UNIVERSITY of York

This is a repository copy of *Post-translational modifications in the Protein Data Bank*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/216797/

Version: Published Version

Article:
Schofield, Lucy C, Dialpuri, Jordan S, Murshudov, Garib N et al. (1 more author) (2024) Post-translational modifications in the Protein Data Bank. Acta crystallographica. Section D, Structural biology. D80. pp. 647-660. ISSN 2059-7983

https://doi.org/10.1107/S2059798324007794

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

STRUCTURAL BIOLOGY

ISSN 2059-7983

Post-translational modifications in the Protein Data Bank

Lucy C. Schofield,[a] Jordan S. Dialpuri,[a] Garib N. Murshudov[b]* and Jon Agirre[a]*

[a]York Structural Biology Laboratory, Department of Chemistry, University of York, York, United Kingdom, and [b]MRC Laboratory of Molecular Biology, University of Cambridge, Cambridge, United Kingdom. *Correspondence e-mail: garib@mrc-lmb.cam.ac.uk, jon.agirre@york.ac.uk

Received 10 May 2024
Accepted 7 August 2024

Edited by N. Pearce, Linköping University, Sweden

Keywords: post-translational modifications; Protein Data Bank; glycosylation; phosphorylation; acetylation.

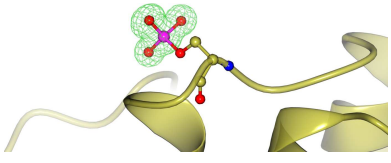
Supporting information: this article has supporting information at journals.iucr.org/d

Proteins frequently undergo covalent modification at the post-translational level, which involves the covalent attachment of chemical groups onto amino acids. This can entail the singular or multiple addition of small groups, such as phosphorylation; long-chain modifications, such as glycosylation; small proteins, such as ubiquitination; as well as the interconversion of chemical groups, such as the formation of pyroglutamic acid. These post-translational modifications (PTMs) are essential for the normal functioning of cells, as they can alter the physicochemical properties of amino acids and therefore influence enzymatic activity, protein localization, protein–protein interactions and protein stability. Despite their inherent importance, accurately depicting PTMs in experimental studies of protein structures often poses a challenge. This review highlights the role of PTMs in protein structures, as well as the prevalence of PTMs in the Protein Data Bank, directing the reader to accurately built examples suitable for use as a modelling reference.

1. Introduction

Protein post-translational modifications (PTMs) are covalent modifications of amino acids that can alter the physicochemical properties, and therefore the function, of a protein. As a result, PTMs are essential for regulating various biochemical processes in cells, such as protein localization, epigenetic regulation and cell signalling (Smotrys & Linder, 2004; Wang *et al.*, 2019; Rocks *et al.*, 2005). Dysregulated PTMs have been shown to have a role in many human diseases, including hyperglycosylation in cancer (Thomas *et al.*, 2021), elevated histone acetylation in diabetes (Wang *et al.*, 2019) and hyperphosphorylation in neurodegenerative diseases (Basheer *et al.*, 2023). Understanding PTMs, particularly in three-dimensional protein structures, is essential for furthering the knowledge of general protein function, the molecular basis of disease, and drug discovery (Bhullar *et al.*, 2018; Dekker *et al.*, 2014; Copeland, 2018).

PTMs can encompass the addition of small molecules, such as phosphorylation and methylation; long-chain modifications, such as glycosylation and lipidation; small proteins, such as in ubiquitination and SUMOylation; as well as the interconversion of chemical groups, such as the formation of isopeptide bonds, pyroglutamic acid and citrulline. The most common targets of post-translational modification are the side chains of amino acid residues, as well as the N- and C-termini of the protein chain. These modifications can alter the chemistry of target amino acids, including size, charge, surface area and hydrophobicity, which can lead to changes in protein properties, such as conformation, protein–protein interactions and enzyme activity (Betts *et al.*, 2017; van den Bedem & Wilson,



OPEN ACCESS
Published under a CC BY 4.0 licence

Table 1
The top 15 PTM sites listed in the dbPTM.

The PTMs with the highest number of total sites are listed. As of February 2024, there were 542 107 putative sites and 2 235 664 experimentally determined PTM sites in the dbPTM (Li *et al.*, 2022). Due to a lack of data in the dbPTM, experimental data for 2-hydroxyisobutyrlation was taken from *DeepKhib*, which used experimentally determined sites to train the predictor algorithm (Zhang *et al.*, 2020).

Modification type	No. of experimental sites	No. of putative sites	Total sites
Phosphorylation	1615150	160490	1775640
Ubiquitination	348308	108349	456657
Acetylation	138171	38530	176701
N-Linked glycosylation	27366	89143	116509
Methylation	16114	16766	32880
2-Hydroxyisobutyrylation	12166	32392	44558
O-Linked glycosylation	16696	8809	25505
Succinylation	17973	6387	24360
Malonylation	12847	145	12992
Sumoylation	5889	5731	11620
S-Palmitoylation	6505	3409	9914
Sulfoxidation	7581	0	7581
Hydroxylation	2404	4543	6947
Amidation	3316	1462	4778
S-Nitrosylation	4172	483	4655

2019; T *et al.*, 2018). These modifications are often reversible and dynamic, with a tendency to occur in disordered regions that are often surface-accessible, allowing amino acids to dynamically fit into the catalytic site of modifying enzymes (Pang *et al.*, 2007; Xie *et al.*, 2007). Many PTMs do not exist in isolation (Venne *et al.*, 2014), as proteins can undergo multiple modifications at various sites that can influence the actions of each other, known as PTM crosstalk (Leutert *et al.*, 2021).

Traditionally, PTMs have been detected and characterized using experimental methods, such as western blotting, and mass spectrometry (Wilkins *et al.*, 1999). As a result of these techniques, the volume of PTM experimental data necessitated the creation of open, easy-access databases. The dbPTM is a database that compiles information on PTMs, which includes both experimentally determined PTM sites as well as putative sites (Li *et al.*, 2022). The experimentally determined sites are derived from existing PTM databases and extracted from research articles; these are mapped to UniProt entries to ensure a nonredundant data set (UniProt Consortium, 2021). Putative sites are derived from the UniProt Knowledgebase (UniProtKB), which predicts these sites based on sequence similarity or evolutionary potential (UniProt Consortium, 2021; Li *et al.*, 2022). The total number of PTM sites in the dbPTM therefore includes a nonredundant list of sites that have been identified either experimentally or putatively (Li *et al.*, 2022). The dbPTM highlights phosphorylation, ubiquitination, acetylation and glycosylation as some of the most common PTM sites (shown in Table 1). The large collection of PTM data in these databases paved the way for PTM prediction tools, such as *NetPhos*, *DeepAcet*, *HydLoc* and *DeepKhib* (Blom *et al.*, 1999; Wu *et al.*, 2019; Huang, Chen *et al.*, 2020; Zhang *et al.*, 2020). These are commonly based on machine-learning approaches trained on experimental data that predict PTM sites based on sequence features, structural properties and evolutionary conservation (Kumar *et al.*, 2017; Wu *et al.*, 2019; Huang, Chen *et al.*, 2020; Blom *et al.*, 1999; Bludau *et al.*, 2022). In parallel, the availability of highly accurate whole-proteome structure predictions has made it possible to map putative phosphorylation, ubiquitination and acetylation PTM sites onto 3D models, allowing algorithms to refine predictions based on factors such as solvent-accessibility (Bludau *et al.*, 2022; Joosten & Agirre, 2022).

Whilst this allows the identification of experimentally determined and putative sites, it does not provide detailed structural information. As PTMs contain information relevant to protein function that is not encoded in the protein sequence, the structure of these modifications is crucial to understand protein structure and function. Macromolecular X-ray crystallography (MX), cryo-electron microscopy (cryo-EM) single-particle analysis (SPA) and nuclear magnetic resonance (NMR) are powerful tools for determining protein structures (Kumar *et al.*, 2020; Atanasova *et al.*, 2020; Reid *et al.*, 2004). However, they face limitations when studying modified proteins, as many post-translational modifications are labile, and the dynamic nature of PTMs leads to non-uniform modification sites, contributing to structural heterogeneity. For a given molecule, each site is either modified or not, and consequently the averaging over all of the unit cells of a crystal in X-ray crystallography, or particles in cryo-EM, can result in a lower apparent occupancy of PTMs, as some molecules may be modified while others are not. PTMs also typically occur in highly flexible solvent-exposed regions and frequently introduce conformational variability into proteins, which can lead to increased disorder at the protein surface (Deller *et al.*, 2016). In fact, protein glycosylation modifications are frequently removed from glycoproteins before performing structural studies in order to avoid these complexities (Agirre, 2017; Deller *et al.*, 2016; Atanasova *et al.*, 2020). While the presence of PTMs may introduce challenges in structure determination, it is essential to consider the functional significance of these modifications, as PTMs play pivotal roles in modulating protein conformation, stability and dynamics, thereby influencing protein function and regulation. As a result, incorporating PTMs into structural studies offers a more comprehensive view into protein structure, dynamics and biological relevance, enhancing the understanding of protein function.

2. Biochemistry of PTMs

When considering the modelling of PTMs within a protein structure, certain biochemical aspects should be considered. Understanding which amino acids can undergo modifications is vital, as most PTMs occur on specific amino acid residues. PTMs can also occur at consensus sequences or motifs, which denote specific amino acid patterns within a protein that act as recognition sites for the enzymes responsible for the modification. These motifs are often essential for the modification to take place, as is seen in *N*-glycosylation, where the consensus motif is N-*X*-S/T (where *X* represents any amino acid other than proline). These can be strict motifs, as for *N*-glycosylation, or weaker motifs, as for phosphorylation (Bludau *et al.*,

2022). Moreover, the wider structural context of the modification site is essential, such as its location within hydrophobic or disordered regions (Pang *et al.*, 2007).

Here, key biochemical information for the most relevant small-molecule PTMs and long-chain PTMs will be explored; only examples where electron density can be calculated from deposited diffraction data and a model will be showcased, along with the Chemical Component Dictionary (CCD) identifiers used by the Protein Data Bank (PDB) to represent them (Westbrook *et al.*, 2015; Berman *et al.*, 2000). For each model, omit maps were calculated to validate the presence of the modification in the electron density. Omit maps were calculated using MTZ files downloaded from the RCSB PDB (Berman *et al.*, 2000) by removing the modification followed by re-refinement with *REFMAC* using randomization, which was used to shake the remaining model (Murshudov *et al.*, 2011). The calculation of omit maps was used to reduce model bias by ensuring that the observed electron density was not influenced by prior assumptions about the presence of the modification or structure. By excluding the region of interest, the resulting omit map provides a more accurate indication of the PTM. All 3D figures were produced using *CCP*4*mg* (McNicholas *et al.*, 2011), which is part of the *CCP*4 software suite (Agirre *et al.*, 2023).

2.1. Phosphorylation

Protein phosphorylation is one of the most abundant and well studied PTMs in the proteome. It involves the reversible addition of a phosphate group from a nucleoside triphosphate, typically ATP, to a polar amino acid side chain via kinase enzymes (Fig. 1). Phosphorylation most commonly occurs on serine (Fig. 1), threonine or tyrosine residues (Supplementary Fig. S1), but can occur on many other amino acids (Li *et al.*, 2022). There is no strict consensus motif for phosphorylation, although individual kinases do recognize specific motifs (Miller & Turk, 2016). Phosphorylation sites often occur in disordered regions, at protein–protein interaction faces and within loop and hinge regions, affecting protein conformation, protein–protein interactions and protein stability (Betts *et al.*, 2017; T *et al.*, 2018; Qin *et al.*, 2021; Durek *et al.*, 2009; Rieloff & Skepö, 2020). The addition of a phosphate group introduces a large, dianionic group that offers a new site to form hydrogen bonds or salt bridges, which can alter protein interactions or create new binding sites (Johnson & Lewis, 2001). Phosphorylation is important for the function of many proteins, including the activation of enzymes, transcription factors and protein receptors (Betts *et al.*, 2017; T *et al.*, 2018; Mayr & Montminy, 2001). It is therefore implicated in several human diseases, including cancer, in which the tyrosine kinase family encompass the largest number of oncoproteins (Singh *et al.*, 2017), as well as in Alzheimer's disease, where the altered protein phosphorylation states of several proteins, including the amyloid-β protein precursor and tau protein, are closely associated with protein aggregation (Kumar *et al.*, 2011; Despres *et al.*, 2017).

