BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Early Events in the Amyloid Formation of the A546T Mutant of Transforming Growth Factor β-Induced Protein in Corneal Dystrophies Compared to the Nonfibrillating R555W and R555Q Mutants

Heidi Koldsø,[†,‡,∥] Ole Juul Andersen,[†,‡] Camilla Lund Nikolajsen,[†,§,⊥] Carsten Scavenius,[†,§] Charlotte S. Sørensen,[†,§] Jarl Underhaug,[†,‡,@] Kasper Runager,[†,§] Niels Chr. Nielsen,[†,‡] Jan J. Enghild,[†,§] and Birgit Schiøtt*[,†,‡]

[†]Center for Insoluble Protein Structures (inSPIN) and Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Aarhus, Denmark

[‡]Department of Chemistry, Aarhus University, Aarhus, Denmark

[§]Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

Supporting Information

ABSTRACT: The human transforming growth factor β-induced protein (TGFBIp) is involved in several types of corneal dystrophies where protein aggregation and amyloid fibril formation severely impair vision. Most disease-causing mutations are located in the last of four homologous fasciclin-1 (FAS1) domains of the protein, and it has been shown that when isolated, the fourth FAS1 domain (FAS1-4) mimics the behavior of full-length TGFBIp. In this study, we use molecular dynamics simulations and principal component analysis to study the wild-type FAS1-4 domain along with three disease-causing mutations (R555W, R555Q, and A546T) to decipher any internal difference in dynamical properties of the domains that may explain their varied stabilities and aggregation properties. In addition, we use a protein−protein docking method in combination with chemical cross-linking experiments and mass spectrometry of the cross-linked species to obtain information about interaction faces between identical FAS1-4 domains. The results show that the pathogenic mutations A546T and R555W affect the packing in the hydrophobic core of FAS1-4 in different directions. We further show that the FAS1-4 monomers associate using their β-rich regions, consistent with peptides observed to be part of the amyloid fibril core in lattice corneal dystrophy patients.

Transforming growth factor β-induced protein (TGFBIp) is an extracellular matrix protein expressed in various tissues, including the cornea.[1,2] It consists of 683 amino acid residues with an N-terminal cysteine-rich EMILIN-1 (EMI) domain and four homologous fasciclin-1 (FAS1) domains.[3] Mutations within the *TGFBI* gene have been linked to corneal deposits leading to visual dysfunction. Such deposits within the cornea are normally divided into two groups: lattice corneal dystrophies (LCD) and granular corneal dystrophies (GCD) giving rise to ordered fibrillar deposits and amorphous deposits, respectively.[3] Within the *TGFBI* gene, more than 50 mutations resulting in either LCD or GCD have been identified;[4] however, no mechanism that explains mutant-induced LCD or GCD has been deciphered. In this study, we focus on the A546T, R555W, and R555Q mutations causing LCD type IIIA, GCD type 1, and Thiel-Behnke corneal dystrophy, respectively.[3]

Studying the behavior of wild-type (wt) and mutant full-length TGFBIp by molecular dynamics (MD) simulations could provide information about the structural and dynamical implications of the specific mutations; however, this is at present not possible as its full-length three-dimensional structure has not yet been determined. Fortunately, a nuclear magnetic resonance (NMR) structure of the fourth FAS1 domain (FAS1-4) has recently been established,[5] and biochemical studies have shown several similarities in the behavior of the isolated FAS1-4 domain and full-length TGFBIp. First, the isolated domain has been shown to display the same stability profile as full-length TGFBIp using urea denaturation (R555W > wt ≥ R555Q > A546T).[6] Second, of the three mutants, only the isolated domain carrying the A546T mutation has been shown to form amyloid fibrils, corresponding nicely with the *in vivo* aggregates.[6,7] Finally, TGFBIp, which is involved in cell adhesion and migration, binds to various integrins.[8] The isolated FAS1-4 domain has been shown to bind to some of the same integrins, block the binding of other proteins to integrins, and mediate cell adhesion and migration, although with varying activity compared to full-length TGFBIp.[8−13] The functional role of the FAS1-4 domain thus appears to be retained when it is studied in isolation, and the

Received: April 29, 2015
Revised: August 21, 2015
Published: August 25, 2015

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556

isolated domain should thus serve as a good model system for studying the effects of the three mutations on the aggregation properties of the full-length protein.

The 135-residue FAS1-4 domain contains an α-helical N-terminal part and a β-rich C-terminal part (Figure 1A).[5] A large hydrophobic core consisting of 39 residues is observed within the protein structure (Figure 1A). This hydrophobic core is believed to stabilize the tertiary structure of the protein through hydrophobic interactions.[14]

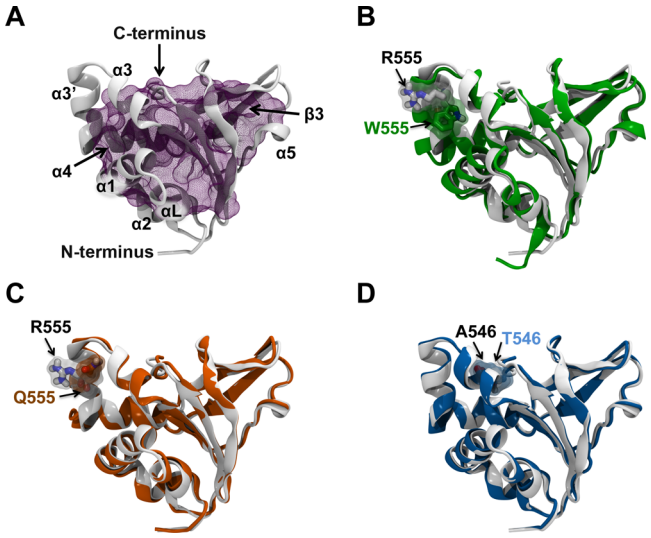

Figure 1. Structures of wt and mutant FAS1-4. The wt structure (white) and the R555W structure (green) are from their respective NMR ensembles, whereas the R555Q (orange) and A546T (blue) structures are from the *in silico* workflow described in Experimental Procedures. (A) Structure of wt FAS1-4 shown in the standard orientation. The protein domain is composed of an α-helical N-terminal part and a β-rich C-terminal part. The termini, α-helices, and β3 are labeled. A large hydrophobic core lines the entire center of the protein and stabilizes the tertiary structure. The hydrophobic core is shown as a purple surface mesh. In panels B–D, the locations of the three pathogenic mutations of FAS1-4 studied herein are indicated by sticks and transparent surfaces. (B) The R555W mutation is located at helix α3′. W555 of the mutant is packed closer in the hydrophobic cavity of α1, α3, and α3′ than the arginine of the wt structure. (C) The other mutation at position 555 is the R555Q mutation. (D) The A546T mutation is located in helix α3.

The mutations analyzed in this study are all located in the vicinity of a small hydrophobic cavity lined by helices α1, α3, and α3′ (Figure 1), which is situated distant from the fibril core (Y571–R588) observed in LCD type 1 fibrils.[15−17] The fibril core region extends from the final part of helix α4 to the first part of β3. The R555W and R555Q mutations are located on the surface of the protein (Figure 1B,C), while the A546T mutation is more buried inside the protein structure (Figure 1D).

MD simulations are widely used for studying the different steps in fibril formation.[18−21] In this study, we use it to explore the differences in dynamical properties between the wt and the A546T, R555W, and R555Q mutants to gain insight into the molecular events leading to either amyloid formation or aggregation of TGFBIp in corneal dystrophies. Furthermore, the homoaggregation of the four variants is investigated with the protein–protein docking tool DOCK/PIERR[22−24] to reveal any differences in protein association patterns induced by the mutations. Finally, the validity of the computational predictions is assessed by chemical cross-linking experiments of the monomers of wt FAS1-4 and each of the three mutants followed by mass spectrometry to identify the cross-linked residues.

