Article

The structure of human thyroglobulin

https://doi.org/10.1038/s41586-020-1995-4

Received: 31 July 2019

Accepted: 16 January 2020

Published online: 5 February 2020

Francesca Coscia[1], Ajda Taler-Verčič[2,3], Veronica T. Chang[1], Ludwig Sinn[4], Francis J. O'Reilly[4], Thierry Izoré[1], Miha Renko[2], Imre Berger[5], Juri Rappsilber[4,6], Dušan Turk[2,3]* & Jan Löwe[1]*

Thyroglobulin (TG) is the protein precursor of thyroid hormones, which are essential for growth, development and the control of metabolism in vertebrates[1,2]. Hormone synthesis from TG occurs in the thyroid gland via the iodination and coupling of pairs of tyrosines, and is completed by TG proteolysis[3]. Tyrosine proximity within TG is thought to enable the coupling reaction but hormonogenic tyrosines have not been clearly identified, and the lack of a three-dimensional structure of TG has prevented mechanistic understanding[4]. Here we present the structure of full-length human thyroglobulin at a resolution of approximately 3.5 Å, determined by cryo-electron microscopy. We identified all of the hormonogenic tyrosine pairs in the structure, and verified them using site-directed mutagenesis and in vitro hormone-production assays using human TG expressed in HEK293T cells. Our analysis revealed that the proximity, flexibility and solvent exposure of the tyrosines are the key characteristics of hormonogenic sites. We transferred the reaction sites from TG to an engineered tyrosine donor–acceptor pair in the unrelated bacterial maltose-binding protein (MBP), which yielded hormone production with an efficiency comparable to that of TG. Our study provides a framework to further understand the production and regulation of thyroid hormones.

Tetra-iodothyronine (also known as thyroxine) (T_4) and tri-iodothyronine (T_3) are iodine-containing thyroid hormones that regulate metabolism and many other fundamental processes in vertebrates[1,5]. T_4, and smaller amounts of T_3, are found in the bloodstream of healthy humans[3], whereas suboptimal levels of thyroid hormones have marked consequences for heart rate, brain function and fetal development. Approximately 5% of the human population suffers from thyroid diseases, but the molecular events behind thyroid-hormone synthesis are yet to be completely understood[6].

Thyroid-hormone synthesis is stimulated by thyroid-stimulating hormone (TSH)—itself produced in pituitary gland—and occurs in the thyroid, from the protein precursor TG. Iodide (I^-) is accumulated in the thyroid both in the cytoplasm and in the lumen of thyroid follicular cells (also called colloid), into which TG is secreted in high amounts[2,7]. Two apical membrane enzymes, a dual oxidase (DUOX, which produces H_2O_2) and thyroid peroxidase (TPO, which oxidizes iodide), allow the extracellular iodination of tyrosine residues within the TG protein substrate[3,8]. TG, a protein dimer of 600 kDa in size, has an unusually high number (about 60) of disulfide bonds per monomer and 17 glycosylation sites, which confer notable stability and solubility to the protein[1,2,9]. Approximately 30 of the 66 tyrosines in the TG monomer are iodinated and, of these, only a small number are hormonogenic (that is, act as a substrate for the formation of thyroid hormones). In the hormonogenic sites, after iodination the aromatic ring of a donor di- or mono-iodotyrosine is transferred to a proximal acceptor di-iodotyrosine; this forms a T_4 (or T_3) hormone that is still connected to the polypeptide backbone, and leaves a dehydroalanine at the donor position[10]. After endocytosis from the colloid to cytoplasmic lysosomes, TG is proteolysed in follicular cells and releases free thyroid hormones[3,11] (Extended Data Fig. 1). Furthermore, a thyroid dehalogenase (DEHAL1) recycles iodide stored in nonhormonogenic iodotyrosines of TG[12].

Previous mass spectrometry analyses of thyroid-extracted TG have identified four or more acceptor tyrosines, but the position of the donors has remained unclear[13].

We produced full-length, noniodinated recombinant human TG (rTG) in HEK293T cells[14]. rTG appears to be indistinguishable by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and cryo-electron microscopy (cryo-EM) from endogenous TG (eTG) purified from thyroid glands of patients with goitres; eTG is partly iodinated (Extended Data Fig. 2a–c). In our cryo-EM structures, TG is bilobed with average dimensions of 120 × 235 Å, which correlates well with previously reported negative-staining electron microscopy data[15]. We collected cryo-EM datasets for both rTG and eTG, and obtained very similar reconstructions at a resolution of about 3.5 Å (Extended Data Fig. 2d–g, Extended Data Table 1). The initial maps obtained by imposing C2 symmetry showed local resolutions that ranged from about 3 Å to about 6 Å. Symmetry expansion and focused refinements improved the quality of the peripheral regions (Extended Data Fig. 2e). Using a combination of de novo and homology modelling, we built an atomic model that covers 93% of the 2,749 amino acids of TG, with variable local quality (Extended Data Fig. 3).

The sequence of TG contains a large number of cysteine-rich domains, which have previously been named type-1, type-2 and type-3 TG-like repeats[16–18] (Fig. 1a). These TG repeats are spaced by linker domains and connected to a C-terminal choline-esterase-like domain (ChEL). Following the domain arrangement in the context of the 3D structure

[1]MRC Laboratory of Molecular Biology, Cambridge, UK. [2]Jožef Stefan Institute, Ljubljana, Slovenia. [3]Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Ljubljana, Slovenia. [4]Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany. [5]Max Planck Bristol Centre for Minimal Biology, University of Bristol, Bristol, UK. [6]Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, UK. *e-mail: dusan.turk@ijs.si; jyl@mrc-lmb.cam.ac.uk

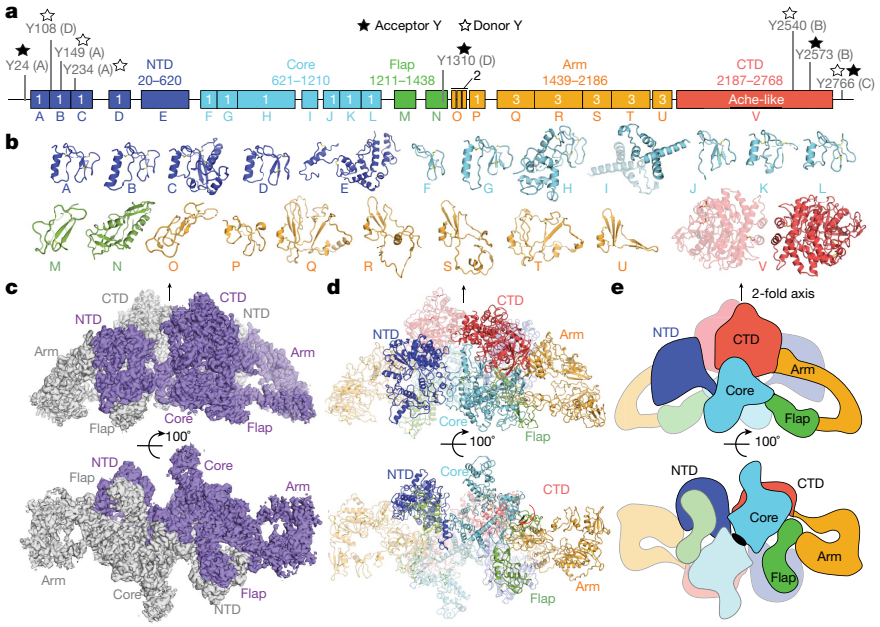


Fig. 1 | The structure of human TG by cryo-EM. a, Domain assignment of human TG. Five regions (NTD, core, flap, arm and CTD) contain domains of type-1 to type-3 TG repeats, as well as the ChEL, labelled as A to V. **b**, Structural gallery of all resolved TG domains. **c**, Cryo-EM map of TG, in which individual subunits are coloured purple or grey. The NTD crosses the major C2 interface. **d**, Ribbon diagram of **c**, coloured as in **a**. **e**, Schematic of TG structure using the same colour scheme as in **a**, **d**.

of TG, we defined five regions of TG: the N-terminal domain (NTD), core, flap, arm and C-terminal domain (CTD) (as indicated in Fig. 1a, b). The dimer interface of TG is very large, at 29,350 Å^2 (Fig. 1c–e, Supplementary Video 1). In the TG monomer, the globular NTD is connected to the core region via a linker (residues 610–620) that crosses the central dimer interface and is partially flexible (Extended Data Fig. 2f). The core contains two triplets of type-1 repeats (among these, the H domain contains a very large insertion) that are separated by the I domain, which is located near the C2 axis. The core is connected to the flap region, which is composed of two Ig-like domains (the M and N domains). The flap extends along the minor axis of the molecule, and the M domain protrudes at the opposite side of the NTD and folds back onto the arm region. The arm comprises a rod-shaped arrangement that is formed by concatenated type-2 TG repeats with a laminin-like fold and by a single type-1 repeat (the P domain). This is followed by a series of type-3 TG repeats that tightly dock onto each other in an arc towards the direction of the C2 axis. The arm is linked to the CTD region (which corresponds to the dimeric ChEL domain), which is located near the C2 axis. Overall, the structure of TG appears entangled and revolves around a central ChEL dimer that interacts with different regions of the arm and core of one chain and—via the E domain—with the NTD of the other chain. Owing to the intertwined nature of the dimer, the NTD interacts with all regions of the other subunit of the dimer. To validate the complex TG architecture, we used crosslinking mass spectrometry and found that the predicted inter- and intramolecular links from our atomic model are in excellent agreement with experimental crosslinks (Extended Data Fig. 4). For example, we detected long-range crosslinks that are consistent with the TG fold, in which the NTD crosses the ChEL dimer interface (residues 539–2524), the N terminus with residue 178 and the arm region (residues 1987–1990) (Extended Data Fig. 4d–f). Most of the disulfide bonds of TG show clear electron microscopy density in our maps (Extended Data Fig. 5a), and we found no intersubunit disulfide bonds, which is consistent with previous observations that TG is a noncovalent dimer[19]. We resolved 12 previously predicted *N*-linked acetylglucosamines (GlcNac) in our maps, and 4 that had not previously been identified (at N110, N484, N1869 and N2122)[20] (Extended Data Table 3). Some glycans are partially buried and are resistant to treatment with deglycosylases, as shown by electron microscopy. Notably, the glycans linked to N2013 mediate the contact between the NTD and CTD and might contribute to dimer stability (Extended Data Fig. 5b, f, Supplementary Video 1).

