STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 70, 2, 2025 (p. 115-150)
doi:10.24193/subbbiol.2025.2.08

Evolutionary patterns of structural disorder and post-translational modifications in the 18.5 kDa myelin basic protein

Ilka Koszorus[1 *] and Ferencz Kósa[1,2]

[1]*Affiliation: Babeş-Bolyai University, Hungarian Department of Biology and Ecology, Cluj-Napoca, Romania;*
[2]*Babeş-Bolyai University, Center for Systems Biology, Biodiversity and Bioresources, Sociobiology and Insect Ecology Lab, Cluj-Napoca, Romania;*
[*] ***Corresponding author, E-mail: koszorusilka@gmail.com.***

Article history: Received 31 August 2025; Revised 4 December 2025; Accepted 05 December 2025; Available online 20 December 2025

Abstract. Myelin basic protein (MBP, 18.5 kDa isoform) is a key structural component of the myelin sheath, where it drives multilayer compaction through electrostatic interactions and dynamic conformational transitions. Despite its functional importance, a comprehensive understanding of MBP's evolutionary patterns of intrinsic disorder, post-translational modifications (PTMs), and sequence-derived properties across vertebrates have been lacking. Here, we analyzed MBP consensus sequences from six major vertebrate clades (*Chondrichthyes, Teleostei, Amphibia, Reptilia, Aves, Mammalia*) using an integrated bioinformatic framework combining intrinsic disorder predictions, Shannon entropy-based complexity profiling, hydrophobic moment (μH) analyses, net charge per residue (NCPR) patterns, and experimentally supported PTM mapping.

Our results reveal that MBP maintains a highly conserved intrinsically disordered architecture characterized by long N- and C-terminal IDRs and several clade-specific central IDRs. Teleosts exhibit a truncated N-terminal, lacking the first 15 residues, but compensate through additional positively charged residues downstream, preserving membrane-binding potential. Amphibians show unique insertions enriched in basic residues, leading to the longest MBPs and potentially enhanced lipid interactions. Shannon entropy and μH profiles demonstrate alternating conserved

α-helices and flexible IDRs that overlap with PTM hotspots, particularly phosphorylation and citrullination sites, suggesting dynamic regulatory roles. NCPR analyses highlight a conserved electrostatic topology composed of alternating basic clusters and acidic/neutral dips, balancing reversible membrane adhesion with controlled aggregation. Together, these findings demonstrate that MBP combines strong structural conservation with lineage-specific adaptations in intrinsic disorder, charge distribution, and PTM patterning. This evolutionary flexibility likely underpins MBP's ability to support functional diversity in myelin architecture while maintaining its essential role in vertebrate nervous system evolution.

Keywords: hydrophobic moment, internally disordered region, myelin basic protein, net charge per residue, sequence complexity.

Introduction

The prevailing paradigm in molecular biophysics and structural biology states that a protein's amino acid sequence determines its three-dimensional structure, and that structure dictates its function - the *one sequence–one structure–one function* model. This view has been challenged by the discovery of intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs). IDPs lack a stable three-dimensional structure under physiological conditions, whereas IDRs are unstructured segments within otherwise folded proteins (Uversky *et al*., 2013). Rather than adopting a fixed conformation, these proteins exist as dynamic ensembles of rapidly interconverting states, allowing them to remain functional in multiple cellular contexts (Oldfield *et al.*, 2019). Some proteins are entirely disordered, while others contain one or more discrete IDRs, often referred to as multi-IDR proteins (Lobley *et al.*, 2005).

IDPs and IDRs are widespread across all domains of life. According to the latest version (2025) of the DisProt database (https://disprot.org), 3,200 IDPs and 9,365 IDRs have been annotated, encompassing 38 distinct biological functions. In eukaryotic genomes, disordered segments are particularly common: for example, 44% of human protein-coding genes contain disordered regions longer than 30 amino acids (van der Lee *et al*., 2014). Computational analyses indicate that 45–50% of eukaryotic proteins possess long disordered segments, whereas in archaea and bacteria, only 7–30% of proteins contain such regions (Xue *et al*., 2013).

The functional significance of intrinsic disorder lies in its ability to expose large and adaptable interaction surfaces that enable recognition of diverse

ligands (Lobley *et al.,* 2005). Consequently, many IDPs are multifunctional and play central roles in molecular recognition, signaling, and regulation (Wright and Dyson, 2014) and they participate in essential cellular processes such as transcriptional and translational control, cell-cycle regulation, and signal transduction (Xue *et al., 2013*).

IDRs are characterized by a high net charge and low overall hydrophobicity, as well as low sequence complexity compared with folded proteins (van der Lee *et al*., 2014). They are strongly depleted in bulky hydrophobic and aromatic residues (tryptophan, tyrosine, phenylalanine) and large aliphatic residues such as valine, while being enriched in charged and polar residues including lysine, arginine, glutamate, glutamine, and serine (Oldfield *et al.,* 2019).

IDPs are also strongly linked to disease, including cancer, cardiovascular disorders, amyloidoses, neurodegenerative diseases, and diabetes (Babu, 2011; Xue *et al*., 2013).

Myelin basic protein (MBP) is the second most abundant component of central nervous system (CNS) myelin, accounting for approximately 30% of its dry protein mass (Kister and Kister, 2023). Its best-established function is to promote adhesion of myelin membrane layers, driving the formation of compact sheaths (Raasakka and Kursula, 2020). Beyond its structural role, MBP has been implicated in the pathogenesis of multiple sclerosis (MS), where increased deimination (citrullination) contributes to myelin destabilization, and MBP may also act as a candidate self-antigen in the autoimmune response (Libich and Harauz, 2008). MBP is generated by alternative splicing of a single mRNA transcript, giving rise to four major isoforms in humans, with molecular masses of 21.5, 20.2, 18.5, and 17.2 kDa (Harauz *et al*., 2004). The classic 18.5 kDa isoform is the predominant form in adult CNS myelin and plays an essential role in maintaining sheath stability (De Avila *et al.*, 2014).

Spectroscopic and sequence-based analyses classify all known MBP isoforms as intrinsically disordered proteins (Libich and Harauz, 2008). This disorder underlies MBP's interactions with a range of partners, including Fyn kinase, cytoskeletal proteins such as actin and tubulin, calmodulin in a Ca^{2+}-dependent manner, and proteins containing SH3 domains (De Avila *et al.,* 2014; Raasakka and Kursula, 2020;). MBP function is further modulated by extensive post-translational modifications (PTMs). Among these, serine/threonine phosphorylation is reversible, whereas arginine deimination is irreversible and can critically alter MBP's stability and interactions (Raasakka and Kursula, 2020).

MBP family members are broadly distributed among vertebrates (Xue *et al.* 2013), and MBP itself is highly conserved across species (Raasakka and Kursula, 2020). Yet, despite intensive research, the relationship between MBP's intrinsic disorder and its strong sequence conservation remains unclear.

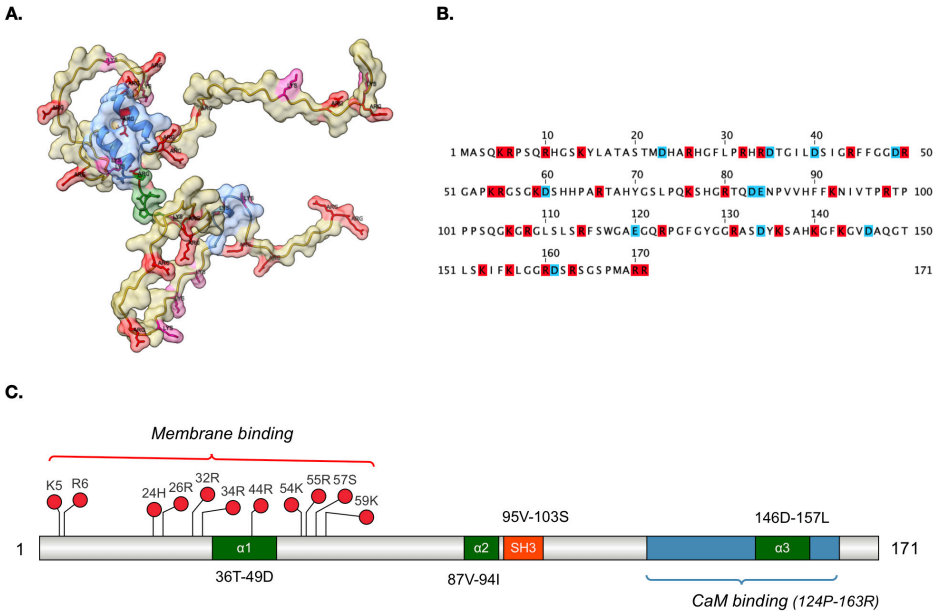

Figure 1. Structural and functional features of the human 18.5 kDa myelin basic protein (MBP, UniProt ID: P02686.3). The 18.5 kDa MBP isoform is depicted in all panels (A–C). (A) Three-dimensional structure predicted by AlphaFold (AF-P02686-F1-model_v4; https://alphafold.ebi.ac.uk/entry/P02686) and visualized with ChimeraX (Pettersen *et al.,* 2021). The α-helices involved in membrane binding (α1–3) are highlighted in blue, arginine residues in red, lysines in purple, and the SH3 domain (Polverini *et al*., 2008) in green. (B) Amino acid sequence of the 18.5 kDa MBP isoform (UniProt ID: P02686.3), with positively charged residues (Arg, Lys) marked in red and negatively charged residues (Asp, Glu) in blue. (C) Schematic representation of the 18.5 kDa MBP isoform sequence with annotated functional regions. Membrane-binding residues are indicated above the sequence, α-helices (α1–3) are marked in green (De Avila, 2014), the SH3-binding motif is shown in orange, and the calmodulin (CaM)-binding region is highlighted in blue (Libich and Harauz, 2008).

Despite numerous proteome-wide studies investigating the overall degree and evolutionary dynamics of intrinsic disorder across species (Ward *et al.*, 2004; Schad *et al.*, 2011; Necci *et al.*, 2016; Zarin *et al*., 2019; Kastano *et al.,* 2020; Singleton and Eisen, 2024; Mughal *et al.*, 2025), relatively few have addressed how the intrinsic disorder of a specific intrinsically disordered protein (IDP) and the parameters governing it have evolved (Siltberg-Liberles, 2011; Xue *et al*., 2013; Siddiqui *et al.*, 2016). In this study, we systematically examine the 18.5 kDa isoform of myelin basic protein (MBP) to explore evolutionary changes

in its disorder-related features, including the degree and extent of intrinsic disorder, the distribution, length, and number of intrinsically disordered regions (IDRs), and key structural characteristics such as sequence complexity, net charge per residue (NCPR), and hydrophobic moment profiles across vertebrates. Additionally, we investigate potential associations between evolutionary shifts in these disorder-related parameters and the distributional and numerical patterns of functionally critical post-translational modifications (PTMs), particularly citrullination and phosphorylation.

Matrials and methods

Retrieval of myelin basic protein amino acid sequence

MBP protein sequences were obtained via BLASTp searches against the non-redundant (nr) protein database using the NCBI BLAST tool (https://blast.ncbi.nlm.nih.gov), with the 18.5 kDa isoform as the query. The search was specifically aimed at identifying orthologous sequences. To ensure lineage-specific coverage, we used distinct representative query sequences for each major vertebrate clade: *Danio rerio* (UniProt ID: XP_005157892.1) for bony fish (*Teleostei*), selecting only mbpa sequences and excluding mbpb isoforms due to their distinct functional roles; *Leucoraja erinacea* (UniProt ID: AAA96756.1) for cartilaginous fish (*Chondrichthyes*), *Xenopus laevis* (UniProt ID: P87346) for amphibians, *Gekko japonicus* (UniProt ID: Q5I1E1) for reptiles, *Gallus gallus* (UniProt ID: P15720-1) for birds, and *Homo sapiens* (UniProt ID: P02686) for mammals. BLASTp searches were performed with the following parameters: maximum number of target sequences: 250, expect threshold (E-value): 0.05, and the BLOSUM62 substitution matrix. Retrieved sequences were filtered based on the following homology criteria: E-value $<10^{-5}$and 100% query coverage (Riley *et al*., 2023). To ensure broad phylogenetic representation, sequences were selected from all major taxonomic orders within each vertebrate group. In total, 199 MBP sequences were curated, with clade-specific distribution (cartilaginous fishes: 15, teleosts: 36, amphibians: 18, reptiles: 31, birds: 49, mammals: 50).

