Wide-Line NMR Melting Diagrams, Their Thermodynamic Interpretation, and Secondary Structure Predictions for A30P and E46K α-Synuclein

Mónika Bokor,* Eszter Házy, and Ágnes Tantos

ABSTRACT: Parkinson’s disease is thought to be caused by aggregation of the intrinsically disordered protein, α-synuclein. Two amyloidogenic variants, A30P, and E46K familial mutants were investigated by wide-line 1H NMR spectrometry as a completion of our earlier work on wild-type and A53T α-synuclein (Bokor, M. et al. WT and A53T α-synuclein systems: melting diagram and its new interpretation. Int. J. Mol. Sci. 2020, 21, 3997.). A monolayer of mobile water molecules hydrates A30P α-synuclein at the lowest potential barriers (temperatures), while E46K α-synuclein has here third as much mobile hydration, insufficient for functionality. According to wide-line 1H NMR results and secondary structure predictions, E46K α-synuclein is more compact than the A30P variant and they are more compact than the wild type (WT) and A53T variants. Linear hydration is potential barrier sections of A30P α-synuclein shows one and E46K shows two slopes. The different slopes of the latter between potential barriers E_{α1} and E_{α2} reflect a change in water–protein interactions. The 31−32% homogeneous potential barrier distribution of the protein–water bonds refers to a non-negligible amount of secondary structures in all four α-synuclein variants. The secondary structures detected by wide-line 1H NMR are solvent-exposed α-helices, which are predicted by secondary structure models. β-sheets are only minor components of the protein structures as three- and eight-state predicted secondary structures are dominated by α-helices and coils.

1. INTRODUCTION
The intrinsically disordered proteins (IDPs) are unfolded under physiological conditions. While this feature is important for many physiological functions,1 disordered sequences are often prone to aggregation and fibril formation with detrimental consequences.2 Several neurodegenerative diseases are caused by the abnormal oligomerization and polymerization of proteins with disordered regions, including Alzheimer’s disease and Parkinson’s disease (PD). The exact molecular event that induces pathological aggregation of these proteins is still waiting to be elucidated,3 but there are certain verified mutations that render disordered proteins more prone to aggregation than their wild-type counterparts. α-Synuclein (α-S) is an IDP under normal physiological conditions (it is the general consensus),4 adopting random coil conformation.5−8 Notwithstanding, there exist variations of wild-type (WT) α-S structural propensity (globular or extended). The suppositions about secondary and tertiary structures of α-S are highly controversial, including conflicting views in the literature.9 The rapid interconversion between conformers impacting α-S and the different experimental methodologies used can be the causes of the controversies. α-S is coded by the SNCA gene, with several point mutations in the gene known to cause familial forms of PD, including A30P, A53T, and E46K mutations.10 A30P mutation retards9 the formation of both oligomers and fibrils and only this mutation affects the overall α-S structure.7 The E46K mutant increases membrane affinity10 and accelerates α-S aggregation and fibril formation.11,12 The A53T mutant considerably accelerates α-S aggregation, and its fibril formation is faster than that of the wild-type (WT) α-S.13−15

Different research groups came to different conclusions about α-S monomers having a globular-like structure or being extended random coils.7 Earlier, the WT and A53T α-S variants were investigated in monomer, oligomer, and amyloid forms.16 The monomers proved to be IDPs and more compact than random coils; about 32(3)% of their solvent-accessible
surface is determined by the secondary structure. They are already functional at the lowest potential barriers with mobile hydration water: a monolayer of mobile hydration water is surrounding them. They realize all possible hydrogen bonds with the solvent water.

Wide-line $^1$H NMR is an accepted method that can provide information on the location and structural environment of hydrogen atoms in proteins as it enables the direct observation of translational and rotational movements of molecules in the condensed phase.

NMR characteristics of aqueous solutions rapidly frozen and then slowly thawed through equilibrium thermal states provide direct information on the immobile and partially or fully mobile parts of the molecules, yielding invaluable insight into the overall structure of the proteins.

Here, our previous work on WT and A53T $\alpha$-S was completed with new results on two other familial mutants, A30P and E46K $\alpha$-S. We measured the melting diagrams (MDs, relative ratios of mobile water as a function of normalized functional temperature; see the Supporting Information in ref 17) by wide-line NMR, i.e., the amount of mobile hydration water as a function of temperature, to get information on the steps and gradience in the development of full hydration of $\alpha$-S. Secondary structure (SS) predictions on the same proteins were calculated and compared to the experimental results by wide-line NMR. These measurements were supplemented with secondary structure predictions.

2. MATERIALS AND METHODS

Expression and purification of recombinant human A30P and E46K mutant variants of $\alpha$-S in a pRK-172-based expression system were performed as described. Briefly, expression of the proteins was performed in Escherichia coli BL21(DE3) in a pT7-7-based expression system, after IPTG induction. Bacterial cell pellets were harvested by centrifugation and resuspended in 10 mM Tris–HCl, pH 8.0, 1 mM EDTA, and 1 mM Complete protease inhibitor cocktail. After cell lysis, streptomycin sulfate-precipitated DNA was removed by centrifugation and an ammonium sulfate precipitation step was performed to selectively precipitate the $\alpha$-S protein. After centrifugation at 13,500g for 30 min at 4 °C, the pellet was dissolved in 10 mM Tris–HCl, pH 7.4, and 1 mM Complete, and filtered using a 0.2 μm mesh. The resulting solution was loaded onto a Resource Q anion exchange column on an Äkta Explorer chromatography system (GE Healthcare). The purity and integrity of the purified proteins were confirmed by SDS-PAGE (for representative gel pictures, see Supporting Figure 1). Peak fractions were collected and dialyzed against double-distilled water before lyophilization.

