amino acids or a deletion or even fragmentation of the protein. For this reason, the choice of the endoprotease enzyme has to be carefully assessed for each single recombinant product, comparing the amino acid sequence of the protein expressed and the enzymatic cleavage site sequence, to ensure that only the cleaving site can be recognized by the endoprotease.

The introduction of a specific genetic signal in the expression cassette can lead the recombinant protein into the periplasm region and allow the formation of disulfide bonds. The periplasm provides an oxidizing environment and helps in forming disulfide bonds, which facilitate the proper folding of recombinant proteins. This occurs when the disulfide bond formations between cysteine amino acids are present in the peptide chain. The protein is produced in a soluble form even if at a relatively low concentration.

E. coli is usually preferred because it offers several advantages, including high growth rate, simple media requirement, ease of handling, high yield, and very cost-effective large-scale production. However, there are also some disadvantages, such as improper eukaryotic protein refolding and low expression level, lack of post-translational modification (glycosylation), endotoxin contamination, and proteolytic digestion of the product. Several genetic modifications of *E. coli* strains have been introduced in order to overcome these problems and some examples are presented below.

N-Linked glycosylation is the most common post-translational modification of proteins in eukaryotes. It has been discovered that the bacterium *Campylobacter jejuni* possesses the capability to glycosylate proteins and it was also shown that a functionally active *N*-glycosylation pathway could be transferred to *E. coli*, obtaining protein glycosylation.[42,43]

The codon usage of the heterologous protein also plays a major role in determining the expression level of recombinant proteins. If the codon usage of the heterologous protein differs significantly from the average codon usage of the *E. coli* host, it could result in very low expression. Usually, the frequency of the codon usage reflects the abundance of their corresponding tRNA.

Expression of heterologous proteins in *E. coli* can be improved by replacing codons that are rarely found in highly expressed *E. coli* genes with more favorable major codons. Similarly, co-expression of the genes encoding for a number of the tRNA for rare codons may enhance the expression of heterologous proteins in *E. coli*.

The use of protease-deficient *E. coli* strains, which carry mutations that eliminate the production of proteases, may also improve the yield of recombinant proteins by reducing proteolytic degradation. Undesired proteolysis, in fact, may lower the product yield or affect biological activity, but also complicates downstream processing as the degradation products will have similar physicochemical properties and will be difficult to eliminate.

The recombinant peptides and proteins are produced and accumulated in the cytoplasm of *E. coli* and their recovery requires cell disruption with release of the lipopolysaccharides, generally known as endotoxins, that compose the bacterial outer membrane. These components are a major concern since they are pyrogenic in humans and can induce immune responses and the production of antidrug antibodies that can reduce the therapeutic efficacy of the recombinant peptide. In the biopharmaceuticals produced using this host, additional purification steps are introduced in the downstream process to guarantee their elimination.[44]

7.5.2 Yeast Expression Systems

Yeast expression systems are used when secretion of the recombinant protein is required. In yeast cells, the compartmentalization into endoplasmic reticulum (ER) and Golgi guarantees the correct folding of the protein along the secretory pathway, allowing the secretion of the recombinant protein in soluble form. The presence of the product in soluble form in the medium helps its recovery and purification, improving the economics of the production process. The natural secretory machinery may not be sufficient under the conditions of high recombinant protein production and might lead to protein degradation and reduction of protein secretion. In this case, metabolic engineering of the secretory pathway is recommended in order to obtain high protein yields.[45] Among yeast strains, *S. cerevisiae* and *P. pastoris* are very commonly used for the production of heterologous proteins.

Yeast is a preferred host for the expression of various proteins that require post-translational modifications for their biological activity. Yeast cells have the ability to carry out numerous post-translational modifications such as phosphorylation, *O*-linked glycosylation, *N*-linked glycosylation, acetylation, and acylation. Recombinant proteins are expressed in yeast in properly folded and functionally active form. Production of biopharmaceuticals using a yeast expression system is also very cost-effective.

However, one major concern with producing therapeutic glycoproteins for human applications is that yeast *N*-glycosylation is of the high-mannose type, which confers a short half-life *in vivo* and hyper-immunogenicity and thus renders the therapeutic glycoprotein less effective. The initial glycosylation that occurs in the ER is the same in yeast and humans, but completely different at the Golgi level where, in humans, the mannosylated proteins are converted to galactosylated and sialylated forms. For this reason, various human therapeutic proteins, including therapeutic monoclonal antibodies, are being produced in mammalian cell lines. Recombinant proteins expressed in mammalian cells are properly folded, glycosylated, and generally produced as functionally active proteins. However, the production cost of biopharmaceuticals using a mammalian expression system is very high and other concerns related to viral contamination and oncogene transmission have to be considered.

Recently, using genetic engineering techniques, some mutant yeasts with deletion of the natural glycosylation pathway and having a humanized glycosylation pattern were produced and used for the production of erythropoietin and antibodies.[40,45] The transformation with expression vectors containing the gene for the recombinant protein production has a high efficiency and, depending on the type of plasmid used, can lead to the integration of the gene in the yeast chromosome or to an autonomously replicating vector. The integration of the gene in the chromosome confers stability to the population without selective pressure. *S. cerevisiae* represents a specific platform for recombinant insulin, insulin analogs,

and GLP-1 agonists used by Novo Nordisk to develop a Crabtree negative and auxotrophic strain with improved secretory function.[7,46] Constitutive glycolytic promoters can be present in the vector [ADH1 (alcohol dehydrogenase), PGK (phosphoglycerate kinase), GAP (GTPase-accelerating protein)], but inducible promoters are preferred because the strong constitutive expression influences the growth rate negatively. The strongest regulated promoters are induced by galactose (Gal) and strongly repressed by glucose.[44] This type of yeast also has some disadvantages because it cannot reach a high cell density, the secretion of proteins is generally poor compared with other yeasts, the tendency to produce ethanol (Crabtree effect) may reduce the fermentation yield, and its glycosylation produces hyper-mannosylated glycans.[40]

