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

Mengfan Lyu[1], Andrey G. Malyutin[3,4,5], and Beth M. Stadtmueller[1,2,*]

[1]Department of Biochemistry, University of Illinois Urbana-Champaign, Urbana, Illinois 61801 USA

[2]Department of Biomedical and Translational Sciences, Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, Urbana, Illinois 61801 USA

[3]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125 USA

[4]Beckman Institute, California Institute of Technology, Pasadena, CA 91125 USA

[5]Present address, Takeda Development Center Americas, San Diego, California 92121

*Address correspondence to bethms@illinois.edu

ABSTRACT

Polymeric (p) immunoglobulins (Igs) serve broad functions during vertebrate immune responses. Typically, pIgs contain between two and six Ig monomers, each with two antigen binding fragments and one fragment crystallization (Fc). In addition, many pIgs assemble with a joining-chain (JC); however, the number of monomers and potential to include JC varies 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 revealed four tIgM Fcs linked through eight C-terminal tailpieces (Tps), which adopt a single β-sandwich-like domain (Tp assembly) located between two Fcs. Remarkably, 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 pIgs have evolved a range of assembly mechanisms and structures, each likely to support unique antibody effector functions.

INTRODUCTION

Polymeric (p) immunoglobulins (Igs) play critical roles in immune system function and are presumed to exist in all jawed vertebrates[1]. Despite the prevalence of pIg in vertebrate species, considerable differences in structure and function are thought to exist. Typically, pIgs 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 pIg assembly and for binding to Fc receptors (FcRs) to elicit downstream functions[2,3]. In mammals, and most birds and reptiles, pIgs are assembled in plasma cells expressing IgA or IgM heavy chains and a protein called the joining chain (JC). Polymeric forms of IgA are typically dimeric (d), containing two IgA monomers and one JC; however functional trimeric, tetrameric and pentameric pIgA have been identified in lower abundance[4–6]. Polymeric forms of IgM are typically pentameric, containing five IgM monomers and one JC, however hexamers lacking a JC have been identified in serum[7,8]. JC is necessary for pIgA and pentameric IgM assembly and required for their secretion into the mucosa by the polymeric Ig Receptor (pIgR). The pIgR is a transcytotic Fc receptor expressed on the basal lateral surface of epithelial cells, which binds to JC-containing pIgA and pIgM and transports them to the apical surface of the cell[9]. On the apical surface the pIgR ectodomain, called secretory component (SC), is proteolytically cleaved, releasing a complex containing SC and the pIg, which is termed a secretory (S) Ig. Whereas pIgM functions in serum and SIgM functions in the mucosa, most pIgA is delivered to the mucosa and functions as SIgA[10]. The pIgM, SIgM and SIgA exhibit unique capabilities compared to monomeric Ig, including high avidity antigen binding, antigen coating or 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[11,12].

While mammalian pIgs have been studied extensively, less is known about pIgs from other vertebrates. Of particular interest are pIgs from teleost (t), or bony fish, which represent an early evolutionary stage of vertebrate adaptive immunity. In contrast to mammals, birds, and most reptiles, teleost species do not encode a JC and express polymeric forms of tIgT and tIgM that are presumed to contain four copies of each respective monomer[13,14]. Whereas both tIgT and tIgM have been found in fish serum and mucus, tIgT is the predominant mucosal antibody, functionally similar to IgA in mammals[14]. Both tIgT and tIgM can be transported to the mucosa by tpIgR, which has distinct domain organization and structure compared to mammalian counterparts and apparently, does not require a JC to bind tIgT or tIgM[15]. The tSC has been isolated in complex with mucosal tIgT and tIgM, indicating that similar to mammalian SC, tSC remains bound to secretory forms of tIgT and tIgM and may play a role in mucosal immune functions[14,16]. Similar to mammalian pIgs, tpIgs have been associated with complement activation and bacterial coating[17].

The structural mechanisms underlying tpIgs' functions, and more broadly, the evolution of pIg structure-function relationships across species, is poorly understood in-part because until now, teleost antibody structures remained unreported. Furthermore, only in the past several years have structures of mammalian pIgs 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[18–20]. In these structures, both the Tps and the JC contribute β-strands to a β-sandwich near the center of the molecule, which stabilizes the assembly. Furthermore, the JC forms additional contacts with two Fcs, provides part of the pIgR (and SC) binding site, and appears to contribute asymmetry to pIg complexes that otherwise contain multiple heavy chains with identical sequence. Given the critical role for the JC in mammalian pIg structures, we hypothesized that tpIgs lacking JC would assemble using different mechanisms and adopt distinct structures. To test this hypothesis and to explore the evolution and diversity of vertebrate pIg structure and function, we determined the cryo-EM structure of the tIgM Fc core at 2.78Å resolution. Together with comparative and

mutational analysis, this structure provides a first glimpse of both a teleost pIg and a pIg lacking a JC, revealing distinct modes of assembly compared to mammalian pIgs 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[20,21], we expressed and purified a truncated teleost IgM heavy chain consisting of CHμ3, CHμ4, and the Tp. We utilized this tIgM Fc (tFcμ) to determine a cryo-EM structure to an average resolution of 2.78Å (Fig. 1, Supplementary Figs. 1, 2). In the structure, most main chain and side chain atoms in CHμ4 domains and the Tps were well resolved (resolution range 2.4-3.0Å). Map resolution was ~3.2Å and above for CHμ3 allowing the positions of main chain atoms and ~50-75% of side chain atoms in each individual chain to be refined; other side chains were built in geometrically reasonable positions (Supplementary Fig. 2).

The refined tFcμ structure contains four tFcμ monomers (Fcμ1 to Fcμ4) that lay approximately flat on one plane and include interfaces between Fcμ1 and Fcμ2, Fcμ2 and Fcμ3, Fcμ3 and Fcμ4. Fcμ1 and Fcμ4 are separated by Tp residues, which form a β-sandwich (hereafter termed *Tp assembly*) comprising two β-strands from each of the four Fcμs (Fig. 1). Overall, the tetramer is ~164Å across in the largest dimension and can be described as two structurally related halves, one containing Fcμ1-Fcμ2 and associated Tps, and the other containing Fcμ3-Fcμ4 and associated Tps; at the center of the tetramer is a solvent accessible hole with a diameter of ~12.2 Å. Despite C2 biological symmetry along an axis running between between Fcμ2, Fcμ3, and the Tp assembly, the complex is not symmetrical with respect to all four Fcs. Rather, adjacent tFcs are related by distinct angles measuring 81°, 72°, 86°, for Fcμ1 and Fcμ2, Fcμ2 and Fcμ3, Fcμ3

110 and Fcμ4, respectively. Fcμ1 and Fcμ4 are related by an angle of 121° and contact the Tp

111 assembly (Fig. 1b, c).

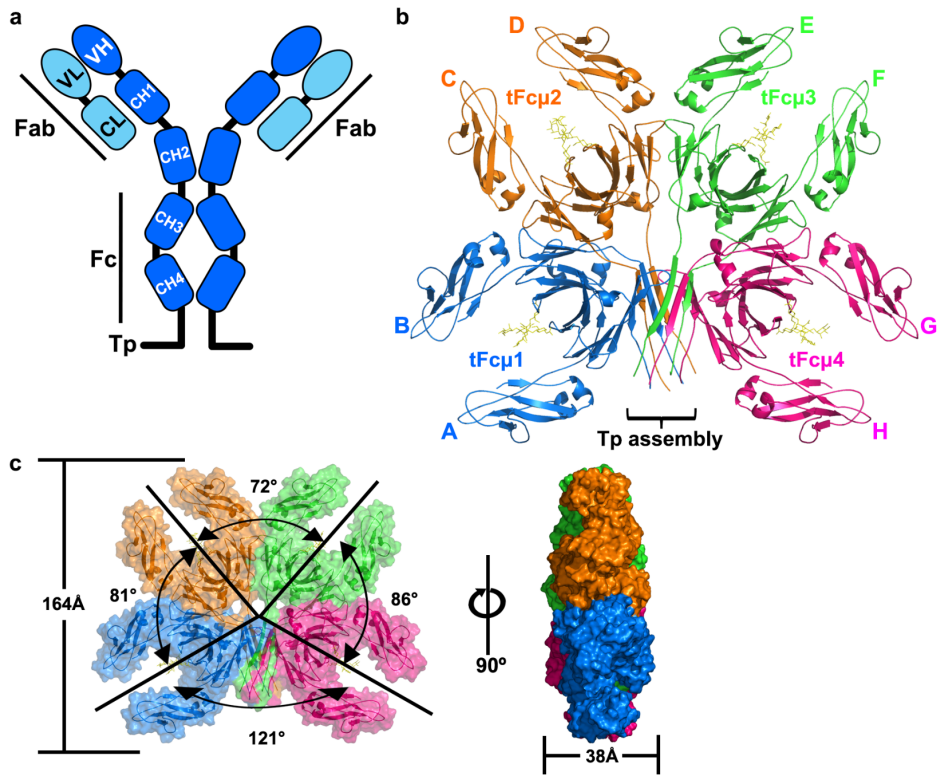


Fig. 1 | Overall structure of teleost IgM-Fc core, tFcμ. a Diagram of tIgM monomer. Constant (C) domains are represented as round corner rectangles and variable (V) domains are represented as ellipses. The heavy (H) is colored marine blue and the light (L) chain is colored light blue. **b** Cryo-EM structure of tFcμ (cartoon representation); chain IDs (A-H) and Fcμ1-4 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 tFcμ tetramer, showing both cartoon representation and molecular surface. Values describing the angles between tFcμ monomers as well as the approximate diameter and thickness of the tFcμ tetramer are indicated. In panels b and c, tFcμ monomers are colored as indicated: *Blue*, tFcμ1. *Orange*, tFcμ2. *Green*, tFcμ3. *Pink*, tFcμ4.

112

113 **tFcμ is formed from Fc monomers containing CHμ4 domains that adopt one of two folds**

114 Given the lack of published structural data for a teleost antibody, we first analyzed the structures

115 of tFcμ monomers and their individual domains, CHμ3 and CHμ4. The tFcμ monomers comprise

116 two copies of CHμ3 and CHμ4, which together form a canonical Fc characterized by an angle of

117 ~87° between CHμ3 and CHμ4 on the same chain and interfaces between adjacent CHμ3 and

118 adjacent CHμ4 domains on different chains. Each domain adopts an Ig-constant fold consisting

119 of two β-sheets linked by a disulfide bond, one sheet containing four β-strands termed A, B, E,
120 and D, and the other sheet containing three β-strands termed C, F, and G; the Tp β-strand follows
121 the CHμ4 G strand. Ordered carbohydrates were apparent on CHμ4 residue Asn 374. Within each
122 tFcμ, two CHμ4 domains interact primarily through residues located in the ABED β-sheets, which
123 mediate the interactions between the C-terminal Ig domain in most Fcs[18–20,22,23]; the interface
124 (including CHμ3, CHμ4 and Tp) is ~924Å^2 and is rich in electrostatic interactions (Fig. 2a and
125 Supplementary Fig. 3a, b).

126 Remarkably, despite sharing identical sequence and adopting Ig constant folds, the eight
127 copies of CHμ4 exhibit structural differences that stabilize the tFcμ tetramer. Specifically, the C-
128 terminal 38 residues (residues 411-448; FG loop, G strand and Tp) adopt one of two folds, with
129 chains A and H adopting an "A-type" fold and chains B-G adopting a "B-type" fold (Fig. 2b, c). In
130 the A-type fold, the FG loop and the G strand include residues 411-416 and 417-422, respectively,
131 and the Tp folds back to form β-sheet interactions with the G strand (Fig. 2b). In the B-type fold,
132 the FG loop and the G strand instead include residues 411-420 and 421-426, respectively,
133 resulting in Tp β-strand (residues 434-441) being directed away from CHμ4 where they form β-
134 sheet interactions with other Tp(s) rather than the G strand (Fig. 2c). Whereas chains A and H
135 are superimposable, the six B-type domains adopt one of three distinct conformations based on
136 the trajectory of the Tp and thus, each B-type domain is superimposable (RMSD of less than
137 0.18Å) with one other B-type domain on the opposing side of the tetramer (Supplementary Fig.
138 5d). In sum, the A-type CHμ4 domains can be described as having a C-F-G-Tp sheet whereas B-
139 type have a C-F-G sheet (with distinct F-G β-sheet interactions) and an extended Tp. Together
140 these two folds facilitate the incorporation of each chain into the Tp assembly (discussed below).

