FEBS Letters

# A common 'aggregation-prone' interface possibly participates in the self-assembly of human zona pellucida proteins

Nikolaos N. Louros[1], Evangelia D. Chrysina[2], Georgios E. Baltatzis[3], Efstratios S. Patsouris[3], Stavros J. Hamodrakas[1] and Vassiliki A. Iconomidou[1]

1 Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Greece
2 Institute of Biology, Medicinal Chemistry and Biotechnology, National Hellenic Research Foundation, Athens, Greece
3 1st Department of Pathology, Medical School, University of Athens, Greece]

**Correspondence**
V. A. Iconomidou, Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 157 01, Greece
Fax: +30 210 7274254
Tel: +30 210 7274871
E-mail: veconom@biol.uoa.gr

(Received 9 December 2015, revised 1 February 2016, accepted 5 February 2016, available online 26 February 2016)

doi:10.1002/1873-3468.12099

Edited by Peter Brzezinski

**Human zona pellucida (ZP) is composed of four glycoproteins, namely ZP1, ZP2, ZP3 and ZP4. ZP proteins form heterodimers, which are incorporated into filaments through a common bipartite polymerizing component, designated as the ZP domain. The latter is composed of two individually folded subdomains, named ZP-N and ZP-C. Here, we have synthesized six 'aggregation-prone' peptides, corresponding to a common interface of human ZP2, ZP3 and ZP4. Experimental results utilizing electron microscopy, X-ray diffraction, ATR FT-IR spectroscopy and polarizing microscopy indicate that these peptides self-assemble forming fibrils with distinct amyloid-like features. Finally, by performing detailed modeling and docking, we attempt to shed some light in the self-assembly mechanism of human ZP proteins.**

**Keywords:** amyloid fibrils; electron microscopy; functional amyloid; homology modeling; peptide-analogs; zona pellucida

## Highlights

- Human ZP proteins form heterodimers that are incorporated into ZP filaments.
- A common 'aggregation-prone' interface of human ZP proteins is predicted.
- Corresponding peptide-analogs present characteristic amyloidogenic properties.
- Computational studies suggest that ZP proteins form dimers via their 'aggregation-prone' interfaces.
- A possible mechanism for ZP protein self-assembly is proposed.

Mammalian oocytes are primarily responsible for synthesis of a protective extracellular coat that controls polyspermy and regulates species-specific recognition of sperm during fertilization [1,2]. In humans, this porous matrix is known as zona pellucida (ZP) and is composed of four individual protein subunits, namely ZP1, ZP2, ZP3 and a fourth protein with high sequence and structural similarities to ZP1, called ZP4

**Abbreviations**
ATR FT-IR spectroscopy, attenuated total reflectance Fourier-transform infrared spectroscopy; TEM, transmission electron microscopy; TFA, trifluoroacetic acid; ZP, zona pellucida.

(ZP1-like) [3,4]. ZP proteins have the ability to form homodimers under nondenaturing conditions [5]. However, detailed evidence suggests that ZP2 and ZP1/ZP4 are incorporated into dimers along with ZP3 forming long interconnected filaments [6–8]. ZP filaments are in turn cross-linked through ZP1 homodimers, proposed to be stabilized by intermolecular disulfide bonds [6,9]. Evolutionary analysis has highlighted that the overall architecture of animal egg coats exhibits remarkable conservation during evolution [9], with similar components found in marine invertebrates [10] and possible homologs located even in yeast species [11].

ZP proteins share a common structure and are incorporated into the heterodimer building blocks of ZP filaments, after cleavage close to their C-terminal by a furin-like protease [12–15]. ZP proteins share a common polymerizing module, designated as ZP domain. This structural unit is composed of approximately 260 residues and is characterized by a conserved pattern of intermolecular disulfide bonds, formed by 8 (type I ZP domain, such as in ZP3) or 10 (type II ZP domain, such as in ZP1, ZP2 and ZP4) invariant Cys residues [16]. ZP1 and ZP4 are also distinguished by an additional feature preceding the ZP domain, namely a single trefoil/P domain [17]. The ZP domain is composed of two individually folded subdomains, known as ZP-N and ZP-C domain, respectively [12]. Several lines of evidence have clearly shown that the former is responsible for polymerization of ZP proteins [18,19]. The mechanism of ZP protein assembly currently remains unknown, however, the crystal structure of the ZP-N domain and the ZP domain as a whole from mouse and chicken ZP3 have been determined, revealing that the ZP-N domain adopts a unique Ig-like β-sandwich fold, which consists of two facing antiparallel β-sheets that are composed of 8 β-strands (A to G) [19,20].

Studies have previously indicated that the protective coats enclosing silkmoth and teleostean eggs possess distinctive features of amyloids [21–23]. Consistent evidence was also derived by detailed structural studies performed on isolated mouse ZP matrices [24]. Amyloids typically form under certain denaturing microenvironmental conditions, deviating from physiological, in which otherwise soluble proteins are converted into insoluble ordered fibrous aggregates, known as amyloid fibrils [25]. Nevertheless, organisms spanning from bacteria to humans express proteins that natively form filamentous arrangements, known as functional amyloid, sharing common features and qualities of amyloid, in order to support fundamental biological processes [21,26–28]. Experimental data has highlighted that the overall self-aggregation tendency of a protein may be directed by the presence of short sequence stretches with an inherent aggregation propensity [29–32]. Recently, peptide segments corresponding to a possible 'aggregation-prone' interaction site, composed of the A and G β-strands of the putative ZP-N domain from hZP1, therein and henceforth called the AG interface, were shown to possess characteristic amyloidogenic properties [33]. Following these results, it was also shown that the homologous region of ZPB (ZP1-like) proteins from teleosts exhibits similar 'aggregation-prone' features [34]. In this work, we investigated the amyloidogenic properties of six 'aggregation-prone' peptides corresponding to the potent AG interfaces of the remaining human ZP proteins. Our experimental findings demonstrate that the peptides self-assemble into filaments with apparent amyloidogenic features, implying the amyloidogenic properties of human ZP protein AG interfaces. Correspondingly, through extensive computational modeling, we attempt to provide structural insights into the formation of ZP filaments by ascertaining the contribution of the AG interface in the formation of the dimeric building blocks composing ZP filaments.

## Materials and methods

### Comparative modeling of human ZP protein ZP-N domains

Sequences of all human ZP proteins, namely ZP1 (hZP1), ZP2 (hZP2), ZP3 (hZP3) and ZP4 (hZP4), in addition to mouse ZP3 (mZP3) (Uniprot ID: P60852, Q05996, P10761, Q12836 and P10761, respectively), were extracted from Uniprot [35]. Sequence alignment of the corresponding ZP-N domains of all hZP proteins to mZP3 was performed using Clustal Omega [36]. Subsequently, three-dimensional models of the ZP-N domain of hZP1, hZP2, hZP3 and hZP4 were derived, based on this alignment, utilizing MODELLER 9v2 [37,38] and the crystal structure of the mZP3 ZP-N domain as template (PDB ID: 3D4C).

