Evolutionary Origins of Amyloidogenicity - Reconstruction and Analysis of Ancestral IAPP Sequences

Ivan Filippov[1], Aria Baghai[1], Alexander Zhyvoloup[1], Lauren Chisholm[2], Michael J. Harms[2], Daniel P. Raleigh[1,3]

[1]Institute of Structural and Molecular Biology, University College London; [2]Institute of Molecular Biology, University of Oregon; [3]Department of Chemistry, Stony Brook University

Abstract

Islet amyloid polypeptide (IAPP) is an amyloidogenic polypeptide contributing to type 2 diabetes. Human IAPP is one of the most amyloidogenic peptides known, but multiple naturally occurring IAPP variants have been shown to be non-amyloidogenic. Furthermore, a closely related paralogue - calcitonin gene related peptide (CGRP), which shares many sequence features with IAPP and is believed to have evolved from a common ancestor, has not been observed to form amyloid fibrils. The evolutionary origin of amyloidogenicity of IAPP is unknown. To gain insight into trends of biophysical qualities over evolutionary time ancestral sequence reconstruction is often used for inferring extinct precursor sequences. Due to the short length and extremely high conservation of IAPP, ordinary ASR methods do not produce useful results. Here we gather an alignment of 308 IAPP sequences and perform ASR of extinct IAPP sequences using a bootstrapping method, reporting high-confidence, unique sequences. Using computational prediction methods and Thioflavin-T assays we show that ancestral IAPP sequences are less amyloidogenic than human IAPP, and amyloidogenicity is likely to decrease with evolutionary time. Our results support the hypothesis that amyloidogenicity was not selected against during evolution and further emphasise that human IAPP is not optimised for solubility. We describe issues related to gathering CGRP sequences for a comparative study and propose a possible solution.

Islet amyloid polypeptide (IAPP), also known as amylin, is a 37-residue signalling hormone produced in pancreatic islet β-cells and co-secreted with insulin (1). Its physiological function is primarily concerned with glucose homeostasis. It is considered a neuropeptide due to an additional function in the brain regulating hunger response (2). Human IAPP (hIAPP) is a known causative agent in type 2 diabetes (T2D), and current research attributes this to its propensity for aggregation into amyloid fibrils (3). *In vitro* amylin is among the most amyloidogenic peptides known to date. Considering that amylin is present in almost all vertebrate taxa, it is of interest that the majority of known variants are highly amyloidogenic, and seemingly no evolutionary pressure has been applied toward increased solubility. This question extends beyond just IAPP, as there are over 40 known amyloidogenic peptides involved in human diseases. IAPP presents an excellent model system for this investigation for a number of reasons: it has a well-defined cellular function, unlike many of the amyloidogenic peptides studied in clinical science; it forms aggregates from its native state; it is believed to have evolved from a common ancestor with other members of the calcitonin/calcitonin gene-related peptide (CT/CGRP) peptide family, which provides a frame of reference for comparative studies (4).

```
              1          10          20          30         O
                                                            ||
hIAPP    H₃N-KCNTATCATQ RLANFLVHSS NNFGAILSST NVGSNTY-C-NH₂
                                                            O
                                                            ||
rIAPP    H₃N-KCNTATCATQ RLANFLVRSS NNLGPVLPPT NVGSNTY-C-NH₂
                                                            O
                                                            ||
hCGRP1   H₃N-ACDTATCVTH RLAGLLSRSG GVVKNNFVPT NVGSKAF-C-NH₂
                                                            O
                                                            ||
hCGRP1   H₃N-ACNTATCVTH RLAGLLSRSG GMVKSNFVPT NVGSKAF-C-NH₂
```
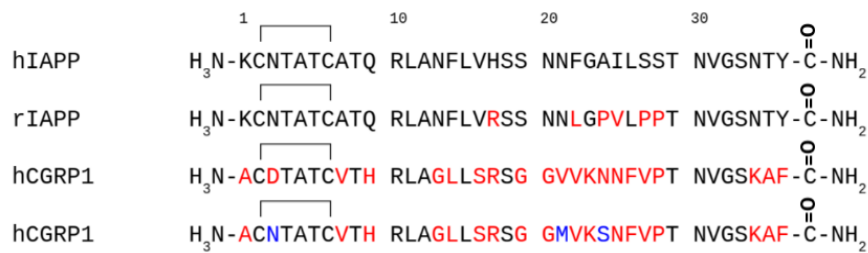
Figure 1. Comparison of hIAPP, rIAPP hCGRP1 and hCGRP2 sequences. Differences from hIAPP are shown in red, differences between hCGRP1 and hCGRP2 are shown in blue. All four polypeptides contain a disulfide bond between residues 2 and 7 and an amidated C-terminus.

CT/CGRP is a family of peptide hormones produced in various locations and with a diverse functionality. It is believed to have originated from a common ancestor, initially forming adrenomedullin and amylin during a duplication event, and further amylin duplications resulting in calcitonin, CGRP and calcitonin receptor stimulating peptide (CRSP) (5). Of these, CGRP presents a particularly interesting opportunity for investigating the evolution of amyloidogenicity. IAPP and CGRP, have high sequence identity and still share many features - as seen in Figure 1, both are 37 residues long, both retain the amidated C-terminus and the N-terminal ring formed by the disulphide bridge between cysteines 2 and 7. IAPP itself remains highly conserved across species in its mature sequence, although some of the naturally occurring variants exhibit almost no amyloidogenic or toxic properties. Mature IAPP and CGRP are both produced by proteolysis of a longer prosequence at highly conserved sites (6, 7).

However, no evidence of CGRP aggregation *in vivo* has been found, and limited studies show much lower aggregation rate *in vitro* (8). The IAPP/CGRP system therefore offers a valuable opportunity to understand how amyloidogenicity either arose during evolution or was selected against in one but not the other peptide. To investigate this, we used a combination of computational and bioinformatics methods to reconstruct the most likely ancestral IAPP sequences and predict their biophysical properties, as well as experimental work with ancestral candidates.

IAPP has been shown to form cytotoxic oligomeric species, with at least two distinct toxicity mechanisms: receptor-mediated, through receptor for advanced glycation products (RAGE); membrane disruption, through yet unknown biophysical mechanisms (9, 10). While IAPP fibrils do not appear to be directly cytotoxic, their build-up leads to pancreatic islet amyloidosis. CGRP is a neuropeptide, best known for its role in the pain phase of a migraine attack. It is involved in multiple pain signalling pathways, and additionally possesses cardioprotective properties. Both peptides have distinct receptors, but show cross-reactivity (11, 12).

Amyloid formation by various peptides has been observed in over 40 human pathologies and has been a longstanding focus of biomedical research, most notably in connection with neurodegenerative diseases such as Alzheimer's. *In vitro* amyloid formation by every peptide studied so far follows a similar kinetic profile, shown in Figure 2 (13). Initially, monomers form oligomeric intermediates, which are often implicated as the primary source of toxicity - this is called the lag phase. Once the concentration of oligomers is sufficient, they begin to assemble into β-sheet fibrils, which continue to elongate during the growth phase, until all of the oligomers have assembled into fibrils and the growth terminates in the saturation phase. Oligomers formed during the lag phase have been implicated as the main toxic

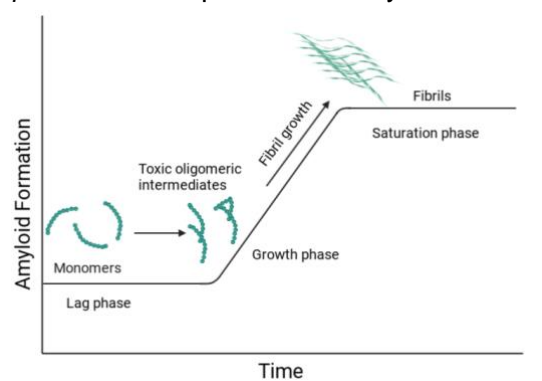species in multiple pathologies, including β-cell death in pancreatic amyloidosis.

Figure 2. A cartoon representation of amyloid formation kinetics.

