scientific reports

Check for updates

OPEN

Evolutionary analysis through structural modeling of FAM222 proteins reveals a novel disordered conserved domain in vertebrates that interacts with NLK

Eduardo Calvario[1], Lorenzo Segovia[2✉] & Mario Zurita[1✉]

Intrinsically disordered regions (IDRs) play essential roles in signaling by mediating protein-protein interactions that are both dynamic and specific. Many IDRs have evolved as kinase-binding modules, particularly in mitogen-activated protein kinase (MAPK) networks, where they facilitate critical regulatory interactions for signaling specificity. However, the evolutionary constraints that shape IDRs and their role in MAPK regulation remain poorly understood. Here, we identify DCD222, a highly conserved IDR within the FAM222 protein family, which interacts with Nemo-like kinase (NLK), an atypical MAPK involved in WNT, NOTCH, VEGF, mTOR, and TGF-β signaling. Phylogenetic analysis reveals that DCD222 has been conserved from cyclostomes to humans, indicating strong evolutionary constraints on its function. Using AlphaFold3, we show that DCD222 binds to NLK through an extended mechanism, in which a reverse docking (revD) motif engages the NLK docking groove in the C-lobe, and β-hairpin-like and β-sheet motifs insert into the N-lobe. We performed three independent atomistic simulations of the protein–protein complex to assess the stability of the interface and found that it remained stable throughout the runs. The interaction does not obstruct the kinase's catalytic site, suggesting that FAM222A and FAM222B are potential modulators of NLK rather than simple substrates. Our findings establish DCD222 as an evolutionarily constrained IDR, providing new insights into how disordered domains contribute to signaling adaptation in vertebrates.

Keywords AlphaFold3, Intrinsically disordered region, MAPK signaling, Protein evolution, Nemo-like kinase.

IDRs play key roles in the regulation of cellular signaling due to their ability to form transient, yet highly specific interactions with other proteins[1,2]. Unlike globular domains, intrinsically disordered proteins (IDPs) and IDRs lack a stable tertiary structure, allowing them to adopt multiple conformations depending on molecular context through mechanisms such as induced folding or partner recognition[3]. More than 30% of eukaryotic proteins have been estimated to contain IDRs[4], and a significant fraction of these interact with regulatory enzymes in different cell signaling pathways[5]. IDRs that interact with kinases have evolved to act as modulators of signaling pathways, either by forming dynamic complexes or by allosterically regulating kinase activity without directly obstructing the active site[6,7].

IDRs involved in MAPK signaling often contain docking motifs, which are crucial for mediating interactions with kinases and assembling highly specific intracellular signaling networks[8]. Well-characterized examples include the scaffold protein JIP1, which regulates JNK activation through a flexible region[9,10], and MEK1, which recruits ERK through a conserved disordered motif[11]. IDRs exhibit high sequence variability, indicating that selective pressures act on their structural plasticity and interaction interfaces rather than on specific residues[12].

[1]Departamento de Genética del Desarrollo y Fisiología Molecular 1, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av Universidad 2001, Col. Chamilpa, Mexico City 62250, Mexico. [2]Departamento de Ingeniería Celular y Biocatálisis 2, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av Universidad 2001, Col. Chamilpa, 62250 Mexico City, México. ✉email: lorenzo.segovia@ibt.unam.mx; mario.zurita@ibt.unam.mx

This raises the question of how new IDR-related proteins have evolved to modulate specific MAPKs and whether these disordered regions have been subjected to selective constraints in vertebrate evolution.

In this study, we investigated the FAM222 protein family, which includes two IDPs in humans: FAM222A (UniProt ID: Q5U5X8) composed of 452 amino acids with a molecular weight of 46.7 kDa[13–15], and FAM222B (Q8WU58) with 562 residues and 59.6 kDa[13,16]. FAM222A has been the most studied member of the family since 2020 and has been implicated in neurodegenerative diseases, including Alzheimer's, where it has been identified in extracellular senile plaques and has been shown to interact with β-amyloid (Aβ) peptides, promoting their aggregation[14]. Additionally, FAM222A has been associated with Nasu-Hakola disease, a rare neurodegenerative disorder that affects both the nervous and skeletal systems[17].

Recent studies suggest that FAM222A may also be involved in angiogenesis[18] and innate immune responses where it has been reported to interact with mitochondrial antiviral signaling protein (MAVS) in mitochondria, promoting its aggregation and activating type I interferon (IFN-I) signaling in response to viral infections[15].

Despite their implication in disease processes, both FAM222A and FAM222B lack recognizable domains and remain poorly understood at the molecular level[14,18]. To address this gap, we carried out an evolutionary and structural analysis of these proteins, and we identify a highly conserved IDR within the FAM222 protein family that we named *D*isordered *C*onserved *D*omain within FAM*222* (DCD222). This disordered domain is conserved in vertebrates, from hagfish to humans, suggesting a strong selective pressure on its function. Using AlphaFold3[19], we demonstrated that DCD222 preferentially interacts with NLK, an atypical MAPK involved in the regulation of multiple signaling pathways, including WNT, NOTCH, VEGF, mTOR, and TGF-β[20–25]. Specifically, DCD222 contains a revD motif that binds to the NLK docking groove and a β-sheet motif that inserts into a previously uncharacterized hydrophobic groove in the N-lobe of NLK. This interaction occurs without obstructing the catalytic site of the kinase, suggesting that FAM222A and FAM222B may act as modulators of NLK.

Given the evolutionary significance of IDRs in MAPK regulation, we hypothesize that the DCD222 domain in FAM222 represents a novel case of an evolutionarily constrained IDR specifically tuned to interact with NLK through a non-canonical mechanism.

Results

Evolutionary origins of FAM222: duplication of genes in gnathostomes

Despite the growing interest in FAM222A and its medical relevance, its phylogenetic history remains unexplored. Furthermore, no homologous domains have been identified within the amino acid sequence, raising fundamental questions about its evolutionary origins and diversification. To address this, we investigated the evolutionary history of FAM222A and FAM222B. We performed a BLASTp search using the human FAM222A (hFAM222A) sequence as a query. This search retrieved 2208 protein sequences annotated as FAM222A and FAM222B, which were exclusively found in 682 vertebrate species. We then selected a representative subset consisting of 20 FAM222A and 18 FAM222B sequences from various vertebrate lineages. To further investigate the evolutionary relationships among these sequences, we performed a multiple sequence alignment (MSA) (Supplementary Fig. 1) followed by phylogenetic reconstruction using the maximum likelihood method (Fig. 1A). The resulting phylogeny revealed that FAM222A and FAM222B cluster into two distinct clades within gnathostomes (jawed vertebrates), supporting the hypothesis that these proteins are paralogs that originated from a gene duplication event in a common ancestor of gnathostomes.

Furthermore, sequence identity analysis showed that hFAM222A and human FAM222B (hFAM222B) share only 34.3% identity, indicating that these proteins have undergone significant evolutionary divergence (Fig. 1B). Additionally, we found that FAM222A maintains 64.6% identity from zebrafish to humans, while FAM222B retains 58% identity over the same evolutionary range.

Interestingly, our phylogenetic analysis identified two FAM222-like proteins that form an outgroup at the base of the vertebrate tree, corresponding to cyclostomes (jawless vertebrates), specifically lampreys and hagfish (Fig. 1A). These sequences exhibit 28% and 26.6% identity with hFAM222A, respectively, and 25.6% and 26% identity with hFAM222B, while sharing 40% identity between themselves (Fig. 1B). This indicates that the cyclostome FAM222-like proteins are single-copy orthologs of the ancestral gene that later duplicated in gnathostomes.

Taken together, our results indicate that FAM222A and FAM222B originated from an ancestral gene that was already present before the divergence of cyclostomes and gnathostomes. However, the gene duplication event in gnathostomes played a crucial role in the functional specialization of FAM222 proteins, which would contribute to lineage-specific adaptations in jawed vertebrates.

FAM222 proteins contain a highly conserved IDR in vertebrates

The conserved regions in proteins can reveal essential biological functions and provide information into specific evolutionary pressures. To identify functionally relevant regions within the FAM222 family, we analyzed the conservation of their sequences in vertebrates. Previous studies have reported that hFAM222A contains a proline-rich region (PRR), while hFAM222B possesses a glycine-rich region in the central part of the protein (GRR)[26]. To further define these regions, we used the amino acid sequences of FAM222 proteins and analyzed them with the scanProsite server[27]. Our results showed that hFAM222A contains a PRR of 152 amino acid residues (147–299), whereas hFAM222B has a glutamine-rich (QRR), rather than a GRR, region consisting of 72 amino acid residues (148–220) (Fig. 2A).

