LETTERS

Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats

Magnus Monné[1]*, Ling Han[1]*, Thomas Schwend[1], Sofia Burendahl[1] & Luca Jovine[1]

Species-specific recognition between the egg extracellular matrix (zona pellucida) and sperm is the first, crucial step of mammalian fertilization[1]. Zona pellucida filament components ZP3 and ZP2 act as sperm receptors, and mice lacking either of the corresponding genes produce oocytes without a zona pellucida and are completely infertile[2]. Like their counterparts in the vitelline envelope of non-mammalian eggs and many other secreted eukaryotic proteins, zona pellucida subunits polymerize using a 'zona pellucida (ZP) domain' module[3–5], whose conserved amino-terminal part (ZP-N) was suggested to constitute a domain of its own[6]. No atomic structure has been reported for ZP domain proteins, and there is no structural information on any conserved vertebrate protein that is essential for fertilization and directly involved in egg–sperm binding. Here we describe the 2.3 ångström (Å) resolution structure of the ZP-N fragment of mouse primary sperm receptor ZP3. The ZP-N fold defines a new immunoglobulin superfamily subtype with a β-sheet extension characterized by an E′ strand and an invariant tyrosine residue implicated in polymerization. The structure strongly supports the presence of ZP-N repeats within the N-terminal region of ZP2 and other vertebrate zona pellucida/vitelline envelope proteins, with implications for overall egg coat architecture, the post-fertilization block to polyspermy and speciation. Moreover, it provides an important framework for understanding human diseases caused by mutations in ZP domain proteins and developing new methods of non-hormonal contraception.

A three-dimensional reconstruction of ZP domain protein endoglin has been recently reported at 25 Å resolution[7]. However, crystallographic studies of this family of proteins have long been hampered by their intrinsic polymerization activity and extensive post-translational modifications, including several intramolecular and intermolecular disulphides, highly heterogeneous N/O-glycosylation, and proteolytic cleavage of precursor forms[5,8]. Expression of isolated ZP-Ns has not resulted in soluble protein, but fusion to N-terminal *Escherichia coli* maltose binding protein (MBP, also known as MalE) allowed partially soluble expression of the ~11 kDa ZP-N fragment of mouse ZP3 (residues Val 42–Arg 143) in bacteria, engineered to form disulphide bonds in the cytoplasm[6]. The carboxy-terminally polyhistidine-tagged protein—which is 71% identical to its human counterpart but does not have glycosylation sites—contained the invariant Cys_1–Cys_4 and Cys_2–Cys_3 disulphide bonds (in which the subscripts number the conserved Cys residues) that are the hallmark of the ZP domain[3,5,6,8,9]. However, it could not be crystallized because it formed filamentous aggregates *in vitro*[6].

ZP domain proteins have a highly mosaic architecture, and very few residues often separate ZP-N sequences from the variable domains that immediately precede them[3,5]. We reasoned that minimization of the linker between MBP and ZP-N, a strategy used to favour crystallization of challenging proteins[10], might shield surfaces causing protein aggregation without affecting the folding of ZP-N. By selectively eluting monomeric molecules from the affinity column used during purification, we obtained milligram amounts of a tri-Ala linker construct (MBP-AAA-ZP-N-6His) that was considerably more soluble and less aggregated than the original fusion (Supplementary Fig. 1a–f). This protein crystallized readily (Supplementary Fig. 1g–i) and its structure was solved in three crystal forms by molecular replacement with MBP (form I), molecular replacement with MBP-AAA-ZP-N (form II) and combined molecular replacement/Zn-single-wavelength anomalous diffraction (SAD) phasing (form III) (Supplementary Table 1 and Supplementary Figs 2 and 3).

ZP-N has approximate dimensions of 52×29×28 Å. Eight β-strands, accounting for 57% of the sequence, form an immunoglobulin-like antiparallel β-sandwich that could not be predicted from the sequence because of its unusual features and topology (Fig. 1). Whereas β-sheet N1 consists of classical immunoglobulin-fold β-strands A-B-E-D[11], strands F and G of β-sheet N2 are much longer than strand C such that they extend beyond the β-sandwich and, together with an extra E′ strand, form a platform (E′-F-G extension) at the bottom of the molecule. The E′ strand, which is preceded by a single-turn α-helix ($\alpha_{EE'}$) and forms a β-bulge with Ser 112 in strand F, extends in the same direction as strand C. A highly structured 11-residue loop (BC) and a single-turn 3_{10} helix ($3_{10\,BC}$) connect β-strands B and C at the top of ZP-N, where a second long loop (FG) is located that is poorly defined in the electron density.

The ZP-N hydrophobic core consists of residues belonging to β-strands B, E, C and the upper half of F and G; β-strand A residues Val 42 and Val 44 also contribute to the core in ZP3 homologues (Fig. 1c). The structure is stabilized by the two invariant disulphides of the ZP domain that clamp both sides of the β-sandwich and, except for part of $Cys_1$46, are completely buried at the bottom of the hydrophobic core (Fig. 2 and Supplementary Fig. 2b). Structure-based alignments indicate that, in other proteins, the ZP-N fold is reinforced by additional disulphides in β-sheet N2; furthermore, they suggest that the ordered loop BC can mediate interactions between ZP-N and preceding domains, because it often contains one extra Cys residue that could form an interdomain disulphide (Supplementary Table 2).

The finding that ZP-N adopts an immunoglobulin-like fold demonstrates that it is a domain of its own. At the same time, ZP-N has only low overall structural similarity to other immunoglobulin-like domains, and even its closest match (D3 domain of ICAM-1 (ref. 12); secondary-structure matching (SSM) *Q*-score 0.29) is considerably different in topology, hydrophobic core and disulphide linkages. Together with the absence of an E′ strand in previously described immunoglobulin-like domains, this suggests that ZP-N belongs to a new immunoglobulin superfamily subtype.

[1]Karolinska Institutet, Department of Biosciences and Nutrition, Hälsovägen 7, SE-141 57 Huddinge, Sweden.
*These authors contributed equally to this work.

