

Systematic Analysis of Post-Translational Modifications for Increased Longevity of Biotherapeutic Proteins

Justin Kim*, Karanveer Sadiora*

Thomas Jefferson High School for Science and Technology, Alexandria, VA, USA

Email: 2026kim2@tjhsst.edu, 2026ksadiora@tjhsst.edu

How to cite this paper: Kim, J. and Sadiora, K. (2024) Systematic Analysis of Post-Translational Modifications for Increased Longevity of Biotherapeutic Proteins. *Computational Molecular Bioscience*, 14, 125-145. <https://doi.org/10.4236/cmb.2024.143005>

Received: July 8, 2024

Accepted: September 7, 2024

Published: September 10, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Protein-based therapeutics (PPTs) are drugs used to treat a variety of different conditions in the human body by alleviating enzymatic deficiencies, augmenting other proteins and drugs, modulating signal pathways, and more. However, many PPTs struggle from a short half-life due to degradation caused by irreversible protein aggregation in the bloodstream. Currently, the most researched strategies for improving the efficiency and longevity of PPTs are post-translational modifications (PTMs). The goal of our research was to determine which type of PTM increases longevity the most for each of three commonly-used therapeutic proteins by comparing the docking scores (DS) and binding free energies (BFE) from protein aggregation and reception simulations. DS and BFE values were used to create a quantitative index that outputs a relative number from -1 to 1 to show reduced performance, no change, or increased performance. Results showed that methylation was the most beneficial for insulin ($p < 0.1$) and human growth hormone ($p < 0.0001$), and both phosphorylation and methylation were somewhat optimal for erythropoietin ($p < 0.1$ and $p < 0.0001$, respectively). Acetylation consistently provided the worst benefits with the most negative indices, while methylation had the most positive indices throughout. However, PTM efficacy varied between PPTs, supporting previous studies regarding how each PTM can confer different benefits based on the unique structures of recipient proteins.

Keywords

Post-Translational Modification, Protein-Based Therapeutics, Therapeutic Half-Life, Protein Aggregation, Protein Reception

*Joint First Authors (In Alphabetical Order).

1. Introduction

Protein-based therapeutics is a rapidly growing field of study that is gaining unprecedented clinical traction worldwide. Currently, over 100 types of therapeutic proteins are approved for use in the European Union and the United States, and the industry surpassed \$100 billion in sales in 2010 [1]. These figures were established just 40 years after the first recombinant protein therapeutic, human insulin, was produced in the 1970s [1]. Therapeutic proteins can be used to treat enzymatic deficiencies, modulate signaling pathways, or assist in delivering other drugs. It gives researchers great flexibility with drug design and allows for direct augmentation of protein levels, resulting in faster, more systemic effects than traditional pharmaceuticals [1]. However, there is still great room for growth. Currently, protein-based therapeutic agents (PPTs) struggle from short half-lives that impair their ability to reach their target tissue and cause lasting effects. Most PPTs have a propensity to aggregate irreversibly during storage, transportation, or injection. Aggregated proteins often change conformations, reducing or modifying their biological activity and creating greater risks of harmful immune responses [2]. Proteins and aggregates are also often rapidly degraded or cleared from the circulatory system through the ubiquitin-proteasome system or glomerular filtration in the kidney [3]. These factors result in low PPT half-life in the circulation and, subsequently, a shorter duration of action. Therefore, there is a clear inverse relationship between aggregation and protein half-life. Consequently, protein-based therapeutic treatments usually require increased doses than their traditional counterparts, which can lead to blood toxicity and immunogenicity concerns, and generally have a lower effectiveness [3]. Our goal is to increase the half-life and longevity of these proteins by decreasing their propensity to aggregate with strategically chosen PTMs to make PPTs a more viable, effective option for drug delivery.

Currently, several different strategies have been explored to increase the success rate of PPTs. For example, post-translational modifications (PTMs) of therapeutic proteins, including PEGylation (covalently attached ethylene glycol to decrease renal clearance) and hyperglycosylation (to mimic endogenous proteins and bypass the immune system), have been developed [4] [5]. A diagram demonstrating PTMs in proteins is shown in **Figure 1**.

There are also PLGA microsphere and lipid delivery strategies that envelop the protein to protect them and increase effectiveness. In addition, new research is being conducted on biodegradable polymers and polypeptides to be used as an alternative to PEGylation, but these require more research to be fully understood and put into use [6]. Similar to PTMs, one study determined the effects of adding a lipidated amino acid group onto the protein as a fatty acid chain. The research determined that *in vivo*, attaching fatty acids to proteins increased their binding to human serum albumin. This was confirmed in mice when GLP1(HepoK), the protein with the fatty acid, showed stronger binding to human serum albumin

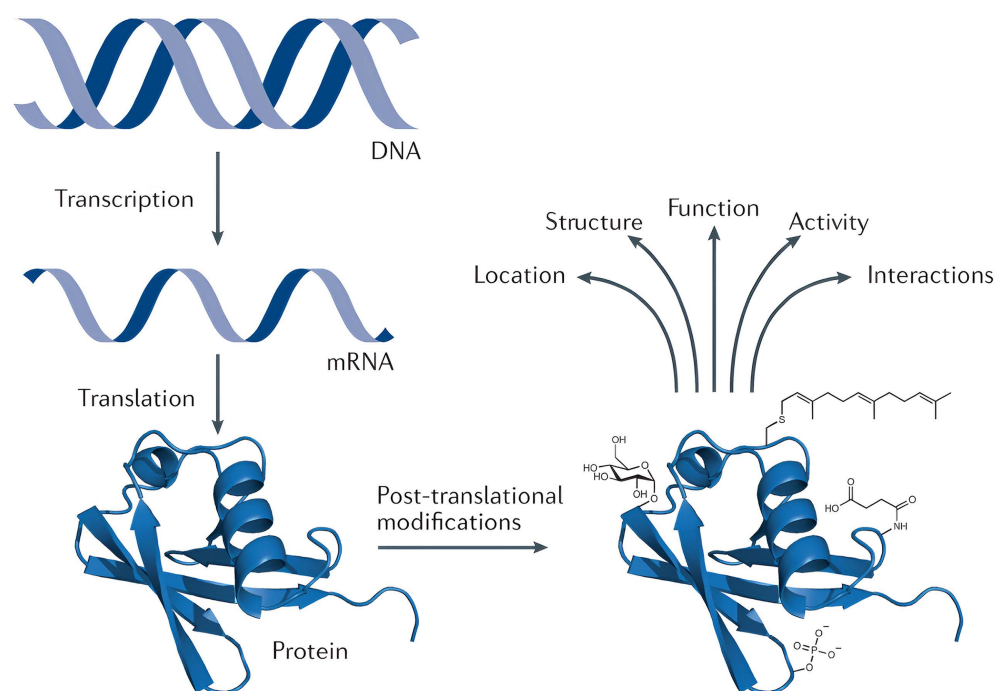


Figure 1. Diagram demonstrating post-translational modifications (PTMs) in proteins and their change in structure & function [5].

than GLP1(WT), the protein without, without impairing the stimulation of the GLP1 receptor in cells, showing that the original function of the protein was not modified [7]. Another variable strategy that has been recently studied is taking advantage of the FcRn-Mediated Recycling Mechanism to increase longevity in contrast to actually modifying the protein itself. Fc fusion proteins have successfully extended the half-lives of therapeutic proteins by utilizing the bivalency of the molecule to improve bioactivity and prolong half-life, and overall utilizes the natural recycling pathway of the FcRn mechanism. In this study, genetic fusion was also utilized to modify the extracellular region of specific proteins, influencing binding with FcRn and increasing longevity as a result [8].

