

Characterization of Glutamine Deamidation in a Long, Repetitive Protein Polymer via Bioconjugate Capillary Electrophoresis

Jong-In Won, Robert J. Meagher, and Annelise E. Barron*

Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois

Received October 31, 2003; Revised Manuscript Received December 12, 2003

We describe a novel method for the determination of glutamine deamidation in a long protein polymer via bioconjugate capillary electrophoresis. Since the current best technique for detection of glutamine (or asparagine) deamidation is mass spectrometry, it is practically impossible to precisely detect the degree of deamidation (i.e., how many residues are deamidated in a polypeptide) in a large protein containing a significant number of glutamine (or asparagine) residues, because the mass difference between native and deamidated residues is just 1 atomic mass unit. However, by covalently attaching polydisperse protein polymers (337 residues) to a monodisperse DNA oligomer (22 bases), the degree of glutamine deamidation, which could not be determined accurately by mass spectrometry, was resolved by free-solution capillary electrophoresis. Electrophoretic separations were carried out after different durations of exposure of the protein to a cyanogen bromide cleavage reaction mixture, which is a general treatment for the purpose of removing an oligopeptide affinity purification tag from fusion proteins. For protein polymers with increasing extents of deamidation, the electromotive force of DNA + polypeptide conjugate molecules increases due to the introduced negative charge of deamidated glutamic acid residues, and consequently CE analysis reveals increasing differences in the electrophoretic mobilities of conjugate molecules, which qualitatively shows the degree of deamidation. Peak analysis of the electropherograms enables quantitative determination of the first four deamidations in a protein polymer. A first-order rate constant of 0.018 h^{-1} was determined for the deamidation of a single glutamine residue in the protein polymer during the cyanogen bromide cleavage reaction.

Introduction

Progress in genetic engineering technology has provided access to various kinds of highly repetitive polypeptides (or “protein polymers”) with well-defined structure, which can serve as analogues to extracellular matrix proteins^{1–4} and spider silk proteins,^{5–8} or as precursors to smectic liquid crystals.^{9,10} Recently, applications of these protein polymers have been expanded to stimuli-responsive silk-elastin-like protein block copolymers for high-throughput protein assays,^{11–13} tunable biopolymers for heavy metal removal,¹⁴ and wheat seed storage proteins.¹⁵ For this reason, we may anticipate that a variety of protein polymers with specific physical properties will be produced for diverse purposes in the future.

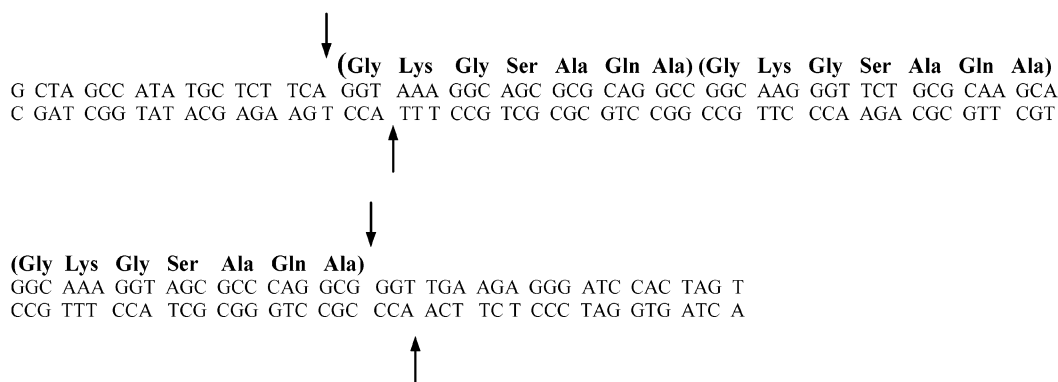
The general procedure for the production of protein polymers is summarized below. Once repetitive DNA concatemers are obtained via the self-ligation of a DNA monomer, they are expressed as protein polymers using a commercial plasmid (e.g., pET plasmids from Novagen or pQE plasmids from Qiagen), which affords a translational

fusion of the target protein with an oligopeptide tag (e.g., polyhistidine) on one terminus of the protein polymer. The presence of this oligopeptide tag on the fusion proteins allows purification of the proteins by immobilized metal affinity chromatography (IMAC), after which this tag is removed from the fusion proteins by treatment with cyanogen bromide^{10,16–19} or with appropriate endoproteinases.

