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 pIgs 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]. Only a subset of Ig heavy chains are assembled into pIgs, 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 pIgs, 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[4–6].

In mammals, 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[7–9]. Polymeric forms of IgM are typically pentameric, containing five IgM monomers and one JC, however hexamers lacking a JC have been identified in serum[10–12]. 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 basolateral surface of epithelial cells, which binds to JC-containing pIgA and pIgM and transports them to the apical surface of the cell[13]. 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[14]. 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[15,16].

While mammalian pIgs have been studied extensively, less is known about pIgs from other jawed vertebrates. Of particular interest are pIgs 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 tIgT and tIgM that are presumed to contain four copies of each respective monomer[6,17,18]. 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[18]. Both tIgT and tIgM can be transported to the mucosa by tpIgR, which has a distinct domain organization and structure compared to mammalian counterparts and does not require a JC to bind tIgT or tIgM[19]. 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[18,20]. Similar to mammalian pIgs, tpIgs have been associated with complement activation and bacterial coating[21].

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[22–24]. 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

82 pIg complexes that otherwise contain multiple heavy chains with identical sequence. Given the
83 critical role for the JC in mammalian pIg structures, we hypothesized that tpIgs lacking JC would
84 assemble using different mechanisms and adopt distinct structures. To test this hypothesis and
85 to explore the evolution and diversity of vertebrate pIg structure and function, we determined the
86 cryo-EM structure of the tIgM Fc core at 2.78Å resolution. Together with comparative and
87 mutational analysis, this structure provides a first glimpse of both a teleost pIg and a pIg lacking
88 a JC, revealing distinct modes of assembly compared to mammalian pIgs and suggesting
89 alternative mechanisms for interacting with antigens and FcRs, including tSC.
90

91 **RESULTS**

92 **Overall structure of tetrameric teleost IgM core**

93 Guided by previous structural studies indicating that the human IgM constant domain CH2 and
94 Fabs are flexible and poorly resolved in cryo-EM maps[24,25], we expressed and purified a truncated
95 teleost IgM heavy chain consisting of CHμ3, CHμ4, and the Tp (Fig. 1a, Supplementary Fig. 1).
96 We utilized this tIgM Fc (tFcμ) to determine a cryo-EM structure to an average resolution of 2.78Å
97 (Fig. 1, Supplementary Fig. 2). In the structure, most main chain and side chain atoms in CHμ4
98 domains and the Tps were well resolved (resolution range 2.4-3.0Å). Map resolution was ~3.2Å
99 and above for CHμ3 allowing the positions of main chain atoms and ~50-75% of side chain atoms
100 in each individual chain to be refined; other side chains were built in geometrically reasonable
101 positions (Supplementary Fig. 2).

102 The refined tFcμ structure contains four tFcμ monomers (Fcμ1 to Fcμ4) that lay
103 approximately flat on one plane and include interfaces between Fcμ1 and Fcμ2, Fcμ2 and Fcμ3,
104 Fcμ3 and Fcμ4. Fcμ1 and Fcμ4 are separated by Tp residues, which form a β-sandwich (hereafter
105 termed *Tp assembly*) comprising two β-strands from each of the four Fcμs (Fig. 1). Overall, the
106 tetramer is ~164Å across in the largest dimension and can be described as two structurally and
107 symmetry related halves, one containing Fcμ1-Fcμ2 and associated Tps, and the other containing

108 Fcμ3-Fcμ4 and associated Tps; at the center of the tetramer is a solvent accessible hole with a
109 diameter of ~12.2 Å. Despite similarity between the two halves of the tetramer, the complex is not
110 symmetrical with respect to all four Fcs; rather, adjacent tFcs are related by distinct angles
111 measuring 81°, 72°, 86°, for Fcμ1 and Fcμ2, Fcμ2 and Fcμ3, Fcμ3 and Fcμ4, respectively. Fcμ1
112 and Fcμ4 are related by an angle of 121° and contact the Tp assembly (Fig. 1b, c).

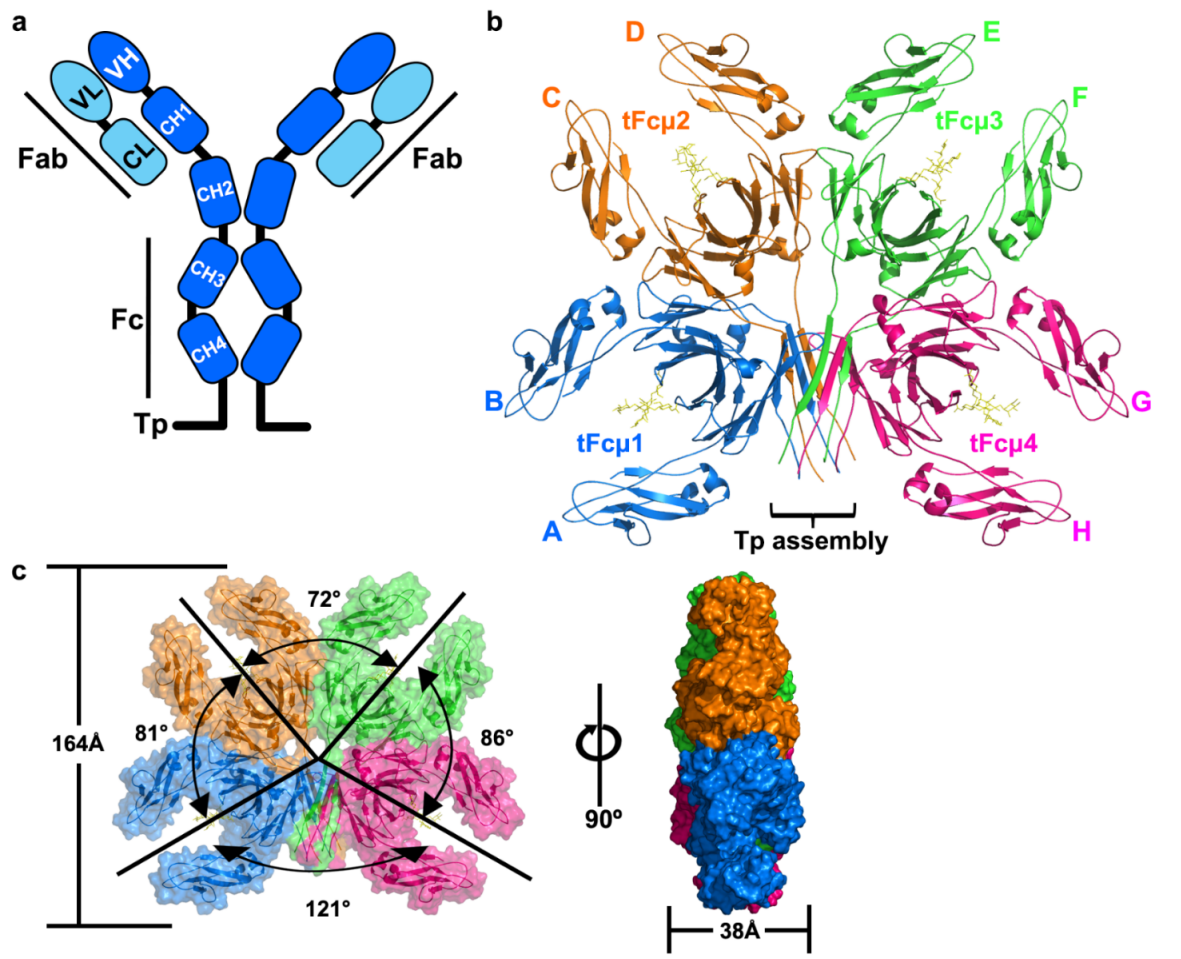


Fig. 1 | Overall structure of teleost IgM-Fc core, tFcμ. a Diagram of tIgM monomer. Constant (C) domains are represented as round corner rectangles and variable (V) domains are represented as ellipses. The heavy (H) is colored marine blue and the light (L) chain is colored light blue. **b** Cryo-EM structure 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.

113

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

116 Given the lack of published structural data for a teleost antibody, we first analyzed the structures
117 of tFcμ monomers and their individual domains, CHμ3 and CHμ4. The tFcμ monomers comprise
118 two copies of CHμ3 and CHμ4, which together form a canonical Fc characterized by an angle of
119 ~87° between CHμ3 and CHμ4 on the same chain and interfaces between adjacent CHμ3 and
120 adjacent CHμ4 domains on different chains (Fig. 2a, Supplementary Fig. 3a). Each domain adopts
121 an Ig-constant fold consisting of two β-sheets linked by a disulfide bond, one sheet containing
122 four β-strands termed A, B, E, and D, and the other sheet containing three β-strands termed C,
123 F, and G; the Tp β-strand follows the CHμ4 G strand. In one chain of each Fcμ monomer (A, D,
124 E, and H) the A strand extends several residues, which we have termed A'. Ordered
125 carbohydrates were apparent on CHμ4 residue Asn 374. Within each tFcμ, two CHμ4 domains
126 interact primarily through residues located in the AA'BED β-sheets, which mediate the interactions
127 between the C-terminal Ig domain in most Fcs[22–24,26,27]; the interface (including CHμ3, CHμ4 and
128 Tp) is ~924Å^2 and includes hydrogen bonds and electrostatic interactions (Supplementary Fig.
129 3b,c).

130 Remarkably, despite sharing identical sequence and adopting Ig constant folds, the eight
131 copies of CHμ4 exhibit structural differences that stabilize the tFcμ tetramer. Specifically, the C-
132 terminal 38 residues (residues 411-448; FG loop, G strand and Tp) adopt one of two folds, with
133 chains A and H adopting an "A-type" fold and chains B-G adopting a "B-type" fold (Fig. 2b, c). In
134 the A-type fold, the FG loop and the G strand include residues 411-416 and 417-426, respectively,
135 and the Tp folds back to form β-sheet interactions with the G strand (Fig. 2b). In the B-type fold,
136 the FG loop and the G strand instead include residues 411-420 and 421-426, respectively,
137 resulting in Tp β-strand (residues 434-441) being directed away from CHμ4 where they form β-
138 sheet interactions with other Tp(s) rather than the G strand (Fig. 2c). Whereas chains A and H
139 are superimposable, the six B-type domains adopt one of three distinct conformations based on

140 the trajectory of the Tp and thus, each B-type domain is superimposable (RMSD of less than
141 0.18Å) with one other B-type domain on the opposing side of the tetramer (Supplementary Fig.
142 4a). In sum, the A-type CHμ4 domains can be described as having a C-F-G-Tp sheet whereas B-
143 type have a C-F-G sheet (with distinct F-G β-sheet interactions) and an extended Tp. Together
144 these two folds facilitate the incorporation of each chain into the Tp assembly (discussed below).

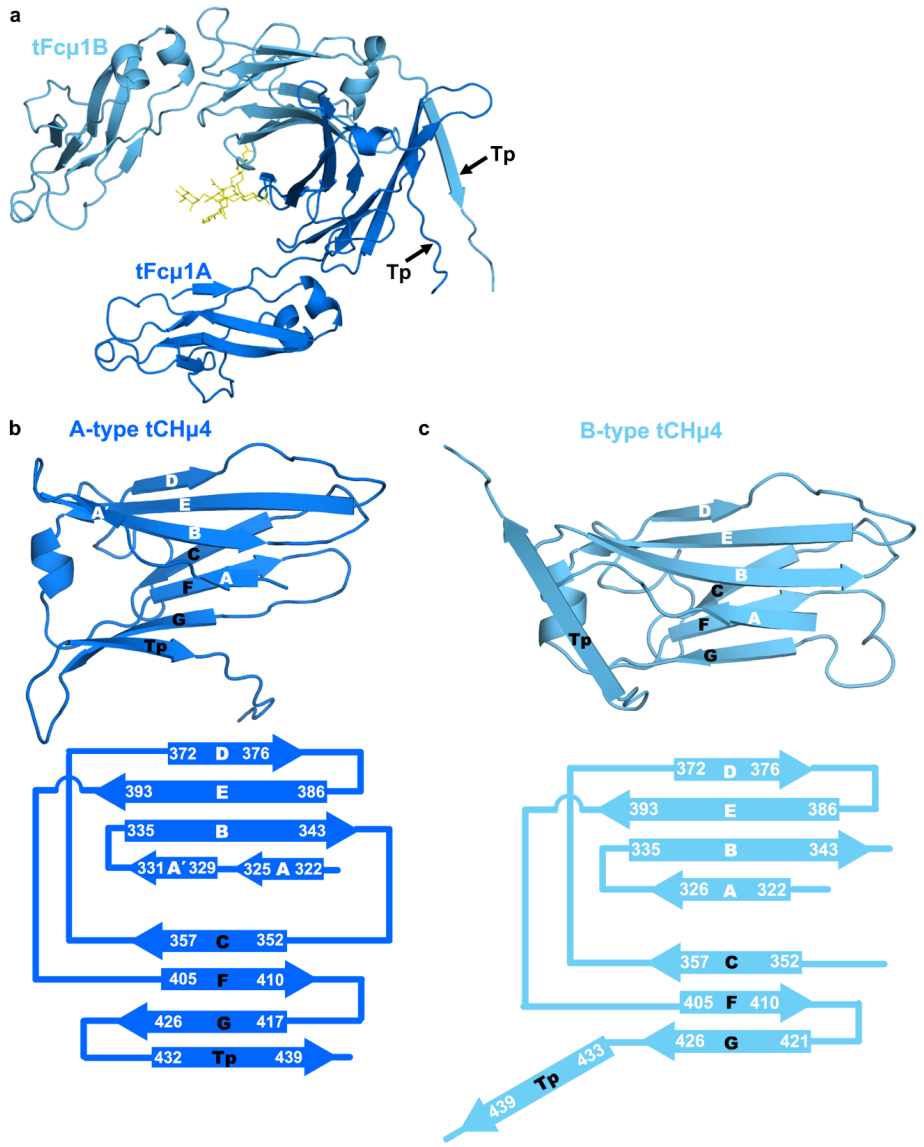


Fig. 2 | tIgM CH4 domains adopt one of two Ig folds. a Cartoon representation of tFcμ1 (chains A and B) with glycans shown in yellow; the orientation is the same as Fig. 1b. **b-c** A-type (panel **b**) and B-type (panel **c**) tCHμ4 folds. *Top*, cartoon representations of A-type tCHμ4 domain and B-type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

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

146 The tFcμ tetramer includes contacts between adjacent Fcμ monomers and Tps. Interactions
147 between adjacent Fcμ monomers (e.g., Fcμ1-Fcμ2, Fcμ2-Fcμ3, and Fcμ3-Fcμ4) are dominated
148 by a common set of residues in the CHμ4 CD loops, FG loops and G-strands, forming an interface
149 of ~1360Å^2 (Fig. 3a). Notable residues mediating these contacts include CHμ4 G strand residues
150 Met422, Thr424, and Arg 423, which form electrostatic interactions and hydrogen bonds with G
151 strand residues, CD loop (Asp359), and F strand (Ser406) residues in the neighboring CHμ4
152 domain. Additionally, FG loop residues Lys416 and Thr418 form interactions with CD loop Arg364
153 and F strand Tyr410, respectively (Fig. 3b, c).