Among the strategies above, PTMs have been the most common and widespread method of furthering PPT efficiency and effectiveness. Specifically, PTMs refer to the addition of chemical groups or small molecules to the amino acids of proteins to modify their structure, function, or some other property [9]. Commonly used PTMs are not limited to just PEGylation and glycosylation; they also include methylation, phosphorylation, acetylation, carboxylation, and others [9]. PTMs happen in the body through enzymatic cleavage and attachments, and similar enzymes are also leveraged in large-scale modification in the protein therapeutics industry.

Although there is significant research on the abilities of certain PTMs, such as phosphorylation, acetylation, and methylation to affect protein behavior in the bloodstream, the results cannot be generalized to many different proteins. For example, phosphorylation at the S96 residue of PolyQ androgen receptor proteins by CDK2 has been found to promote aggregation and subsequent degradation,

while phosphorylation of polyQ huntingtin proteins by Akt has been found to do the opposite [10]. Similarly, methylation has been proven in several studies to both activate and inactivate degradation-mediating regions called degrons in different proteins [10]. Clearly there is wide variability and ambiguity in the effects of PTMs on proteins as each has its own unique structure and function; some modifications or delivery strategies listed above are beneficial for some proteins but not others. Regarding this, our study explores the effects of three common PTMs on three common PPTs each, as explained in further detail in the next section. In addition, research has been done on specific PTMs and methods of addition, but not specifically on the effects on aggregation and longevity in varying PPTs, which our study seeks to measure. How much does each modification affect the propensity of each protein to aggregate during storage or injection? How well is biological function preserved with these modifications? Does each modification confer a net advantage to a specific PPT protein? Answers to these questions can lead to individualized information on aggregation, functional preservation, and longevity for different therapeutic proteins and can be used to analyze trends between PTM methods. They can also prevent life-threatening side-effects of *in vivo* immune responses to aggregation. We hypothesized that PTMs will cause a change in the longevity of the protein by affecting their affinity to aggregate, and that PTMs will also affect biological function and intermolecular interactions by altering their structure.

Our research will be relevant for pharmaceutical companies developing new protein-based drugs and therapies, and for consumers seeking safe treatments. It can provide comprehensive recommendations and insights into optimal combinations of PPTs and PTMs to make currently existing drugs more effective and lower the number of doses that patients need to take in a period of time. This reduces the overall cost of biotherapeutic proteins and also brings experimental proteins that were previously not used due to low half-life or high aggregation to the market, making them available for pharmaceutical applications while simultaneously putting consumers and healthcare providers at ease.

2. Methodology

2.1. Featured Proteins

2.1.1. Insulin & Insulin Receptor

Insulin is a peptide hormone that stimulates uptake of glucose by cells to reduce blood sugar levels and is commonly used for patients suffering with both type 1 and type 2 diabetes to replenish low levels of naturally produced insulin in the body [11]. Insulin is a prominent therapeutic protein. However, synthetic insulin has been shown to aggregate at sites of repeated injection as well as during the production, transportation, and storage process [12], highlighting a need to engineer a mechanism that can reduce insulin aggregation. For our study, we used a crystallized human insulin model [13] (PDB ID: 3I40) (**Appendix A**).

Human insulin binds to insulin receptors present in the membranes of liver and

muscle cells. We modeled the receptor-ligand interaction using the first three domains of an insulin receptor expressed through *C. griseus* [14] (PDB ID: 2HR7).

2.1.2. Erythropoietin (EPO) & EPO Receptor

Erythropoietin is a glycoprotein hormone that stimulates production of red blood cells (erythropoiesis). Recombinant human erythropoietin (rhEPO) is often administered to patients suffering from anemia or those with low hematocrit due to infection or kidney disease [15]. A previous study found that aggregation of rhEPO results in conformational changes that greatly affect its activity and function [16]. By exploring the effects of PTMs on harmful aggregation of rhEPO, we hope to increase its efficacy. We used a human EPO model expressed through *E. coli* in our study [17] (PDB ID: 1BUY) (**Appendix A**).

EPO and rhEPO bind to receptors located on erythroid progenitors in the bone marrow [15]. We modeled this receptor using the extracellular domain of the human EPO receptor [18] (PDB ID: 1ERN).

2.1.3. Human Growth Hormone (HGH) & HGH Receptor

Human growth hormone is a peptide hormone that first received approval by the FDA to treat GH deficiency in children in 1985 [19]. Previous studies conducted by Fradkin *et al.* found that aggregation of growth hormone existing in commercial formulas of the drug stimulated increased levels of immunogenicity in mouse models [20]. To avoid potential side effects of this increased immune response, such as decreased therapeutic half-life and anaphylaxis in the patient [21], methods to reduce aggregation must be found. We used a wild-type HGH model in our study [22] (PDB ID: 1HGU) (**Appendix A**).

To model the receptor-ligand interaction between HGH and its receptor, we used chain B of an HGH-receptor complex [23] (PDB ID: 1A22), which corresponded to just the receptor domain of the modeled complex.

2.2. Featured Post-Translational Modifications

2.2.1. Phosphorylation

Protein phosphorylation refers to the reversible attachment of a phosphate group by a protein kinase (attachment) or phosphatase (detachment) to a side chain of an amino acid. The side chains of serine, threonine, and tyrosine are usually the most commonly phosphorylated. Phosphorylation has the ability to alter the function, structure, and stability of a protein [24], making it an intriguing and ideal PTM to analyze in this study.

2.2.2. Acetylation

Protein acetylation is a type of acylation, which is the addition of an acyl group to a protein's amino acid residues by acyltransferases (attachment) and deacylases (detachment) [25]. Acetylation commonly occurs at serine and lysine residues, and has been implicated in the regulation of protein stability, localization, activity, and affinity for DNA-binding [25]. Because of the ubiquitous functions of

acetylation, many metabolic enzymes are tightly regulated with the addition or removal of acetyl groups [26]. The many abilities of acetylation also make it a valid PTM to analyze in this study.

2.2.3. Methylation

Protein methylation, catalyzed by methyltransferase (attachment) and demethylase (detachment), is the addition of a methyl group to amino acid residues—especially the lysines and arginines—of a protein [27]. Methylation can affect the function and conformation of molecules, making it ideal for modulating cell signaling and DNA repair pathways [27]. The implications of methylation on the characteristics of proteins prompted us to analyze it in this study.

2.3. Materials

We extracted PDB (protein data bank) format files of proteins for insulin, EPO, and HGH from the protein database RCSB.org, an open-source software containing numerous different protein structures and their relevant conformations [28] [29]. We identified the most recently uploaded files for and filtered searches for proteins classified as hormones originating from *Homo sapiens*, allowing us to get the most accurate representation of hormones *in vivo*. In addition, the common receptors for each of these proteins present in the human body were also obtained from the database. In cases where RCSB contained protein-receptor complexes, only the specific chain corresponding to the receptor itself was used.

2.4. Applying Post-Translational Modifications

We utilized an open-source server called Vienna-PTM 2.0 to apply post-translational modifications to the protein files we downloaded from RCSB.org. Vienna-PTM allows users to apply modifications to certain amino acids in uploaded PDB files [30]-[32]. Users can obtain new PDB files and force field parameters for a modified protein that can be used in molecular dynamics simulations such as GROMACS.

For our study, we selected two amino acids that are commonly modified for each group, and modified all instances of those amino acids in each protein to the most neutrally-charged relevant modification in Vienna-PTM to gauge the full effect of these PTMs. The amino acids modified in each group were chosen based on previously conducted studies (see **section 2.2**) and are detailed in **Table 1**.