No positive physiological function has been shown for IAPP amyloidogenicity, nor is it known at which point in the evolution this property has emerged. Among extant species, some show no aggregation *in vivo*, and the concentration required to observe fibrils *in vitro* exceeds any reasonable physiological values. Most notable example of this is the rat IAPP (rIAPP), which differs from hIAPP by only 5 residues in the 20-29 segment, believed to be responsible for a β-sheet conformation and currently the only consistent sequence determinant of non-amyloidogenic peptides (14, 15). In particular, prolines at positions 25 28 and 29, so-called 'rat prolines', which are utilised in a soluble medicinal hIAPP analogue Pramlintide, as together they practically negate aggregation *in vivo* (16).

Ancestral sequence reconstruction (ASR) is a vertical statistical phylogenetic method using maximum likelihood or Bayesian approaches to determine the most likely precursors for each node of a phylogenetic tree (17). To conduct a reconstruction, first a multiple sequence alignment (MSA) is constructed, using homologues of the peptide of interest. The alignment is then used to construct a phylogenetic tree, which can further be reconciled with a species tree if necessary. ASR algorithms usually generate multiple candidate sequences per node, with associated confidence values. Some of the algorithms provide a separate value for each possible residue

at a given position. We use RAxML (Randomised Accelerated Maximum Likelihood), which utilises a parallel maximum likelihood approach for constructing phylogenetic trees and inferring node sequences (18). Ordinarily, ASR is performed on long, highly variable sequences, such as viral proteins, for which extremely large (1,000s-10,000s sequences) MSAs can be built.

IAPP presents challenges to this approach - the peptide is extremely short and only present in vertebrate species, meaning only a relatively small MSA can be assembled. These conditions routinely result in extremely low confidence values, and successful reconstruction was not possible for IAPP until now. This is an important cautionary note for any future work on similar problems, i.e. families of relatively short sequences, which exhibit relatively high conservation. Part of the work presented was to develop an approach that overcomes this issue, done in collaboration with the group of Prof Michael Harms at University of Oregon.

To perform the ASR, we use a bootstrapping method, where prior to the reconstruction, the tree created from the MSA is evaluated against a species tree, reconciled and provided with support values. The ancestors are reconstructed on the tree with supports. The goal of the research described here is to apply this technique to IAPP and CGRP to develop information on early ancestors to investigate trends of biophysical properties over evolutionary time.

Results

We generated an alignment of IAPP sequences using NCBI's basic local alignment search tool (BLAST), initially using mature hIAPP as a query sequence. We used the IAPPs identified during the search as query sequences to collect more distant homologues. Sequences were verified based on three criteria: the annotations; manually checking for processing sites (cleavage signals); and C-terminal glycine at position 33 required

for amidation. Mature IAPP sequences were isolated from the prosequences by removing the flanking regions around cleavage sites, and the resulting mature sequences were aligned using MUSCLE {Edgar 2004}. The alignment consisted of 308 vertebrate species. Analysis of the alignment confirmed an extremely high level of sequence conservation, shown in Figure 3. Compared to hIAPP, the most diverse sequence *Rhinatrema bivittatum* only has 15 substitutions, and *Callorhinchus milii*, so-called "ghost shark", considered to be one of the most slowly evolving organisms, only has 10 substitutions compared to hIAPP, while latimeria *Petromyzon marinus*, used to root the tree, has 19 substitutions and two deletions. Conservation is highest in the 1-12 and 30-34 regions, especially conserved are the cysteines at positions 2 and 7 (forming the disulfide bridge) with no substitutions across the entire alignment, and the C-terminal tyrosine, with only 2 histidine substitutions. Most variable is the 22-29 region, which is described as having significant effects on both amyloidogenicity and toxicity, with the exception of glycine at position 24. This region also includes the FGAIL region in hIAPP, which has been studied as part of the transient β-sheet site during the lag phase oligomer formation (19, 20). We also observed a high number of duplicate sequences, usually within a single clade. For ASR, the duplicate sequences were filtered out using CD-Hit, in order to compensate for overrepresentation in those taxa and reduce the bias caused by identical sequences within a clade. Clades with the highest number of duplicate sequences were birds and rodents.

Resulting alignment of 143 IAPP sequences was used to construct a phylogenetic tree with RaxML, with Lataemaria being used to root the tree. Considering the aforementioned issues with sequence and alignment length, the tree was reconciled using GeneRax (21). The resulting tree can be seen in Figure 4a (full list of sequences and phylogenetic tree can be found in appendices 1 and 2). Reconstruction yielded multiple unique sequences. Of the resulting ancestral sequences, we focused on 3 key nodes: all-mammal ancestor (mIAPP), birds and reptiles ancestor (bIAPP) and the all-species ancestor (aIAPP) (Figure 4b). At each of these nodes there was one sequence which had very high confidence (>0.99) and was the only sequence with significant probability. As expected, the number of substitutions from hIAPP increases from mIAPP to bIAPP to aIAPP, consistent across all ancestors: S29P, one of the 'rat prolines', which is present in 120 out of the 143 unique sequences used for the reconstruction; and H18R, present in 122 sequences. Of all the proline substitutions in rIAPP, S29P is the one with the weakest effect on amyloidogenicity (22). H18R is also present in rIAPP, and has been shown to slow aggregation as a point mutation to hIAPP (23). It is also present in hCGRP1 and hCGRP2. A recent study investigating position 18 shows it to be crucial for IAPP structure during membrane interactions, highlighting their role in slowing fibril formation via stabilisation of the oligomeric species (24). Both bIAPP and aIAPP also have A8V, which is found in both hCGRPs.

Figure 3. Sequence logo of IAPP alignment including 308 vertebrate species.
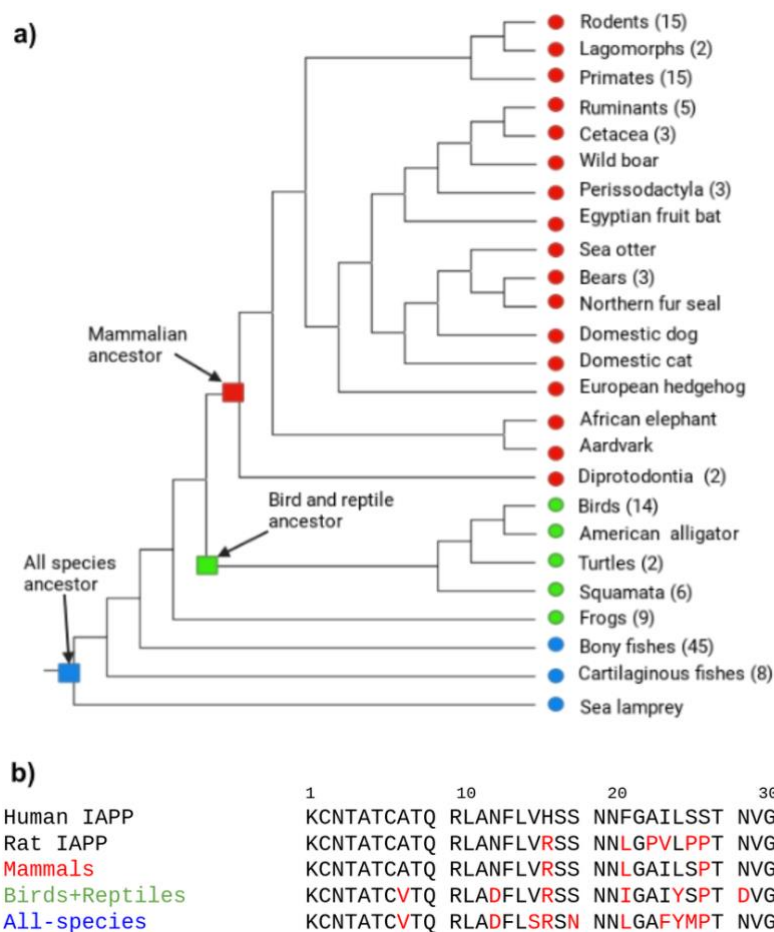
Figure 4. a) Condensed phylogenetic tree constructed from 143 IAPP sequences used for ancestral reconstruction. Branches are labelled with species or clade names, clades include number of species in brackets. Branches are colour coded: red for mammals, green for birds and reptiles, blue for fishes. Ancestral nodes of interest are labelled with squares. b) Alignment showing IAPP sequences of human, rat, mammalian ancestor, birds and reptiles ancestor and all-species ancestor. Residues different from hIAPP are shown in red.