In sample preparation, the mass of lyophilized protein (without any further refinement) was measured, and an appropriate amount of double-distilled water was added to obtain the nominal concentration of 50 mg/mL. Oligomers formed during the process were removed by filtering the solution through a 100 kDa membrane. All measurements were carried out on three identical samples prepared independently.

The wide-line NMR approach we applied is detailed in ref 19. The beginning of the movement (rotation) of water molecules bound to the surface of the protein, the process that is considered melting, is followed by observing motional narrowing in wide-line $^1$H NMR spectroscopy. The motional narrowing is a useful criterion for mobility since dynamics of hydration water happens on a picosecond time scale and NMR is slower by an order of magnitude. Fundamental temperature is $T_i = R \cdot T$ and the normalized fundamental temperature scale is $T_i = R \cdot T/((R T/273.15) + 1)$. The events of the beginning of molecular motion can be characterized on an energy scale by the application of $T_{fi}$.

The number of water molecules in the first hydrate shell of the protein is $n_w$, and the total number of water molecules in the entire heterogeneous hydration region is $n_{w_0}$. Applying these definitions, the total number of hydrating water molecules at 0 °C or $T_{fi} = 1$ is $(n_{w_0} + n_w)$. The number of mobile water molecules per amino acid residue is indicated with $n_{aa}$. It can be calculated from the measured fraction of mobile protons in water $n$ as $n_{aa} = (n/2) \cdot (M_i/(M_{i}/M_0))$, where $a$ is the number of amino acid residues in the protein and $M_i(i)$ is the relative formula mass of compound $i$.

The amount of mobile hydration can be given also by the common measure of hydration (not the same as the term mobile hydration) as the mass of the solvent water divided by the mass of the solute protein, $h = m_{water}/m_{protein}$. In our measurements, it is given by the measured (by wide-line $^1$H NMR) fraction of mobile water $n$ multiplied by the mass of water and divided by the mass of protein, $h = n_{water}/m_{protein}$. Hydration and $n_{aa}$ can be interrelated as $n_{aa} = \sum \frac{m_{H_2O}}{m_{protein}} \cdot M_i(i)$. The hydration can be formally described as a series expansion$^{20,21}$

$$T_{fi} = A + B(T_{fin} - T_{fi}) + C(T_{fin} - T_{fin})^2 + ... \tag{1}$$

where the summation is carried out up to the quadratic term. The cubic term, which was applied in ref 16, was unnecessary. $T_{fi}$ is the lowest temperature where mobile water molecules are detected and $E_{a,0}(T_{fin})$ is the lowest potential barrier with mobile water molecules at the solvent-accessible surface. The parameter $T_{fi}$ gives the temperature where the thermal trend of the MD switches between constant and linearly increasing. Likewise, the trend changes from linear to quadratic at $T_{fi}$. The protein preparation and wide-line NMR measurements were described in former publications (ref 16 and the Supporting Information in ref 17). The applied three-state SS prediction methods are BREwy, Jpred4, Porter S.0, PSIPRED, PSSRX, RaptorX, SCRATCH, and SPIDER3. Further, eight-state predictions were made with the following methods only: BREwy, Porter S.0, RaptorX, SCRATCH, and SPIDER3. Secondary structures made of a minimum of four amino acids were selected. The percentage of secondary structure predictions was calculated as $\sum_{i=1}^{R} \frac{M_i(i)}{M_{i}/M_0}$, where the summation is carried out up to the quadratic term. The cubic term, which was applied in ref 16, was unnecessary. $T_{fi}$ is the lowest temperature where mobile water molecules are detected and $E_{a,0}(T_{fin})$ is the lowest potential barrier with mobile water molecules at the solvent-accessible surface. The parameter $T_{fi}$ gives the temperature where the thermal trend of the MD switches between constant and linearly increasing. Likewise, the trend changes from linear to quadratic at $T_{fi}$.

Protein preparation and wide-line NMR measurements were described in former publications (ref 16 and the Supporting Information in ref 17). The applied three-state SS prediction methods are BREwy, Jpred4, Porter S.0, PSIPRED, PSSRX, RaptorX, SCRATCH, and SPIDER3. Further, eight-state predictions were made with the following methods only: BREwy, Porter S.0, RaptorX, SCRATCH, and SPIDER3. Secondary structures made of a minimum of four amino acids consecutively are considered only.

BREwy$^{22}$ (http://distilldeep.ucd.ie/brewery/) is the state-of-the-art predictor of protein structural annotations (secondary structure in three and eight classes). BREwy is based on ensembles of cascaded BRNNs (bidirectional recurrent neural networks) and Convolutional Neural Networks.

Jpred is a protein secondary structure prediction server and has been in operation since approximately 1998. Jpred incorporates the Jnet algorithm to make predictions that are more accurate. Jpred4$^{23}$ (http://www.compbio.dundee.ac.uk/jpred4/index.html) is its the current version.

Porter S.0 and Porter8 S.0$^{23}$ (http://distilldeep.ucd.ie/porter/) are servers for protein secondary structure prediction in three and eight classes based on ensembles of cascaded BRNNs and Convolutional Neural Networks.

PSIPRED$^{24}$ (http://bioinf.cs.ucl.ac.uk/psipred/) is a simple and accurate secondary structure prediction method, incorpo-
rating two feed-forward neural networks, which perform an analysis on output obtained from PSI-BLAST (Position-Specific Iterated—BLAST).

PSRSM27 (http://qilubio.qlu.edu.cn:82/protein_PSRSM/default.aspx) uses methods based on data partitioning and the semirandom subspace method.

RaptorX Property28 (http://raptorx.uchicago.edu/) is a web server that predicts the structural properties of a protein sequence without using any templates. This server employs a powerful in-house deep-learning model, DeepCNF (Deep Convolutional Neural Fields), to predict the SS.