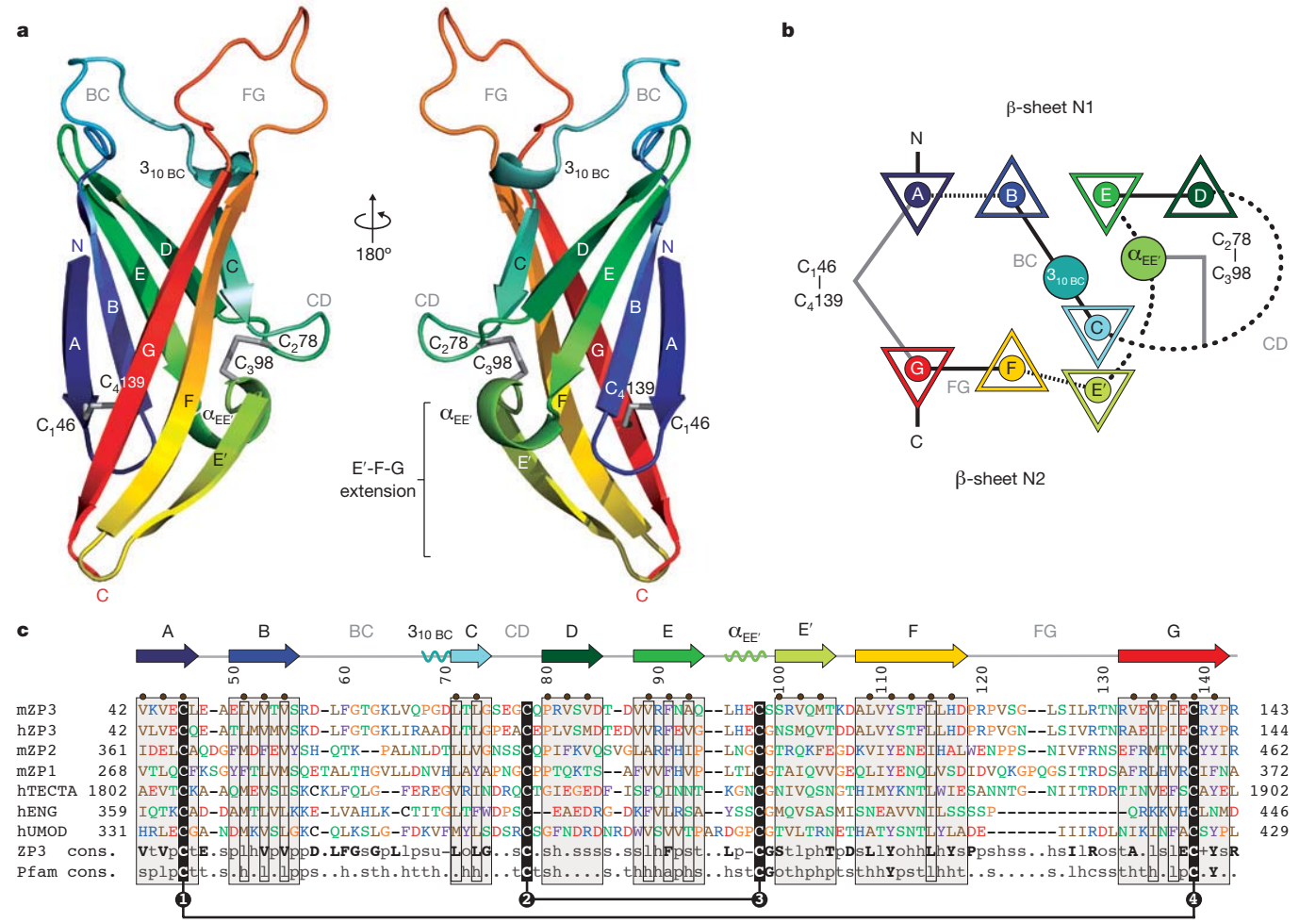

Figure 1 | Overall structure of the ZP-N domain of ZP3. a, Cartoon representation rainbow-coloured from blue (N terminus) to red (C terminus), with conserved disulphides shown as grey sticks. **b**, Topology, with strands as triangles, and helices as circles. Connections between secondary structure elements at the top, middle or bottom of the structure are represented by straight continuous, rounded dashed and straight dashed lines, respectively. **c**, Structure-based alignment of ZP-N sequences of mouse (m) and human (h) zona pellucida proteins, as well as non-egg coat ZP domain proteins α-tectorin (TECTA), endoglin (ENG) and uromodulin (UMOD). Consensus sequences (cons.) for ZP3 homologues and Pfam ZP domain family seed sequences are also shown (for colour-coding and consensus keys, see Methods). Brown circles mark residues in which the side chains lie on the inner side of β-sheets; open and closed black boxes indicate conserved hydrophobic core and Cys residues, respectively.

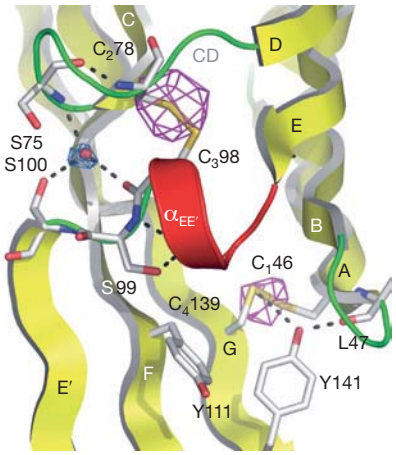

Figure 2 | Key contacts involving conserved ZP-N domain residues. View of the bottom of the hydrophobic core and E′-F-G extension of ZP3 ZP-N, with important residues shown as sticks and hydrogen bonds indicated by dashed lines. Violet mesh is a 4.7 Å resolution phased anomalous difference map contoured at 3.5σ, calculated using diffraction data collected at 8 keV to confirm disulphide bond assignments. Blue mesh is a 2.3 Å resolution composite annealed omit map, contoured at 1.0σ and cropped around a water molecule that mediates interaction between $Cys_3$98, Ser 75 and Ser 100.

Whereas the C-terminal fragment of the ZP domain (ZP-C) of ZP3 has been implicated in sperm binding, ZP-N is involved in zona pellucida filament formation[2,6]. Incorporation of mouse ZP3 into the *Xenopus* vitelline envelope suggests that sequence and structural features important for egg coat assembly are maintained throughout evolution[13]. Two highly conserved regions of ZP-N (Fig. 3a, ovals) expose hydrophobic residues and do not overlap with sites that can be glycosylated or are protease-accessible (Fig. 3b and Supplementary Table 3). Nonspecific juxtaposition of β-strands A, G and the hydrophobic C terminus of MBP (Supplementary Fig. 4) explains why shortening of the MBP/ZP-N linker reduces protein aggregation (Supplementary Fig. 1f). The second region, essentially corresponding to the ZP-N-specific E′-F-G extension, forms intermolecular contacts through antiparallel pairing of symmetry-related E′ β-strands in crystal form III (Fig. 3c) and mediates stacking of molecular layers in the *I*222 and *P*1 crystals (Supplementary Fig. 3). Moreover, it is the part of the structure most strongly predicted to be involved in protein–protein interactions (Supplementary Fig. 5). The extension contains Tyr 111 and Tyr 141 that are involved in an aromatic–aromatic interaction and are the only ZP-N residues, apart from the Cys, to be nearly invariant (Figs 1c and 2). Both tyrosines are missing in ZP domain protein endoglin that does not form polymers[14] (Fig. 1c), and human deafness-associated mutation of the residue corresponding to Tyr 111 in inner ear protein α-tectorin (Tyr1870Cys) disrupts formation of

