Articles

Transient Aggregation and Stable Dimerization Induced by Introducing an Alzheimer Sequence into a Water-Soluble Protein†

Daniel E. Otzen,*,‡ Simona Miron,§,|| Mikael Akke,§ and Mikael Oliveberg⊥

Department of Life Science, Aalborg University, Sohngaardsholmsvej 49, DK-9000 Aalborg, Denmark, Department of Biophysical Chemistry, Lund University, Box 124, SE-221 00 Lund, Sweden, and Department of Biochemistry, Umeå University, SE-90187 Umeå, Sweden

Received July 14, 2004

ABSTRACT: Transient contacts between denatured polypeptide chains are likely to play an important part in the initial stages of protein aggregation and fibrillation. To analyze the nature of such contacts, we have carried out a protein engineering study of the 102-residue protein U1A, which aggregates transiently in the wild-type form during refolding from the guanidinium chloride-denatured state. We have prepared a series of mutants with increased aggregation tendencies by increasing the homology between two β-strands of U1A and the Alzheimer peptide (β-AP). These mutants undergo transient aggregation during refolding, as measured by concentration dependence, double-jump experiments, and binding of ANS, a probe for exposed hydrophobic patches on protein surfaces. The propensity to aggregate increases with increasing homology to β-AP. Further, the degree of transient ANS binding correlates reasonably well with the structural parameters recently shown to play a role in the fibrillation of natively unfolded proteins. Two mutants highly prone to transient aggregation, U1A-J and U1A-G, were also studied by NMR. Secondary structural elements of the U1A-J construct (with lower β-AP homology) are very similar to those observed in U1A-wt. In contrast, the high-homology construct U1A-G exhibits local unfolding of the C-terminal helix, which packs against the β-sheet in the wild-type protein. U1A-G is mainly dimeric according to ^{15}N spin relaxation data, and the dimer interface most likely involves the β-sheet. Our data suggest that the transient aggregate relies on specific intermolecular interactions mediated by structurally flexible regions and that contacts may be formed in different β-strand registers.

Protein aggregation is a ubiquitous phenomenon with consequences ranging from inclusion bodies and poor expression yields in biotechnology to fatalities in protein deposition disorders (*1*−*3*). The process is often very specific (*4*), and a number of studies support a mechanism where aggregation occurs in kinetic competition with productive folding. Partially folded intermediates are suggested to coalesce via complementary "sticky" intermolecular interfaces similar to those found within the monomeric native state (*1*). However, this mechanism is unlikely to encompass all cases, since other studies implicate the unfolded state as the precursor for aggregation (*5*−*8*). Morphologically, aggregates may be amorphous or ordered; in the latter case, they typically form amyloid fibrils with a distinctive X-ray fiber diffraction pattern. The transient nature of the intermediates and the insolubility of the end products are clearly obstacles to detailed structural elucidation of the aggregation process. Aggregation kinetics typically involve a lag, followed by an exponential growth phase. This suggests that aggregation occurs by nucleation-dependent polymerization (*5*, *9*). Nucleation may involve a subset of critical residues (*10*, *11*), consistent with the observation that grafting of a small number of residues from amyloidogenic sequences into otherwise soluble proteins can induce specific oligomerization (*12*) and fibrillation (*13*).

The 39−43 residue Alzheimer peptide (β-AP),[1] which is the major component of the deposits formed in Alzheimer's disease, is one of the most intensely studied amyloidogenic peptides (*14*). β-AP forms an ordered fibril structure composed of β-strands (*5*, *15*−*17*), in which only 19 of the 39 Aβ backbone amides are protected against exchange (*18*). The aggregation of β-AP has also been investigated using synthetic fragments of β-AP (*14*). The hydrophobic C-terminal sequence, comprising residues 29−43, largely determines solubility (*5*, *19*) and is important for nucleation. Small amounts of fragments corresponding to residues 26−42 (β(26−42)) may even be used to seed fibril formation of fragments that are less susceptible to aggregation (*19*). The

† This work was supported by the Danish Technical Research Council (D.E.O.), the Swedish Research Council (M.A. and M.O.), the Swedish Foundation for Strategic Research (M.A.), and a FEBS postdoctoral fellowship (S.M.).

* To whom correspondence should be addressed. E-mail: dao@bio.aau.dk.

‡ Aalborg University.

§ Lund University.

|| Present address: Centre de Biophysique Moleculaire, UPR 4301 CNRS, University of Orleans, Rue Charles Sadron, 45071 Orleans Cedex 02, France.

⊥ Umeå University.

10.1021/bi048509k CCC: $27.50
Published on Web 09/23/2004

more hydrophilic N-terminus probably gives rise to the pronounced pH-dependence of the aggregation process (*20*, *21*). Interestingly, fragments corresponding to different sections of β-AP form different types of aggregates; for example, β(18−28) assembles into ribbonlike structures (*22*), while β(34−42) assumes a filamentous structure (*5*). This suggestion that different modes of interaction or contact patterns may operate is not unexpected, given that β-strands are able to form multiple contact patterns (*14*). For example, when an exposed β-strand in the ribosomal protein S6 was made more homologous to β-AP, this led to the stable formation of a protein tetramer that included well-defined protein−protein interfaces along different strands in addition to the exposed β-strand (*12*).

The early stages of aggregation are likely to be important in any nucleation-driven fibrillation process. In addition, the toxicity of amyloidogenic proteins is now attributed to the soluble oligomeric intermediates formed at early stages of aggregation and fibrillation, rather than the mature full-length fibrils (*23*, *24*). Therefore, to construct a system in which these stages can be investigated in more detail, as well as to gauge the contribution of different residues to this process, we have grafted β-AP-like sequences into the 102-residue RNA-binding domain 1 (RBD1) of protein U1A from the human spliceosome. The protein has a mixed α/β global fold (*25*), formed by a central four-stranded antiparallel β-sheet (residues 11−15, 40−44, 54−59, and 76−86) surrounded by three α-helices (Figure 1A). The A- and B-helices (residues 23−34 and 62−72) are packed against one face of the β-sheet, whereas the C-helix (residues 93−97) packs against the other side of the β-sheet and protects the hydrophobic core of the protein from solvent exposure. U1A folds from the denatured state to the native state without accumulating any intermediate (*26*) but shows an increasing propensity to aggregate transiently when refolding at protein concentrations above 1 μM (*6*).

Our strategy is inspired by the observation that U1A also has weak homology with β-AP. There is limited homology to the hydrophobic C-terminal moiety of β-AP in the 11-residue β-strand 2 of U1A (three residues identical, five homologous), while the last seven residues of β-strand 3 are homologous to residues 16−28 in the hydrophilic moiety of β-AP (three residues identical, two homologous). We have therefore constructed mutants that increase the homology to β-AP in both of these strands (Figure 1B). The exercise is not designed so much to probe how individual residues contribute to the aggregation process, but rather to estimate how a gradual accretion of fibrillogenic sequences of residues and the consequent change in the physicochemical properties of the protein affect the transient aggregation pattern.

Since U1A is fully soluble in its native state, it constitutes a good system for studying these effects. The drive to fold native monomers prevents extensive aggregation by pulling embryonic aggregates apart. We find that increasing β-AP homology gradually increases the propensity to aggregate transiently during refolding, and one mutant even forms dimers under equilibrium conditions. The dimer interface appears to involve hydrophobic interactions between the β-sheets of the monomers. This suggests that there is a threshold for the cumulative effects of amyloidogenic mutations, beyond which new oligomeric states become populated under equilibrium conditions.

MATERIALS AND METHODS

Preparation of Mutants. Wild-type RBD1 of UlA (the N-terminal fragment of the human U1A protein containing residues 1−102) does not contain any tryptophan residues. We have introduced the mutation F56W to provide a convenient spectroscopic probe to follow folding (*6*). For simplicity, we refer to this mutant as U1A-wt. In the following, all other mutations (Figure 1B) are described with U1A-wt as the reference. Mutagenesis was performed with the QuikChange kit (Stratagene) as described previously (*27*). All U1A mutants were purified as described previously (*28*). ^{15}N-labeled U1A mutants were expressed in BL21 (DE3) cells in M9 minimal medium with 1 g/L $^{15}NH_4Cl$. The plasmid encoding the mutant FW248 of RNA-binding domain 2 (RBD2) of U1A was a kind gift from Professor K. Hall. RBD2 was expressed as described (*29*).

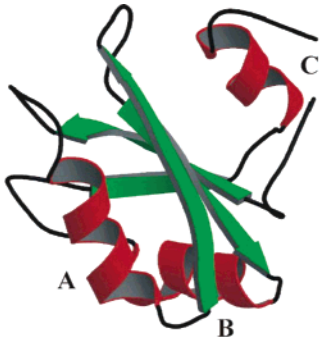
Stopped-Flow Kinetics. Kinetic folding experiments were performed using an SX18-MV stopped-flow instrument (Applied Photophysics, Leatherhead, Surrey, U.K.) as described previously (*27*). All experiments were performed in 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), pH 6.3, at 25 °C. Single-jump aggregation experiments were carried out by 1:10 dilution of the denaturant-unfolded protein into buffer. U1A mutants were unfolded in 5.4 M guanidinium chloride (GdmCl). ANS binding was monitored using 40 μM ANS and a final protein concentration of 10 μM with excitation at 360 nm and emission above 375 nm. All ANS measurements were done at the same photomultiplier voltage to allow direct comparison of the absolute values of the amplitudes. Double-jump experiments were carried out by 1:5 dilution of the denaturant-unfolded protein into buffer (giving a denaturant concentration of 0.9 M and a protein concentration of 10 μM unless otherwise stated), followed by unfolding into 5.0 M GdmCl after various delay times (60 ms to 200 s).

CD Spectroscopy. Spectra were recorded on a Jasco J-715 spectropolarimeter (Japan Spectroscopic Co. Ltd, Hachioji City, Japan) in 5 mM NaOAc, pH 4.8, and 50 mM NaCl. Near-UV spectra were recorded at a concentration of 300 μM using a 0.2 cm path length cuvette, while far-UV spectra were recorded at a concentration of 15 μM using a 0.1 cm path length cuvette. Five scans at a scan rate of 20 nm/min were recorded for each sample.

Binding of ANS and Thioflavin T. Protein (2 μM) was incubated with 40 μM ANS (8-anilino-1-naphthalene-

[1] Abbrevations: $Agg_{50\%}$, protein concentration at which 50% of the protein folds directly from the denatured to the native state; ANS, 8-anilino-1-naphthalene-sulfonate; ANS_{bind}, normalized degree of ANS binding; β-AP, Alzheimer peptide; HSQC, heteronuclear single quantum coherence; k', normalized second-order aggregation rate; NOE, nuclear Overhauser effect; RBD1, RNA binding domain 1; RBD2, RNA binding domain 2; R_1, longitudinal relaxation rate constant; R_2, transverse relaxation rate constant; η_{xy}, transverse cross-correlation relaxation rate constant; U1A-wt, F56W mutant of wild-type RBD1 (residues 1−102 of protein U1A); U1A-A, K50I/M51A mutant of U1A-wt; U1A-B, R47G/K50I/M51A mutant of U1A-wt; U1A-C, R47G/S48V/K50I/M51A mutant of U1A-wt; U1A-D, S46G/R47G/S48V/K50I/M51A mutant of U1A-wt; U1A-E, D42G/I43L/L44M/S46G/R47G/S48V/K50I/M51A mutant of U1A-wt; U1A-F, S46G/R47G/S48V/K50I/M51A/K60A mutant of U1A-wt; U1A-G, S46G/R47G/S48V/K50I/M51A/I58A/K60F mutant of U1A-wt; U1A-H, K60A mutant of U1A-wt; U1A-J, I58A/K60F mutant of U1A-wt.

