

# The structure of the teleost Immunoglobulin M core provides insights on polymeric antibody evolution, assembly, and function

Received: 15 March 2023

Mengfan Lyu<sup>1</sup>, Andrey G. Malyutin<sup>2,3,7</sup> & Beth M. Stadtmueller<sup>1,4,5,6</sup> 

Accepted: 4 November 2023

Published online: 21 November 2023

 Check for updates

Polymeric (p) immunoglobulins (Igs) serve broad functions during vertebrate immune responses. Typically, plgs contain between two and six Ig monomers, each with two antigen binding fragments and one fragment crystallization (Fc). In addition, many plgs assemble with a joining-chain (JC); however, the number of monomers and potential to include JC vary with species and heavy chain class. Here, we report the cryo-electron microscopy structure of IgM from a teleost (t) species, which does not encode JC. The structure reveals four tIgM Fcs linked through eight C-terminal tailpieces (Tps), which adopt a single  $\beta$ -sandwich-like domain (Tp assembly) located between two Fcs. Specifically, two of eight heavy chains fold uniquely, resulting in a structure distinct from mammalian IgM, which typically contains five IgM monomers, one JC and a centrally-located Tp assembly. Together with mutational analysis, structural data indicate that plgs have evolved a range of assembly mechanisms and structures, each likely to support unique antibody effector functions.

Polymeric (p) immunoglobulins (Igs) play critical roles in immune system function and are presumed to exist in all jawed vertebrates<sup>1</sup>. Despite the prevalence of plgs in vertebrate species, considerable differences in structure and function are thought to exist. Typically, plgs are comprised of two to six Ig monomers, each of which contains two heavy chains and two light chains that form one fragment crystallization (Fc) and two antigen-binding fragments (Fabs). The heavy chain includes three to four constant domains (CH1-CH4) and the Fc region is typically formed by the two C-terminal domains (e.g. CH3 and CH4). The Fc is critical for plg assembly and for binding to Fc receptors (FcRs) to elicit downstream functions<sup>2,3</sup>. Only a subset of Ig heavy chains are assembled into plgs, including IgA and IgM in mammals and most birds and reptiles, IgX and IgM in amphibians, IgT and IgM in teleost (t), or bony fish, and IgM in cartilaginous fish. To assemble plgs, plasma cells typically link multiple copies of heavy chains together with a protein called the joining chain (JC); however, incorporation of the JC is variable among species and heavy chain class<sup>4-6</sup>.

In mammals, polymeric forms of IgA are typically dimeric (d), containing two IgA monomers and one JC; however functional trimeric, tetrameric and pentameric plgA have been identified in lower abundance<sup>7-9</sup>. Polymeric forms of IgM are typically pentameric, containing five IgM monomers and one JC, however, hexamers lacking a JC have been identified in serum<sup>10-12</sup>. JC is necessary for plgA and pentameric IgM assembly and required for their secretion into the mucosa by the polymeric Ig receptor (plgR). The plgR is a transcytotic Fc receptor expressed on the basolateral surface of epithelial cells, which binds to JC-containing plgA and plgM and transports them to the apical surface of the cell<sup>13</sup>. On the apical surface the plgR ectodomain, called secretory component (SC), is proteolytically cleaved, releasing a complex containing SC and the plg, which is termed a secretory (S) Ig. Whereas plgM functions in serum and SIgM functions in the mucosa, most plgA is delivered to the mucosa and functions as SIgA<sup>14</sup>. The plgM, SIgM and SIgA exhibit unique capabilities compared to monomeric Ig, including high avidity antigen binding, antigen coating or

<sup>1</sup>Department of Biochemistry, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. <sup>2</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA. <sup>3</sup>Beckman Institute, California Institute of Technology, Pasadena, CA 91125, USA. <sup>4</sup>Department of Biomedical and Translational Sciences, Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. <sup>5</sup>Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. <sup>6</sup>Center for Biophysics and Quantitative Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. <sup>7</sup>Present address: Takeda Pharmaceuticals, Cambridge, MA 02139, USA.  e-mail: [bethms@illinois.edu](mailto:bethms@illinois.edu)

crosslinking, and binding to unique subsets of host and microbial FcRs. The functional outcomes of these interactions are diverse, ranging from complement activation (by IgM), pathogen agglutination and elimination, to commensal microbe colonization and poorly characterized FcR-dependent processes<sup>15,16</sup>.

While mammalian plgs have been studied extensively, less is known about plgs from other jawed vertebrates. Of particular interest are plgs from teleosts, which represent an early evolutionary stage of vertebrate adaptive immunity. Moreover, whereas the vast majority of jawed vertebrates, including in cartilaginous fish and amphibians, encode JC in their genomes, teleosts have lost the JC gene during evolution and express polymeric forms of tlgT and tlgM that are presumed to contain four copies of each respective monomer<sup>6,17,18</sup>. Whereas both tlgT and tlgM have been found in fish serum and mucus, tlgT is the predominant mucosal antibody, functionally similar to IgA in mammals<sup>18</sup>. Both tlgT and tlgM can be transported to the mucosa by tplgR, which has a distinct domain organization and structure compared to mammalian counterparts and does not require a JC to bind tlgT or tlgM<sup>19</sup>. The tSC has been isolated in complex with mucosal tlgT and tlgM, indicating that similar to mammalian SC, tSC remains bound to secretory forms of tlgT and tlgM and may play a role in mucosal immune functions<sup>18,20</sup>. Similar to mammalian plgs, tpls have been associated with complement activation and bacterial coating<sup>21</sup>.

The structural mechanisms underlying tpls' functions, and more broadly, the evolution of plg structure-function relationships across species, are poorly understood in-part because until now, teleost antibody structures remained unreported. Furthermore, only in the past several years have structures of mammalian plgs been reported. Cryo-electron microscopy (cryo-EM) structures of mammalian dIgA, SIgA, and SIgM revealed a central role for the JC, which folded together with C-terminal peptide sequences termed tailpiece (Tps) from the IgA or IgM heavy chains<sup>22–24</sup>. In these structures, both the Tps and the JC contribute  $\beta$ -strands to a  $\beta$ -sandwich near the center of the molecule, which stabilizes the assembly. Furthermore, the JC forms additional contacts with two Fcs, provides part of the plgR (and SC) binding site, and appears to contribute asymmetry to plg complexes that otherwise contain multiple heavy chains with identical sequence. Given the critical role for the JC in mammalian plg structures, we hypothesized that tpls lacking JC would assemble using different mechanisms and adopt distinct structures. To test this hypothesis and to explore the evolution and diversity of vertebrate plg structure and function, we determine the cryo-EM structure of the tlgM Fc core at 2.78 Å resolution. Together with comparative and mutational analysis, this structure provides a glimpse of both a teleost plg and a plg lacking a JC, revealing distinct modes of assembly compared to mammalian plgs and suggesting alternative mechanisms for interacting with antigens and FcRs, including tSC.

## Results

### Overall structure of tetrameric teleost IgM core

Guided by previous structural studies indicating that the human IgM constant domain CH2 and Fabs are flexible and poorly resolved in cryo-EM maps<sup>24,25</sup>, we expressed and purified a truncated teleost IgM heavy chain consisting of CH $\mu$ 3, CH $\mu$ 4, and the Tp (Fig. 1a, Supplementary Fig. 1). We utilized this tlgM Fc (tFc $\mu$ ) to determine a cryo-EM structure to an average resolution of 2.78 Å (Fig. 1, Supplementary Fig. 2). In the structure, most main chain and side chain atoms in CH $\mu$ 4 domains and the Tps were well resolved (resolution range 2.4–3.0 Å). Map resolution was  $\sim$ 3.2 Å and above for CH $\mu$ 3 allowing the positions of main chain atoms and  $\sim$ 50–75% of side chain atoms in each individual chain to be refined; other side chains were built in geometrically reasonable positions (Supplementary Fig. 2e).

The refined tFc $\mu$  structure contains four tFc $\mu$  monomers (Fc $\mu$ 1 to Fc $\mu$ 4) that lay approximately flat on one plane and include interfaces between Fc $\mu$ 1 and Fc $\mu$ 2, Fc $\mu$ 2 and Fc $\mu$ 3, Fc $\mu$ 3 and Fc $\mu$ 4. Fc $\mu$ 1 and Fc $\mu$ 4 are

separated by Tp residues, which form a  $\beta$ -sandwich (hereafter termed *Tp assembly*) comprising two  $\beta$ -strands from each of the four Fc $\mu$ s (Fig. 1). Overall, the tetramer is  $\sim$ 164 Å across in the largest dimension and can be described as two structurally and symmetry-related halves, one containing Fc $\mu$ 1-Fc $\mu$ 2 and associated Tps, and the other containing Fc $\mu$ 3-Fc $\mu$ 4 and associated Tps; at the center of the tetramer is a solvent accessible hole with a diameter of  $\sim$ 12.2 Å. Despite similarity between the two halves of the tetramer, the complex is not symmetrical with respect to all four Fcs; rather, adjacent tFc $\mu$ s are related by distinct angles measuring 81°, 72°, 86°, for Fc $\mu$ 1 and Fc $\mu$ 2, Fc $\mu$ 2 and Fc $\mu$ 3, Fc $\mu$ 3 and Fc $\mu$ 4, respectively. Fc $\mu$ 1 and Fc $\mu$ 4 are related by an angle of 121° and contact the Tp assembly (Fig. 1b, c).

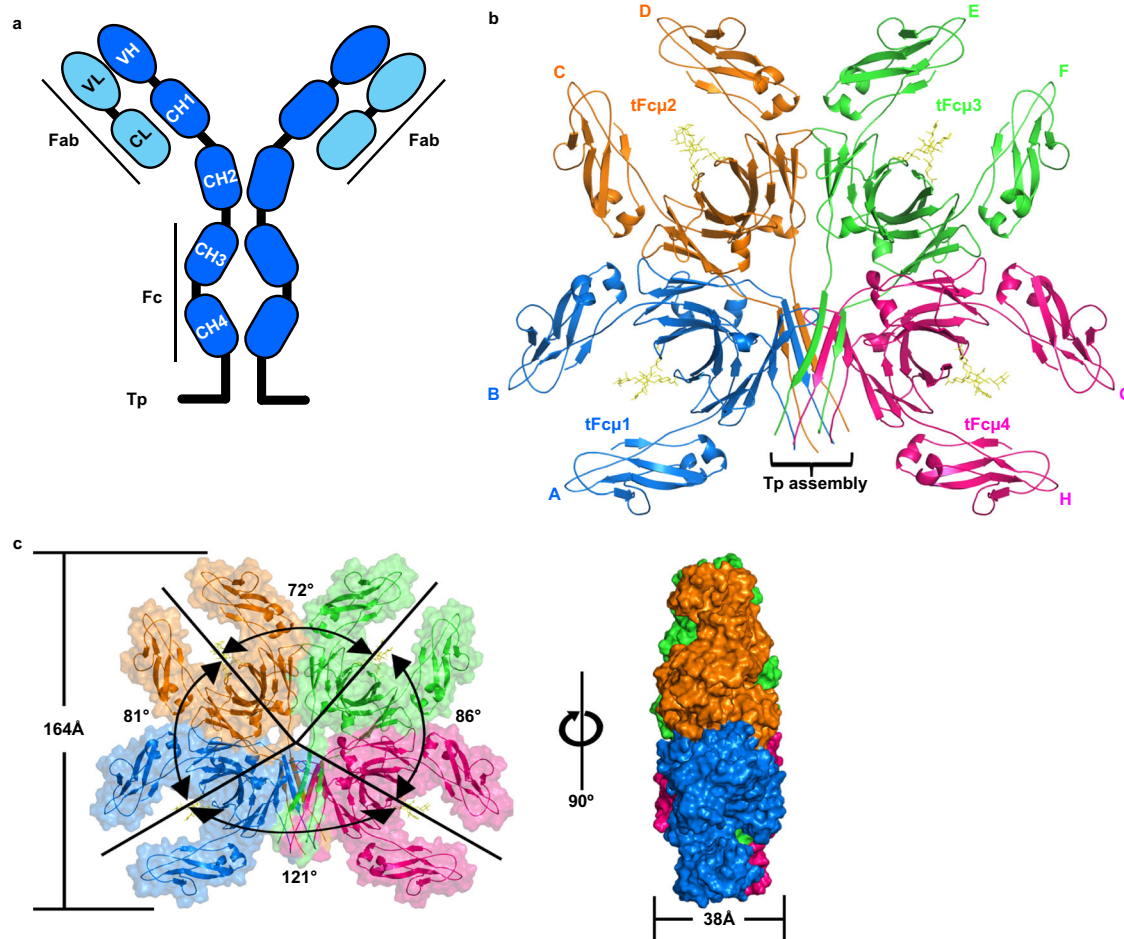
### tFc $\mu$ is formed from Fc monomers containing CH $\mu$ 4 domains that adopt one of two structural folds