A more detailed examination identified a highly conserved region spanning 72 to 77 amino acid residues, present from hagfish to humans, with an identity of 62.5% (Fig. 2A and Supplementary Fig. 2). The strong evolutionary conservation of this region suggests a critical functional role. Additionally, a BLASTp search against the RefSeq Protein database (excluding vertebrates) yielded no hits, indicating that this amino acid region is an exclusive evolutionary signature of this lineage.
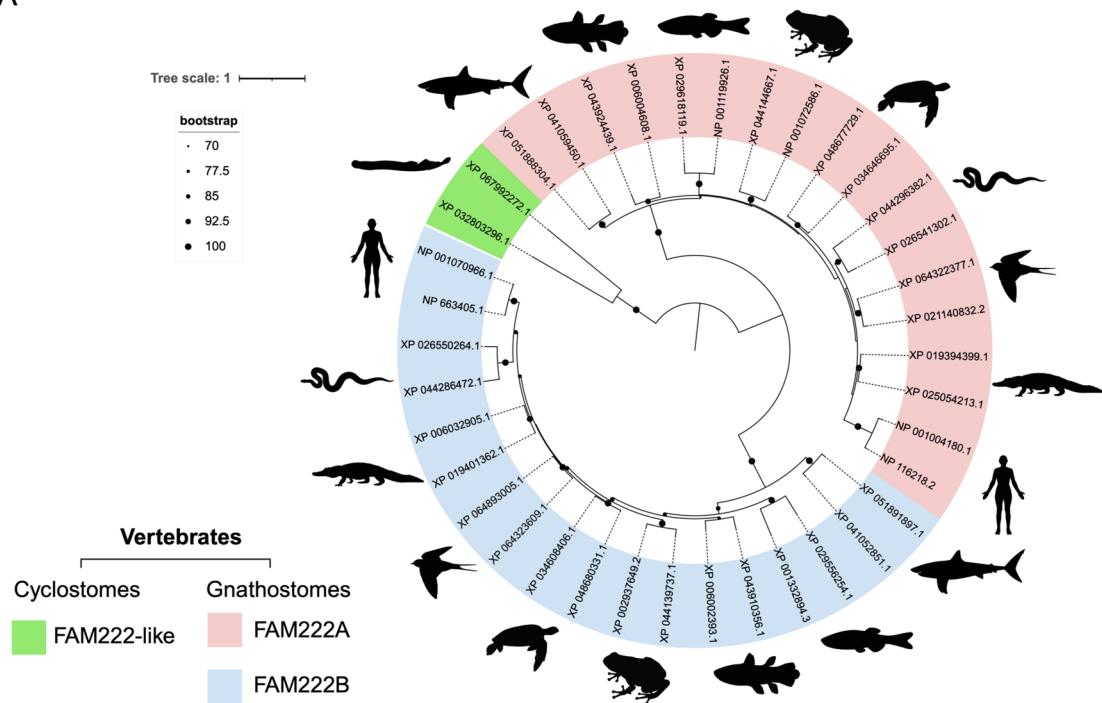
Fig. 1. Evolutionary history of the FAM222 protein family. (**A**) Maximum likelihood phylogenetic tree constructed from 38 full-length protein sequences annotated as FAM222A or FAM222B, representing various vertebrate species, including birds, turtles, alligators, snakes, mammals, amphibians, coelacanths, lungfishes, bony fishes, cartilaginous fishes, lampreys, and hagfish. Only bootstrap values above 70% are displayed. (**B**) Percentage identity matrix generated from the MSA. The matrix shows identity values among FAM222 proteins from representative vertebrates, including human (*H. sapiens*), mouse (*Mus musculus*), frog (*Xenopus tropicalis*), zebrafish (*Danio rerio*), lamprey (*Petromyzon marinus*), and hagfish (*Myxine glutinosa*). Created in BioRender. Calvario, E. (2025) https://BioRender.com/9qf40ns.
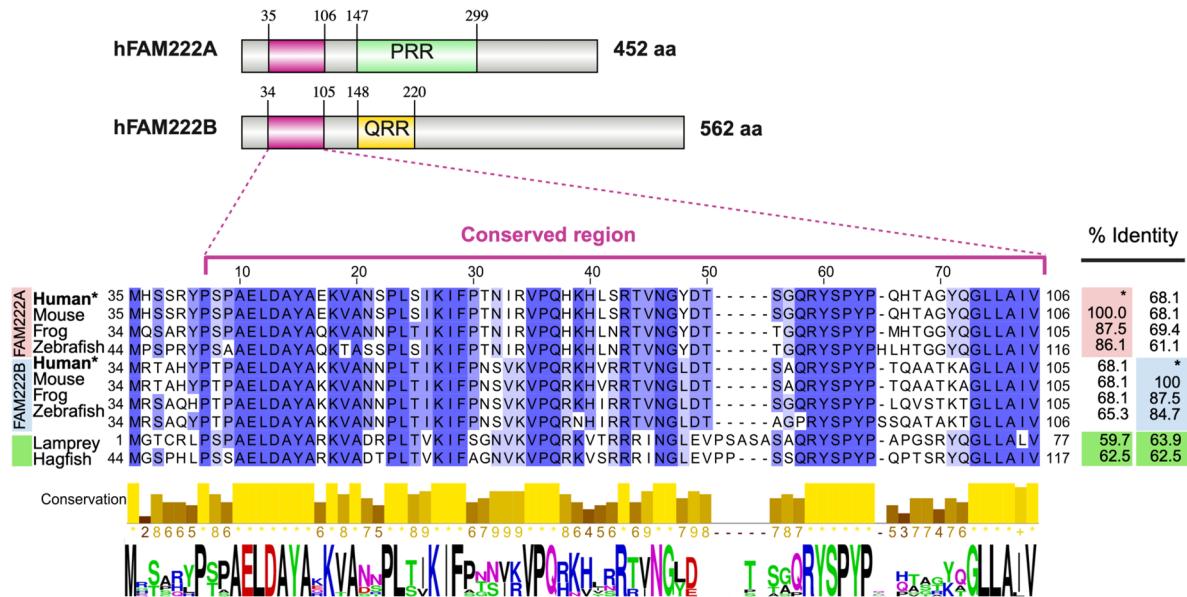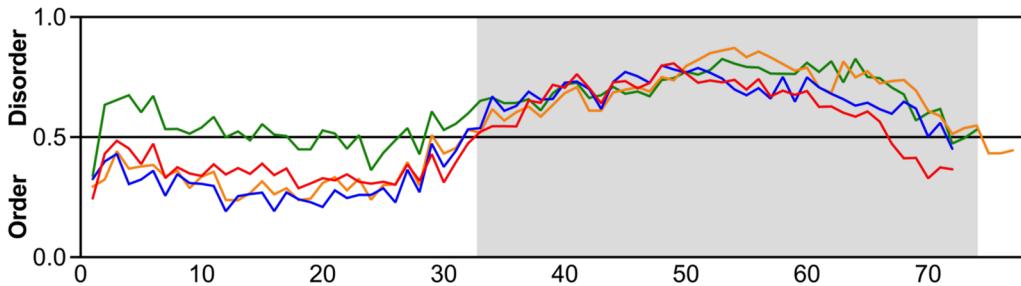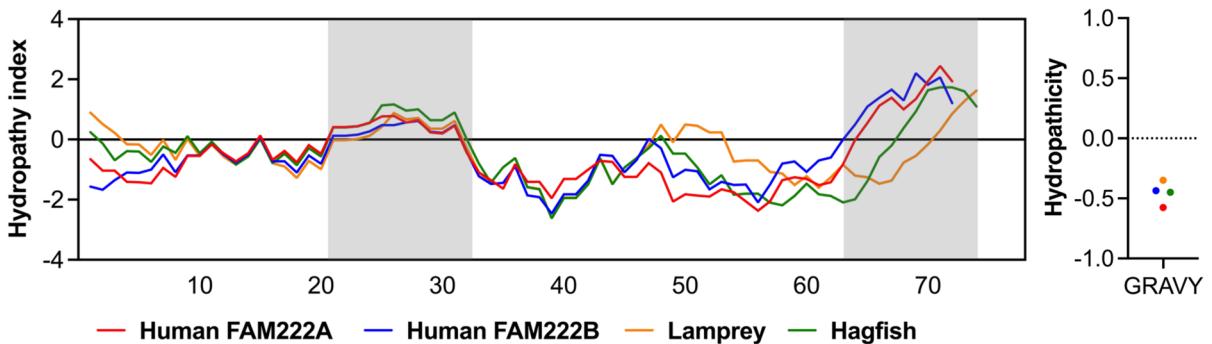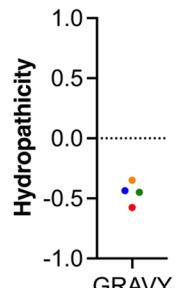
Fig. 2. Analysis of the conserved region in FAM222 proteins. (**A**) Schematic representation of the conserved region in FAM222 proteins generated using IBS 2.0[61]. The lower panel displays the MSA of the most conserved region in this protein family, spanning from hagfish to humans, visualized with Jalview. Amino acid residues are shaded (purple blue) according to their percentage of identity. (**B**) Order-disorder profile of the conserved region, where values greater than 0.5 indicate intrinsic disorder. (**C** and **D**) Hydropathy profile of the conserved region, shown at the residue level and globally. Positive values indicate hydrophobicity, while negative values indicate hydrophilicity.