Table 1. Specific modifications for each PTM.

Group	Modification 1	Modification 2
Phosphorylated	Serine → Phosphoserine (-1)	Threonine → Phosphothreonine (-1)
Acetylated	Serine → Serine-O-acetylglucosamine, N-acetyllysine	Lysine → N-acetyllysine
Methylated	Lysine → Methyllysine (0)	Arginine → Omega-N-methylarginine (0)

After modifying all specified amino acids, we exported and downloaded a new PDB file with the modified protein for analysis. This process was done for all experimentally modified insulin, EPO, and HGH files (See **Appendix A**).

2.5. Characterizing Aggregation

We set out to characterize the affinity for these proteins to aggregate with each other by calculating the DS and BFE between two identical copies of the protein in question. To calculate these values, we used a web server called HawkDock, which specializes in structural prediction and analysis of protein-protein complexes using ATTRACT for global macromolecular docking and HawkRank for scoring [33]-[36]. It takes in one PDB file as a receptor and one PDB file as a ligand. It then outputs the ten most probable complexes, ranked by their docking scores. Docking scores are commonly used by scoring functions of algorithms to represent ligand binding affinities. A more negative DS generally correlates to stronger intermolecular interaction [37]. In the context of our study, we seek to maximize DS and bring it closer to zero, since we are aiming for weaker interactions between proteins and therefore less aggregation.

For our project, we inputted two PDB files of the same, identical protein we were analyzing during that trial into both the receptor and ligand fields of HawkDock. This way, we were able to simulate real interactions between identical proteins administered into the bloodstream through a single injection. The average of the ten complexes was taken in order to equally represent the variability in bonding that may occur in the natural world.

We then ran an MM/GBSA analysis, which is available on HawkDock, on the modeled complexes to obtain BFE values for each corresponding model in kJ/mol [38]-[40]. Binding free energy is a more well-established and practical measure of binding affinity; it quantifies the free energy difference between the bound and unbound states of a complex, often in kJ/mol of complex [41]. Similar to DS, more negative values mean the bound state of a complex is preferred; therefore, we aim to maximize BFE. Additionally, MM/GBSA analysis is a commonly used method to calculate ligand binding affinities computationally, since it does not require large amounts of calculations and boasts remarkable accuracy compared to experimental data [42].

2.6. Analyzing Continuities in Biological Function

In order to minimize loss of function for these proteins from the PTMs, we analyzed the changes in modified proteins' affinities for their native receptors.

We uploaded the respective receptor in the receptor field of HawkDock and the normal or modified protein we were analyzing in the ligand field, and recorded the top model. The top model, which is the most probable, is most likely to be the true receptor-ligand configuration *in vivo*. This assumption was also supported by the fact that the top model had a much lower docking score and binding energy than subsequent, lower-ranked models. Using this methodology, we were able to

add an additional factor to our study's consideration by weighing the change in biological function when making recommendations for most optimal PTMs.

Our overall method design consolidates all of these data points and is summarized in **Figure 2**.

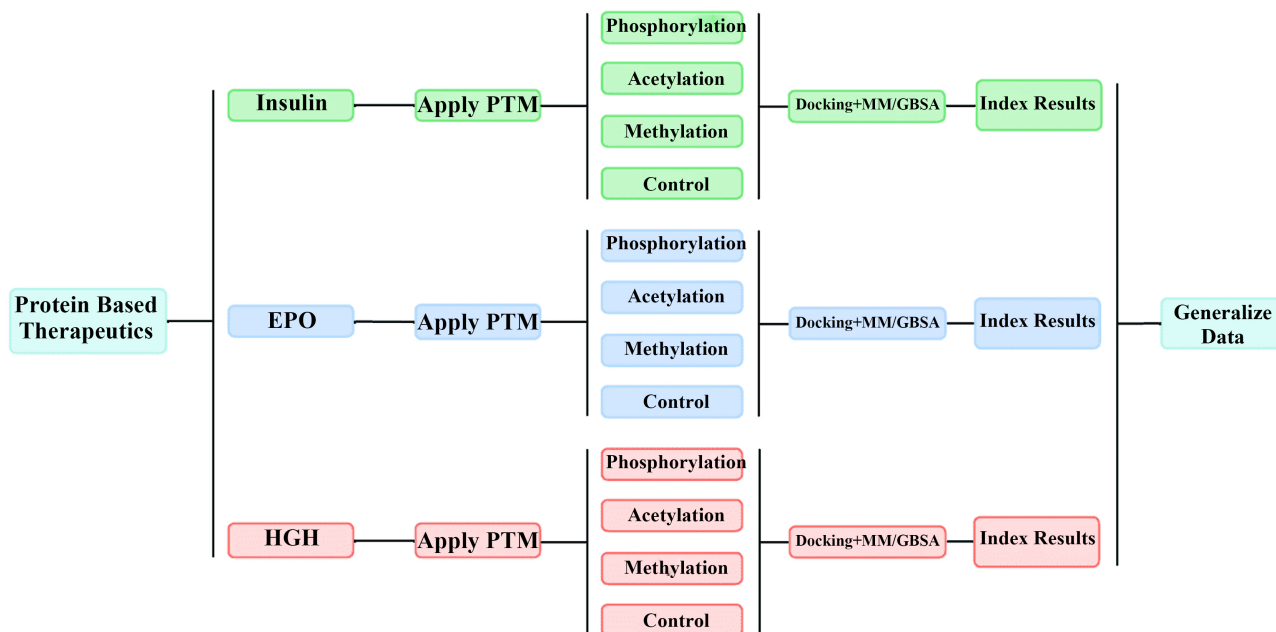


Figure 2. Experimental design flowchart diagram.

3. Results and Data Analysis

3.1. Summary of Results

Since it is difficult to compare improvements in overall protein function (taking into account both aggregation and receptor affinity) without a numerical formula, we will mainly summarize improvements in the reduction of aggregation and comment on significant differences in receptor affinity (functional continuity). For a comparative, numerical analysis between modifications taking both aggregation and reception into account, refer to **Section 3.3**.

The average DS and BFE values for the top 10 models outputted by HawkDock for aggregation are summarized in **Table 2** and **Figure 3(a)** & **Figure 3(c)**, and the DS and BFE values for the top (most optimal) model outputted by HawkDock for receptor-ligand interactions are summarized in **Table 3** and **Figure 3 (b)** & **Figure 3(d)**.

Table 2. Average docking scores and binding free energy values for the top 10 aggregation models outputted by HawkDock.