Coincidentally, the mammalian ancestor is identical to the previously described mammalian consensus sequence of IAPP in (25). Experimental results from that work show that mIAPP has a lower rate of amyloid formation as well as lower toxicity than hIAPP. F23L has been shown previously to slow aggregation, and unpublished results from our laboratory have further shown that the F23L substitution reduces, but does not eliminate amyloid formation and reduces toxicity of hIAPP (26). In the consensus sequence study, H18R hIAPP (primate consensus sequence) has been shown to have an aggregation rate and toxicity lower than hIAPP, but higher than mIAPP.

Together the results imply that mIAPP exhibits a cumulative effect of H18R, F23L and possibly S29P, although the point mutation itself does not affect amyloid formation.

As an intermediate step for analysing the biophysical qualities of the ancestral sequences in absence of experimental data, we used computational tools for predicting amyloidogenic properties from sequence. We first validated each method by comparing the results produced for hIAPP, rIAPP and mIAPP, to confirm that the algorithm correctly predicts the relatively higher amyloidogenicity of hIAPP. We did not use any homology-

based algorithms. Each tool was used to assess 5 sequences: hIAPP, rIAPP (used as non-amyloidogenic control), mIAPP, bIAPP and aIAPP.

ANuPP provides a residue-wise aggregation probability score, but due to each residue being analysed in conjunction with the neighbouring ones, the terminal residues have lower scores, which is emphasised by the authors of the program (27). As seen from the results, all but aIAPP are predicted to have aggregation-prone regions at residues 11-21, with an additional peak at 22-28. Of all peptides, hIAPP has highest aggregation probabilities at both peaks, with the highest overall result (and the only one predicted to form amyloid). bIAPP, mIAPP and rIAPP show similar aggregation profiles, with a reduced second aggregation hotspot around the FGAIL region, and have similar overall scores, although rIAPP is shown as least likely to aggregate. aIAPP, indicated as having the lowest aggregation probability, shows a different profile, with the first peak dramatically reduced, possibly due to the V17S and S20N substitutions, which are not present in mIAPP or bIAPP.
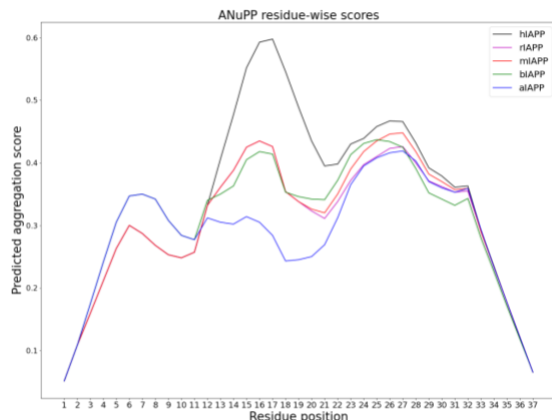

Figure 5. Plots showing residue-wise aggregation probability scores predicted by ANuPP for the 5 IAPP variants studied. X-axis shows predicted aggregation score, Y-axis shows residue position.

ZipperDB uses a sliding hexapeptide window, where each hexapeptide is predicted a Rosetta energy (28, 29). As the algorithm cannot process proline-

containing hexapeptides, the 24-29 region is not possible to assess in all peptides except for hIAPP, in rIAPP the region is 20-29 - this is represented as gaps in the plot. All other gaps represent real results, where the Rosetta energy is below -15kcal/mol. Residues 30-37 are predicted to have identical energies for all of the sequences. Hexapeptides spanning residues 1-10 are consistently scored above the threshold, which is not reflected by ANuPP results, due to its issues with terminal residue processing. Residues 13-20, which have also been indicated as an aggregation hotspot by ANuPP, differ radically between the sequences, with hIAPP scoring as most amylogenic, followed by bIAPP, mIAPP and rIAPP. It is unclear whether bIAPP or mIAPP is more amyloidogenic from the results, but hIAPP clearly scores highest across all regions. This is consistent with known results for rIAPP and mIAPP described earlier and suggests that the bird+reptile and all-species ancestors are also likely more soluble than hIAPP. Interestingly, aIAPP is predicted to have lower aggregation at the 13-20 region than rIAPP, which is non-amyloidogenic.

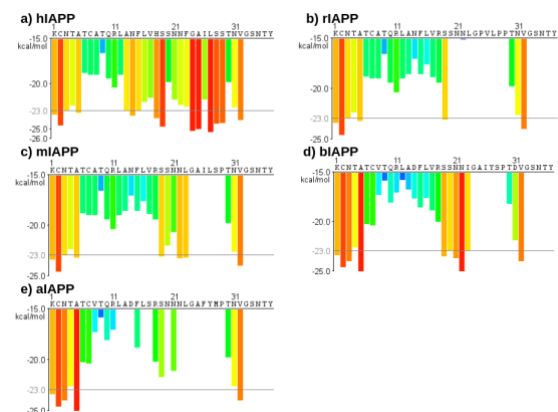*Figure 6. Plot showing results of ZipperDB residue-wise predictions of aggregation for each hexapeptide in the sequences. X-axis represent Rosetta energies, Y-axis represent starting position of the hexapeptide. Gaps between residues 23-29 (19-25 in rIAPP) are due to the presence of proline, which the algorithm cannot process; other gaps are a reflection of real results, meaning the Rosetta energy is below -15kcal/mol.*
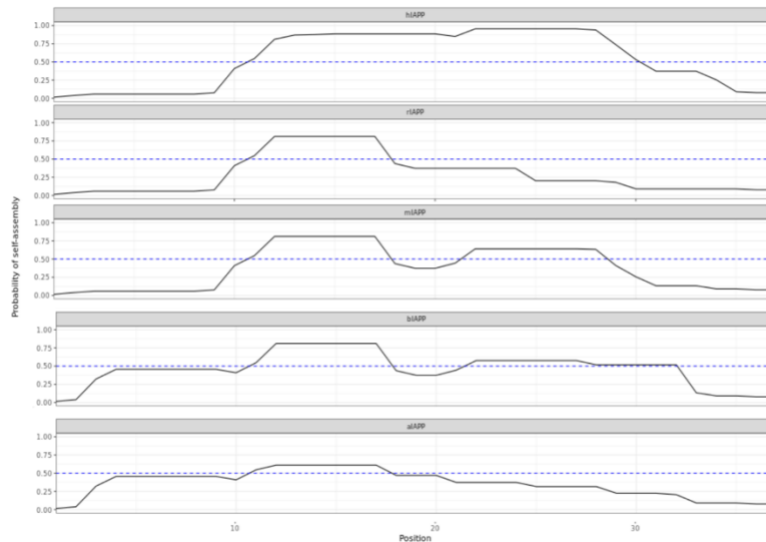
Figure 7. Plot of AmyloGram residue-wise predictions of self-assembly of each sequence. X-axis represents probability, Y-axis represents residue position. Dotted line is a 0.50 probability threshold.

Aylogram is designed to predict amyloidogenic hot spots in the sequence, displaying a probability of amyloidogenicity for each residue (30). Results for IAPP sequences show a similar profile to that of ANuPP, showing a high-probability region at residues 12-18 and 22-28. hIAPP is predicted to have >0.80 probability in the entire 12-28 region, again making it the most amyloidogenic of the sequences. Both mIAPP and bIAPP have a similar profile with two distinct amyloidogenic regions >0.50 probability, although bIAPP shows lower probability in the 12-28 region. rIAPP and aIAPP both lack the second region, likely due to proline substitutions in the case of rIAPP and multiple substitutions in the FGAIL region in aIAPP. 12-18 is also lowest for aIAPP, once again resulting in prediction of lowest mayloidogenicity out of all the sequences.