(A)



(B)

```
Residue no.    1          11         21         31         41         51
Sec. structure         - -β1---      --------α1---     -- ----β2--    -
RBD2       MAPAQPLSENPPNH ILFLTNLPEE T----NELML SMLFNQFPGF KEVRLV--PG RH
U1A wildtype  MAVPETRPNH TIYINNLNEK IKKDELKKSL YAIFSQFGQI LDILVSRSLK MR
β-AP (residues 29-42)                                    GAI IGLMVGGVVI A
U1A-A (KI50/MA51)                                        GQI LDILVSRSLI A
U1A-B (RG47/KI50/MA51)                                   GQI LDILVSGSLI A
U1A-C (RG47/SV48/KI50/MA51)                              GQI LDILVSGVLI A
U1A-D (SG46/RG47/SV48/KI50/MA51)                         GQI LDILVGGVLI A
U1A-E(D42G/I43L/L44M/SG46/RG47/SV48/KI50/MA51)           GQI LGLMVGGVLI A
U1A wildtype  MAVPETRPNH TIYINNLNEK IKKDELKKSL YAIFSQFGQI LDILVSRSLK MR
U1A-F (SG46/RG47/SV48/KI50/MA51/KA60)                    GQI LDILVGGVLI A
U1A-G (SG46/RG47/SV48/KI50/MA51/IF58/KA60)               GQI LDILVGGVLI A
U1A-H (KA60)                                             GQI LDILVSRSLK M
U1A-J(IF58/KA60)                                         GQI LDILVSRSLK M
Residue no.                      61         71          81         91         100
Sec. structure          ---β3--  -----α2--- --          -β4--
RBD2                    DIAWVEFD NEVQAGAARD ALQGFKITQN NAMKISFAKK
U1A wt                  GQAWVIFK EVSSATNALR SMQGFPFYDK PMRIQYAKTD SDIIAKMKGT
β-AP (residues 16-28)     KLVFFA EDVGSNK
U1A wt                  GQAWVIFK EVSSATNALR SMQGFPFYDK PMRIQYAKTD SDIIAKMKGT
U1A-F                   AWVIFA EVSSATN
U1A-G                   AWVFFA EVSSATN
U1A-H                   AWVIFA EVSSATN
U1A-J                   AWVFFA EVSSATN
U1A-K (KI96/MA97)                                                SDIIAIAKGT
```

FIGURE 1: Three-dimensional solution structure (A) of the N-terminal domain of U1A protein (only residues 1−102 displayed) (*25*), represented using MOLSCRIPT (*62*) and sequence comparison (B) between the C-terminal half of the Alzheimer peptide and U1A, as well as the mutants constructed for this study. Note the significant homology in the second half of *β*-strands 2 and 3. The sequence alignment was performed with Multalin (*63*). The notation reflects the order in which the mutants were produced.

sulfonate) or thioflavin T in 5 mM NaOAc, pH 4.8, and 50 mM NaCl at different concentrations of GdmCl. For ANS, excitation was at 360 nm and emission at 475 nm. For thioflavin T, excitation was at 450 nm and emission at 482 nm.

NMR Spectroscopy. All NMR experiments were carried out on a 600 MHz Varian Unity Inova spectrometer at 298 K. U1A-wt and U1A-G/U1A-J samples contained 1 and 0.9 mM protein, respectively. The protein samples were dissolved in 5 mM NaOAc with 50 mM NaCl and contained 0.1 mM sodium 2,2-dimethyl-2-silapentane sulfonate (DSS) and 7% ^2H_2O. The pH was adjusted at 4.8 (uncorrected for isotope effects).

Sensitivity-enhanced ^1H-^{15}N HSQC spectra of U1A-wt, U1A-G, and U1A-J mutants were recorded using gradient selection and water flip-back (*30*−*32*). ^{15}N decoupling during the acquisition was performed using GARP-1 (*33*). The spectra were acquired with 8000 and 1800 Hz spectral widths and 1024 × 128 complex points in the ^1H and ^{15}N dimension, respectively, and 16 transients per free induction decay (FID). Chemical shift assignments were based on 3D ^{15}N total correlation spectroscopy (TOCSY)−HSQC and nuclear Overhauser effect spectrometry (NOESY)−HSQC experiments (*32*) acquired with 80 and 150 ms mixing times, respectively. In both cases, the spectral widths were 8000 and 1500 Hz in the ^1H and ^{15}N dimensions, respectively, and 256, 64, and 1024 complex points were sampled in the t1 (^1H), t2 (^{15}N), and t3 (^1H) dimensions, respectively. Proton chemical shifts were referenced relative to the water signal, which resonates at 4.73 ppm from the DSS signal at 298 K. Nitrogen chemical shifts were referenced indirectly relative to DSS (*34*).

The ^{15}N longitudinal (R_1) and transverse (R_2) autorelaxation rate constants and transverse ^1H-^{15}N dipolar/^{15}N chemical shift anisotropy cross-correlation relaxation rate constants (η_{xy}) were measured using two-dimensional experiments (*35*, *36*). All relaxation data sets were acquired with 2048 × 128 complex points in the ^1H and ^{15}N dimensions, respectively, and the same spectral widths as in the HSQC experiments. In the R_1 and R_2 experiments, eight transients per FID were acquired. In the η_{xy} experiments, 64 and 32 transients were acquired in the cross and reference experiments, respectively.

NMR Data Processing and Analysis. The NMR spectra were processed and analyzed using Felix98 Software (MSI, San Diego, CA). The final size of the processed HSQC spectra was 2048 × 2048 real points. The data were multiplied by exponential (^1H) and $\pi/4$-shifted sine-bell (^{15}N) apodization functions prior to Fourier transformation. The 3D TOCSY−HSQC and NOESY−HSQC spectra were extended by linear prediction in the indirect dimensions, multiplied by $\pi/4$-shifted sine-bell functions, and zero-filled prior to Fourier transformation.

R_1, R_2, and η_{xy} spectra were processed using two protocols to yield spectra optimized for either sensitivity or resolution. For resolution-optimized spectra, Lorentzian−Gaussian (^1H) and sine-squared window functions (^{15}N) were used. Sensitivity-optimized spectra were obtained using Lorentzian and cosine-bell window functions in the ^1H and ^{15}N dimensions, respectively. The data were zero-filled to generate data sets of 1024 × 1024 real points. The R_1 and R_2 values were obtained from 10 experiments with eight relaxation delays: 11 (×2), 77, 155, 255, 455 (×2), 799, 1299, and 1499 ms and 24 (×2), 33, 82, 132, 149 (×2), 215, 298 and 397 ms, respectively. The η_{xy} values were obtained from five experiments with relaxation delays of 10, 20, 30, 50, and 70 ms.

Cross-peak intensities were evaluated as peak heights, and the uncertainties were determined from duplicate spectra and from the standard deviation of the baseplane noise, as described previously (*37*). The R_1 and R_2 values were obtained by fitting two-parameter monoexponential functions to the experimental data. The η_{xy} values were obtained by nonlinear optimization of the function $I_{cross}/I_{auto} = \tanh(\Delta\eta_{I_{xy}})$, where Δ is the relaxation delay and I_{cross} and I_{auto} are the intensities from the cross and reference experiments, respectively (*36*).

RESULTS

Increased Homology to β-AP Generally Increases Transient Aggregation. Refolding of U1A from the GdmCl-denatured state involves several relaxation phases. The fast phase, A_1, corresponds to two-state folding from the denatured state (*6*). The two slow refolding phases, A_2 and A_3, are connected to transient aggregation, since they increase in relative amplitude as the protein concentration increases. The amplitudes of these phases all have a negative sign, showing that the single tryptophan of U1A becomes progressively less quenched as the reaction proceeds. Trp56 is partially solvent-exposed, so both inter- and intramolecular processes (aggregation and folding) can contribute to its fluorescence change.

Since U1A-wt aggregates transiently during refolding (*6*), we must evaluate the change in aggregation tendency induced by increased β-AP homology. We have empirically quantified the degree of transient aggregation as the parameter $Agg_{50\%}$, the protein concentration at which 50% of the protein folds directly from D to N. In practice, this is the protein concentration where the amplitude of A_1 is half that of the total amplitude of all the refolding phases (see Figure 2A). $Agg_{50\%}$, which is 3.8 μM for U1A-wt, varies with the degree of β-AP homology. The mutants U1A-A, U1A-E, U1A-F, U1A-G, U1A-H, and U1A-J, which all have high homologies to β-AP, have low $Agg_{50\%}$ values. On the other hand, U1A-B, U1A-C, and U1A-D all have wild-type-like $Agg_{50\%}$ values. U1A-B, U1A-C, and U1A-D differ from U1A-A in containing the mutation R47G (Figure 1). However, the single mutant R47G shows aggregation tendencies similar to U1A-wt (data not shown), indicating that the R47G mutation in itself does not reduce aggregation tendencies to any greater extent.

The slowest aggregation phase, A_3, disappears for the mutant I58A. This mutation in strand 3 reduces the homology with β-AP, which has a phenylalanine in the corresponding position, see Figure 1. U1A-K is a mutant in which we have introduced the same double mutation as that in U1A-A but in a part of the protein that does not show significant homology to β-AP (Figure 1). The rate of aggregation and $Agg_{50\%}$ value of U1A-K are similar to those of U1A-wt. In most of the other cases, the rate constants decrease upon mutation (Table 1).