EXPERIMENTAL PROCEDURES

Modeling. Four variants of the FAS1-4 domain were studied, namely, wt and three single-point mutants (R555W, R555Q, and A546T). The input for the wt MD simulations was the lowest-energy structure from the NMR ensemble [Protein Data Bank (PDB) entry 2LTB].[5] The R555W MD simulations were also initiated from the lowest-energy structure from the corresponding NMR ensemble (PDB entry 2LTC).[5] The R555W structure contains an additional mutation, as the C-terminal proline residue present in the wt was replaced with an alanine when the sample for NMR was prepared.[5] The C-terminus of TGFBIp has previously been shown to be unstructured,[5] and the substitution of proline with an alanine at the C-terminus of FAS1-4 is thus not expected to largely affect either the structure or the dynamics of the domain. The R555Q and A546T structures were generated by mutating the corresponding side chains of the wt FAS1-4 NMR structure *in silico* utilizing PyMOL and thus do not contain the proline to alanine substitution.[25] The R555W structure was similarly generated on the basis of *in silico* mutation for comparison with results obtained from calculations initiated from the NMR structure, hereby assessing the quality of the mutants generated *in silico*. The rotamer state of the residue mutated *in silico* was chosen to give the most favorable interactions with the surrounding residues as evaluated by minimal steric clashes and optimal nonbonded interactions. Protonation states were assigned by PROPKA2.0.[26] The tautomeric states of H572 and H626 were chosen on the basis of surrounding residues giving rise to H572 being modeled as the Nδ tautomer while H626 was modeled as the Nε tautomer. All other acidic and basic residues were modeled as charged. The proteins were solvated utilizing the TIPS3P water model,[27] and the systems were ionized with NaCl to an ion concentration of 0.1 M to resemble the NMR conditions.[5]

MD Simulations. The systems were minimized by the conjugate gradient method for 15000 steps followed by simulation using the isothermal–isobaric ensemble. The simulations were run in NAMD2.6,[28] utilizing the CHARMM22 force field[29] with CMAP corrections.[30,31] The pressure was maintained at 1 atm by employing the Langevin piston method[32,33] with a piston period of 100 fs. A constant temperature of 310 K was maintained by Langevin dynamics with a damping coefficient of 0.5 ps^{-1}. The particle mesh Ewald method[34] was used to treat full electrostatics, while van der Waals interactions were truncated at a cutoff distance of 12 Å using a switching function from 10 Å.

Each system consisted of approximately 18100 atoms, and the dimensions of the simulation boxes were approximately 61 Å × 61 Å × 52 Å. The simulations were repeated three times, all for 100 ns using a 1 fs time step. A 2 fs time step was used for nonbonded interactions, and a 4 fs time step was used for full electrostatics. Snapshots from every 100 ps were used for analysis, except for cluster analysis where snapshots from every 1 ps were used.

Analysis. The root-mean-square deviation (rmsd) and solvent accessible surface area (SASA) analyses were calculated utilizing VMD1.8.7,[35] while the principle component analysis (PCA) was performed using the built-in *g_covar* and *g_anaeig* functions of GROMACS.[36,37] The degree of solvent exposure was defined as shown in eq 1.

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546–5556

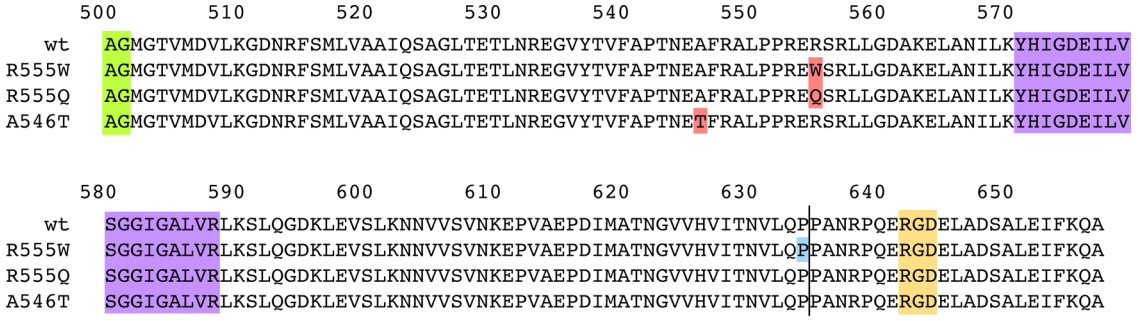

Figure 2. Sequences of the wt, R555W, R555Q, and A546T constructs used for the cross-linking experiments. An Ala-Gly dipeptide (green) was included at the N-terminus to facilitate cleavage by SUMO proteases. Mutations that were present in both the cross-linking and the modeling studies are colored red. The blue proline in the R555W sequence was substituted with an alanine only in the modeling studies. The structures used for the modeling studies ended at P634, which is represented by a vertical line between P634 and P635. The RGD sequence is colored orange, and the fibril core of LCD type 1 fibrils (Y571−R588) is colored purple.

$$\text{degree of solvent exposure (\%)} = \frac{[\text{SASA}_{\text{residue in protein}}]}{[\text{SASA}_{\text{free}}]} \times 100\% \quad (1)$$

The average linkage method[38] available in the *ptraj* analysis package in AmberTools 11[39] was applied for cluster analyses of the snapshots stored during the MD simulations. The clustering cutoff was set to a fixed value of 2.1 Å considering all C_α atoms.[5] The clustering resulted in eight wt clusters, ten R555W clusters, four R555Q clusters, and three A546T clusters, describing the structures sampled during the MD simulations. The representative structure from each cluster was extracted on the basis of a minimal sum of the squared displacements to the other structures within that cluster and used for further analysis.

Protein−Protein Docking. The representative structure obtained from each individual cluster was used as input for protein−protein docking calculations using the DOCK/PIERR method.[22−24] The calculations were performed on the online DOCK/PIERR server. To obtain the docking results, the representative structure from each cluster was docked against itself. Ten poses were stored for each docking calculation.

Chemical Cross-Linking. Four different FAS1-4 constructs (wt, A546T, R555Q, and R555W) consisting of residues 502−657 of human TGFBIp were expressed in *Escherichia coli* and purified as previously described.[6] The decision to use the isolated domain for the cross-linking and mass spectrometry experiments was based on several considerations. First, the modeling studies have been performed on the isolated domain, as necessitated by the lack of a structure of the full-length protein. Using different model systems for the modeling studies and the performed experiments would impose difficulties upon comparison of data from the different methods. Second, the lack of structural information about full-length TGFBIp would complicate the interpretation of the mass spectrometry data. In contrast to the structures used for the modeling studies, the four FAS1-4 constructs also included the C-terminus of TGFBIp. For all constructs, an Ala-Gly dipeptide was included at the N-terminus to facilitate cleavage by SUMO proteases. The sequences of the constructs are provided in Figure 2. The purified recombinant protein was cross-linked using the chemical cross-linker bis(sulfosuccinimidyl)suberate (BS^3) (Pierce). FAS1-4 was titrated with BS^3 to find a conservative yet efficient BS^3 concentration, which was subsequently used to generate cross-linked samples for mass spectrometry. FAS1-4 (~15 μM) was incubated with BS^3 at a molar ratio of 1:200 corresponding to approximately 3 mM BS^3 in 150 mM NaCl and 20 mM NaH_2PO_4 (pH 7.4) for 30 min at room temperature. The cross-linker was dissolved in 150 mM NaCl and 20 mM NaH_2PO_4 (pH 7.4) immediately before use. The reaction was quenched by addition of NH_4HCO_3 to a final concentration of 150 mM, and the sample was incubated for 20 min at room temperature. Cross-linked proteins were boiled for 10 min at 100 °C in sample buffer containing SDS and separated on 5−15% gradient gels (10 cm × 10 cm × 1.5 mm) using the glycine/2-amino-2-methyl-1,3-propanediol/HCl system.[40] Proteins were visualized using Coomassie brilliant blue. Protein bands of interest were excised and digested in-gel by incubation with sequencing grade modified trypsin (Promega) in 50 mM NH_4HCO_3 at 37 °C for 16 h. The resulting peptides were desalted using C18 StageTips (ThermoScientific).

