



Editorial

Recombinant DNA the Bio-Revolution, Between Promise, Hurdles, and Achievements

Mohamed Raafat El-Gewely

Department of Medical Biology, Faculty of Health Sciences, UiT Arctic University of Tromsø,
9037 Tromsø, Norway; raafat.el-gewely@uit.no

Not too long after the discovery of restriction enzymes, by Warner Arbor group in 1962 and 1963 [1,2] followed by the work of Hamilton Smith [3]; Dan Nathan and Hamilton Smith [4]; Wil A. M. Loenen et al. 2014 [5], elegantly reviewed the restriction enzyme type I discovery and history. Noting that Werner Arber, Dan Nathans, and Hamilton Smith, received the Nobel prize in 1978.

In 1973 Stanley Cohen, Anita Chang, Herbert Boyer, and Robert Helling published their first attempt to use restriction enzymes to cut pieces of DNA, ligate it to a plasmid and transform the ligated complex into the *E. coli* host cells, termed the recombinant DNA process. No Nobel prize for this milestone [6].

The first therapeutic recombinant protein on the market was insulin engineered by Genentech in 1978 and further developed and produced on a large scale by Elli Lilly as (Humulin). It was on the market after 5 months' evaluation by FDA on 28 October 1982.

Currently the number of recombinant therapeutic proteins on the market has significantly increased (Recombinant Proteins | DrugBank Online,) and so is its projection (URL last accessed 27 June 2025). The expectations of therapeutic recombinant proteins market are huge. The market, estimated at around USD 3.97 billion in 2025, and is forecasted to hit around USD 11.32 billion by 2034 (<https://www.precedenceresearch.com/recombinant-proteins-market>) (URL last accessed on 27 June 2025). Noting that a parallel market of biosimilars is thriving.

There are several problems in expressing an active desired recombinant protein whether as therapeutic or as a validated active recombinant protein target for drug discovery and development.

Most of the problems were due to lack or reduced expression of the recombinant protein or alternatively producing insoluble protein as inclusion bodies. This special issue is aiming to address such hurdles and current methods to overcome these hurdles. To name the most encountered obstacles, for example the discrepancy of the utilized amino acid tRNA codes between the heterologous host and the original organism. The codon bias in the heterologous gene expression and the original organism as reviewed by Claes Gustafsson et al. in 2004 [7]. Harmonization of the code became the standard as discussed by Evelina Angov et al., 2008 [8]. McDonald et al., 2015 reported that such discrepancy in codon utilization bias between human cells and *E. coli* cells is due to evolution during a recombinant gene expression of human genes affecting the profile of expressed tRNA expression in *E. coli* cells [9]. Interestingly, based on the data of one thousand human genome projects, Parisien et al. 2013 concluded that tRNA genes are more diverse than conventionally perceived [10]. A recurrent neural network (RNN) model of codon optimization achieved higher level of protein expression when transfected transiently or stably into Chinese hamster ovary cells [11]. Demissie et al., 2025 [12] presented an



Received: 4 June 2025

Revised: 24 June 2025

Accepted: 25 June 2025

Published: 10 July 2025

Citation: El-Gewely, M.R.

Recombinant DNA the Bio-Revolution, Between Promise, Hurdles, and Achievements. *Int. J. Mol. Sci.* **2025**, *26*, 6611. <https://doi.org/10.3390/ijms26146611>

Copyright: © 2025 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

integrated approach for the interplay of multiple factors to fine-tune the genetic sequences of the desired recombinant protein to match the translational machinery and codon usage preferences of specific heterologous organism.

The belief that the primary structure/sequence of protein alone determines its tertiary structure independent of the cellular environment is behind most of the problems encountering recombinant protein production. This was based on the works of at least two Nobel laureates; Anfinsen as well as Crick. Of course, it is true, but how? Several recombinant proteins expressed as misfolded inclusion bodies, which would require in vitro manipulations to solubilize [13–15]. Moreover, several synthetic peptides were not active because they did not assume the proper tertiary structure and must be subsequently solubilized [16–20].

The numerous reported difficulties in producing active recombinant proteins lead to the conclusion by El-Gewely, 2009, that the final structure–function of proteins depends on the folding environment, and protein primary structure does not alone guarantee a unique tertiary structure, functionality, or even solubility [21]. Primary structure is a key factor for protein folding, but so is the genetic environment of the cell. As matter of fact, we have managed to compromise the primary structure of the TrpR monomer protein, by site-directed mutagenesis, yet some of these heterodimer's combinations were active in vivo although these active heterodimers contained two complementing mutants [22]. The basic genetic formula can still apply, assuming that protein solubility and activity is the phenotype P, G is the genotype dictating the primary structure, while E = Environment factors and E/G is the interaction between environment and primary structure as well as host genes.

$$P = G + E + E/G$$

However, according to the books (and Anfinsen conclusions) by ignoring the cell environment and its relevant genetics influences: $P = G$

And that could explain most of the observed difficulties.