Double-Jump Experiments Reveal Multiple Folding Routes. Previous work on transient aggregation of U1A did not clearly establish that A_2 and A_3 actually correspond to aggregation involving a gain of native structure, rather than aggregation to a nonnative state (*6*). To resolve this, we

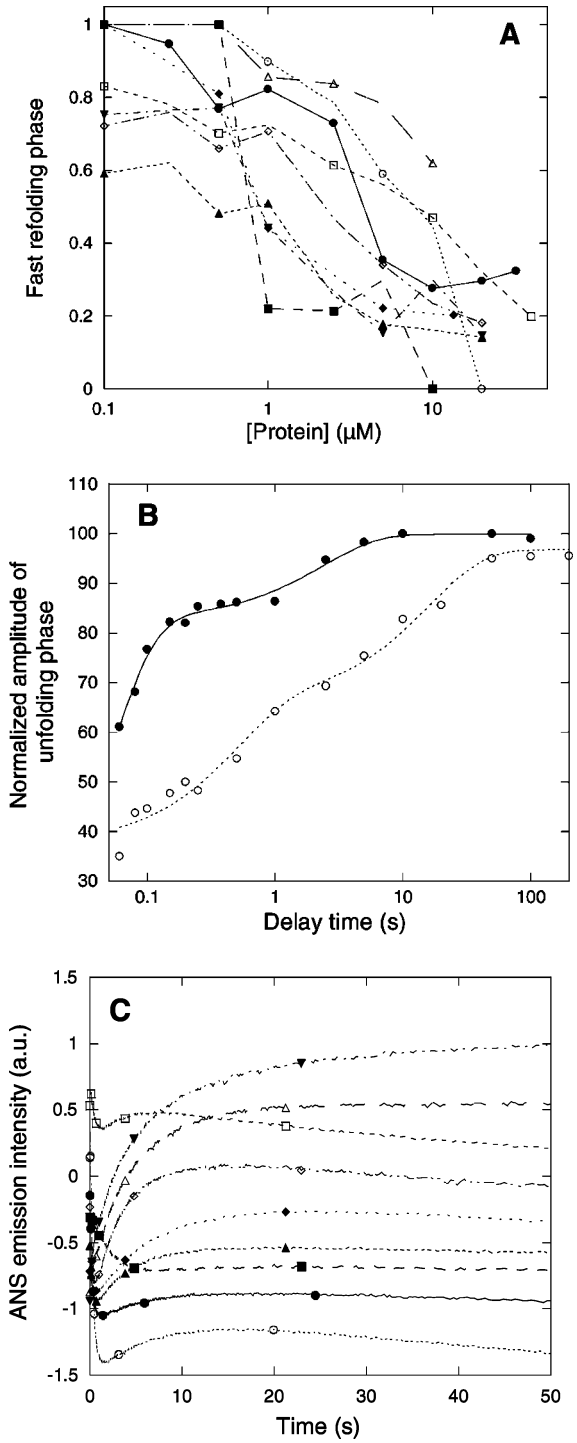


FIGURE 2: Panel A shows the degree of aggregation of U1A mutants versus protein concentration as the amplitude of the fast refolding phase versus the total amplitude of all observed phases: U1A wt (●), I58A (○), U1A-A (■), U1A-C (□), U1A-E (◆), U1A-F (△), U1A-G (▼), U1A-H (◇) and U1A-J (▲). Panel B shows the amplitude of the unfolding phase in double-jump studies for U1A-wt (●) and U1A-G (○) at a final concentration of 10 μM. Unfolded protein was refolded in 0.9 M GdmCl for various periods before being unfolded in 5 M GdmCl. The amplitude of the unfolding phase (which has the rate constant predicted from single-jump unfolding studies) is a measure of the degree of native structure. As can be seen from the amplitudes in Table 1, most of the protein folds from an aggregated state rather than from the denatured state. Panel C shows the time profiles for aggregation of U1A mutants in the presence of ANS. The mutants are represented by symbols as presented in panel A. Note that only U1A-F and U1A-G, which have high homology to β-AP, show a monotonic increase in ANS fluorescence.

carried out double-jump experiments in which unfolded protein was allowed to refold between 60 ms and 200 s before being unfolded again in 5.0 M GdmCl. The amplitude of the unfolding phase is a measure of the amount of native structure formed during the delay time. Since the apparatus does not allow us to measure reliably delay times shorter than 60 ms, we miss the direct refolding rate (around 100 s^{-1} at 0.9 M GdmCl, corresponding to a $t_{1/2}$ of 7 ms for refolding). At 10 μM, U1A-wt shows two slow phases for regain of native structure the rates of which are similar to those from single-jump experiments. A_2 has a rate constant of 12 s^{-1}, while the rate constant of A_3 is an order of magnitude slower and has a much smaller amplitude (Figure 2B, Table 2). At 1 μM, where the aggregation is only marginal, both phases disappear; instead, all native state accumulates within 60 ms (data not shown). The results confirm that A_2 and A_3 represent processes leading to the accumulation of the native state, and the extent of direct folding is low at 10 μM.

All the Alzheimer mutants with higher homology to β-AP than U1A-wt also show clear biphasic behavior (see Figure 2B) with rates in good correspondence to those of A_2 and A_3 from single-jump experiments (Table 2). This suggests that U1A can fold from the denatured state in three different ways, namely, via the direct refolding route, D → N, as well as via two distinct aggregated states, Agg → N and Agg* → N. At this stage, it is not clear whether Agg* forms from Agg or directly from D. The effect of the mutations follows a general pattern: the rate constants for the two slow phases decrease to a similar extent (there is a correlation coefficent of $R = 0.80$ between the two rates, data not shown), but the amplitude of A_3 increases relative to A_2, that is, the relative population of protein involved in the slow-folding aggregate increases with increased Alzheimer homology. The stepwise homology increase from U1A-wt to U1A-G slows down folding in a progressive fashion, in step with a decrease in $Agg_{50\%}$. Furthermore, the lower the value of $Agg_{50\%}$ is, the slower the rate constants are for A_2 and A_3. The three mutants that aggregate to the largest extent, namely, U1A-E, U1A-G, and U1A-J, also fold slower than wild-type in the direct route (Table 1). However, aggregation is not simply linked to slow folding. I58A folds even slower but aggregates less than wild-type. Furthermore, there is no correlation between $Agg_{50\%}$ and log k_f ($R = 0.18$ for 11 points).

Aggregation Can Be Monitored by ANS Binding. Transient aggregation was also probed in the presence of the hydrophobic dye ANS, the fluorescence of which is markedly enhanced upon interaction with a hydrophobic environment (*38*, *39*) as found in loosely structured protein conformations. In the stopped-flow experiments, increased fluorescence is therefore a good indication of ANS binding, while a decrease indicates ANS dissociation. To facilitate direct comparison of the amplitudes, experiments were carried out at 10 μM protein, which is above $Agg_{50\%}$ for nearly all the mutants. Three exponential phases (B_1, B_2, and B_3) are observed (Figure 2C). The very fast phase B_1 has a rate constant generally above 150 s^{-1} (i.e., even faster than direct folding to the native state) and a negative amplitude, which increases with protein concentration. B_1 appears to involve binding to the denatured state, since it is invariant to mutation and—unlike B_2 and B_3—it is also seen when ANS is mixed with denatured U1A under denaturing conditions, although with

Table 1: Equilibrium and Kinetic Data and Aggregation Tendencies for U1A Mutants

mutant	$\Delta\Delta G_{U-F}$ (kcal mol^{-1})[a]	k_f^{1M} (s^{-1})[b]	$Agg_{50\%}$ (μM)[c]	ANS_{bind}[d]	$\langle k_2 \rangle$ (s^{-1})	$\langle k_3 \rangle$ (s^{-1})	ln $\upsilon_{mut}/\upsilon_{wt}$ [f]
U1A-wt	0	97.2 ± 1.0	3.8	0	7.6	0.77	0.00
IA58[e]	1.53 ± 0.10	18.2 ± 0.2	8.4	38	3.2		−2.14
U1A-A	−0.10 ± 0.08	106 ± 1.4	0.8	−12	0.79	0.07	3.03
U1A-B	−0.39 ± 0.02	150 ± 4.5	6.0	−15	8.7	0.63	4.61
U1A-C	−0.14 ± 0.04	61.8 ± 0.60	8.8	−23	9.4	0.58	6.23
U1A-D	−0.23 + 0.03	93.8 ± 2.4	7.5	−11	7.6	0.44	4.92
U1A-E	1.45±0.13	45.8±0.85	0.9	−50	0.39	0.068	2.87
U1A-F	−0.76 ± 0.09	91.2 ± 1.3	2.5	−82	0.76	0.06	7.06
U1A-G	0.20 ± 0.09	25.5 ± 0.2	0.9	−100	1.38	0.074	6.82
U1A-H	−0.17 ± 0.08	103 ± 1.5	2.2	−43	6.9	0.24	2.14
U1A-J	0.47 ± 0.11	35.7 ± 0.4	0.4	−34	2.48	0.31	1.69
U1A-K	−0.24 ± 0.06	87 ± 0.8	6.0		9.9		2.88

[a] Calculated as in ref *61* using an *m*-value of 1.81 ± 0.04 M^{-1}. The value for U1A wt is set to 0. [b] Measured at a concentration of 0.5 μM protein where aggregation is not significant. [c] The protein concentration at which the amplitude of the fast phase is half the total amplitude of the refolding signal. [d] Refers to the amplitude of the two slow phases for refolding of 10 μM protein, measured by ANS fluorescence. The amplitude of wild-type has been subtracted, and the data have been normalized relative to the amplitude of U1A-G, the mutant with the highest amplitude. The value for U1A wt is set to 0. [e] This mutant has reduced homology to β-AP compared to U1A-wt and is therefore expected to aggregate to a smaller extent. [f] υ_{mut} and υ_{wt} are the extension rate constants for fibrillation of mutant and wild-type U1A, respectively, calculated according to ref *8*. These are theoretical values, because U1A does not fibrillate under the experimental conditions used in this study. However, they may be taken as indicators of the general change in fibrillogenicity caused by the mutations.

Table 2: Rate Constants for Formation of the Native State during Refolding Based on Double-Jump Measurements[a]

protein	amplitude phase 1[b]	rate constant phase 1	amplitude phase 2[b]	rate constant phase2
U1A-wt	84 ± 15	12.5 ± 2.3	16 ± 1	0.42 ± 0.09
U1A-A	51 ±6	5.0 ± 1.3	49 ± 4	0.13 ± 0.03
U1A-B	73 ± 19	18 ± 9	27 ± 6	0.61 ± 0.27
U1A-C	58 ± 7	16 ± 5	42 ± 7	0.26 ± 0.11
U1A-D	50 ± 7	6.7 ± 2.4	50 ± 7	0.33 ± 0.08
U1A-E	53 ± 7	1.79 ± 0.48	47 ± 7	0.11 ± 0.03
U1A-F	47 ± 7	2.55 ± 0.91	53 ± 8	0.12 ± 0.03
U1A-G	49 ± 5	1.35 ± 0.30	51 ± 5	0.06 ± 0.01
U1A-H	55 ± 10	8.7 ± 2.4	45 ± 3	0.26 ± 0.02
U1A-J	55 ± 8	9.2 ± 2.4	45 ± 4	0.11 ± 0.03

[a] Prior to the experiment, the protein is unfolded in 5.4 M GdmCl. The protein is then refolded into 0.9 M GdmCl at a concentration of 10 μM and after various delay times unfolded into 5 M GdmCl. All experiments were performed in 50 mM MES, pH 6.3, at 25 °C. [b] For each phase, the relative amplitude is calculated as the amplitude of the phase in question divided by the sum of the two amplitudes.

slightly reduced amplitude (data not shown). Thus, even if aggregates are formed during the deadtime of mixing and subsequently bind ANS, it is difficult to distinguish this process from ANS binding to the denatured state. The rates of B_2 and B_3 correspond relatively closely to those of phases A_2 and A_3. The amplitude of B_2 is negative, indicating ANS binding, while that of B_3 is positive (except for U1A-F and U1A-G), indicating ANS release. This suggests that the second phase represents aggregation as well as folding, see Discussion. We use the sum of the absolute values of the two amplitudes as a measure of the overall extent of ANS binding to the different U1A mutants during the folding process. Inclusion of the amplitude of the fast phase is not warranted because of its inability to distinguish between different mutants and conformational states, as well as the relative uncertainty of its determination. The sum of the amplitudes for B_2 and B_3 at 10 μM protein shows marked variation (normalized in Table 1 as ANS_{bind}). However, the general trend is that the higher the β-AP homology is, the more negative is ANS_{bind} with I58A and U1A-G located at each end of the scale.

