ARTICLE

Molecular basis of egg coat cross-linking sheds light on ZP1-associated female infertility

Kaoru Nishimura [1,4], Elisa Dioguardi [1,4], Shunsuke Nishio [2,3], Alessandra Villa [1], Ling Han [1], Tsukasa Matsuda [2] & Luca Jovine [1]

Mammalian fertilisation begins when sperm interacts with the egg zona pellucida (ZP), whose ZP1 subunit is important for fertility by covalently cross-linking ZP filaments into a three-dimensional matrix. Like ZP4, a structurally-related component absent in the mouse, ZP1 is predicted to contain an N-terminal ZP-N domain of unknown function. Here we report a characterisation of ZP1 proteins carrying mutations from infertile patients, which suggests that, in human, filament cross-linking by ZP1 is crucial to form a stable ZP. We map the function of ZP1 to its ZP-N1 domain and determine crystal structures of ZP-N1 homodimers from a chicken homolog of ZP1. These reveal that ZP filament cross-linking is highly plastic and can be modulated by ZP1 fucosylation and, potentially, zinc sparks. Moreover, we show that ZP4 ZP-N1 forms non-covalent homodimers in chicken but not in human. Together, these data identify human ZP1 cross-links as a promising target for non-hormonal contraception.

[1] Department of Biosciences and Nutrition and Center for Innovative Medicine, Karolinska Institutet, Huddinge SE-141 83, Sweden. [2] Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa, Nagoya 464-8601, Japan. [3] Present address: Department of Biosciences and Nutrition and Center for Innovative Medicine, Karolinska Institutet, Huddinge SE-141 83, Sweden. [4] These authors contributed equally: Kaoru Nishimura, Elisa Dioguardi. Correspondence and requests for materials should be addressed to L.J. (email: luca.jovine@ki.se)

Vertebrate oocytes are surrounded by a specialised extracellular coat that is referred to as zona pellucida in mammals (ZP) and vitelline envelope in non-mammals (VE). This matrix plays essential roles in oogenesis and provides a physical barrier that protects the developing embryo; in tetrapods, it also mediates the initial interaction between gametes at fertilization and, in mammals, contributes to the block to polyspermy[1,2]. Although their fine morphological appearance differs between classes, all vertebrate egg coats consist of filaments made up of a variable number of glycoprotein components that polymerise using a common ZP module[3,4]. This unit, which consists of two structurally related immunoglobulin (Ig)-like domains (ZP-N and ZP-C), is also the building block of the generally more complex coats of invertebrate eggs; moreover, it is contained in many additional extracellular proteins not involved in fertilisation[5].

The human ZP contains four glycoproteins (hZP1-4), only three of which (mZP1-3) are found in the mouse due to pseudogenisation of Zp4[6–8]. While mZP2 and mZP3 are the major components of ZP filaments[6] and have long been implicated in the interaction with sperm[9–11], less abundant mZP1 was shown to covalently cross-link ZP filaments by forming homodimers held together by intermolecular disulphide bridge(s)[12]. This architecture is consistent with the observation that Zp2 or Zp3 null mice lack a ZP[13–15], whereas Zp1 null animals have an egg coat that is, however, looser and fragile[16]. Moreover, the structural similarity between the ZP modules of ZP2 and ZP1[5,17] suggests that cross-link points are introduced into ZP filaments when mZP1 is occasionally incorporated instead of mZP2[5,12].

Analysis of native ZP material suggests that the filament cross-linking function of ZP1 is also conserved in human;[18,19] on the other hand, the biological role of ZP4 (previously referred to as ZPB[8,20,21]) remains unknown[22]. However, the two proteins are structurally related by both containing a trefoil domain immediately before their ZP module;[23,24] moreover, ZP1 and ZP4 have been suggested to harbour a single ZP-N-like domain at their N-termini[25,26]. The possible function of this putative element—which is not found in fish homologues of ZP1[27]—remains nonetheless unclear, considering that multiple copies of isolated ZP-N domains at the N-terminus of ZP2 as well as mollusk vitelline envelope receptor for lysin (VERL) have been shown to regulate sperm binding[28,29].

Notably, a ZP-N signature can also be recognised at the N-terminus of the avian homologues of ZP1 and ZP4[25,26]. In the case of the former, which together with ZP3 as well as a different peripherally associated subunit (ZPD) is the major constituent of the bird VE[30,31], the putative N-terminal ZP-N domain is separated from the trefoil domain by a P/Q-rich repeated sequence region that is also conserved in fish and reptilian homologues of the protein[27,30,32]. Moreover, unlike ZP3 and ZPD, which are secreted by the granulosa cells surrounding the egg, avian ZP1 is produced in the liver and reaches the oocyte via the blood circulation[32]. Interestingly, the soluble precursor of ZP1 is monomeric, suggesting that the protein only forms intermolecular disulphides upon incorporation into the growing VE[33]. On the other hand, avian ZP4 is synthesised by the ovary and only expressed in limited amounts during the early stage of folliculogenesis, so that it remains largely localised within the germinal disc region of the eggs where ZP2 is also found[34].

Because Zp2[−/−] or Zp3[−/−] mice are completely infertile[13–15], whereas Zp1[−/−] animals are subfertile[16], only mZP2 and mZP3 are thought to be crucial for mouse fertility; similarly, although a contribution of hZP1 and hZP4 to human sperm binding cannot be completely ruled out[35], this possibility is not supported by studies of hZP1 rescue mice[36] or transgenic animals expressing hZP4[22]. As a result, the large majority of previous biochemical work focused on ZP2 and ZP3, which have also been investigated by X-ray crystallography[26,29,37,38]. On the contrary, there is no structural information on either ZP1 or ZP4, and the hypothesis that these proteins share a similar function due to their common sequence features remains speculative.

Against this background, the recent identification of different ZP1 mutations in infertile patients[39–43] argues for a much more important role of this glycoprotein in human reproduction than previously recognised. To gain insights into the biological function of ZP1 and its possible relation with that of ZP4, we undertook a biochemical and structural investigation that began with the analysis of a reported case of primary female infertility associated with a frameshift mutation in the human ZP1 gene (I390fs404X)[39]. Our data reveal that the molecular basis of egg coat filament cross-linking by ZP1 is conserved between birds and mammals, whereas the function of ZP4 appears to have diverged among vertebrates. Most importantly, in addition to suggesting that different cross-link conformations may contribute to egg coat dynamics and function, this study provides valuable information to understand ZP1-associated female infertility and identifies human ZP1 as a possible contraceptive target.

Results

ZP1 I390fs404X mutation effect on human ZP subunit secretion.
Although it was shown that oocytes from infertile women homozygous for the ZP1 I390fs404X mutation lack of a ZP[39], the effect of the frameshift at the protein level was not investigated. To address this point, we used human embryonic kidney (HEK) 293T cells to compare the expression of the mutant protein, which is truncated shortly after the ZP-N domain (hZP1Mut), to that of wild-type hZP1 (Fig. 1a). This showed that, whereas hZP1 is secreted into the medium as a mixture of monomer and disulphide-bonded dimer, secretion of hZP1Mut is completely abolished (Fig. 1b, left panel). Although the expression of hZP1Mut is lower than that of hZP1, the truncated mutant protein can be detected in the cell lysate (Fig. 1b, right panel). To investigate whether intracellularly retained hZP1Mut affects secretion of the other human ZP subunits (hZP2-4), we performed co-expression experiments (Fig. 1c, d). As shown in Fig. 1c, secretion of hZP1 increases in the presence of hZP2-4, but co-expression with the latter does not rescue secretion of hZP1Mut. Notably, the secretion levels of hZP2, hZP3 or hZP4 are comparable upon co-expression with either hZP1 or hZP1Mut (Fig. 1d). This data suggest that the ZP1 I390fs404X mutation does not compromise ZP matrix assembly by interfering with the secretion of the other human ZP subunits[39], but rather by affecting the filament cross-linking function of hZP1.

Identification of the ZP1 region cross-linking ZP filaments.
To determine which part of ZP1 is responsible for introducing crosslinks in the ZP, we took advantage of the natural abundance of this subunit in the chicken VE—where it also forms intermolecular disulphides—and the fact that it is specifically degraded upon sperm penetration[44]. Incubation of native VE material with a sperm protease extract releases a soluble ~21 kDa N-terminal fragment of chicken ZP1 (cZP1) that exists both as a monomer and an intermolecularly disulphide-bonded homodimer (Fig. 2a–f). Accordingly, a recombinant construct corresponding to the predicted ZP-N1 domain of cZP1 (cZP1-N1) is secreted both as a monomeric and a disulphide-bonded dimer, whereas a protein encompassing the trefoil domain and C-terminal ZP module of cZP1 (cZP1$_{L600-R958}$; Fig. 2a) does not form intermolecular disulphides (Fig. 2g). Thus, by recapitulating the oligomeric status of full-length cZP1 (Fig. 2h), cZP1-N1 harbours its cross-linking function. Importantly, this conclusion can be extended to the
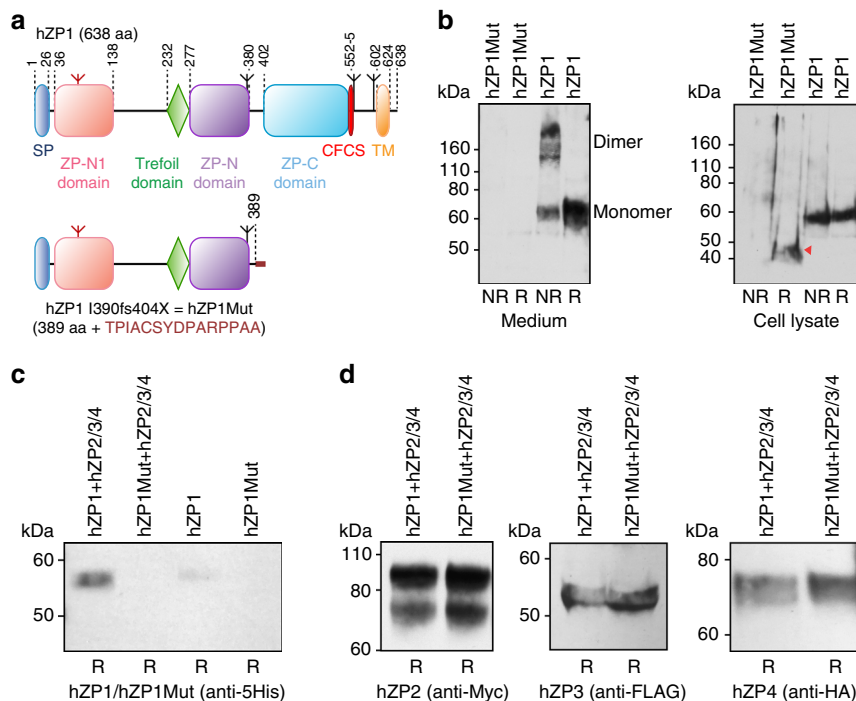
Fig. 1 ZP1 I390fs404X mutant protein does not affect the secretion of hZP2-4. **a** Domain architecture of wild-type and mutant hZP1. SP signal peptide, CFCS consensus furin cleavage site, TM transmembrane domain. Inverted tripods mark N-glycosylation sites, with the conserved N-glycan of the ZP-N1 domain coloured red. Domain boundaries are indicated, and the 15-residue C-terminal extension resulting from the frameshift mutation is highlighted in dark red. **b** Anti-5His immunoblots of medium (left; 2.4 mL per lane) and cell lysate (right; 0.3 mL per lane) samples from HEK293T cells expressing hZP1 constructs show that hZP1Mut is retained intracellularly (red arrow). NR non-reducing conditions, R reducing conditions. **c** Co-expression with human ZP2-4 increases secretion of wild-type hZP1, but does not rescue secretion of hZP1Mut (240 μL medium per lane). **d** The secretion levels of hZP2, hZP3 and hZP4 do not change upon co-expression with either hZP1 or hZP1Mut (240 μL medium per lane)

mammalian homologue of the protein because, mirroring full-length hZP1 (Fig. 1b, left panel), a construct corresponding to hZP1-N1 is also secreted as a homodimer linked by intermolecular disulphide(s) (Fig. 2i). Moreover, size-exclusion chromatography with multi-angle static light-scattering (SEC-MALS) analysis of a construct including the trefoil domain and ZP module of hZP1 (hZP1$_{A195-P590}$; Fig. 1a) excludes the possibility that this region of the molecule contributes to its cross-linking function via non-covalent interactions (Fig. 2j).