Mass Spectrometry. NanoLC−MS/MS was performed using an EASY-nLC II system (ThermoScientific) connected to a TripleTOF 5600 mass spectrometer (AB Sciex). Samples were analyzed in three or more replicate runs. Peptides were dissolved in 0.1% formic acid, injected, trapped, and desalted on a Biosphere C18 column [5 μm, 2 cm × 100 μm inside diameter (Nano Separations)]. The peptides were eluted from the trap column and separated on a 15 cm analytical column (75 μm inside diameter) generated in a P-2000 laser micropipette puller (Sutter Instruments). The column was packed with RP ReproSil-Pur C18-AQ 3 μm resin (Dr. Marisch GmbH, Ammerbuch-Entringen, Germany) and connected in-line to the mass spectrometer. Peptides were eluted at a flow rate of 250 nL/min, using a 50 min gradient from 5 to 35% phase B (0.1% formic acid and 90% acetonitrile). The collected MS data were converted to Mascot generic format (MGF) using the AB SCIEX MS Data Converter beta 1.1 (AB SCIEX) and the "ProteinPilot MGF" parameters. Data analysis was performed with the free software MassAI (MassAI Bioinformatics) that was recently integrated with CrossWork for analysis of cross-linking mass spectrometry samples.[41] The MGF files were filtered with the MGF-filter tool using default values, except that all collected peaks were retained. The files were searched against the FAS1-4 sequence using the following parameters: cross-linked peptides were searched with an MS error tolerance of 0.04 Da and an MS/MS error tolerance of 0.2 Da; three missed cleavages were allowed; oxidized methionine (M-ox) was allowed as a variable modification with up to two modifications per peptide; BS^3 cross-linking was allowed between amines, Lys, Ser, Thr, Tyr, and as dead-end linker. All identifications were manually inspected, and a list of assigned spectra was exported (Supporting Information). Cross-linked peptides were accepted only if high-intensity peaks

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556

were accounted for and if at least three fragment ions from each peptide were observed.

RESULTS

Testing the *in Silico* Models. For the MD simulations performed in this study, the structures of the wt protein and the R555W mutant were taken from the published NMR ensembles,[5] whereas the R555Q and A546T mutants were constructed *in silico* from the wt NMR structure as explained in Experimental Procedures. To examine if the *in silico*-generated starting structures are reliable, a similar approach for the R555W mutant was conducted. The results were compared to the experimental structure, which contains a translation of $\alpha 3'$ corresponding to half a helix turn compared to the wt structure.[5] This translation is also observed in the R555W *in silico* simulations (Figure 3), thereby showing that the NMR ensemble

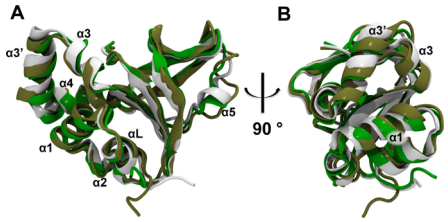

Figure 3. *In silico*-generated R555W structure samples NMR ensemble conformations. Two orientations of an overlay of the wt NMR structure (light gray), the R555W NMR structure (green), and a snapshot from one of the R555W *in silico* trajectories (tan) are shown. The structures have been aligned on all but the four N-terminal residues. The translation of helix $\alpha 3'$ that is observed in the R555W NMR structure is being sampled in the R555W *in silico* simulations. (A) Front view. (B) View from the side of $\alpha 1$, $\alpha 3$, and $\alpha 3'$.

is indeed being sampled during MD when starting from a protein structure generated *in silico*, indicating that the R555Q and A546T *in silico* structures are reliable models. The R555W *in silico* trajectories will not be used for further analysis in the remaining part of the study.

The overall stabilities of the four protein systems were explored by calculating the rmsd for the C_α atoms as shown in Figure 4, where the average from three independent repeats of each system is shown. The data show that there is no significant difference in stability as all systems plateau at an average rmsd of ~2 Å. Even though the R555Q and A546T systems are based on

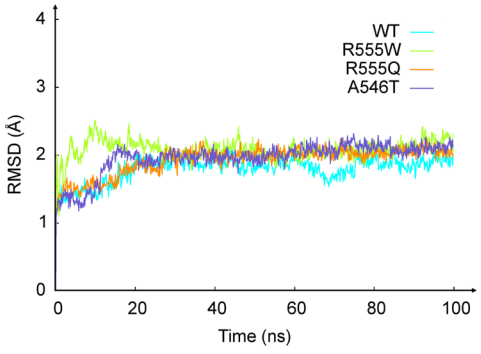

Figure 4. Average rmsd values calculated from three repeats for each of the wt (cyan), R555W (green), R555Q (orange), and A546T (purple) systems. All four systems reach equilibrium structures after ~30 ns, where they plateau at an rmsd of ~2 Å.

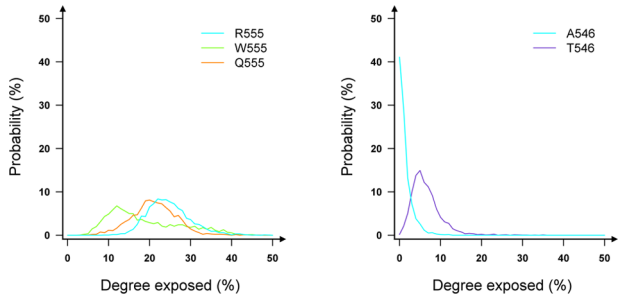

Figure 5. Degree of solvent exposure of residues 555 and 546 in the wt (cyan), R555W (green), R555Q (orange), and A546T (purple) simulations reveals that mutation to a hydrophobic residue results in closer packing toward the hydrophobic core, whereas a mutation to a larger hydrophilic residue results in an increased level of solvent exposure. (Left) Degree of solvent exposure of R555 in the wt, W555 in the R555W mutant, and Q555 in the R555Q mutant. R555 and Q555 show a similar degree of solvent exposure, while the degree of solvent exposure of W555 decreases. (Right) Degree of solvent exposure of A546 in the wt and T546 in the A546T mutant. The level of solvent exposure increases slightly after the introduction of the larger and more hydrophilic T546 into the A546T mutant. A bin size of 1% exposure has been used to bin the data.

in silico mutations, the rmsds of all systems reach a plateau at a level similar to that of the simulations based on NMR structures, indicating that equilibrium structures have been obtained. Equilibrium was reached after ~30 ns for all simulations.

Mutations Alter the Packing of the Hydrophobic Core. Because it has been proposed that a hydrophobic part of the protein becomes accessible during misfolding,[42] we studied the degree of solvent exposure of residues 546 and 555 in the wt protein and the three mutants (Figure 5) as the protein varies at only these positions, with the exception of the previously mentioned C-terminal proline to alanine mutation in the R555W structure. A large decrease in the level of solvent exposure is observed for W555. This indicates that the mutation of the charged arginine in wt FAS1-4 to a more hydrophobic residue in the R555W system results in the packing of this residue toward the hydrophobic core of the protein as was also previously addressed.[5] The mutation to a glutamine at position 555 is more conservative and does not have as large an effect on the degree of solvent exposure. The data for residue 546 in the A546T mutant show a slight increase in the level of solvent exposure, suggesting a more open structure of the A546T mutant compared to that of wt FAS1-4. The substitution with a threonine introduces a larger side chain into $\alpha 3$. The side chain is adjacent to the C-terminal stretch of amino acids in the FAS1-4 domain. To accommodate the larger side chain, the helix is pushed to the side, providing a slightly more open structure (Figure 6).