The co-expression of chaperones together with target recombinant protein in the heterologous host, Improvement of the recombinant protein production, both in quantity and activity [23–25].

Also, by screening synthetic signal peptide library, Jeon et al., 2024 achieved improved production and secretion of target recombinant protein using *Corynebacterium glutamicum* as the heterologous host [26]. See selected slide from one of my presentations on the subject at Bayer Berlin 1n 2015: Bayer-Berlin 7-20-2015-Selected slides.pptx.

It is very crucial to have active and soluble recombinant proteins to produce therapeutic recombinant proteins. Once a protein validated to be a drug discovery target, it is equally important to produce physical amounts of such recombinant protein for the use in drug discovery and development process. McGonigle, 2024, reviewed the subject of drug discovery using recombinant proteins [27].

In recent years, some efforts in using the mRNA directly replace the therapeutic recombinant protein directly instead of going through the efforts of expressing the active therapeutic protein in a heterologous host [28,29].

Interesting recent development is dealing with the prediction of the tertiary protein structure only from its amino acid sequence, as reviewed recently by Meng et al., 2025 [30], thus reducing several required steps in drug discovery, including the expression of active and soluble recombinant protein target for the drug discovery, crystallization of the protein and conducting the essential X-ray crystallography. This would facilitate drug discovery by adding a valuable tool, but this will not impact on the therapeutic recombinant protein field, or in producing active antibodies against undesired expressed protein.

Conflicts of Interest: The author declares no conflicts of interest.

