



# Aggregation-Prone Structural Ensembles of Transthyretin Collected With Regression Analysis for NMR Chemical Shift

Wonjin Yang<sup>1</sup>, Beom Soo Kim<sup>1</sup>, Srinivasan Muniyappan<sup>2</sup>, Young-Ho Lee<sup>3,4,5,6</sup>, Jin Hae Kim<sup>2\*</sup> and Woogyung Yu<sup>1,7\*</sup>

<sup>1</sup>Department of Brain and Cognitive Sciences, DGIST, Daegu, South Korea, <sup>2</sup>Department of New Biology, DGIST, Daegu, South Korea, <sup>3</sup>Research Center for Bioconvergence Analysis, Korea Basic Science Institute, Ochang, South Korea, <sup>4</sup>Department of Bio-analytical Science, University of Science and Technology, Daejeon, South Korea, <sup>5</sup>Graduate School of Analytical Science and Technology, Chungnam National University, Daejeon, South Korea, <sup>6</sup>Research Headquarters, Korea Brain Research Institute, Daegu, South Korea, <sup>7</sup>Core Protein Resources Center, DGIST, Daegu, South Korea

## OPEN ACCESS

### Edited by:

Woonghee Lee,  
University of Colorado Denver,  
United States

### Reviewed by:

Suren Tatulian,  
University of Central Florida,  
United States

Xun Sun,  
Scripps Research Institute,  
United States

### \*Correspondence:

Jin Hae Kim  
jinhaekim@dgist.ac.kr  
Woogyung Yu  
wkyu@dgist.ac.kr

### Specialty section:

This article was submitted to  
Structural Biology,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 30 August 2021

**Accepted:** 05 October 2021

**Published:** 20 October 2021

### Citation:

Yang W, Kim BS, Muniyappan S,  
Lee Y-H Kim JH and Yu W (2021)  
Aggregation-Prone Structural  
Ensembles of Transthyretin Collected  
With Regression Analysis for NMR  
Chemical Shift.  
Front. Mol. Biosci. 8:766830.  
doi: 10.3389/fmolb.2021.766830

Monomer dissociation and subsequent misfolding of the transthyretin (TTR) is one of the most critical causative factors of TTR amyloidosis. TTR amyloidosis causes several human diseases, such as senile systemic amyloidosis and familial amyloid cardiomyopathy/polyneuropathy; therefore, it is important to understand the molecular details of the structural deformation and aggregation mechanisms of TTR. However, such molecular characteristics are still elusive because of the complicated structural heterogeneity of TTR and its highly sensitive nature to various environmental factors. Several nuclear magnetic resonance (NMR) spectroscopy and molecular dynamics (MD) studies of TTR variants have recently reported evidence of transient aggregation-prone structural states of TTR. According to these studies, the stability of the DAGH  $\beta$ -sheet, one of the two main  $\beta$ -sheets in TTR, is a crucial determinant of the TTR amyloidosis mechanism. In addition, its conformational perturbation and possible involvement of nearby structural motifs facilitates TTR aggregation. This study proposes aggregation-prone structural ensembles of TTR obtained by MD simulation with enhanced sampling and a multiple linear regression approach. This method provides plausible structural models that are composed of ensemble structures consistent with NMR chemical shift data. This study validated the ensemble models with experimental data obtained from circular dichroism (CD) spectroscopy and NMR order parameter analysis. In addition, our results suggest that the structural deformation of the DAGH  $\beta$ -sheet and the AB loop regions may correlate with the manifestation of the aggregation-prone conformational states of TTR. In summary, our method employing MD techniques to extend the structural ensembles from NMR experimental data analysis may provide new opportunities to investigate various transient yet important structural states of amyloidogenic proteins.

**Keywords:** transthyretin, nuclear magnetic resonance chemical shift, molecular dynamics computer simulation, protein aggregation, ensemble structure, linear regression

## INTRODUCTION

TTR is a transporter of the thyroid hormone, thyroxine ( $T_4$ ), and holo-retinol binding protein (Ingbar, 1958). It is one of the abundant proteins in human plasma (3–5  $\mu\text{M}$ ) and cerebrospinal fluid (0.25–0.5  $\mu\text{M}$ ) (Stabilini et al., 1968; Schreiber et al., 1990). In its native state, TTR has a  $\beta$ -sandwich structure consisting of two  $\beta$ -sheets, CBEF and DAGH. In addition, this protein maintains a homotetrameric complex, on which two hydrophobic binding pockets for  $T_4$  are constructed (Blake et al., 1978). In addition, TTR is also well known for its amyloidogenic propensity, causing several detrimental human diseases, such as senile systemic amyloidosis and familial amyloid polyneuropathy/cardiomyopathy (Westermarck et al., 1990; Coelho, 1996). Several biophysical analyses have shown that disruption of the tetrameric complex and subsequent release of monomeric species facilitates aggregation including amyloid fibril formation in TTR (Johnson et al., 2012). Dissociation of amyloidogenic monomers can be caused by several factors, including genetic mutations (Adams et al., 2019), post-translational modification (Poltash et al., 2019; Leri et al., 2020), and proteolysis by proteases, (Mangione et al., 2018; Peterle et al., 2020).

Despite its physiological and pathological importance, the molecular details of TTR aggregation remain elusive. A recent solution-state NMR study revealed that monomerization of TTR causes destabilization of the C-terminal  $\beta$ -stand H, making its neighboring  $\beta$ -stand G more accessible and vulnerable to amyloidogenesis (Oroz et al., 2017). It was previously shown that the TTR (105–115) peptide originating from the  $\beta$ -stand G is highly amyloidogenic (Gustavsson et al., 1991). A recent MD study reported a consistent result in which destabilization of the edge at the DAGH  $\beta$ -sheet, namely the  $\beta$ -stands D and H, is responsible for the amyloidogenic propensity of TTR (Zhou et al., 2019; Childers and Daggett, 2020). Furthermore, a series of computational studies have suggested that the DAGH  $\beta$ -sheet may experience structural deformation to reconstruct aggregation-prone  $\alpha$ -sheet-like structures (Steward et al., 2008; Childers and Daggett, 2019). Lim et al. employed solid-state NMR techniques to show that destabilization of the DAGH  $\beta$ -sheet may be caused by the conformational change in the AB loop region (Lim et al., 2016b). From TTR aggregates, they found that the native contact between Leu17 and Pro24 residues in the AB loop was lost, suggesting that non-native distortion of the AB loop may

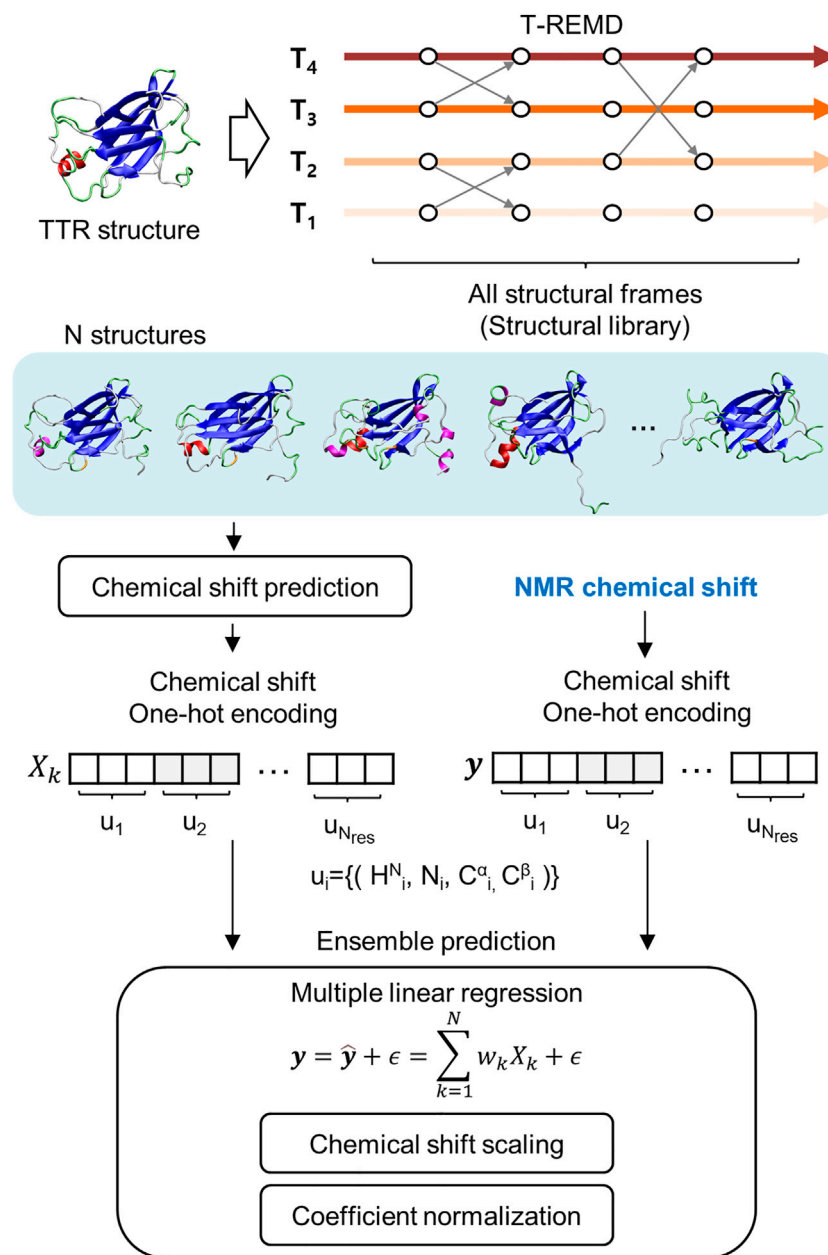
concur with amyloid fibril formation. The structural plasticity of the AB loop has been noted in prior solution-state NMR studies along with unstable structural features of the DAGH  $\beta$ -sheet (Lim et al., 2013; Das et al., 2014). However, there is still a significant gap between direct evidence and theoretical predictions to fully elucidate the molecular details of structural deformation and the resultant aggregation of TTR. In particular, a recent cryo-electron microscopic study of patient-derived TTR amyloid fibrils indicated that TTR should undergo global structural deformation during amyloidogenesis (Schmidt et al., 2019).