An alternative yeast strain used for the large-scale production of biopharmaceuticals is *Pichia pastoris* (*Komagataella phaffii*). *P. pastoris* has two alcohol oxidase enzymes, AOX1 and AOX2, that are transcripts in the presence of methanol and allow the growth of yeast on methanol as the sole carbon and energy source. There are mutant strains defined as Muts with only AOX2 gene and a lower efficiency of growth on methanol and others (Mut–) without any AOX gene.[47] The first application of *P. pastoris* dates back 50 years to the commercial production of single-cell proteins as animal feed additives.[48] This process, owing to an increase in the price of methanol, was abandoned, but the same fermentation method was subsequently used for recombinant protein production. The genomic sequence of *P. pastoris* was published in 2009 and favored the development of biopharmaceuticals using *P. pastoris* as a host.[49] Using genetic engineering, the strain was modified and several customized improved strains were produced. *P. pastoris* has the ability to reach high cell densities with a very high level of expression of heterologous proteins by its robust and tightly regulated methanol-inducible alcohol oxidase 1 (AOX1) promoter and has the advantage of the glycosylation of secreted proteins because it does not hyper-glycosylate the heterologous proteins, so glycoproteins produced in *P. pastoris* may be more suitable for therapeutic use in humans.[50]

P. pastoris can therefore be an attractive alternative for the large-scale production of recombinant insulin and insulin analogs.[6] Heterologous proteins can be expressed at high levels and efficiently secreted. Purification is relatively straightforward owing to the small amount of host proteins. For this reason, the number of recombinant product applications in *P. pastoris* has been increasing over time.

The AOX1 promoter (derived from *P. pastoris* alcohol oxidase 1 gene) is induced on the addition of methanol. The induction system is very efficient, but it has a drawback in scale-up compared with a standard fermentation plant, because of the need for plant adaptation in order to manage the use of an inflammable solvent. Alternative promoters that do not require methanol for induction are also available, such as GAPDH (glyceraldehyde-3-phosphate dehydrogenase), for constitutive expression of recombinant proteins and other methanol-free expression systems.[51]

7.6 Fermentation Process and Sustainability

7.6.1 Sustainability Assessment

The very first concept of "green chemistry" can be traced back to the Pollution Prevention Act of 1990 in the USA, which introduced antipollution generation as a national policy.[52] The term itself was coined in the early 1990s by the staff of the Office of Pollution Prevention and Toxins at the US Environmental Protection Agency (EPA).[53] Many other milestones in the history of green chemistry contributed to its development and diffusion, but the definition of the "12 principles of green chemistry" by Paul Anastas and John Warner in 1998 is considered to be by far the most important one (see Table 7.1).[54] In fact, they were able to define the boundaries of the green chemistry concept, and their work continues to serve as the foundation of the field today.

To really determine if a product, process, or technology could be considered "green", most or all the "12 principles of green chemistry" need to be intrinsically present. Nonetheless, these principles were designed to be the broadest possible, and not all of them could be applied to a specific topic. In the context of peptides production, the 2nd and 10th principles cannot be fully applied, since they refer, at least partially, to the product and not the process. On the other hand, the 9th and 11th principles are specific to industrial plant and engineering technologies, and are outside the scope of a treatise on a chemical topic such as that presented in this chapter. Therefore, we focused our attention on the other principles to assess if the production of peptides by microbial fermentation could be considered a "green" technology. Moreover, we did not consider purification, since it is covered by other technologies presented in this book (*e.g.* SPPS), and its impact is often related to the specific product and not the technology.

Microbial fermentation is an intrinsically safe technology: usually the only solvent employed is water, which is non-toxic and non-flammable, and the temperatures involved are usually low (<50 °C). Moreover, most of the reagents employed are safe, since all we need to do to obtain the desired peptide or nucleotide is to feed energy in the form of sugars to the microbe.

Table 7.1 The 12 principles of green chemistry.[54]

1	Prevent waste
2	Design safer chemicals and products
3	Design less hazardous chemical syntheses
4	Use renewable feedstocks
5	Use catalysts, not stoichiometric reagents
6	Avoid chemical derivatives
7	Maximize atom economy
8	Use safer solvents and reaction conditions
9	Increase energy efficiency
10	Design chemicals and products to degrade after use
11	Analyze in real time to prevent pollution
12	Minimize the potential for accidents

Therefore, we can conclude that microbial fermentation is fully in compliance with the 2nd, 3rd, 8th, and 12th principles of green chemistry.

Almost no fossil-based materials are employed in a microbial fermentation process. In fact, both the microbe and the reagents present in such a process are completely renewable, and water is basically the only solvent needed during the fermentation. The 4th principle perfectly fits the microbial fermentation technology.

The synthesis of peptides is performed by a microbe through the use of specific enzymes, which are the "catalysts of nature". Even if reagents are fed (*e.g.* sugar) to the microbe, they are not considered stoichiometric. Moreover, no intermediate or protection/deprotection steps are needed, since enzymes are extremely selective catalysts. Therefore, the synthesis performed by the microbe is perfectly in line with both the 5th and 6th principles.

Even though it is not trivial to identify a specific chemical transformation in a microbial fermentation process, since the chemical bonds are formed through catalysts (enzymes) and without protection/deprotection steps, its atom economy should be the highest possible for these transformations. For these reasons, we can conclude that the atom economy is maximized if a microbial fermentation process is employed for the synthesis of peptides and nucleotides (7th principle).

