THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 11, Issue of March 18, pp. 10607–10613, 2005
Printed in U.S.A.
© 2005 by The American Society for Biochemistry and Molecular Biology, Inc.

Investigating the Effects of Mutations on Protein Aggregation in the Cell*[S]

Received for publication, November 16, 2004 and in revised form, December 15, 2004
Published, JBC Papers in Press, December 16, 2004, DOI 10.1074/jbc.M412951200

Giulia Calloni‡, Sara Zoffoli‡, Massimo Stefani‡§, Christopher M. Dobson¶‖, and Fabrizio Chiti‡§**

From the ‡Dipartimento di Scienze Biochimiche, Università Degli Studi di Firenze, Viale Morgagni 50, 50134 Firenze, §Centro di Ricerca, Trasferimento e Alta Formazione MCIDNENT, Italy, and the ¶Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

The conversion of peptides and proteins into highly ordered and intractable aggregates is associated with a range of debilitating human diseases and represents a widespread problem in biotechnology. Protein engineering studies carried out *in vitro* have shown that mutations promote aggregation when they either destabilize the native state of a globular protein or accelerate the conversion of unfolded or partially folded conformations into oligomeric structures. We have extended such studies to investigate protein aggregation *in vivo* where a number of additional factors able to modify dramatically the aggregation behavior of proteins are present. We have expressed, in *Escherichia coli* cells, an *E. coli* protein domain, HypF-N. The results for a range of mutational variants indicate that although mutants with a conformational stability similar to that of the wild-type protein are soluble in the *E. coli* cytosol, variants with single point mutations predicted to destabilize the protein invariably aggregate after expression. We show, however, that aggregation of destabilized variants can be prevented by incorporating multiple mutations designed to reduce the intrinsic propensity of the polypeptide chain to aggregate; in the cases discussed here, this is achieved by an increase in the net charge of the protein. These results suggest that the principles being established to rationalize aggregation behavior *in vitro* have general validity for situations *in vivo* where aggregation has both biotechnological and medical relevance.

The conversion of peptides and proteins into highly insoluble fibrillar aggregates is associated with at least 25 well characterized human disorders, including Alzheimer's and Parkinson's diseases and various systemic amyloidoses (1, 2). Recently, many natural proteins that have no known links to human disease have been found to be able to form aggregates similar to the fibrils associated with clinical amyloid diseases (2, 3). This has led to the proposal that both the amino acid sequences of proteins and the components of the cellular machinery dedicated to "housekeeping" functions have evolved to prevent uncontrolled protein aggregation (3, 4). In addition to its significance in cell biology and medicine, protein aggregation is a fundamental problem in biotechnology. For example, the heterologous expression of proteins in bacteria, the *de novo* design of novel proteins or the rational modification of existing proteins, is frequently frustrated by the fact that the polypeptide chains aggregate into large assemblies, including inclusion bodies or amyloid fibrils (5–7).

It is generally accepted that in most cases aggregates such as amyloid fibrils originate from ensembles of partially unfolded conformations rather than from the folded and functional states of proteins (Fig. 1) (1, 3, 8, 9). Consistent with this idea, amino acid substitutions can promote amyloid formation *in vitro* when they destabilize the native state of a protein (10–18). More recently, it has also been shown from *in vitro* experiments that mutations can favor amyloid formation when they facilitate the subsequent step in the aggregation process of globular proteins (or the primary step for natively unfolded proteins or for unstructured peptides), *i.e.* the conversion of unfolded or partially folded states into oligomeric species (19–26). Aggregation in such cases has been found to be facilitated by mutations that either increase the hydrophobicity of the polypeptide chain or its propensity to convert from α-helical to β-sheet structure or decrease the overall net charge on the protein molecule (19–26).

Although the ability to induce amyloid formation *in vitro* under controlled conditions has greatly facilitated the identification of the underlying physicochemical characteristics that govern protein aggregation, the importance of such determinants in the highly complex and crowded intracellular or extracellular environments *in vivo* has not yet been established. Indeed, many aspects of the environment in a living organism can potentially alter dramatically the aggregation behavior of peptides and proteins. These include the effects of macromolecular crowding (27), the presence of partner proteins or small ligands binding preferentially to the native state of a protein (28–30), and the action of molecular chaperones (31, 32), proteases, and other species (31, 33). In addition, ongoing translation provides *in vivo* a continuous supply of unfolded or partially folded protein that facilitates substantially aggregation of the soluble pool that has already attained the native state (34). A complete understanding of protein aggregation in living organisms therefore requires our current knowledge of the chemistry and physics of protein aggregation to be linked specifically with the manifestations of such phenomena in human pathologies and biotechnological systems.

* This work was supported by grants from the European Union (Project HPRN-CT-2002-00241), the Italian Ministero dell'Istruzione, Università e Ricerca (Projects no. RBAU015B47_001, RBNE01S29H_001, 2003025755_003, 2003054414_002, and L. 449/97, Sector "Genomica Funzionale," project "Strutture ed Interazioni Molecolari di Prodotti Genici"), the Ente Cassa di Risparmio di Firenze (Projects 2003.437 and 2003.2029), and the Wellcome Trust. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains Figs. 5 and 6.

‖ To whom correspondence may be addressed. E-mail: cmd44@cam.ac.uk.
** To whom correspondence may be addressed. E-mail: fchiti@scibio.unifi.it.

This is an Open Access article under the CC BY license.

Petitioner Merck, Ex. 1155, p. 10607

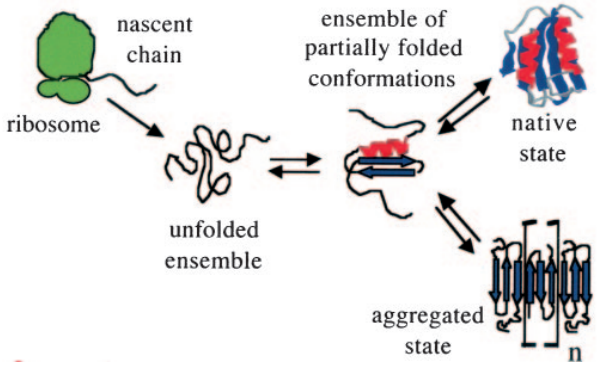

FIG. 1. **Schematic representation of the equilibria existing between different conformational states of a protein in a cell.** As the polypeptide chain is released from the ribosome, the protein is assumed to consist of an ensemble of unstructured conformations. The latter can rapidly collapse into a partially folded state from which either the native conformation or aggregates can form. Both the native state and the early oligomeric aggregates are in equilibrium with the partially folded state. The relative quantities of each that are formed depend on the energetics of these two competing equilibria.