154 Fcμ1 and Fcμ4 do not interface and instead interact with the Tp assembly, which is
155 situated between the two Fcμs (Fig. 1b and Fig. 4a). Notably, Fcμ1 and Fcμ4 each contain one
156 A-type CHμ4 domain and one B-type CHμ4 domain. The A-type domains neighbor the Tp
157 assembly and contribute uniquely to its structure. Specifically, because A-type Tp β-strands form
158 β-sheet interactions with CHμ4 G strands, residues in the pre-Tp loop (residues 434-441) are
159 positioned to contact both the top and bottom Tp β-sheets, effectively clamping the Fcμ1 and
160 Fcμ4 to the Tp assembly. Interactions between the pre-Tp loops and the rest of the Tp assembly
161 are dominated by polar contacts between A-type pre-Tp loop residues and B-type Tp residues
162 (Fig. 4d-f). Together Fc-Fc and Fc-Tp interactions form a band of contacts around the entire
163 tetramer, with A-type CHμ4 domains contributing directly to the Tp assembly (Fig. 4d-f) and B-
164 type CHμ4 domains forming interchain interactions between adjacent Fcμ monomers using CD
165 loops, FG loops and G-strands (Fig. 3b, c).

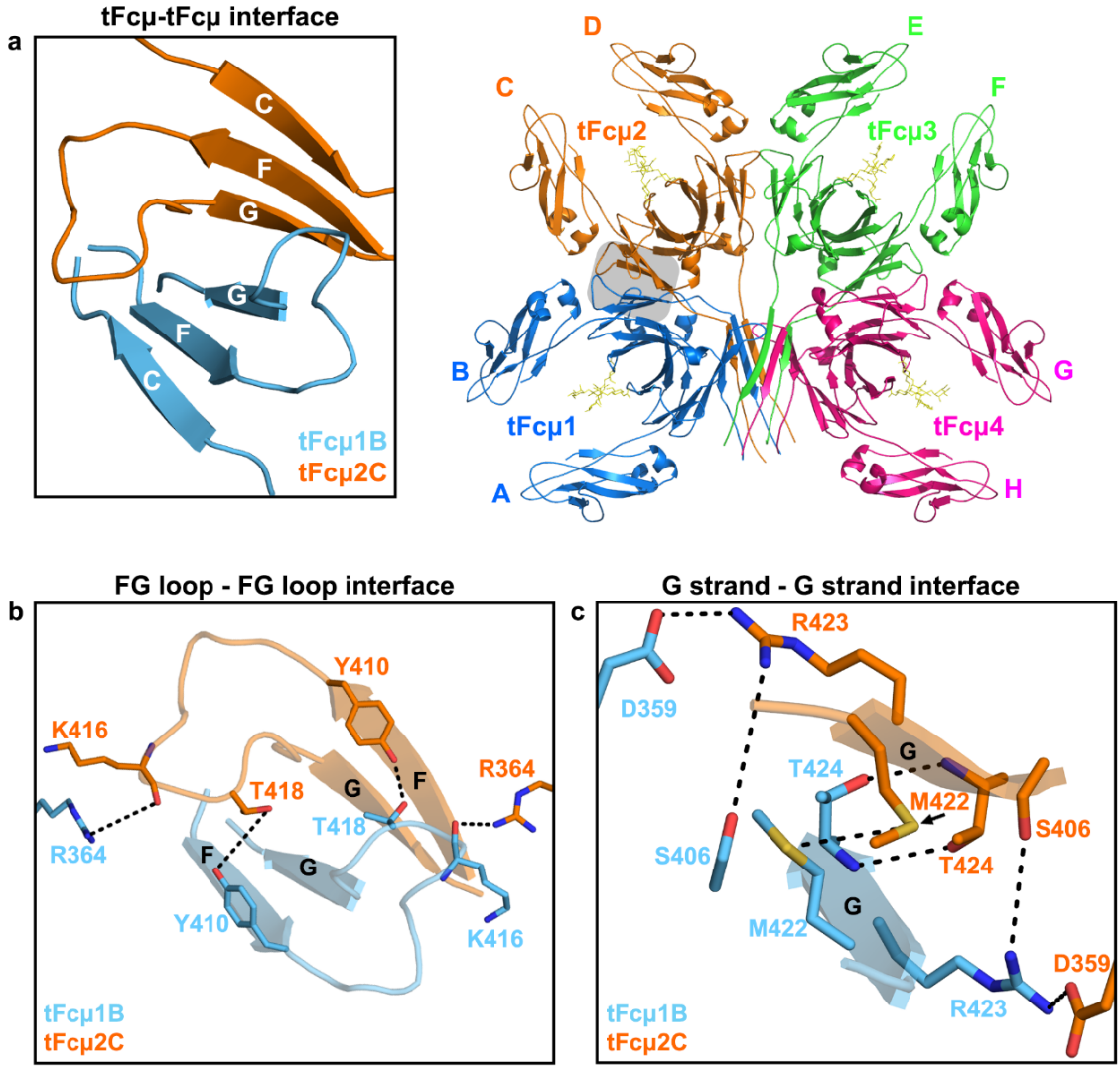


Fig. 3 | The CHμ4 interface between adjacent tFcμ monomers. a. Overview of interactions between tFcμ1B and tFcμ2C at the CHμ4 interface (shaded area in tFcμ tetramer). *Left*, open-book overview of tFcμ-tFcμ interface boxed in the right panel, but shown in a different orientation. Only CHμ4 C-F-G β-sheets and FG loops are shown. *Right*, tFcμ tetramer with tFcμ1B and tFcμ2C at the CHμ4 interface in shaded box. **b** 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.

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

Within the tFcμ tetramer, β-strand pairing in the Tp assembly tightly links individual chains and also promotes stabilizing interactions between the two halves of the tetramer (Fcμ1-Fcμ2 and

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

179 Loop residues 441-448 follow the Tp β-strands. These residues are poorly ordered in the
180 map; however, density suggests potential disulfide bonds between copies of the residue Cys445
181 (Supplementary Fig. 4b-f). Residue 445 is a conserved Cys involved in interchain disulfide
182 bonding in mammalian IgA and IgM[23,24]; despite low local resolution, our map suggests possible
183 disulfide bonding between $Cys445^{Fc\mu1A}$ and $Cys445^{Fc\mu3E}$, between $Cys445^{Fc\mu1B}$ and $Cys445^{Fc\mu2C}$,
184 $Cys445^{Fc\mu2D}$ and $Cys445^{Fc\mu4H}$, and $Cys445^{Fc\mu3F}$ and $Cys445^{Fc\mu4G}$ (see Supplementary Fig. 4b-f).
185 The possibility of inter-chain disulfide bonding is further supported by higher molecular weight
186 bands visible on non-reducing SDS-PAGE upon tFcμ purification (Supplementary Fig. 1b).
187 Because potential Cys445 disulfide bonds are located in the C-terminal loop region following the
188 Tp β-strands, they do not appear to play a direct role in stabilizing the interactions between
189 adjacent Fcs; nevertheless, they may effectively provide a covalent link between the two halves
190 of the tetramer by connecting the top and bottom sheets of the Tp assembly (Fig. 4a and
191 Supplementary Fig. 4b-f). Cysteine to serine mutation at this residue produced a tFcμ monomer
192 in solution and eliminated evidence of inter-chain disulfide bonding, further suggesting a critical
193 role in tIgM polymeric assembly (Supplementary Fig. 4g).

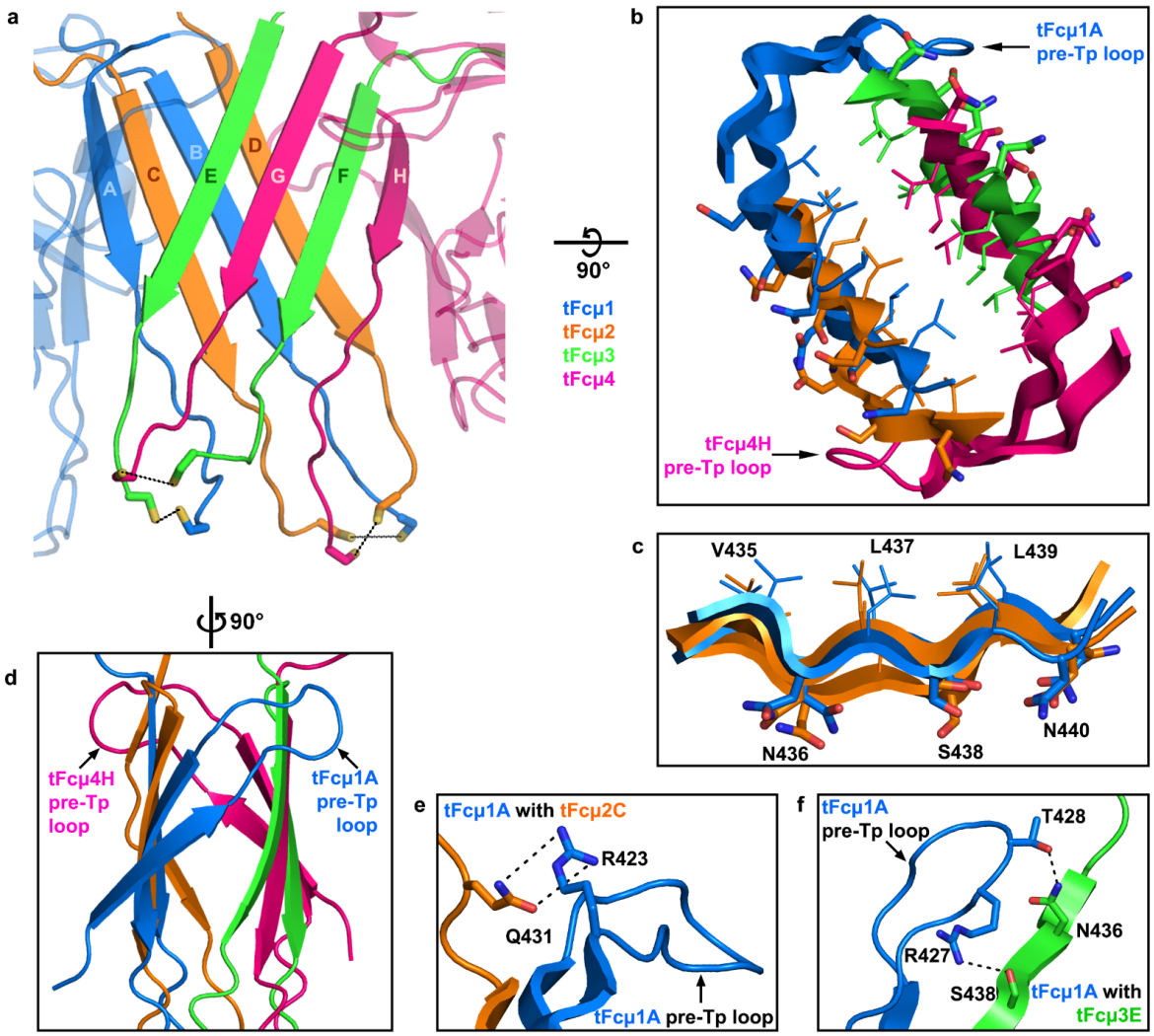


Fig. 4 | Teleost IgM Tp assembly and its interactions with pre-Tp loop of A-type CHμ4. **a** Cartoon representation of the tFcμ Tp assembly. The top β-sheet consists of $Tp^{tFc\mu3E}$, $Tp^{tFc\mu4G}$, $Tp^{tFc\mu3F}$, and $Tp^{tFc\mu4H}$ and bottom β-sheet consists of $Tp^{tFc\mu1A}$, $Tp^{tFc\mu2C}$, $Tp^{tFc\mu1B}$, and $Tp^{tFc\mu2D}$. Cys445 side chains are shown as sticks. Cys445 pairs that may form potential 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.