SCRATCH39 (http://scratch.proteomics.ics.uci.edu/) uses ensembles of bidirectional recurrent neural network architectures, PSI-BLAST-derived profiles, and a large nonredundant training set to derive two new predictors: (a) the second version of the SSpro program for secondary structure classification into three categories and (b) the first version of the SSpro8 program for secondary structure classification into the eight classes produced by the DSSP program.

IUPred2A31-33 was used, which is a combined web interface that allows one to identify disordered protein regions using IUPred2 and disordered binding regions using ANCHOR2. Both IUPred2 and ANCHOR2 indicate a disordered region and a disordered binding site, respectively, with scores above 0.5.

3. RESULTS

3.1. Direct Observations of Structural States through the Melting Diagrams. In an attempt to decipher the molecular background of the differences in the behavior of the two different α-S mutant variants, we conducted wide-line NMR measurements under identical conditions. In wide-line NMR, IDPs and globular proteins are easily distinguishable based on their MDs. While globular proteins show a plateau of mobile hydration water throughout a relatively wide temperature range, disordered proteins are characterized by a constant growth in mobile water with a constant level of mobile hydration water throughout a relatively wide temperature range, disordered proteins are easily distinguishable based on their MDs. While globular proteins show a plateau of mobile hydration water at relatively high temperature and mobile hydration water appears; disordered proteins show the characteristic of IDPs. They show the functional normalized temperature or potential barrier (Δn/ΔTn) values of globular proteins in the Supporting Information of ref 17), and they have an intensely characteristic of IDPs. They show the functional normalized temperature or potential barrier (Δn/ΔTn) values of globular proteins in the Supporting Information of ref 17). They have an intensely

Table 1. Parameter Values for the Polynomial Relation (eq 1 Describes the Mobile Water Fraction, n)a

| α-S variant | A30P | E46K |
|-------------|------|------|
| A = n(T_\text{fl}) = n(T_\text{fl})^b | 0.0156(2) | 0.0058(1) |
| n_{\text{aa}}(E_{\alpha,0}) = n_{\text{aa}}(E_{\alpha,0})^b | 0.90(1) | 0.335(6) |
| h(E_{\alpha,0}) = h(E_{\alpha,0})^b | 0.313(4) | 0.117(2) |
| B = 0.24(1) | 0.07(1), 0.29(2) |
| C = 5(1)-10^7 | 36(5) |
| E_{\alpha,0}/kJ mol^{-1} | 4.95(2) | 4.97(2) |
| T_\text{fl}/°C | -48.2(8) | -47.1(8) |
| E_{\text{fl}}(T_\text{eff})/(kJ mol^{-1}) | 0.879(2) | 0.875(4), 0.856(2) |
| E_{\text{fl}}(T_\text{eff})/(kJ mol^{-1}) | 5.28(1) | 5.26(2), 5.15(1) |
| T_\text{fl}/°C | -33.0(7) | -34(1), -39(6) |
| E_{\text{fl}}/kJ mol^{-1} | 0.959(4) | 0.998(2) |
| E_{\text{fl}}/(kJ mol^{-1}) | 5.76(2) | 5.76(1) |
| E_{\text{fl}}/kJ mol^{-1} | -11(1) | -11.2(5) |
| n(E_{\alpha,0}) = n(E_{\alpha,0})^b | 0.0345(2) | 0.030(2) |
| n_{\text{aa}}(E_{\alpha,0}) | 1.98(1) | 1.7(1) |
| h(E_{\alpha,0}) | 0.689(4) | 0.60(3) |
| n(T_\text{fl} = 1) | 0.13(2) | 0.102(2) |
| n_{\text{aa}}(E_{\alpha,0}) = 6.01 kJ mol^{-1})^c | 7(1) | 5.9(9) |
| h(E_{\alpha,0}) = 6.01 kJ mol^{-1})^c | 2.6(4) | 2.6(3) |

*a Error in the last digit is given in parentheses. n(i) values are given for a 50 mg/mL protein concentration. b, c T_\text{fl} in the case of E46K. Below and above T_\text{fl} = 0.903, respectively.

calculated from the MDs against functional temperature (Δn/ΔTn). The trend changes are more striking on the DMDs.

There is a short constant amount of mobile hydration water at relatively high temperature/fun- ional normalized temperature or potential barrier (Δn/ΔTn) (T_\text{fl}) or E_{\alpha,0} (Table 1) compared to globular proteins (see the Supporting Information of ref 17), and they have an intensely elevating section from E_{t,1} on, in contrast to the constant number of mobile water molecules per amino acid residue, n_{\text{aa}} (see Materials and Methods) values of globular proteins in the same temperature range.

If we want to understand the differences in the behavior of the two mutant proteins, the best way is to create a differential melting diagram (DMD) by plotting the differential values calculated from the MDs against functional temperature (Δn/ΔTn). The trend changes are more striking on the DMDs.

There is a short constant amount of mobile hydration water at relatively high temperature/fun- ional normalized temperature or potential barrier (Δn/ΔTn) (T_\text{fl}) or E_{\alpha,0} (Table 1) compared to globular proteins (see the Supporting Information of ref 17), and they have an intensely elevating section from E_{t,1} on, in contrast to the constant number of mobile water molecules per amino acid residue, n_{\text{aa}} (see Materials and Methods) values of globular proteins in the same temperature range.
mobile hydration increases linearly (constant $\Delta n/\Delta T_{m}$) to reach $277(1)$ H$_2$O/protein at $E_{a,2}$ (Table 1). The increase becomes quadratic (linear $\Delta n/\Delta T_{m}$) at $E_{a,2}$.