Recently, Google DeepMind developed an innovative method, AlphaFold2, which is a machine learning technique to predict the structure of monomeric and globular proteins from a given sequence (Jumper et al., 2021). AlphaFold2 shows a significant accuracy for globular proteins; however, it is still unknown whether AlphaFold2 can determine the structures of highly flexible proteins, such as intrinsically disordered proteins (IDPs) and metamorphic proteins, or investigate their dynamical features. On the other hand, NMR spectroscopy is a useful tool for investigating structural features of dynamic proteins (Kosol et al., 2013). NMR techniques, including nuclear Overhauser effect (NOE)-based techniques, residual dipolar coupling (RDC), paramagnetic relaxation enhancement, and NMR order parameter analysis, provide long-range or short-range contact information and the degree of structural heterogeneity. In addition, several methodologies using the information of inter-atomic distances or NMR J-coupling have been developed to define the structural ensemble of proteins under physiologically relevant conditions (Meng et al., 2018; Shimomura et al., 2019; Shrestha et al., 2019; Ferrie and Petersson, 2020; Lincoff et al., 2020). Our previous study based on NMR chemical shift, MD simulation, and machine learning technique with multiple linear regression provided a reliable ensemble structure of amyloid beta (Yang et al., 2021), the representative pathogenic IDP (Lin et al., 2019). This method provides the expected conformational states of highly mobile proteins at atomic resolution, which is a novel and rigorous approach to investigate various dynamic features of IDPs and intrinsically disordered regions (IDRs) of diverse proteins.

The regression analysis used *de novo* structures calculated from MD simulation and the chemical shift prediction algorithm (Figure 1). However, the previous approach mainly using  $^1\text{H}_\text{N}$  and  $^{15}\text{N}_\text{H}$  chemical shift information was insufficient to distinguish the secondary structural features. This is because the distributions of the chemical shifts of the  $^1\text{H}_\text{N}$  and  $^{15}\text{N}_\text{H}$  atoms for the different

**TABLE 1 |** The regression scores according to the chemical shift prediction algorithms.

	Prediction score (regression)					
	M-TTR			T119M M-TTR		
	UCBSHIFT	SHIFTX2	SPARTA+	UCBSHIFT	SHIFTX2	SPARTA+
$^1\text{H}_\text{N}$	0.8737	0.8429	0.6479	0.8864	0.8793	0.7471
$^{15}\text{N}_\text{H}$	0.9146	0.9363	0.8584	0.9465	0.9289	0.8527
$^{13}\text{C}_\alpha$	0.9686	0.9693	0.9226	0.9370	0.9491	0.9380
$^{13}\text{C}_\beta$	0.9963	0.9966	0.9969	0.9975	0.9968	0.9972
Total	0.7711	0.7623	0.5116	0.7843	0.7728	0.5959



**FIGURE 1** | Schematic flow of the regression approach for NMR chemical shift. It includes the generation of structural library, chemical shift prediction, and the main regression scheme for ensemble prediction.

secondary structures are statistically overlapped (Yu et al., 2011). This study improved the previous regression approach by introducing chemical shift information of the  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  atoms; the chemical shift of  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  show significant correlation with the secondary structure of proteins. We successfully introduced the general scaling process into the regression, irrespective of the type of used atom, which increases the accuracy of the regression approach. We revealed the reliable ensemble structure of two monomeric and highly-dynamic variants of TTR, M-TTR (F87M/L110M), and T119M M-TTR (F87M/L110M/T119M), and identified the minor yet reliable ensemble

structure using the regression approach. The newly determined M-TTR and T119M M-TTR ensembles provide novel and unprecedented insights into TTR aggregation mechanisms.

## METHOD

### Experimental Data Acquisition Circular Dichroism Measurement

Human recombinant TTR samples were prepared as previously described in the prior studies (Kim et al., 2016; Oroz et al., 2017).

For CD measurement, the concentration of the protein samples was adjusted to 20  $\mu\text{M}$  in a buffer consisting of 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 6.5, 100 mM NaCl, and 1 mM dithiothreitol. Cuvettes of 0.5 mm pathlength were used, and the measurement was performed at 25°C. The CD data for protein samples were obtained by subtracting the spectrum of the buffer-only sample.

### Order Parameter Calculation

The NMR chemical shift data for M-TTR and T119M M-TTR were obtained from BMRB entry ID 25986 and 25987, respectively (Kim et al., 2016; Oroz et al., 2017). In this study, the deposited chemical shift datasets were first verified using freshly prepared protein samples. Subsequently, the order parameters were calculated using TALOS-N by feeding the experimental chemical shifts of  $^1\text{H}_\text{N}$ ,  $^{15}\text{N}_\text{H}$ ,  $^{13}\text{CO}$ ,  $^{13}\text{C}_\alpha$ , and  $^{13}\text{C}_\beta$  (Shen and Bax, 2013).

## Molecular Dynamics Simulation

### System Preparation

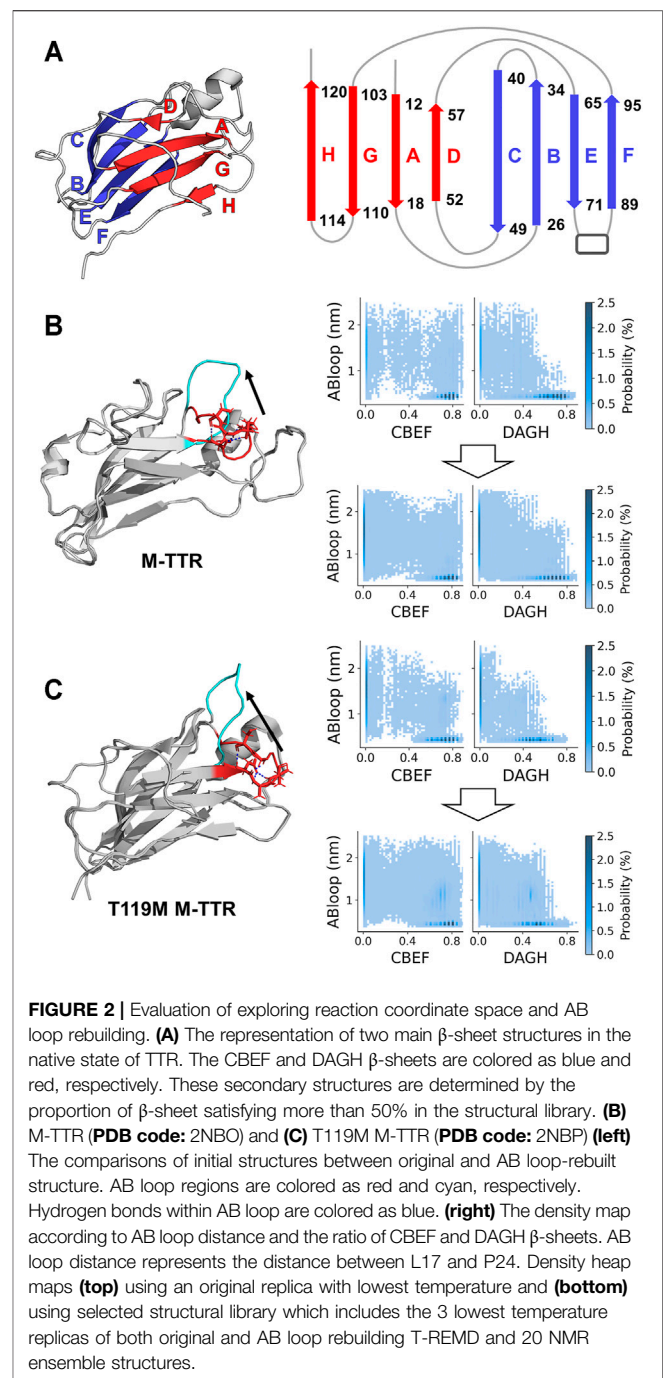
All systems were built using the Leap program, and all simulations were performed using the AMBER20 MD simulation package (Case et al., 2020). The Amber ff99SBildn force field (Lindorff-Larsen et al., 2010) was used for all simulations. Hydrogen atoms were constrained using the SHAKE algorithm (Ryckaert et al., 1977; Miyamoto and Kollman, 1992). The NMR solution structures of M-TTR and T119M M-TTR were used for MD simulations (PDB code: 2NBO (Oroz et al., 2017) and 2NBP (Kim et al., 2016), respectively). The generation of M-TTR and T119M M-TTR structures with AB loop rebuilding were performed using MODELLER 10.0 (Webb and Sali, 2016). The loop refinement process of the AB loop was applied to the positional restraint which makes the distance between the Leu17 and Pro24 residues to be  $20 \pm 1 \text{ \AA}$ . To use ionic strength effects, the salt concentration was 150 mM based on Debye-Hückel screening (Onufriev et al., 2002).

### Minimization and Equilibration

Each system was minimized with 5,000 steepest descent and a maximum of 2,500 conjugate gradient minimization steps. After the minimization steps, the systems were heated for 10 ns with 2 fs time step from 20 K to each target temperature. The temperature was regulated by a Langevin thermostat with  $1.0 \text{ ps}^{-1}$  collision frequency.

### Replica Exchange Molecular Dynamics

To generate an ensemble structure, replica exchange molecular dynamics (REMD) (Sugita and Okamoto, 1999) were performed using the PMEMD program in AMBER20 (Case et al., 2020). Each temperature value for the T-REMD simulation was generated using a temperature generator for REMD simulations (Patriksson and Van Der Spoel, 2008). For M-TTR and M-TTR with AB loop rebuilding, each system with a total of 16 replicas was simulated with a temperature range of 300–507 K. For T119M M-TTR and T119M M-TTR with AB loop rebuilding, each system with a total of 16 replicas was simulated with a temperature range of 300–480 K.



