REVIEW

A glance at post-translational modifications of human thyroglobulin: potential impact on function and pathogenesis

Laura Tosatto[ID] and **Francesca Coscia**[ID]

Human Technopole, Milano, Italy

Correspondence should be addressed to F Coscia: francesca.coscia@fht.org

Abstract

Thyroid hormones are essential for the metabolism of vertebrates and their synthesis, storage and release in the thyroid gland are orchestrated by their large protein precursor thyroglobulin (Tg). Alterations of Tg structure and localisation often correlate with major thyroid disorders. Namely, Tg is the main antigen in autoimmune thyroid diseases, and mutations in its gene are one of the causes of congenital hypothyroidism. Post-translational modifications (PTMs) are crucial for Tg surface properties and may be affected by the disease microenvironment; yet, their role in thyroid homeostasis and pathogenesis remains elusive. The advance of electron cryo-microscopy (cryo-EM) has recently enabled the structure of Tg to be revealed in the un-iodinated and iodinated states. Moreover, *ad hoc* proteomic analyses have lately identified new PTMs in Tg. Here, we provide an overview of the Tg cryo-EM models obtained so far, and we build a three-dimensional map of known PTMs in Tg. Based on their location, we suggest the potential implication of each PTM in hormonogenesis, interactions with cellular partners, colloid cross-linking and hormone release. In addition, several PTMs overlap with immunogenic regions and pathogenic gene mutations. Hence, our analysis reveals a possible cross-talk between PTMs and alteration of Tg function in these disorders. In perspective, multi-omics analyses from patients, interpreted with structural and functional data, may generate more robust models to correlate phenotypes with classes of Tg functional alterations. This integrative approach will likely provide more targeted strategies to restore specific Tg functions in different thyroid pathologies.

## Introduction: thyroglobulin in thyroid hormone homeostasis

The two thyroid hormones (TH) 3,5,3',5'-tetraiodo-L-thyronine (T4) and 3,3′,5-triiodo-L-thyronine (T3) are crucial for human metabolism and development (1, 2). T4 is produced by the thyroid in a much larger amount than T3 (the active hormone) and can be converted to it by de-iodination (3). T3 activates the transcription of a set of pivotal genes in target tissues (e.g. growth factors) upon binding to the nuclear thyroid hormone receptors (THR) (4, 5). TH are synthesised within the thyroid gland follicles, monolayers of polarised thyrocytes enclosing a large extracellular lumen (Fig. 1A and B) (1). The synthesis of TH is stimulated by thyrotropin (or thyroid-stimulating hormone, TSH), a glycoprotein hormone secreted by the pituitary gland. Upon binding to its cognate receptor TSHR on the basolateral membrane of thyroid cells, TSH induces the expression of several thyroid hormone pathway proteins, including thyroglobulin (Tg), the TH glycoprotein
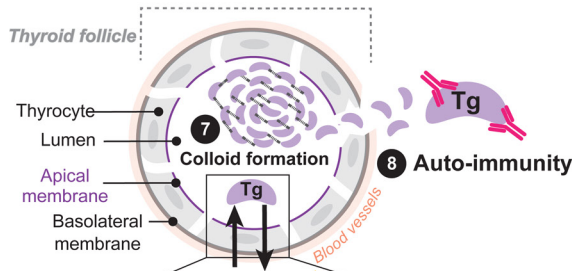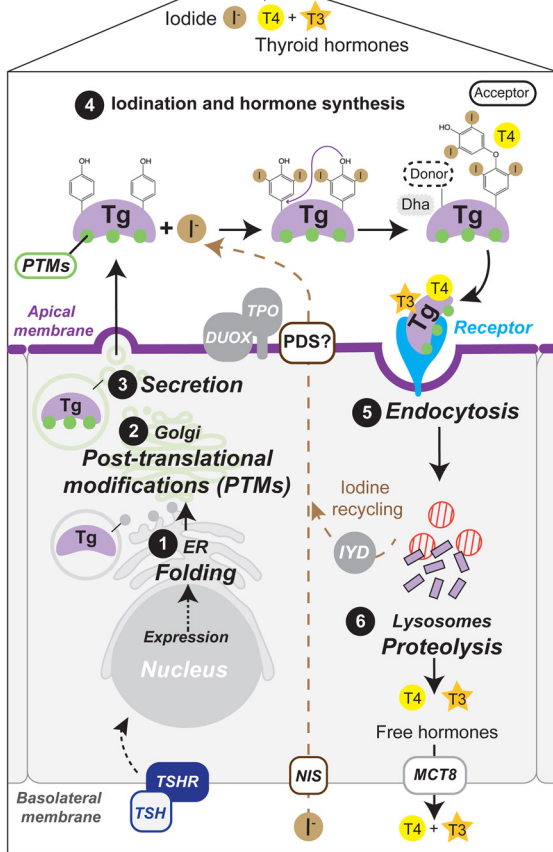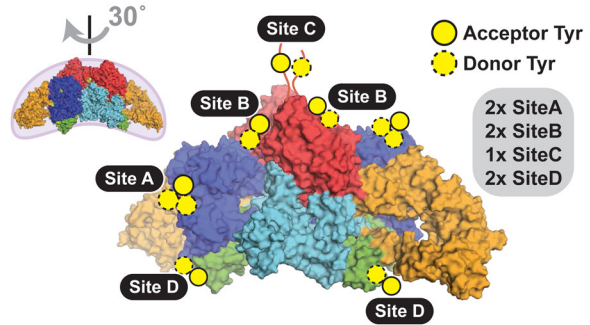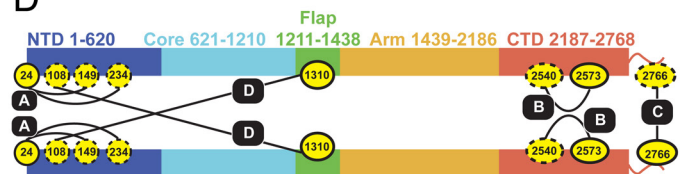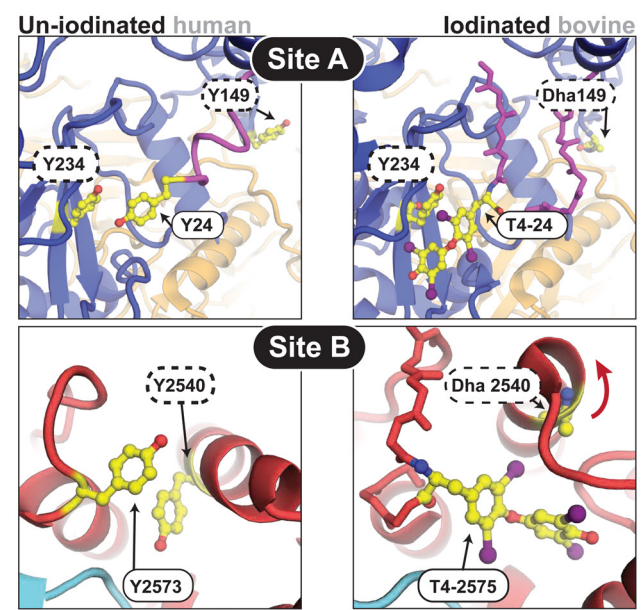
Figure 1