Given that hFAM222A has previously been characterized as an IDP through circular dichroism[14], we evaluated whether this conserved region also exhibited structural disorder. To address this, we used the DR-BERT tool to analyze disorder in the conserved amino acid sequences of hFAM222A (35–106) and hFAM222B (34–105), comparing them with their orthologs in cyclostomes, such as lampreys (1–77) and hagfish (44–117). The results

indicated that this region has an average disorder percentage of 60%, with a more pronounced tendency in its second half, allowing it to be classified as an IDR (Fig. 2B, gray-shaded segment). To further explore the conservation of its physicochemical properties, we assessed the hydropathy profile of this region using the Kyte-Doolittle scale[28]. Two significantly hydrophobic segments were identified in the central and C-terminal regions, leading us to hypothesize that these hydrophobic patches could play a role in protein-protein interaction (Fig. 2C, gray-shaded segments). Since the remainder of the sequence is predominantly hydrophilic, we conducted a Grand Average of Hydropathy (GRAVY) analysis (Fig. 2D). All values obtained were negative, indicating a strong hydrophilic tendency. This property is consistent with the presence of solvent-exposed regions or a disordered conformation[29,30].

The combination of strong evolutionary conservation, structural disorder, and hydrophilic profile suggests that this region may play a key functional role in the FAM222 family. Specifically, its conservation from hagfish to humans supports the hypothesis that it could participate in molecular recognition-mediated interactions.

Structural conservation of the disordered region in FAM222

To gain deeper insight into the structural organization of the conserved region between humans and cyclostomes, we generated structural models using AlphaFold3. Although AlphaFold3 has been extensively validated for high-confidence predictions of globular protein structures, it has also proven to be a valuable tool for identifying intrinsically disordered regions, providing insights into their potential structural organization and conformational tendencies[31,32].

As a first step, we predicted the structures of hFAM222A and hFAM222B from their amino acid sequences. The results indicated that both proteins contain five α-helices embedded within extensive disordered regions, characterized by low pLDDT values (50–70), suggesting a flexible and intrinsically disordered behavior (Fig. 3A and Supplementary Videos 1 and 2).

Unexpectedly, we identified that α-helix H1, present in both proteins, exhibited a pLDDT>90, indicating high confidence in its structural prediction. This structural motif is located at positions 42–53 in hFAM222A and 41–52 in hFAM222B, within the previously identified conserved region. To assess the structural conservation of this structural motif in basal vertebrates, we generated structural models of FAM222-like proteins XP_032803296.1 lamprey (laXP), and XP_067992272.1 hagfish (haXP) (Supplementary Fig. 3 and Supplementary Videos 3 and 4). The results revealed that H1 is also present in both species with a pLDDT comparable to the proteins from human, suggesting that this structural motif has been maintained throughout vertebrate evolution (Fig. 3B and C). In contrast, the remainder of the conserved region remains largely disordered, with pLDDT values between 50 and 70, characteristic of IDRs. However, we observed the formation of small α-helices at the end of the conserved region, except in hFAM222B, and the presence of a possible β-hairpin-like, except in hFAM222A (Fig. 3B and C). Given the low pLDDT values of these elements, it is likely that they represent transient structures that stabilize only under specific conditions.

On the other hand, the conservation of H1 and its high pLDDT in all analyzed species suggests that this structure is stable and plays a key functional role in FAM222 proteins. Its conservation in vertebrates, from cyclostomes to humans, raises the possibility that this α-helix has been retained due to selective pressures, possibly related to protein-protein interactions or molecular regulatory mechanisms.

The disordered conserved region of FAM222 proteins directly binds to NLK

Based on our previous analysis and findings, we have designated the conserved region within FAM222 proteins as DCD222. Given the possibility that this region mediates interactions with other proteins, we explored its interaction network using the STRING database[33], restricting our analysis to experimentally validated interactions.

Our results (Fig. 4A) revealed that hFAM222A physically interacts with seven proteins, including four transcription factors from the MEIS/PBX families (MEIS1, MEIS2, PBX1, PBX2, and PBX3) and with a putative nucleotidyl transferase (MAB21L1) and a MAPK (NLK)[15,34–36]. Similarly, hFAM222B was found to interact with three additional proteins, including an RNA helicase (DDX39A)[37], a ubiquilin protein (UBQLN2)[38], and a transcription factor (SOHLH1)[39], as well as NLK [36]. To assess the relevance of these interactions, we used AlphaFold3 to model protein-protein and protein-DCD222 complexes, as well as to evaluate the effect of DCD222 in silico deletion on the protein-protein interaction (Fig. 4B). Our findings indicated that the strongest interaction was established with NLK, as interface predicted template modelling (ipTM) values exceeded the 0.6 threshold for both hFAM222A and hFAM222B, while all other interactions fell below this cutoff point. However, we observed that ipTM scores for FAM222-NLK were within AlphaFold's gray zone (0.6–0.8), a range where predictions may be prone to false positives or inaccuracies in the predicted interaction sites. To analyze this interaction in more depth, we specifically evaluated the protein-DCD222 interactions. In particular, the ipTM scores for NLK increased above the gray zone, suggesting a more reliable interaction. On the contrary, when DCD222 was deleted, the FAM222-NLK interaction dropped below the 0.6 threshold, indicating that removal of DCD222 significantly weakens or abolishes binding to NLK in both hFAM222A and hFAM222B (Fig. 4B). These results demonstrate that DCD222 is essential for the interaction with NLK, suggesting that this domain plays a pivotal role in the functional association of FAM222 proteins with this kinase.