Sequence alignment and consensus sequence determination

Multiple sequence alignments (MSAs) were performed separately for each clade using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters (Madeira *et al.,* 2024). Sequences with alignment issues were manually removed. Aligned sequences were visualized and analyzed in

Jalview (version 2.11.2.5), and consensus sequences were generated using a 0.5 conservation threshold (Waterhouse *et al.,* 2009). These consensus sequences were then used for downstream analyses, including the prediction of conserved post-translational modification (PTM) motifs and intrinsically disordered regions (IDRs), calculation of Shannon entropy, generation of linear net charge per residue (NCPR) profiles, and hydrophobic moment (μH) analysis. The number of sequences used per clade ranged from 15 to 50. Although most clades met the recommended threshold of 20–50 sequences for high-quality MSAs, some included fewer due to the limited availability of homologous, full-length sequences. Nonetheless, alignments were manually curated to ensure accuracy before consensus generation.

Conservation scores displayed beneath the alignments (Fig. 3) in Jalview were also generated within the software, using the AMAS method of multiple sequence alignment analysis, which quantifies the conservation of physico-chemical properties for each alignment column (Livingstone and Barton, 1993).

A phylogenetic tree was generated using the neighbor-joining (NJ) method in MEGA version 12 for reference purposes: the protein alignment yielded a tree that reflects the known relationships among vertebrates. The tree was not used for subsequent analyses and is provided in the supplementary materials for completeness (see supplementary files: Fig. S1).

Assessment of protein disorder and identification of intrinsically disordered regions

To systematically analyze the intrinsic disorder propensity of MBP sequences across lineages, we applied several disorder prediction tools, each of which rely on distinct principles and vary in their outputs, being optimized for identifying IDRs of different lengths (Liu *et al.*, 2019). Given the difficulty in precisely defining IDR boundaries - even with experimental methods (Jensen *et al*., 2013) - and observed variability in IDR detection across predictors (Yruela *et al.*, 2017), we selected and combined tools to maximize sensitivity across a broad range of disordered features. Current prediction accuracy for IDRs up to 30 residues is estimated at roughly 70% (Monastyrskyy *et al*., 2014).

Our primary analysis was performed using the CAID Prediction Portal (https://caid.idpcentral.org/portal), which runs standardized, benchmarked intrinsic disorder predictors on the input FASTA amino acid sequences (Del Conte *et al.*, 2023). Specifically, we selected a set of complementary predictors from the CAID suite, including AUCpreD (Wang *et al*., 2016), ESpritz-N, ESpritz-X and ESpritz-D (Walsh *et al.,* 2012), IUPred3 (Erdős and Dosztányi, 2020), MobiDB Lite (Necci *et al.,* 2017), PredIDR long and PredIDR short (Xie *et al.,* 2022), to ensure the detection of both short and long intrinsically disordered

regions (IDRs). The CAID portal also generated a consensus annotation of disordered region boundaries, which we used to define the number and extent of disordered regions across the MBP clade-specific consensus sequences. To complement the CAID results, we additionally used the PONDR VLXT predictor (https://www.pondr.com), which is particularly sensitive to short disordered regions and terminal flexibility (Xue *et al.,* 2010). All algorithms output a number between 0 and 1 for each amino acid residue. If the predicted value exceeds or equals 0.5, this residue is considered disordered. All predictions were performed using default parameters, and the output scores were integrated to identify conserved disordered segments across MBP orthologs.

Furthermore, we calculated the predicted percent of disorder, defined as the ratio of disordered residues to the total number of residues, expressed as a percentage (van Bibber *et al.*, 2020).

Sequence complexity profile (Shannon entropy score profile)

Sequence complexity profile (Shannon entropy) for the clade specific consensus sequences was calculated according to Sen *et al.,* 2019 using a custom R script. Then Shannon entropy was computed for each MBP sequence within the six vertebrate clades. Calculations were performed in R 4.5.1 (R Core Team, 2024) using RStudio (Posit Software, 2024) and the seqinr package for sequence handling. A sliding window of 15 residues was applied, and the mean entropy value per sequence was obtained. Resulting Shannon entropy profiles were visualized using custom scripts with the ggplot2 package (Wickham, 2016).

Hydrophobic moment profile

The hydrophobic moment (μH) was calculated to quantify the amphipathic potential of the sequences using the hmoment site from EMBOSS Explorer with the default parameters (https://www.bioinformatics.nl/cgi-bin/emboss/hmoment) (Eisenberg *et al.,* 1984). This initial analysis was performed on representative consensus MBP sequences from each clade. Hydrophobic moment profiles were visualized in R (R Core Team, 2024) using custom scripts and the ggplot2 package (Wickham, 2016).

To complement the hydrophobic moment (μH) calculations we developed a custom R script to analyze a larger dataset of MBP sequences from all available representatives within each clade. FASTA sequences were imported and processed using the seqinr package (Charif & Lobry, 2007) in R version 4.5.1 (R Core Team, 2024) running in RStudio (Posit Software, 2024). The hydrophobic moment was calculated with a sliding window approach, employing a window size of 10 residues and the Eisenberg hydrophobicity scale as reference values (Eisenberg

et al., 1984). Angular periodicities of 100° and 160° per residue were analyzed to capture α-helical and β-strand segment tendencies, respectively. Hydrophobic moments were calculated for each sliding window by vectorially summing hydrophobicity values weighted by their angular positions, then averaged per sequence for statistical analysis.

Net charge per residue (NCPR) profiles

Net charge per residue (NCPR) was calculated for each MBP sequence within each vertebrate clade using R (R Core Team, 2024) in RStudio (Posit Software, 2024). Amino acids were assigned charges at physiological pH (Asp, Glu: –1; Lys, Arg: +1; His: +0.1), and NCPR was computed using a sliding window of five residues (Das and Pappu, 2013). Average NCPR per sequence was subsequently determined.

Prediction of phosphorylation sites

To predict conserved phosphorylation sites in MBP across vertebrate clades, we analyzed the clade-specific consensus sequences using three independent phosphorylation site predictors: DEPP, Musite, and NetPhos (Wang *et al.*, 2017). All predictions focused on serine, threonine, and tyrosine residues, and only sites scoring ≥0.5 in at least two of the three tools were retained.

In addition to *in silico* predictions, we also reviewed the literature for experimentally validated phosphosites in human 18.5 kDa MBP (Kishimoto *et al.* 1985; Harauz *et al.*, 2004). Using these datasets, we tested the performance of our three prediction methods in identifying experimentally confirmed phosphorylation sites. The integration of multiple prediction methods and empirical data allowed for high-confidence identification of functionally relevant phosphorylation motifs.

Prediction of citrullination sites

Although several citrullination site predictors such as ModPred, iCitr-PseAAC, and CKSAAP_CitrSite exist, at the time of writing this manuscript (July 2025), none of these tools were available or functional for our analyses. Citrullinated arginine residues in human MBP were collected from published experimental studies (Harauz and Musse, 2006; Polverini *et al.,* 2010). The human sequence was aligned to clade-specific consensus sequences, and citrullination sites were mapped across clades based on positional correspondence within the multiple sequence alignment. Only aligned arginine residues corresponding to modified positions in human MBP were considered.

Statistical analyses

All sequence-derived features (hydrophobic moment, Shannon entropy, and NCPR) were analyzed in R version 4.5.1 (R Core Team, 2024) using RStudio (Posit Software, 2024). FASTA sequences were processed with the seqinr package (Charif and Lobry, 2007), and data manipulation employed dplyr (Yarberry, 2021). Calculations were performed at both the sequence level (averaging per sequence within each clade) and at the consensus level (using clade-specific consensus sequences). Clade-wise mean values were compared using Welch's one-way ANOVA, appropriate for unequal variances and non-normal distributions (Welch, 1951). When global effects were significant, pairwise post hoc comparisons were carried out with Welch's t-tests and multiple-testing corrections. Data visualization was performed with custom R scripts using *ggplot2* (Wickham, 2016).

Domain structure visualization

Schematic representations of protein domain structures, post-translational modifications (PTMs) and intrinsically disordered regions (IDRs) were created using the IBS 2.0 (Illustrator for Biological Sequences) tool (Liu *et al.*, 2022; https://ibs.renlab.org).

Results and Discussion

Sequence-based analysis of disorder propensity

Disorder propensity profiles (Fig. 2) were predicted using clade-specific consensus sequences derived from multiple sequence alignments of 18.5 kDa myelin basic protein (MBP) orthologs. A total of nine complementary disorder prediction algorithms were employed, and their outputs were integrated based on predictor consensus: a residue was classified as disordered if more than 62.5% of the algorithms consistently identified it as disordered at that position. The combined use of multiple prediction algorithms enabled estimation of the predicted percent disorder per sequence, as well as the prediction of the number, boundaries, lengths, and distribution of conserved intrinsically disordered regions (IDRs) within each clade-specific consensus sequence (Figs. 2 and 3).

The predicted percent disorder was calculated as the ratio of disordered residues to the total number of residues, expressed as a percentage (van Bibber *et al.*, 2020). Across clades, this metric ranged from 52.22% to 86.11% (Table 1). To assess potential evolutionary trends, Spearman rank correlation analyses were conducted and complemented by permutation-based significance testing. While the analysis revealed a negative trend ($\rho = -0.54$), the correlation did not reach statistical significance ($p = 0.30$).

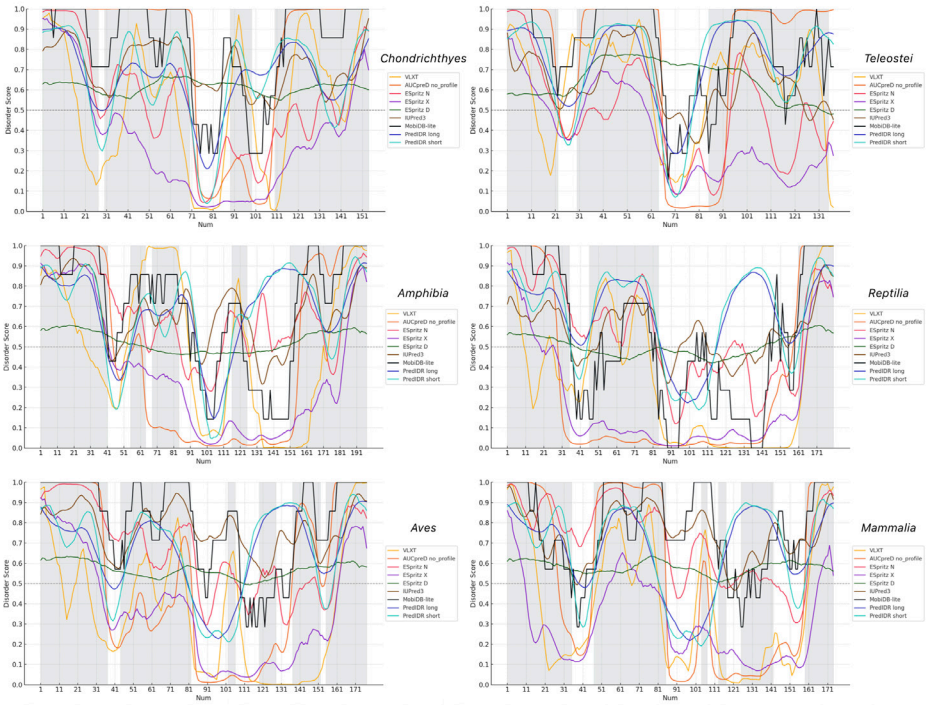

Figure 2. Intrinsically disordered region (IDR) predictions for clade-specific consensus MBP sequences. Disorder propensity was calculated for consensus MBP sequences of six major vertebrate lineages (*Chondrichthyes, Teleostei, Amphibia, Reptilia, Aves,* and *Mammalia*). Per-residue disorder scores (0–1) were obtained using complementary prediction tools from the CAID portal (AUCpreD, ESpritz-N/X/D, IUPred3, MobiDB Lite, PredIDR-long, PredIDR-short) and the PONDR VLXT predictor. A disorder threshold of 0.5 (horizontal reference) was applied to define disordered residues. Shaded background areas indicate regions of consensus disorder (IDRs) across predictors.

The predicted percent disorder has been reported to depend on protein chain length (Xie *et al.*, 2007; Mughal and Caetano-Anollés, 2025), except in viral proteomes, where it decreases with increasing chain length (Xie *et al.*, 2007). In the case of vertebrate MBPs, however, we found no significant correlation between chain length variation and predicted percent disorder values: fishes, which possess the shortest MBP isoforms, exhibited the highest disorder levels. Although the predicted percent disorder generally increases during evolution (Kastano *et al.*, 2020; Mughal and Caetano-Anollés, 2025) and tends to rise with organismal complexity — commonly measured by the number of cell types or overall proteome size (Schad *et al.*, 2011; Kastano *et al.*, 2020) — no such relationships were observed for vertebrate MBPs, likely because the previously reported

correlations were derived from analyses encompassing a much broader taxonomic spectrum.