Overview of thyroglobulin structure and functions in the thyroid. (A and B) Schematic representation of thyroid follicles, composed of a monolayer of polarised thyrocytes enclosing the lumen, where the thyroglobulin (Tg) is abundantly secreted. Upon stimulation of thyroid-stimulating hormone (TSH), binding its cognate receptor TSHR at the basolateral membrane, iodide is uptaken through the sodium/iodide symporter (NIS) and the expression of several thyroid-specific genes is activated, including Tg. Iodide is likely transported from the cytosol to the lumen by pendrin (PDS). Tg is folded in the ER (1), then transported in the Golgi (2) where it undergoes a large series of post-translational modifications (PTMs) to be then secreted in the extracellular lumen of the follicles (3). Here, Tg tyrosines are enzymatically iodinated by dual oxidases (DUOX) and thyroperoxidase (TPO) and seven sites form thyroid hormones (TH) (4). The synthesis occurs by transfer of the iodinated ring of the 'donor' tyrosine to an 'acceptor' tyrosine. The reaction results in either T4 or T3 hormones covalently bound to the Tg polypeptide chain on the acceptor position and a dehydroalanine (Dha) on the donor position. Hormone release to the bloodstream is mediated by endocytosis (5) and subsequent proteolysis from Tg in lysosomes (6). Free hormones are then released via membrane transporters (i.e. monocarboxylate transporter 8, MCT8) to the blood vessels, while iodine bound to other tyrosine residues is recycled via de-iodination by the iodotyrosine deiodinase enzyme (IYD). A large portion of Tg stores iodine and hormones in an extracellular cross-linked matrix called the colloid (7). Tg is also a major antigen in thyroid autoimmune diseases, often due to follicle leaking processes of various origin (8). (C) Overall structure of Tg and location of the seven tyrosine pairs synthesising TH (hormonogenic sites A to D). Tg macro-domains are indicated as follows: N-terminal domain NTD (blue), core (light blue), Flap (green), Arm (orange), C-terminal domain CTD (red) (9). (D) Two-dimensional illustration of the Tg dimer sequence indicating hormonogenic tyrosine positions with a yellow circle (donors: dotted contour, acceptor: plain contour) and connections indicate the pairing in the three-dimensional structure. (E) Close-up of hormonogenic site A and site B in un-iodinated human Tg (9) and mildly iodinated bovine Tg (30), clearly showing T4 and Dha at acceptor and donor positions. The synthesis of the hormones implies local structural rearrangement, such as displacement of a CTD alpha-helix near site B (red arrow).

precursor (6, 7). Tg is a large and complex dimer first folded in the endoplasmic reticulum (ER) then secondly, matured with post-translational modifications (PTMs) in the Golgi apparatus and thirdly, abundantly secreted via the apical thyrocyte membrane into the follicular lumen (7, 8). At the same time, sodium/iodide symporter (NIS), pendrin and other membrane solute carrier proteins allow iodide (I⁻) accumulation from the blood vessels through the follicular cells into the lumen. Here, about 60 tyrosine residues within the Tg dimer are enzymatically iodinated via the combined action of dual oxidases (DUOX) and thyroperoxidase (TPO) (1). As revealed by the recently determined atomic structure (9), most Tg iodotyrosines are far from each other and likely act as an iodine reserve, while seven proximal iodotyrosine pairs give rise to as many THs. Hormonogenesis at these sites occurs via an iodotyrosine side-chain coupling reaction that generates either T4 or T3 on the tyrosine ring 'acceptor' position, while leaving a dehydroalanine (Dha) at a tyrosine ring 'donor' position (Fig. 1B). Receptor-mediated endocytosis allows hormone-embedding Tg to be transferred into the cell and directed to lysosomes (6, 7). Here, mainly by cathepsin-mediated proteolysis, TH are liberated from the Tg scaffold to be then released into the bloodstream via the monocarboxylate transporter 8 (MCT8) (10).

However, a vast majority of Tg remains in the lumen, forming a dense cross-linked matrix, the colloid, which serves as an iodine and hormone reservoir for the organism (11). The colloid is composed of Tg dimers covalently linked by inter-molecular disulfide bonds and N-ε-γ-glutamyl-lysine bonds, the latter established between lysine and glutamine Tg residues (11, 12). It has been reported that cathepsins are not only found intracellularly but are also secreted into the follicular lumen, thereby both releasing TH from colloidal Tg and liberating entire Tg molecules from the matrix, prior to endocytosis (10). The balance between TH storage and release mediated by Tg is central for hormone yields and for the stability of the thyroid tissue (13). Indeed, several thyroid pathologies, some deriving from Tg gene mutations, are associated with anomalous follicular size (e.g. goitre). In addition, Tg leaking and lymphocytic infiltrations are related to the development of anti-Tg antibodies and consequent thyroid tissue damage (Fig. 1A) (14).

Despite extensive research in the field, both physiological regulation and pathological modifications in Tg structure in the hormonogenesis pathway are poorly understood at a molecular level. Hence, it remains arduous to act specifically on the restoration of one of more hormonogenesis steps. Today, the knowledge of Tg

structure offers an excellent tool to investigate and interpret these processes in detail, for example, via the classification of pathogenic aminoacidic mutations (15, 16).

Other fundamental players in Tg folding, regulation and pathogenesis are the many PTMs decorating the protein moiety (17). Recent proteomic and biochemical studies have revealed that beyond disulfide bonds and proteolysis, glycosylation, acetylation, sulfation and phosphorylation may also be important for Tg function (18, 19, 20). An important example is the *de novo* synthesis of T3 from Tg, which was proposed to be regulated by phosphorylation events (18). In addition, PTMs have also been suggested to be important in thyroid pathogenesis, for example, in cancer (17, 21).

In the present review, we compare the recently reported electron cryo-microscopy (cryo-EM) structures of Tg and use the consensus model as a framework to depict the three-dimensional landscape of PTMs. Our analysis suggests how different PTMs may contribute to regulating Tg folding and surface properties and how they may relate to pathogenic mutations and immunogenicity. Therefore, structural interpretation will be a valuable tool in clinical and basic thyroidology to rationally correlate thyroid phenotypes with Tg alterations and to device specific Tg-targeting therapeutics for different thyroid disorders.

Up to date with thyroglobulin structure domains, dimerisation and hormonogenic sites

The thyroid hormone precursor Tg is a giant dimeric glycoprotein (dimer ~600 kDa) orchestrating the synthesis, storage and release of TH within the thyroid gland (6). Moreover, Tg concentration has been proposed as a potential regulator of the follicle refilling–depletion cycle (13) and to have some effects on thyroid cells growth (22). However, more experimental data would be required to confirm these hypotheses. In pathogenesis, Tg is the main antigen in thyroid-related autoimmune diseases (in case of follicle leakage and tissue damage phenomena) and is a serum marker of thyroid cancer, exclusively after total thyroidectomy (23).

These many functions are likely depending on Tg three-dimensional structure, which determines hormonogenesis efficiency through iodotyrosine pairing and interaction with cellular partners (e.g. chaperones, receptors and proteases) (7). After many years of research, the structure of human Tg has been recently resolved by cryo-EM (9, 24). The primary structure of Tg is composed

of 17 repeated cysteine-rich repeats alternated with linker regions and followed by a dimeric acetylcholine esterase-like domain (AChE or C-terminal domain, CTD) (7, 25, 26). Tg monomers fold in an elongated sea-urchin-shaped assembly which creates a large surface contact with the other chain, resulting in an extremely stable dimer (9) (Fig. 1C and D). In Fig. 1C, we depict the Tg macrodomains composing the structure according to a rainbow colour scheme. The AChE interacts with many protein regions far in sequence (core, N-terminal domain (NTD), Flap and partially the Arm). This confirms the previously reported hypothesis of AChE being an internal Tg chaperone and explains why point mutations in the AChE domains severely impair Tg folding, causing ER stress and impaired secretion (27).