AGG	Normal (DS)	Normal (BFE)	P (DS)	P (BFE)	A (DS)	A (BFE)	M (DS)	M (BFE)
Insulin	-2327.37	-26.78	-2064.41	-12.20	-2176.57	-18.59	-2097.45	-17.65

Continued

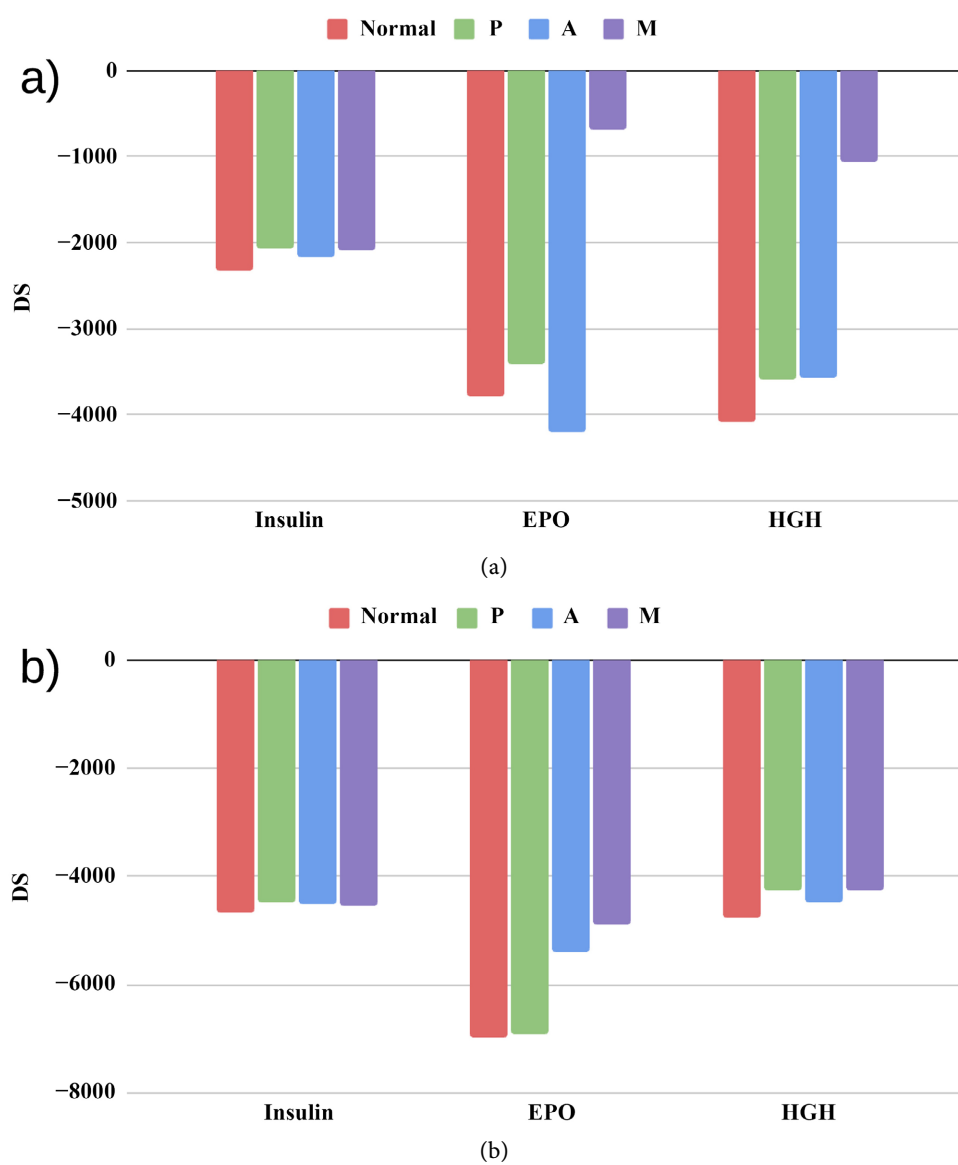
EPO	-3790.09	-23.23	-3412.15	-11.41	-4193.08	-30.24	-684.33	32.86
HGH	-4086.31	-26.12	-3588.18	-10.39	-3570.17	11.38	-1053.03	46.70

More positive is more beneficial.

Table 3. Average docking scores and binding free energy values for the top 10 reception models outputted by HawkDock.

REC	Normal (DS)	Normal (BFE)	P (DS)	P (BFE)	A (DS)	A (BFE)	M (DS)	M (BFE)
Insulin	-4669.43	-7.64	-4475.47	-18.64	-4496.3	-1.75	-4533.07	-28.34
EPO	-6980.32	-43.94	-6905.43	-17.52	-5405.01	-27.3	-4888.91	-9.58
HGH	-4748.43	-16.94	-4250.95	-12.12	-4463.7	1.94	-4248.69	-13.25

More negative is more beneficial.



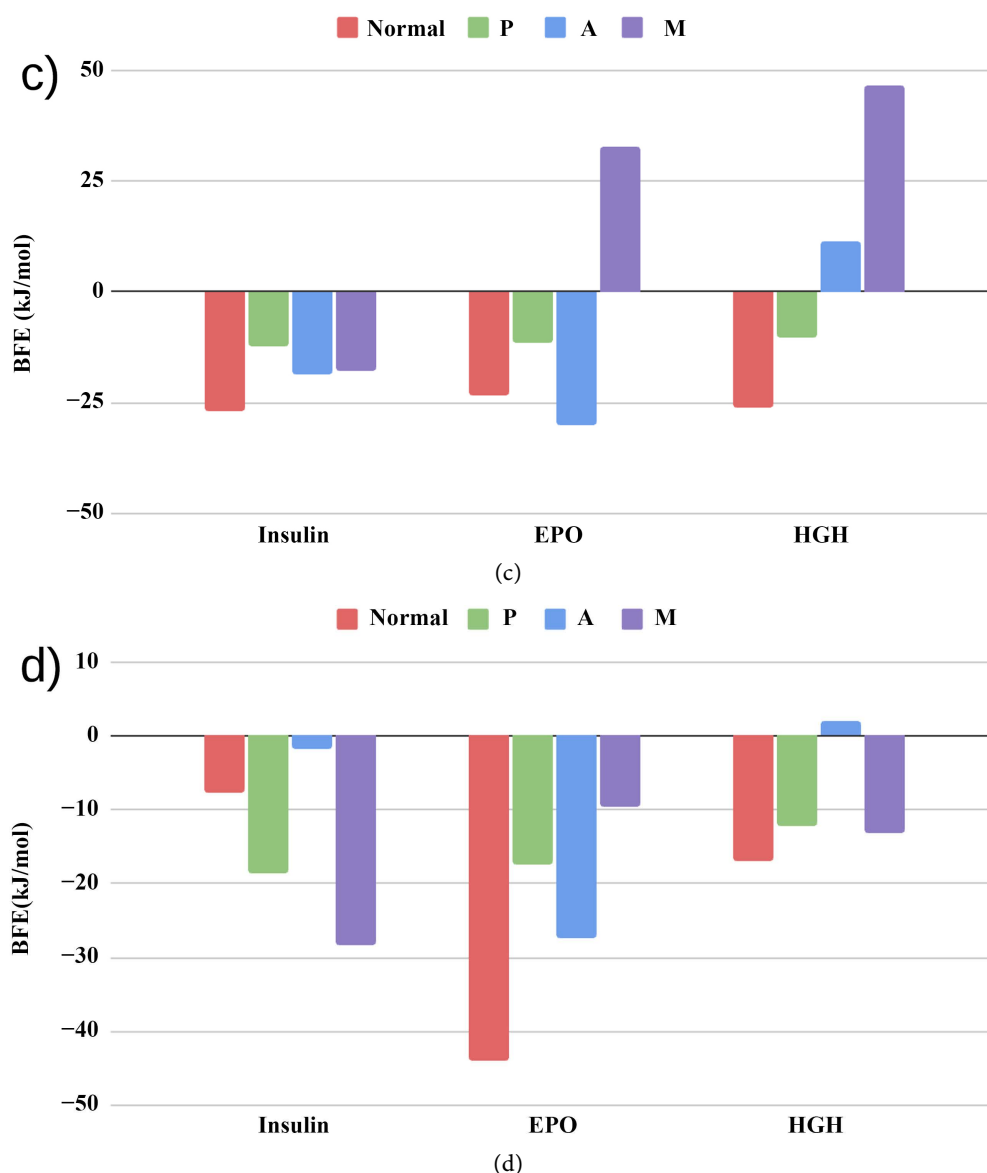


Figure 3. (a) Average docking scores for aggregation. (b) Average docking scores for reception. (c) Average BFE values (in kJ/mol) for aggregation. (d) Average BFE values (in kJ/mol) for reception.