**FIGURE 2 |** Evaluation of exploring reaction coordinate space and AB loop rebuilding. **(A)** The representation of two main  $\beta$ -sheet structures in the native state of TTR. The CBEF and DAGH  $\beta$ -sheets are colored as blue and red, respectively. These secondary structures are determined by the proportion of  $\beta$ -sheet satisfying more than 50% in the structural library. **(B)** M-TTR (PDB code: 2NBO) and **(C)** T119M M-TTR (PDB code: 2NBP) (left) The comparisons of initial structures between original and AB loop-rebuilt structure. AB loop regions are colored as red and cyan, respectively. Hydrogen bonds within AB loop are colored as blue. **(right)** The density map according to AB loop distance and the ratio of CBEF and DAGH  $\beta$ -sheets. AB loop distance represents the distance between L17 and P24. Density heap maps **(top)** using an original replica with lowest temperature and **(bottom)** using selected structural library which includes the 3 lowest temperature replicas of both original and AB loop rebuilding T-REMD and 20 NMR ensemble structures.

During all simulations, exchanges were attempted every 2 ps, and each ensemble was simulated for 1  $\mu\text{s}$ . The total sampling time for each system was 16  $\mu\text{s}$ . The final average exchange ratios were 14.3, 14.2, 18.9, and 18.8%, respectively.

## Molecular Dynamics Trajectory Analysis

All trajectories were processed and analyzed using CPPTRAJ (Roe and Cheatham, 2013) provided by the AMBER20 package (Case et al., 2020). All snapshots of the trajectories were visualized using VMD (Humphrey et al., 1996). For each mutant TTR, the



six trajectories in the three lowest temperature replicas with and without AB loop rebuilding and 20 NMR ensemble structures were used in the contact map and secondary structure analysis. The contact maps were calculated by the distance between Ca-Ca atoms with a threshold of 7.5 Å. The secondary structure was analyzed using the STRIDE program (Frishman and Argos, 1995). The proportion  $P$  of secondary structures for each residue is calculated as followed equation:

$$P^i(\xi) = \frac{\sum_n \delta_{\xi_n=\xi}^i}{N},$$

$$Z = \{\alpha - \text{helix}, \beta - \text{sheet}, 3^{10} - \text{helix}, \text{coil}, \text{tail}\}$$

$Z$  is a set of secondary structure types. The proportion of specific secondary structure  $\xi \in Z$  for residue  $i$  is calculated by the total number of conformations which satisfies the secondary structure of  $n$ th conformation  $\xi_n = \xi$  in  $N$  conformational state of the structural library ( $N = 600,020$ ). And we defined the set  $B$  composed of specific residues which satisfy more than 50% proportion of  $\beta$ -sheet structure for each residue (Figure 2A). Finally, the proportion of  $\beta$ -sheet structure for each conformation is defined as:

$$P_n(\beta - \text{sheet}) = \frac{\sum_{i \in B} \delta_{\xi_n=\beta-\text{sheet}}^i}{|B|}$$

## Chemical Shift Prediction

All visualized results of chemical shift prediction were performed using UCBSHIFT (Li et al., 2020). The other chemical shift prediction algorithms such as SHIFTX2 (Han et al., 2011) and SPARTA+ (Shen and Bax, 2010) were also performed. The chemical shifts of backbone atoms ( $^1\text{H}_\text{N}$ ,  $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$ , and  $^{15}\text{N}_\text{H}$ ) for proteins were selectively used. The prediction was performed at pH 7.5. The residues whose  $^1\text{H}_\text{N}$ ,  $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$ , and  $^{15}\text{N}_\text{H}$  atoms were not fully assigned, including proline and glycine, were selectively eliminated. The input structures for chemical shift prediction were obtained from six replicas: the top three lowest temperatures of T-REMD for each original and AB loop rebuilding system, and the published NMR solution structure. Each trajectory gives us 100,000 structures; therefore, a total of 600,000 and 20 NMR solution structures were used for chemical shift prediction.

## Regression Approach

The multiple linear regression of the predicted chemical shifts was based on a previous study on amyloid beta molecules (Yang et al., 2021). We only used the fully assigned residues for  $^1\text{H}_\text{N}$ ,  $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$ , and  $^{15}\text{N}_\text{H}$  atoms. All regressions were performed using NumPy (Harris et al., 2020) and SciPy (Virtanen et al., 2020) modules, and all visualizations were performed using matplotlib (Hunter, 2007) in Python 3.6.

## Data Scaling

The scale of the chemical shift varies according to nuclear type. The scaling function  $f$  is introduced to solve the scale-difference problem. It is extended by including the additional  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$

atoms compared to the previous method (Yang et al., 2021). The scaling function for each atom is expressed as follows:

$$f: \mathbb{R} \rightarrow \mathbb{R}$$

$$f(x_i) = \log_a \left( \frac{x_i - y_{\min}}{y_{\max} - y_{\min}} \right) + 1 + \theta$$

$x_i$  is the chemical shift of a specific atom.  $y_{\min}$  and  $y_{\max}$  are the minimum and maximum values of the reference chemical shift for the same atom, respectively.  $a$  and  $\theta$  is hyperparameter for minimizing a regression error  $\epsilon$ . The constants  $a$  and  $\theta$  are determined from the regression for the min-max scaled chemical shift. The basis of min-max scaling is the scale of the reference chemical shift for each atom, similar to the above term in the log function.  $a$  is constant in  $[1.1, 10]$ , irrespective of the atom type. The determining constant  $\theta$  is based on a previous hyperparameter fine-tuning method for  $^1\text{H}_\text{N}$  chemical shifts (Yang et al., 2021). Thus,  $\theta$  is negligible for other atoms without  $^1\text{H}_\text{N}$  atoms. The hyperparameter tuning process was performed using the parallelized limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm (Gerber, 2020). The scaling function has an obvious inverse function because it is a one-to-one function. Thus, the coefficients of the multiple linear regression after scaling can be equally applied to the original regression without scaling.

## Multiple Linear Regression

For the interpretation of coefficients as probabilities or appearance of protein structures, we used non-negative least squares (NNLS) regression algorithms that solve the Karush-Kuhn-Tucker (KKT) condition for the non-negative least squares problem with an additional normalization method to make the sum of coefficients to be 1.

$$y = \hat{y} + \epsilon = w_1 X_1 + w_2 X_2 + \dots + w_n X_n + \epsilon$$

$$\sum_i^n w_i = 1$$

$y$  denotes the experimental chemical shift reference.  $\hat{y}$  is the predicted chemical shift among the ensemble of proteins.  $X_i$  is the chemical shift of the protein structure among the ensemble of proteins. The set  $\{X_i\}$  is the generated protein structure library from the MD simulation or other possible methods.  $\{w_i\}$  are the coefficients. In this condition, it represents the appearance or probability of each state.  $n$  is the number of input chemical shifts from the prediction.  $\epsilon$  is error value to be minimized. The minimization process was performed using the sequential least squares programming (SLSQP) algorithm in the SciPy optimization module (Virtanen et al., 2020).

## Feature Selection and Coefficient Normalization

The normalization of coefficients is also used in the previous method (Yang et al., 2021).

$$S = \{w_j\}_{j=1}^N \text{ such that } w_j \geq \epsilon \cdot w_m, w_m = \max_{w \in S} S$$

$S$  is the set of significant coefficients.  $\epsilon$  is a small positive value ( $10^{-5}$ ) as the threshold of selection; the previous study took the same value as a reasonable cutoff (Yang et al., 2021).  $w_m$  is the maximum element of the set  $S$ , which is equal to the maximum

element of the entire coefficient set. An additional regression with  $\sum_i^n w_i = 1$  constraint was performed on set  $S$  using the SLSQP algorithm.

### Data Scoring

The scoring method for the regression was based on the coefficient of determination, denoted  $R^2$ .

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$$

$$\text{RSS} = \sum_{k=1}^r (y_k - \hat{y}_k)^2, \quad \text{TSS} = \sum_{k=1}^r (y_k - \bar{y})^2$$

RSS and TSS are the residual sum of squares and total sum of squares, respectively.  $\bar{y}$  is the mean value of reference.  $\hat{y}_i$  denotes the predicted chemical shift.  $r$  is the number of residues in the protein. The total score combining the scores of all the atoms is as follows:

$$R_{\text{total}}^2 = \sum_{\text{atom} \in A} R_{\text{atom}}^2, \quad A = \{^1\text{H}_N, ^{15}\text{N}_H, ^{13}\text{C}_\alpha, ^{13}\text{C}_\beta\}$$

### Structure Clustering

After the regression, principal component analysis (PCA) and  $k$ -means clustering were performed for the predicted TTR ensemble. The PCA was based on the Ca-Ca contact map with a 7.5 Å cutoff distance.

## Nuclear Magnetic Resonance Order Parameter Analysis

The NMR order parameter  $S^2$  was computed using the isotropic reorientational eigenmode dynamics (iRED) method (Prompers and Brüschweiler, 2002) in CPPTRAJ (Roe and Cheatham, 2013). The input ensemble preparation from the regression approach used copying the feature conformations and duplicating each conformation in proportion to its regression coefficient. The final composition of the input ensemble included 1,000 structures. All analyses, except proline residues, were used to consider N-H atom vectors for each residue. The order parameters of the NMR ensemble for both M-TTR and T119M M-TTR were calculated using NMR solution structures.

## Calculation of Chemical Shift Prediction Error and Nuclear Magnetic Resonance Order Parameter Difference

The Euclidean distance in the projection space of the scaling function between NMR experimental data and the prediction of the regression is calculated as follows:

$$f(\Delta\delta_{\text{atom}}) = f(\delta_{\text{reg,atom}}) - f(\delta_{\text{ref,atom}})$$

$$\|f(\Delta\delta)\| = \sqrt{\sum_{\text{atom} \in A} f(\Delta\delta_{\text{atom}})^2}$$

$\delta_{\text{reg}}$  is the predicted value of the chemical shift from the regression.  $\delta_{\text{ref}}$  is a published chemical shift from the BMRB database for M-TTR and T119M M-TTR.  $\delta_{\text{atom}}$  is the nuclear-wise chemical shift.