We inspected our structure within a 15 Å radius of the reported acceptor tyrosines (designated sites A–D)[21], and identified putative donor tyrosines (Fig. 2a). Acceptor Y24 (which is partially disordered) at site A appears to have two possible donor partners, Y234 and Y149 (which we term donor 1 and donor 2). In site B, the acceptor Y2573 pairs with donor Y2540; at site C, Y2766, which is not resolved in our maps, probably acts both as donor and acceptor across the C2 axis of the dimer[19]. In site D, the acceptor Y1310 pairs with Y108 of the other subunit (Fig. 2a). Therefore, TG probably contains four hormonogenic acceptor tyrosines and five donor tyrosines. We verified this by comparing the amounts of thyroid hormones obtained from unmodified rTG and rTG variants in which the acceptor or donor tyrosines were mutated to phenylalanines, which abolishes hormone formation. Thyroid-hormone synthesis was performed by rTG in vitro iodination, and the T_4 or T_3 concentrations were measured by adapting a commercial enzyme-linked immunosorbent assay (ELISA)[22,23] (Extended Data Fig. 6). Using our reaction conditions, we could detect substantial production only of T_4 and not of T_3 (Extended Data Fig. 6f, g). We used lactoperoxidase in all other reactions, because it showed the same activity as TPO in our assay[8]. As shown in Fig. 2c, when we mutated all acceptors (Y24F, Y2573F, Y2766F and Y1310F)[13,21] we observed no T_4 synthesis, demonstrating that there are no more than four sites in TG that contribute substantially to hormonogenesis. Residual activity, corresponding to about a third of unmodified rTG activity, was detected after mutating all but one of the proposed donors for site A: those at sites B, C and D (Y2540F, Y2766F and Y108F), and either the donor 1 or donor 2 at site A (Y234F and Y149F). When all five proposed donors were mutated, no T_4 formation could be detected, which confirms that site A indeed has two donor tyrosines (Y234 and Y149). We also mutated four exposed tyrosines, which have previously been suggested to be important for hormonogenesis[21,24] and obtained the same activity as for un-mutated rTG. Therefore, we have unequivocally determined and validated the complete set of tyrosines that are involved in the four TG hormonogenic sites (A, B, C and D)—at least under the conditions that we used. The uniqueness of these tyrosine pairs is corroborated by the analysis of all

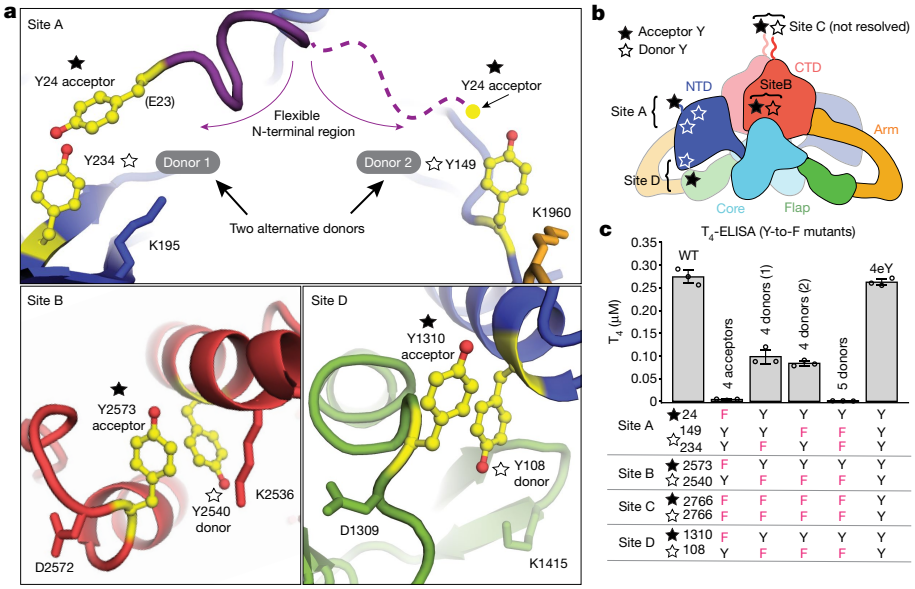


Fig. 2 | Identification and validation of hormonogenic donor–acceptor tyrosine pairs in TG. a, Close-up view of the T_4 hormonogenic sites resolved in the cryo-EM map. Donor and acceptor tyrosines are highlighted in yellow. Site A suggests two donors, Y234 and Y149. **b**, Location of the four hormonogenic sites (A to D) in the TG structure. **c**, T_4 ELISA after in vitro iodination in triplicate. Bar plot and error bars indicate mean and s.d., $n = 3$. Replacing all acceptor tyrosines (Y) with phenylalanines (F) prevents hormone formation (4 acceptor residues: 24, 2573, 2766 and 1310). Replacing all five donors suppresses T_4 synthesis (4 donor residues (1): 2540, 2766, 108 and 234; 4 donor residues (2): 2540, 2766, 108 and 149; 5 donor residues: 2540, 2766, 108, 234 and 149). Replacing other tyrosines at the surface (258, 704, 1467 and 1782) has no effect (denoted as 4eY).

tyrosine pairs at distances of less than 15 Å from each other, calculated from the TG structure (Extended Data Fig. 7a, c). We found that only the tyrosines at the hormonogenic sites of TG appear sufficiently close and exposed to permit T_4 synthesis. Within the overall TG structure, site A is located in the NTD, site B in the CTD, site C at the C terminus across the C2 dimer interface, and site D bridges the NTD and flap (specifically, the M domain) of the other subunit in the dimer (Fig. 2b).

Furthermore, we showed that all acceptor mutants with one hormonogenic site active at a time (Fig. 3a) contribute to hormone formation, and that the sum of the concentrations measured individually recapitulates the total amount of T_4 produced by unmodified rTG. We conclude that TG synthesizes seven molecules of T_4 per dimer, because each TG dimer contains two sites (one per monomer) of A, B and D and only one site C.

The resolved sites A, B and D do not show obvious structural similarities with one another, and active tyrosines are in solvent-accessible and flexible regions (Fig. 2a, Extended Data Fig. 7a, c). A conserved lysine is found in proximity to the donor, and a conserved acidic residue (Glu or Asp) always precedes the acceptor[21]. By mutagenesis of site D, we showed that K1415 is not relevant—and that D1309 is essential—for hormonogenesis (Extended Data Fig. 6d). Inserting an additional Ser-Asp sequence before acceptor Y1310 did not change the amount of T_4 produced. Therefore, overall the presence of the aspartate seems to be more important than a defined distance between the acceptor and the donor tyrosines.

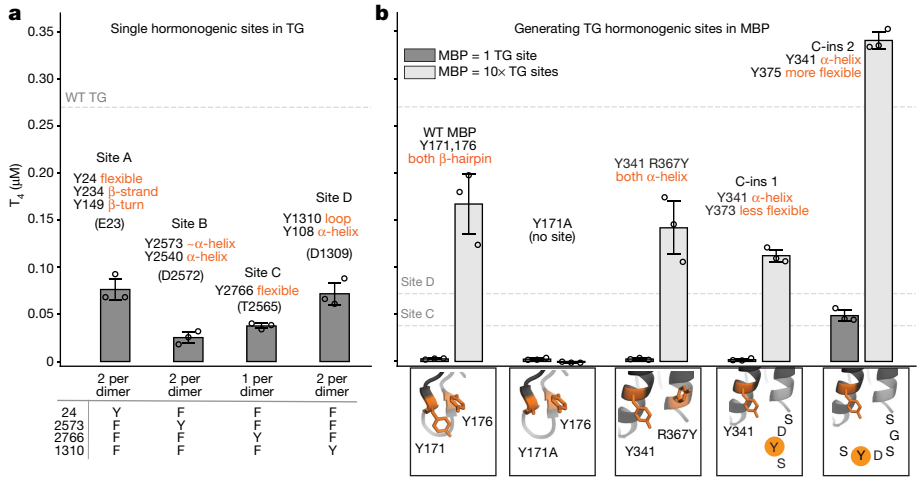


Fig. 3 | Engineering T_4 hormone synthesis by bacterial MBP. a, rTG mutants with only one site active at a time. T_4 release measured with the same T_4 ELISA as shown in Fig. 2c. The sum of T_4 produced by the individual sites recapitulates the T_4 produced by wild-type (WT) TG. ~α-helix, approximately assuming an α-helical fold. **b**, T_4 ELISA after in vitro iodination of engineered MBP to reconstruct the hormonogenic sites of TG. The main text and Supplementary Video 2 provide further details. An MBP version that adds a flexible SGSDYS tail to the C terminus shows activity comparable to a single site on TG (shown in **a**). All measurements were performed in triplicate. Bar plot and error bars indicate mean and s.d., $n = 3$.

Because in sites A, B and D the donor backbone conformation is constrained within secondary structure elements (whereas the acceptor is in more flexible regions), the acidic Asp or Glu residues (which point towards the solvent) could have a role in orienting the following acceptor towards the more-rigid donor. There is no preceding acidic residue in the C-terminal site C, in which both the acceptor and the donor (the Y2766 residues of each of the two chains of the dimer) are in unstructured regions with higher intrinsic flexibility.

Although the reaction mechanism of T_4 synthesis from polypeptide chains remains to be established in more detail, four features seem to be crucial for hormonogenesis: tyrosine pairs must be solvent-exposed to be iodinated, in proximity and in roughly antiparallel orientation, and in highly mobile regions of the protein to allow the considerable bond rearrangement that results in T_4 synthesis.

To validate our list of requirements for hormonogenesis from polypeptide chains, we set out to engineer synthetic T_4 hormonogenic sites into the unrelated bacterial MBP. The MBP of *Escherichia coli* naturally contains 15 tyrosines that are found mostly in its hydrophobic core, except for the Y171-Y176 pair (which are in a solvent-exposed β-hairpin) and Y341 (which is located in a solvent-exposed helix that faces the C-terminal helix) (Extended Data Fig. 7b, d). We engineered a putative partner tyrosine for Y341, first via the mutation R367Y (in a neighbouring helix) and second by inserting a tyrosine-containing peptide at the C terminus: SDYS (C-ins1) or SGSDYS insert (C-ins2). We measured T_4 hormonogenesis with the same ELISA as was used for TG (Extended Data Fig. 7a), at two concentrations—one corresponding to a single TG site and another 10× higher, to allow detection of marginal activities (Fig. 3b). The T_4 hormone production of unmodified MBP was measurable only at the higher concentration. We attribute this basal activity to the Y171-Y176 pair, because it could be suppressed by introducing a Y171A mutation. The low level of T_4 production can be attributed to a lack of flexibility of the backbone near Y171 and Y176 and to their relative orientations, a condition that is different to any of the sites in TG. Equally, the hormonogenic activity was unchanged when introducing the pair Y341-Y367—presumably because both tyrosines were constrained by their rigid α-helical backbone, and the coupling reaction could not occur with any effectiveness. The C-ins1 variant, which produces a putative tyrosine pair with Y341, also showed low activity, possibly because the added linker was designed to be too short for the two tyrosines to come into proximity. Importantly, however, the amount of T_4 that we measured for the two-residue-longer MBP variant C-ins2 was comparable to a single site in TG (Fig. 3b, Supplementary Video 2). Finally, with our ELISA we could also obtain modest T_4 synthesis from random tyrosine copolymers, as previously reported[4] (Extended Data Fig. 6e).