3.2. Statistical Significance between Control and Experimental Groups

To prove statistical significance between the normal (control) group and the experimental (phosphorylated, acetylated, methylated) groups for each protein, we used a single factor ANOVA test conducted through Google Sheets and the “XLMiner Analysis ToolPak” extension.

A single factor ANOVA (analysis of variance) test is a generalization of the two-sample t-test and calculates how much of the variance or discrepancy between data can be attributed to random error or the factor effect [43]. It outputs an F-statistic and P-value. In our study, we used the P-value at a significance level of 0.05 to indicate significance. A lower P-value, in this case $p < 0.05$, means the

variance is likely not due to random chance and instead due to our treatment. Although we can see at a clear glance that the mean DS and BFEs are higher or lower between experimental groups, we applied this test to prove our findings were significant and identify which PTMs on which proteins had significant effects on their aggregation affinity. We also ran an overall ANOVA test with all four groups for each protein to show that the three PTMs we explored caused significant results in general. Since all overall P-values were significant at $\alpha = 0.05$, we rejected our null hypothesis. We did not run an ANOVA test on the receptor affinities since we only had one data point for each group. The P-values on the ANOVA analyses of aggregation affinity data between modified and unmodified proteins are shown in **Table 4** and **Table 5**.

Table 4. P-values for ANOVA tests on docking scores.

DS	Normal x P	Normal x A	Normal x M	Overall
Insulin	0.0121*	0.0813	0.0641	0.0350*
EPO	0.0545	0.0194*	0.0000*	0.0000*
HGH	0.0349*	0.0474*	0.0000*	0.0000*

*Statistically significant ($p < 0.05$).

Table 5. P-values for ANOVA tests on binding free energies.

BFE	Normal x P	Normal x A	Normal x M	Overall
Insulin	0.0028*	0.0861	0.0739	0.0151*
EPO	0.0610	0.1793*	0.0000*	0.0000*
HGH	0.0263*	0.0000*	0.0000*	0.0000*

*Statistically significant ($p < 0.05$).

3.3. Creation of an Index to Assess Protein Viability

In order to quantitatively and specifically measure the improvement or deterioration of each PTM treatment, we created a protein viability index equation by modifying a sigmoid function using the base aggregation and reception values from the normal protein, and the aggregation and reception values from the modified protein:

$$\frac{2}{1 + e^{-\left(0.25 \frac{Value_{agg} - Base_{agg}}{|Base_{agg}|} + 0.75 \frac{Value_{rec} - Base_{rec}}{|Base_{rec}|}\right)}} - 1 \quad (1)$$

The process taken to create the index was to first have a baseline value of comparison for each protein, which was the normal DS and BFE in aggregation and reception for an unmodified protein (DS and BFE are calculated in separate programs, using their specific base values for each protein). Using the base values for a protein, the post-translationally modified protein's average DS or BFE values

were inputted for aggregation and reception for that same protein. By taking the difference between the experimental value and the unmodified base value and dividing by the unmodified base value, the experimental value is normalized for that specific protein which allows other transformations to be put upon the data. A weight of 0.25 was used for the aggregation aspect, and 0.75 was used for the reception aspect. This 1:3 ratio between reception values and aggregation values represents the greater importance of the reception values, exhibiting how the original function of a protein staying constant is not something that can be changed while increasing protein longevity. Changes in the function of the protein as a whole causes longevity to become irrelevant in this setting and therefore must be prioritized in the index. The ratio chosen can vary depending on the difference in importance, however this is something that can be more studied using wet-lab experiments to study how PTMs can change the function of a protein more accurately.

After the normalized and weighted values were calculated, they were put into a sigmoid-shaped logistic function that was modified to output values from -1 to 1 in Python (see **Appendix B**). The logistic function transforms the given data around a midpoint, where the instantaneous rate of change is a vertical line, to a specified range. It is commonly used in data analysis. Here, the point of inflection was 0 , representing no change, and the limits of the left and right sides of the function approached -1 and 1 , respectively. A positive output represented an increase in the overall performance of the modified protein given the weighting of aggregation and reception, and a negative output represented a decrease in the performance of the modified protein. The normal insulin, EPO, and HGH proteins can automatically be assigned a value of 0 on the index as they are the baseline models. The values provided by this index is used as a quantitative measure to determine the change in the DS and BFE caused by the PTM and allows us to objectively and numerically compare the differences in differently modified proteins (see **Appendix B** for index source code). The index values are shown in **Table 6**, **Table 7**, and **Figure 4**.

4. Discussion

In general, our results show that PTMs do have a significant effect on the aggregation and longevity of therapeutic proteins, with all overall P-values < 0.05 and some overall P-values < 0.0001 .

Table 6. Docking score index values for PTMs on each protein.

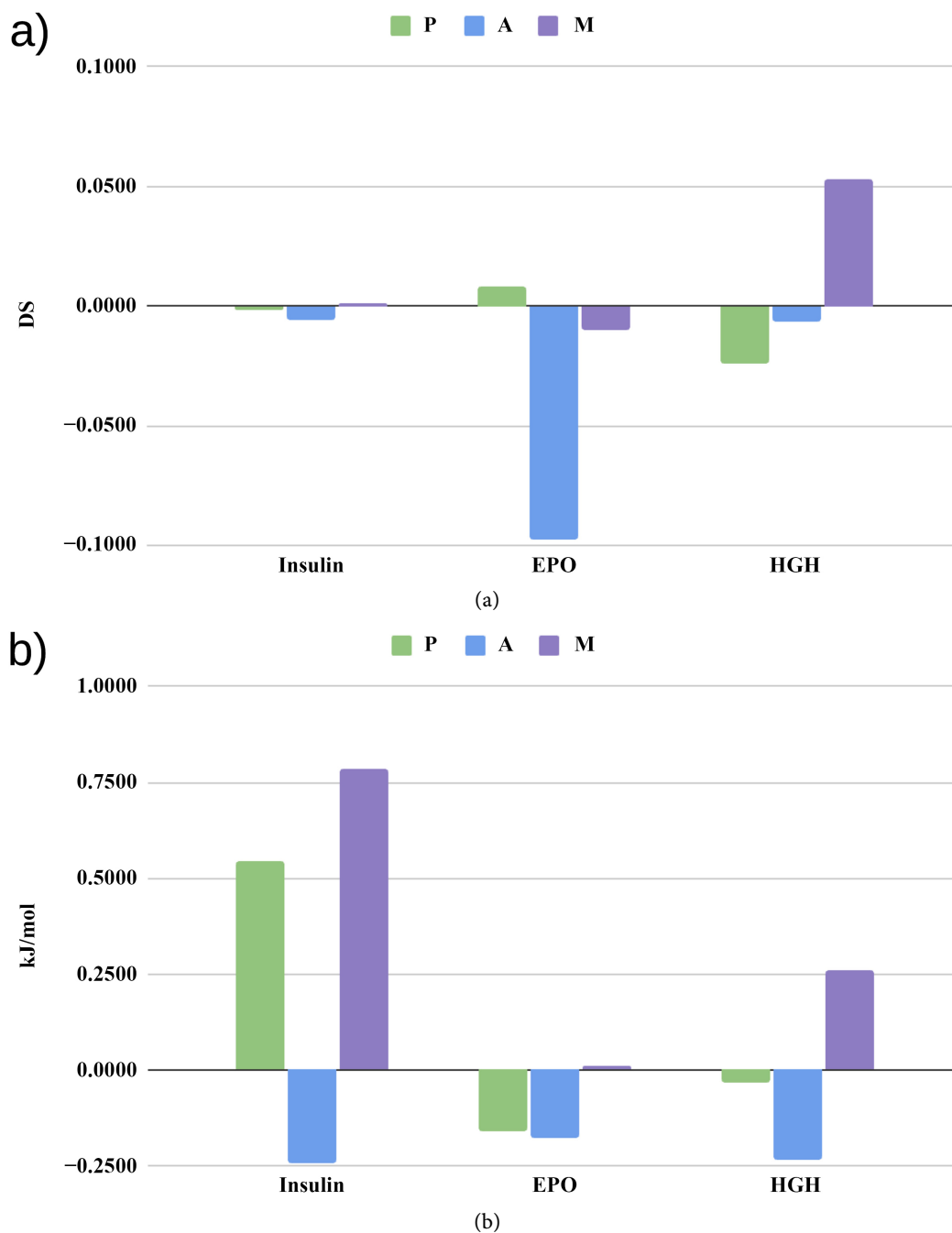
DS	P	A	M
Insulin	-0.0015	-0.0058	0.0014^*
EPO	0.0084^*	-0.0976	-0.0099
HGH	-0.0240	-0.0067	0.0533^*