E46K $\alpha$-S differs from A30P $\alpha$-S mostly in the low potential barrier region (Figures 1 and 2) and behaves unexpectedly at low potential barrier values. The first mobile hydration water molecules for E46K $\alpha$-S were detected at a low potential barrier value ($E_{1} = 4.71(1)$ kJ mol$^{-1}$), that is, at a low temperature value compared to the other $\alpha$-S variants (Figure 1 and ref 16). This potential barrier is as low as in the case of globular proteins (inserted graph in Figure 1).$^{19,20,36}$ The amount of mobile hydration water molecules increases initially, right after the appearance of the first amount of mobile water and before the constant section between $E_{a,0} = 4.97$ kJ mol$^{-1}$ and $E_{a,1} = 5.26(2)$ kJ mol$^{-1}$. The gradual growth of the mobile hydration below $E_{a,0}$ indicates the lack of a step-like change in the potential barriers at $E_{a,0}$ and shows their broad distribution there. The solvent-accessible surface of E46K $\alpha$-S is very heterogeneous regarding protein–water interactions in this potential barrier section. The plateau (constant $naa$ vs $T_{m}$) of E46K has a mobile hydration value of $naa = 0.335(6)$ or $h = 0.117(2)$, which means $46.9(9)$ H$_2$O/protein, and ends at $E_{a,3} = 5.26(2)$ kJ mol$^{-1}$ (Table 1). There is a step, a jump in the magnitude of $naa$, in the MD of E46K at $E_{a,3}$ (Figure 1), which corresponds to a spike in the differential form of the MD (DMD, Figure 2). This is followed by a linearly growing section of MD, but with two different regions with different slopes ($\beta$ parameters, Table 1). This corresponds to constant sections with different magnitudes in the DMD (Figure 2).

The slope of the MD changes at $E_{s} = 5.42$ kJ mol$^{-1}$. The slope change refers to a change in the interaction between water and E46K $\alpha$-S at the slope change, i.e. at $E_{s} = 5.42(6)$ kJ mol$^{-1}$. Different types of water–protein interactions are active below and above the slope change. The hydration at the change is $naa = 0.8(1)$ or $h = 0.27(5)$ or $2.2(4)\times10^{-2}$ H$_2$O/protein. This value is approximately equal to the hydration in the first hydration layer of a protein$^{38,39}$ and the difference between it and $h(E_{a,2})$ is $h = 0.33(8)$ or $2.7(6)\times10^{-2}$ H$_2$O/protein. That is, new types of water molecules in their interactions with proteins become mobile at the change additional to the first hydration layer, which becomes built up until the change.

The two mutant $\alpha$-S variants have nearly parallel MDs at $E_{a,1} \leq E_{s} \leq 6.01$ kJ mol$^{-1}$ with the A30P $\alpha$S having higher mobile hydration (Figure 1). The hydration of A30P $\alpha$-S is greater by $naa = 0.35(2)$ on average than the hydration of E46K $\alpha$-S between $E_{a,1}$ and $E_{s} = 6.01$ kJ mol$^{-1}$.

The $h_{ba}$ and $h_{ha}$ values, i.e., the homogeneously and the heterogeneously bound mobile hydration water amounts for A30P and E46K $\alpha$-S (Table 2), are markedly lower than in WT and A53T variants ($h_{ba} = 0.44(8)$ and $h_{ha} = 2.8(2)$, on the average).$^{16}$ The A30P and E46K $\alpha$-S mutants reach their highest hydration level at the melting point of bulk water with an average of $h = 2.2(3)$, which corresponds to $9(1)\times10^{-2}$ H$_2$O/protein. The WT and A53T $\alpha$-S variants have higher hydration at this point ($h = 3.31(7)$ or $1.32(2)\times10^{-1}$ H$_2$O/protein on the average).$^{16}$

Protein molecules can be characterized and categorized by the homogeneity/heterogeneity of the energy distribution of water binding. This ratio is measured and the defined relation is $HeR = (1 - T_{ba0})/(1 - T_{ha0})$, in which $(1 - T_{ba0})$ and $(1 - T_{ha0})$ give the measured distances from the melting point of ice. The observed heterogeneity ratio for A30P and E46K is $HeR = 0.69(2)$ (Table 2), which corresponds to $31(2)$% homogeneous potential barrier distribution. This distribution width is equal within experimental error with that of WT and A53T $\alpha$-S, being $33(4)$%.$^{16}$ These homogeneity ratios agree with $\alpha$-S being more compact than it is expected for a random coil state,$^{40,41}$ i.e., these proteins have a non-negligible extent of

![Figure 2. Differential form of the melting diagram of A30P (red) and E46K (blue) $\alpha$-synuclein. The differentials ($\Delta n/\Delta T_{m}$) were calculated from the fitted curves shown in Figure 1.](Image)

### Table 2. Dynamic Parameters from the Polynomial Relation Describing the Melting Diagrams

| $\alpha$-S variant | A30P | E46K | A53T | WT |
|-------------------|------|------|------|----|
| $n_{ha}$          | 0.0156(2) | 0.0058(1) | 0.22(4) | 0.22(4) |
| $n_{ba}$          | 0.90(1)   | 0.335(6) | 2.5(5) | 2.5(5) |
| $h_{a}$           | 0.313(4)  | 0.117(2) | 0.44(8) | 0.44(8) |
| $n_{ha}$          | 0.11(2)   | 0.10(2) | 0.18(6) | 0.142(9) |
| $n_{ba}$          | 7(1)      | 5.5(9) | 23(6) | 16(1) |
| $h_{ba}$          | 2.3(4)    | 1.9(3) | 4(1) | 2.8(2) |
| $HeR = (1 - T_{ba0})/(1 - T_{ha0})$ | 0.68(2) | 0.72(3) | 0.65(4) | 0.70(4) |
| $HeR_{a} = n_{ba}/n_{ha}$ | 0.91(1) | 0.92(1) | 0.89(3) | 0.87(3) |
| $HeM = (B + 2C)/(1 - T_{ha0})$ | 8(1)-10$^{-2}$ | 6(1)-10$^{-2}$ | 1.30(6)-10$^{-1}$ | 9.85(5)-10$^{-1}$ |