Movements in the FAS1-4 Domain Favor the Exposure of the Hydrophobic Core. To identify the common dominant movements for all four systems, we performed a PCA on the trajectories. PCA is a useful method for determining the essential dynamics of proteins, and it has been shown that the slow conformational motions in a protein can be identified from relatively short simulations using this approach.[43−45] A better understanding of the dominating motions can be expected to result in improved insight into the underlying mechanism of aggregation of the proteins. A total of 1.2 μs of simulation data was included in the analysis. The first three principal components, accounting for 37% of the motions, are plotted against each other through two-dimensional (2D) projection

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556

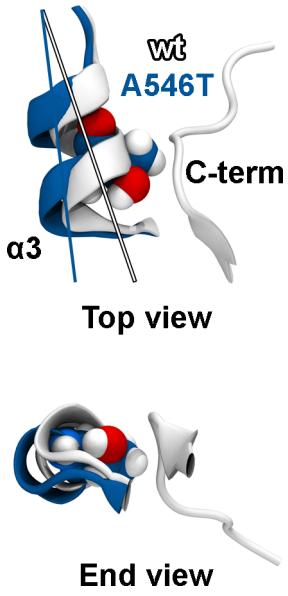

Figure 6. Substitution of an alanine with a threonine in α3 of A546T introduces a larger side chain. The side chain is located next to the C-terminal stretch of amino acids in the FAS1-4 domain. To accommodate the side chain, a slightly more open conformation of this region is reached, as illustrated by lines through the wt and A546T α3 helices.

plots (Figure 7). From these plots, it can be seen that there are no large outliers among the four systems, although some minor differences are apparent.

On the basis of the PCA analysis of the 12 simulations combined, it is possible to project the motions described by principal components onto the protein structure (Figure 8). It is not surprisingly the upper lateral extremities of the FAS1-4 domain (when viewed from the standard orientation in Figure 1) that are involved in the dominating movements. As seen in Figure 8A, the most dominating movement of all systems combined is observed to be the separation of the α-helical region and the β-sheet region, including the movement of $\alpha3'$ observed to be the main difference between the wt and the R555W NMR structures. PC1 of the combined system corresponds to 15% of the movements. The second most dominating common movement is the twisting of α5 with respect to the rest of the protein (Figure 8B). The third largest common movement is the

Figure 8. Dominating movements illustrated by arrows along the three first principal components as observed from the PCA of all trajectories for the four systems. (A) PC1, separation of the α-helical region and the β-sheet region. (B) PC2, twisting of α5 with respect to the rest of the protein. (C) PC3, correlated twisting of $\alpha3'$, α5, and the loops within the β-region.

correlated twisting of $\alpha3'$, α5, and the loops within the β-region (Figure 8C).

FAS1-4 Monomers Interact through Their β-Rich Regions. To explore if the minor structural differences among the four individual FAS1-4 domains cause different aggregation patterns, we set out to perform protein−protein docking calculations, which might reveal possible protein−protein interfaces involved in the aggregation process. Cluster analyses of the MD simulations resulted in eight wt clusters, ten R555W clusters, four R555Q clusters, and three A546T clusters, giving a total of 25 clusters (see Experimental Procedures). The representative structure from each cluster was used for the calculations as described in Experimental Procedures. From the docking results, it is evident that two particular interaction faces are often present among the best ranking poses. In one of the poses, called the Top-to-Top pose (Figure 9A), the two proteins associate via the β-rich region (shown to the right in Figure 1A). In the other pose, the two monomers associate using the same surfaces as in the Top-to-Top pose; however, one of the monomers (shown to the right in Figure 9B) has been rotated by approximately 180° around the horizontal axis such that the bottom of the domain is now pointing upward. This pose is called the Top-to-Bottom pose. Common for both poses is the interesting observation that they display direct interaction between the two Y571−R588 segments that form part of the fibril core in LCD type 1 fibrils.[15−17]

The results of the docking calculations are summarized in Figure 10, showing that in 18 of 25 clusters derived from the MD simulations, the best scoring pose is either a Top-to-Top or a Top-to-Bottom pose. Furthermore, Table 1 reveals that the two

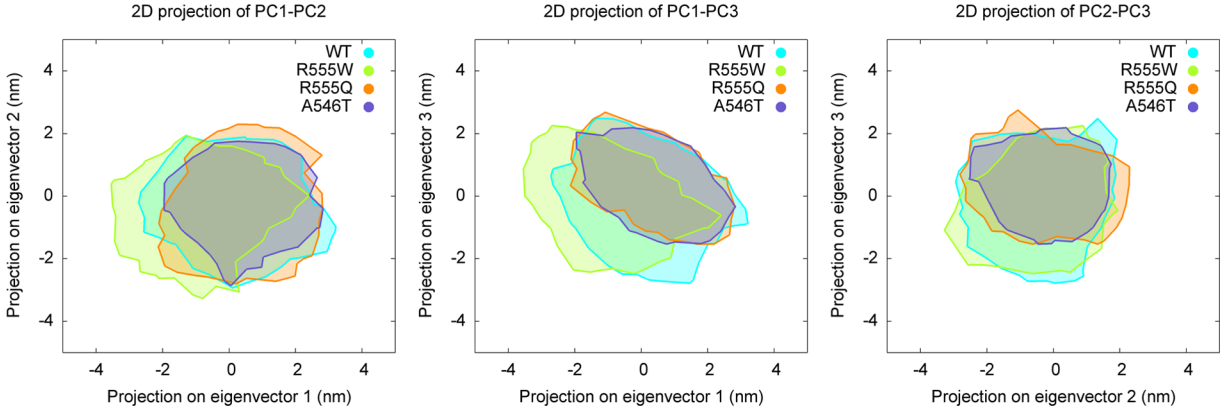

Figure 7. 2D projections of the first three PCA eigenvectors from the merged trajectories back onto the wt (cyan), R555W (green), R555Q (orange), and A546T (purple) trajectories. (Left) Eigenvector 1 on the horizontal axis and eigenvector 2 on the vertical axis. (Middle) Eigenvector 1 on the horizontal axis and eigenvector 3 on the vertical axis. (Right) Eigenvector 2 on the horizontal axis and eigenvector 3 on the vertical axis.

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556

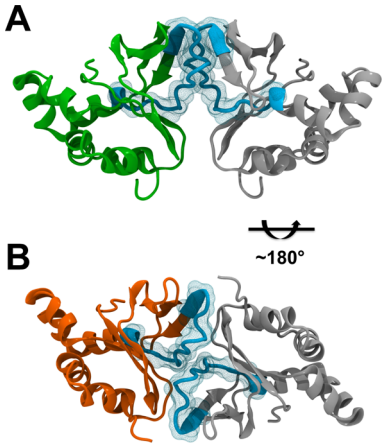

Figure 9. Figures of representative Top-to-Top and Top-to-Bottom poses. The peptide sequence Y571−R588 is colored cyan. Both poses show direct contact between these segments, which have been shown to be part of the fibril core of LCD type 1 fibrils. (A) Top-to-Top pose. (B) Top-to-Bottom pose. The monomer to the right is rotated ∼180° around the horizontal axis with respect to the Top-to-Top pose.

poses have a higher occurrence among the high-ranking poses than among the low-ranking poses, as their presence drops from 72% among the top poses to 45% among all poses.

The docking study indicates a high level of favorable conformations in which the fibril core segments of LCD type 1 fibrils are in direct contact with each other. These might serve as starting points for the fibrillation process in the LCD-causing TGFBIp mutants. However, the study is based on the isolated FAS1-4 domain, and the identified poses might not be accessible for full-length TGFBIp because of steric interactions with other parts of the protein. A crystal structure including the FAS1-3 and FAS1-4 domains of *Drosophila* FAS1 was published in 2003 (PDB entry 1O70).[46] In Figure 11, the FAS1-4 domain from this crystal structure has been aligned with the FAS1-4 domains of both of the two monomers in the Top-to-Top and Top-to-Bottom poses. The alignment reveals that there is no steric clash between the FAS1-3 domains in either of the two poses and that the poses are therefore also likely to be possible for the full protein.

