Computational prediction of PTM in Biologics

Vinod Devaraji, Reshma Mane, Rujuta Shinde, Raghu Rangaswamy

Abstract— Biopharmaceuticals, also known as biologics, are products manufactured from biological sources and used to diagnose and treat diseases. It consists of vaccines, gene therapies, medicinal protein hormones, monoclonal antibodies, and growth factors. A majority of them are intricate mixtures that are not readily known or distinguished and are highly sensitive to their manufacturing and handling conditions. These are often made by genetically modifying cell constituents or cell lines, requiring stable conditions to remain sturdy and intact, which require enormous cost and resources. Even minor alterations lead to cell behavior changes and differences in the structure, stability, or other quality features of the end product. This kind of product has enormous therapeutic potential. It doesn't have any side effects as it is processed from natural resources. There are many kinds of care taken to retain these biologics' potency. Post-translational modification (PTM) can affect the conformation, stability, and function of the protein. Biologics' Manufacturing and preclinical development is the main challenge for biologists compared to traditional small molecular drugs due to their complex structure. Alternatively, bioanalytical approval standards are available for small molecules. Still, it is time-consuming, frequently intricate, or impossible to get an untainted indication criterion for definite biomarkers and get the usual samples without endogenous biomarkers. As a result, rationale measures are frequently not appropriate for biologics studies. In silico methods, they could come in handy as they are pretty quick and less time-consuming. In this study, computational approaches were used to identify, categorize, and analyze the effects of PTM, particularly the methionine oxidation effect on the conformation, stability, and function. Also, an effort has been made to ease biologics' design and development using a computational approach.

Index Terms— Post Translational Modification, Biologics, Computational Biology, Molecular dynamics, Homology modeling, Docking

1 Introduction

Biopharmaceutical is any biological product manufactured or extracted from natural sources [1]. Blood products, as well as human and animal cells, have been classified as biologics. Modern biologics, however, are defined as biotechnology-derived therapeutics which include recombinant proteins like monoclonal antibodies [2], cytokines [3], and tissue growth factors [4].

Traditionally, biologics are derived from living sources such as humans, animals, cells, and micro-organisms [5]. Chemically, biologics are mostly proteins comprising L-amino acids and various sugar molecules. They are present in the form of three-dimensional structures based on secondary, tertiary, and, in some cases, quaternary structures. Today, biologics such as blood and blood components fractionated blood products. Antitoxins to modern biologics are mostly proteins generated by recombinant DNA technology, such as monoclonal antibodies.we can distinguish growing biologics as products that consist of manipulated, refined, or prolonged human cells, vaccines directed against non-infectious disease targets, and gene transfer products. Some of the biologics available in the market are shown in Table 1:

TABLE 1
SHOWS TOP 10 SELLING BIOLOGICS

Sl. No.	Brand	Drug Name	Sales($bn)
1.	Avastin	Bevacizumab	5.75
2.	Epogen	Epoetin alfa	2.56
3.	Neulasta	Pegfilgrastim	3.35
4.	Remicade	Infliximab	3.51
5.	Lantus	Insulin glargine	4.29
6.	Enbrele	Tanercept	3.87
7.	Herceptin	Trastuzumab	4.86
8.	Avonex	Interferon beta-1a	2.32
9.	Humira	Adalimumab	5.48
10.	Rituxan	Rituximab	5.62

The quality of several biologics is still defined by their manufacturing process. However, recent improvements in analytical methods used for characterizing biologics have brought about a change in the way biologics are analyzed [7]. Hence nowadays, biologics are defined by analytical characterization rather than the manufacturing process during the development of innovative products and production methods that mainly encompass antibodies and cytokines. Nonetheless,

- *Vinod Devaraji, Computational Drug Design Lab, School of BioSciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, India, donniv86@gmail.com*
- *Reshma Mane, Department of Biotechnology, RV College of Engineering, RV Vidyanikethan Post, Mysuru Road, Bengaluru 560059, mane.reshma8@gmail.com*
- *Rujuta Shinde, MIT School of Bioengineering Sciences and Research, MIT ADT University, Pune, Maharashtra, India, rujutaashinde@gmail.com*
- *Raghu Rangaswamy, Department of Bioinformatics, Alagappa University, raghu@schrodinger.com*

generics' concept cannot be applied to biologics as it is used for small molecules [8]. Chemically biologics are represented by proteins. The intricate physiochemical properties of proteins make it difficult to express the same biologics' equality if they are produced by two different processes.

Biologics' clinical development has to be performed, similarly to small molecules, based on a proper balancing of risks and benefits. The preclinical development program supports the safety of the proposed clinical studies. However, standard preclinical testing for small molecules is not always appropriate for biologics [9]. They are often highly species-specific in action and immunogenic in test animal species. The mechanism of toxicity is mostly receptor-mediated, including both exaggerated pharmacodynamic responses and biological toxicity.

In many cases, biologics' pharmacokinetics is not only different but much more complex compared to small molecules. The immunogenicity of protein-based therapeutics may influence their pharmacokinetic behavior as well as pharmacodynamics and toxicity [10]. Bioanalytical method development and validation is a prerequisite not exclusively for pharmacokinetic studies but for the whole preclinical and clinical outcome. Due to their unique properties, different assays such as mass assays, activity assays, immunogenicity assays are necessary for early development.

Modern biologics are mainly utilized for the diagnosis, prevention, and treatment of severe and persistent diseases. The majorities of biologics involve several molecular constituents and are a little bigger than small molecules. They have a complex outline of post-translational modification and frequently a complex pattern of process-related impurities. The common modifications include oxidation, glycosylation, phosphorylation, and proteolysis. In proteins, all amino acid residues are prone to oxidation by different reactive oxygen species, but Methionine, Cysteine, Histidine, Tryptophan & Tyrosine residues are mainly responsive to oxidation [11]. Hence biologics require high levels of solubility as well as withholding of activity throughout purification, formulation, storage, and administration.

The last few decades have observed expeditious growth in the detection and efficient analysis of PTM in proteins, during temporal and spatial alteration of proteins by covalent binding of further chemical groups and small proteins, proteolytic cleavage, or intein splicing, PTM significantly develops the proteome multiplicity and performs vital tasks in modulating the firmness and activities of the proteins [12]. Until now, around 350 types of different PTM were experimentally revealed *in vivo*, while consequently efficient methods have identified several exhilarating findings [13].