*For detailed definitions of the parameters, see ref 16. $n_{ha} = A$ is the mobile water fraction bound homogeneously and $n_{ba} = n(T_{m} = 1) - n_{ha}$ is the mobile water fraction bound heterogeneously. The error in the last digit is given in parenthesis.*
secondary structures. The ratio of the amount of heterogeneously bound water to the total number of bound water (heterogeneously plus homogeneously bound) is $HeR_a = \frac{n_{naa}}{(n_{naa} + n_{hoa})}$ for both A30P and E46K variants, approximately as high as $HeR_a = 0.88(1)$ for WT and A53T variants. The $HeR_a$ values indicate high heterogeneity of the bonds formed by the $\alpha$-Ss. The measure of heterogeneity, $HeM = \frac{(B + 2C)/(1 - T_{min})}$, characterizes the degree of heterogeneity of the protein–water interactions close to 0 °C or $E_h = 6.01$ kJ mol$^{-1}$. The A30P and the E46K variants have the same $HeM$ values within experimental error, and these values are significantly smaller than those of the WT (9.81(5)° and the A53T (1.03(6)°) variants. The increase in the hydrations of mutants presented in this work is less intense than in the cases of WT and A53T $\alpha$-Ss, $HeM$(A53T) > $HeM$(WT) > $HeM$(A30P) $\approx$ $HeM$(E46K).

3.2. In Silico Analysis of Secondary Structures. To relate our data from the wide-line NMR results to the disorder and secondary structure content of the two protein variants, we used different in silico structure prediction methods.

Disorder content was predicted with IUPred2,$^{31,32}$ which results in a probability value for each residue of being part of a disordered region, with values above 0.5 indicating disorder tendency. The closer the value is to 1, the higher the probability is of a structurally disordered state at the given position. The average disorder probability scores for the whole length are 0.54(2) for A30P, 0.52(2) for E46K $\alpha$-S (Figure 3).

Since IUPred provides information on the disorder tendency of a protein but not on the structural propensities of the ordered regions, we also applied in silico algorithms to analyze the secondary structure content of the studied proteins.

In the SS predictions, the determinant motifs are coils, helices, and $\beta$-sheets according to three- and eight-state prediction methods. Protein secondary structures are traditionally characterized as three general states: helix (H), strand (E), and coil (C). From these general three states, the DSSP$^{36}$ program proposed a finer characterization of the secondary structures by extending the three states into eight states: helix (G), $\alpha$-helix (H), $\pi$-helix (I), $\beta$-strand (E), bridge (B), turn (T), bend (S), and others (L). These eight secondary structure states are often mapped into the following three states: $H$: $\alpha$-helix, which corresponds to the right- or left-handed cylindrical/helical conformations that include G, H, and I states; $E$: $\beta$-strand, which corresponds to pleated sheet structures that encompass E and B states; $C$: coil, which covers the remaining S, T, and L states. The state-of-the-art methods are currently reaching almost 88% for a three-class prediction and 76.5% for an eight-class prediction.$^{45}$

For the $\alpha$-S variants (Figure 4), the three- and eight-state methods indicate the coil and the helix to be the most dominant.

Three-state predictions use the average of the results of eight different SS prediction methods (Brewery, JPred4, Porter 5.0, PSIPRED, PSRSM, RaptoR, SCRATCH, SPIDER3). The three-state SS prediction of A30P, E46K, and WT $\alpha$-Ss shows that $\alpha$-helices and coils are the main SS elements. $\alpha$-helices form in 47–48% of the protein chain and coils give 39%, while $\beta$-sheets are only present in 9%. The $\beta$-sheet content is minimal in WT $\alpha$-S (4 amino acid residues long) but it is three times greater in the A30P (13 residues) and in the E46K (12 residues) mutants.

The eight-state SS prediction consists of the average of five modeling methods (Brewery, Porter 5.0, RaptoR, SCRATCH, SPIDER3). According to this prediction, A30P and E46K mutants contain two longer $\alpha$-helices (46 and 51% of the whole length, respectively), while the WT version has more pieces of shorter $\alpha$-helix type sections (43% altogether). The WT $\alpha$-S in summary shows a longer $\beta$-sheet SS (13%) than the two mutants (E46K 7%) and A30P has the shortest (5%) such section.

4. DISCUSSION

4.1. Structure of the $\alpha$-Synuclein Mutants. Approximately, one monolayer of mobile water molecules hydrates A30P $\alpha$-S at the lowest potential barriers ($h = 0.313(4)$). E46K has $h = 0.117(2)$ mobile hydration here, which is almost a third of the former and is not enough for the functionality of E46K $\alpha$-S since a protein needs at least $h = 0.2$ to be functional.$^{38,44}$ The lower initial and then overall hydration can be a result of E46K $\alpha$-S being more compact than A30P $\alpha$-S. It appears that every third or fourth hydration site of A30P $\alpha$-S would be occupied in E46K $\alpha$-S. The linear hydration section between $E_{\alpha_1}$ and $E_{\alpha_2}$ for A30P $\alpha$-S can be described with one slope, while the same section for E46K $\alpha$-S has two distinct sections with different slopes. The type of the protein–water interaction changes where the slope changes for E46K at $E_{\alpha} = 5.43(1)$, $naa$ or $h$ increases only slightly before the change; it is almost constant here, and after the change, $naa$ or $h$ increases even more rapidly than A30P $\alpha$-S. The hydration increases from $E_{\alpha_1}^{(id)}$ to $E_{\alpha_2}$ are $\Delta h^{A30P} = 0.48(3)$ and $\Delta h^{E46K} = 0.33(8)$. 