Table 3: Correlations between ANS_{Bind} (Dependent Variable) and Mutational Changes in Physicochemical Parameters

independent variable	intercept	slope	correlation coefficient r [a]	probability of null hypothesis (%)
$\Delta\Delta G_{\beta-coil}$ [b]	−7.2 ± 17.5	7.1 ± 4.3	0.48	>10
$\Delta\Delta G_{coil-\alpha}$ [c]	12.6 ± 13.8	2.2 ± 1.2	0.54	10
Δcharge	10.5 ± 14.1	28.0 ± 8.1	0.76	<0.5
$\Delta\Delta G_{hydrophobicity}$ [d]	6.3 ± 113.0	4.7 ± 1.3	0.76	<0.5
$\Delta\Delta G_{weighted\ sum}$ [e]	2.6 ± 13.1	−9.7 ± 3.0	0.73	<0.5

[a] For comparison, the correlation coefficient between the normalized rate of aggregation *k'* and $\Delta\Delta G_{weighted\ sum}$ is 0.05 (Figure 2A). [b] Calculated according to ref *41*. [c] Calculated according to ref *40*. [d] Calculated according to hydrophobicity (*42*). [e] Calculated according to ref *8*.

Correlation between Aggregation and Primary Structure. A simple empirical equation has recently been developed to predict the aggregation tendencies of mutants relative to the wild-type protein (*8*, *11*). This is based on the observed correlation between aggregation and change in secondary structural propensity, hydrophobicity, and charge caused by the mutation. Using the data for U1A mutants, we have carried out multiple linear regression to search for correlations between different parameters deemed relevant for aggregation. The experimental data are k', $Agg_{50\%}$, and ANS_{bind}. k' is a normalized second-order aggregation rate that compensates for changes in the refolding rate induced by the mutations.[2] ANS_{bind} is the normalized degree of ANS binding, based on the amplitude of the ANS signal obtained during refolding, where the wild-type amplitude is defined as zero and that of the highest mutant U1A-G is defined as −100. The theoretical parameters are changes in charge, α-helix (*40*) and β-sheet propensities (*41*), and hydrophobicity (*42*).

There is no significant correlation between k' (or $Agg_{50\%}$) and the theoretical values (see Figure 3A). The only

[2] k' is derived as follows: at $Agg_{50\%}$, the rate of two-state folding equals the rate of aggregation, that is, $k_f = k_{agg} = k'(Agg_{50\%}) \leftrightarrow k' = k_f/Agg_{50\%}$. In other words, the higher k' is, the more aggregation-prone the mutant is.

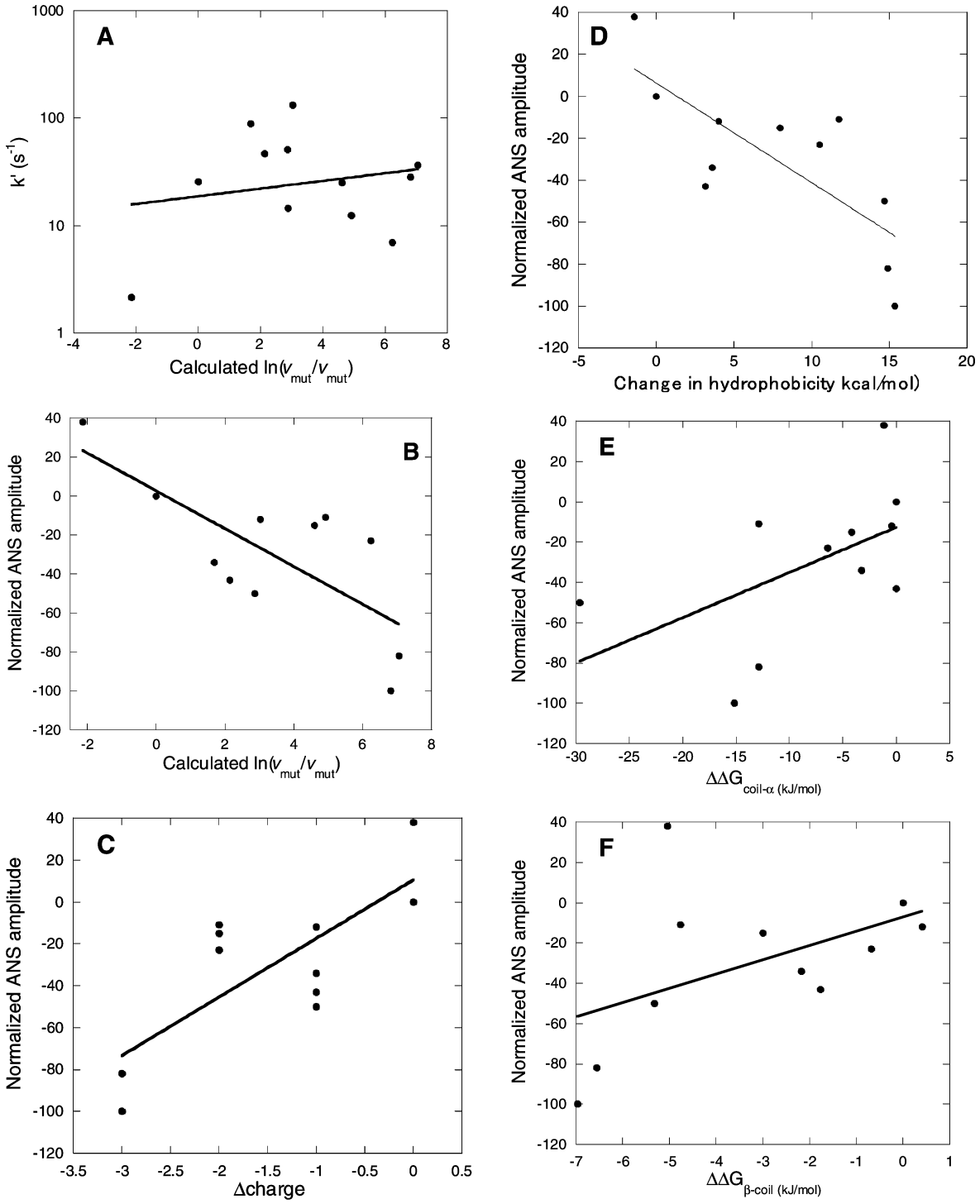


FIGURE 3: Plots of experimental versus theoretical parameters for transient aggregation of U1A mutants. Panel A plots normalized aggregation rate k' ($= k_f^{1M}/Agg_{50\%}$) versus the theoretical effect on fibrillation calculated according to ref *8*. The correlation coefficient is 0.05. The remaining panels show the normalized ANS amplitude versus (B) the theoretical weighted effect on fibrillation and the changes in (C) charge, (D) hydrophobicity, (E) α-helix propensity, and (F) β-sheet propensity, all calculated according to ref *8*. The intercepts, slopes, and correlation coefficients for panels B−F are listed in Table 3.

experimental parameter that correlates significantly with the theoretical ones is ANS_{bind}, as detailed in Table 3 and Figure 3B−D. The correlation is highest with charge and hydrophobicity, whereas the secondary structure propensities correlate less well. Combining the four parameters using the coefficients suggested from previous studies (*8*) also leads to a significant correlation, although it is not higher than those involving charge and hydrophobicity.

Equilibrium Studies. The U1A mutants U1A-G and U1A-J are particularly prone to aggregation during refolding. Is this aggregation only of a transient nature or does it also lead to an equilibrium conformational state that differs from the structure of native U1A-wt? CD far- and near-UV spectra of the mutants and wild-type under equilibrium conditions suggest subtle changes in secondary and tertiary structure (Figure 4A,B). Further insight was provided by ANS, the

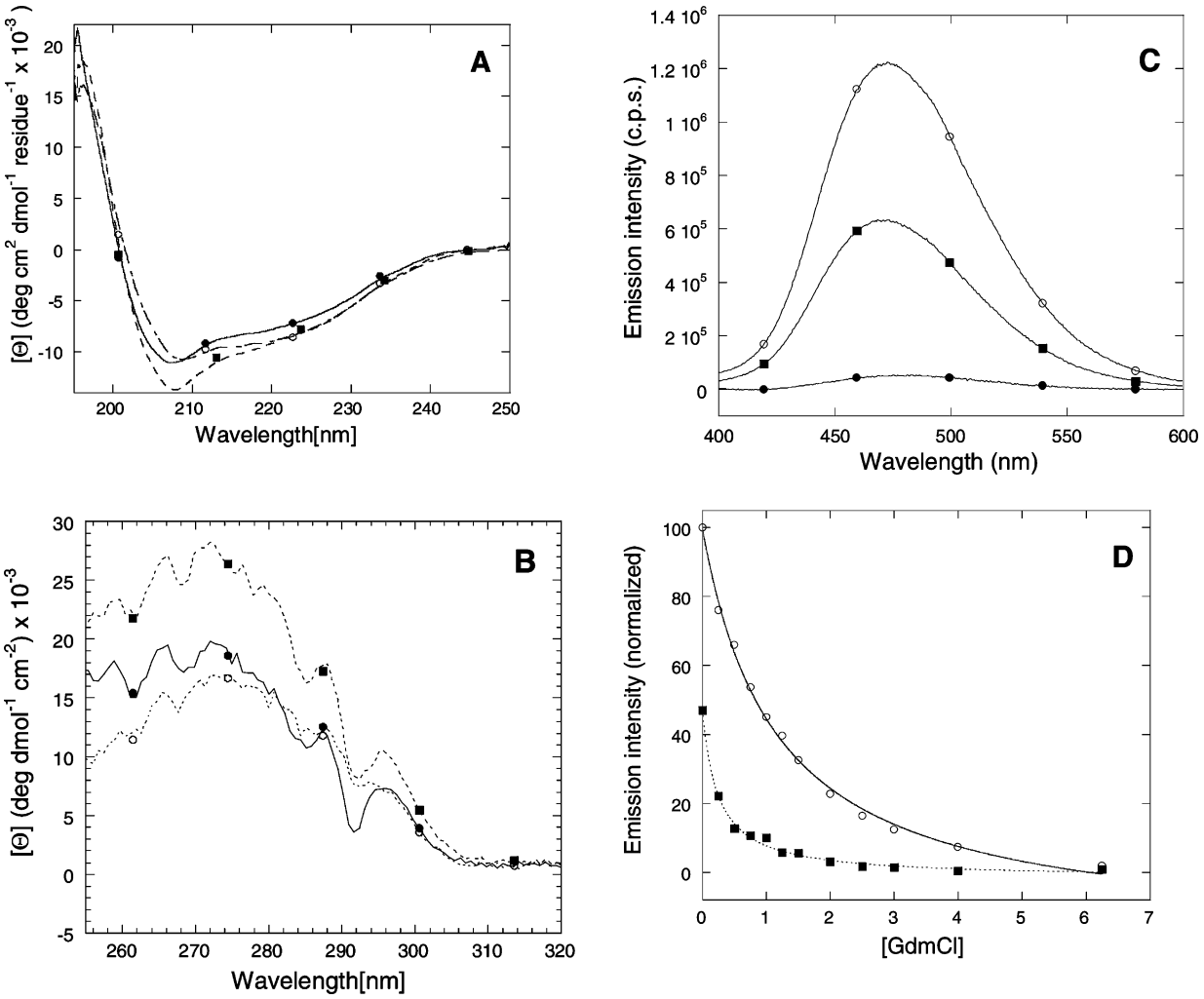


FIGURE 4: (A) Far-UV and (B) near-UV CD spectra of U1A-wt (●), U1A-J (■) and U1A-G (○), (C) binding of ANS to U1A-wt (●), U1A-J (■), and U1A-G (○) with background ANS fluorescence in buffer alone subtracted and excitation at 360 nm, and (D) titration with GdmCl of the ANS-binding state of U1A-G (○) and U1A-J (■). The two lines joining the points are the best fit to a simple binding model, primarily intended to guide the eye and provide an estimate of $[GdmCl]_{1/2}$, the GdmCl concentration at which the ANS fluorescence is reduced to 50% of its intensity in 0 M GdmCl. $[GdmCl]_{1/2}$ is around 0.7 and 0.2 M for U1A-G and U1A-J, respectively.

intrinsic fluorescence of which under native conditions (0 M GdmCl) increases substantially in the presence of U1A-G and to a smaller extent in U1A-J, while U1A-wt has no effect (Figure 4C). As gauged by the decline of the fluorescence intensity in the presence of GdmCl (Figure 4D), the ANS-binding conformation of U1A-G is somewhat more stable than that of U1A-J. The population of the ANS-binding conformation of U1A-J has declined to 50% already at 0.2 M, while the corresponding value for U1A-G is 0.7 M GdmCl. Neither U1A-G nor U1A-J showed any affinity for thioflavin T (data not shown), which tends to bind to fibrillar structures (*43*). Based on these spectroscopic studies, we decided to undertake a more detailed structural analysis using ^{1}H-^{15}N NMR spectroscopy.