Often, biotransformation processes have been considered environmentally friendly, but in many cases these conclusions are not sustained by real metrics.[55] In particular, aqueous-based processes with a relatively low concentration of product generate large amounts of wastes. Even though the reduction of generated wastes should always be the driver to a greener process, peptide microbial fermentation consists mainly of water and not toxic raw materials, as illustrated in Figure 7.5. After the inactivation of the microbe and degradation of the residual peptide, the waste could be directly released into the environment. Therefore, we could even argue whether such processes actually generate wastes (1st principle).

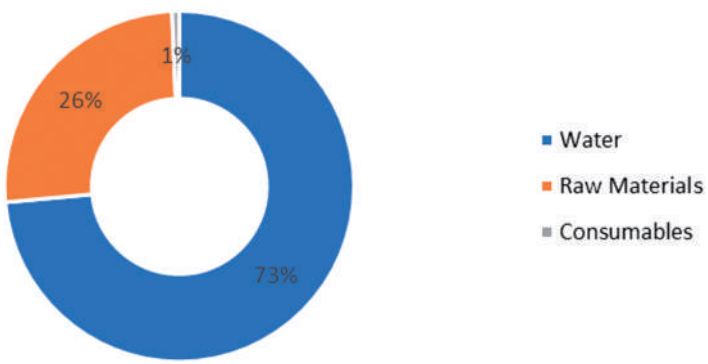


Figure 7.5 Composition of process mass intensity (PMI) from microbial fermentation industrial processes.

Finally, even if the 10th principle mainly refers to the product, most of the materials employed in a microbial fermentation process are biodegradable by definition.

Hence most of the 12 principles of green chemistry are respected in microbial fermentation for the production of peptides and nucleotides, and we can safely conclude that this technology can be considered green.

However, all that glitters is not gold, and much more work needs to be done to improve the greenness of this technology further.

7.6.2 Further Improvements of Sustainability

In order to design a sustainable process, an increase in the amount of output product per unit input of ingredients and energy is mandatory. Mutation of the producing host to create efficient producing strains can lead to a significant improvement in process efficiency and a reduction of environmental impacts. The microorganism is not the only key point of the development activity. Optimization of the cultivation conditions by optimizing the media, temperature, pH, inoculum, and production kinetics is important to achieve the maximum potential of the strain in terms of growth and production of recombinant proteins. The raw material consumed in the largest quantity in fermentation is water, which is used for media preparation, for equipment cleaning, and for cooling during fermentation. The reuse or efficient use of water is of high priority in the scale-up of a fermentation process. The cultivation media are composed of carbon and nitrogen sources and a salt solution containing all the essential elements and some vitamins. The constituents of the media must satisfy the elemental requirements for cell biomass and recombinant protein production giving an adequate supply of energy for biosynthesis and cell maintenance. The selection of appropriate carbon and nitrogen sources and optimization of their concentrations in the media are essential to obtain a high yield and might have a direct impact on process sustainability. Substrates produced using energy- and chemical-intensive processes play a negative role. In contrast, many substrates (starch, proteins, sugar) that are co-produced with other agricultural products and represent a waste of the primary production process have a low environmental impact. For example, glycerol is a cheap carbon source that can be an alternative to glucose.

The carbon source is the energy source of the microorganism. Rapidly metabolized carbon sources lead to a high oxygen demand and high power absorption. Oxygen is a key substrate of the fermentation. If there is a risk of oxygen limitation due to fast metabolism, it may be overcome by decreasing the concentration of the carbon source in the batch medium and adding additional quantities of the same as a continuous or semicontinuous feed during fermentation, or otherwise changing the carbon source with a complex raw material (starch or dextrin) in order to slow the metabolism.

The fermentation kinetics have a great impact on the final product yield and consequently also on process sustainability. Not only the quantity but

also the quality of the recombinant protein could vary depending on the fermentation model implemented. There are several fermentation models that can be adopted, outlined below.

Feed-batch fermentation allows good control of the carbon source availability, which is very important to avoid a shift of metabolism and the production of unwanted by-products. This strategy combines the possibility of separating the growth and production stages, ensuring good cellular activity by means of controlled nutrient utilization.

Typically, in inducible recombinant protein expression systems, there is a high specific growth rate phase, initially, to obtain a large amount of biomass, and a production phase in which the growth rate is decreased to drive the carbon into recombinant protein production. The uncoupling of growth and protein expression can be advantageous for both biomass formation and protein yield.

The reduction of the kinetics during the production phase can improve protein folding and facilitate its exportation outside the cell in yeast.[41]

The feeding profile can vary from linear to exponential, depending on the type of strain and promoter utilized and also on the fermentation phase considered. The use of a defined feeding of substrate allows an ideal growth rate to be maintained, preventing both over-nutrition and cell starvation.

Models of continuous feeding of media with intermittent drawing of broth can also be adopted. Novo Nordisk developed a continuous fermentation process for the production of insulin and GLP-1 analogs.[46] Nevertheless, continuous fermentation is generally considered risky at the industrial scale. In fact, contamination, strain mutations, and product instability may reduce the fermentation yield significantly.

High cell density cultures are a preferred strategy for optimizing recombinant protein volumetric productivity, which is a key parameter of bioprocess cost-effectiveness. It is important to consider that protein production may be negatively affected by a high growth rate, when defining the fermentation model. Impaired folding and subsequent degradation of the product might occur. Although the preferred growth temperature of the host is 28–30 °C, lowering the temperature to 20–25 °C during the induction phase could help to reduce the growth rate and to obtain the correct protein folding and decrease the proteolytic activities responsible for recombinant protein degradation.