Following the classification system proposed by Gsponer *et al. (2008) and* Rajagopalan *et al.* (2011), which distinguishes between highly ordered (0–10% disordered), moderately disordered (11–30%), and highly disordered (31–100%) proteins, our results show that all MBP consensus sequences across vertebrate clades fall into the highly disordered category. Given that MBP exhibited a high degree of disorder even in the earliest vertebrate lineages, these findings support the hypothesis that extensive intrinsic disorder is functionally important for MBP activity.

Identification and distribution of intrinsically disordered regions

Analysis of the intra-sequence distribution patterns of predicted disordered regions (Figs. 2 and 3) revealed a consistent architecture across all vertebrate clades: both an N-terminal and a C-terminal intrinsically disordered region (IDR) were invariably present. This terminal disorder pattern mirrors that of other intrinsically disordered proteins (IDPs) (van der Lee *et al.*, 2014; Necci *et al.*, 2016; de Vries *et al.*, 2024). Additionally, every MBP contains a variable number of internal IDRs, typically ranging from three to six per consensus sequence (Table 1, Fig. 2, 3 and 7).

Table 1. Intrinsic Disorder, Sequence Complexity, Net Charge Per Residue and Post-Translational Modification Predictions for the MBP 18.5 kDa Isoform in Vertebrate Lineages.

	Chondrich-thyes	***Teleostei***	***Amphibia***	***Reptilia***	***Aves***	***Mammalia***
Consensus sequence lenght (aa)	147	137	197	180	177	174
Predicted percent of disorder (%)	86.11	78.83	63.45	52.22	73.44	67.24
Nr of disordered regions (IDR)	4	3	5	3	6	6
Mean length of disordered regions (aa)	31	36	25	31.3	21.66	19.5
Mean hydrophobic moment (μH) at 100°)	0.253	0.264	0.256	0.253	0.248	0.239
Mean hydrophobic moment (μH) at 160°	0.160	0.239	0.183	0.203	0.208	0.198
Mean Shannon entropy	3.067	2.867	2.988	2.977	2.959	3.077
Mean net charge per residue (NCPR)	0.108	0.184	0.114	0.127	0.122	0.118

	Chondrich-thyes	*Teleostei*	*Amphibia*	*Reptilia*	*Aves*	*Mammalia*
Nr. of phosphorylation sites	24	26	35	26	28	22
Nr. of phosphorylation sites in disordered regions	21	22	26	16	21	16
Percent of phospho-rylation sites in disordered regions (%)	87.5	84.61	74.28	61.53	75	72.72
Nr. of citrullination sites	4	4	12	14	14	17
Nr. of citrullination sites in disordered regions	4	3	8	11	11	15
Percent of citrullination sites in disordered regions (%)	100	75	66.66	78.57	78.57	88.23
Consensus sequence lenght (aa)	147	137	197	180	177	174
Predicted percent of disorder (%)	86.11	78.83	63.45	52.22	73.44	67.24
Nr of disordered regions (IDR)	4	3	5	3	6	6

Following Necci *et al.* (2016), IDRs were classified by length into short IDRs (at least five disordered residues) and long IDRs (>20 disordered residues) categories. With the exception of the C-terminal region in mammals, the terminal IDRs of all clade-specific consensus sequences corresponded to long disordered regions, while all short IDRs were located in the central portion of the sequence. While a comprehensive analysis of proteins in the UniProt database demonstrated that long intrinsically disordered regions (IDRs) are predominantly located in central parts of proteins, whereas short IDRs are mostly positioned at the termini (Necci *et al.*, 2016), the opposite pattern is observed for all MBPs. This inversion is likely explained by the fact that several functionally important regions relevant to MBP activity (Fig. 1C) are located within these terminal segments.

According to Necci *et al.*, 2016, proteins with long IDRs are enriched in eukaryotes and short IDRs seem to be uniformly distributed among all domains of life. Notably, short IDRs were more frequent in tetrapods (excluding *Reptilia*) myelin basic proteins, whereas long IDRs predominated in fishes.

While the number of predicted IDRs showed a positive evolutionary trend ($\rho = 0.65$), it was not statistically significant ($p = 0.19$). Conversely, the average IDR length displayed a stronger negative correlation ($\rho = -0.71$) with a lower p-value ($p = 0.13$), though still below the threshold for statistical significance.

To investigate the relationship between intrinsic disorder and amino acid conservation, we analyzed the aligned clade-specific consensus sequences (Fig. 3). We first identified conserved disordered regions (CDRs) as segments consisting of at least four consecutive residues that were predicted to be disordered in >66.66% of the aligned sequences. These CDRs were further categorised based on the framework of Bellay *et al.* (2011), with modifications to distinguish between flexible and constrained disorder. In regions of flexible disorder, positions are disordered in more than 66.66% of sequences, but the mean conservation score of the segment is less than 5. In regions of constrained disorder, positions are disordered in more than 66.66% of sequences, and the mean conservation score exceeds 5.

Most conserved disordered regions, both flexible and constrained, partially overlapped with known functional elements, including α1–α3 helices, lipid membrane binding regions, calmodulin-binding sites, and serine/threonine phosphorylation sites (Figs. 1C, 3, and 7).

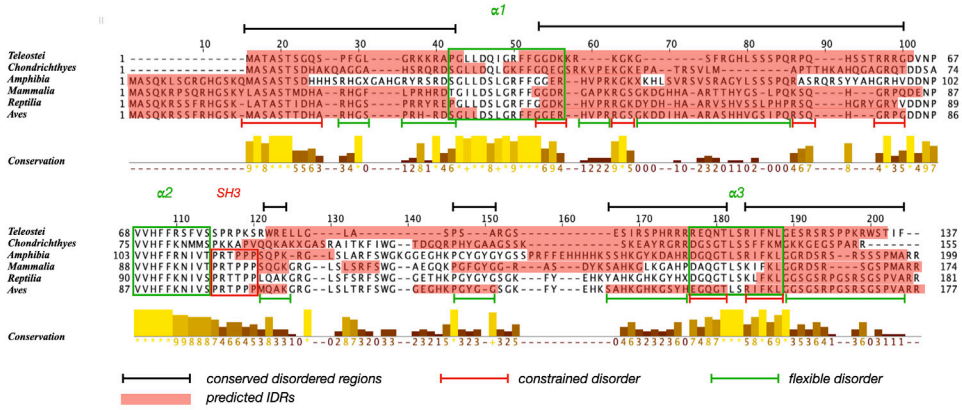

Figure 3. Classification of conserved intrinsically disordered regions (IDRs) in MBP. Multiple sequence alignment of clade-specific MBP consensus sequences with predicted IDRs (brick color) for every sequence. Conserved disordered segments (black) were defined as alignment columns where at least four sequences have a disordered residue. Conserved IDRs were further subdivided into flexible disorder (green) and constrained disorder (red) according to conservation scores, following Bellay *et al.* (2011): segments with an average conservation score ≤5 were classified as flexible disorder, while those with scores >5 were classified as constrained disorder. α-helices are indicated by green rectangles, the SH3-binding motif by a red rectangle. The yellow plot below the alignment shows per-position conservation scores (0–11 scale).

Sequence complexity (Shannon entropy)

The Shannon entropy metric, computed on a sliding 15-amino acid window, reflects the variability and compositional complexity of a sequence (Romero *et al.*, 2001).

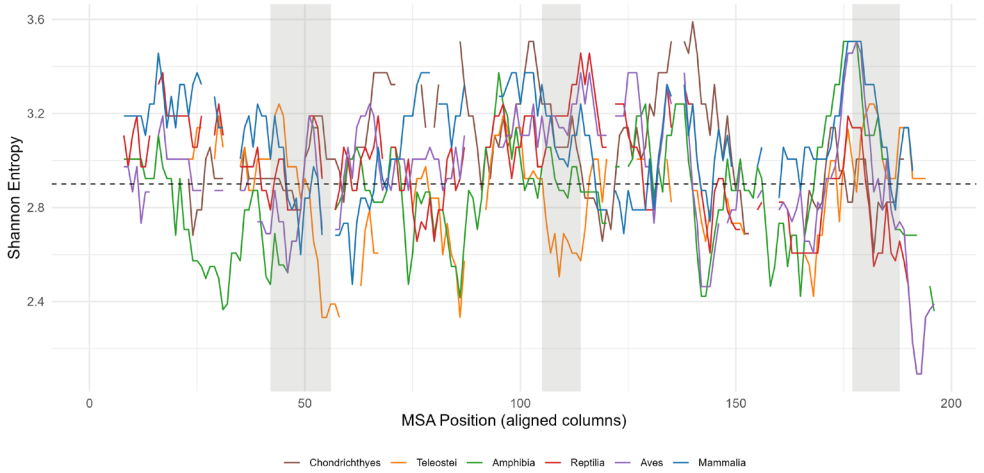

Figure 4. Residue-level aligned profiles of Shannon entropy, with a sliding window of 15 residues, threshold at 2.9. Each plot shows six overlaid lines corresponding to clade-specific MBP consensus sequences: *Chondrichthyes* (brown), *Teleostei* (orange), *Amphibia* (green), *Reptilia* (red), *Aves* (purple), and *Mammalia* (blue). Grey shaded areas in all panels indicate the positions of the three α-helices.

The calculated mean values of the Shannon entropy ranged from 2.287 (*Teleostei*) to 3.077 (*Chondrichthye*s and *Mammalia*) (Table 1, see supplementary files: Fig. S2A). Welch's ANOVA demonstrated that the mean Shannon entropy of MBP differed significantly across vertebrate clades [$F(5, 38.7) \approx 228$, $p < 0.001$, $\eta^2 \approx 0.82$, 96% CI (0.79, 1.00)]. Post-hoc tests revealed clade-specific patterns: cartilaginous fishes (*Chondrichthyes*) differed significantly from all groups except mammals ($p = 0.19$). *Teleosts* (bony fishes) were distinct from all other clades ($p < 0.001$ in all cases). Amphibians did not differ significantly from reptiles ($p = 0.52$), but differed from all others. Reptiles overlapped with amphibians ($p = 0.52$) and birds ($p = 0.16$), but differed from remaining groups. Birds differed from all other clades except reptiles ($p = 0.16$). Mammals differed from all clades except cartilaginous fishes ($p = 0.19$). Overall, entropy displayed strong clade-specific patterns: teleosts consistently had the lowest mean entropy values, cartilaginous fishes the highest, and they overlapped only with mammals.

Figure 4 shows the aligned Shannon entropy profiles of MBP (18.5 kDa isoform) across six vertebrate clades. Entropy values above 2.9 indicate high sequence complexity, strong variability, typical of intrinsically disordered regions (IDRs), while lower values reflect compositional bias or repeats and indicates conservation, often corresponding to structural or functional motifs (e.g., membrane binding) (Romero *et al.*, 2001). The selected entropy value 2.9 threshold is often used in protein disorder and complexity studies as it approximates the point at which variability shifts from constrained to highly flexible (Sen *et al.*, 2019).

Aligned Shannon entropy profiles across *Chondrichthyes, Teleostei, Amphibia, Reptilia, Aves, and Mammalia* reveal both conserved low-complexity domains (entropy values < 2.9) and variable, high-complexity stretches (Fig. 4).

In all clades, three positionally conserved entropy minima can be distinguished, located within the 42–56, 105–114, and 177–188 regions. These correspond to experimentally confirmed α-helices, representing conserved, low-complexity sequence segments (Libich and Harauz, 2008). The N-terminal region (positions 1–15, Fig.3), which is absent in fishes, is characterized by consistently high entropy values, indicative of intrinsic disorder. The entropy profile of this segment shows a gradual increase across vertebrates: it is lowest in amphibians, increases in reptiles and birds, and reaches its highest values in mammals. The elevated intrinsic disorder in this region likely facilitates post-translational modifications, as supported by experimental evidence demonstrating two citrullination sites, two phosphorylation sites, and two membrane-binding residues within this segment in mammals (Harauz *et al.*, 2004). Similarly, the adjacent N-terminal conserved disordered region (positions 16–42, Fig.3), partially overlapping with the α1 helix, exhibits entropy values exceeding 2.9 in all clades except amphibians and *Chondrichthyes* (Fig. 4). The disordered nature of this region suggests an important regulatory role in MBP function, consistent with experimental identification of four phosphorylation sites, three citrullination sites, and four membrane-binding residues in this segment in mammals (Harauz *et al.*, 2004). The conserved disordered region located between the α1 and α2 helices (Fig. 3) shows clade-specific variability in entropy and sequence complexity (Fig. 4): the lowest values are observed in amphibians and teleosts, the highest in *Chondrichthyes* and mammals, whereas birds and reptiles exhibit intermediate levels. This region also contains several post-translational modifications associated with structural flexibility, including four citrullination and three phosphorylation sites, as well as three membrane-binding motifs (Harauz *et al.*, 2004). The 114–140 region (Fig.3), which overlaps with the SH3 motif and contains four phosphorylation and four citrullination sites in mammals (Harauz *et al.*, 2004), is generally characterized by high entropy values (Fig. 4) and a pronounced degree of intrinsic disorder. In contrast, the adjacent 141–170 segment, with the exception of *Chondrichthyes* and mammals, displays relatively

low entropy values on average (Fig. 4), suggesting a tendency toward structural stabilization. Finally, the C-terminal tail (positions 189–200), in contrast to the N-terminal end, shows lower entropy values, indicative of a higher degree of structural order.