Given the lack of published structural data for a teleost antibody, we first analyzed the structures of tFc $\mu$  monomers and their individual domains, CH $\mu$ 3 and CH $\mu$ 4. The tFc $\mu$  monomers comprise two copies of CH $\mu$ 3 and CH $\mu$ 4, which together form a canonical Fc characterized by an angle of  $\sim$ 87° between CH $\mu$ 3 and CH $\mu$ 4 on the same chain and interfaces between adjacent CH $\mu$ 3 and adjacent CH $\mu$ 4 domains on different chains (Fig. 2a, Supplementary Fig. 3a). Each domain adopts an Ig-constant fold consisting of two  $\beta$ -sheets linked by a disulfide bond, one sheet containing four  $\beta$ -strands termed A, B, E, and D, and the other sheet containing three  $\beta$ -strands termed C, F, and G; the Tp  $\beta$ -strand follows the CH $\mu$ 4 G strand. In one chain of each Fc $\mu$  monomer (A, D, E, and H) the A strand extends several residues, which we have termed A'. Ordered carbohydrates were apparent on CH $\mu$ 4 residue Asn 374. Within each tFc $\mu$ , two CH $\mu$ 4 domains interact primarily through residues located in the AA'BED  $\beta$ -sheets, which mediate the interactions between the C-terminal Ig domain in most Fcs<sup>22–24,26,27</sup>; the interface (including CH $\mu$ 3, CH $\mu$ 4 and Tp) is  $\sim$ 924 Å<sup>2</sup> and includes hydrogen bonds (Supplementary Fig. 3b, c).

Remarkably, despite sharing identical sequence and adopting Ig constant folds, the eight copies of CH $\mu$ 4 exhibit structural differences that stabilize the tFc $\mu$  tetramer. Specifically, the C-terminal 38 residues (residues 411–448; FG loop, G strand and Tp) adopt one of two folds, with chains A and H adopting an A-type fold and chains B–G adopting a B-type fold (Fig. 2b, c). In the A-type fold, the FG loop and the G strand include residues 411–416 and 417–426, respectively, and the Tp folds back to form  $\beta$ -sheet interactions with the G strand (Fig. 2b). In the B-type fold, the FG loop and the G strand instead include residues 411–420 and 421–426, respectively, resulting in Tp  $\beta$ -strand (residues 434–441) being directed away from CH $\mu$ 4 where they form  $\beta$ -sheet interactions with other Tp(s) rather than the G strand (Fig. 2c). Whereas chains A and H are superimposable, the six B-type domains adopt one of three distinct conformations based on the trajectory of the Tp and thus, each B-type domain is superimposable (RMSD of less than 0.18 Å) with one other B-type domain on the opposing side of the tetramer (Supplementary Fig. 4a). In sum, the A-type CH $\mu$ 4 domains can be described as having a C-F-G-Tp sheet whereas B-type have a C-F-G sheet (with distinct F-G  $\beta$ -sheet interactions) and an extended Tp. Together these two folds facilitate the incorporation of each chain into the Tp assembly (discussed below).

### Interactions between Fc $\mu$ s are mediated by CH $\mu$ 4-CH $\mu$ 4 and CH $\mu$ 4-Tp interfaces

The tFc $\mu$  tetramer includes contacts between adjacent Fc $\mu$  monomers and Tps. Interactions between adjacent Fc $\mu$  monomers (e.g., Fc $\mu$ 1-Fc $\mu$ 2, Fc $\mu$ 2-Fc $\mu$ 3, and Fc $\mu$ 3-Fc $\mu$ 4) are dominated by a common set of residues in the CH $\mu$ 4 CD loops, FG loops, and G-strands, forming an interface of  $\sim$ 1360 Å<sup>2</sup> (Fig. 3a). Notable residues mediating these contacts include CH $\mu$ 4 G strand residues Met422, Thr424, and Arg423, which form electrostatic interactions and hydrogen bonds with G strand residues, CD loop (Asp359), and F strand (Ser406) residues in the neighboring



**Fig. 1 | Overall structure of teleost IgM-Fc core, tFc $\mu$ .** **a** Diagram of tIgM monomer. Constant (C) domains are represented as round corner rectangles and variable (V) domains are represented as ellipses. The heavy (H) chain is colored marine blue and the light (L) chain is colored light blue. **b** Cryo-EM structure of tFc $\mu$  (cartoon representation); chain IDs (A–H) and Fc $\mu$ 1–4 and the Tp assembly are labeled. Glycans at residue Asn 374 are shown as yellow sticks. **c** Front view (left) and side view

(right) of tFc $\mu$  tetramer, showing both cartoon representation and molecular surface. Values describing the angles between tFc $\mu$  monomers as well as the approximate diameter and thickness of the tFc $\mu$  tetramer are indicated. In panels **b** and **c**, tFc $\mu$  monomers are colored as indicated: Blue, tFc $\mu$ 1. Orange, tFc $\mu$ 2. Green, tFc $\mu$ 3. Pink, tFc $\mu$ 4.

CH $\mu$ 4 domain. Additionally, FG loop residues Lys416 and Thr418 form interactions with CD loop Arg364 and F strand Tyr410, respectively (Fig. 3b, c).

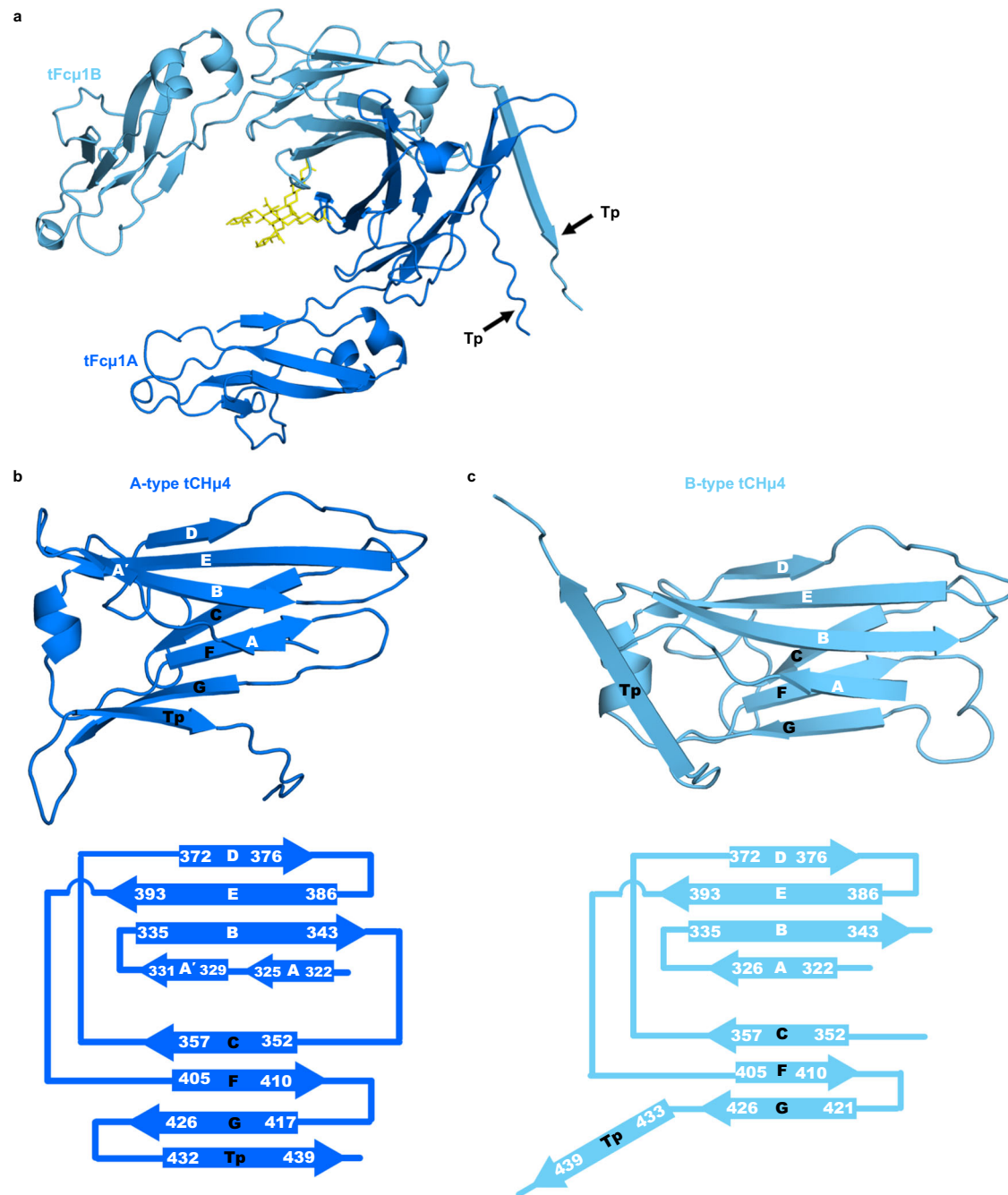
Fc $\mu$ 1 and Fc $\mu$ 4 do not interface and instead interact with the Tp assembly, which is situated between the two Fc $\mu$ s (Figs. 1b and 4a). Notably, Fc $\mu$ 1 and Fc $\mu$ 4 each contain one A-type CH $\mu$ 4 domain and one B-type CH $\mu$ 4 domain. The A-type domains neighbor the Tp assembly and contribute uniquely to its structure. Specifically, because A-type Tp  $\beta$ -strands form  $\beta$ -sheet interactions with CH $\mu$ 4 G strands, residues in the pre-Tp loop (residues 427–431) are positioned to contact both the top and bottom Tp  $\beta$ -sheets, effectively clamping the Fc $\mu$ 1 and Fc $\mu$ 4 to the Tp assembly. Interactions between the pre-Tp loops and the rest of the Tp assembly are dominated by polar contacts between A-type pre-Tp loop residues and B-type Tp residues (Fig. 4d–f). Together Fc–Fc and Fc–Tp interactions form a band of contacts around the entire tetramer, with A-type CH $\mu$ 4 domains contributing directly to the Tp assembly (Fig. 4d–f) and B-type CH $\mu$ 4 domains forming interchain interactions between adjacent Fc $\mu$  monomers using CD loops, FG loops and G-strands (Fig. 3b, c).

#### Tps from adjacent tFc $\mu$ monomers alternate to form two $\beta$ -sheets that stabilize the Tp assembly

Within the tFc $\mu$  tetramer,  $\beta$ -strand pairing in the Tp assembly tightly links individual chains and also promotes stabilizing interactions

between the two halves of the tetramer (Fc $\mu$ 1–Fc $\mu$ 2 and Fc $\mu$ 3–Fc $\mu$ 4). Specifically, we observe that Tp residues 434–440 form  $\beta$ -strands, which together constitute two 4-stranded  $\beta$ -sheets, each comprised of Tps from two adjacent tFc $\mu$  monomers (Fig. 4a). Notably, parallel Tp  $\beta$ -strand pairing alternates between Fc $\mu$ s such that one sheet comprises Tps from chains A–C–B–D and the other comprises Tps from chains E–G–F–H; in other words, two Tp strands from the same Fc $\mu$  are separated by one Tp strand from an adjacent Fc $\mu$  (Fig. 4a). The two  $\beta$ -sheets are stacked, forming two planes rotated at an angle of  $\sim 48^\circ$  relative to each other, and bury a hydrophobic core in-between. Buried residues include Val435, Leu437, and Leu439 from each chain, whereas hydrophilic residues Asn436, Ser438, and Asn440 face outwards (Fig. 4b, c).