Three-dimensional structures of the cZP1-N1 cross-link. To establish the molecular basis of ZP1 cross-linking, we investigated the cZP1-N1 homodimer by X-ray crystallography. cZP1-N1 is chemically heterogeneous due to the presence of two N-glycosylation sites (N65 and N121), only the first of which is highly conserved and also found in hZP1 (Supplementary Fig. 1a). Accordingly, whereas mutation of cZP1-N1 N65 or hZP1-N1 N76 drastically reduces protein secretion, a cZP1 N121Q mutant is secreted as well as the wild-type and—because of its higher homogeneity—was used for crystallisation trials in parallel with hZP1-N1 (Fig. 3a). Whereas the latter did not crystallise, we could grow crystals of cZP1-N1 expressed in human embryonic kidney (HEK) 293S GnTI- cells and treated with Endoglycosidase (Endo) H (Fig. 3b), as well as the same protein expressed in HEK293T cells (Fig. 3c). The structure of deglycosylated cZP1-N1 was determined by gold- and zinc-single-wavelength anomalous dispersion (SAD) phasing; the resulting model was then used to solve by molecular replacement the structure of glycosylated cZP1-N1 (Fig. 3d–g; Supplementary Fig. 2 and Table 1).

Despite very low sequence identity, the protein adopts the same modified immunoglobulin (Ig)-like fold of the ZP2-N1 and ZP3

ZP-N domains[5,26,29,37], including two intramolecular disulphides with 1-4, 2–3 connectivity (C$_1$32-C$_4$124 and C$_2$69-C$_3$90), a Tyr residue located in β-strand F next to C$_4$ (Y99) and an E′ strand (L93-K95) (Fig. 4). However, the cd loop of cZP1-N1 includes an additional β-strand (C′) that forms a β-hairpin with strand C. Together with invariant V63, which interacts hydrophobically with the CD face, this positions an additional Cys (C66) so that it can form an intermolecular disulphide with the same residue from another molecule (Figs 4, 5).

Remarkably, although they all contain the same C66-C66 cross-link, the moieties of the cZP1-N1 homodimers observed in our crystal forms adopt three distinct relative orientations (Fig. 5a). These correspond to an asymmetric interface in the case of deglycosylated cZP1-N1 (whose crystallographic asymmetric unit contains a homodimer) and two related but not identical interfaces in the crystal of glycosylated cZP1-N1 (whose asymmetric unit also contains two copies of the protein (chains A and B) that, however, make intermolecular cross-links with symmetry-related copies of themselves (A′ and B′)). All the cZP1 cross-links have small interface areas that range from 330/412 Å^2 (A-A′/B-B′ dimers of glycosylated cZP1-N1) to 575 Å^2 (deglycosylated cZP1-N1 dimer). Moreover, in addition to C66, they involve a common number of invariant or highly conserved residues, including F61, N65, S67, the C$_2$69-C$_3$90 intramolecular disulphide and H91, as well as I68 (conserved in cZP1, hZP1 and mZP1) (Supplementary Fig. 1a). Distinct sets of interactions mediated by these residues underlie the different homodimer interfaces: for example, in the deglycosylated cZP1-N1 homodimer, C′ strand F61(A) interacts aromatically with W72(B) and I68(A) packs against I68(B) (Fig. 5b); on the other hand, in both glycosylated cZP1-N1 homodimers, the latter interaction is missing because I68 interacts hydrophobically with F61 of the
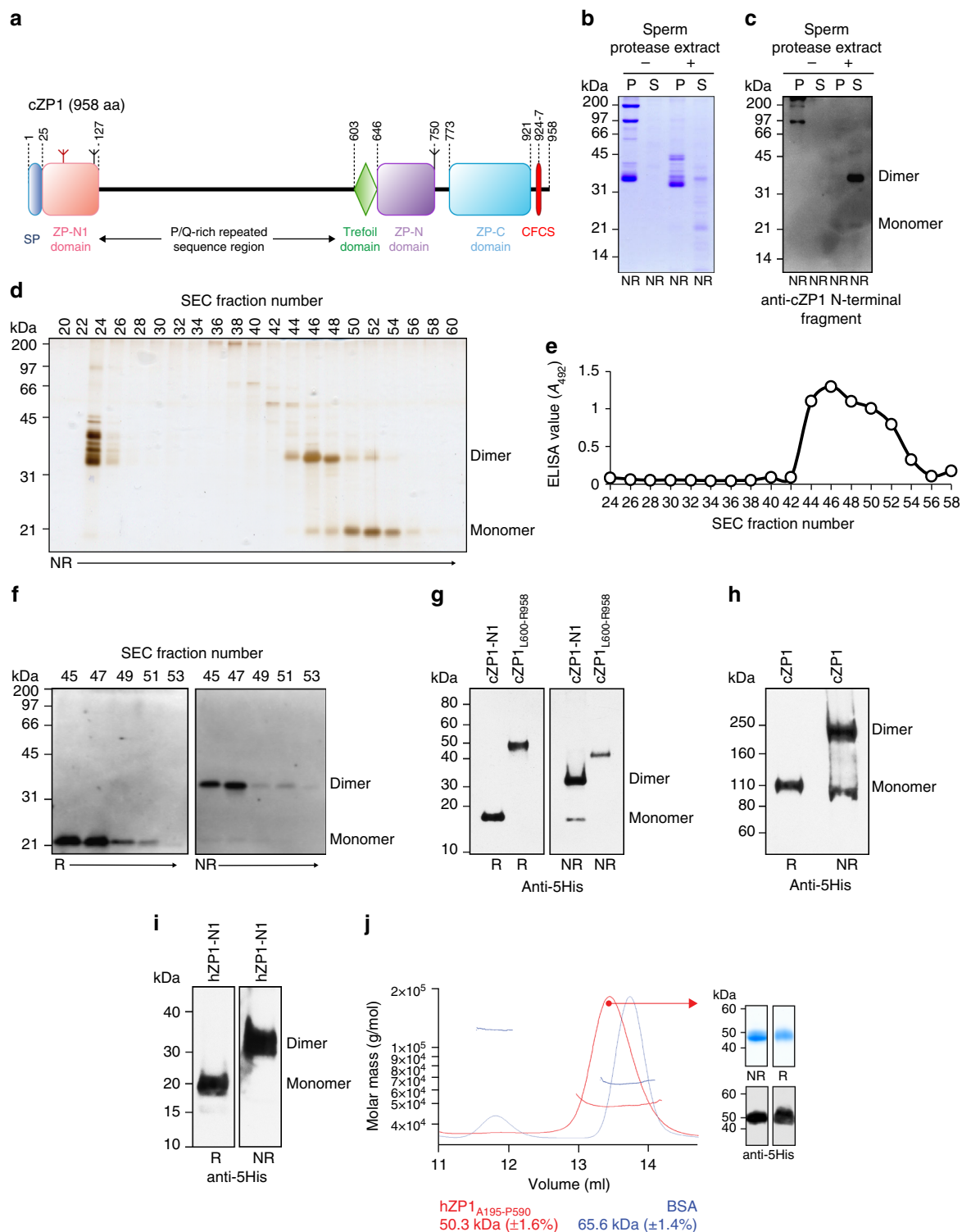
opposite subunit (Fig. 5c). Similarly, while H91(A) stacks against W72(B) in the deglycosylated cross-link structure (Fig. 5d), H91 faces I68 in both of the glycosylated cross-links (Fig. 5c).

Additional contacts between residues in the ab loop/β-strand B region and amino acids in the cd loop stabilises the asymmetric structure of the deglycosylated cZP1-N1 cross-link; in particular, invariant Q38(B) stacks against highly conserved F57(A) and also makes water-mediated hydrogen bonds with the side chains of T59(A) and invariant D55(A) (Fig. 5e). In the case of glycosylated cZP1-N1, on the other hand, the relatively more extensive B/B′ molecule interface is stabilised by a long-range hydrogen bond

between the main-chain carbonyl oxygen of E62 and the side-chain oxygen of conserved S67 (Fig. 5f).

The molecular basis of ZP1-N1 homodimerisation is conserved.
Consistent with the evolutionary conservation of C66 (Fig. 6a; Supplementary Fig. 1a) and the observation that the homodimer interfaces observed in the cZP1-N1 crystals bury a limited fraction of the surface area, SDS-PAGE analysis of SEC fractions show that mutation of C66 abolishes protein homodimerisation (Fig. 6b, c). Similarly, mutation of the corresponding residues of

Fig. 2 The cross-linking function of ZP1 maps to its N-terminal ZP-N1 domain. **a** Domain architecture of cZP1, with feature names abbreviated as in Fig. 1a. **b** Coomassie-stained SDS-PAGE analysis of insoluble (pellet; P) and soluble (S) fractions of native chicken egg coat material that was either untreated or incubated with a sperm protease extract prepared from acrosome-reacted sperm[44]. **c** Immunoblot analysis using anti-cZP1 N-terminal fragment polyclonal[57] confirms the presence of cZP1-N1 in the soluble fraction of egg coat material treated with sperm extract. **d** Silver-stained SDS-PAGE analysis of a size-exclusion chromatography (SEC) separation of the sperm protease-solubilised egg coat material. **e** SEC fractions 44–54 contain cZP1-N1 as evidenced by ELISA analysis using anti-cZP1 N-terminal fragment polyclonal[57]. Bands corresponding to the dimer and monomer of cZP1-N1 released from the egg coat were in-gel digested with chymotrypsin and trypsin, followed by MALDI-MS/MS to confirm their protein sequences. **f** Immunoblot analysis of SEC fractions 45–53 indicates that the ~35 kDa protein in these fractions is a disulphide-linked dimer of the ~21 kDa protein containing cZP1-N1. **g, h** Anti-5His immunoblot of recombinant cZP1-N1 and cZP1$_{L600-R958}$ (panel **g**), as well as full-length cZP1 (panel **h**). Only cZP1-N1 forms a covalent homodimer like full-length cZP1. **i** Anti-5His immunoblot analysis shows that secreted hZP1-N1 also forms a covalent homodimer. **j** SEC-MALS analysis shows that purified hZP1$_{A195-P590}$ produced in HEK293S cells has a molecular mass of ~50 kDa (red profile), consistent with the calculated mass of a monomeric species carrying 2 GlcNac residues (43 kDa). Calibration was carried out using BSA, whose molecular mass detection is shown in blue
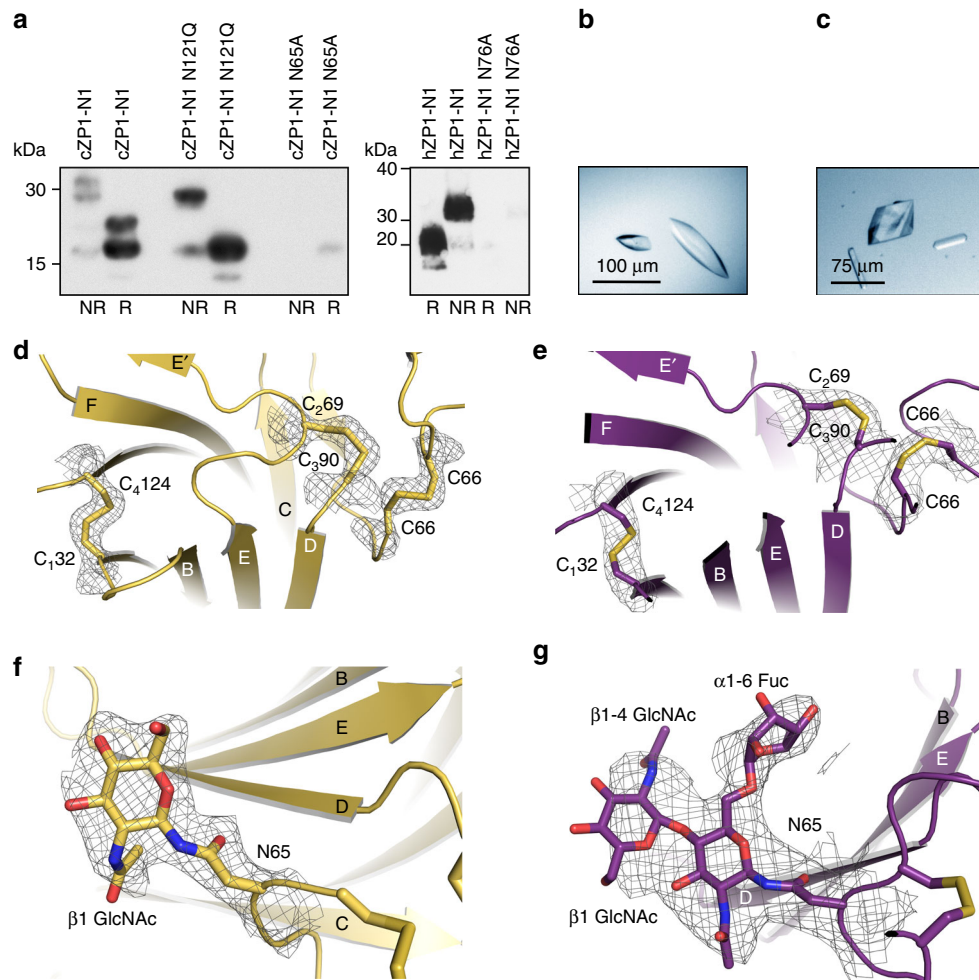