The complex and large scaffold of the TG dimer has been selected to synthesize only seven molecules of hormone, using a chemical reaction involving radicals that could be carried out by unfolded peptides or other less-complex proteins. However, in the context of the thyroid gland and the iodine cycle in vertebrates, the structure of TG effectively combines hormonogenesis with iodination of many other solvent-exposed tyrosine residues for iodine storage[13,21]. The solvent exposure of tyrosines is presumably compensated for by the exceptional solubility and stability that allows TG to persist at high concentrations in the harsh environment of the colloid. Moreover, the complexity of the TG molecule might fulfil further important roles in endocytosis, regulation of the T_3/T_4 ratio, TG proteolytic processing and in the trafficking to lysosomes[25,26]. The atomic structure of human TG presented here will enable further studies towards a deeper understanding of thyroglobulin within the thyroid, and its involvement in thyroid diseases[27–29].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-1995-4.

1. Di Jeso, B. & Arvan, P. Thyroglobulin from molecular and cellular biology to clinical endocrinology. *Endocr. Rev.* **37**, 2–36 (2016).
2. Citterio, C. E., Targovnik, H. M. & Arvan, P. The role of thyroglobulin in thyroid hormonogenesis. *Nat. Rev. Endocrinol.* **15**, 323–338 (2019).
3. Carvalho, D. P. & Dupuy, C. Thyroid hormone biosynthesis and release. *Mol. Cell. Endocrinol.* **458**, 6–15 (2017).
4. Cahnmann, H. J., Pommier, J. & Nunez, J. Spatial requirement for coupling of iodotyrosine residues to form thyroid hormones. *Proc. Natl Acad. Sci. USA* **74**, 5333–5335 (1977).
5. Holzer, G. et al. Thyroglobulin represents a novel molecular architecture of vertebrates. *J. Biol. Chem.* **291**, 16553–16566 (2016).
6. Taylor, P. N. et al. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat. Rev. Endocrinol.* **14**, 301–316 (2018).
7. Sellitti, D. F. & Suzuki, K. Intrinsic regulation of thyroid function by thyroglobulin. *Thyroid* **24**, 625–638 (2014).
8. Xiao, S., Dorris, M. L., Rawitch, A. B. & Taurog, A. Selectivity in tyrosyl iodination sites in human thyroglobulin. *Arch. Biochem. Biophys.* **334**, 284–294 (1996).
9. Heidelberger, M. The molecular weight of thyroglobulin. *Science* **80**, 414 (1934).
10. Gavaret, J. M., Cahnmann, H. J. & Nunez, J. Thyroid hormone synthesis in thyroglobulin. *J. Biol. Chem.* **256**, 9167–9173 (1981).
11. Mondal, S., Raja, K., Schweizer, U. & Mugesh, G. Chemistry and biology in the biosynthesis and action of thyroid hormones. *Angew. Chem. Int. Ed.* **55**, 7606–7630 (2016).
12. Gnidehou, S. et al. Iodotyrosine dehalogenase 1 (DEHAL1) is a transmembrane protein involved in the recycling of iodide close to the thyroglobulin iodination site. *FASEB J.* **18**, 1574–1576 (2004).
13. Dedieu, A., Gaillard, J.-C., Pourcher, T., Darrouzet, E. & Armengaud, J. Revisiting iodination sites in thyroglobulin with an organ-oriented shotgun strategy. *J. Biol. Chem.* **286**, 259–269 (2011).
14. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D* **62**, 1243–1250 (2006).
15. Berg, G., Björkman, U. & Ekholm, R. The structure of newly synthesized intracellular thyroglobulin molecules. *Mol. Cell. Endocrinol.* **20**, 87–98 (1980).
16. 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 J.* **18**, 793–803 (1999).
17. Molina, F., Bouanani, M., Pau, B. & Granier, C. Characterization of the type-1 repeat from thyroglobulin, a cysteine-rich module found in proteins from different families. *Eur. J. Biochem.* **240**, 125–133 (1996).
18. Lee, J. & Arvan, P. Repeat motif-containing regions within thyroglobulin. *J. Biol. Chem.* **286**, 26327–26333 (2011).
19. Citterio, C. E., Morishita, Y., Dakka, N., Veluswamy, B. & Arvan, P. Relationship between the dimerization of thyroglobulin and its ability to form triiodothyronine. *J. Biol. Chem.* **293**, 4860–4869 (2018).
20. Yang, S.-X., Pollock, H. G. & Rawitch, A. B. Glycosylation in human thyroglobulin: location of the *N*-linked oligosaccharide units and comparison with bovine thyroglobulin. *Arch. Biochem. Biophys.* **327**, 61–70 (1996).
21. Lamas, L., Anderson, P. C., Foxy, J. W. & Dunn, J. T. Consensus sequences for early iodination and hormonogenesis in human thyroglobulin. *J. Biol. Chem.* **264**, 13541–13545 (1989).
22. Pommier, J., Deme, D. & Nunez, J. Effect of iodide concentration on thyroxine synthesis catalysed by thyroid peroxidase. *Eur. J. Biochem.* **37**, 406–414 (1973).
23. de Vijlder, J. J. & den Hartog, M. T. Anionic iodotyrosine residues are required for iodothyronine synthesis. *Eur. J. Endocrinol.* **138**, 227–231 (1998).
24. Dunn, J. T., Kim, P. S. & Dunn, A. D. Favored sites for thyroid hormone formation on the peptide chains of human thyroglobulin. *J. Biol. Chem.* **257**, 88–94 (1982).
25. Botta, R. et al. Sortilin is a putative postendocytic receptor of thyroglobulin. *Endocrinology* **150**, 509–518 (2009).
26. Weber, J. et al. Interdependence of thyroglobulin processing and thyroid hormone export in the mouse thyroid gland. *Eur. J. Cell Biol.* **96**, 440–456 (2017).
27. Rivolta, C. M. & Targovnik, H. M. Molecular advances in thyroglobulin disorders. *Clin. Chim. Acta* **374**, 8–24 (2006).
28. Latrofa, F. et al. Thyroglobulin autoantibodies in patients with papillary thyroid carcinoma: comparison of different assays and evaluation of causes of discrepancies. *J. Clin. Endocrinol. Metab.* **97**, 3974–3982 (2012).
29. Fiore, E., Latrofa, F. & Vitti, P. Iodine, thyroid autoimmunity and cancer. *Eur. Thyroid J.* **4**, 26–35 (2015).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

DNA constructs

A gene encoding full-length human TG (Uniprot P01266), additionally containing a 10× histidine tag at the C terminus preceded by a TEV cleavage site, was codon-optimized for mammalian expression and purchased from GenScript. The gene was cloned with EcoRI and XhoI restriction sites (enzymes purchased from NEB) into vector pLEXm[14] for recombinant expression in mammalian cells. Mutations were introduced by PCR and the purified overlapping fragments were assembled with the EcoRI/XhoI linearized vector by Gibson assembly (NEB).

MBP constructs were purchased as gBlocks (IDT) and cloned using BlpI and NdeI restriction sites (enzymes purchased from NEB) into the pHis17 bacterial vector, for the expression of N-terminally 6× histidine-tagged MBP variants. A synthetic gene comprising human *TPO* (UniProt ID P07202) 1–838 with native extracellular signal sequence was cloned into pFastBac1 plasmid adding a C-terminal 6× histidine tag. All translated sequences of recombinant proteins used in this work are listed in Supplementary Information.

Proteins, protein expression and purification

eTG from human thyroid glands was purchased from Biorad. We found that this eTG contains some T_4 (as measured by using our T_4 ELISA), but not T_3 (Extended Data Fig. 6). This eTG is also partially degraded, as shown by SDS–PAGE (Extended Data Fig. 2a). Although the protein behaved well when performing negative-staining electron microscopy, it was challenging to produce specimens amenable to high-resolution cryo-EM. Therefore, eTG was deglycosylated with the addition of 1 µl/200 µg eTG of PNGaseF from NEB for 1 h at 37 °C, and purified by size-exclusion chromatography using a Superose 6 Increase 3.2/300 column (GE Healthcare), equilibrated with buffer T200 (50 mM Tris-HCl, 200 mM NaCl, pH 8.0).

For the production of rTG, HEK293T (ATCC no. CRL-1573) cells were cultured as adherent monolayers until 90% confluency in Dulbecco's modified Eagle's medium (Sigma-Aldrich) supplemented with 10% fetal calf serum (v/v; Sigma-Aldrich), L-glutamine and nonessential amino acids (Invitrogen), and transiently transfected with 2 mg of DNA and 4 mg of polyethyleneimine (PEI; Sigma-Aldrich) per litre of culture. We note that the involvement of glycans in TG dimer formation, as revealed by the structure, might explain the poor expression yields of TG when produced in expression systems with reduced or altered glycosylation, such as insect cells, HEK293S GnT1$^{-/-}$ cells or HEK293T cells in the presence of kifunensine (data not shown).

Five days later, the supernatant containing approximately 0.5 mg/l of secreted TG was collected and filtered (0.22 µm) for protein purification. For smaller cultures (about 125 ml), the supernatant was diluted with an equal volume of buffer T200 containing 20 mM imidazole. For volumes larger than 250 ml, the supernatant was concentrated and buffer-exchanged into buffer T200 using an Äkta Flux system (GE Healthcare). Subsequently, Ni-NTA agarose beads (Qiagen) were added to the supernatant (2 ml per l of supernatant). The mixture was gently stirred at 4 °C for 1 h and the beads collected by centrifugation at 600*g* for 5 min in 50-ml tubes (Falcon, BD Biosciences). The beads were poured into a 10-ml EconoColumn (Bio-Rad Laboratories) and washed with 10 column volumes of buffer T200, supplemented with increasing concentrations of imidazole (20 mM, 50 mM, 80 mM), before elution with 5 column volumes of buffer T200 supplemented with 500 mM imidazole (all buffers adjusted to pH 8.0). Eluting fractions containing TG, according to SDS–PAGE analysis, were pooled and concentrated before further purification by size-exclusion chromatography using a Superose 6 Increase 3.2/300 column (GE Healthcare), or a Superose 6 10/300 GL column for larger amounts of protein. Fractions containing TG were joined, concentrated by ultrafiltration, aliquoted at a concentration of about 0.5 mg/ml, flash-frozen in liquid nitrogen and stored at −80 °C.

For MBP variants, the plasmid was transformed into chemically competent C41(DE3) *E. coli* cells. One-hundred-millilitre bacterial cultures were grown in 2× TY medium at 37 °C in the presence of 100 µg/ml ampicillin, and protein expression was induced at an optical density at 600 nm (OD_{600}) of about 1.0 with 1 mM IPTG for 5 h. Cells were collected by centrifugation at 5,000*g*, resuspended in buffer T200, supplemented with 20 mM imidazole, DNase, RNaseA, lysozyme (Sigma Aldrich) and protease inhibitors (Roche), and lysed by sonication. The soluble fraction was separated by ultracentrifugation at 40,000*g* and MBP protein variants were purified as described for TG, with a size-exclusion chromatography final purification step using a Superdex 200 3.2/300 column.