In the study by Coscia *et al.*, the cryo-EM analysis was performed on recombinant fully glycosylated Tg and tissue-extracted Tg, which was mildly de-glycosylated in native conditions (9). Both structures are equivalent and in both, it is possible to resolve N-glycosylation moieties directly bound to asparagine residues. Hormonogenic tyrosine pairs are located in four site types, which were previously identified by mass spectrometry: sites A, B, C and D (28). While sites A, B and D are replicated in the dimer, site C is formed by two tyrosine residues at the same location in the sequence (Tyr2766), coupling from opposite chains (Fig. 1C and D). Using structure-based biochemical analysis, Coscia and co-workers could uniquely identify all donors and stoichiometry of seven hormone copies synthesised by a Tg dimer. In the cryo-EM model, all sites can be resolved except site C, which is found in an unstructured region on the dimer, close to the C-terminus. The acceptor–donor pairing is summarised in Fig. 1C and D within both the three-dimensional model and onto the linear sequence. While in sites A and B the tyrosine pairs belong to the same subunit, site C and site D tyrosine pairs belong to different chains, at least partially explaining the importance of the dimeric assembly for the complete physiological hormone yields (9, 29).

After the first Tg structure was published, two more Tg structures were resolved by cryo-EM: the fully glycosylated human Tg by Adaixo *et al.* (20) and the bovine poorly iodinated Tg (0.6%) by Kim *et al.* (30). Both structures are very similar to and confirm the findings by Coscia *et al*. Importantly, the bovine Tg cryo-EM from Kim *et al*. resolved for the first time the T4 hormone moieties (on acceptor positions) as well as the Dha (on donor positions) stably embedded in the polypeptide chain (Fig. 1E) (30). Intriguingly, while the overall structure of iodinated bovine Tg is basically identical to the non-iodinated human

homologue, substantial variations in the conformation of the polypeptide chain can be observed in the proximity of the T4-acceptor position (Fig. 1E). While acceptor tyrosine residues are in flexible and exposed regions, the synthesised T4 inserts in a shallow pocket of the Tg surface (30). This may indicate a potential role of conserved surface residues surrounding the hormonogenic tyrosine pairs, which however remain poorly understood.

It is well known that hormonogenic sites are early iodination sites (31), and this is probably why these are clearly visible in a poorly iodinated Tg. Yet, no iodotyrosine densities could be identified in the cryo-EM map (30). This may indicate heterogeneity in iodination of the isolated non-hormonogenic tyrosines, which may average out in the cryo-EM map. At present, it remains unclear what is the sequence of iodination events in Tg and what is the impact of full iodination on the assembly.

Comparing the hormonogenic tyrosine pairs within the atomic model, Coscia *et al*. identified that proximity and flexibility of iodotyrosine pairs are sufficient conditions for the synthesis of T4 to occur from polypeptide chains. Indeed, T4 synthesis can be efficiently recapitulated in other polypeptide scaffolds and polymers containing flexible tyrosine pairs (9, 32). These features are possibly present also in ancestral pre-vertebrate Tg, which could in principle diverge from the full-length Tg sequence (33). It is obscure why the giant structure of Tg has been retained in evolution and why it contains exactly four hormonogenic sites in higher vertebrates. Today, despite the Tg assembly being finally uncovered, the role of the conserved domain architecture in the hormonogenesis cycle is yet to be revealed. However, the integration of biochemical and cellular studies, guided by the knowledge of the Tg dimeric assembly, promises to soon unveil these important mysteries in Tg function (19, 24, 34).

The post-translational modification landscape of human thyroglobulin

PTMs are a key element in Tg maturation, mostly happening in the Golgi apparatus (17). By literally decorating the protein surface, PTMs can drastically change local charge, solubility, stability, oligomerisation and affinity for binding partners, thereby affecting protein function and signalling within the cell. Moreover, pathogenic microenvironments (e.g. tumour cells) are often characterised by changes in PTMs in many proteins (e.g. glycosylation and phosphorylation). Hence, dynamic footprints of PTMs in a set of proteins can potentially be
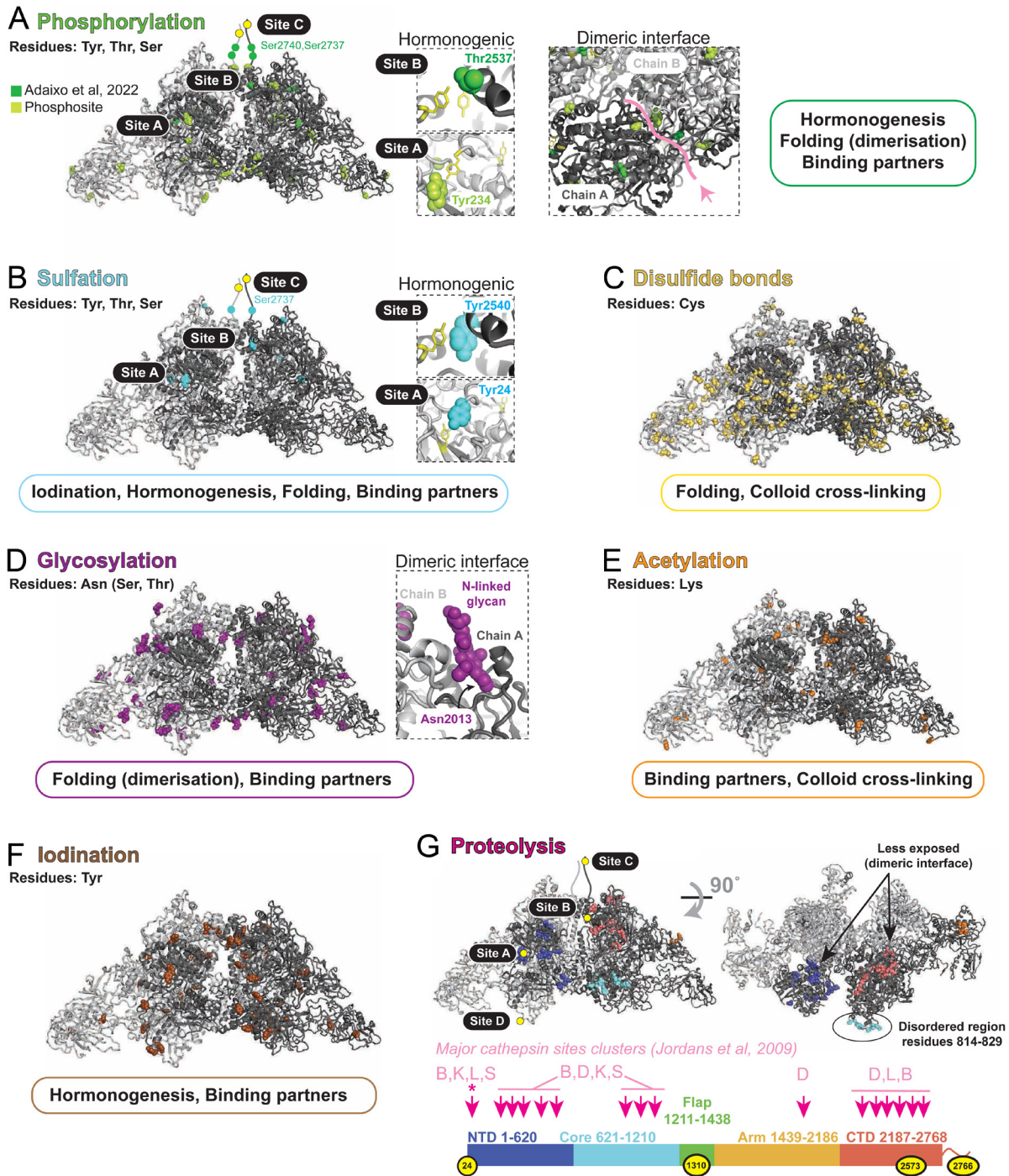
Figure 2