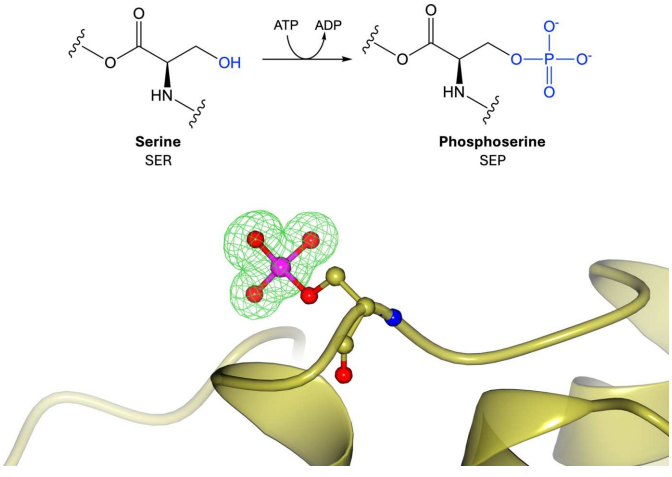


Figure 1
Phosphorylation. Top: phosphorylation of serine involves the addition of a phosphate group donated by ATP to the side-chain hydroxyl group. Bottom: phosphoserine (PDB entry 5n3h; Sadowsky *et al.*, 2011; CCD code SEP). Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by a yellow ribbon model.

2.2. Methylation

Protein methylation involves the reversible addition of a methyl group to the amino group of an amino acid, often donated by *S*-adenosyl-L-methionine via methyltransferase enzymes (Fig. 2; Małecki *et al.*, 2022). Lysine (Fig. 2) and arginine (Supplementary Fig. S2) are the most frequent targets of protein methylation, but it can also occur on other amino acids (Li *et al.*, 2022). The ε-amino group of lysine can accept up to three methyl groups, yielding mono-, di- or trimethylated states (Małecki *et al.*, 2022). The guanidino group of arginine

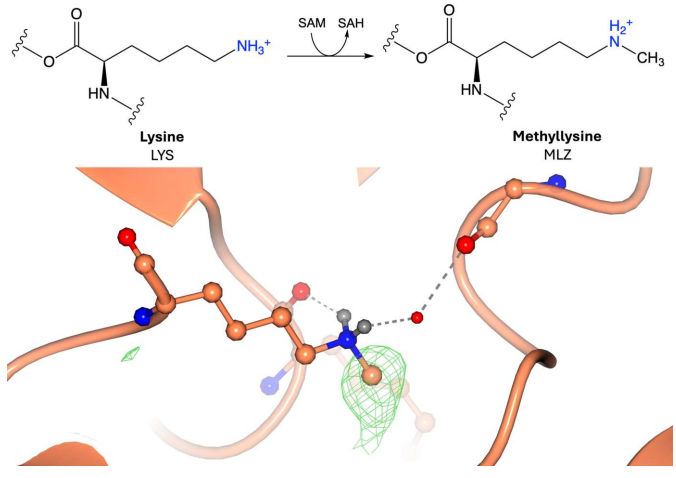


Figure 2
Methylation. Top: methylation of lysine involves the addition of a methyl group donated by *S*-adenosylmethionine (SAM) to the side-chain amino group. Bottom: methyllysine (PDB entry 3kmt; Wei & Zhou, 2010; CCD code MLZ). Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by an orange ribbon model. The environment surrounding the modification is shown, indicating that the NZ atom is protonated (hydrogenation of the NZ atom was performed using *Coot*; Emsley *et al.*, 2010). Hydrogen bonds are displayed as grey dashed lines.

can be methylated on one or on both N atoms, yielding monomethylarginine or dimethylarginine (asymmetric or symmetric) (Małecki *et al.*, 2022). Lysine methylation does not have a well defined consensus sequence, whereas arginine methylation commonly occurs in glycine-rich and arginine-rich regions known as GAR motifs (Wooderchak *et al.*, 2008; Lorton & Shechter, 2019; Daily *et al.*, 2005). The N-terminal methylation of amino acids can also occur by the action of N-terminal methyltransferases, with substrates often containing the consensus motif *X*-P-K/R (*X* = S/P/A/G) after the removal of the initiator methionine (Diaz *et al.*, 2021). Generally, methylation occurs in disordered protein regions, but it can also be found in ordered regions (Narasumani & Harrison, 2018). Methylation increases the bulkiness and alters the hydrogen-bonding capacity of the modified residues, which can affect protein stability, subcellular localization, binding affinity and protein–protein interactions (Liu *et al.*, 2023). Both lysine and arginine methylation are particularly abundant in the N-terminal, flexible tails of histone proteins, and result in an epigenetic mark that controls gene expression and chromatin state (Bannister & Kouzarides, 2011). Arginine methylation can also target several nonhistone proteins that regulate processes such as DNA repair and RNA splicing (Wei *et al.*, 2021; Brobbey *et al.*, 2022). Protein methylation therefore has a regulatory role in many cellular processes, including gene transcription and DNA repair, and can contribute to neurological disorders, cancer and ageing (Liu *et al.*, 2023).

2.3. Hydroxylation

Hydroxylation is an oxidation reaction in which a carbon–hydrogen bond is oxidized into a carbon–hydroxyl bond via hydroxylase enzymes (Fig. 3). Proline is the most frequently hydroxylated residue (Fig. 3), followed by lysine (Supplementary Fig. S3); however, other residues can also undergo hydroxylation (Li *et al.*, 2022). Proline hydroxylation can occur either on the γ-carbon, forming 3-hydroxyproline, or on the β-carbon, forming 4-hydroxyproline, whilst lysine hydroxylation occurs on the δ-carbon, forming 5-hydroxylysine (Tak *et al.*, 2019). There is no known consensus motif for hydroxylation sites, although they tend to occur in surface-accessible, intrinsically disordered regions of proteins (Ismail *et al.*, 2016). Protein hydroxylation increases the hydrophilicity of the amino acids, allowing hydroxylated residues to become more water-soluble, which can impact protein structure and function (Varma *et al.*, 2021). Both hydroxyproline and hydroxylysine are present in collagen, where they play important roles in water solubility and the formation of triple-helical structures found in collagen fibrils, and act as precursors for subsequent PTMs such as glycosylation (Varma *et al.*, 2021; Stawikowski *et al.*, 2014). Hydroxylation also plays a role in hypoxia signalling by regulating hypoxia-inducible factor, where prolyl hydroxylation marks the protein for ubiquitination and subsequent degradation (Bruick & McKnight, 2001); this has been shown to play a role in tumour suppression or promotion (Strocchi *et al.*, 2022).

2.4. Acetylation

Protein acetylation involves the reversible addition of an acetyl group onto an amino acid (Fig. 4). Acetylation most frequently occurs on lysine (Fig. 4) and alanine (Supplementary Fig. S4), as well as methionine and serine, but other amino acids can also be acetylated (Li *et al.*, 2022). Lysine acetylation occurs on the ε-amino group of lysine side chains, which is important for gene transcription through histone acetylation (Davie, 1998). Histone acetyltransferases catalyse the addition of acetyl groups onto histone lysine residues via acetyl-CoA, while histone deacetylases remove

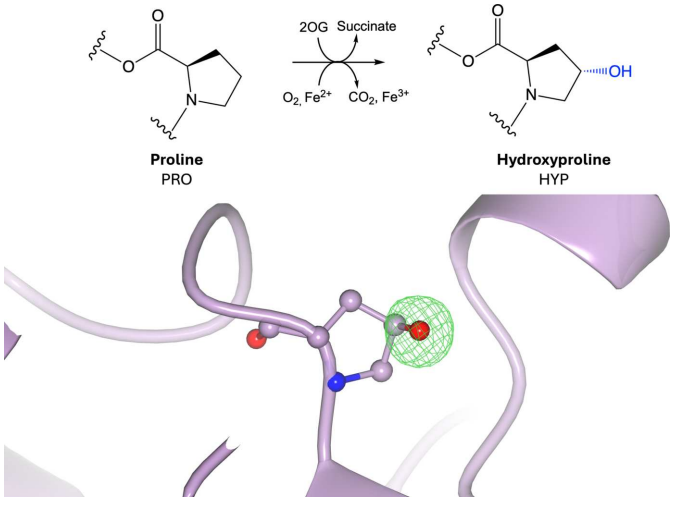


Figure 3
Hydroxylation. Top: hydroxylation of proline involves the addition of a hydroxyl group donated by 2-oxoglutarate (2OG) to the side-chain pyrrolidine ring. Bottom: hydroxyproline (PDB entry 1gk8; Taylor *et al.*, 2001; CCD code HYP). Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by a purple ribbon model.

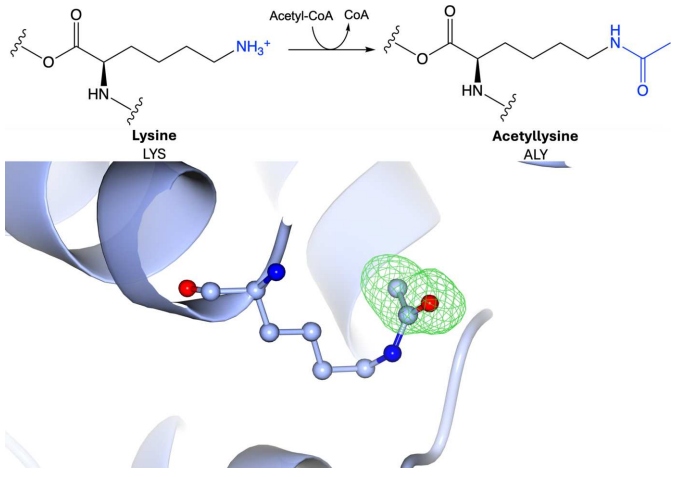


Figure 4
Acetylation. Top: acetylation of lysine involves the addition of an acetyl group donated by acetyl-CoA to the side-chain amino group. Bottom: acetyllysine (PDB entry 5e2f; Y. Kim, G. Joachimiak, M. Endres, G. Babnigg & A. Joachimiak, unpublished work; CCD code ALY). Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by a blue ribbon model.

them, meaning the modification is reversible (Davie, 1998). Protein acetylation can also occur irreversibly on the free α-amino group at the protein N-terminus via N-terminal acetyltransferases (Varland *et al.*, 2015). A global consensus motif for acetylation has not been identified, although lysine-acetylation motifs have been identified in specific organisms (Weinert *et al.*, 2011, 2013; Choudhary *et al.*, 2009; Okanishi *et al.*, 2013; Lundby *et al.*, 2012; Crosby & Escalante-Semerena, 2014; Crosby *et al.*, 2012). Acetylation neutralizes charges on amino acids, affecting the electrostatic properties of proteins. For example, the acetylation of lysine residues on histone proteins weakens the histone–DNA binding affinity, leading to chromatin relaxation, increased accessibility to DNA-binding proteins and subsequent gene transcription (Davie, 1998). As a result, lysine acetylation controls the regulation of proteins involved in many diseases, including cancer and inflammatory conditions (Bai *et al.*, 2016; Hu *et al.*, 2022). Meanwhile, N-terminal acetylation transforms a charged protein N-terminus into a hydrophobic segment, which can impact protein folding, stability, protein–protein interactions and localization (Hwang *et al.*, 2010; Trexler & Rhoades, 2012; Scott *et al.*, 2011; Behnia *et al.*, 2004). In fact, dysregulation of N-terminal acetylation has been shown to have implications in cancer, as well as in developmental disorders, including brain and heart development (Varland, Silva *et al.*, 2023; Varland, Brønstad *et al.*, 2023; Koufaris & Kirmizis, 2020).

2.5. Oxidation

Protein oxidation involves the addition of oxygen-containing groups to amino acid residues via reactive oxygen species (ROS; Fig. 5). This includes sulfoxidation reactions, in which reactive sulfur-containing residues such as cysteine and methionine are the targets of oxidative stress (Li *et al.*, 2022).