Comparisons between teleost and human IgM

IgM is considered to be one of the evolutionarily oldest[1]. It is omnipresent in all jawed vertebrates, except coelocanths where IgM has been lost, and teleost fish are the second oldest class

197 expressing IgM, after cartilaginous fish[1,28]. To gain insights on pIg evolution we compared teleost
198 and human (h) IgM structures[24,25]. Whereas tIgM is a tetramer, hIgM forms a JC-associated
199 pentamer. Available human IgM structures contain the Fc region (hFcμ), JC and human secretory
200 component (hSC) (Fig. 5a). Overall, tFcμ and hFcμ adopt structures characterized by adjacent
201 Fc monomers linked through Fcμ-Fcμ contacts and a Tp assembly; however, the structures are
202 not globally superimposable and the relative orientation of the Fcμs, residues involved in Fcμ-Fcμ
203 contacts, and the location and structure of the Tp assembly differ markedly, indicating that
204 mechanisms of pIg assembly and function have evolved from teleost to humans.

205 A fundamental difference between tFcμ and hFcμ is the geometric relationship among
206 adjacent Fcμ monomers. For example, adjacent tFcμs are related by angles of 80-85° and ~120°
207 (non-adjacent tFcμ1 and tFcμ4 separated by the Tp assembly) while the five adjacent hFcμs are
208 related by angles between ~53° and ~62° (non-adjacent hFcμ1 and hFcμ5 separated by hJC)
209 (Fig. 1b and Fig. 5a). The inter-Fc contacts that stabilize these geometries in both tFcμ and hFcμ
210 include CHμ4 C-F-G β-sheets and FG loops; however participating residues are not conserved,
211 resulting in an interface of ~400Å^2 between hCHμ4s and a larger interface of ~800Å^2 between
212 tCHμ4s. In addition, adjacent hFcμ monomer CH3 domains form contacts and are covalently
213 linked by disulfide bonds between Cys414 residues[24,25] whereas this cysteine residue is not
214 conserved in tFcμ and thus does not form inter-Fc disulfide bonds between tFcμ CH3 domains.

215 Structural comparisons between CH4 domains and Tps from humans and fish IgM are
216 especially relevant to understanding pIg evolution and the functional differences between species
217 because these domains form the majority of contacts that stabilize polymeric forms of IgM. Human
218 and teleost IgM heavy chain constant domains share 28% sequence identity and 33% similarity
219 among CH4 domains (Fig. 5b). A structural alignment between the human and teleost CHμ4
220 domains (chain B, excluding Tp residues) reveals a RMSD of 1.386Å when aligned to A-type
221 domains, and RMSD of 1.047Å when aligned to B-type domains, suggesting that that B-type
222 domains share greater structural similarity to human IgM CHμ4. Structural variability in the main

223 chain arises primarily from differences in the length and conformations of the CD and DE loops
224 (Fig. 5c), which contain residues that mediate contacts within and between Fcμs and are not well
225 conserved from teleost to mammalian IgM. Even in regions that share high structurally similarity,
226 such as the A-B-D-E β-sheet, some residues mediating contacts are not conserved, including
227 those that stabilize the CH4-CH4 interface within each Fc (Fig. 5b).

228 The Tp has long been implicated in pIg assembly; it is presumed to be present in all pIgs[2,3]
229 and is conserved among vertebrate pIg sequences, although residues near the C-terminus are
230 variable, including the position of a conserved cysteine relative to the C-terminus (Supplementary
231 Fig. 5). Mammalian IgM and IgA structures revealed Tps folding together with the JC to form β
232 sandwich-like domains near the center of the complexes[22–24]. In human SIgM, residues 562-568
233 of each hFcμ Tp form a β-strand, and 10 copies form two parallel β-sheets that lie near-
234 perpendicular to the central axis of the pentamer, positioning the distal β-strand from each sheet
235 to pair with one β-strand from JC (Fig. 5a). In contrast, the tFcμ Tp assembly lacks a JC and is
236 located toward the side of the complex between Fcμ1 and Fcμ4 with β-sheets positioned near-
237 parallel to the central axis (Fig. 1b). In both structures, conserved hydrophobic residues Leu437
238 and Leu439 (Val564 and Leu566 in hFcμ) face the core of the β-sandwich and conserved
239 hydrophilic residues Asn436 and Ser438 (Asn563 and Ser565 in hFcμ) face outward (Fig. 4b, c).
240 However, despite the observation that conserved residues mediate hydrophobic interactions at
241 the core of both hFcμ and tFcμ Tp assemblies, the relationship between the two Tp assembly β-
242 sheets differs; the two tFcμ β-sheets are related by an angle of 48° whereas the two hFcμ β-
243 sheets are related by an angle of 158° (Fig. 4a and Fig. 5a).

244

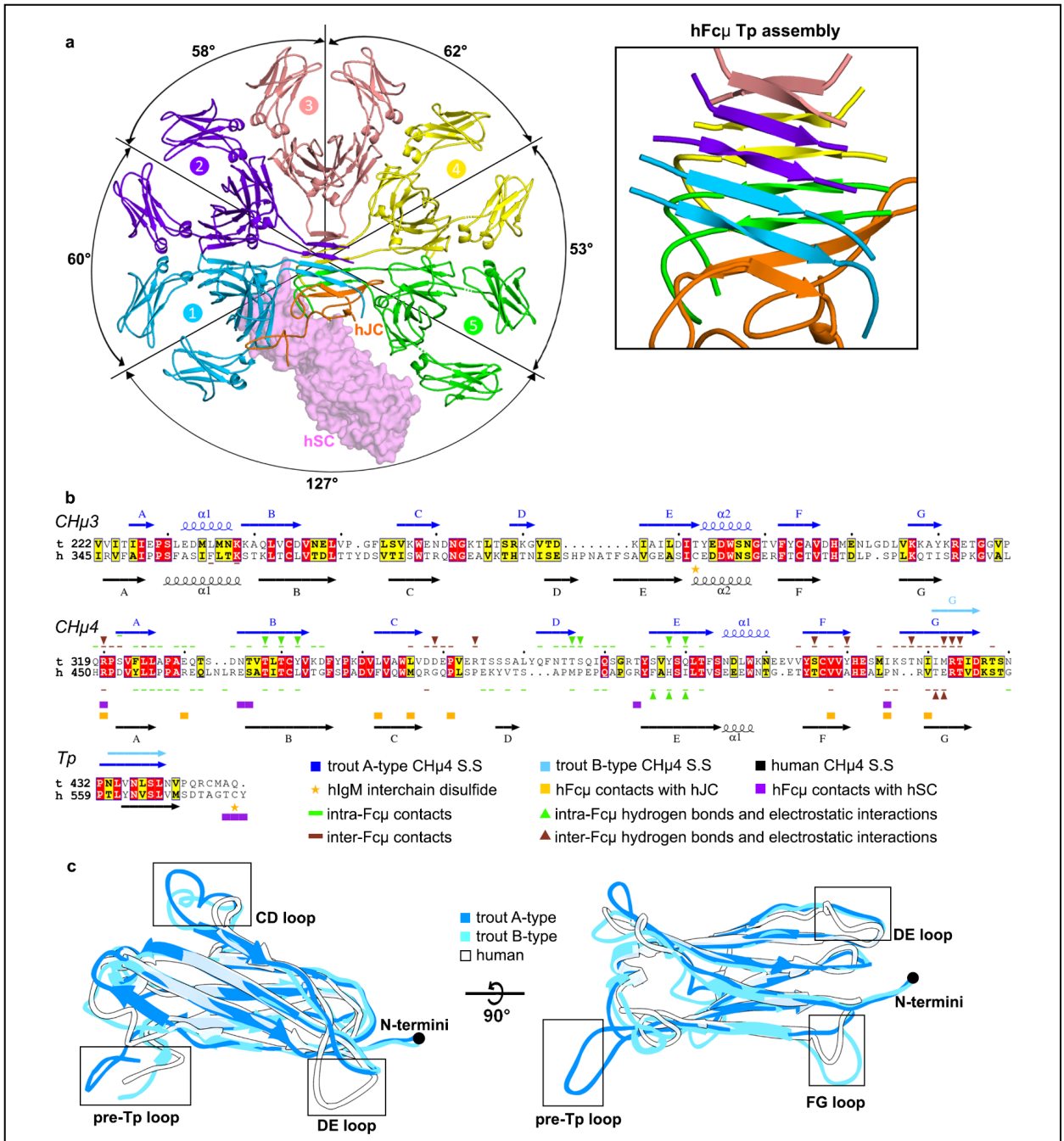


Fig. 5 | Comparison of IgM structures from trout and humans. a Human SFcμ structure, containing five hFcμ monomers, hJC, and hSC (PDB ID: 6kxs); hFcμ and hJC are shown as cartoon representations and hSC is shown as molecular surface representation. Angles between Fcμ monomers are indicated in the structure. Inset, 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. **b** Sequence alignment of IgM CHμ3, CHμ4, and Tp from *Oncorhynchus mykiss* (rainbow trout, t) with *Homo sapiens* (human, h), 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. **c** Structural alignment of hFcμ CHμ4 domain with tFcμ A-type CHμ4 domain (chain A) and tFcμ B-type CHμ4 domain (chain B). All domains are shown as cartoon representations.

245 The most striking difference between tFcμ and hFcμ Tp assemblies is a difference in β-
246 strand pairing. In tFcμ, each Tp pairs with a Tp originating from an adjacent Fc monomer while in
247 hFcμ, Tp β-strands pair with a second Tp from the same hFcμ monomer. The exception to this is
248 hFcμ3, which contributes one Tp to each β-sheet in the hFcμ Tp assembly (Fig. 4a and Fig. 5a).
249 The hFcμ Tp strand-pairing pattern is also observed for Tps in polymeric forms of IgA[23]. Despite
250 differences in strand pairing, the length from the last residues of CHμ4 (Asp426 in tIgM; Asp 553
251 in hIgM) to the first residue in the FcμTp β-strand (Leu434 in tIgM; Tyr562 in hIgM) is conserved
252 between the tFcμ B-type fold and hFcμ. Also conserved among hFcμ and tFcμ is a potential N-
253 linked glycosylation site (PNGS) including residues Asn436 and Ser438. Earlier studies indicated
254 that glycosylation at this site can influence hIgM polymerization and may prevent
255 aggregation[24,29,30]. However, we did not observe ordered glycosylation at this site in the tFcμ
256 structure (Supplementary Fig. 4h).

257 Finally, both tFcμ and hFcμ contain a conserved cysteine residue in the C-terminal loop
258 following the Tp β-strand. While only four out of ten copies of Cys575 are ordered in published
259 hFcμ structures[24], and density at this conserved cysteine (Cys445) is poor in the tFcμ structure
260 (Supplementary Fig. 4b-f), data point toward an essential role for a cysteine near the C-terminus
261 of IgMs. As noted, we observe putative disulfide bond formation linking tFcμ $Cys445^{Fc\mu1A}$ with
262 $Cys445^{Fc\mu3E}$ and $Cys445^{Fc\mu2D}$ with $Cys445^{Fc\mu4H}$ and deletion of this residue inhibits tetramer
263 formation (Supplementary Fig. 4g). In the case of hFcμ, this cysteine (Cys575) forms disulfides
264 with the JC or with another Tp located in the same sheet and is reportedly necessary for hIgM
265 assembly[24,25]. The pattern of disulfide bonding appears to be different and to link different regions
266 in tIgM versus hIgM. In sum, structural analysis of tFcμ and hFcμ reveals that despite conservation
267 in sequence and length, IgM Tps can assemble into markedly different structures to promote
268 assembly of pIgs.
269
270

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

272 Armed with new structural insights and a desire to better understand pIg evolution and assembly
273 mechanisms, we explored the ability of tFcμ motifs to facilitate the assembly of pIgs from other
274 heavy chains including human mu (IgM), alpha-1 (IgA1) and gamma-1 (IgG1). The tFcμ structure
275 suggests that both A-type and B-type tCHμ4 domains are essential for tFcμ assembly. Given that
276 tFcμ A-type and B-type tCHμ4 domains, despite having the same sequence, differ structurally in
277 the FG loop, G strand, and pre-Tp loop regions, we hypothesized that the mechanisms of tFcμ
278 assembly rely heavily on residues in these motifs. To test this hypothesis, we designed twelve
279 chimeric expression constructs, in which residues found in the A-type or B-type tCHμ4 G strand
280 and/or Tp residues are replaced with endogenous counterparts in human Fcμ, Fcα1, and Fcγ1
281 (Fig. 6a). Chimeric Fc constructs and associated controls were transiently transfected into human
282 cells and the resulting proteins were purified and the polymeric state was evaluated using size
283 exclusion elution chromatography with in-line multi-angle light scattering (SEC-MALS), which
284 reports on the absolute molecular weight of target proteins and complexes[31].