PTM act as a biological switch to activate or deactivate molecules by the signal transduction process. Protein phosphorylation is one of the most studied PTM, which accounts for around 30% of all PTM. PhosphoSitePlus is one of the biggest archives of PTM data, which consists of data for different types of PTM of massive proteins from in vivo and in vitro, obtained from various vertebrates and invertebrates [14]. The latest development in mass spectrometry has emerged in an expanding growth in the annotation of PTM. Just in the SwissProt Knowledgebase, 89,931 of a total of 27 distinguished PTM types reported experimentally. Phosphorylation and glycosylation are major PTM; however, other PTMs also perform a crucial part in several cellular activities; to list a small number of acetylation, ubiquitination, sulfonation, myristoylation, prenylation, glycosyl-phosphatidylinositol anchoring, etc. [15]. Though these PTM have been studied in detail there are only some databases and sources to analyze these PTMs. Some of the databases are listed in Table 2:

TABLE 2
SHOWS DATABASES AVAILABLE FOR PTM STUDIES

Sl. No.	Database	Type of PTM information
1.	DSDBASE [16]	Disulfide Linkages
2.	O-GLYCBASE [17]	Glycosylation
3.	GlycoSuiteDB [18]	Glycosylation
4.	UniPep [19]	Glycosylation
5.	dbOGAP [20]	Glycosylation
6.	KiNG [21]	Phosphorylation
7.	ProMEX [22]	Phosphorylation
8.	PhosSNP [23]	Phosphorylation
9.	LymPHOS [24]	Phosphorylation
10.	PhosphoNet	Phosphorylation
11.	PhosphoregDB [25]	Phosphorylation
12.	PhosphoPOINT [26]	Phosphorylation
13.	Protein kinase resource [27]	Phosphorylation
14.	PHOSIDA [28]	Phosphorylation
15.	Kinomer [29]	Phosphorylation
16.	Phospho3D [30]	Phosphorylation
17.	PhosPhAt [31]	Phosphorylation
18.	The Phosphorylation Site DB [32]	Phosphorylation

19.	PhosphoSitePlus [14]	Phosphorylation
20.	PhosphoGRID [33]	Phosphorylation
21.	PhosphoPep [34]	Phosphorylation

With the indifference in traditional experimental techniques, computational studies of PTM have been in great discussion among researchers, and scientists had been in great debate as to how and when to use these methods. These methods can identify the number of potential candidates and quickly produce useful data for additional experimental analysis faster and efficiently. From the literature survey, it was found that several tools are available for PTM analysis, the majority of them are freely available and few of them are commercial tools. Some of the currently available software for the PTM analysis are listed in Table 3:

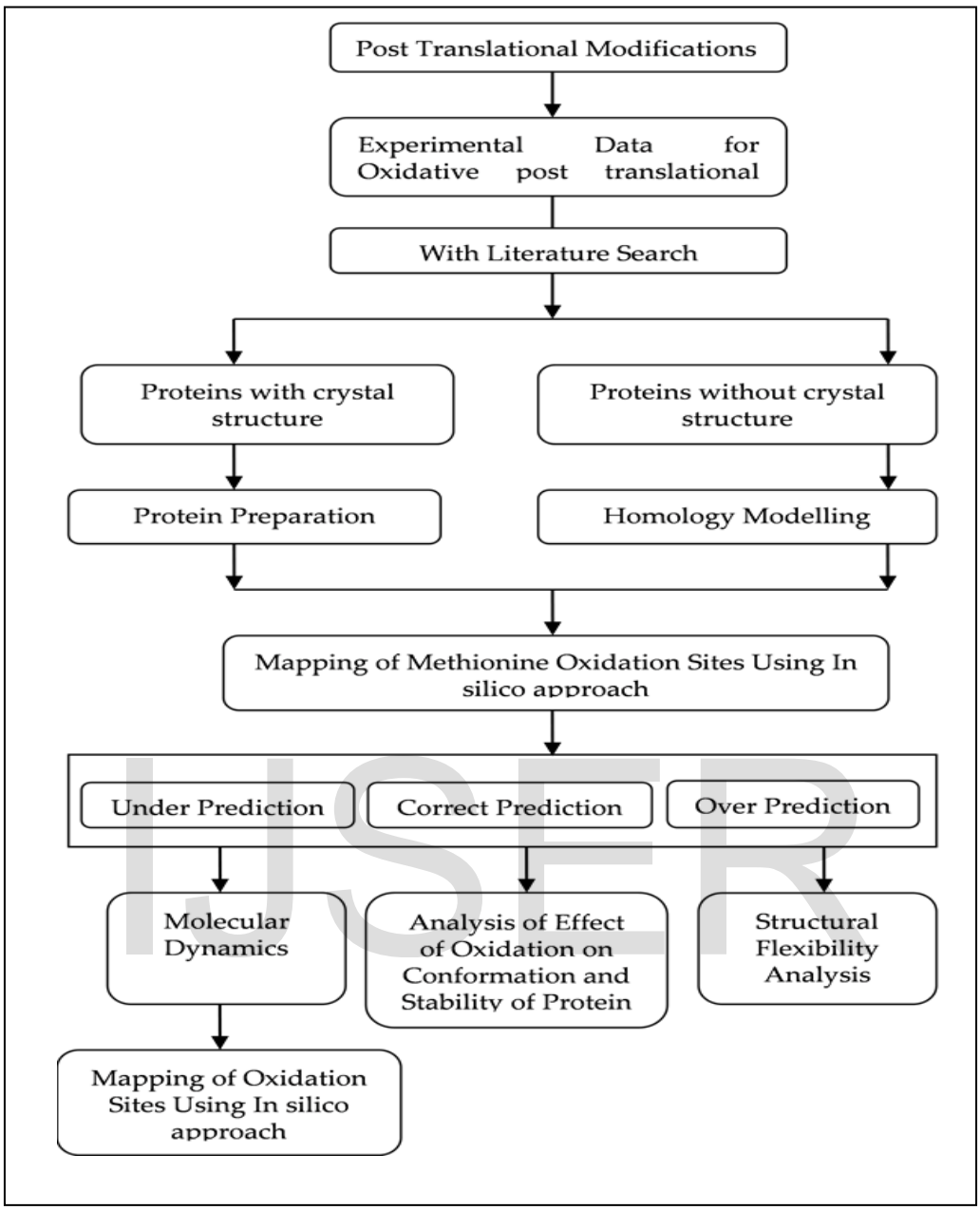
TABLE 3
SHOWS SOFTWARE AVAILABLE FOR PTM STUDIES

Sl. No.	Software	Type of PTM Analyzed
1	BioLuminate [35]	Oxidation, Deamidation, Glycosylation & Proteolysis
2	PHOXTRACK [36]	Phosphorylation
3	Isobar^{PTM} [37]	Phosphorylation
4	Scansite [38]	Phosphorylation
5	PredPhospho [39]	Acetylation
6	KinasePhos [40]	Phosphorylation
7	LysAcet [41]	Lysine acetylation
8	NetAcet [42]	Acetylation
9	NetPhosK [43]	Glycosylation and Phosphorylation
10	SUMOsp [44]	Sumoylation
11	NetNGlyc [45]	Glycosylation
12	NetOGlyc [46]	O-glycosylation

With this aim, literature that contains experimental data for oxidative techniques of PTM was collected. Data found in the literature divided into two parts, proteins with crystal structures and proteins without crystal structures. For the proteins with crystal structures, protein refinement was carried out, and oxidation sites were mapped using BioLuminate. Homology Modeling was carried out for the proteins without crystal structures, and molecular simulation studies were performed to stabilize the system. Later mapping of Methionine oxidation sites was done using BioLuminate. The mapping results were divided into 3 parts; over prediction, correct prediction, and under prediction. For the proteins with under prediction, molecular simulation studies were carried out for different simulation timings. If the simulation timing is more, more stabilized the structure will be. Biopharmaceuticals are very intricate and hence a PTM has a huge influence on them because the PTM of amino acids can affect the protein conformation and indirectly the protein function. Hence PTM was induced manually in the proteins and simulation studies for 10ns were performed to check how Methionine oxidation can affect the conformation of the protein. In the current study, we aimed to do a comparative study between experimentally reported and in silico approaches to check the efficiency of these predictions and also to check the strategies to look over as to how to overcome such prediction errors using different in silico approaches such as loop refinement, Molecular dynamics, Homology modeling, etc. We also aimed to check the effects of PTM on the stability of the biologics using the computational approach that can assist in the development and commercialization of the biologics. This extensive study would provide a blueprint on a complete protocol using in silico reactive residue prediction and MD simulation approaches to check the post-translational modification issues related to biologics products.