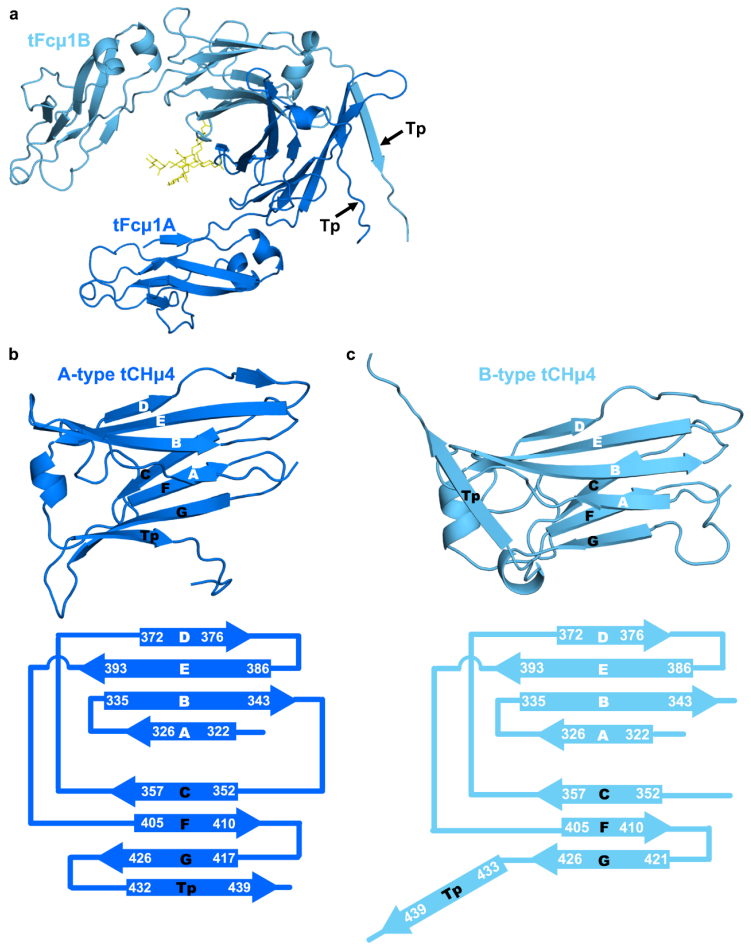


Fig. 2 | tIgM CH4 domains adopt one of two Ig folds. a Cartoon representation of tFcμ1 with glycans shown in yellow; the orientation is the same as Fig. 1b. **b-c** A-type and B-type tCHμ4 folds. *Top*, cartoon representations of A-type tCHμ4 domain and B-type tCHμ4 domain. *Bottom*, topology diagram showing the secondary structure of tFcμ1A and tFcμ1B; terminal residues of each β-strand are numbered.

Interactions between Fcμs are mediated by CHμ4-CHμ4 and CHμ4-Tp interfaces

The tFcμ tetramer includes contacts between adjacent Fcμ monomers and Tps. Interactions between adjacent Fcμ monomers (e.g., Fcμ1-Fcμ2, Fcμ2-Fcμ3, and Fcμ3-Fcμ4) are dominated by a common set of residues in the CHμ4 CD loops, FG loops and G-strands, forming an interface of ~1360Å^2 (Fig. 3a). Notable residues mediating these contacts include CHμ4 G strand residues Met422, Thr424, and Arg 423, which form electrostatic interactions and hydrogen bonds with G strand residues, CD loop (Asp359), and F strand (Ser406) residues in the neighboring CHμ4

domain. Additionally, FG loop residues Lys417 and Thr418 form electrostatic interactions with CD loop Arg364 and F strand Tyr410, respectively (Fig. 3b, c).

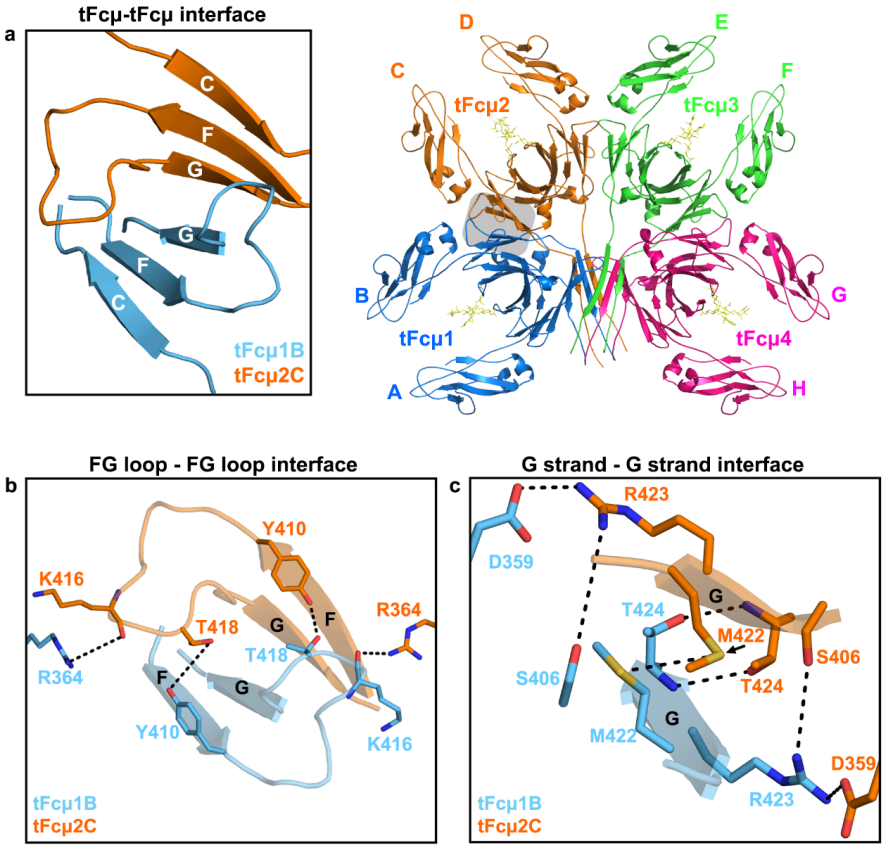


Fig. 3 | The CHμ4 interface between adjacent tFcμ monomers. A Overview of interactions between tFcμ1B and tFcμ2C at the CHμ4 interface (shaded area in tFcμ tetramer). *Left*, open-book overview of tFcμ-tFcμ interface boxed in the right panel, but shown in a different orientation. Only CHμ4 C-F-G β-sheets and FG loops are shown. *Right*, tFcμ tetramer with tFcμ1B and tFcμ2C at the CHμ4 interface in shaded box. **b** Electrostatic interactions and hydrogen bonding mediated by FG loops between tFcμ1B and tFcμ2C at the CHμ4 interface. Cartoon representations are shown for the F strand, G strand, and FG loop. Residues participating in interactions are shown as sticks. **c** Electrostatic and hydrogen bonding mediated by G strands between tFcμ1B and tFcμ2C at the CHμ4 interface. A cartoon representation of G strand is shown and interfacing residues are shown as sticks.

Fcμ1 and Fcμ4 do not interface and instead interact with the Tp assembly, which is situated between the two Fcμs (Fig. 1b and Fig. 4a). Notably, Fcμ1 and Fcμ4 each contain one A-type CHμ4 domain and one B-type CHμ4 domain. The A-type domains neighbor the Tp assembly and contribute uniquely to its structure. Specifically, because A-type Tp β-strands form

156 β-sheet interactions with CHμ4 G strands, residues in the pre-Tp loop (residues 434-441) are
157 positioned to contact both the top and bottom Tp β-sheets, effectively clamping the Fcμ1 and
158 Fcμ4 to the Tp assembly. Interactions between the pre-Tp loops and the rest of the Tp assembly
159 are dominated by polar contacts between A-type pre-Tp loop residues and B-type Tp residues
160 (Fig. 4d-f). Together Fc-Fc and Fc-Tp interactions form a band of contacts around the entire
161 tetramer, with A-type CHμ4 domains contributing directly to the Tp assembly (Fig. 4d-f) and B-
162 type CHμ4 domains forming interchain interactions between adjacent Fcμ monomers using CD
163 loops, FG loops and G-strands (Fig. 3b, c).
164

165 **Tps from adjacent tFcμ monomers alternate to form two β-sheets that stabilize the Tp**
166 **assembly**

167 Within tFcμ tetramer, β-strand pairing in the Tp assembly tightly links individual chains and also
168 promotes stabilizing interactions between the two halves of the tetramer (Fcμ1-Fcμ2 and Fcμ3-
169 Fcμ4). Specifically, we observe that Tp residues 434-440 form β-strands, which together
170 constitute two 4-stranded β-sheets, each comprised of Tps from two adjacent tFcμ monomers
171 (Fig. 4a). Notably, parallel Tp β-strand pairing alternates between Fcμs such that one sheet
172 comprises Tps from chains A-C-B-D and the other comprises Tps from chains E-G-F-H; in other
173 words, two Tp strands from the same Fcμ are separated by one Tp strand from an adjacent Fcμ
174 (Fig. 4a). The two β-sheets are stacked, forming two planes rotated at an angle of ~48° relative
175 to each other, and bury a hydrophobic core in-between. Buried residues include Val435, Leu437,
176 and Leu439 from each chain, whereas hydrophilic residues Asn436, Ser438, and Asn440 face
177 outwards (Fig. 4b, c).

178 Loop residues 441-448 follow the Tp β-strands. These residues are poorly ordered in the
179 map, however density suggests potential disulfide between copies of the residue Cys445
180 (Supplementary Fig. 5b). Residue 445 is a conserved Cys involved in interchain disulfide bonding
181 in mammalian IgA and IgM[19,20]; despite low local resolution, our map suggests possible disulfide

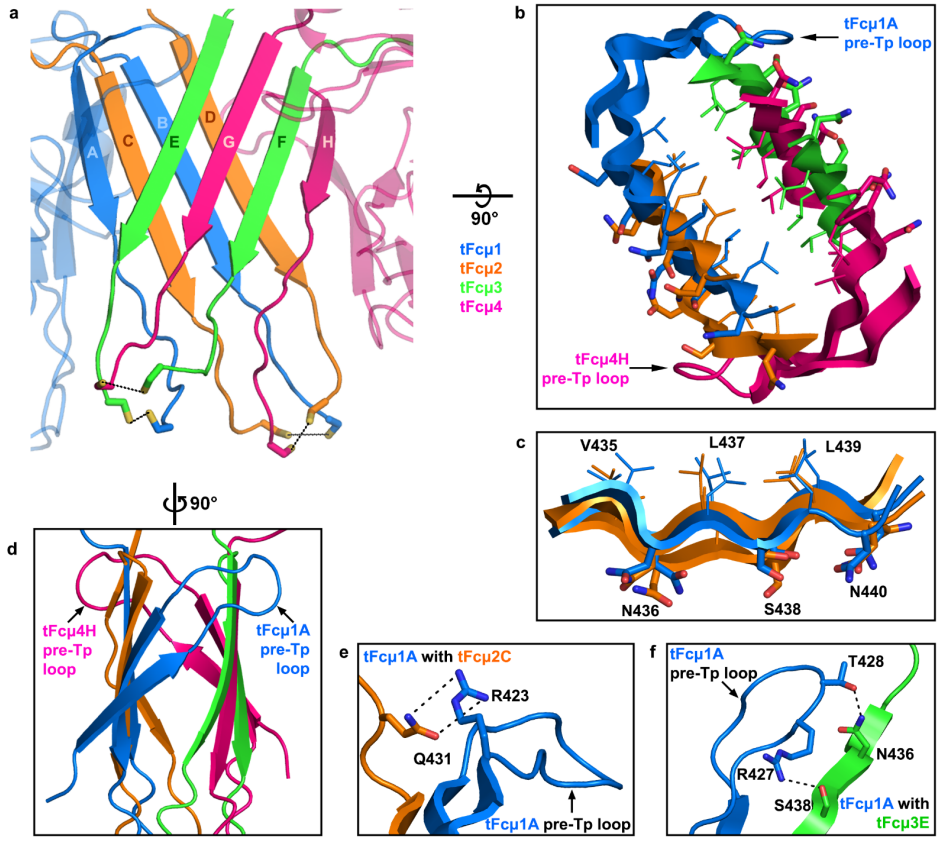


Fig. 4 | Teleost IgM Tp assembly and its interactions with pre-Tp loop of A-type CHμ4. **a** Cartoon representation of the tFcμ Tp assembly. The top β-sheet consists of $Tp^{tFc\mu3E}$, $Tp^{tFc\mu4G}$, $Tp^{tFc\mu3F}$, and $Tp^{tFc\mu4H}$ and bottom β-sheet consists of $Tp^{tFc\mu1A}$, $Tp^{tFc\mu2C}$, $Tp^{tFc\mu1B}$, and $Tp^{tFc\mu2D}$. Cys445 side chains are shown as sticks. Cys445 pairs that may form disulfides are linked using dashed lines. **b** Closeup view of the tFcμ Tp assembly hydrophobic core, which is depicted from the Tp C-termini looking toward the center of the tFcμ tetramer. **c** Closeup view of the tFcμ Tp assembly top β-sheet comprising β-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 tFcμ Tp assembly. **e** The tFcμ1A pre-Tp loop residues Arg423, Arg427, and Thr428 interact with $Tp^{tFc\mu2C}$ residues Gln431, Asn436 and Ser438; only chain A and chain C are shown. **f** The tFcμ1A pre-Tp loop residues Arg427 and Thr428 interact with $Tp^{tFc\mu3E}$ residues Asn436 and Ser438; only chain A and chain E are shown.

bonding between $Cys445^{Fc\mu1A}$ and $Cys445^{Fc\mu3E}$, between $Cys445^{Fc\mu1B}$ and $Cys445^{Fc\mu2C}$, $Cys445^{Fc\mu2D}$ and $Cys445^{Fc\mu4H}$, and $Cys445^{Fc\mu3F}$ and $Cys445^{Fc\mu4G}$ (see Supplementary Fig. 5b). Because these potential disulfides are located in the C-terminal loop region following the 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. 3a and Supplementary