285 As noted, human IgM is naturally expressed as a hexamer or a JC-containing pentamer,
286 and the Tp is reportedly essential for the assembly of both complexes[30]. Consistent with published
287 studies, in our experimental system the removal of Tp residues resulted in only monomeric hFcμ
288 (Fig. 6b and Supplementary Fig. 6a). To determine if the tFcμ Tp could promote pIg assembly in
289 hFcμ constructs lacking the endogenous Tp (and the JC), we measured the molecular weight of
290 four chimeric proteins, hFcμ-tFcμ-chi1 to chi4 (Fig. 6b and Supplementary Fig. 6a). The hFcμ-
291 tFcμ-chi1 and hFcμ-tFcμ-chi2 were monomeric; however, hFcμ-tFcμ-chi3 and hFcμ-tFcμ-chi4
292 were polymeric with average molecular weights consistent with tetrameric complexes (Fig. 6b and
293 Supplementary Table 3). The hFcμ-tFcμ-chi3 contains the tFcμ Tp and hFcμ-tFcμ-chi4 contains
294 the B-type tCHμ4 G strand and tFcμ Tp residues; thus, data indicate that substitution of hFcμ
295 residues with these motifs, but not the A-type tCHμ4 G strand and tFcμ Tp can promote pIg

assembly. However, in contrast to the hexamers observed when full-length hFcμ is expressed without JC, chimeric pIgs adopted probable tetramers.

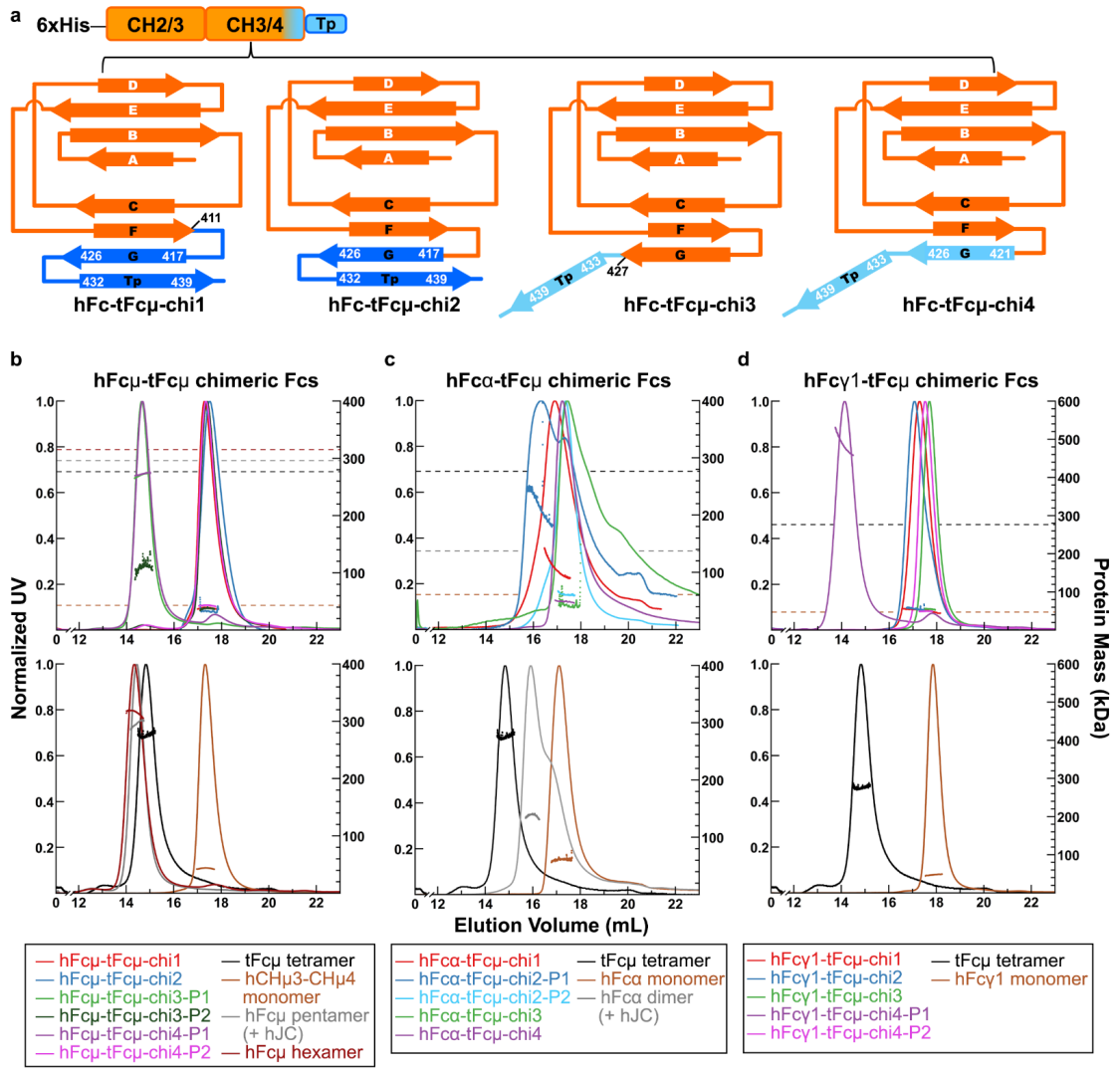


Fig. 6 | Design and characterization of chimeric Fcs. a Top: the design of chimeric Fc constructs; bottom: schematic showing topology of chimeric Fc constructs. Each chimeric Fc consists of an N-terminal hexa-histidine tag, partial hFc sequence (orange), and tFcμ C-terminal sequence (marine or light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

298 We also sought to determine if the tCHμ4 G strand and/or Tp residues could promote pIg
299 assembly in the context of hFcα, which is monomeric in the absence of the JC and typically
300 dimeric when expressed with a JC, or in the context of Fcγ1, which is naturally monomeric and
301 does not contain a Tp. In contrast to hFcμ-tFcμ chimeras, we found that hFcα-tFcμ-chi1 and hFcα-
302 tFcμ-chi2 formed pIgs whereases hFcα-tFcμ-chi3 and hFcα-tFcμ-chi4 did not. The molecular
303 weight of resulting pIgs was consistent with a mix of species ranging from dimers to tetramers
304 (Figs. 6c and Supplementary Fig. 6b). Data signify that in the context of the human alpha-1 heavy
305 chain, the G strand and Tp residues found in A-type tCHμ4 fold, can promote pIg assembly in the
306 absence of JC, whereas G strand and Tp residues found in the B-type tCHμ4 fold, or Tp residues
307 alone, cannot. Among hFcγ1-tFcμ chimeras, only hFcγ1-tFcμ-chi4 promoted pIg assembly,
308 suggesting that in the context of the human gamma-1 heavy chain, substituting G strand and Tp
309 residues found in the B-type tCHμ4 fold could promote assembly whereas others could not. For
310 all chimeric Fc constructs that promoted pIg assembly, pIgs were isolated from mixtures
311 containing mIgs before SEC-MALS analysis (Supplementary Fig. 6). Among hFcμ-tFcμ chimeras,
312 the pIgs represented more than 50% of the total protein purified whereas pIgs isolated from hFcα-
313 tFcμ and hFcγ1-tFcμ chimeras represented less than 30% of the total protein purified, with mIgs
314 making up the rest of the total (Supplementary Fig. 6). This suggests that in our experimental
315 system, pIg assembly promoted by tFcμ motifs was most efficient for hFcμ-tFcμ chimeras.
316

317 **DISCUSSION**

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

324 Specifically, the Ig domains and Tps provide a common set of building blocks, while differences
325 in heavy chain amino acid sequence and the presence or absence of JC are variables adding
326 plasticity to the process.

327 As noted, most jawed vertebrates, including mammals, birds, amphibians, reptiles and
328 cartilaginous fish, reportedly encode JC and can incorporate it into pIgM, whereas teleost fish
329 have lost JC and assemble the Fc regions of pIgs using only Ig domains and Tps[4–6]. We find that
330 this unique teleost characteristic is associated with distinct tFcμ structural folding patterns when
331 compared to mammalian pIgM structures containing the JC[24,25]. This raises the possibility that
332 JC-independent, and JC-dependent assembly evolved to employ distinct biophysical
333 mechanisms. However, we also find that tFcμ C-terminal residues (and presumably associated
334 structural motifs) are able to promote JC-independent pIg assembly of otherwise human heavy
335 chains, suggesting some degree of interchangeability among the sequences of pIg-forming heavy
336 chains and assembly mechanisms. For example, chimeric hFcμ-tFcμ revealed that a subset of
337 tFcμ motifs were able to promote pIg assembly of an otherwise human IgM heavy chain in the
338 absence of JC. It is intriguing that this effect was found only when the tFcμ CH4 G strand and/or
339 Tp residues found in the B-type fold, but not the A-type fold, were used. While this may reflect the
340 higher degree of similarity shared between the B type fold and human IgM domains, the resulting
341 pIgs appeared to be tetramers rather hexamers, the predominant polymeric state associated with
342 human IgM lacking the JC in our experiments (and also in human serum)[10]. It remains unknown
343 whether hFcμ-tFcμ chimera tetramers assemble into structures more similar to tIgM or hIgM with
344 respect to the Tp assembly structure, contacts, and geometric relationships between the
345 incorporated Fcs. While pIg assembly of hFcα-tFcμ, and hFcγ1-tFcμ chimeras was notably less
346 efficient than hFcμ-tFcμ chimeras, it is significant that tFcμ components could promote some pIg
347 assembly in the context of human IgA1 and IgG1 heavy chains, neither of which naturally form
348 pIgs in the absence of JC. While we cannot rule out the possibility that assembly in plasma cells
349 differs from the HEK cells used in our experiments, experimental cell lines assembled controls of

350 expected MW and provided a uniform system to evaluate how sequence differences among
351 chimeras influences pIg assembly.

352 Despite the prevalence of JC in the genomes of most jawed vertebrates (other than
353 teleosts), some species assemble pIgs in the absence of JC. For example human IgM can
354 assemble without JC (typically hexameric) and amphibian IgX (polymeric) also appears to lack
355 the JC [5,6,10–12]. This raises the question of whether or not pIgs lacking the JC from species other
356 than teleost fish have structures most similar to hexameric hIgM or to tetrameric tIgM. At this time,
357 a high resolution structure detailing the Tp assembly of hexameric IgM has not been reported;
358 however, negative-stain EM and small-angle light scattering (SAXS) data suggest all six hFcμs
359 form Fc-Fc contacts, and that the Tp assembly remains in the center of the complex as seen in
360 pentameric IgM, rather than bridging two Fcs as we observed in tetrameric tFcμ[32,33]. The
361 structures of unliganded Tps (e.g., those found in monomeric IgA or IgM) are also unknown and
362 therefore it remains unclear if and how Tps might change conformation during mammalian pIg
363 assembly in absence of JC and/or if they adopt conformations distinct from what is observed in
364 mammalian pIg and SIg structures. Without such information, it is difficult to judge whether the Tp
365 β-strands would alternate in hIgM hexamers, as observed in tFcμ tetramers, or if they would pair
366 with another Tp β-strand from the same Fc as observed in hFcμ pentamer and mammalian IgA
367 polymers.

368 While our structural data provide considerable insight on tIgM structure and assembly
369 mechanisms, the stages of the assembly process remain unclear. Identification of two distinct
370 tCHμ4 folds evokes the question of how tIgM heavy chains fold and if they adopt A- and/or B-type
371 CHμ4 domains prior to the assembly of pIgM. In isolation, the A-type fold appears likely to be
372 more stable than the B-type fold because the β-strand region of the Tp forms β-sheet interactions
373 with the G strand, rather than being extended away from the core of the Ig domain in a potentially
374 flexible location (see Fig. 2b, c). While further study is needed to define the stages of tFcμ
375 assembly, one possibility is that heavy chains first adopt the A-type conformation and that

376 polymeric assembly is triggered when two tFcμ monomers come into proximity and interactions
377 at the Fc-Fc interface displace Tp β-strands from both Fcs, effectively kicking them out of the Ig-
378 fold and allowing them to form parallel β-sheet interactions with each other. It is unclear if there
379 is sufficient energy in the tFcμ-tFcμ interface for its formation to break Tp-G strand hydrogen
380 bonding and thus if this model is correct, it is conceivable that molecular chaperones, which have
381 been implicated in mammalian pIgA and pIgM assembly processes, are involved[34,35]. Regardless,
382 sequential interactions between Fcs could add strands to the Tp assembly until the tetramer was
383 closed by the last remaining A-type fold. In a related model, sequential addition of heavy chains
384 might culminate in dimers (tetramer halves), two of which subsequently come together to form a
385 tetramer. These models are consistent with the alternating Tp β-strand pairing that we observe in
386 tFcμ, but that is not seen in mammalian IgM or IgA containing JC[22–24]. Furthermore, given that
387 Cys445Ser mutation abolished tFcμ tetramer formation (Supplementary Fig. 4g), it is possible
388 that disulfide bonds linking non-adjacent Fcs, and/or the top and bottom β-sheets in the Tp
389 assembly (see Fig. 4a, Supplementary Fig. 4b-f) stabilize some part of this process. A critical role
390 for Cys445 is consistent with published work indicating that disulfide crosslinking within tIgM
391 influences antibody assembly, stability and half-life[17,36,37]; however additional studies will be
392 needed to work out its precise role(s) in these processes.

