Post-Translational Modifications and Their Applications in Biopharmaceutics

Keywords: post-translational modifications, biopharmaceuticals, glycosylation, proteolysis, disease pathogenesis

Abstract: Post-Translational Modifications(PTMs) of proteins are a key aspect of the increase in protein diversity within cells. They also play an important role in controlling many biological processes and hence can be used to design therapeutic drugs to control and prevent the progression of certain diseases. This paper will provide a brief overview of the mechanism of action of PTMs as well as the role they play in disease pathogenesis. It will discuss in detail the use of PTMs in the biopharmaceutical industry by looking at two specific examples, namely glycosylation and proteolysis, and consider the challenges faced by pharmacologists in their pursuit to manufacture more efficient drugs.

1. Introduction

Identifying and understanding cellular processes at a molecular level is essential for the understanding of the working of cells and on a larger scale, entire organisms. One of the major developments that helped understand these processes was the Human Genome Project, in which over 3 billion nucleotides of the human genome were sequenced, and the genes which were made of these nucleotides mapped. The human genome consists of 20,000 to 25,000 genes and is constant from one cell to the another. It also rarely undergoes any changes

during the lifespan of the cell. On the other hand, there are over 1 million proteins in human cells[1]. The proteins that are or can be expressed by a cell at a certain time -the proteome - varies greatly from one cell to another and even during the lifetime of a cell. This expansion in number from the genome to the proteosome is due to two processes; Alternative Splicing, which occurs after transcription, and Post-Translational Modifications, which occur after translation. Transcription is the process by which DNA segments are copied into complementary pre-RNA segments. In alternative splicing, different combinations of coding sequences present in the pre-mRNA are selected to form variable mature mRNA transcripts, which in turn are used to synthesize numerous isoforms of proteins through translation. Translation is the process by which the polypeptide chains of proteins are synthesized by ribosomes, which read the information present in the spliced mature mRNA.

Post-Translational Modifications (PTMs) are chemical modifications that occur to the polypeptide chain after translation has occurred. These changes can include the addition of functional groups, carbohydrates, lipids or other proteins to the amino acids of proteins, or the cleavage of polypeptide bonds in a pro-peptide so as to convert it to a mature protein. Some modifications even act as markers to choose protein targets for degradation. It is known that there are only 20 amino acids present in human cells, and that the DNA codes for only these. However, PTMs occur on these 20 amino acids to produce over 140 variations

of these. Thus, there is a large increase in the diversity of the structure and function of proteins due to PTMs[2]. PTMs also play an important role in the biological processes occurring in cells such as gene expression regulation, interprotein interactions, protein localization, signalling and degradation.