Fig. 3 Structure determination of Endo H-deglycosylated and glycosylated cZP1-N1. **a** Anti-5His immunoblot shows that mutation of N-glycosylation site N65, but not N121, significantly reduces the expression of cZP1-N1 (left panel). Similarly, hZP1-N1 with a N76A mutation is essentially not secreted (right panel). All lanes correspond to 20 μL medium. **b** Crystals of Endo H-deglycosylated cZP1-N1. **c** Crystals of glycosylated cZP1-N1. **d, e** Details of the Endo H-deglycosylated (panel **d**) and glycosylated (panel **e**) cZP1-N1 structures, depicted in cartoon representation with canonical ZP-N domain disulphides and the C66 cross-link shown as sticks. Relevant portions of the *2mFo-DFc* electron density maps contoured at 1.0 σ are shown as black meshes. **f, g** Detailed views of the N65-linked glycans of Endo H-treated (panel **f**) and glycosylated cZP1-N1 (panel **g**). Electron density maps are depicted as in panels **d** and **e**

hZP1-N1 and mZP1-N1 (C77 and C69, respectively) completely disrupts their ability to homodimerise (Fig. 6d–g). Notably, whereas the conditioned medium of cells expressing cZP1-N1 contains a small fraction of monomeric protein in addition to the prevalent homodimeric species, secreted hZP1-N1 and mZP1-N1 are entirely dimeric (compare Fig. 6c with Fig. 6e–g). Together, these data indicate that formation of the conserved inter-molecular disulphide is required for ZP1-N1 homodimerisation

in both birds and mammals, although the efficiency of this process may vary between the two.

ZP1-N1 fucosylation and infertility-associated mutation W83R. C66 is part of an almost invariant sequon (Fig. 6a; Supplementary Fig. 1a), whose glycosylation is crucial for cZP1-N1 and hZP1-N1 secretion (Fig. 3a). Interestingly, heterogeneous

Table 1 X-ray data collection and refinement statistics

	cZP1 ZP-N1 (HEK293S) [PDB: 6GF6]	cZP1 ZP-N1 Zn²⁺ (HEK293S) [PDB: 6GF7]	cZP1 ZP-N1 (HEK293T) [PDB: 6GF8]
Data collection			
Resolution range (Å)	37.0–2.30 (2.40–2.30)	53.3–2.70 (2.82–2.70)	24.7–3.10 (3.24–3.10)
Space group	$P\,4_1 2_1 2$	$P\,4_1 2_1 2$	$C\,2\,2\,2_1$
Unit cell			
a, b, c (Å)	75.98, 75.98, 101.88	75.41, 75.41, 100.78	90.77, 125.73, 105.00
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Total reflections	267258 (11835)	92642 (6475)	45635 (5509)
Unique reflections	13549 (1420)	15254 (1896)	11096 (1326)
Multiplicity	19.7 (8.3)	6.1 (3.4)	4.1 (4.2)
Completeness (%)	97.9 (84.5)	99.9 (99.2)	98.7 (96.4)
Mean I/sigma (I)	23.81 (1.49)	13.63 (1.06)	10.49 (1.18)
Wilson B-factor (Å²)	38.90	70.30	111.38
R_{meas} (%)	11.6 (149.8)	10.3 (162.7)	9.5 (111.2)
R_{pim} (%)	2.6 (49.5)	4.1 (85.2)	4.4 (51.9)
$CC_{1/2}$	1.00 (0.55)	1.00 (0.51)	1.00 (0.64)
CC^*	1.00 (0.84)	1.00 (0.82)	1.00 (0.89)
Refinement			
Reflections	13547 (1419)	15254 (1891)	11096 (1319)
Free reflections	1349 (137)	1528 (188)	1070 (122)
R_{work} (%)	19.22 (29.53)	22.18 (36.53)	21.04 (38.88)
R_{free} (%)	22.92 (31.86)	27.18 (43.14)	23.52 (43.51)
CC_{work}	0.960 (0.746)	0.927 (0.610)	0.947 (0.816)
CC_{free}	0.935 (0.743)	0.905 (0.438)	0.930 (0.696)
Number of non-H atoms	1834	1723	1736
	1712	1660	1638
Macromolecules			
Ligands	34	37	98
Solvent	88	26	0
Protein residues	216	208	217
RMS deviations			
Bonds (Å)	0.005	0.004	0.003
Angles (°)	0.72	0.72	0.65
Average B-factor	47.54	74.13	140.69
	47.29	73.51	138.49
Macromolecules			
Ligands	68.86	97.41	177.50
Solvent	44.32	80.40	–

amino acid of human ZP1 (W83R) was recently found in an infertile patient lacking the ZP (H. Zhao, personal communication and reference[40]). To investigate the effect of this pathogenic substitution, we compared the expression of a hZP1-N1 W83R mutant to that of the wild-type protein. Consistent with an important role of W72 in the folding and dimerisation of cZP1-N1, this experiment showed that W83R not only decreases the secretion of hZP1-N1 but also significantly reduces its ability to form disulphide-bonded homodimers (Fig. 7b). This suggests that the mutation leads to lack of the ZP by hindering human ZP filament cross-linking.

Two Zn²⁺-binding His of cZP1 affect cross-linking of hZP1-N1. Examination of the zinc-SAD data used for phasing revealed that asymmetric interaction of the cZP1-N1 homodimer moieties creates a negatively charged pocket where H71(A) coordinates a Zn²⁺ ion that faces the $C_2$69(A)-$C_3$90(A) intramolecular disulphide, as well as the C66 cross-link (Fig. 8a). The anomalous difference map of the data also shows a peak near H71(B) but, consistently with the much lower accessibility of the latter due to crystal packing, this site is weaker than the previous (~6 σ vs. 11 σ). In the native electron density maps of cZP1-N1, the Zn²⁺ is missing but the same His residue, together with closely located H91, is part of the homodimer interface in both its asymmetric and symmetric conformations. In the former case, the pocket is occupied by a molecule of glycerol from the cryoprotectant solution and H91(A) forms an aromatic residue cluster with W72 (B) and F61(A) (Figs 5d, 8b). The imidazole ring of H71(A) stacks against the $C_2$69(A)–$C_3$90(A) disulphide, and is also hydrogen bonded to the main chain carbonyl group of $C_3$90(A); at the same time, the main chain atoms of H71(A) stabilise the intermolecular cross-link by hydrogen bonding to the backbone of C66(A) (Fig. 8b). In the case of the symmetric interface, H71 is also H-bonded to $C_3$90 and C66 and, together with H91 from the same moiety of the homodimer, lines the wall of a cavity whose edge interacts hydrophobically with I68 from the other subunit (Fig. 8c). Notably, the two His residues and the W72 are almost completely conserved in ZP1 homologs from other species, including mouse and human (Fig. 6a; Supplementary Fig. 1a).

In the Zn²⁺-soaked crystal, where H71(A) adopts a different orientation and is too far from $C_3$90 (A) to make an hydrogen bond, the $C_2$69(A)-$C_3$90(A) disulphide is partially reduced (most likely due to radiation damage) and H91(A) adopts two alternative conformations, one of which no longer stacks against W72(B) and chelates the Zn²⁺ together with H71(A) itself (Fig. 8d). Accordingly, molecular dynamics (MD) simulations suggest that, in solution, the two His residues can stably coordinate a zinc ion together with four water molecules (Fig. 8e).

Considering that physiological release of zinc from activated oocytes has been suggested to remodel the mammalian ZP[47,48], we evaluated the effect of mutating the conserved His residues on the formation of the intermolecular cross-link of hZP1-N1. Whereas mutation of H102 (corresponding to cZP1 H91) does not hinder protein homodimerisation, a construct carrying a mutation of H82 (corresponding to cZP1 H71) is secreted as a mixture of disulphide-bonded dimers and tetrameric aggregates; combination of the two mutations, on the other hand, drastically interferes with cross-linking of hZP1-N1 (Fig. 8f–h).

Taken together, these observations indicate that homodimerisation of hZP1 depends on the same His residues that form a Zn²⁺-binding site in cZP1-N1.

Oligomerisation properties of human and chicken ZP4-N1. Considering that ZP subunit ZP4 has the same domain architecture as ZP1, its ZP-N1 domain could in principle also harbour

core fucosylation of native mouse and human ZP carbohydrates was observed by mass spectrometry[45,46], which also suggests that the same heterogeneity exists in native cZP1-N1 (Supplementary Fig. 4); however, the relative domain arrangement found in the crystals of deglycosylated cZP1-N1 (Fig. 5a) would not be compatible with the presence of an α1-6-linked fucose (Supplementary Fig. 5). These observations suggest that the structure of the conserved N-glycan of ZP1 could modulate the conformation of ZP filament cross-links by favouring a specific type of ZP1/ZP1 interface.

In the structure of glycosylated cZP1-N1 as well as chain A of the deglycosylated protein, the cd loop containing C66 packs against W72, a highly conserved residue in β-strand D (Fig. 6a; Supplementary Fig. 1a). In chain A of glycosylated cZP1-N1, this residue also stacks against the ring of the α1-6-linked fucose (Fig. 7a); on the contrary, as mentioned above, W72(B) stacks against H91(A) in the asymmetric interface of the deglycosylated homodimer (Fig. 5d). Notably, a mutation of the corresponding
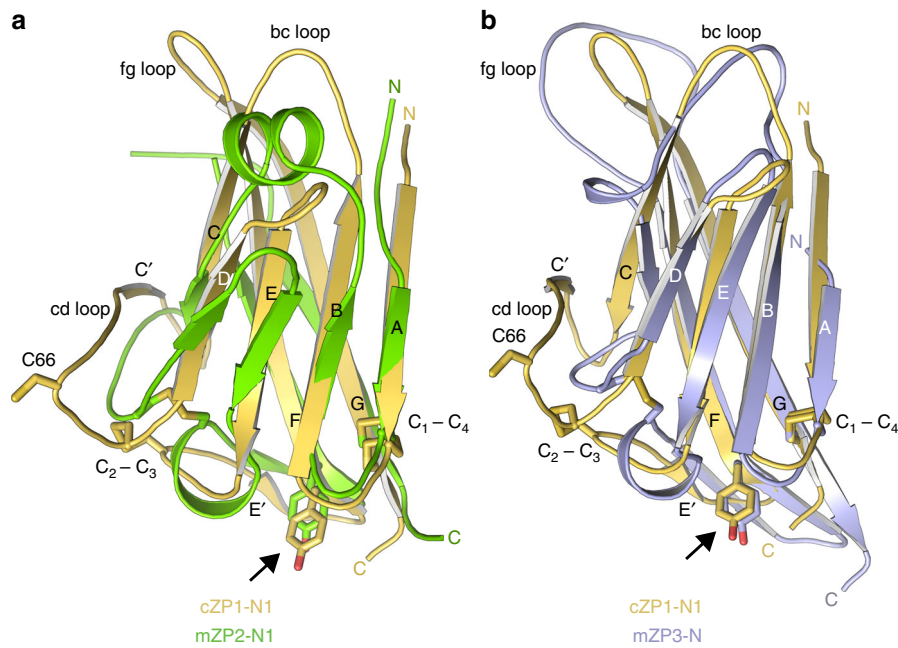
Fig. 4 Structural comparison of ZP-N domains from vertebrate ZP1, ZP2 and ZP3. **a** Superimposition of cZP1-N1 (chain B of the highest resolution structure, PDB 6GF6) and mZP2-N1 (PDB 5II6), with a root-mean-square deviation (RMSD) of 2.1 Å over 74 aligned residues (13% sequence identity). Cross-linking C66, conserved intramolecular disulphides and β-strand F Tyr (arrow) are shown as sticks. **b** Superimposition of cZP1-N1 and mZP3-N (chain A of PDB 5OSQ), with a RMSD of 2.1 Å over 75 aligned residues (11% sequence identity). Note how the bc and fg loops of cZP1-N1 are significantly shorter than those of mZP3-N

a cross-linking function. However, SEC analysis of purified hZP4-N1 (Fig. 9a) as well as pull-down experiments of Myc-tagged hZP4-N1 using an equivalent histidine-tagged construct (Fig. 9b) indicate that the protein is entirely monomeric. On the other hand, parallel experiments using cZP4-N1 indicate that, in the chicken, ZP4 forms a non-covalent homodimer (Fig. 9a–c).