Loop residues 441–448 follow the Tp  $\beta$ -strands. These residues are poorly ordered in the map; however, density suggests potential disulfide bonds between copies of the residue Cys445 (Supplementary Fig. 4b–f). Residue 445 is a conserved Cys involved in interchain disulfide bonding in mammalian IgA and IgM<sup>23,24</sup>, despite low local resolution, our map suggests possible disulfide bonding between Cys445<sup>Fc $\mu$ 1A</sup> and Cys445<sup>Fc $\mu$ 3E</sup>, Cys445<sup>Fc $\mu$ 1B</sup> and Cys445<sup>Fc $\mu$ 2C</sup>, Cys445<sup>Fc $\mu$ 2D</sup> and Cys445<sup>Fc $\mu$ 4H</sup>, and Cys445<sup>Fc $\mu$ 3F</sup> and Cys445<sup>Fc $\mu$ 4G</sup> (see Supplementary Fig. 4b–f). The possibility of inter-chain disulfide bonding is further supported by higher molecular weight bands visible on non-reducing SDS-PAGE upon tFc $\mu$  purification (Supplementary Fig. 1b). Because potential Cys445 disulfide bonds are located in the C-terminal loop region following the



**Fig. 2 | tlgM CH4 domains adopt one of two Ig folds.** **a** Cartoon representation of tFcμ1 (chains A and B) with glycans shown in yellow; the orientation is the same as Fig. 1b. **b, c** A-type (panel **b**) and B-type (panel **c**) tCHμ4 folds. *Top*, cartoon representations of A-type tCHμ4 domain and B-type tCHμ4 domain. *Bottom*,

topology diagram showing the secondary structure of tFcμ1A and tFcμ1B; terminal residues of each β-strand are numbered. B-type chains D and E contain A and A' strands (not shown in this figure).

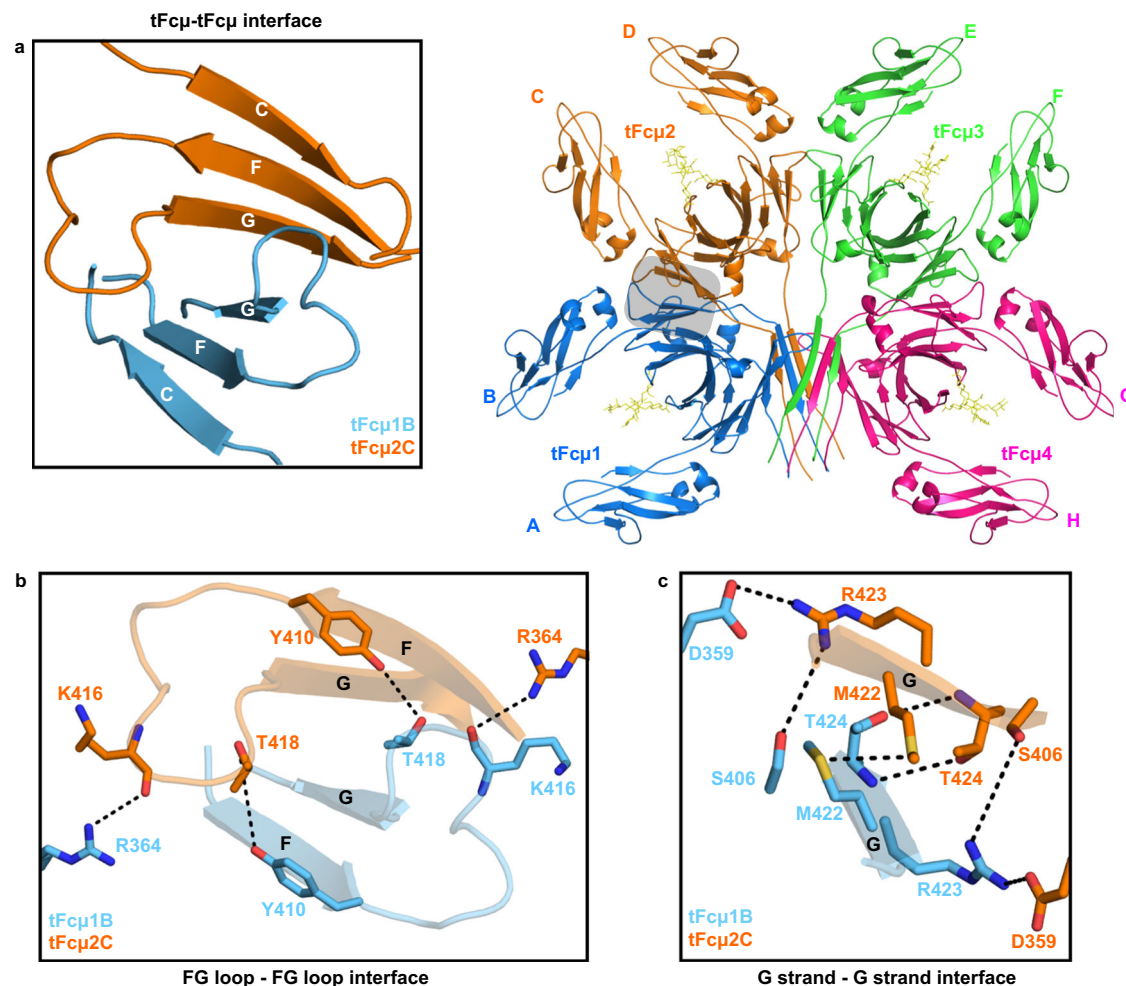
Tp β-strands, they do not appear to play a direct role in stabilizing the interactions between adjacent Fcs; nevertheless, they may effectively provide a covalent link between the two halves of the tetramer by connecting the top and bottom sheets of the Tp assembly (Fig. 4a and Supplementary Fig. 4b–f). Cysteine to serine mutation at this residue produced a tFcμ monomer in solution and eliminated evidence of inter-chain disulfide bonding, further suggesting a critical role in tlgM polymeric assembly (Supplementary Fig. 4g).

### Comparisons between teleost and human IgM

IgM is considered to be one of the evolutionarily oldest Igs<sup>1</sup>. It is omnipresent in all jawed vertebrates, except coelocanths where IgM

has been lost, and teleost fish are the second oldest class expressing IgM, after cartilaginous fish<sup>1,28</sup>. To gain insights on plg evolution we compared teleost and human (h) IgM structures<sup>24,25</sup>. Whereas tlgM is a tetramer, hIgM forms a JC-associated pentamer. Available human IgM structures contain the Fc region (hFcμ), JC (hJC), and human secretory component (hSC) (Fig. 5a). Overall, tFcμ and hFcμ adopt structures characterized by adjacent Fc monomers linked through Fcμ–Fcμ contacts and a Tp assembly; however, the structures are not globally superimposable and the relative orientation of the Fcμs, residues involved in Fcμ–Fcμ contacts, and the location and structure of the Tp assembly differ markedly, indicating that mechanisms of plg assembly and function have evolved from teleost to humans.





**Fig. 3 | The CH $\mu$ 4 interface between adjacent tFc $\mu$  monomers. a** Overview of interactions between tFc $\mu$ 1B and tFc $\mu$ 2C at the CH $\mu$ 4 interface (shaded area in tFc $\mu$  tetramer). *Left*, open-book overview of tFc $\mu$ -tFc $\mu$  interface boxed in the right panel, but shown in a different orientation. Only CH $\mu$ 4 C-F-G  $\beta$ -sheets and FG loops are shown. *Right*, tFc $\mu$  tetramer with tFc $\mu$ 1B and tFc $\mu$ 2C at the CH $\mu$ 4 interface in shaded box. **b** Hydrogen bonding mediated by FG loops between tFc $\mu$ 1B and tFc $\mu$ 2C at the

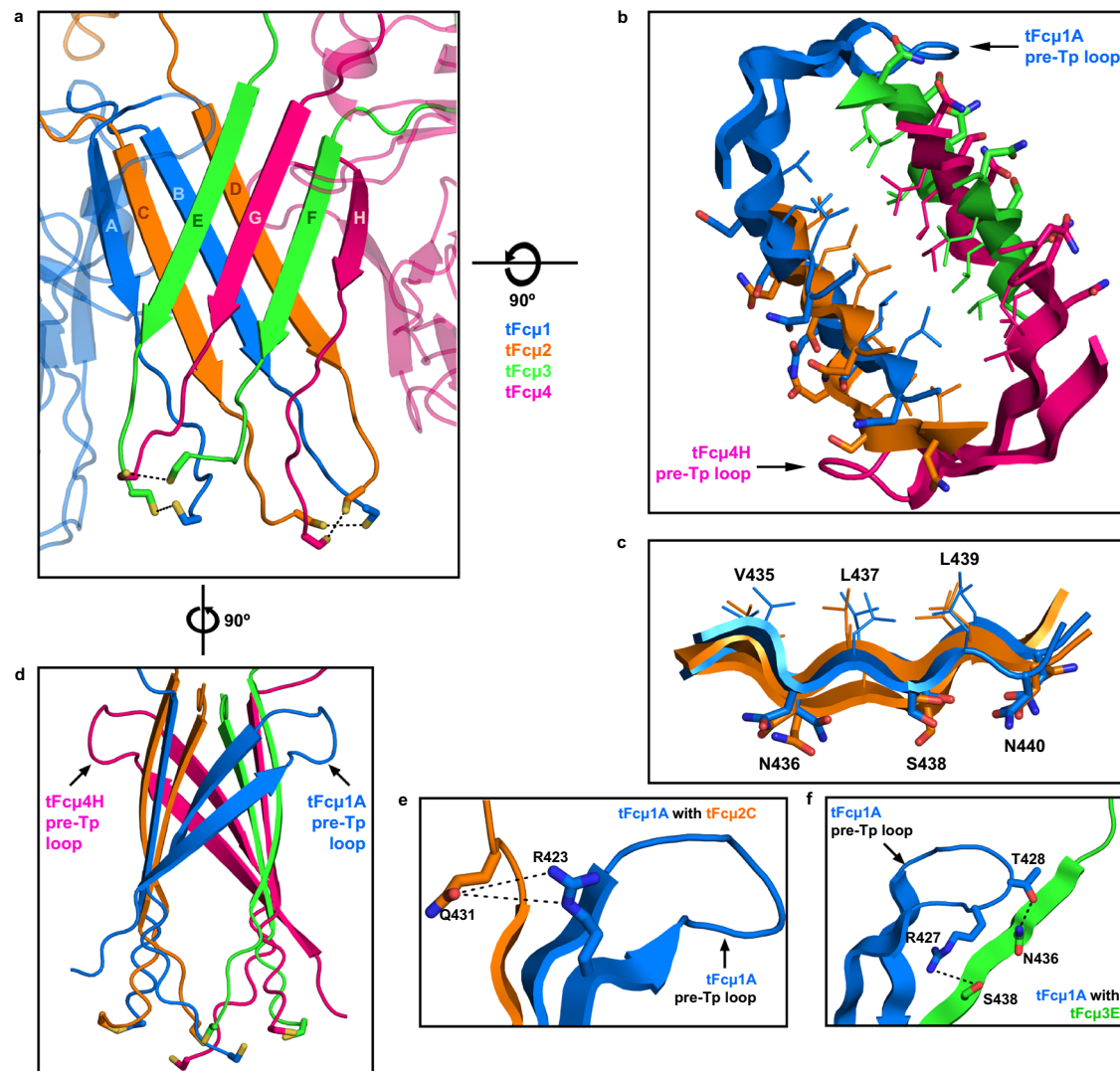
CH $\mu$ 4 interface. Cartoon representations are shown for the F strand, G strand, and FG loop. Residues participating in interactions are shown as sticks. **c** Electrostatic interactions and hydrogen bonding mediated by G strands between tFc $\mu$ 1B and tFc $\mu$ 2C at the CH $\mu$ 4 interface. Cartoon representations of the G strands are shown, and interfacing residues are shown as sticks.