### Identification of potential ZP-N polymerization interfaces

Sequences of all well-reviewed zona pellucida proteins were extracted from Uniprot (Uniprot ID: P60852, Q62005, O54766, I6M4H4, Q05996, P20239, O54767, P48829, Q9BH10, P42099, P47983, P47984, P21754, P10761, P48833, P97708, P48830, P42098, P48831, P48832, P53785, P23491, P53786, P79762, Q12836, Q8CH34, Q00193, Q9BH11, P48834 and Q07287). Multiple sequence alignment was performed utilizing Clustal Omega [36]. The derived alignment and the ZP-N modeled structures were used as an input to implement WHISCY [39] and the consensus predictor meta-PPISP [40], both algorithmic approaches towards the identification of protein–protein interactive surfaces for hZP1, hZP2, hZP3 and hZP4.

## Peptide synthesis and sample preparation

‘Aggregation-prone’ peptides representing the A and G β-strands of the ZP-N domain from human ZP2, ZP3 and ZP4, namely TGELCT (HZP2_A), FRMTVKC (HZP2_G), VLVECQ (HZP3_A), AEIPIEC (HZP3_G), VTLHCT (HZP4_A) and FRLHVSC (HZP4_G), were synthesized based on our detailed computational analysis and the respective homology to their hZP1 equivalents, previously shown to possess amyloidogenic properties [33], with a simple substitution of their cysteine (C) residues by alanine (A), to avoid formation of undesired disulfide bonds, at neutral pH. Peptides were synthesized by GeneCust Europe, Luxembourg (purity > 98%, free N- and C-terminals). All synthesized peptides were dissolved in distilled water (pH 5.75, concentration 10 mg·mL$^{-1}$) and were incubated for a time period of 1–2 weeks. All peptides were found to fold and self-assemble into amyloid-like fibrils, producing gels. Mature amyloid-like fibrils were formed after an incubation period of 1–2 weeks, at ambient temperatures.

## Negative staining and transmission electron microspopy (TEM)

Fibril suspensions from all peptide solutions were applied to glow-discharged 400-mesh carbon and plastic-coated copper grids for 60 s. The grids were stained with a drop of 2% (w/v) aqueous uranyl acetate for 60 s. Removal of excess stain was performed in air, by blotting with a filter paper. The grids were air dried and examined with a Morgagni™ 268 transmission electron microscope, operated at 80 kV. Digital acquisitions were performed with an 11 megapixel side-mounted Morada CCD camera (Soft Imaging System, Muenster, Germany).

## X-ray fiber diffraction

Suspensions (10 μL) of each peptide solution were placed between siliconized rods, spaced 2 mm apart. The droplets were air dried slowly at ambient conditions to form oriented fibers suitable for X-ray diffraction. X-ray diffraction patterns for HZP2_A, HZP4_A and HZP4_G peptides were collected, using a SuperNova-Agilent Technologies X-ray generator equipped with a 135-mm ATLAS CCD detector and a 4-circle kappa goniometer, at the Institute of Biology, Medicinal Chemistry and Biotechnology, National Hellenic Research Foundation (CuK$_\alpha$ high intensity X-ray microfocus source, $\lambda = 1.5418$ Å), operated at 50 kV, 0.8 mA. The specimen-to-film distance was set at 52 mm. The exposure time was set to 300 s. The X-ray diffraction patterns were initially viewed using the program CRYSALISPRO [41] and subsequently displayed and measured with the aid of the program IMOSFLM [42]. X-ray diffraction patterns for HZP2_G, HZP3_A and HZP3_G were collected at the P14 beamline, at a wavelength of 1.23953 Å (Petra III, EMBL-Hamburg, Germany) using a PILATUS 6 M detector. The detector distance was set to 225.11 mm and exposure times were set to 1 s. X-ray diffraction patterns were displayed and measured using IMOSFLM.

## Attenuated total reflectance Fourier-transform infrared spectroscopy and post-run spectra computations

Droplets of all peptide fibril solutions were cast on flat stainless-steel plates coated with an ultrathin hydrophobic layer (SpectRIM, Tienta Sciences, Inc. Indianapolis, USA). The drops were allowed to air dry at ambient conditions to form hydrated thin films. IR spectra were obtained at a resolution of 4 cm$^{-1}$, utilizing an IR microscope (IRScope II, BrukerOPTICS, Bruker Optik GmbH, Ettlingen, Germany), equipped with a Ge ATR objective lens (20×) and attached to a FT-IR spectrometer (Equinox 55, BrukerOPTICS). Ten 32-scan spectra were collected from each sample and averaged to improve the signal to noise ratio. All spectra are shown in the absorption mode after correction for the wavelength-dependence of the penetration depth ($d_p$ analogous λ).

## Congo red staining and polarized light stereomicroscopy

Fibril suspensions of each peptide-analog solution were applied to glass slides and air dried at ambient conditions. The films produced, containing amyloid-like fibrils, were stained with a 1% Congo red solution in distilled water (pH 5.75) for 20 min [43]. Excess stain was removed through tap water washes [43]. The samples were observed under bright field illumination and between crossed polars, using a Leica MZ75 polarizing stereomicroscope equipped with a JVC GC-X3E camera.

## Docking procedures

The derived ZP-N domain models of human ZP proteins were utilized in driven docking experiments with HADDOCK version 2.2 [44]. CNS1.2 was utilized for structure calculations [45]. Nonbonded interactions were calculated with the OPLS force field [46] using a cutoff of 8.5 Å. The solvated docking protocol [47] explicitly accounting for solvent during the docking procedure was preferred to unsolvated docking since it may yield higher quality docking predictions [48]. Interaction restraints to drive the docking were set unambiguously and were not subjected to random removal. Residues corresponding to the A and G β-strands of the ZP-N domains were defined as active residues. The electrostatic potential ($E_{elec}$) was calculated using a shift function, while a switching function (between 6.5 and 8.5 Å) was used to define the van der Waals potential ($E_{vdw}$). HADDOCK score, being the weighted sum of intermolecular electrostatic ($E_{Elec}$), van der Waals ($E_{vdW}$), desolvation ($\Delta G_{solv}$) and ambiguous interaction restraint (AIR) energies with default weight factors of 0.2, 1.0, 1.0 and 0.1, respectively, was used to rank the generated poses.

## Results

Synthesized peptides corresponding to the homologous A and G β-strands (AG interface) of ZP2 (HZP2_A and HZP2_G), ZP3 (HZP3_A and HZP3_G) and ZP4 (HZP4_A and HZP4_G) were examined thoroughly and they were found to self-assemble forming fibril-containing gels, after an incubation period of 1–2 weeks. Negative staining indicated that all peptides self-assemble forming fibrous arrangements with the basic characteristics of amyloid fibrils, since they appear straight, long and unbranched with a tendency to coalesce laterally (Fig. 1). More specifically, the 'aggregation-prone' peptides HZP2_A and HZP3_G self-assemble into amyloid-like fibrils with a diameter of 30–50 Å that coalesce laterally forming striated ribbons of various thickness (Fig. 1A,D). The HZP2_G peptide forms amyloid-like fibrils with a diameter of 70–100 Å (Fig. 1B), whereas the HZP3_A,