Protein polymers containing glutamine or asparagine residues in their repeating sequences must be used with caution, because these residues can undergo deamidation to form the charged amino acids glutamic acid and aspartic acid.²⁰ The deamidation rate of glutamine and asparagine has been reported to be strongly influenced by the neighboring amino acid residues, pH, temperature, and ionic strength of the buffer system.^{21–25} In particular, if protein polymers that include glutamine or asparagine residues are reacted with cyanogen bromide to remove the oligopeptide tag, they may be unstable because the reaction is generally performed at very low pH (70% formic acid/30% water) and the glutamine or asparagine residues are rapidly deamidated under such acidic conditions.^{21,26}

However, there have been no reports to date about glutamine or asparagine deamidation in a long protein polymer. Previously published studies of deamidation were performed only with small peptides (from 3 residues to 13

* To whom correspondence should be addressed. Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Rm. E136, Evanston IL 60208. Phone: (847) 491-2778. Fax: (847) 491-3728. E-mail: a-barron@northwestern.edu.

Scheme 1. DNA Sequence of a *de Novo* Designed Gene, and the Amino Acid Repeat that the Central Gene Region Encodes

residues).^{25,27,28} Since the current best technique for detection of deamidation in polypeptides is mass spectrometry, it is practically impossible to precisely detect the degree of deamidation (i.e., exactly how many residues are deamidated) in a long polypeptide containing a large number of glutamine or asparagine residues, because the mass difference between native and deamidated residues is just 1 atomic mass unit. Thus, another detection method is needed to ascertain the degree of deamidation in a long protein polymer.

Here, we present a new method for the detection of glutamine deamidation in a long polypeptide (337 residues). By attaching the protein polymer to a monodisperse DNA oligonucleotide and analyzing the electrophoretic mobility of the conjugate molecule by free-solution capillary electrophoresis, we were able to observe the degree of glutamine deamidation in a long protein polymer. This method is similar in principle to end-labeled free-solution electrophoresis (ELFSE) or free-solution conjugate electrophoresis (FSCE)^{29–31} in that polydisperse and water-soluble polymers, which are conjugated to monodisperse DNA, can be separated by free-solution electrophoresis in a microchannel. However, the method described here enables electrophoretic separation of the conjugates according to charge variations, whereas FSCE has previously been demonstrated for size-dependent separation of uncharged polymer samples.

Glutamine is a neutral amino acid, and thus does not impart an electromotive driving force in an electric field. Upon deamidation to glutamic acid, a negative charge is introduced, which introduces an electromotive force toward the anode during electrophoresis in neutral buffer. The magnitude of this force is increased in protein polymers with greater extents of deamidation. Differences in this force result in different and easily distinguishable electrophoretic mobilities for DNA–polypeptide conjugate molecules with different degrees of deamidation. Following capillary-based electrophoretic separation of the conjugate molecules, the degree of glutamine deamidation in the protein polymer is observed, with the extent of the first several deamidations measured quantitatively. We believe that this method can be used for high-resolution characterization of any long, repetitive protein polymer containing glutamine or asparagine residues and potentially for other proteins, if they are analyzed in the denatured, random-coil state, as long as there is an appropriate method for unique conjugation of the protein polymer to a single oligonucleotide.

Materials and Methods

Materials. *E. coli* strain BLR(DE3) and plasmid pET-19b were purchased from Novagen (Madison, WI). *E. coli* strain Top10, plasmid pUC18, and Ni-chelating resin were obtained from Invitrogen, Inc. (Carlsbad, CA). *Pfu* DNA polymerase and restriction endonuclease *Eam*1104 I were purchased from Stratagene (La Jolla, CA) and *Sap* I restriction enzyme was obtained from New England Biolabs (Beverly, MA). *Taq* DNA polymerase, T4 DNA ligase, T4 DNA kinase, and all other restriction enzymes were purchased from Promega (Madison, WI). Synthetic oligonucleotides were supplied by Oligos Etc. (Wilsonville, OR) and all other primers were obtained from the Northwestern University Biotechnology Laboratory (Chicago, IL).