393 Another remarkable finding of our work is that the majority of pIg assembly mechanisms
394 appear likely to culminate in the release of an asymmetric pIg complex, despite being assembled
395 from multiple copies of identical Ig heavy chains with identical sequence. Asymmetry in
396 mammalian pIgM and pIgA was verified with the publication of pIgA and pIgM structures, in which
397 the JC was bound to two Igs uniquely, and is thus postulated to template the addition of monomers
398 during assembly and to confer asymmetry on the resulting complex[22–24]. However, the tFcμ
399 structure, which lacks a JC, also reveals elements of asymmetry, with four Fc monomers directed
400 away from the Tp assembly such that the relative geometric relationships between each tFcμ are
401 not identical. As noted, this arrangement appears to be dependent upon the existence of A-type

402 and B-type CHμ4 domains, yet it is unclear if asymmetry in pIgs evolved as an efficient means to
403 assemble a pIg and/or if asymmetry provides a functional advantage.

404 IgM is a critical Ig heavy chain class found in all jawed vertebrates. Studies in mammals
405 demonstrate that its effector functions are mediated by high avidity interactions with antigen (from
406 five sets of Fabs), complement activation, binding to IgM-specific Fc receptors (FcRs), and
407 mucosal functions associated with pIgR-dependent secretion. While relatively few reports have
408 documented its function in teleost fish, evidence suggests that tIgM also participates in
409 complement activation, mucosal functions such as bacterial coating and immune exclusion and
410 binds IgM-specific FcRs[21,36,38,39]. The tetrameric state of tIgM we observe implies that it has the
411 potential to bind multiple copies of an epitope simultaneously, which is likely to support high-
412 avidity antigen binding and/or antigen crosslinking. The structural relationship between tFcμ CH2
413 and Fab domains remains uncharacterized and thus modeling potential positions of tIgM Fabs is
414 non-trivial; however, based recent structures detailing the hIgM CH2 domain we estimate they
415 might extend ~125Å from the center of the complex[40]. Furthermore, our structure indicates that
416 Fabs are likely to be directed away from the Tp assembly, which separates Fcμ1 and Fcμ4,
417 leaving portions of the Tp assembly and its adjacent Fcs sterically accessible. It is functionally
418 intriguing that each copy of tFcμ is related to another by a different angle, raising the possibility
419 that different geometric relationships between Fabs within a pIg provide unique binding
420 advantages for different types of antigen distribution. For example, Fcs separated by 120° might
421 bind epitopes that are separated by larger distances compared to Fcs separated by 80° (Fig. 1c).
422 We also hypothesize that the 120° separation between tFcμ1 and tFcμ4 could promote FcR
423 binding to tFcμ1 and/or tFcμ4. We note that the two Fcs in dimeric forms of mammalian IgA are
424 related by angles of 97°; in human IgM, two Fcs are separated by 53°-62°, with exception of Fcμ1
425 and Fcμ5 which are separated by an angle of 127° (see Fig. 5a). The 97° bend between two IgA
426 monomers in the SIgA structure is predicted to increase accessibility to FcR binding sites and
427 promote binding away from Fabs[22]. Future studies will be necessary to determine if the geometric

428 relationship between Ig monomers in pIgs have been selected for because of a functional
429 advantage, an assembly advantage, or both.

430 Although teleost SIgT has been identified as the immunoglobulin specialized in teleost
431 mucosal immunity, substantial amounts of tSIgM are also found in gut and skin mucus before and
432 after immunization[18,21,41]. SIgs are crucial to mucosal immunity because of their ability to bind and
433 exclude microbes before they breach epithelial cells. This is especially important to fish since their
434 skin is covered in mucus and constantly exposed to an external aquatic environment. In both
435 mammals and teleost fish, the pIgR binds and transports pIgM across epithelial cells and is
436 cleaved from the cell surface upon delivery to the mucosa, leaving its ectodomain bound to the
437 pIg where it is referred to as secretory component (SC). In mammals, the JC is necessary for this
438 process and forms direct contacts with SC, which contains five Ig-like domains[24,25]. In contrast,
439 JC is not required for delivery of tSIgM to the mucosa and tSC contains just two Ig-like domains
440 that share low sequence conservation with mammalian counterparts; these observations indicate
441 an alternative tSC-tIgM binding mechanism[19,42]. Consistent with these observations, we found
442 that only two hIgM residues that bind hSC are conserved and solvent-accessible on tFcμ (Fig.
443 5b). It is yet unknown where the tSC binding site is located and if that site overlaps with other FcR
444 binding sites.

445

446 **MATERIALS AND METHODS**

447 **Construct design and protein expression**

448 The gene encoding the *Oncorhynchus mykiss* IgM heavy chain secretory form (GenBank:
449 AAW66974.1) was obtained from the NCBI database. DNA sequence encoding a signal peptide
450 (MFPASLLLLLAAASCVHC), a hexa-histidine tag, CHμ3 (residues 222-318), CHμ4 (residues
451 319-432), and Tp (residues 433-448) was codon optimized for human cell expression,
452 synthesized (Integrated DNA Technologies, Inc.) and cloned into mammalian transient
453 expression vector pD2610-v1 (Atum). Resulting plasmid DNA was transiently transfected into

454 HEK Expi293-F cell line using the ExpiFectamine™ 293 Transfection Kit (Thermo Fisher
455 Scientific). Six days after transfection, supernatants were harvested and His-tagged tIgM-Fc (tFcμ)
456 protein was purified using Ni-NTA Agarose (Qiagen) and subsequently, Superose 6 Increase
457 10/300 column (Cytiva) size exclusion chromatography (SEC). Elution fractions corresponding to
458 the predicted molecular weight of tFcμ tetramer were collected and stored in TBS buffer
459 (containing 20mM Tris-HCl and 150mL NaCl, pH 7.8) for cryo-EM.

460 To design chimeric Fcs, published structures containing hFcα, hFcγ1, and hFcμ (PDB IDs:
461 6UE7, 1MCO, 6KXS, respectively) were structurally aligned to A-type and B-type tCHμ4
462 structures. Constructs hFc-tFcμ-chi1 and hFc-tFcμ-chi2 were designed based on structural
463 alignment to A-type tCHμ4 and constructs hFc-tFcμ-chi3 and hFc-tFcμ-chi4 were designed based
464 on structural alignment to B-type tCHμ4. In construct hFc-tFcμ-chi1, A-type tCHμ4 residues in FG
465 loop, G strand, pre-Tp loop, and Tp (residues 412-448) were swapped into the C-terminal CH
466 domain of hFcs (CH3 or CH4). In the hFc-tFcμ-chi2 construct, A-type tCHμ4 residues in G strand,
467 pre-Tp loop, and Tp (residues 417-448) were swapped into the C-terminal CH domain of hFcs
468 (CH3 or CH4). In the hFc-tFcμ-chi3 construct, B-type tCHμ4 residues in pre-Tp loop, and Tp
469 (residues 427-448) were swapped into the C-terminal CH domain of hFcs (CH3 or CH4). In the
470 hFc-tFcμ-chi4 construct, B-type tCHμ4 residues in G strand, pre-Tp loop, and Tp (residues 421-
471 448) were swapped into the C-terminal CH domain of hFcα and hFcμ, or fused after the C-
472 terminus of hFcγ1. Chimeric Fc constructs were cloned, transfected, and purified using the same
473 protocol described for the tFcμ construct.

474

475 **SEC-Multiangle light scattering (SEC-MALS) analysis**.

476 Proteins samples were prepared as described. After purification by SEC (see Supplementary Figs.
477 1&6), fractions in each peak were pooled together, concentrated to approximately 2 g/L, filtered,
478 and sent to University of Pennsylvania for SEC-MALS analysis. SEC-MALS experiments were

479 performed on tFcμ and mutant variants. 100uL injections of purified protein in TBS buffer was
480 loaded onto a Superose 6 Increase 10/300 column (Cytiva) at 0.5 ml/min at 25°C. Absolute
481 molecular weights were determined using MALS. The scattered light intensity of the column eluant
482 was recorded at 18 different angles using a DAWN-HELEOS MALS detector (Wyatt Technology
483 Corp.) operating at 658 nm after calibration with the monomer fraction of Type V BSA (Sigma).
484 Protein concentration of the eluant was determined using an in-line Optilab T-rex interferometric
485 refractometer (Wyatt Technology Corp.). The weight-averaged molecular weight of species within
486 defined chromatographic peaks was calculated using the ASTRA software version 8.0 (Wyatt
487 Technology Corp.), by construction of Debye plots ($KC/R\theta$ *versus* $\sin^2[\theta/2]$) at 1-s data intervals.
488 The weight-averaged molecular weight was then calculated at each point of the chromatographic
489 trace from the Debye plot intercept, and an overall average molecular weight was calculated by
490 averaging across the peak. Data were further analyzed using the protein conjugate analysis
491 module as implemented in Astra for protein with carbohydrate modifier. For all samples, the
492 predicted molecular weight from the protein sequence, the average molecular weight determined
493 by MALS, and the number and potential of PNGS are listed in Supplementary Table 4.
494

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

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

505

506 **Cryo-EM data processing**

507 Movies were imported into cryoSPARC and binned by 2 during patch motion correction[43,44].
508 Images were curated for CTF fits, ice thickness, and contamination resulting in 3131 remaining
509 images. Blob picker was used with about half of the images to generate an initial particle stack.
510 After 2D classification, best classes were selected as templates for Template picker on the full
511 dataset, resulting in ~1.3 M picks. Particles were binned by 2 during extraction and subject to
512 several rounds of 2D classification and heterogeneous refinement. Remaining 698k particles were
513 re-extracted at the full pixel size. After an additional round of heterogeneous refinement, 610k
514 particles were used in non-uniform refinement (with C2 symmetry), resulting in a 2.9 Å map[45].
515 After a cycle of local and global CTF refinement, an improved map at 2.78 Å was obtained. A non-
516 uniform refinement with C1 symmetry was also computed, resulting in a 2.9 Å map. C1 and C2
517 maps were essentially superimposable (not shown) and the C2 map was used for structure
518 refinement.

519

520 **Atomic model building, refinement, and validation**

521 A homology model for a single tCHμ4 domain was made with SWISS-MODEL using the human
522 IgM-Fc core cryo-EM structure (PDB ID: 6KXS) as template; the C-terminal Tp was not included
523 in the homology model[46]. The tCHμ4 homology model was docked into real-space density of
524 $CH\mu4^{Fc\mu1A}$ using UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and
525 Informatics at the University of California, San Francisco, with support from NIH P41-GM103311[47].
526 Initial inspection of the docked homology model revealed several regions poorly fitted into density.
527 To improved model fitting, less structured regions (residues Val362 to Gln378, C-terminal
528 sequence after residue Arg427) were deleted from the homology model and resulting model was
529 automatically fitted into real-space density using Coot Molecular Graphics Package[48].
530 Subsequently, residues 397-427 were deleted from the model due to mis-fitting and then rebuilt

531 manually. The outcome was inspected and manually rebuilt in Coot before being subjected to
532 Phenix real-space refinement[49]. After several refinement iterations, the model of CHμ4^{Fcμ1A} and
533 its Tp agreed well with the cryo-EM density map and a copy of the model was docked into density
534 of CHμ4^{Fcμ1B}. Due to the structural differences between CHμ4^{Fcμ1A} and CHμ4^{Fcμ1B}, adjustments to
535 the C-terminal sequence of CHμ4^{Fcμ1B} were made manually in Coot to optimize model fitting into
536 cryo-EM map. Models of CHμ4^{Fcμ1A} and CHμ4^{Fcμ1B} were combined and refined using Phenix and
537 Coot Molecular Graphics Package[48,49]. Copies of CHμ4^{Fcμ1B} were docked into the density of
538 CHμ4$^{Fcμ2C-4G}$ and one copy of CHμ4^{Fcμ1A} was docked into the density of CHμ4^{Fcμ4H} using UCSF
539 Chimera; these newly docked CHμ4 models were first refined individually using Phenix and then
540 combined with previously refined CHμ4^{Fcμ1A} and CHμ4^{Fcμ1B} domains[47,49]. Residues linking CHμ4
541 domains to Tps and residues in the Tp assembly were built manually into the combined model.
542 Cys445 residues were built into density with the most fit rotamers that assume preferred
543 orientation to form interchain disulfides. The resulting model was subjected to iterations of
544 automatic refinement in Phenix and manual adjustment in Coot Molecular Graphics Package.
545 The refined CHμ4 model was used as template to make a homology model for the tCHμ3 domain
546 using SWISS-MODEL[46,48,49]. The CHμ3 homology model was docked into the density of
547 CHμ3^{Fcμ2D} and refined using similar methods as described above. Subsequently, copies of
548 CHμ3^{Fcμ2D} were docked into the density of remaining CHμ3 domains. CHμ3 domains were joined
549 with the previously refined CHμ4 domains and Tp assembly; the entire model is refined using
550 Phenix and Coot Molecular Graphics Package[48,49]. The final tFcμ structure was evaluated by
551 Phenix EM Validation, MolProbity, and EMRinger[49,50]; results are summarized in Supplementary
552 Table 1.
553
554 **Sequence and structural alignments and structural analysis**
555 NCBI Genbank accession numbers for Ig heavy chain sequences shown in figures are:
556 *Oncorhynchus mykiss* (trout) IgM secretory form (AAW66974.1), *Salmo Salar* (salmon) IgM