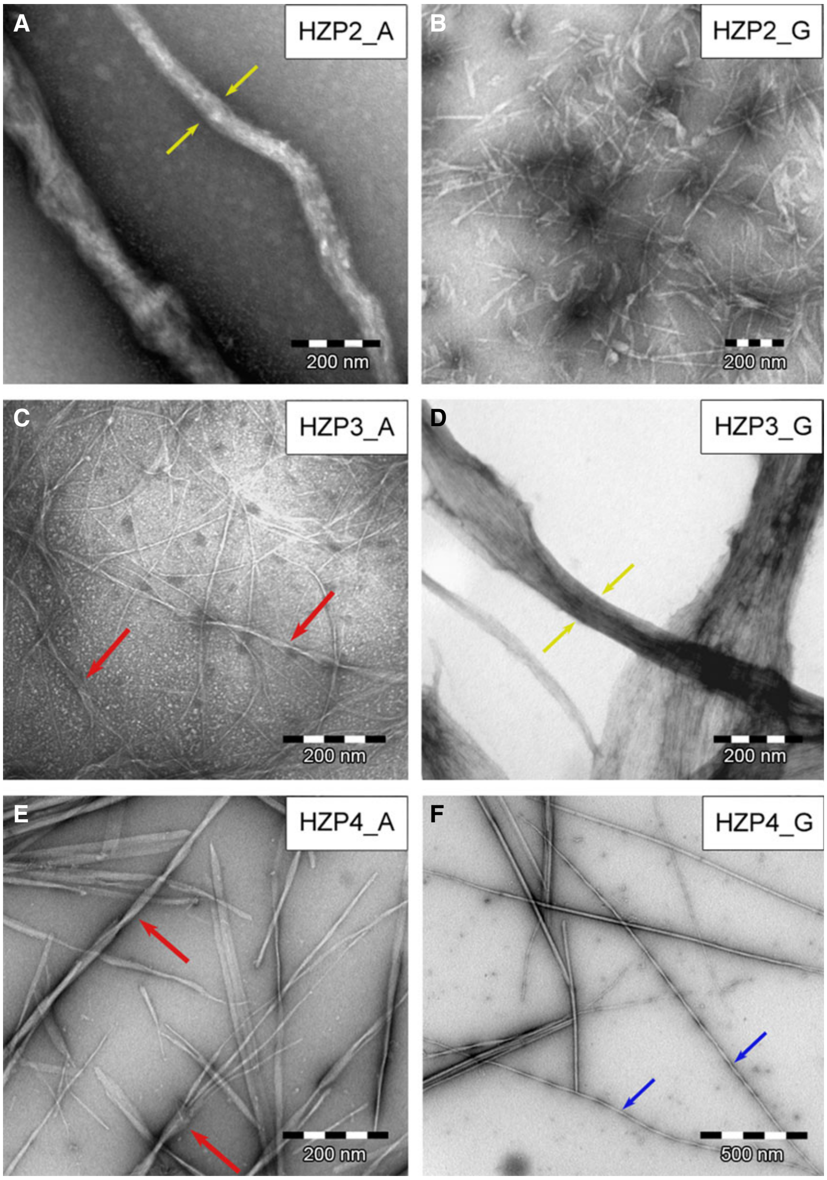


**Fig. 1.** Electron micrographs of amyloid-like fibrils formed by the (A) HZP2_A, (B) HZP2_G, (C) HZP3_A, (D) HZP3_G, (E) HZP4_A and (F) HZP4_G 'aggregation-prone' peptides. In all cases, the peptides self-assemble forming straight and unconnected amyloid-like fibrils. The fibrils exhibit evident polymorphism, since the HZP2_A and HZP3_G fibrils form striated ribbons of variable thickness (shown in yellow arrows), whereas the HZP3_A, HZP4_A and HZP4_G fibrils coalesce laterally forming superhelical arrangements composed of two (shown in blue arrows) to several individual fibrils (shown in red arrows).

HZP4_A and HZP4_G peptides form superhelical arrangements composed of two to several individual fibrils that wound around each other, with diameters that vary between 70 and 120 Å, respectively (Fig. 1C, E,F). Such an apparent morphological polymorphism of amyloid fibrils has been exhibited by many different amyloid-forming proteins or peptides [31,34,49,50].

X-ray diffraction experiments were performed on fibers containing amyloid-like fibrils, derived by self-assembly of each ‘aggregation-prone’ peptide. The X-ray diffraction patterns indicate that amyloid fibrils formed in every case possess the typical ‘cross-β’-like architecture of amyloid fibrils (Fig. 2). Specifically, a strong meridional reflection corresponding to a structural repeat of 4.6–4.7 Å is clearly evident along the axis of the fiber, resembling the distance between successively hydrogen bonded β-strands which are aligned in a perpendicular fashion to the fibril main axis. Moreover, all patterns exhibit a second strong reflection along the equator, corresponding to a d-spacing of 7.7–11.9 Å. This structural repeat most probably corresponds to the packing distance between β-sheets that are aligned parallel to the fibril axis. The differences observed in the equatorial reflections indicate