A fundamental difference between tFc $\mu$  and hFc $\mu$  is the geometric relationship among adjacent Fc $\mu$  monomers. For example, adjacent tFc $\mu$ s are related by angles of 80–85° and -120° (non-adjacent tFc $\mu$ 1 and tFc $\mu$ 4 separated by the Tp assembly) while the five adjacent hFc $\mu$ s are related by angles between -53° and -62° (non-adjacent hFc $\mu$ 1 and hFc $\mu$ 5 separated by hJC) (Figs. 1b and 5a). The inter-Fc contacts that stabilize these geometries in both tFc $\mu$  and hFc $\mu$  include CH $\mu$ 4 C-F-G  $\beta$ -sheets and FG loops; however, participating residues are not conserved, resulting in an interface of -400Å<sup>2</sup> between hCH $\mu$ 4s and a larger interface of -800Å<sup>2</sup> between tCH $\mu$ 4s. In addition, adjacent hFc $\mu$  monomer CH3 domains form contacts and are covalently linked by disulfide bonds between Cys414 residues<sup>24,25</sup> whereas this cysteine residue is not conserved in tFc $\mu$  and thus does not form inter-Fc disulfide bonds between tFc $\mu$  CH3 domains.

Structural comparisons between CH4 domains and Tps from human and teleost IgM are especially relevant to understanding plg evolution and the functional differences between species because these domains form the majority of contacts that stabilize polymeric forms of IgM. Human and teleost IgM heavy chain constant domains share 28% sequence identity and 33% similarity among CH4 domains (Fig. 5b). A structural alignment between the human (chain B, excluding Tp residues) and teleost CH $\mu$ 4 domains revealed a RMSD of 1.386 Å when aligned to A-type domains, and RMSD of 1.047 Å when

aligned to B-type domains, suggesting that B-type domains share greater structural similarity to human IgM CH $\mu$ 4. Structural variability in the main chain arises primarily from differences in the length and conformations of the CD and DE loops (Fig. 5c), which contain residues that mediate contacts within and between Fc $\mu$ s and are not well conserved from teleost to mammalian IgM. Even in regions that share high structural similarity, such as the A-B-D-E  $\beta$ -sheet, some residues mediating contacts are not conserved, including those that stabilize the CH4-CH4 interface within each Fc (Fig. 5b).

The Tp has long been implicated in plg assembly; it is presumed to be present in all plgs<sup>2,3</sup> and is conserved among vertebrate plg sequences, although residues near the C-terminus are variable, including the position of a conserved cysteine relative to the C-terminus (Supplementary Fig. 5). Mammalian IgM and IgA structures revealed Tps folding together with the JC to form  $\beta$  sandwich-like domains near the center of the complexes<sup>22–24</sup>. In human SIgM, residues 562–568 of each hFc $\mu$  Tp form a  $\beta$ -strand, and 10 copies form two parallel  $\beta$ -sheets that lie near-perpendicular to the central axis of the pentamer, positioning the distal  $\beta$ -strand from each sheet to pair with one  $\beta$ -strand from JC (Fig. 5a). In contrast, the tFc $\mu$  Tp assembly lacks a JC and is located toward the side of the complex between Fc $\mu$ 1 and Fc $\mu$ 4 with  $\beta$ -sheets positioned near-parallel to the central axis (Fig. 1b). In both structures, conserved hydrophobic residues Leu437 and



**Fig. 4 | Teleost IgM Tp assembly and its interactions with pre-Tp loop of A-type CH $\mu$ 4.** **a** Cartoon representation of the tFc $\mu$  Tp assembly. The top  $\beta$ -sheet consists of Tp<sup>tFc $\mu$ 3E</sup>, Tp<sup>tFc $\mu$ 4C</sup>, Tp<sup>tFc $\mu$ 3F</sup>, and Tp<sup>tFc $\mu$ 4H</sup> and bottom  $\beta$ -sheet consists of Tp<sup>tFc $\mu$ 1A</sup>, Tp<sup>tFc $\mu$ 2C</sup>, Tp<sup>tFc $\mu$ 4B</sup>, and Tp<sup>tFc $\mu$ 2D</sup>. Cys445 side chains are shown as sticks. Cys445 pairs that may form potential disulfides are linked using dashed lines. **b** Closeup view of the tFc $\mu$  Tp assembly hydrophobic core, which is depicted from the Tp C-termini looking toward the center of the tFc $\mu$  tetramer. **c** Closeup view of the tFc $\mu$

Tp assembly top  $\beta$ -sheet comprising  $\beta$ -strands from chains A-D. The side chains of hydrophobic residues are shown as lines while side chains of hydrophilic residues are shown as sticks. **d** Sideview of the tFc $\mu$  Tp assembly. **e** The tFc $\mu$ 1A pre-Tp loop residues Arg423 interacts with Tp<sup>tFc $\mu$ 2C</sup> residues Gln431; only chain A and chain C are shown. **f** The tFc $\mu$ 1A pre-Tp loop residues Arg427 and Thr428 interact with Tp<sup>tFc $\mu$ 3E</sup> residues Asn436 and Ser438; only chain A and chain E are shown.

Leu439 (Val564 and Leu566 in hFc $\mu$ ) face the core of the  $\beta$ -sandwich and conserved hydrophilic residues Asn436 and Ser438 (Asn563 and Ser565 in hFc $\mu$ ) face outward (Fig. 4b, c). However, despite the observation that conserved residues mediate hydrophobic interactions at the core of both hFc $\mu$  and tFc $\mu$  Tp assemblies, the relationship between the two Tp assembly  $\beta$ -sheets differs; the two tFc $\mu$   $\beta$ -sheets are related by an angle of 48° whereas the two hFc $\mu$   $\beta$ -sheets are related by an angle of 158° (Fig. 4a and Fig. 5a).

The most striking difference between tFc $\mu$  and hFc $\mu$  Tp assemblies is a difference in  $\beta$ -strand pairing. In tFc $\mu$ , each Tp pairs with a Tp originating from an adjacent Fc monomer while in hFc $\mu$ , Tp  $\beta$ -strands pair with a second Tp from the same hFc $\mu$  monomer. The exception to this is hFc $\mu$ 3, which contributes one Tp to each  $\beta$ -sheet in the hFc $\mu$  Tp assembly (Figs. 4a and 5a). The hFc $\mu$  Tp strand-pairing pattern is also observed for Tps in polymeric forms of IgA<sup>23</sup>. Despite differences in strand pairing, the length from the last residue of CH $\mu$ 4 (Asp426 in tIgM; Asp 553 in hIgM) to the first residue in the Fc $\mu$ Tp  $\beta$ -strand (Leu434 in tIgM; Tyr562 in hIgM) is conserved between the tFc $\mu$  B-type

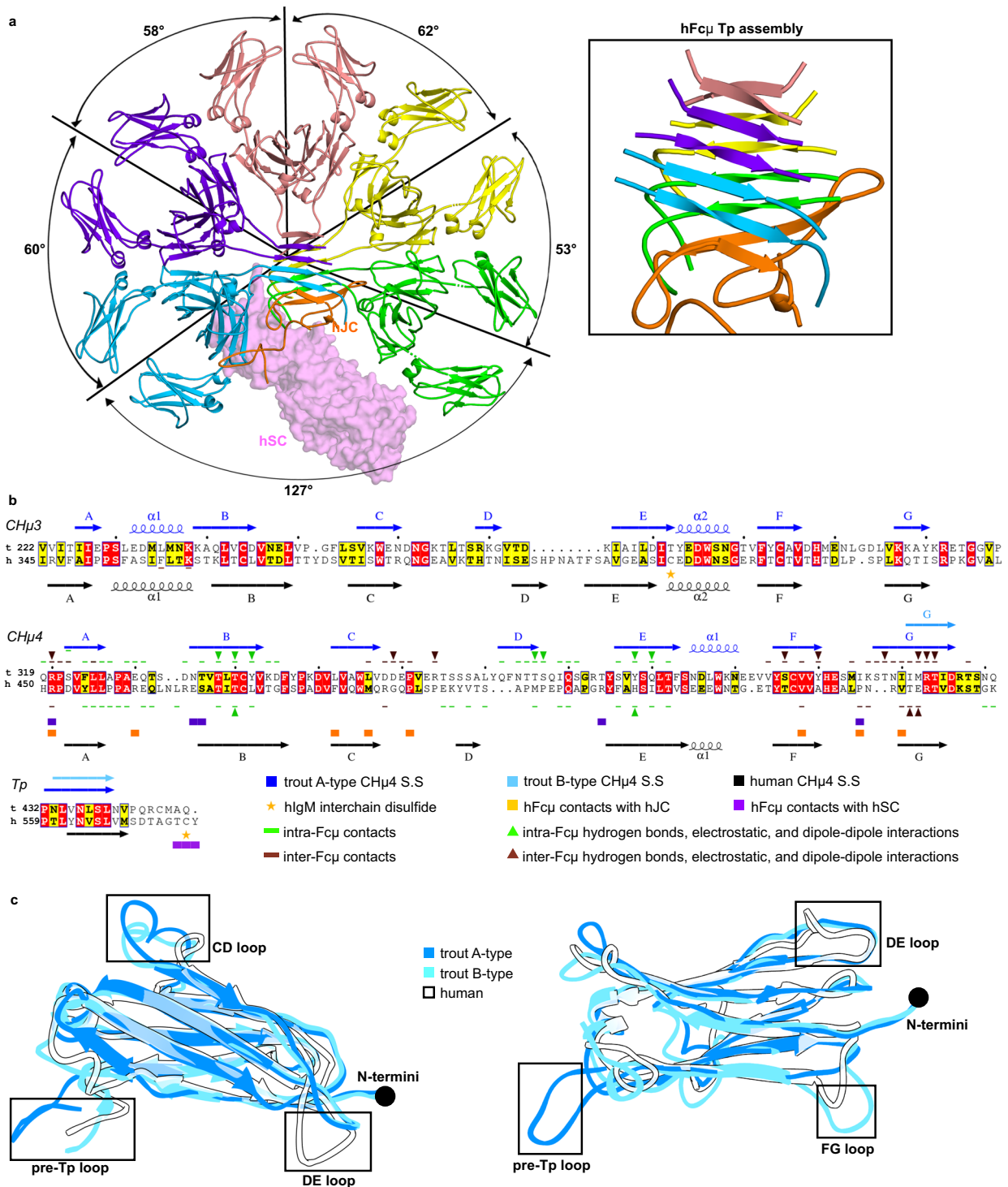
fold and hFc $\mu$ . Also conserved among hFc $\mu$  and tFc $\mu$  is a potential N-linked glycosylation site (PNGS) including residues Asn436 and Ser438. Earlier studies indicated that glycosylation at this site can influence hIgM polymerization and may prevent aggregation<sup>24,29,30</sup>. However, we did not observe ordered glycosylation at this site in the tFc $\mu$  structure (Supplementary Fig. 4h).

Finally, both tFc $\mu$  and hFc $\mu$  contain a conserved cysteine residue in the C-terminal loop following the Tp  $\beta$ -strand. While only four out of ten copies of Cys575 are ordered in published hFc $\mu$  structures<sup>24</sup>, and density at this conserved cysteine (Cys445) is poor in the tFc $\mu$  structure (Supplementary Fig. 4b–f), data point toward an essential role for a cysteine near the C-terminus of IgMs. As noted, we observe putative disulfide bond formation linking tFc $\mu$  Cys445<sup>Fc $\mu$ 1A</sup> with Cys445<sup>Fc $\mu$ 3E</sup> and Cys445<sup>Fc $\mu$ 2D</sup> with Cys445<sup>Fc $\mu$ 4H</sup> and mutation of this residue inhibits tetramer formation (Supplementary Fig. 4g). In the case of hFc $\mu$ , this cysteine (Cys575) forms disulfides with the JC or with another Tp located in the same sheet and is reportedly necessary for hIgM assembly<sup>24,25</sup>. The pattern of

disulfide bonding appears to be different and to link different regions in tIgM versus hIgM. In sum, structural analysis of tFc $\mu$  and hFc $\mu$  reveals that despite conservation in sequence and length, IgM Tps can assemble into markedly different structures to promote assembly of plgs.