We investigated whether this association has been evolutionarily conserved among vertebrates. To assess this hypothesis, we replicated our analysis using the laXP and haXP proteins of cyclostomes, employing AlphaFold3 to model their interactions with NLK. The results were highly consistent with those observed for hFAM222A and hFAM222B, revealing similar interaction patterns and structural stability (Fig. 4C). To further explore the structural conservation of DCD222 in its NLK-bound state, we performed a structural alignment of DCD222 from hFAM222A, hFAM222B, laXP, and haXP (Fig. 4D). Surprisingly, this alignment revealed an average root mean square deviation (RMSD) of 0.570±0.047 Å, indicating that despite being an IDR, DCD222 adopts a
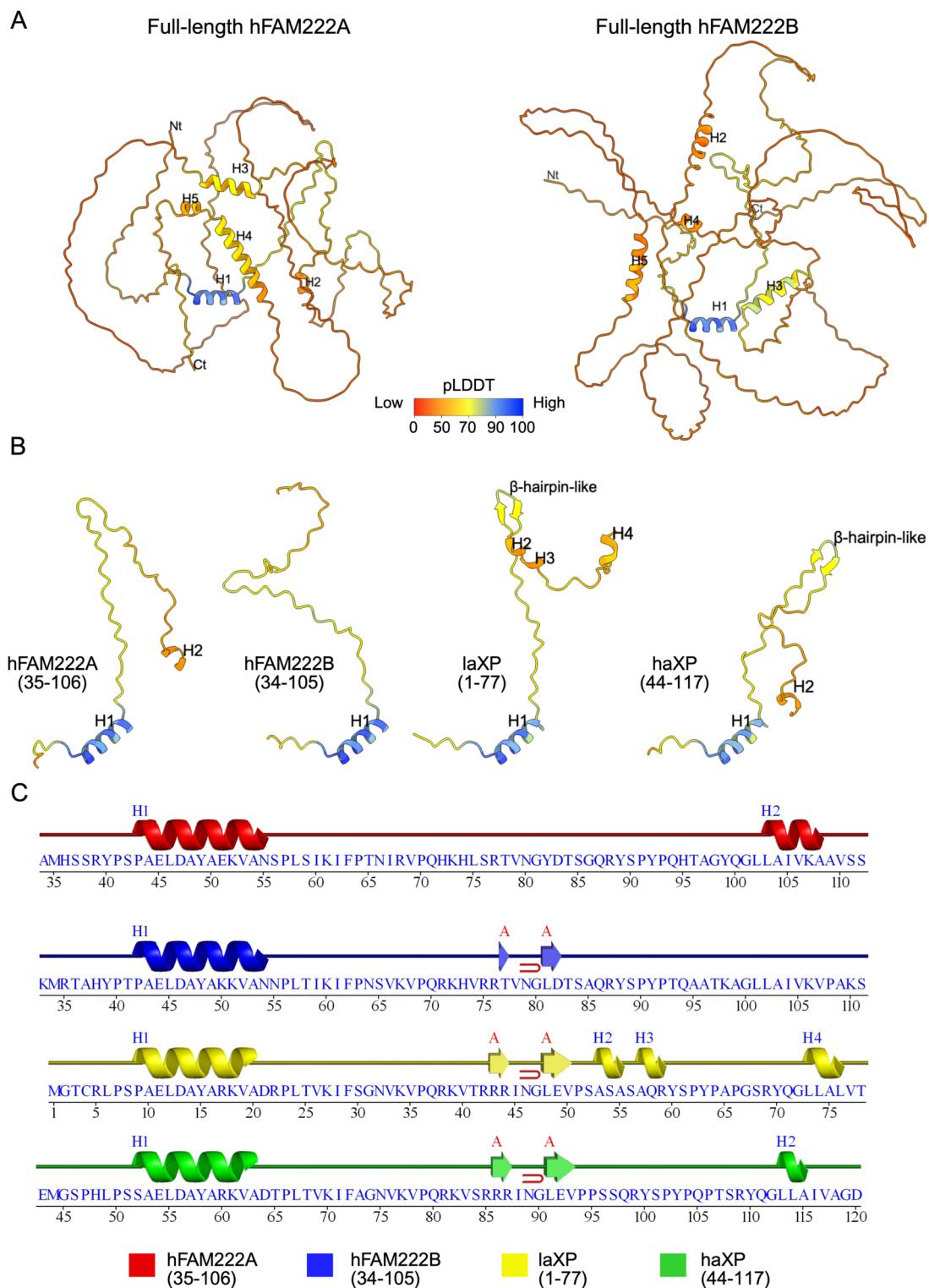
Fig. 3. Structural modeling of the conserved region in FAM222 proteins. (**A**) Representative structural models of hFAM222A and hFAM222B generated using AlphaFold3. Amino acid residues are colored based on their pLDDT scores, where values near 100 indicate high confidence and values near 0 indicate low confidence, as visualized in ChimeraX. (**B** and **C**) Structural models of the conserved domain in gnathostomes and cyclostomes, with their corresponding linear representations of secondary structure.
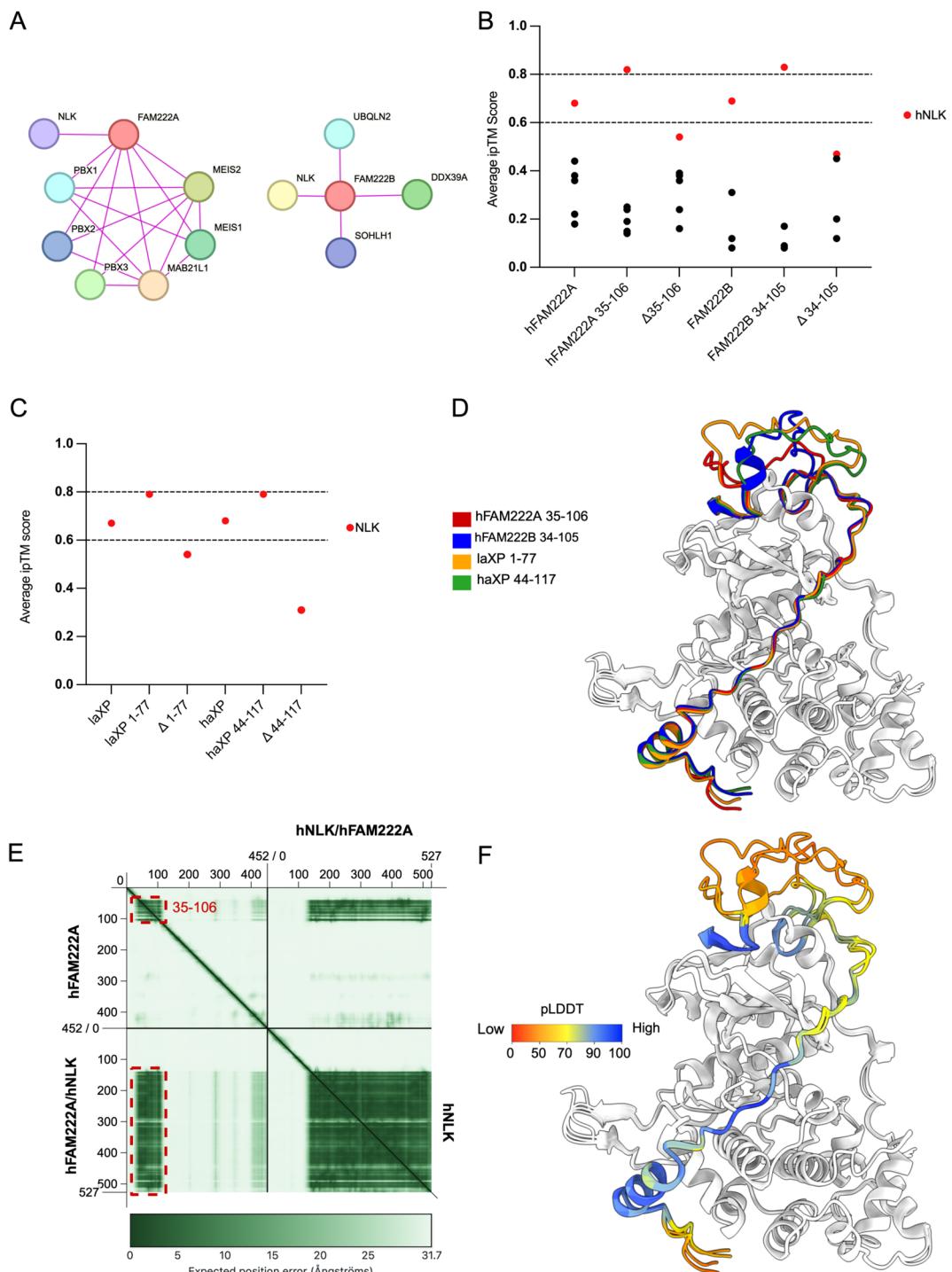
Fig. 4. Interaction of the DCD222 region with protein partners. (**A**) Experimentally validated protein-protein interactions of hFAM222A and hFAM222B retrieved from the STRING database. (**B** and **C**) Average ipTM values for protein-protein interactions predicted using AlphaFold3. Values between 0.6 and 0.8 (gray zone) indicate moderate confidence with some uncertainty in binding affinity, while values above 0.8 suggest a highly confident interaction prediction. (**D** and **F**) Structural alignment of the DCD222 region of hFAM222A, hFAM222B, laXP, and haXP in complex with the kinase domain of NLK (RMSD=0.570±0.047 Å), colored according to their pLDDT scores. (**E**) Representative PAE plot of the full-length interaction of hFAM222A with hNLK. Values close to 0 Å indicate high confidence in the spatial relationship between residues or domains. Red dashed boxes highlight the DCD222 region.

stable and highly conserved conformation upon binding to NLK, reinforcing its functional relevance. Moreover, this structural stability is further supported by a pLDDT above 90 across most of the domain, suggesting high confidence in the predicted local structure (Fig. 4F and Supplementary Fig. 4). Additionally, Predicted Aligned Error (PAE) values close to 0 Å indicate high accuracy in the structural alignment between DCD222 from humans and cyclostomes in its NLK-bound state (Fig. 4E and Supplementary Fig. 4). These findings suggest that the FAM222-NLK interaction is not only structurally and functionally significant but also ancestral and evolutionarily preserved across vertebrates, highlighting its potential role in conserved regulatory processes.

The DCD222 region contains a RevD motif

IDRs often contain short linear motifs (SLiMs), short amino acid sequences (8–23 residues) that mediate interactions with other proteins[40]. Given this, we hypothesized that the DCD222 domain might harbor SLiMs involved in its interaction with NLK. To investigate this, we analyzed the DCD222 sequence of hFAM222A as a query in both orientations: N-terminal to C-terminal and C-terminal to N-terminal, using the eukaryotic linear motif resource (ELM) (Fig. 5A)[41]. The results revealed that DCD222 contains a revD motif, an 8 to 12 residue sequence that binds to the docking groove and the common docking (CD) region of MAPKs[42]. These regions correspond to a hydrophobic groove and a negatively charged surface, respectively, and are exclusive to this family of kinases[11,43].

Interestingly, the revD motif in DCD222 is highly conserved from hagfish to humans and contains the signature sequence $\Phi X\Phi XXXXXXXR/K$, where Φ represents a hydrophobic residue, X any amino acid, and R/K an arginine or lysine, respectively (Fig. 5A). To assess how this motif interacts with NLK, we analyzed its structural integration within the DCD222 region. We found that L58, S59, I60, K61, and I62 directly interact directly with the docking groove, forming four hydrogen bonds. Furthermore, L58 and I60 insert deeply into hydrophobic pockets of the docking groove with pLDDT values ranging from 90 to 100, indicating high-confidence structural predictions (Fig. 5B). We found that the revD motif does not interact with the CD region or any negatively charged region in NLK, since our analysis confirmed that this MAPK lacks such a region (Supplementary Fig. 5).

Interestingly, we observed that this motif binds to NLK in the same manner across its orthologs, suggesting that this interaction mechanism has been evolutionarily conserved across all vertebrates (Fig. 5C).

Since the docking groove is an exclusive feature of MAPKs, we further tested whether the DCD222 domain could mediate interactions with other human MAPKs (MAPK1, MAPK3, MAPK4, MAPK6, MAPK7, MAPK8, MAPK9, MAPK10, MAPK11, MAPK12, MAPK13, and MAPK14) using AlphaFold3. The results showed that all ipTM values were below 0.6 (Fig. 5D and Supplementary Fig. 6), suggesting that FAM222 proteins interact specifically with NLK and not with other MAPKs through the DCD222 domain. To further support the biological relevance of this interaction, we examined RNA expression patterns at the single-cell level using the Human Protein Atlas[44] and found that human NLK, hFAM222A, and hFAM222B are expressed at high levels in overlapping populations of neuronal, glial, and germline cells (Supplementary Fig. 7). This expression pattern indicates that these proteins are present in the same cellular contexts, supporting their potential involvement in shared signaling pathways.