Discussion

Although ZP1 remains by far the less studied component of the mammalian ZP, our characterisation of the hZP1 I390fs404X mutant protein (Fig. 1) suggests that infertile patients carrying the corresponding mutation lack the ZP due to an impairment of hZP1 secretion that in turn causes absence of filament cross-linking, rather than interference with hZP3 secretion as previously hypothesised[39]. Indeed, even if mutated hZP1 could cause intracellular retention of hZP3 or other ZP subunits, this would only affect a small fraction of the latter. This is because hZP2-4 are expressed in significantly higher amounts than hZP1[21,49,50] and—in agreement with observations in the mouse[51]—can be secreted by mammalian cells independently from each other (ref. [19], where hZP4 was misannotated as hZP1[8,20,21]). Such a situation would thus be unlikely to cause complete absence of the ZP, also considering that $Zp3^{+/-}$ heterozygous mice that lack ~50% of ZP3 are still able to assemble a ZP of reduced thickness and, like $Zp2^{+/-}$ animals, retain normal fertility[13–15,52].

Consistent with an essential role of hZP1 and/or hZP4 in human ZP matrix assembly, transgenic mice lacking mouse ZP subunits but expressing only hZP2 and hZP3 cannot form a ZP;[22] moreover, despite its structural similarity to hZP1, hZP4 is clearly not sufficient to rescue the biogenesis or stability of the ZP in ZP1 I390fs404X homozygous patients[39]. The latter observation, together with the fact that the trefoil/ZP module region of hZP1 does not form non-covalent homodimers (Fig. 2j), suggest that hZP1-N1 cross-link formation is essential for the assembly or maintenance of the human ZP and, hence, fertility. Notably, two other ZP1 variants have been found together with I390fs404X in infertile patients with empty follicle syndrome[41,43]. Similarly to I390fs404X itself, both of these mutations are expected to impair the cross-linking ability of hZP1 by either replacing an exposed residue of hZP1-N1 β-strand C with an additional Cys (R61C)[41] or truncating the protein at the level of the hZP1-N1 bc loop (P54fs64X)[43]. Also taking into account the identification of additional hZP1-truncating variants associated with lack of a ZP[42], a picture is starting to emerge whereby different constellations of homozygous or compound heterozygous ZP1 mutations cause primary infertility by hindering human ZP cross-linking and/or oocyte growth. Moreover, in relation with the first aspect, it is apparent that ZP1 mutations can have a damaging effect not only by completely abolishing ZP cross-linking due to impairment of protein secretion (I390fs404X; Fig. 1), but also by specifically affecting hZP1 residues (such as W83R; Fig. 7) whose structural importance affects cross-link formation. Notably, such effects can co-exist: for example, the infertile patient carrying the ZP1 W83R mutation itself was also found to harbour a W471X allele that, similarly to I390fs404X, encodes a C-terminally truncated ZP1 product[40]. Because of the significant sequence identity between cZP1-N1 and its mouse and human counterparts (53% and 51%, respectively; Supplementary Fig. 1a), the crystallographic information reported in this paper provides a solid framework for understanding the effect of the aforementioned mutations on human ZP cross-linking, as well as interpreting additional infertility-associated hZP1-N1 variants that will be reported in the future. So far, the identification of such mutations has been complicated by the fact that they only compromised fertility in the homozygous or compound heterozygous state and, by affecting reproduction, intrinsically limited genetic analyses that depend on family history. However, this situation may rapidly change as a result of the increasing availability of individual whole-exome/genome sequencing.

From a biological point of view, our structural studies validate the hypothesis[25,26] that—as also recently seen in the case of
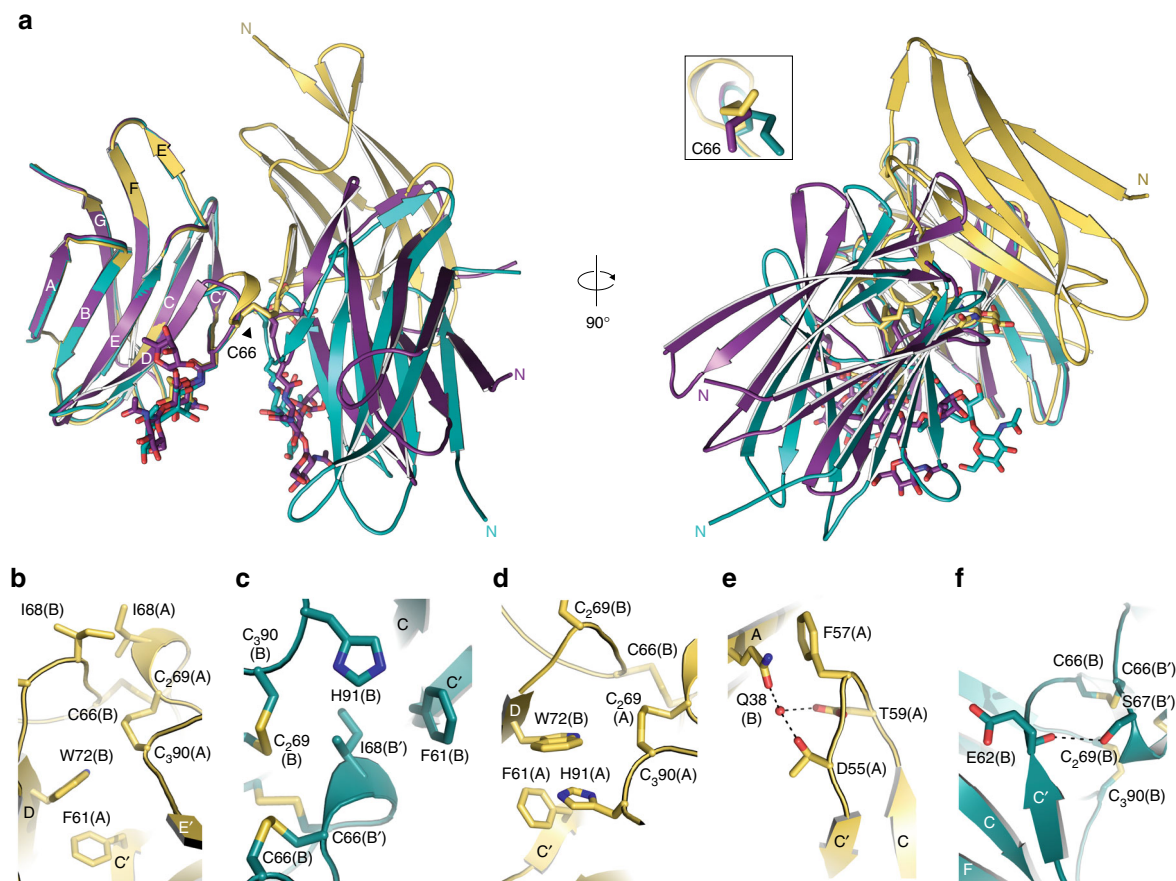
Fig. 5 Conformational variability of the ZP1 cross-link. **a** Different relative arrangements of the subunits of the Endo H-treated (chains A/B, gold) and glycosylated (chains A/A′, violet purple; chains B/B′, teal) cZP1-N1 homodimers. The N-glycan attached to conserved N65 and the C66 cross-link are shown as sticks. **b–f** Details of the asymmetric and symmetric interfaces of the Endo H-treated and glycosylated homodimers, respectively. Structures are coloured as in panel **a**, with selected amino acids shown as sticks and hydrogen bonds indicated by dashed lines

ZP2[29]—the N-terminal region of ZP1 (and, by extension, ZP4) also adopts a ZP-N fold (Fig. 4). However, the cZP1-N1 domain also contains a short C′ β-strand (R60–E62) that contributes to its homodimerisation (Fig. 5b–f). Because such a strand had previously only been found in ZP-C domains as part of a C′–C″ β-hairpin[5], its presence within cZP1-N1 brings additional support to the idea that ZP-N and ZP-C originated by duplication of an ancestral Ig-like domain[37].

Most importantly, analysis of the cZP1-N1 cross-link crystals reveals the presence of different homodimer interfaces, stabilised by alternative interactions involving a set of largely conserved residues (Fig. 5; Supplementary Fig. 1). This suggests that the invariant intermolecular disulphide of ZP1—whose mutation abolishes the homodimerisation of chicken, human and mouse ZP1-N1 (Fig. 6)—allows the protein to introduce a cross-link that is not only mechanically and chemically resistant but also able to join ZP filaments with highly variable relative orientations. By requiring multiple interactions to be comparably strong, a non-covalent interface would be much less plastic and thus enforce significant constraints on the geometry of filament cross-linking. Interestingly, sequence covariation considerations support the interfaces observed in the cZP1-N1 crystals by showing that, in some species from the suborder Serpentes, both members of the H91(A)/W72(B) stacking pair (and in one case also F61(A)) are replaced by non-aromatic residues, whereas I68 is substituted with a more hydrophobic Phe; at the same time, the conserved sequon preceding the cross-linking Cys has been replaced by another within the de loop (Supplementary Fig. 1b). As a result of

this kind of changes, one type of homodimer interface may be favoured over the other in particular subsets of species.

Notwithstanding the molecular contortionism of the cZP-N1 homodimer, structural considerations suggest that core fucosylation of the essential N-glycan that immediately precedes the intermolecular disulphide (Fig. 3; Supplementary Fig. 1) may favour the symmetric interface of the cZP1-N1 cross-link over its more extensive asymmetric alternative (Supplementary Fig. 5). Because cortical granule exocytosis releases small amounts of α-fucosidase[53], an enzyme that has also been detected in the mammalian oviductal and uterine fluid[54,55], defucosylation of hZP1 could influence the biological function of the egg coat during fertilisation or pre-implantation development by altering the mechanical properties of ZP cross-links.