Chemical Cross-Linking of FAS1-4 Supports Docking Experiments. The docking calculations indicate that specific interactions between the FAS1-4 monomers exist. The chemical

Figure 10. Classification of poses from protein−protein docking calculations. The docking poses are ranked by their docking scores with the best scoring poses placed in the top of each column. The poses have been classified on the basis of visual inspection. Besides the Top-to-Top (green) and Top-to-Bottom (orange) poses, the results have been divided into Helix-to-Helix (purple) poses and a single unphysical pose (red) in which part of one of the proteins has crossed a peptide strand in the other protein, which in nature would require bond breakage and subsequent re-formation of the peptide strand. This pose is suspected to be an artifact of the first step in the DOCK/PIERR algorithm, which consists of docking one protein onto another using a coarse, residue-based potential, where such strand crossings would not receive a penalty large enough to remove the pose from the rest of the top-ranking poses. All other poses (blue) did not show a clear trend and were not further characterized. A pose has been classified as a Top-to-Top pose or a Top-to-Bottom pose on the basis of the symmetry observed in the poses in panels A and B of Figure 9, respectively. A pose has been classified as a Helix-to-Helix pose if the two proteins show binding between their respective α-helical regions.

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556

Table 1. Presence of Top-to-Top and Top-to-Bottom Poses in Docking Results[a]

no. of poses per cluster included in the calculation, starting with the highest ranking pose in each cluster	1	2	3	4	5	6	7	8	9	10
presence of Top-to-Top and Top-to-Bottom poses	72%	66%	60%	59%	57%	53%	51%	49%	46%	45%

[a]The first number (72%) corresponds to the presence among only the best scoring pose in each cluster. The last number (45%) corresponds to the presence among all 10 poses in each cluster. It is evident that the two poses are generally ranked higher than other poses, as their abundance drops when more low-ranking poses are included in the calculation.

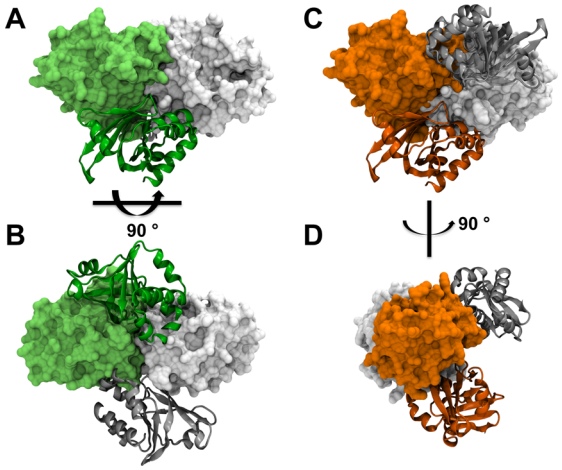

Figure 11. Alignment of the FAS1-4 domain (surface representation) from *Drosophila* FAS1 on the (A and B) Top-to-Top (gray and green) and (C and D) Top-to-Bottom (gray and orange) poses reveals that there is no steric hindrance between the FAS1-3 domains (ribbon representation), thus making it likely that the two poses are also accessible for full-length TGFBIp. The alignment was performed using residues 472−597 from *Drosophila* FAS1-4 and residues 504−629 from TGFBIp FAS1-4.

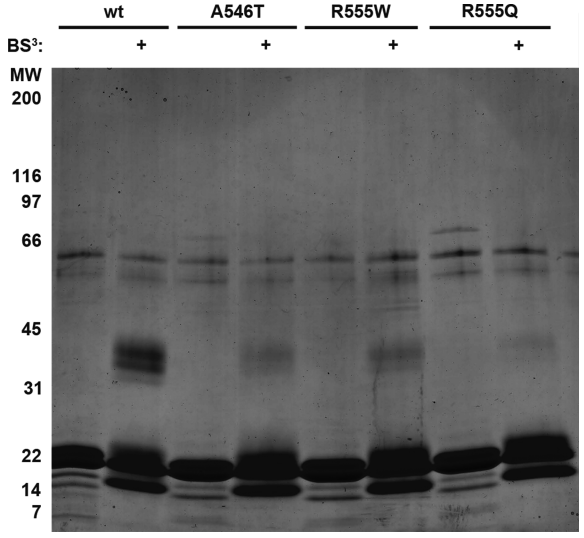

Figure 12. Incubation of the four FAS1-4 variants with BS^3 resulted in the formation of a covalently cross-linked product of the same expected molecular weight (MW) as a FAS1-4 dimer (~34 kDa). The cross-linking furthermore resulted in the formation of a band with a MW lower than that of the monomer. This is presumably due to intramolecular cross-linking in the monomeric domain, preventing full unfolding under the SDS conditions. Bands with MWs of ~60 kDa are present for all lanes on the gel, which is consistent with keratin contamination.

cross-linker BS^3 was therefore employed on all four FAS1-4 variants to examine if similar interactions also occur *in vitro*. BS^3 is a homobifunctional *N*-hydroxysulfosuccinimide (NHS) ester-based cross-linker that reacts predominantly with primary amines, e.g., lysine residues. The products from the cross-linking experiment of FAS1-4 were separated by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) (Figure 12). The analysis revealed that for all four variants, a fraction of the monomers was cross-linked into dimers. The cross-linking reaction also resulted in the formation of a band with a molecular weight lower than that of un-cross-linked FAS1-4. As this band was not present in the control lane, it was presumably the result of intramolecular cross-links within FAS1-4 monomers preventing complete unfolding of FAS1-4 during SDS−PAGE. Cross-linking of the entire FAS1-4 monomer population into the dimeric form was not possible even at very high cross-linker concentrations.

The cross-linked proteins were excised from the gel and digested with trypsin for subsequent LC−MS/MS analysis. All samples were analyzed in triplicate. The experiment resulted in 220 spectra assigned to a cross-link of FAS1-4 (Table 2 and Supporting Information). All 220 spectra represented a cross-link of the N-termini between two monomeric FAS1-4 domains (Figure 13). Cross-linking of the N-termini of two FAS1-4 monomers is consistent with the Top-to-Top pose; however, cross-links consistent with the Top-to-Bottom pose were not observed.

Table 2. Identified Cross-Linked Peptides from LC−MS/MS Experiments[a]

FAS1-4 variant	X-link	best score	spectral count
wt	A500 × A500	108.0	77
A546T	A500 × A500	98.4	26
R555Q	A500 × A500	119.6	41
R555W	A500 × A500	108.6	76

[a]Cross-linked proteins were isolated by SDS−PAGE and digested with trypsin. The generated peptides were analyzed by LC−MS/MS, and the data were searched against the FAS1-4 sequence using MassAI. A summary of identified cross-linked residues for each of the FAS1-4 variants is listed together with the scores for the top scoring spectra (MassAI), and the total number of spectra assigned to the cross-link from three replicates. An example of an annotated spectrum can be seen in Figure 13. Detailed MS information about all spectra is available in the Supporting Information.

DISCUSSION

Protein stability and protein−protein interactions are key factors in protein misfolding diseases such as the *TGFBI*-linked corneal dystrophies, and information about these is crucial for understanding the mechanisms of protein aggregation. In this study, we have used MD simulations, protein−protein docking, and chemical cross-linking to obtain new insight into the processes governing TGFBIp aggregation in corneal disease, including the initial stages of amyloid formation.