For human TPO ectodomain expression in insect cells, bacmids were prepared using DH10EMBacY competent cells (Geneva Biotech), following the Bac-to-Bac Baculovirus Expression System user guide by Invitrogen. Sf9 cells were transfected with FuGENE Transfection reagent (Promega). V_1 virus was used for large-scale protein production in 400 ml, and cells were supplemented with 5-aminolevulinic acid, which is a precursor of the haem prostetic group and has previously been reported to increase specific activity of TPO[30]. Four days after infection, cells were collected via centrifugation (800*g*, and then 10,000*g*, both at 4 °C) and the supernatant was concentrated by tangential flow filtration. The concentrated supernatant was dialysed against PBS, pH 7.4 and TPO was purified by nickel affinity purification (GE Healthcare HisTrap HP, 1-ml column) using as binding buffer PBS plus 20 mM imidazole, pH 7.4 and as elution buffer PBS plus 400 mM imidazole, also at pH 7.4. Relevant fractions were pooled, concentrated by ultrafiltration and TPO was further purified using a Superdex 200 10/300 column (GE Healthcare), pre-equilibrated with PBS, pH 7.4. TPO was concentrated to 1 mg/ml and flash-frozen in liquid nitrogen for storage at −80 °C. The haem occupancy of TPO was estimated by measuring the ratio of absorbance at 412 nm and 280 nm, which was 0.2 (which roughly equates to 20% occupancy)[30].

Electron microscopy sample preparation and data collection

Graphene oxide grids were prepared following a published procedure[31] using as support Quantifoil Cu/Rh 200 mesh R2/2 grids. Three microlitres of TG sample was applied to the graphene oxide grids at a concentration of approximately 0.05 mg/ml and plunge-frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher). Whereas eTG behaved better when deglycosylated, there was no difference between glycosylated and deglycosylated rTG. Therefore, we collected datasets of deglycosylated eTG and glycosylated rTG.

Images were acquired on a K2 Summit detector (Gatan) in counting mode using a Titan Krios G2 (Thermo Fisher) electron microscope at 300 kV. A Quantum GIF energy filter (Gatan) was used with a slit width of 20 eV to remove inelastically scattered electrons. The eTG dataset was collected at eBIC (Diamond Light Source) and the rTG dataset was collected at MRC-LMB. For the eTG dataset, 40 movie frames were recorded, using a fluency of 1.18 electrons per Å^2 per frame, for a total accumulated dose of 47.2 electrons per Å^2 at a pixel size of 1.043 Å on the specimen. For the rTG dataset, 52 movie frames were recorded, using a fluency of 0.91 electrons per Å^2 per frame, for a total accumulated dose of 36.3 electrons per Å^2 at a pixel size of 1.149 Å on the specimen. Further details are presented in Extended Data Table 1.

Cryo-EM image processing

Movie frames were corrected for gain using a reference, motion-corrected and dose-weighted using MOTIONCOR2[32]. Aligned micrographs were used to estimate the contrast transfer function (CTF) in Gctf[33]. All subsequent image-processing steps were performed using

single-particle reconstruction methods in RELION 2.1 or 3.0[34,35]. Poor-quality images were discarded after manual inspection. Particles were initially manually picked to generate 2D class references for automated picking in RELION. After picking the whole dataset automatically, particles were extracted with 4 × 4 binning and two rounds of reference-free 2D classifications were performed. The particles belonging to the best 2D classes were extracted un-binned (400-pixel box size) and used for 3D reconstruction, applying C2 symmetry. The resolution was estimated with a Fourier shell correlation (FSC) criterion of 0.143 and *B*-factor sharpening was applied, both using the relion_postprocess routine. Local map resolution was estimated with RELION. Bayesian polishing and per particle CTF and tilt correction were performed, but did not provide a substantial improvement in resolution.

The central regions of the maps, corresponding to the C-terminal ChEL domains, were well-defined (about 3 Å in resolution), whereas peripheral regions corresponding to the arms were noisy and at lower resolutions (about 6 Å). To improve resolution (especially in the arm regions), we expanded the dataset using the relion_symmetry_expand routine by the C2 symmetry and re-extracted the particles centred at the least-resolved part of the map[36] (Extended Data Fig. 2d). After applying a soft mask to one half of the TG dimer, several cycles of refinement and 3D classification were run using solvent flattening. Although the overall resolution of the expanded and re-centred map was very similar to the C2 counterpart, the local resolution and continuity of the density in the arm region was considerably improved (Extended Data Fig. 2e). The same procedure was applied to both eTG and rTG datasets, which at the end looked virtually identical (as shown in Extended Data Fig. 2g). The best final maps, which we used for model building, were the C2 map from the eTG dataset (overall resolution of 3.39 Å) and the expanded map from the rTG dataset (overall resolution of 3.67 Å). To allow simultaneous model building in both maps, the expanded rTG map was aligned and resampled onto the C2 map, and subsequently C2-symmetrized for the final dimer TG model.

Model building

Following the domain annotations reported in UniprotKB for TG (P01266), we generated homology models with SWISSMODEL[37], as reported in Extended Data Table 2. We visually inspected the map and localized matching domains for some of the models, which we fitted by correlation in Chimera[38]. First, the large ChEL domain at the dimeric TG interface was identified, and then triplets of TG type-1 repeats were fitted[16,17]. Their correct order was determined by the differences in side-chain densities surrounding the CWC motif and the characteristic loop insertions of each domain. The best-resolved type-3 repeat was the U domain, which we built de novo in MAIN[39] and used as a template to generate homology models for the other type-3 repeats (Q, R, S and T domains). The connecting regions with unknown folds were build de novo in MAIN[39] and COOT[40], with help of secondary structure predictions (HHpred[41] and Jpred[42]). The correct assignment of the polypeptide chain register was often helped by the presence of large aromatic side chains and disulfide bond pairs, as well as the presence of glycosylation sites at Asn residues (12, plus 4 additional sites that we identified) (Extended Data Fig. 5a–f, Extended Data Table 3). Residues 24 (acceptor site A) to 29 were built tentatively into a map filtered to lower resolution using the MAIN-score map-density modification procedure[39]; however, the coordinates of these residues are only indicative, and they are therefore included in the model as poly-alanine (Extended Data Fig. 6g).

Initially, the full model was built in one half of the rTG map (expanded, re-centred and masked) and refined using Phenix.real_space_refinement[43]. Subsequently, the full dimeric structure was generated by applying C2 symmetry and further refined in the C2-symmetrized rTG map. Final statistics and validation of the model are reported in Extended Data Fig. 2 and Extended Data Table 1. The resolution of the map is non-uniform and, consequently, the model has variable quality depending on the map region. To illustrate this, per residue *B*-factor and per-residue map-to-model cross correlation plots as calculated in Phenix are provided in Extended Data Fig. 3.

Crosslinking and mass spectrometry analysis

An rTG aliquot of 100 µl at 0.5 mg/ml was buffer-exchanged using a Superose 6 Increase 3.2/300 column (GE Healthcare), pre-equilibrated with 20 mM Hepes, 200 mM NaCl at pH 7.5. Pooled fractions were incubated for 2 h on ice at 0.5 mg/ml with or without 1 mM bis(sulfosuccinimidyl suberate (BS^3) (Thermo Fisher). The crosslinking reaction was quenched by adding 50 mM ammonium bicarbonate and the product was analysed using a Superose 6 Increase 3.2/300 column pre-equilibrated in buffer T200 (Extended Data Fig. 4). Gel bands corresponding to crosslinked rTG were excised and digested with trypsin (Pierce) following an in-gel digestion protocol[44]. The resulting tryptic peptides were extracted and desalted using C18 StageTips[45].

The enrichment of crosslinked peptides was accomplished by size-exclusion chromatography using a Superdex Peptide 3.2/300 column (GE Healthcare). The mobile phase consisted of 30% (v/v) acetonitrile and 0.1% trifluoroacetic acid, running at a flow rate of 10 µl/min. The earliest five peptide-containing fractions (50 µl each) were collected and dried in a vacuum concentrator.

Liquid chromatography–tandem mass spectrometry analysis was performed using an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific), connected to an Ultimate 3000 RSLCnano system (Dionex, Thermo Fisher Scientific). Samples were resuspended in 1.6% v/v acetonitrile 0.1% v/v formic acid and injected onto an EASY-Spray column of 50-cm length (Thermo Fisher) running at 300 nl/min with mobile phases A (0.1% formic acid) and B (80% acetonitrile, 0.1% formic acid). Samples were eluted by applying a gradient ranging from 2% to 45% B over 90 min. Each gradient was optimized for the corresponding size-exclusion chromatography fraction. After this, a washing step was applied in which the content of B was ramped to 55% and 95% within 2.5 min each, followed by 5 min at 95% B. Each fraction was analysed in duplicate. The settings of the mass spectrometer were as follows: data-dependent mode with 3-s top-speed setting; MS1 scan in Orbitrap at 120,000 resolution over 400 to 1,600 m/z; MS2 scan trigger only on precursors with $z = 3–6+$; fragmentation by higher-energy collisional dissociation using a decision tree logic with optimized collision energies; MS2 scan in Orbitrap at resolution of 30,000; dynamic exclusion was enabled upon single observation for 60 s. Generation of fragment spectra peak lists from raw mass spectrometric data used msConvert (version 3.0.11729), operating under default settings. Precursor m/z values were recalibrated and the crosslink search was performed using Xi[46] using the isoform-1 sequence of TG without N-terminal signal peptide sequence and extra residues from TEV cleavage. Decoy sequences were generated by reversing the protein sequence. For the search MS1 and MS2, accuracies were set to 3 and 10 ppm, respectively. Tryptic peptides (full trypsin specificity) with up to four missed cleavages were allowed. The reaction specificity of BS^3 was restricted to the side chains of lysine, serine, threonine and tyrosine, as well as the protein N termini. Carbamidomethylation on cysteine was set as fixed; oxidation on methionine, hydrolysed or aminolysed BS^3 from hydrolysis or ammonia quenching on a free crosslinker end were set as variable modifications. Identified crosslinked peptide candidates were filtered to a false discovery rate (FDR) of 2% on residue-pair-level using XiFDR[47]. A list of all experimental crosslinks is reported in Supplementary Table 1. Inter- and intramolecular theoretical crosslinking pairs were calculated from our TG atomic structure and overlapped with the experimental pairs with a Xi score >12 and an estimated FDR = 0 (Extended Data Fig. 4c).