The observation that a Zn^{2+} ion can be bound by two His residues of cZP1-N1 (Fig. 8), one of which is invariant (cZP1 H71/hZP1 H82) and the other almost completely conserved (cZP1 H91/hZP1 H102) (Supplementary Fig. 1), also raises a possible connection with the zinc spark response of mammalian oocytes to activation. By co-ordinately releasing intracellular zinc into the extracellular space, this exocytotic event—which has also been described in human[47]—allows resumption of the cell cycle[56]. At the same time, zinc sparks have been suggested to contribute to the block to polyspermy by inducing physiochemical changes into ZP[48], but the molecular basis of this effect is unknown. The presence of a conserved Zn^{2+}-binding site in ZP1 (Fig. 8d, e), together with the finding that mutation of the two Zn^{2+}-binding His severely impairs homodimerisation of
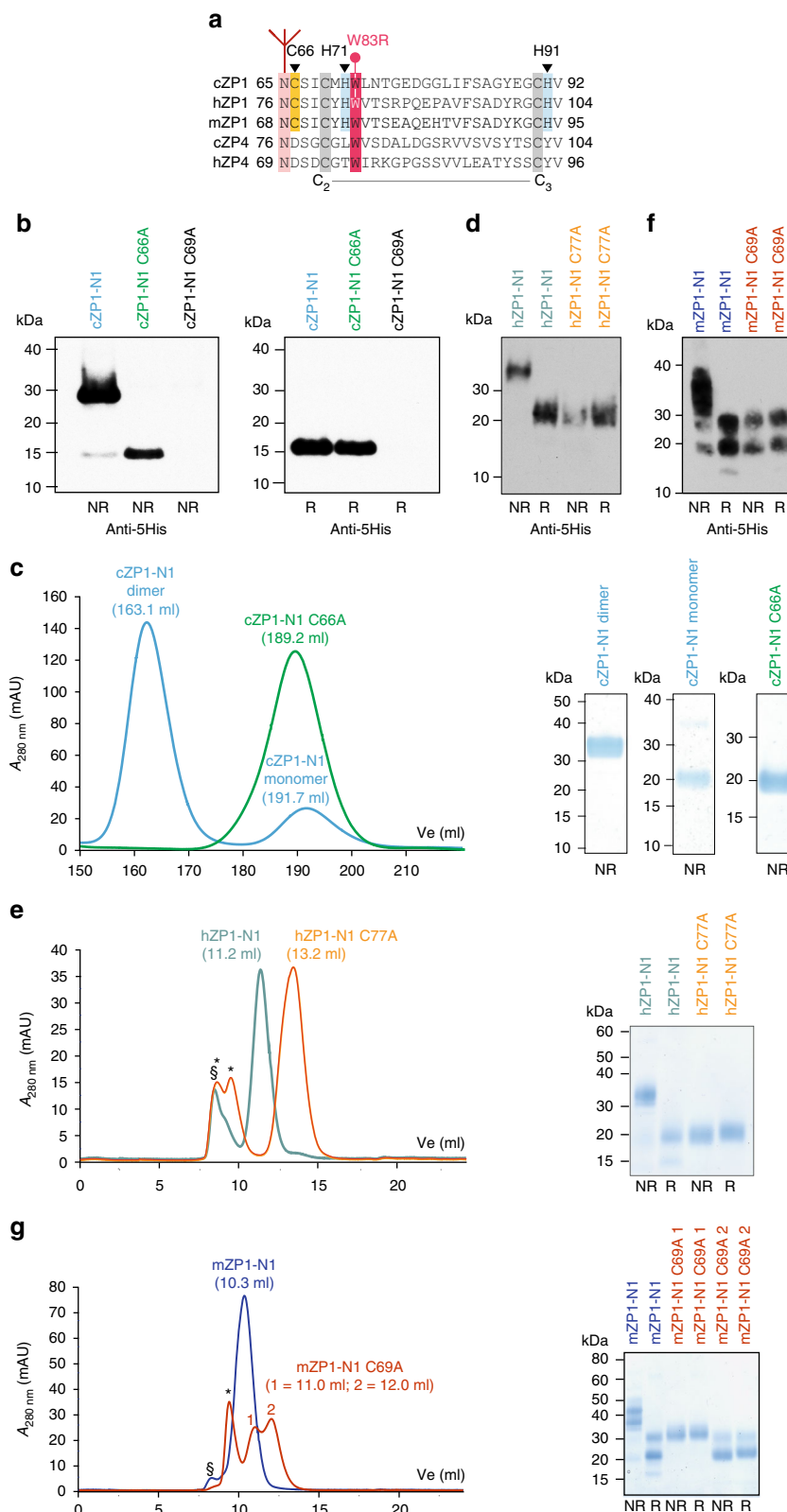
hZP1-N1 (Fig. 8f–h), suggests that zinc ions released by the zygote could modulate the architecture of the ZP by altering the conformation of its ZP1 cross-links. This could in turn hinder sperm penetration by facilitating the compaction of ZP filaments that is observed upon egg activation or exposure of unfertilised oocytes to Zn^{2+} (ref. [48]). In agreement with such a mechanism, which would not exclude the presence of additional binding sites

in other ZP subunits[48], H71 directly stabilises the cZP1-N1 intermolecular disulphide via two main-chain H-bonds in both asymmetric and symmetric homodimer interfaces (Fig. 8b–d). H91, on the other hand, stabilises the asymmetric interface by stacking against W72 from the opposite molecule (Figs 5d, 8b) whereas, in the case of the symmetric homodimer, it contributes to the I68-binding pocket together with H71 (Fig. 5c and left

Fig. 6 ZP1-N1 homodimerisation in chicken, human and mouse. **a** Sequence alignment of selected ZP1/4 homologues. The conserved N-glycosylation site and non-canonical Cys of ZP1 are highlighted in pink and yellow, respectively, whereas canonical ZP-N domains Cys 2 and 3 are shaded grey. hZP1 W83, mutated in infertile patients (Fig. 7), and conserved ZP1-N1 His residues involved in Zn^{2+} binding (Fig. 8) are shaded in red and cyan, respectively. **b** Immunoblots with anti-5His monoclonal show that, whereas mutation of $C_2$69 abolishes protein expression, mutation of non-canonical C66 does not affect secretion of cZP1-N1 but completely prevents its covalent homodimerisation. **c** Coomassie-stained SDS-PAGE analysis of SEC experiments performed with a HiLoad 26/600 Superdex 75 -pg column confirms that immobilised metal affinity chromatography (IMAC)-purified wild-type cZP1-N1 consists of two species, corresponding to dimeric and monomeric proteins. On the contrary, cZP1-N1 C66A is entirely monomeric. **d–g** Parallel analyses show that homodimerisation of the human (panels **d**, **e**) and mouse (panels **f**, **g**) ZP1-N1 domains is mediated by C77 and C69, respectively. SEC runs were performed using a Superdex 75 Increase 10/300 GL column; void volume and contaminant peaks are indicated by § and *. As supported by deglycosylation experiments with PNGase F (Supplementary Fig. 3), heterogeneous glycosylation of the two sequons of mZP1-N1 (N49 and N68)[17] results in two distinct bands on SDS-PAGE
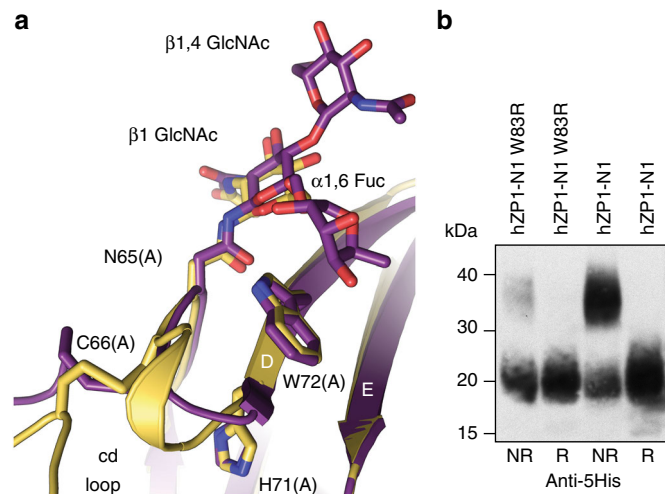

Fig. 7 ZP1 mutation W83R affects secretion and homodimerisation of hZP1-N1. **a** W72 stacks between the cd loop and the N65 glycan α1,6 fucose. **b** Anti-5His immunoblot of hZP1-N1 W83R mutant (20 µL medium) compared with wild-type (10 µL medium) 72 h post transfection

panels of Fig. 8c). Notably, like the Zn^{2+}-binding pocket of the asymmetric homodimer (Fig. 8a, inset), the I68-binding pocket also has a negatively charged potential (Fig. 8c, right panels). This suggests that exocytosed zinc may not only affect the asymmetric interface by interfering with the stacking between H91(A) and W72(B) (Fig. 8d), but also alter the symmetric interface by competing with I68. Of direct relevance to these considerations and consistent with the fact that zinc spark-induced ZP architecture changes are triggered by relatively high concentrations of the ion[48], we do not observe bound Zn^{2+} in native crystals of cZP1-N1 (Fig. 8b).

The finding that hZP4-N1 is entirely monomeric (Fig. 9a, b) explains why—as mentioned above—hZP4 does not functionally complement ZP1 in infertile patients carrying ZP1 mutations[39]. While mammalian ZP4 could be involved in increasing the thickness of the ZP (ref. 28 and Supplementary Table 1), cZP4-N1 is markedly different from its human counterpart in that it is entirely secreted as a homodimer (Fig. 9a–c). This observation supports the recent suggestion that cZP4 functionally substitutes cZP1 in the VE of the immature oocyte, which is assembled before cZP1 starts being expressed[30,34]. Consistent with the role of cZP1 as a major component of the VE of mature oocytes[30,34], the P/Q-rich repeated region of this subunit (absent in cZP4) may act as a molecular spring that stretches while the egg coat dramatically expands over the course of oocyte maturation. This would allow the cZP1-N1 cross-links to be maintained, until they are released (Supplementary Fig. 4) when penetrating sperm degrades the P/Q-rich region of cZP1[30,44]. While the very low

abundance of mammalian ZP1 makes it difficult to determine whether this is also proteolyzed upon sperm penetration, it is interesting to note that ZP1-N1 domains are not found in fish, where sperm penetrates the egg coat through the micropyle. However, fish VE subunits also contain N-terminal P/Q-rich regions that are covalently cross-linked by transglutaminase during post-fertilisation hardening[27].

In summary, our studies provide the first structural information on ZP1 and reveal the molecular basis of egg coat cross-linking from birds to mammals. Considering that the intermolecular cross-link of hZP1 is only formed upon its secretion into the extracellular space (Fig. 1b), the identification of this previously neglected subunit as critical for human fertility also highlights it as a possible target for non-hormonal contraception.

Methods

DNA constructs. cDNA fragments encoding 6His-tagged hZP1 and hZP1Mut (M1-S389 followed by the sequence TPIACSYDPARPPAA), Myc-tagged hZP2, FLAG-tagged hZP3 and HA-tagged hZP4 were synthesised (GenScript; GeneArt/Thermo Fisher Scientific) and subcloned into pHLsec3[29]. A synthetic construct encompassing cZP1-N1 (cZP1$_{L25-G149S}$) was subcloned into pJ609 (DNA2.0/ATUM) in frame with sequences coding for an IgK signal peptide and a 6His-tag inserted within the C-terminal linker fragment following conserved $C_4$124; the same DNA was combined with a PCR fragment amplified from cZP1full/pGEM[57] to generate a 6His-tagged full-length cZP1 ORF in pSI (Promega). The pHLsec3 construct expressing C-terminally 6His-tagged mZP1-N1$_{M1-A141}$ was generated by PCR, using as template Addgene plasmid 14644 (Mouse ZP1; JD#264[58]). Constructs expressing C-terminally His- or Myc-tagged versions of hZP1-N1 (hZP1$_{M1-A149}$), hZP4-N1 (hZP4$_{M1-T140}$) and cZP4-N1 (cZP4$_{M1-S148}$) were generated by PCR from the above constructs or pTargeT/cZP4[34]. cDNAs encoding conserved His mutants of cZP1-N1 and hZP1-N1 were synthesised (ATUM) and cloned into pJ609 and pHLsec3, respectively, as described above; the pHLsec3 construct expressing C-terminally 8His-tagged hZP1$_{A195-P590}$ was generated by overlap extension PCR. Oligonucleotides were ordered from Sigma-Aldrich and constructs were verified by DNA sequencing (Eurofins Genomics).

Mammalian protein expression. Proteins were transiently expressed in HEK293 transfected using 25 kDa branched polyethylenimine and cultured in serum-free media[59], with the exception of hZP1$_{A195-P590}$ whose production was carried out using media supplemented with 2% foetal bovine serum (Biological Industries). Co-expression experiments were performed using equivalent ratios of plasmid DNAs; for control single-expression experiments, expression plasmids were supplemented with empty vector DNA (pSI; Promega) to maintain the same ZP subunit/total DNA ratio.