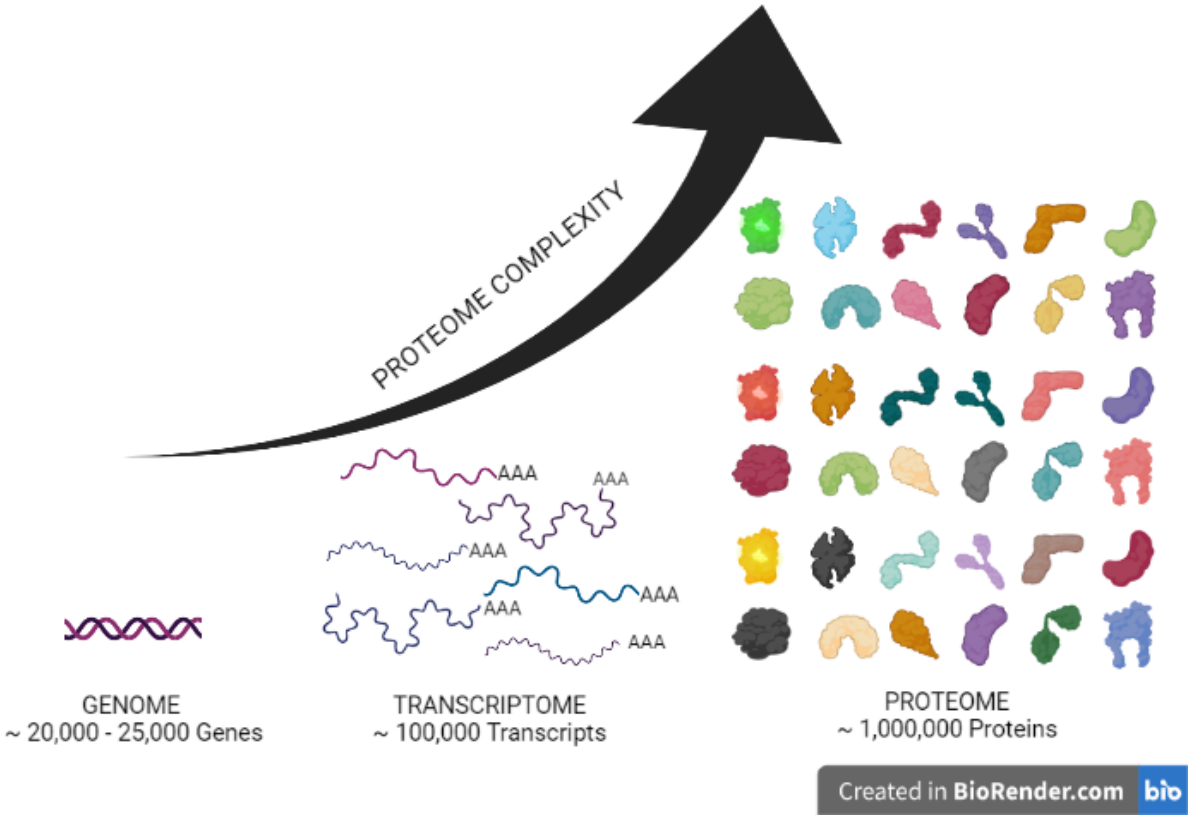

Figure 1. Showing the exponential increase in complexity from genome to transcriptome to proteome. Post-translational modifications are key to increasing proteome diversity. Created with BioRender.com

2. Post-Translational Modifications

2.1. Mechanism of Action of PTMs

Post-translational modifications can occur at any phase during protein synthesis. Early PTMs are performed on the polypeptide chain formed as soon as translation occurs and before it gets folded into its final conformation. They may play a role in the efficiency with which the protein gets folded later, the stability of the final conformation as well as the targeting of the polypeptide chain to different locations in the cell. Late stage PTMs occur on fully folded proteins and are important for activating or inactivating the catalytic ability of these proteins or regulating any other function of these proteins[2].

While, post-translational modifications have been discovered in prokaryotic organisms[3], they are far more common in eukaryotes and mainly occur on histones, secretory and membrane proteins. PTMs are performed either on the amino acid side chains of polypeptides or on the peptide bonds linking amino acids themselves and mostly occur through enzymatic catalysis[4]. Almost 5% of the eukaryotic genome codes for enzymes which recognize specific signal sequences present in the proteins to be modified and bind to them to catalyse over 400 post-translational modifications. However, not all proteins require enzymes in order to be modified. Some proteins possess autocatalytic domains, and can thus perform their own modifications[5]. Since modifications can be occur at any point in the life cycle of a protein, they can take place in various cell organelles, from the cytoplasm to the endoplasmic reticulum and the Golgi apparatus.

PTMs can be either reversible or irreversible depending on the type of bond formed during modification. Modifications which involve the formation of covalent bonds are reversible, while modifications which involve proteolytic cleavage of peptide bonds are irreversible. Reversible PTMs are extremely useful, as they help in controlling the functions of proteins quickly and while using less energy. The degradation of an unnecessary protein and its synthesis from scratch when needed is much more time-consuming and costly than simply toggling the activity of the protein between “on” and “off” by modifying and de-modifying it[4].

The most common post-translational modifications include phosphorylation, glycosylation, ubiquitination, methylation, acetylation, lipidation and proteolysis. Though over 400 post-translational modifications have been identified in the human genome, the function of less than 200 of these has been discovered.