T_4 and T_3 ELISA

We produced thyroid hormones from recombinant (noniodinated) proteins following the procedures reported previously for poorly iodinated

eTG[22], and quantified the T_4 and T_3 products via ELISA designed to work in blood serum (Abcam, ab108661, human thyroxine ELISA kit; Abcam, ab108664, human tri-iodothyronine ELISA kit (free + total T_3)). To perform the assays, protein at a concentration of 0.1 µM was added to 1 mM KI, 24 mM glucose, 2 µg/ml glucose oxidase and 3 µg/ml lactoperoxidase or TPO (Extended Data Fig. 6c). All commercial reagents were purchased from Sigma. The iodination reaction was allowed to proceed for 10 min at 37 °C and then Pronase protease mix (Roche) was added at about 2.5 µg/ml to digest all enzymes, and TG to release thyroid hormones. This was followed by heat inactivation for 15 min at 95 °C, a step that does not affect the stability of free T_4 or T_3[48]. The reaction product was diluted to measure thyroid-hormone production in dynamic range compatible with the ELISA kits. Subsequently, we added BSA at a concentration of 60 mg/ml to the mixture as an essential blocking agent, and added the mixture to ELISA plates for detection, following the manufacturer's instructions. To check whether any of the iodination components were interfering with the ELISA (which is optimized for thyroid-hormone detection in serum) we performed the assay without iodide but adding known amounts of T_4 or T_3 (Sigma Aldrich). This yielded calibration curves that were similar to the ones provided by the manufacturer, and also showed good dynamic range (Extended Data Fig. 6). Only T_4 was produced in detectable amounts in our in vitro thyroid-hormone production assay. Considering that TG has 3–4 sites, 0.1 µM was determined to be the optimal starting concentration for the assay. Positive controls were performed with eTG (containing some hormones and partly iodinated) and negative controls with T_3 (Sigma), lysozyme and with FtsZ from *Staphylococcus aureus*, which does not contain tyrosines. Using our calibration curve, we converted the absorbance into T_4 concentration, performed each measurement three times independently and reported the average and s.d. values (Figs. 2,3, Extended Data Fig. 6). Recombinant TPO was added at a five-times-higher concentration, compensating for its haem content of only approximately 20%. Tyrosine copolymers were purchased from Sigma Aldrich (P4659, P0151, P4409 and P1800) and dissolved in T200 buffer.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Datasets generated during the current study are available from the RCSB Protein Data Bank (PDB) accession code 6SCJ, Electron Microscopy Data Bank (EMDB) accession code EMD-10141 and ProteomeXchange accession code PXD014821. All other data generated or analysed during this study are included in this published article and its Supplementary Information.

30. Guo, J., McLachlan, S. M., Hutchison, S. & Rapoport, B. The greater glycan content of recombinant human thyroid peroxidase of mammalian than of insect cell origin facilitates purification to homogeneity of enzymatically protein remaining soluble at high concentration. *Endocrinology* **139**, 999–1005 (1998).
31. Bokori-Brown, M. et al. Cryo-EM structure of lysenin pore elucidates membrane insertion by an aerolysin family protein. *Nat. Commun*. **7**, 11293 (2016).
32. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
33. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol*. **193**, 1–12 (2016).
34. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol*. **180**, 519–530 (2012).
35. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
36. Li, Y. et al. Mechanistic insights into caspase-9 activation by the structure of the apoptosome holoenzyme. *Proc. Natl Acad. Sci. USA* **114**, 1542–1547 (2017).
37. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. **46**, W296–W303 (2018).
38. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem*. **25**, 1605–1612 (2004).
39. Turk, D. MAIN software for density averaging, model building, structure refinement and validation. *Acta Crystallogr. D* **69**, 1342–1357 (2013).
40. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
41. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. **33**, W244–W248 (2005).
42. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res*. **43**, W389–W394 (2015).
43. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
44. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc*. **1**, 2856–2860 (2006).
45. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem*. **75**, 663–670 (2003).
46. Mendes, M. L. et al. An integrated workflow for crosslinking mass spectrometry. *Mol. Syst. Biol*. **15**, e8994 (2019).
47. Fischer, L. & Rappsilber, J. Quirks of error estimation in cross-linking/mass spectrometry. *Anal. Chem*. **89**, 3829–3833 (2017).
48. Ledeți, I. et al. Thermal stability of synthetic thyroid hormone l-thyroxine and l -thyroxine sodium salt hydrate both pure and in pharmaceutical formulations. *J. Pharm. Biomed. Anal*. **125**, 33–40 (2016).

Acknowledgements We thank C. Savva, G. Cannone and S. Chen for help with electron microscopes; S. Scheres, R. F. Leiro, J. Zivanov, P. Emsley, V. Chandrasekaran and S. Masilius for image processing and model building advice; R. Aricescu for help with the expression of TG in mammalian cells; C. Heroven and D. Laverty for help with mammalian tissue culture; F. Bürmann and G. Slodkowicz for help with data analysis; F. van den Ent and T. Nierhaus for advice on protein work; T. Darling and J. Grimmett for computing support; and D. Clare for electron microscopy data collection. We acknowledge the Diamond Light Source for access and support of the cryo-EM facilities at the UK's National Electron Bio-imaging Centre (eBIC), funded by the Wellcome Trust, MRC and BBRSC. This work was funded by the Medical Research Council (U105184326 to J.L.), the Wellcome Trust (202754/Z/16/Z to J.L., 203149 to J.R.) and by the Slovenian Research Agency (ARRS; P1-0048, IO-0048 and J1-7479 to D.T.). This work was supported by the Wellcome Trust through a Senior Research Fellowship (103139 to J.R.) and by the DFG, German Research Foundation (329673113 and EXC 2008/1 – 390540038 to J.R.).

Author contributions F.C. performed TG mammalian expression, cryo-EM, TG biochemistry, ELISA design and data analysis. F.C., J.L. and D.T. built and refined the TG model. A.T.-V. and I.B. performed TPO biochemistry. V.T.C. performed expression of TG in mammalian cells with F.C. T.I. collected initial negative staining data on eTG. M.R. developed initial eTG purifications. L.S., F.J.O. and J.R. performed crosslinking mass-spectrometry analysis. F.C. and J.L. wrote the manuscript with assistance of other authors. J.L., F.C., D.T. and A.T.-V. were responsible for project strategy and data interpretation.

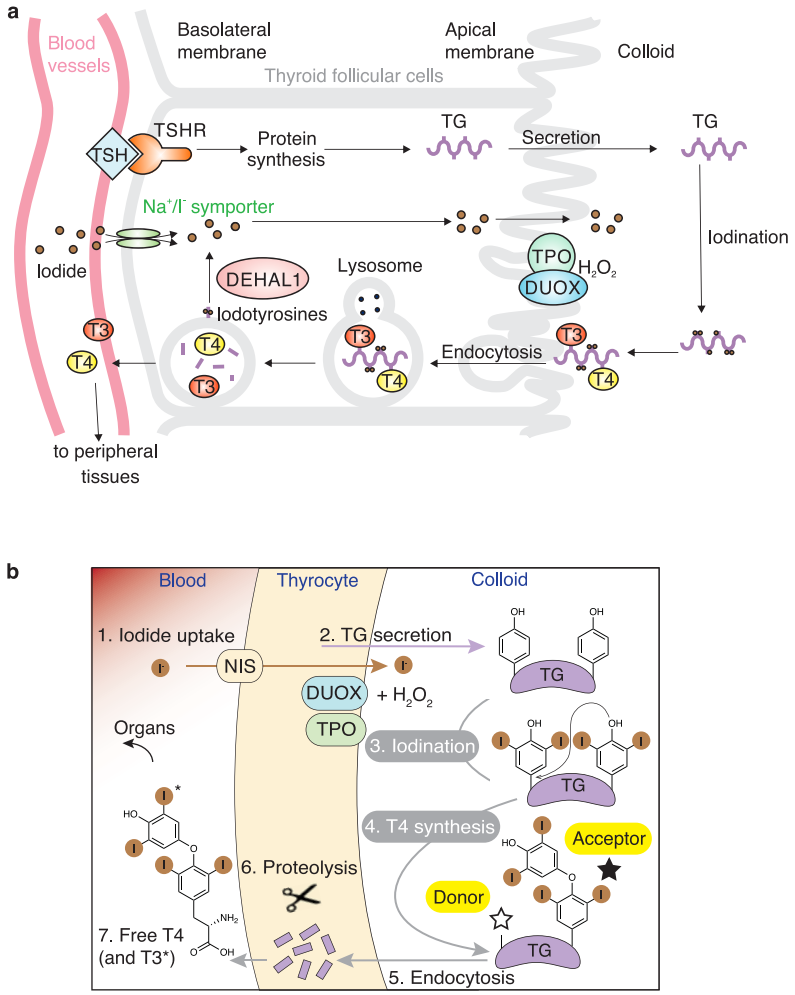
Competing interests The authors declare no competing interests.

Additional information

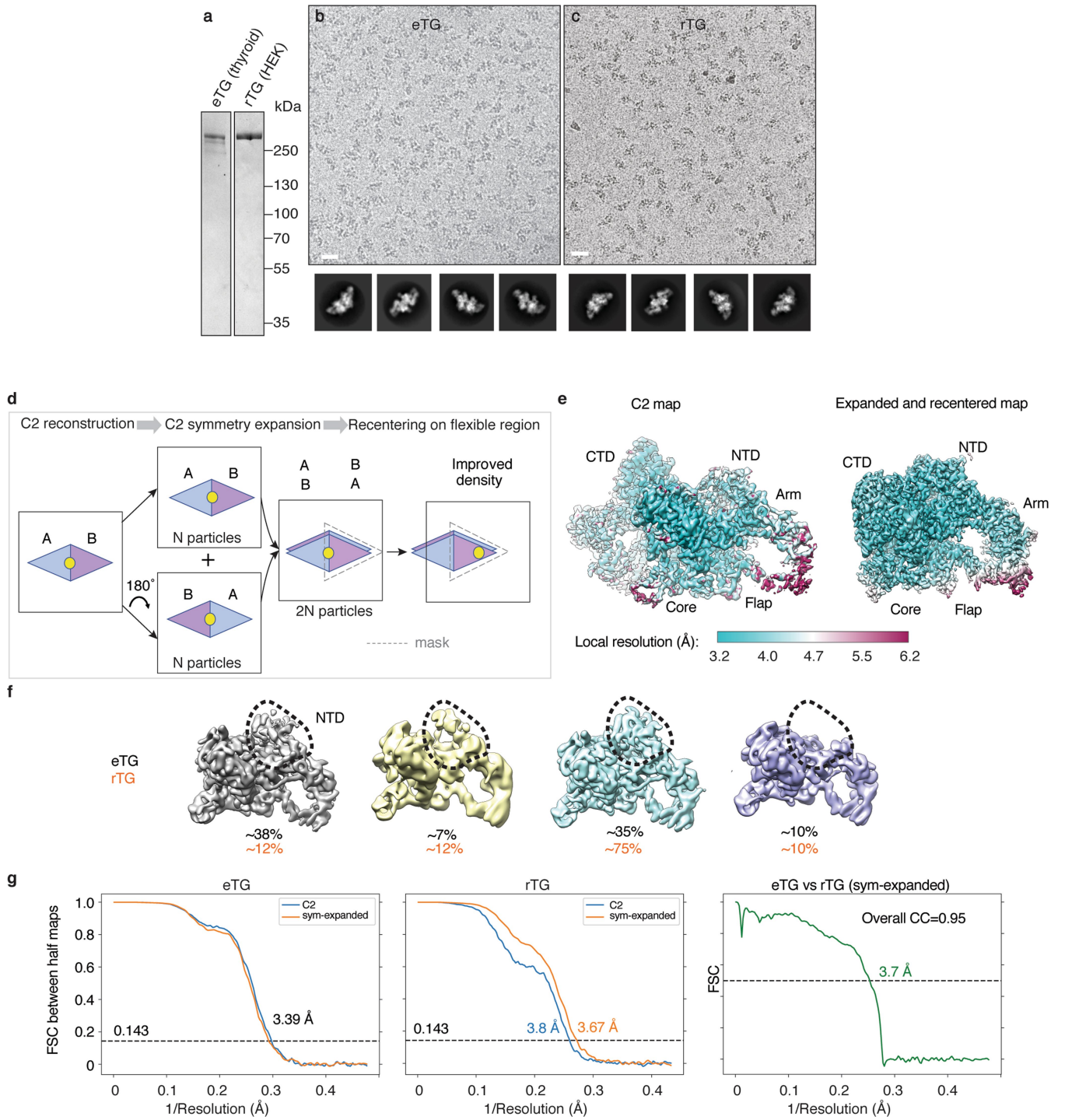
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-1995-4.