189 Fig. 5b). Cysteine to serine mutation at this residue produced a tFcμ monomer, further suggesting
190 a critical role in tIgM polymeric assembly (Supplementary Fig. 5c).
191
192 **Comparisons between teleost and human IgM**
193 IgM is considered to be one of the evolutionarily oldest Igs. It is omnipresent in all jawed species
194 and teleost fish are the second oldest class expressing IgM, after cartilaginous fish[1]. To gain
195 insights on pIg evolution we compared teleost and human (h) IgM structures[20,21]. Whereas tIgM
196 is a tetramer, hIgM forms a JC-associated pentamer. Available human IgM structures contain the
197 Fc region (hFcμ), JC and human secretory component (hSC) (Fig. 5a). Overall, tFcμ and hFcμ
198 adopt structures characterized by adjacent Fc monomers linked through Fcμ-Fcμ contacts and a
199 Tp assembly; however, the structures are not globally superimposable and the relative orientation
200 of the Fcμs, residues involved in Fcμ-Fcμ contacts, and the location and structure of the Tp
201 assembly differ markedly, indicating that mechanisms of pIg assembly and function have evolved
202 from teleost to humans.
203 A fundamental difference between tFcμ and hFcμ is the geometric relationship among
204 adjacent Fcμ monomers. For example, adjacent tFcμs are related by angles of 80-85° and ~120°
205 (non-adjacent tFcμ1 and tFcμ4 separated by the Tp assembly) while the five adjacent hFcμs are
206 related by angles between ~53° and ~62° (non-adjacent hFcμ1 and hFcμ5 separated by hJC)
207 (Fig. 1b and Fig. 5a). The inter-Fc contacts that stabilize these geometries in both tFcμ and hFcμ
208 include CHμ4 C-F-G β-sheets and FG loops; however participating residues are not conserved,
209 resulting in an interface of ~400Å^2 between hCHμ4s and a larger interface of ~800Å^2 between
210 tCHμ4s. In addition, adjacent hFcμ monomer CH3 domains form contacts and are covalently
211 linked by disulfide bonds between Cys414 residues[20,21] whereas this cysteine residue is not
212 conserved in tFcμ and thus does not form inter-Fc disulfide bonds between tFcμ CH3 domains.
213 Structural comparisons between CH4 domains and Tps from humans and fish IgM are
214 especially relevant to understanding pIg 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. 5d). A structural alignment between the human and teleost CHμ4 domains reveals a RMSD of 1.269Å when aligned to A-type domains and RMSD of 1.179Å when aligned to B-type domains, suggesting that that B-type domains share greater structural similarity to human IgM CHμ4. 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 Fcμs and are not well conserved from teleost to mammalian IgM. Even in regions that share high structurally similarity, such as the A-B-D-E β-sheet, some residues mediating contacts are not conserved, including those that stabilize the CH4-CH4 interface within each Fc (Fig. 5d).

The Tp has long been implicated in pIg assembly; it is presumed to be present in all pIgs[2,3] and is moderately conserved among vertebrate pIg sequences (Supplementary Fig. 4). Mammalian IgM and IgA structures revealed Tps folding together with the JC to form β sandwich-like domains near the center of the complexes[18–20]. In human SIgM, residues 562-568 of each hFcμ Tp form a β-strand, and 10 copies form two parallel β-sheets that lie near-perpendicular to the central axis of the pentamer, positioning the distal β-strand from each sheet to pair with one β-strand from JC (Fig. 5b). In contrast, the tFcμ Tp assembly lacks a JC and is located toward the side of the complex between Fcμ1 and Fcμ4 with β-sheets positioned near-parallel to the central axis (Fig. 1b). In both structures, conserved hydrophobic residues Leu437 and Leu439 (Val564 and Leu566 in hFcμ) face the core of the sandwich and conserved hydrophilic residues Asn436 and Ser438 (Asn563 and Ser565 in hFcμ) face outward (Fig. 4b, c). However, despite the observation that conserved residues mediate hydrophobic interactions at the core of both hFcμ and tFcμ Tp assemblies, the relationship between the two Tp assembly β-sheets differs; the two tFcμ β-sheets are related by an angle of 48° whereas the two hFcμ β-sheets are related by an angle of 158° (Fig. 4a and Fig. 5b).

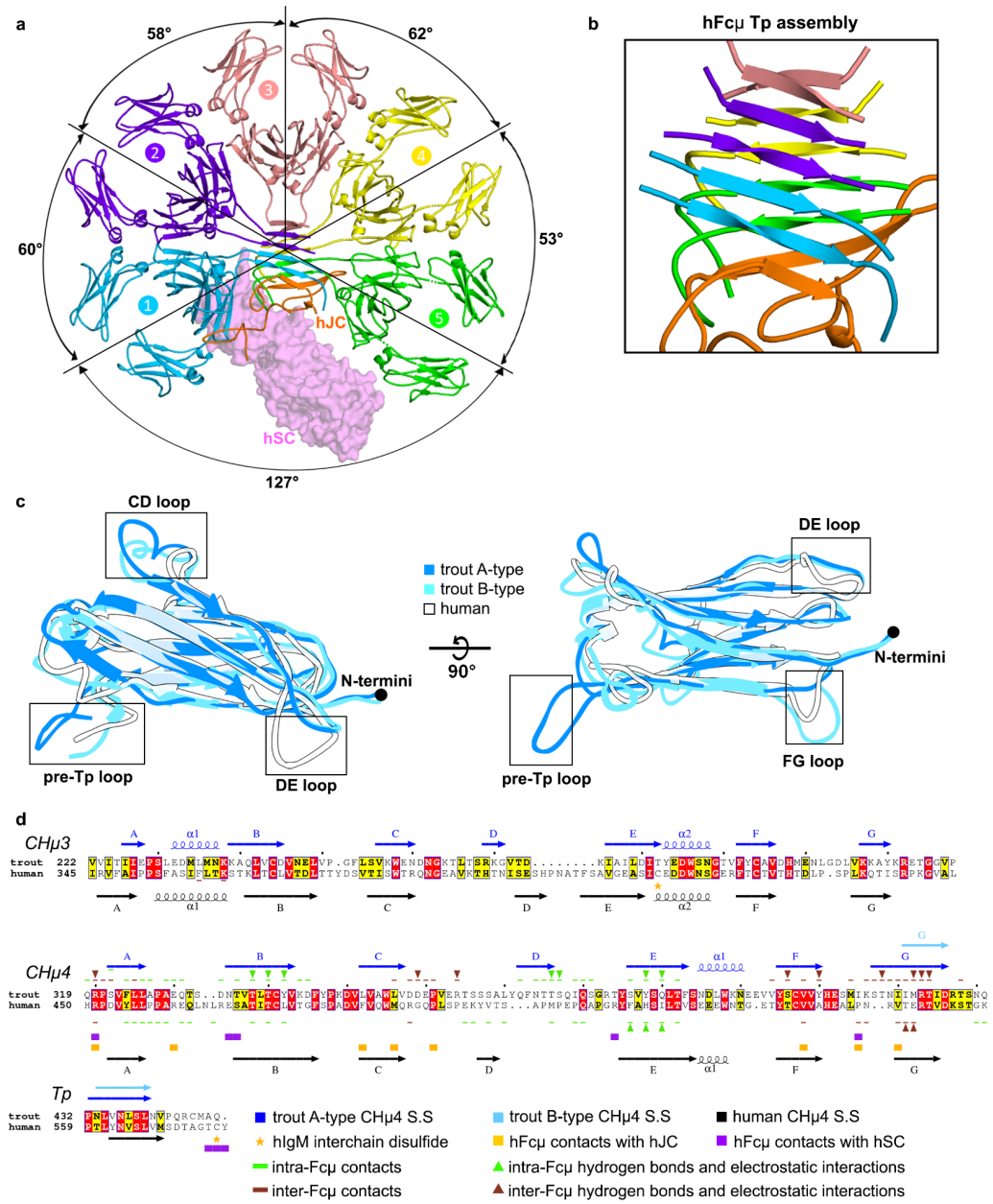


Fig. 5 | Comparison of IgM structures from trout and humans. a Human SFcμ structure, containing five hFcμ monomers, hJC, and hSC (PDB ID: 6kxs); hFcμ and hJC are shown as cartoon representations and hSC is shown as molecular surface representation. Angles between Fcμ monomers are indicated in the structure. **b** Overview of hFcμ Tp assembly. Residues 560-576 of hFcμ heavy chain and part of hJC are shown and colored using the same strategy as seen in panel a. **c** Structural alignment of hFcμ CHμ4 domain with tFcμ A-type CHμ4 domain and tFcμ B-type CHμ4 domain. All domains are shown as cartoon representations. **d** Sequence alignment of IgM CHμ3, CHμ4, and Tp from *Oncorhynchus mykiss* (rainbow trout) with *Homo sapiens* (human), with secondary structure (S.S) and interchain contacts shown at top (tFcμ) and bottom (hSFcμ) of the sequences. For B-type tCHμ4, secondary structure features identical to A-type are not indicated. In all panels, figures are annotated and colored according to the key.

242 The most striking difference between tFcμ and hFcμ Tp assemblies is a difference in β-
243 strand pairing. In tFcμ, each Tp pairs with a Tp originating from an adjacent Fc monomer while in
244 hFcμ, Tp β-strands pair with a second Tp from the same hFcμ monomer. The exception to this is
245 hFcμ3, which contributes one Tp to each β-sheet in the hFcμ Tp assembly (Fig. 4a and Fig. 5b).
246 The hFcμ Tp strand-pairing pattern is also observed for Tps in polymeric forms of IgA[19]. Despite
247 differences in strand pairing, the length from the last residues of CHμ4 (Asp426 in tIgM; Asp 553
248 in hIgM) to the first residue in the FcμTp β-strand (Leu434 in tIgM; Tyr562 in hIgM) is conserved
249 between the tFcμ B-type fold and hFcμ. Also conserved among hFcμ and tFcμ is a potential N-
250 linked glycosylation site (PNGS) including residues Asn436 and Ser438. Earlier studies indicated
251 that glycosylation at this site can influence hIgM polymerization and may prevent
252 aggregation[20,24,25]. However, we did not observe ordered glycosylation at this site in the tFcμ
253 structure (Supplementary Fig. 5a).
254 Finally, both tFcμ and hFcμ contain a conserved cysteine residue in the C-terminal loop
255 following the Tp β-strand. While only four out of ten copies of Cys575 are ordered in published
256 hFcμ structures[20], and density at this conserved cysteine (Cys445) is poor in the tFcμ structure
257 (Supplementary Fig. 5b), data point toward an essential role for a cysteine near the C-terminus
258 of IgMs. As noted, we observe putative disulfide bond formation linking tFcμ $Cys445^{Fc\mu1A}$ with
259 $Cys445^{Fc\mu3E}$ and $Cys445^{Fc\mu2D}$ with $Cys445^{Fc\mu4H}$ and deletion of this residue inhibits tetramer
260 formation (Supplementary Fig. 5c). In the case of hFcμ, this cysteine (Cys575) forms disulfides
261 with the JC or with another Tp located in the same sheet and is reportedly necessary for hIgM
262 assembly[20,21]. The pattern of disulfide bonding appears to be different and to link fundamentally
263 different regions in tIgM versus hIgM. In sum, structural analysis of tFcμ and hFcμ reveals that
264 despite conservation in sequence and length, IgM Tps can assemble into markedly different
265 structures to promote assembly of pIgs.
266
267

268 **tFcμ motifs introduced into human heavy chains can promote the assembly of pIgs**

269 Armed with new structural insights and a desire to better understand pIg evolution and assembly
270 mechanisms, we explored the ability of tFcμ motifs to facilitate the assembly of pIgs from other
271 heavy chains including human mu (IgM), alpha-1 (IgA1) and gamma-1 (IgG1). The tFcμ structure
272 suggests that both A-type and B-type tCHμ4 domains are essential for tFcμ assembly. Given that
273 tFcμ A-type and B-type tCHμ4 domains differed in the FG loop, G strand, and pre-Tp loop regions,
274 we hypothesized that the mechanisms of tFcμ assembly rely heavily on residues in these motifs.
275 To test this hypothesis, we designed twelve chimeric expression constructs, in which residues
276 found in the A-type or B-type tCHμ4 G strand and/or Tp residues are replaced with endogenous
277 counterparts in human Fcμ, Fcα1, and Fcγ1 (Fig. 6a). Chimeric Fc constructs and associated
278 controls were transiently transfected into human cells and the resulting proteins were purified and
279 the polymeric state was evaluated using size exclusion elution chromatography with in-line multi-
280 angle light scattering (SEC-MALS), which reports on the absolute molecular weight of target
281 proteins and complexes[26].

282 As noted, human IgM is naturally expressed as a hexamer or a JC-containing pentamer,
283 and the Tp is reportedly essential for the assembly of both complexes[25]. Consistent with published
284 studies, in our experimental system the removal of Tp residues resulted in only monomeric hFcμ
285 (Fig. 6b and Supplementary Fig. 6a). To determine if the tFcμ Tp could promote pIg assembly in
286 hFcμ constructs lacking the endogenous Tp (and the JC), we measured the molecular weight of
287 four chimeric proteins, hFcμ-tFcμ-chi1 to chi4 (Fig. 6b and Supplementary Fig. 6a). The hFcμ-
288 tFcμ-chi1 and hFcμ-tFcμ-chi2 were monomeric; however, hFcμ-tFcμ-chi3 and hFcμ-tFcμ-chi4
289 were polymeric with average molecular weights consistent with tetrameric complexes (Fig. 6b and
290 Supplementary Table 3). The hFcμ-tFcμ-chi3 contains the tFcμ Tp and hFcμ-tFcμ-chi4 contains
291 the B-type tCHμ4 G strand and tFcμ Tp residues; thus, data indicate that substitution of hFcμ
292 residues with these motifs, but not the A-type tCHμ4 G strand and tFcμ Tp can promote pIg

293 assembly. However, in contrast to the hexamers observed when full-length hFcμ is expressed

294 without JC, chimeric pIgs adopted probable tetramers.