Gene Design, Synthesis, and Multimerization. A single-stranded, synthetic oligonucleotide (102 bases, see Scheme 1), which encodes three tandem repeats of -(Gly-Ala-Gly-Gln-Gly-Ser-Ala)-, was amplified via PCR with two corresponding oligonucleotide primers [5'-ATATAGAATTCC-TCTTCAGGT-3'; 5'-TATATGAATTCCTCTTCTACC-3']. This reaction was performed using an MJ Research DNA Thermal Cycler with 30 cycles at 95 °C for 1 min (denaturing), 55 °C for 1 min (annealing), and 72 °C for 2 min (elongation). A high concentration of dsDNA was obtained from the PCR, and subsequently digested with *Eam*1104 I endonuclease (37 °C, 12 h) to produce DNA monomers with cohesive termini. The 63-bp DNA monomer was then fractionated and recovered via 3% agarose gel electrophoresis onto a diethylaminoethyl (DEAE)–cellulose membrane. The isolated DNA monomer was purified with standard phenol/chloroform extraction and ethanol precipitation. Higher order multimers were constructed by the previously reported strategy developed in our laboratory, which we have termed the “controlled cloning method”,³² and the gene encoding 16 repeats of monomer (1008 bp) was obtained. DNA sequence analyses were performed in our laboratory with a MegaBACE 96-capillary DNA sequencing instrument (Amersham-Pharmacia BioSciences, Sunnyvale, CA).

Protein Expression and Purification. A pUC18 plasmid with an ampicillin resistance gene was used as a cloning vector, and a pET-19b plasmid with ampicillin resistance was used for protein expression. Protein expression was performed with two 2 L Erlenmeyer flasks. Single colonies of *E. coli* expression strain BLR(DE3)(pET-19b) were used to inoculate a seed culture of LB medium (50 mL) containing

both carbenicillin (50 $\mu\text{g}/\text{mL}$) and tetracycline (30 $\mu\text{g}/\text{mL}$) antibiotics. The seed culture was grown to saturation and used to inoculate 2×500 mL of the same medium in two 2 L culture flasks. The main culture was incubated for several hours at 37 °C with agitation at 250 rpm until the OD_{600} reached 0.6–0.8. At that time, IPTG was added to a final concentration of 1 mM to initiate recombinant protein production. Cells were cultured for an additional 3 h and then harvested by centrifugation at 10 000 g and 4 °C for 15 min.

The cell paste was resuspended in 30 mL of the lysis buffer (8 M urea, 20 mM sodium phosphate, 500 mM sodium chloride; pH 7.8). The cells were lysed by several freeze (–80 °C)/thaw (37 °C) cycles and subsequently sonicated at 30 s intervals for 3 min. The lysate was centrifuged at 10 000 g and 4 °C for 30 min to pellet the cell debris, and the supernatant was loaded onto Ni-chelating resin (Probond, Invitrogen, CA). The target proteins were selectively eluted at low pH (pH 4.0) after the endogenous host proteins had been washed off from the resin extensively with higher pH buffers (pH 7.8, pH 6.0, and pH 5.3 buffers). The eluents were analyzed by 12% discontinuous SDS–PAGE. Purified proteins were dialyzed against deionized water (MWCO = 10 000) (Pierce, IL) for 3 days and lyophilized to a dry powder.

Cyanogen Bromide Cleavage Reaction. The fusion polypeptide (total 4 mg) was dissolved in 70% formic acid/water (Aldrich, WI) (final concentration of protein ≈ 1 mg/mL), and the solution was purged with N_2 gas for 3 min to remove dissolved oxygen. Cyanogen bromide (Sigma, MO) was added to the mixture at a final concentration of about 5 mg/mL. The reaction mixture was purged again with N_2 gas for 5 min. The reaction tubes were capped, covered with aluminum foil, and stirred at room temperature. The volatiles, cyanogen bromide, and the *N*-terminal oligopeptide tag were later removed by centrifugal ultrafiltration using Microcon (MWCO = 10 000, Millipore, MA).

MALDI-TOF Analyses. A Voyager-DE PRO mass spectrometer (Analytical Services Laboratory, Northwestern University) was used for MALDI-TOF analysis, and sinapinic acid was used as the matrix.