Correspondence and requests for materials should be addressed to D.T. or J.L.

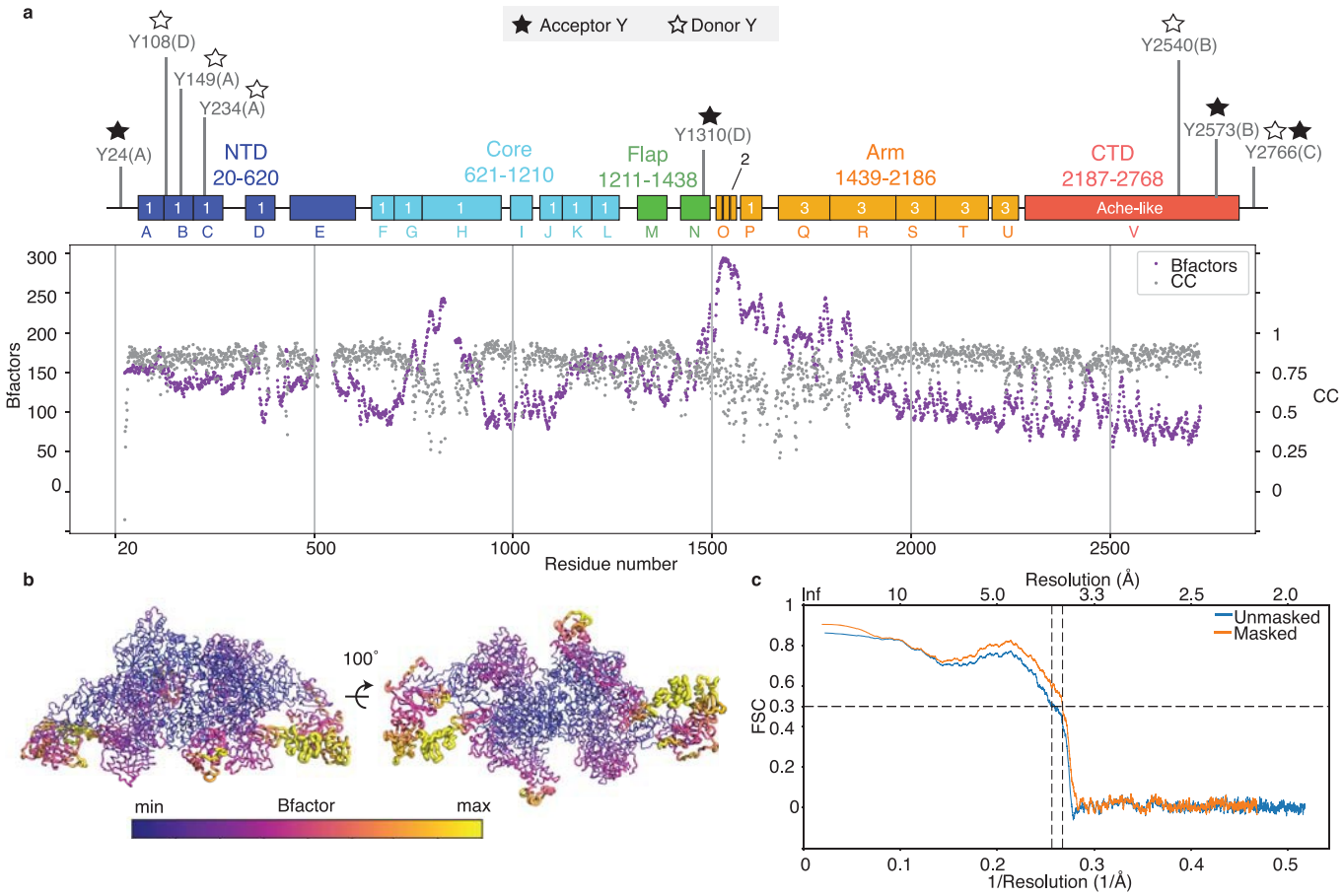
Peer review information *Nature* thanks Ulrich Schweizer, Janet Vonck and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



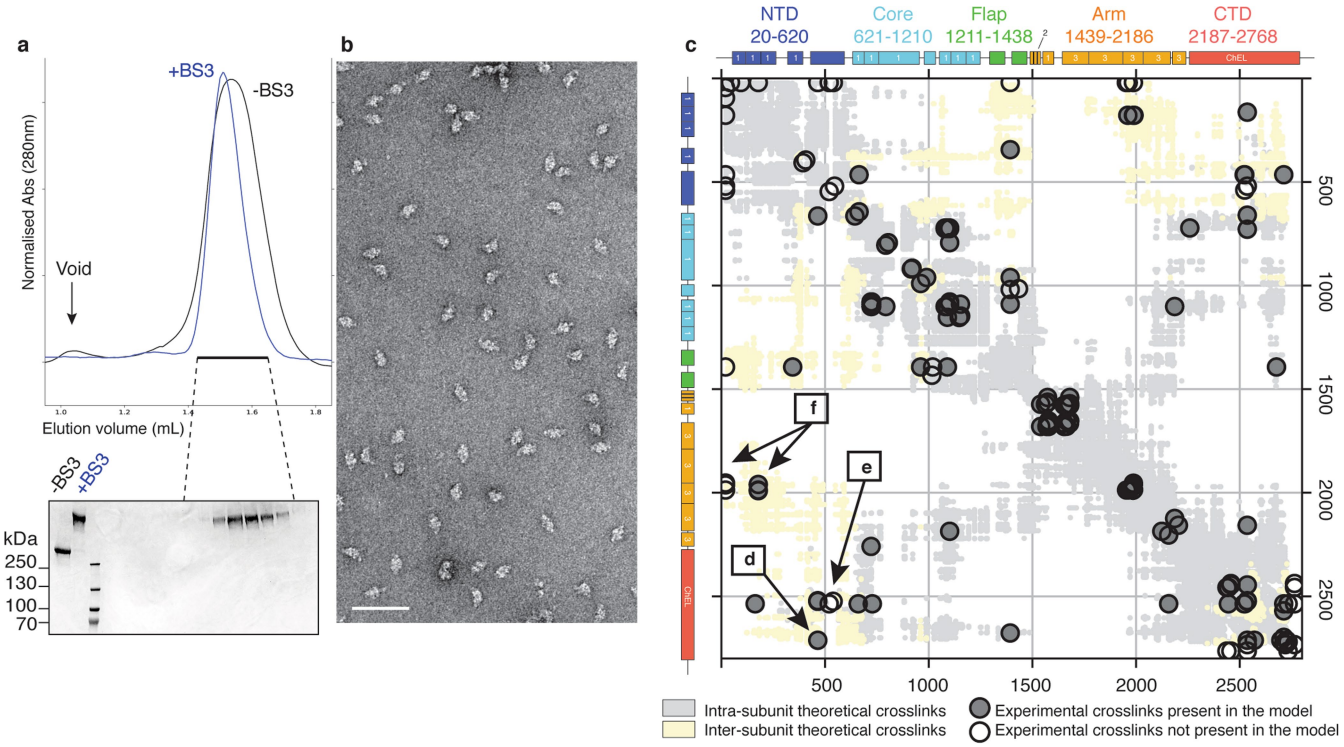
Extended Data Fig. 1 | The iodine cycle in the thyroid gland and the chemistry of thyroid-hormone formation. a, Iodide is extracted from the blood vessels and into the thyroid cells via the Na^+ and I^- symporter. TSH binds TSH receptor (TSHR) to induce the expression of TG. TG is secreted into the extracellular lumen of follicular cells (also called colloid). DUOX and TPO catalyse the iodination of TG; therefore, T_4 (or T_3) hormones are formed on the TG polypeptide chain. After hormonogenesis, TG is reimported and proteolysed in lysosomes to release T_4/T_3 into the blood. DEHAL1 de-iodinates iodotyrosines to recycle iodide in thyroid cells. **b**, T_4 (or T_3) synthesis from TG in the thyroid gland.

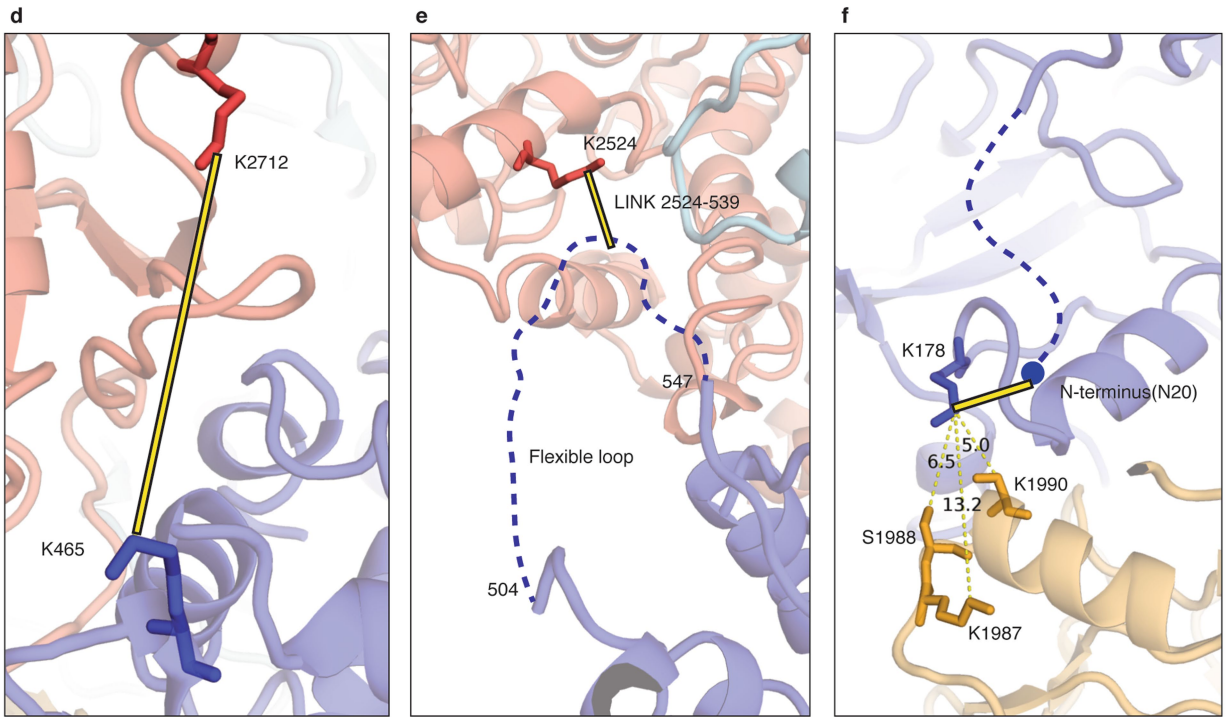


Extended Data Fig. 2 | Cryo-EM reconstruction of eTG and rTG. a, SDS–PAGE of eTG from extracts from goitrous thyroid, and rTG expressed in HEK293T cells. **b**, Cryo-EM micrograph of eTG with calculated reference-free 2D class averages below. Scale bar, 200 Å. **c**, Cryo-EM micrograph of recombinant rTG with 2D class averages, showing the two proteins to be structurally identical at this level of analysis. Scale bar, 200 Å. **d**, Schematic illustrating the C2 symmetry expansion and recentring procedure, which was used to enhance TG map quality in peripheral regions. For a detailed procedure see 'Cryo-EM image processing' in Methods. **e**, Local resolution of the C2 and symmetry-expanded and recentred eTG maps. **f**, Flexibility of NTD results in varying map quality and occupancy of this region in a number of 3D class averages (calculated in RELION). **g**, FSC between RELION 'gold standard' half-maps and between the final eTG and rTG maps, showing their strong similarity.