An improvement in recombinant protein production may be obtained by reducing the metabolic load. The metabolic load corresponds to the proportion of resources devoted to the basic cellular metabolism (cell maintenance) and dedicated to recombinant gene maintenance and expression. The metabolic load increases with increase in the genetic load, gene size or gene copy number, and the expression level. An increase in the gene copy number may not result in increased recombinant protein production if an overload problem has been determined. The effect produced by the metabolic load is a decrease in the growth rate of the host and mistranslation of the recombinant protein. To reduce the metabolic load, integration of the

recombinant gene in the chromosome might be helpful, because maintenance of plasmids inside the cell requires a large amount of energy. Optimization of the formula of the medium to guarantee the availability of amino acids for recombinant protein production is also an essential factor.[56]

Contributions to global warming are to a large extent driven by energy consumption.[5] Fermentations require steam sterilization of the equipment and stirring of the media during cultivation, which has a certain energy consumption.[3] Energy consumption is an important aspect of the fermentation process and its minimization is a crucial factor to improve sustainability. Host strains with high growth rates, such as *E. coli*, represent an advantage in terms of energy depletion during cultivation because their fermentation cycle time is short. An *E. coli* fermentation usually lasts 24–48 h, whereas a yeast fermentation normally takes 5–7 days. Even though the peak of energy absorption may be higher, a shortened fermentation time has a positive effect on the total energy consumption.

To guarantee good oxygen transfer in high cell density fermentations, a higher stirring speed is applied and an increase in power absorption is observed. Nevertheless, the increase in the process yield can result in an improvement in PMI and lead to an overall positive life-cycle assessment.

The power consumption of the fermentation is not only related to the fermentation time. A decrease in energy consumption during cultivation can be obtained also by choosing efficient mixing equipment in the bioreactor. There are innovative impeller types that were designed to improve mixing and allow the conduction of the fermentation at a lower stirring speed compared with classical Rushton turbines. The result is a decrease in power absorption.

7.7 Scale-up of a Recombinant Product

To ensure the reliability of the scale-up, it is important to adopt a suitable scale-up criterion. A constant oxygen mass transfer coefficient (k_La), constant specific power input (P/V), constant impeller tip speed, and constant dissolved oxygen concentration are four commonly used scale-up criteria used in the fermentation industry.[57]

One of the major substrates of the fermentation is oxygen, which is provided by aeration. For optimal oxygen transfer, the air flow and agitation are adjusted automatically to maintain a minimum dissolved oxygen set point, which is usually fixed at 20%. Oxygen transfer capacity coupled with excellent cooling facilities is crucial to achieve the maximum potential of fermentation. As the culture scale increases, it becomes more difficult to ensure dissolved oxygen availability throughout the culture and oxygen often becomes the limiting nutrient for the fermentation process. A constant k_La (volumetric oxygen transfer coefficient) is the best option for scale-up.[58] A minimum dissolved oxygen set point can be considered as an alternative to k_La.

The internal asset of the fermenter and the type and dimensions of the impellers and their reciprocal distances are key factors. To guarantee

maximum success in scaling up, it is important to define a good scale-down model to be used in the development of the fermentation process. If methanol is used in the fermentation, specific safety plant adaptations have to be introduced to minimize the fire risk.

In general, *E. coli* fermentation can be easily scaled up and the risk of failure is very low. Using this bacterial strain, huge amounts of recombinant protein can be obtained in a short time at a low cost with an environmentally sustainable process. The modulation of the growth rate during production is the main key point. If this aspect is not controlled, the risk of plasmid loss or formation of inclusion bodies would be very high. Genetic stability is higher if yeast is selected as a host, because the recombinant gene is generally integrated in the chromosome. Eukaryotic cells require a fermentation model of greater complexity and consequently the scale-up is more tricky. Oxygen limitation could have a severe effect on the fermentation yield. Ethanol formation in *S. cerevisiae* could completely compromise fermentation.

Strict control of the feed rates and the study of the optimal feed profile are mandatory in yeast fermentations and the fermentation kinetics have to be well defined to guarantee the secretion of a properly folded product. Once the upstream phase is complete, where the purpose is to have the expressing organism to produce the target molecule at the highest possible concentration, the ensuing part is the downstream process. Here the target molecule must be isolated and purified, to meet the required quality criteria for the drug substance. This is achieved in a series of steps, each one increasing the stream quality from the previous step, in a continuous enhancement from the broth to the API. The first goal of the downstream process is to have the peptide in aqueous solution or in the supernatant phase of the broth. This is more easily achieved when yeasts (*i.e. S. cerevisiae* or *P. pastoris*) are employed to express the API, as these organisms generally are able to produce the desired molecule in an extracellular fashion and the target compound is already in the liquid mixture. In contrast, when employing other types of bacteria that express the target peptide in an intramolecular fashion (*i.e. E. coli*), a cell disruption step is required. In any case, most often a volume reduction is useful before any other operation; centrifugation or membrane filtration are common operations. After this, the batch volume is reduced, making the process less expensive and less time consuming, at the same time some of the most dissimilar components of the broth are removed from the broth itself. In the case that a lysis is necessary, two main types of cell disruption are available to the process chemist to perform the operation: chemical lysis and mechanical lysis. Based on several factors, such as stability of the cell, economic considerations, and technological feasibility, many options can be considered. Chemical lysis can be brought about by (among others) osmotic or temperature shocks, enzyme digestions, or lipid membrane dissolution. Mechanical lysis, on the other hand, can be realized by homogenization, grinding, or sonication, all methods that are able to rupture the cell membrane and release the peptide in solution.