DCD222 binds to NLK through an extended conformation

To further investigate the interaction between the intrinsically disordered DCD222 domain and NLK, we conducted a detailed structural analysis. Despite its overall disordered nature, DCD222 consistently wraps around NLK with high pLDDT values across all AlphaFold3 models, from hagfish to humans (Supplementary Fig. 8 and Supplementary Videos 5 and 6). Within this flexible domain, the preformed α-helix H1 directly interacts with the C-lobe of NLK (Fig. 6A). In addition, the most hydrophobic segment, corresponding to the revD motif, engages the docking groove of NLK (Fig. 6B), highlighting how specific disordered regions can establish stable interfaces through SLiMs embedded within a flexible structural context.

The remaining portion of the domain associates with the N-lobe without directly engaging the active site. This interaction induces the formation and stabilization of a β-hairpin-like structure, with a pLDDT of 90, indicating high confidence in this prediction (Fig. 6C). In addition, we identified a previously unreported hydrophobic groove in the upper region of the N-lobe (Fig. 6D). Within this groove, DCD222 forms and inserts a short β-sheet composed of three residues from the conserved GLLAIV motif (Fig. 6E and Supplementary Fig. 9), which directly interacts with the C-terminal extension of NLK. This interaction, not observed with other MAPKs (Supplementary Fig. 6), reinforces the specificity of the DCD222-NLK complex. We also examined whether other regions of NLK contribute to binding, including the CMGC insert, but found no interaction in any of the AlphaFold3 models generated (Supplementary Fig. 9).

Based on these findings, we propose that DCD222 binds to NLK through an extended conformation, a model in which little to no structural changes occur upon binding, allowing the formation of a specific complex[45], a mechanism clearly observed in the DCD222-NLK interaction. These findings suggest that this extended conformation has been evolutionarily conserved in vertebrates.

The DCD222 region establishes an evolutionarily conserved interaction with NLK through multiple types of interactions

Our analysis identified the formation of 20 to 26 hydrogen bonds, 265 to 297 non-bonded contacts (including hydrophobic, electrostatic, and aromatic interactions) and one salt bridge between the DCD222 region and the kinase domain of NLK (Fig. 7A). The combination of these interactions suggests that the DCD222-NLK interaction is highly stable and structurally regulated.

Through the MSA (Fig. 7B), we determined that only 13 out of the 40 to 44 amino acid residues form hydrogen bonds and are highly conserved from hagfish to humans. These conserved residues are in five distinct
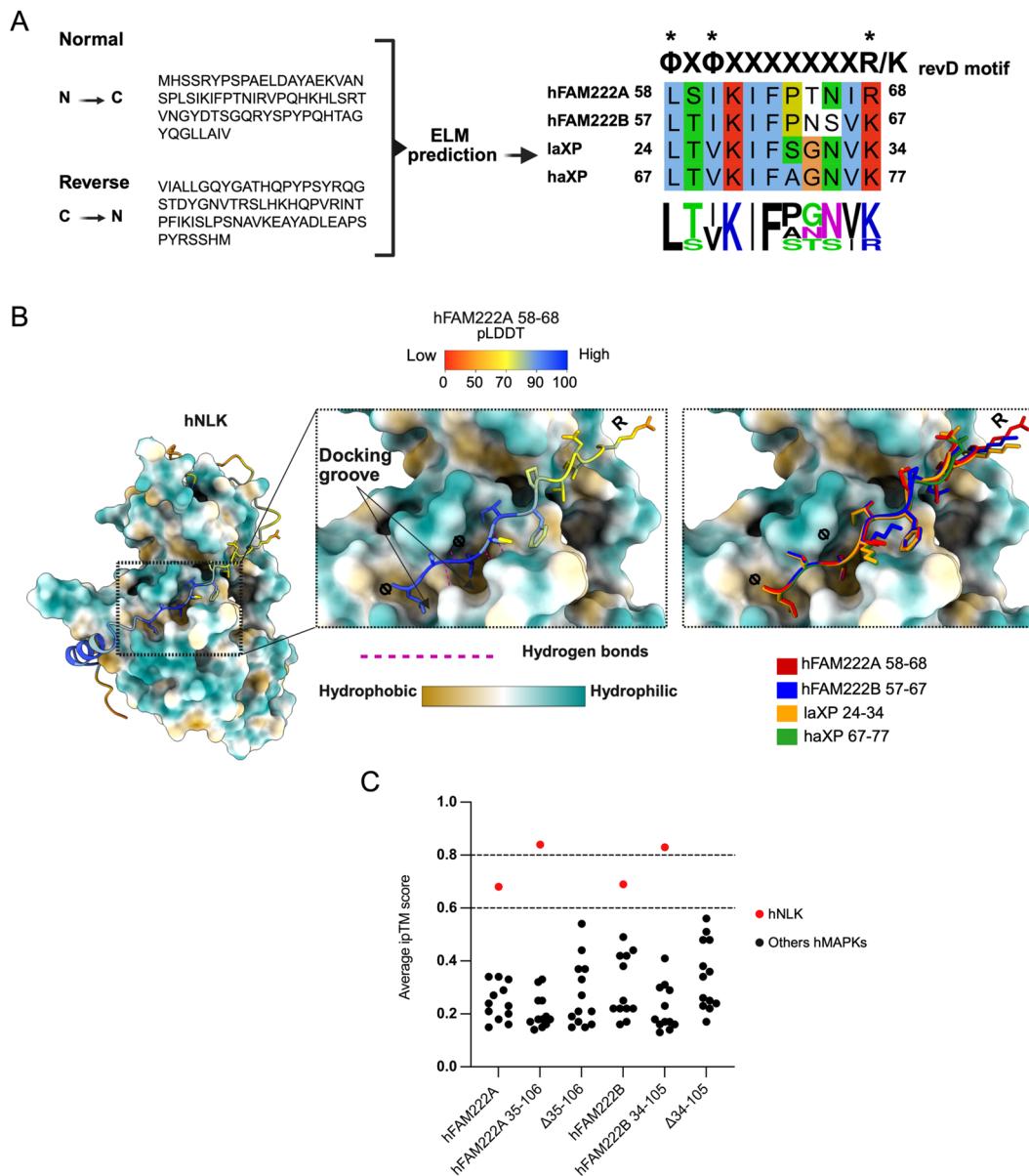
Fig. 5. Identification of the revD motif in the DCD222 region. (A) Prediction of SLiMs within the DCD222 domain of hFAM222A using the ELM server, along with their conservation in hFAM222B, laXP, and haXP. (**B** and **C**) Interaction of the revD-motif with the docking groove of NLK, where the surface hydrophobicity is represented using a color scale. (**D**) Average ipTM values for interactions between hFAM222 proteins, the DCD222 region, and its mutants with 12 additional human MAPKs, including NLK.

regions within DCD222, which may be essential for its interaction with NLK: the α-helix H1, the revD motif, the β-hairpin-like, the β-sheet-forming region, and the PQH/R motif. We observed that 8 of the 13 conserved residues form hydrogen bonds with NLK through the peptide backbone, while the remaining 5 residues interact via their side chains. Recently, it was reported that NLK can phosphorylate S59 of hFAM222A[15]. Interestingly, we found that phosphorylated S59 is part of the revD motif, however, its side chain does not form hydrogen bonds with NLK, whereas its peptide backbone does. This could suggest that this post-translational modification may not directly affect the binding of the DCD222 region to NLK (Supplementary Fig. 5).

We also observed that the amino acids within the kinase domain of NLK that interact with DCD222 are highly conserved. However, their conservation is comparable to that of the rest of the domain from hagfish to humans (Supplementary Fig. 10). Therefore, we conclude that there is no evidence of a co-evolutionary process between the DCD222 region and NLK.