NMR Resonance Assignments. Resonance assignments of the U1A-G and U1A-J mutants, as well as U1A-wt, were obtained based on 3D ^{1}H-^{15}N heteronuclear NMR spectra. U1A-wt and U1A-J have NMR spectra with better sensitivity and spectral resolution than the U1A-G mutant (Figure 5). Both U1A-J and U1A-G exhibit chemical shift changes relative to U1A-wt, but these are more accentuated for the U1A-G construct.

All amide groups (corresponding to all residues, except four prolines) were observed in the HSQC spectra of U1A-wt and U1A-J. Analysis of the heteronuclear (^{1}H-^{15}N) 2D HSQC, 3D TOCSY−HSQC, and 3D NOESY−HSQC spectra yielded assignment of almost all $^1H^N$ and ^{15}N backbone resonances in U1A-wt. Sequential assignments were based on the observation of consecutive $d_{NN}(i,i+1)$ or $d_{\alpha N}(i,i+1)$ NOEs or both in the 3D ^{15}N NOESY−HSQC spectrum. These assignments were subsequently used as the starting point for the assignment of resonances in the U1A-J and U1A-G mutants. Out of 96 $^1H^N/^{15}$N cross-peaks observable in the spectra of U1A-wt and U1A-J, 93 have been assigned for U1A-wt and 86 for U1A-J. The nonassigned resonances correspond to residues R47, V57, and K98 for U1A-wt, and I33, I43, R47, L49, Q54, W56, V57, F59, F60, and E61 for U1A-J. The nonassigned resonances are predominantly located in α-A, *β*-2, *β*-3, and α-B. In all cases, the nonassigned resonances are significantly broadened.

In the HSQC spectrum of the U1A-G mutant, only 81 $^1H^N/^{15}$N cross-peaks are observable. Of these, 17 cross-peaks exhibit very narrow line widths (about 5−6 Hz). All others are substantially broadened with line widths ranging from 15 to 27 Hz. The large line widths render the assignment process difficult, but nevertheless 55 resonances were unambiguously assigned. The unassigned residues comprise H10−N18, I33−F34, L41−V62, G74−F77, and R83−K88.

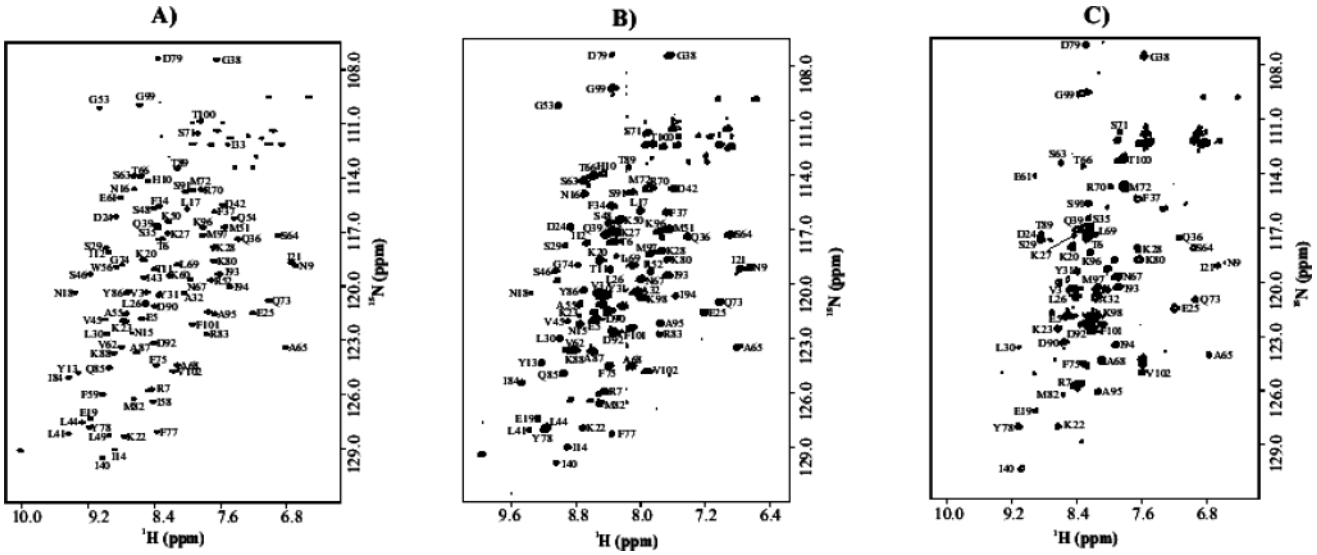


FIGURE 5: Two-dimensional ^{1}H-^{15}N HSQC spectra of (A) U1A-wt, (B) U1A-J, and (C) U1A-G mutants recorded at 298 K. The assigned ^{1}HN/^{15}N cross-peaks are labeled according to the residue number.

With the exception of I33−F34, located in helix A, the unassigned residues are all located in the β-sheet and the loops connecting the β-strands. Note that segments I33−F34 and L41−V62 include all residues that are unassigned also in U1A-J. This observation suggests that similar perturbations of the structure occur in the two mutants. In sharp contrast to the remainder of the residues in U1A-G, 17 residues comprising V3−R7 and D90−V102 (residue 4 is a proline) have very intense cross-peaks, indicating that the N- and C-terminal segments exhibit high flexibility in the U1A-G protein and are largely unstructured (see further below).

NMR Structural Analysis. Chemical shifts are exquisitely sensitive to conformational changes (*44*). The chemical shift differences between U1A-wt and U1A-J/U1A-G indicate that conformational perturbations are caused by the mutations. Moderate chemical shift differences among the three forms are observed throughout the sequence (Figures 5 and 6), suggesting relatively small rearrangements of the tertiary structure. Larger $^{1}H^{N}$ and ^{15}N chemical shift deviations ($\Delta\delta$ > 0.3 and 2 ppm, respectively) are observed only for residues that are located in the C-terminal helix of the U1A-G mutant (Figure 6).

NOE cross-peak patterns provide information about the secondary structure elements in proteins and β-sheet topologies (*45*). The medium- and short-range NOEs identified for U1A-wt and U1A-J indicate that the secondary structural elements and fold topology are conserved between these two proteins, and that they agree with the structure of F56 wild-type U1A (*25*). The α-helices were identified by sequential $d_{NN}(i,i+1)$ NOEs between amide protons. Analysis of strong sequential $d_{\alpha N}(i,i+1)$ and weak intraresidue $d_{\alpha N}(i,i)$ NOEs, as well as long-range $d_{NN}(i,j)$ and $d_{\alpha N}(i,j)$ NOEs, identified the four-stranded antiparallel β-sheet in U1A-wt and U1A-J. Since the 55 residues assigned in U1A-G are located predominantly in the α-helices and loops, no information about the preservation of the β-sheet region of the U1A-G mutant could be obtained from the NMR data. The A- and B-helices were identified based on characteristic $d_{NN}(i,i+1)$ NOE connectivities. In the C-terminal part of the U1A-G variant, residues D90−K98, corresponding to helix C in U1A, are characterized by strong $d_{\alpha N}(i,i+1)$ NOEs and the absence of $d_{NN}(i,i+1)$ NOEs, indicating that this segment has an extended and unstructured conformation in U1A-G.

NMR Relaxation Data. NMR relaxation measurements provide a sensitive measure of the rotational diffusion correlation time, which is dependent on molecular weight and shape. Using ^{15}N spin relaxation data, we studied the tendency of the U1A-wt, U1A-J, and U1A-G constructs to self-associate.

It has been shown previously that U1A-RBD1 has a tendency to aggregate above 1 mM concentration (*29*). We performed all ^{15}N relaxation experiments reported here at a concentration of 0.9 mM for U1A-J and U1A-G and 1 mM for U1A-wt. Comparison of relaxation data obtained for the 0.9 mM U1A-G sample (the mutant harboring the greatest tendency to aggregate) with an equivalent data set obtained at 0.6 mM confirms that the aggregation state is unaffected by the difference in concentration over this range (data not shown). ^{15}N R_1 relaxation rates of U1A-wt, U1A-J, and U1A-G are shown in Figure 7A. As observed, the R_1 values are highly similar for U1A-wt and U1A-J, while those for U1A-G are dramatically smaller. Notably, U1A-G exhibits significantly larger dispersion among its values of R_1 as compared to U1A-wt and U1A-J; in particular, the C-terminal segment deviates from helices A and B. The weighted mean R_1 values for U1A-wt and U1A-J are 1.4 ± 0.2 and 1.5 ± 0.1 s^{-1}, respectively. These values agree well with what is expected for a nearly spherical monomeric protein of this size. For U1A-G, the core residues (9−88) with R_1 < 1.0 s^{-1} have a weighted mean of $\langle R_1 \rangle = 0.85 \pm 0.07$ s^{-1}, which is significantly smaller than the values obtained for U1A-wt and U1A-J. As shown in Figure 7B, the lower value of R_1 indicates that U1A-G has a larger rotational correlation time, τ_m, than U1A-wt and U1A-J. Furthermore, the gradual increase and decrease in R_1 observed for segments 89−94 and 94−102, respectively, indicate that the effective rotational diffusion correlation times decrease as the distance from the protein core increases (Figure 7A,B).