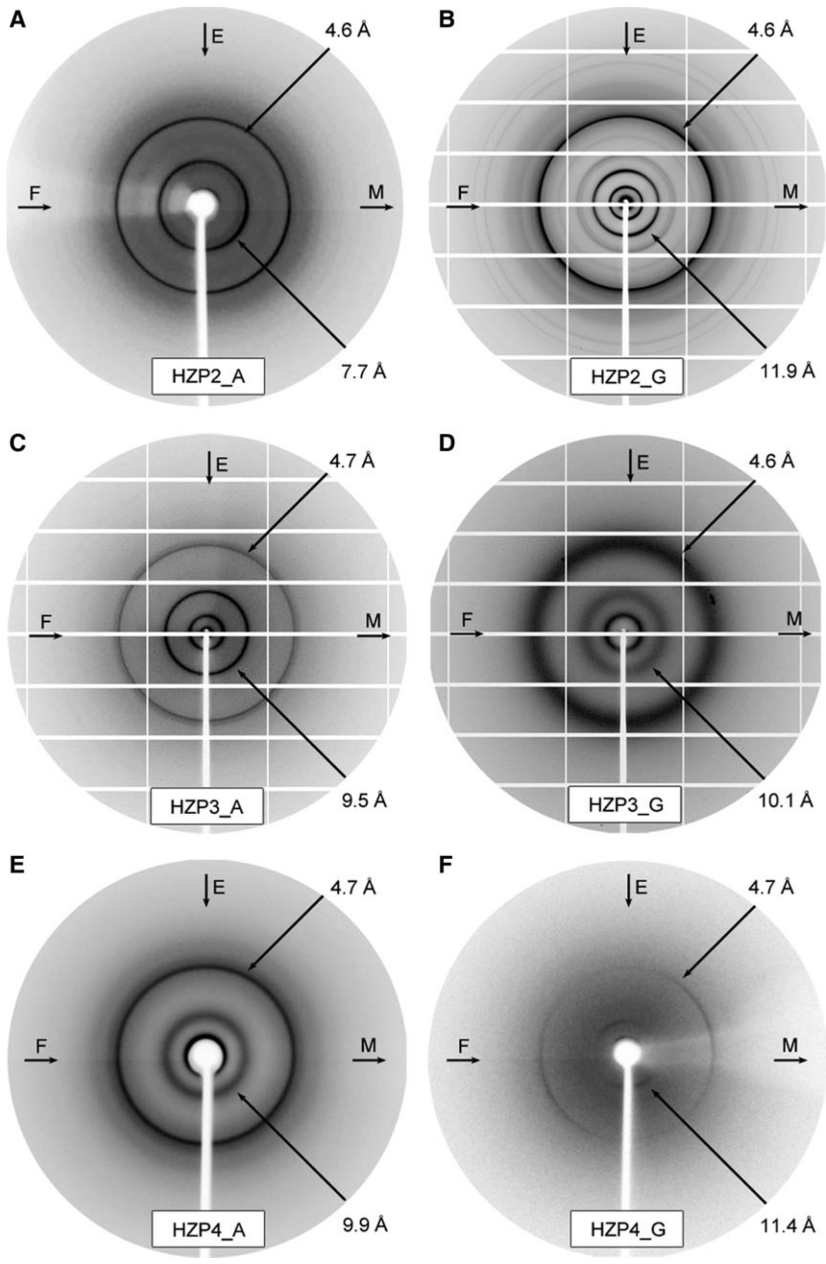


**Fig. 2.** X-ray diffraction patterns produced by oriented fibers containing (A) HZP2_A, (B) HZP2_G, (C) HZP3_A, (D) HZP3_G, (E) HZP4_A and (F) HZP4_G derived amyloid fibrils. The patterns are indicative of a ‘cross-β’ structure displaying both a meridional (M) 4.6–4.7 Å and an equatorial (E) 7.7–11.9 Å reflection, corresponding to the distance between consecutive β-strands and β-sheets that are aligned perpendicularly or along the fiber (F) axis, respectively.

dissimilarities in the packing distance of each peptide and arise from the variable sizes of the side chains that are interlocked in the steric-zippers formed between the β-sheets in each case. Model hexapeptides corresponding to the 'aggregation-prone' peptides were obtained by scanning ZipperDB [51]. The coordinates are derived after threading and energetic evaluation utilizing Rosetta-Design [52]. A detailed analysis of the derived models indicated that the accommodation of the side chains in a quarter-staggered fashion is sterically possible, in similar packing distances as the d-spacings observed. Peptide-analogs composed of residues with shorter side chains, such as HZP2_A or HZP4_A, have closer packing arrangements, in comparison to peptides with larger side chains, such as HZP2_G or HZP4_G. At certain cases, reflections may appear as rings due to poor alignment of the oriented fiber constituent fibrils.

In an effort to supplement our X-ray diffraction data, concomitant evidence was derived with the use

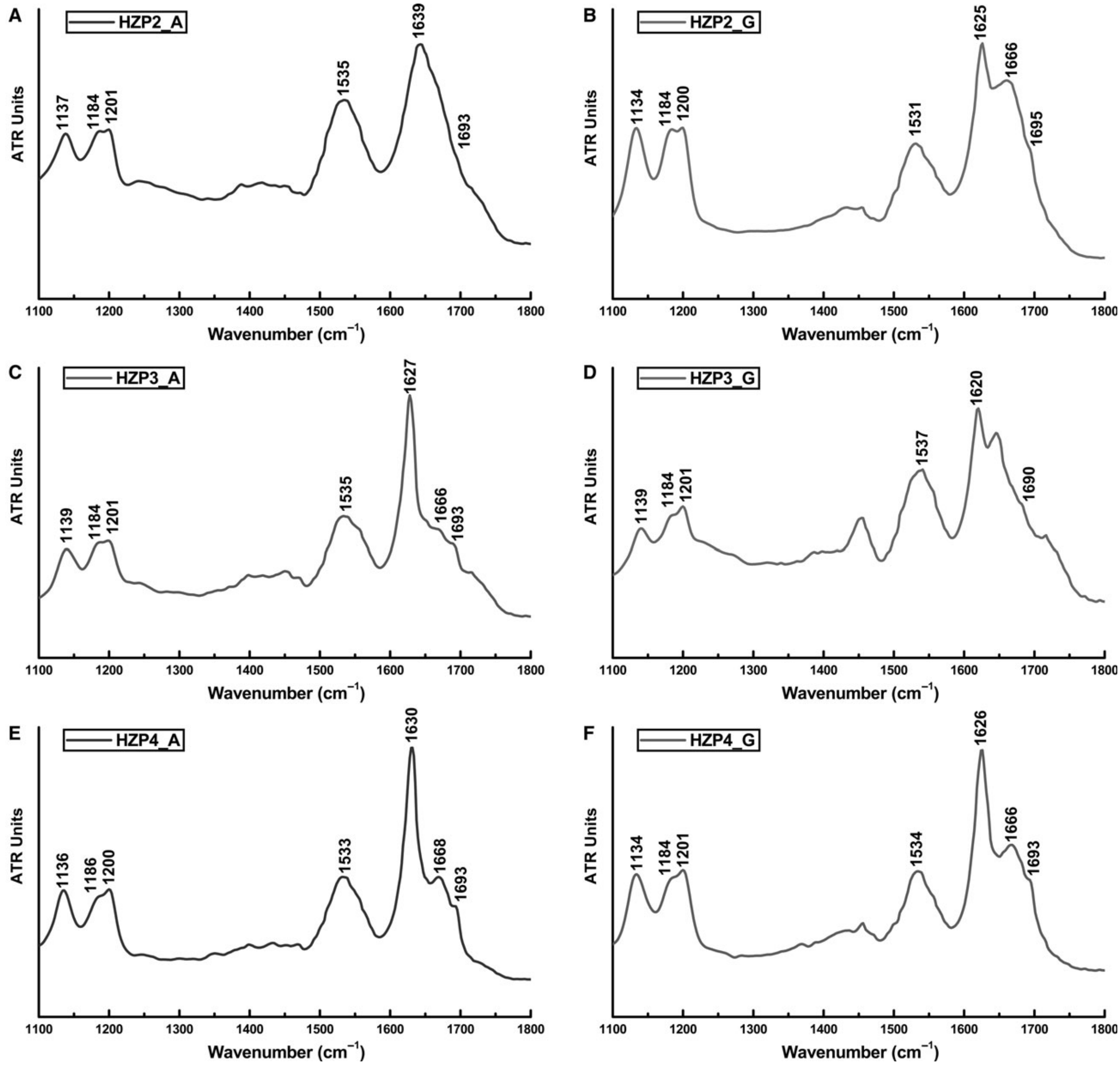


**Fig. 3.** ATR FT-IR spectra produced from thin hydrated films containing amyloid fibrils derived by the self-polymerizing (A) HZP2_A, (B) HZP2_G, (C) HZP3_A, (D) HZP3_G, (E) HZP4_A and (F) HZP4_G peptides. Fibrils derived by each amyloidogenic peptide possess an antiparallel β-sheet secondary structure, as it is clearly evident by the presence of strong amide I and II bands, in addition to a characteristic amide I shoulder within the range of 1690–1697 cm$^{-1}$, respectively (Table 1).

of ATR FT-IR spectral acquisitions. ATR FT-IR spectra obtained from thin hydrated films containing amyloid-like fibrils from the solution samples of the 'aggregation-prone' peptides indicate that they adopt an antiparallel β-sheet conformation, supporting our X-ray diffraction results (Fig. 3). More specifically, a preponderant amide I β-sheet band is viewed for the HZP2_A, HZP2_G, HZP3_A, HZP3_G, HZP4_A and HZP4_G peptides (Table 1), whereas similar assignments arise based on the presence of an amide II peak, viewed for spectra derived from all peptides, respectively. Finally, an additional shoulder determined by the second derivative spectra (Fig. S3), located in the 1690–1697 $cm^{-1}$ region in all cases, suggests that the peptides form antiparallel β-sheets (Table 1).