The difference in the NMR order parameter  $S^2$  between the predicted ensemble from the regression and the NMR ensemble is defined as follows:

$$|\Delta S^2| = |S_{\text{reg}}^2 - S_{\text{ref}}^2|$$

## RESULTS

### Generation of Structural Library

The aim of our regression approach is to select the important conformations that comprise the protein ensemble in the possible conformation pool. In the free-energy landscape mapping of the conformational states of a system, a globular protein commonly has a sharp, stable state. Therefore, the major ensemble of globular proteins shows a similar conformation, including the backbone positions of the  $\alpha$ -helix or  $\beta$ -sheet secondary structures. Thus, the major ensemble of globular proteins can sufficiently represent the entire conformation of the protein. However, this statement is not equally true to dynamic soluble proteins, including IDPs or locally spanned IDRs. The ensemble of the dynamic protein includes many minor state conformations. As a result, the conventional approach often fails to sufficiently reflect minor yet important conformational states. In contrast, our regression approach incorporating NMR chemical shift information can provide a reliable protein ensemble with extended consideration for minor states. NMR chemical shifts collectively reflect appearances and mobile features of all ensemble states at atomic resolution. Therefore, our approach is an accurate and efficient strategy to constrain the MD-based ensemble pool without inadequate removal of relevant conformational states.

To consider important conformations in a sufficient size, we performed MD simulations with enhanced sampling and temperature replica exchange molecular dynamics (T-REMD) to fully explore the conformational space of the protein (Figure 1). The number of input conformations for the regression should be sufficiently large to obtain important conformations. In this process, we refer to a set of possible conformational states as a structural library. T-REMD makes the structural library for M-TTR and T119M M-TTR. However, we considered the possibility of insufficient exploration of the MD simulation because of the large size of TTR. Therefore, we analyzed the amount of exploration of the reaction coordinate space with the proportion of CBEP and DAGH  $\beta$ -sheets and related AB loops. In the present study, we put a particular focus on the AB loop, because it was proposed as an important structural element contributing to the stability of the DAGH  $\beta$ -sheet. A prior solid-state NMR study evidenced that TTR lost the native-like AB loop conformation (Lim et al., 2016b); we consider this observation important and appropriate because it is made with actual aggregates of TTR, not in an un-aggregated soluble state. The contact between the AB loop and DAGH  $\beta$ -sheet is calculated by the distance between Leu17 and Pro24 residues, and we call this term the short AB loop distance. From this analysis, we confirmed that a single simulation could not

sufficiently explore the states in which the DAGH  $\beta$ -sheet was stable, and the AB loop distance was large (**Figure 2**). Considering the previous evidence, we generated the artificial structures of M-TTR and T119M M-TTR to satisfy the AB loop distance, and the average proportion of DAGH  $\beta$ -sheets became large (**Figure 2**). We used the top three lowest temperature replicas from the T-REMD simulation with the AB loop rebuilding structure for each M-TTR (**Supplementary Figures S1**). In addition, we considered the NMR ensemble structures of M-TTR and T119M M-TTR as the major ensemble states. After combining all structures into the structural library, we prepared the input structures to map the possible conformational space.

## Determination of Transthyretin Ensemble Using the Multiple Linear Regression for Nuclear Magnetic Resonance Chemical Shift

After generating the structural library, we predicted the NMR chemical shift from each structure in the library. A previous study showed that UCBSHIFT (Li et al., 2020), a chemical shift prediction algorithm, provides better regression quality for IDP-like protein, e.g., amyloid-beta (Yang et al., 2021). UCBSHIFT might be a good prediction algorithm for the mobile C-terminal region of TTR to perform the regression. We prepared the prediction dataset of NMR chemical shifts for all conformations in the structural library using UCBSHIFT. The regression used a total of 600,020 chemical shift sets. The previous study used only the chemical shift values of  $^1\text{H}_\text{N}$  and  $^{15}\text{N}_\text{H}$  atoms for the regression. We have added the chemical shift of the additional  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  atoms. During the regression, we neglected the partially assigned residues for  $^1\text{H}_\text{N}$ ,  $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$ , and  $^{15}\text{N}_\text{H}$  atoms, such as proline and glycine residues. The regression solves the minimization problem on the KKT condition with an additional constraint satisfying the sum of coefficients to be one, which uses a non-negative least squares (NNLS) algorithm and additional optimization. To satisfy the revised KKT condition, we interpreted the regression coefficients as the appearance of the conformations in the ensemble. During the minimization of regression, we introduced a hyper-parameter  $\theta$  to optimize the  $^1\text{H}_\text{N}$  chemical shift, which depends on the reference chemical shift. The hyper-parameters  $\theta$  of M-TTR and T119M M-TTR were set to 0.0661 and 0.0431 ppm, respectively. The constants  $a$  as the logarithm base in the scaling function were selected in [1.1, 10]. As a result,  $a$  values were set to 10. Finally, multiple linear regression analysis provided the probability of each conformational state in the structural library.

As a result, the regression provided a fitted NMR chemical shift to the experimental reference. We used simple regression analysis to score the regression: the coefficient of determination ( $R^2$ ), which is a statistical measure that shows the proportion of variation. The atom scores for  $^1\text{H}_\text{N}$ ,  $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$ , and  $^{15}\text{N}_\text{H}$  atoms were 0.8737, 0.9686, 0.9963, and 0.9146 in M-TTR, and 0.8864, 0.9370, 0.9975, and 0.9465 in T119M M-TTR, respectively. The  $^{13}\text{C}-^{15}\text{N}_\text{H}$  and  $^1\text{H}_\text{N}-^{15}\text{N}_\text{H}$  chemical shift plots comparing the experimental and predicted data are shown in **Figure 3**. The chemical shift error for each atom is also shown in

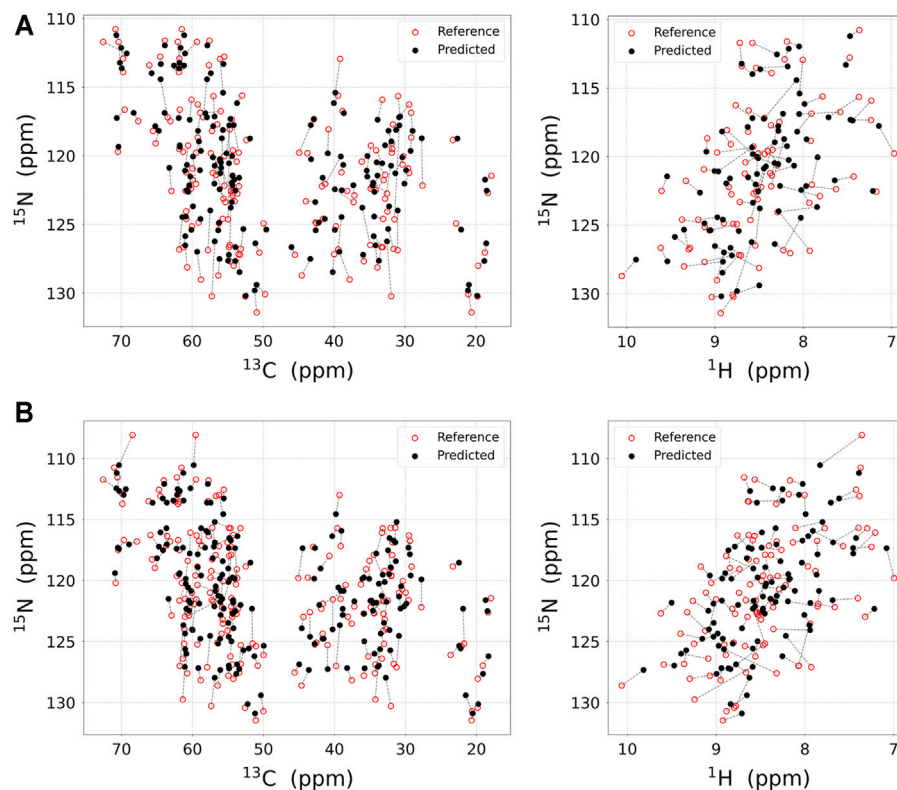
**Supplementary Figure S2**. After the regression, we compared the original NMR ensemble with the predicted TTR ensemble using regression (**Figure 4**). The regression provided possible local conformations. We partitioned the conformational ensemble from the regression into clusters using  $k$ -means clustering in the two-dimensional principal component plane based on the Ca-Ca contact map (**Supplementary Figure S3**). The major ensembles of both M-TTR and T119M M-TTR include the NMR ensemble conformations as the center of cluster. The second predominant M-TTR cluster show the rigid H  $\beta$ -stand with an increased conformational homogeneity. Notably, the less populated M-TTR cluster exhibit the relaxed  $\beta$ -barrel-like conformation, which satisfies the long AB loop distance between Leu17 and Pro24 residues. These  $\beta$ -barrel conformations have a native-like  $\beta$ -sandwich template including the CBEF and AG  $\beta$ -sheets except that it has a very long D  $\beta$ -strand which connects the CBEF and AG  $\beta$ -sheets into a circular barrel shape. This observation raises an intriguing possibility that disruption of the AB loop may correlate with overall structural perturbation and subsequent aggregation of TTR. On the other hand, the prediction ensembles of T119M M-TTR are similar to those of the previously established with NMR spectroscopy except for a slight difference in the EF loop and the D  $\beta$ -strand. This observation is consistent with the previous studies where T119M M-TTR maintains more homogeneous structural states than M-TTR (Lim et al., 2013; Kim et al., 2016).

Finally, to strictly consider the implication of choosing different algorithms for chemical shift prediction, we performed the regression approach with SHIFTX2 and SPARTA+ (**Supplementary Figure S4**). Upon comparing the regression quality using  $R^2$  score (**Table 1**), we found that the total regression scores for the four atoms are 0.7711, 0.7623 and 0.5116 in M-TTR, and 0.7843, 0.7728 and 0.5959 in T119M M-TTR with the order of UCBSHIFT, SHIFTX2, and SPARTA+, respectively. In the perspective of the regression score, UCBSHIFT is slightly more appropriate than the others for our regression approach, as we concluded in our previous study (Yang et al., 2021).