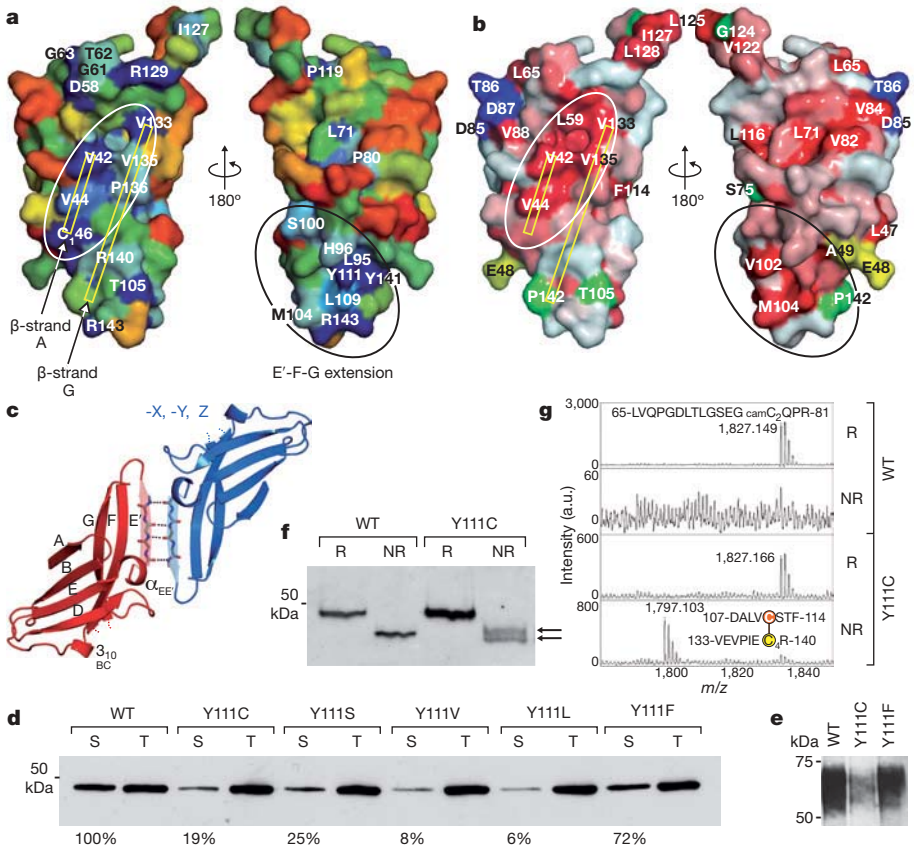

Figure 3 | ZP-N domain interaction sites. **a**, **b**, Surface representation of ZP3 ZP-N, with residues coloured from red to blue by increasing conservation (**a**) or from white to red by increasing hydrophobicity (**b**). N-glycosylation, protease cleavage and RGD sites are green, yellow and blue, respectively (see Supplementary Table 3). Views are from the MBP interface (left) and opposite side (right). **c**, Intermolecular backbone interactions between ZP-N-specific E′ β-strands connect N2 β-sheets in crystal form III. **d**, Immunoblot analysis of soluble (S) and total (T) expression of wild-type (WT) MBP-AAA-ZP-N-6His and Tyr 111 mutants with monoclonal anti-5His antibody. Percentages indicate the amount of soluble protein relative to the wild type. **e**, Immunoblot analysis with polyclonal anti-ZP3 antibody of conditioned medium from Chinese hamster ovary (CHO)-K1 cells expressing full-length ZP3 constructs. **f**, Immunoblot analysis of soluble wild-type and Tyr111Cys MBP-AAA-ZP-N-6His with anti-5His antibody in reducing (R) and non-reducing (NR) conditions. For comparison, tenfold more extract was loaded for the mutant than the wild type. **g**, Matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) analysis of trypsin/chymotrypsin-digested samples from **f** detects an abnormal Cys 111–$Cys_4$139 disulphide in non-reduced Tyr111Cys mutant protein only. The intense peak at *m*/*z* 1827, corresponding to carbamidomethylated (cam) ZP3 peptide Leu 65–Arg 81, confirms efficient disulphide reduction after dithiothreitol treatment.

striated-sheet matrix filaments in the sulcal region of the tectorial membrane[15]. As also observed on substitution with other non-aromatic residues, mutation of Tyr 111 to Cys reduces solubility of MBP-AAA-ZP-N-6His (Fig. 3d), and a corresponding decrease in secretion is seen in full-length ZP3 Tyr111Cys expressed in mammalian cells (Fig. 3e). On the other hand, Tyr111Phe mutants of MBP-AAA-ZP-N-6His and full-length ZP3 are ~70% soluble (Fig. 3d) and almost as efficiently secreted (Fig. 3e) as the respective wild-type proteins, in agreement with the observation that the Tyr is naturally substituted by Phe in a very small number of ZP domain sequences. Impairment of protein secretion is a general feature of most other characterized ZP-N domain mutations, which cause disease by haploinsufficiency owing to effects on folding that can be explained on the basis of the structure (Supplementary Fig. 6 and Supplementary Table 4). Unlike these mutations, the Tyr-to-Cys substitution must however also directly affect filament assembly, because the tectorial membrane is not disrupted in normal-hearing animals heterozygous for a targeted deletion in α-tectorin (*Tecta*$^{+/\Delta ENT}$)[15]. Tyr 111 is adjacent to $Cys_4$139 in the structure (Fig. 2), suggesting that the mutation could interfere with disulphide bond formation. Analysis of MBP-AAA-ZP-N-6His Tyr111Cys shows that this consists of two disulphide isoforms (Fig. 3f, arrows), one of which contains an abnormal Cys111–$Cys_4$139 bridge (Fig. 3g). In molecular dynamics simulations both isoforms are as stable as the wild type when folded and show only local differences where the Tyr side chain is absent (Supplementary Fig. 7). In light of studies on the effect of the mutation on protein assembly into the zona pellucida[4], these data collectively indicate that, although only a small fraction of α-tectorin Tyr1870Cys folds and is secreted, this acts as a dominant negative by incorporating into nascent tectorial membrane filaments and preventing their elongation due to structural alterations around the conserved Tyr. Thus, crystallographic, mutational and biochemical evidence demonstrates a crucial role for the ZP-N-specific E′-F-G extension in polymerization.