Deposits containing amyloid-like fibrils derived from the self-polymerizing HZP2_A, HZP2_G, HZP3_A, HZP3_G, HZP4_A and HZP4_G peptides were stained with the amyloid-specific Congo red dye, in order to validate the amyloidogenic properties of the peptides. Congo red has been shown to selectively bind on amyloid fibrils and produces a characteristic yellow/green birefringence, when viewed under crossed polars of a polarizing microscope [53,54]. As clearly seen under bright field illumination, fibril-containing gels of all the peptides selectively bind to the Congo red dye (Fig. 4). Additionally, the characteristic apple/green birefringence of amyloid deposits is clearly seen under crossed polars of a polarizing microscope (Fig. 4).

Three-dimensional models of the ZP-N domain for all ZP proteins were derived, namely hZP1, hZP2, hZP3 and hZP4, by performing homology modeling (Fig. S1). Structural analysis of the derived models elucidated apparent structural similarities for the ZP-N domain of all ZP proteins, since in all cases it adopts an Ig-like antiparallel β-sandwich fold, composed of two facing β-sheets, held together by two conserved intramolecular disulfide bonds, formed by four invariant Cys residues. The derived models were utilized as input for the prediction of possible protein–protein interaction sites, combined with conservation information derived by a multiple sequence alignment of ZP protein sequences (Fig. S2). Two distinct algorithms were used, namely the WHISCY web server [39], in addition to a metapredictor, meta-PPISP [40], which utilizes a linear regression method, built on three individual web servers, namely cons-PPISP [55], Promate [56], and PINUP [57]. Impressively, both algorithms support our experimental evidence, indicating that for all human ZP proteins, the majority of possible protein–protein interaction site hot-spots more or less coincide with the exposed AG interface surface (Fig. 5). Moreover, although other segments were also sparsely predicted for different ZP proteins, the only surface-exposed segment that is commonly predicted simultaneously by both algorithms for all ZP proteins as a possible interaction site is the AG interface, thus further supporting our experimental results. In this aspect, rounds of driven docking experiments were held. Particularly, following recent evidence, suggesting that ZP3–ZP2 and ZP3–ZP1/ZP4 dimers form the main axis of mammalian ZP filaments [8,58], we investigated whether these building blocks could be formed based on the aggregation propensity of the AG interface of ZP proteins. The docking results, introducing reasonably favorable energies with no violation restraints, yet weaker desolvation energies that do not contribute significantly to the favorable interaction (Table 2), suggest that protein–protein interactions taking place between the AG interfaces of ZP proteins could possibly promote the formation of dimers (Fig. 6). The docking computations are suggestive but not conclusive; therefore future experiments are essential, such as series of point mutations or FRET analysis of human full length ZP proteins, in order to experimentally verify their assembly mechanism and the implication of the identified 'aggregation-prone' segments in this process.

**Table 1.** Bands observed in the ATR FT-IR spectra produced from hydrated films of the 'aggregation-prone' peptides. As it is clearly evident by their tentative assignments, fibrils formed after self-assembly of the 'aggregation-prone' peptides adopt an antiparallel β-sheet secondary structure (Fig. 3).

| Bands ($cm^{-1}$) | | | | | | |
|---|---|---|---|---|---|---|
| HZP2_A | HZP2_G | HZP3_A | HZP3_G | HZP4_A | HZP4_G | Assignment |
| 1137 | 1134 | 1138 | 1139 | 1136 | 1134 | TFA |
| 1184 | 1184 | 1184 | 1184 | 1186 | 1184 | TFA |
| 1201 | 1200 | 1201 | 1201 | 1200 | 1201 | TFA |
| 1535 | 1531 | 1535 | 1537 | 1533 | 1534 | β-sheet (Amide II) |
| 1639 | 1625 | 1627 | 1620 | 1630 | 1626 | β-sheet (Amide I) |
| 1668 | 1666 | 1666 | 1664 | 1668 | 1666 | TFA |
| 1693 | 1695 | 1693 | 1690 | 1693 | 1693 | Antiparallel β-sheet |

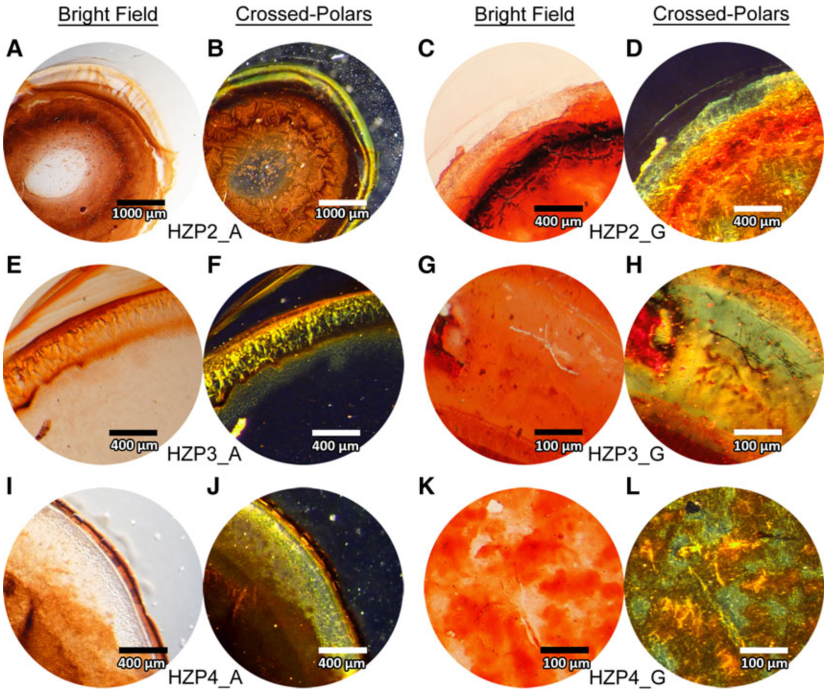


**Fig. 4.** Congo red staining of (A, B) HZP2_A, (C, D) HZP2_G, (E, F) HZP3_A, (G, H) HZP3_G, (I, J) HZP4_A and (K, L) HZP4_G deposits. (A, C, E, G, I, K) All peptide gels selectively bind the Congo red dye, as seen under bright field illumination. (B, D, F, H, J, L) Furthermore, all peptide gels exhibit the amyloid-specific apple/green birefringence under crossed polars of a polarizing microscope.