Experimental data was obtained for bIAPP amyloid formation rate using a Thioflavin-T (ThT) assay. ThT is a small fluorescent molecule known to bind to amyloid fibrils, and has been commonly used for amyloid kinetics studies of IAPP (31). Plotted results for both peptides at 30 μm concentration can be seen in Figure 8. Both peptides reached a plato during the course of the experiment, confirming that bIAPP does form amyloid, although the

rate of formation is ~3.5 times lower than that of hIAPP. hIAPP T_{50} = 3.3 hours; bIAPP T_{50}=11.5 hours. This finding is consistent with the computational predictions and known sequence determinants of lower amyloidogenicity present in bIAPP. Due to technical difficulties, it was not possible to directly compare the ancestral peptides and experimentally confirm the predicted differences between them.
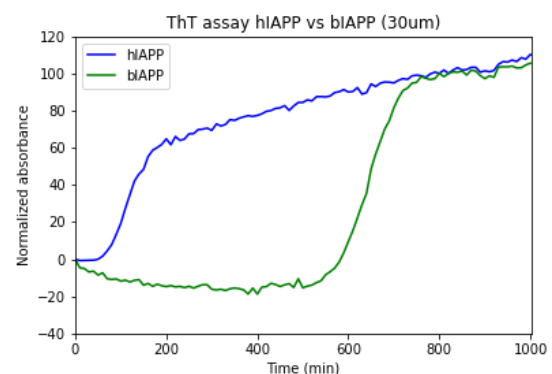

Figure 8. ThT assay plot, showing hIAPP in blue and bIAPP in green.

We attempted to collect a similar alignment of CGRP, but came across interesting technical problems caused by CGRP evolutionary history. Except for fishes, birds and marsupialia, CGRP exists in 2 forms - CGRP-1 (carried on the CT/CGRP gene) and a paralogue CGRP-

2, which differs slightly in mature form and does not carry a functional CT exon (32). Distinguishing between the two peptides in the mature form is extremely difficult: human CGRP-1 (hCGRP-1) and human CGRP-2 (hCGRP-2) only differ in 4 residues. While euarchontoglires mammals only have the two CGRPs, there is an additional form of CGRP paralogues - calcitonin receptor-stimulating peptide (CRSP) in laurasiatherian (33). CRSP further has multiple forms, CRSP-1 is 38 residues long, and therefore differentiable, but other CRSPs (up to 5 in a single species), are indistinguishable by mature sequence from CGRPs, and can only be identified with pharmacological experiments, as their target receptor is different to that of CGRP. As CRSP has only recently been discovered, many of the annotated CGRP sequences are in fact mislabelled CRSPs, and therefore are not fit to use in ASR, if the goal is determining a CGRP ancestor to be compared with IAPP. Another complication arises from CRSP gene structure, which sometimes includes a functional CT exon, and in some species the calcitonin is only carried on a CRSP gene and not on either of the two CGRP genes(32). Since the two mature sequences are extremely closely related, it is unknown whether these CT/CRSP genes are a direct lineage from CT/CGRP. Chromosomal localisation of the paralogues relative to each other is not consistent, so cannot be used to distinguish them either.

Fortunately, all of the genes exist within a well-defined locus between CYP2R1 and INSC1 (34, 35). To collect CGRP sequences, we isolate the locus and perform a local BLAST search against human CGRP-1 to find homologous sequences. We use NCBIs GenomeViewer to determine whether the sequence is part of a gene with a functional CT exon if the species has multiple CGRP-like sequences. Furthermore, we assess whether the gene has necessary promoter sequences, as it is known that primates possess a CGRP pseudogene (pCGRP), which does not get expressed (36). We then separate CGRP-

1 sequences to be used for ASR. As the process requires a high degree of manual validation, we have to date only been able to collect and verify 20 sequences (not included), which is insufficient for ASR. Since the clades which have distinguishable CGRP-1 are closer to the root of the tree, we believe the resulting ancestors will be representative for comparison with IAPP, although a further confirmation of IAPP sequences exclusively from the species used for CGRP reconstruction will provide higher confidence in the validity of the comparison.

Discussion

To the best of our knowledge, this is the first rigorously annotated alignment of IAPP, as well as the first successful ancestral reconstruction not just of a peptide from the CT/CGRP family, but any peptide of comparable length that yielded high-confidence ancestor sequence candidates. Although it was not possible to complete ASR for CGRP, we have developed an approach to solve this issue in further studies. Additionally, the ASR method used here provides an opportunity to use alignments that include paralogs, which can unite all IAPPs, CGRPs and CRSPs. Reconstructing the same ancestors using this method can further confirm the validity of the sequences we obtained. Performing this with only CGRPs and CRSPs can provide an ancestral IAPP node, since CGRP is believed to have come from an IAPP duplication event.

As we expected from the ancestral sequences, both bIAPP and mIAPP are less amyloidogenic than hIAPP. While we were not able to test aIAPP amyloid formation kinetics, based on the sequence features present in the other ancestral peptides we expect it to be less amyloidogenic. In particular, H18R and F23L, which have been shown to slow fibril formation as point mutations and have cumulative effect in mIAPP sequence, are present in aIAPP, suggesting that it is likely to be less

amyloidogenic than hIAPP. Our computational analysis supports this prediction, as all the programs used show that of the ancestors, aIAPP is either same as or lower in amyloidogenicity than bIAPP, which in turn is lower than hIAPP. Should this prediction be confirmed by experimental data, the additional substitutions in aIAPP will be of interest to study as point mutations. Particularly interesting are the S17V and S20N, as the algorithms show a reduction in amyloidogenicity in that region of the sequence, and have not, to our knowledge, been previously studied.

These results confirm that hIAPP is not optimised for solubility, and further suggest that amyloidogenicity has decreased over evolutionary time. However, it is unclear what pressures may have acted upon the gene for this effect to occur. One explanation for this gain-of-function is simply accidental. In addition to being post-childbearing, T2D development is associated with persistent high blood glucose concentrations, which is not a condition typically achieved by most animals, and would therefore likely not have significant selection pressures associated with it. On the other hand, many of the substitutions that appear in the ancestral sequences and have been shown to reduce amyloid formation as hIAPP point mutations are in the 18-27 region, particularly the FGAIL region, which has been implicated in being responsible for the specificity of IAPP to its receptor. Since members of the CT/CGRP family have similar receptors (CLR/CTR and RAMPs), and often show cross-reactivity, it is possible that IAPP amyloidogenicity has increased due to specificity sequence determinants. Investigating this possibility would require further experimentation on the binding kinetics of IAPP variants to CT/CGRP family receptors. However, receptors used for pharmacological studies are usually human or rat, and binding affinities of ancestral sequences to these receptors will not be an accurate representation of specificity. Ancestral reconstruction of the receptors is extremely complicated, as both the IAPP and CGRP receptors are composite of a G-protein coupled receptor (GPCR) and a RAMP, and it is possible that ancestral receptor complexes had different compositions to extant ones (37).

Even with a lack of selection against amyloidogenicity it is therefore interesting to determine which substitutions lead to the increased amyloidogenicity and how they relate to the toxicity and specificity of the peptides. Once a CGRP reconstruction is complete, it will be possible to identify which residues, if any, were shared by the IAPP and CGRP ancestors and later diverged during evolution, possibly leading to the differential specificities. As CRSPs are extremely similar to CGRPs, while often being highly specific to the CT receptor, similar analysis of CRSPs can be used to further identify CGRP-specific residues. If CGRP ancestors are shown to be non-amyloidogenic, and share a higher sequence identity with IAPP ancestors than do extant CGRP-IAPP pairs, it would support the explanation that amyloidogenicity is a by-product of emergent specificity. It can still be said in this case that amyloidogenicity is accidental, insofar as there is no positive pressure for amyloid formation itself, but the sequence determinants of specificity and aggregation are shared.