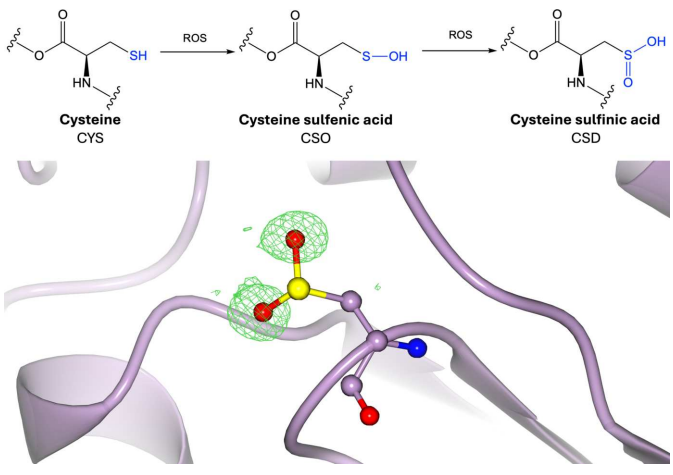


Figure 5
Oxidation. Top: oxidation of cysteine involves the reaction between the side-chain thiol group of cysteine and reactive oxygen species (ROS) to form cysteine sulfenic, then cysteine sulfinic acid and cysteine sulfonic acid (Supplementary Fig. S5). Bottom: cysteine sulfinic acid (PDB entry 1soa; Canet-Avilés *et al.*, 2004; CCD code CSD). Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by a purple ribbon model.

One-electron oxidation of cysteine forms thiyl radicals, which react either with other thiols to form disulfide bonds or with O_2 to generate thiyl peroxyl radicals (Wardman & von Sonntag, 1995). The two-electron oxidation of cysteine by oxidants forms unstable sulfenic acid, sulfinic acid (Fig. 5) and sulfonic acid species (Supplementary Fig. S5), which either yield oxyacids by hydrolysis reactions or disulfide bonds by reacting with other thiol groups (Claiborne *et al.*, 1999; Turell *et al.*, 2008). Methionine residues can also be oxidized to form methionine sulfoxide, which can be further oxidized to methionine sulfone (Hoshi & Heinemann, 2001). Furthermore, oxidant species such as superoxide can react with nitric oxide to form nitrosating species, which can lead to nitrosylated cysteine residues or nitrated tyrosine residues (Nedospasov *et al.*, 2000; Berlett *et al.*, 1996). Oxidation lacks a strict motif, although it often occurs in surface-accessible regions (Sanchez *et al.*, 2008; Yang *et al.*, 2017; Garrido Ruiz *et al.*, 2022). Oxidation of protein residues introduces a polar group, which can affect the structural properties of proteins, and cysteine oxidation has been shown to affect cell signalling and enzyme activity (van den Bedem & Wilson, 2019), whilst methionine oxidation has been shown to destabilize proteins and affect enzyme function, with implications in neurological disorders (Liu *et al.*, 2008; Mulinacci *et al.*, 2011; Chandran & Binninger, 2023).

2.6. Pyroglutamic acid

Pyroglutamic acid (also known as pyrrolidone carboxylic acid or 5-oxoproline) is an amino acid modification involving the cyclic lactam of glutamic acid or glutamine (Fig. 6, Supplementary Fig. S6; Kumar & Bachhawat, 2012; Connell & Hanes, 1956). It is formed by the cyclization of N-terminal

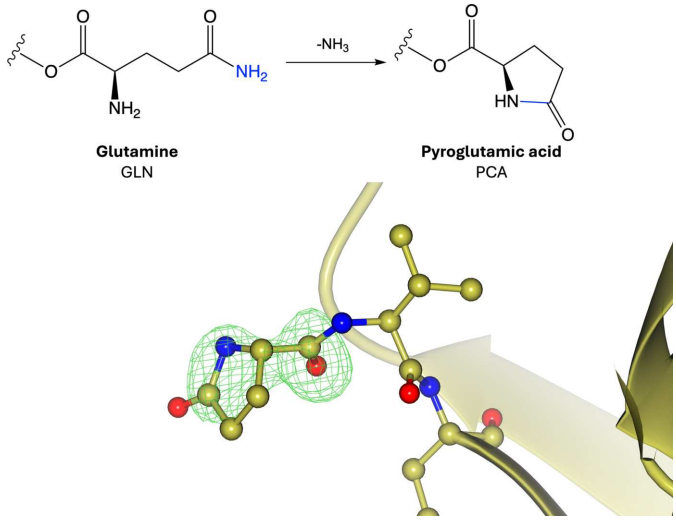


Figure 6
Pyroglutamic acid. Top: formation of pyroglutamic acid involves the cyclization of an N-terminal glutamine or glutamic acid. Bottom: pyroglutamic acid (PDB entry 8ojt; Davies *et al.*, 2023; CCD code PCA). The first three N-terminal residues are shown. Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by a yellow ribbon model.

glutamine or glutamic acid residues through enzymatic catalysis by glutaminyl cyclase, and most commonly occurs on glutamine residues (Connell & Hanes, 1956; Kumar & Bachhawat, 2012; Li *et al.*, 2022). Pyroglutamic acid can also form spontaneously under acidic conditions, which has been shown to occur in immunoglobulin structures *in vitro* (Chelius *et al.*, 2006). The lack of an identified consensus sequence for this modification can be attributed to limited research in this area, although it has been suggested that pyroglutamic acid is more likely to occur within coiled regions (Pang *et al.*, 2007). Many proteins, including antibodies, enzymes and structural proteins, have been shown to exhibit an N-terminal pyroglutamic acid (Brandt *et al.*, 1984; Chelius *et al.*, 2006). This modification has been shown to increase the half-life of antibodies and structural proteins by blocking the action of aminopeptidases, thereby reducing protein degradation (Cummins & O'Connor, 1998; Chelius *et al.*, 2006; Brandt *et al.*, 1984). In addition, it has been shown to affect protein–receptor binding and has been linked to protein aggregation in Alzheimer's disease by increasing hydrophobicity and encouraging β-sheet formation (Gunn *et al.*, 2010; Hinkle & Tashjian, 1973).

2.7. Glycosylation

One of the most commonly occurring and well studied PTMs is glycosylation, where an oligosaccharide moiety is covalently attached to an amino acid via a glycosidic bond, forming a glycoprotein. Protein glycosylation is often categorized into two major types: N-linked and O-linked glycosylation, with C- and S-linked glycosylation being far less frequent (Varki *et al.*, 2022). N-linked glycosylation is one of the most well studied PTMs and occurs on the side-chain N atom of an asparagine residue, with a strict consensus sequence N-*X*-S/T (where *X* is any amino acid except proline; Fig. 7, Supplementary Fig. S7), whereas *O*-glycosylation occurs on the side-chain O atom of a serine or threonine residue with no known consensus sequence (Fig. 8, Supplementary Fig. S8). Glycosylation reactions are diverse and are catalysed by numerous different enzymes that attach specific glycans to specific amino acids. Glycosylation sites are typically surface-accessible and coat the surface of many proteins, including antibodies, enzymes and cell-surface receptors (Suga *et al.*, 2018). It is estimated that at least 50% of human proteins are glycosylated (An *et al.*, 2009). Glycosylation is crucial in many biological processes, including molecular recognition and cell signalling, meaning that glycans play a critical role in human health and disease. For example, deregulation of glycosylation can contribute to several hallmarks of cancer through increased proliferation and metastasis (Purushothaman *et al.*, 2023) and play a role in viral infections such as in SARS-CoV-2, where spike glycoproteins protruding from the cell surface are necessary for viral host-cell entry (Huang, Yang *et al.*, 2020).

The study of the glycoproteome is challenging due to the multitude and diversity of glycoprotein isoforms. This is a

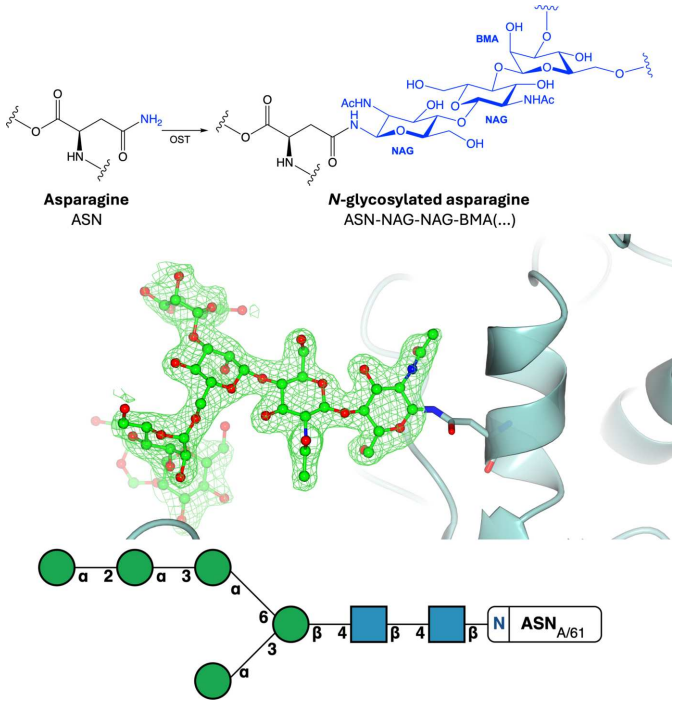


Figure 7
N-Glycosylation. Top: *N*-glycosylation of asparagine involves the addition of an *N*-glycan to the side-chain amino group via oligosaccharide transferase (OST). Middle: N-linked glycosylation (PDB entry 5fji; Agirre *et al.*, 2016). Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by a blue ribbon model. Bottom: the symbol nomenclature for glycans (SNFG) representation is shown and was generated using the *Privateer Web App* (Dialpuri, Bagdonas, Schofield, Pham, Holland, Bond *et al.*, 2024).

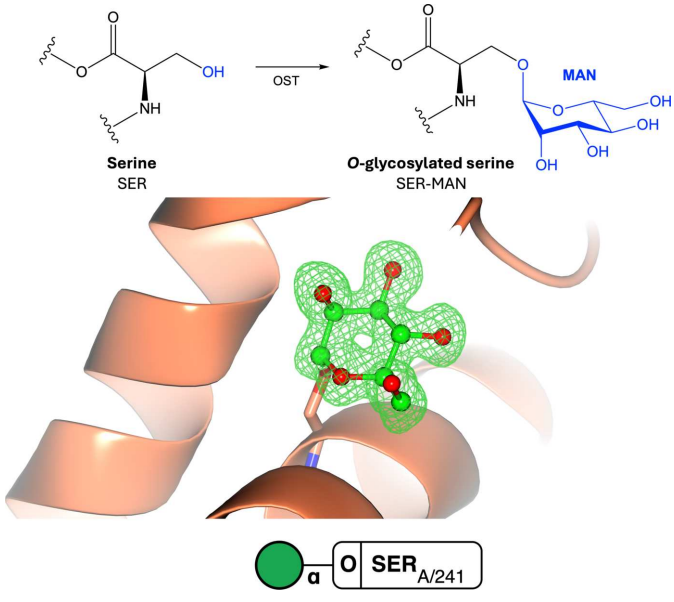


Figure 8
O-Glycosylation. Top: *O*-glycosylation of serine or threonine involves the addition of an *O*-glycan to the side-chain hydroxyl group via oligosaccharide transferase (OST). Middle: O-linked glycosylation (PDB entry 2ciw; Kühnel *et al.*, 2006). Omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by an orange ribbon model. Bottom: the symbol nomenclature for glycans (SNFG) representation is shown and was generated using the *Privateer Web App* (Dialpuri, Bagdonas, Schofield, Pham, Holland, Bond *et al.*, 2024).

result of complex glycan-processing events that involve subsequent trimming and modification events that occur according to the available cellular enzymes: glycoside hydrolases, glycosyl transferases and oligosaccharyl transferases. Understanding the three-dimensional structure of sugars is challenging due to their various stereochemical and anomeric conformations (Agirre, 2017; Atanasova *et al.*, 2020). Previously, the production of a correct three-dimensional structure of a glycoprotein was difficult as many refinement and validation processes relied on software written for proteins and nucleic acids, and the libraries of restraints had become outdated (Agirre, 2017; Agirre *et al.*, 2017; Atanasova *et al.*, 2020). In addition, obtaining a high-resolution structure is generally more difficult for proteins containing sugars due to glycan heterogeneity and mobility, which lead to poorer experimental data than for sugar-free structures (Agirre *et al.*, 2017; van Beusekom *et al.*, 2018).