Once the target peptide is in solution, the first stage of purification starts, namely the removal of destabilizing or degrading impurities. As one can imagine, the broth as is, or the lysate broth in the case of lysis, contains many undesired or even deleterious contaminants, such as pyrogens, viruses, proteases, and other high molecular weight protein-based materials. Once again, a few process options are available: filtrations are common choices, and also capture and release chromatography techniques. Filtrations are often employed to improve the quality of the mixture under cheap conditions: with proper choice of the filtration media, one can remove solid particles from a complex mixture, or even separate soluble compounds. In addition to conventional filtrations, which are based on flowing the mixture in a perpendicular direction through a filtration medium to build up a cake of filtered material, tangential flow filtration (TFF) (sometimes referred to as membrane filtration) is commonly performed as a bioseparation step. In TFF, the flow of the mixture is parallel to the filtration membrane, running over its surface at a defined flow-rate. By regulating the operating parameters, the process developer can manipulate the system and apply an apparent pressure (defined as transmembrane pressure) that forces the mixture onto the membrane surface and, consequently, the filtered solution flows through it (see Figure 7.6). The main advantage of this technique is the avoidance of the buildup of a cake on the filtration medium, as the continuous flow removes any deposited particles from the surface and recycles them back to the feed tank; the result is a steady filtration flow-rate.[59] Membranes range greatly in size, type of filtration material, porosity, and physical disposition, giving the process developer a vast range of choices to select the most suitable one for the process at hand. In many cases when dealing with complex and inherently inhomogeneous mixture such as broths, a flocculation or sedimentation operation can help to keep the phase efficient.

At this stage of the downstream process, the target is to remove most of the compounds with high molecular weight and improve the purity of the stream. Microfiltration TFF techniques are employed, where filtration

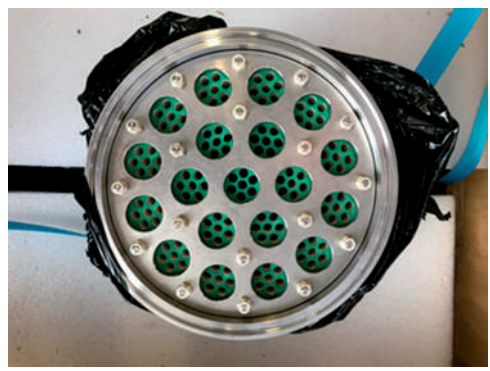


Figure 7.6 Head-on view of the interior of a microfiltration membrane.

media ranging from 0.1 to 0.45 μm allow the efficient filtration of smaller compounds (which generally are the target product) away from the bulk of contaminants. The recovered permeate from this phase is a mostly clear solution, which often needs to be concentrated to a more reasonable volume to achieve process efficiency. Once again TFF techniques are employed in this phase, where smaller pore membranes are used to retain this time the target compound in the retentate solution (see Figure 7.7).

Another technique commonly applied to upgrade the quality of the stream is capture and release chromatography (also called affinity chromatography). In this purification step, it is possible to achieve both a volume reduction of the batch and removal of bulk impurities, such as endotoxins and DNA-related



Figure 7.7 Microfiltration apparatus: at (a) industrial scale, and (b) pilot scale.

residues, among others. The methodology is based on a specific binding between the target molecule and the adsorbent, which is functionalized with specific functionalities. Therefore, the starting solution is loaded into the system (usually a packed column), then the system undergoes a wash phase, in which all the unbound product is physically eluted from the column, and finally an eluent able to displace the bound product from the resin is introduced and the eluates can be selectively collected to obtain a solution with highly improved quality and reduced volumes. Significantly, high yields and material recovery are easily achievable with capture and release purification. There are several possible interactions and bindings that can be exploited to carry out an affinity purification on proteins: enzyme–ligand type of interactions, receptor–agonist systems, and also metal–ligand complexations. The last system is denoted immobilized metal affinity chromatography (IMAC) and exploits the ability of some amino acids, the most utilized being histidine, to chelate transition metals such as nickel or copper (see Figure 7.8).

These methods of stream upgrading can be so efficient that target peptides sometimes are strategically expressed as a fusion protein, including a tail or 'tag' specially tailored to have a strong affinity with the selected adsorbent and allow for a clean capture stage. Obviously, the ease of purification is not the only advantage of expressing a fusion protein, it also comes with

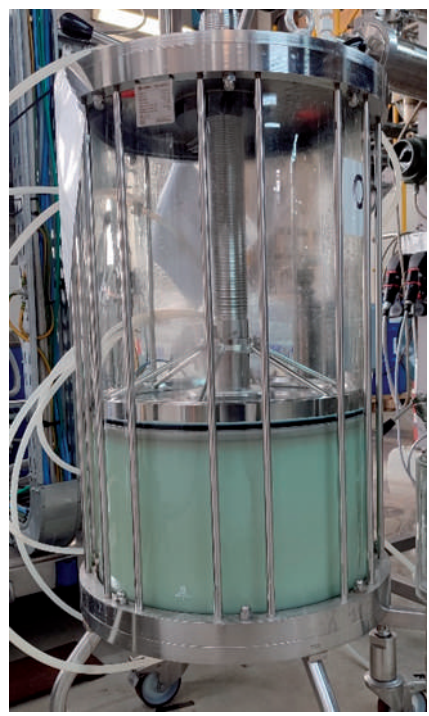


Figure 7.8 Pilot-scale IMAC column.

increased stability of the fusion protein itself in the fermentation broth and in the lysis step. Substantial research has been reported on this topic over the years, and its use is a relatively standard option in the arsenal of a process developer. The down side of this strategy is plainly the need to excise the 'extra' part of the molecule, but, again, the tag is generally engineered in such a manner that the removal is fairly straightforward and does not impair the process. Very defined amino acid sequences are inserted where required, and these sequences are the target for a protease-type enzyme that performs a clean hydrolysis of the peptidic bond, finally releasing the peptide in its correct primary structure and removing the fusion tag.

In most processes, an intermediate purification step is required before moving to the final stages of API manufacture. Often an ion-exchange resin is used, but affinity chromatography or size-exclusion chromatography can also prove useful. This stage of purification is aimed at removing the last remnants of host cell proteins and process-related impurities and to achieve virus clearance. In cases where a bioconversion was performed, the added enzyme and the excised fusion tag also need to be removed.