*Positive, beneficial change (value > 0).

Table 7. Binding free energy index values for PTMs on each protein.

BFEE	P	A	M
Insulin	0.5427*	-0.2457	0.7851*
EPO	-0.1605	-0.1778	0.0085*
HGH	-0.0314	-0.2341	0.2606*

*Positive, beneficial change (value > 0).

**Figure 4.** An index of 0 signifies no change and serves as a baseline for comparison. (a) Docking score index values. (b) Binding free energy index values.

For insulin, phosphorylation had the most beneficial effect on DS and BFE for aggregation (significant with $p < 0.05$, $p < 0.01$ respectively). However, for receptor affinity, it did result in a slight increase in DS while BFE decreased, suggesting ambiguity for phosphorylation's effect in terms of functional continuity. With the objective index, phosphorylation had a substantially positive score for DS, but a slightly negative score for BFE, ranking it second best behind methylation which had substantially positive scores for both DS and BFE. Surprisingly, in terms of aggregation, methylation was beneficial but not as much as phosphorylation, but it received the highest DS and BFE index scores when incorporating receptor affinity. Acetylation was the least beneficial and resulted in slightly negative index scores for both DS and BFE, signaling that it may actually have detrimental effects on insulin function and half-life, although the differences were not statistically significant.

In terms of EPO, methylation conferred a significant increase in DS and BFE ($p = 0$ for both), demonstrating that it had the greatest effect in reducing aggregation and subsequently increasing half-life. In fact, BFE actually became positive (32.86 kJ/mol). It did, however, also have the most negative effect on receptor affinity. Therefore, index scores for methylation were close to 0, with one positive and one negative. Phosphorylation was the next best for aggregation and also had one positive and one negative index, putting it in similar standing with methylation in terms of overall benefit. Acetylation had the worst indexes, with an extremely negative DS score and a negative BFE score (significant with $p < 0.05$). Both phosphorylation and acetylation conferred more moderate effects on aggregation and receptor affinity than methylation.

Finally, with HGH, methylation once again had the most beneficial effect on aggregation by increasing DS and BFE ($p = 0$ for both). It also resulted in just slight decreases in receptor affinity, giving it positive index scores for both DS and BFE and making it the most beneficial PTM. Phosphorylation and acetylation had less drastic results compared to methylation for aggregation, with phosphorylation having a slightly higher DS than acetylation. However, it had a much lower BFE compared to acetylation for aggregation (-10.392 kJ/mol vs. 11.381 kJ/mol), and results from both phosphorylation and acetylation were significant ($p < 0.05$ for both). Using the index, phosphorylation seems to rank higher than acetylation since it has a greatly less negative score for BFE. All modifications resulted in little change to receptor affinity. The only notable difference is that acetylation resulted in a positive BFE for receptor affinity, suggesting that it improves receptor-ligand interactions slightly.

Interestingly, even though we cannot generalize due to the small scope of this experiment, acetylation seemed to consistently confer the least benefit to therapeutic proteins; in fact, all six of its index scores were negative. This may be due to the charged nature of the acetyl groups that may have made the proteins more negative and therefore promoted polar or charged interactions and subsequent aggregation. Additionally, methylation greatly reduced aggregation (by increasing

DS and BFE) in EPO and HGH. We infer that this may be due to the nonpolar properties of the group interfering with aggregation. In general, phosphorylation seemed to have moderate effects. The varying beneficial and detrimental effects of phosphorylation and methylation in particular between the three proteins supports Lee *et al.*'s findings regarding the ambiguity and variability in the effects of PTMs on both related and unrelated protein groups [10].

For some proteins, there was a discrepancy between DS and BFE (i.e. DS increased but BFE decreased). This is still a valid result since the definitions of DS and BFE are slightly different, since the DS takes into account the structure of the complex and the intermolecular interactions while the BFE takes into account specifically the free energy differences between the bound and unbound states of the complex. In other words, DS incorporated BFE into its calculation and is a more generalized value regarding the protein interactions.

5. Conclusions

In this study, we analyzed the docking scores and free energy values of post-translationally modified proteins bound to themselves and their specific receptors. Using this information, we determined which post-translational modifications were best for insulin, erythropoietin, and human growth hormone by comparing the change in aggregation and reception values using a normalized index that we created. We found methylation to be the most beneficial for insulin and HGH, and both phosphorylation and methylation to be somewhat optimal for EPO. Post-translational modifications have been a common field of research in recent times, however existing research had the issue of not explicitly comparing the changes in protein-protein interaction between different PTMs on different proteins. This study, although done on a smaller scale, allowed the impacts of PTMs to be quantified and systematically compared, where different PTMs can be seen to have a different effect on each different protein.