The entropy-based complexity analysis reveals evolutionary stratification of MBP disorder across vertebrates. *Mammalia*, *Aves*, *Reptilia* consistently exhibit higher mean entropy values across the protein, suggesting increased sequence variability and functional adaptability compared to amphibians and *Teleostei*. Evolutionarily, the diversification of entropy patterns supports the hypothesis that MBP underwent clade-specific functional tuning to accommodate differences in myelin architecture.

Hydrophobic moment profiles across vertebrate clades

Because protein disorder scores show the strongest correlations with overall polarity and hydrophobicity (Singleton and Eisen, 2024), we analyzed the aligned hydrophobic moment (μH) profiles (Fig. 5 A, B) of the 18.5 kDa myelin basic protein (MBP) across six vertebrate clades (*Chondrichthyes*, *Teleostei, Amphibia, Reptilia, Aves,* and *Mammalia*) at helical angles of 100° and 160° (Eisenberg *et al.*, 1984). The hydrophobic moment quantifies the amphipathicity of a sequence segment: a high μH indicates clustering of hydrophobic side chains on one face of an α-helix or β-strand (Eisenberg *et al.*, 1982; Phoenix and Harris, 2002). Such amphipathic segments frequently correspond to membrane-interacting domains (Gulsevin and Meiler, 2021).

The calculated mean hydrophobic moment (μH) values for amphipathic α-helices (100°) ranged from 0.323 to 0.369, and for amphipathic β-sheets (160°) from 0.219 to 0.342 (Table 1, see supplementary files: Fig. S 2C, D). The hydrophobic moment profiles exhibited clear clade-specific differences. For amphipathic α-helices (100°), Welch's ANOVA revealed a large effect [$F(5, 38.1) = 81.62$, $p < 0.001$, $\eta^2 \approx 0.52$, 95% CI (0.44, 1)]. *Teleostei* (mean = 0.264) had significantly higher α-helical μH values than all other clades ($p \leq 0.01$). Mammals (0.239) exhibited the lowest values, significantly lower than every other clade ($p < 0.001$). *Chondrichthyes* (0.253) did not differ significantly from *Amphibia* (0.256, $p = 0.42$) or *Reptilia* (0.253, $p = 0.44$), but differed from *Aves* (0.248, $p = 0.010$) and *Mammalia* ($p < 0.001$). *Amphibia* overlapped with *Reptilia* and *Chondrichthyes* (both $p > 0.1$), but differed from *Aves* and *Mammalia* (both $p < 0.001$). *Reptilia* overlapped with *Amphibia* ($p = 0.14$) and *Aves* ($p = 0.057$), but were significantly different from *Mammalia* ($p < 0.001$). Aves differed significantly from *Teleostei*, *Chondrichthyes*, *Amphibia*, and *Mammalia* (all $p \leq 0.01$), but not from *Reptilia* ($p = 0.057$). Overall, *Teleostei* exhibited the highest α-helical μH values, *Mammalia* the lowest, and the remaining clades formed an intermediate but partially overlapping group.

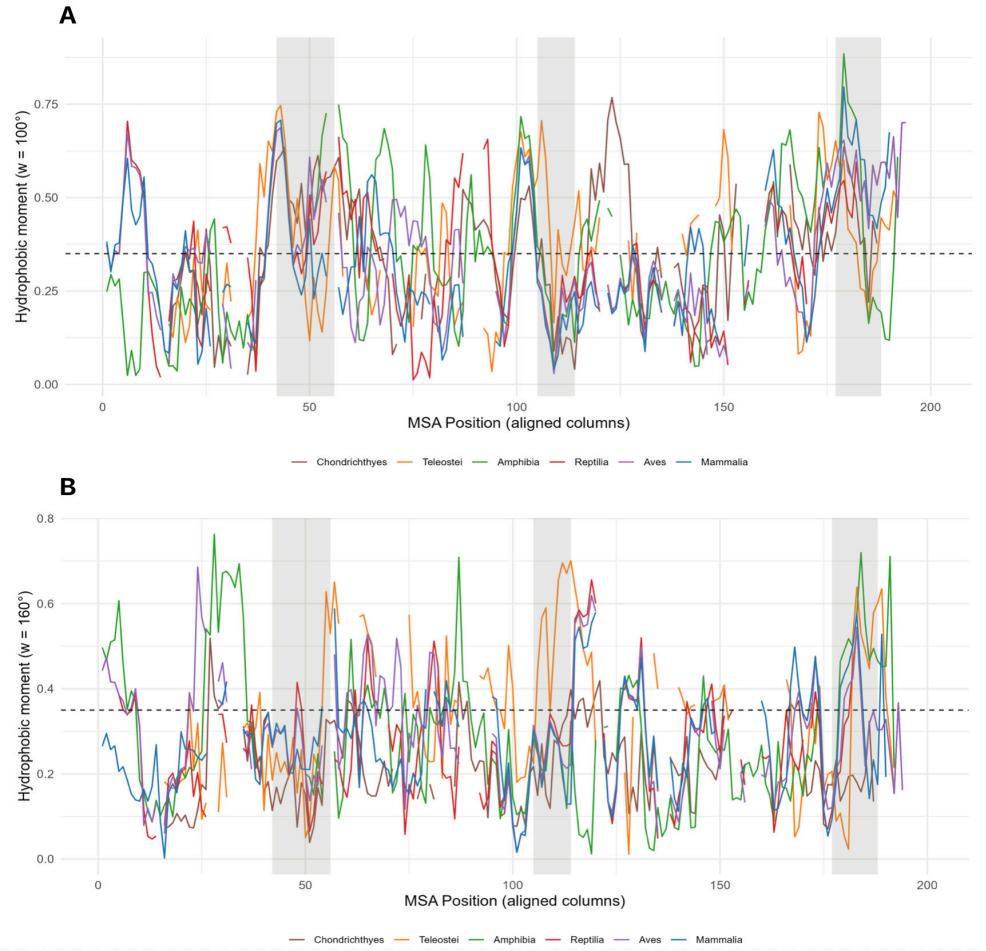

Figure 5. Residue-level aligned profiles of hydrophobic moment. The plots show in order: (A) Hydrophobic moment at 100° (α-helical periodicity), with a sliding window of 10 residues, threshold at 0.35. (B) Hydrophobic moment at 160° (β-strand periodicity), with a sliding window of 10 residues, threshold at 0.35. Each plot shows six overlaid lines corresponding to clade-specific MBP consensus sequences: *Chondrichthyes* (brown), *Teleostei* (orange), *Amphibia* (green), *Reptilia* (red), *Aves* (purple), and *Mammalia* (blue). Grey shaded areas in all panels indicate the positions of the three α-helices.

For amphipathic β-sheets (160°), the differences were even stronger [$F(5, 40.5) = 502.70$, $p < 0.001$, $\eta^2 \approx 0.88$, 95% CI (0.86, 1.00)]. All clades differed significantly ($p < 0.01$), with *Chondrichthyes* showing the lowest and *Teleostei* the highest μH means.

Thus, hydrophobic moment analyses reveal robust clade-specific variation: α-helical values indicate partial overlaps among intermediate clades (*Amphibia, Reptilia, Chondrichthyes*), β-sheet values demonstrate complete separation between all groups.

In addition to comparing clade-specific μH means, comparative hydrophobic moment profiling offers deeper structural insights. Several studies have proposed threshold ranges for μH values, enabling structure-dependent interpretations (Eisenberg *et al.*, 1989; Segrest *et al.*, 1990; Phoenix and Harris, 2002): low μH (< 0.35) values are indicative of weak amphipathicity, typically associated with disordered or flexible regions unlikely to form stable helices, intermediate μH (0.35–0.5) values represents transient or interfacial helices, which may dynamically interact with membranes or partner proteins, and high μH (> 0.5) values are characteristic of strongly amphipathic membrane-binding helices.

Examining hydrophobic moment (μH) profiles (Fig. 5A, B) corresponding to conserved disordered regions (CDRs) (Fig. 3) revealed notable evolutionary differences among clades. In the segment 1–10 of the N-terminal IDR (positions 1–42, Fig. 3) *Reptilia*, *Aves*, and *Mammalia* exhibit μH profiles indicative of a propensity for amphipathic α-helix formation (high 100°, low 160°, Fig. 5A, B). By contrast, *Amphibia* display a preference for β-sheet formation at the same region, likely due to the presence of three adjacent glycines, which may destabilize α-helices. Similar cases where the same conserved disordered region favors different secondary structures across taxa have been reported for other proteins (Siltberg-Liberles, 2011).

Chondrichthyes and *Teleostei* lack the first 15 N-terminal residues (Fig, 3). In mammals, this missing segment contains two experimentally confirmed membrane-binding residues, two citrullination sites and two phosphorylation sites (Harauz *et al.*, 2004). Region 11–25 partially overlapping a constrained CDR (Fig. 3), shows low μH for both angles in all clades (Fig. 5A, B), consistent with disordered character. In region 26–35 in *Chondrichthyes*, *Amphibia*, and *Aves*, the profile shows low 100° but high 160° values (Fig. 5A, B), suggesting a β-sheet propensity. Other clades exhibit low μH at both angles, indicative of disorder.

The central conserved region (positions 53–100, Fig. 3) overlapping the α1 helix, alternates between high 100° and high 160° peaks, implying structural plasticity. It can adopt both shallowly membrane-inserting α-helices and amphipathic β-sheets (Fig. 5A, B). Disorder predictors uniformly classify this region as disordered across all clade consensus sequences (Fig. 2). Amphibians exhibit higher μH values here than other tetrapods, suggesting a possible adaptive increase in membrane-binding capacity during the transition from aquatic to terrestrial environments. In mammals, this region includes four experimentally confirmed membrane-binding residues, three phosphorylation sites, and four

citrullination sites (Harauz *et al.*, 2004). Amphibians possess an inserted sequence segment here that contains multiple positively charged residues, which likely enhance the protein's membrane-binding affinity.

In the C-terminal IDRs region 120–130, partially overlapping a flexible conserved region (positions 121–124, Fig. 3), exhibits high α-helical μH only in *Reptilia*, whereas all other clades show μH values consistent with disorder (Fig. 5A). In region 131–160, all clades generally exhibit low μH values (indicative of disorder), except for *Teleostei*, which display a localized high 100° μH peak near position 150, suggesting potential amphipathic α-helix formation. The short C-terminal tail (positions 181–184) overlapping the α3 helix exhibits simultaneously high 100° and 160° μH values, indicating marked structural plasticity (Fig. 5A, B).

Together, these findings reveal that MBP conserved disordered regions exhibit clade-specific amphipathic profiles, with structural adaptability likely linked to membrane interactions, post-translational regulation, and functional diversification across vertebrate evolution.

Net charge per residue (NCPR) across vertebrate MBP 18.5 kDa orthologs

The net charge per residue (NCPR) differed significantly among vertebrate clades ($F(5, 34.3) = 971.46$, $p < 0.001$), with an exceptionally large effect size ($\eta^2 \approx 0.97$, 95% CI [0.96, 1.00]). Post-hoc comparisons revealed distinct clade-specific patterns: teleosts showed the highest NCPR (mean ≈ 0.184), significantly greater than all others ($p < 0.001$). At the other extreme, cartilaginous fishes exhibited the lowest values (mean ≈ 0.108), while amphibians (mean ≈ 0.114), mammals (mean ≈ 0.118), birds (mean ≈ 0.122), and reptiles (mean ≈ 0.127) formed an intermediate cluster with modest but statistically significant pairwise differences (e.g., amphibians vs. mammals, $p = 0.037$; chondrichthyans vs. amphibians, $p = 0.030$; chondrichthyans vs. mammals, $p < 0.001$). Thus, the NCPR rank order was: cartilaginous fishes (lowest) < amphibians < mammals < birds < reptiles < teleosts (highest) (see supplementary files: Fig. S2B).

Aligned NCPR (net charge per residue) profiles of clade-specific consensus sequences (Fig. 6) revealed an electrostatic architecture that is largely conserved across vertebrates: positively charged peaks — reflecting lysine- and arginine-rich segments — occur predominantly at the N- and C-terminal IDRs, whereas negative troughs correspond to central conserved disordered regions (CDRs) (Fig. 3). These positively charged segments are known to mediate electrostatic binding to the negatively charged cytoplasmic leaflet of myelin, thereby promoting multilayer compaction (Raasakka *et al.*, 2017). In contrast, neutral or acidic regions appear mainly at conserved positions. The alternating pattern of

strongly positive peaks and neutral/acidic dips in MBP's NCPR profile reflects the characteristic electrostatic signature of intrinsically disordered proteins.