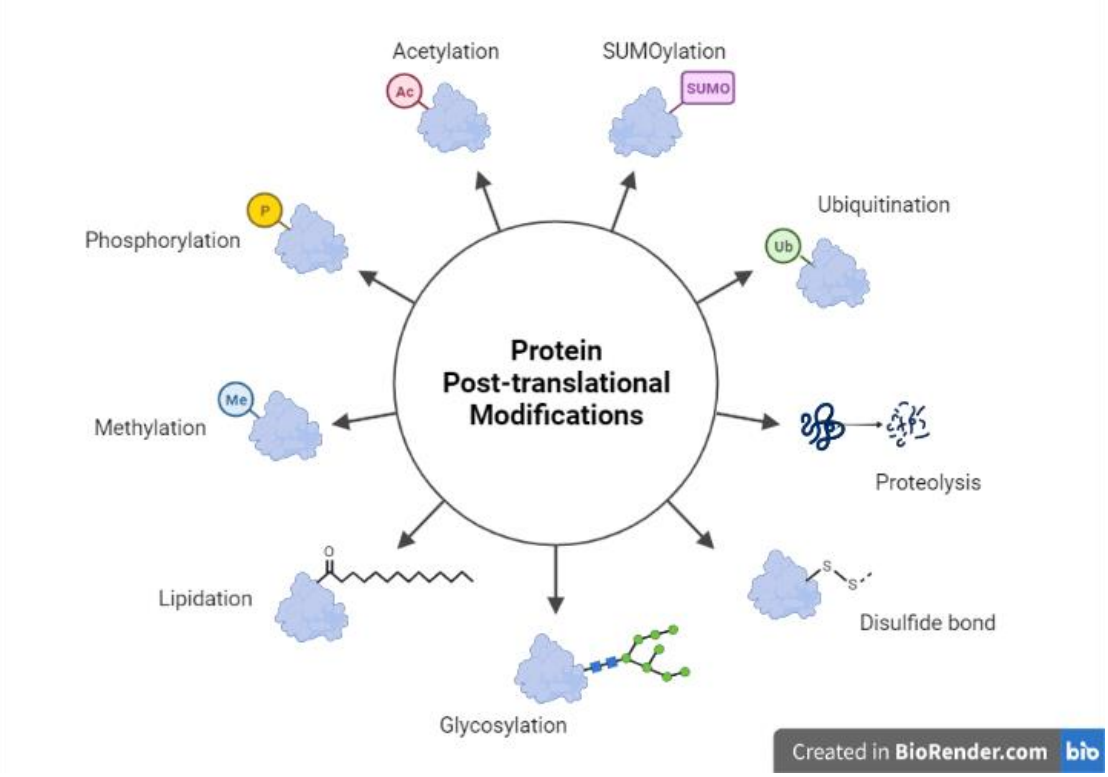

Figure 2. Most common types of Post-Translational modifications. Created with BioRender.com

2.2. Post-Translational Modifications and Diseases

As PTMs play a major role in determining the folding, structure and function of proteins, any error or disruption in a PTM will disrupt the biological processes involving the protein and may lead to the emergence of a disease. The main challenge in the study of post-transitionally modified proteins has been in their detection and isolation methods. However, recent advances in proteomic technologies(define) such as two-dimensional gel electrophoresis and mass spectrometry have helped us understand exactly how certain PTMs play a direct

role in the pathogenesis of diseases. It has been found that mutations in the potential PTM sites are usually directly involved in various disorders.

For example, histone proteins undergo various post-translational modifications such as acetylation and methylation, which alter the structure of the chromatin network. Some modifications also act as signals to ensure that DNA binds to certain proteins and not others. Thus, any mutation in these PTMs will affect the way histone interacts with different regions of DNA like the coding sequences, promoter regions and repetitive sequences, which may lead to cancer. Another example of a mutation in a PTM which causes a disease is the hyperphosphorylation of tau proteins, which are associated with microtubules in cells. This leads to the formation of paired helical filaments and neurofibrillary tangles, which are key pathological features of many neurodegenerative disorders, such as Alzheimer's Disease.