2.8. Lipidation

Protein lipidation describes a series of modifications in which a lipid molecule is covalently attached to a protein. Several types of lipidation exist, including the addition of fatty acids, isoprenoids, sterols, phospholipids and glycosylphosphatidylinositol anchors. Lipidation is crucial for cell signalling, as it modulates protein function in reaction to stimuli by increasing protein hydrophobicity. This can alter protein–membrane binding affinities, change subcellular localization, and impact protein folding, stability and protein–protein interactions (Rocks *et al.*, 2005; Tanaka *et al.*, 1995). The most extensively studied lipidation types include the addition of fatty-acid chains (palmitoylation and myristoylation) and prenylation (addition of isoprenoids, including farnesylation and geranylgeranylation). Whilst proteins are more or less ordered and structured, lipids are flexible and may have

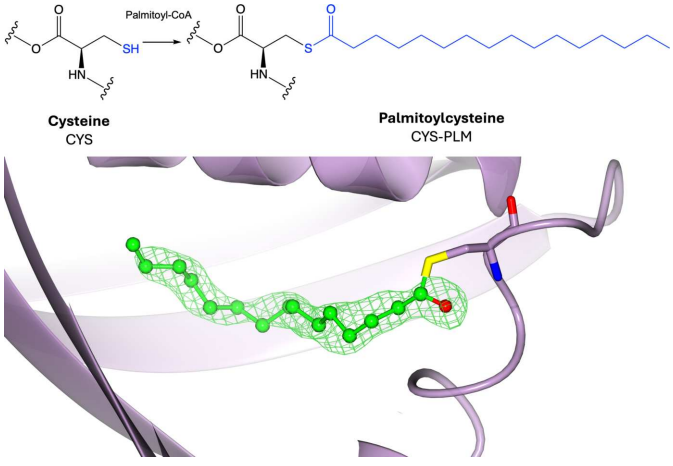


Figure 9
Palmitoylation. Top: palmitoylation of cysteine involves the addition of a palmitoyl group donated by palmitoyl-CoA to the side-chain thiol group. Bottom: palmitoylcysteine (PDB entry 2w3y; Quevillon-Cheruel *et al.*, 2009; CCD code PLM). Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by a purple ribbon model.

partial occupancy. In addition, lipids have complex conformations and complex torsion angles. Whilst saturated lipids are extremely flexible due to free rotation around their many single bonds, unsaturated lipids have *cis–trans* stereochemistry that should be considered; recently, it has been shown that many PDB structures containing lipids have incorrect *cis–trans* stereochemistry (Waibl *et al.*, 2022).

2.9. Lipidation: palmitoylation

Palmitoylation describes the covalent attachment of a 16-carbon fatty-acid palmitoyl group to a protein (Fig. 9, Supplementary Fig. S9). The most common type of palmitoylation is *S*-palmitoylation, which involves the reversible attachment of palmitate from palmitoyl-CoA to the thiol group of a cysteine residue via a thioester linkage by palmitoyl acyltransferases (Li *et al.*, 2022). Less commonly, irreversible *N*-palmitoylation can occur at the N-terminus of Hedgehog proteins (Buglino & Resh, 2008), as well as *O*-palmitoylation, where palmitate is irreversibly added to serine or threonine hydroxyl groups (Gao & Hannoush, 2014). No strict consensus sequence has been identified for palmitoylation; however, *S*-palmitoylated cysteine residues in yeast often exist adjacent to myristoylation or prenylation sites, and are frequently located in the cytoplasmic regions flanking, or within, transmembrane domains (Roth *et al.*, 2006; Salaun *et al.*, 2010). Palmitoylation enhances the hydrophobicity of amino acid residues, which can affect their membrane association (Rocks *et al.*, 2005). Due to the reversible nature of *S*-palmitoylation through palmitoyl acyltransferases and palmitoyl protein thioesterases, proteins have been shown to use cycles of palmitoylation/depalmitoylation to translocate intracellularly from one membrane to another (Rocks *et al.*, 2005). Palmitoylation therefore plays a significant role in protein trafficking, membrane localization and cellular signalling (Smotrys & Linder, 2004). Dysregulated palmitoylation has been shown to play a role in several human diseases, including cancer, neurological disorders and cardiovascular diseases (Ramzan *et al.*, 2023; Kong *et al.*, 2023; Baldwin *et al.*, 2023).

2.10. Lipidation: myristoylation

Myristoylation involves the addition of a 14-carbon saturated fatty-acid myristoyl group to the α-amino group of an N-terminal glycine residue via an amide bond (Fig. 10; Wolven *et al.*, 1998). *N*-Myristoylated proteins generally contain the N-terminal consensus sequence M-G-*X*-*X*-*X*-S/T, where the initiator methionine is removed and the myristate is added to the exposed N-terminal glycine (Wolven *et al.*, 1998). This modification is irreversible and can be added either co-translationally or post-translationally, catalysed by *N*-myristoyltransferases (Wolven *et al.*, 1998). *N*-Myristoylation can also occur on the ε-amino group of internal lysine residues (Supplementary Fig. S10), although this is less common (Kosciuk *et al.*, 2020). Due to the hydrophobic nature of this modification, myristoylation plays an essential role in membrane targeting, protein–protein interactions and regulates a number of signal transduction pathways (Adam *et al.*,

2007; Hu *et al.*, 2010; Maurer-Stroh *et al.*, 2004; Timms *et al.*, 2019). However, myristoylation alone is often insufficient to stably anchor a protein to a membrane; instead a second signal is required, which involves either a cluster of hydrophobic or positively charged amino acids or a covalently attached palmitate moiety (Seykora *et al.*, 1996; Tanaka *et al.*, 1995; Gaffarogullari *et al.*, 2011). The orientation of this modification can be highly dynamic, in which the myristoylated site is either located in a hydrophobic pocket or flipped out and surface-exposed for membrane binding, which is known as a 'myristoyl switch' (Tanaka *et al.*, 1995). Protein myristoylation plays a role in a number of diseases, including cancer, where it has been shown to promote tumorigenesis (Tan *et al.*, 2023), as well as the virulence of HIV infections (Socas & Ambroggio, 2018; Bryant & Ratner, 1990).

2.11. Lipidation: *S*-prenylation

Prenylation involves the irreversible addition of a 15-carbon farnesyl or a 20-carbon geranylgeranyl group to the thiol group of a cysteine residue at the carboxy-terminus via a thioester linkage (Fig. 11, Supplementary Fig. S11). A consensus motif, known as the C*AAX* box, has been identified in most prenylated proteins; in this motif, *A* stands for any aliphatic residue, while *X* represents an amino acid that determines whether the protein undergoes farnesylation or geranylgeranylation (Reid *et al.*, 2004; Seabra *et al.*, 1991). Farnesyl transferase prefers *X* to be methionine, serine, glutamine or cysteine, whereas geranylgeranyl transferase-1 prefers *X* to be leucine or isoleucine, although these rules are not absolute (Reid *et al.*, 2004; Lebowitz & Prendergast, 1998; Boutin *et al.*, 1998; Seabra *et al.*, 1991). Like palmitoylation and myristoylation, *S*-prenylation facilitates membrane association. Whilst geranylgeranylation is sufficiently hydrophobic to facilitate membrane anchoring, farnesylated proteins require a second signal for stable membrane interaction, typically palmitoylation or a cluster of positive amino acids (Cuiffo & Ren, 2010; Hancock *et al.*, 1990). Protein prenylation has implications in diseases such as cancer and diabetes (Borini Etichetti *et al.*, 2020; Gendaszewska-Darmach *et al.*, 2021).

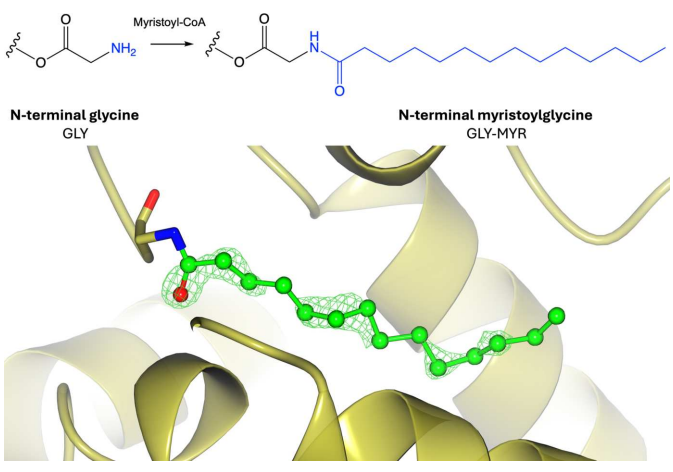


Figure 10
Myristoylation. Top: myristoylation of N-terminal glycine involves the addition of a myristoyl group donated by myristoyl-CoA to the N-terminal amino group. Bottom: myristoylglycine (PDB entry 4zv5; Doležal *et al.*, 2016; CCD code MYR). Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by a yellow ribbon model.

3. PTMs in the Protein Data Bank

3.1. PTM annotation in macromolecular crystallographic files

PTMs have versatile annotations in macromolecular crystallographic files that are dependent on the type of PTM. Small-molecule PTMs are defined as modifications that include ten or fewer atoms (for example phosphorylation, hydroxylation and methylation), whereas long-chain PTMs include greater than ten atoms (glycosylation and lipidation) (wwPDB Processing Procedures and Policies Document, Section A, 2014). For small-molecule PTMs, both the modification and the residue atoms are listed within the same three-letter CCD code and are part of the polymeric sequence of the protein. For example, *SEP* includes both the atoms of the serine residue and the phosphate group involved in the modification. In contrast, for protein glycosylation and most cases of lipidation, the CCD code does not include the residue atoms; instead the modification is covalently linked to the polymer. For example, *FAR* includes atoms of a farnesyl group and does not include the atoms of the cysteine residue to which it is linked.

Within legacy PDB and mmCIF files, the CCD codes of small-molecule PTMs are listed in the modified residue data items (PDB, `MODRES`; mmCIF, `_pdbx_struct_mod_residue`), which include any modified polymer components. For protein glycosylation, the modified amino acid residue is

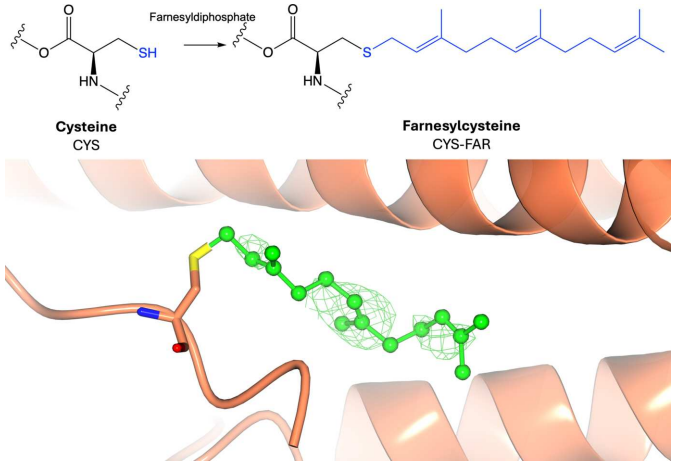


Figure 11
Prenylation. Top: prenylation of cysteine involves the addition of a farnesyl group (or geranylgeranyl group; Supplementary Fig. S9) donated by farnesyl diphosphate (or geranylgeranyl diphosphate) to the side-chain thiol group. Bottom: farnesylcysteine (PDB entry 6k1z; Ji *et al.*, 2019; CCD code FAR). Positive omit density is shown in green at 3σ for the modified residue. The rest of the protein chain is represented by an orange ribbon model.

also listed here, *e.g. ASN*, *SER*, *THR* are listed, without the sugar. In mmCIF files, protein glycosylation is clearly annotated due to unique data items for glycans. Firstly, `_pdbx_entity_branch_link` contains the sugar-chain information, and secondly `_struct_conn.pdbx_role`, contains information regarding the glycan type, with allowed values of *C*-mannosylation, *N*-glycosylation, *O*-glycosylation or *S*-glycosylation. Annotation in crystallographic files differs for lipids; in mmCIF files, lipid CCD codes can be found in a number of data items that highlight special features or structurally relevant sites, including `_struct_site`, `_struct_site_gen` and `_pdbx_entity_instance_feature`, as well as in `_pdbx_entity_nonpoly` and `pdbx_nonpoly_scheme`, which provide information about nonpolymeric components.