The N-terminal portion of the NCPR profile (residues 1–42; Fig. 3) largely overlaps with the N-terminal conserved disordered region (CDR) and is characterized by moderately to strongly positive net charge values (approximately +0.25 to +0.80). In teleost fishes, this segment is truncated by about 15 amino acids compared to tetrapod MBPs. *Teleostei* appear to compensate for the loss of lipid-binding, positively charged residues within positions 1–15 by featuring two additional positively charged residues near the end of the N-terminus, which accounts for the pronounced positive peak observed between positions 31 and 41.

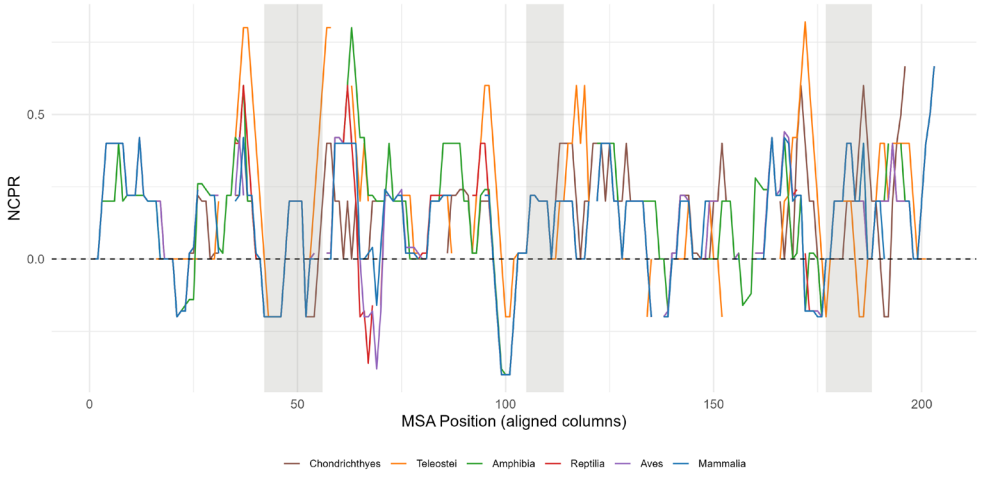

Figure 6. Aligned profiles of net charge per residue (NCPR), with a sliding window of 5 residues, threshold at 0. Each plot shows six overlaid lines corresponding to clade-specific MBP consensus sequences: *Chondrichthyes* (brown), *Teleostei* (orange), *Amphibia* (green), *Reptilia* (red), *Aves* (purple), and *Mammalia* (blue). Grey shaded areas in all panels indicate the positions of the three α-helices.

The central region of MBP (residues 60–119, Fig. 3) exhibits fluctuating net charge — from approximately −0.4 to +0.8 — with a generally positive bias, which mitigates but does not fully eliminate electrostatic repulsion, thereby creating transient windows for intermolecular contact. Notably, this charge variability coincides with an amyloid-prone segment shown to mediate the formation of amyloid fibrils in myelin sheaths (Sysoev *et al.*, 2025), where transient reduction in repulsion could facilitate adhesive MBP stacking, yet also increase the risk of aggregation under altered conditions.

The C-terminal half (residues 124-184, Fig. 3) shows oscillating charge (Fig. 6) and several conserved disordered regions. The segment spanning residues 161–164 contains a histidine-rich insertion in amphibians, introducing a pH-sensitive motif that may fine-tune MBP assembly under variable environmental conditions (Alibardi, 2002; Buchko *et al.*, 2022). The segment spanning residues 166–181, overlapping partially with α3, remains strongly basic until residue 174, where most clades insert an acidic residue; teleosts diverge with consecutive arginines (171–173), raising NCPR above +0.8. Residues 191–184 (Thr, Ser, Arg) are conserved as potential phosphorylation/citrullination sites regulating assembly (Harauz and Boggs, 2013; Smirnova *et al*., 2021).

The extended C-terminal tail (residues 184–204) adds further regulatory capacity. Lys188 is widely conserved but missing in teleosts, Arg194 remains conserved across clades, except cartilaginous fishes, reinforcing positive charge. Overall, the tail balances strong positive clusters with acidic sites, ensuring both adhesion and regulatory potential (Homchaudhuri *et al*., 2010; Raasakka and Kursula, 2022).

Despite this conserved alternating topology, clade-specific adaptations are evident. Mammals and birds share nearly identical high-charge profiles at both termini, consistent with tightly compacted myelin and fast conduction (Müller *et al*., 2013; Sysoev *et al*., 2025). Reptiles largely retain this pattern but with subtle charge reductions and neutral extensions, reflecting conservation with minor adaptation to ectothermy (Müller-Späth *et al.*, 2010; Li and Buck, 2022). Amphibians modify the central region with histidine insertions, creating pH-sensitive motifs that allow environmental responsiveness (Alibardi, 2002). Teleosts markedly intensify terminal charge through dense arginine clusters, likely compensating for lower temperatures by enhancing electrostatics (Aponte-Santamaría *et al*., 2017; Dreier *et al*., 2018). Cartilaginous fish retain lower overall charge intensity, consistent with early marine environments where extreme polycationicity was unnecessary (Tai *et al*., 1985; Bellard, 2016).

Thus, MBP's electrostatic architecture appears both conserved and adaptable. MBP emerged alongside the evolution of myelin in early vertebrates (Nawaz et al., 2013) and has retained its characteristic alternating charge topology under strong selective pressure. In teleosts, the amplification of MBP isoforms may reflect adaptive responses following whole-genome duplication events (Glasauer and Neuhauss, 2014). Amphibians and reptiles appear to have introduced more flexible modifications, whereas birds and mammals have converged on high-charge conservation to optimize saltatory conduction (Stämpfli, 1954). By contrast, cartilaginous fishes retain a more ancestral electrostatic configuration (Bellard, 2016). Overall, the NCPR landscape highlights MBP's dual nature: intrinsically disordered yet functionally encoded through

finely tuned charge distribution. Conserved basic clusters, interspersed with acidic and neutral dips, enable reversible membrane adhesion and provide platforms for post-translational regulation (Bianchi *et al.,* 2022). This tunable electrostatic framework underpins MBP's central role in myelin compaction, balancing deep evolutionary stability with lineage-specific adaptations.

Distribution of predicted phosphorylation and citrullination sites

As previous studies have identified a notable correlation between the distribution of post-translational modifications (PTMs) and intrinsically disordered regions (IDRs) (Iakoucheva *et al.*, 2004; Kurotani *et al.*, 2014; Zarin *et al.*, 2019,), we aimed to investigate whether vertebrate evolution has influenced the number of functionally critical PTMs (specifically, phosphorylation and citrullination) within MBP, and whether their distribution correlates with predicted IDRs. Because experimentally validated phospho- and citrullination sites in MBP are limited to only a few species (Zand *et al.*, 2001; Kim *et al.*, 2009; Zang *et al.*, 2012), we supplemented these data with predictive methods.

We used computational predictors to identify phosphorylation sites for each clade (Figure 7, see supplementary files: Table S1), which revealed a high but variable proportion of sites located in disordered regions: 87.5% in cartilaginous fish, 84.6% in bony fish, 74.3% in amphibians, 61.5% in reptiles, 75.0% in birds, and 72.2% in mammals (Table 1).

Separately, using experimentally determined human MBP phosphorylation sites (n = 20), we examined the aligned consensus sequences across the six vertebrate clades (*Chondrichthyes, Teleostei, Amphibia, Reptilia, Aves,* and *Mammalia*) (Fig. 3). Six of these sites were conserved across all clades (18S/T, 20S, 21T, 114S/T, 181T, 183S). Additional positions were retained only in subsets of lineages: 8S, 13S, 42S/T, 65S, 79Y, 81S, 117T, 121S, 136S, 159S/Y, 166S/Y, 193S, 199S, and 201S. 11 of the 20 sites were located within conserved disordered regions. According to this predictions phosphorylation sites accumulated gradually, with six positions conserved across all vertebrates and numerous additional sites appearing in amphibians and higher lineages. Under normal conditions, MBP functions as a phosphoprotein in the myelin sheath (Turner *et al.,* 1982). Phosphorylation is critical for regulating protein function, localization, and protein–protein interactions. By partially neutralizing MBP's charge, phosphorylation modulates membrane-binding affinity and participation in intracellular signaling pathways (Harauz and Boggs, 2013). According to our data, across vertebrates, predicted phosphorylation sites are generally associated with intrinsically disordered regions, although the strength of this association varied among clades. This suggests that in early vertebrates, phosphorylation

was already tightly linked to disorder, but over evolutionary time, it expanded into more structured regions, particularly in reptiles.

In dogfish, experimental methods identified six phosphorylation sites (Zand *et al.*, 2001), whereas our combined predictive approach estimated 24 sites across the consensus MBP sequence for cartilaginous fishes, and the alignment-based method predicted seven. These apparent discrepancies likely arise both from differences between predictive methodologies and from limitations in the available experimental data.

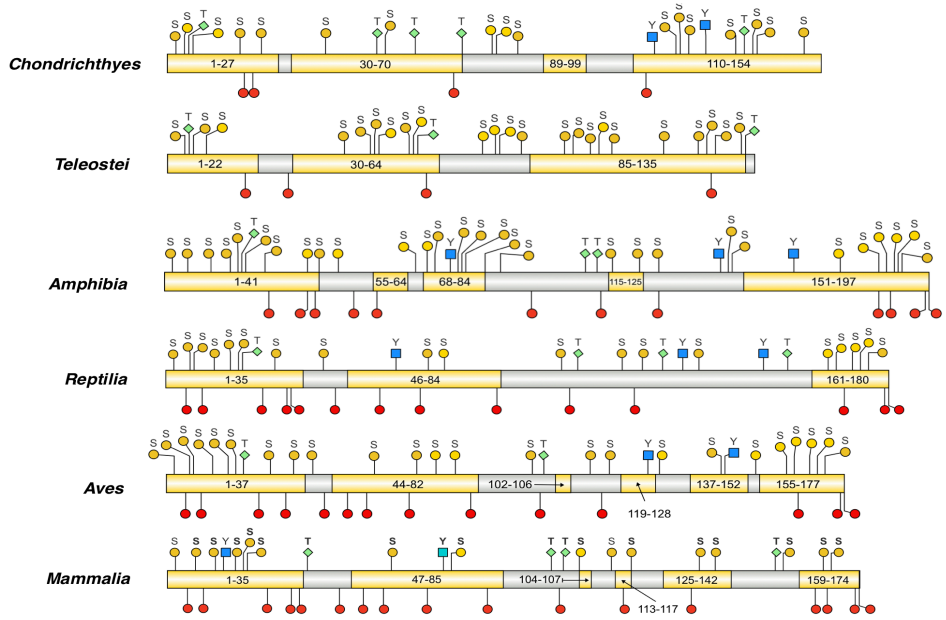

Figure 7. Predicted post-translational modifications (PTMs) of MBP consensus sequences across vertebrate lineages. For each of the six vertebrate clades (*Chondrichthyes, Teleostei, Amphibia, Reptilia, Aves, Mammalia*), the consensus MBP sequence is shown with annotated PTM sites. Phosphorylation sites were identified using three independent predictors (DEPP, NetPhos, MusiteDeep) and are indicated only when supported by at least two predictors: yellow circles (serine), green diamonds (threonine), and blue squares (tyrosine). In the *Mammalia* panel, phosphorylation sites that are bolded have been experimentally confirmed (Harauz and Boggs, 2013). Citrullination sites (red circles) were compiled from published experimental studies (Harauz *et al.*, 2004; Kishimoto *et al.,* 1985). Intrinsically disordered regions are highlighted in yellow along the sequences (numbers indicate the start and end positions of each disordered region within the sequence).

The number of predicted citrullination sites gradually increased across vertebrate clades, from 4 sites in cartilaginous and bony fishes to a maximum of 17 sites in mammals (Table 1, and see supplementary files: Table S2). Several positions were highly conserved across clades (e.g., R6, R40, R50, R63, R98) and include residues also implicated in membrane binding, notably R40 and R63. Eleven of these sites in the aligned sequences were located within conserved disordered regions. On the individual consensus sequences (Figure 7), the proportion of citrullination sites in disordered regions varied (Table 1.): 100% in cartilaginous fish, 75% in bony fish, 66.66% in amphibians, 78.57% in reptiles, 78.57% in birds, and 88.23% in mammals.