557 (AAB24064.1), *Xenopus laevis* (frog) IgM (AAA49774.1), *Trachemys scripta elegans* (turtle) IgM
558 (AFR90255.1), *Gallus gallus* (chicken) IgM (P01875), *Mus musculus* (mouse) IgM (P01872), and
559 *Homo sapiens* (human) IgM (P01871). Sequences alignments were made using ClustalOmega
560 and corresponding figures were made using Espript3[51,52]. Structural alignments between human
561 IgM CH4 of (residues 450-555 of chain A; PDB ID: 6KXS) and teleost IgM-CH4 (residues 316-
562 428 of chain A or chain B) were made using the "align" function in PyMOL; structural alignments
563 between teleost IgM CH3-CH4-Tp (residues 225-445) were made using the "align" function in
564 PyMOL[53]. Interfaces were calculated using PISA, "Protein Interfaces, surfaces and assemblies"
565 service at the European Bioinformatics Institute[54]; results are summarized in Supplementary
566 Table 2.
567
568 **Fc angle and distance measurements**
569 To measure radius of a polymeric forms of Fcμ, the tFcμ (this study) and human SIgM-Fc
570 core (PDB ID: 6KXS) structures were imported into UCSF Chimera[47]. For each model, a plane
571 was defined using built-in "structure measurements" tools and the radius was calculated
572 automatically. To measure angles between adjacent Fcs, Tp sequences were removed from
573 models. An axis was defined using mass weighting and helical correction for each Fc monomer
574 and then angles between the defined axes were measured and defined as the angles between
575 adjacent Fcs. To measure the diameter of the solvent accessible hole at the center of tFcμ
576 tetramer, tFcμ model was imported into UCSF Chimera[47]. Nine atom pairs located on the surface
577 were chosen around the hole. Distance between two atoms in each pair was measured and the
578 diameter of this hole was determined to be the average of these distances. Chosen atom pairs
579 and the distance measurements from each pair are summarized in Supplementary Table 3.
580 To measure the angle between the two tFcμ-Tp β-sheets, the tFcμ tetramer structure was
581 imported into PyMOL[53]. The angle between $Tp^{tFc\mu 1A}$ and $Tp^{tFc\mu 3E}$ was measured as the dihedral
582 angle between Cα of $Leu434^{tFc\mu 1A}$, $Asn440^{tFc\mu 1A}$, $Asn440^{tFc\mu 3E}$, and $Leu434^{tFc\mu 3E}$; angles between

583 $Tp^{tFc\mu1B}$ and $Tp^{tFc\mu3F}$, $Tp^{tFc\mu2C}$ and $Tp^{tFc\mu4G}$, $Tp^{tFc\mu2D}$ and $Tp^{tFc\mu4H}$ were measured using the same
584 method. The angle between two Tp β-sheets was determined to be the average of these
585 measurements. To measure the angle between the two hFcμ-Tp β-sheets the hSIgM-Fc core
586 structure (PDB ID: 6KXS) was imported into PyMOL[53]. The angle between $Tp^{Fc\mu1A}$ and $Tp^{Fc\mu5L}$
587 was measured as the dihedral angle between Cα of $Tyr562^{Fc\mu1A}$, $Val567^{Fc\mu1A}$, $Val567^{Fc\mu5L}$, and
588 $Tyr562^{Fc\mu5L}$; angles between $Tp^{Fc\mu1B}$ and $Tp^{Fc\mu5K}$, $Tp^{Fc\mu2C}$ and $Tp^{Fc\mu4H}$, $Tp^{Fc\mu2D}$ and $Tp^{Fc\mu4G}$, $Tp^{Fc\mu3E}$
589 and $Tp^{Fc\mu3F}$ were measured using the same method. An average of these measurements was
590 taken and the angle between two Tp β-sheets was determined to be the supplement angle of the
591 average.

592 To measure the angle between tFcμ CH3 and CH4, the tFcμ tetramer model was imported
593 into PyMOL[53]. For each copy of heavy chain, the angle between tFcμ CH3 and tFcμ CH4 was
594 measured as the dihedral angle between the Cα atoms of His298, Lys238, Glu412, and Thr332.
595 The average of eight angle measurements was reported as the final value of tFcμ CH3-CH4 angle.
596

597 **Figures**

598 Structural figures were made using UCSF Chimera, PyMOL[47,53]. SEC data and SEC-MALS data
599 were replotted using Origin. All figures were assembled using Adobe Photoshop.
600

601 **DATA AVAILABILITY.**

602 Cryo-EM density map has been deposited in the EM databank (www.ebi.ac.uk/emdb) with the
603 accession code EMD-40054 and the refined coordinate has been deposited in the Protein Data
604 Bank (www.rcsb.org) with accession code 8GHZ.
605

606 **AUTHOR CONTRIBUTIONS.**

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

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. Cryogenic electron microscopy was done in the Beckman Institute cryo-EM resource center at California Institute of Technology. This work was supported by NIH grant 1R01AI165570 and University of Illinois start-up funding to B.M.S.

COMPETING INTERESTS.

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

MATERIALS AND CORRESPONDENCE.

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

REFERENCES

1. 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).
2. Chapuis, R. M. & Koshland, M. E. Linkage and assembly of polymeric IgA immunoglobulins. *Biochemistry* **14**, 1320–1326 (1975).
3. Wiersma, E. J. & Shulman, M. J. Assembly of IgM. Role of disulfide bonding and noncovalent interactions. *J Immunol* **154**, 5265–5272 (1995).
4. Hohman, V. S. *et al.* J Chain in the Nurse Shark: Implications for Function in a Lower Vertebrate. *J Immunol* **170**, 6016–6023 (2003).

634 5. Hadjiazimi, I. & micheahamzehpour, M. Xenopus-laevis 19s immunoglobulin -
635 ultrastructure and J-chain isolation. *Immunology* **30**, 587–591 (1976).
636 6. Castro, C. D. & Flajnik, M. F. Putting J Chain Back on the Map: How Might Its
637 Expression Define Plasma Cell Development? *The Journal of Immunology* **193**,
638 3248–3255 (2014).
639 7. Saito, S. *et al.* IgA tetramerization improves target breadth but not peak potency of
640 functionality of anti-influenza virus broadly neutralizing antibody. *PLoS Pathog* **15**,
641 e1007427 (2019).
642 8. Suzuki, T. *et al.* Relationship of the quaternary structure of human secretory IgA to
643 neutralization of influenza virus. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7809–7814
644 (2015).
645 9. Terauchi, Y. *et al.* IgA polymerization contributes to efficient virus neutralization on
646 human upper respiratory mucosa after intranasal inactivated influenza vaccine
647 administration. *Human Vaccines & Immunotherapeutics* **14**, 1351–1361 (2018).
648 10. Eskeland, T. & Christensen, T. B. IgM Molecules with and without J Chain in Serum
649 and after Purification, Studied by Ultra-centrifugation, Electrophoresis, and Electron
650 Microscopy. *Scand J Immunol* **4**, 217–228 (1975).
651 11. Randall, T. D., Brewer, J. W. & Corley, R. B. Direct evidence that J chain regulates
652 the polymeric structure of IgM in antibody-secreting B cells. *Journal of Biological*
653 *Chemistry* **267**, 18002–18007 (1992).
654 12. Erlandsson, L., Andersson, K., Sigvardsson, M., Lycke, N. & Leanderson, T. Mice
655 with an inactivated joining chain locus have perturbed IgM secretion. *Eur. J.*
656 *Immunol.* **28**, 2355–2365 (1998).

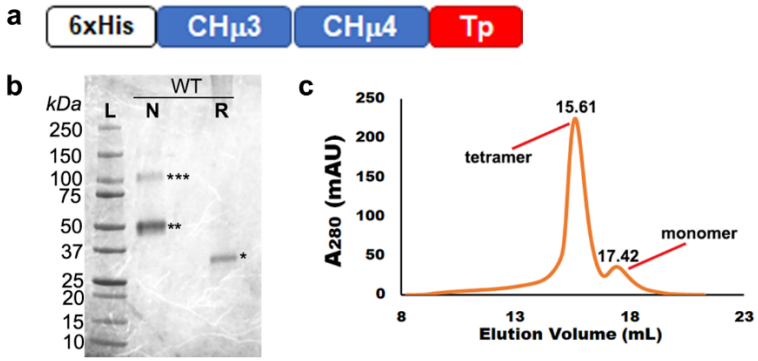
657 13. Kaetzel, C. S. The polymeric immunoglobulin receptor: bridging innate and adaptive
658 immune responses at mucosal surfaces. *Immunol Rev* **206**, 83–99 (2005).
659 14. Macpherson, A. J., Yilmaz, B., Limenitakis, J. P. & Ganal-Vonarburg, S. C. IgA
660 Function in Relation to the Intestinal Microbiota. *Annu. Rev. Immunol.* **36**, 359–381
661 (2018).
662 15. Keyt, B. A., Baliga, R., Sinclair, A. M., Carroll, S. F. & Peterson, M. S. Structure,
663 Function, and Therapeutic Use of IgM Antibodies. *Antibodies* **9**, 53 (2020).
664 16. Pabst, O. New concepts in the generation and functions of IgA. *Nat Rev Immunol*
665 **12**, 821–832 (2012).
666 17. Kaattari, S. *et al.* Varied redox forms of teleost IgM: an alternative to isotypic
667 diversity? *Immunol Rev* **166**, 133–142 (1998).
668 18. Zhang, Y.-A. *et al.* IgT, a primitive immunoglobulin class specialized in mucosal
669 immunity. *Nat Immunol* **11**, 827–835 (2010).
670 19. Stadtmueller, B. M. *et al.* The structure and dynamics of secretory component and
671 its interactions with polymeric immunoglobulins. *eLife* **5**, e10640 (2016).
672 20. Hamuro, K., Suetake, H., Saha, N. R., Kikuchi, K. & Suzuki, Y. A Teleost Polymeric
673 Ig Receptor Exhibiting Two Ig-Like Domains Transports Tetrameric IgM into the
674 Skin. *J Immunol* **178**, 5682–5689 (2007).
675 21. Salinas, I., Fernández-Montero, Á., Ding, Y. & Sunyer, J. O. Mucosal
676 immunoglobulins of teleost fish: A decade of advances. *Developmental &*
677 *Comparative Immunology* **121**, 104079 (2021).
678 22. Kumar Bharathkar, S. *et al.* The structures of secretory and dimeric immunoglobulin
679 A. *eLife* **9**, e56098 (2020).

680 23. Kumar, N., Arthur, C. P., Ciferri, C. & Matsumoto, M. L. Structure of the secretory
681 immunoglobulin A core. *Science* **367**, 1008–1014 (2020).
682 24. Li, Y. *et al.* Structural insights into immunoglobulin M. *Science* **367**, 1014–1017
683 (2020).
684 25. Kumar, N., Arthur, C. P., Ciferri, C. & Matsumoto, M. L. Structure of the human
685 secretory immunoglobulin M core. *Structure* **29**, 564-571.e3 (2021).
686 26. Ahmed, A. A., Keremane, S. R., Vielmetter, J. & Bjorkman, P. J. Structural
687 characterization of GASDALIE Fc bound to the activating Fc receptor FcγRIIIa.
688 *Journal of Structural Biology* **194**, 78–89 (2016).
689 27. Wurzburg, B. A., Garman, S. C. & Jardetzky, T. S. Structure of the Human IgE-Fc
690 Cε3-Cε4 Reveals Conformational Flexibility in the Antibody Effector Domains.
691 *Immunity* **13**, 375–385 (2000).
692 28. Amemiya, C. T. *et al.* The African coelacanth genome provides insights into tetrapod
693 evolution. *Nature* **496**, 311–316 (2013).
694 29. Lalla, C. de, Fagioli, C., Cessi, F. S., Smilovich, D. & Sitia, R. Biogenesis and
695 function of IgM: the role of the conserved μ-chain tailpiece glycans. *Molecular*
696 *Immunology* **35**, 837–845 (1998).
697 30. Pasalic, D. *et al.* A peptide extension dictates IgM assembly. *Proc Natl Acad Sci*
698 *USA* **114**, E8575–E8584 (2017).
699 31. Wyatt, P. J. Light scattering and the absolute characterization of macromolecules.
700 *Analytica Chimica Acta* **272**, 1–40 (1993).
701 32. Hiramoto, E. *et al.* The IgM pentamer is an asymmetric pentagon with an open
702 groove that binds the AIM protein. *Sci. Adv.* **4**, eaau1199 (2018).