## Validation of the Transthyretin Ensemble Predictions

### Analysis of the Secondary Structures

To quantitatively measure the composition of the secondary structures of M-TTR, we used CD spectroscopy with BeStSel (Beta Structure Selection) analysis (Micsonai et al., 2015), which provides the proportion of the secondary structures from the CD spectra using a machine learning approach (**Figure 5A**). After BeStSel analysis, we compared the secondary structure composition of the predicted ensemble from the regression with the BeStSel results (**Figure 5B**). Combining all results, we confirmed that our prediction for the TTR ensemble give the most similar results to the BeStSel analysis (**Figure 5** and **Table 2**). In particular, our prediction provided a more reliable  $\beta$ -sheet proportion analysis result for M-TTR than that from the previous NMR ensemble. Subsequent detailed



**FIGURE 3 |** The result of the chemical shift regression approach. Experimental NMR chemical shift (red open circle) and predicted chemical shift (black filled circle) using regression approach for (A) M-TTR and (B) T119M M-TTR are displayed in 2D plane. Chemical shifts for the same residues are linked with dotted line. The chemical shift prediction for  $^{15}\text{N}_{\text{H}}$ ,  $^{13}\text{C}_{\alpha}$ , and  $^{13}\text{C}_{\beta}$  atoms (left) and for  $^{15}\text{N}_{\text{H}}$  and  $^1\text{H}_{\text{N}}$  atoms (right) are respectively plotted. All visualized regression results are obtained with UCBSHIFT. The regression score for each atom is represented in Table 1.

analysis of the results for M-TTR identified that the major difference between our prediction and NMR ensemble is the existence of the C-terminal H  $\beta$ -strand and two short  $\beta$ -strands in residues 21–22 and 54–57. This corroborates that our approach is effective to provide an additional conformational ensemble, which NMR-based ensemble failed to accommodate. In contrast, T119M M-TTR showed that the  $\beta$ -sheet residues from the prediction are mostly consistent with the  $\beta$ -sheet residues of the NMR ensemble. **Supplementary Figure** provide a comparison between before and after the regression of the secondary structure (**Supplementary Figure S5**) and the contact map analysis (**Supplementary Figure S6**). After the regression, the composition of the  $\beta$ -sheet was more manifested than before the regression. This indicates that our regression procedure can efficiently select the relevant and meaningful conformational features.

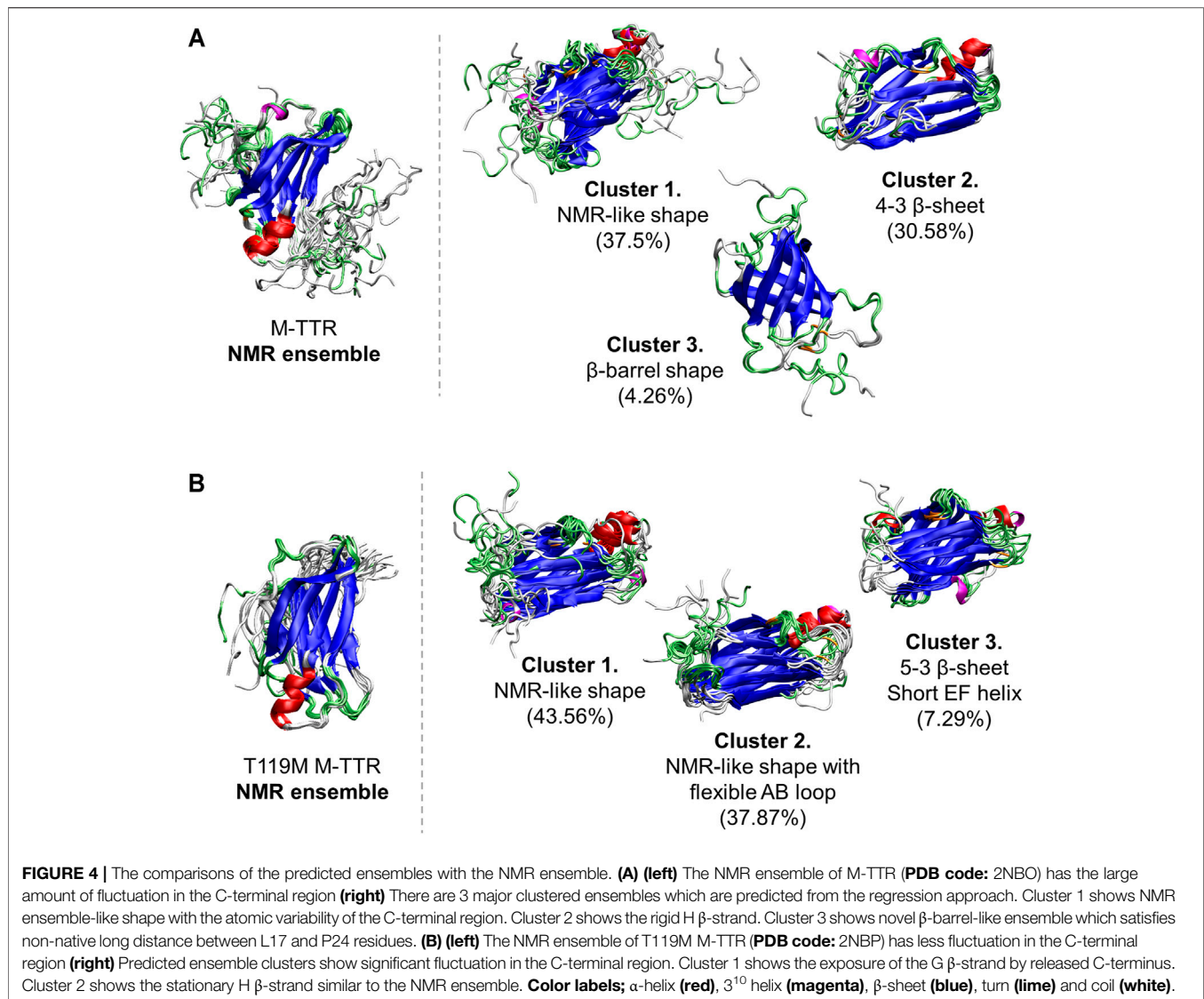
### Analysis of the Nuclear Magnetic Resonance Order Parameter

To verify the predicted M-TTR and T119M M-TTR ensembles, we performed a comparative analysis of the NMR order parameter, representing the amount of fluctuation of the N-H bond vector. We prepared the regression ensemble by copying the feature

conformations and duplicating each conformation in proportion to its regression coefficient composed of 1,000 structures. We calculated the NMR order parameter for each residue using N-H vectors, except for proline residues. Combining all NMR order parameter data, we could check the highly conserved regions which compose the specific secondary structures in the solution NMR ensemble (**Figure 6A**). Our regression ensemble is more consistent to the experimental order parameter except for a few loop regions of both M-TTR and T119M M-TTR. The order parameter analysis for the NMR ensemble showed that the order parameter for the residues around the secondary structure was close to 1. Therefore, the NMR ensemble has more stationary secondary structures and consistent alignment of N-H vectors than the experimental conditions. Each mutant TTR ensemble derived from our prediction was more consistent to its experimental order parameter than the NMR ensemble. In particular, the C-terminal regions of both mutant TTRs maintain the similarities of NMR order parameters between experimental data and prediction. These observations again support the superiority of our approach to reflect the actual structural conformations than the NMR-based ensemble selection.

We verified the correlation between the difference in the NMR order parameters and the chemical shift prediction error (**Figure 6B** and **Supplementary Figure S2**). The chemical



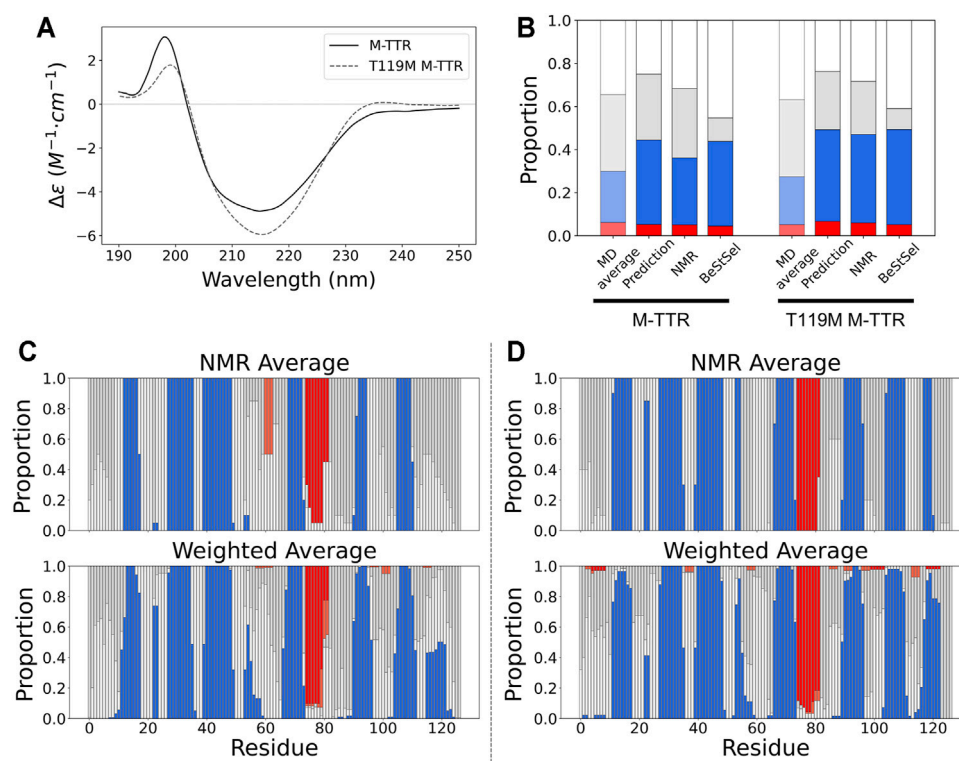


shift error was calculated using the Euclidean metric in the projection space of the scaling function. The projection space maintains the relative position of each residue in the chemical shift for each atom. From this analysis, we acknowledged that the tendency of the absolute difference of NMR order parameter and the tendency of the chemical shift error in the projection space of the scaling function are similar to each other (**Figure 6B**), although the quality of regression according to the structural library is robust (**Supplementary Figure S7**). This observation implies that the chemical shift prediction error is a major limiting factor to have more accurate results with our approach.