Our structure has implications for animal egg coat architecture by strongly supporting the recent hypothesis that zona pellucida/vitelline envelope proteins contain additional N-terminal copies of ZP-N[16]. Using structure-based distant homologue recognition, we identified three isolated ZP-N domains in the N-terminal region of secondary sperm receptor ZP2 and built a three-dimensional model of its entire ZP-N1–N4 array (Fig. 4a, b). The model is structurally sound because it buries conserved hydrophobic residues and, consistent with the absence of free Cys residues in ZP2 (ref. 8), involves most Cys in canonical Cys_1–Cys_4, Cys_2–Cys_3 ZP-N bridges that agree with available disulphide bond assignments[8,17,18]. The remaining two Cys are positioned so that they can form an extra disulphide within ZP-N3 (Fig. 4c). By exposing sites known to be glycosylated, targeted by proteases, or recognized by monoclonal antibodies, the model is also entirely consistent with a large amount of other experimental data available on ZP2

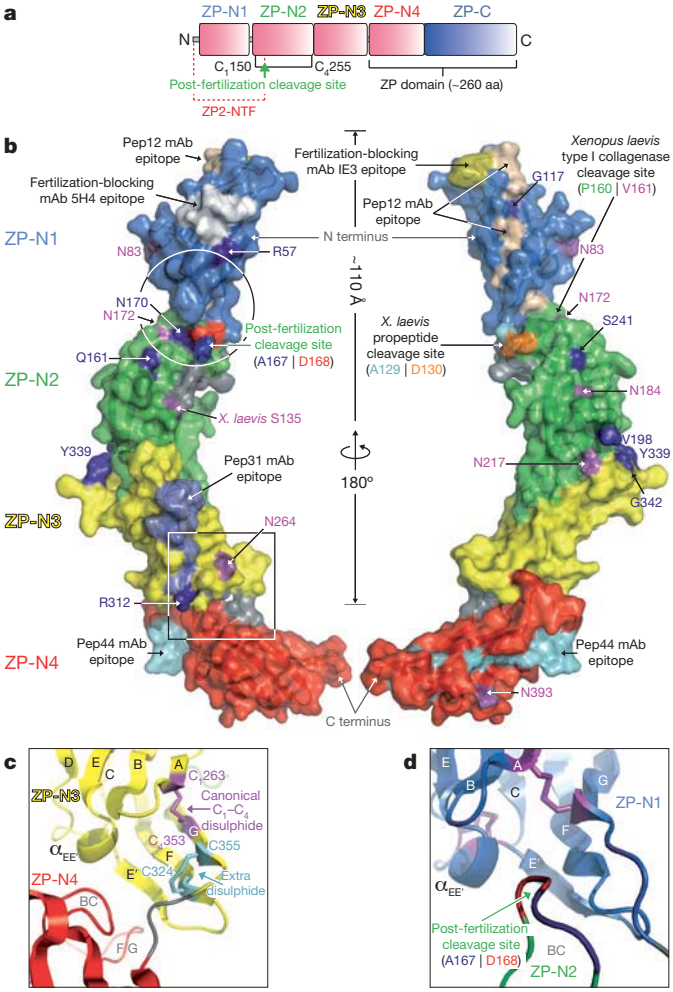

Figure 4 | Model of the ZP-N domain repeat region of ZP2. **a**, **b**, Domain architecture of ZP2 (**a**) and three-dimensional surface model of its ZP-N1–N4 region (**b**), with relevant features indicated (see Supplementary Table 5). aa, amino acids. Glycosylation and positively selected sites are dark blue and violet, respectively. Unless otherwise specified, the residue numbers refer to mouse ZP2 sequence. mAb, monoclonal antibody. **c**, Detail of the region boxed in **b**, highlighting a further disulphide bond between non-canonical Cys residues of ZP-N3. **d**, View of the region circled in **b**, suggesting that post-fertilization cleavage of ZP-N2 BC loop could affect the position of ZP-N1 relative to the rest of ZP2.

(Fig. 4b and Supplementary Table 5). Furthermore, owing to short or absent interdomain linkers, ZP-N1–N4 assumes an extended rod-like structure, the ~110 Å span of which matches the size of protrusions extending from isolated zona pellucida filaments imaged by electron microscopy[19]. This agrees with the observation that the C-terminal ZP domain of ZP2 is responsible for its polymerization into filaments, from which the N-terminal fragment (ZP-N1–N3) could protrude[4]; accordingly, the latter was shown to bind acrosome-reacted sperm[20], providing a target for contraceptive vaccines[21,22].

After gamete fusion, site-specific proteolysis of ZP2 produces an N-terminal fragment (ZP2-NTF) that remains attached to the rest of the protein via a single disulphide bond involving its fifth Cys[17,18,23]. Cleavage of ZP2 triggers hardening of the zona pellucida as part of the block to polyspermy[17,23,24] and was recently proposed to regulate mammalian sperm binding to the zona pellucida by modulating the supramolecular structure of the matrix[24]. On the basis of our model, the ZP2 cleavage site lies within the long BC loop of ZP-N2, whose canonical Cys_1 corresponds to the fifth Cys of ZP2-NTF; thus, after fertilization the latter remains linked to the rest of ZP2 through the Cys_1–Cys_4 disulphide of the second repeat (Fig. 4a, b, d). Proteolytic processing could alter the orientation of ZP-N1 relative to ZP-N2 because the substrate loop lies at the interface between the two repeats, in proximity to the ZP-N1/ZP-N2 linker (Fig. 4d). At the same time, cleavage of the only loop connecting the two β-sheets at the top of ZP-N2 would make the underlying region available for interaction. Considering that other extracellular adhesion molecules consisting of several immunoglobulin superfamily domains often establish cell–cell contacts through the formation of adhesive zippers[12,25], this combination of events could trigger contacts between ZP-N repeats protruding from different zona pellucida filaments. This would result in the observed post-fertilization three-dimensional compaction of filaments in the inner layer of the zona pellucida[26], generating a physical barrier to sperm penetration as well as mask binding sites for sperm on ZP-N1–N3.

Extension of the structure-based analysis that we applied to ZP2 to other zona pellucida/vitelline envelope proteins indicates that the ZP-N fold described in this study accounts for >50% of the structure of all egg coats from frog to human. However, we could not identify extra copies of ZP-N in ZP domain proteins not involved in fertilization, or in vitelline envelope components from fish whose only function is structural. Thus ZP-N repeats are only found in egg coat subunits of vertebrate classes where the zona pellucida/vitelline envelope is directly responsible for species-restricted gamete interaction (Supplementary Fig. 8). ZP2 and ZP3 are among the 10% most divergent proteins in mammals[27] and, on the basis of our model, 11 out of 12 sites suggested to be positively selected in the sequence of mature ZP2 (refs 28 and 29) are located within ZP-N repeats 1–3. Most of these residues are predicted to be exposed and could potentially be available for interaction with sperm (Fig. 4b and Supplementary Table 5). Taken together, these observations raise an unexpected parallel with invertebrate gamete recognition proteins, where rapid evolution of repeated domains is implicated in speciation[30].