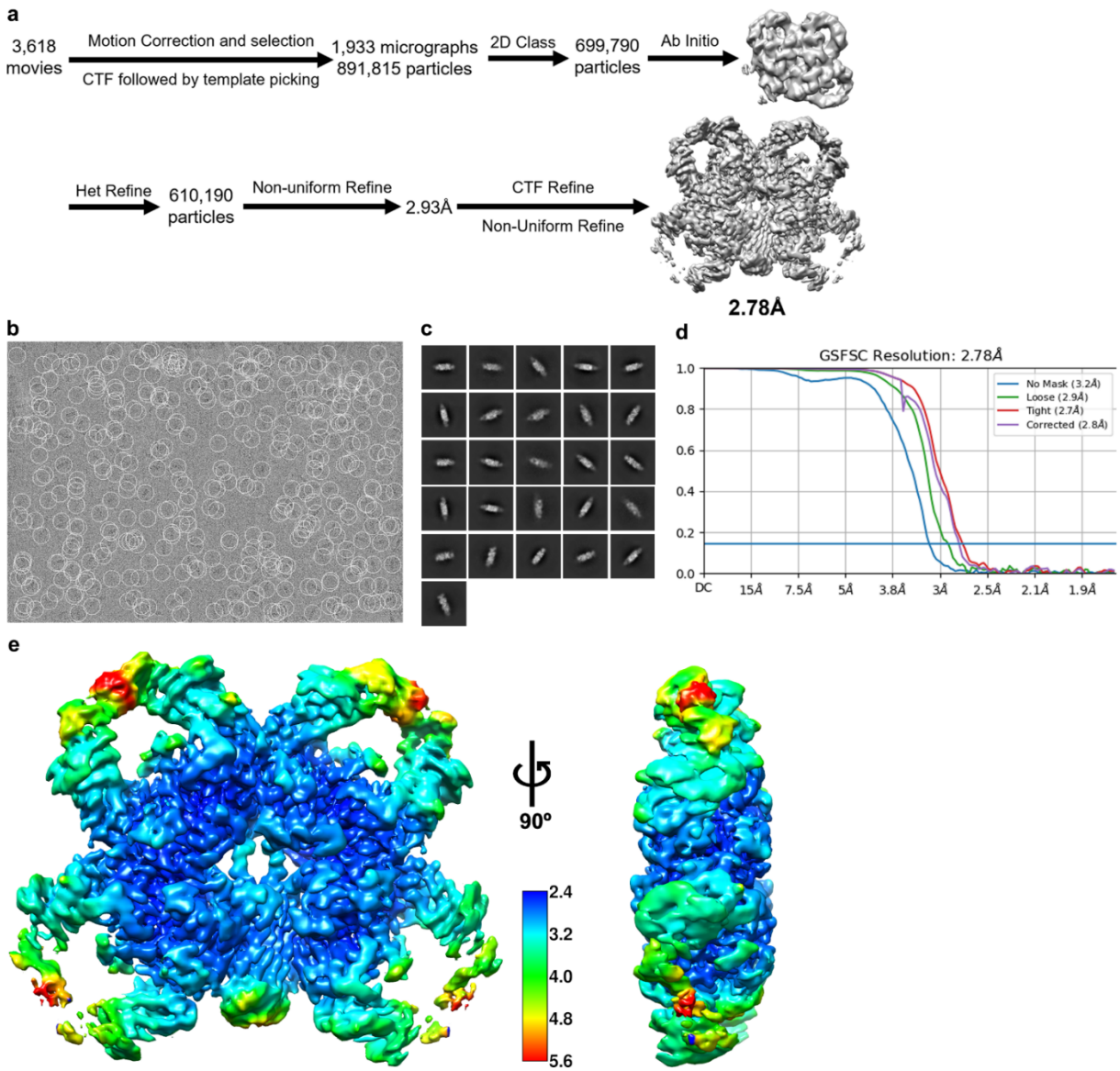
703 33. Muller, R. *et al.* High-resolution structures of the IgM Fc domains reveal principles of
704 its hexamer formation. *Proceedings of the National Academy of Sciences* **110**,
705 10183–10188 (2013).
706 34. Anelli, T. *et al.* Sequential steps and checkpoints in the early exocytic compartment
707 during secretory IgM biogenesis. *EMBO J* **26**, 4177–4188 (2007).
708 35. van Anken, E. *et al.* Efficient IgM assembly and secretion require the plasma cell
709 induced endoplasmic reticulum protein pERp1. *Proc. Natl. Acad. Sci. U.S.A.* **106**,
710 17019–17024 (2009).
711 36. Ye, J., Kaattari, I. M., Ma, C. & Kaattari, S. The teleost humoral immune response.
712 *Fish & Shellfish Immunology* **35**, 1719–1728 (2013).
713 37. Ye, J., Bromage, E. S. & Kaattari, S. L. The Strength of B Cell Interaction with
714 Antigen Determines the Degree of IgM Polymerization. *The Journal of Immunology*
715 **184**, 844–850 (2010).
716 38. Nayak, D. K., Tang, A., Wilson, M., Miller, N. W. & Bengtén, E. Channel catfish
717 soluble FcμR binds conserved linear epitopes present on Cμ3 and Cμ4. *Molecular*
718 *Immunology* **47**, 1306–1316 (2010).
719 39. Sheng, X., Qian, X., Tang, X., Xing, J. & Zhan, W. Polymeric Immunoglobulin
720 Receptor Mediates Immune Excretion of Mucosal IgM–Antigen Complexes Across
721 Intestinal Epithelium in Flounder (Paralichthys olivaceus). *Front. Immunol.* **9**, 1562
722 (2018).
723 40. Chen, Q., Menon, R., Calder, L. J., Tolar, P. & Rosenthal, P. B. Cryomicroscopy
724 reveals the structural basis for a flexible hinge motion in the immunoglobulin M
725 pentamer. *Nat Commun* **13**, 6314 (2022).

726 41. Hu, Y.-L., Xiang, L.-X. & Shao, J.-Z. Identification and characterization of a novel
727 immunoglobulin Z isotype in zebrafish: Implications for a distinct B cell receptor in
728 lower vertebrates. *Molecular Immunology* **47**, 738–746 (2010).
729 42. Feng, L.-N. *et al.* Molecular cloning and functional analysis of polymeric
730 immunoglobulin receptor gene in orange-spotted grouper (Epinephelus coioides).
731 *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular*
732 *Biology* **154**, 282–289 (2009).
733 43. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms
734 for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296
735 (2017).
736 44. Rubinstein, J. L. & Brubaker, M. A. Alignment of cryo-EM movies of individual
737 particles by optimization of image translations. *Journal of Structural Biology* **192**,
738 188–195 (2015).
739 45. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization
740 improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214–1221
741 (2020).
742 46. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and
743 complexes. *Nucleic Acids Research* **46**, W296–W303 (2018).
744 47. Pettersen, E. F. *et al.* UCSF Chimera--A visualization system for exploratory
745 research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
746 48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of
747 *Coot*. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).

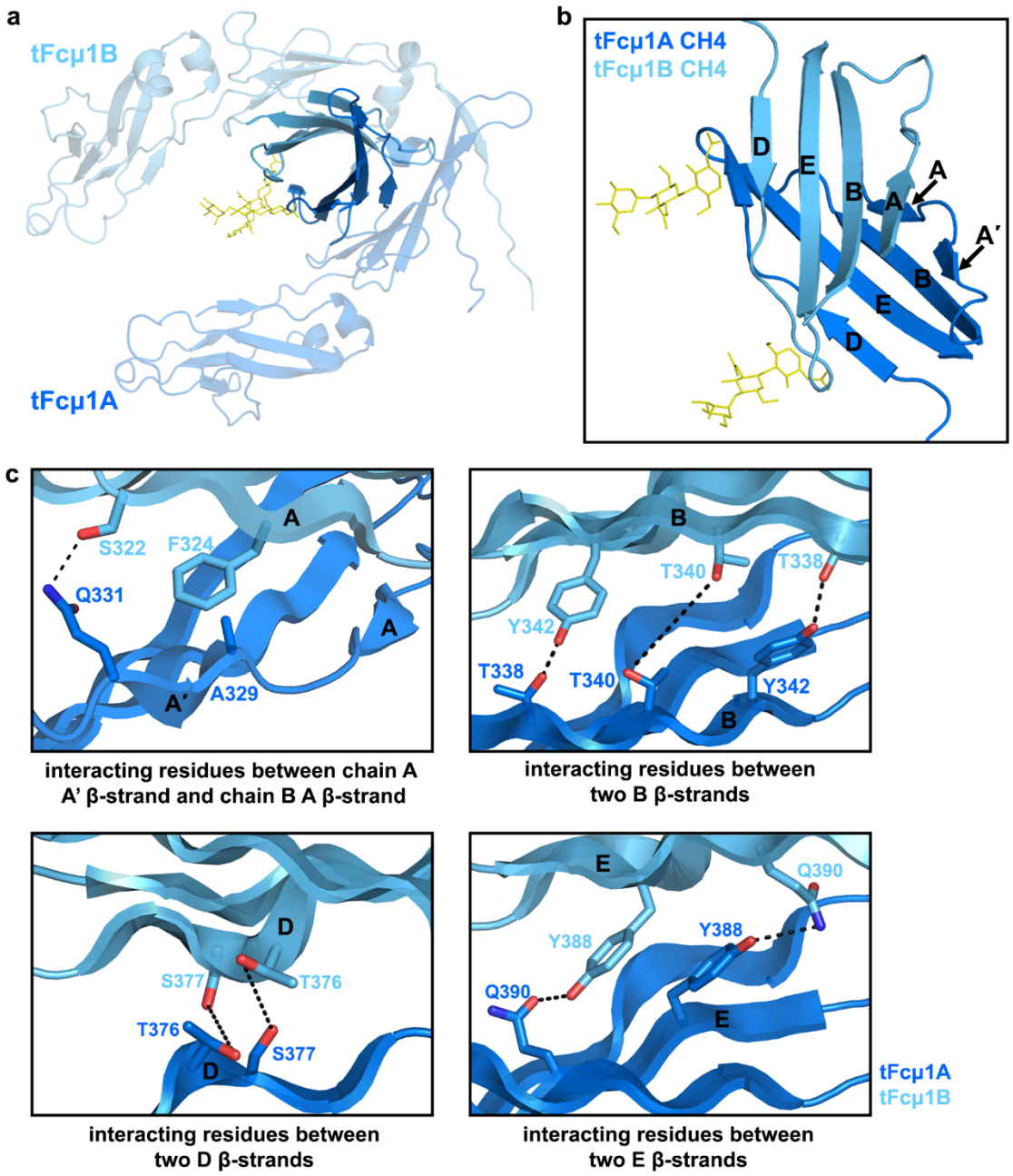
748 49. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons
749 and electrons: recent developments in *Phenix*. *Acta Crystallogr D Struct Biol* **75**,
750 861–877 (2019).
751 50. Barad, B. A. *et al.* EMRinger: side chain–directed model and map validation for 3D
752 cryo-electron microscopy. *Nat Methods* **12**, 943–946 (2015).
753 51. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence
754 alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
755 52. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new
756 ENDscript server. *Nucleic Acids Research* **42**, W320–W324 (2014).
757 53. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
758 54. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline
759 State. *Journal of Molecular Biology* **372**, 774–797 (2007).
760 55. Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and
761 the correlation to protein function. *Pac Symp Biocomput* 310–322 (2002).
762