## DISCUSSIONS

Protein aggregation and amyloidosis are among the most critical events associated with various detrimental pathological processes in humans (Chiti and Dobson, 2017). Although several studies

have been conducted to understand the related mechanisms, their mechanistic details are still lacking. This is attributed to the highly heterogeneous and dynamic structural states of proteins in their aggregation-prone states (Kelly, 1996; Yang et al., 2021). To overcome these challenges, NMR spectroscopy (Daskalov et al., 2021; Dyson and Wright, 2021) and MD simulation techniques (Prabakaran et al., 2021; Strodel, 2021) are the two major methodologies that significantly contributed to advancing our understanding of the aggregation and amyloidosis mechanisms of various proteins. Indeed, NMR spectroscopy has been a major technique for investigating the mobile structural features of IDPs and amyloidogenic proteins, such as amyloid beta (Crescenzi et al., 2002), tau (Mukrasch et al., 2009), and  $\alpha$ -synuclein (Ulmer et al., 2005). In contrast, MD simulations provides physical movements of atoms with the evolution of femtosecond dynamics. It determines the forces and potential energies of interatomic interactions by solving Newton's equations of motion, which can determine the



**FIGURE 5 |** The comparison of secondary structures between NMR ensemble and predicted ensemble from the regression approach. **(A)** The CD spectra for M-TTR (solid line) and T119M M-TTR (dotted line) **(B)** The secondary structure proportion for all MD trajectories (MD average), predicted ensemble (prediction), NMR ensemble (NMR) and BeStSel prediction using CD spectra (BeStSel). The residual secondary structures of both **(C)** M-TTR and **(D)** T119M M-TTR. **Color labels**  $\alpha$ -helix (red),  $\beta$ -sheet (blue), turn (gray) and coil (white).

**TABLE 2 |** The secondary structure proportion for each TTR ensemble.

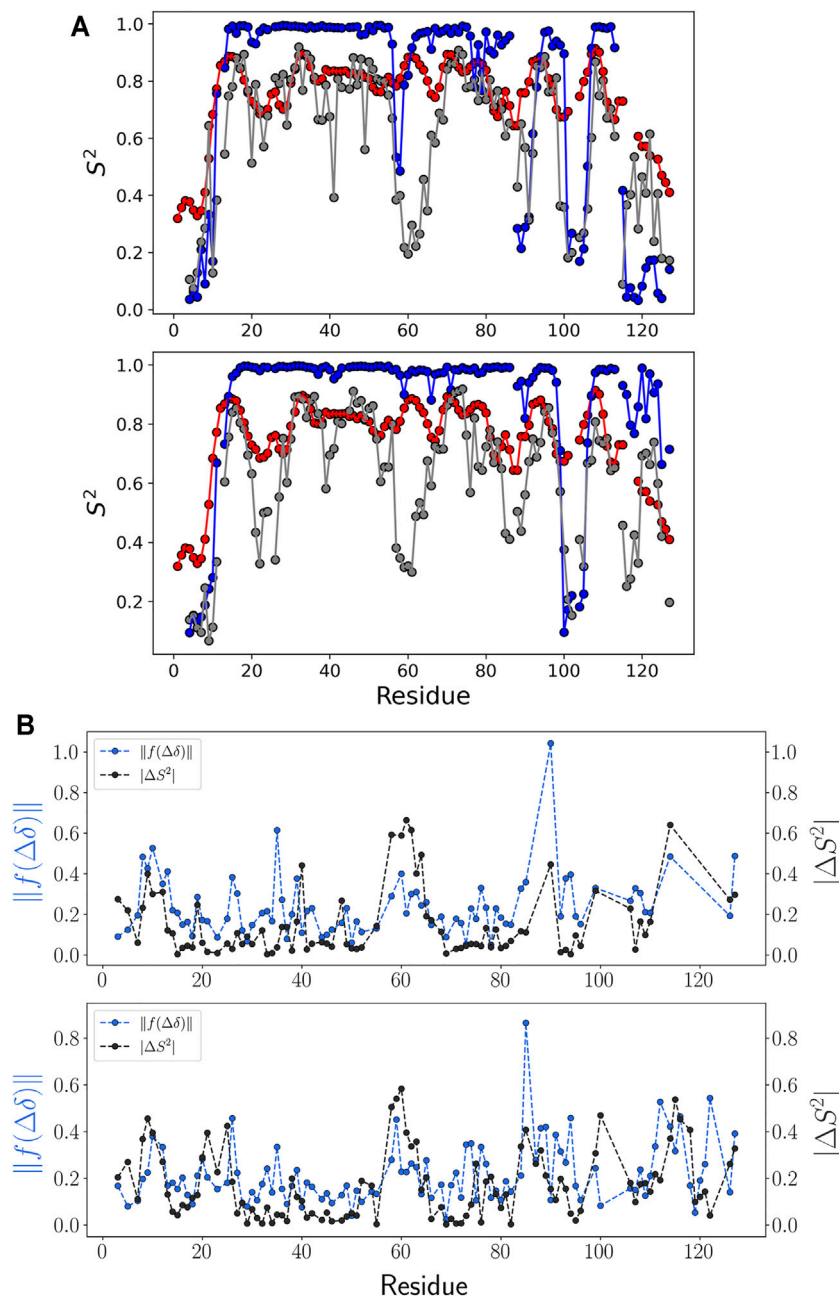
Secondary structure	Secondary structure proportion (%)							
	M-TTR				T119M M-TTR			
	Before regression	After regression	NMR ensemble	BeStSel	Before regression	After regression	NMR ensemble	BeStSel
$\alpha$ -helix	6.2	5.7	5.1	4.5	5.2	6.2	6.0	5.2
$\beta$ -sheet	23.7	41.5	31.0	39.3	22.2	43.4	41.0	44.1
Turn	35.6	28.3	32.2	10.9	35.7	29.0	24.7	9.7
Coil	34.5	24.5	31.7	45.3	36.9	21.3	28.3	41.0

thermodynamically stable structure of the protein. Several MD-based studies have investigated the dynamics and aggregation mechanisms of amyloidogenic proteins, such as amyloid beta (Tran and Ha-Duong, 2015), tau (Leonard et al., 2021),  $\alpha$ -synuclein (Otaki et al., 2018), and other proteins (Meli et al., 2008; Spagnoli et al., 2020).

TTR has been an important target of various structural studies because of its physiological and pathological importance. The native tetrameric structure of TTR is maintained to exert its physiological role as a carrier of thyroid hormones and retinol-binding proteins, whereas the amyloidogenic propensities manifest upon its monomerization (Johnson et al., 2012). Recent NMR spectroscopic

studies have shown that M-TTR stabilizes heterogeneous states, in which the C-terminal  $\beta$ -strand becomes highly mobile (Oroz et al., 2017). However, this study could not exclude the possible multiple structural states of this  $\beta$ -strand and the subsequent structural rearrangement. Moreover, it is not clear how its disordered nature correlates with the aggregation-prone property of TTR. It is also noteworthy that the NMR data for determining the M-TTR structural models were obtained under a pressurized condition (0.5 kbar) (Oroz et al., 2017), implying that even more diverse structural heterogeneity may manifest in a physiological condition.

There are many *in silico* methods to determine the structure of IDPs, such as amyloid-beta (Saravanan et al., 2020). Meng et al.



**FIGURE 6 |** NMR order parameter and chemical shift prediction error. **(A)** NMR order parameter  $S^2$  for each ensemble of **(top)** M-TTR and **(bottom)** T119M M-TTR is colored as follows: NMR experiment (**red**), calculation using NMR ensemble (**blue**), and calculation using the predicted ensemble (**gray**). **(B)** Comparison between NMR order parameter error and chemical shift prediction error. Errors were calculated by the difference between the regression result and NMR experiment data. Especially, the error calculation of chemical shift prediction was calculated in the projection space by scaling function  $f$ .

(2018) attempted to obtain molecular-level insight and a distinguishable conformational ensemble of the IDP-like protein using MD simulation with single-molecule Förster resonance energy transfer spectroscopy. Other studies used the interatomic distance information from NOEs and RDCs or scalar coupling information to reveal the ensemble of soluble proteins (Meng et al., 2018; Shimomura et al., 2019; Shrestha et al., 2019; Ferrie and Petersson, 2020; Lincoff et al., 2020). Our studies used

NMR chemical shift data to determine the conformational state of the protein ensemble. Several methodologies have been developed to use chemical shift data with fragment-based approaches (Cavalli et al., 2007; Nerli et al., 2018; Chandy et al., 2020). Our novel method uses a different approach to extend the experimental observation of NMR spectroscopy using MD simulations. It gives each conformation of the selected ensemble, which is not restricted to any structural constraints.

As a result, we can obtain diverse conformational states at atomic resolution in the solution. Moreover, we would like to stress that this regression methodology may be further improved by incorporating additional experimental data, such as secondary structure contents from CD spectroscopy, NOE-based distance information, and J coupling-based torsion angle data. Finally, the present study efficiently expands the exploration range of MD simulation by reflecting the previous experimental observation of the non-native AB loop distance in aggregated TTR. We think that the similar strategy can be effective for MD simulation to explore additional structural abnormality, whose correlation with aggregation propensity was proposed, e.g., the CD loop (Klimtchuk et al., 2018; Dasari et al., 2020), the EF helix/loop (Sun et al., 2018; Ferguson et al., 2021), and the H  $\beta$ -strand (Oroz et al., 2017).

The significant difference between the structural models determined from NMR experimental data and the extended ensembles of M-TTR reported in this study indicates that the C-terminal  $\beta$ -stand, which was determined to be disordered in the NMR models, is still highly mobile. However, it also appears that at least some population of M-TTR may stabilize a native-like  $\beta$ -stand structure in the C-terminal region. NMR structure determination procedures are highly dependent on the accurate analysis of NOE signals, thus limiting the observation of dominant conformations even in the presence of coexisting multiple states. It has been suggested that M-TTR may have several distinctive conformations under native conditions, as observed in the tetrameric conformation of the X-ray crystallographic study (Ulmer et al., 2005), the monomeric conformation of the pressurized NMR study (Oroz et al., 2017), and the distinctive monomeric conformation of the T119M M-TTR NMR study (Kim et al., 2016). In particular, the extended ensembles of this study correlate well with the NMR relaxation dispersion results of WT TTR and M-TTR (Lim et al., 2013; Das et al., 2014), supporting the superiority of our novel methodology for characterizing structural heterogeneity. Finally, the extended ensembles exhibited that non-native loosening of the AB loop accompanies with universal and significant structural perturbation. The AB loop was previously proposed as a region whose structural changes are related to the aggregation of TTR (Lim et al., 2016a). The ensembles indicate that monomerization of TTR may incur structural deformation in the C-terminal  $\beta$ -stand and the AB loop, followed by further structural rearrangement to facilitate amyloid generation.