In the present study we have expressed, in *Escherichia coli* cells, the N-terminal domain of the *E. coli* protein HypF (HypF-N). HypF-N has been found to be able to convert *in vitro*, under appropriate conditions, into fibrils that are morphologically and tinctorially indistinguishable from those associated with disease (35, 36). The fibril formation process of HypF-N consists of a number of steps during which pre-fibrillar aggregates precede formation of structured fibrillar species (4, 36). In this paper we describe how the expression of a range of mutants of this protein in the cytosol of *E. coli* cells has provided us with an opportunity to explore the aggregation of a protein within the living organism in which it is naturally expressed. Analysis of a range of mutations and a comparison with a structurally homologous protein with very different aggregation behavior *in vitro* have enabled us to show that a number of important principles that we and others have put forward from experiments *in vitro* are also of general validity *in vivo*.

MATERIALS AND METHODS

Protein Expression and Purification—Protein expression was performed using the glutathione *S*-transferase (GST)[1] fusion system. Cultures of *E. coli* XL-1 Blue cells harboring the pGEX-2T/HypF-N plasmid were grown overnight at 37 °C in LB medium supplemented with 100 μg/ml ampicillin with vigorous shaking. These were then diluted (1:10) in fresh LB medium and grown until the A_{600} reached ~1.0 under the same conditions. Protein expression was induced for 2 h at 37 °C using 0.2 mM isopropyl β-D-thiogalactoside. Cell lysis and protein purification by affinity chromatography (glutathione-agarose, Sigma) was performed as described previously (37). HypF-N cleaved from GST was directly eluted from the column after overnight incubation with bovine thrombin (50 units, Sigma). Protein purity was checked by SDS-PAGE, and the final protein concentration was determined by UV absorption measurements (ϵ_{280} = 1.22 ml mg^{-1} cm^{-1}). Mutated genes of HypF-N were obtained using the QuikChange site-directed mutagenesis kit from Stratagene (La Jolla, CA). The presence of the desired mutation was assessed by sequencing the entire gene. All mutational variants were expressed in *E. coli* XL-1 Blue cells and purified similarly to the wild-type protein. The Protein Data Bank code 1GXT was used for the design of the mutations.

Cloning and Expression of Isolated HypF-N Variants—The genes of wild-type, F22A, L58A, and L84A HypF-N were cloned in the pET-11a plasmid (Novagen, EMD Biosciences Inc., Madison, WI). *E. coli* BL21 cells were transformed with the recombinant vectors and grown with vigorous shaking at 37 °C in LB medium containing 100 μg/ml ampicillin until they reached an absorbance of 0.5 at 600 nm. 3-ml aliquots of the cultures were diluted in 100 ml of fresh LB medium and grown for an additional 4 h under the same conditions of shaking and temperature. No induction was performed in order to maintain low levels of protein expression.

SDS-PAGE—Cell growth and protein expression for XL-1 Blue cells expressing each variant fused to GST, and for BL21 cells expressing some of the variants in their isolated forms, were carried out as described above. Cells were harvested by centrifugation of 40 ml of bacterial growths, resuspended in 4 ml of phosphate-buffered saline, and lysed using 1 mg ml^{-1} hen lysozyme (Sigma) and 5 cycles of sonication for 30 s. 1.5 ml were centrifuged at 14,000 × *g* for 10 min. The pellet fraction was resuspended in 1.5 ml of 1% SDS and boiled for 10 min. Aliquots of 100 μl of both the supernatant and the solubilized pellet were mixed with 50 μl of 3× sample buffer, and the volume resulting from the following formula was applied to the gel: 270 μl/(concentration factor × A_{600} at the end of protein expression), where the concentration factor indicates how many times the cells increase in concentration after harvesting (10 in our experiments). A 15% SDS-PAGE was performed at 200 V and 25 mA per gel. Proteins were visualized by Coomassie Blue staining (0.1% Coomassie Blue, 10% acetic acid, 40% methanol).

Western Blotting—Proteins separated using SDS-PAGE were transferred from the gel onto polyvinylidene difluoride membrane (Immobilon-P, Millipore) for 1 h at 400 mA and 25 V using a miniVE Blot Module (Hoefer). Immunochemical detection of the HypF-N-GST fusion protein was achieved by incubation with serum collected from rabbits immunized with purified HypF-N (antibody production was carried out by Primm s.r.l., Milan, Italy). The serum was diluted (1:1000) in TBS buffer (10 mM Tris-HCl, 100 mM NaCl, pH 7.5) containing 5% nonfat dried milk (bovine), 0.05% Tween 20 and incubated overnight at 4 °C with immunoblots. After washing (0.1% Tween 20 in TBS buffer) the membranes were incubated with secondary anti-rabbit antibodies conjugated with horseradish peroxidase and developed with the enhanced chemiluminescence kit (both from Amersham Biosciences).

Urea Unfolding at Equilibrium—For each purified protein variant, the intrinsic fluorescence of 25–30 pre-equilibrated samples containing 0.02 mg ml^{-1} protein in varying concentrations of urea was measured. All experiments were carried out in 50 mM acetic acid, 2 mM dithiothreitol, pH 5.5, 28 °C. Measurements were made on an LS 55 spectrofluorimeter (PerkinElmer Life Sciences) using excitation and emission wavelengths of 280 and 335 nm, respectively. For each protein variant, the dependence of intrinsic fluorescence on denaturant concentration was analyzed as described by Santoro and Bolen (38) to yield the free energy change of the unfolding reaction in water ($\Delta G_{U\text{-}F}^{H_2O}$), the dependence of ΔG on denaturant concentration (m value), and the midpoint of denaturation (C_m).

Thioflavin T (ThT) Fluorescence—Wild-type and mutated HypF-N were incubated at a concentration of 0.4 mg ml^{-1} in 50 mM acetate buffer, pH 5.5, 30% (v/v) trifluoroethanol, 25 °C, at urea concentrations ranging from 0 to 3 M. After 30 min, 60 μl of each protein sample were mixed with 440 μl of 25 μM ThT, 25 mM phosphate buffer, pH 6.0. The resulting fluorescence was measured using the LS 55 fluorimeter (PerkinElmer Life Sciences). The excitation and emission wavelengths were 440 and 485 nm, respectively.

RESULTS

Many Mutants of HypF-N Aggregate after in Vivo Expression—Wild-type HypF-N and 18 mutational variants having single amino acid substitutions were expressed in *E. coli* cell cultures (Table I). Three groups of variants can be identified from this set: (i) 8 mutants in which an alanine residue replaces a buried hydrophobic residue; (ii) 7 mutants in which an alanine or a glycine residue replaces a proline residue; and (iii) 3 mutants in which a positively charged amino acid replaces a hydrophilic or a negatively charged amino acid exposed to the solvent (Table I). To facilitate purification of the proteins from *E. coli* cell lysates, all variants were initially expressed as fusion proteins in which the N-terminal residue of HypF-N follows the C-terminal residue of GST. Purification of 12 of the HypF-N mutants failed after expression in *E. coli* cells (indicated in Table I with a minus or plus/minus sign). By contrast, wild-type HypF-N and six mutants (indicated in Table I with a *plus* sign) could be purified with reasonable yields (~2 mg per liter of growth medium).