Figure 3. Prediction of protein disorder and disordered binding sites for A30P and E46K $\alpha$-synuclein by the IUPred2 (A30P cyan, E46K blue solid lines) and ANCHOR2 (A30P dark green, E46K green solid lines) programs. The results for WT $\alpha$-synuclein are also given for comparison (IUPred2 orange, ANCHOR2 red dashed lines). A score above 0.5 predicts protein disorder or disordered binding site.
This amount of increase suggests a newer layer of mobile hydration to build up.\(^{38}\) The rate of expansion of the hydration layers is the greatest at potential barriers greater than \(E_a\). The difference in the hydration of A30P and E46K \(\alpha\)-Ss in favor of A30P is indicative again of a more compact structure of E46K \(\alpha\)-S. The \(\alpha\)-S mutants A30P and E46K have markedly lower hydration \((h(E_a = 6.01 \text{ kJ mol}^{-1}))\), Table 1\) than the mutant A53T \((h^{\text{A53T}} = 4.0(2))\) and the WT \((h^{\text{WT}} = 3.3(2))\) \(\alpha\)-S at the melting point of bulk water \((E_a = 6.01 \text{ kJ mol}^{-1})\). It can be deduced from these data that E46K \(\alpha\)-S is the most compact structurally with the smallest solvent-accessible surface and A53T \(\alpha\)-S has the most open structure.

A recent cryoelectron microscopy study revealed that the E46K mutant \(\alpha\)-S forms structurally distinct, more compact amyloid fibrils than the wild type.\(^{45}\) The authors attributed this feature to a misfolding pathway of the mutant, where the salt bridge between E46 and K80 is disrupted by the electrostatic repulsion in the mutant fibril. This altered structural tendency might be reflected already in the monomer form, as indicated by our measurements.

The homogeneously and the heterogeneously bound mobile hydration water amounts for A30P and E46K \(\alpha\)-S \((h_{\text{ho}}\) and \(h_{\text{ho}}\) Table 2\) also show that the WT and A53T \(\alpha\)-Ss are more open structurally than the present mutants.

The heterogeneity ratio, \(H/e\), from the dynamic MD parameters shows non-negligible secondary structures in A30P and E46K \(\alpha\)-S by the 31(2) % homogeneous potential barrier distribution of the protein–water bonds.

### 4.2. Comparison of the Measured and Predicted Structures

\(\alpha\)-S has been mainly considered to contain \(\alpha\)-helicies with a small number of isolated \(\beta\)-sheets\(^{46-49}\) or to be an \(\alpha + \beta\) protein,\(^{50,51}\) but the prediction of an all-\(\beta\) structure with some peripheral small \(\alpha\)-helicies for \(\alpha\)-S is also a valid possibility.\(^{52}\)

Wide-line NMR measurements provide valuable information on the overall solvent accessibility and structural states of the proteins studied but no detailed structural information. Combining our measured data with structure predictions enables us to understand and interpret the structural and physiological consequences of the studied mutations.

IUPred2\(\alpha\)\(^{31,32}\) is a combined web interface that allows one to identify disordered protein regions using IUPred2 and disordered binding regions using ANCHOR2. The algorithm identifies disordered protein regions and it is found that A30P \(\alpha\)-S has longer such regions than E46K \(\alpha\)-S. Based on the IUPred results, the disorder tendencies of WT, A30P, and E46K \(\alpha\)-Ss are very close to each other, but A30P \(\alpha\)-S has a more open structure than E46K \(\alpha\)-S. ANCHOR2 predicts regions that undergo a disorder-to-order transition upon binding to another protein. Based on ANCHOR2 results, A30P and E46K, just as WT \(\alpha\)-S, have a disordered binding site spanning 30 residues at their C terminus and a transitional but binding region of 10 residues before it.

These results agree with the finding that A30P and E46K \(\alpha\)-Ss are also IDPs, as seen by wide-line \(^1H\) NMR. More precisely, 70(3) % of their solvent-accessible surface is heterogeneous/disordered (Table 2).

Even though the majority of the proteins appears to be in a solvent-accessible, disordered state, they also contain a significant amount of regions with secondary structures.

The three- and eight-state-predicted SSs of the here studied \(\alpha\)-S mutants are dominated by \(\alpha\)-helicies and coils and \(\beta\)-sheets are only minor components of the structures. From this, we can deduce that the secondary structures detected by wide-line NMR are solvent-exposed \(\alpha\)-helicies in these proteins. The predicted SSs for the three \(\alpha\)-S variants (WT, A30P, E46K) are very similar to each other, with the biggest difference being in the extent of \(\beta\)-sheets in the three-state predictions. This method predicts almost no \(\beta\)-sheets in the WT \(\alpha\)-S but a markedly increased \(\beta\)-sheet content for the two mutants (Figure 4A). On the contrary, the eight-state predictions forecast the largest \(\beta\)-sheet content for WT \(\alpha\)-S, which is in accordance with our earlier secondary structure calculations.\(^{17}\)

At the sites of the mutations, the predictions show no special features, while the disorder prediction clearly showed an increased disorder tendency of the A30P variant around the mutation site.

A comparison of these predictions with the results of the wide-line NMR reveals that the measured structural states are clearly different from the predicted ones. The mutations induce measurable changes in the secondary structure content of the protein, resulting in alterations of the overall structures in the case of the mutants. This observation also highlights the limits of structure prediction algorithms in detecting structural changes caused by single amino acid changes.
ASSOCIATED CONTENT

Accession Codes

WT α-S: P37840; A30P α-S: VAR_007957; E46K α-S: VAR_022703; A53T α-S: VAR_007454; BSA: P02769; β-casein: P02666; lysozyme: P86383.

AUTHOR INFORMATION

Corresponding Author

Mónika Bokor — Institute for Solid State Physics and Optics, Wigner Research Centre for Physics, 1121 Budapest, Hungary; orcid.org/0000-0001-8883-1316; Email: bokor.monika@wigner.hu

Authors

Eszter Házy — Institute of Enzymology, Research Centre for Natural Sciences, 1117 Budapest, Hungary
Ágnes Tantos — Institute of Enzymology, Research Centre for Natural Sciences, 1117 Budapest, Hungary

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c00477

Author Contributions

This manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors have no acknowledgment to declare.