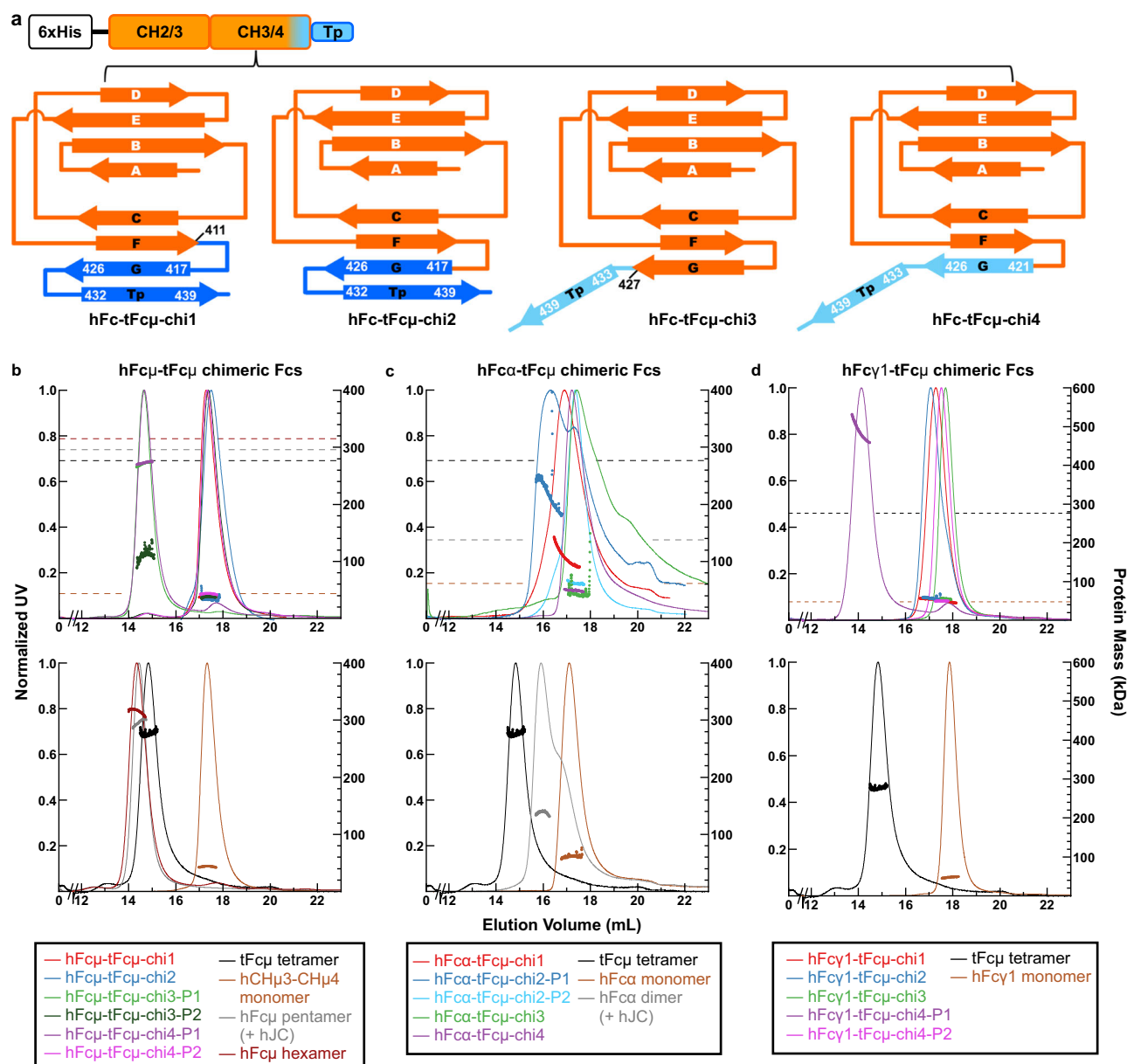
### tFc $\mu$ motifs introduced into human heavy chains can promote the assembly of plgs

Armed with additional structural insights and a desire to better understand plg evolution and assembly mechanisms, we explored the ability of tFc $\mu$  motifs to facilitate the assembly of plgs from other heavy



**Fig. 5 | Comparison of IgM structures from trout and humans. a** Human SFc $\mu$  structure, containing five hFc $\mu$  monomers, hJc, and hSc (PDB ID: 6KXS<sup>24</sup>); hFc $\mu$  and hJc are shown as cartoon representations and hSc is shown as molecular surface representation. Angles between Fc $\mu$  monomers are indicated in the structure. Inset, overview of hFc $\mu$  Tp assembly. Residues 560–576 of hFc $\mu$  heavy chain and part of hJc are shown and colored as in panel a. **b** Sequence alignment of IgM CH $\mu$ 3, CH $\mu$ 4, and Tp from *Oncorhynchus mykiss* (rainbow trout, t) and *Homo sapiens* (human, h),

with secondary structure (S.S) and interchain contacts shown at top (tFc $\mu$ ) and bottom (hFc $\mu$ ) of the sequences. For B-type tCH $\mu$ 4, secondary structure features identical to A-type tCH $\mu$ 4 are not indicated. In all panels, figures are annotated and colored according to the key. **c** Structural alignment of hFc $\mu$  CH $\mu$ 4 domain with tFc $\mu$  A-type CH $\mu$ 4 domain (chain A) and tFc $\mu$  B-type CH $\mu$ 4 domain (chain B). All domains are shown as cartoon representations.



**Fig. 6 | Design and characterization of chimeric Fcs.** **a** *Top*, the design of chimeric Fc constructs; *bottom*, schematic showing topology of chimeric Fc constructs. Each chimeric Fc consists of an N-terminal hexa-histidine tag, partial hFc sequence (orange), and tFc $\mu$  C-terminal sequence (marine or light blue). **b–d** SEC-MALS chromatograms for the indicated chimeric Fcs (top) and associated control complexes (bottom). SEC elution profiles are shown as solid curves (left axis; normalized UV signal) and light scattering data, indicating protein mass, is shown as horizontal dots (right axis; protein mass in kDa). All data are colored according to

the key below. In top panels, average protein mass of controls is indicated by horizontal dashed lines. **b** hFc $\mu$ -tFc $\mu$  chimeric Fcs and relevant controls. **c** hFc $\alpha$ -tFc $\mu$  chimeric Fcs and relevant controls. **d** hFc $\gamma$ 1-tFc $\mu$  chimeric Fcs and relevant controls. For all samples, the predicted molecular weight from the protein sequence, the average molecular weight determined by MALS, and the number and potential of PNGS are listed in Supplementary Table 4. The amino acid sequences of chimeric Fc used in this study are provided in Supplementary Table 5. Source data are provided as a Source Data file.

chains including human  $\mu$  (IgM),  $\alpha$ 1 (IgA1) and  $\gamma$ 1 (IgG1). The tFc $\mu$  structure suggests that both A-type and B-type tCH $\mu$ 4 domains are essential for tFc $\mu$  assembly. Given that tFc $\mu$  A-type and B-type tCH $\mu$ 4 domains, despite having the same sequence, differ structurally in the FG loop, G strand, and pre-Tp loop regions, we hypothesized that the mechanisms of tFc $\mu$  assembly rely heavily on residues in these motifs. To test this hypothesis, we designed twelve chimeric expression constructs, in which residues found in the A-type or B-type tCH $\mu$ 4 G strand and/or Tp replaced endogenous counterparts in human Fc $\mu$ , Fc $\alpha$ 1, and Fc $\gamma$ 1 (Fig. 6a). Chimeric Fc constructs and associated controls were transiently transfected into human cells and the resulting protein complexes were purified and their polymeric

state(s) evaluated using size exclusion elution chromatography with in-line multi-angle light scattering (SEC-MALS), which reports on the absolute molecular weight of target proteins and complexes<sup>31</sup>.

As noted, human IgM is naturally expressed as a hexamer or a JC-containing pentamer, and the Tp is reportedly essential for the assembly of both complexes<sup>30</sup>. Consistent with published studies, in our experimental system the removal of Tp residues resulted in only monomeric hFc $\mu$  (Fig. 6b and Supplementary Fig. 6a). To determine if the tFc $\mu$  Tp could promote plg assembly in hFc $\mu$  constructs lacking the endogenous Tp (and the JC), we measured the molecular weight of four chimeric proteins, hFc $\mu$ -tFc $\mu$ -chi1 to chi4 (Fig. 6b and Supplementary Fig. 6a). The hFc $\mu$ -tFc $\mu$ -chi1 and hFc $\mu$ -tFc $\mu$ -chi2 were



monomeric; however, hFc $\mu$ -tFc $\mu$ -chi3 and hFc $\mu$ -tFc $\mu$ -chi4 were polymeric with average molecular weights consistent with tetrameric complexes (Fig. 6b and Supplementary Table 4). The hFc $\mu$ -tFc $\mu$ -chi3 contains the tFc $\mu$  Tp and hFc $\mu$ -tFc $\mu$ -chi4 contains the B-type tCH $\mu$ 4 G strand and tFc $\mu$  Tp residues; thus, data indicate that substitution of hFc $\mu$  residues with these motifs, but not the A-type tCH $\mu$ 4 G strand residues linked to the tFc $\mu$  Tp, can promote plg assembly. However, in contrast to the hexamers observed when full-length hFc $\mu$  is expressed without JC, chimeric plgs adopted probable tetramers.

We also sought to determine if the tCH $\mu$ 4 G strand and/or Tp residues could promote plg assembly in the context of hFc $\alpha$ , which is monomeric in the absence of the JC and typically dimeric when expressed with a JC, or in the context of Fc $\gamma$ 1, which is naturally monomeric and does not contain a Tp. In contrast to hFc $\mu$ -tFc $\mu$  chimeras, we found that hFc $\alpha$ -tFc $\mu$ -chi1 and hFc $\alpha$ -tFc $\mu$ -chi2 formed plgs whereas hFc $\alpha$ -tFc $\mu$ -chi3 and hFc $\alpha$ -tFc $\mu$ -chi4 did not. The molecular weight of resulting plgs was consistent with a mix of species ranging from dimers to tetramers (Fig. 6c and Supplementary Fig. 6b). Data signify that in the context of the human alpha-1 heavy chain, the G strand and Tp residues found in A-type tCH $\mu$ 4 fold can promote plg assembly in the absence of JC, whereas G strand and Tp residues found in the B-type tCH $\mu$ 4 fold, or Tp residues alone, cannot. Among hFc $\gamma$ 1-tFc $\mu$  chimeras, only hFc $\gamma$ 1-tFc $\mu$ -chi4 promoted plg assembly, suggesting that in the context of the human gamma-1 heavy chain, substituting G strand and Tp residues found in the B-type tCH $\mu$ 4 fold could promote assembly whereas others could not. For all chimeric Fc constructs that promoted plg assembly, plgs were isolated from mixtures containing mIgs before SEC-MALS analysis (Supplementary Fig. 6). Among hFc $\mu$ -tFc $\mu$  chimeras, the plgs represented more than 50% of the total protein purified whereas plgs isolated from hFc $\alpha$ -tFc $\mu$  and hFc $\gamma$ 1-tFc $\mu$  chimeras represented less than 30% of the total protein purified, with mIgs making up the rest of the total (Supplementary Fig. 6). This suggests that in our experimental system, plg assembly promoted by tFc $\mu$  motifs was most efficient for hFc $\mu$ -tFc $\mu$  chimeras.

## Discussion

Throughout evolution, plgs have evolved critical, yet incompletely characterized functions in jawed vertebrates. Data reported here, together with recently published mammalian plg structures, demonstrate that plgs are a structurally diverse group of antibodies varying in composition, size, and conformation. For the host, assembly of these complexes is a critical first step to implementing a vast array of downstream effector functions and we find evidence that across species, this can be accomplished through several distinct yet overlapping mechanisms. Specifically, the Ig domains and Tps provide a common set of building blocks, while differences in heavy chain amino acid sequence and the presence or absence of JC are variables adding plasticity to the process.