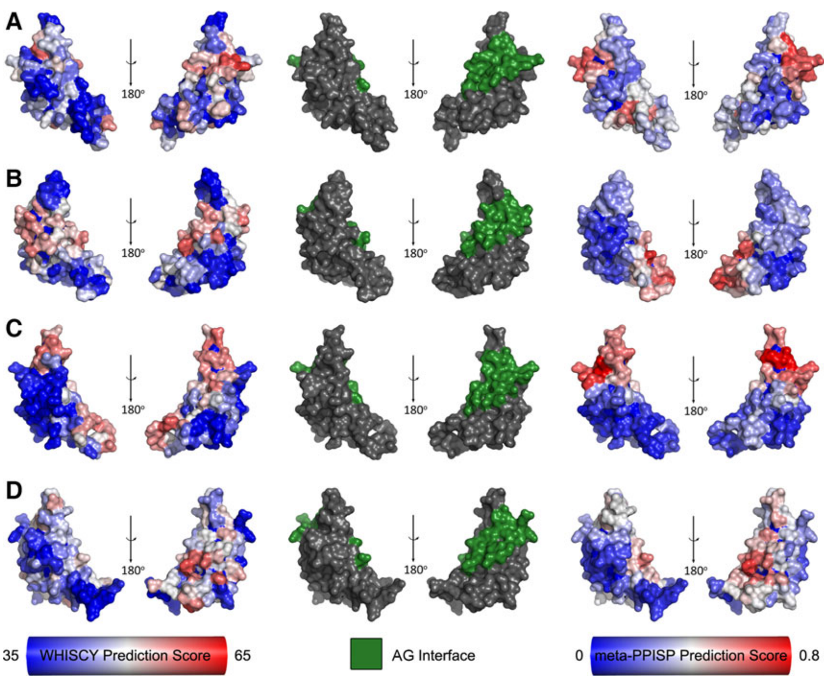


**Fig. 5.** Protein–protein interface prediction by WHISCY [39] and metaPPISP [40] for the ZP-N domain of (A) hZP1, (B) hZP2, (C) hZP3 and (D) hZP4. Surface representation of the ZP-N domains of hZP1, hZP2, hZP3 and hZP4 are shown in gray, with the corresponding AG interfaces highlighted in green. The corresponding WHISCY predictions are also shown in surface representations on the left side, whereas the meta-PPISP results are shown on the right. At both cases, scoring is highlighted with a blue to red gradient, from lower to higher score (scale bars below indicate minimum/maximum score for WHISCY and meta-PPISP, respectively). The majority of interface hot-spots (red-colored surfaces) predicted both by WHISCY and meta-PPISP more or less overlap to the corresponding AG interface surfaces of the ZP-N domains, with the single exception of the meta-PPISP results for the hZP2 ZP-N domain.

## Discussion

Our experimental results show that the peptide-analogs of the A and G β-strands of the ZP-N domain from hZP2, hZP3 and hZP4 proteins, in a similar manner to hZP1, clearly possess distinct amyloidogenic properties exhibiting the typical structural and tinctorial characteristics of amyloids [59,60].

Previous extensive studies have been focused on the ZP-N domain, which is considered responsible for ZP protein polymerization [18,19]. Detailed analysis of the solved crystal structure of the ZP-N domain and the complete ZP domain of mouse and chicken ZP3 protein, respectively, revealed that the AG interface may be a prominent 'aggregation-prone' interface for ZP proteins [19,20,58]. Similar evidence was also derived in the case

**Table 2.** Results of driven docking experiments. Energies of the most reliable results for all ZP dimers according to HADDOCK are shown (Fig. 5).

| Docked Dimers | HADDOCK score | Van der Waals Energy (Kcal·mol$^{-1}$) | Electrostatic Energy (Kcal·mol$^{-1}$) | Desolvation Energy (Kcal·mol$^{-1}$) | Restraints Violation Energy (Kcal·mol$^{-1}$) | Total Buried Surface Area (Å$^2$) |
|---|---|---|---|---|---|---|
| ZP1-ZP3 | −123.9 | −64.7 | −352.6 | 11.4 | 0.0 | 1768.5 |
| ZP2-ZP3 | −65.5 | −45.8 | −213.0 | 22.9 | 0.0 | 1554.1 |
| ZP4-ZP3 | −126.0 | −67.6 | −442.2 | 26.0 | 0.0 | 1975.3 |

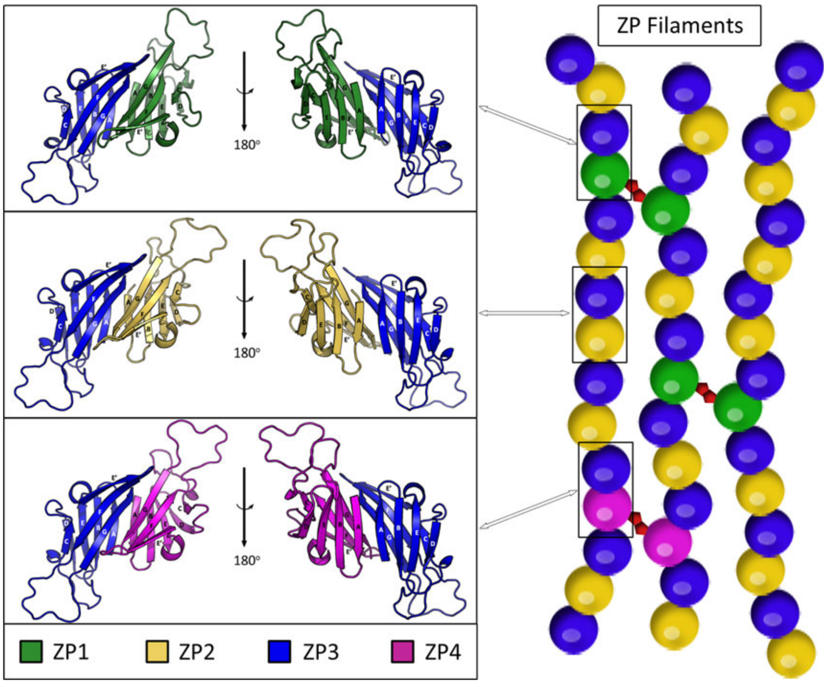


**Fig. 6.** Human ZP protein ZP-N domain dimers derived from driven docking, performed by HADDOCK. (A) ZP1-ZP3, (B) ZP2-ZP3 and (C) ZP4-ZP3 dimers are formed with similar interactions formulated between their corresponding AG interfaces (Table 2). (D) The dimers formed based on the aggregation properties of the AG interfaces, could possibly represent the successive building blocks brought together by head-to-tail interactions. Taking into account other predicted amyloidogenic segments [24], or segments that have been reported to take part in protein-protein interactions of ZP proteins [12,19,20], it is possible that the aggregation-propensity of the AG interface along with the above forces lead towards the formation of a perpetual β-helical structure along the axis of the ZP filaments. ZP filament interconnection may occur based on interactions between the N-terminal extensions or the trefoil domain of ZP1/ZP4 (shown in red pentagons), as recent studies have previously suggested [9,65]. ZP-N domain monomers from ZP1, ZP2, ZP3 and ZP4 are shown in green, yellow, blue and magenta, respectively.

of the ZP-N domain of hZP1 [33]. Our results corroborate this notion, by elucidating the amyloidogenic properties of peptide-analogs corresponding to the AG interfaces of all human ZP proteins. Based on this assumption, we sought to investigate whether this polymerization interface could contribute in the self-assembly mechanism of ZP proteins. Sequence conservation analysis revealed that the AG interface is a prominent interaction site for all human ZP proteins (Fig. 5). These results are strengthened by previous elaborate studies suggesting that amyloidogenic proteins with an Ig-like fold have a tendency to polymerize via their edged β-strands, forming dimers or higher order polymers [61]. This analysis elucidated that strategically placed bulky moieties, such as β-bulges, protect soluble monomers from polymerizing via complementary H-bonds between edged β-strands. Similar head-to-tail polymerization mechanisms have been suggested for other amyloid-forming proteins with an Ig-like fold, such as transthyretin or superoxide dismutase [62,63].