Overview of known thyroglobulin post-translational modifications and potential implication on its function. (A, B, C, D, E and F) Post-translational modifications (PTMs) occurring on definite thyroglobulin (Tg) residue types and locations, mapped as coloured spheres on the Tg dimeric structure (grey ribbon). The represented Tg model indicates the atoms resolved in the cryo-EM map (only glycosylation and hormones PTMs are directly visible in the structure). Dotted line insets: exemplary close-ups on region of the structure, suggesting implication of the given PTMs in the regulation of Tg functions (outlined in Fig. 1A and B), summarised in the rounded rectangles for each PTM type. (G) Cathepsins (B, D, K, L and S) proteolytic sites mapped on Tg structure and their position with respect to acceptor tyrosine residues (hormonogenic site A to B) where thyroid hormones are embedded (spheres coloured by domain); some of these sites are derived from digestion of rabbit Tg (10, 53). Proteolysis close to the exposed hormonogenic sites A at Tg N-terminus may contribute to extracellular hormone release even within the colloid (indicated with *). Other exposed sites may release Tg from the colloid or, together with buried cleavage sites (some at dimeric interface), may be involved in later hormone release from Tg and in its complete dissociation in the lysosomes.

involved in increased pathogenicity and be markers of specific disease progression (35).

Multiple PTMs within Tg have been identified by high-throughput mass spectrometry screening methods and reported in online databases (36). In addition, recent *ad hoc* proteomic studies on Tg have revealed unanticipated PTMs on the protein scaffold, such as acetylation (20). We collected the majority of the known PTMs detected on Tg and reported them in Supplementary Table 1 (see section on supplementary materials given at the end of this article). The Tg structure and position of hormonogenic sites offer a 'molecular cartography' to map in three-dimensions known PTMs and infer potential implications in the regulation of Tg function. In Fig. 2, we compose an atlas of PTMs on the Tg dimeric ribbon diagram, focusing on some interesting PTMs and describing the potential impact of single modifications on colloid cross-linking, dimerisation, hormonogenesis, folding and interaction with cellular partners. It is important to highlight that while glycosylation and disulfide bonds are constitutive PTMs in Tg, iodination, phosphorylation, sulfation and acetylation can be variable in relative stoichiometry and abundance, potentially leading to several forms of Tg. Moreover, it is possible that different PTMs may trigger or exclude other PTMs in specific areas of Tg and influence the relative abundance of Tg populations in different thyroid states. A big challenge remains to experimentally determine what stimuli induce these PTMs, to eventually understand and control the underlying molecular mechanisms.

Phosphorylation

Phosphorylation is a negatively charged PTM added to serine, threonine and tyrosine residues (at physiological pH). It is a reversible PTM introduced by several kinase enzymes and can be hydrolysed by the action of phosphatases. Therefore, it can mark a specific cellular or protein state in time and space. Mapping the Tg phosphorylation sites reported by Adaixo *et al.* (20) and on the Server Phosphosite (36), we noticed that several ones correspond to or are located nearby hormonogenic tyrosine pairs, for example, in site B Thr2537 and in site A Tyr234, the latter corresponding to a donor site (Figs 1A and 2A) (9). Citterio *et al.* highlighted that *de novo* T3 synthesis within the AChE domain is enhanced by phosphorylation, likely promoted by the casein kinase FAM20C (18). This kinase is localised in the Golgi apparatus and ubiquitously expressed in tissues (37). It is still unclear what is the chemical influence of phosphate groups on the increase of T3 to T4 synthesis from hormonogenic sites. Moreover, several phosphorylation sites are found close

to the Tg dimeric interface and might strongly affect the stability of the dimer. However, this hypothesis should be verified experimentally. Taken together, phosphorylation could be a key Tg function regulator either before or after secretion and potentially also after cell internalisation, influencing trafficking and degradation. Furthermore, phosphorylation signalling is often altered in cancer (35), and Tg could be one of the proteins affected in the process (17, 35). Additional biochemical validation and structure-guided biochemical investigations in this direction promise to better clarify the impact of this important PTM on thyroid function and disease.

Sulfation

Sulfation is a negatively charged PTM at physiological pH; it can be added to the same residue types as phosphorylation but is mostly found on tyrosines (38). It is catalysed by sulfotransferase enzymes mainly in the Golgi apparatus and it is considered an irreversible PTM *in vivo* up to now. Sulfation plays regulatory role in many biochemical processes. It is often found near clusters of acidic residues on the protein surface, and in some cases, it co-localises with O-glycosylation (39). In Tg, sulfation sites are mainly located in the AChE domain and importantly close to site C, on the site B donor Tyr2540 and on the most reactive hormonogenic site A acceptor 24 (Fig. 2B) (20, 40, 41, 42). As for phosphorylation, it remains unclear what is the impact of sulfation on hormonogenesis; therefore, it would be of great interest to collect more experimental data on this aspect and its regulation.

Disulfide bonds

Disulfide bonds derive from the oxidative coupling of two cysteine residues close in space. The Tg dimer presents an impressive amount of disulfide bonds (about 120) which are only found within the same subunit (Fig. 2C) (9). Being a well-known player in Tg folding and stability, mutations of oxidised cysteine residues severely hamper Tg folding and are associated with ER stress and congenital hypothyroidism (CH) (15, 19). In addition, disulfide bonds are a major inter-molecular Tg cross-link stabilizing the colloid; these reactions are likely catalysed by the internal thioredoxin region in the 'Arm' macrodomain (43, 44). Possibly, modulation of oxidation states of cysteine residues on the surface play an important role in the formation and disassembly of the matrix, but it is still unclear what are the specific inter-molecular cross-links enabling colloid packing in the follicle lumen (11).

Glycosylation

Glycosylation consists of the addition of glycan moieties onto either asparagine (N-glycosylation) and more rarely on serine and threonine residues (O-glycosylation). It is reported that Ser-2749 bears O-glycosylation which protects the C-terminus from proteolysis and enhances hormone formation (45); yet more experimental evidence would be required to confirm the location and role of this modification. On the other hand, it is well known that Tg is abundantly N-glycosylated up to a tenth of its weight, with approximately 16 identified sites including N-acetylglucosamine (GlcNAc), followed by a variable number of mannose glycosylation units (Fig. 2D) (9). Apart from conferring high solubility to the abundantly secreted Tg, glycosylation importantly contributes to the correct folding of the dimer. In particular, as showed by the cryo-EM map reported by Coscia *et al.* (9) and also confirmed by proteomics and site-directed mutagenesis studies (19), Asn2013 bears an N-glycosylation moiety which is buried in a pocket between the two Tg chains, and it also resists de-glycosylation in native conditions. Indeed, mutations nearby this site are shown to be detrimental to Tg folding, degradation and secretion and are related to congenital hypothyroidism (CH) (19). In addition, it has been reported that Tg also carries mannose-6 phosphate groups, well-known targeting signals for the transport to lysosomes. However, the role of these modifications in Tg endocytosis and trafficking is yet to be confirmed (46, 47).