2 MATERIALS AND METHODS

Tools Used: Protein Preparation Wizard, Prime, BioLuminate, Desmond
Techniques Used: Protein Preparation, Homology Modeling, Molecular Dynamics Simulation, Reactive Residue Prediction.
Analysis Workflow:



2.1 Protein Preparation

All the crystal structures obtained from the protein data bank were prepared with the help of protein preparation wizard because usually structures have issues such as missing hydrogen, missing amino acids, etc. which need to be fixed. Protein structures were imported and processed to assign bond orders, to add hydrogen atoms, to create zero-order bond metal ions, to create disulfide bonds, and to delete water beyond 5Å from HET groups. Then the protein structures were reviewed and modified by selecting chains carefully and removing unwanted chains, waters and hetero atoms such as ligand and metal. Later the protein structures were refined by optimizing H-bond to remove the steric clash caused due to addition of hydrogen with sample water orientations with the use of PROPKA at neutral pH. Later structures were minimized to restrain the heavy atoms to RMSD 0.30Å with the OPLS_2005 force field.

2.2 Homology modeling

Homology modeling was carried out for the proteins without crystal structure using the structure prediction wizard panel of prime software. Using the file option, sequence was imported and BLAST homology search was done using BLOSUM62 similarity matrix. Based on the score, identities, and protein families, a suitable template was selected. ClustalW was used for high sequence identity and secondary structure prediction was carried out. By using the knowledge-based model building method, models were built.

2.3 MD Simulation

Modeled proteins and proteins with under prediction of oxidation sites were studied by molecular dynamics simulations

using Desmond. Model complexes were built using the system builder panel of Desmond. Individual complexes were solvated by predefined SPC water molecules and orthorhombic simulation box was used with OPLS_2005 force field for all complexes. Each complex was minimized using 2000 maximum iterations and convergence threshold of 1.0 kcal/mol/Å. The periodic boundary condition with the NPT ensemble and default recording interval (energy 1.2ps and trajectory 4.8ps) was used. All energy minimizations and MD simulations were performed at temperature 300K and 1.01325 bar pressure.

2.4 Mapping of oxidation sites

Mapping of oxidation sites was done using the BioLuminate Tool. BioLuminate is generally used for examining biologics and protein systems. Oxidation sites were predicted using the reactive residue option of BioLuminate. Reactive residues were identified by matching residue patterns in the sequence. Four patterns were provided by default, for the common reactions: deamidation, oxidation, glycosylation, and proteolysis. This tool analyzes the structure of a given protein to identify the residues that match the patterns.

3 RESULTS AND DISCUSSION

A major part of the present study involves understanding the output obtained at each stage of the analysis protocol used and validating the results based on the parameters. These results are then used to draw logical conclusions keeping in mind the set of objectives of the project. In this project, a set of samples on oxidative post-translational modification techniques were studied to understand their impact on production of biopharmaceuticals using the bioinformatic computational approach. A detailed report on the results obtained during the analysis along with the understanding derived from their outcomes is illustrated in this chapter.

3.1 Extraction of sample data

The main objective was to collect a set of samples on oxidative post-translational modification techniques. A set of proteins containing oxidized methionine, cysteine, histidine, tyrosine, and tryptophan were selected from recent PubMed publications. The full text of the publications was reviewed. The decisive factor used in the review process is to check if the locations of oxidative sites are given or not. Finally, 48 proteins from 39 publications were selected. The list of selected proteins is given in Table 4 & 5 below.

TABLE 4
TEST SET OF PROTEINS WITH CRYSTAL STRUCTURE

Sl. No	Proteins With Crystal Structure	PDB Ids	Enzyme Class
1	RAS P21	1CTQ, 1GNR, 1IOZ,1Q21, 1JAH, 1LF0, 1RVD, 1XCM, 1XJ0, 2Q21,4Q21,6Q21	Hydrolase
2	DJ-I	2RK3, 2RK6, 3CZA, 4BTE, 1SOA, 1P5F, 2OR3, 3CY6, 4P2G, 1PE0, 2RK4, 3B36, 2R1T, 3B38	Hydrolase
3	Antithrombin	1ANT,1ATH	Hydrolase
4	CaM	1CLL	Transferase
5	S100A9	1XK4, 1IRJ, 4GGF, 4XJK, 1MR8	Transferase
6	APOA-I	3R2P, 1AV1, 2A01	Transferase
7	APOA-II	1L6L, 2OU1	Transferase
8	Staphylococcal Nuclease	1EY0, 1EY4, 1EYC, 1EZ8, 1F2Y, 1II3, 1KAA	Hydrolase

9	UCH-L1	4DM9, 4JKJ, 3IFW, 3IRT, 3KVF, 3KW5	Hydrolase
10	Actin	1ATN	Hydrolase
11	Cystatin C	1R4C, 3GAX, 3NX0, 3QRD, 3S67, 3SVA, 1G96, 1TIJ	Hydrolase
12	Alpha 1-Antitrypsin	3DRM, 3DRU, 3NDD, 3NE4, 1EZX, 1ATU, 1QLP, 2QUG	Hydrolase
13	Alpha A Crystallin	3L1E,3L1F	Oxidoreductase
14	Alpha B Crystallin	4M5S	Oxidoreductase
15	Actin Trx1	1ERT, 1ERV, 1ERW, 2HSH, 2HXK, 4POM, 4LL1, 1AIU, 1AUC, 3KD0, 2IFQ	Transferase
16	H-Ras	121P, 1CLU, 1CTQ, 1GNR, 1JAH, 1JAI, 1LF0, 1LF5, 1QRA, 1RVD, 1XCM, 1XJ0	Hydrolase
17	AhpC Peroxidase	4MA9, 4MAB, 1YF1, 1N8J, 1YEP, 1YEX, 1YF0	Oxidoreductase
18	Archaeal Peroxiredoxin	2ZCT, 2CV4, 2E2M, 2E2G, 2NVL	Oxidoreductase
19	AhpD	1GU9, 1KNC, 1LW1, 1ME5	Oxidoreductase
20	Tropomyosin	2TMA, 2EFR, 2EFS, 2D3E, 2Z5H, 2Z5I	Transferase
21	Tyrosine Phosphotase	1LYV, 3F99, 3F9A, 3F9B, 3U96, 4GF3, 1PA9, 2I42	Hydrolase
22	PerR	2FE3, 2RGV, 3F8N	Hydrolase
23	Pseudomonas aeruginosa azurin	3U25	Oxidoreductase
24	Angiotensin I	2WXX, 2WXY, 2WY0	Hydrolase
25	Cytoglobin	1UMO, 1URV, 1URY, 1UT0, 1UX9, 1V5H, 2DC3, 3AG0, 4B3W	Oxidoreductase
26	Choline Oxidase	2JBV, 3LJP, 3NNE, 4MJW	Oxidoreductase
27	Aryl alcohol Oxidase	3FIM	Oxidoreductase
28	Bovine Cytochrome c Oxidase	2DYR, 1V54,1V55, 2Y69, 3ABM, 3AG1, 3AG2, 3AG3, 3WG7, 1OCC, 1OCR	Oxidoreductase
29	human Cu,Zn-superoxide dismutase	1AZV, 1HL4,1HL5, 1FUN, 1MFM, 1N18, 1N19, 1OZU, 1P1V, 1PTZ, 1PU0	Oxidoreductase