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556

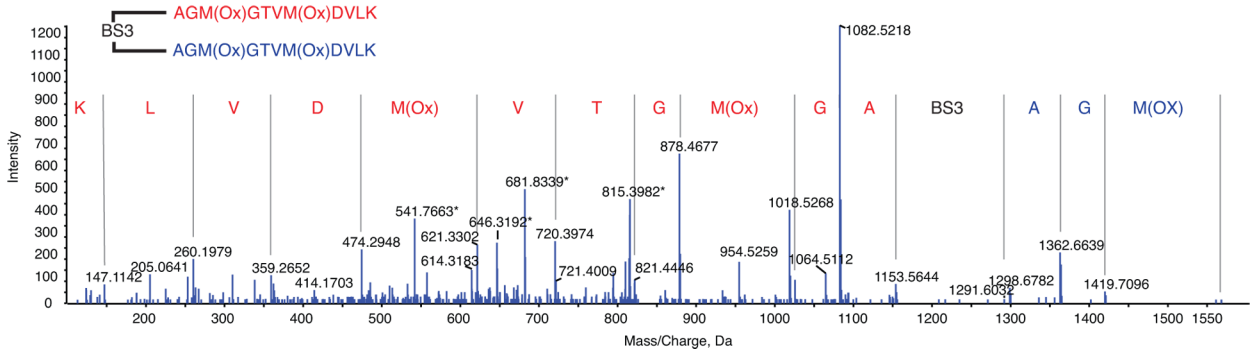

Figure 13. Assigned MS/MS spectrum for N-termini cross-linked by BS^3. Cross-linked proteins were isolated by SDS−PAGE and digested with trypsin. The generated peptides were analyzed by LC−MS/MS, and the data were searched against the FAS1-4 sequence using MassAI. The search resulted in 220 spectra all representing an N-terminal cross-link. A representative MS/MS spectrum of the N-terminal cross-link is shown with assignment of the y-ion series. Asterisks denote doubly charged y-ions and the precursor ion (~815).

The multiple MD simulations of wt and three pathogenic mutations of the FAS1-4 domain of TGFBIp reveal delicate control of hydrophobic core stability. The substitution of the charged R555 with the hydrophobic aromatic W increases the packing of this residue toward the hydrophobic core, whereas the R555Q mutant is very similar to the wt with regard to side chain packing. In contrast, the mutation of the hydrophobic A546 to the hydrophilic T increases the amount of solvent exposure at this site. This corroborates nicely the findings from *in vitro* experiments stating that R555W is more stable than wt whereas A546T is less stable.[6] The stability of the protein structure for the four variants can accordingly be hypothesized to be dictated by the hydrophobic packing within the protein core.

On the basis of PCA, it is possible to extract the dominating movements within the systems. The first three PCs of all four systems combined illustrate that, in general, the largest movements involve localized motions within the outer part of the α-helical region, the β-sheet region, or both combined. It is therefore easy to imagine that this movement of the two parts of the domain forms a basis for the structural unfolding. The α and β regions are connected by a large hydrophobic core, which will be exposed upon the movement described by PC1. The pathogenic mutants, all forming part of either the moving part of the α region or the interface between the α and β regions, could have an influence on the PC1 movements; mutation to a more hydrophobic residue may stabilize the hydrophobic packing preventing the structural breathing from unfolding the protein, whereas mutation to a more hydrophilic residue is expected to destabilize the hydrophobic core and decrease the activation barrier of protein unfolding, allowing for solvent exposure of this region upon fluctuations as described by PC1-3.

To further explore the possible effect of the observed dominating movements and the effects of the hydrophobic stabilization, we performed protein−protein docking calculations. Docking of the isolated FAS1-4 domain revealed a large abundance of conformations in which the fibril core segments, as observed in LCD type 1 fibrils (Y571−R588),[15−17] are in direct contact with each other. Comparison with the crystal structure of *Drosophila* FAS1 indicates that both of the interaction patterns will be possible also for the full-length TGFBIp structure, assuming that the FAS1-3 and FAS1-4 domains have the same respective orientation in the full-length protein as in *Drosophila* FAS1.

The protein−protein interaction sites observed from the protein−protein docking calculations were further validated by cross-linking experiments. In these experiments, it was found that BS^3 was able to cross-link both wt FAS1-4 and the three mutants in solution, and that cross-linking occurs between the N-termini. Full-length TGFBIp, although considered to be primarily monomeric in solution, may form both dimers and trimers.[47] The cross-linking experiments showed that FAS1-4 has the ability to form dimers in solution. Generally, cross-linking is considered to catch even very low-affinity protein interactions through the formation of covalent bonds between side chains in spatial proximity. Cross-linking mass spectrometry therefore provides information about both interaction partners and distance constraints for the interaction.[48] The maximal length of the BS^3 spacer is 11.4 Å, allowing C_α atoms of cross-linked lysine residues to be 26−30 Å apart in a proposed structure.[49] The docking experiments predict that, on average, five to seven lysine pairs having their C_α atoms within 26 Å are present in the two poses. In the Top-to-Top pose, the flexible N-termini are expected to be within the experimentally predicted proximity. As opposed to this, the N-termini are located on opposite sides of the dimer in the Top-to-Bottom pose, and cross-linking them is thus not possible. The experiments are therefore consistent only with the Top-to-Top pose. From fibrillation assays, it is known that no significant dimer population exists at protein concentrations similar to those used in the cross-linking experiments.[7] The observed cross-linked dimers are thus more likely to be transient encounters that can evolve into oligomers than they are to be sequestrations protecting the monomers from partial unfolding and fibrillation.

Some of the most well-known amyloid diseases, such as Parkinson's disease, Alzheimer's disease, and type II diabetes, originate from intrinsically disordered proteins (IDPs), namely, α-synuclein, amyloid-β, and amylin, respectively.[50,51] Despite their classification, IDPs are known to sample partially structured states.[50,52] It is believed that these states could be important in the formation of oligomers en route to the final fibril.[50] However, many amyloid-related proteins, such as TGFBIp, have well-characterized tertiary structures. Remarkably, some of them, such as insulin[53] and transthyretin,[54] contain mainly α-helical structure in their native state. Many hypotheses regarding mutation-induced oligomer formation of these proteins exist; misfolding, altered proteolytic processing, increased structural flexibility, altered electrostatics, abnormal protein turnover, and changes in extracellular matrix interactions have all been suggested as possible culprits.[55] Another interesting possibility is the existence of so-called native-like states, as documented in several cases.[53,56−59] Examples include insulin,[53] S6 from *Thermus thermophilus*,[56] and acylphosphatase from *Sulfolobus*

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556

solfataricus (Sso AcP).[57] The fibrillation of these proteins is initiated from oligomers consisting of conformational states that are thermodynamically distinct from the native state but structurally similar to it.[58] Notably, extensive enzymatic activity was maintained in the Sso Acp oligomers.[60] The energy of the native-like state is higher than that of the native state, but a low-energy barrier allows transitions between them driven by thermal fluctuations. Mutations that destabilize the native state generally increase the population of non-native states, including unfolded, intermediate, and native-like states.[58] As the Top-to-Top pose is observed for both the wt, A546T, and the nonfibrillating mutants, it does not provide an explanation for the differences in aggregate morphology. However, as the A546T mutant has previously been shown to be more flexible and less stable than the wt,[7] we propose that the identified pose plays a role in the fibrillation process. When two monomers meet in a Top-to-Top orientation, direct contact is made between the fibril cores of the two monomers. From here, the flexible and destabilized monomers could undergo a transition into a native-like state using movements as described by PC1-3, thereby forming destabilized oligomers that are en route to the formation of small protofibrils. The time scale related to these events is beyond that available using molecular dynamics simulations, and a detailed description of this pathway is thus not readily available. Another finding supporting the proposed sequence of events is the observation that wt FAS1-4 is able to form fibrils, although to a much lesser extent.[6] This fibrillation process might likewise be initiated from the pose, and the decreased extent of fibrillation could be linked to a smaller population of the native-like state due to lower flexibility and higher stability. In a recent study, it was shown that the presence of fibrils formed by the Y571−R588 core peptide was able to seed the fibrillation of both isolated A546T FAS1-4 and native TGFBIp.[61] This can within the proposed mechanism accordingly be ascribed to immediate binding of the formed protofibrils to the seeds and elongation of these to mature fibrils, thereby eliminating the stage of the fibrillation process related to the initial formation of the fibril scaffold.