ABBREVIATIONS

α-S, α-synuclein; BRNN, bidirectional recurrent neural network; DeepCNF, deep convolutional neural fields; DMD, differential melting diagram; DSSP, dictionary of secondary structure of proteins; HeM, measure of heterogeneity; HeR, heterogeneity ratio; HeRr, second ratio of heterogeneity; IDP, intrinsically disordered protein; MD, melting diagram; PSI-BLAST, position-specific iterated—BLAST; SS, secondary structure; PD, Parkinson’s disease; WT, wild type

REFERENCES

(1) Uversky, V. N. The mysterious unfoldomer: structureless, underappreciated, yet vital part of any given proteome. Proteomics 2010, No. 568068.
(2) Shigemitsu, Y.; Hiroaki, H. Common molecular pathogenesis of disease-related intrinsically disordered proteins revealed by NMR analysis. J. Biochem. 2018, 163, 11–18.
(3) Gómez-Benito, G. N.; Garcia-Sanz, P.; Michel, A.; Dumoulin, M. R.; et al. Modeling Parkinson’s disease with the α-synuclein protein. Front. Pharmacol. 2020, 11, No. 356.
(4) Fauvet, B.; Mbefo, M. K.; Fares, M.-B.; Desobry, C.; Michael, S.; Ardah, M. T.; Tsika, E.; Coune, P.; Prudent, M.; Lion, N.; Eliezer, D.; Moore, D. J.; Schneider, B.; Aebischer, P.; El-Agnaf, O. M.; Masliah, E.; Lashuel, H. A. α-Synuclein in central nervous system and from erythrocytes, mammalian cells, and Escherichia coli exists predominantly as disordered monomer. J. Biol. Chem. 2012, 287, 15345–15364.
(5) Wu, K.-P.; Kim, S.; Fela, D. A.; Baum, J. Characterization of conformational and dynamic properties of natively unfolded human and mouse α-synuclein ensembles by NMR: Implication for aggregation. J. Mol. Biol. 2008, 378, 1104–1115.
(6) Berrocal, R.; Vasquez, S. V.; Krs, R.; Gadad, B. S.; Rao, K. S. α-Synuclein misfolding versus aggregation relevance to Parkinson’s disease: Critical assessment and modeling. Mol. Neurobiol. 2015, 51, 1417–91431.
(7) Bisi, N.; Feni, L.; Paeqini, K.; Pérez-Peña, H.; Onger, S.; Pieraccini, S.; Pellegrino, S. α-Synuclein: An all-inclusive trip around its structure, influencing factors and applied techniques. Front. Chem. 2021, 9, No. 66685.
(8) Breydo, L.; Wu, J. W.; Uversky, V. N. α-Synuclein misfolding and Parkinson’s disease. Biochim. Biophys. Acta, Mol. Basis Dis. 2012, 1822, 261–285.
(9) Tosatto, L.; Horrocks, M. H.; Dear, A. J.; Knowles, T. P. J.; Dalla Serra, M.; Cremades, N.; Dobson, C. M.; Klenerman, D. Single-molecule FRET studies on α-synuclein oligomerization of Parkinson’s disease genetically related mutants. Sci. Rep. 2015, 5, No. 16696.
(10) Giannakos, E.; Pacifico, J.; Smith, D. P.; Hung, L. W.; Masters, C. L.; Cappai, R.; Wade, J. D.; Barnham, K. J. Dimeric structures of α-synuclein bind preferentially to lipid membranes. Biochim. Biophys. Acta, Biomembr. 2008, 1778, 1112–1119.
(11) Fredenburg, R. A.; Rospigliosi, C.; Meray, R. K.; Kessler, J. C.; Lashuel, H. A.; Eliezer, D.; Lansbury, P. T. The impact of the E46K mutation on the properties of α-synuclein in its monomeric and oligomeric states. Biochemistry 2007, 46, 7107–7118.
(12) Pandey, N.; Schmidt, R. E.; Galvin, J. E. The α-synuclein mutation E46K promotes aggregation in cultured cells. Exp. Neurol. 2006, 197, 515–520.
(13) Li, J.; Uversky, V. N.; Fink, A. L. Effect of familial Parkinson’s disease point mutations A30P and A53T on the structural properties, aggregation, and fibrillation of human α-synuclein. Biochemistry 2001, 40, 11604–11613.
(14) Lemkau, K. L.; Cornellas, G.; Lee, S. W.; Ricardson, L. K.; Woods, W. S.; George, J. M.; Rienstra, C. M. Site-specific perturbations of α-synuclein fibril structure by the Parkinson’s disease associated mutations A53T and E46K. PLoS One 2013, 8, No. e49750.
(15) Conway, K. A.; Harper, J. D.; Lansbury, P. T. Accelerated in vitro fibril formation by a mutant α-synuclein linked to early-onset Parkinson disease. Nat. Med. 1998, 4, 1318–1320.
(16) Bokor, M.; Tantos, A.; Tompa, P.; Han, K.-H.; Tompa, K. WT and A53T α-synuclein systems: melting diagram and its new interpretation. Int. J. Mol. Sci. 2020, 21, No. 3997.
(17) Bokor, M.; Tantos, A. Secondary structures of proteins: a comparison of models and experimental results. J. Proteome Res. 2021, 20, 1802–1808.
(18) van Raaij, M. E.; Segers-Nolten, I. M. J.; Subramaniam, V. Quantitative morphological analysis reveals ultrastructural diversity of amyloid fibrils from alpha-synuclein mutants. Biophys. J. 2006, 91, 1.96–1.98.
(19) Tompa, K.; Bokor, M.; Tompa, P. The melting diagram of protein solutions and its thermodynamic interpretation. Int. J. Mol. Sci. 2018, 19, No. 3571.
(20) Tompa, K.; Bokor, M.; Verebélyi, T.; Tompa, P. Water rotation barriers on protein molecular surfaces. Chem. Phys. 2015, 448, 15–25.
(21) Tompa, K.; Bokor, M.; Ágner, D.; Iván, D.; Kovács, D.; Verebélyi, T.; Tompa, P. Hydrogen mobility and protein-water interactions in proteins in the solid state. ChemPhysChem 2017, 18, 677–682.
(22) Torrisi, M.; Pollastri, G. Protein Structure Annotations. In Essentials of Bioinformatics; Shaik, N.; Hakeem, K.; Banagapappali, B.; Elango, R., Eds.; Springer: Chambridge, 2019; Vol. 1, pp 201–234.
(23) Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G. J. PRed4: a protein secondary structure prediction server. Nucleic Acids Res. 2015, 43, W389–W394.
(24) Torrisi, M.; Kaleel, M.; Pollastri, G. DeeProfs and Cascaded Recurrent and Convolutional Neural Networks for state-of-the-art Protein Secondary Structure Prediction. Sci. Rep. 2019, 9, No. 12374.
(25) Torrisi, M.; Kaleel, M.; Pollastri, G. Porter S: state-of-the-art ab initio prediction of protein secondary structure in 3 and 8 classes. BioRxiv 2018, No. 289033.