Preparation of the Oligonucleotide. A 22-base M13 oligonucleotide primer [5'- $\text{X}_1\text{GGX}_2\text{TTTCCCAGTCACGACGTTG}$ -3'] (Oligo, Etc., Wilsonville, OR) was prepared for conjugation to the protein polymer. The DNA oligomers included an Oregon Green fluorescent label on their 4th residue (X_2) from the 5' terminus to allow for laser-induced fluorescence (LIF) detection, and a thiol (–SH) functionality on the 5' terminus (X_1) to allow for covalent conjugation with the maleimide functionality introduced at the amino terminus of the polypeptide. Before conjugation, the thiol functionality of the oligonucleotide was deprotected and reduced, to eliminate any intramolecular dimers formed as a result of oxidation between the thiols, by first dissolving 12.8 nmol of the oligonucleotide in 23 μL of 1 X triethylammonium acetate buffer and 3.45 μL of 1 M AgNO_3 , and incubating at 25 °C for 30 min. Next, 4.6 μL of 1 M dithiothreitol (DTT) was added and reacted for 5 min. The sample was then centrifuged to remove the Ag–DTT

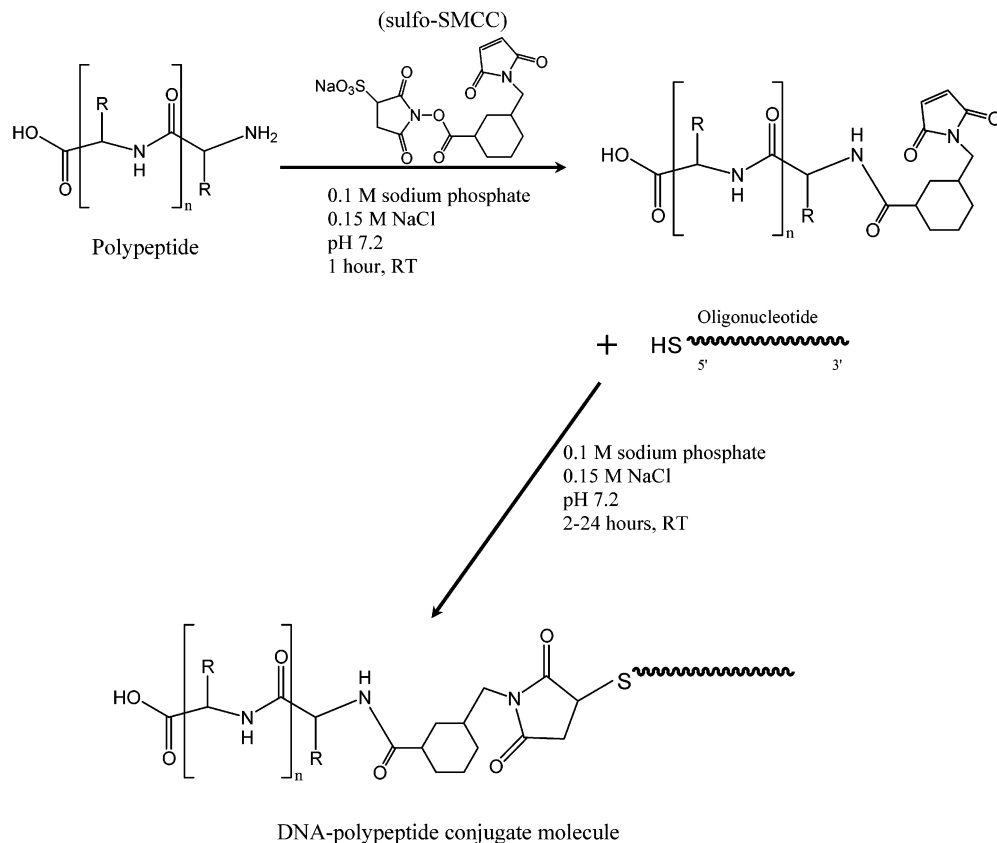
precipitate, and the liquid phase was aspirated and placed in a separate tube. The Ag–DTT precipitate was washed twice with 23 μL of buffer and centrifuged, and all three liquid phases were combined. The liquid containing the DNA was then gel-filtered on Centri-Sep columns (Princeton Separations, Princeton, NJ) to remove any residual DTT and buffer salts. The eluent from these columns was then immediately frozen in liquid nitrogen and stored at –80 °C for later use.