30	catalase-peroxidase (KatG)	1SJ2, 2CCA, 2CCD, 4C50, 4C51	Oxidoreductase
31	MEDI-493	1ZA6	Transferase
32	Acetylcholine	4ZK4	Hydrolase
33	MopE	2VOV, 2VOW, 2VOX	Lyase
34	human growth harmone	1AXI, 3HHR	Hydrolase
35	Bacterioferritin	2Y3Q, 2VXI, 3E2C, 4CVR, 4CVS, 4CVT, 2HTN, 3E1P, 3E1Q, 4CVP	Oxidoreductase
36	Heme Oxygenase	1N45, 1OYL, 1OZL, 1OZR, 1OZW, 1XJZ, 3CZY, 1NI6, 1OZE, 1S8C	Oxidoreductase
37	AGO2	4W5O, 4W5T	Hydrolase

TABLE 5

TEST SET OF PROTEINS WITHOUT CRYSTAL STRUCTURE

Sl. No	**Proteins Without Crystal Structure**	**UniProt Ids**	**Enzyme Class**
1	Beta A1 Crystallin	P05813	Oxidoreductase
2	Beta B1 Crystallin	P53674	Oxidoreductase
3	Beta B2 Crystallin	P43320	Oxidoreductase
4	Beta B3 Crystallin	P26998	Oxidoreductase
5	Gamma B Crystallin	P07316	Oxidoreductase
6	Gamma C Crystallin	P07315	Oxidoreductase
7	Gamma S Crystallin	S55263	Oxidoreductase
8	Alpha Synuclein	P37840	Ligase

9	DAAO	Q99042	Oxidoreductase
10	D1 protein	P14660	Hydrolase
11	Gamma D Crystallin	P07320	Oxidoreductase

3.2 Mapping of oxidation sites for proteins with crystal structures

Out of 48 proteins, 37 proteins were with the crystal structure. From the protein data bank, X-ray crystallographic structure for proteins was obtained. Protein preparation was carried out for the proteins and mapping of the oxidation sites was done using BioLuminate (Fig.1). The mapping of oxidation sites is divided into three parts: correct prediction, overprediction, and under prediction. Results of mapping of oxidation sites for proteins with crystal structures using BioLuminate are given in Table 6.

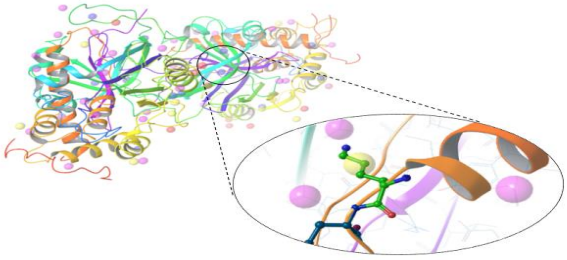


Fig.1. Shows mapping of oxidation sites using BioLuminate (PDB ID: 1ANT)

TABLE 6

SHOWS RESULTS FOR THE CORRECT PREDICTION OF OXIDATION SITE USING BIOLUMINATE

Sl.NO	Protein Name	Experimental Residues	Prediction of Oxidation Sites Using BioLuminate				
			Methionine	Cysteine	Histidine	Tyrosine	Tryptophan
1	Antithrombin	17M, 20M, 314M, 315M	17M, 20M, 314M, 315M	NA	NA	NA	NA
2	ACTIN	44M, 47M, 176M, 190M,227M, 269M, 355M	44M, 47M, 176M, 190M,227M, 269M, 355M	NA	NA	NA	NA
3	CaM	36M, 51M, 71M, 76M, 109M, 124M, 145M, 146M	36M, 51M, 71M, 76M, 109M, 124M, 145M, 146M	NA	NA	NA	NA
4	Cystatin C	14M, 41M	14M, 41M	NA	NA	NA	NA
5	S100A9	63M, 81M, 83M, 42C	63M, 81M, 83M	ND	NA	NA	NA
6	Alpha 1-Antitrypsin	1M, 226M, 242M, 351M, 358M, 232C	1M, 226M, 242M, 351M, 358M	232C	NA	NA	NA
7	UCH-L1	1M,12M, 124M, 179M	1M,12M, 124M, 179M	NA	NA	NA	NA
8	RAS P21	1M, 67M, 72M, 111M, 4Y, 40Y, 96Y, 137Y, 157Y	1M, 67M, 72M, 111M	NA	NA	4Y, 40Y, 96Y, 137Y, 157Y	NA
9	DJ-I	17M, 26M, 133M, 134M	17M, 26M, 133M, 134M	NA	NA	NA	NA
10	APOA-I	86M, 112M,	86M, 112M,	NA	NA	NA	NA

		148M	148M				
11	APOA-II	26M	26M	NA	NA	NA	NA
12	Staphylococcal Nuclease	26M, 32M, 65M, 98M	26M, 32M, 65M, 98M	NA	NA	NA	NA
13	Alpha A Crystallin	138M	138M	NA	NA	NA	NA
14	Alpha B Crystallin	68M	68M	NA	NA	NA	NA
15	Actin Trx1(Transferase)	32C,35C,62C, 69C	NA	32C,35C,62C, 69C	NA	NA	NA
16	H-Ras	118C	NA	118C	NA	NA	NA
17	AhpC Peroxidase	165C	NA	165C	NA	NA	NA
18	Archaeal Peroxiredoxin	50C	NA	50C	NA	NA	NA
19	AhpD	130C, 133C, 137H	NA	130C, 133C	137H	NA	NA
20	Tropomyosin	190C	NA	190C	NA	NA	NA
21	Tyrosine Phosphotase	403C, 402H	NA	403C	ND	NA	NA
22	PerR	37H, 91H	NA	NA	37H, 91H	NA	NA
23	Pseudomonas aeruginosa azurin	20H, 48Y	NA	NA	20H	78Y	NA
24	Angiotensin I	6H, 9H	NA	NA	ND	NA	NA
25	Cytoglobin	81H	NA	NA	81H	NA	NA
26	Choline Oxidase	351H	NA	NA	ND	NA	NA
27	Aryl Alcohol Oxidase	502H	NA	NA	ND	NA	NA
28	Bovine Cytochrome *c* Oxidase	334W	NA	NA	NA	NA	334W
29	Human Cu,Zn-Superoxide Dismutase	32W	NA	NA	NA	NA	32W
30	Catalase-Peroxidase (KatG)	107W, 229Y	NA	NA	NA	229Y	107W
31	MEDI-493	105W	NA	NA	NA	NA	105W
32	Acetylcholine	86W, 93Y	NA	NA	NA	93Y	86W
33	MopE	130W	NA	NA	NA	NA	130W
34	Human Growth Harmone	86W	NA	NA	NA	NA	86W
35	Bacterioferritin	25Y, 45Y, 58Y	NA	NA	NA	25Y, 45Y, 58Y	NA

36	Heme Oxy-genase	58Y	NA	NA	NA	58Y	NA
37	AGO2	393Y	NA	NA	NA	393Y	NA

NOTE: ND: Not Detected, NA: Not Applicable

3.3 Mapping of oxidation sites for proteins without crystal structures

Out of 48 proteins, 11 proteins did not have a crystal structure. Homology Modeling had been carried out for those proteins. Sequential information for those proteins was collected from the UniProt database using their UniProt IDs (Fig. 2). After modeling, protein refinement was carried to refine the structure, and then the mapping of oxidation sites was done. Results of mapping of oxidation sites for modeled proteins using BioLuminate is given in Table 7.