It should be noted that not all mutations linked to fibrillation are destabilizing. The G623D mutation linked to LCD is shown to be more stable than the wt,[55] and this mutant likely follows one of the other previously mentioned pathways suggested for mutation-induced oligomer formation. The findings in this paper can thus not be imposed on FAS1-4 variants that have not formed part of the study.

In conclusion, our simulations show how a single point mutation might affect the protein stability through stabilization or destabilization of the hydrophobic core. The dominating movements from the MD simulations further support how the mutations could have an effect on the overall dynamics of FAS1-4. Additionally, we observe a possible protein−protein interface with contact between the β-rich regions of the two FAS1-4 domains that might play an important role during the fibrillation process of A546T. An elaborate explanation for the aggregation of the R555Q and R555W mutants is not available from the obtained data. The results indicate that these mutants are also able to form the Top-to-Top dimer pose, which is not surprising considering that the location of R555 is opposite the dimer face. The resistance to amyloid formation is thought to arise from the absence of a significant population of a destabilized native-like state.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b00473.

A list of all annotated MS/MS spectra (XLS)

■ AUTHOR INFORMATION

Corresponding Author

*Department of Chemistry, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark. Telephone: +45 8715 5975. Fax: +45 8619 6199. E-mail: birgit@chem.au.dk.

Present Addresses

∥H.K.: D. E. Shaw Research, 120 W. 45th St., 39th Floor, New York, NY 10036.

⊥C.L.N.: Chr. Hansen A/S, Hørsholm, Denmark.

@J.U.: Department of Chemistry, University of Bergen, Bergen, Norway.

Author Contributions

H.K. and O.J.A. contributed equally and should be considered co-first authors.

Funding

This work was supported by grants from the Danish Council for Independent Research | Technology and Production Sciences (FTP 11−105010), the Danish National Research Foundation (DNRF59), and by a grant from the National Eye Institute (R01 EY012712). Computations were possible through allocations of time at the Centre for Scientific Computing, Aarhus. O.J.A. was supported by a grant from the Novo Scholarship Program.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

BS^3, bis(sulfosuccinimidyl)suberate; EMI, EMILIN-1; FAS1, fasciclin-1; FAS1-4, fourth FAS1 domain of TGFBIp; GCD, granular corneal dystrophy; IDP, intrinsically disordered protein; LC−MS/MS, liquid chromatography and tandem mass spectrometry; LCD, lattice corneal dystrophy; MD, molecular dynamics; MGF, Mascot generic format; M-ox, oxidized methionine; MW, molecular weight; NHS, *N*-hydroxysulfosuccinimide; PCA, principal component analysis; rmsd, root-mean-square deviation; SASA, solvent accessible surface area; Sso AcP, *S. solfataricus* acylphosphatase; SUMO, small ubiquitin-like modifier; TGFBIp, transforming growth factor β-induced protein.

■ REFERENCES

(1) Runager, K., Enghild, J. J., and Klintworth, G. K. (2008) Focus on molecules: Transforming growth factor beta induced protein (TGFBIp). *Exp. Eye Res. 87*, 298−299.

(2) Rawe, I. M., Zhan, Q., Burrows, R., Bennett, K., and Cintron, C. (1997) Beta-ig: Molecular cloning and in situ hybridization in corneal tissues. *Invest. Ophthalmol. Visual Sci. 38*, 893−900.

(3) Kannabiran, C., and Klintworth, G. K. (2006) TGFBI gene mutation in corneal dystrophies. *Hum. Mutat. 27*, 615−625.

(4) Yang, J., Han, X., Huang, D., Yu, L., Zhu, Y., Tong, Y., Zhu, B., Li, C., Weng, M., and Ma, X. (2010) Analysis of *TGFBI* gene mutations in Chinese patients with corneal dystrophies and review of the literature. *Mol. Vision 16*, 1186−1193.

(5) Underhaug, J., Koldsø, H., Runager, K., Nielsen, J. T., Sørensen, C. S., Kristensen, T., Otzen, D. E., Karring, H., Malmendal, A., Schiøtt, B., Enghild, J. J., and Nielsen, N. C. (2013) Mutation in transforming

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556

growth factor beta induced protein associated with granular corneal dystrophy type 1 reduces the proteolytic susceptibility through local structural stabilization. *Biochim. Biophys. Acta, Proteins Proteomics 1834*, 2812−2822.
(6) Runager, K., Basaiawmoit, R. V., Deva, T., Andreasen, M., Valnickova, Z., Sørensen, C. S., Karring, H., Thøgersen, I. B., Christiansen, G., Underhaug, J., Kristensen, T., Nielsen, N. C., Klintworth, G. K., Otzen, D. E., and Enghild, J. J. (2011) Human phenotypically distinct *TGFBI* corneal dystrophies are linked to the stability of the fourth FAS1 domain of TGFBIp. *J. Biol. Chem. 286*, 4951−4958.
(7) Andreasen, M., Nielsen, S. B., Runager, K., Christiansen, G., Nielsen, N. C., Enghild, J. J., and Otzen, D. E. (2012) Polymorphic fibrillation of the destabilized fourth fasciclin-1 domain mutant A546T of the transforming growth factor-β-induced protein (TGFBIp) occurs through multiple pathways with different oligomeric intermediates. *J. Biol. Chem. 287*, 34730−34742.
(8) Son, H., Nam, J., Kim, S., and Kim, I. (2013) Multiple FAS1 domains and the RGD motif of TGFBI act cooperatively to bind $\alpha v\beta 3$ integrin, leading to anti-angiogenic and anti-tumor effects. *Biochim. Biophys. Acta, Mol. Cell Res. 1833*, 2378−2388.
(9) Nam, J., Son, H., Jun, E., Cha, K., Lee, B., Park, R., and Kim, I. (2012) FAS1 domain protein inhibits $VEGF_{165}$-induced angiogenesis by targeting the interaction between VEGFR-2 and $\alpha v\beta 3$ integrin. *Mol. Cancer Res. 10*, 1010−1020.
(10) Kim, H., Kim, P., Bae, S. M., Son, H., Thoudam, D. S., Kim, J., Lee, B., Park, R., and Kim, I. (2009) Transforming growth factor-ß−induced protein (TGFBIp/ß ig-h3) activates platelets and promotes thrombogenesis. *Blood 114*, 5206−5215.
(11) Kim, H., and Kim, I. (2008) Transforming growth factor-ß-induced gene product, as a novel ligand of integrin $\alpha_M ß_2$, promotes monocytes adhesion, migration and chemotaxis. *Int. J. Biochem. Cell Biol. 40*, 991−1004.
(12) Nam, J., Kim, J., Jeong, H., Lee, S., Lee, B., Choi, J., Park, R., Park, J. Y., and Kim, I. (2003) Identification of the $\alpha_v ß_3$ integrin-interacting motif of ßig-h3 and its anti-angiogenic effect. *J. Biol. Chem. 278*, 25902−25909.
(13) Kim, J., Jeong, H., Nam, J., Lee, B., Choi, J., Park, R., Park, J. Y., and Kim, I. (2002) Identification of motifs in the fasciclin domains of the transforming growth factor-ß-induced matrix protein ßig-h3 that interact with the $\alpha v ß 5$ integrin. *J. Biol. Chem. 277*, 46159−46165.
(14) Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins. *FASEB J. 10*, 75−83.
(15) Karring, H., Runager, K., Thøgersen, I. B., Klintworth, G. K., Højrup, P., and Enghild, J. J. (2012) Composition and proteolytic processing of corneal deposits associated with mutations in the *TGFBI* gene. *Exp. Eye Res. 96*, 163−170.
(16) Karring, H., Poulsen, E. T., Runager, K., Thøgersen, I. B., Klintworth, G. K., Højrup, P., and Enghild, J. J. (2013) Serine protease HtrA1 accumulates in corneal transforming growth factor beta induced protein (TGFBIp) amyloid deposits. *Mol. Vision 19*, 861−876.
(17) Poulsen, E. T., Runager, K., Risør, M. W., Dyrlund, T. F., Scavenius, C., Karring, H., Praetorius, J., Vorum, H., Otzen, D. E., Klintworth, G. K., and Enghild, J. J. (2014) Comparison of two phenotypically distinct lattice corneal dystrophies caused by mutations in the transforming growth factor beta induced (TGFBI) gene. *Proteomics: Clin. Appl. 8*, 168−177.
(18) Morriss-Andrews, A., and Shea, J.-E. (2015) Computational studies of protein aggregation: Methods and applications. *Annu. Rev. Phys. Chem. 66*, 643−666.
(19) Dupuis, N. F., Wu, C., Shea, J.-E., and Bowers, M. T. (2011) The amyloid formation mechanism in human IAPP: Dimers have β-strand monomer-monomer interfaces. *J. Am. Chem. Soc. 133*, 7240−7243.
(20) Lemkul, J. A., and Bevan, D. R. (2010) Assessing the stability of Alzheimer's amyloid protofibrils using molecular dynamics. *J. Phys. Chem. B 114*, 1652−1660.
(21) Ma, B., and Nussinov, R. (2006) Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. *Curr. Opin. Chem. Biol. 10*, 445−452.
(22) Viswanath, S., Ravikant, D. V. S., and Elber, R. (2013) Improving ranking of models for protein complexes with side chain modeling and atomic potentials. *Proteins: Struct., Funct., Bioinf. 81*, 592−606.
(23) Ravikant, D. V. S., and Elber, R. (2011) Energy design for protein-protein interactions. *J. Chem. Phys. 135*, 065102.
(24) Ravikant, D. V. S., and Elber, R. (2010) PIE - Efficient filters and coarse grained potentials for unbound protein-protein docking. *Proteins: Struct., Funct., Bioinf. 78*, 400−419.
(25) *The PyMOL Molecular Graphics System*, Schrödinger, LLC, New York.
(26) Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pK_a values. *Proteins: Struct., Funct., Bioinf. 61*, 704−721.
(27) Neria, E., Fischer, S., and Karplus, M. (1996) Simulation of activation free energies in molecular systems. *J. Chem. Phys. 105*, 1902−1921.
(28) Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 1781−1802.
(29) MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B 102*, 3586−3616.
(30) MacKerell, A. D., Jr., Feig, M., and Brooks, C. L., III (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem. 25*, 1400−1415.
(31) MacKerell, A. D., Jr., Feig, M., and Brooks, C. L., III (2004) Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc. 126*, 698−699.
(32) Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys. 101*, 4177−4189.
(33) Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys. 103*, 4613−4621.
(34) Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald: An $N{\cdot}\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089−10092.
(35) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics 14*, 33−38.
(36) Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model. 7*, 306−317.
(37) Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput. 4*, 435−447.
(38) Shao, J., Tanner, S. W., Thompson, N., and Cheatham, T. E., III (2007) Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput. 3*, 2312−2334.
(39) Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2010) *AMBER 11*, University of California, San Francisco.