Detailed studies have shown that motifs composed of 6–8 residues, derived from β-continuous interfaces of oligomeric proteins have an intrinsic self-aggregation propensity by retaining the polymerizing properties of their parental protein interfaces [64]. In close analogy, since all the peptides of our study clearly have the ability to self-aggregate forming amyloid-like fibrous structures, it is possible that their parental AG interfaces are capable of driving ZP proteins towards the formation of dimers. This notion is supported by rounds of driven docking experiments suggesting that human ZP proteins could be able to form dimers by interacting via their AG interfaces.

Other amyloidogenic sites have been previously predicted in ZP proteins across species, utilizing evolutionary conservation analysis [24]. Possible interaction sites include, among others, the E′-F-G extension (composed of the E′, F and G β-strands of the ZP-N domain), which has been previously suggested to be involved in protein–protein interactions [20], or a conserved hydrophobic stretch, known as the internal hydrophobic patch (IHF) [12], a segment that is essential for the regulation of mouse ZP protein assembly [12,19]. In this line, the E′ β-strand of the ZP-N domain of mouse ZP3 was also found to serve as a crystal contact leading to the formation dimers [20]. Furthermore, ZP filament interconnection has been suggested to occur via the presence of the trefoil domains that diversify ZP1/ZP4 from the other ZP proteins, as it has been previously proposed [9,65]. As a result, it appears that the intrinsic aggregation propensity of the AG interface might contribute as a possible driving force for the formation of the fundamental dimeric building blocks composing ZP filaments, although interactions between sites other than the AG interface are most certainly essential for the overall ZP amyloid fibril formation. Identifying the self-assembly process of a functional amyloid, such as human zona pellucida, may have significant pharmacological applications by providing invaluable information towards the design of effective contraceptive drugs, by targeting regions important for the assembly or overall structure of mature ZP, thus constituting the oocyte infertile, or the production of potent biomaterials, by taking advantage of the impressive physicochemical properties of such structures.

## Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement N°283570). The infrastructure that the authors used for the present work is associated with the Greek National Research Infrastructure in Structural Biology, Instruct-EL. We thank the Institute of Biology, Medicinal Chemistry and Biotechnology at National Hellenic Research Foundation for hospitality. We also thank the University of Athens for support.

## Author contributions

NNL, SJH and VAI designed experiments; NNL performed the experiments; NNL, EDC, GEB, ESP, SJH and VAI analyzed data; EDC, GEB, ESP, SJH and VAI provided tool and reagents; NNL, SJH and VAI wrote the manuscript. All authors approved the final manuscript.

## References

1 Wassarman PM (1983) Fertilization. In Cell Interactions and Development (Yamada KM, ed), pp. 1–27. John Wiley and Sons, New York.

2 Wassarman PM (1999) Fertilization in animals. *Dev Genet* **25**, 83–86.

3 Bleil JD, Beall CF and Wassarman PM (1981) Mammalian sperm-egg interaction: fertilization of mouse eggs triggers modification of the major zona pellucida glycoprotein, ZP2. *Dev Biol* **86**, 189–197.

4 Dumont JN and Brummett AR (1985) Egg envelopes in vertebrates. *Dev Biol* **1**, 235–288.

5 Litscher ES, Janssen WG, Darie CC and Wassarman PM (2008) Purified mouse egg zona pellucida glycoproteins polymerize into homomeric fibrils under non-denaturing conditions. *J Cell Physiol* **214**, 153–157.

6 Greve JM and Wassarman PM (1985) Mouse egg extracellular coat is a matrix of interconnected filaments possessing a structural repeat. *J Mol Biol* **181**, 253–264.

7 Sasanami T, Ohtsuki M, Ishiguro T, Matsushima K, Hiyama G, Kansaku N, Doi Y and Mori M (2006) Zona Pellucida Domain of ZPB1 controls specific binding of ZPB1 and ZPC in Japanese quail (*Coturnix japonica*). *Cells Tissues Organs* **183**, 41–52.

8 Rankin TL, O'Brien M, Lee E, Wigglesworth K, Eppig J and Dean J (2001) Defective zonae pellucidae in Zp2-null mice disrupt folliculogenesis, fertility and development. *Development* **128**, 1119–1126.

9 Monné M, Han L and Jovine L (2006) Tracking down the ZP domain: from the mammalian zona pellucida to the molluscan vitelline envelope. *Semin Reprod Med* **24**, 204–216.

10 Aagaard JE, Vacquier VD, MacCoss MJ and Swanson WJ (2009) ZP domain proteins in the abalone egg coat include a paralog of VERL under positive selection that binds lysin and 18-kDa sperm proteins. *Mol Biol Evol* **27**, 193–203.

11 Swanson WJ, Aagaard JE, Vacquier VD, Monne M, Sadat Al Hosseini H and Jovine L (2011) The molecular basis of sex: linking yeast to human. *Mol Biol Evol* **28**, 1963–1966.

12 Jovine L, Qi H, Williams Z, Litscher ES and Wassarman PM (2004) A duplicated motif controls assembly of zona pellucida domain proteins. *Proc Natl Acad Sci USA* **101**, 5922–5927.

13 Jovine L, Darie CC, Litscher ES and Wassarman PM (2005) Zona pellucida domain proteins. *Annu Rev Biochem* **74**, 83–114.

14 Litscher ES, Qi H and Wassarman PM (1999) Mouse zona pellucida glycoproteins mZP2 and mZP3 undergo carboxy-terminal proteolytic processing in growing oocytes. *Biochemistry* **38**, 12280–12287.

15 Boja ES, Hoodbhoy T, Fales HM and Dean J (2003) Structural characterization of native mouse zona pellucida proteins using mass spectrometry. *J Biol Chem* **278**, 34189–34202.

16 Bork P and Sander C (1992) A large domain common to sperm receptors (ZP2 and ZP3) and TGF-β type III receptor. *FEBS Lett* **300**, 237–240.

17 Bork P (1993) A trefoil domain in the major rabbit zona pellucida protein. *Protein Sci* **2**, 669–670.

18 Jovine L, Janssen WG, Litscher ES and Wassarman PM (2006) The PLAC1-homology region of the ZP domain is sufficient for protein polymerisation. *BMC Biochem* **7**, 11.

19 Han L, Monne M, Okumura H, Schwend T, Cherry AL, Flot D, Matsuda T and Jovine L (2010) Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. *Cell* **143**, 404–415.

20 Monné M, Han L, Schwend T, Burendahl S and Jovine L (2008) Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. *Nature* **456**, 653–657.

21 Iconomidou VA, Vriend G and Hamodrakas SJ (2000) Amyloids protect the silkmoth oocyte and embryo. *FEBS Lett* **479**, 141–145.

22 Iconomidou VA, Chryssikos GD, Gionis V, Vriend G, Hoenger A and Hamodrakas SJ (2001) Amyloid-like fibrils from an 18-residue peptide analogue of a part of the central domain of the B-family of silkmoth chorion proteins. *FEBS Lett* **499**, 268–273.