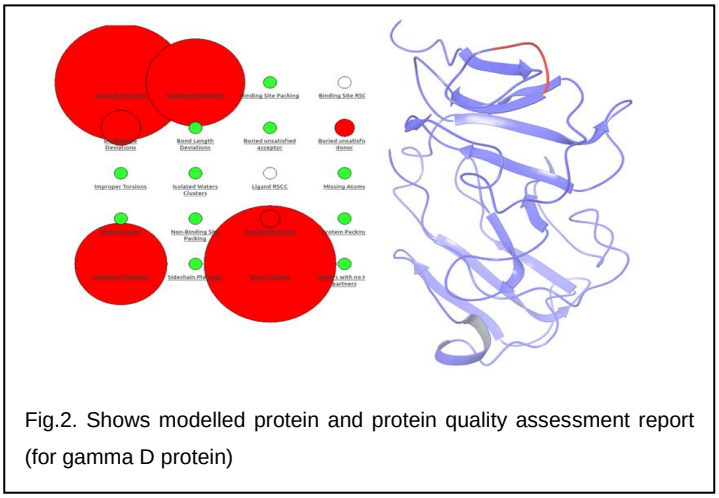


Fig.2. Shows modelled protein and protein quality assessment report (for gamma D protein)

TABLE 7

SHOWS RESULTS FOR THE CORRECT PREDICTION OF OXIDATION SITE FOR PROTEINS WITHOUT CRYSTAL STRUCTURE USING BIOLUMINATE

SL.NO	Protein Name	Experimental Residues	Prediction of Oxidation Sites for proteins without crystal structure Using BioLuminate				
			Methionine	Cysteine	Histidine	Tyrosine	Tryptophan
1	Alpha Synu-clein	1M, 5M, 116M, 127M,39Y, 125Y, 133Y, 136Y, 50H	1M, 5M, 116M, 127M	NA	50H	39Y, 125Y, 133Y, 136Y	NA
2	Beta A1 Crys-tallin	126M	126M	NA	NA	NA	NA
3	Beta B1 Crys-tallin	226M, 216W	226M	NA	NA	NA	216W
4	Beta B2 Crys-tallin	122M, 59W	122M	NA	NA	NA	59W
5	Beta B3 Crys-tallin	129M	129M	NA	NA	NA	NA
6	Gamma B crystallin	70M, 69W	70M	NA	NA	NA	69W
7	Gamma S Crystallin	41M, 101M, 106M	41M, 101M, 106M	NA	NA	NA	NA
8	Gamma C Crystallin	70M, 56Y	70M	NA	NA	56Y	NA
9	Gamma D Crystallin	46Y	NA	NA	NA	46Y	NA
10	DAAO	108C	NA	108C		NA	NA
11	D1 Protein	190H	NA	NA	190H	NA	NA

NOTE: ND: Not Detected, NA: Not Applicable

3.4 Mapping of oxidation sites after simulation for the proteins without crystal structures

After modeling of the proteins, simulation studies were carried out for the stabilization of the structures. Molecular dynamics simulation was carried out for 11 proteins for 2ns and 5ns simulation time. Simulation studies were carried out using Desmond. After molecular dynamics simulation, average structure and last frame calculation was carried out for the stabilized structure of modeled proteins. Here are the results of mapping of oxidation sites for modeled proteins after simulation using BioLuminate given in Table 8.

TABLE 8

SHOWS RESULTS FOR THE PREDICTION OF OXIDATION SITE FOR PROTEINS WITHOUT CRYSTAL STRUCTURE USING BIOLUMINATE

Sl.NO	Protein Name	Experimental Residues	Prediction of Oxidation Sites for protein without crystal structure Using BioLuminate after MD				
			Methionine	Cysteine	Histidine	Tyrosine	Tryptophan
Simulation Time 2ns							
1	Alpha Synuclein	1M, 5M, 116M, 127M,39Y, 125Y, 133Y, 136Y, 50H	1M, 5M, 116M, 127M	NA	50H	39Y, 125Y, 133Y, 136Y	NA
2	Beta A1 Crystallin	126M	126M	NA	NA	NA	NA
3	Beta B1 Crystallin	226M, 216W	226M	NA	NA	NA	NA
4	Beta B2 Crystallin	122M, 59W	122M	NA	NA	NA	NA
5	Beta B3 Crystallin	129M	129M	NA	NA	NA	NA
6	Gamma B Crystallin	70M, 69W	70M	NA	NA	NA	NA
7	Gamma S Crystallin	41M, 101M, 106M	41M, 101M, 106M	NA	NA	NA	NA
8	Gamma C Crystallin	70M, 56Y	70M	NA	NA	56Y	NA
Simulation Time 5ns							
9	Gamma D Crystallin	46Y	NA	NA	NA	46Y	NA
10	DAAO	108C	NA	108C	NA	NA	NA
11	D1 Protein	190H	NA	NA	190H	NA	NA

NOTE: ND: Not Detected, NA: Not Applicable

3.5 Analysis of structural flexibility for the proteins with over prediction of oxidation sites

During mapping of oxidation sites, it was found that BioLuminate had predicted oxidation sites present in the proteins other than the oxidation sites mentioned in experimental data (Table 9). This is because of the structural flexibility of the Methionine residues.

TABLE 9

SHOWS RESULTS FOR THE OVER PREDICTION OF OXIDATION SITE USING BIOLUMINATE

SL. NO	Protein Name	Experi-mental Residues	Prediction of Oxidation Sites Using BioLuminate				
			Methionine	Cysteine	Histidine	Tyrosine	Tryptophan
1	Antithrombin	17M, 20M, 314M, 315M	89M, 103M,251, 252M, 281M, 320M,338M, 423M	NA	NA	NA	NA
2	ACTIN	44M, 47M, 176M, 190M,227 M, 269M, 355M	82M, 119M, 123M, 132M, 283M, 299M, 305M, 313M, 325M	NA	NA	NA	NA
3	CaM	36M, 51M, 71M, 76M, 109M, 124M, 145M, 146M	72M, 144M	NA	NA	NA	NA
4	Cystatin C	14M, 41M	110M	NA	NA	NA	NA
5	S100A9	63M, 81M, 83M, 42C	1M, 5M,78M,94M	ND	NA	NA	NA
6	Alpha 1-Antitrypsin	1M, 226M, 242M, 351M, 358M, 232C	63M, 220M,221M, 374M, 385M	ND	NA	NA	NA
7	UCH-L1	1M,12M, 124M, 179M	82M	NA	NA	NA	NA
8	Alpha Synu-clein	1M, 5M, 116M, 127M,39Y, 125Y, 133Y, 136Y, 50H	1M, 5M, 116M, 127M	NA	50H	39Y, 125Y, 133Y, 136Y	NA
9	RAS P21	1M, 67M, 72M, 111M, 4Y, 40Y, 96Y, 137Y, 157Y	1M, 67M, 72M, 111M	NA	NA	4Y, 40Y, 96Y, 137Y, 157Y	NA
10	DJ-I	17M, 26M, 133M, 134M	ND	NA	NA	NA	NA
11	APOA-I	86M, 112M, 148M	ND	NA	NA	NA	NA

