Structural flexibility and heterogeneity of recombinant human glial fibrillary acidic protein (GFAP)

Dea Gogishvili[*1,2], Eva Illes-Toth[3], Matthew J Harris[3], Christopher Hopley[3], Charlotte E Teunissen[4,5], and Sanne Abeln[1,2]

[1]Bioinformatics, Computer Science Department, Vrije Universiteit Amsterdam, The Netherlands
[2]AI Technology for Life, Department of Computing and Information Sciences, Department of Biology, Utrecht University
[3]National Measurement Laboratory at Laboratory of the Government Chemist (LGC), Teddington, United Kingdom
[4]Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands
[5]Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands
[*]d.gogishvili@vu.nl

ABSTRACT

Glial fibrillary acidic protein (GFAP) is a promising biomarker for brain and spinal cord disorders. Recent studies have highlighted the differences in the reliability of GFAP measurements in different biological matrices. The reason for these discrepancies is poorly understood as our knowledge of the protein's 3-dimensional conformation, proteoforms, and aggregation remains limited. Here, we investigate the structural properties of GFAP under different conditions. For this, we characterised recombinant GFAP proteins from various suppliers and applied hydrogen-deuterium exchange mass spectrometry (HDX-MS) to provide a snapshot of the conformational dynamics of GFAP in artificial CSF compared to the phosphate buffer. Our findings indicate that recombinant GFAP exists in various conformational species. Furthermore, we show that GFAP dimers remained intact under denaturing conditions. HDX-MS experiments show an overall decrease in H-bonding and an increase in solvent accessibility of GFAP in aCSF compared to the phosphate buffer, with clear indications of mixed EX2 and EX1 kinetics. To understand possible structural interface regions and the evolutionary conservation profiles, we combined HDX-MS results with the predicted GFAP-dimer structure by AlphaFold-Multimer. We found that deprotected regions with high structural flexibility in artificial CSF overlap with predicted conserved dimeric interface regions. Our results suggest that GFAP exists in several conformational forms. Structural property predictions combined with the HDX data show an overall deprotection and signatures of aggregation in aCSF. We anticipate that the outcomes of this research will contribute to a deeper understanding of the structural flexibility of GFAP and ultimately shed light on its behaviour in different biological matrices.

Introduction

Glial fibrillary acidic protein (GFAP) is a type-III intermediate filament protein, a well-recognised marker for astrocytes in the central nervous system[1]. Like other intermediate filament proteins vimentin, desmin, and peripherin, GFAP is a structural protein involved in organising the cytoskeleton, shaping and maintaining the mechanical strength of cells. Additionally, GFAP is involved in various other cellular processes, including cell motility and migration[2], cell proliferation[3,4], exocytosis and vesicle mobility[5], synapse formation[6], neuronal plasticity and neurite outgrowth[6–8]. Furthermore, GFAP is essential for the integrity of BBB[9] and maintenance of myelination[9,10]. GFAP is similar to other intermediate filament proteins with a shared four α-helical 1A, 1B, 2A, and 2B segments flanked by the disordered amino- (head) and carboxy-terminal (tail) domains that largely vary in sequence[11,12]. Under physiological conditions, type III intermediate filament proteins assemble into large oligomers visible by electron microscopy. The GFAP rod 1B domain was proposed to form a homotetrameric structure of two parallel coiled coils stabilised by salt bridges and hydrophobic interactions[12]. However, the 1B domain is a relatively small fraction of the protein and the native assembly of the full-length GFAP remains unclear.

GFAP is a biomarker for reactive astrogliosis and is involved in a wide range of central nervous system pathologies[13,14]. Traumatic brain injury and glial scar formation have been linked with increased concentration of GFAP[15–17]. Multiple studies demonstrated that GFAP and vimentin knockout mice are more susceptible to brain injuries and ischemic brain damage[18,19] suggesting a protective astrocytic function of GFAP. Various gain-of-function mutations in GFAP and its unique cytoplasmic

inclusions with small HSPs are well-known characteristics of Alexander's disease[20–23]. GFAP concentration in the brain has been found to negatively correlate with cognitive function and is considered a potential early marker of neurodegeneration[24]. Elevated GFAP levels in serum have been observed in various neurodegenerative disorders, including Alzheimer's disease (AD), dementia with Lewy bodies, and Parkinson's disease[25]. Astrogliosis was shown to correlate with AD pathology as GFAP mRNA levels are increased together with amyloid-β deposits[26]. Correlations have been found between blood GFAP levels and amyloid-β burden[27,28], as well as cognitive deterioration and disease severity in AD[29]. Moreover, increased levels of plasma GFAP have been detected in cognitively healthy individuals at risk of AD, indicating early astrocytic damage in the presymptomatic stages of the disease[30].

Typically, CSF biomarkers are more reliable indicators of cerebral pathologies than blood biomarkers[31,32]. Recent studies in AD samples indicate that GFAP measurements in different body fluids do not agree with each other and that GFAP measurements in plasma are more accurate in discriminating amyloid-β profiles[28,32]. Furthermore, CSF GFAP was shown to be susceptible to freeze-thaw cycles compared to blood GFAP[32]. A potential reason for this discrepancy could be linked to the flexibility and structural dynamics and the formation of higher-order species. However, none of these structural characteristics are currently well understood. Building on the existing knowledge of GFAP, here we pose the following questions: Are there any signs of oligomerisation or structural changes to be found under different conditions, and what are the structural determinants of those? To address this, first, we characterised recombinant GFAP proteins from different commercial sources using intact protein analysis as well as proteolytic digestion and LC-MS analysis. Second, we use hydrogen-deuterium exchange mass spectrometry (HDX-MS) to investigate the solvent accessibility and hydrogen bonding of GFAP in artificial CSF (aCSF) compared to the phosphate buffer. Additionally, we investigated how well the deprotection signatures from the HDX experiments agreed with the predicted monomeric and dimeric structures using AlphaFold. Overall, we observed a strong deprotection (decrease in H-bonding and increase in solvent accessibility) of GFAP in aCSF with additional evidence of aggregation in distinct regions.

Results

Figure 1 illustrates our step-by-step approach to characterising recombinant GFAP. First, we compared recombinant proteins obtained from three different commercial sources. For this, we conducted intact protein analyses and performed proteolytic digestion to confirm the reported sequences. After choosing the optimal protein preparation we proceeded to explore GFAP structure in different matrices using hydrogen-deuterium exchange mass spectrometry and used computational tools to understand the implications of structural regions and their potential functional roles.

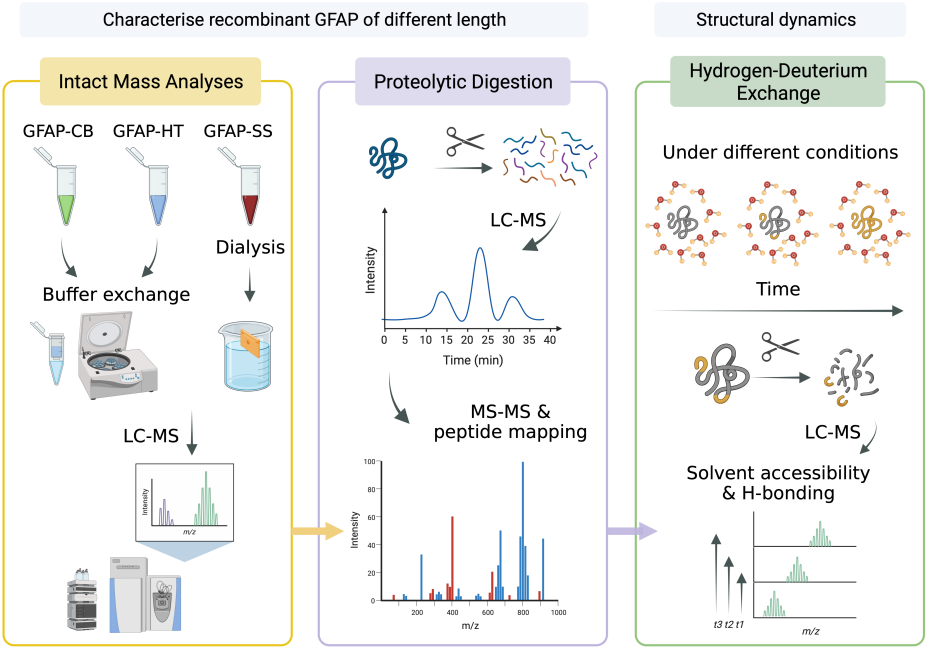

Figure 1. Outline of the study. In this study we first investigated recombinant GFAP from different sources using intact mass analysis and proteolytic digestion to confirm reported sequences. After selecting the optimal preparation, based on our results, we moved on to investigate the structural changes of GFAP under different conditions.

Intact protein analysis by LC-MS

We conducted intact protein analyses of GFAP to investigate and compare recombinant protein preparations obtained from three different commercial sources: Cambridge BioScience (CB), HyTest (HT), and Synaptic Systems (SS). We refer to these recombinant proteins as GFAP-CB, GFAP-HT, and GFAP-SS respectively. In line with the gel electrophoresis results, the intact LC-MS analysis of the recombinant GFAP revealed the presence of two co-existing protein species, spanning monomeric and dimeric forms for all preparations (Figures 2, S1, S2, and S3). During the ionisation process, proteins can accommodate electric charges, resulting in different charge states. The charge state distribution (CSD) refers to the varying number of electric charges associated with ionised forms of the protein. The deconvolution algorithm assigned monomeric and dimeric species GFAP-CB exhibiting a CSD ranging from 10^{+}- 25^{+} and 32^{+}- 50^{+}, respectively (Figure S2A-B). Importantly, urea occurs in the formulation of GFAP-CB and it is known to spontaneously dissociate to form cyanate and ammonia in aqueous solutions leading to a chemical modification known as carbamylation[33–35]. Inspection of the high-resolution intact masses revealed the presence of three mass shifts by approximately 43 Da as compared with the non-modified form of the protein with a corresponding deconvoluted mass of 19950.0144 Da (relative abundance 33.6%). The most abundant form was assigned to be a carbamylated form of the protein, with a corresponding mass of 19993.0226 Da (relative abundance 100%), followed by the addition of two carbamyl groups corresponding to 20036.0289 Da (abundance 82.8%) (Figure S2D).