PTMs therefore have versatile annotations in macromolecular crystallographic files, with no specific data item dedicated to PTMs, apart from for glycosylation. Even for small-molecule PTMs, the modified-residue data category includes any modified polymer components, which can include coenzymes and synthetic chemical modifications (such as chromophores). Furthermore, there are issues with redundant labelling in lipidation. While most cases are listed as covalently linked to a polymer, there are instances where lipid modifications are annotated similarly to small-molecule PTMs, with both the lipid molecule and residue atoms included within the CCD code. For example, lysine myristoylation is represented by *MYK*, which includes the atoms for the lysine residue and the myristoyl group and is found in the modified-residue category.

3.2. Frequency of PTMs in the PDB

Currently, there is no single data item in mmCIF files that allows for exclusive searches of PTMs in the PDB. As a result, this analysis used individual CCD codes to search for PTMs in the PDB. Initially, a list of PTM-related keywords was constructed using the UniProtKB PTM list, which describes the chemical nature of protein modifications using a controlled vocabulary (UniProt Consortium, 2021); this includes annotations of both RESID (Reference Sequence Identifier; Garavelli, 2004) and CHEBI (Chemical Entities of Biological Interest; Hastings *et al.*, 2016) identifiers. Entries were removed if they were not defined as a PTM by comparison with the dbPTM, although analysis was still performed (Supplementary Fig. S12). Following this, PTMs were matched to their CCD codes, primarily using a substructure search between the CCD and ChEBI databases, using *RDKit* (https://www.rdkit.org). In addition, the RESID was used to search the PDB via the RCSB PDB Search API (Bittrich *et al.*, 2023; Rose *et al.*, 2021) to generate a dataset of individual RESIDs and associated protein structures; subsequently, a compilation of covalently linked CCD codes was identified and linked to each RESID. This allowed the generation of an exhaustive PTM list with corresponding CCD codes, which was then used to search the PDB via the RCSB PDB Search API (Bittrich *et al.*, 2023; Rose *et al.*, 2021).

In total, over 23 000 examples of PDB entries containing post-translationally modified residues were identified. The most common PTM type identified was *N*-glycosylation, which makes up nearly half of the identified PTMs (Fig. 12). This is followed by phosphorylation, methylation and acetylation. Generally, these trends align well with the number of PTM sites reported in the dbPTM (Table 1), with *N*-glycosylation and phosphorylation being the most well studied. In fact, phosphoserine, phosphotyrosine and phosphothreonine are among the most common small-molecule PTMs identified in the PDB (Fig. 13). Methylation and hydroxylation modifications were also identified as some of the most common PTM types in the PDB, with dimethyllysine, trimiethyllysine, acetyllysine and hydroxyproline amongst the most common small-molecule PTMs to be identified. In addition, oxidation has been identified as a common PTM in the PDB, with hydroxycysteine, cysteine sulfinic acid and cysteine sulfonic acid in high abundance. When analysing protein oxidation in the context of crystallography, it is key to recognize that radiation damage can promote cysteine oxidation, meaning that oxidized cysteines may not always be biologically relevant (Close & Bernhard, 2019; Garrido Ruiz *et al.*, 2022). Furthermore, there is a prevalence of acetylation, hydroxylation and *O*-glycosylation, further highlighting these modifications as some of the most well studied, as suggested in the dbPTM (Table 1).

Whilst pyroglutamic acid, carboxylation and formylation were identified as common PTM types in the PDB, they are less common both putatively and experimentally as reported in the dbPTM (Li *et al.*, 2022). For pyroglutamic acid, this may be explained by the spontaneous formation of pyroglutamic acid *in vitro*, where its formation has been attributed to being an artefact during sample preparation and storage of immunoglobulins, which may account for the increased occurrence of this modification in the PDB (Chelius *et al.*, 2006).

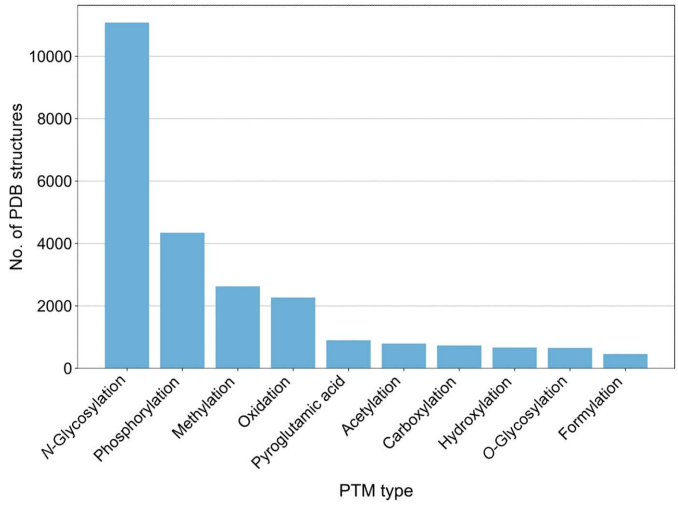


Figure 12
Most common types of PTM in the PDB. The top ten PTM types identified in the PDB are shown. Data were obtained by searching the RCSB PDB Search API using identified CCD codes corresponding to PTMs. Glycosylation data were obtained from the *Privateer* Database (Dialpuri, Bagdonas, Schofield, Pham, Holland, Bond *et al.*, 2024; Dialpuri, Bagdonas, Schofield, Pham, Holland & Agirre, 2024).

Furthermore, the extent of lysine carboxylation, a spontaneous PTM, is not fully understood; however, it has been found in a number of key enzymes, including pyruvate carboxylase, urease, RuBisCo and β-lactamase (Jimenez-Morales *et al.*, 2014; Sheng *et al.*, 2019). Furthermore, *N*-formyl methionine is an amino acid derivative that is key for prokaryotic protein synthesis; it is encoded by the AUG codon, which is the start codon for protein synthesis. It is therefore the N-terminal amino acid of nearly all proteins in prokaryotic systems, which explains its prevalence in the PDB. As a result, it is typically considered to be a pre-translational modification rather than a PTM. Whilst these results are interesting, it is key to consider that as the PDB was used as the primary data source, redundant depositions may also explain the prevalence of less common PTMs.

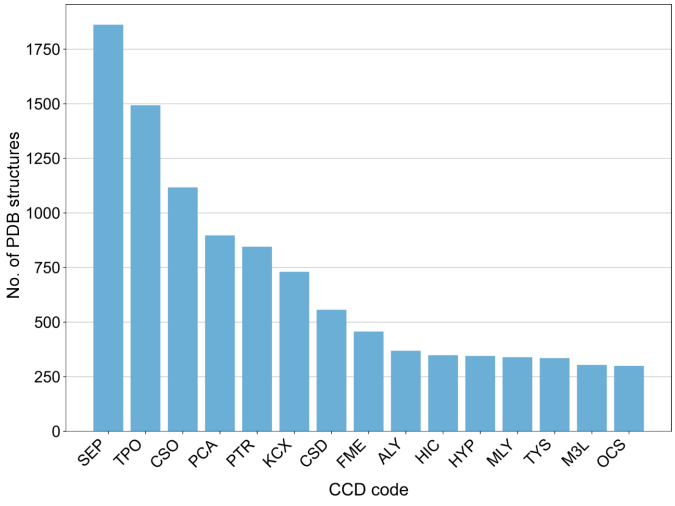


Figure 13
Most common small-molecule post-translationally modified residues in the PDB. The top 15 small-molecule PTMs identified in the PDB are shown. Data were obtained by searching the RCSB PDB Search API using identified CCD codes corresponding to PTMs. These data include CCD codes which are located in the polymeric sequence.

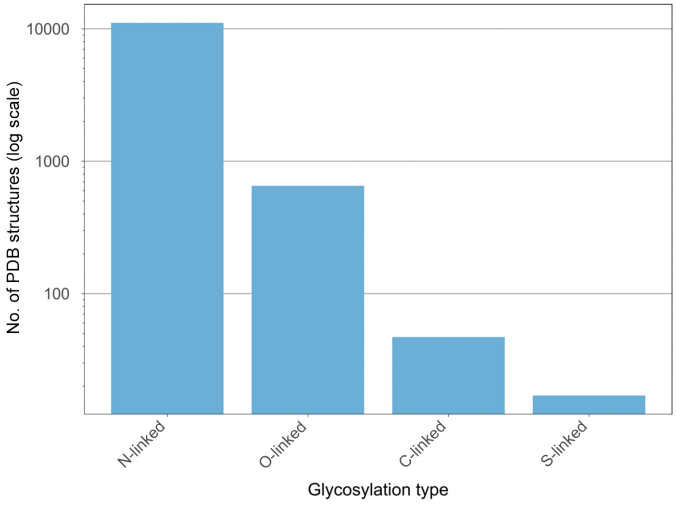


Figure 14
Glycosylation types in the PDB. Bars show the number of PDB structures containing each glycosylation type. The *y* axis is shown in log count. Data were obtained from the *Privateer* database (Dialpuri, Bagdonas, Schofield, Pham, Holland & Agirre, 2024; Dialpuri, Bagdonas, Schofield, Pham, Holland, Bond *et al.*, 2024).

It is well known that *N*-glycans are the most frequent type of glycosylation, followed by *O*-glycans, with fewer examples of *C*- and *S*-glycans (Fig. 14), and the trends shown in the PDB seem to reflect this well. Whilst glycans are often hard to deal with in crystallography, *Privateer* (Agirre *et al.*, 2015; Dialpuri, Bagdonas, Schofield, Pham, Holland, Bond *et al.*, 2024), a software package that identifies and rectifies model errors in protein–glycan structures, has facilitated the modelling of sugars. *Privateer* is able to produce torsion restraints (Atanasova *et al.*, 2022), as well as linkage torsion analysis (Dialpuri *et al.*, 2023), allowing refinement to produce more accurate and representative models with ease. Lipidation PTMs, however, were identified less frequently in the PDB, likely due to their less ordered, flexible structures that can lead to partial occupancy (Fig. 15). Myristoylation and palmitoylation are the most common forms of lipidation identified in the PDB, reflecting well the most common putative and experimental lipidation sites reported in the dbPTM (Table 1).

One factor that should be considered when analysing PTMs in the PDB is the difficulty of searching for isopeptides, a PTM which involves the formation of bonds between amino acid side chains, which are not well annotated in PDB files and have therefore not been captured in this analysis. Another factor that should be considered when analysing PTMs in the PDB is the possibility that protein modifications have been captured by experimental techniques but have not been modelled. *PreLysCar*, a prediction tool for lysine carboxylation, highlights an example (Fig. 16) where a lysine residue is

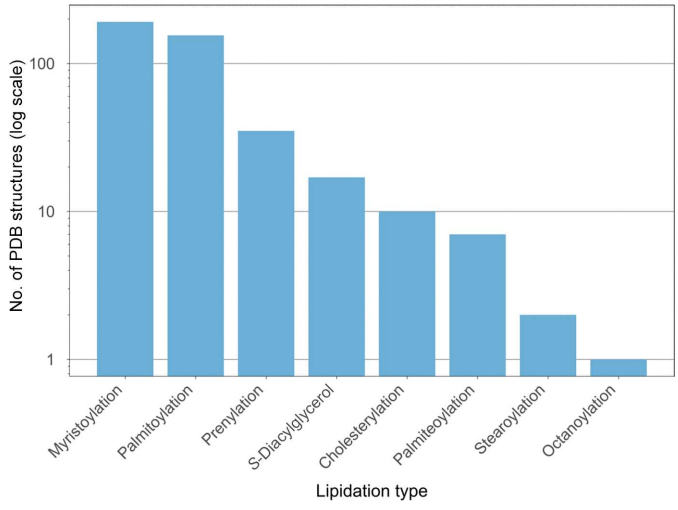


Figure 15
Lipidation types in the PDB. Bars show the number of PDB structures containing each detected lipidation type. The *y* axis is shown in log count. Data were obtained by searching the RCSB PDB Search API using identified CCD codes corresponding to lipid PTMs.

predicted to be carboxylated (Jimenez-Morales *et al.*, 2014). Having inspected the difference density ($mF_o - DF_c$) surrounding the lysine residue, a region of positive difference density surrounding the NZ atom can be identified, into which a carboxyl group can be modelled (Jimenez-Morales *et al.*, 2014). This highlights the idea that there may be unmodelled PTMs in the PDB, where there is density present but they have not been modelled. This suggests that this analysis is almost certainly an underestimation of the number of PTMs in the PDB.