12	APOA-II	26M	ND	NA	NA	NA	NA
13	Staphylococcal Nuclease	26M, 32M, 65M, 98M	ND	NA	NA	NA	NA
14	Alpha A Crystallin	138M	ND	NA	NA	NA	NA
15	Alpha B Crystallin	68M	ND	NA	NA	NA	NA
16	Beta A1 Crystallin	126M	46M, 107M, 111M, 151MET, 161M	NA	NA	NA	NA
17	Beta B1 Crystallin	226M, 216W	113M, 137M, 144M, 124W, 127W, 219W	NA	NA	124W, 127W, 219W	216W
18	Beta B2 Crystallin	122M, 59W	193M	NA	NA	82W, 85W	59W
19	Beta B3 Crystallin	129M	ND	NA	NA	NA	NA
20	Gamma B Crystallin	70M, 69W	44M, 91M, 103M, 137M, 172M	NA	NA	43W, 132W, 158W	69W
21	Gamma S Crystallin	41M, 101M, 106M	56M,90M	NA	NA	NA	NA
22	Gamma C Crystallin	70M, 56Y	44M,102M, 103M, 160M	NA	NA	7Y, 46Y, 63Y, 66Y, 93Y, 134Y, 139Y, 144Y	NA
23	Gamma D Crystallin	46Y	NA	NA	NA	7Y, 17Y, 29Y, 51Y, 56Y, 63Y, 93Y, 98Y, 134Y, 139Y, 144Y, 151Y, 154Y	NA
24	Actin Trx1(Transferase)	32C,35C,62C, 69C	NA	73C	NA	NA	NA
25	H-Ras	118C	NA	ND	NA	NA	NA
26	AhpC Peroxidase	165C	NA	46C	NA	NA	NA
27	Archaeal Peroxiredoxin	50C	NA	213C	NA	NA	NA
28	AhpD	130C, 133C, 137H	NA	ND	62H	NA	NA
29	Tropomyosin	190C	NA	ND	NA	NA	NA
30	Tyrosine Phosphotase	403C, 402H	NA	221C, 234C, 259C, 418C	ND	NA	NA
31	DAAO	108C	NA	145C, 159C, 193C, 257C, 298C		NA	NA

32	PerR	37H, 91H	NA	NA	25H, 119H, 127H	NA	NA
33	Pseudomonas Aeruginosa Azurin	20H, 48Y	NA	NA	83H	ND	NA
34	Angiotensin I	6H, 9H	NA	NA	ND	NA	NA
35	Cytoglobin	81H	NA	NA	65H, 161H	NA	NA
36	D1 Protein	190H	NA	NA	92H, 195H, 198H, 215H, 252H, 332H	NA	NA
37	Choline Oxi-dase	351H	NA	NA	ND	NA	NA
38	Aryl Alcohol Oxidase	502H	NA	NA	190H	NA	NA
39	Bovine Cyto-chrome *c* Oxi dase	334W	NA	NA	NA	NA	6W, 25W, 81W, 103W, 126W, 186W, 236W, 275W, 288W, 323W, 334W, 396W, 409W, 450W, 473W,494W
40	Human Cu, Zn-Superoxide Dismutase	32W	NA	NA	NA	NA	ND
41	Catalase-Peroxidase (KatG)	107W, 229Y	NA	NA	NA	28Y, 64Y, 95Y, 98Y, 113Y, 155Y, 197Y,210Y, 304Y, 337Y, 339Y, 353Y, 390Y, 413Y, 426Y, 597Y, 608Y, 638Y, 678Y, 711Y,	38W, 39W, 90W, 91W, 135W, 149W, 161W, 191W, 198W, 204W, 300W, 341W, 351W, 397W, 412W, 438W, 477W, 505W, 668W, 689W, 728W
42	MEDI-493	105W	NA	NA	NA	NA	56W
43	Acetylcholine	86W, 93Y	NA	NA	NA	20Y, 54Y , 93 Y ,149Y, 168Y,169Y 174Y,182Y, 188Y, 195Y	55W, 60W, 67W, 147W
44	MopE	130W	NA	NA	NA	NA	112W, 127W, 136W, 171W,180W, 211W, 305W, 320W, 328W
45	Human Growth Har-mone	86W	NA	NA	NA	NA	80W
46	Bacterioferri-tin	25Y, 45Y, 58Y	NA	NA	NA	10Y, 107Y, 114Y, 149Y	NA
47	Heme Oxy-genase	58Y	NA	NA	NA	55Y, 78Y, 97Y, 107Y, 134Y, 137Y, 182Y	NA

48	AGO2	393Y	NA	NA	NA	55Y, 57Y, 101Y, 174Y, 225Y, 311Y, 322Y, 338Y, 420Y, 494Y, 519Y, 529Y, 625Y, 654Y, 667Y, 698Y, 749Y, 765Y, 784Y, 790Y, 804Y, 805Y, 815Y	NA

NOTE: ND: Not Detected, NA: Not Applicable

3.6 Molecular Dynamics Simulation and mapping of oxidation sites for Proteins with Under Prediction

The mapping of oxidation sites was divided into three parts: correct prediction, underprediction, and overprediction. In total, 7 proteins with under prediction of oxidation sites were observed. Molecular dynamics simulation was carried out for those proteins. After the simulation, 3 proteins with correct prediction and 4 proteins again with under prediction were observed. Here are the mapping of oxidation sites for proteins with under predictions after simulation using BioLuminate given in Table 10.

TABLE 10

SHOWS RESULTS FOR THE PREDICTION OF OXIDATION SITE FOR PROTEINS WITH UNDER PREDICTION USING BIOLUMINATE

SL.NO	Protein Name	Experimental Residues	Prediction of Oxidation Sites for protein with under prediction Using BioLuminate after MD				
			Methionine	Cysteine	Histidine	Tyrosine	Tryptophan
Simulation Time 5ns							
1	S100A9	42C	NA	42C	NA	NA	NA
2	Choline Oxidase	351H	NA	NA	ND	NA	NA
3	Aryl Alcohol Oxidase	502H	NA	NA	ND	NA	NA
4	Cytoglobin	81H	NA	NA	81H	NA	NA
5	Tyrosine Phosphotase	402H	NA	NA	ND	NA	NA
6	Angiotensin I	6H, 9H	NA	NA	ND	NA	NA
7	DAAO	108C	NA	108C	NA	NA	NA

NOTE: ND: Not Detected, NA: Not Applicable

3.7 Analysis of effects on the protein conformation due to methionine oxidation

The majority of the biopharmaceuticals are the results of post-translational modifications of the protein. Post-translational modifications play a pivotal role in managing protein activity. Therefore recognizing the type of variation and its position may be crucial for understanding the function of a given protein and eventually the cell as a whole. These post-translational modifications have distinctive consequences on correct folding, druggability, receptor binding, immunogenicity, pharmacokinetics, and stability of the biopharmaceuticals. The protein staphylococcal nuclease was selected to check the consequence of Methionine oxidation on the conformation. From the literature, it was found that in staphylococcal nuclease Methionine 26 and 32 were not predominantly sensitive to either mutation or oxidation. Certainly, changeover in these residues can be more disruptive than oxidation. It was found that oxidation of Methionine 98 destabilized the protein by 0.2kcal/mol, but the effects of substitutions were as bad or even worse. Hence though the substitution of Methionine prohibited oxidative damage in case of positions such as Methionine 98, it will also be required to make substitutions at other sites in protein to locate a stable structural resolution. This is in contrast to the behavior of Methionine 65. To check this, Methionine 98 of the protein Staphylococcal nuclease (PDB ID: 1EY0) was changed to Methionine sulfone and molecular dynamics was done for 10ns (Fig. 3). Similarly, molecu-

lar dynamics was carried out for the wild Staphylococcal nuclease (PDB ID: 1EY0) for 10ns. After molecular dynamics, RMSF (Root Mean Square Fluctuation) plots for both structures were generated and analyzed. RMSF plots are shown in Fig. 4 and 5.