Intact protein analyses of the recombinant GFAP-HT by LC-MS showed the presence of mono- and multimeric forms of the protein (Figure 2). The results of our analyses showed three distinct peaks, with peak I representing the most accessible state of the protein and exhibiting a charge state distribution (CSD) ranging from 35^{+}- 50^{+} (Figure 2B). Peak II corresponded to the intermediate state of GFAP, with a CSD of 41^{+}- 50^{+} (Figure 2C), while peak III represented the most compact conformation with a CSD of $M43^{+}$- 50^{+} (Figure 2D) according to the Xtract algorithm, we observed multimodal distributions, in particular in the mass spectrum of peak III suggested the presence of higher-order species. Using mass deconvolution, we identified a possible mass shift of ~1 Da between the most abundant monoisotopic masses for peak I and peak III ((37882.2304 (100%); 37882.3158 (100%) respectively) and peak II (37881.2015 (100%)). The mass shift of ~1 Da could be the result of a chemical modification. For instance, a possible cause of a ~1 Da loss modification is amidation, which is the formation of an amide on the C-terminus (-0.9840 Da). Another possibility is oxidative deamination of lysine, in which α-aminoadipic-delta-semialdehyde forms inter- and intra-molecular cross-links with other allysine or lysine residues[36].

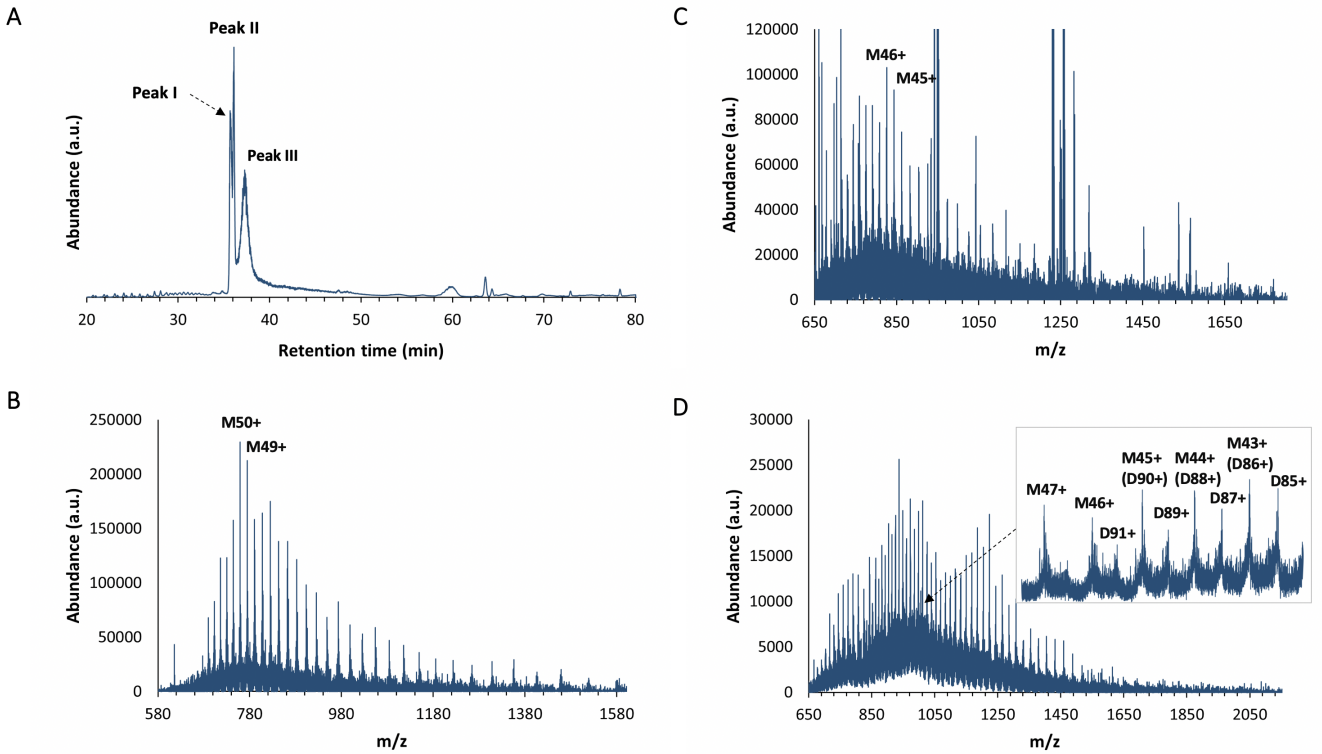

Figure 2. Chromatogram and mass spectra of recombinant GFAP from HyTest. Chromatogram (**A**), and mass spectra of peak I (**B**), peak II (**C**) and peak III (**D**). Charge state distributions (CSD) for peak I (35^{+}- 50^{+}), peak II (41^{+}- 50^{+}), and peak III ($M43^{+}$- 50^{+}) were automatically assigned by the software (Xtract).

Prior to the LC-MS analyses experimental procedure for the full-length GFAP-SS the protein was dialysed. Nevertheless, the chromatogram did not provide distinct, well-resolved peaks. We considered the three most dominant peaks: Peak I (Figure S3) represented the most open conformation, while peaks II and III (Figure S3 C-D) corresponded to more compact conformations based on the spread of charge state distributions. It is important to note that all mass spectra exhibited clear multimodal distributions, suggesting the presence of different species. Nevertheless, it was challenging to deconvolve the mass spectra, possibly due to the complexity of the sample and the overlapping multiple protein species. Peak I of the chromatogram contained a dominant protein species with a monoisotopic mass of 37960.0 Da and its dimers as indicated by deconvolution in UniDec[37]. This mass is inconsistent with both the expected mass of the full-length protein (54070 Da) and the corresponding protein bands observed on the gel (Figure S1). The observed species might be the fragment of the full-length protein that dominates the mass spectra. Additionally, we performed intact protein LC-MS analysis on buffer-exchanged GFAP-CB and GFAP-HT in water (Figure S4). For GFAP-CB peak II became dominant and peak I was absent from the profile. Similarly, for GFAP-HT peak III was present and peaks I and II were absent from the profile. These results suggest, that when we buffer exchange the proteins, the protein species with the most extended monomeric conformations disappear as they are no longer stabilised by the reagents such as urea or TWEEN20.

Proteolytic digestion and LC-MS analysis

To facilitate the comparison of protein sources and select the optimal preparation for the subsequent structural analysis, it is crucial to confirm the supplied sequences. Therefore, we performed proteolytic digestion on recombinant GFAP obtained from three different sources followed by LC-MS/MS analysis. Proteins were enzymatically digested with chymotrypsin and Asp-N separately and analysed with LC-MS/MS to confirm the sequence reported by the manufacturers.

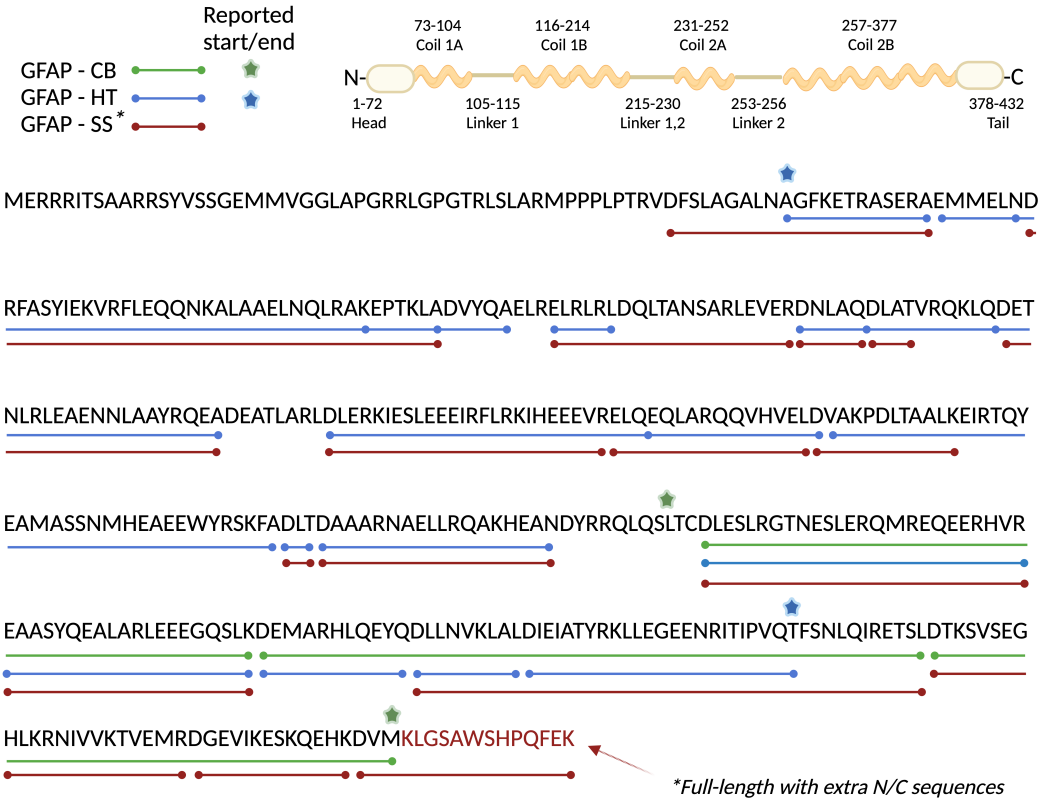

Figure 3. Peptide mapping of the recombinant GFAP from CB, HT, and SS. A peptide map of the proteins was obtained following the enzymatic digestion of the three GFAP proteins. The peptides were separated by a UHPLC Vanquish system on a C18 column coupled to an Orbitrap Q-Exactive Plus. The identified peptides are mapped against the supplied sequences from the manufacturers (see supporting information). The identified proteolytic peptides of the GFAP-CB are depicted in green, GFAP-HT in blue, and GFAP-SS in red. An additional profiled sequence is annotated in red at the end of the full-length GFAP-SS. The green and blue stars indicate the reported start and the end of recombinant proteins, GFAP-CB and GFAP-HT.