Our results indicate that T119M M-TTR may have more homogeneous structural states than M-TTR. This is consistent with a series of studies in which T119M substitution increased the overall structural stability of TTR (Hammarström et al., 2003; Lim et al., 2013; Das et al., 2014). In addition, the present ensembles provide a couple of intriguing predictions. First, the C-terminal  $\beta$ -stand harbors reduced yet still significant dynamic features, explaining why T119M M-TTR is more amyloidogenic than WT TTR (Lim et al., 2013). Moreover, our results indicate that the M119 sidechain exhibits several distinctive directions in the ensembles. In previous NMR structural models, the M119 sidechain was positioned inward, suggesting that hydrophobic interaction of the M119 sidechain with other nearby residues may stabilize the C-terminal  $\beta$ -stand structures of TTR (Kim et al., 2016). However, the present structural ensembles show that this residue may have some residual,

dynamic features. Subsequent investigation is evidently necessary to appreciate how the mobility of M119 (or T119 of WT TTR) contributes to the aggregation propensity of TTR; yet, our results imply that this residue- or region-specific dynamics may represent structural heterogeneity of TTR in its monomeric and aggregation-prone states. We envision that the models from this study may provide unprecedented insights to design subsequent experimental strategies and to advance our understanding to the aggregation mechanism of TTR.

In summary, these observations support the strength of the current approach in that the calculated ensembles better represent the residual structural flexibility and the amyloidogenic propensity of M-TTR and T119M M-TTR. Although the conformational shape of “real” amyloidogenic species is of great interest to elucidate the mechanisms of amyloidogenesis in detail, its direct experimental observation is challenging due to its heterogeneous and aggregation-prone nature. We expect that our novel methodology may provide a powerful and efficient way to appreciate the dynamic features of amyloidogenic proteins and to reveal the related mechanistic details regarding their physiology or pathology.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## FUNDING

This work was supported by RandD Programs of DGIST (21-CoE-BT-01) funded by the Ministry of Science and ICT of Korea (W-K.Y.), the National Research Foundation funded by the Ministry of Science and ICT, Republic of Korea (NRF-2021R1F1A1056456 to W-K.Y., NRF-2018R1C1B6008282 to J.H.K., and NRF-2019R1A2C1004954 to Y.-H.L.), and KBSI (C130000, C180310, C140130 and C170100 to Y.-H.L.).

## ACKNOWLEDGMENTS

We thank the DGIST supercomputing and big data center for the allocation of dedicated supercomputing time.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.766830/full#supplementary-material>



## REFERENCES

- Adams, D., Koike, H., Slama, M., and Coelho, T. (2019). *Hereditary transthyretin amyloidosis: A model of medical progress for a fatal disease*. Berlin: Nature Publishing Group.
- Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rérat, B., and Rérat, C. (1978). Structure of prealbumin: Secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. *J. Mol. Biol.* 121, 339–356. doi:10.1016/0022-2836(78)90368-6
- Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., Iii, Cruzeiro, V.W.D., et al. (2020). *Amber 2020*. San Francisco: University of California.
- Cavalli, A., Salvatella, X., Dobson, C. M., and Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci.* 104, 9615–9620. doi:10.1073/pnas.0610313104
- Chandy, S. K., Thapa, B., and Raghavachari, K. (2020). Accurate and cost-effective NMR chemical shift predictions for proteins using a molecules-in-molecules fragmentation-based method. *Phys. Chem. Chem. Phys.* 22, 27781–27799. doi:10.1039/d0cp05064d
- Childers, M. C., and Daggett, V. (2019). Drivers of  $\alpha$ -Sheet Formation in Transthyretin under Amyloidogenic Conditions. *Biochemistry* 58, 4408–4423. doi:10.1021/acs.biochem.9b00769
- Childers, M. C., and Daggett, V. (2020). Edge strand dissociation and conformational changes in transthyretin under amyloidogenic conditions. *Biophysical J.* 119, 1995–2009. doi:10.1016/j.bpj.2020.08.043
- Chiti, F., and Dobson, C. M. (2017). Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu. Rev. Biochem.* 86, 27–68. doi:10.1146/annurev-biochem-061516-045115
- Coelho, T. (1996). Familial amyloid polyneuropathy. *Curr. Opin. Neurol.* 9, 355–359. doi:10.1097/00019052-199610000-00007
- Crescenzi, O., Tomaselli, S., Guerrini, R., Salvadori, S., D'Ursi, A. M., Temussi, P. A., et al. (2002). Solution structure of the Alzheimer amyloid  $\beta$ -peptide (1–42) in an apolar microenvironment. *Eur. J. Biochem.* 269, 5642–5648. doi:10.1046/j.1432-1033.2002.03271.x
- Das, J. K., Mall, S. S., Bej, A., and Mukherjee, S. (2014). Conformational flexibility tunes the propensity of transthyretin to form fibrils through non-native intermediate states. *Angew. Chem. Int. Ed.* 53, 12781–12784. doi:10.1002/anie.201407323
- Dasari, A. K. R., Arreola, J., Michael, B., Griffin, R. G., Kelly, J. W., and Lim, K. H. (2020). Disruption of the CD Loop by Enzymatic Cleavage Promotes the Formation of Toxic Transthyretin Oligomers through a Common Transthyretin Misfolding Pathway. *Biochemistry* 59, 2319–2327. doi:10.1021/acs.biochem.0c00079
- Daskalov, A., El Mammeri, N., Lends, A., Shenoy, J., Lamon, G., Fichou, Y., et al. (2021). Structures of pathological and functional amyloids and prions, a solid-state NMR perspective. *Front. Mol. Neurosci.* 14, 670513. doi:10.3389/fnmol.2021.670513
- Dyson, H. J., and Wright, P. E. (2021). NMR illuminates intrinsic disorder. *Curr. Opin. Struct. Biol.* 70, 44–52. doi:10.1016/j.sbi.2021.03.015
- Ferguson, J. A., Sun, X., Dyson, H. J., and Wright, P. E. (2021). Thermodynamic Stability and Aggregation Kinetics of EF Helix and EF Loop Variants of Transthyretin. *Biochem.* 00071c00073. doi:10.1021/acs.biochem.1c00073
- Ferrie, J. J., and Petersson, E. J. (2020). A unified de novo approach for predicting the structures of ordered and disordered proteins. *J. Phys. Chem. B* 124, 5538–5548. doi:10.1021/acs.jpcc.0c02924
- Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins* 23, 566–579. doi:10.1002/prot.340230412
- Gerber, F. (2020). *florafauna/optimParallel-python v0.1.1 (v0.1.1)*. Zenodo.
- Gustavsson, Å., Engström, U., and Westermark, P. (1991). Normal transthyretin and synthetic transthyretin fragments from amyloid-like fibrils *in vitro*. *Biochem. Biophysical Res. Commun.* 175, 1159–1164. doi:10.1016/0006-291x(91)91687-8
- Hammarström, P., Wiseman, R. L., Powers, E. T., and Kelly, J. W. (2003). Prevention of Transthyretin Amyloid Disease by Changing Protein Misfolding Energetics. *Science* 299, 713–716. doi:10.1126/science.1079589
- Han, B., Liu, Y., Ginzinger, S. W., and Wishart, D. S. (2011). SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* 50, 43–57. doi:10.1007/s10858-011-9478-4
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi:10.1038/s41586-020-2649-2
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graphics* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi:10.1109/mcse.2007.55
- Ingbar, S. H. (1958). Pre-albumin: A thyroxine-binding protein of human plasma. *Endocrinology* 63, 256–259. doi:10.1210/endo-63-2-256
- Johnson, S. M., Connelly, S., Fearn, C., Powers, E. T., and Kelly, J. W. (2012). The transthyretin amyloidoses: From delineating the molecular mechanism of aggregation linked to pathology to a regulatory-agency-approved drug. *J. Mol. Biol.* 421, 185–203. doi:10.1016/j.jmb.2011.12.060
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Kelly, J. W. (1996). Alternative conformations of amyloidogenic proteins govern their behavior. *Curr. Opin. Struct. Biol.* 6, 11–17. doi:10.1016/s0959-440x(96)80089-3
- Kim, J. H., Oroz, J., and Zweckstetter, M. (2016). Structure of monomeric transthyretin carrying the clinically important T119M mutation. *Angew. Chem. Int. Ed.* 55, 16168–16171. doi:10.1002/anie.201608516
- Klimtchuk, E. S., Prokava, T., Frame, N. M., Abdullahi, H. A., Spencer, B., Dasari, S., et al. (2018). Unusual duplication mutation in a surface loop of human transthyretin leads to an aggressive drug-resistant amyloid disease. *Proc. Natl. Acad. Sci. USA* 115, E6428–E6436. doi:10.1073/pnas.1802977115
- Kosol, S., Contreras-Martos, S., Cedeño, C., and Tompa, P. (2013). Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* 18, 10802–10828. doi:10.3390/molecules180910802
- Leonard, C., Phillips, C., and McCarty, J. (2021). Insight into seeded tau fibril growth from Molecular Dynamics simulation of the Alzheimer's disease protofibril core. *Front. Mol. Biosciences* 8, 109. doi:10.3389/fmolb.2021.624302
- Leri, M., Rebuzzini, P., Caselli, A., Luti, S., Natalello, A., Giorgetti, S., et al. (2020). S-homocysteinylation effects on transthyretin: Worsening of cardiomyopathy onset. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1864, 129453. doi:10.1016/j.bbagen.2019.129453
- Li, J., Bennett, K. C., Liu, Y., Martin, M. V., and Head-Gordon, T. (2020). Accurate prediction of chemical shifts for aqueous protein structure on “Real World” data. *Chem. Sci.* 11, 3180–3191. doi:10.1039/c9sc06561j
- Lim, K. H., Dasari, A. K. R., Hung, I., Gan, Z., Kelly, J. W., and Wemmer, D. E. (2016a). Structural changes associated with transthyretin misfolding and amyloid formation revealed by solution and solid-state NMR. *Biochemistry* 55, 1941–1944. doi:10.1021/acs.biochem.6b00164
- Lim, K. H., Dasari, A. K. R., Hung, I., Gan, Z., Kelly, J. W., Wright, P. E., et al. (2016b). Solid-State NMR Studies Reveal Native-like  $\beta$ -Sheet Structures in Transthyretin Amyloid. *Biochemistry* 55, 5272–5278. doi:10.1021/acs.biochem.6b00649
- Lim, K. H., Dyson, H. J., Kelly, J. W., and Wright, P. E. (2013). Localized structural fluctuations promote amyloidogenic conformations in transthyretin. *J. Mol. Biol.* 425, 977–988. doi:10.1016/j.jmb.2013.01.008
- Lin, Y., Sahoo, B. R., Ozawa, D., Kinoshita, M., Kang, J., Lim, M. H., et al. (2019). Diverse Structural Conversion and Aggregation Pathways of Alzheimer's Amyloid- $\beta$  (1–40). *ACS Nano* 13, 8766–8783. doi:10.1021/acsnano.9b01578
- Lincoff, J., Haghighatdari, M., Krzeminski, M., Teixeira, J. M. C., Gomes, G.-N. W., Gradinaru, C. C., et al. (2020). Extended experimental inferential structure determination method in determining the structural ensembles of disordered protein states. *Commun. Chem.* 3, 74. doi:10.1038/s42004-020-0323-0
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–1958. doi:10.1002/prot.22711
- Mangione, P. P., Verona, G., Corazza, A., Marcoux, J., Canetti, D., Giorgetti, S., et al. (2018). Plasminogen activation triggers transthyretin amyloidogenesis *in vitro*. *J. Biol. Chem.* 293, 14192–14199. doi:10.1074/jbc.ra118.003990
- Meli, M., Morra, G., and Colombo, G. (2008). Investigating the mechanism of peptide aggregation: Insights from mixed Monte Carlo-Molecular Dynamics simulations. *Biophysical J.* 94, 4414–4426. doi:10.1529/biophysj.107.121061
- Meng, F., Bellaiche, M. M. J., Kim, J.-Y., Zerze, G. H., Best, R. B., and Chung, H. S. (2018). Highly Disordered Amyloid- $\beta$  Monomer Probed by Single-Molecule