[1] The abbreviations used are: GST, glutathione *S*-transferase; AcP, human muscle acylphosphatase; ThT, thioflavin T.

TABLE I
Conformational stabilities of wild-type HypF-N and 18 mutational variants having single amino acid substitutions

Mutant	Solubility[a]	Experimental $\Delta G_{U\text{-}F}^{H_2O}$ [b]	Predicted $\Delta G_{U\text{-}F}^{H_2O}$ [c]	Experimental $\Delta\Delta G_{U\text{-}F}^{H_2O}$ [d]	Predicted $\Delta\Delta G_{U\text{-}F}^{H_2O}$ [d]
		kJ mol^{-1}	*kJ mol*$^{-1}$	*kJ mol*$^{-1}$	*kJ mol*$^{-1}$
Wild-type	+	29 ± 3			
F22A	−	ND[e]	10.7 ± 3.4	ND	18.3 ± 3.4
V26A	+/−	ND	21.6 ± 3.4	ND	7.4 ± 3.4
L29A	−	ND	17.0 ± 3.4	ND	12.0 ± 3.4
L33A	−	ND	19.6 ± 3.4	ND	9.4 ± 3.4
L58A	−	ND	20.8 ± 3.4	ND	8.2 ± 3.4
I80A	+/−	ND	25.4 ± 3.4	ND	3.6 ± 3.4
L84A	+/−	ND	25.1 ± 3.4	ND	3.9 ± 3.4
F88A	−	ND	8.8 ± 3.4	ND	20.2 ± 3.4
P54A	+	26 ± 3	22.8 ± 3.4	2.8 ± 1.0	6.5 ± 3.4
P66A	−	ND	15.1 ± 3.4	ND	13.9 ± 3.4
P66G	−	ND	8.3 ± 3.4	ND	20.7 ± 3.4
P67A	+	28 ± 3	26.0 ± 3.4	0.9 ± 1.0	3.3 ± 3.4
P78A	+	29 ± 3	22.6 ± 3.4	0.2 ± 1.0	6.7 ± 3.4
P85A	−	ND	17.6 ± 3.4	ND	11.4 ± 3.4
P85G	−	ND	14.4 ± 3.4	ND	14.6 ± 3.4
Q28K	+	31 ± 3	26.5 ± 3.4	−2.2 ± 1.0	2.8 ± 3.4
E55K	+	26 ± 3	31.5 ± 3.4	3.6 ± 1.0	−2.2 ± 3.4
E87K	+	29 ± 3	30.8 ± 3.4	0.7 ± 1.0	−1.5 ± 3.4

[a] Plus, minus, and plus/minus signs indicate HypF-N variants that are fully soluble, fully aggregated, and partially aggregated after expression *in vivo*, respectively. Only mutants with the plus signs could be purified.

[b] Free energy change of the unfolding transition in the absence of denaturant ($\Delta G_{U\text{-}F}^{H_2O}$). Thermodynamic parameters ($\Delta G_{U\text{-}F}^{H_2O}$, m, and C_m) were obtained experimentally for the purified HypF-N proteins from the analysis of urea denaturation curves according to a two-state model (38). In order to reduce the error in $\Delta G_{U\text{-}F}^{H_2O}$, this value was calculated for each variant from the product of the C_m value (midpoint of denaturation) and the average m value (determined from the dependence of ΔG on denaturant concentration), according to the procedure described previously (39). The indicated error values correspond to an estimate of the experimental S.D. (σ) based on a large set of data obtained previously with HypF-N and other proteins. These amount to about 10% of the $\Delta G_{U\text{-}F}^{H_2O}$ values.

[c] $\Delta G_{U\text{-}F}^{H_2O}$ values for the variants that could not be purified were calculated by subtracting the $\Delta\Delta G_{U\text{-}F}^{H_2O}$ values obtained using the Fold-X algorithm (40) from the experimentally determined $\Delta G_{U\text{-}F}^{H_2O}$ value of the wild-type protein. The error corresponds to the S.D. (σ).

[d] Stability differences between mutated and wild-type HypF-N. For each mutant the value indicated was obtained by subtracting $\Delta G_{U\text{-}F}^{H_2O}$ value of the mutant from that of wild-type HypF-N. The error corresponds to the S.D. (σ).

[e] ND, not determined.

To determine the behavior of the HypF-N variants following expression as fusion proteins, in each case we performed an SDS-PAGE analysis of both the supernatant (soluble cytosolic proteins) and the pellet (insoluble proteins) obtained after cell lysis and centrifugation. Fig. 2 shows the Western blots obtained from the SDS-polyacrylamide gels of the bacterial clones expressing the hydrophobic core mutants (the corresponding gels run in parallel without Western blotting are reported in the supplemental Fig. 5). A large band corresponding to the molecular weight of the GST-HypF-N fusion protein (37 kDa) was detected by HypF-N-directed antibodies in the soluble fraction of the *E. coli* cells producing the wild-type protein (Fig. 2). This indicates that wild-type HypF-N remains, to a considerable extent, soluble after expression.

For some variants carrying mutations within the hydrophobic core, an intense band corresponding to the fusion protein can be observed only in the pellet lanes (Fig. 2). These mutants (F22A, L29A, L33A, L58A, and F88A) appear to aggregate completely after expression *in vivo*. For other mutants (V26A, I80A, and L84A), the GST-HypF-N fusion band was clearly visible in both the soluble and insoluble fractions, indicating that significant quantities of these expressed proteins aggregated after expression, whereas another fraction remained soluble. The latter mutants could not be purified because the pool remaining in solution after expression aggregated rapidly during the purification procedure. It was shown that the soluble pool of a protein that initially partitioned between an aggregated and a soluble state continued to migrate to the insoluble form (34). A number of attempts to purify the aggregating variants from inclusion bodies using suitable protocols were performed. All of them failed because of the inexorable aggregation of the mutated proteins during the purification procedure after their solubilization from the inclusion bodies (see supplemental Fig. 6). Overall, this analysis shows three different typologies of behavior of the HypF-N variants: fully soluble (indicated with a *plus* sign in Table I), partially aggregated (indicated with a *plus/minus* sign in Table I), and fully aggregated (indicated with a *minus* sign in Table I) after expression. The difficulty in purifying many of the HypF-N variants arises from the high propensities of the variants involved to aggregate after expression.