Figure 3 shows the peptide coverage obtained for GFAP from the three different sources. BioPharma Finder was used

to analyse the data and compare it to theoretical coverage, which is the fraction of the full-length GFAP sequence covered by the respective recombinant GFAP. The sequence coverage was found to be 28% (Asp-N) and 27% (chymotrypsin) for the GFAP-CB (theoretical coverage 33%), 67% (Asp-N), and 73% (chymotrypsin) for the GFAP-HT (theoretical coverage 75%) and 68% (Asp-N) and 12% (chymotrypsin) for the full-length GFAP-SS (theoretical coverage 100%). The GFAP-SS also contains a theoretical sequence of 23 amino acids and a His-tag on the N-terminus. However, this fragment was not identified under any of the digestion conditions reported here. Moreover, the N-terminal sequence GFAP-SS could not be fully confirmed and the N-terminal of GFAP-CB possibly contained an extra sequence that was not reported. Based on the obtained results from intact LC-MS, proteolytic digestion, and subsequent LC-MS, we selected GFAP-HT for further investigations.

Hydrogen-Deuterium-eXchange (HDX)-MS analysis

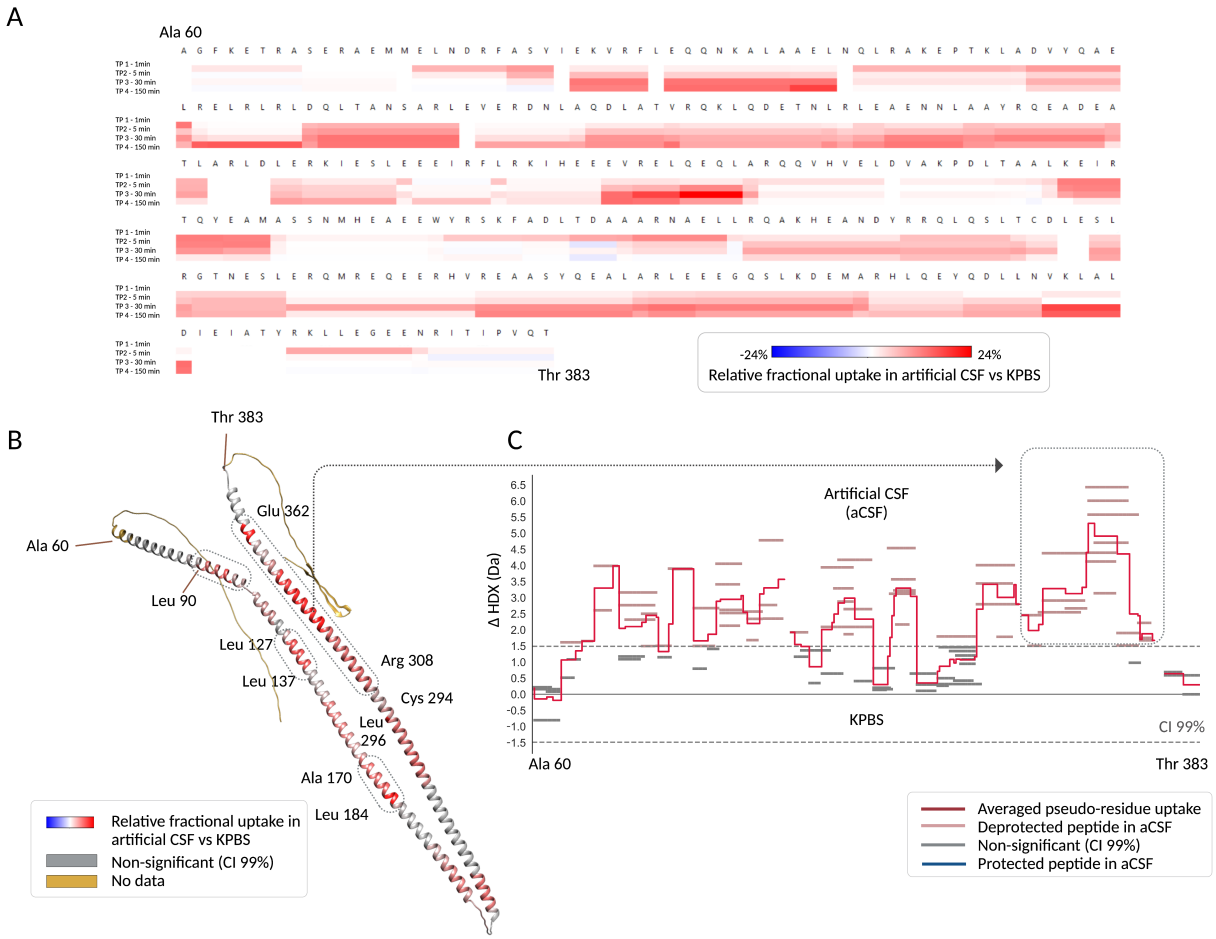

Figure 4. Continuous HDX-MS data of GFAP in artificial CSF compared to phosphate buffer (KPBS). (**A**) Heatmap showing identified peptides across the sequence of GFAP-HT and relative fractional uptake (RFU) at each time point. (**B**) The average deuterium uptake in aCSF versus KPBS mapped onto the full-length AlphaFold structure of GFAP. (**C**) Woods plot depicting the mass changes of individual residues of GFAP-HT in aCSF versus KPBS. Experimental data are shown on the peptide level where every bar is a discrete peptide. Calculated data on the residue level are shown with the line in red where each data point is the average deuterium uptake value per residue spanning the corresponding peptide. Detailed RFU information in the format of a butterfly plot, difference plot, and Woods plots for each individual time point are depicted in Figures S5 and S6.

We proceeded to investigate the structural dynamics of GFAP-HT under three conditions, namely KPBS buffer and two setups in aCSF. Hydrogen-Deuterium-eXchange (HDX-MS) analysis was carried out to investigate the structural changes of GFAP-HT invoked by KPBS and aCSF. HDX-MS provides temporally resolved snap-shots about the solvent accessibility and H-bonding of proteins at a regional peptidic level. There are multiple other factors that influence HDX, including but

not limited to local primary sequence, temperature, and pH[38]. This technique is often performed as a differential experiment between two or more states including alternate conditions and/or binding partners. A peptide with less H-bonding and with high solvent accessibility indicates a more exposed region accessible to deuterated buffer, whereas another one with lower solvent accessibility or decreased solvent accessibility suggests protection from an exchange. Secondary structure has an important influence on the level of uptake; unstructured (less H-bonding, more solvent accessibility) regions often exhibit dynamic motions and high deuterium uptake, in contrast to tightly folded (more H-bonding, less solvent accessibility), buried regions with low uptake. To achieve residue-level information, we used a pseudo-residue averaging technique during data processing[39].

GFAP-HT shows a strong overall deprotection in aCSF versus KPBS

HDX-MS results showed a strong overall deprotection of GFAP-HT in aCSF vs KPBS buffer conditions (Figures 4, S5). We were intrigued by these findings because we expected aggregation to occur which should have caused an increase in H-bonding and a decrease in solvent accessibility, therefore an increase in protection. We explored further by investigating and modifying the artificial CSF conditions. Note that in one condition, the protein was in the original formulation and a KPBS buffer, whereby a large amount of aCSF was added during the course of the HDX-MS experiment as part of the labelling buffer compared to the control, where GFAP-HT was in KPBS buffer at all times. We next explored if buffer exchanging GFAP into aCSF would reveal further information about the conformational rearrangements. In this case, the initial KPBS buffer is exchanged with aCSF prior to the experiment. Therefore, we subsequently explore the third condition, which we refer to as the equilibrated aCSF (eq-aCSF) state. When we compared GFAP in aCSF versus eq-aCSF, we could observe more protection at the earlier time points (1 min and 5 min exposure) and deprotection at the later time points (30 min and 150 min exposure), however, most of the peptides are non-significant (Figure S7). Notably, we saw that there was no substantial difference in the overall protection/deprotection between aCSF and eq-aCSF (Figure S7) and the difference between KPBS and eq-aCSF shows a similar pattern as compared to aCSF (Figure S6).