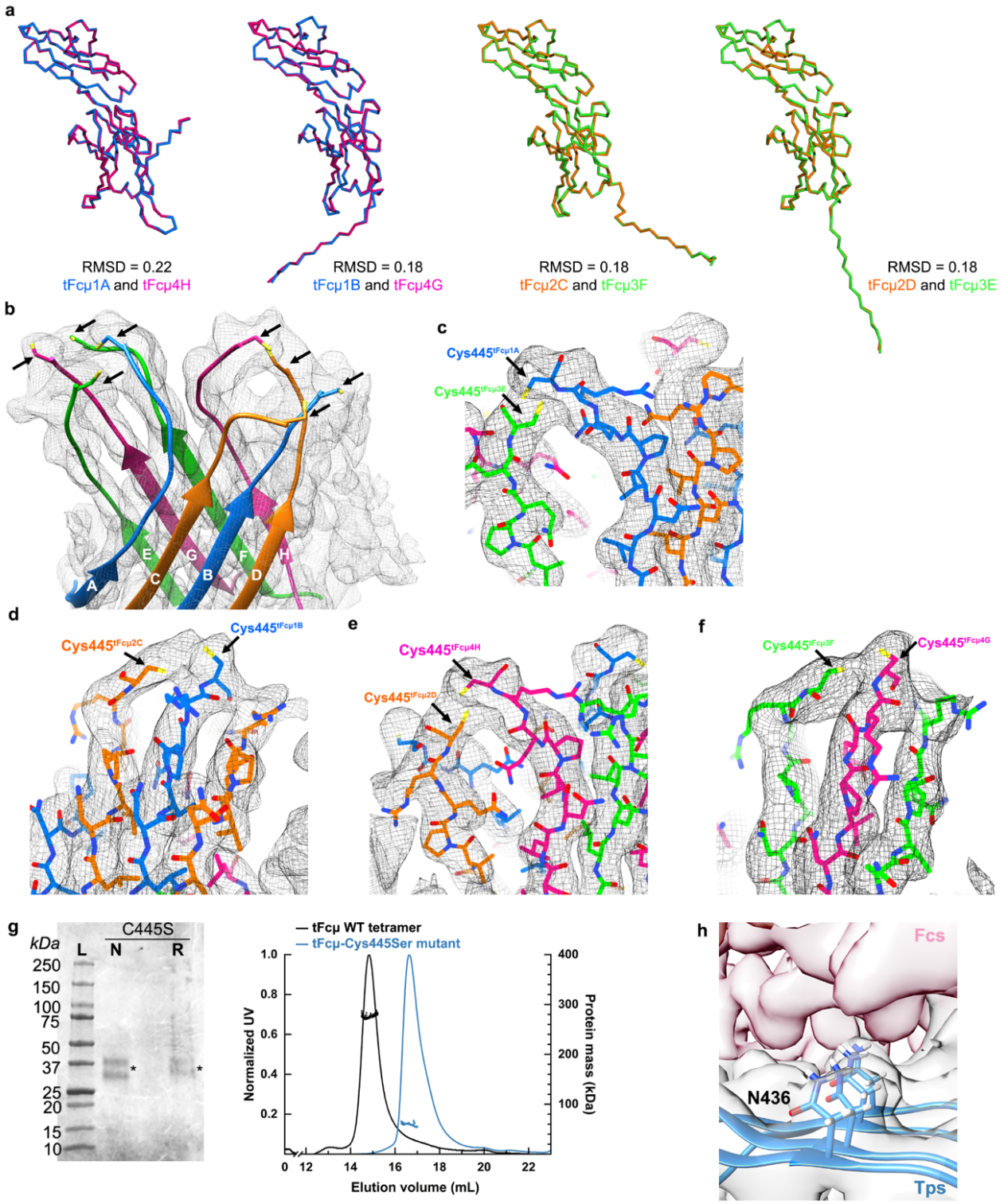
Supplementary Fig. 1 | Design of tFcμ construct and characterization of recombinant tFcμ expression. a Schematic of the tIgM heavy chain domains included in the **tFcμ** expression construct, including an N-terminal hexa-histidine tag. **b** SDS-PAGE gel analysis of wildtype tFcμ tetramer. *WT*, wildtype recombinant tFcμ protein. *L*, molecular weight standard. *N*, non-reducing conditions. *R*, reducing conditions. Putatively, bands correspond to *, single heavy chain. **, two disulfide-linked heavy chains. ***, higher order disulfide-linked heavy chains. **c** Size exclusion column chromatogram of recombinant tFcμ protein.



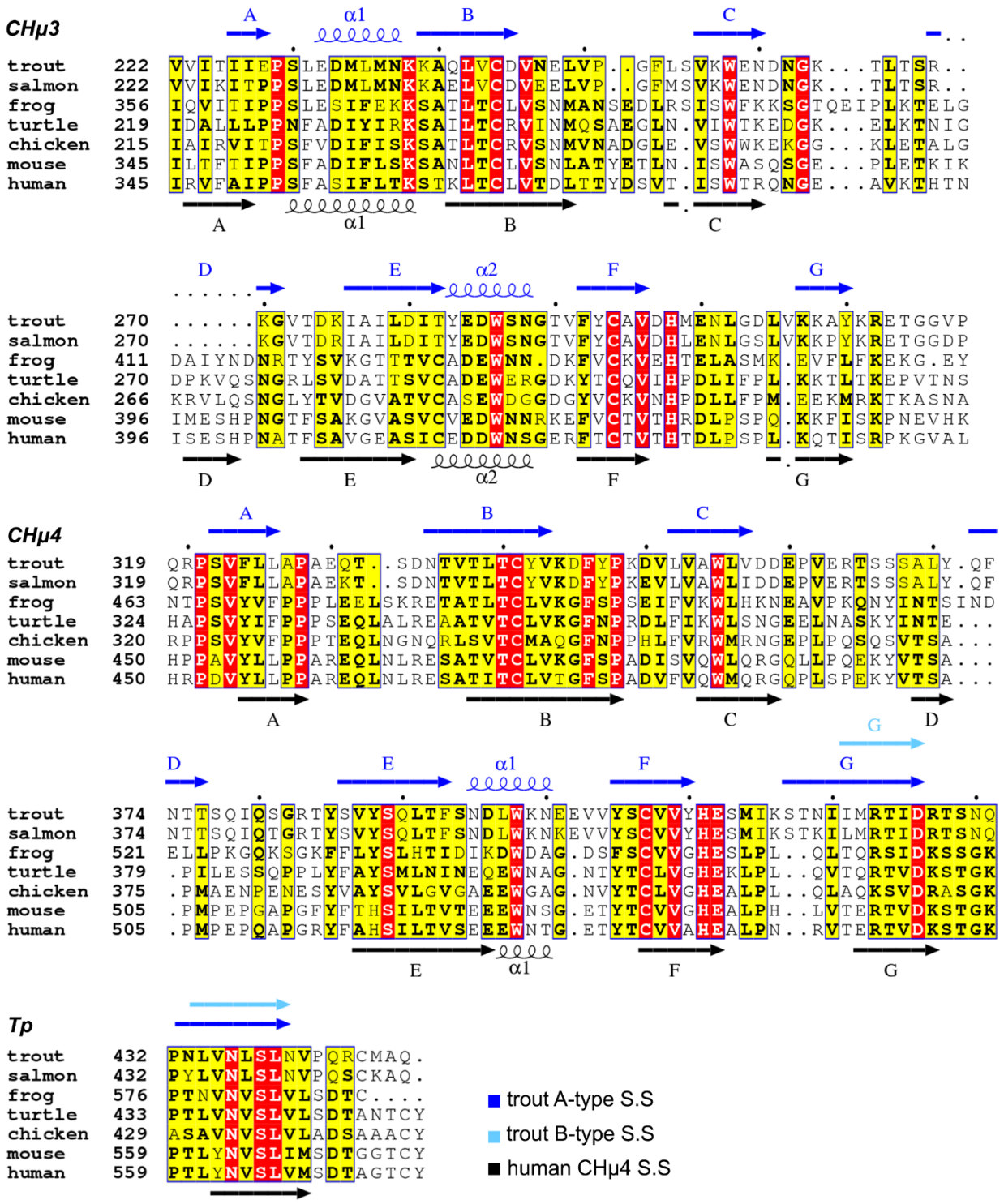
Supplementary Fig. 2 | tFcμ cryo-EM data collection, processing, and validation. a Schematic summary of tFcμ cryo-EM data processing pipeline in CryoSparc. **b** Representative micrograph after motion correction and selection. **c** Representative 2D class averages. **d** FSC curves for the final reconstruction 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).



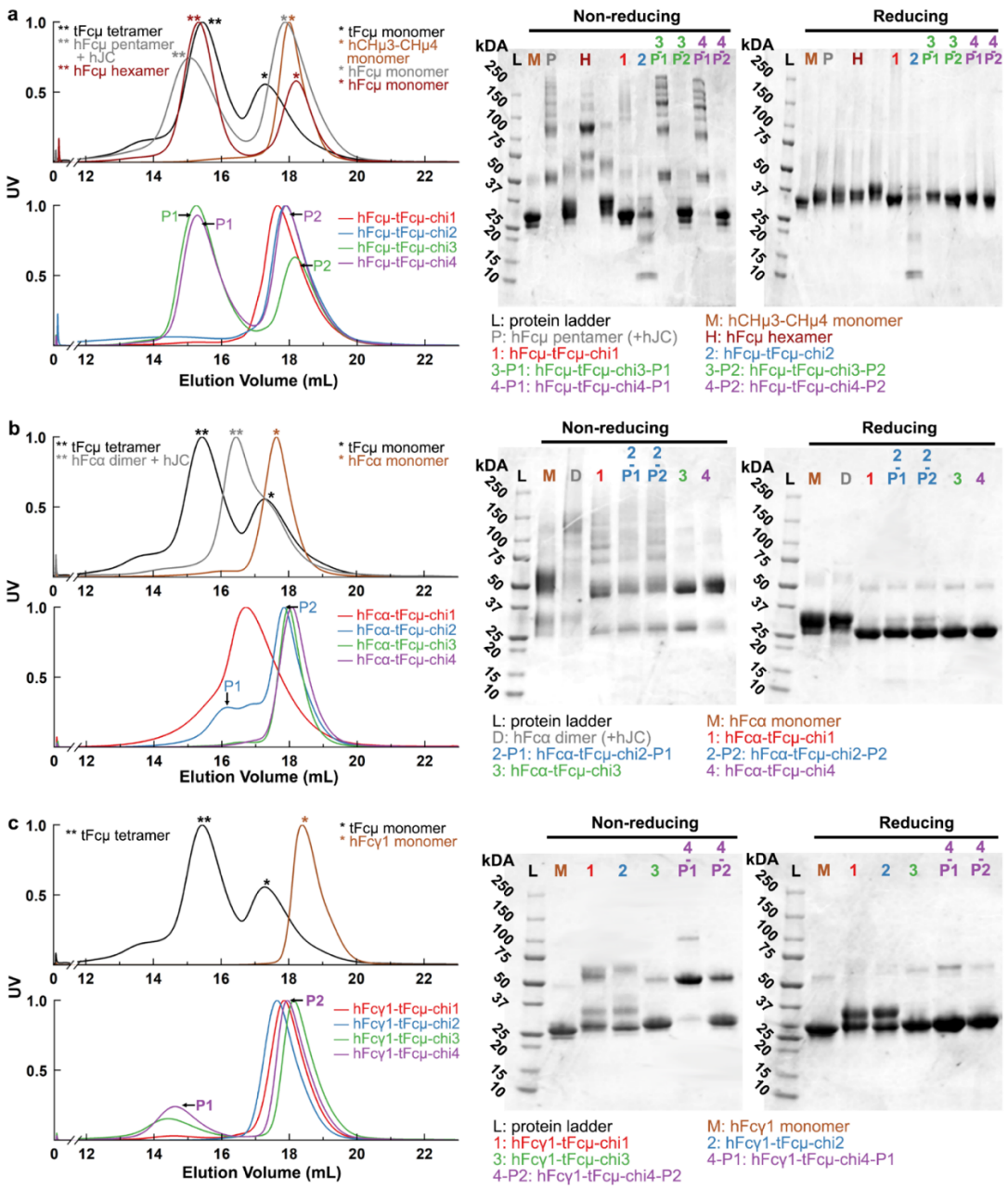
Supplementary Fig. 3 | Interface between two CHμ4 domains within Fcμ monomer. **a** Cartoon representation of tFcμ1 with glycans shown as yellow sticks. The intra-Fc tCHμ4-tCHμ4 interface is rendered using solid colors whereas the rest of the structure is rendered using transparent color. The orientation 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. Chain A contains A and A' strands whereas chain B contains only an A strand followed by a loop. **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; putative hydrogen bonds are shown as black, dashed lines.


a
RMSD = 0.22
tFcμ1A and tFcμ4H
RMSD = 0.18
tFcμ1B and tFcμ4G
RMSD = 0.18
tFcμ2C and tFcμ3F
RMSD = 0.18
tFcμ2D and tFcμ3E
b
A
B
C
D
E
F
G
H
c
Cys445tFcμ1A
Cys445tFcμ3E
d
Cys445tFcμ2C
Cys445tFcμ1B
e
Cys445tFcμ4H
Cys445tFcμ2D
f
Cys445tFcμ3F
Cys445tFcμ4G
g
kDa
L
C445S
N
R
250
150
100
75
50
37
25
20
15
10
tFcμ WT tetramer
tFcμ-Cys445Ser mutant
Normalized UV
Protein mass (kDa)
Elution volume (mL)
h
Fcs
N436
Tps

Supplementary Fig.4 | Local cryo-EM density of areas of interest and structural alignment of tFcμ chains. a Structural alignments between tFcμ chain A and chain H, chain B and chain G, chain C and chain F, chain D and chain E. **b-f** Cryo-EM density near tFcμ C-termini. The coloring scheme is the same as Fig. 1b. **b** Overview of C-termini density with side chains of Cys445 shown as sticks. **c-f** Density around potential interchain disulfide between $Cys445^{Fc\mu1A}$ and $Cys445^{Fc\mu3E}$, between $Cys445^{Fc\mu1B}$ and $Cys445^{Fc\mu2C}$, between $Cys445^{Fc\mu2D}$ and $Cys445^{Fc\mu4H}$, and between $Cys445^{Fc\mu3F}$ and $Cys445^{Fc\mu4G}$, respectively. **g** Characterization of tFcμ-Cys445Ser monomer. *Left*, SDS-PAGE gel. *WT*, wildtype recombinant tFcμ protein. C445S, tFcμ with Cys445Ser mutation. *L*, molecular weight standard. *N*, non-reducing condition. *R*, reducing condition. *, single heavy chain. The doublet band is attributed to heterogeneous glycosylation. *Right*, 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). **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. 5 | 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. 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.

826 **Supplementary Table 1. Cryo-EM data collection and refinement statistics associated**
827 **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		

828

829

830 **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

831

832

833 **Supplementary Table 3. Diameter measurements for the solvent accessible hole at the**
834 **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

835

836 **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

837 **PNGS stands for "potential N-linked glycosylation site".*
838 ***Prediction potential is calculated using NetNGlyc 1.0 developed by DTU Health Tech[52].*
839