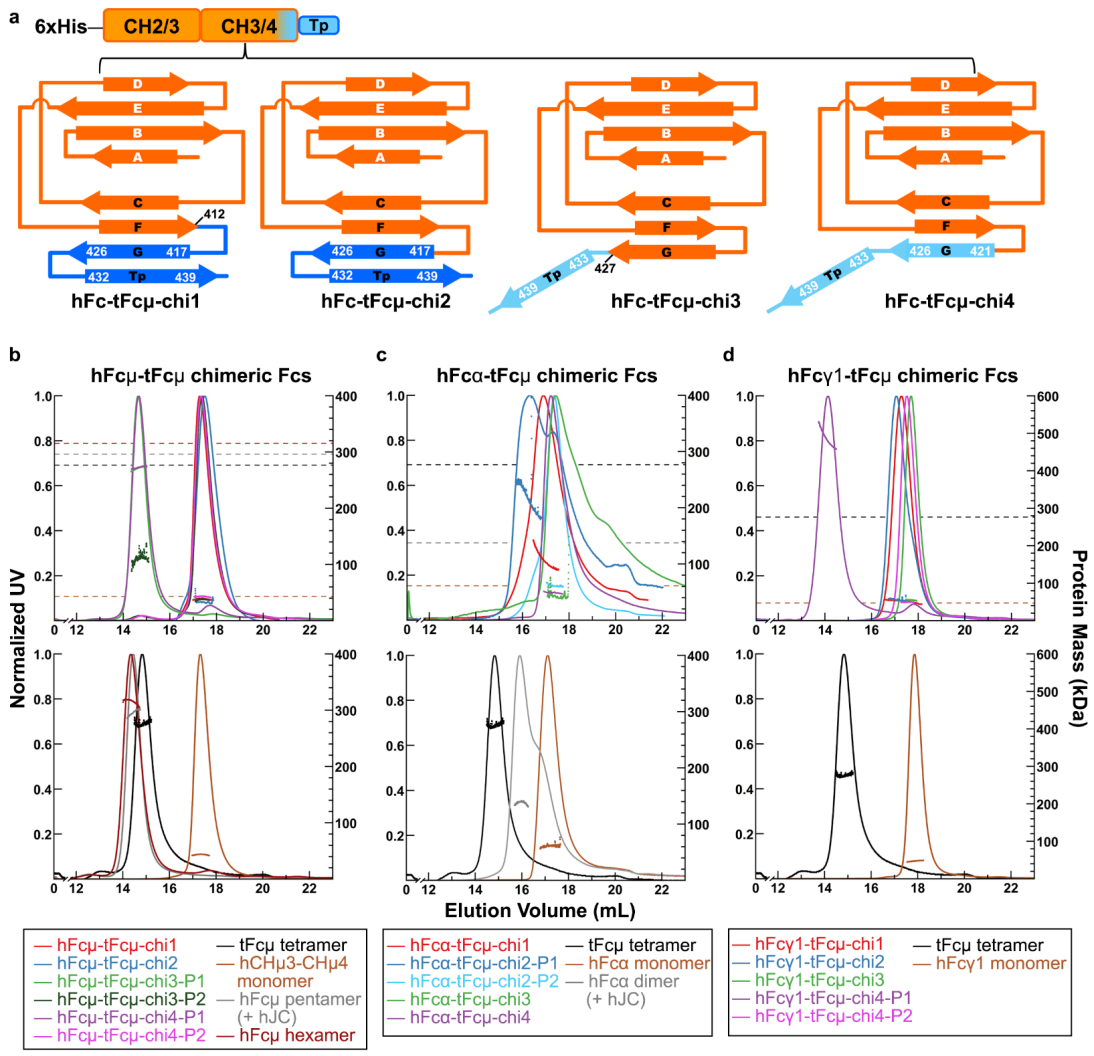


Fig. 6 | Design and characterization of chimeric Fcs | a Top: the design of chimeric Fc constructs; bottom: schematic showing topology of chimeric Fc constructs. Each chimeric Fc consists of an N-terminal hexa-histidine tag, partial hFc sequence (orange), and tFcμ C-terminal sequence (marine or light 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** hFcμ-tFcμ chimeric Fcs and relevant controls. **c** hFcα-tFcμ chimeric Fcs and relevant controls. **d** hFcγ1-tFcμ 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.

295

296 We also sought to determine if the tCHμ4 G strand and/or Tp residues could promote pIg
297 assembly in the context of hFcα, which is monomeric in the absence of the JC and typically
298 dimeric when expressed with a JC, or in the context of Fcγ1, which is naturally monomeric and
299 does not contain a Tp. In contrast to hFcμ-tFcμ chimeras, we found that hFcα-tFcμ-chi1 and hFcα-
300 tFcμ-chi2 formed pIgs whereases hFcα-tFcμ-chi3 and hFcα-tFcμ-chi4 did not. The molecular
301 weight of resulting pIgs was consistent with a mix of species ranging from dimers to tetramers
302 (Figs. 6c and Supplementary Fig. 6b). Data signify that in the context of the human alpha-1 heavy
303 chain, the G strand and Tp residues found in A-type tCHμ4 fold, can promote pIg assembly in
304 the absence of JC, whereas G strand and Tp residues found in the B-type tCHμ4 fold, or Tp
305 residues alone, cannot. Among hFcγ1-tFcμ chimeras, only hFcγ1-tFcμ-chi4 promoted pIg
306 assembly, suggesting that in the context of the human gamma-1 heavy chain, substituting G
307 strand and Tp residues found in the B-type tCHμ4 fold could promote assembly whereas others
308 could not. For all chimeric Fc constructs that promoted pIg assembly, pIgs were isolated from
309 mixtures containing mIgs before SEC-MALS analysis (Supplementary Fig. 6). Among hFcμ-tFcμ
310 chimeras, the pIgs represented more than 50% of the total protein purified whereas pIgs isolated
311 from hFcα-tFcμ and hFcγ1-tFcμ chimeras represented less than 30% of the total protein purified,
312 with mIgs making up the rest of the total (Supplementary Fig. 6). This suggests that in our
313 experimental system, pIg assembly promoted by tFcμ motifs was most efficient for hFcμ-tFcμ
314 chimeras.
315

316 **DISCUSSION**

317 Throughout evolution, pIgs have evolved critical, yet incompletely characterized functions in
318 jawed vertebrates. Data reported here, together with recently published mammalian pIg
319 structures, demonstrate that pIgs are a structurally diverse group of antibodies varying in
320 composition, size, and conformation. For the host, assembly of these complexes is a critical first
321 step to implementing a vast array of downstream effector functions and we find evidence that

322 across species, this can be accomplished through several distinct yet overlapping mechanisms.
323 Specifically, the Ig domains and Tps provide a common set of building blocks, while differences
324 in heavy chain amino acid sequence and the presence or absence of JC are variables adding
325 plasticity to the process.

326 In teleost fish, the Fc regions of pIgs are assembled from Ig domains and Tps formed by
327 heavy chain amino acid sequence, and the tFcμ structure reported in this study revealed a folding
328 pattern distinct from mammalian pIg structures containing the JC[27–29]. This points to the possibility
329 that JC-independent, and JC-dependent assembly evolved to employ distinct biophysical
330 mechanisms. However, we find that motifs from tFcμ are able to promote JC-independent pIg
331 assembly of otherwise human heavy chains, suggesting some degree of interchangeability
332 among the C-terminal sequences of pIg-forming heavy chains. For example, chimeric hFcμ-tFcμ
333 revealed that a subset of tFcμ motifs were able to promote pIg assembly of an otherwise human
334 IgM heavy chain in the absence of JC. It is intriguing that this effect was found only when the tFcμ
335 CH4 G strand and/or Tp residues found in the B-type fold, but not the A-type fold, were used.
336 While this may reflect the higher degree of similarity shared between the B type fold and human
337 IgM domains, the resulting pIgs appeared to be tetramers rather hexamers, the predominant
338 polymeric state associated with human IgM lacking the JC. It remains unknown whether hFcμ-
339 tFcμ chimera tetramers assemble into structures more similar to tIgM or hIgM with respect to the
340 Tp assembly structure, contacts, and geometric relationships between the incorporated Fcs.
341 While pIg assembly of hFcα-tFcμ, and hFcγ1-tFcμ chimeras was notably less efficient than hFcμ-
342 tFcμ chimeras, it is significant that tFcμ components could promote some pIg assembly in the
343 context of human IgA1 and IgG1 heavy chains, neither of which naturally form pIgs in the absence
344 of JC. While we cannot rule out the possibility that assembly in plasma cells differs from the HEK
345 cells used in our experiments, experimental cell lines assembled controls of expected MW and
346 provided a uniform system to evaluate how sequence differences among chimeras influences pIg
347 assembly.

348 The JC is found in the genomes of cartilaginous fish, reptiles, amphibians, birds ,and
349 mammals; however it is unclear if it is functionally incorporated into pIg in all these species and
350 as noted, hexameric hIgM can assemble in the absence of JC[27–29]. This raises the question of
351 whether or pIg lacking the JC from species other than teleost fish have structures most similar to
352 hexameric hIgM or to tetrameric tIgM? At this time, a high resolution structure detailing the Tp
353 assembly of hexameric IgM has not been reported; however, negative-stain EM and small-angle
354 light scattering (SAXS) data suggest all six hFcμs form Fc-Fc contacts, and that the Tp assembly
355 remains in the center of the complex as seen in pentameric IgM, rather than bridging two Fcs as
356 we observed in tetrameric tFcμ[30,31]. The structures of unliganded Tps (e.g. those found in
357 monomeric IgA or IgM) are also unknown and therefore it remains unclear if and how Tps might
358 change conformation during mammalian pIg assembly in absence of JC and/or if they adopt
359 conformations distinct from what is observed in mammalian pIg and SIg structures. Without such
360 information, it is difficult to judge whether the Tp β-strands would alternate in hIgM hexamers, as
361 observed in tFcμ tetramers, or if they would pair with another Tp β-strand from the same Fc as
362 observed in hFcμ pentamer and mammalian IgA polymers.

363 While our structural data provide considerable insight on tIgM structure and assembly
364 mechanisms, the stages of the assembly process remain unclear. Identification of two distinct
365 tCHμ4 folds evokes the question of how tIgM heavy chains fold and if they adopt A- and/or B-type
366 CHμ4 domains prior to the assembly of pIgM. In isolation, the A-type fold appears likely to be
367 more stable than the B-type fold because the β-strand region of the Tp forms β-sheet interactions
368 with the G strand, rather than being extended away from the core of the Ig domain in a potentially
369 flexible location (see Fig. 2b). While further study is needed to define the stages of tFcμ assembly,
370 one possibility is that heavy chains first adopt the A-type conformation and that polymeric
371 assembly is triggered when two tFcμ monomers come into proximity and interactions at the Fc-
372 Fc interface displace Tp β-strands from both Fcs, effectively kicking them out of the Ig-fold and
373 allowing them to form parallel β-sheet interactions with each other. It is unclear if there is sufficient

374 energy in the tFcμ-tFcμ interface for its formation to break Tp-G strand hydrogen bonding and
375 thus if this model is correct, it is conceivable that molecular chaperones, which have been
376 implicated in mammalian pIgA and pIgM assembly processes, are involved[32,33]. Regardless,
377 sequential interactions between Fcs could add strands to the Tp assembly until the tetramer was
378 closed by the last remaining A-type fold. This model is consistent with the alternating Tp β-strand
379 pairing that we observe in tFcμ, but that is not seen in mammalian IgM or IgA containing JC[18–20].
380 Furthermore, given that Cys445Ser mutation abolished tetramer formation, it is possible that
381 disulfide bonds linking non-adjacent Fcs, and/or the top and bottom β-sheets in the Tp assembly
382 (see Fig. 4a, Supplementary Fig. 5b, c) stabilize some part of this process.

383 Another remarkable finding of our work is that the majority of pIg assembly mechanisms
384 appear likely to culminate in the release of an asymmetric pIg complex, despite being assembled
385 from multiple copies of identical Ig heavy chains with identical sequence. Asymmetry in
386 mammalian pIgM and pIgA was verified with the publication of pIgA and pIgM structures, in which
387 the JC was bound to two Igs uniquely, and is thus postulated to template the addition of monomers
388 during assembly and to confer asymmetry on the resulting complex[18–20]. However, the tFcμ
389 structure, which lacks a JC, also reveals elements of asymmetry, with four Fc monomers directed
390 away from the Tp assembly such that the relative geometric relationships between each tFcμ are
391 not identical. As noted, this arrangement appears to be dependent upon the existence of A-type
392 and B-type CHμ4 domains, yet it is unclear if asymmetry in pIgs evolved as an efficient means to
393 assemble a pIg and/or if asymmetry provides a functional advantage.