R_2 and η_{xy} values for U1A-G are shown in Figure 7C. Due to extensive line broadening of residues in the β-sheet, relaxation rates could be measured only for residues in the

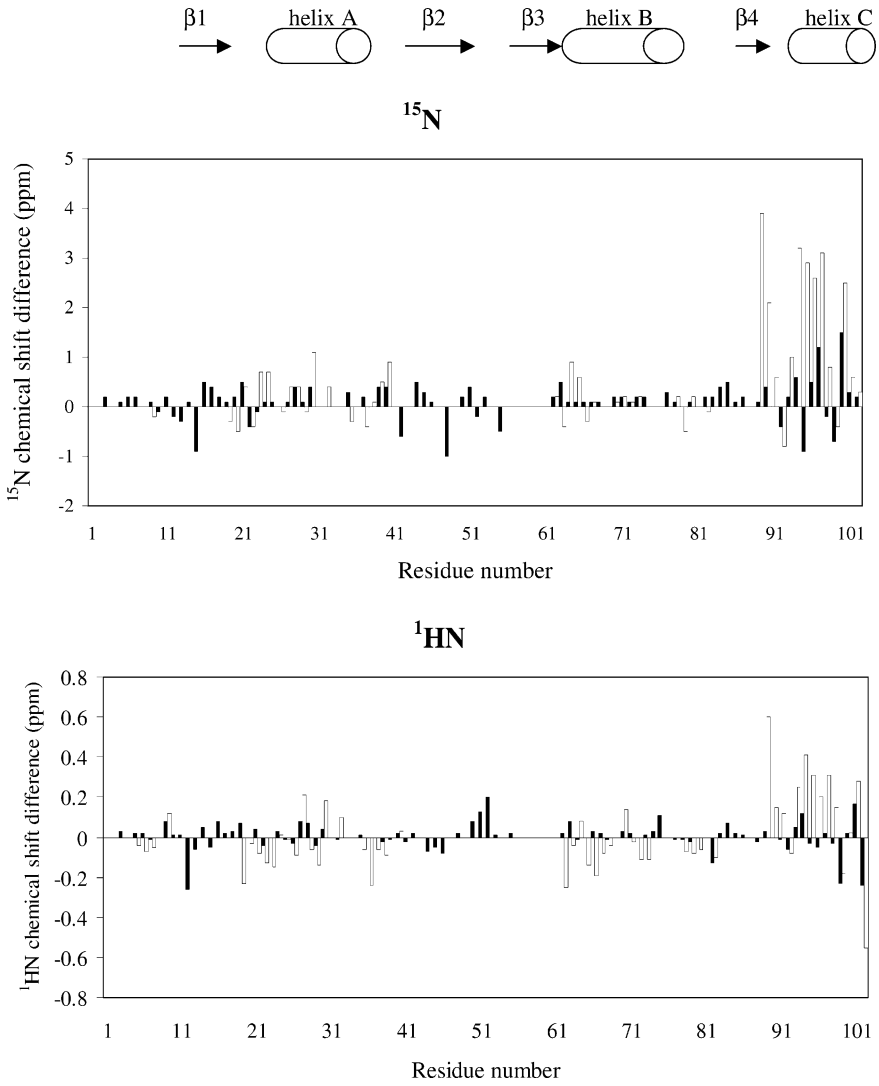


FIGURE 6: ^{15}N and $^{1}H^{N}$ chemical shift difference between U1A-wt and U1A-G (open bars) and U1A-J (filled bars) for all assigned residues along the polypeptide sequence. The secondary structure elements are indicated at the top of the figure.

helices and near the termini. Resonances from helices A and B are also severely line-broadened, which compromises the precision in the measurements, and especially so for the inherently insensitive η_{xy} experiment. U1A-G shows a large dispersion also in the R_2 and η_{xy} relaxation rates with the lowest values observed for residues 89−102. Again, these results show that the segment corresponding to the C-terminal α-helix in the wild-type protein has transverse relaxation rates typical of a highly flexible peptide chain. The weighted mean values of R_2 and η_{xy} for core residues with rates greater than 10 s^{-1} are 17 ± 3 and 14 ± 4 s^{-1}, respectively. In the absence of exchange contributions to R_2, the expected ratio between R_2 and η_{xy} is 1.3 ± 0.1 (*46*). The measured ratio (1.2 ± 0.4) agrees with the expected result within experimental uncertainties, indicating that exchange does not make a dominant contribution to R_2 for these residues. For comparison, the average R_2 is 8.3 ± 0.2 s^{-1} (at 298 K and a static magnetic field strength of 11.7 T) for the predominantly monomeric wild-type N-terminal domain of U1A (RBD1) (*29*). Since R_2 is approximately proportional to the rotational correlation time and only weakly dependent on static magnetic field strength, comparison of the results for wild-type and U1A-G directly indicates that U1A-G is predominantly dimeric. Estimating the rotational correlation time from the weighted average $\langle R_2/R_1 \rangle = 17$ yields a value of $\tau_m = 13$ ns, which is approximately twice as big as the value obtained for the monomeric N-terminal domain of U1A, $\tau_c = 7.6 \pm 0.5$ ns (*29*). The agreement between $\langle R_2/\eta_{xy} \rangle$ and the expected value indicates that the extracted value of τ_m is not compromised by conformational exchange. In contrast to residues at the N- and C-terminal segments and in helices A and B, the extensive broadening of resonances from the β-sheet suggests that residues in this region experience significant exchange between different conformations.

DISCUSSION

Correlation of Aggregation with properties of the β-AP Peptide. We undertook this study on the transient aggregation of U1A mutants to obtain a better understanding of the events involved in the early stages of protein aggregation and oligomer assembly. Our strategy was to introduce sequences that were expected to increase fibrillogenicity into a soluble scaffold system, thus eliminating solubility issues as well as preventing the aggregation reaction from passing beyond the oligomer stage. As is apparent from Table 1, increasing the homology to β-AP is also predicted to lead to a general increase in fibrillation propensity relative to U1A-wt, using the empirical relationships reported by Chiti et al. (*8*).

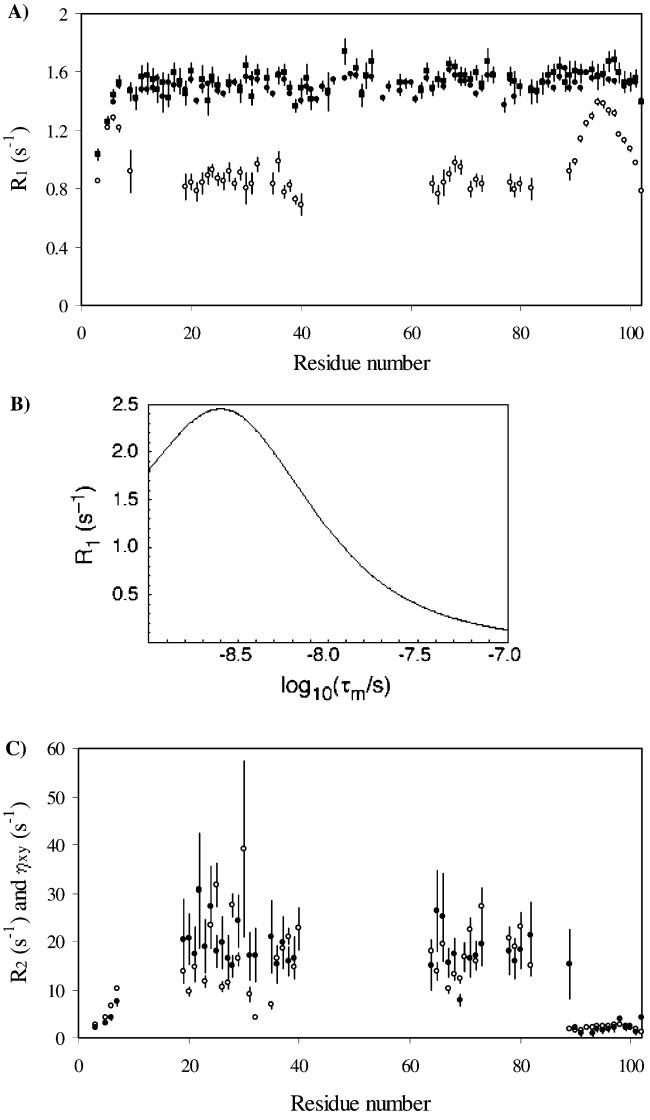


FIGURE 7: (A) ^{15}N R_1 data for U1A-wt (●), U1A-J (■), and U1A-G (○), plotted as a function of residue number, (B) variation of ^{15}N R_1 relaxation rates with rotational correlation time τ_m, and (C) ^{15}N R_2 data (○) and transverse cross-correlation rate constants η_{xy} (●), shown as a function of amino acid number for U1A-G.

Before we embark on a general discussion of our findings, it is worth noting that several residue-specific properties of the β-AP peptide appear to be retained when sequences of increasing homology to β-AP are engineered into U1A-wt. For example, the two mutated residues in U1A-A (I50/A51; located between strands β-2 and β-3) correspond to I41/A42 in the hydrophobic C-terminal moiety of β-AP. These two residues are critical for the nucleation step in amyloid formation (*19*), although they do not affect the stability of the amyloid once formed. U1A-A is markedly more aggregation-prone than U1A-wt; it also forms inclusion bodies in *E. coli* even when grown at 30 °C, in contrast to U1A-wt (data not shown). In contrast, U1A-K, which contains a similar double mutation in a part of the protein (the C-terminal α-helix) that does not show homology to β-AP, does not aggregate significantly more than U1A-wt. Thus, the aggregation tendency of U1A-A cannot simply be ascribed to the increased hydrophobicity of this mutant compared to U1A-wt. Further, increasing the β-AP homology in strand β-3, which is homologous to residues 16−28 in the N-terminal sequence of β-AP (Figure 1B), also stimulates aggregation. Peptides corresponding to residues 16−28 are reported to form ribbons but no amyloids (*22*), while the sequence 1−28 produces insoluble β-pleated sheet structures (*22*). Finally, I58 in U1A corresponds to F19 in β-AP. The U1A mutation I58A further decreases homology to β-AP and reduces aggregation tendencies. Similarly, the mutation F19I/F20G in a 10−23 fragment of β-AP is much more soluble than the wild-type 10−23 fragment and does not yield filaments (*20*).

In nature, RBD1-U1A is linked to another domain, RBD2, with which it shares 25% amino acid identity and the same overall topology (*29*). RBD2 has limited homology to β-AP but in a different part of the protein, corresponding to the second α-helix (Figure 1). RBD2 shows a very limited degree of aggregation between 0.03 and 12 μM (the fast phase decreases from ca. 105 to 85 s^{-1}), but, unlike U1A-wt, it has no distinct aggregation phase at high protein concentrations. It has one slow refolding phase, which is concentration-independent, and double-jump experiments strongly suggest that this is due to proline isomerization (data not shown). The low level of aggregation is consistent with the hypothesis that transient aggregation can be mediated by β-AP homology primarily in β-strand sequences.

U1A may represent a good soluble model system to study the early processes in amyloid formation. The reasonable correlation between Chiti's parameters (*8*) and the ANS data (Figure 3) suggests that this parameter reveals aspects reflecting the associative behavior of the U1A mutants. One may question the validity of comparing the experimental parameters from the present type of work with the empirical relationships derived from certified fibrillation studies (*8*). In the latter case, focus has been on a late stage in fibrillation, namely, the elongation phase, which unlike the lag phase may not be critically dependent on early events such as nucleation (*10*) but typically occurs after many hours or days. This contrasts with the millisecond−minute time scale of aggregation, which has to compete with folding. Here only small aggregation interfaces between a small number of polypeptide chains ("preclusters") are formed; these are then dissociated by the rapid folding of U1A. Nevertheless, such preclusters may still be relevant to the fibrillation process as soluble precursors to protofibrils (*47*) and macroscopic aggregates; as such, they may well occur during aggregation of β-AP but may be missed due to the limited time resolution of light-scattering methods. Their formation may be encouraged by particular conditions of the stopped-flow experiment, in which the structural flexibility of the denatured state as it folds (see ref *48*) and the turbulent flow associated with mixing (*49*, *50*) will encourage sampling of intermolecular contacts.