In early vertebrates, almost all predicted citrullination sites were located in disordered regions (Fig. 7), but they were few in number (four in cartilaginous fish and four in bony fish). In amphibians, their number increased, with a majority expected to fall within disordered regions. Reptiles and birds had more predicted sites, most of which were predicted to be in disordered regions, and mammals carried the highest number, many within disordered segments. This pattern shows that even though the proportion of predicted disordered residues dropped from 100% in cartilaginous fish to ~67–88% in tetrapods, the absolute number of predicted disorder-associated citrullination sites more than tripled. Many of these residues coincide with membrane-binding arginines, highlighting an increasing potential for regulatory conflict between citrullination and membrane adhesion as MBP diversified (Gogól, 2013). The trajectory of citrullination differs from phosphorylation by combining strong disorder association with numerical expansion. Citrullination modifies specific arginine residues in MBP, converting them to citrulline via calcium-dependent PAD (peptidylarginine deiminase) enzymes. This reduces MBP's positive charge, weakening its electrostatic binding to negative membrane lipids and potentially disrupting myelin structure (Moscarello *et al*., 2006). It may also reduce interactions with other proteins, affecting myelin organization (Carrillo-Vico *et al.*, 2010).

Together, these findings suggest that MBP evolution was shaped by a progressive accumulation of PTMs within disordered regions, which not only expanded the number of modification sites, but also increased the dynamic potential of the protein. Phosphorylation acts as a flexible, conserved regulatory mechanism, partially neutralizing MBP charge to fine-tune membrane binding and signaling, while citrullination selectively targets membrane-binding arginines, potentially disrupting MBP-membrane interactions and protein-protein contacts. By localizing within intrinsically disordered regions, these PTMs likely create a "regulatory code," enabling switch-like modulation of interactions, scaffold assembly, and fuzzy binding without requiring rigid structural changes (Bah and Forman-Kay, 2016; Darling and Uversky, 2018).

Conclusion

This study presents an integrative evolutionary analysis of the 18.5 kDa myelin basic protein (MBP) isoform across major vertebrate clades, combining sequence-based intrinsic disorder predictions, hydrophobic moment and net charge per residue (NCPR) profiling, and post-translational modification (PTM) mapping. By employing consensus sequence alignments and multi-predictor computational approaches, we identified both deeply conserved and lineage-specific features that contribute to the structural and functional versatility of MBP.

Our results demonstrate that MBP exhibits an intrinsically disordered architecture across all vertebrate groups, with predicted disorder percentages ranging from 52.22% to 86.11%. The spatial distribution of intrinsically disordered regions (IDRs) follows a conserved pattern: long N- and C-terminal IDRs are universally present, while the number and length of central IDRs vary among clades. Short IDRs dominate in tetrapods, whereas teleost fishes exhibit longer, more extended disordered segments. These conserved disordered regions (CDRs) overlap with functional elements such as membrane-binding regions, α-helices, and PTM motifs, underscoring their dual role as structural scaffolds and regulatory platforms.

Shannon entropy profiles and hydrophobic moment distributions reveal that MBP's low-complexity segments are closely associated with structural motifs. Three conserved entropy minima correspond to experimentally verified α-helical segments, likely serving as structural anchors within otherwise flexible regions. In contrast, high-entropy segments align with disordered zones enriched in phosphorylation and citrullination sites, supporting the hypothesis that intrinsic disorder enhances regulatory plasticity. Despite overarching conservation, our analyses reveal lineage-specific fine-tuning. Hydrophobic moment profiles show that reptiles, birds, and mammals display a tendency toward amphipathic α-helix formation in the N-terminal CDR, whereas amphibians favor β-sheet formation. Central disordered regions alternate between α-helical and β-sheet propensities, reflecting structural plasticity, while teleosts and reptiles uniquely exhibit C-terminal helical propensity, absent in other clades.

Although the overall electrostatic topology of MBP — alternating basic clusters and neutral or acidic dips that facilitate reversible membrane adhesion — has remained stable since the emergence of myelin in early vertebrates, subtle inter-clade differences reflect functional divergence. Mammals and birds exhibit enhanced conservation of positively charged domains, consistent with the requirements of rapid saltatory conduction, while amphibians and reptiles display greater flexibility in their electrostatic architecture. Teleost MBPs, which lack the first 15 N-terminal residues, exhibit compensatory enrichment

of positively charged residues nearby, partially restoring membrane-binding potential. Amphibians show unique insertions that extend the protein and introduce additional basic residues, likely enhancing membrane affinity and supporting clade-specific adaptations in myelin organization.

Mapping experimentally validated and computationally predicted PTMs onto aligned consensus sequences revealed a strong association between phosphorylation and citrullination sites and disordered regions, although the strength of this association varied across clades. In early vertebrates, PTMs were primarily localized within IDRs, while in reptiles and birds their distribution expanded into more ordered regions, suggesting evolutionary broadening of regulatory control. PTMs were frequently co-located with conserved helices and membrane-binding sites, indicating a synergistic relationship between structural flexibility and post-translational regulation.

In summary, MBP exemplifies the dual nature of an intrinsically disordered protein: highly conserved in its core architecture yet dynamically adaptable through clade-specific modifications. The preservation of structural and electrostatic features across vertebrate evolution underscores strong selective pressure to maintain MBP's essential role in myelin compaction and stability. Simultaneously, adaptive changes in disorder profiles, sequence complexity, secondary structure, charge distribution, and PTM positioning reflect evolutionary tuning to diverse myelin architectures and functional requirements. Extending this comparative framework to other intrinsically disordered proteins may deepen our understanding of how disorder-mediated regulation shapes protein evolution and cellular plasticity.

References

Alibardi, L., 2002. Immunocytochemical localization of keratins, associated proteins and uptake of histidine in the epidermis of fish and amphibians. *Acta Histochem. 104,* 297–310. https://doi.org/10.1078/0065-1281-00651

Aponte-Santamaría, C., Fischer, G., Båth, P., Neutze, R. & De Groot, B.L., 2017. Temperature dependence of protein-water interactions in a gated yeast aquaporin. *Sci. Rep. 7*, 4016. https://doi.org/10.1038/s41598-017-04180-z

Babu, M.M., Van Der Lee, R. & De Groot, N.S., Gsponer, J., 2011. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol. 21,* 432–440. https://doi.org/10.1016/j.sbi.2011.03.011

Bah, A & Forman-Kay, J.D., 2016. Modulation of intrinsically disordered protein function by post-translational modifications. J. *Biol. Chem. 291,* 6696–6705. https://doi.org/10.1074/jbc.R115.695056

Bates, I.R., Feix, J.B., Boggs, J.M. & Harauz, G., 2004. An immunodominant epitope of myelin basic protein is an amphipathic α-helix. *J. Biol. Chem. 279,* 5757–5764. https://doi.org/10.1074/jbc.M311504200

Bedja-Iacona, L., Richard, E., Marouillat, S., Brulard, C., Alouane, T., Beltran, S., Andres, C.R., Blasco, H., Corcia, P., Veyrat-Durebex, C. & Vourc'h, P., 2024. Post-translational variants of major proteins in amyotrophic lateral sclerosis provide new insights into the pathophysiology of the disease. *IJMS 25,* 8664. https://doi.org/10.3390/ijms25168664

Bellay, J., Han, S., Michaut, M., Kim, T., Costanzo, M., Andrews, B.J., Boone, C., Bader, G.D., Myers, C.L. & Kim, P.M., 2011. Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol. 12,* R14. https://doi.org/10.1186/gb-2011-12-2-r14

Bianchi, G., Mangiagalli, M., Barbiroli, A., Longhi, S., Grandori, R., Santambrogio, C. & Brocca, S., 2022. Distribution of charged residues affects the average size and shape of intrinsically disordered proteins. *Biomolecules 12,* 561. https://doi.org/10.3390/biom12040561

Buchko, G.W., Mergelsberg, S.T., Tarasevich, B.J. & Shaw, W.J., 2022. Residue-specific insights into the intermolecular protein–protein interfaces driving amelogenin self-assembly in solution. *Biochemistry 61,* 2909–2921. https://doi.org/10.1021/acs.biochem.2c00522

Carrillo-Vico, A., Leech, M.D. & Anderton, S.M., 2010. Contribution of myelin autoantigen citrullination to t cell autoaggression in the central nervous system. *J. Immunol. 184,* 2839–2846. https://doi.org/10.4049/jimmunol.0903639

Charif, D. & Lobry, J.R., 2007. Seqinr 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis, in: bastolla, u., Porto, M., Roman, H.E., Vendruscolo, M. (Eds.), structural approaches to sequence evolution, biological and medical physics, biomedical engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 207–232. https://doi.org/10.1007/978-3-540-35306-5_10

Darling, A.L. & Uversky, V.N., 2018. Intrinsic disorder and posttranslational modifications: the darker side of the biological dark matter. *Front. Genet. 9,* 158. https://doi.org/10.3389/fgene.2018.00158

Das, R.K. & Pappu, R.V., 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A. 110,* 13392–13397. https://doi.org/10.1073/pnas.1304749110

De Avila, M. & Vassall, K.A., Smith, G.S.T., Bamm, V.V., Harauz, G., 2014. The proline-rich region of 18.5 kDa myelin basic protein binds to the SH3-domain of Fyn tyrosine kinase with the aid of an upstream segment to form a dynamic complex *in vitro*. *Biosci. Rep. 34,* e00157. https://doi.org/10.1042/BSR20140149

De Bellard, M.E., 2016. Myelin in cartilaginous fish. *Brain Res. 1641,* 34–42. https://doi.org/10.1016/j.brainres.2016.01.013

De Vries, I., Bak, J., Salmoral, D.Á., Xie, R., Borza, R., Konijnenberg, M. & Perrakis, A., 2024. Disentangling the CHAOS of intrinsic disorder in human proteins. *bioRxiv* https://doi.org/10.1101/2024.10.26.620428

Del Conte, A., Bouhraoua, A., Mehdiabadi, M., Clementel, D., Monzon, A.M., CAID predictors, Holehouse, A.S., Griffith, D., Emenecker, R.J., Patil, A., Sharma, R., Tsunoda, T., Sharma, A., Tang, Y.J., Liu, B., Mirabello, C., Wallner, B., Rost, B., Ilzhöfer, D., Littmann, M., Heinzinger, M., Krautheimer, L.I.M., Bernhofer, M., McGuffin, L.J., Callebaut, I., Feildel, T.B., Liu, J., Cheng, J., Guo, Z., Xu, J., Wang, S., Malhis, N., Gsponer, J., Kim, C.-S., Han, K.-S., Ma, M.-C., Kurgan, L., Ghadermarzi, S., Katuwawala, A., Zhao, B., Peng, Z., Wu, Z., Hu, G., Wang, K., Hoque, M.T., Kabir, M.W.U., Vendruscolo, M., Sormanni, P., Li, M., Zhang, F., Jia, P., Wang, Y., Lobanov, M.Y., Galzitskaya, O.V., Vranken, W., Díaz, A., Litfin, T., Zhou, Y., Hanson, J., Paliwal, K., Dosztányi, Z., Erdős, G., Tosatto, S.C.E. & Piovesan, D., 2023. CAID prediction portal: a comprehensive service for predicting intrinsic disorder and binding regions in proteins. *Nucleic Acids Res.* 51, W62–W69. https://doi.org/10.1093/nar/gkad430

Dreier, L.B., Nagata, Y., Lutz, H., Gonella, G., Hunger, J., Backus, E.H.G. & Bonn, M., 2018. Saturation of charge-induced water alignment at model membrane surfaces. *Sci. Adv. 4,* eaap7415. https://doi.org/10.1126/sciadv.aap7415

Eisenberg, D., Weiss, R.M. & Terwilliger, T.C., 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci. U.S.A. 81,* 140–144. https://doi.org/10.1073/pnas.81.1.140

Eisenberg, D., Weiss, R.M., Terwilliger, T.C. & Wilcox, W., 1982. Hydrophobic moments and protein structure. Faraday Symp. *Chem. Soc. 17,* 109. https://doi.org/10.1039/fs9821700109

Eisenberg, D., Wesson, M. & Wilcox, W., 1989. Hydrophobic moments as tools for analyzing protein sequences and structures, in: fasman, g.d. (Ed.), prediction of protein structure and the principles of protein conformation. Springer US, Boston, MA, pp. 635–646. https://doi.org/10.1007/978-1-4613-1571-1_16

Erdős, G. & Dosztányi, Z., 2020. Analyzing protein disorder with IUPred2A. *Curr. Protoc. Bioinform. 70,* e99. https://doi.org/10.1002/cpbi.99

Glasauer, S.M.K. & Neuhauss, S.C.F., 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics 289,* 1045–1060. https://doi.org/10.1007/s00438-014-0889-2

Gogól, M., 2013. Citrullination – small change with a great consequence. *Acta Universitatis Lodziensis. Folia Biologica Et Ecologica 9,* 17–25. https://doi.org/10.2478/fobio-2013-0003

Gsponer, J., Futschik, M.E., Teichmann, S.A. & Babu, M.M., 2008. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science 322,* 1365–1368. https://doi.org/10.1126/science.1163581