Conjugation of the Protein Polymer to a DNA Oligomer. Conjugation was performed using the heterobifunctional linker sulfosuccinimidyl 4-(*N*-maleimidomethyl)-1-cyclohexane carboxylate (sulfo-SMCC) (Pierce, Rockford, IL) as a coupling reagent.³³ While the *N*-hydroxysulfosuccinimide (S–NHS) ester end of this reagent can react with the primary amine group on proteins to form stable amide bonds, the maleimide end of sulfo-SMCC is specific for coupling to thiolated DNA fragments. The *N*-terminus of the protein was activated by reaction with sulfo-SMCC in an aqueous solution containing 0.1 M sodium phosphate and 0.15 M NaCl, pH 7.2, for 1 h at room temperature with continuous mixing. A protein concentration of roughly 10 mg/mL and a 10-fold molar excess of the sulfo-SMCC were used. The excess sulfo-SMCC was then removed by centrifugal ultrafiltration using Microcon filtration devices (MWCO = 10 000). To complete the conjugation reaction, 5'-thiolated DNA was added directly to the retentate containing the activated protein, which was dissolved in the same pH 7.2 phosphate buffer above and allowed to react for 2–24 h at room temperature. The entire procedure is diagrammed in Scheme 2.

Electrophoretic Analysis of Polypeptide–Oligonucleotide Conjugates. The polypeptide–oligonucleotide conjugates were analyzed via capillary electrophoresis. Capillary electrophoresis separations were conducted in 25- μm inner-diameter fused silica capillaries with various lengths (Polymicro Technologies, Phoenix, AZ) filled with a 1X TTE, 7 M urea buffer (50 mM Tris, 50 mM TAPS, 2 mM EDTA, 7 M urea, pH = 8.4). 7 M urea was used to keep the protein in an unstructured state. The internal surface of the capillary was coated with an adsorbed layer of POP-5 polymer (Applied Biosystems, CA) to reduce electroosmotic flow to negligible levels. All analyses were performed with a BioFocus 3000 Capillary Electrophoresis system equipped with a laser-induced fluorescence (LIF) detector (BioRad, Hercules, CA). Immediately prior to sample injection, the injection end of the capillary was briefly dipped into pure water to remove any residual buffer salts on the outer surface of the capillary. The samples were introduced into the capillary by a pressure injection, with pressure–time constants of 30–60 psia s. Electrophoresis was conducted at 400 V/cm until all peaks had eluted. Detection of the analytes was accomplished by excitation of the fluorescent label using the 488 nm line of an argon-ion laser, with emission detected at 520 nm.

Results and Discussion

Gene Construction and Multimerization. A 102-base synthetic oligonucleotide (ssDNA), which encodes three tandem repeats of GAGQGSA, was designed and PCR-

Scheme 2. Schematic Diagram of the Strategy Employed for Linking Polypeptides to Thiolated DNA, Using Sulfo-SMCC as a Coupling Reagent

amplified with two primers to generate a large amount of dsDNA (Scheme 1). The coding sequence of the gene (shown between the arrows) was flanked by two *Eam*1104 I recognition sites, the cleavage sites of which are indicated by arrows in Scheme 1.

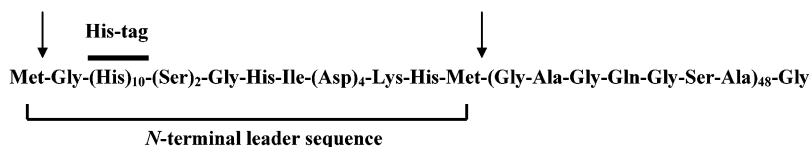
After PCR amplification, the 102-bp gene was digested with the *Eam*1104 I restriction enzyme to generate a 63-bp DNA “monomer”. This DNA fragment was isolated from the reaction mixture by agarose gel electrophoresis onto a diethylaminoethyl (DEAE)-cellulose membrane. The isolated monomers were then multimerized through a particular controlled cloning process, which was described previously,³² and the final gene encoding 16 repeats of the gene was constructed. The forward and reverse sequences of each recombinant plasmid were verified through 3 repeats in both sense and antisense directions using automated DNA sequence analysis by capillary array electrophoresis (data not shown).