Finally, to assess the dynamic stability of the complex, we performed three independent atomistic simulations of the hFAM222A–hNLK system using GROMACS 2025.3 (CUDA-enabled)[46,47] under standard isothermal–isobaric (NPT) conditions. The simulated systems remained thermodynamically stable throughout, with mean temperature and pressure fluctuations closely centered around target values (T ≈ 300.0 ± 0.33 K; P ≈ 1.0 ±
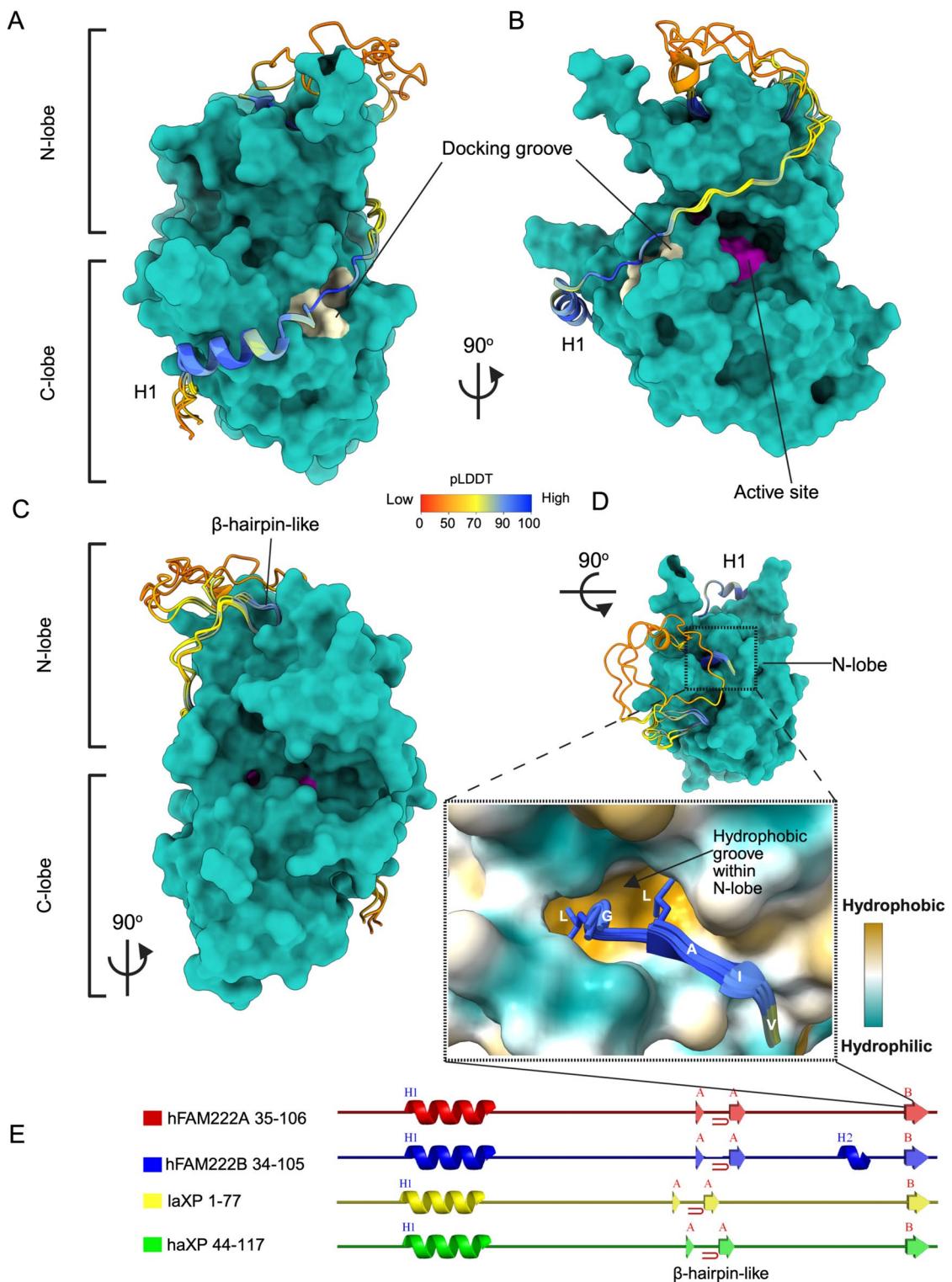
Fig. 6. Interaction of the DCD222 region with NLK. (**A–E**) Structural representations of the DCD222 region from hFAM222A, hFAM222B, laXP, and haXP (aligned structures) in complex with the kinase domain of NLK, including both the N-lobe and C-lobe. The lower panels display their secondary structure in a linear representation and their bound state to NLK. (**D** and **E**) Interaction of the GLLAIV motif with the hydrophobic groove located in the N-lobe of NLK.
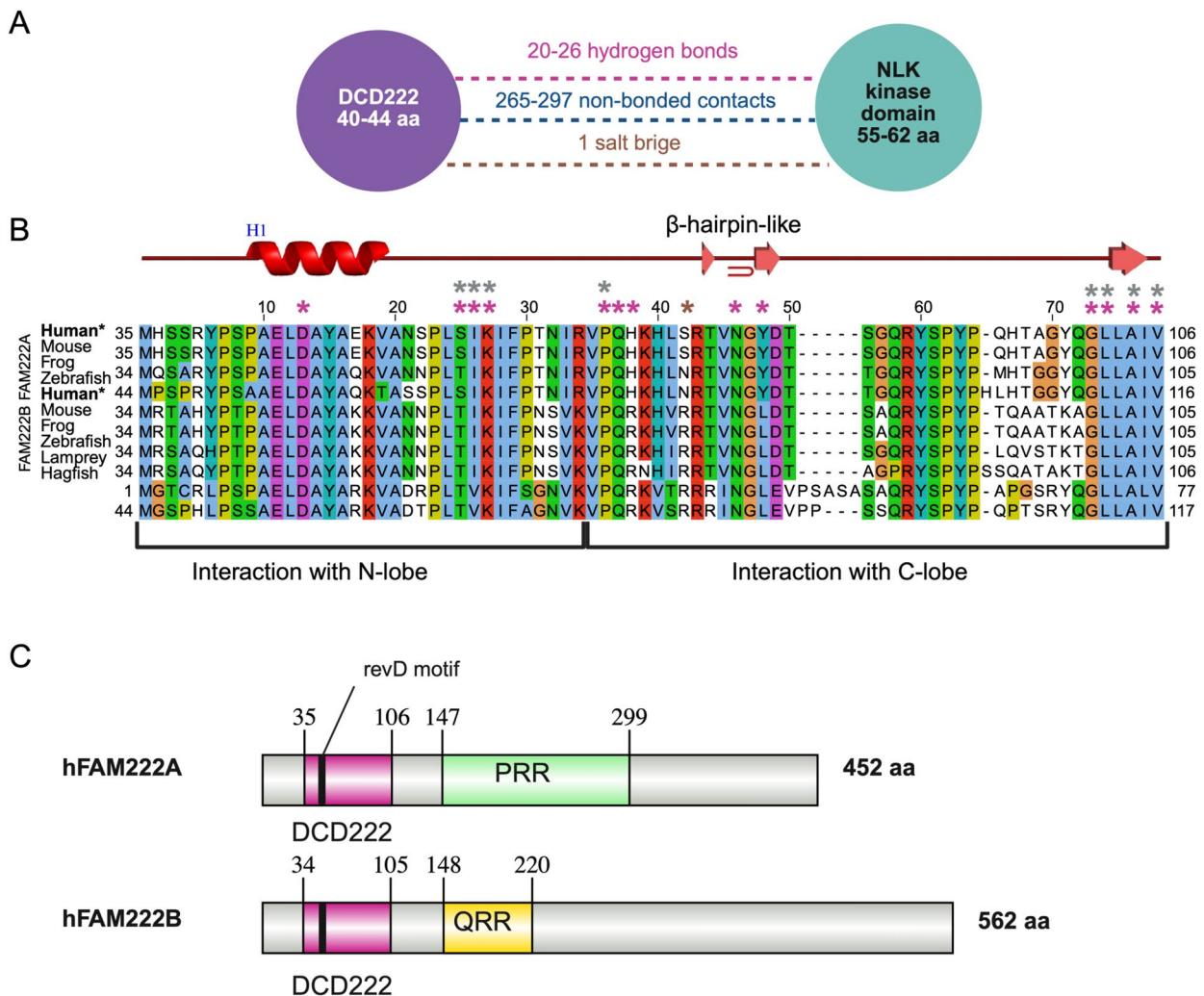
Fig. 7. Types of interactions between the DCD222 region and the NLK kinase domain. (**A**) Types of interactions observed in the DCD222-NLK complex, analyzed using PDBsum. (**B**) MSA of DCD222, including its secondary structure in the NLK-bound state. Pink asterisks indicate DCD222 residues that form hydrogen bonds with NLK and are conserved from hagfish to humans. Gray asterisks denote residues that form hydrogen bonds through their peptide backbone. In this scheme, amino acids are shown according to Jalview based on their chemical properties and sequence conservation: blue=hydrophobic residues, red=positively charged residues, magenta=negatively charged residues, green=polar residues, cyan=aromatic residues, pink=cysteines, orange=glycines, yellow=prolines, and white=non-conserved residues. (**C**) Linear representation of human FAM222 proteins, showing the DCD222 region, proline-rich regions, and glutamine-rich regions. This was generated using IBS 2.0[61].

30 bar). To quantify inter-chain stability and interface persistence, we computed the center-of-mass distance between the chains, the minimum heavy-atom distance, heavy-atom contact counts (< 0.45 nm), and inter-chain hydrogen-bond counts and lifetimes (Supplementary Table S1). Per-residue contact occupancies (hFAM222A) and residue-to-residue minimum-distance heatmaps were also generated. Additional analyses were performed on the intrinsically disordered hFAM222A, focusing on residues 43–53 (α-helix H1) and 35–106 (DCD222). Across all simulations, the complex remained continuously associated (bound fraction ca. 1.0; mean minimum distance of 0.162–0.170 nm), exhibiting extensive inter-chain hydrogen bonding (54, 50, and 56 bonds per frame in the three replicates, with lifetimes around 2.5–2.8 ns). The DCD222 segment formed the principal and most persistent interface (2.7–2.9 contacting residues on average) with low variance in contact count and long contact survival, consistent with a robust but moderately dynamic interface characteristic of IDP interactions. By contrast, α-helix H1 remained spatially proximal (minimum distance of 0.186–0.197 nm) but contributed relatively few direct contacts (all analyses were performed using PBC-corrected trajectories sampled every 1 ns). Overall, these simulations reveal that the hFAM222A–hNLK complex is maintained through a stable yet flexible interface dominated by the DCD222 region, a configuration emblematic of functional interactions mediated by IDPs.