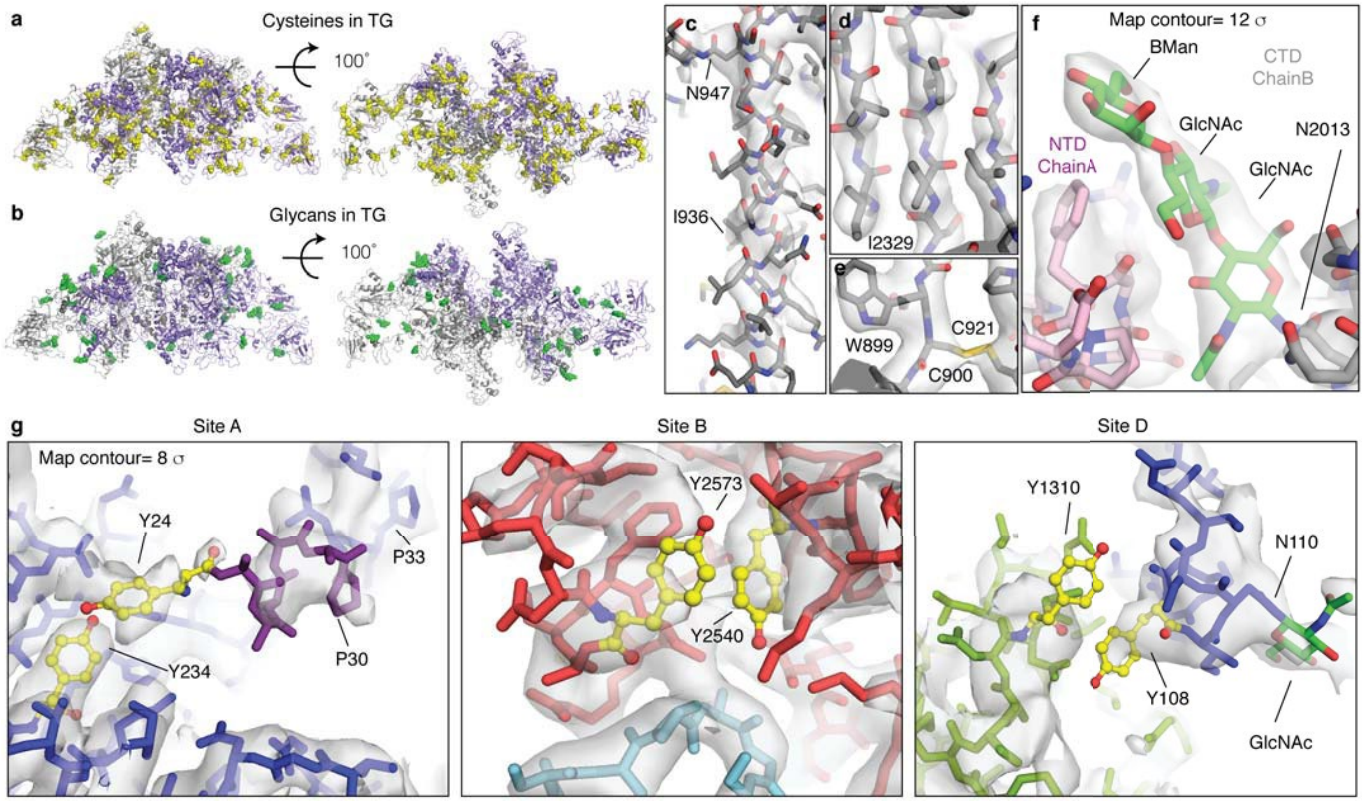


Extended Data Fig. 3 | Local properties of the atomic TG model. a, Per-residue atomic *B*-factor and cross-correlation with the rTG map, plotted per residue number. **b**, Local *B*-factor colour-coded onto the surface of the TG structure. **c**, FSC between the map and model calculated for rTG. FSC 0.5 is indicated.