"EX1-like" kinetics indicate aggregation hotspots

Next, we assessed the structural changes of GFAP-HT under aCSF conditions by examining the exchange profiles of individual peptides (Figures 5, S8, S9, S10, S11). Figure S9 illustrates an example of a peptide manifesting in "EX2 kinetics" where deuterium is gradually incorporated over time via breathing motions in every state, most commonly observed in continuous HDX-MS experiments. Such gradual uptake suggests the peptide is undergoing dynamic motions occurring faster than the labelling rate[40]. On the other hand, Figure S10 represents an example of "EX1-like" kinetics under conditions of the eq-aCSF. Such patterns typically occur when dynamic conformational changes via global unfolding happen slower than the labelling rate or multiple conformers exist under equilibrium[40]. We observe a clear multimodal distribution emerging at 1 min labelling time in the case of eq-aCSF, which remains stable until the last time point of labelling, suggesting the persistent presence of both protected and deprotected (more exposed) structures from the beginning of the labelling until the longer time points. Intact analyses revealed that the GFAP dimers were highly resistant to chaotropic agents and organic solvents. It is possible that some exchange could have been due to monomers derived from dimeric or higher order GFAP species with an alternate conformation yielding EX1-like kinetics contributing to a mixed pattern following 1 min deuteration in eq-aCSF. Figure 5 manifests in EX1-like kinetics for both the aCSF and eq-aCSF conditions with emerging bimodal HDX patterns at the longer labelling times. Another example (Figure S11) of the HDX exchange pattern shows that there were different conformational forms even in KPBS. This is in line with our results of the intact LC-MS analysis in which we observed dimers that appeared to be resistant to urea. This is supported by the presence of the dimers both observed by intact LC-MS mass analysis (Figure 2) and gel electrophoresis (Figure S1).

Observing such multimodal HDX exchange patterns of eq-aCSF is especially interesting concomitantly with an overall decrease in H-bonding and an increase in the solvent accessibility of GFAP-HT. Importantly, the overall differential uptake is calculated from the centroids of the isotopic envelopes of the peptides that the software assigns leading to erroneous/inaccurate values. Our data show that GFAP-HT is more exposed overall in aCSF and eq-aCSF but simultaneously exhibits conformational hotspots of aggregation and features multiple conformational forms that may derive from pre-formed aggregates in solution e.g. dimers, exhibiting an EX1-like exchange pattern throughout all time points.

Predicted GFAP-dimer combined with deuterium uptake results reveal deprotected interface regions

Due to the recent development of the AlphaFold[44], it is also possible to predict multi-chain complexes protein interfaces with the AlphaFold-Multimer[41]. GFAP-HT sequences were searched using ColabFold[45]. Figure 6A shows high confidence in structured regions, while the flanking and inner loop regions are predicted with lower confidence. Figure 6B shows GFAP-HT dimer colour-coded based on the RFU across all time points in aCSF versus KPBS. Our HDX results combined with the AlphaFold-Multimer predicted structure revealed the deprotected interface fragments of the protein in aCSF. Notably, we observe an overall deprotection of the GFAP-HT in aCSF and simultaneously hotspots of aggregation. These results suggest

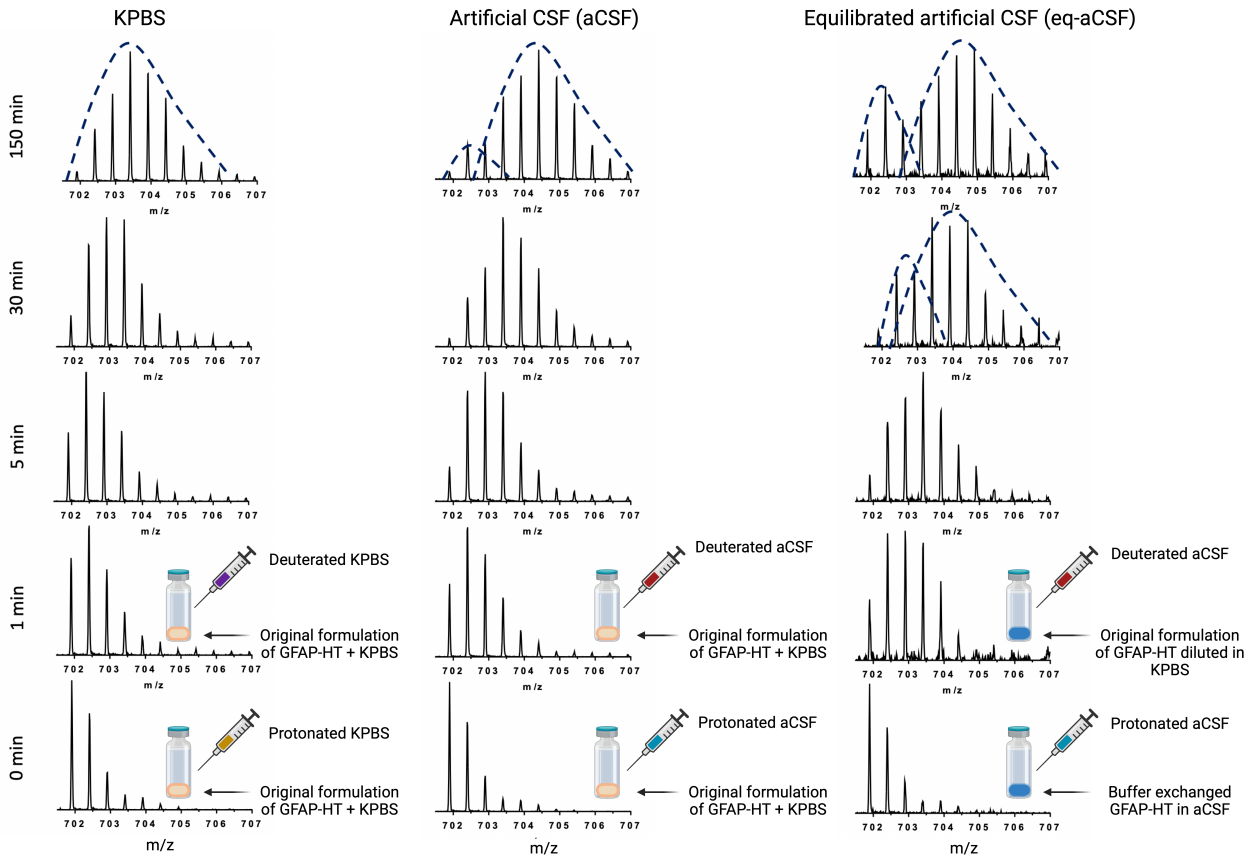

Figure 5. Temporally resolved HDX-MS mass spectra of GFAP – case example. Relative deuterium uptake of the peptide -ATVRQKLQDETN (90-101, 1402.728 Da, 2^{+}) and respective individual mass spectra at each time point and conditions. GFAP-HT in KPBS displays EX2 kinetics and a dynamic profile. In contrast, GFAP in aCSF and eq-aCSF shows evidence also for EX1-like kinetics, with two main species, a more protected and an exposed, at later labelling times. For detailed information see Figure S8.

that, when more monomeric subunits are present, the interface regions that are typically protected by dimers can show less H-bonding and be more accessible to the solvent leading to higher relative fractional uptake.

Deprotected regions of GFAP-HT in artificial CSF are highly conserved and overlap with the coiled-coil region

We were also interested in the relationship between deprotected regions and evolutionary conservation profiles. To reveal important structural and functional regions, we used ConSurf tool[46], which uses multiple sequence alignment to reconstruct a phylogenetic tree. We further overlaid the conservation predictions with the secondary structural property predictions (helix, strand, coil, disorder, exposed/buried) by NetSurfP3.0[42] and compared this to the relative deuterium uptake data of aCSF versus KPBS. Figure 6C shows a clear similarity between the regions that are highly conserved and also show deprotection in aCSF versus KPBS. Furthermore, since coiled-coil structural domains typically mediate various biological interactions, we used a deep neural network-based tool DeepCoil2 to predict the probability of forming such structures on GFAP[47–49]. Figures 6C and S12 suggest, that structurally and functionally important regions, that typically participate in interactions, either forming dimers/tetramers or interacting with other partners, become exposed under aCSF conditions.

Discussion

GFAP has shown potential as a biomarker for various neurodegenerative diseases, but its implementation in the clinical setting is limited by several factors. One of these factors is our incomplete understanding of the biomarker's structure and conformational forms under different conditions. In this study, we sought to investigate and compare the properties of recombinant GFAP protein preparations obtained from three different commercial sources. We confirmed the sequences with proteolytic digestion and investigated the original formulations by intact protein analysis, ultimately selecting GFAP-HT for obtaining a picture of