Acetylation

Acetylation occurs on lysine residues and suppresses its positive charge. Mostly known as an important gene expression regulator, for example, on histone proteins (48, 49), unexpectedly 15 acetylation sites have been recently identified on several exposed residues of Tg (Fig. 2E) (20). The degree and function of acetylation in Tg are yet to be elucidated, but we suggest that it may affect both surface charge properties and reduce the ability to form colloid N-ε-γ-glutamyl-lysine bonds which stabilise the colloid matrix (12).

Iodination

Iodination occurs in the extracellular thyroid follicle lumen on basically all exposed tyrosine residues in Tg (Fig. 2F), and it is the defining modification on the TH precursor (28). Apart from determining the synthesis of TH in specific flexible iodotyrosine pairs, the ability of Tg to incorporate many iodine atoms in a single molecule is essential for the storage of this essential trace element in the body. It has been proposed that iodination affects hydrophobicity and conformation of Tg, possibly triggering endocytosis (50, 51), as well as influencing auto-antibody epitopes (52), yet these aspects are yet to be completely understood at a structural level.

Proteolysis

Proteolysis by cathepsins is essential to enable extracellular Tg molecules release from the colloid and TH release both from the colloid and in the lysosomes (Fig. 1A and B) (1, 53). In Fig. 2G, we map the identified cathepsin sites (cathepsin B, D, K, L and S) on Tg three-dimensional structure and represent them using the same macrodomains colour scheme reported in Fig. 1D (10, 53).

Our analysis reveals that the release of TH by cathepsin cleavage within the colloid, without major Tg structural alterations, could only occur at the exposed cleavage sequence close to site A (Fig. 2G, lower panel). Indeed, the last cleavage site at residues 2677–2680 would allow TH release from site C, however also withdrawing a central part of the dimerisation interface at the AChE domain. Other cleavage sites within the AChE are partially buried, and we propose that the associated proteolysis events might happen only later in the lysosomes rather than extracellularly. While digestion of AChE and core most certainly causes severe disruption of the Tg dimer, proteolysis at NTD, Arm and other sites may probably just 'nick' Tg in presence of the many disulfide bonds. Interestingly, two very close cleavage sites are in the core macrodomain and are located near the disordered region 814–829 containing a glutamine residue (represented in cyan in Fig. 2G). Being quite exposed and not contributing to the dimeric interface, this region may be a putative site for scission of full-length Tg molecules from the colloid. These hypotheses should be certainly verified by experimental evidence, but our analysis provides a starting point to rationally dissect the importance of each cleavage site within Tg processing and TH homeostasis.

Potential cross-talk of thyroglobulin post-translational modifications and pathogenesis

Being a pivotal protein in thyroid homeostasis, Tg is involved directly or indirectly in a variety of thyroid diseases, such as autoimmune thyroid disease (AITD) and CH.

Anti-Tg antibodies are present in several thyroid-related autoimmune diseases (e.g. in Hashimoto

A Overlap between immunogenic regions and PTMs
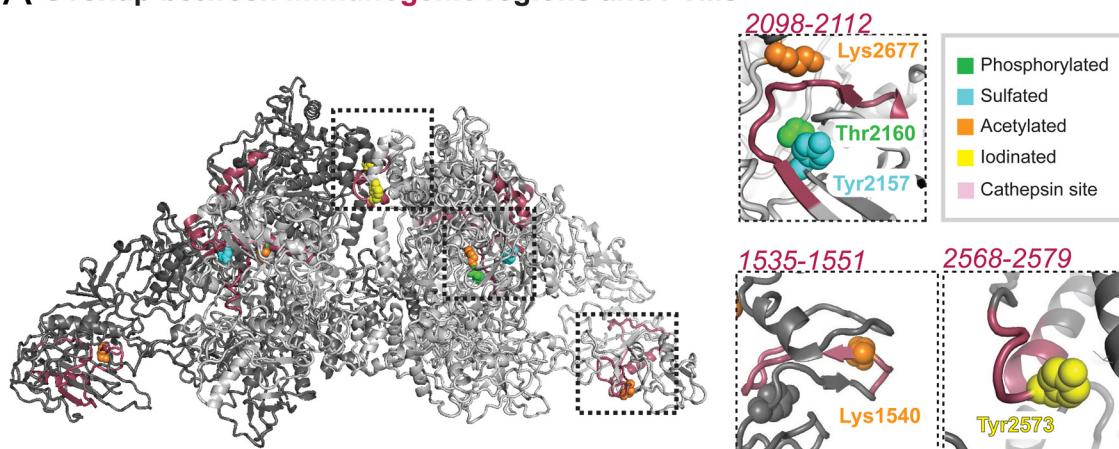
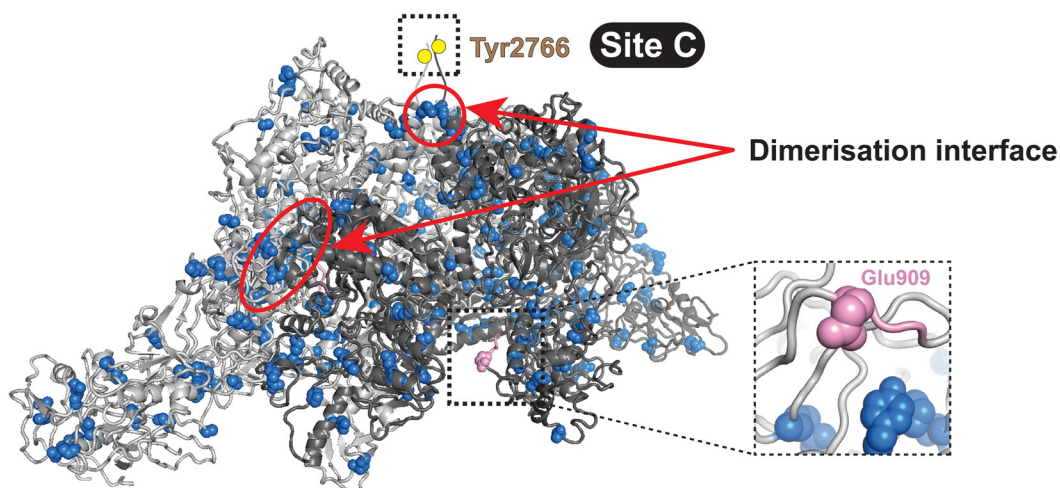## B Overlap between missense mutations and PTMs

Figure 3
Potential cross-talk of thyroglobulin post-translational modifications (PTMs) and pathogenesis. (A) Overlap between immunogenic regions (epitopes identified in experimental autoimmune thyroiditis, EAT) and PTMs, suggesting a potential influence on auto-antibody recognition and regulation. (B) Overlap between Tg missense mutations and PTMs, suggesting an impairment of PTMs-related Tg functions when such mutations occur in congenital hypothyroidism. Several missense mutations are found at dimeric interfaces within the three-dimensional structure of the Tg (red arrows).

thyroiditis) and often in thyroid cancer patients (54, 55). Some of the Tg regions of immunogenicity were identified in experimental autoimmune thyroiditis (EAT) (56, 57), and we show their location in space on the Tg structure in Fig. 3A. The mechanisms and inheritance of AITD, together with evolution of human Tg epitopes and antibody paratopes, are still an active area of research (58, 59). Interestingly, it has been shown that Tg iodination could affect the recognition by auto-antibodies, and this may be related to a major change of epitope protein surface (52). We systematically analysed the overlap and proximity of all the PTMs reported for Tg (see the 'The post-translational modification landscape of human thyroglobulin' section)

and proposed that iodinated hormonogenic site B could be involved in this mechanism, overlapping with immunogenic region 2568–2579. In addition, we revealed that also phosphorylation, sulfation and acetylation may affect the surface in close proximity to other epitopes, potentially reducing or enhancing the interaction with existing auto-antibodies and participating in the evolution of the epitope–paratope pairs (Fig. 3A).