HypF-N Mutants Aggregating in Vivo Are Less Stable Than Soluble Variants—The HypF-N variants that could be purified in reasonable yields, including the wild-type protein, were characterized in further detail. The conformational stabilities, after cleavage from the GST protein, were measured by acquiring urea titration curves at equilibrium, using fluorescence spectroscopy to detect the conformational changes of the proteins accompanying denaturation (Fig. 3). A single sharp transition was observed for all variants, and urea denaturation curves were analyzed using a two-state model (38). The values of the free energy change of the unfolding transition ($\Delta G_{U\text{-}F}^{H_2O}$) for all the variants analyzed are listed in Table I. The wild-type protein was found to have a conformational stability of 29 ± 3 kJ mol^{-1}. Most interestingly, all the purified variants have stabilities similar to the wild-type protein, with none displaying a $\Delta G_{U\text{-}F}^{H_2O}$ value lower than 25 ± 3 kJ mol^{-1} (Table I).

The conformational stabilities of the variants that could not be purified because of their aggregation behavior were estimated using Fold-X, an algorithm based on an empirical formula derived from the experimentally determined $\Delta G_{U\text{-}F}^{H_2O}$ values of over 1000 single mutants from different proteins (40). Fold-X has been shown to predict the change in the conformational stability of a protein as a result of specific mutations

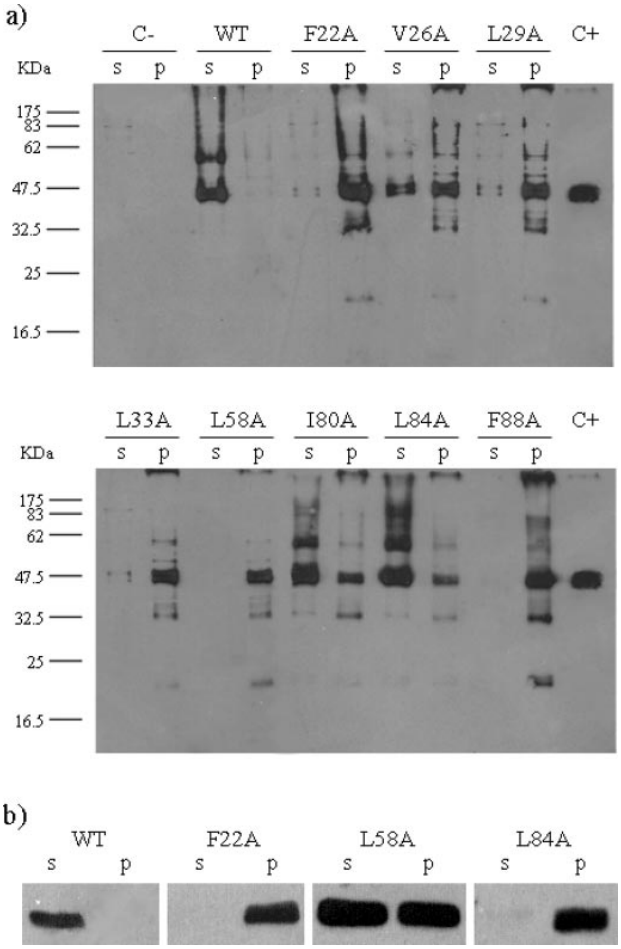

FIG. 2. *a,* Western-blotted SDS-polyacrylamide gels obtained from lysates of *E. coli* cells expressing GST-HypF-N fusion proteins. For each protein variant, the electrophoretic separations of total protein in both the soluble (*s*) and insoluble (*p*) fractions are shown. Cells expressing no fusion protein and expressing wild-type HypF-N were used as controls (*C*− and *WT lanes*, respectively). Purified wild-type HypF-N was also used as a positive control (*C*+ *lanes*). Molecular weight standards are in kDa units. Nonblotted SDS-polyacrylamide gels performed in parallel from the same protein samples are shown in supplemental Fig. 5. *b,* Western-blotted SDS-polyacrylamide gels were obtained from lysates of *E. coli* cells expressing basic levels (noninduced) of HypF-N with no fused GST. The wild-type protein and three representative variants are shown; the symbols *s* and *p* are the same as in *a*.

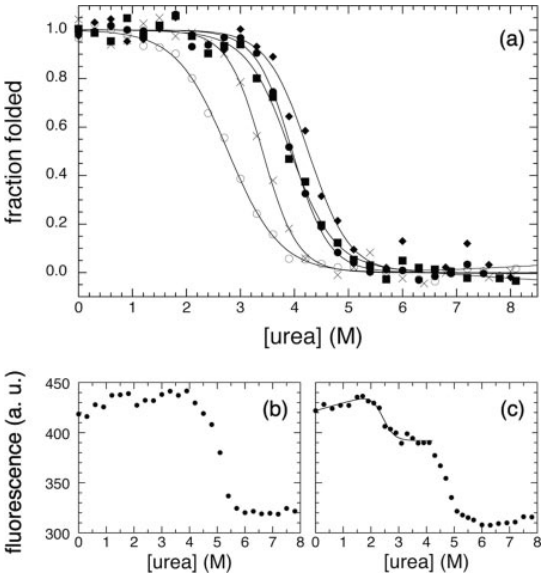

FIG. 3. *a,* representative equilibrium urea-denaturation curves for wild-type (●), Q28K (♦), E55K (×), P78A (■), and the quadruple mutant E55K/V59R/E77R/E87K (○) HypF-N. The curves were obtained at pH 5.5 and at 28 °C by intrinsic fluorescence emission measurements. *Solid lines* represent best fits to the two-state model described by Santoro and Bolen (38). The data points are reported as the fraction of the folded protein present at each denaturant concentration, defined as $(y_{obs} - y_u)/(y_f - y_u)$, where y_f and y_u are the intrinsic fluorescence signals of the folded and unfolded states, respectively, and y_{obs} is the fluorescence signal at the relevant urea concentration; the thermodynamic parameters obtained from the analysis of the curves are reported in Table I. *b* and *c*, equilibrium urea-denaturation curves for GST fused to wild-type HypF-N (*b*) and the quadruple mutant E55K/V59R/E77R/E87K (*c*). The experimental conditions were the same as in *a*.

($\Delta\Delta G_{U\text{-}F}^{H_2O}$) within a standard deviation (σ) of 3.4 kJ mol^{-1} and an error (2σ) of 6.8 kJ mol^{-1} from the real value (40). All of the predicted $\Delta\Delta G_{U\text{-}F}^{H_2O}$ values of the purified variants are similar, within a 2σ value of 6.8 kJ mol^{-1}, to those determined experimentally (Table I). All the nine variants that aggregated completely after expression are predicted to have $\Delta\Delta G_{U\text{-}F}^{H_2O}$ values that are destabilized relative to the wild-type by more than 2σ (6.8 kJ mol^{-1}) (Table I). The three variants that were found to be aggregated only partially after expression were found to be destabilized to a lower degree, ranging from 3.6 to 7.4 kJ mol^{-1} (Table I).