Protein analysis. For immunoblotting, proteins separated by SDS-PAGE were transferred to a nitrocellulose membrane (GE Healthcare Life Sciences) and probed with primary antibodies anti-5His monoclonal (1:1000; Penta•His BSA-free, QIA-GEN catalogue number 34660), anti-Myc monoclonal (1:1000; clone 9E10, Sigma-Aldrich catalogue number M4439), anti-FLAG monoclonal (1:1000; clone M2, Sigma-Aldrich catalogue number F1804), anti-HA monoclonal (1:1000; clone HA-7, Sigma-Aldrich catalogue number H3663) or anti-cZP1 N-terminal fragment polyclonal[57] (1:3000). Secondary antibodies were horseradish peroxidase-conjugated goat anti-mouse IgG Fc (1:10,000; Life Technologies/Thermo Fisher Scientific catalogue number A16084) or horseradish peroxidase-conjugated horse anti-mouse IgG (1:10,000; Cell Signalling Technology catalogue number 7076). Chemiluminescence detection was performed with Western Lightning ECL Plus (Perkin Elmer).

For cell lysate analysis, cells transfected for 72 h were washed twice with 1 mL of cold phosphate-buffered saline (PBS). In total, 200 µL of lysis buffer containing
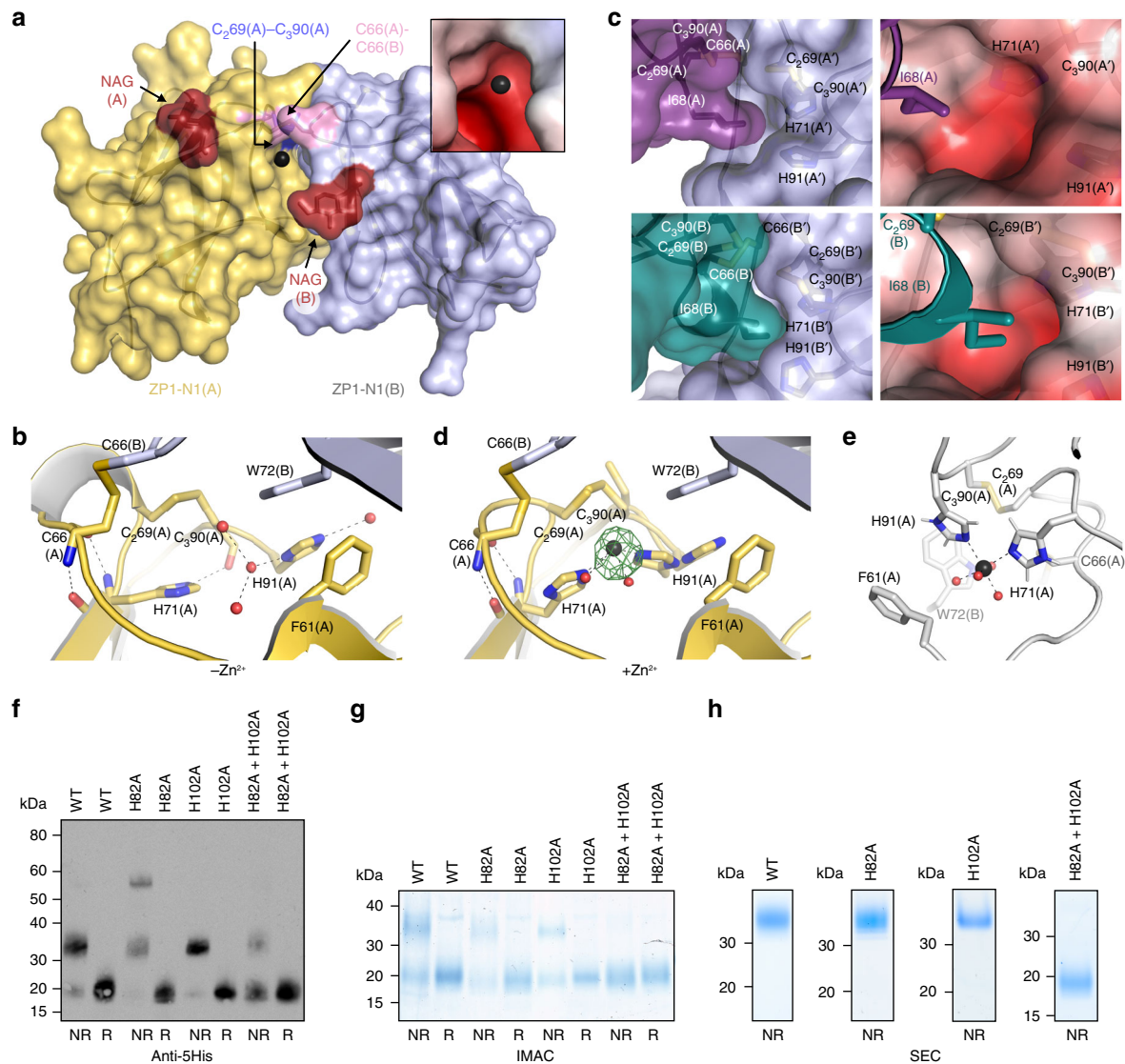
Fig. 8 Role of a conserved ZP1-N1 His pair in Zn^{2+} binding and homodimerisation. **a** Surface representation of the structure of the deglycosylated cZP1-N1 homodimer in the crystal soaked with $Zn(OAc)_2$, highlighting the position of the Zn^{2+} ion bound to H71(A) and H91(A) (black sphere) relative to the protein subunits, two proximal disulphide bonds and N-acetylglucosamine (NAG) residues attached to N65. The A subunit of the homodimer is coloured gold according to Fig. 5, whereas subunit B is shown in lilac. The inset shows a detailed view of the region around the Zn^{2+}, coloured by electrostatic potential using a gradient from red ($-4\,kT\,e^{-1}$) to blue ($+4\,kT\,e^{-1}$) through white ($0\,kT\,e^{-1}$). **b** Details of the asymmetric interface of the cZP1-N1 homodimer, as observed in the native crystal of the deglycosylated protein. Selected amino acids are shown as sticks, with hydrogen bonds indicated by dashed lines. **c** Surface representation of the symmetric interface of glycosylated cZP1-N1 (chain A, top left panel; chain B, bottom left panel), showing the position of H71 and H91 relative to I68. Notably, the H71/H91 pocket of the A' subunit is more accessible than that of B', due to apparent flexibility of the relatively small symmetric interface. The right panels, which are rotated by 30 degrees over the y-axis compared with the left ones, show zoomed views of the symmetric interfaces around I68(A) and I68(B), with the surface of the A' and B' molecules coloured by electrostatic potential as in the inset of panel **a**. **d** Details of the asymmetric interface of deglycosylated cZP1-N1, as observed in the Zn^{2+}-soaked crystal. The green mesh is an anomalous difference map calculated at $\lambda = 1.2825$ Å and contoured at $6\,\sigma$. **e** 100-ns MD simulation snapshot of the Zn^{2+}-binding site shown in panel **a**. H71 and H91, together with four water molecules, coordinate the zinc ion during the whole simulation time (mean Zn-His N distance 2.22 ± 0.06 Å; mean Zn–H_2O O distance 2.10 ± 0.05 Å). **f** Anti-5His immunoblot analysis of wild-type (WT) and His mutant hZP1-N1 constructs secreted by HEK293T cells (100 µL medium). **g**, **h** Coomassie-stained SDS-PAGE analysis of IMAC (panel **g**) and SEC (panel **h**) fractions from the purification of the same constructs shown in panel **f**. The SEC profiles corresponding to panel **h** are reported in Supplementary Fig. 6

50 mM Na-HEPES pH 7.8, 150 mM NaCl, 0.1% (v v^{-1}) SDS, 0.5% (v v^{-1}) sodium deoxycholate, 1% (v v^{-1}) Triton X-100 and protease inhibitors (cOmplete mini EDTA-free protease inhibitor cocktail; Roche) were added to one well and the plate incubated overnight at 4 °C. Lysate samples were recovered by scraping cells in the presence of SDS-PAGE loading buffer. In all, 40 µL of cell lysate was boiled and separated by SDS-PAGE for immunoblot analysis.

Deglycosylation experiments in reducing conditions were performed with PNGase F (New England Biolabs), according to the manufacturer's instructions; for analysing deglycosylation products in non-reducing conditions, DTT-containing Glycoprotein Denaturing Buffer was replaced with 10 mM Na-HEPES pH 7.5, 1% (v v^{-1}) SDS.

Preparation of soluble ZP glycoproteins from proteolytically degraded chicken egg coat. Egg coats were isolated from the F1 oocytes of laying white leghorn hens; sperm and sperm protease were prepared from ejaculated semen using gentle centrifugation and nitrogen cavitation, respectively[44]. In total, 100 mg of isolated egg coat (wet weight) was suspended in 1 mL of PBS by ultrasonication[31] and then mixed with the sperm protease preparation. The reaction mixture was kept at 39 °C for 20 h and centrifuged at 16,000 x g for 5 min. The supernatant was 0.45-µm filtered and injected into a HiPrep 16/60 Sephacryl S-300 HR column (GE Healthcare). SEC elution profiles were monitored by measuring absorbance at $\lambda = 280$ nm, and both the ~35 kDa and ~21 kDa N-terminal fragments of cZP1
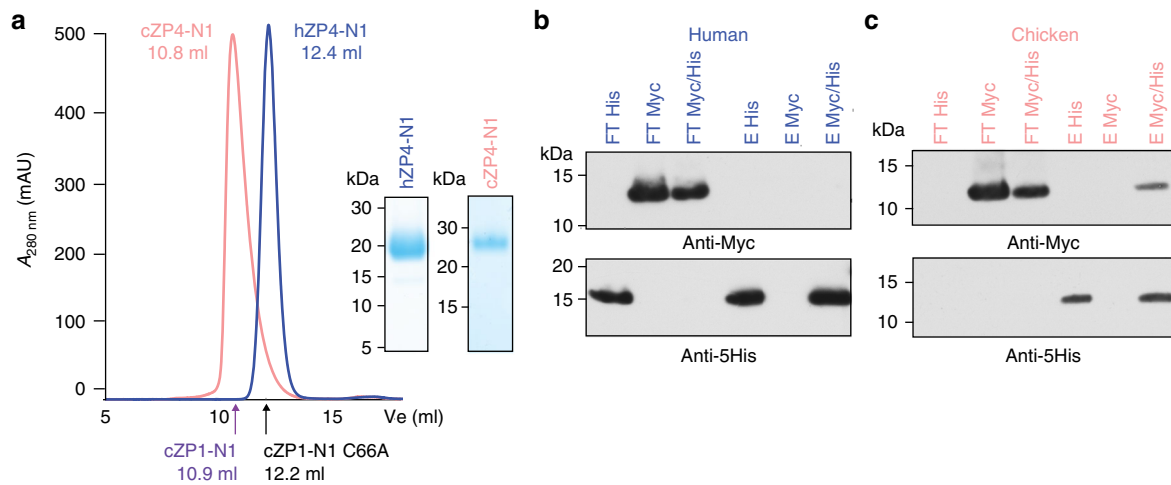
Fig. 9 ZP4-N1 forms a non-covalent homodimer in chicken, but not in human. **a** SEC analysis of purified hZP4-N1 and cZP4-N1 and non-reducing Coomassie-stained SDS-PAGE analysis of the corresponding peak fractions. Peaks are normalised to 500 mAU, and elution volumes of the cZP1-N1 homodimer and the cZP1-N1 C66A monomer are indicated. **b, c** Immunoblot analysis in reducing conditions of His-pull-down experiments of co-expressed hZP4-N1-His/Myc or cZP4-N1-His/Myc. Samples were PNGase F-treated before SDS-PAGE. FT flow-through, E elution

were analysed by mass spectrometry following in-gel chymotryptic peptide digestion with chymotrypsin (Promega)[60].

Ethics. Animal experiments were performed in accordance to the local ethics committee's approval (Committee for Animal Experiments, Graduate School of Bioagricultural Sciences, Nagoya University, approval number 2016030218).

Mass spectrometry. For LC-MALDI-TOF MS/MS analysis, N-terminal fragments of native cZP1 were boiled in SDS in the presence of 0.75% (v v^{-1}) β-mercaptoethanol and treated with PNGase F. Both deglycosylated and glycosylated samples were reduced, alkylated and digested with chymotrypsin in the presence of 0.01% (w v^{-1}) ProteaseMAX™ Surfactant (Promega). A part of these peptide preparations was subjected to LC-MALDI-TOF MS/MS analysis. Chymotryptic glycopeptides from the sample without PNGase F treatment were captured using RCA-1 lectin, and the resulting complexes concentrated with 30 kDa cut-off centrifugal filtration devices. Bound glycopeptides were released using 0.1% (v v^{-1}) trifluoroacetic acid and analysed by MALDI-TOF/MS.