4. Challenges and future directions

Having understood the importance of PTMs in protein structure dynamics, it is clear that they are essential for the understanding of protein structure and function. It is key to recognize the importance of integrating structural data with functional and biochemical data, which provides a comprehensive understanding of how PTMs regulate protein function at the molecular level. This will aid in the understanding of the mechanisms underlying PTM-mediated regulation, which is key for many biochemical processes. A significant challenge facing PTMs in protein structures is the lack of consistent, exclusive annotation within macromolecular structure files. In this study, identifying PTMs in the PDB was primarily performed using a substructure search using *RDKit* (https://www.rdkit.org). Whilst this was effective, some PTMs may have been missed. Fortunately, the wwPDB is in the process of standardizing PTM annotation in mmCIF files for both PDB and CCD entries, making it easier to search for all PTM types by the beginning of 2025 (wwPDB, 2024).

Going forward, several advancements in protein structure-determination techniques will aid in the investigation of PTMs. We can now capture heterogeneous protein populations, which is particularly useful in the study of PTMs. Specifically, this can be performed using methods such as cryo-EM SPA, which allows the reconstruction of protein structures from thousands of individual images, as well as serial femtosecond crystallography, which allows diffraction patterns to be obtained for individual molecules. As techniques for protein structure determination improve, it is important to develop semi-automatic tools for PTM identification, modelling and validation. This will allow users to edit their structures quickly and will speed up a once time-consuming task. In conclusion, while there are several challenges in the study of PTMs at the structural level, advancements in structure-determination techniques, the improvement of PTM annotation in PDBx/mmCIF model files, and the increasing awareness of their importance will facilitate the study of PTMs going forward.

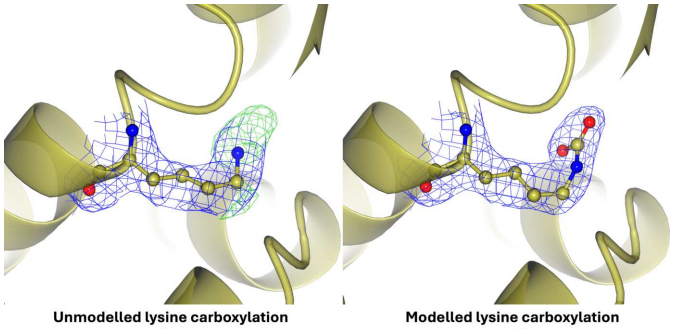


Figure 16
Unmodelled PTMs in the PDB. Density for protein modifications may be present but left unmodelled. (*a*) Lys84 has positive difference density surrounding the side-chain amino group (PDB entry 2jc7; Santillana *et al.*, 2007). (*b*) Addition of a carboxyl group to the NZ atom of Lys84 [using *Coot* (Emsley *et al.*, 2010) and *AceDRG* (Long *et al.*, 2017)], followed by refinement with *REFMAC* (Murshudov *et al.*, 2011). $2mF_o - DF_c$ electron density is shown in blue at 1σ for the residue. Positive difference density ($mF_o - DF_c$) is shown in green contoured to 4σ and clipped within 4 Å of Lys84. The rest of the protein chain is represented by a yellow ribbon model.

5. Related literature

The following references are cited in the supporting information for this article: Bokhovchuk *et al.* (2023), Ferrara *et al.* (2011), Gokulan *et al.* (2013), Hilgers & Ludwig (2001), Huang *et al.* (2023), Kumar *et al.* (2022), Lee & Paetzel (2011), Merő *et al.* (2019), Oakley *et al.* (2002) and Yang *et al.* (2013).

Acknowledgements

We are grateful to the members of our research teams (York and MRC–LMB Cambridge) for their input. We would also like to thank Sameer Velankar, Deborah Harrus and Marcus Bage at the PDBe, EMBL–EBI, Cambridge. We would like to thank Nicholas Pearce, Helen Ginn and Clemens Vonrhein for the kind invitation to speak at the 2024 CCP4 Study Weekend and contribute to this proceedings issue.

Funding information

Lucy Schofield is funded by STFC/CCP4 PhD studentship agreement 4462290 (York)/S2 2024 012 (STFC). Jordan Dialpuri is funded by the BBSRC (grant No. BB/T0072221), Jon Agirre is a Royal Society University Research Fellow (awards UF160039 and URF\R\221006) and Garib Murshudov is funded by MRC grant MC_UP_A025_1012.

References

Adam, R. M., Mukhopadhyay, N. K., Kim, J., Di Vizio, D., Cinar, B., Boucher, K., Solomon, K. R. & Freeman, M. R. (2007). *Cancer Res.* **67**, 6238–6246.
Agirre, J. (2017). *Acta Cryst.* D**73**, 171–186.
Agirre, J., Ariza, A., Offen, W. A., Turkenburg, J. P., Roberts, S. M., McNicholas, S., Harris, P. V., McBrayer, B., Dohnalek, J., Cowtan, K. D., Davies, G. J. & Wilson, K. S. (2016). *Acta Cryst.* D**72**, 254–265.
Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C. B., Baslé, A., Beilsten-Edmands, J., Borges, R. J., Brown, D. G., Burgos-Mármol, J. J., Berrisford, J. M., Bond, P. S., Caballero, I., Catapano, L., Chojnowski, G., Cook, A. G., Cowtan, K. D., Croll, T. I., Debreczeni, J. É., Devenish, N. E., Dodson, E. J., Drevon, T. R., Emsley, P., Evans, G., Evans, P. R., Fando, M., Foadi, J., Fuentes-Montero, L.,

Garman, E. F., Gerstel, M., Gildea, R. J., Hatti, K., Hekkelman, M. L., Heuser, P., Hoh, S. W., Hough, M. A., Jenkins, H. T., Jiménez, E., Joosten, R. P., Keegan, R. M., Keep, N., Krissinel, E. B., Kolenko, P., Kovalevskiy, O., Lamzin, V. S., Lawson, D. M., Lebedev, A. A., Leslie, A. G. W., Lohkamp, B., Long, F., Malý, M., McCoy, A. J., McNicholas, S. J., Medina, A., Millán, C., Murray, J. W., Murshudov, G. N., Nicholls, R. A., Noble, M. E. M., Oeffner, R., Pannu, N. S., Parkhurst, J. M., Pearce, N., Pereira, J., Perrakis, A., Powell, H. R., Read, R. J., Rigden, D. J., Rochira, W., Sammito, M., Sánchez Rodríguez, F., Sheldrick, G. M., Shelley, K. L., Simkovic, F., Simpkin, A. J., Skubak, P., Sobolev, E., Steiner, R. A., Stevenson, K., Tews, I., Thomas, J. M. H., Thorn, A., Valls, J. T., Uski, V., Usón, I., Vagin, A., Velankar, S., Vollmar, M., Walden, H., Waterman, D., Wilson, K. S., Winn, M. D., Winter, G., Wojdyr, M. & Yamashita, K. (2023). *Acta Cryst.* D**79**, 449–461.

Agirre, J., Davies, G. J., Wilson, K. S. & Cowtan, K. D. (2017). *Curr. Opin. Struct. Biol.* **44**, 39–47.

Agirre, J., Iglesias-Fernández, J., Rovira, C., Davies, G. J., Wilson, K. S. & Cowtan, K. D. (2015). *Nat. Struct. Mol. Biol.* **22**, 833–834.

An, H. J., Froehlich, J. W. & Lebrilla, C. B. (2009). *Curr. Opin. Chem. Biol.* **13**, 421–426.

Atanasova, M., Bagdonas, H. & Agirre, J. (2020). *Curr. Opin. Struct. Biol.* **62**, 70–78.

Atanasova, M., Nicholls, R. A., Joosten, R. P. & Agirre, J. (2022). *Acta Cryst.* D**78**, 455–465.

Bai, A. H. C., Wu, W. K. K., Xu, L., Wong, S. H., Go, M. Y., Chan, A. W. H., Harbord, M., Zhang, S., Chen, M., Wu, J. C. Y., Chan, M. W. Y., Chan, M. T. V., Chan, F. K. L., Sung, J. J. Y., Yu, J., Cheng, A. S. L. & Ng, S. C. (2016). *J. Crohns Colitis*, **10**, 726–734.

Baldwin, T. A., Teuber, J. P., Kuwabara, Y., Subramani, A., Lin, S. J., Kanisicak, O., Vagnozzi, R. J., Zhang, W., Brody, M. J. & Molkentin, J. D. (2023). *J. Biol. Chem.* **299**, 105426.

Bannister, A. J. & Kouzarides, T. (2011). *Cell Res.* **21**, 381–395.

Basheer, N., Smolek, T., Hassan, I., Liu, F., Iqbal, K., Zilka, N. & Novak, P. (2023). *Mol. Psychiatry*, **28**, 2197–2214.

Bedem, H. van den & Wilson, M. A. (2019). *J. Synchrotron Rad.* **26**, 958–966.

Behnia, R., Panic, B., Whyte, J. R. C. & Munro, S. (2004). *Nat. Cell Biol.* **6**, 405–413.

Berlett, B. S., Friguet, B., Yim, M. B., Chock, P. B. & Stadtman, E. R. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 1776–1780.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Betts, M. J., Wichmann, O., Utz, M., Andre, T., Petsalaki, E., Minguez, P., Parca, L., Roth, F. P., Gavin, A.-C., Bork, P. & Russell, R. B. (2017). *PLoS Comput. Biol.* **13**, e1005462.

Beusekom, B. van, Lütteke, T. & Joosten, R. P. (2018). *Acta Cryst.* F**74**, 463–472.

Bhullar, K. S., Lagarón, N. O., McGowan, E. M., Parmar, I., Jha, A., Hubbard, B. P. & Rupasinghe, H. P. V. (2018). *Mol. Cancer*, **17**, 48.

Bittrich, S., Bhikadiya, C., Bi, C., Chao, H., Duarte, J. M., Dutta, S., Fayazi, M., Henry, J., Khokhriakov, I., Lowe, R., Piehl, D. W., Segura, J., Vallat, B., Voigt, M., Westbrook, J. D., Burley, S. K. & Rose, Y. (2023). *J. Mol. Biol.* **435**, 167994.

Blom, N., Gammeltoft, S. & Brunak, S. (1999). *J. Mol. Biol.* **294**, 1351–1362.

Bludau, I., Willems, S., Zeng, W.-F., Strauss, M. T., Hansen, F. M., Tanzer, M. C., Karayel, O., Schulman, B. A. & Mann, M. (2022). *PLoS Biol.* **20**, e3001636.

Bokhovchuk, F., Mesrouze, Y., Meyerhofer, M., Fontana, P., Zimmermann, C., Villard, F., Erdmann, D., Kallen, J., Scheufler, C., Velez-Vega, C. & Chène, P. (2023). *Protein Sci.* **32**, e4545.

Borini Etichetti, C. M., Arel Zalazar, E., Cocordano, N. & Girardini, J. (2020). *Front. Oncol.* **10**, 595034.

Boutin, J. A., Marande, W., Goussard, M., Loynel, A., Canet, E. & Fauchere, J.-L. (1998). *Arch. Biochem. Biophys.* **354**, 83–94.