These results are summarized in Fig. 4*a*, which shows a clear distinction between the conformational stabilities measured for the fully soluble variants, the conformational stabilities predicted for the mutants that partition initially between the soluble and aggregated fractions, and those predicted for the fully aggregating variants. Thus, destabilization of the native state of HypF-N as a result of mutation results in a dramatic reduction in the ability of this protein domain to remain soluble following expression in *E. coli* cells. Most interestingly, aggregation does not require a degree of destabilization of the native state to such an extent that the latter is only marginally populated under the conditions present in the *E. coli* cells. Indeed, we observe aggregation *in vivo* of mutants having $\Delta G_{U\text{-}F}^{H_2O}$ values significantly higher than zero, although lower than that of the wild-type protein (Fig. 4*a*).

The Solubility of Destabilized HypF-N Variants Can Be Restored by Decreasing the Aggregation Propensity of Their Partially Unfolded States—In an additional set of experiments, we explored the possibility of producing destabilized variants of HypF-N that can effectively escape aggregation as a result of a reduction in the propensity of the partially unfolded state to aggregate. Because mutations that increase the net charge have been found to reduce the intrinsic tendency of a partially unfolded polypeptide chain to aggregate regardless of the position of the mutated residue in the sequence (23), three variants of HypF-N were produced with substitutions that introduce a substantial change in charge (see legend to Fig. 4*b*). All substitutions involved residues that are highly exposed to the solvent in the native state. At physiological pH (7.4), the quadruple and the two quintuple mutational variants have net charges of +6 and +7, respectively, compared with a net charge of −1 for the wild-type protein. These mutants were indeed found not to aggregate significantly *in vivo* and could be purified in reasonable yield. The experimentally determined $\Delta G_{U\text{-}F}^{H_2O}$ values of these variants are significantly lower than that of the wild-type protein (Fig. 4*b*) and are well below the threshold required for full solubility in the case of the mutations listed in Table I (*cf.* Fig. 4, *a* and *b*). In the presence of 30%

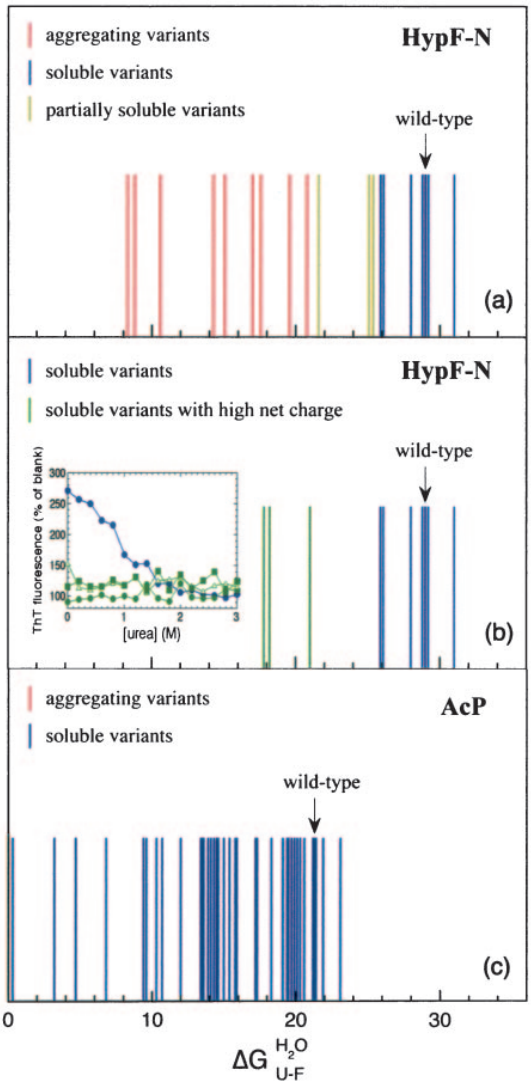

FIG. 4. *a,* conformational stabilities of wild-type HypF-N and 18 variants, each with a single mutation (reported in Table I). *Blue, gold,* and *red bars* indicate mutants that were found to be completely soluble, partially aggregated, or fully aggregated after expression *in vivo*, respectively. *b*, conformational stabilities for wild-type and all soluble single variants (*blue bars*) and three multiple mutants that were soluble despite their relatively low stability (*green bars*). The latter are a quadruple mutant containing the E55K/V59R/E77R/E87K substitutions ($\Delta G_{U\text{-}F}{}^{H_2O} = 21 \pm 3$ kJ mol^{-1}) and two quintuple mutants with the additional T5R or Q32K mutations, respectively (both of which have $\Delta G_{U\text{-}F}{}^{H_2O} = 18 \pm 3$ kJ mol^{-1}). The *inset* shows the ThT fluorescence present after incubation for 30 min under denaturing conditions (30% trifluoroethanol) as a function of the urea concentration for the wild-type protein (*blue circles*) and for the quadruple mutant (*green circles*) and quintuple mutants (*green triangles* and *squares*) HypF-N. *c,* conformational stabilities for wild-type AcP and 42 variants, each with a single mutation. The mutants illustrated in the figure and their conformational stabilities have been described previously (23, 41). *Blue bars* indicate soluble mutants that could be purified. In all three panels $\Delta G_{U\text{-}F}{}^{H_2O}$ errors (corresponding to the standard deviation) are 3.0 and 3.4 kJ mol^{-1} for soluble and aggregated variants, respectively.

(v/v) trifluoroethanol, solvent conditions found to promptly denature HypF-N after a few milliseconds, the three charged variants were found to be soluble after 30 min, whereas the wild-type protein converted into ThT-positive aggregates under these conditions (Fig. 4*b*, *inset*). These results indicate that the three charged variants are substantially more soluble than wild-type HypF-N in their partially unfolded states and can escape aggregation in *E. coli* despite the destabilization of their native states.

Evaluation of the Stabilities and Aggregation Propensities of HypF-N Variants in the GST-fused and Isolated Forms—Wild-type, F22A, L58A, and L84A HypF-N variants were also expressed in *E. coli* as single domains rather than fused to GST (see "Materials and Methods" for details). Expression was maintained at relatively low levels, with no isopropyl β-D-thiogalactoside induction, in order to mimic more closely the physiological situation. Similarly to the corresponding GST-fused proteins, the three destabilized variants were found to aggregate after expression, whereas wild-type HypF-N remained largely soluble (Fig. 2*b*). The presence of GST therefore does not reverse the aggregation behavior of the variants investigated here.