Since Tg protein function revolves around molecular surface properties, it is expected that nonsense and missense mutations in Tg gene affect folding and hormonogenesis and often lead to CH with clear phenotypes (15). The advancement of next-generation

sequencing (NGS) and pre-natal or new-born screening allowed to identify more than 200 Tg genetic variants associated with thyroid disfunction (15). It can be anticipated that cysteine mutations or drastic aminoacidic changes in the AChE or core domains directly impact secretion, ER stress, dimerisation and hormonogenesis. A few CH mutations have been systematically studied by cellular and proteomic approaches and were shown to impact not only folding but also degradation pathways and glycosylation (19). However, other mutations result in almost conservative aminoacidic changes and are located in peripheral regions not necessarily involved in dimer stability. Hence, at present, we are still far from a complete understanding of the impact of all Tg genetic mutations on protein function and on associated thyroid phenotypes.

In Fig. 3B, we represent aminoacidic changes deriving from new missense mutations recently identified via NGS (15) and we analyse the overlap with the known PTMs (see the 'The post-translational modification landscape of human thyroglobulin' section). What the mapping on the three-dimensional structure revealed, compared to previous representation on the linear sequence, is a strong overlap with many parts of the dimeric interface. Moreover, we report an overlay with the unstructured iodination and hormonogenesis Tyr2766 and Glu909, belonging to a cathepsin proteolytic site. Thence, it is possible that aminoacidic mutations are also affecting already folded and secreted Tg solubility, processing by cathepsins and potentially other functions ruled by PTMs. Moreover, as PTMs may occur or change over time, we hypothesise that the effects of modified Tg on these positions could also appear noticeable in a later stage or transiently within a disease condition.

Overall, the present structural analysis suggests a potential and interesting cross-talk between autoimmunity and mutations with functions regulated by PTMs. We envisage that a wider genomic and proteomic high-throughput screening of phenotypic patients, guided by the structure of Tg, will enable more in-depth determination of the molecular implications of PTMs in the near future.

Conclusions and perspectives

Tg is a large and pivotal protein responsible for the synthesis of TH within the thyroid gland. Variations in its abundance, gene sequence and localisation are often associated with thyroid diseases and used in diagnostics. The many types of PTMs are an integrative part of Tg structure, and their regulatory roles in function and pathogenesis have been highlighted since the 1980s, but in absence of a molecular structure, it has been impossible to rationally predict their impact (17). Here, we revisit the TH biosynthetic pathway orchestrated by Tg and we map the known PTMs on the recently determined Tg structure (9). Based on their spatial position we hypothesise that these modifications could affect many molecular mechanisms beyond Tg folding: colloid cross-linking, hormonogenesis, endocytosis and auto-antibody epitopes (7). Importantly, some mutations cause CH overlap with PTMs and may impair additional functions ruled by these chemical modifications that could be targeted and restored in thyroid treatment.

In particular for what concerns CH, modern DNA sequencing methods now enable the collection of large phenotype–mutation databases (60, 61), yet deriving and targeting their specific impact on thyroid homeostasis remain challenging. We suggest that it would be effective to integrate Tg proteomic profiles, structural localisation and functional cellular studies with genomic data, in order to produce a more rational categorisation of phenotypic mutations linked to altered Tg functions. For example, we hypothesise that mutations clustering on regions mediating colloid cross-linking or interaction with trafficking factors may reduce TH release and concentration in the bloodstream, possibly associating with enlarged follicular size phenotypes.

In perspective, these broader correlative models may lead to more specific treatment in thyroid diseases, targeting identified classes of Tg mutations or altered PTM profiles. Moreover, they may provide predictive tools on the effects of newly occurring mutations, based on their location on Tg structure. Additionally, these comprehensive databases may reveal disease-dependent Tg PTM patterns that could become valuable diagnostic tools. As for Tg, this methodology could be extended to other proteins involved in the development of CH and of other thyroid pathologies.

References

1 Carvalho DP & Dupuy C. Thyroid hormone biosynthesis and release. *Molecular and Cellular Endocrinology* 2017 **458** 6–15. (https://doi.org/10.1016/j.mce.2017.01.038)

2 Mullur R, Liu YY & Brent GA. Thyroid hormone regulation of metabolism. *Physiological Reviews* 2014 **94** 355–382. (https://doi.org/10.1152/physrev.00030.2013)

3 Luongo C, Dentice M & Salvatore D. Deiodinases and their intricate role in thyroid hormone homeostasis. *Nature Reviews: Endocrinology* 2019 **15** 479–488. (https://doi.org/10.1038/s41574-019-0218-2)

4 Bernal J. Thyroid hormone receptors in brain development and function. *Nature Clinical Practice: Endocrinology and Metabolism* 2007 **3** 249–259. (https://doi.org/10.1038/ncpendmet0424)

5 Iwen KA, Oelkrug R & Brabant G. Effects of thyroid hormones on thermogenesis and energy partitioning. *Journal of Molecular Endocrinology* 2018 **60** R157–R170. (https://doi.org/10.1530/JME-17-0319)

6 Citterio CE, Targovnik HM & Arvan P. The role of thyroglobulin in thyroid hormonogenesis. *Nature Reviews: Endocrinology* 2019 **15** 323–338. (https://doi.org/10.1038/s41574-019-0184-8)

7 Di Jeso B & Arvan P. Thyroglobulin from molecular and cellular biology to clinical endocrinology. *Endocrine Reviews* 2016 **37** 2–36. (https://doi.org/10.1210/er.2015-1090)

8 Kim PS & Arvan P. Folding and assembly of newly synthesized thyroglobulin occurs in a pre-Golgi compartment. *Journal of Biological Chemistry* 1991 **266** 12412–12418. (https://doi.org/10.1016/S0021-9258(18)98913-1)

9 Coscia F, Taler-Verčič A, Chang VT, Sinn L, O'Reilly FJ, Izoré T, Renko M, Berger I, Rappsilber J, Turk D, *et al.* The structure of human thyroglobulin. *Nature* 2020 **578** 627–630. (https://doi.org/10.1038/s41586-020-1995-4)

10 Jordans S, Jenko-Kokalj S, Kühl NM, Tedelind S, Sendt W, Brömme D, Turk D & Brix K. Monitoring compartment-specific substrate cleavage by cathepsins B, K, L, and S at physiological pH and redox conditions. *BMC Biochemistry* 2009 **10** 23. (https://doi.org/10.1186/1471-2091-10-23)