23 Podrabsky JE, Carpenter JF and Hand SC (2001) Survival of water stress in annual fish embryos: dehydration avoidance and egg envelope amyloid fibers. *Am J Physiol Regul Integr Comp Physiol* **280**, R123–R131.

24 Egge N, Muthusubramanian A and Cornwall GA (2015) Amyloid properties of the mouse egg zona pellucida. *PLoS ONE* **10**, e0129907.

25 Chiti F and Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333–366.

26 Hammer ND, Wang X, McGuffie BA and Chapman MR (2008) Amyloids: friend or foe? *J Alzheimers Dis* **13**, 407–419.

27 Iconomidou VA and Hamodrakas SJ (2008) Natural protective amyloids. *Curr Protein Pept Sci* **9**, 291–309.

28 Shewmaker F, McGlinchey RP and Wickner RB (2011) Structural insights into functional and pathological amyloid. *J Biol Chem* **286**, 16533–16540.

29 Lopez de la Paz M and Serrano L (2004) Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci USA* **101**, 87–92.

30 Teng PK and Eisenberg D (2009) Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng Des Sel* **22**, 531–536.

31 Louros NN, Tsiolaki PL, Griffin MD, Howlett GJ, Hamodrakas SJ and Iconomidou VA (2015) Chameleon ‘aggregation-prone’ segments of apoA-I: a model of amyloid fibrils formed in apoA-I amyloidosis. *Int J Biol Macromol* **79**, 711–718.

32 Louros NN, Tsiolaki PL, Zompra AA, Pappa EV, Magafa V, Pairas G, Cordopatis P, Cheimonidou C, Trougakos IP, Iconomidou VA *et al*. (2015) Structural studies and cytotoxicity assays of “aggregation-prone” IAPP8-16 and its non-amyloidogenic variants suggest its important role in fibrillogenesis and cytotoxicity of human amylin. *Biopolymers* **104**, 196–205.

33 Louros NN, Iconomidou VA, Giannelou P and Hamodrakas SJ (2013) Structural analysis of peptide-analogues of human Zona Pellucida ZP1 protein with amyloidogenic properties: insights into mammalian Zona Pellucida formation. *PLoS ONE* **8**, e73258.

34 Louros NN, Petronikolou N, Karamanos T, Cordopatis P, Iconomidou VA and Hamodrakas SJ (2014) Structural studies of “aggregation-prone” peptide-analogues of teleostean egg chorion ZPB proteins. *Biopolymers* **102**, 427–436.

35 Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M *et al*. (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**, D115–D119.

36 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J *et al*. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

37 Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U and Sali A (2006). Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit 5.6, 1–31.

38 Sali A and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779–815.

39 de Vries SJ, van Dijk AD and Bonvin AM (2006) WHISCY: what information does surface conservation yield? Application to data-driven docking. *Proteins* **63**, 479–489.

40 Qin S and Zhou HX (2007) meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics* **23**, 3386–3387.

41 Oxford Diffraction (2009) Chrysalis Promotions. Oxford Diffraction Ltd., Abingdon, Oxfordshire, England.

42 Leslie AGW and Powell HR (2007) Processing diffraction data with Mosflm. In Evolving Methods for Macromolecular Crystallography (Read R and Sussman JL, eds), pp. 41–51. Springer, Dordrecht, The Netherlands.
43 Romhanyi G (1971) Selective differentiation between amyloid and connective tissue structures based on the collagen specific topo-optical staining reaction with congo red. *Virchows Arch A Pathol Pathol Anat* **354**, 209–222.
44 van Zundert GC, Rodrigues JP, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond AS, van Dijk M, de Vries SJ and Bonvin AM (2015) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol*, doi: 10.1016/j.jmb.2015.09.014.
45 Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905–921.
46 Jorgensen WL and Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* **110**, 1657–1666.
47 van Dijk AD and Bonvin AM (2006) Solvated docking: introducing water into the modelling of biomolecular complexes. *Bioinformatics* **22**, 2340–2347.
48 Kastritis PL, van Dijk AD and Bonvin AM (2012) Explicit treatment of water molecules in data-driven protein-protein docking: the solvated HADDOCKing approach. *Methods Mol Biol* **819**, 355–374.
49 Fandrich M (2007) On the structural definition of amyloid fibrils and other polypeptide aggregates. *Cell Mol Life Sci* **64**, 2066–2078.
50 Kodali R and Wetzel R (2007) Polymorphism in the intermediates and products of amyloid assembly. *Curr Opin Struct Biol* **17**, 48–57.
51 Goldschmidt L, Teng PK, Riek R and Eisenberg D (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci USA* **107**, 3487–3492.
52 Kuhlman B and Baker D (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci USA* **97**, 10383–10388.
53 Benhold H (1922) Eine spezifische Amyloidfarbung mit Kongorot. *Munch Med Wochenschr* **69**, 1537–1538.
54 Divry D and Florkin M (1927) Sur les proprietes optiques de l'amyloide. *C R Soc Biol* **97**, 1808–1810.
55 Chen H and Zhou HX (2005) Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. *Proteins* **61**, 21–35.
56 Neuvirth H, Raz R and Schreiber G (2004) ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol* **338**, 181–199.
57 Liang S, Zhang C, Liu S and Zhou Y (2006) Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res* **34**, 3698–3707.
58 Monne M and Jovine L (2011) A structural view of egg coat architecture and function in fertilization. *Biol Reprod* **85**, 661–669.
59 Sunde M and Blake C (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem* **50**, 123–159.
60 Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB and Blake CC (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* **273**, 729–739.
61 Richardson JS and Richardson DC (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci USA* **99**, 2754–2759.
62 Correia BE, Loureiro-Ferreira N, Rodrigues JR and Brito RM (2006) A structural model of an amyloid protofilament of transthyretin. *Protein Sci* **15**, 28–32.
63 Elam JS, Taylor AB, Strange R, Antonyuk S, Doucette PA, Rodriguez JA, Hasnain SS, Hayward LJ, Valentine JS, Yeates TO *et al.* (2003) Amyloid-like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS. *Nat Struct Biol* **10**, 461–467.
64 Valery C, Pandey R and Gerrard JA (2013) Protein beta-interfaces as a generic source of native peptide tectons. *Chem Commun* **49**, 2825–2827.
65 Braun BC, Ringleb J, Waurich R, Viertel D and Jewgenow K (2009) Functional role of feline zona pellucida protein 4 trefoil domain: a sperm receptor or structural component of the domestic cat zona pellucida? *Reprod Domest Anim* **44** (Suppl 2), 234–238.

## Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

**Fig. S1.** Comparative modeling of the ZP-N domain from human (A) ZP1, (B) ZP2, (C) ZP3 and (D) ZP4.

**Fig. S2.** Multiple sequence alignment of the corresponding ZP-N domain of ZP proteins.

**Fig. S3.** Second derivative ATR FT-IR spectra (1500–1800 $cm^{-1}$) of the (A) HZP2_A, (B) HZP2_G, (C) HZP3_A, (D) HZP3_G, (E) HZP4_A and (F) HZP4_G 'aggregation-prone' peptides.