Because of the difficulty in handling the aggregating variants in both their GST-fused and isolated forms, it was not possible to assess whether the GST component affected the stability of the HypF-N domain in these variants. However, we could measure the urea denaturation curves of both the GST-fused and isolated domains for two soluble variants with significantly different conformational stabilities, *i.e.* the wild-type and the quadruple E55K/V59R/E77R/E87K mutant (Fig. 3, *b* and c). The wild-type GST-HypF-N fused protein shows a single transition, indicating that the two domains denature within similar ranges of urea concentrations (Fig. 3*b*). This curve also shows that the presence of the GST component does not confer upon HypF-N a stability that is significantly different from that of the isolated domain (*cf.* Fig. 3, *a* and *b*). The quadruple mutant showed two well defined transitions, the first occurring approximately between 2 and 3 M urea and the second between 4 and 5 M urea (Fig. 3*c*). Analysis of the first transition yields a $\Delta G_{U\text{-}F}{}^{H_2O}$ value of 18 ± 4 kJ mol^{-1}, in good agreement with the value of 21 ± 3 kJ mol^{-1} determined for the isolated quadruple mutant. The mutations therefore appear to have similar effects on the conformational stability of the HypF-N domain whether or not it is covalently attached to GST.

Comparison between Mutants of HypF-N and AcP—A comparison between the present results for HypF-N and the findings of a mutational study of the homologous protein human muscle acylphosphatase (AcP) sheds further light on the importance of the intrinsic aggregation propensity in determining *in vivo* behavior. Wild-type AcP and 42 single mutants have been expressed previously in *E. coli* cells, and were then purified and analyzed using procedures identical to those described here for HypF-N (23, 37, 41). The experimentally determined $\Delta G_{U\text{-}F}{}^{H_2O}$ values range from 0 to 24 kJ mol^{-1}, the wild-type value being 19 ± 1 kJ mol^{-1} (Fig. 4*c*). All 42 mutants appear to be completely soluble within the cells, and all could readily be purified. The mutational variants include those with substitutions of residues both in the hydrophobic core and on the surface and of prolines, *i.e.* very much the same types of mutations that were analyzed in the present work with HypF-N (23, 41). The very different aggregation behavior of the AcP and HypF-N mutants can, however, be rationalized in the light of the fact that the two proteins have aggregation rates that differ by 3 order of magnitudes under conditions in which both are denatured (42). This can be mainly attributed to the difference of hydrophobicity and charge between the two protein sequences (42). Thus, just as three multiple mutations of HypF-N render it soluble in the expression system, AcP can tolerate destabilizing mutations and remain soluble after expression *in vivo* because the partially unfolded state has a low propensity to aggregate.

DISCUSSION

Fig. 1 shows a simplified scheme describing the conformational opportunities for a protein such as HypF-N after biosynthesis in a bacterial cell (3, 43). The protein is synthesized and released from the ribosome as an unfolded polypeptide chain. In some cases, ensembles of partially folded conformations can then be formed rapidly. That this is the case for HypF-N is indicated by recent *in vitro* findings showing that under conditions close to physiological the unfolded state of wild-type HypF-N collapses on the sub-millisecond time scale into a partially folded state from which the folded state forms within ~50 ms (44). The first early event of aggregation of HypF-N, which also occurs from the partially folded state, is also relatively rapid and takes place within 5 s (35, 42). Changes of protein concentrations in an *E. coli* cell as a result of biosynthesis and degradation occur much more slowly, on the time scale of hours (43). This implies that although the total amount of HypF-N within a cell changes with time after induction of expression, the native, unfolded, partially folded, and initial aggregated states are in constant dynamic equilibrium.

Wild-type HypF-N does not accumulate in inclusion bodies after expression, presumably because its rapid folding rate and high native state stability prevent any significant concentration of partially folded protein from being present within the cell. By contrast, destabilization of the native state following mutation increases the population of the partially folded ensemble to an extent that is sufficient to convert effectively the entire pool of the protein into inclusion bodies. Moreover, when the propensity of such a partially folded state to aggregate increases as a result of mutation, considerable aggregation can also take place even if the equilibrium between the partially folded and native states is not altered. Part of the study described here was carried out by expressing HypF-N as a protein domain fused to GST. When the propensity of a mutational variant to aggregate is high enough, it is likely that the GST component is also incorporated in the inclusion bodies. It has been shown for example that the fusion of the green fluorescent protein to a protein of interest does not affect the solubility of the latter (45). More importantly, fusion of proteins of interest with green fluorescent protein is utilized as a means to distinguish soluble and insoluble variants (45). Our control experiments indeed indicate that solubilities of the HypF-N destabilized variants remain low when these variants are expressed without GST.

The results presented here suggest that a relatively high conformational stability of the native structure, in which most of the hydrophobic residues and amide and carboxyl groups are buried or involved in intramolecular interactions, is a very effective strategy to enable a protein to escape aggregation within the cell. The results indicate that, as with aggregation *in vitro*, aggregation occurs in the cell from a partially folded state. Furthermore, self-assembly *in vivo* appears to be promoted by the same relatively nonspecific driving forces as those that act *in vitro*, such as hydrophobic interactions and formation of extensive β-sheet structure. On the other hand, electrostatic repulsions resulting from a high charge on the protein molecules hinder this process. Aggregation of a protein *in vivo* can be prevented, independently of the stability of the native state, by mutations that reduce the intrinsic aggregation rate of the fully or partially unfolded states, for example by increasing its net charge. Such a strategy could be of very great value in the context of the production of proteins in biotechnology.

Our observations show that aggregation of a protein can occur when natural buffering and protective mechanisms are overwhelmed either by a decrease in native state stability or an increase in the intrinsic propensity of the partially unfolded state to aggregate. This conclusion emphasizes the observed correlation between the effects of mutations and the onset of diseases (12, 17, 23, 26). It is a reassuring result in the context of studies of this type that the physicochemical principles observed through experiments involving simplified noncellular environments are so relevant to the systems *in vivo*.

In the familial amyloidotic neuropathy associated with transthyretin, a protein normally existing as a tetramer, correlations have been observed between disease severity and the extent of the destabilization of the native tetramer and folded monomer and the associated acceleration of tetramer dissociation as a consequence of mutation (15, 17). In addition, natural mutations of a range of peptides and proteins that are associated with disease have been shown, in very large majority of cases, to reduce rather than increase the net charge of the polypeptide chain (23). Furthermore, natural mutations associated with familial diseases and involving peptides and proteins that are natively unfolded generally speed up aggregation via one of the three mechanisms mentioned in the Introduction (26). These observations suggest that the mechanisms of action by which natural mutations mediate pathogenesis are related to an increased susceptibility to aggregate the corresponding polypeptide chains either as a result of a destabilization of the native state or an acceleration of the subsequent conversion of the partially or totally unfolded state into aggregates. It therefore appears that, at least in some cases, the physicochemical factors that determine protein aggregation *in vitro* and in the cell also govern rather directly the pathogenesis of the human diseases with which protein deposition is associated.