Gulsevin, A. & Meiler, J., 2021. Prediction of amphipathic helix—membrane interactions with Rosetta. *PLoS. Comput. Biol. 17,* e1008818. https://doi.org/10.1371/journal.pcbi.1008818

Harauz, G. & Boggs, J.M., 2013. Myelin management by the 18.5-kDa and 21.5-kDa classic myelin basic protein isoforms. *J. Neurochem. 125,* 334–361. https://doi.org/10.1111/jnc.12195

Harauz, G., Ishiyama, N., Hill, C.M.D., Bates, I.R., Libich, D.S. & Farès, C., 2004. Myelin basic protein—diverse conformational states of an intrinsically unstructured protein and its roles in myelin assembly and multiple sclerosis. *Micron 35,* 503–542. https://doi.org/10.1016/j.micron.2004.04.005

Harauz, G. & Musse, A.A., 2007. A tale of two citrullines—structural and functional aspects of myelin basic protein deimination in health and disease. *Neurochem. Res. 32,* 137–158. https://doi.org/10.1007/s11064-006-9108-9

Holehouse, A.S., Das, R.K., Ahad, J.N., Richardson, M.O.G. & Pappu, R.V., 2017. Cider: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *Biophys. J. 112,* 16–21. https://doi.org/10.1016/j.bpj.2016.11.3200

Homchaudhuri, L., De Avila, M., Nilsson, S.B., Bessonov, K., Smith, G.S.T., Bamm, V.V., Musse, A.A., Harauz, G. & Boggs, J.M., 2010. Secondary structure and solvent accessibility of a calmodulin-binding c-terminal segment of membrane-associated myelin basic protein. *Biochemistry 49,* 8955–8966. https://doi.org/10.1021/bi100988p

Houben, B., Michiels, E., Ramakers, M., Konstantoulea, K., Louros, N., Verniers, J., Van Der Kant, R., De Vleeschouwer, M., Chicória, N., Vanpoucke, T., Gallardo, R., Schymkowitz, J. & Rousseau, F., 2020. Autonomous aggregation suppression by acidic residues explains why chaperones favour basic residues. *EMBO J. 39,* e102864. https://doi.org/10.15252/embj.2019102864

Iakoucheva, L.M., 2004. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res. 32,* 1037–1049. https://doi.org/10.1093/nar/gkh253

Jensen, M.R., Ruigrok, R.W. & Blackledge, M., 2013. Describing intrinsically disordered proteins at atomic resolution by NMR. *Cur. Opin. Struct. Biol. 23,* 426–435. https://doi.org/10.1016/j.sbi.2013.02.007

Kastano, K., Erdős, G., Mier, P., Alanis-Lobato, G., Promponas, V.J., Dosztányi, Z. & Andrade-Navarro, M.A., 2020. Evolutionary study of disorder in protein sequences. *Biomolecules 10,* 1413. https://doi.org/10.3390/biom10101413

Kim, J., Zhang, R., Strittmatter, E.F., Smith, R.D. & Zand, R., 2009. Post-translational modifications of chicken myelin basic protein charge components. *Neurochem. Res. 34,* 360–372. https://doi.org/10.1007/s11064-008-9788-4

Kishimoto, A., Nishiyama, K., Nakanishi, H., Uratsuji, Y., Nomura, H., Takeyama, Y. & Nishizuka, Y., 1985. Studies on the phosphorylation of myelin basic protein by protein kinase C and adenosine 3‘:5‘-monophosphate-dependent protein kinase. *J. Biol. Chem. 260,* 12492–12499. https://doi.org/10.1016/S0021-9258(17)38898-1

Kister, A. & Kister, I., 2023. Overview of myelin, major myelin lipids, and myelin-associated proteins. *Front. Chem. 10,* 1041961. https://doi.org/10.3389/fchem.2022.1041961

Kurotani, A., Tokmakov, A.A., Kuroda, Y., Fukami, Y., Shinozaki, K. & Sakurai, T., 2014. Correlations between predicted protein disorder and post-translational modifications in plants. *Bioinformatics 30,* 1095–1103. https://doi.org/10.1093/bioinformatics/btt762

Li, Z. & Buck, M., 2023. A proteome-scale analysis of vertebrate protein amino acid occurrence: Thermoadaptation shows a correlation with protein solvation but less so with dynamics. *Proteins 91,* 3–15. https://doi.org/10.1002/prot.26404

Libich, D.S. & Harauz, G., 2008. Backbone dynamics of the 18.5kDa isoform of myelin basic protein reveals transient α-helices and a calmodulin-binding site. *Biophys. J. 94,* 4847–4866. https://doi.org/10.1529/biophysj.107.125823

Liu, Y., Wang, X. & Liu, B., 2019. A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *Brief. Bioinform. 20,* 330–346. https://doi.org/10.1093/bib/bbx126

Livingstone, C.D. & Barton, G.J., 1993. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Bioinformatics 9,* 745–756. https://doi.org/10.1093/bioinformatics/9.6.745

Lobley, A., Orengo, C.A., Swindells, M.B. & Jones, D., 2005. Inferring function using patterns of native disorder in proteins. *PLoS. Comput. Biol. preprint,* e162. https://doi.org/10.1371/journal.pcbi.0030162.eor

Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A.R.N., Lopez, R. & Butcher, S., 2024. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res. 52,* W521–W525. https://doi.org/10.1093/nar/gkae241

Monastyrskyy, B., Kryshtafovych, A., Moult, J., Tramontano, A. & Fidelis, K., 2014. Assessment of protein disorder region predictions in CASP10. *Proteins 82,* 127–137. https://doi.org/10.1002/prot.24391

Moscarello, M.A., Mastronardi, F.G. & Wood, D.D., 2007. The role of citrullinated proteins suggests a novel mechanism in the pathogenesis of multiple sclerosis. *Neurochem. Res. 32,* 251–256. https://doi.org/10.1007/s11064-006-9144-5

Mughal, F. & Caetano-Anollés, G., 2025. Evolution of intrinsic disorder in the structural domains of viral and cellular proteomes. *Sci. Rep. 15,* 2878. https://doi.org/10.1038/s41598-025-86045-4

Müller, C., Bauer, N.M., Schäfer, I. & White, R., 2013. Making myelin basic protein -from mRNA transport to localized translation. *Front. Cell. Neurosci. 7,* 1-7. https://doi.org/10.3389/fncel.2013.00169

Müller-Späth, S., Soranno, A., Hirschfeld, V., Hofmann, H., Rüegger, S., Reymond, L., Nettels, D. & Schuler, B., 2010. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A. 107,* 14609–14614. https://doi.org/10.1073/pnas.1001743107

Nawaz, S., Schweitzer, J., Jahn, O. & Werner, H.B., 2013. Molecular evolution of myelin basic protein, an abundant structural myelin component: phylogeny of myelin basic protein. *Glia 61,* 1364–1377. https://doi.org/10.1002/glia.22520

Necci, M., Piovesan, D., Dosztányi, Z. & Tosatto, S.C.E., 2017. MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics 33,* 1402–1404. https://doi.org/10.1093/bioinformatics/btx015

Necci, M., Piovesan, D. & Tosatto, S.C.E., 2016. Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. *Prot. Sci 25,* 2164–2174. https://doi.org/10.1002/pro.3041

Oldfield, C.J., Uversky, V.N., Dunker, A.K. & Kurgan, L., 2019. Introduction to intrinsically disordered proteins and regions, in: Intrinsically Disordered Proteins. *Elsevier,* pp. 1–34. https://doi.org/10.1016/B978-0-12-816348-1.00001-6

Phoenix, D.A. & Harris, F., 2002. The hydrophobic moment and its use in the classification of amphiphilic structures (Review). *Mol. Membr. Biol. 19,* 1–10. https://doi.org/10.1080/09687680110103631

Polverini, E., Coll, E.P., Tieleman, D.P. & Harauz, G., 2011. Conformational choreography of a molecular switch region in myelin basic protein — Molecular dynamics shows induced folding and secondary structure type conversion upon threonyl phosphorylation in both aqueous and membrane-associated environments. *Biochim. Biophys. Acta Biomembr. 1808,* 674–683. https://doi.org/10.1016/j.bbamem.2010.11.030

Raasakka, A. & Kursula, P., 2020. Flexible players within the sheaths: the intrinsically disordered proteins of myelin in health and disease. *Cells 9,* 470. https://doi.org/10.3390/cells9020470

Raasakka, A., Ruskamo, S., Kowal, J., Barker, R., Baumann, A., Martel, A., Tuusa, J., Myllykoski, M., Bürck, J., Ulrich, A.S., Stahlberg, H. & Kursula, P., 2017. Membrane Association Landscape of Myelin Basic Protein Portrays Formation of the Myelin Major Dense Line. *Sci. Rep. 7,* 4974. https://doi.org/10.1038/s41598-017-05364-3

Rajagopalan, K., Mooney, S.M., Parekh, N., Getzenberg, R.H. & Kulkarni, P., 2011. A majority of the cancer/testis antigens are intrinsically disordered proteins. *J. Cell. Biochem. 112,* 3256–3267. https://doi.org/10.1002/jcb.23252

Riley, A.C., Ashlock, D.A. & Graether, S.P., 2023. The difficulty of aligning intrinsically disordered protein sequences as assessed by conservation and phylogeny. *PLoS. ONE 18,* e0288388. https://doi.org/10.1371/journal.pone.0288388

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J. & Dunker, A.K., 2001. Sequence complexity of disordered protein. *Proteins 42,* 38–48. https://doi.org/10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3

Schad, E., Tompa, P. & Hegyi, H., 2011. The relationship between proteome size, structural disorder and organism complexity. *Genome Biol. 12,* R120. https://doi.org/10.1186/gb-2011-12-12-r120

Segrest, J.P., De Loof, H., Dohlman, J.G., Brouillette, C.G. & Anantharamaiah, G.M., 1990. Amphipathic helix motif: classes and properties. *Proteins 8,* 103–117. https://doi.org/10.1002/prot.340080202

Sen, S., Dey, A., Chowdhury, S., Maulik, U. & Chattopadhyay, K., 2019. Understanding the evolutionary trend of intrinsically structural disorders in cancer relevant proteins as probed by Shannon entropy scoring and structure network analysis. *BMC Bioinformatics 19,* 549. https://doi.org/10.1186/s12859-018-2552-0

Siddiqui, I.J., Pervaiz, N. & Abbasi, A.A., 2016. The Parkinson disease gene SNCA: evolutionary and structural insights with pathological implication. *Sci. Rep. 6,* 24475. https://doi.org/10.1038/srep24475

Siltberg-Liberles, J., 2011. Evolution of structurally disordered proteins promotes neostructuralization. *Mol. Biol. Evol. 28*, 59–62. https://doi.org/10.1093/molbev/msq291

Singleton, M.D. & Eisen, M.B., 2024. Evolutionary analyses of intrinsically disordered regions reveal widespread signals of conservation. *PLoS. Comput. Biol. 20,* e1012028. https://doi.org/10.1371/journal.pcbi.1012028

Smirnova, E.V., Rakitina, T.V., Ziganshin, R.H., Arapidi, G.P., Saratov, G.A., Kudriaeva, A.A. & Belogurov, A.A., 2021. Comprehensive atlas of the myelin basic protein interaction landscape. *Biomolecules 11,* 1628. https://doi.org/10.3390/biom11111628

Stämpfli, R., 1954. Saltatory conduction in nerve. *Phy. Rev.* https://journals.physiology.org/doi/abs/10.1152/physrev.1954.34.1.101?journalCode=physrev

Sysoev, E.I., Shenfeld, A.A., Belashova, T.A., Valina, A.A., Zadorsky, S.P. & Galkin, A.P., 2025. Amyloid fibrils of the myelin basic protein are an integral component of myelin in the vertebrate brain. *Sci. Rep. 15,* 29053. https://doi.org/10.1038/s41598-025-13524-z

Tai, F.L., Smith, R., Bernard, C.C.A. & Hearn, M.W.T., 1986. Evolutionary divergence in the structure of myelin basic protein: comparison of chondrichthye basic proteins with those from higher vertebrates. *J. Neurochem. 46,* 1050–1057. https://doi.org/10.1111/j.1471-4159.1986.tb00617.x

Turner, R.S., Chou, C. -H. J., Kibler, R.F. & Kuo, J.F., 1982. Basic protein in brain myelin is phosphorylated by endogenous phospholipid-sensitive Ca^{2+} -dependent protein kinase. *J. Neurochem. 39,* 1397–1404. https://doi.org/10.1111/j.1471-4159.1982.tb12583.x

Uversky, V.N., 2019. Intrinsically disordered proteins and their “mysterious” (meta)physics. *Front. Phys. 7,* 10. https://doi.org/10.3389/fphy.2019.00010