Some future directions of this research include expanding the scope of the study to include more commonly used therapeutic proteins, such as clotting factor XIII and interferons, to explore the effects of PTMs on more proteins' longevity. We can also explore and analyze more PTMs beyond the three used in this study; for example, carboxylation, glycosylation, and PEGylation are other commonly used PTMs in the industry. Combined with future research dealing with a larger pool of proteins and PTMs, as well as a more sophisticated and wet-lab test of the relation between aggregation versus reception to get better index values, this information can be used to create a large-scale database of PTM impacts on different protein biotherapeutics. A resource of such scale would be an essential piece of information for use in pharmaceutical and medical fields to modify existing drug treatments and improve the world of drug development.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Dimitrov, D.S. (2012) Therapeutic Proteins. In: Voynov, V. and Caravella, J.A., Eds., *Therapeutic Proteins: Methods and Protocols*, Humana Press, 1-26. https://doi.org/10.1007/978-1-61779-921-1_1
- [2] Roberts, C.J. (2014) Therapeutic Protein Aggregation: Mechanisms, Design, and Control. *Trends in Biotechnology*, **32**, 372-380. <https://doi.org/10.1016/j.tibtech.2014.05.005>
- [3] Pisal, D.S., Kosloski, M.P. and Balu-Iyer, S.V. (2010) Delivery of Therapeutic Proteins. *Journal of Pharmaceutical Sciences*, **99**, 2557-2575. <https://doi.org/10.1002/jps.22054>
- [4] Zaman, R., Islam, R.A., Ibnat, N., Othman, I., Zaini, A., Lee, C.Y., *et al.* (2019) Current Strategies in Extending Half-Lives of Therapeutic Proteins. *Journal of Controlled Release*, **301**, 176-189. <https://doi.org/10.1016/j.jconrel.2019.02.016>
- [5] Conibear, A.C. (2020) Deciphering Protein Post-Translational Modifications Using Chemical Biology Tools. *Nature Reviews Chemistry*, **4**, 674-695. <https://doi.org/10.1038/s41570-020-00223-8>
- [6] Su, Y., Zhang, B., Sun, R., Liu, W., Zhu, Q., Zhang, X., *et al.* (2021) PLGA-Based Biodegradable Microspheres in Drug Delivery: Recent Advances in Research and Application. *Drug Delivery*, **28**, 1397-1418. <https://doi.org/10.1080/10717544.2021.1938756>
- [7] Fu, C., Chen, Q., Zheng, F., Yang, L., Li, H., Zhao, Q., *et al.* (2018) Genetically Encoding a Lipidated Amino Acid for Extension of Protein Half-Life *in Vivo*. *Angewandte Chemie International Edition*, **58**, 1392-1396. <https://doi.org/10.1002/anie.201811837>
- [8] Tan, H., Su, W., Zhang, W., Wang, P., Sattler, M. and Zou, P. (2019) Recent Advances in Half-Life Extension Strategies for Therapeutic Peptides and Proteins. *Current Pharmaceutical Design*, **24**, 4932-4946. <https://doi.org/10.2174/1381612825666190206105232>
- [9] Zhong, Q., Xiao, X., Qiu, Y., Xu, Z., Chen, C., Chong, B., *et al.* (2023) Protein Post-translational Modifications in Health and Diseases: Functions, Regulatory Mechanisms, and Therapeutic Implications. *MedComm*, **4**, e261. <https://doi.org/10.1002/mco2.261>
- [10] Lee, J.M., Hammarén, H.M., Savitski, M.M. and Baek, S.H. (2023) Control of Protein Stability by Post-Translational Modifications. *Nature Communications*, **14**, Article No. 201. <https://doi.org/10.1038/s41467-023-35795-8>
- [11] Landgraf, W. and Sandow, J. (2016) Recombinant Human Insulins—Clinical Efficacy and Safety in Diabetes Therapy. *European Endocrinology*, **12**, 12-17.
- [12] Das, A., Shah, M. and Saraogi, I. (2022) Molecular Aspects of Insulin Aggregation and Various Therapeutic Interventions. *ACS Bio & Med Chem Au*, **2**, 205-221. <https://doi.org/10.1021/acsbiochemau.1c00054>
- [13] Timofeev, V.I., Chuprov-Netochin, R.N., Samigina, V.R., Bezuglov, V.V., Miroshnikov, K.A. and Kuranova, I.P. (2010) X-Ray Investigation of Gene-Engineered Human Insulin Crystallized from a Solution Containing Polysialic Acid. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, **66**, 259-263. <https://doi.org/10.1107/s1744309110000461>
- [14] Lou, M., Garrett, T.P.J., McKern, N.M., Hoyne, P.A., Epa, V.C., Bentley, J.D., *et al.* (2006) The First Three Domains of the Insulin Receptor Differ Structurally from the Insulin-Like Growth Factor 1 Receptor in the Regions Governing Ligand Specificity.

- Proceedings of the National Academy of Sciences*, **103**, 12429-12434.
<https://doi.org/10.1073/pnas.0605395103>
- [15] Jacob, J., John, M., Jaison, V., Jain, K. and Kakkar, N. (2012) Erythropoietin Use and Abuse. *Indian Journal of Endocrinology and Metabolism*, **16**, 220-227.
<https://doi.org/10.4103/2230-8210.93739>
- [16] Ghezlou, M., Mokhtari, F., Kalbasi, A., Riazi, G., Kaghazian, H., Emadi, R., *et al.* (2020) Aggregate Forms of Recombinant Human Erythropoietin with Different Charge Profile Substantially Impact Biological Activities. *Journal of Pharmaceutical Sciences*, **109**, 277-283. <https://doi.org/10.1016/j.xphs.2019.05.036>
- [17] Cheetham, J.C., Smith, D.M., Aoki, K.H., Stevenson, J.L., Hoeffel, T.J., Syed, R.S., *et al.* (1998) NMR Structure of Human Erythropoietin and a Comparison with Its Receptor Bound Conformation. *Nature Structural Biology*, **5**, 861-866.
<https://doi.org/10.1038/2302>
- [18] Livnah, O., Stura, E.A., Middleton, S.A., Johnson, D.L., Jolliffe, L.K. and Wilson, I.A. (1999) Crystallographic Evidence for Preformed Dimers of Erythropoietin Receptor before Ligand Activation. *Science*, **283**, 987-990.
<https://doi.org/10.1126/science.283.5404.987>
- [19] Danowitz, M. and Grimberg, A. (2022) Clinical Indications for Growth Hormone Therapy. *Advances in Pediatrics*, **69**, 203-217.
<https://doi.org/10.1016/j.yapd.2022.03.005>
- [20] Fradkin, A.H., Carpenter, J.F. and Randolph, T.W. (2009) Immunogenicity of Aggregates of Recombinant Human Growth Hormone in Mouse Models. *Journal of Pharmaceutical Sciences*, **98**, 3247-3264. <https://doi.org/10.1002/jps.21834>
- [21] Kuriakose, A., Chirmule, N. and Nair, P. (2016) Immunogenicity of Biotherapeutics: Causes and Association with Posttranslational Modifications. *Journal of Immunology Research*, **2016**, Article ID: 1298473. <https://doi.org/10.1155/2016/1298473>
- [22] Chantalat, L., Jones, N.D., Korber, F., Navaza, J. and Pavlovsky, A.G. (1995) The Crystal Structure of Wild-Type Growth Hormone at 2.5 Å Resolution. *Protein & Peptide Letters*, **2**, 333-340. <https://doi.org/10.2174/092986650202220524124754>
- [23] Clackson, T., Ultsch, M.H., Wells, J.A. and de Vos, A.M. (1998) Structural and Functional Analysis of the 1:1 Growth Hormone: Receptor Complex Reveals the Molecular Basis for Receptor Affinity. *Journal of Molecular Biology*, **277**, 1111-1128.
<https://doi.org/10.1006/jmbi.1998.1669>
- [24] Seok, S. (2021) Structural Insights into Protein Regulation by Phosphorylation and Substrate Recognition of Protein Kinases/Phosphatases. *Life*, **11**, Article No. 957.
<https://doi.org/10.3390/life11090957>
- [25] Shang, S., Liu, J. and Hua, F. (2022) Protein Acylation: Mechanisms, Biological Functions and Therapeutic Targets. *Signal Transduction and Targeted Therapy*, **7**, Article No. 396. <https://doi.org/10.1038/s41392-022-01245-y>
- [26] Christensen, D.G., Xie, X., Basisty, N., Byrnes, J., McSweeney, S., Schilling, B., *et al.* (2019) Post-Translational Protein Acetylation: An Elegant Mechanism for Bacteria to Dynamically Regulate Metabolic Functions. *Frontiers in Microbiology*, **10**, Article No. 1604. <https://doi.org/10.3389/fmicb.2019.01604>
- [27] Clarke, S.G. (2013) Protein Methylation at the Surface and Buried Deep: Thinking Outside the Histone Box. *Trends in Biochemical Sciences*, **38**, 243-252.
<https://doi.org/10.1016/j.tibs.2013.02.004>
- [28] Berman, H.M. (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**, 235-242.
<https://doi.org/10.1093/nar/28.1.235>