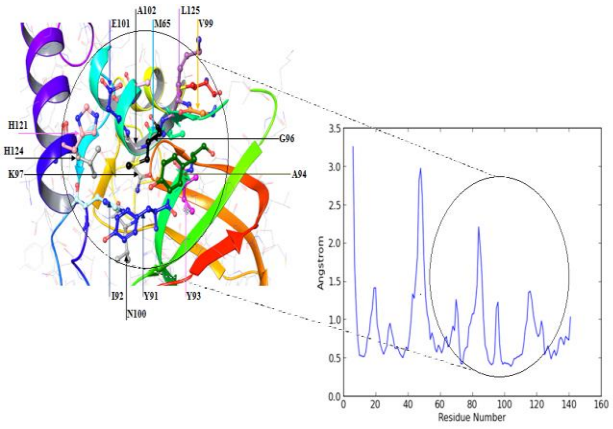


Fig.3. Shows change in the conformation of protein structure due to PTM (PDB ID: 1EY0).

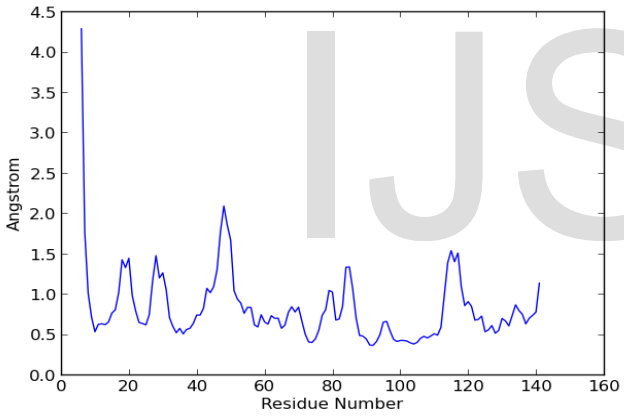


Fig.4. Shows RMSF Plot for wild Staphylococcal nuclease (PDB ID: 1EY0)

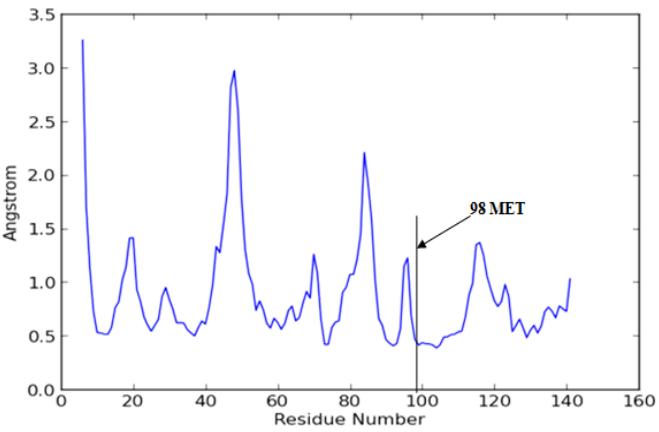


Fig.5. Shows RMSF Plot for mutated Staphylococcal nuclease (PDB ID: 1EY0)

4 Conclusion

We wish to conclude that our computational predictions of reactive hotspot sites reported experimentally predicted the closest proximities. The protein structures whose reactive hotspots (PTM) were not expected as per experiments can be refined using molecular dynamics and Loop refinements calculations. Suppose an X-ray crystal structure is not available for a particular protein of interest. In that case, we could perform homology modeling and then process PTM predictions. All the above predictions performed in the study shall be used as a benchmarking to get more confidence in computational works. We could initiate biological design in a rational,cost-effectively manner from the computational predictions. The computational biology results confirm that most of the predictions could be achieved computationally. Using molecular modeling tools shall give more significant insights at the atomic level.

REFERENCES

[1] P. J. Declerck, "Biologicals and biosimilars: a review of the science and its implications," *Generics Biosimilars Initiat. J.*, 2012, doi: 10.5639/gabij.2012.0101.005.

[2] P. P. Catterall, "Monoclonal antibodies: A witness seminar in contemporary medical history," *Med. Hist.*, 1994, doi: 10.1017/S0025727300036632.

[3] J. (2007). Zhang, J.-M., & An, "Cytokines, Inflammation, and Pain. International," *Int Anesth. Clin.*, 2007.

[4] S. Ivkovic *et al.*, "Connective tissue growth factor coordinates chondrogenesis and angiogenesis during skeletal development," *development*. 2003, doi: 10.1242/dev.00505.

[5] R. Rajagopalan, "An Overview of Similar Biologics," *J. Sci. Innov. Res.*, 2013.

[6] D. J. A. Crommelin, G. Storm, R. Verrijk, L. De Leede, W. Jiskoot, and W. E. Hennink, "Shifting paradigms: Biopharmaceuticals versus low molecular weight drugs," *Int. J. Pharm.*, 2003, doi: 10.1016/S0378-5173(03)00376-4.

[7] A. M. Dempster, "Nonclinical safety evaluation of biotechnologically derived pharmaceuticals," *Biotechnol. Annu. Rev.*, 2000, doi: 10.1016/S1387-2656(00)05037-7.

[8] A. Baumann, "Early Development of Therapeutic Biologics - Pharmacokinetics," *Curr. Drug Metab.*, 2005, doi: 10.2174/138920006774832604.

[9] G. R. Galluppi *et al.*, "Integration of pharmacokinetic and pharmacodynamic studies in the discovery, development, and review of protein therapeutic agents: A conference report," in *Clinical Pharmacology and Therapeutics*, 2001, doi: 10.1067/mcp.2001.115455.

[10] I. Mahmood, "Interspecies Scaling of Protein Drugs: Prediction of Clearance from Animals to Humans," *J. Pharm. Sci.*, 2004, doi: 10.1002/jps.10531.

[11] K. Xu, V. N. Uversky, and B. Xue, "Local Flexibility Facilitates Oxidization of Buried Methionine Residues," *Protein Pept. Lett.*, 2012, doi: 10.2174/092986612800494084.

[12] M. Mann and O. N. Jensen, "Proteomic analysis of post-translational modifications," *Nature Biotechnology*. 2003, doi: 10.1038/nbt0303-255.

[13] N. Kresge, R. D. Simoni, and R. L. Hill, "The process of reversible phosphorylation: the work of Edmond H. Fischer.," *J. Biol. Chem.*, 2011, doi: 10.1074/jbc.O110.000242.

[14] P. V. Hornbeck *et al.*, "PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse," *Nucleic Acids Res.*, 2012, doi: 10.1093/nar/gkr1122.