As noted, most jawed vertebrates, including mammals, birds, amphibians, reptiles, and cartilaginous fish, reportedly encode JC and can incorporate it into plgM, whereas teleost fish have lost JC and assemble the Fc regions of plgs using only Ig domains and Tps<sup>4–6</sup>. We find that this unique teleost characteristic is associated with distinct tFc $\mu$  structural folding patterns when compared to mammalian plgM structures containing the JC<sup>24,25</sup>. This raises the possibility that JC-independent, and JC-dependent assembly evolved to employ distinct biophysical mechanisms. However, we also find that tFc $\mu$  C-terminal residues (and presumably associated structural motifs) are able to promote JC-independent plg assembly of otherwise human heavy chains, suggesting some degree of interchangeability among the sequences of plg-forming heavy chains and assembly mechanisms. For example, chimeric hFc $\mu$ -tFc $\mu$  revealed that a subset of tFc $\mu$  motifs were able to promote plg assembly of an otherwise human IgM heavy chain in the absence of JC. It is intriguing that this effect was found only when the tFc $\mu$  CH4 G strand and/or Tp residues found in the B-type fold, but

not the A-type fold, were used. While this may reflect the higher degree of similarity shared between the B-type fold and human IgM domains, the resulting plgs appeared to be tetramers rather than hexamers, the predominant polymeric state associated with human IgM lacking the JC in our experiments (and also in human serum)<sup>10</sup>. It remains unknown whether hFc $\mu$ -tFc $\mu$  chimera tetramers assemble into structures more similar to tIgM or hIgM with respect to the Tp assembly structure, contacts, and geometric relationships between the incorporated Fcs. While plg assembly of hFc $\alpha$ -tFc $\mu$ , and hFc $\gamma$ 1-tFc $\mu$  chimeras was notably less efficient than hFc $\mu$ -tFc $\mu$  chimeras, it is significant that tFc $\mu$  components could promote some plg assembly in the context of human IgA1 and IgG1 heavy chains, neither of which naturally form plgs in the absence of JC. While we cannot rule out the possibility that assembly in plasma cells differs from the HEK cells used in our experiments, experimental cell lines assembled controls of expected MW and provided a uniform system to evaluate how sequence differences among chimeras influences plg assembly.

Despite the prevalence of JC in the genomes of most jawed vertebrates (other than teleosts), some species assemble plgs in the absence of JC. For example, human IgM can assemble without JC (typically hexameric) and amphibian IgX (polymeric) also appears to lack the JC<sup>5,6,10–12</sup>. This raises the question of whether or not plgs lacking the JC from species other than teleost fish have structures most similar to hexameric hIgM or to tetrameric tIgM. At this time, a high-resolution structure detailing the Tp assembly of hexameric IgM has not been reported; however, negative-stain EM and small-angle light scattering (SAXS) data suggest all six hFc $\mu$ s form Fc-Fc contacts, and that the Tp assembly remains in the center of the complex as seen in pentameric IgM, rather than bridging two Fcs as we observed in tetrameric tFc $\mu$ <sup>32,33</sup>. The structures of unliganded Tps (e.g., those found in monomeric IgA or IgM) are also unknown and therefore it remains unclear if and how Tps might change conformation during mammalian plg assembly in absence of JC and/or if they adopt conformations distinct from what is observed in mammalian plg and Slg structures. Without such information, it is difficult to judge whether the Tp  $\beta$ -strands would alternate in hIgM hexamers, as observed in tFc $\mu$  tetramers, or if they would pair with another Tp  $\beta$ -strand from the same Fc as observed in hFc $\mu$  pentamer and mammalian plgA.

While our structural data provide considerable insight on tIgM structure and assembly mechanisms, the stages of the assembly process remain unclear. Identification of two distinct tCH $\mu$ 4 folds evokes the question of how tIgM heavy chains fold and if they adopt A- and/or B-type CH $\mu$ 4 domains prior to the assembly of plgM. In isolation, the A-type fold appears likely to be more stable than the B-type fold because the  $\beta$ -strand region of the Tp forms  $\beta$ -sheet interactions with the G strand, rather than being extended away from the core of the Ig domain in a potentially flexible location (see Fig. 2b, c). While further study is needed to define the stages of tFc $\mu$  assembly, one possibility is that heavy chains first adopt the A-type conformation and that polymeric IgM assembly is triggered when two tFc $\mu$  monomers come into proximity and interactions at the Fc-Fc interface displace Tp  $\beta$ -strands from both Fcs, effectively kicking them out of the Ig-fold and allowing them to form parallel  $\beta$ -sheet interactions with each other. It is unclear if there is sufficient energy in the tFc $\mu$ -tFc $\mu$  interface for its formation to break Tp-G strand hydrogen bonding and thus if this model is correct, it is conceivable that molecular chaperones, which have been implicated in mammalian plgA and plgM assembly processes, are involved<sup>34,35</sup>. Regardless, sequential interactions between Fcs could add strands to the Tp assembly until the tetramer was closed by the last remaining A-type fold. In a related model, sequential addition of heavy chains might culminate in dimers (tetramer halves), two of which subsequently come together to form a tetramer. These models are consistent with the alternating Tp  $\beta$ -strand pairing that we observe in tFc $\mu$ , but that is not seen in mammalian IgM or IgA containing JC<sup>22–24</sup>. Furthermore, given that Cys445Ser mutation abolished tFc $\mu$  tetramer

formation (Supplementary Fig. 4g), it is possible that disulfide bonds linking non-adjacent Fcs, and/or the top and bottom  $\beta$ -sheets in the Tp assembly (see Fig. 4a, Supplementary Fig. 4b–f) stabilize some part of this process. A critical role for Cys445 is consistent with published work indicating that disulfide crosslinking within tIgM influences antibody assembly, stability, and half-life<sup>17,36,37</sup>; however additional studies will be needed to work out its precise role(s) in these processes.

Another remarkable finding of our work is that the majority of plg assembly mechanisms appear likely to culminate in the release of an asymmetric plg complex, despite being assembled from multiple copies of identical Ig-heavy chains with identical sequence. Asymmetry in mammalian plgM and plgA was verified with the publication of plgA and plgM structures, in which the JC was bound to two Igs uniquely, and is thus postulated to template the addition of monomers during assembly and to confer asymmetry on the resulting complex<sup>22–24</sup>. However, the tFc $\mu$  structure, which lacks a JC, also reveals elements of asymmetry, with four Fc monomers directed away from the Tp assembly such that the relative geometric relationships between each tFc $\mu$  are not identical. As noted, this arrangement appears to be dependent upon the existence of A-type and B-type CH $\mu$ 4 domains, yet it is unclear if asymmetry in plgs evolved as an efficient means to assemble a plg and/or if asymmetry provides a functional advantage.

IgM is a critical Ig heavy chain class found in all jawed vertebrates. Studies in mammals demonstrate that its effector functions are mediated by high avidity interactions with antigen (from five sets of Fabs), complement activation, binding to IgM-specific Fc receptors (FcRs), and mucosal functions associated with plgR-dependent secretion. While relatively few reports have documented its function in teleost fish, evidence suggests that tIgM also participates in complement activation, mucosal functions such as bacterial coating and immune exclusion, and binds IgM-specific FcRs<sup>21,36,38,39</sup>. The tetrameric state of tIgM we observe implies that it has the potential to bind multiple copies of an epitope simultaneously, which is likely to support high-avidity antigen binding and/or antigen crosslinking. The structural relationship between tFc $\mu$  CH2 and Fab domains remains uncharacterized and thus modeling potential positions of tIgM Fabs is non-trivial; however, based on recent structures detailing the hIgM CH2 domain we estimate they might extend  $\sim 125$  Å from the center of the complex<sup>40</sup>. Furthermore, our structure indicates that Fabs are likely to be directed away from the Tp assembly, which separates Fc $\mu$ 1 and Fc $\mu$ 4, leaving portions of the Tp assembly and its adjacent Fcs sterically accessible. It is functionally intriguing that each copy of tFc $\mu$  is related to another by a different angle, raising the possibility that different geometric relationships between Fabs within a plg provide unique binding advantages for different types of antigen distribution. For example, Fcs separated by  $120^\circ$  might bind epitopes that are separated by larger distances compared to Fcs separated by  $80^\circ$  (Fig. 1c). We also hypothesize that the  $120^\circ$  separation between tFc $\mu$ 1 and tFc $\mu$ 4 could promote FcR binding to tFc $\mu$ 1 and/or tFc $\mu$ 4. We note that the two Fcs in dimeric forms of mammalian IgA are related by angles of  $97^\circ$ ; in human IgM, two Fcs are separated by  $53^\circ$ – $62^\circ$ , with exception of Fc $\mu$ 1 and Fc $\mu$ 5 which are separated by an angle of  $127^\circ$  (see Fig. 5a). The  $97^\circ$  bend between two IgA monomers in the SIgA structure is predicted to increase accessibility to FcR binding sites and promote binding away from Fabs<sup>22</sup>. Future studies will be necessary to determine if the geometric relationship between Ig monomers in plgs have been selected for because of a functional advantage, an assembly advantage, or both.

Although tSIgT has been identified as the immunoglobulin specialized in teleost mucosal immunity, substantial amounts of tSIgM are also found in gut and skin mucus before and after immunization<sup>18,21,41</sup>. SIgs are crucial to mucosal immunity because of their ability to bind and exclude microbes before they breach epithelial cells. This is especially important to fish since their skin is covered in mucus and constantly exposed to an external aquatic environment. In both

mammals and teleosts, the plgR binds and transports plgM across epithelial cells and is cleaved from the cell surface upon delivery to the mucosa, leaving its ectodomain bound to the plg where it is referred to as secretory component (SC). In mammals, the JC is necessary for this process and forms direct contacts with SC, which contains five Ig-like domains<sup>24,25</sup>. In contrast, JC is not required for delivery of tSIgM to the mucosa and tSC contains just two Ig-like domains that share low sequence conservation with mammalian counterparts; these observations indicate an alternative tSC-tIgM binding mechanism<sup>19,42</sup>. Consistent with these observations, we found that only two hIgM residues that bind hSC are conserved and solvent-accessible on tFc $\mu$  (Fig. 5b). It is yet unknown where the tSC binding site is located and if that site overlaps with other FcR binding sites.

## Methods

### Construct design and protein expression

The gene encoding the *Oncorhynchus mykiss* IgM heavy chain secretory form (GenBank: [AAW66974.1](https://www.ncbi.nlm.nih.gov/)) was obtained from the NCBI database (<https://www.ncbi.nlm.nih.gov/>). DNA sequence encoding a signal peptide (MFPASLLLLLAAASCVHC), a hexa-histidine tag, CH $\mu$ 3 (residues 222–318), CH $\mu$ 4 (residues 319–432), and Tp (residues 433–448) was codon optimized for human cell expression, synthesized (Integrated DNA Technologies, Inc.), and cloned into mammalian transient expression vector pD2610-v1 (Atum). Resulting plasmid DNA was transiently transfected into HEK Expi293-F cell line (Gibco, A14528) using the ExpiFectamine 293 Transfection Kit (Gibco). Six days after transfection, supernatants were harvested and His-tagged tIgM-Fc (tFc $\mu$ ) protein was purified using Ni-NTA Agarose (Qiagen) and subsequently, Superose 6 Increase 10/300 column (Cytiva) size exclusion chromatography (SEC). Elution fractions corresponding to the predicted molecular weight of tFc $\mu$  tetramer were collected and stored in TBS buffer (containing 20 mM Tris-HCl and 150 mM NaCl, pH 7.8) for cryo-EM.

To design chimeric Fcs, published structures containing hFc $\alpha$ , hFc $\gamma$ 1, and hFc $\mu$  (PDB IDs: 6UE7<sup>23</sup>, 1MCO<sup>43</sup>, 6KXS<sup>24</sup>, respectively) were structurally aligned to A-type and B-type tCH $\mu$ 4 structures. Constructs hFc-tFc $\mu$ -chi1 and hFc-tFc $\mu$ -chi2 were designed based on structural alignment to A-type tCH $\mu$ 4 and constructs hFc-tFc $\mu$ -chi3 and hFc-tFc $\mu$ -chi4 were designed based on structural alignment to B-type tCH $\mu$ 4. In construct hFc-tFc $\mu$ -chi1, A-type tCH $\mu$ 4 residues in FG loop, G strand, pre-Tp loop, and Tp (residues 412–448) were swapped into the C-terminal CH domain of hFcs (CH3 or CH4). In the hFc-tFc $\mu$ -chi2 construct, A-type tCH $\mu$ 4 residues in G strand, pre-Tp loop, and Tp (residues 417–448) were swapped into the C-terminal CH domain of hFcs (CH3 or CH4). In the hFc-tFc $\mu$ -chi3 construct, B-type tCH $\mu$ 4 residues in the pre-Tp loop, and Tp (residues 427–448) were swapped into the C-terminal CH domain of hFcs (CH3 or CH4). In the hFc-tFc $\mu$ -chi4 construct, B-type tCH $\mu$ 4 residues in G strand, pre-Tp loop, and Tp (residues 421–448) were swapped into the C-terminal CH domain of hFc $\alpha$  and hFc $\mu$ , or fused after the C-terminus of hFc $\gamma$ 1. Chimeric Fc constructs were cloned, transfected, and purified using the same protocol described for the tFc $\mu$  construct.