The structure of ZP3 ZP-N offers a first glimpse into the basis of mammalian fertilization at atomic resolution. Because many pathological mutations in other ZP domain proteins fall within ZP-N, this work will not only be relevant for reproductive medicine, but also for severe human diseases such as non-syndromic deafness or renal and vascular disorders.

METHODS SUMMARY

Proteins were co-expressed with a leaderless version of disulphide isomerase DsbC and TrxA mutant Gly33Pro/Pro34Tyr in *E. coli* Origami B (DE3)[6]. Monomeric fusions purified by affinity and size-exclusion chromatography were crystallized by vapour diffusion with sodium acetate (crystal form I), PEG 6000 (crystal form II) and PEG 6000/$ZnCl_2$ (crystal form III) as precipitants (Supplementary Fig. 1g–i). Diffraction data was collected at beamlines BM14 (crystal forms I and II) and ID23-1 (crystal form III) of the European Synchrotron Radiation Facility (ESRF), Grenoble. Crystal form I was refined to 2.9 Å resolution in space group *I*222, with a crystallographic *R*-factor (R_{work}) of 19.0% and a free *R*-factor (R_{free}) of 22.7% (Supplementary Table 1). Triclinic crystal form II was refined to 2.3 Å resolution with $R_{work} = 20.1\%$ and $R_{free} = 22.6\%$ (Supplementary Table 1 and Supplementary Fig. 2). Crystal form III was refined to 3.1 Å resolution in space group $P2_12_12$, with $R_{work} = 20.8\%$ and $R_{free} = 25.6\%$ (Supplementary Table 1). The final models for crystal forms I and II, that have essentially identical structure (average ZP-N main chain root mean square (r.m.s.) distance = 0.5 Å) as well as very similar cell dimensions and crystal packing (Supplementary Fig. 3), include all ZP-N residues and have excellent stereochemistry, with no Ramachandran plot outliers. Crystal form III has also similar cell dimensions and structure (r.m.s. distance = 0.8 Å), but lacks electron density for the FG loop and differs in packing (Fig. 3c).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 September; accepted 29 October 2008.

1. Nixon, B., Aitken, R. J. & McLaughlin, E. A. New insights into the molecular mechanisms of sperm-egg interaction. *Cell. Mol. Life Sci.* **64**, 1805–1823 (2007).
2. Wassarman, P. M. Zona pellucida glycoproteins. *J. Biol. Chem.* **283**, 24285–24289 (2008).

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. *Nature Cell Biol.* **4,** 457–461 (2002).
5. Jovine, L., Darie, C. C., Litscher, E. S. & Wassarman, P. M. Zona pellucida domain proteins. *Annu. Rev. Biochem.* **74,** 83–114 (2005).
6. Jovine, L., Janssen, W. G., Litscher, E. S. & Wassarman, P. M. The PLAC1-homology region of the ZP domain is sufficient for protein polymerisation. *BMC Biochem.* **7,** 11 (2006).
7. Llorca, O., Trujillo, A., Blanco, F. J. & Bernabeu, C. Structural model of human endoglin, a transmembrane receptor responsible for hereditary hemorrhagic telangiectasia. *J. Mol. Biol.* **365,** 694–705 (2007).
8. 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).
9. Kanai, S. *et al.* Disulfide linkage patterns of pig zona pellucida glycoproteins ZP3 and ZP4. *Mol. Reprod. Dev.* **75,** 847–856 (2008).
10. Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* **12,** 1313–1322 (2003).
11. Halaby, D. M., Poupon, A. & Mornon, J. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng.* **12,** 563–571 (1999).
12. Yang, Y. *et al.* Structural basis for dimerization of ICAM-1 on the cell surface. *Mol. Cell* **14,** 269–276 (2004).
13. Doren, S. *et al.* Incorporation of mouse zona pellucida proteins into the envelope of *Xenopus laevis* oocytes. *Dev. Genes Evol.* **209,** 330–339 (1999).
14. ten Dijke, P., Goumans, M. J. & Pardali, E. Endoglin in angiogenesis and vascular diseases. *Angiogenesis* **11,** 79–89 (2008).
15. Legan, P. K. *et al.* A deafness mutation isolates a second role for the tectorial membrane in hearing. *Nature Neurosci.* **8,** 1035–1042 (2005).
16. 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).
17. Lindsay, L. L. & Hedrick, J. L. Proteolysis of *Xenopus laevis* egg envelope ZPA triggers envelope hardening. *Biochem. Biophys. Res. Commun.* **324,** 648–654 (2004).
18. Tian, J., Gong, H. & Lennarz, W. J. *Xenopus laevis* sperm receptor gp69/64 glycoprotein is a homolog of the mammalian sperm receptor ZP2. *Proc. Natl Acad. Sci. USA* **96,** 829–834 (1999).
19. Wassarman, P. M. & Mortillo, S. Structure of the mouse egg extracellular coat, the zona pellucida. *Int. Rev. Cytol.* **130,** 85–110 (1991).
20. Tsubamoto, H. *et al.* Expression of recombinant human zona pellucida protein 2 and its binding capacity to spermatozoa. *Biol. Reprod.* **61,** 1649–1654 (1999).
21. Hasegawa, A., Hamada, Y., Shigeta, M. & Koyama, K. Contraceptive potential of synthetic peptides of zona pellucida protein (ZPA). *J. Reprod. Immunol.* **53,** 91–98 (2002).
22. Sun, W., Lou, Y. H., Dean, J. & Tung, K. S. A contraceptive peptide vaccine targeting sulfated glycoprotein ZP2 of the mouse zona pellucida. *Biol. Reprod.* **60,** 900–907 (1999).
23. Bleil, J. D., Beall, C. F. & Wassarman, P. M. Mammalian sperm-egg interaction: fertilization of mouse eggs triggers modification of the major zona pellucida glycoprotein, ZP2. *Dev. Biol.* **86,** 189–197 (1981).
24. Rankin, T. L. *et al.* Fertility and taxon-specific sperm binding persist after replacement of mouse sperm receptors with human homologs. *Dev. Cell* **5,** 33–43 (2003).
25. Aricescu, A. R. & Jones, E. Y. Immunoglobulin superfamily cell adhesion molecules: zippers and signals. *Curr. Opin. Cell Biol.* **19,** 543–550 (2007).
26. Familiari, G., Heyn, R., Relucenti, M. & Sathananthan, H. Structural changes of the zona pellucida during fertilization and embryo development. *Front. Biosci.* **13,** 6730–6751 (2008).
27. Makalowski, W. & Boguski, M. S. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc. Natl Acad. Sci. USA* **95,** 9407–9412 (1998).
28. Swanson, W. J., Yang, Z., Wolfner, M. F. & Aquadro, C. F. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl Acad. Sci. USA* **98,** 2509–2514 (2001).
29. Turner, L. M. & Hoekstra, H. E. Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Mol. Biol. Evol.* **23,** 1656–1669 (2006).
30. Galindo, B. E., Vacquier, V. D. & Swanson, W. J. Positive selection in the egg receptor for abalone sperm lysin. *Proc. Natl Acad. Sci. USA* **100,** 4639–4643 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgments This work was supported by Karolinska Institutet, the Swedish Research Council (grant 2005-5102) and European Community (Marie Curie ERG 31055). We thank P. Nordlund and G. Schneider for access to the Stockholm SGC robotic crystallization facility and beamtime; H. Belrhali, J. McCarthy, R. Ravelli and M. Walsh for assistance at ESRF; P. Afonine, R. Grosse-Kunstleve and P. Zwart for help with phenix.refine; G. Murshudov for help with REFMAC; C. Chothia, F. Cotelli, J.-Å. Gustafsson, R. Herbst-Irmer, A. Kohl, R. Ladenstein, M. Letarte, E. Litscher, E. Morgunova, A. Murzin, K. Nagai, L. Nilsson, D. Rhodes, R. Toftgård and P. Wassarman for discussions and comments.