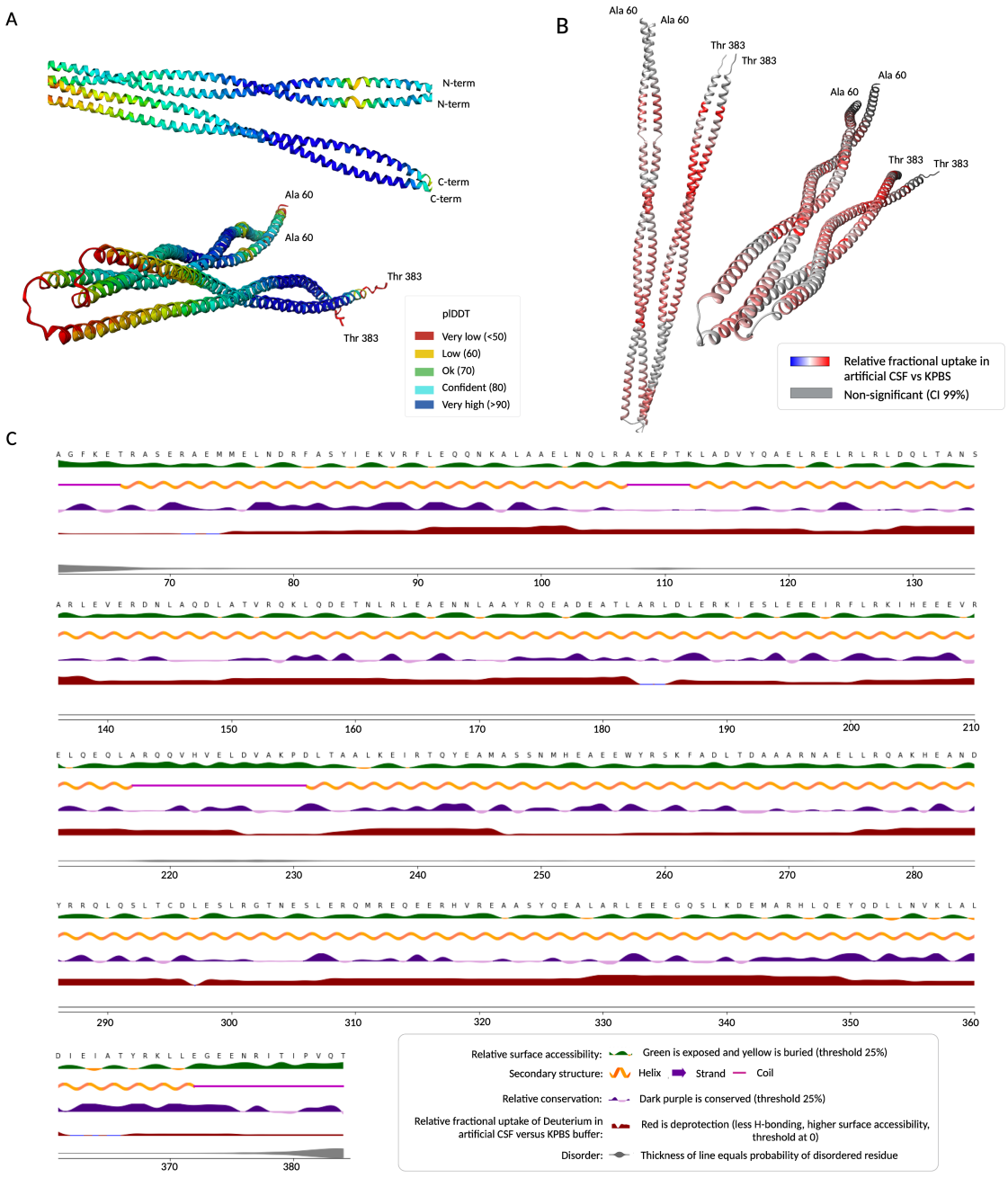

Figure 6. Predictions of the protein complex, secondary structural components, and conservation profiles. (**A**) Predicted structure of a recombinant GFAP-HT dimer by AlphaFold-Multimer[41] and colour-coded with the prediction confidence. (**B**) Predicted structure of a recombinant GFAP-HT dimer colour-coded based on relative fractional uptake (RFU) of GFAP monomers in aCSF vs KPBS across all time points. (**C**) Secondary structural elements, relative surface accessibility, and disorder displayed on a sequence of GFAP-HT predicted by NetSurfP 3.0[42]. Conservation is predicted by ConSurf[43]. Experimentally obtained pseudo-residue relative fractional uptake values in aCSF versus KPBS buffer are displayed in red.

structural flexibility by HDX-MS. HDX-MS data showed strong overall deprotection of the GFAP-HT in aCSF compared to KPBS in addition to region-specific EX1-like behaviour pointing towards aggregation.

The investigation of GFAP at the regional level revealed individual peptides, which exhibited multimodal HDX exchange patterns in all states, in particular under eq-aCSF conditions. In essence, we observed a decrease in H-bonding and an increase in solvent accessibility of GFAP-HT, while simultaneous detection of EX1-like exchange kinetics suggested the

formation and existence of pre-formed aggregates that became more observable under the conditions of eq-aCSF. As shown in Figure 5 under eq-aCSF conditions, the protein is dissolved in artificial CSF resulting in the formation of aggregates that are visible at earlier time points. Significant changes in the overall structure of a protein typically occur gradually and can be detected in HDX data by distinct HDX populations with multimodal isotopic distributions[50]. This behaviour of interconversion between conformations, referred to as EX1 kinetics, is indicative of slow large-scale structural rearrangements and/or multiple conformers[40]. However, detecting EX1 kinetics is not straightforward and requires careful inspection of the data. While centroid fitting is a common method for analysing HDX data and is suitable for EX2 kinetics, it is insufficient for analysing EX1 kinetics. Instead, binomial fitting is more sensitive and can detect the presence of multiple similar HDX states and characterise their exchange profiles through bimodal deconvolution[40,51]. Additionally, HDX-MS is an averaging method and is unable to discern the contribution of individual conformers without an internal standard or targeted pre-fractionation isolating individual populations. Application of pulsed HDX-MS with a more specific focus on different species could elucidate more about GFAP's self-assembly mechanism[40]. Our HDX data combined with the structural property predictions indicate that we have different conformational species of GFAP and the deprotected regions overlap with the conserved regions. Simultaneously we see higher order species and aggregation hotspots with the complex EX1 kinetics. One possible explanation for this could be a shift in the conformational equilibrium caused by monomers that undergo large structural changes with dynamic movements and motions resulting in increased solvent accessibility of the helical chains and simultaneous aggregation (Figure 7). However, this is a hypothesis and needs to be investigated by further experiments. Our findings provide initial insights into the conformational landscape of GFAP-HT under different in vitro conditions, which with further experimental validation may inform the development of clinical assays.

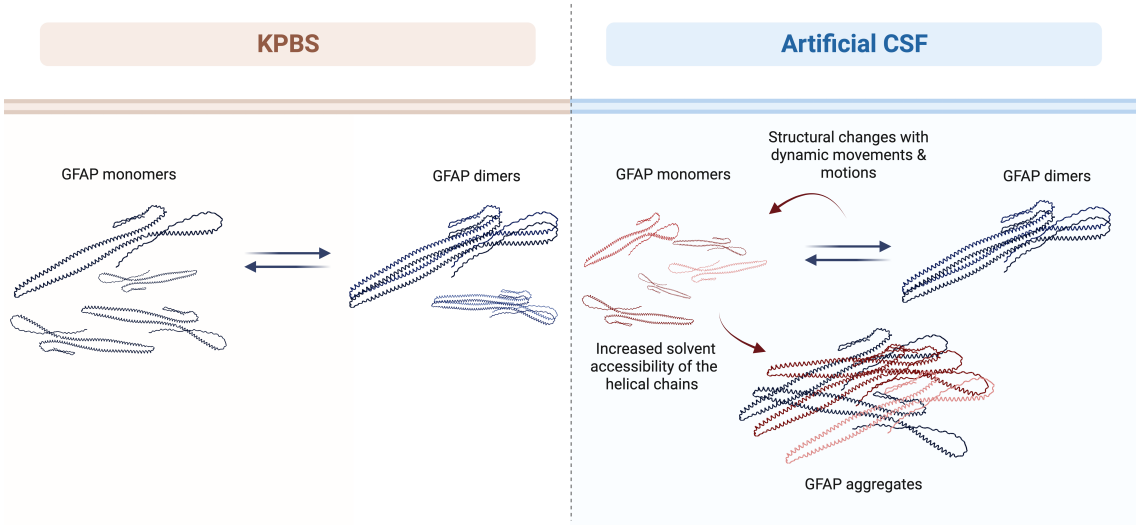

Figure 7. The hypothesis figure The hypothesis figure represents a plausible theory for the structural dynamics of GFAP in different matrices. Under the conditions of aCSF, GFAP appears to be generally deprotected as a result of dynamic refolding events, exhibiting increased solvent accessibility. Simultaneously, HDX-MS data shows EX1-like kinetics in localised regions suggesting aggregation. Conservation profiles and the predicted dimer structure suggest that the rearranged monomeric population in aCSF leads to the formation of higher-order species and amorphous aggregates.

CSF has been recently proposed to be an unreliable matrix for measuring GFAP in AD[32], which is unusual since CSF biomarkers better reflect on cerebral pathologies than blood biomarker[31,32]. Interestingly, blood measurements of GFAP outperform CSF measurements in detecting early amyloid-β pathology[28]. One of the reasons was proposed to root from the stability differences in blood- and CSF GFAP[32] and that GFAP concentration in CSF is more susceptible to the effects of freeze-thaw cycles. This still does not fully explain the discrepancy as GFAP measurements in serum were also superior in fresh samples. Other underlying reasons for the matrix differences were proposed to be due to the discharge mechanisms of GFAP, and its breakdown products, into blood[32]. Based on our results, we hypothesise that the potential unreliability of GFAP measurements in CSF may be attributed to the structural properties and remarkable flexibility of GFAP. We believe that our limited knowledge of the structural dynamics of GFAP, its rearrangement, and the formation of higher-order species under different conditions is a major hurdle for measuring GFAP in different biological matrices. Several open questions could build upon our findings. For example, investigating the full-length protein using HDX-MS analysis could provide a more comprehensive understanding of the conformational changes occurring in GFAP. Furthermore, mimicking conditions

resembling blood, for instance by incorporating serum albumin, could more closely replicate the physiological environment and help to distinguish the conformational differences of GFAP as a biomarker in a clinical setting. More in-depth analysis of GFAP's interactions with other proteins or small molecules could provide insights into potential therapeutic targets for neurodegenerative diseases as intermediate filament-associated proteins play an important role in filament stability or facilitate links to other structures within the cell[14]. E.g. small heat shock protein (HSP) HSP27 was found in GFAP inclusions in the brains of patients suffering from Alexander's disease[20], as well as in mice overexpressing GFAP[52]. Therefore, interacting partners, such as small HSPs, including HSP27, αB-crystallin[53,54], along with other partners, such as Plectin[55], and a family of regulatory proteins 14-3-3[56] could be an interesting avenue to explore.