Samples were injected into a DiNa Nano LC system equipped with a DiNa MaP autospotter (KYA technologies) and eluted using a linear acetonitrile gradient. MS spectra were obtained using a MALDI-TOF/TOF 5800 Proteomic Analyser mass spectrometer (Applied Biosystems) and analysed using Mascot[61].

Protein purification. Seventy-two hours after transfection, the conditioned media from mammalian cells was harvested, 0.22-μm filtered (Sarstedt) and adjusted to 20 mM Na-HEPES pH 7.8, 500 mM NaCl, 5–10 mM imidazole (IMAC buffer). In all, 10 mL pre-equilibrated nickel agarose slurry (Ni-NTA; QIAGEN) was added per litre of medium and incubated for either 1 h at room temperature (RT) or overnight at 4 °C. After washing the beads with 100 column volumes IMAC buffer, proteins were batch-eluted with five column volumes elution buffer (20 mM Na-HEPES pH 7.8, 150 mM NaCl, 500 mM imidazole). HEK293S-expressed cZP1-N1 was deglycosylated with Endo H (1:10 mass ratio) for 1 h at 37 °C in 120 mM Na/K phosphate pH 6.0 prior to elution.

Proteins were concentrated with appropriate centrifugal filtration devices (Amicon) and further purified by SEC at 4 °C using an ÄKTA$_{FPLC}$ chromatography system (GE Healthcare). For crystallographic studies, cZP1-N1 was injected into a HiLoad 26/600 Superdex 200 -pg column (GE Healthcare) pre-equilibrated with 20 mM Na-HEPES pH 8.0, 500 mM NaCl. Protein-containing fractions were pooled and concentrated to 5 mg mL^{-1} (Endo H-deglycosylated cZP1-N1) or 3.5 mg mL^{-1} (glycosylated cZP1-N1), respectively; prior to crystallisation setup, samples were dialysed against 20 mM Na-HEPES pH 8.0, 200 mM NaCl. For analytical SEC and SEC-MALS experiments, 20 mM Na-HEPES pH 7.8, 150 mM NaCl was used as running buffer.

Sequence analysis. To create a non-redundant database of ZP1-N1 sequences, a redundant collection of 510 entries was first created by combining the results of parallel BLAST[62] searches performed using as queries a selection of ZP1 homologs from HomoloGene[63] entry 33483. After pruning of sequences identical to or contained within others, 128 entries with 5 Cys were processed with CD-HIT[64], using a 0.8 sequence identity threshold. Following manual restore of the mZP1-N1 sequence (belonging to the same CD-HIT cluster as hZP1-N1) and deletion of two

sequences whose full-length counterparts did not contain a trefoil domain[65,66], this resulted in a non-redundant database of 28 sequences from amniotes which were aligned with MAFFT[67] as implemented in SnapGene (GSL Biotech).

SEC-MALS analysis. Measurements were performed using an Ettan LC high-performance liquid chromatography system with a UV-900 detector (Amersham Pharmacia Biotech; λ = 280 nM), coupled with a miniDawn Treos MALS detector (Wyatt Technology; λ = 658 nm) and an Optilab T-rEX dRI detector (Wyatt Technology; λ = 660 nm). Separation was performed using a Superdex 75 Increase 10/300 GL column (GE Healthcare), using a flow rate of 0.4 mL min^{-1} and a mobile phase consisting of 10 mM Na-HEPES pH 7.8, 150 mM NaCl. The data processing and weight-averaged molecular mass calculation were performed using the ASTRA 7.1.3 software (Wyatt Technology).

Protein crystallisation. Hanging drop vapour diffusion experiments were set up by hand at 20 °C, by mixing protein and reservoir solutions at a 1:1 ratio and equilibrating against 1 mL of reservoir solution. Endo H-treated cZP1-N1 N121Q from HEK293S cells and untreated cZP1-N1 from HEK293T cells crystallised in 0.6 M LiCl, 0.1 M tri-sodium citrate pH 5.5 and 29% (v v^{-1}) MPD, 0.1 M MES pH 6.5, respectively. For heavy-atom derivatisation, crystals of the Endo H-treated protein were soaked in reservoir solutions supplemented with 5 mM KAu(CN)$_2$ or 1 mM Zn(OAc)$_2$ for 20 h before cryocooling.

X-ray data collection. All data sets were collected at 100 K from single crystals cryoprotected by supplementing the mother liquor with MPD to a final concentration of 35% (v v^{-1}) and subsequently flash-frozen in liquid nitrogen.

Data sets for Endo H-treated cZP1-N1 were collected at beamline BL14.1 of BESSY II (Helmholtz-Zentrum, Berlin) using a MarMosaic 225 CCD detector (Mar) (native data, λ = 1.7710 Å) or beamline I03 of Diamond Light Source (DLS, England) using a a PILATUS 6M-F detector (DECTRIS) (gold-derivative data, λ = 1.03970 Å; zinc-derivative data, λ = 1.2825 Å). The native data set for the glycosylated protein was collected at European Synchrotron Radiation Facility (ESRF, Grenoble) beamline ID23-1[68] (λ = 0.9791 Å), using a PILATUS 6 M detector (DECTRIS).

The data collection statistics generated using phenix.merging_statistics[69] are reported in Table 1.

Data processing and structure determination. X-ray diffraction data sets were processed using XDS[70].

The structure of the gold derivative of Endo H-treated cZP1-N1 was solved at 2.75 Å resolution by single-wavelength anomalous dispersion (SAD) phasing with PHENIX AutoSol[71], which located six sites (FOM 0.35, BAYES-CC 53.5). The resulting initial model, which included both moieties of a ZP1 cross-link for a total of 155 residues (R$_{work}$ 42%, R$_{free}$ 44%, model-map correlation 62%), was manually rebuilt in Coot[72] and refined against the 2.3 -Å resolution native data set using phenix.refine;[73] protein geometry was validated with MolProbity[74] and carbohydrate structure validation was carried out using Privateer[75]. The structure of the same protein was also independently solved by Zn-SAD phasing at 2.7 Å resolution (FOM 0.31, BAYES-CC 45.1), yielding an initial model of 163 residues (R$_{work}$ 35%, R$_{free}$ 0.41%, model-map correlation 73%) that was manually completed

and led to a refined set of coordinates that includes nine Zn^{2+} ions. The structure of glycosylated cZP1-N1 was determined by molecular replacement with Phaser[76], using as search model the structure of a monomer of Endo-H-treated cZP1-N1.

The MolProbity Ramachandran plots of the structures described in this study show the following percentages of favoured, allowed and outlier residues: 98.6, 1.4, 0.0 (Endo H-treated cZP1-N1); 99.5, 0.5, 0.0 (Endo H-treated cZP1-N1 with Zn^{2+}); 100.0, 0.0, 0.0 (glycosylated cZP1-N1). Refinement and validation statistics generated using phenix.table_one[69] are reported in Table 1.

Structure analysis. Structural alignments were performed using UCSF Chimera;[77] protein–protein interfaces were analysed manually as well as using FoldX[78], PDBsum[79], PIC[80] and PISA[81]. Electrostatic surface potential calculations were performed with PDB2PQR[82] and APBS[83], via the APBS Tools plugin of PyMOL (Schrödinger, LLC). Figures were created with PyMOL.

Molecular dynamics simulations. Simulations were performed using GROMACS version 2016[84] and the CHARMM36 force field[85,86], based on the crystallographic coordinates of the Zn^{2+}-bound cZP1-N1 homodimer. Simulation conditions were set up to be appropriate for pH 7.0, solvating the proteins using a 8-nm dodecahedric box filled with TIP3P water molecules and 0.1 M Na^+/Cl^- ion concentrations. The particle mesh Ewald method was employed to treat Coulomb interactions (switching distance 1.2 nm/grid 0.12 nm); Lennard–Jones interactions were switched off between 0.8 and 1.2 nm; pressure and temperature were maintained constant at 1 bar and 298 K, using the Parrinello–Rahman barostat (tau = 5.0 ps) and velocity rescaling thermostat (tau = 0.5 ps), respectively. Bond distances and angles of water were constrained using the SETTLE algorithm, whereas other bond distances were constrained with the LINCS algorithm. A leap-frog integrator (time step = 2 fs) was used. Three independent 100 ns replicas were performed; before data generation, 20-ns position-restrained simulations were performed to relax the positions of Zn^{2+}-proximal water molecules.

Pull-down analysis of protein–protein interaction. For pull-down experiments of 6His-tagged hZP4-N1 and cZP4-N1 co-transfected with Myc-tagged counterparts, 2 mL conditioned media were harvested, equilibrated with 20 mM Na-HEPES pH 7.8, 100 mM NaCl, 10 mM imidazole and incubated with 50 μL Ni-NTA beads for 1 h at RT. Beads were collected by centrifugation at 100 x *g* and washed four times with 500 μL binding buffer followed by centrifugation at 100 x *g* for 5 min; bound proteins were eluted with 100 μL 20 mM Na-HEPES pH 7.8, 150 mM NaCl, 500 mM imidazole. In total, 100 μL flow-through and 5 μL elution fractions were treated with PNGase F using standard procedures and analysed by immunoblotting with anti-5His and anti-Myc.

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

Data availability

The source data underlying Figs 1b–d, 2b–d, 2f–j, 3a, 6b–g, 7b, 8f–h, 9 and Supplementary Fig. 3 are provided as a Source Data file. Atomic coordinates and structure factors have been deposited with the Protein Data Bank (PDB) under accession codes 6GF6 (Endo H-treated cZP1-N1 homodimer, high-resolution native), 6GF7 (Endo H-treated cZP1-N1 homodimer, Zn^{2+} derivative) and 6GF8 (glycosylated cZP1-N1 homodimer). Mass spectrometric data has been deposited with the Japan ProteOme STandard Repository (jPOST) under ProteomeXchange Consortium ID PXD013994.