Author Contributions M.M., L.H. and L.J. generated constructs and performed protein expression and purification. T.S. performed mass spectrometric analysis. S.B. carried out molecular dynamics simulations. L.J. and M.M. crystallized proteins, determined structures and wrote the manuscript.

Author Information Atomic coordinates and structure factors have been deposited with the Protein Data Bank under accession codes 3D4C (crystal form I), 3D4G (crystal form II) and 3EF7 (crystal form III). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.J. (luca.jovine@ki.se).

METHODS

DNA constructs. A PCR fragment encoding ZP3 ZP-N was cloned into vectors pLJMBP4c (ref. 6) or pLJMBP6c to generate constructs expressing cytoplasmic MBP Glu385Ala/Lys388Ala/Asp389Ala/Arg393Asn (ref. 31), fused via an Ala-Ala-Ala linker to C- or N-terminally histidine-tagged ZP3 ZP-N. Point mutations were introduced using the QuickChange protocol (Stratagene). All constructs were verified by DNA sequencing.

Protein expression and purification. Proteins were expressed essentially as described[6]. Soluble fusions were captured by immobilized-metal affinity chromatography using HisTrap HP columns (GE Healthcare), and monomeric protein was eluted separately from oligomeric and aggregated species with a 2 column volume linear gradient to 86 mM imidazole (~9.4 mS cm^{-1}), followed by an 18 column volume gradient to 500 mM imidazole (Supplementary Fig. 1b, d, e). Concentrated peak fractions from the first gradient were incubated with 10 mM D-(+)-maltose, further purified by size-exclusion chromatography using a HiLoad 26/60 Superdex 75 prep grade column (GE Healthcare) (Supplementary Fig. 1c–e) and concentrated to 15 mg ml^{-1}. The identity of the purified product and the presence of native disulphide bond connectivity were confirmed by mass spectrometry. Immunoblotting experiments were performed as previously described[6].

Mass spectrometry. Samples were prepared as described[6] and analysed by MALDI–TOF and MALDI–TOF/TOF (Ultraflex II, Bruker Daltonics). Spectra were annotated with the Flexanalysis software (Bruker Daltonics).

Crystallization and diffraction data collection. Crystal form I of MBP-AAA-ZP-N-6His grew in 3–10 days at 20 °C in 1.0 M sodium acetate, 100 mM Na-HEPES, pH 7.5, 50 mM $CdSO_4$. Macroseeded specimens (~250×200×100 μm; Supplementary Fig. 1g) were cryoprotected by gradual transfer into 1.5 M magnesium acetate, 100 mM Na-HEPES, pH 7.5, 10 mM Tris-HCl, pH 7.2, 50 mM $CdSO_4$, 1 mM D-(+)-maltose, mounted in cryoloops and flash frozen in liquid nitrogen. They diffracted to 2.9 Å at synchrotron and belonged to space group *I*222, with one fusion protein per asymmetric unit (Supplementary Table 1). Essentially identical crystals could also be grown in the same condition using an N-terminally tagged fusion construct (6His-MBP-AAA-ZP-N); however, because the diffraction quality of these crystals was not superior to those of MBP-AAA-ZP-N-6His, they were not pursued further.

Crystal form II of MBP-AAA-ZP-N-6His grew over the course of 3 weeks in 8–24% PEG 6000, 100 mM Na-HEPES, pH 7.0–8.0, 125–250 mM $CaCl_2$ at 20 °C. Crystals (~250×125×125 μm; Supplementary Fig. 1h) were cryoprotected by stepwise transfer into a solution equivalent to the mother liquor plus 20% PEG 400. They diffracted to a Bragg spacing of 2.3 Å at synchrotron, with eight molecules per asymmetric unit in a primitive triclinic lattice (Supplementary Table 1). Also in the case of this crystal form, crystals of construct 6His-MBP-AAA-ZP-N could be grown in related conditions that had the same symmetry and very similar unit cell parameters. A 2.7 Å resolution data set (R_{sym} = 3.7 (43.2)) was collected to assess whether the position of the histidine tag within the fusion constructs had any effect on the structure of ZP-N.

Crystal form III of MBP-AAA-ZP-N-6His grew into feather-shaped aggregates in 2 days in 12% PEG 6000, 100 mM Tris-HCl, pH 8.2, 150 mM $CaCl_2$, 2.5 mM $ZnCl_2$ at 20 °C (Supplementary Fig. 1i). Pieces of a crushed aggregate were cryoprotected in 12% PEG 6000, 100 mM Na-HEPES, pH 7.8, 150 mM $CaCl_2$, 5 mM $ZnCl_2$ supplemented with 20% PEG 400 and screened at synchrotron for presence of a single lattice. A 3.1 Å resolution data set was obtained, which suggested that the protein crystallized in space group $P2_12_12$, with two molecules per asymmetric unit. All data sets were processed with the XDS package[32].