3. PTMs in Pharmaceuticals:

3.1. History of the Biopharmaceutical Industry

The Biopharmaceutical industry, now valued at approximately 70-80 billion dollars on the global market, has grown rapidly over the past 30 years. The first biopharmaceutical which was approved for human use in 1982 was the recombinant human insulin protein, known as *humulin* and produced in E. coli bacteria. Biopharmaceutical products today include nucleic-acid based products, engineered cell and tissue-based products as well as recombinant proteins.

However, recombinant proteins still form the majority of therapeutic drugs sold. These proteins are commonly eukaryotic proteins engineered in laboratories and modified through post-translational modifications[6].

3.2. PTMs Commonly used in Therapeutic Drug Manufacturing

Though there are many PTMs which have been discovered, those such as phosphorylation and acetylation, which are involved in intracellular regulation (gene expression, movement through the endoplasmic system, intracellular transduction) are not useful in the production of therapeutic proteins. The most common modifications associated with biopharmaceuticals include glycosylation, amidation, carboxylation and hydroxylation. Other PTMs, such as disulphide bond formation and proteolytic cleavage are necessary to ensure the stability of the protein product or are used to convert an inactive product to an active one. They are performed on proteins after they have been fully synthesized.

3.3. Glycosylation and its Use in Manufacturing Monoclonal Antibodies

Glycosylation involves the addition of a carbohydrate moiety to specific residues present on a protein chain. The process is reversible and catalysed by the enzyme glycosyltransferase, and commonly occurs on the serine, threonine, asparagine, and tryptophan amino acid side chains of proteins present on the cell surface as well as secretory proteins. Based on where the carbohydrate is

added, there are two types of glycosylation: N-linked Glycosylation, where the oligosaccharide is added to the nitrogen atom in the side chain of asparagine residues of proteins in the Endoplasmic Reticulum, and O-linked Glycosylation, where the oligosaccharide is added to the oxygen atom in the side chain of serine or threonine residues of proteins in the Golgi Apparatus [4].

Glycosylation is one of the most common and complex post-translational modifications. Almost 1-2% of the human genome codes for enzymes that catalyse glycosylation. It is the most widely used PTM in therapeutic drugs, with glycoproteins making up almost 40% of all biopharmaceuticals sold in the market[7]. Most of these are monoclonal antibodies, which have become a fast-growing branch of biopharmaceuticals and are used to treat a variety of diseases, including cancer.

Monoclonal antibodies are those antibodies which bind to only one specific antigen. The antibodies used as therapeutics are of the immunoglobin G(IgG) class. These antibodies consist of 2 heavy chains and 2 light chains, bound together to form 3 protein moieties. Two of these moieties are called Fab regions, and have specific antigen-binding sites, while the third is called the Fc region, which has ligand-binding sites to bind to ligands which in turn activate the antibody's effector functions. Effector functions play an important in the body's humoral immunity[8]. During its synthesis in the cell, IgG-Fc region gets post-transitionally modified, and an oligosaccharide gets linked to the protein

through N-glycosylation. This oligosaccharide is essential in modulating the binding affinity of IgG to its corresponding antigen and thus activating effector functions which will result in the destruction of the antigens.

One specific example of how understanding oligosaccharide functions in antibodies helps in the engineering of cancer drugs is the removal of the fucose sugar from the glyco-component that binds to the IgG antibody. IgG is normally glycosylated at the asparagine residue 297 of the antibody's heavy chain, and this modification aids in triggering antibody-dependent cellular cytotoxicity (ADCC), which is an effector function used to mediate the destruction of cancerous cells. It was discovered that the presence of the fucose sugar on the glycol-component reduced the ADCC activity by 100 times. Therefore, cells which did not possess the gene to code for the fucosyltransferase enzyme(FUT8 gene) which attached the fucose group to rest of the carbohydrate were engineered. These cells were found to produce antibodies with higher cancer-killing ability[6](see Fig. 3).