In this study, we evaluated different recombinant GFAP proteins with the combination of experimental and computational tools for unravelling the structural dynamics and motion of GFAP in different conditions. Overall, our study sheds light on the structural rearrangements and multimeric states of GFAP under various conditions, providing insights into the challenges of developing clinical assays using GFAP as a biomarker. Moreover, we show that GFAP dimers are present even under highly denaturing conditions. Our findings demonstrate the importance of careful evaluation of GFAP protein preparations, and caution manufacturers dedicated to antibody development for different epitopes of isoforms that may be differentially buried or exposed in alternative oligomeric forms. These data highlight the need for a better understanding of GFAP conformational changes in different biological environments. Despite the limitations of this study, including the use of recombinant GFAP protein and the absence of physiological conditions, our results provide a way for future studies to investigate GFAP's structure and dynamics in a more complex and relevant biological context and set the ground for exploratory epitope mapping studies.

Materials and Methods

Computational analysis

HDX-MS data pre-processing and figure generation were performed using Python version 3.9.7. For evolutionary conservation predictions, the ConSurf tool was used[46]. NetSurfP3.0[42] was used to predict residue-level secondary structure, surface accessibility, and disorder. The code was modified to add relative deuterium uptake results as an extra layer. DeepCoil2 to predict the probability of forming coiled-coil structures on full-length GFAP[47–49]. The illustrations in Figures 1 and 7 were created with Biorender.com.

Materials

Chemicals such as urea, and deuterium oxide, sodium chloride were purchased from Honeywell (Seelze, Germany); sodium bicarbonate, potassium chloride, sodium phosphate, monobasic, magnesium chloride, glucose, and calcium chloride were purchased from Merck, formerly known as Sigma-Aldrich (Gillingham, United Kingdom). Potassium phosphate (monobasic, dibasic), formic acid, tris (2-carboxy-ethyl) phosphine (TCEP), dithiothreitol (DTT), ammonium bicarbonate, iodoacetamide (IAA) were purchased from Fisher Scientific (Loughborough, United Kingdom). RapiGest^{TM} (SF Surfactant) and Formic acid (98/100%) were purchased from Fisher Scientific (Loughborough, United Kingdom). Optigrade HPLC Special Grade acetonitrile and ultra-pure water (18 MΩ cm^{-1}) were used. Enzymes chymotrypsin (V1061) and Asp-N (V1621) were purchased from Promega Corporation (Madison, USA). Recombinant Human Glial Fibrillary Acidic Protein (GFAP) derived from E. coli (230-00034, Glu281-Met432) in monobasic sodium phosphate, sodium chloride, imidazole, and 6 M urea was purchased from Cambridge BioScience (Cambridge, United Kingdom). Recombinant GFAP (22/07-11-8G45, 60-383) expressed in E. coli was purchased from HyTest (Turku, Finland) lyophilised in Sodium Hydrogencarbonate, CHAPS, and TWEEN20. Recombinant full-length GFAP (AWB 150675145011) in PBS and 4 M urea was purchased from Synaptic Systems (Goettingen, Germany). Sequence information of supplied recombinant GFAP is provided in the supporting information.

Buffer exchange and dialysis

In total three GFAP solutions of proteins were prepared: 455 pM (GFAP-Cambridge BioScience, GFAP-CB), 900 pM (GFAP-HyTest, GFAP-HT), and 740 pM (GFAP-Synaptic Systems, GFAP-SS). For desalting and buffer exchange and dialysis centrifugal filters, Amicon Ultra 0.5 mL, regenerated cellulose 3,000 NMWL were purchased from Merck Millipore Ltd (Burlington, Massachusetts, United States). Slide-A-Lyzer Dialysis Cassette (10,000 MWCO 0.1-0.5 mL Capacity) was purchased from Thermo Scientific (Bremen, Germany). Protein solutions were loaded in centrifugal filter units and centrifuged at 13,000 rpm for 15 min at room temperature. In total, after 5 runs of buffer exchange, 45 μL of protein solutions were obtained and the concentration was calculated using their respective extinction coefficients (4470 M^{-1} cm^{-1} GFAP-CB, 18910 M^{-1} cm^{-1} GFAP-HT, 25900 M^{-1} cm^{-1} GFAP-SS) using a Nanodrop 2000 spectrophotometer. The final concentrations obtained for GFAP were 3.4 μM (GFAP-CB), and 22.2 μM (GFAP-HT). GFAP-SS was dialysed using a Slide-A-Lyzer Dialysis Cassette (10,000 MWCO 0.1-0.5 mL Capacity) purchased from Thermo Scientific (Bremen, Germany). The cassette was immersed in water for 2 min, and the 100 μL aliquot of protein solution was injected. The cassette was dialysed at room temperature under constant spinning for 5 hours after which the injected protein solution was collected.

Protein gel electrophoresis

For investigating recombinant GFAP proteins from three different sources, protein gel electrophoresis was performed using proteins in their original formulations. 10% Bis-Tris gel electrophoresis NuPAGE™ (Invitrogen) 10% Bis-Tris Gel, NuPAGE (Invitrogen) LDS Sample Buffer (4x), Novex (Life Technologies) MES SDS Running Buffer (20x), NuPAGE™ (Invitrogen) Sample Reducing Agent (10x), Colloidal blue staining kit, Stainer A and Stainer B (Invitrogen), PageRuler™ Plus Prestained Protein Ladder, Mini Gel Tank (Invitrogen Novex) XCELL SureLock™ were purchased from Thermo Fisher Scientific (Bremen, Germany). In total 2.4 μM stocks of GFAP protein samples were prepared and NuPAGE loading dye was added. A 10% Bis-Tris gel was run at 200V for 35 minutes. The fixing and staining procedure was performed, and the gel was destained in 200 mL of deionised water overnight.

Intact protein analysis by LC-MS

To compare the properties of recombinant protein preparations obtained from three different commercial sources, intact protein LC-MS analyses were performed. GFAP solutions were diluted to concentrations of 3.5 μM (GFAP-CB), 0.5 μM (GFAP-HT), and 3.4 μM (GFAP-SS). Samples were injected onto an Acclaim PepMap 100 1.0 mm x 15cm nanoViper C18 column, 3 μm, 100 Å, purchased from Thermo Fisher Scientific (Bremen, Germany). B was ramped from 3% to 60% in 60 min at a flow rate of 0.1 mL/min. A was composed of H_2O with 0.5% FA, and B contained acetonitrile with 0.5% formic acid. MS experiments were performed using a Q-Exactive Plus Orbitrap instrument with a HESI-II probe source (Thermo Fisher Scientific, Bremen, Germany) in positive ionization (nESI) mode. Intact protein mode and high vacuum were switched on. Parameters were as follows: source temperature: 320°C; capillary voltage: 3.5 kV; S-lens RF level: 50%; sheath gas: 25 a.u.; auxiliary gas: 5 a.u.; sweep gas: 2 a.u.; maximum injection time: 250 ms; AGC target value: $3E^6$; Microscan: 1. Data were acquired in 140K resolution mode over a range of 400–4,000 m/z, in intact protein mode. Protein chromatograms and mass spectra were analysed using Thermo Scientific Xcalibur (version 4.2.47). Protein deconvolution was performed using Freestyle software (version 1.5.93.34) using the in-built Xtract algorithm function that determines the monoisotopic mass. Deconvolution was performed in the 10–180 kDa range, considering 25–80 charge states, and a target mass of 19 kDa, 37 kDa, and 54 kDa for GFAP-CB, GFAP-HT, and GFAP-SS respectively. Deconvolution for GFAP-SS was additionally performed using UniDec Engine software (version 3.2.0).

Proteolytic digestion and LC-MS analysis

To confirm the sequence of the purchased proteins, a nominal amount of 0.35 mg/mL (Cambridge BioScience), 0.1 mg/mL (HyTest), and 0.1 mg/mL (Synaptic Systems) recombinant GFAP proteins were first treated withRapiGest (10-30:1 ratio reagent: protein), reduced with DTT at 4mM final concentration and equilibrated for 30 minutes at 56°C using a thermoshaker (Camlab Microtherm, Norman Way Industrial Estate, England). The protein solutions were afterward alkylated with iodoacetamide (IAA) at 10 mM final concentration for 45 min in the dark. Samples were digested for 4 hours by adding Asp-N and Chymotrypsin (ratio 1:50 protein:enzyme for Cambridge BioScience and 1:5.5 for HyTest respectively) using a shaker (Grant-bio-PMS 1000i, Cambridge, United Kingdom) placed into the incubator (VELP Scientifica Srl, Usmate, Italy). Digestion was stopped with 1% formic acid and samples were centrifuged at 13,000 rpm for 10 min. 10 μL of digested GFAP samples were separated with a Vanquish Ultra-High-Performance Liquid Chromatography (UHPLC) on an Acclaim PepMap 100 1.0 mm x 15 cm nanoViper C18 column, 3 μm, 100 Å coupled with an Orbitrap Q Exactive Plus (Thermo Fischer Scientific, Bremen, Germany). A was composed of H_2O with 0.1% FA, and B contained acetonitrile with 0.1% FA. The mobile phases were composed of water and acetonitrile with 0.1% formic acid. The chromatographic gradient was ramped from 3% B to 50% B in 80 minutes and at a flow rate of 60 μL/min. Analysis was performed under standard pressure in peptide mode, the high vacuum was switched off. Peptides were ionised by heated electrospray ionization (HESI) positive using the following parameters: sheath gas flow rate: 25 μL/min; auxiliary gas flow rate: 10 μL/min; AGC target value 3E6; capillary temperature 320°C; spray voltage; 2.5 kV, Microscan: 1. Data were acquired at 140K resolution over a range of 150–2,000 m/z, exclusion time and charges, stepped collision energy voltages of 27, 30 and 33 V. Data analyses were carried out with BioPharma Finder (Thermo Scientific, v2.0) by using the supplied recombinant GFAP sequences (see supporting information). Parameters for peptide mapping were as follows: Absolute MS Signal Threshold: 5000; MS noise level: 250; S/N Threshold: 20; Peak width (min): 2; Time limits: 3-80; Maximum retention time shift (min): 2.53; Maximum peptide mass: 7000; Mass Accuracy (ppm): 5; Minimum confidence 0.8. A maximum number of modifications for a peptide: 1; Specificity: high; Minimum and maximum MS peak width: 1.2-4.2; Proteases Asp-N and chymotrypsin respectively.