11 Berndorfer U, Wilms H & Herzog V. Multimerization of thyroglobulin (TG) during extracellular storage: isolation of highly cross-linked TG from human thyroids. *Journal of Clinical Endocrinology and Metabolism* 1996 **81** 1918–1926. (https://doi.org/10.1210/jcem.81.5.8626858)

12 Saber-Lichtenberg Y, Brix K, Schmitz A, Heuser JE, Wilson JH, Lorand L & Herzog V. Covalent cross-linking of secreted bovine thyroglobulin by transglutaminase. *FASEB Journal* 2000 **14** 1005–1014. (https://doi.org/10.1096/fasebj.14.7.1005)

13 Sellitti DF & Suzuki K. Intrinsic regulation of thyroid function by thyroglobulin. *Thyroid* 2014 **24** 625–638. (https://doi.org/10.1089/thy.2013.0344)

14 Morshed SA, Latif R & Davies TF. Delineating the autoimmune mechanisms in Graves' disease. *Immunologic Research* 2012 **54** 191–203. (https://doi.org/10.1007/s12026-012-8312-8)

15 Citterio CE, Rivolta CM & Targovnik HM. Structure and genetic variants of thyroglobulin: pathophysiological implications. *Molecular and Cellular Endocrinology* 2021 **528** 111227. (https://doi.org/10.1016/j.mce.2021.111227)

16 Stern E, Schoenmakers N, Nicholas AK, Kassif E, Hamiel OP & Yeshayahu Y. A novel mutation in the thyroglobulin gene resulting in neonatal goiter and congenital hypothyroidism in an Eritrean infant. *Journal of Clinical Research in Pediatric Endocrinology* 2021 **14** 221–226. (https://doi.org/10.4274/jcrpe.galenos.2021.2020.0278)

17 Xavier AC, Maciel RM, Vieira JG, Dias-da-Silva MR & Martins JR. Insights into the posttranslational structural heterogeneity of thyroglobulin and its role in the development, diagnosis, and management of benign and malignant thyroid diseases. *Archives of Endocrinology and Metabolism* 2016 **60** 66–75. (https://doi.org/10.1590/2359-3997000000103)

18 Citterio CE, Veluswamy B, Morgan SJ, Galton VA, Banga JP, Atkins S, Morishita Y, Neumann S, Latif R, Gershengorn MC, *et al.* De novo triiodothyronine formation from thyrocytes activated by thyroid-stimulating hormone. *Journal of Biological Chemistry* 2017 **292** 15434–15444. (https://doi.org/10.1074/jbc.M117.784447)

19 Wright MT, Kouba L & Plate L. Thyroglobulin interactome profiling defines altered proteostasis topology associated with thyroid dyshormonogenesis. *Molecular and Cellular Proteomics* 2021 **20** 100008. (https://doi.org/10.1074/mcp.RA120.002168)

20 Adaixo R, Steiner EM, Righetto RD, Schmidt A, Stahlberg H & Taylor NMI. Cryo-EM structure of native human thyroglobulin. *Nature Communications* 2022 **13** 61. (https://doi.org/10.1038/s41467-021-27693-8)

21 Yamamoto K, Tsuji T, Tarutani O & Osawa T. Phosphorylated high mannose-type and hybrid-type oligosaccharide chains of human thyroglobulin isolated from malignant thyroid tissue. *Biochimica et Biophysica Acta* 1985 **838** 84–91. (https://doi.org/10.1016/0304-4165(85)90253-3)

22 Suzuki K, Kawashima A, Yoshihara A, Akama T, Sue M, Yoshida A & Kimura HJ. Role of thyroglobulin on negative feedback autoregulation of thyroid follicular function and growth. *Journal of Endocrinology* 2011 **209** 169–174. (https://doi.org/10.1530/JOE-10-0486)

23 Spencer CA. Clinical review: clinical utility of thyroglobulin antibody (TgAb) measurements for patients with differentiated thyroid cancers (DTC). *Journal of Clinical Endocrinology and Metabolism* 2011 **96** 3615–3627. (https://doi.org/10.1210/jc.2011-1740)

24 Coscia F & Taler-Verčič A. Cryo-EM: a new dawn in thyroid biology. *Molecular and Cellular Endocrinology* 2021 **531** 111309. (https://doi.org/10.1016/j.mce.2021.111309)

25 Gunčar G, Pungerčič G, Klemenčič I, Turk V & Turk D. Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO Journal* 1999 **18** 793–803. (https://doi.org/10.1093/emboj/18.4.793)

26 Vassart G, Bacolla A, Brocas H, Christophe D, de Martynoff G, Leriche A, Mercken L, Parma J, Pohl V & Targovnik H. Structure, expression and regulation of the thyroglobulin gene. *Molecular and Cellular Endocrinology* 1985 **40** 89–97. (https://doi.org/10.1016/0303-7207(85)90162-5)

27 Lee J, Di Jeso B & Arvan P. The cholinesterase-like domain of thyroglobulin functions as an intramolecular chaperone. *Journal of Clinical Investigation* 2008 **118** 2950–2958. (https://doi.org/10.1172/JCI35164)

28 Dedieu A, Gaillard JC, Pourcher T, Darrouzet E & Armengaud J. Revisiting iodination sites in thyroglobulin with an organ-oriented shotgun strategy. *Journal of Biological Chemistry* 2011 **286** 259–269. (https://doi.org/10.1074/jbc.M110.159483)

29 Citterio CE, Morishita Y, Dakka N, Veluswamy B & Arvan P. Relationship between the dimerization of thyroglobulin and its ability to form triiodothyronine. *Journal of Biological Chemistry* 2018 **293** 4860–4869. (https://doi.org/10.1074/jbc.RA118.001786)

30 Kim K, Kopylov M, Bobe D, Kelley K, Eng ET, Arvan P & Clarke OB. The structure of natively iodinated bovine thyroglobulin. *Acta Crystallographica: Section D, Structural Biology* 2021 **77** 1451–1459. (https://doi.org/10.1107/S2059798321010056)

31 Lamas L, Anderson PC, Fox JW & Dunn JT. Consensus sequences for early iodination and hormonogenesis in human thyroglobulin. *Journal of Biological Chemistry* 1989 **264** 13541–13545. (https://doi.org/10.1016/S0021-9258(18)80031-X)

32 Cahnmann HJ, Pommier J & Nunez J. Spatial requirement for coupling of iodotyrosine residues to form thyroid hormones. *PNAS* 1977 **74** 5333–5335. (https://doi.org/10.1073/pnas.74.12.5333)

33 Di Fiore MM, Perrone L & D'Aniello A. Presence of a human-like thyroid stimulating hormone (TSH) in Ciona intestinalis. *Life Sciences* 1997 **61** 623–629. (https://doi.org/10.1016/s0024-3205(97)00437-2)

34 Antonica F, Kasprzyk DF, Opitz R, Iacovino M, Liao XH, Dumitrescu AM, Refetoff S, Peremans K, Manto M, Kyba M, *et al.* Generation of functional thyroid from embryonic stem cells. *Nature* 2012 **491** 66–71. (https://doi.org/10.1038/nature11525)

35 Singh V, Ram M, Kumar R, Prasad R, Roy BK & Singh KK. Phosphorylation: implications in cancer. *Protein Journal* 2017 **36** 1–6. (https://doi.org/10.1007/s10930-017-9696-z)

36 Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V & Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research* 2015 **43** D512–D520. (https://doi.org/10.1093/nar/gku1267)