References

- Arber, W.; Dussoix, D. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage λ . *J. Mol. Biol.* **1962**, *5*, 18–36. [[CrossRef](#)] [[PubMed](#)]
- Arber, W.; Hattman, S.; Dussoix, D. On the host-controlled modification of bacteriophage λ . *Virology* **1963**, *21*, 30–35. [[CrossRef](#)] [[PubMed](#)]
- Kelly, T.J., Jr.; Smith, H.O. A restriction enzyme from *Hemophilus influenzae*. II. Base sequence of the recognition site. *J. Mol. Biol.* **1970**, *51*, 393–409. [[CrossRef](#)] [[PubMed](#)]
- Nathans, D.; Smith, H.O. Restriction endonucleases in the analysis and restructuring of DNA. *Annu. Rev. Biochem.* **1975**, *44*, 273–289. [[CrossRef](#)]
- Loenen, W.A.M.; Dryden, D.T.F.; Raleigh, E.A.; Wilson, G.G.; Murray, N.E. Highlights of the DNA cutters: A short history of the restriction enzymes. Type I restriction enzymes and their relatives. *Nucleic Acids Res.* **2014**, *42*, 13–19.
- Cohen, S.N.; Chang, A.C.; Boyer, H.W.; Helling, R.B. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3240–3244. [[CrossRef](#)]
- Claes, G.; Govindarajan, S.; Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol.* **2004**, *22*, 346–353.
- Angov, E.; Hillier, C.J.; Kincaid, R.L.; Lyon, J.A. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE* **2008**, *3*, e2189. [[CrossRef](#)]
- McDonald, M.J.; Chou, C.H.; Swamy, K.B.; Huang, H.-D.a.; Leu, J.-Y. The evolutionary dynamics of tRNA-gene copy number and codon-use in *E. coli*. *BMC Evol. Biol.* **2015**, *15*, 63. [[CrossRef](#)]
- Parisien, M.; Wang, X.; Pan, T. Diversity of human tRNA genes from the 1000-genomes project. *RNA Biol.* **2013**, *10*, 1853–1867. [[CrossRef](#)]
- Goulet, D.R.; Yan, Y.; Agrawal, P.; Waight, A.B.; Mak, A.N.; Zhu, Y. Codon optimization using a recurrent neural network. *J. Comput. Biol.* **2023**, *30*, 70–81. [[CrossRef](#)]
- Demissie, E.A.; Park, S.Y.; Moon, J.H.; Lee, D.Y. Comparative Analysis of Codon Optimization Tools: Advancing toward a Multi-Criteria Framework for Synthetic Gene Design. *J. Microbiol. Biotechnol.* **2025**, *35*, e2411066. [[CrossRef](#)]
- Nabiel, A.; Yosua, Y.; Sriwidodo, S.; Maksum, I.P. Overview of refolding methods on misfolded recombinant proteins from *Escherichia coli* inclusion bodies. *J. App. Biol. Biotech.* **2023**, *11*, 47–52. [[CrossRef](#)]
- Markossian, K.A.; Kurganov, B.I. Protein Folding, Misfolding, and Aggregation. Formation of Inclusion Bodies and Aggregates. *Biochemistry (Moscow)* **2004**, *69*, 971–984. [[CrossRef](#)]
- Guisse, A.D.; West, S.M.; Chaudhuri, J.B. Protein folding in vivo and renaturation of recombinant proteins from inclusion bodies. *Mol. Biotechnol.* **1996**, *6*, 53–64. [[CrossRef](#)]
- Mori, T.; Kataoka, H.; Tanabe, G.; Into, T. Solubility affects IL-1 β -producing activity of the synthetic candidalysin peptide. *PLoS ONE* **2022**, *17*, e0273663. [[CrossRef](#)]
- Boutin, J.A.; Tartar, A.L.; Dorselaer, A.V.; Vaudry, H. General lack of structural characterization of chemically synthesized long peptides. *Protein Sci.* **2019**, *28*, 857–867. [[CrossRef](#)]
- Burra, G.; Thakur, A.K. Insights into the molecular mechanism behind solubilization of amyloidogenic polyglutamine-containing peptides. *Pept. Sci.* **2018**, *111*, e24094. [[CrossRef](#)]
- Haney, E.F.; Wu, B.C.; Lee, K.; Hilchie, A.L.; Hancock, R.E.W. Aggregation, and Its Influence on the Immunomodulatory Activity of Synthetic Innate Defense Regulators. *Cell Chem. Biol.* **2017**, *24*, 969–980.e4. [[CrossRef](#)]
- Zapadka, K.L.; Becher, F.J.; Gomes dos Santos, A.L.; Jackson, S.E. Factors affecting the physical stability (aggregation) of peptide therapeutics. *Interface Focus* **2017**, *7*, 20170030. [[CrossRef](#)]
- El-Gewely, M.R. Shorter is better. *Nat. Biotechnol.* **1999**, *17*, 210.
- Storbakk, N.; Fenton, C.; Riise, H.M.F.; Nilsen, I.W.; El-Gewely, M.R. In Vivo Interaction Between Mutated Tryptophan Repressors of *Escherichia coli*. *J. Mol. Biol.* **1996**, *256*, 889–896. [[CrossRef](#)] [[PubMed](#)]
- Fatima, K.; Naqvi, F.; Younas, H. A Review: Molecular Chaperone-mediated Folding, Unfolding and Disaggregation of Expressed Recombinant Proteins. *Cell Biochem. Biophys.* **2021**, *79*, 153–174. [[CrossRef](#)] [[PubMed](#)]
- Raschmanová, H.; Weninger, A.; Knejzlík, Z.; Melzoch, K.; Kovar, K. Engineering of the unfolded protein response pathway in *Pichia pastoris*: Enhancing production of secreted recombinant proteins. *Appl. Microbiol. Biotechnol.* **2021**, *105*, 4397–4414. [[CrossRef](#)] [[PubMed](#)]
- Zhou, Y.; Lu, Z.; Wang, X.; Selvaraj, J.N.; Zhang, G. Genetic engineering modification and fermentation optimization for extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 1545–1556. [[CrossRef](#)]
- Jeon, E.J.; Lee, S.M.; Hong, H.S.; Jeong, K.J. Design of fully synthetic signal peptide library and its use for enhanced secretory production of recombinant proteins in *Corynebacterium glutamicum*. *Microb. Cell Fact.* **2024**, *23*, 252. [[CrossRef](#)]
- McGonigle, P. How Biologics Have Changed the Drug Discovery Landscape. *Annu. Rev. Pharmacol. Toxicol.* **2025**, *65*, 29–46. [[CrossRef](#)]

28. Vavilis, T.; Stamoula, E.; Ainatzoglou, A.; Sachinidis, A.; Lamprinou, M.; Dardalas, I.; Vizirianakis, I.S. mRNA in the Context of Protein Replacement Therapy. *Pharmaceutics* **2023**, *15*, 166. [[CrossRef](#)]
29. Rohner, E.; Yang, R.; Foo, K.S.; Goedel, A.; Chien, K.R. Unlocking the promise of mRNA therapeutics. *Nat. Biotechnol.* **2022**, *40*, 1586–1600. [[CrossRef](#)]
30. Meng, Y.; Zhang, Z.; Zhou, C.; Tang, X.; Hu, X.; Tian, G.; Yang, J.; Yao, Y. Protein structure prediction via deep learning: An in-depth review. *Front. Pharmacol.* **2025**, *16*, 1498662. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.