The main reason glycosylation is so useful in the pharmaceutical industry is that it produces a large amount of heterogeneity. The variety in glycoproteins arises from the addition of different types of oligosaccharides to any site and also due to the fact that multiple sugar moieties can be added to both the N- and O-glycosylation sites. This diversity allows for the customization of Monoclonal

antibodies, such that specific antibodies can be synthesized in laboratories for each type of antigen.

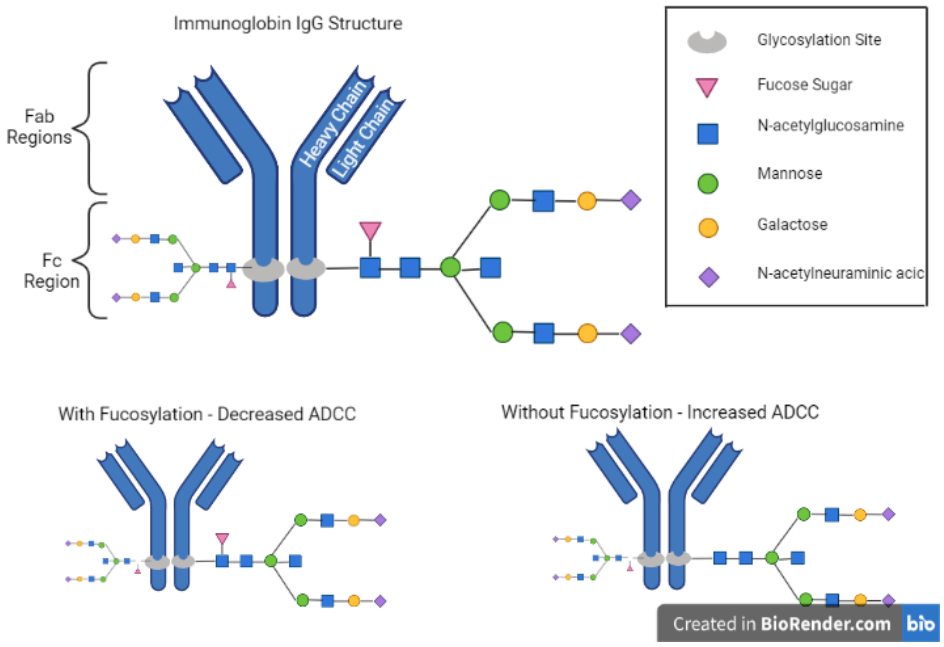

Figure 3. Structure of IgG antibodies. The removal of the fucose group on the sugar sidechain increases the cancer-killing ability of antibodies by up to 100 times. Created with BioRender.com

3.4. Proteolytic Cleavage and its Use in Manufacturing Insulin

Proteolytic processing is an irreversible post-translational modification which involves the hydrolysis of peptide bonds by a set of enzymes called proteases. This process cleaves large proteins into smaller ones, exposing new N- and C-terminals, and thus produces new proteins with modified functions[9]. There are mainly two types of proteolysis based on what function they perform. Degradative proteolysis is essential for the removal of misfolded proteins and

unassembled protein subunits. Biosynthetic proteolysis involves the cleavage of signal peptide sequences from nascent proteins as well as the cutting of zymogens(inactive forms of enzymes) at specific sites in order to convert them to their active, mature forms[4].

One application of biosynthetic proteolysis is in the production of human insulin. Insulin is a protein used in the regulation of fat and carbohydrate metabolism. Pre-proinsulin is the nascent polypeptide form of insulin which is synthesized in the beta pancreatic cells and possesses a signal sequence directing it to the Endoplasmic reticulum. Once pre-proinsulin enters the ER, the signal sequence is cleaved, allowing it to be folded into the inactive zymogen form of the enzyme, known as proinsulin. Proinsulin consists of 3 polypeptide chains - A, B and C - held together by 2 disulphide bonds in the order B-C-A. Besides this, A also has interchain disulphide bonds. Since the C chain is biologically inactive, it is removed by proteolytic cleavage in the Golgi apparatus, and A and B are fused back together to produce the active insulin enzyme(see Fig. 4).