Variable Registers. What can our model system tell us about the aggregation process? The fact that individual side-chain mutations affect aggregation suggests specific interactions. This is consistent with the observation that U1A aggregation is not affected by the presence of unrelated denatured proteins (D.E.O. and M.O., unpublished observations). It therefore seems plausible that the polypeptide chains align in a well-defined manner during transient aggregation. Several indirect lines of evidence suggest that the intermolecular registers vary from one mutant to another, so the exact pattern is unlikely to be critical for aggregation. First, U1A-A

and U1A-J have increased homology to β-AP in different strands (and therefore probably have different intermolecular contact surfaces), yet they both have low $Agg_{50\%}$ values. Second, transient aggregation generally increases with increasing β-AP homology in a progressive fashion (Table 1). This shows that no individual residue is essential for aggregation, that is, there is a certain level of degeneracy. Different residues contribute to different extents; for example, the double mutations K50I/M51A and K58I/K60F make major contributions. The variation in contact patterns is also consistent with the large variation in the different ANS affinities of the aggregates. Third, the existence of alternative registers is consistent with the observations that peptides comprising various fragments of β-AP can aggregate into different types of nonamyloid-like β-sheet structures rather than fibrils (*5*, *20*, *22*, *51*), can form transient microcrystalline arrays (*52*), and—when grafted into other proteins—can form tetrameric assemblies with multiple β-strand interfaces, including regions outside the β-AP region (*12*).

Thus the homology of U1A to β-AP in both β-strands 2 and 3 may give rise to different alignments and a heterogeneous population of oligomeric molecules. As a consequence, it is possible that replacement of individual amino acids in U1A will lead to register swaps. A similar effect has been observed on the macroscopic level for the protein associated with light-chain deposition diseases, where two mutations sufficed to change the appearance of the aggregate from amorphous to fibrillar (*53*). Such register swaps may veil any correlations between aggregation propensities and secondary structure propensities, explaining the lack of correlation between $Agg_{50\%}$ and the aggregation propensities predicted using the parameters of Chiti et al. (*8*). The ANS-binding amplitudes appear to be a more appropriate parameter to correlate with the Chiti parameters, possibly because the sites on the protein that bind ANS are also those that are involved in intermolecular contacts (see discussion below).

A Model for Transient Aggregation. On the basis of the suggested variation in register, we imagine that the aggregation process involves an ensemble of interconverting species, progressing in parallel rather than following a highly specific aggregation pathway. As the participating protein molecules become more organized, fewer and more potent contact patterns are sampled, analogous to the funnel view of protein folding (*54*). Despite this complexity, however, there appears to be some general features that can be encompassed in a simple kinetic scheme. Since three phases are involved, of which the two slowest are sensitive to protein concentration, a minimalist scheme (Figure 8) must contain at least four states, namely, the denatured state, D, the native state, N, and two different aggregate species, Agg and Agg*, with high ANS affinity due to hydrophobic patches exposed in intermolecular contacts. These states probably all interconvert and enable N to be formed in several different ways. Since we have not been able to identify any partially folded monomeric states during folding of U1A according to standard analyses of kinetic data (*6*, *26*), our model postulates that aggregation of U1A takes place from the denatured state, D, without the involvement of any "sticky" intermediates (*6*, *55*). This is consistent with observations on acyl phosphatase A (*7*) and myoglobin (*56*). Agg is probably formed in a diffusion-limited manner within the deadtime of the experiment (*6*) in competition with folding. Since the native

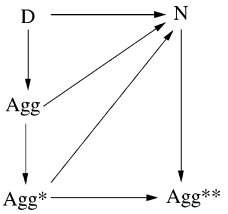


FIGURE 8: Proposed transient aggregation model for U1A. Upon transfer to native conditions, the denatured state D may fold directly to N or aggregate to Agg and Agg*. It is unclear whether U1A is released from Agg and Agg* in a completely or partially folded state or whether it has to unfold completely before reaching the native state. Under some conditions, a stable aggregate Agg**, such as a dimer, may form.

state is formed in two different slow phases, Agg can probably rearrange to Agg*, as well as dissociate and fold to the native state. D, Agg, and Agg* are unlikely to coexist under the unfolding conditions used here (5.2 M GdmCl) prior to transferring them to refolding conditions, since folding from 7.4 M GdmCl, which would be expected to destabilize the more aggregated species, did not alter the aggregation behavior (data not shown). Furthermore, the same aggregation characteristics were seen when refolding was carried out from the SDS-denatured state (data not shown).

Both the rearrangement from Agg to Agg* and the subsequent dissociation and folding to N may give rise to the release of ANS observed in the B_2 phase. It appears contradictory that our double-jump experiments suggest that the second phase corresponds to folding while ANS data suggest that further aggregation takes place. However, it is entirely possible that both events take place during this phase, leading to kinetic partitioning between folding from Agg to N and further aggregation from Agg to Agg*, see Figure 8. The only kinetic requirement for these parallel tracks is that the rate constants are of the same magnitude.

Specific interactions must be involved in both Agg and Agg*, since they are equally sensitive to side-chain mutations. However, the protein molecules in Agg* are probably more structured, making dissociation and folding from this state slower. In the final folding step from Agg* to N (covering a time scale of 10−100 s, corresponding to phase B_3), ANS probably dissociates from the native state, leading to the decline in ANS signal seen for nearly all the mutants. However, U1A-G, which has the highest potential to aggregate and dimerize, shows no ANS release, since the ANS signal continues to increase throughout the experiment. This is consistent with the ability of ANS to bind to oligomeric structures as postulated in our model. Our NMR experiments suggest that the accumulation of mutations beyond a certain threshold leads to the formation of stable dimers. We now turn to a discussion of the additional information that this provides.

Mechanism of Dimerization of U1A. Our data for U1A-wt and U1A-J agree well with other structural reports on U1A. Dynamical properties of the N-terminal domain of the U1A protein have been studied previously using ^{15}N NMR relaxation data (*29*, *57*, *58*). The rotational correlation time of the N-terminal domain has been determined for the F56 U1A protein and the F56Y mutant. The values of $\tau_m = 7.6 \pm 0.5$ and 6.0 ± 0.6 ns, respectively, at 298 K are highly

similar to our data for U1A-wt and U1A-J. Further evidence for the conservation of structure by the F56W mutation in U1A is that it does not change the affinity and specificity of the protein for RNA (*57*) nor the CD spectrum of the protein (*59*).

However, U1A-G, in particular, and U1A-J to a smaller extent have undergone structural changes compared to the wild-type. What part of the protein do they affect? The crystal structure of U1A (*25*) revealed two hydrophobic cores. The major hydrophobic core (core 1) comprises residues 12, 14, 17, 21, 26, 30, 33, 34, 41, 43, 45, 57, 59, 65, 69, 72, 77, 78, 82, 84, and 86, while the "satellite" hydrophobic core (core 2) involves residues L44, I55, F56, I93, I94, and M97 and is formed by interactions between the C-terminal helix C and the β-sheet surface. Residue 56 is not exposed to the solvent, judging from steady-state fluorescence and quenching studies on Trp56 (*59*). Our data suggest that both cores are altered by the mutations.

The NOE connectivities clearly indicate that the solution structure is conserved in both U1A-wt and U1A-J. Furthermore, the ^{15}N relaxation data on U1A-wt and U1A-J reveal that the C-terminal helix shows approximately the same degree of flexibility (or lack thereof) as the core of the protein, corroborating the view that the C-terminal helix is intact in these proteins. However, the nonassigned residues in U1A-J have significantly broadened cross-peaks, which suggest exchange among multiple conformations. These nonassigned residues are located in the region of core 1, indicating that core 1 has become more flexible although the protein remains as stable as U1A-wt (Table 1).

The amino acid substitutions of U1A-G lead to dimerization of the protein. The self-association is most likely mediated by the β-sheet and is associated with the unfolding of helix C and disruption of core 2. The dimer interaction surface thus involves residues in the β-sheet, most of which are broadened significantly, indicating extensive conformational averaging on the micro- to millisecond time scales, possibly between different registers. Together, the present results suggest a model where the β-AP homology mutations destabilize native contacts between the β-sheet and the C-terminal α-helix, thereby increasing the relative propensity to form intermolecular β-sheet contacts that are averaged dynamically.

The difference between U1A-G and U1A-J in ANS binding can be rationalized in light of the NMR results. ANS is released from both proteins at increasing GdmCl concentrations under conditions where the protein remains native. It is possible that the perturbation of core 1 enables ANS to bind to some extent to U1A-J. ANS release from U1A-J is presumably simply due to its solvation by GdmCl. The increased binding of ANS to U1A-G, as manifested by the higher concentrations of GdmCl needed to displace it from the protein, is consistent with destabilization of core 2 in addition to core 1. Furthermore, the formation of dimers with a poorly defined interface directly suggests the possibility of additional binding sites for ANS in the U1A-G dimer interface, thus coupling binding to dimer formation. In principle, the increased line widths in U1A-J may be caused by exchange between monomeric and dimeric states; however, the good agreement between the R_1 rates for U1A-J and U1A-wt indicate that the population of dimers must be very small (approximately a few percent) in this case.

Our data are consistent with the notion that intermolecular contacts occurring during aggregation are mediated by regions of structural flexibility. The conformational "flexing" experienced during refolding enables transient sampling of intermolecular interfaces. This is linked to the formation of transient patches of solvent-exposed hydrophobic regions, as seen by the tendency of the U1A mutants to bind ANS during this process. ANS binding tendencies and presumably the degree of flexibility in the native state increase with homology to β-AP. Only U1A-G, which is sufficiently perturbed by mutations to retain extensive regions of flexibility in the compact folded state, is found to maintain the dimer structure under equilibrium conditions and to bind ANS permanently. If flexibility, and thus the presence of partially disordered and solvent-exposed hydrophobic regions, is linked to the ability to aggregate, then this explains why our empirical ANS-binding parameter, ANS_{bind}, is the experimental parameter that correlates best with the structural parameters shown to predict aggregation propensities for other proteins (*8*).

It is not a novel observation that flexibility or plasticity favors certain types of protein−protein interactions. Proteases preferentially degrade unstructured regions of their protein substrate, in some cases forming β-sheet structures with the substrate (*60*), and fibrillation often proceeds from partially folded states that accumulate under mildly denaturing conditions (*2*). However, it is interesting that the flexibility does not come at the expense of global stability; U1A-G has essentially wild-type stability. None of the mutations in U1A-J and U1A-G directly involve those residues identified as part of the two hydrophobic cores (*22*), yet conformational changes are apparently transmitted to the cores such that residues in these regions experience significant line broadening, consistent with increased flexibility.

ACKNOWLEDGMENT

We thank Professor K. Hall for providing the plasmid expressing RBD2.