Van Bibber, N.W., Haerle, C., Khalife, R., Xue, B. & Uversky, V.N., 2020. Intrinsic disorder in tetratricopeptide repeat proteins. *IJMS 21,* 3709. https://doi.org/10.3390/ijms21103709

Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R.V., Tompa, P., Uversky, V.N., Wright, P.E. & Babu, M.M., 2014. Classification of intrinsically disordered regions and proteins. *Chem. Rev. 114,* 6589–6631. https://doi.org/10.1021/cr400525m

Walsh, I., Martin, A.J.M., Di Domenico, T. & Tosatto, S.C.E., 2012. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics 28,* 503–509. https://doi.org/10.1093/bioinformatics/btr682

Wang, D., Zeng, S., Xu, C., Qiu, W., Liang, Y., Joshi, T. & Xu, D., 2017. MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction. *Bioinformatics 33,* 3909–3916. https://doi.org/10.1093/bioinformatics/btx496

Wang, S., Ma, J. & Xu, J., 2016. AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. *Bioinformatics 32,* i672–i679. https://doi.org/10.1093/bioinformatics/btw446

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. & Jones, D.T., 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol. 337,* 635–645. https://doi.org/10.1016/j.jmb.2004.02.002

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. & Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics 25,* 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Welch, B.L., 1951. On the comparison of several mean values: an alternative approach. *Biometrika 38,* 330. https://doi.org/10.2307/2332579

Wickham, H., 2016. Programming with ggplot2, in: ggplot2, use R! *Springer International Publishing, Cham,* pp. 241–253. https://doi.org/10.1007/978-3-319-24277-4_12

Wright, P.E. & Dyson, H.J., 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol. 16,* 18–29. https://doi.org/10.1038/nrm3920

Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Uversky, V.N. & Obradovic, Z., 2007. Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions. *J. Proteome Res. 6,* 1882–1898. https://doi.org/10.1021/pr060392u

Xie, Y., Li, Huiqin, Luo, X., Li, Hongyu, Gao, Q., Zhang, L., Teng, Y., Zhao, Q., Zuo, Z. & Ren, J., 2022. IBS 2.0: an upgraded illustrator for the visualization of biological sequences. *Nucleic Acids Res. 50,* W420–W426. https://doi.org/10.1093/nar/gkac373

Xue, B., Brown, C.J., Dunker, A.K. & Uversky, V.N., 2013. Intrinsically disordered regions of p53 family are highly diversified in evolution. *Biochim. Biophys. Acta Proteins and Proteomics*, *1834,* 725–738. https://doi.org/10.1016/j.bbapap.2013.01.012

Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K. & Uversky, V.N., 2010. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta Proteins and Proteomics*, *1804,* 996–1010. https://doi.org/10.1016/j.bbapap.2010.01.011

Yarberry, W., 2021. *CRAN Recipes: DPLYR, Stringr, Lubridate, and RegEx in R.* Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6876-6

Yruela, I., Oldfield, C.J., Niklas, K.J. & Dunker, A.K., 2017. Evidence for a strong correlation between transcription factor protein disorder and organismic complexity. *Genome Biol. Evol. 9,* 1248–1265. https://doi.org/10.1093/gbe/evx073

Zand, R., Jin, X., Kim, J., Wall, D.B., Gould, R. & Lubman, D.M., 2001. Studies of posttranslational modifications in spiny dogfish myelin basic protein. *J. Neurosci.* https://link.springer.com/article/10.1023/A:1010921230859

Zarin, T., Strome, B., Nguyen Ba, A.N., Alberti, S., Forman-Kay, J.D. & Moses, A.M., 2019. Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *eLife 8,* e46883. https://doi.org/10.7554/eLife.46883

Zavialova, M.G., Zgoda, V.G. & Nikolaev, E.N., 2017. Analysis of the role of protein phosphorylation in the development of diseases. *Biochem. Moscow Suppl. Ser. B 11,* 203–218. https://doi.org/10.1134/S1990750817030118

Zhang, C., Walker, A.K., Zand, R., Moscarello, M.A., Yan, J.M. & Andrews, P.C., 2012. Myelin basic protein undergoes a broader range of modifications in mammals than in lower vertebrates. *J. Proteome Res. 11,* 4791–4802. https://doi.org/10.1021/pr201196e

Supplementary materials

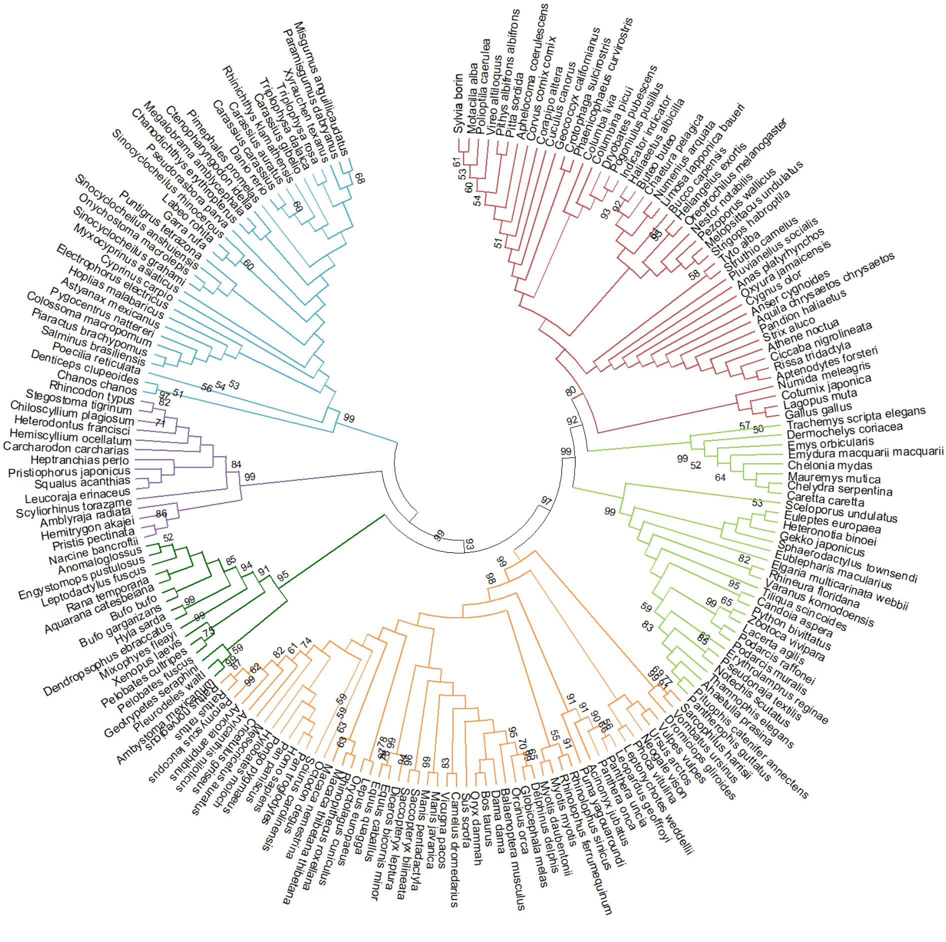

Figure S1. Neighbor-joining phylogenetic tree of representative vertebrate species based on protein sequence alignment. The tree was constructed using the Neighbor-Joining (NJ) method in MEGA version 12. Support for each node was assessed with 500 bootstrap replicates, with values shown at key nodes. Clades are color-coded as follows: purple – *Chondrichthyes*, blue – *Teleostei*, dark green – *Amphibia*, light green – *Reptilia*, red – *Aves*, orange – *Mammalia*.

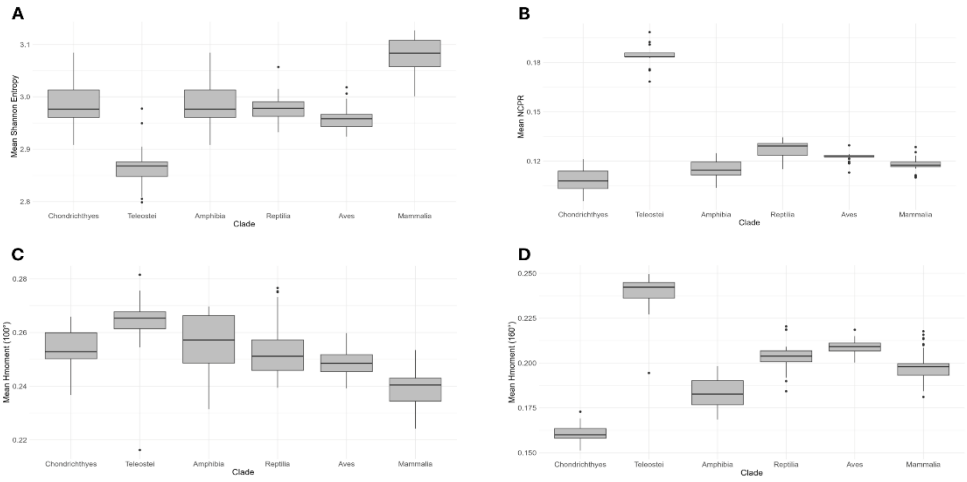

Figure S2. Comparison of biophysical properties of the 18.5 kDa MBP across six vertebrate clades (*Chondrichthyes, Teleostei, Amphibia, Reptilia, Aves,* and *Mammalia*). Boxplots show (A) mean Shannon entropy, (B) net charge per residue (NCPR), (C) mean hydrophobic moment at 100°, and (D) mean hydrophobic moment at 160°.

Table S1. Predicted phosphorylation sites of the MBP 18.5 kDa isoform across vertebrate clades (confirmed by ≥2 predictors: PONDR DEPP, Musite, NetPhos). In the table, letters represent amino acid one-letter codes, while numbers indicate the position within the consensus sequence.

Chondrichthyes	***Teleostei***	***Amphibia***	***Reptilia***	***Aves***	***Mammalia***
3S	5S	3S	3S	3S	3S
5S	6T	5S	7S	7S	8S
6T	7S	6T	8S	8S	13S
7S	10S	7S	13S	13S	15Y
18S	42S	18S	17T	17S	18S
23S	48S	23S	19S	19S	20S
38S	49S	38S	20T	21T	21S
50T	50S	41S	28S	28S	36T
52S	57S	46S	40S	34S	57S
59T	58S	66S	58Y	39S	70Y
70T	59T	69S	66S	55S	72S
115Y	74S	71S	70S	66S	97T
120S	77S	75Y	99S	71S	100T
123S	78S	77S	103T	76S	104S
127Y	83S	78S	114S	96S	112S
134S	93S	79S	119S	99T	117S
136T	95S	84S	124T	111S	134S

Chondrichthyes	*Teleostei*	*Amphibia*	*Reptilia*	*Aves*	*Mammalia*
138S	99S	88S	129Y	116S	138S
139S	101S	110T	133S	126Y	153T
150S	104S	113T	149Y	129S	155S
	116S	117S	155T	145S	165S
	124S	124S	165S	146Y	167S
	126S	129S	167S	154S	169S
	128S	145Y	171S	162S	
	134S	147S	173S	164S	
	135T	148S	175S	168S	
		164Y		170S	
		176S		172S	
		186S			
		188S			

Table S2. Conservation of experimentally identified citrullination sites of the human MBP 18.5 kDa isoform across aligned vertebrate consensus sequences. In the table, letters represent amino acid one-letter codes, while numbers indicate the position within the consensus sequence. Conserved residues are shown in bold, and substitutions that remain positively charged are highlighted in red.

Homo sapiens	*Chondrich-thyes*	*Teleostei*	*Amphibia*	*Reptilia*	*Aves*	*Mammalia*
R6	0	0	L6	**R6**	**R6**	**R6**
R10	0	0	G10	**R10**	**R10**	**R10**
R26	A13	P11	**R28**	**R25**	**R25**	**R26**
R32	**R19**	K17	**R38**	**R31**	H31	**R32**
R34	**R21**	**R19**	**R40**	**R32**	**R32**	**R34**
R44	K31	**R29**	**R50**	**R43**	**R42**	**R44**
R50	G37	K35	**R56**	K49	**R48**	**R50**
R55	K44	K38	K61	**R54**	**R53**	**R55**
R66	L54	S42	S73	**R64**	**R64**	**R66**
R80	**R68**	**R61**	**R96**	**R83**	**R80**	**R81**
R98	K86	P79	**R114**	**R101**	**R98**	**R99**
R114	K104	0	**R129**	**R117**	**R114**	**R115**
R123	**R113**	S93	K140	K126	K123	K124
R131	S122	S99	S149	K135	K131	**R132**
R160	K147	S124	**R186**	S166	S162	**R163**
R163	G150	**R127**	**R189**	**R169**	**R165**	**R166**
R170	0	0	**R198**	**R180**	**R176**	**R173**
R171	0	0	**R199**	**R181**	**R177**	**R174**