REFERENCES

1. Selkoe, D. J. (2003) *Nature* **426,** 900–904
2. Stefani, M., and Dobson, C. M. (2003) *J. Mol. Med.* **81,** 678–699
3. Dobson, C. M. (2003) *Nature* **426,** 884–890
4. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M., and Stefani, M. (2002) *Nature* **416,** 507–511
5. West, M. W., Wang, W., Patterson, J., Mancias, J. D., Beasley, J. R., and Hecht, M. H. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11211–11216
6. Carrìo, M. M., and Villaverde, A. (2002) *J. Biotechnol.* **96,** 3–12
7. Schlieker, C., Bukau, B., and Mogk, A. (2002) *J. Biotechnol.* **96,** 13–21
8. Uversky, V. N., and Fink, A. L. (2004) *Biochim. Biophys. Acta* **1698,** 131–153
9. Rochet, J. C., and Lansbury, P. T., Jr. (2000) *Curr. Opin. Struct. Biol.* **10,** 60–68
10. Hurle, M. R., Helms, L. R., Li, L., Chan, W., and Wetzel, R. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 5446–5450
11. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C. F., and Pepys, M. B. (1997) *Nature* **385,** 787–793
12. Stathopulos, P. B., Rumfeldt, J. A., Scholz, G. A., Irani, R. A., Frey, H. E., Hallewell, R. A., Lepock, J. R., and Meiering, E. M. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 7021–7026
13. Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G., and Dobson, C. M. (2000) *EMBO J.* **19,** 1441–1449
14. Ramirez-Alvarado, M., Merkel, J. S., and Regan, L. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 8979–8984
15. Quintas, A., Vaz, D. C., Cardoso, I., Saraiva, M. J., and Brito, R. M. (2001) *J. Biol. Chem.* **276,** 27207–27213
16. Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., Robinson, C. V., and Dobson, C. M. (2002) *Nat. Struct. Biol.* **9,** 308–315
17. Hammarstrom, P., Jiang, X., Hurshman, A. R., Powers, E. T., and Kelly, J. W. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16427–16432
18. Smith, D. P., Jones, S., Serpell, L. C., Sunde, M., and Radford, S. E. (2003) *J. Mol. Biol.* **330,** 943–954
19. Otzen, D. E., Kristensen, O., and Oliveberg, M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 9907–9912
20. Villegas, V., Zurdo, J., Filimonov, V. V., Aviles, F. X., Dobson, C. M., and Serrano, L. (2000) *Protein Sci.* **9,** 1700–1708
21. Giasson, B. I., Murray, I. V., Trojanowski, J. Q., and Lee, V. M. (2001) *J. Biol. Chem.* **276,** 2380–2386
22. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G., and Dobson, C. M. (2002) *Nat. Struct. Biol.* **9,** 137–143
23. Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G., and Dobson, C. M. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16419–16426
24. Lopez De La Paz, M., Goldie, K., Zurdo, J., Lacroix, E., Dobson, C. M., Hoenger, A., and Serrano, L. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16052–16057
25. Tjernberg, L., Hosia, W., Bark, N., Thyberg, J., and Johansson, J. (2002) *J. Biol. Chem.* **277,** 43243–43246
26. Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003) *Nature* **424,** 805–808
27. Minton, A. P. (2000) *Curr. Biol.* **10,** R97–R99

28. Klabunde, T., Petrassi, H. M., Oza, V. B., Raman, P., Kelly, J. W., and Sacchettini, J. C. (2000) *Nat. Struct. Biol.* **7,** 312–321
29. Chiti, F., Taddei, N., Stefani, M., Dobson, C. M., and Ramponi, G. (2001) *Protein Sci.* **10,** 879–886
30. Green, N. S., Palaninathan, S. K., Sacchettini, J. C., and Kelly, J. W. (2003) *J. Am. Chem. Soc.* **125,** 13404–13414
31. Dougan, D. A., Mogk, A., and Bukau, B. (2002) *Cell. Mol. Life Sci.* **59,** 1607–1616
32. Hartl, F. U., and Hayer-Hartl, M. (2002) *Science* **295,** 1852–1858
33. Goldberg, A. L. (2003) *Nature* **426,** 895–899
34. Klein, J., and Dhurjati, P. (1995) *Appl. Environ. Microbiol.* **61,** 1220–1225
35. Chiti, F., Bucciantini, M., Capanni, C., Taddei, N., Dobson, C. M., and Stefani, M. (2001) *Protein Sci.* **10,** 2541–2547
36. Relini, A., Torrassa, S., Rolandi, R., Gliozzi, A., Rosano, C., Canale, C., Bolognesi, M., Plakoutsi, G., Bucciantini, M., Chiti, F., and Stefani, M. (2004) *J. Mol. Biol.* **338,** 943–957
37. Modesti, A., Taddei, N., Bucciantini, M., Stefani, M., Colombini, B., Raugei, G., and Ramponi, G. (1995) *Protein Expression Purif.* **6,** 799–805
38. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry* **27,** 8063–8068
39. Clarke, J., and Fersht, A. R. (1993) *Biochemistry* **32,** 4322–4329
40. Guerois, R., Nielsen, J. E., and Serrano, L. (2002) *J. Mol. Biol.* **320,** 369–387
41. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M., and Dobson, C. M. (1999) *Nat. Struct. Biol.* **6,** 1005–1009
42. Calamai, M., Taddei, N., Stefani, M., Ramponi, G., and Chiti, F. (2003) *Biochemistry* **42,** 15078–15083
43. Hoffmann, F., Posten, C., and Rinas, U. (2001) *Biotechnol. Bioeng.* **72,** 315–322
44. Calloni, G., Taddei, N., Plaxco, K. W., Ramponi, G., Stefani, M., and Chiti, F. (2003) *J. Mol. Biol.* **330,** 577–591
45. Waldo, G. S. (2003) *Curr. Opin. Chem. Biol.* **7,** 33–38

SDS-PAGE for Comassie staining procedure

6 ml of XL-1 bacterial growths expressing each HypF-N variant were collected, lysed and centrifuged as described in the *Materials and methods* section of this work. The soluble (supernatant) and insoluble (pellet) fractions were separated on a 15% SDS-PAGE. After the centrifugation step the supernatant was recovered and the entire protein concentration was determined by the Bradford assay. A supernatant volume containing 20 μg of total protein was mixed with 5 μl of sample buffer containing 3% SDS and 5% β-mercaptoethanol, boiled for 5 min and applied to the gel. The pellet was solubilised in 60 μl of the sample buffer by boiling for 15 min. A volume of 3 μl for each sample was applied to the gel. The electrophoresis was performed at 200 V and 25 mA per gel. Proteins were visualised by Coomassie Blue staining (0.1% Coomassie Blue, 10% acetic acid, 40% methanol).