### SEC-multiangle light scattering (SEC-MALS) analysis

Protein samples were prepared as described. After purification by SEC (see Supplementary Figs. 1 and 6), fractions in each peak were pooled together, concentrated to approximately 2 g/L, filtered, and sent to University of Pennsylvania for SEC-MALS analysis. SEC-MALS experiments were performed on tFc $\mu$  and mutant variants. 100  $\mu$ l injections of purified protein in TBS buffer was loaded onto a Superose 6 Increase 10/300 column (Cytiva) at 0.5 ml/min at 25 °C. Absolute molecular weights were determined using MALS. The scattered light intensity of the column eluant was recorded at 18 different angles using a DAWN-HELEOS MALS detector (Wyatt Technology Corp.) operating at 658 nm after calibration with the monomer fraction of Type V BSA (Sigma).

The protein concentration of the eluant was determined using an inline Optilab T-rex interferometric refractometer (Wyatt Technology Corp.). The weight-averaged molecular weight of species within defined chromatographic peaks was calculated using the ASTRA software version 8.0 (Wyatt Technology Corp.), by construction of Debye plots ( $KC/R\theta$  versus  $\sin^2(\theta/2)$ ) at 1-s data intervals. The weight-averaged molecular weight was then calculated at each point of the chromatographic trace from the Debye plot intercept, and an overall average molecular weight was calculated by averaging across the peak. Data were further analyzed using the protein conjugate analysis module as implemented in Astra for protein with carbohydrate modifier. For all samples, the predicted molecular weight from the protein sequence, the average molecular weight determined by MALS, and the number and potential of PNGS are listed in Supplementary Table 4<sup>44</sup>.

### Cryo-EM grid preparation and data collection

UltraAuFoil R1.2/1.3 300 mesh grids (Quantifoil) were glow discharged in a Pelco easiGlow (TedPella Inc) for 90 seconds at 25 mA current. Using a Vitrobot Mark IV (Thermo Fisher) grids were prepared at 4 °C, 100% RH, blot force of 5, 0 s wait and drain time, and with blot time ranging from 2 to 8 second. Sample was diluted from 1.8 mg/mL to 0.1 mg/mL prior to freezing. Movies were collected using SerialEM on a Titan Krios G3 cryoTEM (Thermo Fisher) operating at 300 kV, equipped with BioQuantum Energy Filter (20 eV slit width, Gatan) and a K3 direct electron detector (Gatan) using beam image-shift in a 3x3x3 pattern. A total of 3618 movies were collected at 105,000 magnification in super-resolution mode with the corrected pixel size of 0.427 Å/pix, -60 electros/Å<sup>2</sup> total dose, and 40 frames per movie.

### Cryo-EM data processing

Movies were imported into cryoSPARC and binned by 2 during patch motion correction<sup>45,46</sup>. Images were curated for CTF fits, ice thickness, and contamination resulting in 3131 remaining images. Blob picker was used with about half of the images to generate an initial particle stack. After 2D classification, best classes were selected as templates for the Template picker on the full dataset, resulting in ~1.3 M picks. Particles were binned by 2 during extraction and subject to several rounds of 2D classification and heterogeneous refinement. The remaining 698k particles were re-extracted at the full pixel size. After an additional round of heterogeneous refinement, 610k particles were used in non-uniform refinement (with C2 symmetry), resulting in a 2.9 Å map<sup>47</sup>. After a cycle of local and global CTF refinement, an improved map at 2.78 Å was obtained. A non-uniform refinement with C1 symmetry was also computed, resulting in a 2.9 Å map. C1 and C2 maps were essentially superimposable and the C2 map was used for structure refinement.

### Atomic model building, refinement, and validation

A homology model for a single tCH $\mu$ 4 domain was made with SWISS-MODEL using the human IgM-Fc core cryo-EM structure (PDB ID: 6KXS<sup>24</sup>) as template; the C-terminal Tp was not included in the homology model<sup>48</sup>. The tCH $\mu$ 4 homology model was docked into real-space density of CH $\mu$ 4<sup>Fc $\mu$ IA</sup> using UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311<sup>49</sup>. Initial inspection of the docked homology model revealed several regions poorly fitted into density. To improved model fitting, less structured regions (residues Val362 to Gln378, C-terminal sequence after residue Arg427) were deleted from the homology model and resulting model was automatically fitted into real-space density using Coot Molecular Graphics Package<sup>50</sup>. Subsequently, residues 397–427 were deleted from the model due to mis-fitting and then rebuilt manually. The outcome was inspected and manually rebuilt in Coot before being subjected to Phenix real-space refinement<sup>51</sup>. After several

refinement iterations, the model of CH $\mu$ 4<sup>Fc $\mu$ IA</sup> and its Tp agreed well with the cryo-EM density map and a copy of the model was docked into density of CH $\mu$ 4<sup>Fc $\mu$ IB</sup>. Due to the structural differences between CH $\mu$ 4<sup>Fc $\mu$ IA</sup> and CH $\mu$ 4<sup>Fc $\mu$ IB</sup>, adjustments to the C-terminal sequence of CH $\mu$ 4<sup>Fc $\mu$ IB</sup> were made manually in Coot to optimize model fitting into cryo-EM map. Models of CH $\mu$ 4<sup>Fc $\mu$ IA</sup> and CH $\mu$ 4<sup>Fc $\mu$ IB</sup> were combined and refined using Phenix and Coot Molecular Graphics Package<sup>50,51</sup>. Copies of CH $\mu$ 4<sup>Fc $\mu$ IB</sup> were docked into the density of CH $\mu$ 4<sup>Fc $\mu$ 2C-4G</sup> and one copy of CH $\mu$ 4<sup>Fc $\mu$ IA</sup> was docked into the density of CH $\mu$ 4<sup>Fc $\mu$ 4H</sup> using UCSF Chimera; these newly docked CH $\mu$ 4 models were first refined individually using Phenix and then combined with previously refined CH $\mu$ 4<sup>Fc $\mu$ IA</sup> and CH $\mu$ 4<sup>Fc $\mu$ IB</sup> domains<sup>49,51</sup>. Residues linking CH $\mu$ 4 domains to Tps and residues in the Tp assembly were built manually into the combined model. Cys445 residues were built into density with rotamers that assumed the preferred orientation to form interchain disulfides. The resulting model was subjected to iterations of automatic refinement in Phenix and manual adjustment in the Coot Molecular Graphics Package. The refined CH $\mu$ 4 model was used as template to make a homology model for the tCH $\mu$ 3 domain using SWISS-MODEL<sup>48,50,51</sup>. The CH $\mu$ 3 homology model was docked into the density of CH $\mu$ 3<sup>Fc $\mu$ 2D</sup> and refined using similar methods as described above. Subsequently, copies of CH $\mu$ 3<sup>Fc $\mu$ 2D</sup> were docked into the density of remaining CH $\mu$ 3 domains. CH $\mu$ 3 domains were joined with the previously refined CH $\mu$ 4 domains and Tp assembly; the entire model is refined using Phenix and Coot Molecular Graphics Package<sup>50,51</sup>. The final tFc $\mu$  structure was evaluated by Phenix EM Validation, MolProbity, and EMRinger<sup>51,52</sup>; results are summarized in Supplementary Table 1.

### Sequence and structural alignments and structural analysis

NCBI Genbank accession numbers for Ig heavy chain sequences shown in figures are: *Oncorhynchus mykiss* (trout) IgM secretory form (AAW66974.1), *Salmo Salar* (salmon) IgM (AAB24064.1), *Xenopus laevis* (frog) IgM (AAA49774.1), *Trachemys scripta elegans* (turtle) IgM (AFR90255.1), *Gallus gallus* (chicken) IgM (P01875), *Mus musculus* (mouse) IgM (P01872), and *Homo sapiens* (human) IgM (P01871). Sequences alignments were made using ClustalOmega and corresponding figures were made using Esprict3<sup>53,54</sup>. Structural alignments between human IgM CH4 of (residues 450–555 of chain A; PDB ID: 6KXS<sup>24</sup>) and teleost IgM-CH4 (residues 316–428 of chain A or chain B) were made using the “align” function in PyMOL; structural alignments between teleost IgM CH3-CH4-Tp (residues 225–445) were made using the “align” function in PyMOL<sup>55</sup>. Interfaces were calculated using PISA, “Protein Interfaces, surfaces and assemblies” service at the European Bioinformatics Institute<sup>56</sup>; results are summarized in Supplementary Table 2.

### Fc angle and distance measurements

To measure the radius of polymeric forms of Fc $\mu$ , the tFc $\mu$  (this study) and human SIgM-Fc core (PDB ID: 6KXS<sup>24</sup>) structures were imported into UCSF Chimera<sup>49</sup>. For each model, a plane was defined using built-in “structure measurements” tools and the radius was calculated automatically. To measure angles between adjacent Fcs, Tp sequences were removed from models. An axis was defined using mass weighting and helical correction for each Fc monomer and then angles between the defined axes were measured and defined as the angles between adjacent Fcs. To measure the diameter of the solvent-accessible hole at the center of tFc $\mu$  tetramer, tFc $\mu$  model was imported into UCSF Chimera<sup>49</sup>. Nine atom pairs located on the surface were chosen around the hole. The distance between two atoms in each pair was measured and the diameter of this hole was determined to be the average of these distances. Chosen atom pairs and the distance measurements from each pair are summarized in Supplementary Table 3.



To measure the angle between the two tFc $\mu$ -Tp  $\beta$ -sheets, the tFc $\mu$  tetramer structure was imported into PyMOL<sup>55</sup>. The angle between Tp<sup>Fc $\mu$ 1A</sup> and Tp<sup>Fc $\mu$ 3E</sup> was measured as the dihedral angle between C $\alpha$  atoms of Leu434<sup>Fc $\mu$ 3E</sup>, Asn440<sup>Fc $\mu$ 1A</sup>, Asn440<sup>Fc $\mu$ 3E</sup>, and Leu434<sup>Fc $\mu$ 3E</sup>; angles between Tp<sup>Fc $\mu$ 1B</sup> and Tp<sup>Fc $\mu$ 3F</sup>, Tp<sup>Fc $\mu$ 2C</sup> and Tp<sup>Fc $\mu$ 4G</sup>, Tp<sup>Fc $\mu$ 2D</sup> and Tp<sup>Fc $\mu$ 4H</sup> were measured using the same method. The angle between two Tp  $\beta$ -sheets was determined to be the average of these measurements. To measure the angle between the two hFc $\mu$ -Tp  $\beta$ -sheets the hSIgM-Fc core structure (PDB ID: 6KXS<sup>24</sup>) was imported into PyMOL<sup>55</sup>. The angle between Tp<sup>Fc $\mu$ 1A</sup> and Tp<sup>Fc $\mu$ 5L</sup> was measured as the dihedral angle between C $\alpha$  atoms of Tyr562<sup>Fc $\mu$ 1A</sup>, Val567<sup>Fc $\mu$ 1A</sup>, Val567<sup>Fc $\mu$ 5L</sup>, and Tyr562<sup>Fc $\mu$ 5L</sup>; angles between Tp<sup>Fc $\mu$ 1B</sup> and Tp<sup>Fc $\mu$ 5K</sup>, Tp<sup>Fc $\mu$ 2C</sup> and Tp<sup>Fc $\mu$ 4H</sup>, Tp<sup>Fc $\mu$ 2D</sup> and Tp<sup>Fc $\mu$ 4G</sup>, Tp<sup>Fc $\mu$ 3E</sup> and Tp<sup>Fc $\mu$ 3F</sup> were measured using the same method. An average of these measurements was taken and the angle between two Tp  $\beta$ -sheets was determined to be the supplement angle of the average.