Taken together, these findings support our proposal that the DCD222 region represents an intrinsically disordered domain (IDD) that specifically mediates the interaction with NLK (Fig. 7C). This interaction appears to be both ancestral and functionally relevant, suggesting that DCD222 has played a conserved role in regulating NLK function throughout vertebrate evolution.

Discussion

Our results reveal the evolution and functionality of the FAM222 protein family, highlighting the structural and functional conservation of the IDD named DCD222 and its interaction with NLK. The identification of FAM222A and FAM222B as products of a gene duplication event in the common ancestor of jawed vertebrates, together with the presence of a single copy ortholog in cyclostomes, supports the hypothesis that this protein family evolved through duplication events followed by functional specialization. The divergence between FAM222A and FAM222B in humans, with only 34.3% identity, suggests that these proteins have been subjected to different selective pressures following gene duplication. However, the conservation of the DCD222 region, with 62.5% identity from hagfish to humans, suggests strong evolutionary constraints that indicate a key function maintained throughout vertebrate evolution.

The characterization of DCD222 as a highly conserved IDD suggests that this region plays a role within the FAM222 family, facilitating dynamic interactions with other proteins in specific cellular processes. Additionally, the strong hydrophilic tendency observed in DCD222, along with the presence of key hydrophobic segments, indicates that this region could act as a flexible interface for molecular binding, allowing conformational adaptation that favors transient but specific interactions.

Structural modeling with AlphaFold3 revealed that, despite its disordered nature, DCD222 contains a highly conserved α-helix, designated H1, with pLDDT values above 90 across all species analyzed. The presence of this stable helix within a predominantly disordered region and our data suggest that it may act as an anchoring point in protein-protein interactions. This type of structural organization has been documented in other IDRs, where short helices and inducible secondary structures play a crucial role in binding specificity such as the NR-box motif[48]. The conservation of H1 from cyclostomes to humans indicates that this structure has been retained due to its functional importance, possibly providing stability to the interaction with NLK or facilitating its molecular recognition.

Recent studies have demonstrated that FAM222A and NLK interact through co-immunoprecipitation, co-immunofluorescence, and *in vitro* assays, where NLK was observed to phosphorylate FAM222A at S59[15]. The experimental confirmation of this interaction reinforces the model proposed in our structural analysis, suggesting that DCD222 is the region that binds to NLK. The conservation of this interaction in cyclostome orthologs indicates that the FAM222-NLK association is ancestral and has been maintained throughout vertebrate evolution. This evolutionary conservation is particularly relevant in the context of MAPKs, as many regulatory proteins in this kinase family contain highly specific IDRs, whose evolutionary retention is often related to essential functions in signaling pathway regulation[49].

A detailed analysis of the interaction between DCD222 and NLK revealed an extended conformation mechanism in which the IDD contacts the N-lobe forming an antiparallel β-sheet, and the C-lobe, interacting with the docking groove without engaging the active site. Consistent with this model, molecular dynamics simulations confirmed that the representative hFAM222A–hNLK complex remains stably associated through a persistent yet flexible interface dominated by the DCD222 region, a configuration typical of functional interactions mediated by IDPs. This interaction closely resembles that observed in the yeast STE5/FUS3 complex, where the 288–316 amino acid region of STE5 (a yeast scaffold protein) adopts the same topology upon interacting with FUS3 (a yeast MAPK), promoting allosteric activation[50]. Thus, our model could suggest that DCD222 not only provides binding specificity but may also modulate NLK activity.

The ability of IDRs to act as flexible binding modules has been extensively documented in the literature, and this type of interaction has been proposed to be particularly relevant in proteins involved in cellular signaling[45]. The identification of the revD motif within DCD222 and its specific interaction with the docking groove of NLK reinforces this hypothesis, suggesting that DCD222 could act as a specific regulator of NLK kinase activity. On the other hand, the FAM222 proteins could also function as scaffold proteins, with DCD222 mediating the interaction with NLK, while the other four α-helices observed in AlphaFold models of the FAM222 proteins (Fig. 3A) could facilitate interactions with other partners in signaling pathways.

Together, our findings suggest that the FAM222 protein family has evolved to specifically interact with NLK through the DCD222 region, whose flexible but functionally significant structure allows it to play a key role in the regulation of this kinase. The conservation of this interaction throughout vertebrate evolution indicates that DCD222 may modulate critical functions in NLK-dependent signaling pathways, which could have implications in various cellular processes, such as differentiation, stress response, and transcriptional regulation. Supporting this notion, data from The Human Protein Atlas indicate that NLK, FAM222A, and FAM222B could be co-expressed in neural, glial, and germline cells, with particularly strong expression observed in oligodendrocytes. This expression pattern suggests that, in addition to their physical interaction, NLK and FAM222 proteins may functionally converge in specific biological contexts. Furthermore, this study provides a solid foundation for future explorations of the evolution of IDRs and their contribution to the specificity of protein–protein interactions in vertebrate cellular signaling.

Methods
Phylogenetic analysis
The FAM222A sequence from *Homo sapiens* was retrieved from the NCBI database (NP_116218.2) and used as a query. A BLASTp search[51] was then performed on the NCBI server to identify homologous sequences, using

the RefSeq Protein database, with a maximum of 5000 target sequences, a BLOSUM45 substitution matrix, and an E-value threshold of 1×10^{-6}. From the retrieved sequences, a representative subset of 38 sequences from various vertebrates was selected (Supplementary material), including 20 annotated as FAM222A and 18 as FAM222B. These sequences were aligned with MAFFT on the EMBL web server[52] under default parameters. A phylogenetic tree was subsequently constructed using the maximum likelihood method in IQ-TREE[53], with an ultrafast bootstrap of 10,000 replicates to assess branch support. The resulting tree and MSA were visualized using iTOL[54] and Jalview 2.11.4.1[55].

Structural prediction and protein-protein interaction analysis

For protein structure modeling and evaluation of protein-protein interactions, we used AlphaFold3 through its web server[19], employing FASTA sequences obtained from NCBI and UniProt (Supplementary material). All predictions were made using default parameters. Additionally, we calculated the average ipTM from the five models generated for each interaction (Supplementary material) and subsequently visualized the data using GraphPad Prism 10.3.1. All structural models were analyzed using PDBsum[56], while Predicted Aligned Error (PAE) graphs were visualized with the PAE Viewer tool[57]. Protein structures and protein-protein interactions were further examined using ChimeraX 1.9[58]. For visualization purposes, the N-terminal residues 1–131 of NLK were omitted in all structural representations.

Disorder prediction

To predict intrinsically disordered regions, we employed the DR-BERT[59] tool using the amino acid sequence corresponding to the DCD222 domain (**Supplementary material**). The predicted disorder profiles were then compared across different sequences.

Hydropathy analysis

Hydropathy calculations were performed using the Kyte-Doolittle and GRAVY algorithms, based on FASTA sequences. The resulting hydropathy profiles were compared among the sequences analyzed.

Molecular dynamics simulations

The protein–protein complex was modelled using GROMACS 2025.3 (CUDA-enabled)[46,47] under standard isothermal–isobaric (NPT) conditions across three independent atomistic simulations.

Protein parameters were assigned according to the AMBER99SB-ILDN force field[60], while water molecules were represented using the TIP3P model. All systems were prepared with constraints applied to X–H bonds using the leap-frog integrator ($\Delta t = 2$ fs), Particle-Mesh Ewald electrostatics for long-range interactions, and a Verlet cutoff scheme (rcoulomb = rvdw = 1.0 nm). Temperature was maintained at 300 K using a V-rescale thermostat with separate coupling groups for the protein and the solvent/non-protein components ($\tau t = 0.1$ ps). Pressure was controlled using the Berendsen barostat during equilibration ($\tau p = 2$ ps), followed by the Parrinello–Rahman barostat during production ($\tau p = 5$ ps, Pref = 1 bar; isotropic coupling). Equilibration consisted of 0.5 ns in the NVT ensemble and 1.0 ns in the NPT ensemble. Each production trajectory was extended for 100 ns. The simulated systems remained thermodynamically stable throughout, with mean temperature and pressure fluctuations closely centered around target values (T ≈ 300.0 ± 0.33 K; P ≈ 1.0 ± 30 bar).

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. The AlphaFold model files generated in this work are available upon request from Dr. Mario Zurita ([mario.zurita@ibt.unam.mx].