Figure 5

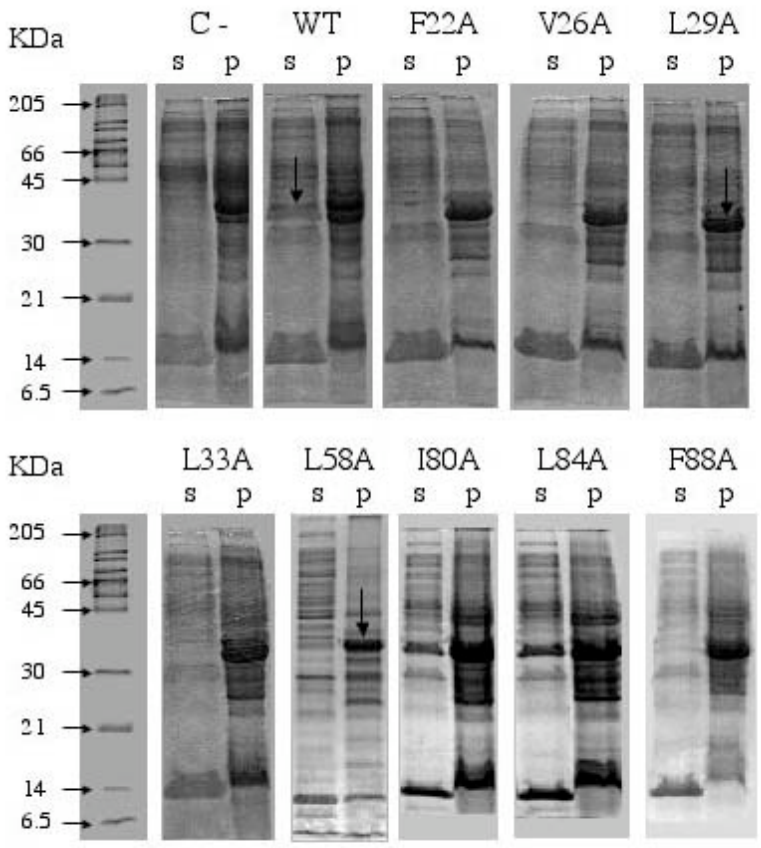

Figure 5, legend

SDS-PAGE analysis of lysates obtained from *E. coli* cells expressing HypF-N mutants carrying single hydrophobic residue substitutions. For each protein variant the electrophoretic separations of total protein in both the soluble (s) and insoluble (p) fractions are shown. Cells expressing no fusion protein and expressing wild-type HypF-N were used as controls (C- and WT lanes, respectively). Molecular weight standards (in kDa units) are shown on the left side of each panel. The band corresponding to the HypF-N-GST fusion protein (37 kDa) is indicated in the cytosolic fraction of cells expressing the wild-type protein, and in the pellet lanes of two representative mutants, by the arrows. This band, which is absent in the negative control and in the pellet lane of the wild-type variant, is clearly visible in the pellet fraction of all mutated forms.

Purification of mutated variants from inclusion bodies

Purification of the aggregating mutants from the inclusion bodies using dedicated protocols (Da-Wei et al, 2004, *Acta Biochim. Biophys. Sinica*, **36**: 141-146) had been tried hard, with the aim of determining the $\Delta\Delta G_{U\text{-}F}^{H_2O}$ values experimentally. While the mutated proteins were easily re-solubilised from the inclusion bodies, they inexorably aggregated again during the purification procedure (Figure 6).

Figure 6

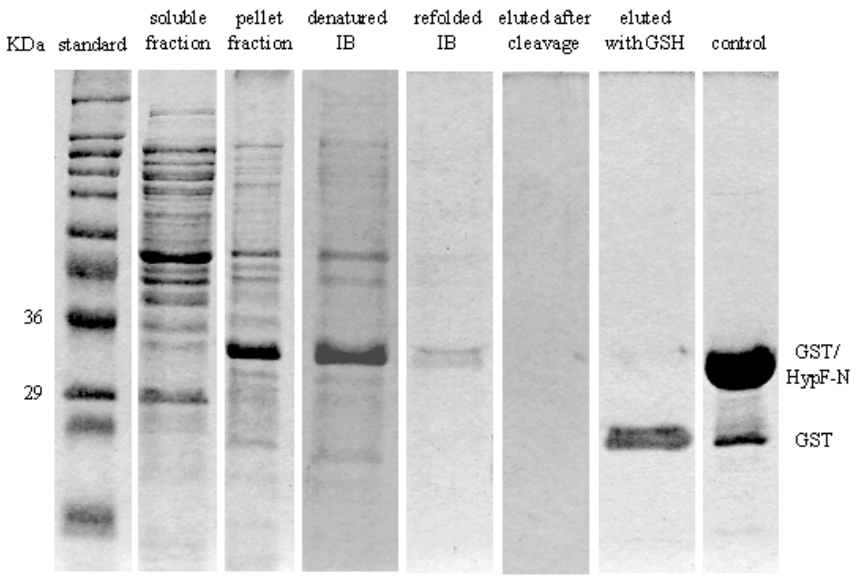

Figure 6, legend

SDS-PAGE analysis of the fractions collected at the various steps involved in the purification from inclusion bodies of the F22A HypF-N mutant.

Soluble and pellet lanes correspond to the supernatant (soluble proteins) and the pellet (insoluble, aggregating proteins) fractions separated by centrifugation of cell lysates.

Denatured IB: pellet fraction solubilised for 24 h in buffer, pH 8.0, containing 8M urea.

Refolded IB: sample obtained after dialysis against phosphate buffer, pH 7.3, and filtration (0.45 μm filter) steps of solubilised IB diluted (1:10) and incubated for 3 d in buffer, pH 8.0, containing 2M urea. This sample, still containing the GST/F22A HypF-N fusion protein, was then applied to the affinity chromatography column.

Eluted after cleavage: fraction eluted after overnight cleavage with bovine thrombin. There is no visible band corresponding to the isolated HypF-N.

Eluted with GSH: sample eluted from the column by a washing step involving a solution containing 5 mM reduced glutathione at the end of the purification procedure. A band corresponding to the isolated GST indicates that binding of the fusion protein to the resin has occurred and suggests that after thrombin cleavage the isolated HypF-N mutant has aggregated in the column.

Control lane shows bands corresponding to purified GST/HypF-N fusion protein and GST alone, respectively.