Brandt, A., Glanville, R. W., Hörlein, D., Bruckner, P., Timpl, R., Fietzek, P. P. & Kühn, K. (1984). *Biochem. J.* **219**, 625–634.

Brobbey, C., Liu, L., Yin, S. & Gan, W. (2022). *Int. J. Mol. Sci.* **23**, 9780.

Bruick, R. K. & McKnight, S. L. (2001). *Science*, **294**, 1337–1340.

Bryant, M. & Ratner, L. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 523–527.

Buglino, J. A. & Resh, M. D. (2008). *J. Biol. Chem.* **283**, 22076–22088.

Canet-Avilés, R. M., Wilson, M. A., Miller, D. W., Ahmad, R., McLendon, C., Bandyopadhyay, S., Baptista, M. J., Ringe, D., Petsko, G. A. & Cookson, M. R. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 9103–9108.

Chandran, S. & Binninger, D. (2023). *Antioxidants*, **13**, 21.

Chelius, D., Jing, K., Lueras, A., Rehder, D. S., Dillon, T. M., Vizel, A., Rajan, R. S., Li, T., Treuheit, M. J. & Bondarenko, P. V. (2006). *Anal. Chem.* **78**, 2370–2376.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V. & Mann, M. (2009). *Science*, **325**, 834–840.

Claiborne, A., Yeh, J. I., Mallett, T. C., Luba, J., Crane, E. J., Charrier, V. & Parsonage, D. (1999). *Biochemistry*, **38**, 15407–15416.

Close, D. M. & Bernhard, W. A. (2019). *J. Synchrotron Rad.* **26**, 945–957.

Connell, G. E. & Hanes, C. S. (1956). *Nature*, **177**, 377–378.

Copeland, R. A. (2018). *Phil. Trans. R. Soc. B*, **373**, 20170080.

Crosby, H. A. & Escalante-Semerena, J. C. (2014). *J. Bacteriol.* **196**, 1496–1504.

Crosby, H. A., Pelletier, D. A., Hurst, G. B. & Escalante-Semerena, J. C. (2012). *J. Biol. Chem.* **287**, 15590–15601.

Cuiffo, B. & Ren, R. (2010). *Blood*, **115**, 3598–3605.

Cummins, P. M. & O'Connor, B. (1998). *Biochim. Biophys. Acta*, **1429**, 1–17.

Daily, K. M., Radivojac, P. & Dunker, A. K. (2005). *2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pp. 1–7. PIscataway: IEEE.

Davie, J. R. (1998). *Curr. Opin. Genet. Dev.* **8**, 173–178.

Davies, A. M., Beavil, R. L., Barbolov, M., Sandhar, B. S., Gould, H. J., Beavil, A. J., Sutton, B. J. & McDonnell, J. M. (2023). *Mol. Immunol.* **159**, 28–37.

Dekker, F. J., van den Bosch, T. & Martin, N. I. (2014). *Drug Discov. Today*, **19**, 654–660.

Deller, M. C., Kong, L. & Rupp, B. (2016). *Acta Cryst.* F**72**, 72–95.

Despres, C., Byrne, C., Qi, H., Cantrelle, F.-X., Huvent, I., Chambraud, B., Baulieu, E.-E., Jacquot, Y., Landrieu, I., Lippens, G. & Smet-Nocca, C. (2017). *Proc. Natl Acad. Sci. USA*, **114**, 9080–9085.

Dialpuri, J. S., Bagdonas, H., Atanasova, M., Schofield, L. C., Hekkelman, M. L., Joosten, R. P. & Agirre, J. (2023). *Acta Cryst.* D**79**, 462–472.

Dialpuri, J. S., Bagdonas, H., Schofield, L. C., Pham, P. T., Holland, L. & Agirre, J. (2024). *Beilstein J. Org. Chem.* **20**, 931–939.

Dialpuri, J. S., Bagdonas, H., Schofield, L. C., Pham, P. T., Holland, L., Bond, P. S., Sánchez Rodríguez, F., McNicholas, S. J. & Agirre, J. (2024). *Acta Cryst.* F**80**, 30–35.

Diaz, K., Meng, Y. & Huang, R. (2021). *Curr. Opin. Chem. Biol.* **63**, 115–122.

Doležal, M., Zábranský, A., Dostál, J., Vaněk, O., Brynda, J., Lepšík, M., Hadravová, R. & Pichová, I. (2016). *Retrovirology*, **13**, 2.

Durek, P., Schudoma, C., Weckwerth, W., Selbig, J. & Walther, D. (2009). *BMC Bioinformatics*, **10**, 117.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Ferrara, C., Grau, S., Jäger, C., Sondermann, P., Brünker, P., Waldhauer, I., Hennig, M., Ruf, A., Rufer, A. C., Stihle, M., Umaña, P. & Benz, J. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 12669–12674.

Gaffarogullari, E. C., Masterson, L. R., Metcalfe, E. E., Traaseth, N. J., Balatri, E., Musa, M. M., Mullen, D., Distefano, M. D. & Veglia, G. (2011). *J. Mol. Biol.* **411**, 823–836.

Gao, X. & Hannoush, R. N. (2014). *Nat. Chem. Biol.* **10**, 61–68.

Garavelli, J. S. (2004). *Proteomics*, **4**, 1527–1533.
Garrido Ruiz, D., Sandoval-Perez, A., Rangarajan, A. V., Gunderson, E. L. & Jacobson, M. P. (2022). *Biochemistry*, **61**, 2165–2176.
Gendaszewska-Darmach, E., Garstka, M. A. & Błażewska, K. M. (2021). *J. Med. Chem.* **64**, 9677–9710.
Gokulan, K., O'Leary, S. E., Russell, W. K., Russell, D. H., Lalgondar, M., Begley, T. P., Ioerger, T. R. & Sacchettini, J. C. (2013). *J. Biol. Chem.* **288**, 16484–16494.
Gunn, A. P., Masters, C. L. & Cherny, R. A. (2010). *Int. J. Biochem. Cell. Biol.* **42**, 1915–1918.
Hancock, J. F., Paterson, H. & Marshall, C. J. (1990). *Cell*, **63**, 133–139.
Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P. & Steinbeck, C. (2016). *Nucleic Acids Res.* **44**, D1214–D1219.
Hilgers, M. T. & Ludwig, M. L. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 11169–11174.
Hinkle, P. M. & Tashjian, A. H. Jr (1973). *J. Biol. Chem.* **248**, 6180–6186.
Hoshi, T. & Heinemann, S. (2001). *J. Physiol.* **531**, 1–11.
Hu, M., He, F., Thompson, E. W., Ostrikov, K. K. & Dai, X. (2022). *Cancers* **14**, 346.
Hu, T., Li, C., Cao, Z., Van Raay, T. J., Smith, J. G., Willert, K., Solnica-Krezel, L. & Coffey, R. J. (2010). *J. Biol. Chem.* **285**, 13561–13568.
Huang, Q., Chen, X., Wang, Y., Li, J., Liu, H., Xie, Y., Dai, Z., Zou, X. & Li, Z. (2020). *Chemom. Intell. Lab. Syst.* **202**, 104035.
Huang, R., Warner Jenkins, G., Kim, Y., Stanfield, R. L., Singh, A., Martinez-Yamout, M., Kroon, G. J., Torres, J. L., Jackson, A. M., Kelley, A., Shaabani, N., Zeng, B., Bacica, M., Chen, W., Warner, C., Radoicic, J., Joh, J., Dinali Perera, K., Sang, H., Kim, T., Yao, J., Zhao, F., Sok, D., Burton, D. R., Allen, J., Harriman, W., Mwangi, W., Chung, D., Teijaro, J. R., Ward, A. B., Dyson, H. J., Wright, P. E., Wilson, I. A., Chang, K.-O., McGregor, D. & Smider, V. V. (2023). *Proc. Natl Acad. Sci. USA*, **120**, e2303455120.
Huang, Y., Yang, C., Xu, X.-F., Xu, W. & Liu, S.-W. (2020). *Acta Pharmacol. Sin.* **41**, 1141–1149.
Hwang, C.-S., Shemorry, A. & Varshavsky, A. (2010). *Science*, **327**, 973–977.
Ismail, H. D., Newman, R. H. & Kc, D. B. (2016). *Mol. Biosyst.* **12**, 2427–2435.
Ji, C., Du, S., Li, P., Zhu, Q., Yang, X., Long, C., Yu, J., Shao, F. & Xiao, J. (2019). *PLoS Pathog.* **15**, e1007876.
Jimenez-Morales, D., Adamian, L., Shi, D. & Liang, J. (2014). *Acta Cryst.* D**70**, 48–57.
Johnson, L. N. & Lewis, R. J. (2001). *Chem. Rev.* **101**, 2209–2242.
Joosten, R. P. & Agirre, J. (2022). *PLoS Biol.* **20**, e3001673.
Kong, Y., Liu, Y., Li, X., Rao, M., Li, D., Ruan, X., Li, S., Jiang, Z. & Zhang, Q. (2023). *J. Transl. Med.* **21**, 826.
Kosciuk, T., Price, I. R., Zhang, X., Zhu, C., Johnson, K. N., Zhang, S., Halaby, S. L., Komaniecki, G. P., Yang, M., DeHart, C. J., Thomas, P. M., Kelleher, N. L., Fromme, J. C. & Lin, H. (2020). *Nat. Commun.* **11**, 1067.
Koufaris, C. & Kirmizis, A. (2020). *Cancers* **12**, 2631.
Kühnel, K., Blankenfeldt, W., Terner, J. & Schlichting, I. (2006). *J. Biol. Chem.* **281**, 23990–23998.
Kumar, A. & Bachhawat, A. K. (2012). *Curr. Sci.* **102**, 288–297.
Kumar, A., Narayanan, V. & Sekhar, A. (2020). *Biochemistry*, **59**, 57–73.
Kumar, D., Jha, B., Bhatia, I., Ashraf, A., Dwivedy, A. & Biswal, B. K. (2022). *Proteins*, **90**, 3–17.
Kumar, P., Joy, J., Pandey, A. & Gupta, D. (2017). *PLoS One*, **12**, e0183318.
Kumar, S., Rezaei-Ghaleh, N., Terwel, D., Thal, D. R., Richard, M., Hoch, M., Mc Donald, J. M., Wüllner, U., Glebov, K., Heneka, M. T., Walsh, D. M., Zweckstetter, M. & Walter, J. (2011). *EMBO J.* **30**, 2255–2265.
Lebowitz, P. F. & Prendergast, G. C. (1998). *Oncogene*, **17**, 1439–1445.
Lee, J. & Paetzel, M. (2011). *Acta Cryst.* F**67**, 188–192.
Leutert, M., Entwisle, S. W. & Villén, J. (2021). *Mol. Cell. Proteomics*, **20**, 100129.
Li, Z., Li, S., Luo, M., Jhong, J.-H., Li, W., Yao, L., Pang, Y., Wang, Z., Wang, R., Ma, R., Yu, J., Huang, Y., Zhu, X., Cheng, Q., Feng, H., Zhang, J., Wang, C., Hsu, J. B.-K., Chang, W.-C., Wei, F.-X., Huang, H.-D. & Lee, T.-Y. (2022). *Nucleic Acids Res.* **50**, D471–D479.
Liu, D., Ren, D., Huang, H., Dankberg, J., Rosenfeld, R., Cocco, M. J., Li, L., Brems, D. N. & Remmele, R. L. Jr (2008). *Biochemistry*, **47**, 5088–5100.
Liu, R., Zhao, E., Yu, H., Yuan, C., Abbas, M. N. & Cui, H. (2023). *Signal Transduct. Target. Ther.* **8**, 310.
Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A. & Murshudov, G. N. (2017). *Acta Cryst.* D**73**, 112–122.
Lorton, B. M. & Shechter, D. (2019). *Cell. Mol. Life Sci.* **76**, 2933–2956.
Lundby, A., Lage, K., Weinert, B. T., Bekker-Jensen, D. B., Secher, A., Skovgaard, T., Kelstrup, C. D., Dmytriyev, A., Choudhary, C., Lundby, C. & Olsen, J. V. (2012). *Cell Rep.* **2**, 419–431.
Małecki, J. M., Davydova, E. & Falnes, P. Ø. (2022). *J. Biol. Chem.* **298**, 101791.
Maurer-Stroh, S., Gouda, M., Novatchkova, M., Schleiffer, A., Schneider, G., Sirota, F. L., Wildpaner, M., Hayashi, N. & Eisenhaber, F. (2004). *Genome Biol.* **5**, R21.
Mayr, B. & Montminy, M. (2001). *Nat. Rev. Mol. Cell Biol.* **2**, 599–609.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
Merő, B., Radnai, L., Gógl, G., Tőke, O., Leveles, I., Koprivanacz, K., Szeder, B., Dülk, M., Kudlik, G., Vas, V., Cserkaszky, A., Sipeki, S., Nyitray, L., Vértessy, B. G. & Buday, L. (2019). *J. Biol. Chem.* **294**, 4608–4620.
Miller, C. J. & Turk, B. E. (2016). *Methods Mol. Biol.* **1360**, 203–216.
Mulinacci, F., Capelle, M. A. H., Gurny, R., Drake, A. F. & Arvinte, T. (2011). *J. Pharm. Sci.* **100**, 451–463.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Narasumani, M. & Harrison, P. M. (2018). *PLoS Comput. Biol.* **14**, e1006349.
Nedospasov, A., Rafikov, R., Beda, N. & Nudler, E. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 13543–13548.
Oakley, A. J., Prokop, Z., Boháč, M., Kmuníček, J., Jedlička, T., Monincová, M., Kutá-Smatanová, I., Nagata, Y., Damborský, J. & Wilce, M. C. J. (2002). *Biochemistry*, **41**, 4847–4855.
Okanishi, H., Kim, K., Masui, R. & Kuramitsu, S. (2013). *J. Proteome Res.* **12**, 3952–3968.
Pang, C. N. I., Hayen, A. & Wilkins, M. R. (2007). *J. Proteome Res.* **6**, 1833–1845.
Purushothaman, A., Mohajeri, M. & Lele, T. P. (2023). *J. Biol. Chem.* **299**, 102935.
Qin, W., Ugur, E., Mulholland, C. B., Bultmann, S., Solovei, I., Modic, M., Smets, M., Wierer, M., Forné, I., Imhof, A., Cardoso, M. C. & Leonhardt, H. (2021). *Nucleic Acids Res.* **49**, 7406–7423.
Quevillon-Cheruel, S., Leulliot, N., Muniz, C. A., Vincent, M., Gallay, J., Argentini, M., Cornu, D., Boccard, F., Lemaître, B. & van Tilbeurgh, H. (2009). *J. Biol. Chem.* **284**, 3552–3562.
Ramzan, F., Abrar, F., Mishra, G. G., Liao, L. M. Q. & Martin, D. D. O. (2023). *Front. Physiol.* **14**, 1166125.
Reid, T. S., Terry, K. L., Casey, P. J. & Beese, L. S. (2004). *J. Mol. Biol.* **343**, 417–433.
Rieloff, E. & Skepö, M. (2020). *J. Chem. Theory Comput.* **16**, 1924–1935.
Rocks, O., Peyker, A., Kahms, M., Verveer, P. J., Koerner, C., Lumbierres, M., Kuhlmann, J., Waldmann, H., Wittinghofer, A. & Bastiaens, P. I. H. (2005). *Science*, **307**, 1746–1752.
Rose, Y., Duarte, J. M., Lowe, R., Segura, J., Bi, C., Bhikadiya, C., Chen, L., Rose, A. S., Bittrich, S., Burley, S. K. & Westbrook, J. D. (2021). *J. Mol. Biol.* **433**, 166704.