Finally, the stage is set to perform a fine chromatographic purification (see Figure 7.9), the aim being to remove the product-related impurities and achieve the quality target profile. Broadly, two types of purification are used: low–medium-pressure chromatography and high-pressure chromatography, both of which are based on hydrophobic interactions working in reverse phase. Low-pressure purification is the most appealing from the process point of view, owing to both the easier operation and the lower overall cost



Figure 7.9 Example of an industrial chromatographic system.

of operation. In low-pressure purifications, several resins with various particle sizes and pore distributions can be used, whereas high-pressure purifications generally rely on a silica matrix functionalized with alkyl chains (C_4, C_8, and C_{18} are the most common).

Following the purification step, one can proceed to isolate the API. In bioprocesses the target peptide is a relatively sensitive product, and heat and physical shocks can be critical, hence normal drying in a standard heating apparatus can be deleterious to the quality of the product. Freeze-drying is applied instead: the high-purity eluates obtained from the purification stage are frozen at cryogenic temperatures and then lyophilized. A treatment of the eluted solution is generally necessary before freeze-drying, as organic solvents have to be removed, in addition to pH corrections and eventually addition of stabilizers. When complete, the downstream process results in the isolation of an amorphous powder of pure peptide for commercialization.

The ultimate goal of the development phase is to scale up the process from the laboratory scale to the pilot scale or even directly to the plant scale. When doing this, additional considerations must be taken into account. The product quality is clearly the main factor, but other parameters also come into play: economics and instrumental availability are critical to a successful process definition. In light of this, volume reduction and time efficiency play a crucial role in the process development. Moreover, it is important to match upstream and downstream phases of the same process: the timelines need to align to achieve the required efficiency. A single fermentation broth can sustain more than one downstream process; at the same time, the completion of a downstream process can require less time than an upstream process. To avoid any plant downtime, prior planning is critical.

When developing a new process in R&D, correct downscaling is important: knowledge and experience of the plant are essential and the involvement of all the necessary personnel can help to anticipate many issues. The time scale to consider in plant operations is on a different level compared with the laboratory, and product stability must be correctly assessed to avoid incurring any unexpected surprises. Most of the instrumentation used in plants can be correctly reproduced in the laboratory, and a robust process design considers the scalability of all parameters, such as longer operating times and more complex unit operations. This implies a thorough knowledge of product stability in any phase of the process, both chemical stability, to control the impurity profile, and physical stability, to prevent undesired phenomena such as precipitation, aggregation, or clogging, or even more subtle fibrillation, all of which cause losses of material. Many other parameters can change dramatically with an increase in batch size during the scale-up phase, and all of them must be kept under control, *e.g.* pressure, local concentration inhomogeneity, or pH variations.

A powerful tool in a process chemist's hands is the design of experiment (DoE), a tool highly supported by the "quality by design" (QbD) concept expressed by International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) guidelines (ICH Q8).[60] By

applying this statistical instrument one can define an area of feasibility of the process with focused experiments; this results in a much safer scalability of the process itself: rather than a working on a defined value per parameter, a wider working area is defined as safe. This in turn allows one to modify, and most importantly retain total control over, any parameter during the scale-up without concerns for the product quality or yield. This is a "win–win" situation where the process guarantees the output quality of the product and eventual issues normally encountered during the scale-up phase can be treated under control.

7.8 Future Directions: Beyond the Canonical 20 Amino Acids

The new frontier in recombinant protein production is represented by the possibility of introducing unnatural amino acids into proteins. A possible strategy used to achieve this aim employs the three stop codons Amber, Ochre, and Opal to specify the unnatural amino acid. The first experiments were carried out with Amber codon in a cell-free system. Subsequently, the translation machinery was introduced in recombinant cells to achieve the direct incorporation of unnatural amino acids into the protein during the protein synthesis.[61]

The Amber codon is the least-used stop codon and its reassignment to unnatural amino acids was expected to have no impact on the synthesis of other proteins. An aminoacyl tRNA synthetase that specifically aminoacylates the suppressor tRNA with the unnatural amino acid was introduced in the recombinant strain genome. Feeding the fermentation with the unnatural amino acid, its incorporation into the protein was achieved.[62] The efficiency of incorporation is still low because the Amber stop codon is also recognized by the cellular release factor, which determines the termination of the protein synthesis.[63] This research, which is in an early phase, will potentially allow the creation of new proteins that harbor one or more unnatural amino acids and reduce the gap between biotechnology and chemical synthesis.

A very promising new tool, with great potential to streamline the development of new projects and scale-up effectiveness, among many other applications, is the coming of age of bioinformatic and chemoinformatic tools. The ability to rationally design new ideas *a priori* and test them *in silico*, and the possibility to exploit the capacity of algorithms to direct the research in a more focused manner, are all areas with great impact. When applied to the field of enzymes, for example, control and prediction of the activity of enzymes become a real possibility, leading to shorter development times, more efficient results, and overall more economical solutions, coupled with less waste. This is already taking place in academic laboratories, where these abilities are being combined with the direct evolution of enzymes with excellent results, opening up the way to improve the robustness and reactivity of enzymes, together with a broader selectivity for a wider range of substrates.

These advances can positively impact the industrial environment with effects on development timelines and material waste.