The inability to produce sufficient insulin leads to a serious condition called diabetes mellitus. Recombinant human insulin is produced in E. coli bacteria in a manner very similar to how it is synthesized within the human cells and given to these patients in order to make up for the deficiency. In this process, DNA which codes for proinsulin is inserted into E. coli bacteria and proinsulin is

produced in large quantities. It is then isolated, and proteolytic cleavage is performed on it to remove the C-chain[10]. This method is more efficient than the previous methods of synthesizing the A and B chains separately and incubating them together in order to produce mature insulin and has allowed for the large-scale production of human insulin. Thus, understanding how proteolytic cleavage occurs and its importance in the functioning of insulin has allowed the pharmaceutical industry to keep up with the increasing demand for insulin.

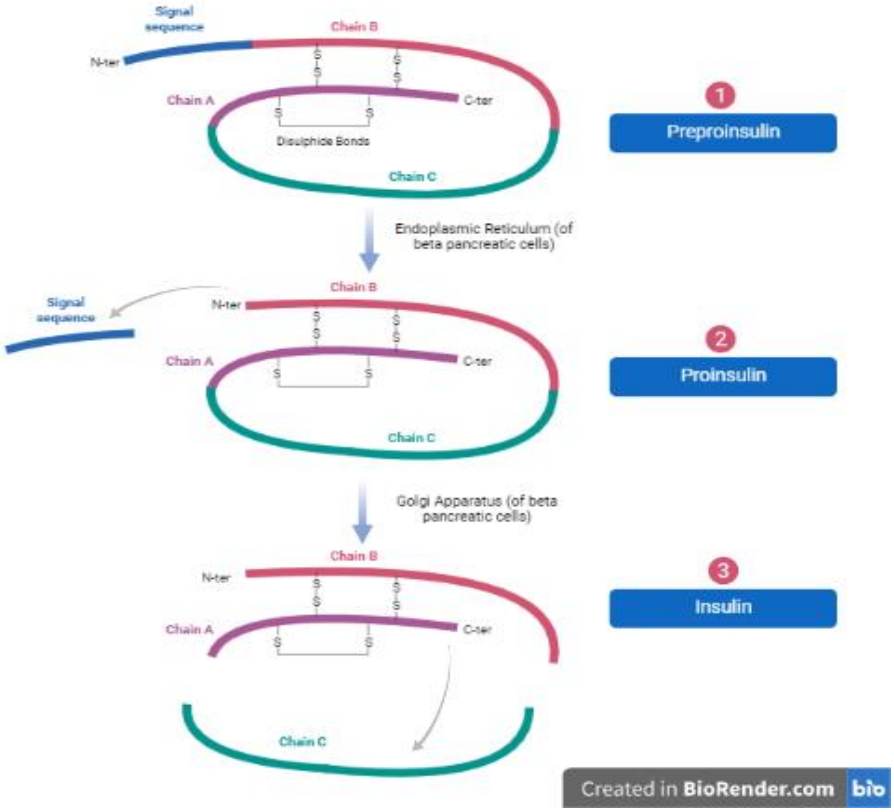

Figure 4. Shows the conversion of Pre-proinsulin to Insulin by removal of inactive C sidechain. Created with BioRender.com

3.5. Other Examples of PTMs Performed on Biopharmaceuticals

Amidation: This process involves the replacement of the carboxyl group present on the C-terminal of a protein with an amide (-$CONH_2$) group. One of the applications of amidation in the biopharmaceutical industry is in the production of recombinant salmon calcitonin to treat Paget's disease. Calcitonin is a hormone used in the regulation of calcium and potassium levels in the blood, and it has been discovered that salmon calcitonin is much more efficient than human calcitonin. However, E. coli can only produce a precursor peptide which has not undergone any post-translational modification and thus, amidation is performed on this peptide after it has been isolated to convert it into an active hormone.

γ-Carboxylation and β-Hydroxylation: γ-Carboxylation involves the conversion of specific glutamate residues on protein chains to γ-carboxyglutamate and β-hydroxylation involves the conversion of target aspartate residues into β-hydroxyaspartate. These PTMs are found in activated protein C(APC), which is an antithrombotic molecule, i.e., prevents the formation of blood clots. The native C protein possesses nine carboxylation sites and one hydroxylation site, all of which must be modified in order for the protein to become active. The deficiency of APCs is characteristic of sepsis, and in many cases, may even be a predictor of death. Thus, the active protein C is manufactured, where all 10

PTM sites have been modified, and these are sold as a drug to treat severe sepsis[6].