[15] K. S. Kamath, M. S. Vasavada, and S. Srivastava, "Proteomic databases and tools to decipher post-translational modifications," *Journal of Proteomics*. 2011, doi: 10.1016/j.jprot.2011.09.014.

[16] A. Vinayagam, G. Pugalenthi, R. Rajesh, and R. Sowdhamini, "DSDBASE: A consortium of native and modelled disulphide bonds in proteins," *Nucleic Acids Res.*, 2004.

[17] J. E. Hansen, O. Lund, J. O. Nielsen, J. S. Hansen, and S. Brunak, "O-GLYCBASE: A revised database of O-glycosylated proteins," *Nucleic Acids Research*. 1996, doi: 10.1093/nar/24.1.248.

[18] C. A. Cooper, H. J. Joshi, M. J. Harrison, M. R. Wilkins, and N. H. Packer, "GlycoSuiteDB: A curated relational database of glycoprotein glycan structures and their biological sources. 2003 update," *Nucleic Acids Research*. 2003, doi: 10.1093/nar/gkg099.

[19] H. Zhang *et al.*, "UniPep - A database for human N-linked glycosites: A resource for biomarker discovery," *Genome Biol.*, 2006, doi: 10.1186/gb-2006-7-8-r73.

[20] J. Wang, M. Torii, H. Liu, G. W. Hart, and Z. Z. Hu, "dbOGAP - An Integrated Bioinformatics Resource for Protein O-GlcNAcylation," *BMC Bioinformatics*, 2011, doi: 10.1186/1471-2105-12-91.

[21] A. Krupa, K. R. Abhinandan, and N. Srinivasan, "KinG: A database of protein kinases in genomes," *Nucleic Acids Res.*, 2004, doi: 10.1093/nar/gkh019.

[22] J. Hummel *et al.*, "ProMEX: A mass spectral reference database for proteins and protein phosphorylation sites," *BMC Bioinformatics*, 2007, doi: 10.1186/1471-2105-8-216.

[23] J. Ren *et al.*, "PhosSNP for systematic analysis of genetic polymorphisms that influence protein phosphorylation," *Mol. Cell. Proteomics*, 2010, doi: 10.1074/mcp.M900273-MCP200.

[24] D. Ovelleiro, M. Carrascal, V. Casas, and J. Abian, "LymPHOS: Design of a phosphosite database of primary human T cells," *Proteomics*, 2009, doi: 10.1002/pmic.200800701.

[25] A. R. R. Forrest *et al.*, "PhosphoregDB: The tissue and sub-cellular distribution of mammalian protein kinases and phosphatases," *BMC Bioinformatics*. 2006, doi: 10.1186/1471-2105-7-82.

[26] C. Y. Yang *et al.*, "PhosphoPOINT: A comprehensive human kinase interactome and phospho-protein database," in *Bioinformatics*, 2008, doi: 10.1093/bioinformatics/btn297.

[27] C. Petretti and C. Prigent, "The Protein Kinase Resource: everything you always wanted to know about protein kinases but were afraid to ask," *Biol. Cell*, 2005, doi: 10.1042/bc20040077.

[28] F. Gnad, J. Gunawardena, and M. Mann, "PHOSIDA 2011: The post-translational modification database," *Nucleic Acids Res.*, 2011, doi: 10.1093/nar/gkq1159.

[29] D. M. A. Martin, D. Miranda-Saavedra, and G. J. Barton, "Kinomer v. 1.0: A database of systematically classified eukaryotic protein kinases," *Nucleic Acids Res.*, 2009, doi: 10.1093/nar/gkn834.

[30] A. Zanzoni *et al.*, "Phospho3D 2.0: An enhanced database of three-dimensional structures of phosphorylation sites," *Nucleic Acids Res.*, 2011, doi: 10.1093/nar/gkq936.

[31] M. Zulawski, R. Braginets, and W. X. Schulze, "PhosPhAt goes kinases-searchable protein kinase target information in the plant phosphorylation site database PhosPhAt," *Nucleic Acids Res.*, 2013, doi: 10.1093/nar/gks1081.

[32] S. M. Wurgler-Murphy, D. M. King, and P. J. Kennelly, "The Phosphorylation Site Database: A guide to the serine-, threonine-, and/or tyrosine-phosphorylated proteins in prokaryotic organisms," *Proteomics*, 2004, doi: 10.1002/pmic.200300711.

[33] C. Stark *et al.*, "PhosphoGRID: a database of experimentally verified in vivo protein phosphorylation sites from the budding yeast Saccharomyces cerevisiae.," *Database (Oxford).*, 2010, doi: 10.1093/database/bap026.

[34] B. Bodenmiller *et al.*, "PhosphoPep - A database of protein phosphorylation sites in model organisms," *Nature Biotechnology*. 2008, doi: 10.1038/nbt1208-1339.

[35] BioLuminate v, Schrödinger, LLC, New York, NY, 2015.

[36] C. Weidnery, C. Fischery, and S. Sauer, "PHOXTRACK-a tool for interpreting comprehensive datasets of post-translational modifications of proteins," *Bioinformatics*, 2014, doi: 10.1093/bioinformatics/btu572.

[37] F. P. Breitwieser and J. Colinge, "IsobarPTM: A software tool for the quantitative analysis of post-translationally modified proteins," *J. Proteomics*, 2013, doi: 10.1016/j.jprot.2013.02.022.

[38] J. C. Obenauer, L. C. Cantley, and M. B. Yaffe, "Scansite 2.0: Proteome-wide prediction of cell signalling interactions using short sequence motifs," *Nucleic Acids Res.*, 2003, doi: 10.1093/nar/gkg584.

[39] J. H. Kim, J. Lee, B. Oh, K. Kimm, and I. S. Koh, "Prediction of phosphorylation sites using SVMs," *Bioinformatics*, 2004, doi: 10.1093/bioinformatics/bth382.

[40] H. Da Huang, T. Y. Lee, S. W. Tzeng, and J. T. Horng, "KinasePhos: A web tool for identifying protein kinase-specific phosphorylation sites," *Nucleic Acids Res.*, 2005, doi: 10.1093/nar/gki471.

[41] S. Li, H. Li, M. Li, Y. Shyr, L. Xie, and Y. Li, "Improved Prediction of Lysine Acetylation by Support Vector Machines," *Protein Pept. Lett.*, 2009, doi: 10.2174/092986609788923338.

[42] L. Kiemer, J. D. Bendtsen, and N. Blom, "NetAcet: Prediction of N-terminal acetylation sites," *Bioinformatics*, 2005, doi: 10.1093/bioinformatics/bti130.

[43] N. Blom, T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft, and S. Brunak, "Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence," *Proteomics*. 2004, doi: 10.1002/pmic.200300771.

[44] J. Ren *et al.*, "Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0," *Proteomics*, 2009, doi: 10.1002/pmic.200800646.

[45] R. Gupta and S. Brunak, "Prediction of glycosylation across the human proteome and the correlation to protein function.," *Pac. Symp. Biocomput.*, 2002, doi: 10.1142/9789812799623_0029.

[46] J. E. Hansen, O. Lund, N. Tolstrup, A. A. Gooley, K. L. Williams, and S. Brunak, "NetOglyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility," *Glycoconj. J.*, 1998, doi: 10.1023/A:1006960004440.