List of Abbreviations

ADH1	Alcohol dehydrogenase
AOX1, AOX2	Aldehyde oxidase enzymes
API	Active pharmaceutical ingredient
cDNA	Complementary DNA
DoE	Design of experiment
DPP-4	Dipeptidyl peptidase-4
emp	Enzyme-modified porcine
ER	Endoplasmic reticulum
FDA	US Food and Drug Administration
GAP	GTPase-accelerating protein
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GDP	Gross domestic product
GLP-1	Glucagon-like peptide 1
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
lacUV	Promoter from the *Escherichia coli* lac operon
LCA	Life-cycle assessment
LPPS	Liquid-phase peptide synthesis
MCS	Multiple cloning site
NEP 24.11	Neutral endopeptidase
PGK	Phosphoglycerate kinase
pUC	Plasmid cloning vector
QbD	Quality by design
rDNA	Recombinant DNA
rh-insulin	Recombinant human insulin
RP-HPLC	Reversed-phase high-performance liquid chromatography
SNAC	Sodium 8-[(2-hydroxybenzoyl)amino]octanoate
SPPS	Solid-phase peptide synthesis
TFA	Trifluoroacetic acid
TFF	Tangential flow filtration

References

1. L. Ferrazzano, M. Catani, A. Cavazzini, G. Martelli, D. Corbisiero, P. Cantelmi, T. Fantoni, A. Mattellone, C. De Luca, S. Feletti, W. Cabri and A. Tolomelli, *Green Chem.*, 2022, **24**, 975.
2. K. Wowra, E. Hegel, A. Scharf, A. Gruenberger and K. Rosenthal, *Trends Biotechnol.*, 2023, **41**, 1199.

3. M. Becker, N. Kolheb, S. Hunger, R. Ecshrich, R. Mueller and A. Aurich, *Eng. Life Sci.*, 2020, **20**, 90.
4. S. Feijoo, S. Gonzalz-Garcia, J. M. Lema and M. T. Moreira, *J. Cleaner Prod.*, 2017, **165**, 204.
5. H. Nielsen, K. Oxenboll and H. Wenzel, *Int. J. Life Cycle Assess.*, 2007, **12**, 432.
6. N. A. Baeshen, M. N. Baeshen, A. Sheikh, R. S. Bora, M. M. M. Ahmed, H. A. Ramadan, K. S. Saini and E. M. Redwan, *Microb. Cell Fact.*, 2014, **13**, 141.
7. T. Kjeldsen, P. Balschmidt, I. Diers, M. Hach, N. C. Kaarsholm and S. Ludvigsen, *Biotechnol. Genet. Eng. Rev.*, 2001, **18**, 89.
8. R. E. Chance, N. B. Glazer and K. L. Wisher, in *Biopharmaceuticals An Industrial Perspective*, 1999, p. 149.
9. F. G. Banting, C. H. Best and J. J. R. Macleod, *Am. J. Physiol.*, 1922, **59**, 479.
10. D. Mohajan and H. K. Mohajan, *Stud. Soc. Sci. Humanit.*, 2023, **2**, 8.
11. G. Walsh, *Appl. Microbiol. Biotechnol.*, 2005, **67**, 151.
12. L. Fryklund, *Med. Res. Arch.*, 2023, 11.
13. I. Vecchio, C. Tornali, N. L. Bragazzi and M. Martini, *Front. Endocrinol.*, 2018, **9**, 1.
14. G. A. Homandberg, J. A. Mattis and M. Laskowski, *Biochemistry*, 1978, **17**, 5220.
15. A. Riggs, *Endocr. Rev.*, 2021, **42**, 374.
16. W. Landgraf and J. Sandow, *Diabetes*, 2016, **12**, 12.
17. J. Alyas, A. Rafiq, H. Amir, S. U. Khan, T. Sultana, A. Ali, A. Hameed, I. Ahmad, A. Kazmi, T. Sajid and A. Ahmad, *Biomed. Res. Ther.*, 2021, **8**, 4540.
18. G. Walsh and E. Walsh, *Nat. Biotechnol.*, 2022, **40**, 1722.
19. P. Fonjallaz and E. Loumaye, *J. Biotechnol.*, 2000, **79**, 185.
20. Editorial, *Nat. Rev. Endocrinol.*, 2023, **19**, 311.
21. I. Mullerup, T. Kornfelt, P. Oreskov and H. Tottrup, *Glucagon III*, Springer Berlin, Heidelberg, 1996.
22. B. Manandhar and J. M. Ahn, *J. Med. Chem.*, 2015, **58**, 1020.
23. L. B. Knudsen and J. Lay, *Front. Endocrinol.*, 2019, **10**, 1.
24. J. L. Lau and M. K. Dunn, *Bioorg. Med. Chem.*, 2017, **26**, 2700.
25. I. Guryanov, A. Bondesan, D. Visentini, A. Orlandin, B. Biondi, C. Toniolo, F. Formaggio, A. Ricci, J. Zanon and W. Cabri, *J. Pept. Sci.*, 2016, 471.
26. D. Carbajo Lopez and F. Albericio Palomera, Method for preparation of Liraglutide using BAL linker, *WO pat.*, WO138855, 2017.
27. M. L. Puppala and V. Mudii, Process for preparation of Liraglutide using recombinant *Saccharomyces cerevisiae*, *WO pat.*, WO 082138, 2019.
28. D. L. Dunweber, I. H. Jensen and L. B. Hansen, Method for producing acylated peptides, *WO pat.*, WO029077, 2004.
29. T. Moks, L. Abrahmsen, B. Osterlof, S. Josephson, M. Ostlig, S. O. Enfors, I. Persson, B. Nilsson and e M. Uhlen, *Nature*, 1987, **5**, 379.
30. S. Gupta, S. S. Salunkhe, B. Varshney and R. S. Mody, Expression and large-scale production of peptides, *WO pat.*, WO172921, 2018.