4. Challenges faced by the Biopharmaceutical Industry in Production of Therapeutic Drugs

Although we now understand much more about PTMs than we did 15-20 years ago, one commonly occurring phenomenon, called PTM crosstalk, is still greatly under-studied. Crosstalk between PTMs occurs when multiple PTMs are performed on the same protein and contribute equally to the functioning of the protein[11]. For example, multiple O-linked and N-linked glycoforms occurring on the same polypeptide. Thus, in the future, there must be greater efforts to understand the precise relationship between the structure and function of proteins which possess these complex post-translational modifications so that they can be utilized to produce more accurate biopharmaceutical drugs[7].

One of the main challenges faced by the biopharmaceutical industry in the production of therapeutic drugs is the fact that they use a reductionist philosophy. In this, the main function of a protein is identified *in vitro* and no effort is made to understand the other roles the protein may play. Therefore, when post-translational modifications are performed on a certain protein in order to optimize it for a single function, it may adversely affect other, equally important functions performed by the protein *in vivo*(within the organism). This may lead to patients experiencing side effects on receiving such a drug. Hence,

there is a need to develop better methods to identify exactly which patients will react favourably and which will experience adverse reactions when given a particular drug. There is also the larger challenge of coming up with methods to study the effect of biopharmaceutical drugs *in vivo*. Once we have understood this, we can easily determine why certain patients react unfavourably to drugs and other do not[7].

5. Conclusion

For a long time, great efforts were being made to study the human genome and its implications in detail, while the large-scale study of the protein function and interactions was lagging behind. It is only in the past 15-20 years that technological advances have been made to characterize and understand the functions of PTMs. This has allowed for a more logical approach towards the engineering of PTMs to produce biopharmaceuticals which show enhanced specificity in their function and therapeutic effect and are thus more efficient and of better quality[7]. However, while a broad picture of post-translational medications has started to emerge, there is still much that is unknown about them.

References:

1. Virág, D. *et al.* Current Trends in the Analysis of Post-translational Modifications. *Chromatographia* **83**, 1–10 (2020).

2. Uversky, V. N. Posttranslational Modification. *Brenner's Encycl. Genet. Second Ed.* 425–430 (2013) doi:10.1016/B978-0-12-374984-0.01203-1.

3. Macek, B. *et al.* Protein post-translational modifications in bacteria. *Nat. Rev. Microbiol. 2019 1711* **17**, 651–664 (2019).

4. Ramazi, S. & Zahiri, J. Post-translational modifications in proteins: resources, tools and prediction methods. *Database* **2021**, (2021).

5. Overview of Post-Translational Modification | Thermo Fisher Scientific - IN. https://www.thermofisher.com/in/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-post-translational-modification.html.

6. Walsh, G. Post-translational modifications of protein biopharmaceuticals. *Drug Discov. Today* **15**, 773–780 (2010).

7. Walsh, G. & Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat. Biotechnol. 2006 2410* **24**, 1241–1252 (2006).

8. Van Erp, E. A., Luytjes, W., Ferwerda, G. & Van Kasteren, P. B. Fc-mediated antibody effector functions during respiratory syncytial virus

infection and disease. *Front. Immunol.* **10**, 548 (2019).

9. Rogers, L. D. & Overall, C. M. C. Proteolytic Post-translational Modification of Proteins: Proteomic Tools and Methodology. *Mol. Cell. Proteomics* **12**, 3532 (2013).

10. Baeshen, N. A. *et al.* Cell factories for insulin production. *Microb. Cell Fact.* **13**, (2014).

11. Beltrao, P., Bork, P., Krogan, N. J. & Van Noort, V. Evolution and functional cross-talk of protein post-translational modifications. *Mol. Syst. Biol.* **9**, (2013).