Protein Expression and Purification. The protein of interest was produced in *E. coli* strain BLR(DE3) (Novagen, Madison, WI) using a modified pET-19b expression vector (Novagen), which encodes the target proteins as a translational fusion to a 24mer polypeptide that includes an

N-terminal decahistidine sequence. The 10 consecutive histidine residues of the fusion proteins allow their purification by immobilized metal affinity chromatography. Isolation of fusion protein was accomplished under denaturing conditions with a pH gradient, and the purity and molecular weight of the protein were characterized by SDS-PAGE. The purified protein was dialyzed for 3 days to remove salts and then lyophilized. The final fusion polypeptide afforded unoptimized yields of ~10 mg/L. The purified polypeptide was analyzed by reversed phase HPLC to ensure that the protein was pure (data not shown).

Cyanogen Bromide Cleavage Reaction and MALDI-TOF Analyses. Cyanogen bromide (CNBr) is a reagent that is commonly used for cleaving the peptide bond selectively on the carboxyl terminus of a methionine residue.³⁴ Since a methionine residue is located immediately adjacent to the repetitive protein sequence in the expressed fusion protein, and the repetitive sequence does not contain a methionine residue (see Figure 1), cyanogen bromide could be used in this system to remove the *N*-terminal leader oligopeptide.

Eight samples were taken at various time intervals (from 2.5 to 311 h, each with a volume of 350 μ L) during the CNBr reaction, and 150 μ L of dH₂O was added to each sample to

**Figure 1.** Amino acid sequence of an expressed fusion protein, which has an *N*-terminal leader sequence containing a decahistidine tag (arrows represent the cleavage sites of cyanogen bromide).