- FRET and MD Simulation. *Biophysical J.* 114, 870–884. doi:10.1016/j.bpj.2017.12.025
- Micsonai, A., Wien, F., Kerna, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., et al. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA* 112, E3095–E3103. doi:10.1073/pnas.1500851112
- Miyamoto, S., and Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962. doi:10.1002/jcc.540130805
- Mukrasch, M. D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., et al. (2009). Structural polymorphism of 441-residue Tau at single residue resolution. *Plos Biol.* 7, e34. doi:10.1371/journal.pbio.1000034
- Nerli, S., Mcshan, A. C., and Sgourakis, N. G. (2018). Chemical shift-based methods in NMR structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* 106–107, 1–25. doi:10.1016/j.pnmrs.2018.03.002
- Onufriev, A., Case, D. A., and Bashford, D. (2002). Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.* 23, 1297–1304. doi:10.1002/jcc.10126
- Oroz, J., Kim, J. H., Chang, B. J., and Zweckstetter, M. (2017). Mechanistic basis for the recognition of a misfolded protein by the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* 24, 407–413. doi:10.1038/nsmb.3380
- Otaki, H., Taguchi, Y., and Nishida, N. (2018). *Molecular dynamics simulation reveals that switchable combinations of  $\beta$ -sheets underlie the prion-like properties of  $\alpha$ -synuclein amyloids*. US: bioRxiv, 326462.
- Patriksson, A., and Van Der Spoel, D. (2008). A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* 10, 2073–2077. doi:10.1039/b716554d
- Peterle, D., Pontarollo, G., Spada, S., Brun, P., Palazzi, L., Sokolov, A. V., et al. (2020). A serine protease secreted from *Bacillus subtilis* cleaves human plasma transthyretin to generate an amyloidogenic fragment. *Commun. Biol.* 3, 764. doi:10.1038/s42003-020-01493-0
- Poltash, M. L., Shirzadeh, M., McCabe, J. W., Moghadamchargari, Z., Laganowsky, A., and Russell, D. H. (2019). New insights into the metal-induced oxidative degradation pathways of transthyretin. *Chem. Commun.* 55, 4091–4094. doi:10.1039/c9cc00682f
- Prabakaran, R., Rawat, P., Thangakani, A. M., Kumar, S., and Gromiha, M. M. (2021). Protein aggregation: In silico algorithms and applications. *Biophys. Rev.* 13, 71–89. doi:10.1007/s12551-021-00778-w
- Prompers, J. J., and Brüschweiler, R. (2002). General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc.* 124, 4522–4534. doi:10.1021/ja012750u
- Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of Molecular Dynamics trajectory data. *J. Chem. Theor. Comput.* 9, 3084–3095. doi:10.1021/ct400341p
- Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular Dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi:10.1016/0021-9991(77)90098-5
- Saravanan, K. M., Zhang, H., Zhang, H., Xi, W., and Wei, Y. (2020). On the Conformational Dynamics of  $\beta$ -Amyloid Forming Peptides: A Computational Perspective. *Front. Bioeng. Biotechnol.* 8, 532. doi:10.3389/fbioe.2020.00532
- Schmidt, M., Wiese, S., Adak, V., Engler, J., Agarwal, S., Fritz, G., et al. (2019). Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. *Nat. Commun.* 10, 5008. doi:10.1038/s41467-019-13038-z
- Schreiber, G., Aldred, A. R., Jaworowski, A., Nilsson, C., Achen, M. G., and Segal, M. B. (1990). Thyroxine transport from blood to brain via transthyretin synthesis in choroid plexus. *Am. J. Physiology-Regulatory, Integr. Comp. Physiol.* 258, R338–R345. doi:10.1152/ajpregu.1990.258.2.r338
- Shen, Y., and Bax, A. (2013). Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* 56, 227–241. doi:10.1007/s10858-013-9741-y
- Shen, Y., and Bax, A. (2010). SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* 48, 13–22. doi:10.1007/s10858-010-9433-9
- Shimomura, T., Nishijima, K., and Kikuchi, T. (2019). A new technique for predicting intrinsically disordered regions based on average distance map constructed with inter-residue average distance statistics. *BMC Struct. Biol.* 19, 3. doi:10.1186/s12900-019-0101-3
- Shrestha, U. R., Juneja, P., Zhang, Q., Gurumoorthy, V., Borreguero, J. M., Urban, V., et al. (2019). Generation of the configurational ensemble of an intrinsically disordered protein from unbiased molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA* 116, 20446–20452. doi:10.1073/pnas.1907251116
- Spagnoli, G., Rigoli, M., Novi Inverardi, G., Codeseira, Y. B., Biasini, E., and Requena, J. R. (2020). Modeling PrPSc generation through deformed templating. *Front. Bioeng. Biotechnol.* 8, 590501. doi:10.3389/fbioe.2020.590501
- Stabilini, R., Vergani, C., Agostoni, A., and Agostoni, R. P. V. (1968). Influence of age and sex on prealbumin levels. *Clinica Chim. Acta* 20, 358–359. doi:10.1016/0009-8981(68)90173-3
- Steward, R. E., Armen, R. S., and Daggett, V. (2008). Different disease-causing mutations in transthyretin trigger the same conformational conversion. *Protein Eng. Des. Selection* 21, 187–195. doi:10.1093/protein/gzm086
- Strodel, B. (2021). Energy landscapes of protein aggregation and conformation switching in intrinsically disordered proteins. *J. Mol. Biol.* 167182. doi:10.1016/j.jmb.2021.167182
- Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314, 141–151. doi:10.1016/s0009-2614(99)01123-9
- Sun, X., Jaeger, M., Kelly, J. W., Dyson, H. J., Wright, P. E., Kelly, W., et al. (2018). Mispacking of the Phe87 Side Chain Reduces the Kinetic Stability of Human Transthyretin. *Biochemistry* 57, 6919–6922. doi:10.1021/acs.biochem.8b01046
- Tran, L., and Ha-Duong, T. (2015). Exploring the Alzheimer amyloid- $\beta$  peptide conformational ensemble: A review of molecular dynamics approaches. *Peptides* 69, 86–91. doi:10.1016/j.peptides.2015.04.009
- Ulmer, T. S., Bax, A., Cole, N. B., and Nussbaum, R. L. (2005). Structure and Dynamics of Micelle-bound Human  $\alpha$ -Synuclein. *J. Biol. Chem.* 280, 9595–9603. doi:10.1074/jbc.m411805200
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi:10.1038/s41592-019-0686-2
- Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* 54, 5–375.6.37. doi:10.1002/cpbi.3
- Westermarck, P., Sletten, K., Johansson, B., and Cornwell, G. G. (1990). Fibril in senile systemic amyloidosis is derived from normal transthyretin. *Proc. Natl. Acad. Sci.* 87, 2843–2845. doi:10.1073/pnas.87.7.2843
- Yang, W., Kim, B. S., Lin, Y., Ito, D., Kim, J. H., Lee, Y.-H., et al. (2021). Exploring ensemble structures of Alzheimer's amyloid  $\beta$  (1–42) monomer using linear regression for the MD simulation and NMR chemical shift. *bioRxiv* 2008, 2023.
- Yu, W., Lee, W., Lee, W., Kim, S., and Chang, I. (2011). Uncovering symmetry-breaking vector and reliability order for assigning secondary structures of proteins from atomic NMR chemical shifts in amino acids. *J. Biomol. NMR* 51, 411–424. doi:10.1007/s10858-011-9579-0
- Zhou, S., Cheng, J., Yang, T., Ma, M., Zhang, W., Yuan, S., et al. (2019). Exploration of the misfolding mechanism of transthyretin monomer: Insights from hybrid-resolution simulations and Markov state model analysis. *Biomolecules* 9, 889. doi:10.3390/biom9120889

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yang, Kim, Muniyappan, Lee, Kim and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.