It is also worth noting that although the ancestors chosen in this study cover a long period of evolutionary time, they are extremely distant from each other, and so may not reflect the local trends within particular clades. We found that some clades had more duplicates than others, and certain sequences were more prevalent, which may indicate a different distribution of biophysical properties. To confirm that the trend we observe is indeed real, it is necessary to study it at multiple evolutionary timescales. It is further necessary to compare the clades to each other, to determine if their extant members have the same relation as the ancestors.

Methodology

Sample preparation

Peptides were synthesised using solid-phase Fmoc (9-fluorenylemethyloxycarbonyl) synthesis according to standard protocols used in previous work by our group. We performed the synthesis using a CEM microwave-assisted automatic LibertyBlue synthesiser, using CEM ProTide rink amide 0.18mmol load resin for C-terminal amide production and CEM Fmoc-protected amino acids. We used cyclised dipeptide 'pseudoprolines' Fmoc-L-Asp(tBu)-L-Thr[PSI(Me,Me)Pro]-OH and L-Ser[PSI(Me,Me)Pro]-OH, supplied by Iris Biotech GMBH to create a steric barrier for nascent chain aggregation during synthesis. Residues 1-17 from the N-terminus were double coupled, all arginine residues were double coupled at 75°C, all cysteine residues were single coupled at 50°C and all pseudoprolines were double coupled. To cleave the peptide product off the resin, as well as reduce the cyclised pseudoprolines and scavenge the side chain protecting groups, we used a cleavage mixture based on trifluoroacetic acid (TFA) - 90% TFA, 2.5% Phenol, 2.5% Thioanisole, 2.5% 6-Dioxa-1,8-octanedithiol, 2.5% H_2O. Peptide was cleaved in the mixture for 35 minutes at 38°C, after which it was diluted with diethyl ether and washed 3 times with centrifugation and resuspension. Dried peptide was dissolved in dimethyl sulfoxide (DMSO) to a concentration of 10mg/ml and purified using reverse-phase high performance liquid chromatography (RF-HPLC), using a C18 column, with water and acetonitrile as mobile phase with 0.1% TFA as the ion pairing agent. Purified product was dried and redissolved to a concentration of approximately 2.5nm and oxidised by contact with air for 3 days to form a disulfide bond between C2 and C7 residues. Oxidised peptide was then purified further, dried, dissolved in hexafluoro-2-propanol (HFIP), then lyophilised. Product identity was confirmed with time-of-flight mass spectrometry.

Thioflavin-T amyloid formation assay

Thioflavin-T (ThT) assay was conducted using a CLARIOstar Plus plate reader (BMG Labtech, Cary, NC) with 450nm excitation and 485nm emission wavelengths. Lyophilised peptides were dissolved in HFIP, aliquoted and re-lyophilised. Aliquoted dry peptide was then dissolved in 30um ThT containing buffer immediately prior to measurement. Peptides were measured at 16, 30, 60, 120, 250 and 300 μm. Measurements were allowed to continue for 14 days.

Alignment preparation and ancestral sequence reconstruction

IAPP sequences were gathered using NCBI's basic local alignment search tool (BLAST), resulting in a total of 308 sequences (38, 39). Sequences were verified first by the annotation, then by identifying cleavage signals necessary for the mature IAPP to be produced from the prosequence, as well as the G32 necessary for C-terminal amidation. Mature sequences were isolated where necessary by removing flanking regions around the cleavage sites. Resulting sequences were aligned using a local installation of MUSCLE (40). Duplicate sequences were filtered out using a local installation of CD-Hit and remaining alignment was used to construct a gene tree with RAxML (18, 41). GeneRax was then used to construct a species tree, and evaluate the gene tree with bootstrapping (21). RAxML was then used to reconstruct the ancestors on the resulting reconciled tree.
RAxML and GeneRax processes were performed on University of Oregon HPC cluster Talapas. Full method can be found at https://github.com/harmslab/asr-protocol.

Computational predictions of amyloidogenicity

5 mature IAPP sequences (human, rat, mammalian ancestor, birds and reptiles ancestor and all species ancestor) were evaluated. ANuPP, Amylogram and ZipperDB prediction programs were used via dedicated web servers (27, 28, 29, 30).

Author contribution – D.P.R provided scientific mentorship and conceived ideas for the project. A.B and A.Z performed experiments and prepared samples. I.F performed bioinformatics work constructing alignment and computational analysis. M.J.H and L.C performed ancestral sequence reconstruction.

References

1. Raleigh, D., Zhang, X., Hastoy, B., and Clark, A. (2017) The β-cell assassin: IAPP cytotoxicity. *J. Mol. Endocrinol.* **59**, R121–R140
2. Boyle, C. N., Lutz, T. A., and Le Foll, C. (2018) Amylin – Its role in the homeostatic and hedonic control of eating and recent developments of amylin analogs to treat obesity. *Mol. Metab.* **8**, 203–210
3. Zhang, X.-X., Pan, Y.-H., Huang, Y.-M., and Zhao, H.-L. (2016) Neuroendocrine hormone amylin in diabetes. *World J. Diabetes.* **7**, 189
4. Wimalawansa, S. J. (1997) Amylin, calcitonin gene-related peptide, calcitonin, and adrenomedullin: a peptide superfamily. *Crit. Rev. Neurobiol.* **11**, 167–239
5. Sekiguchi, T. (2018) The calcitonin/calcitonin gene-related peptide family in invertebrate deuterostomes. *Front. Endocrinol. (Lausanne).* **9**, 695
6. Akter, R., Cao, P., Noor, H., Ridgway, Z., Tu, L. H., Wang, H., Wong, A. G., Zhang, X., Abedini, A., Schmidt, A. M., and Raleigh, D. P. (2016) Islet Amyloid Polypeptide: Structure, Function, and Pathophysiology. *J. Diabetes Res.* 10.1155/2016/2798269
7. Russell, F. A., King, R., Smillie, S. J., Kodji, X., and Brain, S. D. (2014) Calcitonin Gene-Related Peptide: Physiology and Pathophysiology. *Physiol. Rev.* **94**, 1099
8. Cooper, G. J. S. (1994) Amylin Compared with Calcitonin Gene-Related Peptide: Structure, Biology, and Relevance to Metabolic Disease. *Endocr. Rev.* **15**, 163–201
9. Abedini, A., Cao, P., Plesner, A., Zhang, J., He, M., Derk, J., Patil, S. A., Rosario, R., Lonier, J., Song, F., Koh, H., Li, H., Raleigh, D. P., and Schmidt, A. M. (2018) RAGE binds preamyloid IAPP intermediates and mediates pancreatic β cell proteotoxicity. *J. Clin. Invest.* **128**, 682–698
10. Cao, P., Abedini, A., Wang, H., Tu, L. H., Zhang, X., Schmidt, A. M., and Raleigh, D. P. (2013) Islet amyloid polypeptide toxicity and membrane interactions. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19279–19284
11. Chantry, A., Leighton, B., and Day, A. J. (1991) Cross-reactivity of amylin with calcitonin-gene-related peptide binding sites in rat liver and skeletal muscle membranes. *Biochem. J.* **277**, 139–143
12. Watkins, H. A., Rathbone, D. L., Barwell, J., Hay, D. L., and Poyner, D. R. (2013) Structure-activity relationships for α-calcitonin gene-related peptide. *Br. J. Pharmacol.* **170**, 1308–1322
13. Murphy, R. M. (2007) Kinetics of amyloid formation and membrane interaction with amyloidogenic proteins. *Biochim. Biophys. Acta - Biomembr.* **1768**, 1923–1934
14. Westermark, P., Engstrom, U., Johnson, K. H., Westermark, G. T., and Betsholtz, C. (1990) Islet amyloid polypeptide: Pinpointing amino acid residues linked to amyloid fibril formation. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 5036–5040
15. Cao, P., Meng, F., Abedini, A., and Raleigh, D. P. (2010) The Ability of Rodent Islet Amyloid Polypeptide to Inhibit Amyloid Formation by Human Islet Amyloid Polypeptide Has Important Implications For the Mechanism of Amyloid Formation and the Design of Inhibitors.