(26) Buchan, D. W. A.; Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. 2019, 47, W402–W407.

(27) Ma, Y.; Liu, Y.; Cheng, J. Protein Secondary Structure Prediction Based on Data Partition and Semi-Random Subspace Method. Sci. Rep. 2018, 8, No. 9856.

(28) Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-based protein structure modeling using the RaptorX web server. Nat. Protoc. 2012, 7, 1511–1522.

(29) Pollastri, G.; Przybylski, D.; Rost, B.; Baldi, P. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins 2002, 47, 228–235.

(30) Hefferman, R.; Yang, Y.; Paliwal, K.; Zhou, Y. Capturing nonlocal interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility. Bioinformatics 2017, 33, 2842–2849.

(31) Mészáros, B.; Erdős, G.; Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 2018, 46, W329–W337.

(32) Erdős, G.; Dosztányi, Z. Analyzing Protein Disorder with IUPred2A. Curr. Protoc. Bioinf. 2020, 70, No. e99.

(33) Dosztányi, Z.; Czórmók, V.; Tompa, P.; Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J. Mol. Biol. 2005, 347, 827–839.

(34) Mészáros, B.; Simon, I.; Dosztányi, Z. Prediction of protein binding regions in disordered proteins. PLoS Comput. Biol. 2009, 5, No. e1000376.

(35) Dosztányi, Z. Prediction of protein disorder based on IUPred. Protein Sci. 2018, 27, 331–340.

(36) Tompa, K.; Bánski, P.; Bokor, M.; Kamasa, P.; Lasanda, G.; Tompa, P. Interfacial water at protein surfaces: wide-line NMR and DSC characterization of hydration in ubiquitin solutions. Biophys. J. 2009, 96, 2789–2798.

(37) Rupley, J. A.; Gratton, E.; Careri, G. Water and globular proteins. Trends Biol. Sci. 1983, 8, 18–22.

(38) Nakasako, M. Water-protein interactions from high-resolution protein crystallography. Philos. Trans. R. Soc., B 2004, 359, 1191–1206.

(39) Durchschlag, H.; Zipper, P. Comparative investigations of biopolymer hydration by physicochemical and modeling techniques. Biophys. Chem. 2001, 93, 141–157.

(40) Waudby, C. A.; Camilloni, C.; Fitzpatrick, A. W. P.; Cabrita, L. D.; Dobson, C. M.; Vendruscolo, M.; Christodoulou, J. In-cell NMR characterization of the secondary structure populations of a disordered conformation of α-synuclein within E. coli cells. PLoS One 2013, 8, No. e72286.

(41) Cremades, N.; Chen, S. W.; Dobson, C. M. Structural Characteristics of α-Synuclein Oligomers. In International Review of Cell and Molecular Biology; 2017; Vol. 329, pp 79–143 DOI: 10.1016/bse.ircmb.2016.08.010.

(42) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983, 22, 2577–2637.

(43) Smolarczyk, T.; Roterman-Konieczna, L.; Stapor, K. Protein secondary structure prediction: a review of progress and directions. Curr. Bioinform. 2020, 15, 90–107.

(44) Frauenfelder, H.; Gratton, E. Protein dynamics and hydration. Methods Enzymol. 1986, 127, 207–216.

(45) Boyer, D. R.; Li, B.; Sun, Ch.; Fan, W.; Zhou, K.; Hughes, M. P.; Sawaya, M. R.; Jiang, L.; Eisenberg, D. S. The α-synuclein hereditary mutation E46K unlocks a more stable, pathogenic fibril structure. Proc. Natl. Acad. Sci. U.S.A. 2020, 117, 3592–3602.

(46) Bhattacharya, S.; Xu, L.; Thompson, D. Molecular Simulations reveal terminal group mediated stabilization of helical conformers in both amyloid-β42 and α-synuclein. ACS Chem. Neurosci. 2019, 10, 2830–2842.

(47) Cartelli, D.; Aliverti, A.; Barbiroli, A.; Santambrogio, C.; Ragg, E. M.; Casagrande, F. V. M.; Cantele, F.; Beltramone, S.; Marangon, J.; De Gregorio, C.; Pandini, V.; Emanuele, M.; Chieregatti, E.; Pieraccini, S.; Holmqvist, S.; Babucco, L.; Roybon, L.; Pezzoli, G.; Grandori, R.; Arnl, I.; Cappelletti, G. α-Synuclein is a novel microtubule dynamase. Sci. Rep. 2016, 6, No. 33289.