Structure determination. Crystal form I: The structure of MBP-AAA-ZP-N-6His was first solved in space group *I*222 by molecular replacement with Phaser[33], using an ensemble of selected MBP structures as search model. Correctness of the solution was confirmed by strong difference electron density (>6σ) for a molecule of D-(+)-maltose (coordinates for which were not included in the search ensemble) within the MBP ligand-binding pocket. After statistical phase improvement in Pirate[34], clear density could also be observed for fragments of ZP-N which, as expected, were positioned on the C-terminal side of the MBP moiety. The ZP-N model was built manually with Coot[35] and the backrub tool of KiNG[36], using 3.2 Å resolution SigmaA-weighted difference Fourier maps[37], composite annealed omit maps calculated with CNS[38] using a starting temperature of 4,500 K to minimize model bias, and X-ray holographic image reconstruction in Eden[39]. The position of the two conserved disulphide bridges of ZP-N was confirmed by sulphur anomalous signal in phased difference maps generated using 5.9-fold redundant low resolution data (Fig. 2). The structure was refined with phenix.refine[40] using a maximum likelihood (ML) target. Bulk solvent correction was applied throughout, and Translation/Libration/Screw refinement of B factors was performed during final rebuilding cycles, on the basis of TLSMD[41] analysis of individually refined atomic displacement parameters.

Crystal form II: A set of coordinates partially refined against the *I*222 data was used to solve the high resolution structure of MBP-AAA-ZP-N-6His by molecular replacement with Phaser. Initial refinement was complicated by the presence of strong pseudo-orthorhombic symmetry, including a translational non-crystallographic symmetry (NCS) component. However, a series of parallel test refinement jobs in all possible space groups clearly indicated that this crystal form belongs to the triclinic system and allowed structure determination to proceed. The structure was refined in space group *P*1 with phenix.refine and REFMAC[42], essentially as described above; partial NCS restraints were also applied, except in the final cycles of refinement. NCS editing features of Coot were used during rebuilding, but molecules within the asymmetric unit were also individually rebuilt at two different stages of refinement. FG loops were built iteratively on the basis of traces generated with Xpleo[43]; solvent structure was assembled by phenix.refine and manually edited in Coot. Comparison of the final model (Supplementary Fig. 2) to partially refined coordinates of 6His-MBP-AAA-ZP-N (R_{work} = 19.3%, R_{free} = 26.0%) revealed essentially no difference between ZP-N moieties (average main chain r.m.s. distance excluding FG loop residues = 0.13 Å), indicating that the structure of the domain is not influenced by presence of a C-terminal histidine tag.

Crystal form III: The structure was solved by molecular replacement with Phaser, using chain A of the refined *P*1 coordinates as search model. Ordered Zn^{2+} ions were identified in phased anomalous electron density maps and used to obtain experimental phase information using the EP_SAD mode of Phaser (figure of merit = 0.280 (0.132)). Refinement against a 7.6-fold redundant data set collected at the Zn K-edge (λ = 1.28235 Å) was carried out with phenix.refine, using a phased maximum likelihood (MLHL) target and anomalous scattering factors (f′ = −8.21, f″ = 4.75) derived from X-ray fluorescence spectra with CHOOCH[44]. Because visual inspection of initial maps revealed that one of the two molecules within the asymmetric unit (chain B) was relatively disordered (presumably due to the lack of β-strand E'-mediated contacts), no NCS restraints were used in this case.

Models were validated with MolProbity[45] and ValLigURL[46]. The side chains of ZP3 residues Glu 45 ($P2_12_12$ only); Glu 48 (*P*1 chain C, $P2_12_12$ chain A); Lys 43; Leu 59 ($P2_12_12$ chain B); Lys 64; Gln 67 ($P2_12_12$); Leu 73 ($P2_12_12$ chain B); Ser 75 (*I*222); Glu 76; Gln 79 ($P2_12_12$ chain B); Arg 81 and Arg 90 ($P2_12_12$); Arg 101 ($P2_12_12$ chain B); Lys 106; Asp 107, Arg 132 and Arg 140 ($P2_12_12$ chain B) are partially disordered. Side chains of FG loop residues Arg 120–Arg 129 are poorly resolved, and backbone density for Ser 123–Ser 126 (*I*222) or Pro 121–Ser 126 (*P*1) is weak; nevertheless, the latter regions could be modelled because their position is constrained by the better defined residues that flank them within the loop. No electron density could be observed for FG loop residues Pro 119–Asn 131 in the $P2_12_12$ map.

Sequence and structure analysis. Structure-based multiple sequence alignments were assembled using STRAP[47] and 3D-Coffee[48], starting from iteratively optimized pairwise alignments generated by MODELLER[49]. Residues in Fig. 1c are colour-coded as follows: brown, hydrophobic; purple, aromatic; orange, [Pro, Gly]; white, Cys; green, polar; blue, positively charged; red, negatively charged. The 70% threshold consensus sequences for ZP3 homologues with ≤90% identity (ZP3 cons.) and for Pfam ZP domain family seed sequences (Pfam cons.) were calculated with Consensus[50]; lowercase characters correspond to the following amino acid sets: a/aromatic = [Phe, Tyr, Trp, His]; l/aliphatic = [Ile, Val, Leu]; h/hydrophobic = [Phe, Tyr, Trp, His, Ile, Val, Leu, Ala, Gly, Met, Cys, Lys, Arg, Thr]; +/positive = [His, Lys, Arg]; −/negative = [Asp, Glu]; c/charged = [His, Lys, Arg, Asp, Glu]; p/polar = [His, Lys, Arg, Asp, Glu, Gln, Asn, Ser, Thr, Cys]; o/alcohol = [Ser, Thr]; u/tiny = [Gly, Ala, Ser]; s/small = [Gly, Ala, Ser, Val, Thr, Asp, Asn, Pro, Cys]; t/turnlike = [Gly, Ala, Ser, His, Lys, Arg, Asp, Glu, Gln, Asn, Ser, Thr, Cys]; (.) = any residue; uppercase letters indicate the specific amino acids with the same one-letter code. Evolutionary conservation was analysed with ConSurf[51], using alignments of non-redundant sequence data sets generated with 90% identity threshold. Interfaces were studied with PISA[52], cavities were analysed with CAVER[53] and Poisson-Boltzmann electrostatic calculations were performed using APBS[54]. Structural similarity was evaluated with SSM[55] and Dali[56]. Protein–protein interaction sites were predicted with PINUP[57], ISIS[58], ProMate[59], SPPIDER[60] and InterProSurf[61]. Figures were prepared using PyMOL[62].

Homology recognition and modelling of isolated ZP-N domains. Non-ZP domain ZP-Ns were identified using Fugue[63] and a HOMSTRAD database[64] that consisted of 8,724 folds and included a profile for ZP-N. MODELLER-generated homology models of the hits were then used to generate extra profiles for further scanning of ZP domain protein sequences, and the procedure was repeated until no more copies of ZP-N could be found. Using this method, all ZP-N repeats of mouse ZP2 were identified with >95% confidence. Parallel searches were also performed with a profile hidden Markov model created using HMMER[65], on the basis of ZP3 ZP-N secondary structure and surface accessibility information.