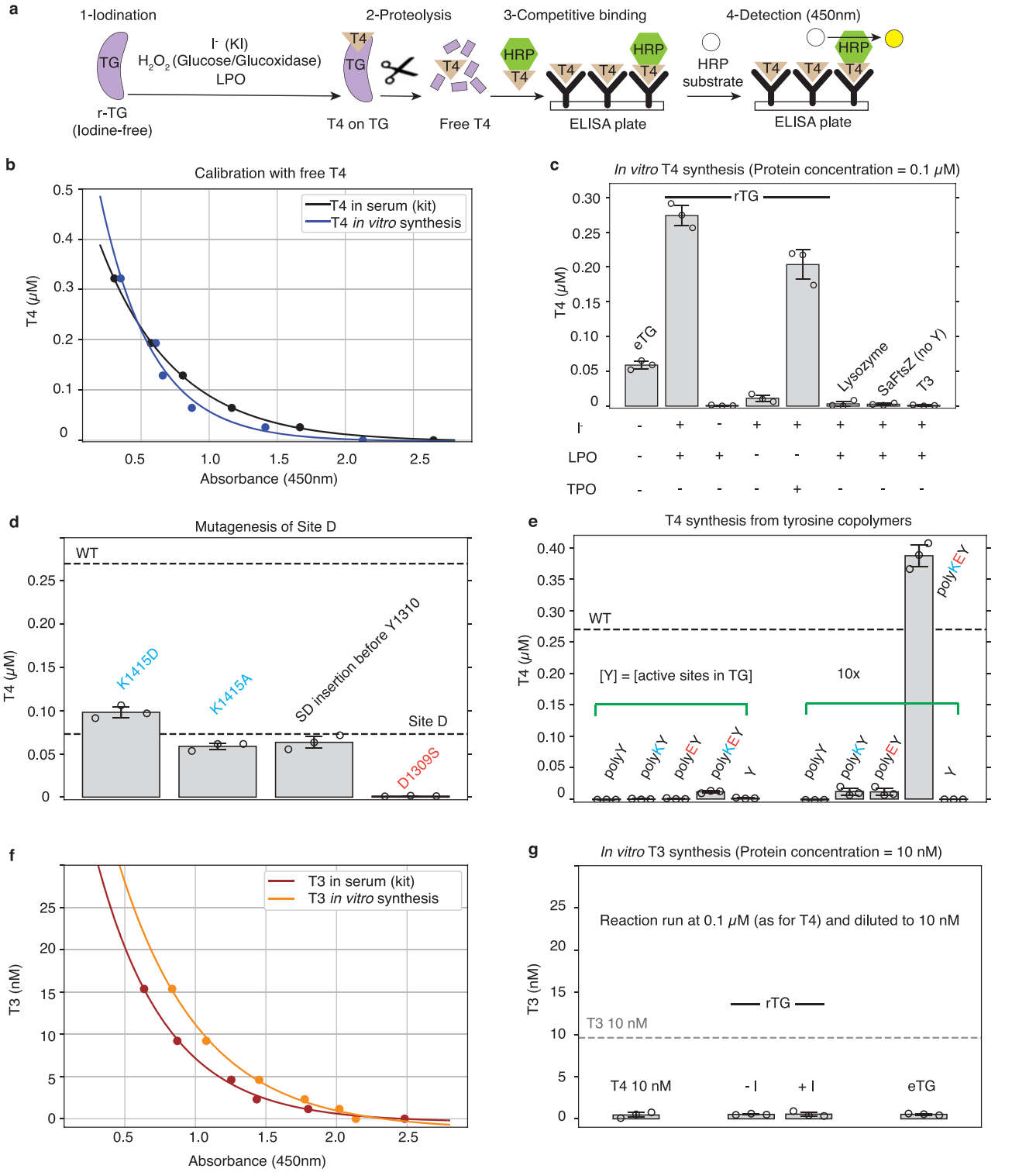




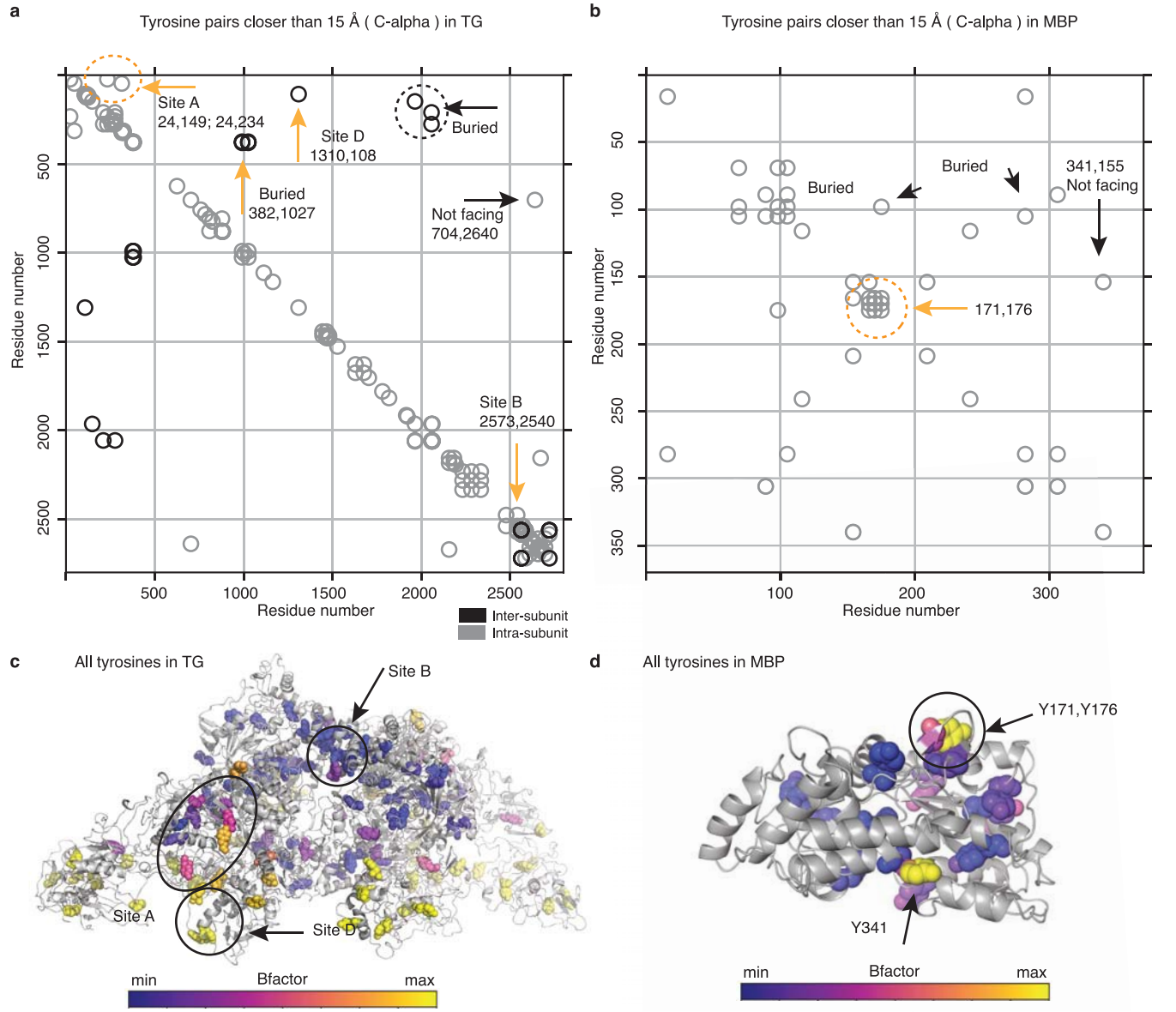
Extended Data Fig. 4 | Validation of the 3D architecture of TG by mass spectrometry crosslinking. a, Size-exclusion chromatograms of rTG before and after BS^3 crosslinking and subsequent SDS–PAGE (Coomassie-stained). **b**, Negative-staining micrograph of crosslinked rTG, showing the absence of higher-order structures caused by unwanted inter-dimer crosslinks. **c**, Plot representing experimental crosslinks (circles) overlapping with predicted crosslinks, calculated from the structure determined here. **d**–**f**, Detail of key TG interfaces confirmed by the crosslinking.



Extended Data Fig. 5 | Details of TG cryo-EM map. a, All disulfide bonds in TG included in the model (yellow spheres). **b**, Glycans detected in the cryo-EM maps and included in the TG atomic model (green spheres). **c**, Close-up view of a typical α-helix in the TG cryo-EM map (part of the core region). **d**, Close-up of a β-sheet in TG (part of the ChEL domain). **e**, Close-up of the disulfide bond C900–C921 (core region). **f**, Map details of N2013 and N-linked GlcNAc between two TG subunits. **g**, Close-up views showing conformational disorder within the hormonogenic sites (making precise side-chain placements difficult), but the backbone positions are resolved (rTG map).



Extended Data Fig. 6 | Quantitative thyroid-hormone ELISA assays. **a**, Schematic summarizing in vitro thyroid-hormone synthesis and quantification via ELISA assays. **b**, T_4 assay calibration curves with added T_4, under manufacturer-recommended and modified (T_4 synthesis as performed here) conditions. **c**, Validation of the T_4 ELISA assay. eTG presumably contains already-reacted tyrosine side chains. rTG produces T_4. Addition of iodide is required for the reaction to occur. Lactoperoxidase is as active as TPO, taking the reduced 20% haem content in our TPO into account. Lysozyme (some tyrosines), FtsZ from *S. aureus* (no tyrosines) and T_3 produce no T_4 signal. **d**, Mutating residues in hormonogenic site D in a version of TG that is active only in site D shows that a conserved lysine residue is not important for the reaction. Adding an extra Ser-Asp before Y1310 has no effect, but the mutation D1309S abolished activity. **e**, Synthesis of T_4 from tyrosine copolymers as measured by the T_4 ELISA assay. Only a polymer in which tyrosines are spaced apart and preceded by Lys-Asp produces some T_4. Activity is lower than in a single site of TG (or MBP, compare with Fig. 3). **f**, T_3 assay calibration curves with added T_3 under recommended and modified (as for T_4) conditions. **g**, No substantial T_3 production was detected from iodinated rTG or eTG from goitre. In **c**–**e**, **g**, bar plot and error bars indicate mean and s.d., $n = 3$.



Extended Data Fig. 7 | Tyrosine-pair proximity plots for TG and MBP. **a**, **b**, Proximity plots of tyrosine residues closer than 15 Å to each other, calculated from TG (**a**) and MBP (**b**) atomic models (TG, this study; MBP, PDB accession code 1ANF). The coordinates of each point in the plot represent a tyrosine-pair position (residue number). For the TG dimer, the distance between tyrosines from the same or the other subunit in the dimer are shown in grey or black, respectively. In TG, there are no more than five pairs that are exposed and in <15 Å proximity at the same time, predicting the absence of other sites important for hormonogenesis. In MBP, only one pair closer than 15 Å is sufficiently exposed to be a candidate for hormonogenesis. **c**, **d**, Ribbon diagram of TG and MBP in which tyrosine residues are represented as spheres and coloured by *B*-factor, which largely indicates solvent-exposed residues.

Extended Data Table 1 | TG cryo-EM and model statistics

Data collection and processing	eTG (C2 map)	rTG (expanded map)
Magnification	134,615x	105,000x
Voltage (kV)	300	300
Electron exposure ($e^-/Å^2$)	45	40
Defocus range (μm)	0.5-3.5	0.5-3.5
Pixel size (Å)	1.043	1.149
Symmetry imposed	C2	C1
Initial particle images	422,087	240,194
Final particle images	181,830	151,601
Map resolution (Å) – (FSC 0.143)	3.39	3.67
Map resolution range (Å)	2.8-7.0	3.3-5.0
Map sharpening B factor ($Å^2$)	-138.641	-113.164
Refinement		
Model resolution (Å) (FSC 0.5)	3.4 (3.8)	
PDB ID	6SCJ	
EMDB	EMD-10141	
Model composition		
Non-hydrogen atoms	39,886	
Protein residues	2,569	
RMS deviations		
Bond lengths (Å)	0.005	
Bond angles (°)	1.274	
Validation		
Molprobity score (percentile)	1.85 (83^{rd})	
Clashscore	4.88	
Poor rotamers (%)	0.14	
Ramachandran plot		
Favored (%)	86.69	
Allowed (%)	13.31	
Disallowed (%)	0	

Extended Data Table 2 | TG domain annotation

Region (residue range)	Region name	Domain	Domain type	Template for homology model (PDB ID)
31-92	NTD	A	type-1 TG repeat	1ICF.I
93-160	NTD	B	type-1 TG repeat	1ICF.I
161-297	NTD	C	type-1 TG repeat	1ICF.I
298-358	NTD	D	type-1 TG repeat	1ICF.I
359-620	NTD	E	helical	*de novo*
621-658	Core	F	type-1 TG repeat	1ICF.I
659-726	Core	G	type-1 TG repeat	1ICF.I
727-921	Core	H	type-1 TG repeat	1ICF.I
922-1008	Core	I	similar to type-1 (dimer)	*de novo*
1023-1073	Core	J	type-1 TG repeat	1ICF.I
1074-1145	Core	K	type-1 TG repeat	1ICF.I
1146-1211	Core	L	type-1 TG repeat	1ICF.I
1210-1283	Flap	M	Fv from an IgM	1IGM
1284-1438	Flap	N	Fv from an antibody	2E27
1439-1510	Arm	O	TNF/EGF/laminin-like	5wiw
1511-1565	Arm	P	type-1 TG repeat	1ICF.I
1603-1723	Arm	Q	type-3 TG repeat	U domain
1724-1892	Arm	R	type-3 TG repeat	U domain
1893-1995	Arm	S	type-3 TG repeat	U domain
1996-2129	Arm	T	type-3 TG repeat	U domain
2130-2186	Arm	U	type-3 TG repeat	*de novo*
2187-2728	CTD	V	ChEL (dimer)	5HQ3

Extended Data Table 3 | List of *N*-linked GlcNac in TG structure

Asn	GlcNAc	GlcNAc	Bman	
76	3617			
110	3620			Newly identified
198	3618			
484	3619			Newly identified
529				not visible in the map (flexible)
748				not visible in the map (flexible)
816				not visible in the map (flexible)
947	2733			
1220	2734			
1348				not present
1349	2735			
1365	2736			
1716	2737			
1774	2748			
1869	2747			Newly identified
2013	2740	2741	2742	
2122	2746			Newly identified
2250	2743			
2295	2744			
2582	2745			

nature research

Corresponding author(s): Jan Löwe
Dušan Turk

Last updated by author(s): 18/12/19

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	FEI TEM user interface, EPU, DigitalMicrograph (Gatan)
Data analysis	Python 3, Phenix 1.14rc2-3191, RELION 2.1, RELION 3.0, Coot (ccp4-7.0), MAIN 2019, Chimera 1.10 , Gctf 1.06, MotionCor 2.1 , XiSEARCH 1.6.743, XiFDR 1.1.26.58, SWISS MODEL 2019-05.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Experimental EM maps and model coordinates have been deposited in the Protein Data Bank (PDB) with accession codes EMD-10141, 6SCJ. The mass spectrometry proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD014821.
T4 concentrations measured with ELISA commercial kit and other data are available upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size were not predetermined: the amount of initially collected EM images (2 days automated data collection) was sufficient to achieve a reasonable resolution for rTG model building. Therefore a similarly sized dataset was collected for rTG.
Data exclusions	No data were excluded from the analysis apart badly picked or low resolution particles from EM data, as listed in the ms tables.
Replication	ELISA assays were performed in triplicate, as well as the expression and purification procedures. Preliminary small EM preliminary datasets were collected on thyroglobulin with reproducible outcomes and the largest datasets were used for final image processing.
Randomization	Randomization was not relevant for this study, since data were collected automatically and did not involve choosing.
Blinding	Blinding was not relevant to the study, as all the data were collected automatically and systematically

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Abcam, ab108661: Human Thyroxine ELISA Kit(Lot:GR3257329-1); Abcam, ab108664: Human Triiodothyronine ELISA Kit (Lot:GR3300688-1)
Validation	The antibodies were verified using purified T4 and T3 hormones, yielding concentration standard curves as expected.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Mammalian HEK 293T cells: ATCC no. CRL-3216, Insect cells: Invitrogen (Sf9)
Authentication	Used as expression strains only, independent verification after purchase not required. The over-expressed and purified TG was verified independently by MS.
Mycoplasma contamination	All cells were negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.