Hydrogen-Deuterium-eXchange (HDX)-MS analysis

To investigate structural characteristics of GFAP-HT under different conditions, HDX-MS experiments were performed on a Synapt G2Si HDMS in tandem with an Acquity UPLC M-Class system with HDX and automation (Waters Corporation, Manchester, UK) and a LEAP PAL autosampler (Trajan Scientific Europe Ltd, Milton Keynes, UK) for sample management. The mass spectrometer was calibrated against NaI and sample data were acquired without initial lock-mass correction. ESI-MS

experiments were performed in positive nanoelectrospray ionisation mode, sensitivity mode, 100°C source temperature, 3 kV capillary voltage, 80 V cone voltage, 80 V source offset, and 250°C desolvation temperature. Data acquisition in MSE mode was carried out over a range of 50–2000 m/z, with a scan time of 0.3 s. Data was lock-mass corrected against m/z 556.2771 of LeuEnk (100 fmol solution in 50% methanol, 0.1% formic acid), applied post-acquisition.

The recombinant human protein GFAP from HyTest was solubilised at 1 mg/mL (100 μM) in water and was stored separately at 4°C. GFAP sample was diluted to a final protein concentration of 10 μM with 50 mM KPBS - potassium phosphate buffer, pH 7.4. The first condition of interest for the HDX experiments comprised of protonated 50 mM potassium phosphate (equilibration) buffer (pH 7.4 in H_2O), and 50 mM potassium phosphate (labelling) buffer (pD 7.0, prepared in D_2O respectively). The second condition mimicking CSF conditions was prepared as described above in H_2O (pH 7.4) and in D_2O (pD 7.0) respectively in addition to the control (GFAP in KPBS). Quench solution was prepared with 1 M TCEP and 8 M urea in 100 mM potassium phosphate buffer (pH 2.5). For the direct comparison between the two conditions, artificial CSF was prepared. 5 μl of a protein sample at 10 μM was mixed with 57 μl equilibration buffer or labelling buffer at 20°C. After the appropriate incubation time has elapsed, 50 μl of the sample is transferred into 70 μl quench for 30 sec at 1°C to prevent further deuteration. 80 μl of the sample was then injected and digested on-line by a Waters Enzymate BEH pepsin column (300 Å, 5 μm, 2.1 × 30 mm, Waters, Milford, MA) at 20°C and the subsequent peptides were trapped on a Waters BEH C18 VanGuard pre-column (5 μm, 2.1 mm) for 3 minutes at a flow rate of 70 μl/min in buffer A (H_2O + 0.2% formic acid, pH 2.5). Afterwards, peptides were eluted by a linear gradient of organic solvent, buffer B (acetonitrile + 0.2% formic acid, pH 2.5), from 5-40% over 7 minutes and then separated by UPLC using a Waters BEH C -18 analytical column before being transferred into the mass spectrometer by ESI. All trapping and chromatography were performed at 0°C to minimise the back exchange of deuterium for hydrogen. Last, the sample entered the ToF mass analyser for m/z separation and determination. Each sample was run with a 17-minute acquisition time with most peptides eluting between 6 and 14 minutes. Each run included a wash of the LC with a saw-tooth gradient of buffer B from 5-95% and back again over 2 min and repeated a second time to eliminate carry-over into the next sample runs.

A summary table of the HDX data and the HDX data table are included in the supporting information according to the consensus guidelines[57]. For the generation of peptide maps, MS^E data were analysed with the ProteinLynx Global Server (PLGS) software v3.0.3 (Waters, Milford, MA) with 135 counts for low energy and 30 counts for high energy. Peak lists were searched against the recombinant GFAP sequence from HyTest (see supporting information), pepsin, and randomised sequences. To identify a protein, at least 3 fragment ions per peptide, 7 fragment ions per protein, and 1 peptide per protein should be detected. Enzymatic settings were set to non-specific cleavages and the allowance of 1 missed cleavage. The false discovery rate was maintained below 2%. PLGS output files were imported to DynamX v3.0.0 (Waters, Milford, MA). To generate peptide maps with reproducibly identified peptides, peptides must have been observed in at least 5 out of 6 data files, and have a minimum number of 0.1 product ions identified per amino acid residue. Cluster data files (provided in the supplements) were submitted to Deuteros[58] (v. 2.0) which outputs the differential data into the format of Woods plots. Settings were as follows: deuterium (%): 91.3%, alpha: 0.01. Hybrid statistical filtering was performed to identify peptides that show significant uptake differences between the two states. An in-built blanket confidence interval of 98% (given as 0 ± DU - Deuterium uptake threshold in Daltons) was applied across all peptides of each time point. Significant peptides for each time point were then exported to project the differential uptake data onto a 3D model of the protein. To calculate the uptake differences, peptide data for all time points were summed up and plotted against the pseudo-residue position on the Woods plot. In order to determine peptides with significant differences in deuterium uptake, the standard error of the mean (SEM) was calculated and the CI 99% value of the t-test at 2 degrees of freedom was used to calculate the threshold value in Da.

Data availability

HDX-MS cluster, state data, and summary tables are available as supplementary files. Additional information and/or code related to this work can be obtained by request for purposes of replicating results from the corresponding author.

Acknowledgements

We would like to thank Dr. Milena Quaglia for helping with the conceptualisation of this project, Francesca Robertson for the critical review of the manuscript, and colleagues at LGC Group, namely Luise Luckau, Valentina Faustinelli, and Salomé Coppens for instrument training, technical support, interesting ideas, and suggestions.

Funding

Research of DG, CT, and SA are supported by the European Commission (Marie Curie International Training Network, grant agreement No 860197 (MIRIADE). CT is supported by JPND (bPRIDE)), Health Holland, the Dutch Research Council (ZonMW), Alzheimer Drug Discovery Foundation, The Selfridges Group Foundation, Alzheimer Netherlands, Alzheimer Association.

Competing interests

CT is the recipient of ABOARD, which is a public-private partnership receiving funding from ZonMW (#73305095007) and HealthHolland, Topsector Life Sciences & Health (PPP-allowance; #LSHM20106). More than 30 partners participate in ABOARD. ABOARD also receives funding from Edwin Bouw Fonds and Gieskes-Strijbisfonds. CT has a collaboration contract with ADx Neurosciences, Quanterix and Eli Lilly, performed contract research or received grants from AC-Immune, Axon Neurosciences, Biogen, Brainstorm Therapeutics, Celgene, EIP Pharma, Eisai, Olink, PeopleBio, Roche, Toyama, Vivoryon. CT serves on editorial boards of Medidact Neurologie/Springer, Alzheimer Research and Therapy, Neurology: Neuroimmunology & Neuroinflammation, and is editor of a Neuromethods book Springer. The rest of the authors do not have any competing interests to declare.

Author contributions statement

Conceptualisation DG, EIT, CH, SA; Data collection DG, EIT, MH; Data curation DG, EIT, MH; Funding acquisition: CT, SA; Investigation: DG, EIT, MH; Visualization DG, EIT; Writing – Original draft preparation: DG, EIT, SA; Writing – review & editing: DG, EIT, MH, CH, CT, SA.