Modelling of the N-terminal fragment of ZP2. Individual models of ZP-N domains 1–4 of mouse ZP2 (ZP-N1, Gly 47–Gln 138; ZP-N2, Glu 146–Asp 258; ZPN-3, Leu 259–Ser 357; ZP-N4, Ile 361–Arg 462) were created with MODELLER and merged into a single chain using Coot. Relative orientation of the domains was constrained by the short loop sequences connecting adjacent ZP-Ns. This was particularly evident in the case of ZP-N2 and ZP-N3, in which, owing to the absence of an interdomain linker, the first β-strand of ZP-N3 is positioned so that it extends in the same direction as the last β-strand of ZP-N2; as a result, the BC and FG loops of ZP-N3 wrap around the lower part of ZP-N2. A similar arrangement is observed in the case of ZP-N4 and ZP-N3, respectively. Relative positioning of ZP-N1 and ZP-N2 depends on interactions between their longer intradomain linker and both the BC and FG loops of ZP-N2; the tip of the former loop, which is particularly long owing to insertion of the post-fertilization cleavage site, also directly contacts the bottom part of ZP-N1. Coordinates were energy minimized using CHARMM[66], resulting in a final model with 96% of the residues in allowed Ramachandran plot regions. Accessibility of sites experimentally determined to be N-glycosylated[8] was assessed with GlyProt[67].

Molecular dynamics simulations. Coordinates of ZP3 residues from chain A of the *P*1 structure were used to set up a wild-type model of ZP-N in CHARMM. A model of mutant Tyr111Cys was constructed from the wild-type coordinates by modifying the Tyr 111 side chain with CHARMM and set up with two alternative disulphide bond connectivities, one in which the mutated residue was free (Tyr111Cys*) and the other one where Cys 111 was disulphide-bonded to Cys_4139, leaving Cys_146 unpaired ($Tyr111Cys^X$). Models were solvated using TIP3P water[68] in a 65×40×40 Å box. The system was minimized with the steepest descent and ABNR method, initially with Cα atom constraints (force constant 20/50 kcal mol^{-1} Å^{-2}) and then without constraints. A standard simulation protocol was used[69] with temperature set to 300 K and allowed to vary within ±10 K, as well as periodic boundary conditions and weak constraints on Cα atoms (1 kcal mol^{-1} Å^{-2}) to maintain biochemical properties. Molecular dynamics simulations were performed for 10 ns and systems were equilibrated for 400 ps before data collection started.

Mammalian cell culture and transient expression experiments. Complementary DNA fragments encoding full-length wild-type mouse ZP3 and mutants Tyr111Cys and Tyr111Phe were cloned into vector pSI (Promega) and transfected into CHO-K1 cells (American Type Culture Collection) as previously described[4]. Immunoblotting was performed with anti-ZP3 polyclonal antibody 8818 (1:10,000 dilution; Pocono Rabbit Farm) and peroxidase-conjugated AffiniPure goat anti-rabbit IgG (H+L) (1:10,000; Jackson ImmunoResearch Europe).

31. Center, R. J. *et al.* Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. *Protein Sci.* **7,** 1612–1619 (1998).
32. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26,** 795–800 (1993).
33. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
34. Cowtan, K. General quadratic functions in real and reciprocal space and their application to likelihood phasing. *Acta Crystallogr. D* **56,** 1612–1621 (2000).
35. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
36. Davis, I. W., Arendall, W. B. III, Richardson, D. C. & Richardson, J. S. The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure* **14,** 265–274 (2006).
37. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42,** 140–149 (1986).
38. Brünger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2,** 2728–2733 (2007).
39. Somoza, J. R. *et al.* Holographic methods in X-ray crystallography. IV. A fast algorithm and its application to macromolecular crystallography. *Acta Crystallogr. A* **51,** 691–708 (1995).
40. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. The Phenix refinement framework. *CCP4 Newsletter* **42,** contribution 8 (2005).
41. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62,** 439–450 (2006).
42. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
43. van den Bedem, H., Lotan, I., Latombe, J. C. & Deacon, A. M. Real-space protein-model completion: an inverse-kinematics approach. *Acta Crystallogr. D* **61,** 2–13 (2005).
44. Evans, G. & Pettifer, R. F. CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Cryst.* **34,** 82–86 (2001).
45. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).
46. Kleywegt, G. J. & Harris, M. R. ValLigURL: a server for ligand-structure comparison and validation. *Acta Crystallogr. D* **63,** 935–938 (2007).
47. Gille, C. & Frommel, C. STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics* **17,** 377–378 (2001).
48. O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D. G. & Notredame, C. 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *J. Mol. Biol.* **340,** 385–395 (2004).
49. Fiser, A. & Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **374,** 461–491 (2003).
50. Bork, P. Consensus. <http://coot.embl.de/Alignment/consensus.html>.
51. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33,** W299–W302 (2005).
52. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
53. Petrek, M. *et al.* CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* **7,** 316 (2006).
54. 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).
55. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60,** 2256–2268 (2004).
56. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20,** 478–480 (1995).
57. Liang, S., Zhang, C., Liu, S. & Zhou, Y. Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res.* **34,** 3698–3707 (2006).
58. Ofran, Y. & Rost, B. Protein-protein interaction hotspots carved into sequences. *PLOS Comput. Biol.* **3,** e119 (2007).
59. Neuvirth, H., Raz, R. & Schreiber, G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J. Mol. Biol.* **338,** 181–199 (2004).
60. Porollo, A. & Meller, J. Prediction-based fingerprints of protein-protein interactions. *Proteins* **66,** 630–645 (2007).
61. Negi, S. S. & Braun, W. Statistical analysis of physical-chemical properties and prediction of protein-protein interfaces. *J. Mol. Model.* **13,** 1157–1167 (2007).
62. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, 2002).
63. Shi, J., Blundell, T. L. & Mizuguchi, K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310,** 243–257 (2001).
64. Stebbings, L. A. & Mizuguchi, K. HOMSTRAD: recent developments of the Homologous Protein Structure Alignment Database. *Nucleic Acids Res.* **32,** D203–D207 (2004).
65. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14,** 755–763 (1998).
66. Brooks, B. R. *et al.* CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4,** 187–217 (1983).
67. Bohne-Lang, A. & von der Lieth, C. W. GlyProt: in silico glycosylation of proteins. *Nucleic Acids Res.* **33,** W214–W219 (2005).
68. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).
69. Burendahl, S., Treuter, E. & Nilsson, L. Molecular dynamics simulations of human LRH-1: the impact of ligand binding in a constitutively active nuclear receptor. *Biochemistry* **47,** 5205–5215 (2008).