Biochemistry. **49**, 872

16. Pullman, J., Darsow, T., and Frias, J. P. (2006) Pramlintide in the Management of Insulin-Using Patients with Type 2 and Type 1 Diabetes. *Vasc. Health Risk Manag.* **2**, 203

17. Selberg, A. G. A., Gaucher, E. A., and Liberles, D. A. (2021) Ancestral Sequence Reconstruction: From Chemical Paleogenetics to Maximum Likelihood Algorithms and Beyond. *J. Mol. Evol.* **89**, 157–164

18. Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* **30**, 1312

19. Abedini, A., Plesner, A., Cao, P., Ridgway, Z., Zhang, J., Tu, L. H., Middleton, C. T., Chao, B., Sartori, D. J., Meng, F., Wang, H., Wong, A. G., Zanni, M. T., Verchere, C. B., Raleigh, D. P., and Schmidt, A. M. (2016) Time-resolved studies define the nature of toxic IAPP intermediates, providing insight for anti-amyloidosis therapeutics. *Elife.* **5**, 1–28

20. Buchanan, L. E., Dunkelberger, E. B., Tran, H. Q., Cheng, P. N., Chiu, C. C., Cao, P., Raleigh, D. P., De Pablo, J. J., Nowick, J. S., and Zanni, M. T. (2013) Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient β-sheet. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19285–19290

21. Morel, B., Kozlov, A. M., Stamatakis, A., and Szoll}osi, G. J. (2020) GeneRax: A Tool for Species-Tree-Aware Maximum Likelihood-Based Gene Family Tree Inference under Gene Duplication, Transfer, and Loss. *Mol. Biol. Evol.* **37**, 2763–2774

22. Ridgway, Z., Eldrid, C., Zhyvoloup, A., Ben-Younis, A., Noh, D., Thalassinos, K., and Raleigh, D. P. (2020) Analysis of Proline Substitutions Reveals the Plasticity and Sequence Sensitivity of Human IAPP Amyloidogenicity and Toxicity. *Biochemistry.* **59**, 742–754

23. Abedini, A., and Raleigh, D. P. (2005) The role of His-18 in amyloid formation by human islet amyloid polypeptide. *Biochemistry.* **44**, 16284–16291

24. Khemtemourian, L., Fatafta, H., Davion, B., Lecomte, S., Castano, S., and Strodel, B. (2022) Structural Dissection of the First Events Following Membrane Binding of the Islet Amyloid Polypeptide. *Front. Mol. Biosci.* **9**, 225

25. Noh, D., Bower, R. L., Hay, D. L., Zhyvoloup, A., and Raleigh, D. P. (2020) Analysis of Amylin Consensus Sequences Suggests That Human Amylin Is Not Optimized to Minimize Amyloid Formation and Provides Clues to Factors That Modulate Amyloidogenicity. *ACS Chem. Biol.* **15**, 1408–1416

26. Tu, L. H., and Raleigh, D. P. (2013) Role of aromatic interactions in amyloid formation by islet amyloid polypeptide. *Biochemistry.* **52**, 333–342

27. Prabakaran, R., Rawat, P., Kumar, S., and Michael Gromiha, M. (2021) ANuPP: A Versatile Tool to Predict Aggregation Nucleating Regions in Peptides and Proteins. *J. Mol. Biol.* **433**, 166707

28. Thompson, M. J., Sievers, S. A., Karanicolas, J., Ivanova, M. I., Baker, D., and Eisenberg, D. (2006) The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4074–4078

29. Goldschmidt, L., Teng, P. K., Riek, R., and Eisenberg, D. (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3487–3492

30. Burdukiewicz, M., Sobczyk, P., Rödiger, S., Duda-Madej, A., MacKiewicz, P., and Kotulska, M. (2017) Amyloidogenic motifs revealed by n-gram analysis. *Sci. Rep.* **7**, 1–10

31. Xue, C., Lin, T. Y., Chang, D., and Guo, Z. (2017) Thioflavin T as an amyloid dye: Fibril quantification,
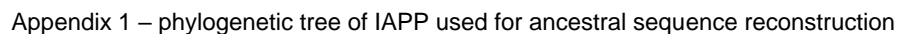
optimal concentration and effect on aggregation. *R. Soc. Open Sci.* 10.1098/rsos.160696

32. Katafuchi, T., Yasue, H., Osaki, T., and Minamino, N. (2009) Calcitonin receptor-stimulating peptide: Its evolutionary and functional relationship with calcitonin/calcitonin gene-related peptide based on gene structure. *Peptides.* **30**, 1753–1762

33. Katafuchi, T., Kikumoto, K., Hamano, K., Kangawa, K., Matsuo, H., and Minamino, N. (2003) Calcitonin receptor-stimulating peptide, a new member of the calcitonin gene-related peptide family: Its isolation from porcine brain, structure, tissue distribution, and biological activity. *J. Biol. Chem.* **278**, 12046–12054

34. Kittur, S. D., Hoppener, J. W. M., Antonarakis, S. E., Daniels, J. D., Meyers, D. A., Maestri, N. E., Jansen, M., Korneluk, R. G., Nelkin, B. D., and Kazazian, H. H. (1985) Linkage map of the short arm of human chomosome 11: Location of the genes for catalase, calcitonin, and insulin-like growth factor II. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 5064–5067

35. Steenbergh, P. H., Höppener, J. W. M., Zandberg, J., Lips, C. J. M., and Jansz, H. S. (1985) A second human calcitonin/CGRP gene. *FEBS Lett.* **183**, 408–412

36. Höppener, J. W. M., Steenbergh, P. H., Zandberg, J., Adema, G. J., van Kessel, A. H. M. G., Lips, C. J. M., and Jansz, H. S. (1988) A third human CALC (pseudo)gene on chromosome 11. *FEBS Lett.* **233**, 57–63

37. Hay, D. L., Garelja, M. L., Poyner, D. R., and Walker, C. S. (2018) Update on the pharmacology of calcitonin/CGRP family of peptides: IUPHAR Review 25. *Br. J. Pharmacol.* **175**, 3–17

38. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9

39. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410

40. Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797

41. Li, W., and Godzik, A. (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* **22**, 1658–1659

Appendix 1 – IAPP alignment used for ancestral sequence reconstruction

>Ailuropoda melanoleuca
KCNTATCATQRLANFLVRSSNNLGAILSPTNVGSNTY
>Alligator mississippiensis
KCNTATCVTQRLADFLVRSSNHIGAIYSPTNVGSNTY
>Ambystoma mexicanum
KCNTATCVTQRLADYLVRSSNNIGTIYSPTNVGSYTY
>Anaxyrus baxteri
KCNTATCVTQRLADFLVRSSNSHGPIYAPTNVGSYTY
>Anolis carolinensis
RCNTATCVTQRLADFLVRSSNTIGAIYSPTNVGSNTY
>Aotus nancymaae
KCNIATCSMHRLADFLGRSGNNFGAIISPTNVGSNTY
>Astyanax mexicanus
RCNTATCVTQRLADYLVRSSNTMGTVYSPTNVGSSTY
>Boleophthalmus pectinirostris
KCNTATCVTQRLADFLVRSSNTIGTVYAPTNVGAAAY
>Bos indicus
KCGTATCETQRLANFLAPSSNKLGAIFSPTKMGSNTY
>Bos mutus
KCGTATCETQRLANFLAPSSNKLGAISSPTKMGSNTY
>Bubalus bubalis
KCGTATCVTQRLANFLAPSGNKLGAIFSPTKMGSNTY
>Calidris pugnax
KCNTATCVTQRLADFLVRSSGNIGAIYSPTNVGSNTY
>Callorhinchus milii
TCNTATCVTQRLADFLSRSDKNLGPFYIPTNVGSNTY
>Callorhinus ursinus
KCNTATCATQRLANFLVRSSNNLGAILSHTNVGSNTY
>Canis lupus familiaris
KCNTATCATQRLANFLVRTSNNLGAILSPTNVGSNTY
>Capra hircus
KCGTATCATQRLANFLAPSGNKLGAVFSSTKMGSNTH
>Carassius auratus
KCNTATCVTQRLADFLVRSSNTRGTVYAPTNVGANTY
>Carcharodon carcharias
KCNTATCVTQRLADFLSRSNNNWGAFYIPTNVGSNTY
>Carlito syrichta
KCNTATCAMQNLAHFLVRSSNNFGAILSPTNVGSNTY
>Castor canadensis
KCNTATCATHRLANFLVHSSNNLGAVLPLTNVGSNTY
>Cavia porcellus
KCNTATCATQRLTNFLVRSSHNLGAALLPTDVGSNTY
>Cebus capucinus imitator
KCNTATCSMHRLADFLGRSGNNFGAILSPTNVGSNTY
>Ceratotherium simum simum
KCNTATCSTQCLKNFLVHSSNNLGAILSPTNVGSNMY
>Chaetura pelagica
KCNTATCVTQRLADFLVRSGSNIGAIYSPTNVGSNTY
>Channa punctata
KCNTATCVTQRLADFLVRSSNTVGTVYAPTNVGSATY
>Chanos chanos
KCNTATCVTQRLADFLIRSSNTIGTVYAPTNVGSSAY
>Chiloscyllium plagiosum
RCNTATCVTQRLADFLSRSNNNLGAFYLPTNVGSNTY
>Chinchilla lanigera
KCNTATCATQRLTHFLVRSSHHLGAVLPPTNVGSNTY
>Chlamydotis macqueenii
KCNTATCVTQRLVRSSSHIGAIYSPTNVGSNTY
>Chlorocebus sabaeus
KCNTATCATQRLANFLVRSSNNFGTILSSTDVGSNTY
>Coilia nasus
KCNTATCVTQRLADFLVRSSNTIGTVYAPTNVGASTY
>Cottoperca gobio
KCNTATCVTQRLSDFLVRSSNTIGTIYVPTNVGSSTY
>Cricetulus griseus
KCNTATCATQRLTNFLVHSSNNLGPVLSPTNVGSNTY
>Cuculus canorus
KCNTATCVTQRLADFLVRSSNSIGAIYPPTNVGSNTY
>Cynoglossus semilaevis
KCNTATCVTQRLADFLVRSSNTIGTVYAPTNVGSTAY
>Cyprinodon variegatus
KCNTATCVTQRLADFLVRSSNTIGAVYVPTNVGSSTY
>Cyprinus carpio
KCNTATCVTQRLADFLIRSSNTLGTVYAPTNVGANTY
>Danio rerio
KCNTATCVTQRLADFLIRSSNTIGTVYAPTNVGSNTY
>Dipodomys ordii
VCNTATCATHRLASFLLHSSHTLGALPPTNVGSNTY
>Electrophorus electricus
KCNTATCVTQRLADFLIRSSNAIGTVYAPTNVGSSTY
>Enhydra lutris kenyoni
KCNTATCVTQRLANFLVRSSNNLGAILSPTDVGSNTY
>Equus asinus
KCDTATCVTQRLANFLVHSSNNLGAILSPTNVGSNTY

>Equus caballus
KCDTATCVTQRLANFLVHSSNNLGAILSPTSVGSNTY
>Erinaceus europaeus
RCNTATCATQRLVNFLSRSSNNLGAILSPTDVGSNTY
>Erpetoichthys calabaricus
KCNTATCVTQRLADYLIRSSSNKGMLYAPTNVGSNTY
>Esox lucius
KCNTATCVTQRLADFLTRSSNTIGTVYTPTNVGSSTY
>Falco peregrinus
KCNTATCVTQRLADFLVRSSNNIGAIYSPTNVGSNTY
>Felis catus
KCNTATCATQRLANFLIRSSNNLGAILSPTNVGSNTY
>Fukomys damarensis
KCNTATCATQRLTTFLVRSSHNLGAALLPTDVGSNTY
>Gallus gallus
KCNTATCVTQRLADFLVRSSSNIGAIYSPTNVGSNTY
>Gekko japonicus
KCNTATCVTQRLADFLVRSSNTIGTIYSPTNVGANTY
>Geospiza fortis
KCNTATCVTQRLADFLVRSSGSAGALYPPTNVGSNTY
>Globicephala melas
KCYTATCATQCLENFLVWSSNNLGTIFSPTKVGSNTY
>Gopherus evgoodei
KCNTATCVTQRLADYLVRSSNNIGAIYSPTDVGSNTY
>Gorilla gorilla gorilla
KCNTVTCATQRLANFLVRSSNNFGAILSSTNVGSNTY
>Gouania willdenowi
KCNTATCVTQRLADFLVRSSNTVGTVYIPTNVGSSTY
>Gymnocypris eckloni
TCNTATCMTQRLADFLIRSSNTLGTVYAPTNVGANTY
>Gymnocypris selincuoensis
TCNTATCVTQRLADFLIRSSNTLGTVYAPTNVGANTY
>Haliaeetus albicilla
KCNTATCVTQRLADFLVRSSNSIGAIYSPTNVGSNTY
>Heterocephalus glaber
KCNTATCTIQRLTNFLVRSSHNLGAVLLPTDVGSNTY
>Heterodontus zebra
KCNTATCVTQRLADFLSRSNHNLGAIYMPTNVGSNTY
>Homo sapiens
KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTY
>Hynobius retardatus
KCNTATCVTQRLADFLVRSSNNIGAIYSPTNVGSYTY
>Jaculus jaculus
KCNTATCATQRLANFLVRSSSSLGVILPATNVGSNTY
>Labrus bergylta
KCNTATCVTQRLADFLVRSSNTIGTVYAPTNVGSSTY
>Latimeria chalumnae
KCNTATCVTQRLADFLIRANGNMRALHSPTNVGAYTY
>Lepisosteus oculatus
KCNTATCVTQRLADFLIRSSNTVGTVYSPTNVGSNTY
>Lepus americanus
KCNTVTCATQRLANFLIHSSNNFGAIFSPPSVGSNSY
>Lipotes vexillifer
ECYTATCATQHLENFLVWSSNNLGAIFSPTKVGPNTY
>Loxodonta africana
KCNTATCVRQHLANFLHHSSDSLESVLPATNVGSNTY
>Macaca nemestrina
KCNTATCATQRLANFLVRSSNNFGTILSSTNVGSDTY
>Manacus vitellinus
KCNTATCVTQRLADFLVRSSNSIGAVYSPTNVGSNTY
>Mandrillus leucophaeus
KCNTATCATQRLANFLVRSSNNFGTILSSTNVGSNTY
>Marmota marmota marmota
KCNTATCATQRLANFLVRSSHNLGAILSTTNVGSNTY
>Mastacembelus armatus
RCNTATCVTQRLADFLVRSSNIIGTVYTPTNVGSATY
>Mesocricetus auratus
KCNTATCATQRLANFLVHSNNNLGPVLSPTNVGSNTY
>Microcebus mittermeieri
TCSTATCTIQRLANFLVRYSNNFGAVHSPTNVRSNGY
>Microtus ochrogaster
ECNTATCATQRLANFLVHSSNGLGPVLSSTNVGSNTY
>Micrurus surinamensis
KCNTATCVTQRLADFLVRSSNTIGTIYAPTNVGSNTY
>Monodon monoceros
KCYTATCATQRLENFLVWSSNNLGVIFSPTKVGSNTY
>Mus musculus
KCNTATCATQRLANFLVRSSNNLGPVLPPTNVGSNTY
>Myripristis murdjan
KCNTATCVTQRLADFLVRSSNTIGTVYAPTNVGSSAY
>Nanorana parkeri
KCNTATCVTQRLADFLVRSNHHGPLYGPTNVGSFTY
>Nestor notabilis
KCNTATCVTQRLADFLVRSGNNIGAIYSPTNVGSNTY
>Notamacropus eugenii
KCNTATCVTQRLADFLVRSNNNMGAIFSPTNVGSNTY
>Nothobranchius furzeri
KCNTATCVTQRLADFLVRSSNTIGTVYVPTNVGSSTY

```
>Nothobranchius korthausae
KCNTATCVTQRLADFLVRSSNTVGTVYVPTNVGSSTY
>Notothenia coriiceps
KCNTATCVTQRLSDFLVRSSNTIGTVYAPTNVGSGTY
>Ochotona princeps
KCNTITCATQRLANFLVHSSNNFGAIFSPVNLGSKSY
>Octodon degus
KCNTATCATQRLTNFLVRSSHNLGAALPPTKVGSNTY
>Odorrana tormota
KCNTATCVTQRLADFLVRSNNHHGPIHSPTNVGSFTY
>Oncorhynchus nerka
KCNTATCVTQRLADFLTRSSNTIGTVYAPTNVGSNTY
>Orycteropus afer afer
KCNTATCATQRLTNFLTRSSNNIGAIPPSTNVGSNTY
>Oryzias melastigma
KCNTATCVTQRLADFLVRSSNTIGTVYVPTNVGSATY
>Otolemur garnettii
KCNTATCATQRLANFLVRSSNNFGAVHSPTNVGSNTY
>Ovis aries
KCGTATCATQRLANFLAPSGNKLGAVFSPRKMGSNTH
>Pan troglodytes
KCNTATCATQRLANFLVRSSNNFGAILSSTNVGSNTY
>Papio anubis
ICNTATCATQRLANFLVRSSNNFGTILSSTNVGSNTY
>Paramormyrops kingsleyae
RCNTATCVTQRLADFLIRSSNTIGTVYAPTNVGSNTY
>Parus major
KCNTATCVTQRLADFLVRSSNTIGAVYSPTNVGSNTY
>Peromyscus maniculatus bairdii
KCNTATCATQRLTNFLVRSSNNLGPVLPPTNVGSNTY
>Petromyzon marinus
ECKASTCAIQRLVSMMSR--NRVSGDSPRTDVGSATY
>Phaethon lepturus
KCNTATCVTQRLADFLVRSSNNIGAVYSPTNVGSNTY
>Piliocolobus tephrosceles
KCNTATCATQRLANFLVRSSNNFGPILSSTNVGSNTY
>Poecilia reticulata
KCNTATCVTQRLADFVVRSSNTIGTVYVPTNVGSSTY
>Pogona vitticeps
RCNTATCVTQRLADFLVRSSNTFGAIYTPTNVGSNTY
>Porichthys notatus
KCNTATCVTQRLADFLVRSSNTIGTVYAPTNVGSGAY
>Ptychobarbus kaznakovi
TCNTATCVTQRLADFLIRSSNTLGTIYAPTNVGANTY
>Pygocentrus nattereri
RCNTATCMTQRLADFLVRSSNTIGTVYIPTNVGSSAY
>Python bivittatus
KCNTATCVTQRLADFLVRSSNTIGAIYSPTNVGSNTY
>Rana temporaria
RCNTATCVTQRLADFLVRSNNHHDAIHSPTNVGSFTY
>Rhinatrema bivittatum
KCNTATCVTQRLADFLVRSSNISELFIHQPMWGSYTY
>Rhincodon typus
RCNTATCVTQRLADFLSRSNSNLGAFYLPTNVGSNTY
>Rhinopithecus bieti
KCNTATCATQRLANFLVRSSNNFGSILSSTNVGSNTY
>Rousettus aegyptiacus
KCNTATCAVQWLANFLVHSSNNFSVSLSLTNVGSNTY
>Salarias fasciatus
KCNTATCVTQRLADFLVRSSNTIGTVYVPTNVGSAAY
>Salmo salar
KCNTATCVTQRLADFLTRSSNTIGTMYAPTNVGSSTY
>Salmo trutta
KCNTATCVTQRLADFLTRSSNTIGTVYAPTNVGSSTY
>Salvelinus alpinus
KCNTATCVTQRLADFLTRSSNTIGTVYAPTNVGSSAY
>Sander lucioperca
KCNTATCVTQRLSDFLVRSSNTIGTVYAPTNVGSATY
>Schizothorax oconnori
KCNTATCVTQRLADFLIRSSNTLGTVYASTDVGANTY
>Scleropages formosus
RCNTATCVTQRLADFLIRSSNTIGTVYAPTNVGSSAY
>Scyliorhinus torazame
MCNTATCVTQRLADFLSRSNNKLGAFYMPTNVGSNTY
>Serinus canaria
KCNTATCVTQRLADFLVRSSGSVGALYPPTNVGSHTY
>Sinocyclocheilus anshuiensis
```