To measure the angle between tFc $\mu$  CH3 and CH4, the tFc $\mu$  tetramer model was imported into PyMOL<sup>55</sup>. For each copy of heavy chain, the angle between tFc $\mu$  CH3 and tFc $\mu$  CH4 was measured as the dihedral angle between the C $\alpha$  atoms of His298, Lys238, Glu412, and Thr332. The average of eight angle measurements was reported as the final value of tFc $\mu$  CH3-CH4 angle.

## Figures

Structural figures were made using UCSF Chimera and PyMOL<sup>49,55</sup>. SEC data and SEC-MALS data were replotted using Origin. All figures were assembled using Adobe Photoshop and Adobe Illustrator.

## Reporting summary

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

## Data availability

The tFc $\mu$  tetramer Cryo-EM data generated in this study have been deposited in the EM databank ([www.ebi.ac.uk/emdb](http://www.ebi.ac.uk/emdb)) with the accession code EMD-40054. The refined coordinates generated in this study have been deposited in the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)) with accession code 8GHZ. Previously published protein structure data used for analysis in this study are available in the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)) under PDB ID: 6KXS<sup>24</sup> (human hSIgM-Fc core), 6UE7<sup>23</sup> (human hSIgA-Fc core), 1MCO<sup>43</sup> (human IgG1-Fc). Source data are provided with this paper. The protein sequences used for sequence alignment are available in NCBI database (<https://www.ncbi.nlm.nih.gov/>) under these accession codes: *Oncorhynchus mykiss* (trout) IgM secretory form (AAW66974.1), *Salmo Salar* (salmon) IgM (AAB24064.1), *Xenopus laevis* (frog) IgM (AAA49774.1), *Trachemys scripta elegans* (turtle) IgM (AFR90255.1), *Gallus gallus* (chicken) IgM (P01875), *Mus musculus* (mouse) IgM (P01872), and *Homo sapiens* (human) IgM (P01871). All other data are available in the article and its Supplementary files or from the corresponding author upon request. Source data are provided with this paper.

## References

- Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* **11**, 47–59 (2010).
- Chapuis, R. M. & Koshland, M. E. Linkage and assembly of polymeric IgA immunoglobulins. *Biochemistry* **14**, 1320–1326 (1975).
- Wiersma, E. J. & Shulman, M. J. Assembly of IgM. Role of disulfide bonding and noncovalent interactions. *J. Immunol.* **154**, 5265–5272 (1995).
- Hohman, V. S. et al. J Chain in the Nurse Shark: Implications for Function in a Lower Vertebrate. *J. Immunol.* **170**, 6016–6023 (2003).
- Hadjiazimi, I. & Micheahamzhepour, M. XENOPUS-LAEVIS 19S IMMUNOGLOBULIN - ULTRASTRUCTURE AND J-CHAIN ISOLATION. *IMMUNOLOGY* **30**, 587–591 (1976).
- Castro, C. D. & Flajnik, M. F. Putting J Chain Back on the Map: How Might Its Expression Define Plasma Cell Development? *J. Immunol.* **193**, 3248–3255 (2014).
- Saito, S. et al. IgA tetramerization improves target breadth but not peak potency of functionality of anti-influenza virus broadly neutralizing antibody. *PLoS Pathog.* **15**, e1007427 (2019).
- Suzuki, T. et al. Relationship of the quaternary structure of human secretory IgA to neutralization of influenza virus. *Proc. Natl Acad. Sci. USA* **112**, 7809–7814 (2015).
- Terauchi, Y. et al. IgA polymerization contributes to efficient virus neutralization on human upper respiratory mucosa after intranasal inactivated influenza vaccine administration. *Hum. Vacc. Immunotherap.* **14**, 1351–1361 (2018).
- Eskeland, T. & Christensen, T. B. IgM Molecules with and without J Chain in Serum and after Purification, Studied by Ultra-centrifugation, Electrophoresis, and Electron Microscopy. *Scand. J. Immunol.* **4**, 217–228 (1975).
- Randall, T. D., Brewer, J. W. & Corley, R. B. Direct evidence that J chain regulates the polymeric structure of IgM in antibody-secreting B cells. *J. Biol. Chem.* **267**, 18002–18007 (1992).
- Erlandsson, L., Andersson, K., Sigvardsson, M., Lycke, N. & Leanderson, T. Mice with an inactivated joining chain locus have perturbed IgM secretion. *Eur. J. Immunol.* **28**, 2355–2365 (1998).
- Kaetzl, C. S. The polymeric immunoglobulin receptor: bridging innate and adaptive immune responses at mucosal surfaces. *Immunol. Rev.* **206**, 83–99 (2005).
- Macpherson, A. J., Yilmaz, B., Limenitakis, J. P. & Ganai-Vonarburg, S. C. IgA Function in Relation to the Intestinal Microbiota. *Annu. Rev. Immunol.* **36**, 359–381 (2018).
- Key, B. A., Baliga, R., Sinclair, A. M., Carroll, S. F. & Peterson, M. S. Structure, Function, and Therapeutic Use of IgM Antibodies. *Antibodies* **9**, 53 (2020).
- Pabst, O. New concepts in the generation and functions of IgA. *Nat. Rev. Immunol.* **12**, 821–832 (2012).
- Kaattari, S. et al. Varied redox forms of teleost IgM: an alternative to isotypic diversity? *Immunol. Rev.* **166**, 133–142 (1998).
- Zhang, Y.-A. et al. IgT, a primitive immunoglobulin class specialized in mucosal immunity. *Nat. Immunol.* **11**, 827–835 (2010).
- Stadtmueller, B. M. et al. The structure and dynamics of secretory component and its interactions with polymeric immunoglobulins. *eLife* **5**, e10640 (2016).
- Hamuro, K., Suetake, H., Saha, N. R., Kikuchi, K. & Suzuki, Y. A Teleost Polymeric Ig Receptor Exhibiting Two Ig-Like Domains Transports Tetrameric IgM into the Skin. *J. Immunol.* **178**, 5682–5689 (2007).
- Salinas, I., Fernández-Montero, Á., Ding, Y. & Sunyer, J. O. Mucosal immunoglobulins of teleost fish: A decade of advances. *Dev. Comp. Immunol.* **121**, 104079 (2021).
- Kumar Bharathkar, S. et al. The structures of secretory and dimeric immunoglobulin A. *eLife* **9**, e56098 (2020).
- Kumar, N., Arthur, C. P., Ciferri, C. & Matsumoto, M. L. Structure of the secretory immunoglobulin A core. *Science* **367**, 1008–1014 (2020).
- Li, Y. et al. Structural insights into immunoglobulin M. *Science* **367**, 1014–1017 (2020).
- Kumar, N., Arthur, C. P., Ciferri, C. & Matsumoto, M. L. Structure of the human secretory immunoglobulin M core. *Structure* **29**, 564–571.e3 (2021).
- Ahmed, A. A., Keremane, S. R., Vielmetter, J. & Bjorkman, P. J. Structural characterization of GASDALIE Fc bound to the activating Fc receptor Fc $\gamma$ R11a. *J. Struct. Biol.* **194**, 78–89 (2016).



27. Wurzburg, B. A., Garman, S. C. & Jardetzky, T. S. Structure of the Human IgE-Fc Cε3-Cε4 Reveals Conformational Flexibility in the Antibody Effector Domains. *Immunity* **13**, 375–385 (2000).
28. Amemiya, C. T. et al. The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 311–316 (2013).
29. Lalla, C., de Fagioli, C., Cessi, F. S., Smilovich, D. & Sitia, R. Biogenesis and function of IgM: the role of the conserved μ-chain tailpiece glycans. *Mol. Immunol.* **35**, 837–845 (1998).
30. Pasalic, D. et al. A peptide extension dictates IgM assembly. *Proc. Natl Acad. Sci. USA* **114**, E8575–E8584 (2017).
31. Wyatt, P. J. Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **272**, 1–40 (1993).
32. Hiramoto, E. et al. The IgM pentamer is an asymmetric pentagon with an open groove that binds the AIM protein. *Sci. Adv.* **4**, eaau1199 (2018).
33. Muller, R. et al. High-resolution structures of the IgM Fc domains reveal principles of its hexamer formation. *Proc. Natl Acad. Sci.* **110**, 10183–10188 (2013).
34. Anelli, T. et al. Sequential steps and checkpoints in the early exocytic compartment during secretory IgM biogenesis. *EMBO J.* **26**, 4177–4188 (2007).
35. van Anken, E. et al. Efficient IgM assembly and secretion require the plasma cell induced endoplasmic reticulum protein pERp1. *Proc. Natl Acad. Sci. USA* **106**, 17019–17024 (2009).
36. Ye, J., Kaattari, I. M., Ma, C. & Kaattari, S. The teleost humoral immune response. *Fish. Shellfish Immunol.* **35**, 1719–1728 (2013).
37. Ye, J., Bromage, E. S. & Kaattari, S. L. The Strength of B Cell Interaction with Antigen Determines the Degree of IgM Polymerization. *J. Immunol.* **184**, 844–850 (2010).
38. Nayak, D. K., Tang, A., Wilson, M., Miller, N. W. & Bengtén, E. Channel catfish soluble FcμR binds conserved linear epitopes present on Cμ3 and Cμ4. *Mol. Immunol.* **47**, 1306–1316 (2010).
39. Sheng, X., Qian, X., Tang, X., Xing, J. & Zhan, W. Polymeric Immunoglobulin Receptor Mediates Immune Excretion of Mucosal IgM–Antigen Complexes Across Intestinal Epithelium in Flounder (*Paralichthys olivaceus*). *Front. Immunol.* **9**, 1562 (2018).
40. Chen, Q., Menon, R., Calder, L. J., Tolar, P. & Rosenthal, P. B. Cryomicroscopy reveals the structural basis for a flexible hinge motion in the immunoglobulin M pentamer. *Nat. Commun.* **13**, 6314 (2022).
41. Hu, Y.-L., Xiang, L.-X. & Shao, J.-Z. Identification and characterization of a novel immunoglobulin Z isotype in zebrafish: Implications for a distinct B cell receptor in lower vertebrates. *Mol. Immunol.* **47**, 738–746 (2010).
42. Feng, L.-N. et al. Molecular cloning and functional analysis of polymeric immunoglobulin receptor gene in orange-spotted grouper (*Epinephelus coioides*). *Comp. Biochem. Physiol. Part B: Biochem. Mol. Biol.* **154**, 282–289 (2009).
43. Guddat, L. W., Herron, J. N. & Edmundson, A. B. Three-dimensional structure of a human immunoglobulin with a hinge deletion. *Proc. Natl Acad. Sci. USA* **90**, 4271–4275 (1993).
44. Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput.* 310–322 (2002).
45. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
46. Rubinstein, J. L. & Brubaker, M. A. Alignment of cryo-EM movies of individual particles by optimization of image translations. *J. Struct. Biol.* **192**, 188–195 (2015).
47. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
48. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
49. Pettersen, E. F. et al. UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D. Biol. Crystallogr* **66**, 486–501 (2010).
51. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D. Struct. Biol.* **75**, 861–877 (2019).
52. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
53. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
54. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
55. Schrödinger, L. L. C. The PyMOL Molecular Graphics System. *Version 1*, 8 (2015).
56. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).

## Acknowledgements

We thank Kushol Gupta (University of Pennsylvania) for assistance with SEC-MALS data collection and analysis and thank members of the Stadtmueller Laboratory for insightful conversations and suggestions related to this work. (Cryo) Electron microscopy was performed in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. This work was supported by NIH grant 1R01AI165570 and University of Illinois start-up funding to B.M.S, and the Lowell P. Hager Fellowship in Biochemistry to M.L.

## Author contributions

The study was conceived by B.M.S and M.L.; experiments were conducted by M.L. and A.G.M.; all authors contributed to data analysis and manuscript writing.

## Competing interests

BMS and ML are listed as inventors on a patent application that includes the design, production, and use of chimeric antibodies, some of which include teleost Ig heavy chain motifs. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-43240-z>.

**Correspondence** and requests for materials should be addressed to . Present address: Takeda Pharmaceuticals, Cambridge, MA 02139, USABeth M. Stadtmueller.

**Peer review information** *Nature Communications* thanks Martin Flajnik, Xinhua Chen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023