31. N. Cheng, L. Yang, N. Dai, Z. Hu, F. Yang, R. Chen, J. Chen, Y. Zhou, Y. Huang and Z. Su, *Process Biochem.*, 2008, **43**, 1253.
32. S. Ahmadi, M. B. Shahsavani, Z. Tavaf, R. M. Albaghlany, A. Kumar, A. A. Moosavi-Movahedi and R. Yousefi, *PLoS One*, 2022, **17**, 1.
33. J. Lau, P. Bloch, L. Schaffer, I. Petterson, J. Speltzer, J. Kofoed, K. Madsen, L. B. Knudsen, J. McGuire, D. B. Steensgaard, H. M. Strauss, D. X. Gram, S. M. Knudsen, F. S. Nielsen, P. Thygesen, S. Reedtz-Runge and T. Fruse, *J. Med. Chem.*, 2015, **58**, 7370.
34. J. F. Lau, A. S. Andersen, P. Bloch, J. Lau, P. W. Garibay, T. Kruse, I. S. N. Noerby, C. U. Jessen, C. Christensen and J. C. Norrild, Semi-recombinant preparation of GLP-1 analogues, *WO pat.*, WO083549, 2009.
35. M. Lenharo, *Nature*, 2023, **620**, 480.
36. A. M. Vieira Gomes, T. S. Carmo, L. S. Carvalho, F. M. Bahia and N. S. Parachin, *Microorganisms*, 2018, **6**, 1.
37. M. Pouresmaeil and S. Aziz-Dargahlou, *Arch. Microbiol.*, 2023, **205**, 1.
38. T. W. Overton, *Drug Discovery Today*, 2014, **19**, 590.
39. G. L. Rosano and E. A. Ceccarelli, *Front. Microbiol.*, 2013, **40**, 257.
40. A. Berlec and B. Strukelj, *J. Ind. Microbiol. Biotechnol.*, 2013, **40**, 257.
41. A. Madhavan, K. B. Arun, R. Sindhu, J. Krishnamoorthy, R. Reshmy, R. Sirohi, A. Pugazhendi, M. K. Awasthi, G. Szakacs and P. Binod, *Microb. Cell Fact.*, 2021, **20**, 1.
42. M. Wacker, D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren and M. Aebi, *Science*, 2002, **298**, 1790.
43. N. Ferrer-Miralles, J. Domingo-Espin, J. L. Corchero, E. Vazquez and A. Villaverde, *Microb. Cell Fact.*, 2009, **8**, 1.
44. S. Wegmueller and S. Schimd, *Curr. Org. Chem.*, 2014, **18**, 1005.
45. G. Wang, M. Huang and J. Nielsen, *Curr. Opin. Biotechnol.*, 2017, **48**, 77.
46. A. Srivastava, J. Piskur, J. Nielsen and M. Egel-Mitani, Method for the production of heterologous polypeptides in transformed yeast cells, *US pat.*, US6190883, 2001.
47. M. Karbalaei, S. A. Rezaee and H. Farsiani, *J. Cell. Physiol.*, 2020, **235**, 5867.
48. M. Ahmad, M. Hirz, H. Pichler and H. Schwab, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 5301.
49. K. De Schutter, Y. Lin, P. Tiels, A. Van Hecke, S. Glinka, J. Weber Lehmann, P. Rouze, Y. Van de Peer and N. Callewaert, *Nat. Biotechnol.*, 2009, **27**, 561.
50. N. K. Tripathi and A. Shrivastava, *Front. Bioeng. Biotechnol.*, 2019, **7**, 1.
51. S. Takagi, N. Tsutsumi, Y. Terui, X. King, H. Yurimoto and Y. Sakai, *FEMS Yeast Res.*, 2019, **19**, 1.
52. Summary of the Pollution Prevention Act: 42 U.S.C. §13101 et seq, 1990.
53. Green Chemistry History, available at: https://www.acs.org/greenchemistry/what-is-green-chemistry/history-of-green-chemistry.html.
54. P. T. Anastas and J. C. Warner, *Green Chemistry: Theory and Practice*, Oxford University Press, New York, 1998.

55. P. D. De Maria, *Curr. Opin. Green Sustainable Chem.*, 2021, **31**, 100514.
56. M. Vandermies and P. Frickers, *Microorganisms*, 2019, **7**, 1.
57. M. Sun, A. X. Gao, X. Liu, Y. Yang, R. Ledesma-Amaro and Z. Bai, *Microb. Cell Fact.*, 2023, **22**, 1.
58. A. H. M. Gameil, F. Yusuf, A. S. Azmi and N. I. M. Puad, *Biol. Nat. Res. Eng. J.*, 2021, **5**, 37.
59. P. Agrawal, K. Wilkstein, E. Giunn, M. S.-M. C. I. Mason and J. Sayle, *Org. Process Res. Dev.*, 2023, **27**, 571.
60. ICH Q8 (R2), Pharmaceutical development – Scientific guideline, available at: https://www.ema.europa.eu/en/ich-q8-r2-pharmaceutical-development-scientific-guideline.
61. T. S. Young and P. G. Schultz, *J. Biol. Chem.*, 2010, **285**, 11039.
62. D. De la Torre and J. W. Chin, *Nat. Rev. Genet.*, 2021, **22**, 169.
63. S. Hans, N. Kumar, N. Gohil, K. Khambhati, G. Bhattacharjee, S. S. Deb, R. Maurya, V. Kumar, S. M. S. Reshamwala and V. Singh, *Microb. Cell Factories*, 2022, **21**, 1.
64. S. Ludvigsen, D. B. Steensgaard, J. K. Thomsen, H. Strauss and M. Normann, Liraglutide, 2012, DOI: 10.2210/pdb4apd/pdb.
65. X. Zhang, M. J. Belousoff, R. Danev, P. M. Sexton and D. Wootten, Semaglutide-bound Glucagon-Like Peptide-1 (GLP-1) Receptor in Complex with Gs protein, 2020, DOI: 10.2210/pdb7ki0/pdb.
66. X. Zhang, M. J. Belousoff, R. Danev, P. M. Sexton and D. Wootten, Structure and dynamics of semaglutide- and taspoglutide-bound GLP-1R-Gs complexes, *Cell Rep.*, 2021, **36**, 109374.