Appendix 1 – phylogenetic tree of IAPP used for ancestral sequence reconstruction

Mus musculus
Peromyscus maniculatus bairdii
Mesocricetus auratus
Cricetulus griseus
Microtus ochrogaster
Jaculus jaculus
Dipodomys ordii
Castor canadensis
Octodon degus
Chinchilla lanigera
Cavia porcellus
Fukomys damarensis
Heterocephalus glaber
Marmota marmota marmota
Urocitellus parryi
Lepus americanus
Ochotona princeps
Rhinopithecus bieti
Piliocolobus tephrosceles
Chlorocebus sabaeus
Mandrillus leucophaeus
Papio anubis
Theropithecus gelada
Macaca nemestrina
Homo sapiens
Pan troglodytes
Gorilla gorilla gorilla
Aotus nancymaae
Cebus capucinus imitator
Carlito syrichta
Microcebus murinus
Otolemur garnettii
Capra hircus
Ovis aries
Bubalus bubalis
Bos indicus
Bos mutus
Lipotes vexillifer
Globicephala melas
Monodon monoceros
Sus scrofa
Equus asinus
Equus caballus
Ceratotherium simum simum
Rousettus aegyptiacus
Enhydra lutris kenyoni
Ursus maritimus
Ursus arctos horribilis
Ailuropoda melanoleuca
Callorhinus ursinus
Canis lupus familiaris
Felis catus
Erinaceus europaeus
Loxodonta africana
Orycteropus afer afer
Notamacropus eugenii
Vombatus ursinus
Taeniopygia guttata
Geospiza fortis
Serinus canaria
Parus major
Manacus vitellinus
Nestor notabilis
Falco peregrinus
Haliaeetus albicilla
Calidris pugnax
Phaethon lepturus
Chlamydotis macqueenii
Cuculus canorus
Chaetura pelagica
Gallus gallus
Alligator mississippiensis
Gopherus evgoodei
Terrapene carolina triunguis
Micrurus surinamensis
Python bivittatus
Pogona vitticeps
Anolis carolinensis
Gekko japonicus
Sphenodon punctatus
Odorrana tormota
Rana temporaria
Nanorana parkeri
 Anaxyrus baxteri
Xenophrys sangzhiensis
Xenopus laevis
Hynobius retardatus
Ambystoma mexicanum
Rhinatrema bivittatum
Schizothorax oconnori
Gymnocypris eckloni
Gymnocypris selincuoensis
Ptychobarbus kaznakovi
Cyprinus carpio
Carassius auratus
Sinocyclocheilus rhinocerous
Sinocyclocheilus anshuiensis
Danio rerio
Electrophorus electricus
Tachysurus fulvidraco
Astyanax mexicanus
Pygocentrus nattereri
Chanos chanos
Colia nasus
Oncorhynchus nerka
Salvelinus alpinus
Salmo trutta
Salmo salar
Esox lucius
Gouania willdenowi
Tripterygion delaisi
Salarias fasciatus
Poecilia reticulata
Cyprinodon variegatus
Nothobranchius furzeri
Nothobranchius korthausae
Oryzias melastigma
Cynoglossus semilaevis
Channa punctata
Mastacembelus armatus
Tetraodon nigroviridis
Takifugu rubripes
Labrus bergylta
Notothenia coriiceps
Cottoperca gobio
Sander lucioperca
Boleophthalmus pectinirostris
Sphaeramia orbicularis
Poricthys notatus
Myripristis murdjan
Paramormyrops kingsleyae
Scleropages formosus
Lepisosteus oculatus
Erpetoichthys calabaricus
Tetronarce californica
Heterodontus zebra
Scyliorhinus torazame
Chiloscyllium plagiosum
Rhincodon typus
Carcharodon carcharias
Squalus acanthias
Callorhinchus milii
Petromyzon marinus