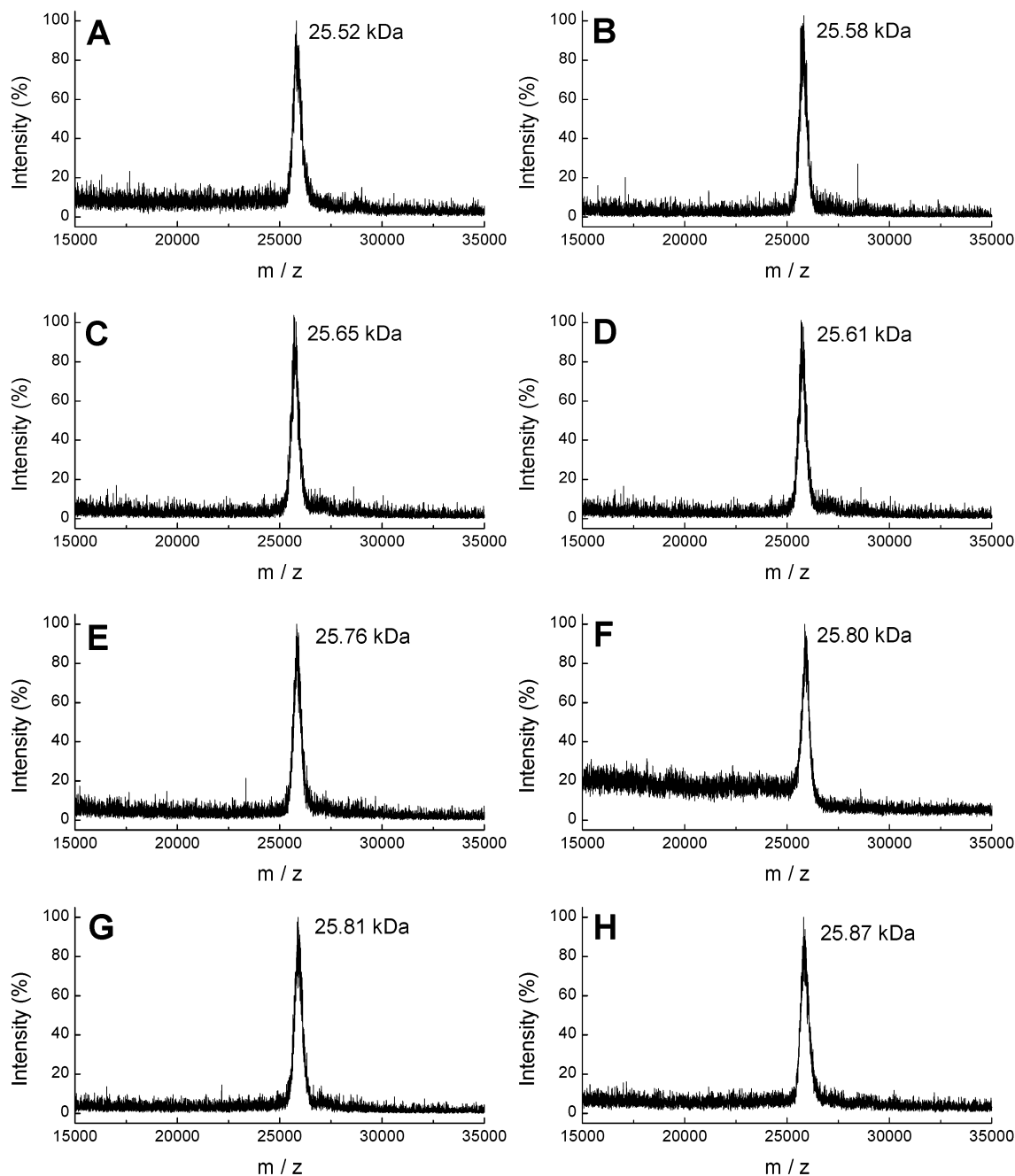


Figure 2. Mass spectra of the protein samples cleaved with CNBr for 2.5 (A), 6 (B), 11 (C), 23 (D), 47 (E), 71 (F), 143 (G), and 311 h (H). The expected mass of each protein is 25.45 kDa, if all glutamine residues are in native form (not deamidated). An FFT filtering algorithm was used to smooth the raw MALDI data, for presentation purposes only.

stabilize the filter membrane during the centrifugal filtration. The volatiles, cyanogen bromide, and the cleaved *N*-terminal oligopeptide fragments were removed by centrifugal ultrafiltration using a Microcon filter (MWCO = 10 000, Millipore, MA) and by several subsequent washes with distilled, deionized water. The final products were identified with MALDI-TOF mass spectrum analysis, as shown in Figure 2. The analytical MALDI-MS results confirmed the purity and identity of each target protein. A nearly constant mass was maintained for all durations of the CNBr cleavage reaction, indicating no significant degradation of the backbone of the protein polymer during the cleavage reaction. Traces of cationic ions binding more tightly to the deamidated glutamic acid residues may be responsible for the slight increase of molecular weight as the CNBr cleavage progresses.

Electrophoretic Analysis of Polypeptide–Oligonucleotide Conjugates. Each purified polypeptide sample was conjugated to the 22-base oligonucleotide using sulfo-SMCC as a coupling reagent. The polypeptides were first “activated” with a maleimide group by using an excess of the bifunctional linker sulfo-SMCC. The excess sulfo-SMCC was then removed by Microcon filtration (MWCO = 10 000), and the retentates were collected. The 22-base M13 primer with a 5′ thiol and an Oregon Green fluorescent label was then added to the activated polypeptides. The resulting conjugates were then diluted in DI water and analyzed by capillary electrophoresis (CE). Electropherograms are shown in Figure 3. The residual, unconjugated DNA peaks elute first around 7.5 min, followed by the elution of DNA–polypeptide conjugate peaks, in order of *decreasing* charge. Two or more