394 IgM is a critical Ig heavy chain class found in all jawed vertebrates. Studies in mammals
395 demonstrate that its effector functions are mediated by high avidity interactions with antigen (from
396 five sets of Fabs), complement activation, binding to IgM-specific Fc receptors (FcRs), and
397 mucosal functions associated with pIgR-dependent secretion. While relatively few reports have
398 documented its function in teleost fish, evidence suggests that tIgM also participates in
399 complement activation, mucosal functions such as bacterial coating and immune exclusion and

binds IgM-specific FcRs[17,34–36]. 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 tFcμ CH2 and Fab domains remains uncharacterized and thus modeling potential positions of tIgM Fabs is non-trivial; however, based recent structures detailing the hIgM CH2 domain we estimate they might extend ~125Å from the center of the complex[37]. Furthermore, our structure indicates that Fabs are likely to be directed away from the Tp assembly, which separates Fcμ1 and Fcμ4, leaving portions of the Tp assembly and its adjacent Fcs sterically accessible. It is functionally intriguing that each copy of tFcμ is related to another by a different angle, raising the possibility that different geometric relationships between Fabs within a pIg provide unique binding advantages for different types of antigen distribution. For example, Fcs separated by 120° might bind epitopes that are separated by larger distances compared to Fcs separated by 80° (Fig. 1c). We also hypothesize that the 120° separation between tFcμ1 and tFcμ4 could promote FcR binding to tFcμ1 and/or tFcμ4. We note that the two Fcs in dimeric forms of mammalian IgA are related by angles of 97°; in human IgM, two Fcs are separated by 53°-62°, with exception of Fcμ1 and Fcμ5 which are separated by an angle of 127° (see Fig. 5a). The 97° bend between two IgA monomers in the SIgA structure is predicted to increase accessibility to FcR binding sites and promote binding away from Fabs[18]. Future studies will be necessary to determine if the geometric relationship between Ig monomers in pIg have been selected for because of a functional advantage, an assembly advantage, or both.

Although teleost SIgT 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[14,17,38]. SIg species 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 teleost fish, the pIgR binds and transports pIgM across epithelial cells and

426 is cleaved from the cell surface upon delivery to the mucosa, leaving its ectodomain bound to the
427 pIg where it is referred to as secretory component (SC). In mammals, the JC is necessary for this
428 process and forms direct contacts with SC, which contains five Ig-like domains[20,21]. In contrast,
429 JC is not required for delivery of tSIgM to the mucosa and tSC contains just two Ig-like domains
430 that share low sequence conservation with mammalian counterparts; these observations indicate
431 an alternative tSC-tIgM binding mechanism[15,39]. Consistent with these observations, we found
432 that only two hIgM residues that bind hSC are conserved and solvent-accessible on tFcμ (Fig.
433 5d). It is yet unknown where the tSC binding site is located and if that site overlaps with other FcR
434 binding sites.
435

436 **MATERIALS AND METHODS**

437 **Construct design and protein expression**

438 The gene encoding the *Oncorhynchus mykiss* IgM heavy chain secretory form (GenBank:
439 AAW66974.1) was obtained from the NCBI database. DNA sequence encoding a signal peptide
440 (MFPASLLLLLAAASCVHC), a hexa-histidine tag, CHμ3 (residues 222-318), CHμ4 (residues
441 319-432), and Tp (residues 433-448) was codon optimized for human cell expression,
442 synthesized (Integrated DNA Technologies, Inc.) and cloned into mammalian transient
443 expression vector pD2610-v1 (Atum). Resulting plasmid DNA was transiently transfected into
444 HEK Expi293-F cell line using the ExpiFectamine™ 293 Transfection Kit (Thermo Fisher
445 Scientific). Six days after transfection, supernatants were harvested and His-tagged tIgM-Fc (tFcμ)
446 protein was purified using Ni-NTA Agarose (Qiagen) and subsequently, Superose 6 Increase
447 10/300 column (Cytiva) size exclusion chromatography (SEC). Elution fractions corresponding to
448 the predicted molecular weight of tFcμ tetramer were collected and stored in TBS buffer
449 (containing 20mM Tris-HCl and 150mL NaCl, pH 7.8) for cryo-EM.

450 To design chimeric Fcs, published structures containing hFcα, hFcγ1, and hFcμ (PDB IDs:
451 6UE7, 1MCO, 6KXS, respectively) were structurally aligned to A-type and B-type tCHμ4

452 structures. Constructs hFc-tFcμ-chi1 and hFc-tFcμ-chi2 were designed based on structural
453 alignment to A-type tCHμ4 and constructs hFc-tFcμ-chi3 and hFc-tFcμ-chi4 were designed based
454 on structural alignment to B-type tCHμ4. In construct hFc-tFcμ-chi1, A-type tCHμ4 residues in FG
455 loop, G strand, pre-Tp loop, and Tp (residues 412-448) were swapped into the C-terminal CH
456 domain of hFcs (CH3 or CH4). In the hFc-tFcμ-chi2 construct, A-type tCHμ4 residues in G strand,
457 pre-Tp loop, and Tp (residues 417-448) were swapped into the C-terminal CH domain of hFcs
458 (CH3 or CH4). In the hFc-tFcμ-chi3 construct, B-type tCHμ4 residues in pre-Tp loop, and Tp
459 (residues 427-448) were swapped into the C-terminal CH domain of hFcs (CH3 or CH4). In the
460 hFc-tFcμ-chi4 construct, B-type tCHμ4 residues in G strand, pre-Tp loop, and Tp (residues 421-
461 448) were swapped into the C-terminal CH domain of hFcα and hFcμ, or fused after the C-
462 terminus of hFcγ1. Chimeric Fc constructs were cloned, transfected, and purified using the same
463 protocol described for the tFcμ construct.
464
465 **SEC-Multiangle light scattering (SEC-MALS) analysis**.
466 Proteins samples were prepared as described. After purification by SEC (see Supplementary Figs.
467 1&6), fractions in each peak were pooled together, concentrated to approximately 2 g/L, filtered,
468 and sent to University of Pennsylvania for SEC-MALS analysis. SEC-MALS experiments were
469 performed on tFcμ and mutant variants. 100uL injections of purified protein in TBS buffer was
470 loaded onto a Superose 6 Increase 10/300 column (Cytiva) at 0.5 ml/min at 25°C. Absolute
471 molecular weights were determined using MALS. The scattered light intensity of the column eluant
472 was recorded at 18 different angles using a DAWN-HELEOS MALS detector (Wyatt Technology
473 Corp.) operating at 658 nm after calibration with the monomer fraction of Type V BSA (Sigma).
474 Protein concentration of the eluant was determined using an in-line Optilab T-rex interferometric
475 refractometer (Wyatt Technology Corp.). The weight-averaged molecular weight of species within
476 defined chromatographic peaks was calculated using the ASTRA software version 8.0 (Wyatt
477 Technology Corp.), by construction of Debye plots ($KC/R\theta$ *versus* $\sin^2[\theta/2]$) at 1-s data intervals.

478 The weight-averaged molecular weight was then calculated at each point of the chromatographic
479 trace from the Debye plot intercept, and an overall average molecular weight was calculated by
480 averaging across the peak. Data were further analyzed using the protein conjugate analysis
481 module as implemented in Astra for protein with carbohydrate modifier. For all samples, the
482 predicted molecular weight from the protein sequence, the average molecular weight determined
483 by MALS, and the number and potential of PNGS are listed in Supplementary Table 4.
484

485 **Cryo-EM grid preparation and data collection**

486 UltrAuFoil R1.2/1.3 300 mesh grids (Quantifoil) were glow discharged in a Pelco easiGlow
487 (TedPella Inc) for 90 seconds at 25 mA current. Using a Vitrobot Mark IV (ThermoFisher) grids
488 were prepared at 4 ^{o}C, 100 % RH, blot force of 5, 0 s wait and drain time, and with blot time
489 ranging from 2 to 8 second. Sample was diluted from 1.8 mg/mL to 0.1 mg/mL prior to freezing.
490 Movies were collected using SerialEM on a Titan Krios G3 cryoTEM (Thermo Fisher) operating
491 at 300 kV, equipped with BioQuantum Energy Filter (20 eV slit width, Gatan) and a K3 direct
492 electron detector (Gatan) using beam image-shift in a 3x3x3 pattern. A total of 3618 movies were
493 collected at 105,000 magnification in super resolution mode with the corrected pixel size of 0.427
494 Å/pix, ~60 electros/Å^2 total dose, and 40 frames per movie.
495

496 **Cryo-EM data processing**

497 Movies were imported into cryoSPARC and binned by 2 during patch motion correction[40,41].
498 Images were curated for CTF fits, ice thickness, and contamination resulting in 3131 remaining
499 images. Blob picker was used with about half of the images to generate an initial particle stack.
500 After 2D classification, best classes were selected as templates for Template picker on the full
501 dataset, resulting in ~1.3 M picks. Particles were binned by 2 during extraction and subject to
502 several rounds of 2D classification and heterogeneous refinement. Remaining 698k particles were
503 re-extracted at the full pixel size. After an additional round of heterogeneous refinement, 610k

504 particles were used in non-uniform refinement (with C2 symmetry), resulting in a 2.9 Å map[42].
505 After a cycle of local and global CTF refinement, an improved map at 2.78 Å was obtained. A non-
506 uniform refinement with C1 symmetry was also computed, resulting in a 2.9 Å map. C1 and C2
507 maps were essentially superimposable (not shown) and the C2 map was used for structure
508 refinement.
509

510 **Structure determination, refinement, and validation**

511 A homology model for a single tCHμ4 domain was made with SWISS-MODEL using the human
512 IgM-Fc core cryo-EM structure (PDB ID: 6KXS) as template; the C-terminal Tp was not included
513 in the homology model[43]. The tCHμ4 homology model was docked into real-space density of
514 $CH\mu4^{Fc\mu1A}$ using UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and
515 Informatics at the University of California, San Francisco, with support from NIH P41-GM103311[44].
516 Initial inspection of the docked homology model revealed several regions poorly fitted into density.
517 To improved model fitting, less structured regions (residues Val362 to Gln378, C-terminal
518 sequence after residue Arg427) were deleted from the homology model and resulting model was
519 automatically fitted into real-space density using Coot Molecular Graphics Package[45].
520 Subsequently, residues 397-427 were deleted from the model due to mis-fitting and then rebuilt
521 manually. The outcome was inspected and manually rebuilt in Coot before being subjected to
522 Phenix real-space refinement[46]. After several refinement iterations, the model of $CH\mu4^{Fc\mu1A}$ and
523 its Tp agreed well with the cryo-EM density map and a copy of the model was docked into density
524 of $CH\mu4^{Fc\mu1B}$. Due to the structural differences between $CH\mu4^{Fc\mu1A}$ and $CH\mu4^{Fc\mu1B}$, adjustments to
525 the C-terminal sequence of $CH\mu4^{Fc\mu1B}$ were made manually in Coot to optimize model fitting into
526 cryo-EM map. Models of $CH\mu4^{Fc\mu1A}$ and $CH\mu4^{Fc\mu1B}$ were combined and refined using Phenix and
527 Coot Molecular Graphics Package[45,46]. Copies of $CH\mu4^{Fc\mu1B}$ were docked into the density of
528 $CH\mu4^{Fc\mu2C\text{-}4G}$ and one copy of $CH\mu4^{Fc\mu1A}$ was docked into the density of $CH\mu4^{Fc\mu4H}$ using UCSF
529 Chimera; these newly docked CHμ4 models were first refined individually using Phenix and then

530 combined with previously refined CHμ4^{Fcμ1A} and CHμ4^{Fcμ1B} domains[44,46]. Residues linking CHμ4
531 domains to Tps and residues in the Tp assembly were built manually into the combined model.
532 Cys445 residues were built into density with rotamers that assume preferred orientation to form
533 interchain disulfides. The resulting model was subjected to iterations of automatic refinement in
534 Phenix and manual adjustment in Coot Molecular Graphics Package. The refined CHμ4 model
535 was used as template to make a homology model for the tCHμ3 domain using SWISS-
536 MODEL[43,45,46]. The CHμ3 homology model was docked into the density of CHμ3^{Fcμ2D} and refined
537 using similar methods as described above. Subsequently, copies of CHμ3^{Fcμ2D} were docked into
538 the density of remaining CHμ3 domains. CHμ3 domains were joined with the previously refined
539 CHμ4 domains and Tp assembly; the entire model is refined using Phenix and Coot Molecular
540 Graphics Package[45,46]. The final tFcμ structure was evaluated by Phenix EM Validation,
541 MolProbity, and EMRinger[46,47]; results are summarized in Supplementary Table 1.
542
543 **Sequence and Structural alignments and structural analysis**
544 NCBI Genbank accession numbers for Ig heavy chain sequences shown in figures are:
545 *Oncorhynchus mykiss* IgM secretory form (AAW66974.1), *Salmo Salar* IgM (AAB24064.1),
546 *Xenopus laevis* IgM (AAA49774.1), *Trachemys scripta elegans* IgM (AFR90255.1), *Gallus gallus*
547 IgM (P01875), *Mus musculus* IgM (P01872), and *Homo sapiens* IgM (P01871). Sequences
548 alignments were made using ClustalOmega and corresponding figures were made using
549 Espript3[48,49]. Structural alignments between human and teleost IgM were made using the "super"
550 function in PyMOL[50]. Interfaces were calculated using PISA, "Protein Interfaces, surfaces and
551 assemblies" service at the European Bioinformatics Institute[51]; results are summarized in
552 Supplementary Table 2.
553
554
555

556 **Fc angle and distance measurements**

557 To measure radius of a polymeric forms of Fcμ, the tFcμ (this study) and human SIgM-Fc
558 core (PDB ID: 6KXS) structures were imported into UCSF Chimera[44]. For each model, a plane
559 was defined using built-in "structure measurements" tools and the radius was calculated
560 automatically. To measure angles between adjacent Fcs, Tp sequences were removed from
561 models. An axis was defined using mass weighting and helical correction for each Fc monomer
562 and then angles between the defined axes were measured and defined as the angles between
563 adjacent Fcs. To measure the diameter of the solvent accessible hole at the center of tFcμ
564 tetramer, tFcμ model was imported into UCSF Chimera[44]. Nine atom pairs located on the surface
565 were chosen around the hole. Distance between two atoms in each pair was measured and the
566 diameter of this hole was determined to be the average of these distances. Chosen atom pairs
567 and the distance measurements from each pair are summarized in Supplementary Table 3.