- [29] Berman, H., Henrick, K. and Nakamura, H. (2003) Announcing the Worldwide Protein Data Bank. *Nature Structural & Molecular Biology*, **10**, 980-980. <https://doi.org/10.1038/nsb1203-980>
- [30] Margreitter, C., Petrov, D. and Zagrovic, B. (2013) Vienna-PTM Web Server: A Toolkit for MD Simulations of Protein Post-Translational Modifications. *Nucleic Acids Research*, **41**, W422-W426. <https://doi.org/10.1093/nar/gkt416>
- [31] Margreitter, C., Reif, M.M. and Oostenbrink, C. (2017) Update on Phosphate and Charged Post-Translationally Modified Amino Acid Parameters in the GROMOS Force Field. *Journal of Computational Chemistry*, **38**, 714-720. <https://doi.org/10.1002/jcc.24733>
- [32] Petrov, D., Margreitter, C., Grandits, M., Oostenbrink, C. and Zagrovic, B. (2013) A Systematic Framework for Molecular Dynamics Simulations of Protein Post-Translational Modifications. *PLOS Computational Biology*, **9**, e1003154. <https://doi.org/10.1371/journal.pcbi.1003154>
- [33] Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D., *et al.* (2019) Hawkdock: A Web Server to Predict and Analyze the Protein-Protein Complex Based on Computational Docking and MM/GBSA. *Nucleic Acids Research*, **47**, W322-W330. <https://doi.org/10.1093/nar/gkz397>
- [34] Zacharias, M. (2003) Protein-Protein Docking with a Reduced Protein Model Accounting for Side-Chain Flexibility. *Protein Science*, **12**, 1271-1282. <https://doi.org/10.1110/ps.0239303>
- [35] Feng, T., Chen, F., Kang, Y., Sun, H., Liu, H., Li, D., *et al.* (2017) Hawkrank: A New Scoring Function for Protein-Protein Docking Based on Weighted Energy Terms. *Journal of Cheminformatics*, **9**, Article No. 66. <https://doi.org/10.1186/s13321-017-0254-7>
- [36] Hou, T., Qiao, X., Zhang, W. and Xu, X. (2002) Empirical Aqueous Solvation Models Based on Accessible Surface Areas with Implicit Electrostatics. *The Journal of Physical Chemistry B*, **106**, 11295-11304. <https://doi.org/10.1021/jp025595u>
- [37] Kinnings, S.L., Liu, N., Tonge, P.J., Jackson, R.M., Xie, L. and Bourne, P.E. (2011) A Machine Learning-Based Method to Improve Docking Scoring Functions and Its Application to Drug Repurposing. *Journal of Chemical Information and Modeling*, **51**, 408-419. <https://doi.org/10.1021/ci100369f>
- [38] Hou, T., Wang, J., Li, Y. and Wang, W. (2010) Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*, **51**, 69-82. <https://doi.org/10.1021/ci100275a>
- [39] Sun, H., Li, Y., Tian, S., Xu, L. and Hou, T. (2014) Assessing the Performance of MM/PBSA and MM/GBSA Methods. 4. Accuracies of MM/PBSA and MM/GBSA Methodologies Evaluated by Various Simulation Protocols Using Pdbbind Data Set. *Physical Chemistry Chemical Physics*, **16**, 16719-16729. <https://doi.org/10.1039/c4cp01388c>
- [40] Chen, F., Liu, H., Sun, H., Pan, P., Li, Y., Li, D., *et al.* (2016) Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 6. Capability to Predict Protein-protein Binding Free Energies and Re-Rank Binding Poses Generated by Protein-protein Docking. *Physical Chemistry Chemical Physics*, **18**, 22129-22139. <https://doi.org/10.1039/c6cp03670h>
- [41] Hata, H., Phuoc Tran, D., Marzouk Sobeh, M. and Kitao, A. (2021) Binding Free Energy of Protein/Ligand Complexes Calculated Using Dissociation Parallel Cascade Selection Molecular Dynamics and Markov State Model. *Biophysics and Physicobiology*, **18**, 305-316. <https://doi.org/10.2142/biophysico.bppb-v18.037>

- [42] Genheden, S. and Ryde, U. (2015) The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opinion on Drug Discovery*, **10**, 449-461. <https://doi.org/10.1517/17460441.2015.1032936>
- [43] NIST/SEMATECH e-Handbook of Statistical Methods. <http://www.itl.nist.gov/div898/handbook/>

Appendix A

Protein Structure Images

Note: Proteins structures shown below come from the RSCB.org database (**Figures A1(a)-(c)**) and the HawkDock software visual output (**Figures A1(d)-(e)**).

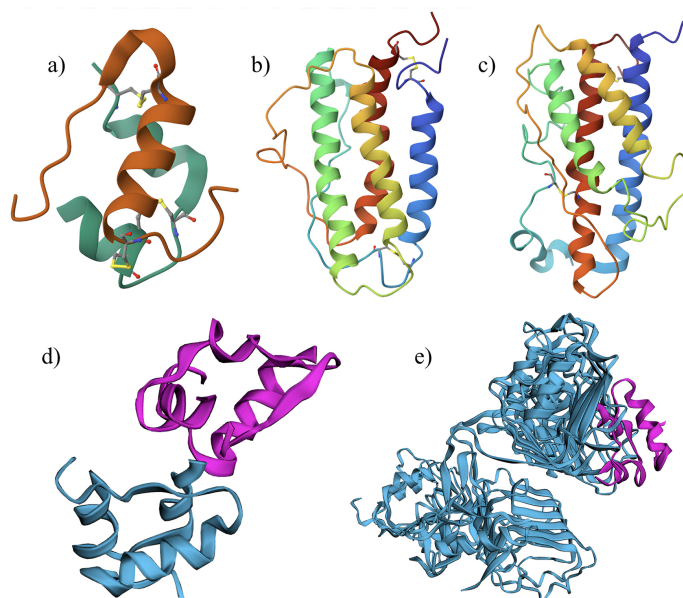


Figure A1. (a) Insulin. (b) Erythropoietin. (c) Human Growth Hormone. (d) Insulin Aggregation Complex. (e) Insulin-Receptor Binding Complex.

Appendix B

Source Code for Index Calculations

Note. A different file was used for each protein and for DSI/BFE which had the baseline values unique to each. The file shown below is hghindex.py, specifically calculating BFE.

```
1 import numpy as np
2
3 def sigmoid(x, k=1, c=0):
4     #Generalizes data on a scale from -1 to 1 using sigmoid mathematical function
5     return (2 * (1 / (1 + np.exp(-k * (x - c))))) - 1
6
7     #Inputted values from experimental groups(variable)
8     #The following shows data from acetylated HGH BFE
9     value_agg = 46.695
10    value_rec = -13.25
11
12    #Baseline values from normal group(constant)
13    #The following shows the baseline values from normal HGH BFE
14    base_agg = -26.124
15    base_rec = -16.94
16
17    #Weighting of aggregation vs reception
18    #Shows a 1:3 ratio between importance of aggregation vs reception
19    weight_agg = 0.25
20    weight_rec = 1 - weight_agg
21
22    def normalize(value, base):
23        #Normalizes data to a constant scale using baseline values
24        return (value - base) / abs(base)
25
26    norm_agg = (normalize(value_agg, base_agg))
27    norm_rec = -1 * (normalize(value_rec, base_rec))
28
29    protein_combined = weight_agg * norm_agg + weight_rec * norm_rec
30
31    #Outputs a final score by inputting normalized values into the sigmoid function
32    final_score = sigmoid(protein_combined)
33
34    print(final_score)
35
```