Roth, A. F., Wan, J., Bailey, A. O., Sun, B., Kuchar, J. A., Green, W. N., Phinney, B. S., Yates, J. R. & Davis, N. G. (2006). *Cell*, **125**, 1003–1013.
Sadowsky, J. D., Burlingame, M. A., Wolan, D. W., McClendon, C. L., Jacobson, M. P. & Wells, J. A. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 6056–6061.
Salaun, C., Greaves, J. & Chamberlain, L. H. (2010). *J. Cell Biol.* **191**, 1229–1238.
Sanchez, R., Riddle, M., Woo, J. & Momand, J. (2008). *Protein Sci.* **17**, 473–481.
Santillana, E., Beceiro, A., Bou, G. & Romero, A. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 5354–5359.
Scott, D. C., Monda, J. K., Bennett, E. J., Harper, J. W. & Schulman, B. A. (2011). *Science*, **334**, 674–678.
Seabra, M. C., Reiss, Y., Casey, P. J., Brown, M. S. & Goldstein, J. L. (1991). *Cell*, **65**, 429–434.
Seykora, J. T., Myat, M. M., Allen, L. A., Ravetch, J. V. & Aderem, A. (1996). *J. Biol. Chem.* **271**, 18797–18802.
Sheng, X., Hou, Q. & Liu, Y. (2019). *Theor. Chem. Acc.* **138**, 17.
Singh, V., Ram, M., Kumar, R., Prasad, R., Roy, B. K. & Singh, K. K. (2017). *Protein J.* **36**, 1–6.
Smotrys, J. E. & Linder, M. E. (2004). *Annu. Rev. Biochem.* **73**, 559–587.
Socas, L. B. P. & Ambroggio, E. E. (2018). *Langmuir*, **34**, 6051–6062.
Stawikowski, M. J., Aukszi, B., Stawikowska, R., Cudic, M. & Fields, G. B. (2014). *J. Biol. Chem.* **289**, 21591–21604.
Strocchi, S., Reggiani, F., Gobbi, G., Ciarrocchi, A. & Sancisi, V. (2022). *Oncogene*, **41**, 3665–3679.
Suga, A., Nagae, M. & Yamaguchi, Y. (2018). *Glycobiology*, **28**, 774–785.
T, D., Venkatraman, P. & Vemparala, S. (2018). *Sci. Rep.* **8**, 12976.
Tak, I.-R., Ali, F., Dar, J. S., Magray, A. R., Ganai, B. A. & Chishti, M. Z. (2019). *Protein Modificomics: From Modifications to Clinical Perspectives*, edited by T. A. Dar & L. R. Singh, pp. 1–35. New York: Academic Press.
Tan, X.-P., He, Y., Yang, J., Wei, X., Fan, Y.-L., Zhang, G.-G., Zhu, Y.-D., Li, Z.-Q., Liao, H.-X., Qin, D.-J., Guan, X.-Y. & Li, B. (2023). *Signal Transduct. Target. Ther.* **8**, 14.
Tanaka, T., Amest, J. B., Harvey, T. S., Stryer, L. & lkura, M. (1995). *Nature*, **376**, 444–447.
Taylor, T. C., Backlund, A., Bjorhall, K., Spreitzer, R. J. & Andersson, I. (2001). *J. Biol. Chem.* **276**, 48159–48164.
Thomas, D., Rathinavel, A. K. & Radhakrishnan, P. (2021). *Biochim. Biophys. Acta*, **1875**, 188464.
Timms, R. T., Zhang, Z., Rhee, D. Y., Harper, J. W., Koren, I. & Elledge, S. J. (2019). *Science*, **365**, eaaw4912.
Trexler, A. J. & Rhoades, E. (2012). *Protein Sci.* **21**, 601–605.
Turell, L., Botti, H., Carballal, S., Ferrer-Sueta, G., Souza, J. M., Durán, R., Freeman, B. A., Radi, R. & Alvarez, B. (2008). *Biochemistry*, **47**, 358–367.
UniProt Consortium (2021). *Nucleic Acids Res.* **49**, D480–D489.
Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L. & Seeberger, P. H. (2022). Editors. *Essentials of Glycobiology*. New York: Cold Spring Harbor Laboratory Press.
Varland, S., Brønstad, K. M., Skinner, S. J. & Arnesen, T. (2023). *Am. J. Med. Genet. A*, **191**, 2402–2410.
Varland, S., Osberg, C. & Arnesen, T. (2015). *Proteomics*, **15**, 2385–2401.
Varland, S., Silva, R. D., Kjosås, I., Faustino, A., Bogaert, A., Billmann, M., Boukhatmi, H., Kellen, B., Costanzo, M., Drazic, A., Osberg, C., Chan, K., Zhang, X., Tong, A. H. Y., Andreazza, S., Lee, J. J., Nedyalkova, L., Ušaj, M., Whitworth, A. J., Andrews, B. J., Moffat, J., Myers, C. L., Gevaert, K., Boone, C., Martinho, R. G. & Arnesen, T. (2023). *Nat Commun*, **14**, 6774.
Varma, S., Orgel, J. P. R. O. & Schieber, J. D. (2021). *Int. J. Mol. Sci.* **22**, 9068.
Venne, A. S., Kollipara, L. & Zahedi, R. P. (2014). *Proteomics*, **14**, 513–524.
Waibl, F., Liedl, K. R. & Rupp, B. (2022). *FEBS J.* **289**, 2793–2804.
Wang, Y., Hou, C., Wisler, J., Singh, K., Wu, C., Xie, Z., Lu, Q. & Zhou, Z. (2019). *J. Diab. Invest*, **10**, 51–61.
Wardman, P. & von Sonntag, C. (1995). *Methods Enzymol.* **251**, 31–45.
Wei, H. & Zhou, M.-M. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 18433–18438.
Wei, H.-H., Fan, X.-J., Hu, Y., Tian, X.-X., Guo, M., Mao, M.-W., Fang, Z.-Y., Wu, P., Gao, S.-X., Peng, C., Yang, Y. & Wang, Z. (2021). *Sci. Bull. (Beijing)*, **66**, 1342–1357.
Weinert, B. T., Iesmantavicius, V., Wagner, S. A., Schölz, C., Gummesson, B., Beli, P., Nyström, T. & Choudhary, C. (2013). *Mol. Cell*, **51**, 265–272.
Weinert, B. T., Wagner, S. A., Horn, H., Henriksen, P., Liu, W. R., Olsen, J. V., Jensen, L. J. & Choudhary, C. (2011). *Sci. Signal.* **4**, ra48.
Westbrook, J. D., Shao, C., Feng, Z., Zhuravleva, M., Velankar, S. & Young, J. (2015). *Bioinformatics*, **31**, 1274–1278.
Wilkins, M. R., Gasteiger, E., Gooley, A. A., Herbert, B. R., Molloy, M. P., Binz, P. A., Ou, K., Sanchez, J. C., Bairoch, A., Williams, K. L. & Hochstrasser, D. F. (1999). *J. Mol. Biol.* **289**, 645–657.
Wolven, A., van't Hof, W. & Resh, M. D. (1998). *Methods Mol. Biol.* **84**, 261–266.
Wooderchak, W. L., Zang, T., Zhou, Z. S., Acuña, M., Tahara, S. M. & Hevel, J. M. (2008). *Biochemistry*, **47**, 9456–9466.
Wu, M., Yang, Y., Wang, H. & Xu, Y. (2019). *BMC Bioinformatics*, **20**, 49.
wwPDB (2014). *wwPDB Processing Procedures, Policies Document Section A*. https://www.wwpdb.org/documentation/policy
wwPDB (2024). *wwPDB: 2024 News*. https://www.wwpdb.org/news/news?year=2024&662fa1a565d8a8cacaf76f9c#662fa1a565d8a8cacaf76f9c.
Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z. & Uversky, V. N. (2007). *J. Proteome Res.* **6**, 1917–1932.
Yang, D., Fang, Q., Wang, M., Ren, R., Wang, H., He, M., Sun, Y., Yang, N. & Xu, R.-M. (2013). *Nat. Struct. Mol. Biol.* **20**, 1116–1118.
Yang, R., Jain, T., Lynaugh, H., Nobrega, R. P., Lu, X., Boland, T., Burnina, I., Sun, T., Caffry, I., Brown, M., Zhi, X., Lilov, A. & Xu, Y. (2017). *mAbs*, **9**, 646–653.
Zhang, L., Zou, Y., He, N., Chen, Y., Chen, Z. & Li, L. (2020). *Front. Cell Dev. Biol.* **8**, 580217.