REFERENCES

1. Fink, A. L. (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid, *Folding Des. 3*, R9−R29.
2. Rochet, J.-C., and Lansbury, P. T. (2000) Amyloid fibrillogenesis: themes and variations, *Curr. Opin. Struct. Biol. 10*, 60−68.
3. Kelly, J. W. (1998) The alternative conformations of amyloidogenic proteins and their multistep assembly pathways, *Curr. Opin. Struct. Biol. 8*, 101−106.
4. London, J., Skrzynia, C., and Goldberg, M. E. (1974) Renaturation of *Escherichia coli* tryptophanase after exposure to 8M urea. Evidence for the existence of nucleation centers, *Eur. J. Biochem. 47*, 409−415.
5. Halverson, K., Fraser, P. E., Kirschner, D. A., and Lansbury, P. T., Jr. (1990) Molecular determinants of amyloid deposition in Alzheimer's Disease: Conformational studies of synthetic β-protein fragments, *Biochemistry 29*, 2639−2644.
6. Silow, M., and Oliveberg, M. (1997) Transient aggregates in protein folding are easily mistaken for folding intermediates, *Proc. Natl. Acad. Sci. U.S.A. 94*, 6084−6086.
7. Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G., and Dobson, C. M. (2000) Mutational analysis of the propensity for amyloid formation by a globular protein, *EMBO J. 19*, 1441−1449.
8. Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates, *Nature 424*, 805−808.
9. Harper, J. D., and Lansbury, P. T. J. (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths

and physiological consequences of the time-dependent solubility of amyloid proteins, *Annu. Rev. Biochem. 66*, 385−407.
10. Pedersen, J. S., Christiansen, G., and Otzen, D. E. (2004) Modulation of S6 fibrillation by unfolding rates and gatekeeper residues, *J. Mol. Biol.*, in press.
11. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G., and Dobson, C. M. (2002) Kinetic partitioning of protein folding and aggregation, *Nat. Struct. Biol. 9*, 137−143.
12. Otzen, D. E., Kristensen, P., and Oliveberg, M. (2000) Designed protein tetramer zipped together with an Alzheimer sequence: a structural clue to amyloid assembly, *Proc. Natl. Acad. Sci. U.S.A. 97*, 9907−9912.
13. Ventura, S., Zurdo, J., Narayanan, S., parreno, M., Mangues, R., Reif, B., Chiti, F., Giannoni, E., Dobson, C. M., Aviles, F. X., and Serrano, L. (2004) Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7258−7263.
14. Iversen, L. L., Mortishire-Smith, R. J., Pollack, S. J., and Shearman, M. S. (1995) The toxicity in vitro of β-amyloid protein, *Biochem. J. 311*, 1−16.
15. Benzinger, T. L. S., Gregory, D. M., Burkoth, T. S., Miller-Auer, H., Lynn, D. G., Botto, R. E., and Meredith, S. C. (1998) Propagating structure of Alzheimer's β-amyloid(10−35) is parallel β-sheet with residues in exact register, *Proc. Natl. Acad. Sci. U.S.A. 95*, 13407−13412.
16. Kirschner, D. A., Inouye, H., Duffy, L. K., Sinclair, A., Lind, M., and Selkoe, D. J. (1987) Synthetic peptide homologous to β protein from Alzheimer disease forms amyloid-like fibrils *in vitro*, *Proc. Natl. Acad. Sci. U.S.A. 84*, 6953−6957.
17. Lansbury, P. T. J., Costa, P. R., Griffiths, J. M., Simon, E. J., Auger, M., Halverson, K. J., Kocisko, D. A., Hendsch, Z. S., Ashburn, T. T., Spencer, R. G. S., Tidor, B., and Griffin, R. G. (1995) Structural model for the β-amyloid fibril based on interstrand alignment of an antiparallel-sheet comprising a C-terminal peptide, *Nat. Struct. Biol. 2*, 990−998.
18. Kheterpal, I., Wetzel, R., and Cook, K. D. (2003) Enhanced correction methods for hydrogen exchange-mass spectrometric studies of amyloid fibrils, *Protein Sci. 12*, 635−643.
19. Jarrett, J. T., Berger, E. P., and Lansbury, P. T., Jr. (1993) The carboxy terminus of the β amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's Disease, *Biochemsitry 32*, 4693−4697.
20. Hilbich, C., Kisters-Woike, B., Reed, J., Masters, C., and Beyreuther, K. (1991) Aggregation and secondary structure of synthetic amyloid βA4 peptides of Alzheimer's disease, *J. Mol. Biol. 218*, 149−163.
21. Zagorski, M. G., and Barrow, C. J. (1992) NMR studies of amyloid β-peptides: proton assignments, secondary structure, and mechanism of an α-helix−β-sheet conversion for a homologous, 28-residue, N-terminal fragment, *Biochemistry 31*, 5621−5631.
22. Gorevic, P. D., Castano, E. M., Sarma, R., and Frangione, B. (1987) Ten to fourteen residue peptides of Alzheimer's Disease protein are sufficient for amyloid fibril formation and its characteristic X-ray diffraction pattern, *Biochem. Biophys. Res. Comm. 147*, 854−862.
23. Volles, M. J., and Lansbury, P. T. (2003) Zeroing in on the pathogenic form of alpha-synuclein and its mechanism of neurotoxicity in Parkinson's Disease, *Biochemistry 42*, 7871−7878.
24. Hardy, J., and Selkoe, D. J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics, *Science 297*, 353−356.
25. Avis, J. M., Allain, F. H., Howe, P. W., Varani, G., Nagai, K., and Neuhaus, D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding, *J. Mol. Biol. 257*, 398−411.
26. Silow, M., and Oliveberg, M. (1997) High-energy channelling in protein folding, *Biochemistry 36*, 7633−7637.
27. Otzen, D. E., Kristensen, O., Proctor, M., and Oliveberg, M. (1999) Structural changes in the transition state of protein folding: an alternative interpretation of curved chevron plots, *Biochemistry 38*, 6499−6511.
28. Nagai, K., Oubridge, C., Jessen, T.-H., Li, J., and Evans, P. R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A, *Nature 348*, 515−520.
29. Lu, J., and Hall, K. B. (1997) Tertiary structure of RBD2 and backbone dynamics of RBD1 and RBD2 of the human U1A protein determined by NMR spectroscopy, *Biochemistry 36*, 10393−10405.
30. Palmer, A. G., Cavanagh, J., Wright, P. E., and Rance, M. (1991) Sensitivity Improvement in Proton-Detected 2-Dimensional Heteronuclear Correlation NMR Spectroscopy, *J. Magn. Reson. 93*, 151−170.
31. Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc. 114*, 10663−10665.
32. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques, *J. Biomol. NMR 4*, 845−858.
33. Shaka, A. J., Barker, P. B., and Freeman, R. (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation, *J. Magn. Reson. 64*, 547−552.
34. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J., and Sykes, B. D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR, *J. Biomol. NMR 6*, 135−140.
35. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation, *Biochemistry 33*, 5984−6003.
36. Kroenke, C. D., Loria, J. P., Lee, L. K., Rance, M., and Palmer, A. G. (1998) Longitudinal and Transverse ^{1}H-^{15}N Dipolar/^{15}N Chemical Shift Anisotropy Relaxation Interference: Unambiguous Determination of Rotational Diffusion Tensors and Chemical Exchange Effects in Biological Macromolecules, *J. Am. Chem. Soc. 120*, 7905−7915.
37. Skelton, N. J., Palmer, A. G., Akke, M., Kordel, J., Rance, M., and Chazin, W. J. (1993) Practical aspects of two-dimensional proton-detected 15N spin relaxation measurements, *J. Magn. Reson., Ser. B 102*, 729−739.
38. Mulqueen, P. M., and Kronman, M. J. (1982) Binding of naphthalene dyes to the N and A conformsers of bovine alpha-lactalbumin, *Arch. Biochem. Biophys. 215*, 28−39.
39. Semisotnov, G. V., Rodinova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas', A. F., and Gilmanshin, R. I. (1991) Study of the "Molten Globule" Intermediate State in Protein Folding by a Hydrophobic Fluorescent Probe, *Biopolymers 31*, 119−128.
40. Lacroix, E., Viguera, A. R., and Serrano, L. (1998) Elucidating the folding problem of alpha-helices: local motifs, long-range elctrostatics, ionic-strength dependence and prediction of NMR parameters, *J. Mol. Biol. 284*, 173−191.
41. Street, A. G., and Mayo, S. L. (1999) Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone, *Proc. Natl. Acad. Sci. U.S.A. 96*, 9074−9076.
42. Roseman, M. A. (1988) Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds, *J. Mol. Biol. 200*, 513−522.
43. Levine, H. I. (1999) Quantification of β-sheet amyloid fibril structures with thioflavin T, *Methods Enzymol. 309*, 274−284.
44. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure, *J. Mol. Biol. 222*, 311−333.
45. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
46. Åkerud, T., Thulin, E., Van Etten, R. L., and Akke, M. (2002) Intramolecular dynamics of low-molecular weight protein tyrosine phosphatase in monomer-dimer equilibrium studied by NMR: a model for changes in dynamics upon target binding, *J. Mol. Biol. 322*, 137−152.
47. Harper, J. D., Wong, S. S., Lieber, C. M., and Lansbury, P. T. J. (1999) Assembly of Aβ amyloid protofibrils: An in vitro model for a possible early event in Alzheimer's disease, *Biochemistry 38*, 8972−8980.
48. Tsai, C.-J., Kumar, S., Ma, B., and Nussinov, R. (1999) Folding funnels, binding funnels and protein function, *Protein Sci. 8*, 1181−1190.
49. Souillac, P. O., Uversky, V. N., Millett, I. S., Khurana, R., Doniach, S., and Fink, A. L. (2002) Elucidation of the molecular mechanism during the early events in immunoglobulin light chain amyloid fibrillation, *J. Biol. Chem. 277*, 12666−12679.
50. Nielsen, L., Khurana, R., Coats, A., Frøkjær, S., Brange, J., Vyas, S., Uversky, V. N., and Fink, A. L. (2001) Effect of environmental

factors on the kinetics of insulin fibril formation: Elucidation of the molecular mechanism, *Biochemistry 40*, 6036−6046.
51. Barrow, C. J., and Zagorski, M. G. (1991) Solution structures of β peptide and its constituent fragments: relation to amyloid deposition, *Science 253*, 179−182.
52. Otzen, D. E., and Oliveberg, M. (2004) Transient formation of nanocrystalline structures during fibrillation of an Alzheimer-like peptide, *Protein Sci. 13*, 1417−1421.
53. Helms, L. R., and Wetzel, R. (1996) Specificity of abnormal assembly in immunoglobulin light chain deposition disease and amyloidosis, *J. Mol. Biol. 257*, 77−86.
54. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat. Struct. Biol. 4*, 10−19.
55. Silow, M., and Oliveberg, O. (2003) High concentrations of viscogens decrease the protein folding rate constant by prematurely collapsing the coil, *J. Mol. Biol. 326*, 263−271.
56. Fandrich, M., Forge, V., Buder, K., Kittler, M., Dobson, C. M., and Diekmann, S. (2003) Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments, *Proc. Natl. Acad. Sci. U.S.A. 100*, 15463−15468.
57. Kranz, J. K., and Hall, K. B. (1999) RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface, *J. Mol. Biol. 285*, 215−231.
58. Kranz, J. F., Lu, J., and Hall, K. B. (1996) Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain, *Protein Sci. 5*, 1567−1583.
59. Jean, J. M., Clerte, C., and Hall, K. B. (1999) Global and local dynamics of the human U1A protein determined by tryptophan fluorescence, *Protein Sci. 8*, 2110−2120.
60. Branden, C. I., and Tooze, J. (1999) *Introduction to Protein Structure*, 2nd ed., Garland, New York.
61. Clarke, J., and Fersht, A. R. (1993) Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation, *Biochemistry 32*, 4322−4329.
62. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure, *J. Appl. Crystallogr. 24*, 946−950.
63. Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering, *Nucleic Acids Res. 16*, 10881−10890.

BI048509K