568 To measure the angle between the two tFcμ-Tp β-sheets, the tFcμ tetramer structure was
569 imported into PyMOL[50]. The angle between $Tp^{tFc\mu 1A}$ and $Tp^{tFc\mu 3E}$ was measured as the dihedral
570 angle between Cα of $Leu434^{tFc\mu 1A}$, $Asn440^{tFc\mu 1A}$, $Asn440^{tFc\mu 3E}$, and $Leu434^{tFc\mu 3E}$; angles between
571 $Tp^{tFc\mu 1B}$ and $Tp^{tFc\mu 3F}$, $Tp^{tFc\mu 2C}$ and $Tp^{tFc\mu 4G}$, $Tp^{tFc\mu 2D}$ and $Tp^{tFc\mu 4H}$ were measured using the same
572 method. The angle between two Tp β-sheets was determined to be the average of these
573 measurements. To measure the angle between the two hFcμ-Tp β-sheets the hSIgM-Fc core
574 structure (PDB ID: 6KXS) was imported into PyMOL[50]. The angle between $Tp^{Fc\mu 1A}$ and $Tp^{Fc\mu 5L}$
575 was measured as the dihedral angle between Cα of $Tyr562^{Fc\mu 1A}$, $Val567^{Fc\mu 1A}$, $Val567^{Fc\mu 5L}$, and
576 $Tyr562^{Fc\mu 5L}$; angles between $Tp^{Fc\mu 1B}$ and $Tp^{Fc\mu 5K}$, $Tp^{Fc\mu 2C}$ and $Tp^{Fc\mu 4H}$, $Tp^{Fc\mu 2D}$ and $Tp^{Fc\mu 4G}$, $Tp^{Fc\mu 3E}$
577 and $Tp^{Fc\mu 3F}$ were measured using the same method. An average of these measurements was
578 taken and the angle between two Tp β-sheets was determined to be the supplement angle of the
579 average.

580 To measure the angle between tFcμ CH3 and CH4, the tFcμ tetramer model was imported
581 into PyMOL[50]. For each copy of heavy chain, the angle between tFcμ CH3 and tFcμ CH4 was

582 measured as the dihedral angle between the Cα atoms of His298, Lys238, Glu412, and Thr332.
583 The average of eight angle measurements was reported as the final value of tFcμ CH3-CH4 angle.
584
585 **Figures**
586 Structural figures were made using UCSF Chimera, PyMOL[44,50]. SEC data and SEC-MALS data
587 were replotted using Origin. All figures were assembled using Adobe Photoshop.
588
589 **DATA AVAILABILITY.**
590 Cryo-EM density map has been deposited in the EM databank (www.ebi.ac.uk/emdb) with the
591 accession code EMD-40054 and the refined coordinate has been deposited in the Protein Data
592 Bank (www.rcsb.org) with accession code 8GHZ.
593
594 **AUTHOR CONTRIBUTIONS.**
595 The study was conceived by B.M.S and M.L.; experiments were conducted by M.L. and A.G.M.;
596 all authors contributed to data analysis and manuscript writing.
597
598 **ACKNOWLEDGEMENTS.**
599 We thank Kushol Gupta (University of Pennsylvania) for assistance with SEC-MALS data
600 collection and analysis and thank members of the Stadtmueller Laboratory for insightful
601 conversations and suggestions related to this work. Cryogenic electron microscopy was done in
602 the Beckman Institute cryo-EM resource center at California Institute of Technology. This work
603 was supported by NIH grant 1R01AI165570 and University of Illinois start-up funding to B.M.S.
604
605
606
607

608 **COMPETING INTERESTS.**

609 BMS and ML are listed as inventors on a draft provisional patent application that includes the
610 design, production and use of chimeric antibodies, some of which include teleost Ig heavy chain
611 motifs.

612

613 **MATERIALS AND CORRESPONDENCE.**

614 Correspondence and requests for materials should be addressed to Beth Stadtmueller,
615 bethms@illinois.edu.

616

617 **REFERENCES**

618 1. Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic
619 events and selective pressures. *Nat Rev Genet* **11**, 47–59 (2010).
620 2. Chapuis, R. M. & Koshland, M. E. Linkage and assembly of polymeric IgA immunoglobulins.
621 *Biochemistry* **14**, 1320–1326 (1975).
622 3. Wiersma, E. J. & Shulman, M. J. Assembly of IgM. Role of disulfide bonding and
623 noncovalent interactions. *J Immunol* **154**, 5265–5272 (1995).
624 4. Saito, S. *et al.* IgA tetramerization improves target breadth but not peak potency of
625 functionality of anti-influenza virus broadly neutralizing antibody. *PLoS Pathog* **15**,
626 e1007427 (2019).
627 5. Suzuki, T. *et al.* Relationship of the quaternary structure of human secretory IgA to
628 neutralization of influenza virus. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7809–7814 (2015).
629 6. Terauchi, Y. *et al.* IgA polymerization contributes to efficient virus neutralization on human
630 upper respiratory mucosa after intranasal inactivated influenza vaccine administration.
631 *Human Vaccines & Immunotherapeutics* **14**, 1351–1361 (2018).
632 7. Davis, A. C., Roux, K. H. & Shulman, M. J. On the structure of polymeric IgM. *Eur. J.*
633 *Immunol.* **18**, 1001–1008 (1988).
634 8. Eskeland, T. & Christensen, T. B. IgM Molecules with and without J Chain in Serum and
635 after Purification, Studied by Ultra-centrifugation, Electrophoresis, and Electron Microscopy.
636 *Scand J Immunol* **4**, 217–228 (1975).
637 9. Kaetzel, C. S. The polymeric immunoglobulin receptor: bridging innate and adaptive
638 immune responses at mucosal surfaces. *Immunol Rev* **206**, 83–99 (2005).

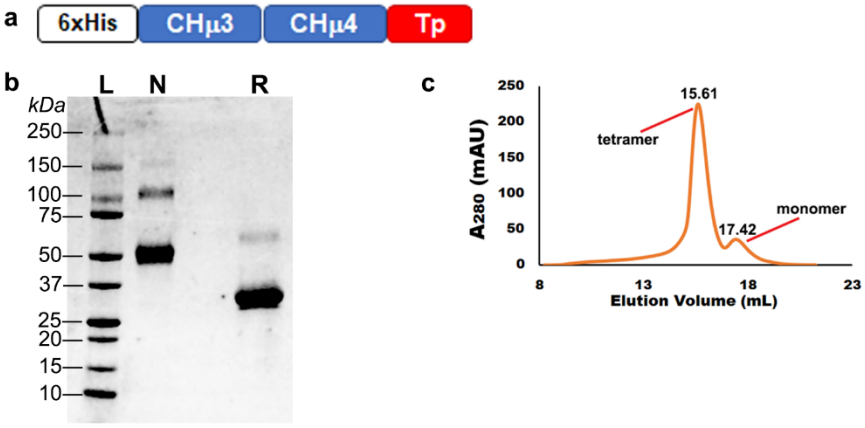
639 10. Macpherson, A. J., Yilmaz, B., Limenitakis, J. P. & Ganal-Vonarburg, S. C. IgA Function in
640 Relation to the Intestinal Microbiota. *Annu. Rev. Immunol.* **36**, 359–381 (2018).
641 11. Keyt, B. A., Baliga, R., Sinclair, A. M., Carroll, S. F. & Peterson, M. S. Structure, Function,
642 and Therapeutic Use of IgM Antibodies. *Antibodies* **9**, 53 (2020).
643 12. Pabst, O. New concepts in the generation and functions of IgA. *Nat Rev Immunol* **12**, 821–
644 832 (2012).
645 13. Kaattari, S. *et al.* Varied redox forms of teleost IgM: an alternative to isotypic diversity?
646 *Immunol Rev* **166**, 133–142 (1998).
647 14. Zhang, Y.-A. *et al.* IgT, a primitive immunoglobulin class specialized in mucosal immunity.
648 *Nat Immunol* **11**, 827–835 (2010).
649 15. Stadtmueller, B. M. *et al.* The structure and dynamics of secretory component and its
650 interactions with polymeric immunoglobulins. *eLife* **5**, e10640 (2016).
651 16. Hamuro, K., Suetake, H., Saha, N. R., Kikuchi, K. & Suzuki, Y. A Teleost Polymeric Ig
652 Receptor Exhibiting Two Ig-Like Domains Transports Tetrameric IgM into the Skin. *J*
653 *Immunol* **178**, 5682–5689 (2007).
654 17. Salinas, I., Fernández-Montero, Á., Ding, Y. & Sunyer, J. O. Mucosal immunoglobulins of
655 teleost fish: A decade of advances. *Developmental & Comparative Immunology* **121**,
656 104079 (2021).
657 18. Kumar Bharathkar, S. *et al.* The structures of secretory and dimeric immunoglobulin A. *eLife*
658 **9**, e56098 (2020).
659 19. Kumar, N., Arthur, C. P., Ciferri, C. & Matsumoto, M. L. Structure of the secretory
660 immunoglobulin A core. *Science* **367**, 1008–1014 (2020).
661 20. Li, Y. *et al.* Structural insights into immunoglobulin M. *Science* **367**, 1014–1017 (2020).
662 21. Kumar, N., Arthur, C. P., Ciferri, C. & Matsumoto, M. L. Structure of the human secretory
663 immunoglobulin M core. *Structure* **29**, 564-571.e3 (2021).
664 22. Ahmed, A. A., Keremane, S. R., Vielmetter, J. & Bjorkman, P. J. Structural characterization
665 of GASDALIE Fc bound to the activating Fc receptor FcγRIIIa. *Journal of Structural Biology*
666 **194**, 78–89 (2016).
667 23. Wurzburg, B. A., Garman, S. C. & Jardetzky, T. S. Structure of the Human IgE-Fc Cε3-Cε4
668 Reveals Conformational Flexibility in the Antibody Effector Domains. *Immunity* **13**, 375–385
669 (2000).
670 24. Lalla, C. de, Fagioli, C., Cessi, F. S., Smilovich, D. & Sitia, R. Biogenesis and function of
671 IgM: the role of the conserved μ-chain tailpiece glycans. *Molecular Immunology* **35**, 837–
672 845 (1998).

673 25. Pasalic, D. *et al.* A peptide extension dictates IgM assembly. *Proc Natl Acad Sci USA* **114**,
674 E8575–E8584 (2017).
675 26. Wyatt, P. J. Light scattering and the absolute characterization of macromolecules. *Analytica*
676 *Chimica Acta* **272**, 1–40 (1993).
677 27. Davis, A. C. & Shulman, M. J. IgM - molecular requirements for its assembly and function.
678 *Immunology Today* **10**, 118–128 (1989).
679 28. HADJIAZIMI, I. & MICHEAHAMZEHPOUR, M. XENOPUS-LAEVIS 19S
680 IMMUNOGLOBULIN - ULTRASTRUCTURE AND J-CHAIN ISOLATION. *IMMUNOLOGY*
681 **30**, 587–591 (1976).
682 29. Hohman, V. S. *et al.* J Chain in the Nurse Shark: Implications for Function in a Lower
683 Vertebrate. *J Immunol* **170**, 6016–6023 (2003).
684 30. Hiramoto, E. *et al.* The IgM pentamer is an asymmetric pentagon with an open groove that
685 binds the AIM protein. *Sci. Adv.* **4**, eaau1199 (2018).
686 31. Muller, R. *et al.* High-resolution structures of the IgM Fc domains reveal principles of its
687 hexamer formation. *Proceedings of the National Academy of Sciences* **110**, 10183–10188
688 (2013).
689 32. Anelli, T. *et al.* Sequential steps and checkpoints in the early exocytic compartment during
690 secretory IgM biogenesis. *EMBO J* **26**, 4177–4188 (2007).
691 33. van Anken, E. *et al.* Efficient IgM assembly and secretion require the plasma cell induced
692 endoplasmic reticulum protein pERp1. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17019–17024
693 (2009).
694 34. Nayak, D. K., Tang, A., Wilson, M., Miller, N. W. & Bengtén, E. Channel catfish soluble
695 FcμR binds conserved linear epitopes present on Cμ3 and Cμ4. *Molecular Immunology* **47**,
696 1306–1316 (2010).
697 35. Sheng, X., Qian, X., Tang, X., Xing, J. & Zhan, W. Polymeric Immunoglobulin Receptor
698 Mediates Immune Excretion of Mucosal IgM–Antigen Complexes Across Intestinal
699 Epithelium in Flounder (Paralichthys olivaceus). *Front. Immunol.* **9**, 1562 (2018).
700 36. Ye, J., Kaattari, I. M., Ma, C. & Kaattari, S. The teleost humoral immune response. *Fish &*
701 *Shellfish Immunology* **35**, 1719–1728 (2013).
702 37. Chen, Q., Menon, R., Calder, L. J., Tolar, P. & Rosenthal, P. B. Cryomicroscopy reveals the
703 structural basis for a flexible hinge motion in the immunoglobulin M pentamer. *Nat Commun*
704 **13**, 6314 (2022).

38. 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. *Molecular Immunology* **47**, 738–746 (2010).
39. Feng, L.-N. *et al.* Molecular cloning and functional analysis of polymeric immunoglobulin receptor gene in orange-spotted grouper (Epinephelus coioides). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **154**, 282–289 (2009).
40. 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).
41. Rubinstein, J. L. & Brubaker, M. A. Alignment of cryo-EM movies of individual particles by optimization of image translations. *Journal of Structural Biology* **192**, 188–195 (2015).
42. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214–1221 (2020).
43. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296–W303 (2018).
44. Pettersen, E. F. *et al.* UCSF Chimera--A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
46. 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).
47. Barad, B. A. *et al.* EMRinger: side chain–directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* **12**, 943–946 (2015).
48. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
49. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320–W324 (2014).
50. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
51. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* **372**, 774–797 (2007).
52. Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* 310–322 (2002).