References

1. Wassarman, P. M. & Litscher, E. S. Chapter Ten - The mouse egg's zona pellucida. In *Current Topics in Developmental Biology* (eds. Litscher, E. S. & Wassarman, P. M.) Vol. 130, 331–356 (Academic Press, Cambridge, MA, 2018).
2. Litscher, E. S. & Wassarman, P. M. Evolution, structure, and synthesis of vertebrate egg-coat proteins. *Trends Dev. Biol.* **8**, 65–76 (2014).
3. Bork, P. & Sander, C. A large domain common to sperm receptors (Zp2 and Zp3) and TGF-β type III receptor. *FEBS Lett.* **300**, 237–240 (1992).
4. Jovine, L., Qi, H., Williams, Z., Litscher, E. & Wassarman, P. M. The ZP domain is a conserved module for polymerization of extracellular proteins. *Nat. Cell Biol.* **4**, 457–461 (2002).
5. Bokhove, M. & Jovine, L. Structure of zona pellucida module proteins. In *Current Topics in Developmental Biology* (eds. Litscher, E. S. & Wassarman, P. M.) Vol. 130, 413–442 (Academic Press, Cambridge, MA, 2018).
6. Bleil, J. D. & Wassarman, P. M. Structure and function of the zona pellucida: identification and characterization of the proteins of the mouse oocyte's zona pellucida. *Dev. Biol.* **76**, 185–202 (1980).
7. Sacco, A. G., Yurewicz, E. C., Subraminian, M. G. & DeMayo, F. J. Zona pellucida composition: species cross reactivity and contraceptive potential of antiserum to a purified pig zona antigen (PPZA). *Biol. Reprod.* **25**, 997–1008 (1981).
8. Conner, S. J., Lefièvre, L., Hughes, D. C. & Barratt, C. L. R. Cracking the egg: increased complexity in the zona pellucida. *Hum. Reprod.* **20**, 1148–1152 (2005).
9. Bleil, J. D. & Wassarman, P. M. Mammalian sperm-egg interaction: identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm. *Cell* **20**, 873–882 (1980).
10. Bleil, J. D., Greve, J. M. & Wassarman, P. M. Identification of a secondary sperm receptor in the mouse egg zona pellucida: role in maintenance of binding of acrosome-reacted sperm to eggs. *Dev. Biol.* **128**, 376–385 (1988).
11. Gahlay, G., Gauthier, L., Baibakov, B., Epifano, O. & Dean, J. Gamete recognition in mice depends on the cleavage status of an egg's zona pellucida protein. *Science* **329**, 216–219 (2010).
12. Greve, J. M. & Wassarman, P. M. Mouse egg extracellular coat is a matrix of interconnected filaments possessing a structural repeat. *J. Mol. Biol.* **181**, 253–264 (1985).
13. Rankin, T. L. et al. Defective zonae pellucidae in *Zp2*-null mice disrupt folliculogenesis, fertility and development. *Development* **128**, 1119–1126 (2001).
14. Rankin, T. et al. Mice homozygous for an insertional mutation in the *Zp3* gene lack a zona pellucida and are infertile. *Development* **122**, 2903–2910 (1996).
15. Liu, C. et al. Targeted disruption of the *mZP3* gene results in production of eggs lacking a zona pellucida and infertility in female mice. *Proc. Natl Acad. Sci. USA* **93**, 5431–5436 (1996).
16. Rankin, T., Talbot, P., Lee, E. & Dean, J. Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. *Development* **126**, 3847–3855 (1999).
17. Boja, E. S., Hoodbhoy, T., Fales, H. M. & Dean, J. Structural characterization of native mouse zona pellucida proteins using mass spectrometry. *J. Biol. Chem.* **278**, 34189–34202 (2003).
18. Moos, J., Faundes, D., Kopf, G. S. & Schultz, R. M. Composition of the human zona pellucida and modifications following fertilization. *Hum. Reprod.* **10**, 2467–2471 (1995).
19. Kiefer, S. M. & Saling, P. Proteolytic processing of human zona pellucida proteins. *Biol. Reprod.* **66**, 407–414 (2002).
20. Hughes, D. C. & Barratt, C. L. Identification of the true human orthologue of the mouse *Zp1* gene: evidence for greater complexity in the mammalian zona pellucida. *Biochim. Biophys. Acta* **1447**, 303–306 (1999).
21. Lefièvre, L. Four zona pellucida glycoproteins are expressed in the human. *Hum. Reprod.* **19**, 1580–1586 (2004).
22. Yauger, B., Boggs, N. A. & Dean, J. Human ZP4 is not sufficient for taxon-specific sperm recognition of the zona pellucida in transgenic mice. *Reproduction* **141**, 313–319 (2011).
23. Bork, P. A trefoil domain in the major rabbit zona pellucida protein. *Protein Sci.* **2**, 669–670 (1993).
24. Harris, J. D. et al. Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: the ZPA, ZPB and ZPC gene families. *DNA Seq.* **4**, 361–393 (1994).
25. Callebaut, I., Mornon, J.-P. & Monget, P. Isolated ZP-N domains constitute the N-terminal extensions of Zona Pellucida proteins. *Bioinformatics* **23**, 1871–1874 (2007).
26. Monné, M., Han, L., Schwend, T., Burendahl, S. & Jovine, L. Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. *Nature* **456**, 653–657 (2008).
27. Litscher, E. S. & Wassarman, P. M. The fish egg's zona pellucida. In *Current Topics in Developmental Biology* (eds. Litscher, E. S. & Wassarman, P. M.) Vol. 130, 275–305 (Academic Press, Cambridge, MA, 2018).
28. Avella, M. A., Baibakov, B. & Dean, J. A single domain of the ZP2 zona pellucida protein mediates gamete recognition in mice and humans. *J. Cell Biol.* **205**, 801–809 (2014).
29. Raj, I. et al. Structural basis of egg coat-sperm recognition at fertilization. *Cell* **169**, 1315–1326.e17 (2017).
30. Nishio, S., Okumura, H. & Matsuda, T. Egg-coat and zona pellucida proteins of chicken as a typical species of aves. In *Current Topics in Developmental Biology* (eds. Litscher, E. S. & Wassarman, P. M.) Vol. 130, 307–329 (Academic Press, Cambridge, MA, 2018).
31. Okumura, H. et al. A newly identified zona pellucida glycoprotein, ZPD, and dimeric ZP1 of chicken egg envelope are involved in sperm activation on sperm-egg interaction. *Biochem. J.* **384**, 191–199 (2004).
32. Bausek, N., Waclawek, M., Schneider, W. J. & Wohlrab, F. The major chicken egg envelope protein ZP1 is different from ZPB and is synthesized in the liver. *J. Biol. Chem.* **275**, 28866–28872 (2000).
33. Okumura, H., Okajima, T., Nadano, D. & Matsuda, T. Association of chicken zona pellucida glycoprotein (ZP) B1 with ZPC induces formation of

ZPB1-ZPC fibrous aggregates containing disulphide-bridged ZPB1 dimer. *Biochem. Biophys. Res. Commun.* **364**, 682–688 (2007).

34. Nishio, S. et al. Glycosylated chicken ZP2 accumulates in the egg coat of immature oocytes and remains localized to the germinal disc region of mature eggs. *Biol. Reprod.* **91**, 107 (2014).

35. Ganguly, A. et al. In humans, zona pellucida glycoprotein-1 binds to spermatozoa and induces acrosomal exocytosis. *Hum. Reprod.* **25**, 1643–1656 (2010).

36. Baibakov, B., Boggs, N. A., Yauger, B., Baibakov, G. & Dean, J. Human sperm bind to the N-terminal domain of ZP2 in humanized zonae pellucidae in transgenic mice. *J. Cell Biol.* **197**, 897–905 (2012).

37. Han, L. et al. Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. *Cell* **143**, 404–415 (2010).

38. Bokhove, M. et al. A structured interdomain linker directs self-polymerization of human uromodulin. *Proc. Natl Acad. Sci. USA* **113**, 1552–1557 (2016).

39. Huang, H.-L. et al. Mutant ZP1 in familial infertility. *N. Engl. J. Med.* **370**, 1220–1226 (2014).

40. Yang, P. et al. Novel zona pellucida gene variants identified in patients with oocyte anomalies. *Fertil. Steril.* **107**, 1364–1369 (2017).

41. Yuan, P. et al. Novel mutation in the *ZP1* gene and clinical implications. *J. Assist. Reprod. Genet.* **36**, 741-747 (2019).

42. Zhang, Z., Shangguan, T., Li, Y.-Y. & He, W. Infertility due to lack of zona pellucida caused by a compound heterozygous mutation in *ZP1* gene. *Reprod. Dev. Med.* **2**, 183–186 (2018).

43. Sun, L. et al. Compound heterozygous *ZP1* mutations cause empty follicle syndrome in infertile sisters. *bioRxiv* 363911, https://doi.org/10.1101/363911 (2018).

44. Takeuchi, Y. et al. Morphological and biochemical changes of isolated chicken egg-envelope during sperm penetration: degradation of the 97-kilodalton glycoprotein is involved in sperm-driven hole formation on the egg-envelope. *Biol. Reprod.* **64**, 822–830 (2001).

45. Easton, R. L. et al. Structural analysis of murine zona pellucida glycans. Evidence for the expression of core 2-type O-glycans and the Sdᵃ antigen. *J. Biol. Chem.* **275**, 7731–7742 (2000).

46. Pang, P.-C. et al. Human sperm binding is mediated by the Sialyl-Lewis^X oligosaccharide on the zona pellucida. *Science* **333**, 1761–1764 (2011).

47. Duncan, F. E. et al. The zinc spark is an inorganic signature of human egg activation. *Sci. Rep.* **6**, 24737 (2016).

48. Que, E. L. et al. Zinc sparks induce physiochemical changes in the egg zona pellucida that prevent polyspermy. *Integr. Biol.* **9**, 135–144 (2017).

49. Gupta, S. K. The human egg's zona pellucida. In *Current Topics in Developmental Biology* (eds. Litscher, E. S. & Wassarman, P. M.) Vol. 130, 379–411 (Academic Press, Cambridge, MA, 2018).

50. Chiu, P. C. N. et al. Native human zona pellucida glycoproteins: purification and binding properties. *Hum. Reprod.* **23**, 1385–1393 (2008).

51. Hoodbhoy, T. et al. ZP2 and ZP3 traffic independently within oocytes prior to assembly into the extracellular zona pellucida. *Mol. Cell. Biol.* **26**, 7991–7998 (2006).

52. Wassarman, P. M., Liu, C., Chen, J., Qi, H. & Litscher, E. S. Ovarian development in mice bearing homozygous or heterozygous null mutations in zona pellucida glycoprotein gene mZP3. *Histol. Histopathol.* **13**, 293–300 (1998).

53. Miller, D. J., Gong, X., Decker, G. & Shur, B. D. Egg cortical granule N-acetylglucosaminidase is required for the mouse zona block to polyspermy. *J. Cell Biol.* **123**, 1431–1440 (1993).

54. Tulsiani, D. R., Chayko, C. A., Orgebin-Crist, M. C. & Araki, Y. Temporal surge of glycosyltransferase activities in the genital tract of the hamster during the estrous cycle. *Biol. Reprod.* **54**, 1032–1037 (1996).

55. Coy, P., García-Vázquez, F. A., Visconti, P. E. & Avilés, M. Roles of the oviduct in mammalian fertilization. *Reproduction* **144**, 649–660 (2012).

56. Kim, A. M. et al. Zinc sparks are triggered by fertilization and facilitate cell cycle resumption in mammalian eggs. *ACS Chem. Biol.* **6**, 716–723 (2011).

57. Okumura, H. et al. Identification of distinctive interdomain interactions among ZP-N, ZP-C and other domains of zona pellucida glycoproteins underlying association of chicken egg-coat matrix. *FEBS Open Bio* **5**, 454–465 (2015).

58. Epifano, O., Liang, L. F., Familari, M., Moos, M. C. Jr & Dean, J. Coordinate expression of the three zona pellucida genes during mouse oogenesis. *Development* **121**, 1947–1956 (1995).

59. Bokhove, M. et al. Easy mammalian expression and crystallography of maltose-binding protein-fused human proteins. *J. Struct. Biol.* **194**, 1–7 (2016).

60. Okumura, H., Aoki, N., Sato, C. & Nadano, D. Heterocomplex formation and cell-surface accumulation of hen's serum zona pellucida B1 (ZPB1) with ZPC expressed by a mammalian cell line (COS-7): a possible initiating step of egg-envelope matrix construction. *Biol. Reprod.* **76**, 9–18 (2007).

61. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).

62. Johnson, M. et al. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).

63. Sayers, E. W. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **39**, D38–D51 (2011).

64. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

65. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2018).

66. Marchler-Bauer, A. & Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32**, W327–W331 (2004).

67. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

68. Nurizzo, D. et al. The ID23-1 structural biology beamline at the ESRF. *J. Synchrotron Radiat.* **13**, 227–238 (2006).

69. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).

70. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).

71. Terwilliger, T. C. et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D. Biol. Crystallogr.* **65**, 582–601 (2009).

72. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).

73. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 352–367 (2012).

74. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).

75. Agirre, J. et al. Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834 (2015).

76. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

77. Pettersen, E. F. et al. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

78. Schymkowitz, J. et al. The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).

79. Laskowski, R. A. et al. PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* **22**, 488–490 (1997).

80. Tina, K. G., Bhadra, R. & Srinivasan, N. PIC: protein interactions calculator. *Nucleic Acids Res.* **35**, W473–W476 (2007).

81. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

82. Dolinsky, T. J. et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).

83. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).

84. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015).

85. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71 (2016).

86. Guvench, O. et al. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. *J. Chem. Theory Comput.* **7**, 3162–3180 (2011).

Acknowledgements

Author contributions

K.N. expressed, purified and crystallised proteins; determined and refined structures together with L.J. and performed part of the biochemical assays. E.D. carried out all other

biochemical and biophysical studies. S.N. and T.M. analysed native cZP1 biochemically and by mass spectrometry. L.H. and S.N. helped with mammalian expression. A.V. performed and analysed MD simulations. L.J. directed the study and wrote the paper with E.D. and K.N.

Additional information

Competing interests: The authors declare no competing interests.

Peer review information: *Nature Communications* thanks Christopher Barratt, Christine Gourier and other anonymous reviewer(s) for their contribution to the peer review of this work.

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