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556

(40) Bury, A. F. (1981) Analysis of protein and peptide mixtures: Evaluation of three sodium dodecyl sulphate-polyacrylamide gel electrophoresis buffer systems. *J. Chromatogr. A 213*, 491−500.
(41) Rasmussen, M. I., Refsgaard, J. C., Peng, L., Houen, G., and Højrup, P. (2011) CrossWork: Software-assisted identification of cross-linked peptides. *J. Proteomics 74*, 1871−1883.
(42) Dobson, C. M. (1999) Protein misfolding, evolution and disease. *Trends Biochem. Sci. 24*, 329−332.
(43) Lange, O. F., and Grubmüller, H. (2006) Can principal components yield a dimension reduced description of protein dynamics on long time scales? *J. Phys. Chem. B 110*, 22842−22852.
(44) David, C. C., and Jacobs, D. J. (2014) Principal component analysis: A method for determining the essential dynamics of proteins. *Methods Mol. Biol. 1084*, 193−226.
(45) Hayward, S., and de Groot, B. L. (2008) Normal modes and essential dynamics. *Methods Mol. Biol. 443*, 89−106.
(46) Clout, N. J., Tisi, D., and Hohenester, E. (2003) Novel fold revealed by the structure of a FAS1 domain pair from the insect cell adhesion molecule fasciclin I. *Structure 11*, 197−203.
(47) Basaiawmoit, R. V., Oliveira, C. L. P., Runager, K., Sørensen, C. S., Behrens, M. A., Jonsson, B. H., Kristensen, T., Klintworth, G. K., Enghild, J. J., Pedersen, J. S., and Otzen, D. E. (2011) SAXS models of TGFBIp reveal a trimeric structure and show that the overall shape is not affected by the Arg124His mutation. *J. Mol. Biol. 408*, 503−513.
(48) Rappsilber, J. (2011) The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol. 173*, 530−540.
(49) Merkley, E. D., Rysavy, S., Kahraman, A., Hafen, R. P., Daggett, V., and Adkins, J. N. (2014) Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci. 23*, 747−759.
(50) Wu, C., and Shea, J.-E. (2012) The structure of intrinsically disordered peptides implicated in amyloid diseases: Insights from fully atomistic simulations. In *Computational Modeling of Biological Systems* (Dokholyan, N. V., Ed.) 1st ed., pp 215−227, Springer, New York.
(51) Soto, C. (2003) Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci. 4*, 49−60.
(52) He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N., and Dunker, A. K. (2009) Predicting intrinsic disorder in proteins: An overview. *Cell Res. 19*, 929−949.
(53) Bouchard, M., Zurdo, J., Nettleton, E. J., Dobson, C. M., and Robinson, C. V. (2000) Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci. 9*, 1960−1967.
(54) Sørensen, J., Hamelberg, D., Schiøtt, B., and McCammon, J. A. (2007) Comparative MD analysis of the stability of transthyretin providing insight into the fibrillation mechanism. *Biopolymers 86*, 73−82.
(55) Stenvang, M., Andreasen, M., Enghild, J. J., and Otzen, D. E. (2014) The molecular basis for TGF Blp-related corneal dystrophies. In *Bio-nanoimaging: Protein Misfolding and Aggregation* (Uversky, V. N., and Lyubchenko, Y. L., Eds.) 1st ed., pp 179−188, Academic Press, New York.
(56) Pedersen, J. S., Christensen, G., and Otzen, D. E. (2004) Modulation of S6 fibrillation by unfolding rates and gatekeeper residues. *J. Mol. Biol. 341*, 575−588.
(57) Plakoutsi, G., Taddei, N., Stefani, M., and Chiti, F. (2004) Aggregation of the acylphosphatase from *Sulfolobus solfataricus*: The folded and partially unfolded states can both be precursors for amyloid formation. *J. Biol. Chem. 279*, 14111−14119.
(58) Chiti, F., and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol. 5*, 15−22.
(59) Relini, A., Marano, N., and Gliozzi, A. (2014) Misfolding of amyloidogenic proteins and their interactions with membranes. *Biomolecules 4*, 20−55.
(60) Plakoutsi, G., Bemporad, F., Calamai, M., Taddei, N., Dobson, C. M., and Chiti, F. (2005) Evidence for a mechanism of amyloid formation involving molecular reorganisation within native-like precursor aggregates. *J. Mol. Biol. 351*, 910−922.
(61) Sørensen, C. S., Runager, K., Scavenius, C., Jensen, M. M., Nielsen, N. S., Christiansen, G., Petersen, S. V., Karring, H., Sanggaard, K. W., and Enghild, J. J. (2015) Fibril core of transforming growth factor beta-induced protein (TGFBIp) facilitates aggregation of corneal TGFBIp. *Biochemistry 54*, 2943−2956.

DOI: 10.1021/acs.biochem.5b00473
Biochemistry 2015, 54, 5546−5556