739 **SUPPLEMENTARY FIGURES AND TABLES**

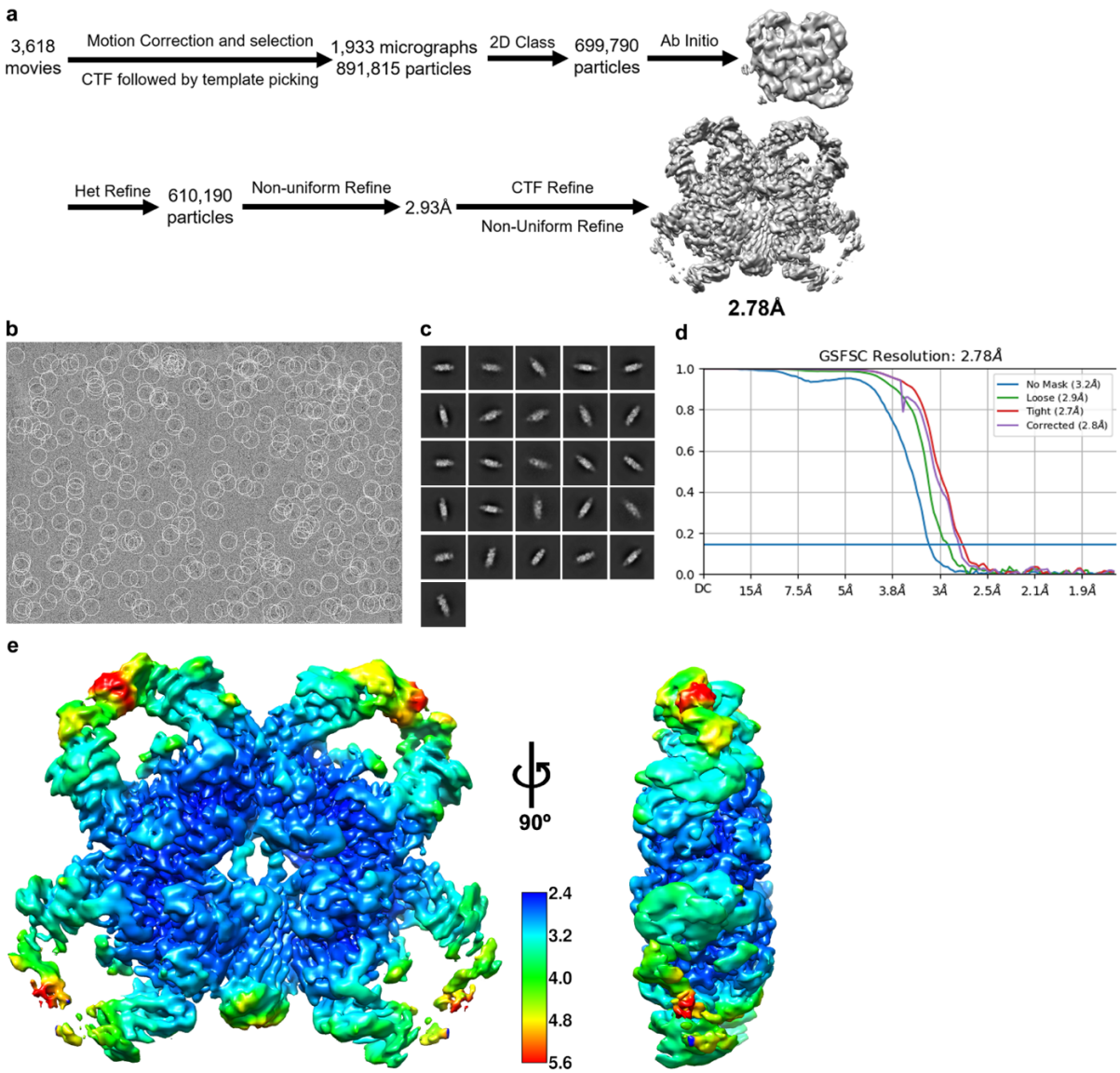
740



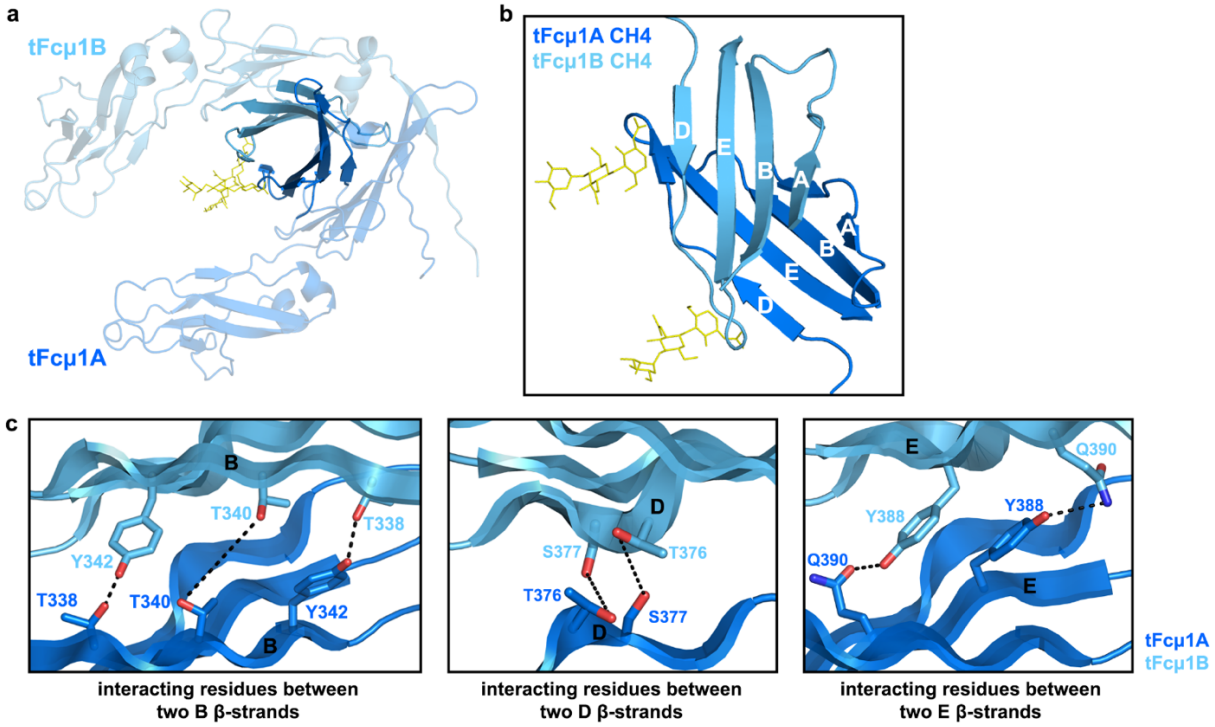
741

742 **Supplementary Fig. 1 | Design of tFcμ construct and characterization of recombinant tFcμ**
743 **expression. a** Schematic of the tFcμ expression construct, including an N-terminal signal peptide
744 and hexa-histidine tag. **b** SDS-PAGE gel analysis of recombinant tFcμ protein. *NR*, non-reducing
745 condition. *R*, reducing condition. The higher molecular weight band in R lane is attributed to
746 incomplete reduction of the tFcμ monomer. **c** Size exclusion column chromatogram of
747 recombinant tFcμ protein.

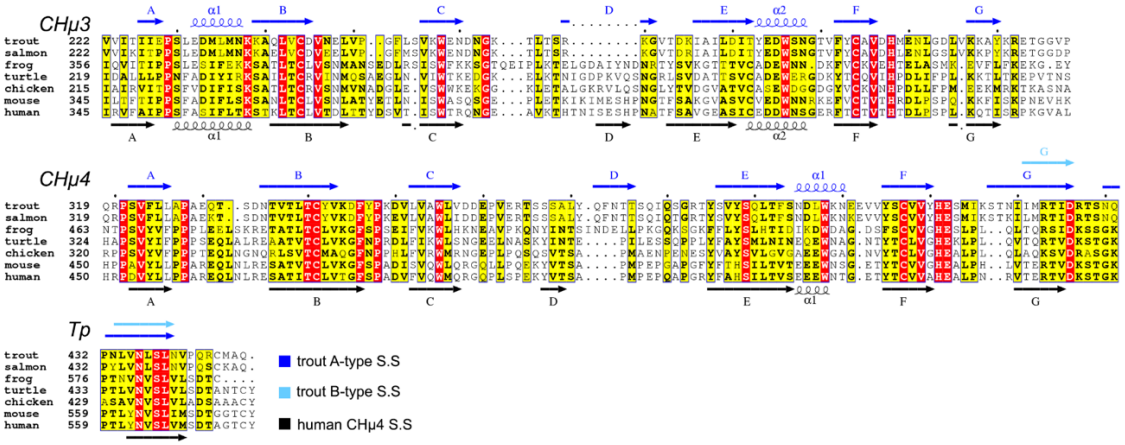
748



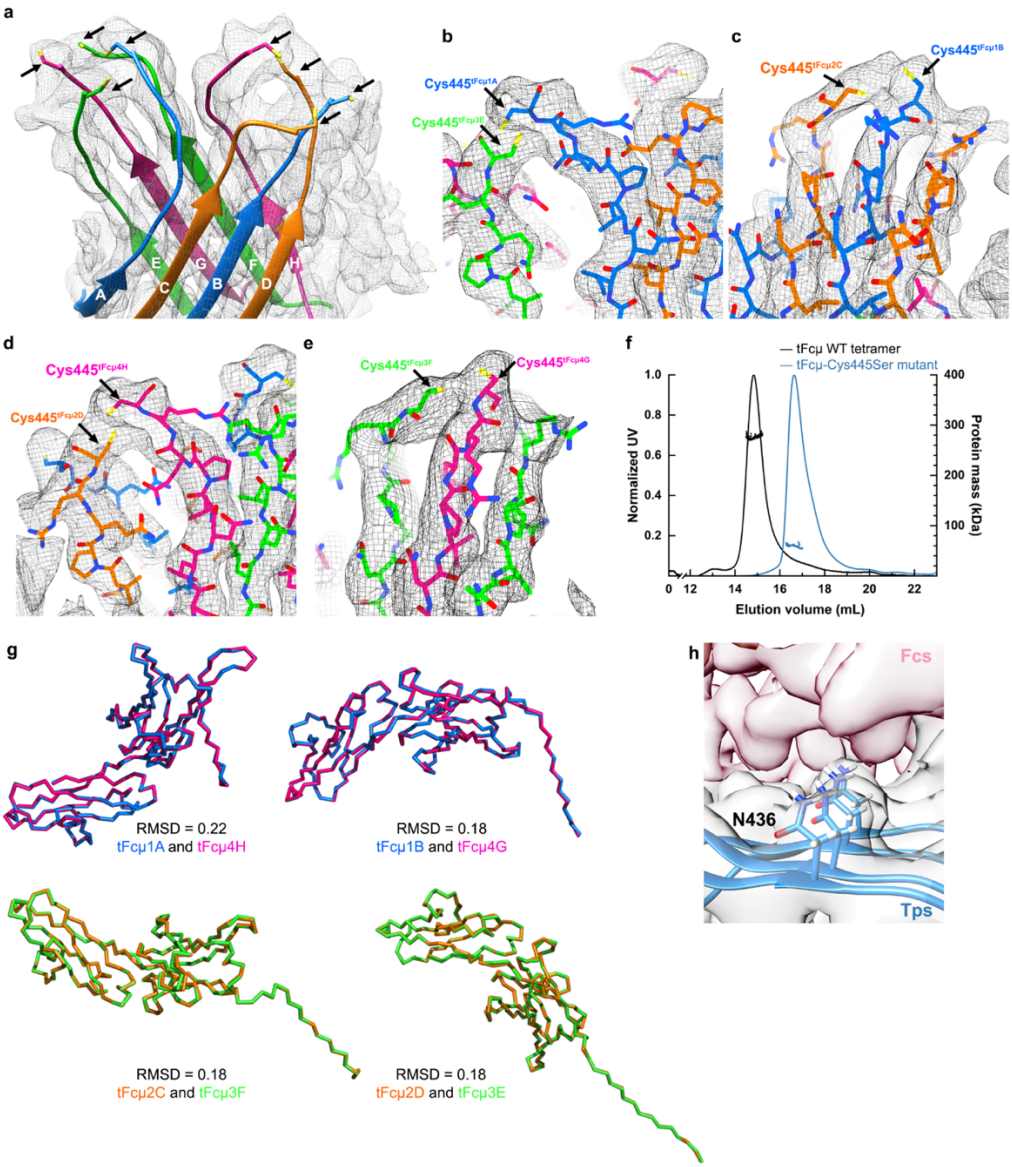
Supplementary Fig. 2 | tFcμ cryo-EM data collection, processing, and validation. a Schematic summary of tFcμ cryo-EM data processing pipeline in CryoSparc. **b** Representative micrograph after motion correction and selection. **c** Representative 2D class averages. **d** FSC curves for the final reconstruction with reported resolution at FSC = 0.143 shown by the blue horizontal line. **e** Local resolution map of the final reconstruction calculated in CryoSparc and rendered in UCSF Chimera; the unit of scale bar is given in Angstroms (Å).