References

1. Lowery, J., Kuczmarski, E. R., Herrmann, H. & Goldman, R. D. Intermediate filaments play a pivotal role in regulating cell architecture and function. *J. Biol. Chem.* **290**, 17145–17153 (2015).
2. Yoshida, T. *et al.* The functional alteration of mutant gfap depends on the location of the domain: morphological and functional studies using astrocytoma-derived cells. *J. Hum. Genet.* **52**, 362–369 (2007).
3. Rutka, J. T. *et al.* Effects of antisense glial fibrillary acidic protein complementary dna on the growth, invasion, and adhesion of human astrocytoma cells. *Cancer Res.* **54**, 3267–3272 (1994).
4. Messing, A. *et al.* Fatal encephalopathy with astrocyte inclusions in gfap transgenic mice. *The Am. journal pathology* **152**, 391 (1998).
5. Potokar, M. *et al.* Cytoskeleton and vesicle mobility in astrocytes. *Traffic* **8**, 12–20 (2007).
6. Emirandetti, A., Zanon, R. G., Sabha Jr, M. & de Oliveira, A. L. R. Astrocyte reactivity influences the number of presynaptic terminals apposed to spinal motoneurons after axotomy. *Brain research* **1095**, 35–42 (2006).
7. Rozovsky, I. *et al.* Estradiol (e2) enhances neurite outgrowth by repressing glial fibrillary acidic protein expression and reorganizing laminin. *Endocrinology* **143**, 636–646 (2002).
8. Finch, C. E. Neurons, glia, and plasticity in normal brain aging. *Neurobiol. aging* **24**, S123–S127 (2003).
9. Liedtke, W. *et al.* Gfap is necessary for the integrity of cns white matter architecture and long-term maintenance of myelination. *Neuron* **17**, 607–615 (1996).
10. Giménez y Ribotta, M., Langa, F., Menet, V. & Privat, A. Comparative anatomy of the cerebellar cortex in mice lacking vimentin, gfap, and both vimentin and gfap. *Glia* **31**, 69–83 (2000).
11. Chernyatina, A. A., Guzenko, D. & Strelkov, S. V. Intermediate filament structure: the bottom-up approach. *Curr. opinion cell biology* **32**, 65–72 (2015).
12. Kim, B., Kim, S. & Jin, M. S. Crystal structure of the human glial fibrillary acidic protein 1b domain. *Biochem. biophysical research communications* **503**, 2899–2905 (2018).
13. Pekny, M. & Pekna, M. Astrocyte intermediate filaments in cns pathologies and regeneration. *The J. Pathol. A J. Pathol. Soc. Gt. Br. Irel.* **204**, 428–437 (2004).
14. Middeldorp, J. & Hol, E. Gfap in health and disease. *Prog. neurobiology* **93**, 421–443 (2011).
15. Eddleston, M. & Mucke, L. Molecular profile of reactive astrocytes—implications for their role in neurologic disease. *Neuroscience* **54**, 15–36 (1993).
16. Papa, L. *et al.* Elevated levels of serum glial fibrillary acidic protein breakdown products in mild and moderate traumatic brain injury are associated with intracranial lesions and neurosurgical intervention. *Annals emergency medicine* **59**, 471–483 (2012).
17. Bogoslovsky, T. *et al.* Increases of plasma levels of glial fibrillary acidic protein, tau, and amyloid β up to 90 days after traumatic brain injury. *J. neurotrauma* **34**, 66–73 (2017).

18. Nawashiro, H., Messing, A., Azzam, N. & Brenner, M. Mice lacking gfap are hypersensitive to traumatic cerebrospinal injury. *Neuroreport* **9**, 1691–1696 (1998).

19. Nawashiro, H., Brenner, M., Fukui, S., Shima, K. & Hallenbeck, J. M. High susceptibility to cerebral ischemia in gfap-null mice. *J. Cereb. Blood Flow & Metab.* **20**, 1040–1044 (2000).

20. Tomokane, N., Iwaki, T., Tateishi, J., Iwaki, A. & Goldman, J. Rosenthal fibers share epitopes with alpha b-crystallin, glial fibrillary acidic protein, and ubiquitin, but not with vimentin. immunoelectron microscopy with colloidal gold. *The Am. journal pathology* **138**, 875 (1991).

21. Tang, G., Xu, Z. & Goldman, J. E. Synergistic effects of the sapk/jnk and the proteasome pathway on glial fibrillary acidic protein (gfap) accumulation in alexander disease. *J. Biol. Chem.* **281**, 38634–38643 (2006).

22. Quinlan, R. A., Brenner, M., Goldman, J. E. & Messing, A. Gfap and its role in alexander disease. *Exp. cell research* **313**, 2077–2087 (2007).

23. Jany, P. L. *et al.* Csf and blood levels of gfap in alexander disease. *Eneuro* **2** (2015).

24. Wharton, S. *et al.* Mrc cognitive function and ageing neuropathology study group. population variation in glial fibrillary acidic protein levels in brain ageing: relationship to alzheimer-type pathology and dementia. *Dement Geriatr Cogn Disord* **27**, 465–73 (2009).

25. Oeckl, P. *et al.* Glial fibrillary acidic protein in serum is increased in alzheimer's disease and correlates with cognitive impairment. *J. Alzheimer's Dis.* **67**, 481–488 (2019).

26. Le Prince, G. *et al.* Alterations of glial fibrillary acidic protein mrna level in the aging brain and in senile dementia of the alzheimer type. *Neurosci. letters* **151**, 71–73 (1993).

27. Asken, B. M. *et al.* Plasma glial fibrillary acidic protein levels differ along the spectra of amyloid burden and clinical disease stage. *J. Alzheimer's Dis.* **78**, 265–276 (2020).

28. Benedet, A. L. *et al.* Differences between plasma and cerebrospinal fluid glial fibrillary acidic protein levels across the alzheimer disease continuum. *JAMA neurology* **78**, 1471–1483 (2021).

29. Benussi, A. *et al.* Serum glial fibrillary acidic protein (gfap) is a marker of disease severity in frontotemporal lobar degeneration. *J. Alzheimer's Dis.* **77**, 1129–1141 (2020).

30. Chatterjee, P. *et al.* Plasma glial fibrillary acidic protein is elevated in cognitively normal older adults at risk of alzheimer's disease. *Transl. psychiatry* **11**, 1–10 (2021).

31. Palmqvist, S. *et al.* Discriminative accuracy of plasma phospho-tau217 for alzheimer disease vs other neurodegenerative disorders. *Jama* **324**, 772–781 (2020).

32. Simrén, J. *et al.* Differences between blood and cerebrospinal fluid glial fibrillary acidic protein levels: The effect of sample stability. *Alzheimer's & Dementia* **18**, 1988–1992 (2022).

33. Stark, G. R., Stein, W. H. & Moore, S. Reactions of the cyanate present in aqueous urea with amino acids and proteins. *J. Biol. Chem.* **235**, 3177–3181 (1960).

34. Marier, J. & Rose, D. Determination of cyanate, and a study of its accumulation in aqueous solutions of urea. *Anal. biochemistry* **7**, 304–314 (1964).

35. Hagel, P., Gerding, J., Fieggen, W. & Bloemendal, H. Cyanate formation in solutions of urea: I. calculation of cyanate concentrations at different temperature and ph. *Biochimica et Biophys. Acta (BBA)-Protein Struct.* **243**, 366–373 (1971).

36. Ahmed, F., Kleffmann, T. & Husain, M. Acetylation, methylation and allysine modification profile of viral and host proteins during influenza a virus infection. *Viruses* **13**, 1415 (2021).

37. Marty, M. T. *et al.* Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. chemistry* **87**, 4370–4376 (2015).

38. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).

39. Nyíri, K. *et al.* Structural model of human dutpase in complex with a novel proteinaceous inhibitor. *Sci. Reports* **8**, 4326 (2018).

40. Hodge, E. A., Benhaim, M. A. & Lee, K. K. Bridging protein structure, dynamics, and function using hydrogen/deuterium-exchange mass spectrometry. *Protein Sci.* **29**, 843–855 (2020).

41. Evans, R. *et al.* Protein complex prediction with alphafold-multimer. *BioRxiv* 2021–10 (2021).

42. Høie, M. H. *et al.* Netsurfp-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic acids research* **50**, W510–W515 (2022).

43. Yariv, B. *et al.* Using evolutionary data to make sense of macromolecules with a "face-lifted" consurf. *Protein Sci.* **32**, e4582 (2023).

44. Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).

45. Mirdita, M. *et al.* Colabfold: making protein folding accessible to all. *Nat. methods* **19**, 679–682 (2022).

46. Ashkenazy, H. *et al.* Consurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* **44**, W344–W350 (2016).

47. Zimmermann, L. *et al.* A completely reimplemented mpi bioinformatics toolkit with a new hhpred server at its core. *J. molecular biology* **430**, 2237–2243 (2018).

48. Gabler, F. *et al.* Protein sequence analysis using the mpi bioinformatics toolkit. *Curr. Protoc. Bioinforma.* **72**, e108 (2020).

49. Ludwiczak, J., Winski, A., Szczepaniak, K., Alva, V. & Dunin-Horkawicz, S. Deepcoil—a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics* **35**, 2790–2795 (2019).

50. Weis, D. D., Wales, T. E., Engen, J. R., Hotchko, M. & Ten Eyck, L. F. Identification and characterization of ex1 kinetics in h/d exchange mass spectrometry by peak width analysis. *J. Am. Soc. for Mass Spectrom.* **17**, 1498–1509 (2006).

51. Guttman, M., Weis, D. D., Engen, J. R. & Lee, K. K. Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra. *J. Am. Soc. for Mass Spectrom.* **24**, 1906–1912 (2013).

52. Eng, L. F., Lee, Y. L., Kwan, H., Brenner, M. & Messing, A. Astrocytes cultured from transgenic mice carrying the added human glial fibrillary acidic protein gene contain rosenthal fibers. *J. neuroscience research* **53**, 353–360 (1998).

53. Graceffa, P. Hsp27-actin interaction. *Biochem. research international* **2011** (2011).

54. Haslbeck, M., Weinkauf, S. & Buchner, J. Small heat shock proteins: simplicity meets complexity. *J. Biol. Chem.* **294**, 2121–2132 (2019).

55. Tian, R., Gregor, M., Wiche, G. & Goldman, J. E. Plectin regulates the organization of glial fibrillary acidic protein in alexander disease. *The Am. journal pathology* **168**, 888–897 (2006).

56. Li, H. *et al.* 14-3-3γ affects dynamics and integrity of glial filaments by binding to phosphorylated gfap. *J. cell science* **119**, 4452–4461 (2006).

57. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (hdx-ms) experiments. *Nat. methods* **16**, 595–602 (2019).

58. Lau, A. M., Claesen, J., Hansen, K. & Politis, A. Deuteros 2.0: peptide-level significance testing of data from hydrogen deuterium exchange mass spectrometry. *Bioinformatics* **37**, 270–272 (2021).