References

1. Holehouse, A. S. & Kragelund, B. B. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat. Rev. Mol. Cell. Biol.* **25**, 187–211 (2024).
2. Orand, T. & Jensen, M. R. Binding mechanisms of intrinsically disordered proteins: insights from experimental studies and structural predictions. *Curr. Opin. Struct. Biol.* **90**, 102958 (2025).
3. van der Lee, R. et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
4. Pancsa, R. & Tompa, P. Structural disorder in eukaryotes. *PLoS One.* **7**, e34687 (2012).
5. Bondos, S. E., Dunker, A. K. & Uversky, V. N. Intrinsically disordered proteins play diverse roles in cell signaling. *Cell. Communication Signal.* **20**, 20 (2022).
6. Gógl, G., Kornev, A. P., Reményi, A. & Taylor, S. S. Disordered protein kinase regions in regulation of kinase domain cores. *Trends Biochem. Sci.* **44**, 300–311 (2019).
7. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell. Biol.* **16**, 18–29 (2015).
8. Gaestel, M. MAPK-Activated protein kinases (MKs): novel insights and challenges. *Front Cell. Dev. Biol* **3**, (2016).
9. Orand, T. et al. Bipartite binding of the intrinsically disordered scaffold protein JIP1 to the kinase JNK1. *Proceedings of the National Academy of Sciences* 122, e2419915122 (2025).
10. Mariño Pérez, L. et al. Structural basis of homodimerization of the JNK scaffold protein JIP2 and its heterodimerization with JIP1. *Structure* **32**, 1394–1403e5 (2024).
11. Tanoue, T., Yamamoto, T. & Nishida, E. Modular structure of a Docking surface on MAPK Phosphatases *. *J. Biol. Chem.* **277**, 22942–22949 (2002).
12. Singleton, M. D. & Eisen, M. B. Evolutionary analyses of intrinsically disordered regions reveal widespread signals of conservation. *PLoS Comput. Biol.* **20**, e1012028 (2024).
13. Consortium, T. U. UniProt: the universal protein knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617 (2025).

14. Yan, T. et al. FAM222A encodes a protein which accumulates in plaques in alzheimer's disease. *Nat. Commun.* **11**, 411 (2020).
15. Gao, J. et al. Aggregatin is a mitochondrial regulator of MAVS activation to drive innate immunity. *J. Immunol.* **214**, 238–252 (2025).
16. Spiegler, S. et al. FAM222B is not a likely novel candidate gene for cerebral cavernous malformations. *Mol. Syndromol.* **7**, 144–152 (2016).
17. Satoh, J., Kino, Y., Yanaizu, M., Ishida, T. & Saito, Y. Reactive astrocytes express Aggregatin (*FAM222A*) in the brains of Alzheimer's disease and Nasu-Hakola disease. *Intractable Rare Dis Res* 9, 217–221 (2020).
18. Tzani, A. et al. FAM222A, part of the BET-Regulated basal endothelial Transcriptome, is a novel determinant of endothelial biology and Angiogenesis—Brief report. *Arterioscler. Thromb. Vasc Biol.* **44**, 143–155 (2024).
19. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* **630**, 493–500 (2024).
20. Ishitani, T. et al. Nemo-like kinase suppresses Notch signalling by interfering with formation of the Notch active transcriptional complex. *Nat. Cell. Biol.* **12**, 278–285 (2010).
21. Masoumi, K. C. et al. NLK-mediated phosphorylation of HDAC1 negatively regulates Wnt signaling. *Mol. Biol. Cell.* **28**, 346–355 (2016).
22. Ke, H. et al. Nemo-like kinase regulates the expression of vascular endothelial growth factor (VEGF) in alveolar epithelial cells. *Sci. Rep.* **6**, 23987 (2016).
23. Ohkawara, B. et al. Role of the TAK1-NLK-STAT3 pathway in TGF-β-mediated mesoderm induction. *Genes Dev.* **18**, 381–386 (2004).
24. Ota, S. et al. NLK positively regulates Wnt/β-catenin signalling by phosphorylating LEF1 in neural progenitor cells. *EMBO J.* **31**, 1904–1915 (2012).
25. Yuan, H. X. et al. NLK phosphorylates raptor to mediate stress-induced mTORC1 Inhibition. *Genes Dev.* **29**, 2362–2376 (2015).
26. Besli, N. & Yenmis, G. Assessment of the interaction of Aggregatin protein with Amyloid-Beta (Aβ) at the molecular level via in Silico analysis. *Acta Chim. Slov.* **67**, 1262–1272 (2020).
27. de Castro, E. et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365 (2006).
28. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
29. Basu, S. & Bahadur, R. P. Do sequence neighbours of intrinsically disordered regions promote structural flexibility in intrinsically disordered proteins? *J. Struct. Biol.* **209**, 107428 (2020).
30. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. & Obradović, Z. Intrinsic disorder and protein function. *Biochemistry* **41**, 6573–6582 (2002).
31. Alderson, T. R., Pritišanac, I., Kolarić, Đ., Moses, A. M. & Forman-Kay, J. D. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proceedings of the National Academy of Sciences* 120, e2304302120 (2023).
32. Ruff, K. M & Pappu, R. V. AlphaFold and implications for intrinsically disordered proteins. *J. Mol. Biol.* **433**, 167208 (2021).
33. Szklarczyk, D. et al. The STRING database in 2025: protein networks with directionality of regulation. *Nucleic Acids Res.* **53**, D730–D737 (2025).
34. Huttlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).
35. Huttlin, E. L. et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022–3040e28 (2021).
36. Varjosalo, M. et al. The protein interaction landscape of the human CMGC kinase group. *Cell. Rep.* **3**, 1306–1320 (2013).
37. Shi, P. et al. SUMOylation of DDX39A alters binding and export of antiviral transcripts to control innate immunity. *J. Immunol.* **205**, 168–180 (2020).
38. Whiteley, A. M. et al. Global proteomics of < em > Ubqln 2-based murine models of ALS. *Journal Biol. Chemistry* **296**, (2021).
39. Luck, K. et al. A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
40. Cermakova, K. & Hodges, H. C. Interaction modules that impart specificity to disordered protein. *Trends Biochem. Sci.* **48**, 477–490 (2023).
41. Kumar, M. et al. ELM—the eukaryotic linear motif resource—2024 update. *Nucleic Acids Res.* **52**, D442–D455 (2024).
42. Garai, Á. et al. Specificity of linear motifs that bind to a common Mitogen-Activated protein kinase Docking groove. *Sci. Signal.* **5**, ra74–ra74 (2012).
43. Nguyen, T., Ruan, Z., Oruganty, K., Kannan, N. & Co-Conserved, M. A. P. K. Features couple D-Domain Docking groove to distal allosteric sites via the C-Terminal flanking tail. *PLoS One.* **10**, e0119636 (2015).
44. Digre, A. & Lindskog, C. The human protein atlas—Integrated omics for single cell mapping of the human proteome. *Protein Science* 32, (2023).
45. Milles, S., Salvi, N., Blackledge, M. & Jensen, M. R. Characterization of intrinsically disordered proteins and their dynamic complexes: from in vitro to cell-like environments. *Prog Nucl. Magn. Reson. Spectrosc.* **109**, 79–100 (2018).
46. Pronk, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
47. Abraham, M. J. et al. High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1–2.* **GROMACS**, 19–25 (2015).
48. Plevin, M. J., Mills, M. M. & Ikura, M. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem. Sci.* **30**, 66–69 (2005).
49. Zeke, A. et al. Systematic discovery of linear binding motifs targeting an ancient protein interaction surface on MAP kinases. *Mol. Syst. Biol.* **11**, 837 (2015).
50. Bhattacharyya, R. P. et al. The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Sci. (1979).* **311**, 822–826 (2006).
51. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
52. Madeira, F. et al. The EMBL-EBI job dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* **52**, W521–W525 (2024).
53. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
54. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
55. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
56. Laskowski, R. A. et al. Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134 (2018).
57. Elfmann, C. & Stülke, J. PAE viewer: a webserver for the interactive visualization of the predicted aligned error for multimer structure predictions and crosslinks. *Nucleic Acids Res.* **51**, W404–W410 (2023).
58. Meng, E. C. et al. UCSF chimerax: tools for structure Building and analysis. *Protein Sci.* **32**, e4792 (2023).
59. Nambiar, A., Forsyth, J. M., Liu, S. & Maslov, S. DR-BERT: A protein Language model to annotate disordered regions. *Structure* 32, 1260–1268e3 (2024).
60. Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the amber ff99SB protein force field. *Proteins Struct. Funct. Bioinform.* **78**, 1950–1958 (2010).

61. Xie, Y. et al. IBS 2.0: an upgraded illustrator for the visualization of biological sequences. *Nucleic Acids Res.* **50**, W420–W426 (2022).

Acknowledgements

Author contributions

Eduardo Calvario generated the models, analyzed all the data presented, and wrote the manuscript. Lorenzo Segovia assisted with data analysis and manuscript review. Mario Zurita conceived the project, reviewed the manuscript, and funded the project.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-28838-1.

Correspondence and requests for materials should be addressed to L.S. or M.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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