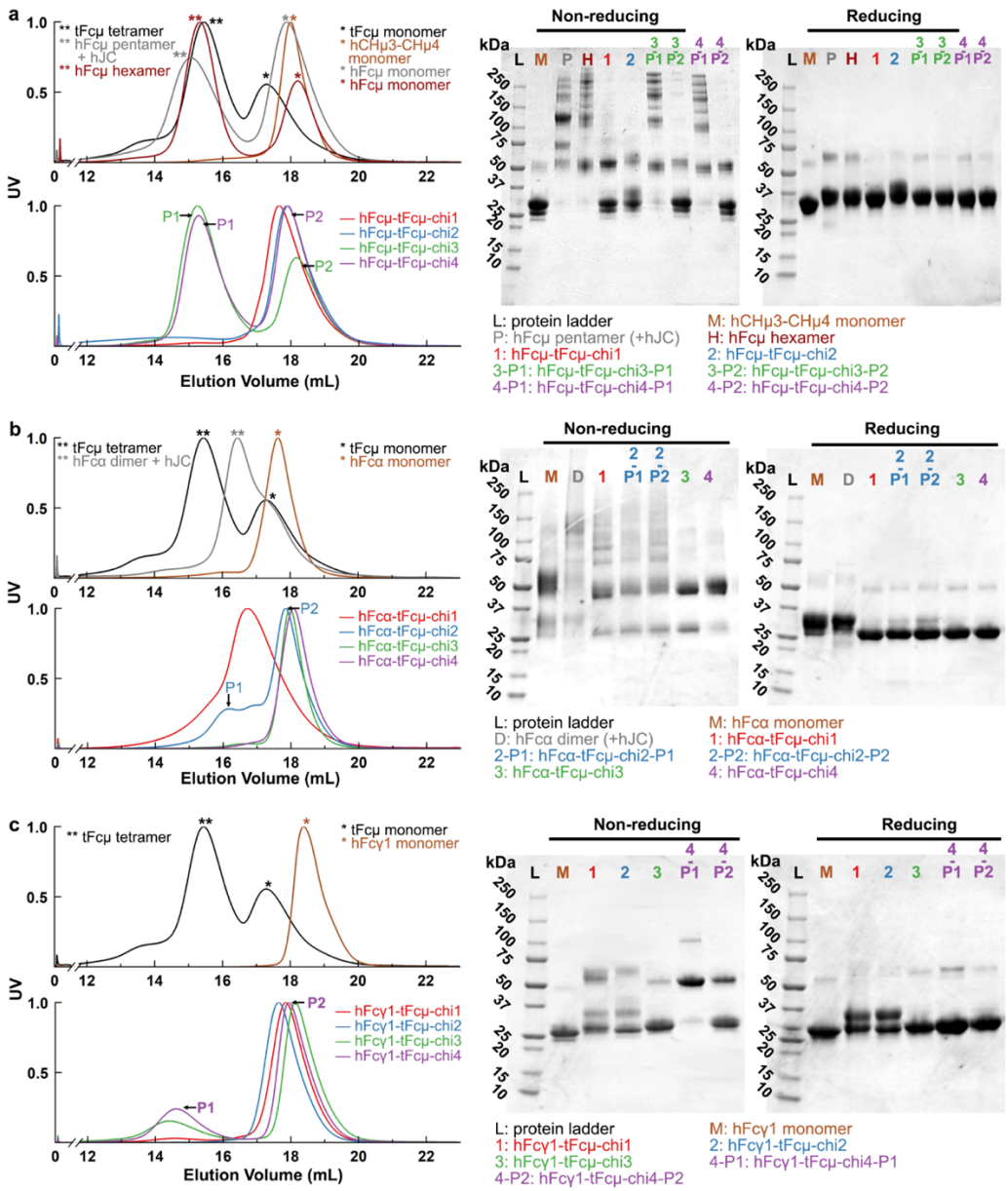
Supplementary Fig. 3 | Interface between two CHμ4 domains within Fcμ monomer. **a** Cartoon representation of tFcμ1 with glycans shown as yellow sticks. The intra-Fc tCHμ4-tCHμ4 interface is rendered using solid colors whereas the rest of the structure is rendered using transparent color. The orientation is the same as Fig. 1b. **b** Closeup view of intra-Fc tCHμ4-tCHμ4 interface between chain A and chain B with Ig-fold strand nomenclature labeled (i.e. A, B, E, D β-strands). **c** Interactions at intra-Fc tCHμ4-tCHμ4 interface between chain A and chain B. From left to right: interacting residues between two B, β-strands, two D, β-strands, and two E, β-strands.



Supplementary Fig. 4 | Sequence alignment of CHμ3, CHμ4, and Tp residues. Sequence alignment of CHμ3, CHμ4 and Tp residues from trout, salmon, frog, turtle, chicken mouse, and human. Secondary structure (S.S) information for CHμ3, CHμ4, and the Tp are annotated. For the B-type tCHμ4 register, elements identical to A-type are not included.



Supplementary Fig.5 | Local cryo-EM density and structural alignment of tFcμ chains. a-e Cryo-EM density near tFcμ C-termini. The coloring scheme is the same as Fig. 1b. **a** Overview of C-termini density with side chains of Cys445 shown as sticks. **b-e** Density around potential interchain disulfide bonds between Cys445$^{Fc\mu1A}$ and Cys445$^{Fc\mu3E}$, between Cys445$^{Fc\mu1B}$ and Cys445Fcμ2C, between Cys445Fcμ2D and Cys445$^{Fc\mu4H}$, and between Cys445$^{Fc\mu3F}$ and Cys445$^{Fc\mu4G}$, respectively. **f** SEC-MALS chromatograms for wildtype tFcμ tetramer and tFcμ-Cys445Ser monomer. 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). **g** Structural alignments between tFcμ chain A and chain H, chain B and chain G, chain C and chain F, chain D and chain E. **h** Density of tFcμ Tp assembly around residue Asn436. Cartoon representation of Tp$^{tFc\mu1A}$, Tp$^{tFc\mu1B}$, Tp$^{tFc\mu2C}$, and Tp$^{tFc\mu2D}$ are shown. Density associated with non-Tp residues is shaded red.



Supplementary Fig. S6 Purification of hFc-tFcμ chimeric Fcs | a-b The **left** graph in each panel includes SEC chromatographs of hFc-tFcμ chimeras and relevant controls obtained during purification and prior to SEC-MALS experiments. For controls, peaks containing presumptive monomers are indicated with a single asterisk and peaks containing higher order polymers are indicated with two asterisks. For chimeras with more than one peak, "P1" designates higher order polymers and "P2" designates presumptive monomers. For each sample, elution fractions located at the center of each peak were and run on the SDS-PAGE gels under non-reducing and reducing condition, the results of which are shown in the **right** graph in each panel. **a** hFcμ-tFcμ chimeras and relevant controls. **b** hFcα-tFcμ chimeras and relevant controls. **c** hFcγ1-tFcμ chimeras and relevant controls. The two bands observed for reduced samples hFcγ1-tFcμ-chi1 and hFcγ1-tFcμ-chi2 are attributed to heterogeneous glycosylation.

797 **Supplementary Table 1. Cryo-EM data collection and refinement statistics associated**
798 **with tFcμ**

Accession codes		**Model**	
EM Databank	EMD-40054	Refinement software	PHENIX[46,47] and Coot[45]
Protein Databank	8GHZ	Validation software	PHENIX[46,47]
		Chains	16
Data Collection		Total atoms	28,463
Magnification	105,000	Non-hydrogen atoms	14,390
Pixel size (Å)	0.427	Residues	1,768
Frames per movie	40	Bond RMSD	
Movies selected	3,618	Length (Å)	0.004
Defocus range (μm)	0.5-5	Angle (°)	1.036
Dosage (e/Å^2)	60	MolProbity score	1.95
Initial particles	1,928,186	Clash score	6.95
Final particles	610,190	Ramachandran plot (%)	
		Favored	91.89
Resolution estimate (unmasked data)		Allowed	8.11
Software	PHENIX[46,47]	Outliers	0
FSC (model)=0.143 (Å)	2.79	Rotamer outliers (%)	1.28
FSC (half map 1,2)=0.143 (Å)	2.86	Cβ outliers (%)	0
		Peptide plane (%)	
Density and Fit		Cis proline	0
Model/Data CC		Twisted proline	0
Mask	0.79	CaBLAM outliers (%)	3.57
Box	0.78	EMRinger score	3.09
Peaks	0.65		
Volume	0.78		

799

800

801 **Supplementary Table 2. tFcμ and hFcμ PISA calculations**

Chain 1	Chain 2	Interface Area (Å^2)	Interface ∆G (kcal/mol)	Interface ∆G P -value
Interfaces between teleost IgM CH4 domain				
A	B	917.5	-7.54	0.36
C	D	928.5	-7.19	0.31
E	F	904.3	-9.60	0.19
G	H	914.9	-8.60	0.31
B	C	856.8	-2.73	0.61
D	E	729.4	-3.60	0.43
F	G	812.2	-3.61	0.55
Interfaces between human IgM CH4 domain				
A	B	1216.5	-18.92	0.06
C	D	1167.7	-17.20	0.11
E	F	1159.7	-20.98	0.04
G	H	1194.1	-14.29	0.22
K	L	1182.8	-17.45	0.12
B	C	447.5	-0.24	0.70
D	E	400.9	-0.75	0.66
F	G	372.7	-2.54	0.50
H	K	382.2	-2.40	0.55

802

803

804 **Supplementary Table 3. Diameter measurements for the solvent accessible hole at the**
805 **center of tFcμ structure**

Atom 1	Atom 2	Distance
Thr428$^{Fc\mu 2D}$/CB	Asn433$^{Fc\mu 3E}$/HB2	12.07Å
Thr428$^{Fc\mu 3E}$/CB	Asn433$^{Fc\mu 2D}$/HB3	12.89Å
Asn430$^{Fc\mu 2D}$/CA	Pro432$^{Fc\mu 3E}$/C	12.36Å
Ser429$^{Fc\mu 2D}$/CB	Asn433$^{Fc\mu 4G}$/HB2	12.11Å
Thr428$^{Fc\mu 3E}$/O	Asn433$^{Fc\mu 2D}$/HB3	11.47Å
Asn430$^{Fc\mu 3E}$/CA	Pro432$^{Fc\mu 2D}$/C	13.52Å
Asn430$^{Fc\mu 3E}$/H	Asn433$^{Fc\mu 2D}$/CB	11.99Å
Asn433$^{Fc\mu 3E}$/N	Ser429$^{Fc\mu 2D}$/O	11.15Å
Asn433$^{Fc\mu 3E}$/CB	Asn430$^{Fc\mu 2D}$/H	12.40Å
		Avg.: 12.2Å

806

807

808 **Supplementary Table 4. MALS data summary**

Sample description	Protein MW calculated from sequence (kDa)	Average protein MW from MALS data (Da)	TOTAL number of PNGS* (prediction potential**> 0.7)	TOTAL number of PNGS (prediction potential > 0.5)	Number of PNGS on each HC: (prediction potential > 0.7)	Number of PNGS on each HC: (prediction potential > 0.5)
hCHμ3-CHμ4	49.6	4.338×10^4 (±0.543%)	0	4	0	2
hFcμ pentamer +	285.1	2.959×10^5 (±0.106%)	11	31	1	3
hFcμ hexamer	340.4	3.156×10^5 (±0.073%)	12	36	1	3
hFcμ-tFcμ-chi1	54.0	3.859×10^4 (±0.499%)	0	6	0	3
hFcμ-tFcμ-chi2	54.5	3.310×10^4 (±3.953%)	0	6	0	3
hFcμ-tFcμ-chi3-P1	267.5	2.708×10^5 (±0.077%)	0	24	0	3
hFcμ-tFcμ-chi3-P2	53.5	3.748×10^4 (±0.387%)	0	6	0	3
hFcμ-tFcμ-chi4-P1	214.3	2.725×10^5 (±0.076%)	0	24	0	3
hFcμ-tFcμ-chi4-P2	53.6	4.283×10^4 (±0.388%)	0	6	0	3
hFcα dimer +hJC	67.5	1.377×10^5 (±0.180%)	9	9	1	2
hFcα monomer	25.9	6.117×10^4 (±1.187%)	4	4	1	2
hFcα-tFcμ-chi1	52.7	1.054×10^5 (±1.097%)	4	8	1	2
hFcα-tFcμ-chi2-P1	210.4	2.161×10^5 (±1.352%)	8	16	1	2
hFcα-tFcμ-chi2-P2	52.6	6.119×10^4 (±8.826%)	2	4	1	2
hFcα-tFcμ-chi3	52.1	4.519×10^4 (±16.598%)	2	4	1	2
hFcα-tFcμ-chi4	52.1	4.888×10^4 (±2.909%)	2	4	1	2
hFcγ1	49.6	4.723×10^4 (±0.996%)	2	2	1	1
hFcγ1-tFcμ-chi1	54.0	5.217×10^4 (±3.975%)	2	4	1	2
hFcγ1-tFcμ-chi2	54.5	5.598×10^4 (±5.051%)	2	4	1	2
hFcγ1-tFcμ-chi3	53.5	5.481×10^4 (±0.419%)	2	4	1	2
hFcγ1-tFcμ-chi4-P1	106.7	4.837×10^5 (±0.091%)	16	32	1	2
hFcγ1-tFcμ-chi4-P2	53.4	4.917×10^4 (±0.381%)	2	4	1	2

809 **PNGS stands for "potential N-linked glycosylation site".*
810 ***Prediction potential is calculated using NetNGlyc 1.0 developed by DTU Health Tech[52].*

811

812