37 Tagliabracci VS, Pinna LA & Dixon JE. Secreted protein kinases. *Trends in Biochemical Sciences* 2013 **38** 121–130. (https://doi.org/10.1016/j.tibs.2012.11.008)

38 Maxwell JWC & Payne RJ. Revealing the functional roles of tyrosine sulfation using synthetic sulfopeptides and sulfoproteins. *Current Opinion in Chemical Biology* 2020 **58** 72–85. (https://doi.org/10.1016/j.cbpa.2020.05.007)

39 Mehta AY, Heimburg-Molinaro J, Cummings RD & Goth CK. Emerging patterns of tyrosine sulfation and O-glycosylation cross-talk and co-localization. *Current Opinion in Structural Biology* 2020 **62** 102–111. (https://doi.org/10.1016/j.sbi.2019.12.002)

40 Venot N, Nlend MC, Cauvi D & Chabaud O. The hormonogenic tyrosine 5 of porcine thyroglobulin is sulfated. *Biochemical and Biophysical Research Communications* 2002 **298** 193–197. (https://doi.org/10.1016/s0006-291x(02)02425-7)

41 Nlend MC, Cauvi D, Venot N, Desruisseau S & Chabaud O. Thyrotropin regulates tyrosine sulfation of thyroglobulin. *European Journal of Endocrinology* 1999 **141** 61–69. (https://doi.org/10.1530/eje.0.1410061)

42 Nlend MC, Cauvi DM, Venot N & Chabaud O. Role of sulfated tyrosines of thyroglobulin in thyroid hormonosynthesis. *Endocrinology* 2005 **146** 4834–4843. (https://doi.org/10.1210/en.2005-0197)

43 Klein M, Gestmann I, Berndorfer U, Schmitz A & Herzog V. The thioredoxin boxes of thyroglobulin: possible implications for intermolecular disulfide bond formation in the follicle lumen. *Biological Chemistry* 2000 **381** 593–601. (https://doi.org/10.1515/BC.2000.076)

44 Atkinson HJ & Babbitt PC. An atlas of the thioredoxin fold class reveals the complexity of function-enabling adaptations. *PLoS Computational Biology* 2009 **5** e1000541. (https://doi.org/10.1371/journal.pcbi.1000541)

45 Conte M, Arcaro A, D'Angelo D, Gnata A, Mamone G, Ferranti P, Formisano S & Gentile F. A single chondroitin 6-sulfate oligosaccharide unit at Ser-2730 of human thyroglobulin enhances hormone formation and limits proteolytic accessibility at the carboxyl terminus. Potential insights into thyroid homeostasis and autoimmunity. *Journal of Biological Chemistry* 2006 **281** 22200–22211. (https://doi.org/10.1074/jbc.M513382200)

46 Herzog V, Neumüller W & Holzmann B. Thyroglobulin, the major and obligatory exportable protein of thyroid follicle cells, carries the lysosomal recognition marker mannose-6-phosphate. *EMBO Journal* 1987 **6** 555–560. (https://doi.org/10.1002/j.1460-2075.1987.tb04790.x)

47 Lemansky P & Herzog V. Endocytosis of thyroglobulin is not mediated by mannose-6-phosphate receptors in thyrocytes. Evidence for low-affinity-binding sites operating in the uptake of thyroglobulin. *European Journal of Biochemistry* 1992 **209** 111–119. (https://doi.org/10.1111/j.1432-1033.1992.tb17267.x)

48 Shvedunova M & Akhtar A. Modulation of cellular processes by histone and non-histone protein acetylation. *Nature Reviews: Molecular Cell Biology* 2022 **23** 329–349. (https://doi.org/10.1038/s41580-021-00441-y)

49 Verdone L, Caserta M & Di Mauro E. Role of histone acetylation in the control of gene expression. *Biochemistry and Cell Biology* 2005 **83** 344–353. (https://doi.org/10.1139/o05-041)

50 Botta R, Lisi S, Pinchera A, Giorgi F, Marcocci C, Taddei AR, Fausto AM, Bernardini N, Ippolito C, Mattii L, *et al.* Sortilin is a putative postendocytic receptor of thyroglobulin. *Endocrinology* 2009 **150** 509–518. (https://doi.org/10.1210/en.2008-0953)

51 Lisi S, Pinchera A, McCluskey RT, Willnow TE, Refetoff S, Marcocci C, Vitti P, Menconi F, Grasso L, Luchetti F, *et al.* Preferential megalin-mediated transcytosis of low-hormonogenic thyroglobulin: a control mechanism for thyroid hormone release. *PNAS* 2003 **100** 14858–14863. (https://doi.org/10.1073/pnas.2432267100)

52 Latrofa F, Fiore E, Rago T, Antonangeli L, Montanelli L, Ricci D, Provenzale MA, Scutari M, Frigeri M, Tonacchera M, *et al.* Iodine contributes to thyroid autoimmunity in humans by unmasking a cryptic epitope on thyroglobulin. *Journal of Clinical Endocrinology and Metabolism* 2013 **98** E1768–E1774. (https://doi.org/10.1210/jc.2013-2912)

53 Dunn AD, Crutchfield HE & Dunn JT. Thyroglobulin processing by thyroidal proteases. Major sites of cleavage by cathepsins B, D, and L. *Journal of Biological Chemistry* 1991 **266** 20198–20204. (https://doi.org/10.1016/S0021-9258(18)54909-7)

54 Fiore E, Latrofa F & Vitti P. Iodine, thyroid autoimmunity and cancer. *European Thyroid Journal* 2015 **4** 26–35. (https://doi.org/10.1159/000371741)

55 Latrofa F, Ricci D, Montanelli L, Rocchi R, Piaggi P, Sisti E, Grasso L, Basolo F, Ugolini C, Pinchera A, *et al.* Thyroglobulin autoantibodies in patients with papillary thyroid carcinoma: comparison of different assays and evaluation of causes of discrepancies. *Journal of Clinical Endocrinology and Metabolism* 2012 **97** 3974–3982. (https://doi.org/10.1210/jc.2012-2406)

56 Carayanniotis G. Molecular parameters linking thyroglobulin iodination with autoimmune thyroiditis. *Hormones* 2011 **10** 27–35. (https://doi.org/10.14310/horm.2002.1290)

57 Carayanniotis G. Recognition of thyroglobulin by T cells: the role of iodine. *Thyroid* 2007 **17** 963–973. (https://doi.org/10.1089/thy.2007.0199)

58 Latrofa F, Ricci D, Grasso L, Vitti P, Masserini L, Basolo F, Ugolini C, Mascia G, Lucacchini A & Pinchera A. Characterization of thyroglobulin epitopes in patients with autoimmune and non-autoimmune thyroid diseases using recombinant human monoclonal thyroglobulin autoantibodies. *Journal of Clinical Endocrinology and Metabolism* 2008 **93** 591–596. (https://doi.org/10.1210/jc.2007-1199)

59 Bogusławska J, Godlewska M, Gajda E & Piekiełko-Witkowska A. Cellular and molecular basis of thyroid autoimmunity. *European Thyroid Journal* 2022 **11** e210024. (https://doi.org/10.1530/ETJ-21-0024)

60 Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, *et al.* The UK biobank resource with deep phenotyping and genomic data. *Nature* 2018 **562** 203–209. (https://doi.org/10.1038/s41586-018-0579-z)

61 Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine* 2015 **12** e1001779. (https://doi.org/10.1371/journal.pmed.1001779)