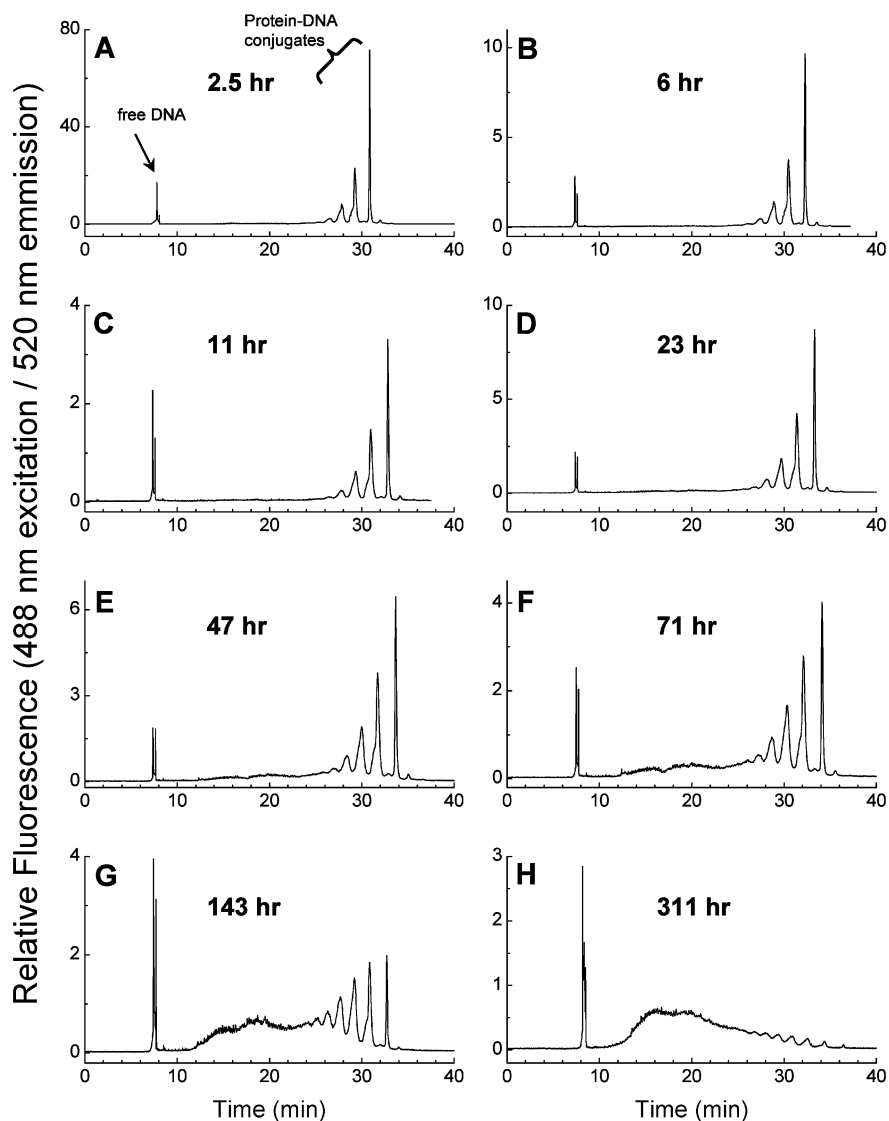


Figure 3. Capillary electropherograms of protein-DNA conjugates, with protein samples cleaved with CNBr for 2.5 (A), 6 (B), 11 (C), 23 (D), 47 (E), 71 (F), 143 (G), and 311 h (H). The fused silica capillary was 25 μm ID \times 44 cm long (39.6 cm to detection window), with the inner surface coated with POP5 polymer. The running buffer was 50 mM Tris, 50 mM TAPS, 2 mM EDTA with 7 M urea, pH 8.4, with 3% v/v POP5 solution added for wall-coating purposes. Injection was hydrodynamic, with a pressure-time constant of 30 psi s. Electric field strength was 17.6 kV (400 V/cm).

peaks are often observed around 7.5 min for the unconjugated DNA; this could represent a combination of free, unconjugated ssDNA, DNA conjugated to sulfo-SMCC with no polypeptide, two strands of ssDNA dimerized end-to-end by disulfide bond formation, or other structures that may result from interaction between the ssDNA and the fluorescent label.³⁵ In Figure 3A, with a 2.5 h CNBr cleavage, three major, well-resolved conjugate peaks can be observed between 27 and 31 min. As the length of the cleavage reaction is extended, numerous other small peaks appear in the region between about 10–27 min, growing in magnitude as the cleavage reaction proceeds. We hypothesize that these peaks represent DNA–polypeptide conjugates with different degrees of glutamine deamidation.

Since glutamine is a neutral amino acid, it does not contribute to the total charge of a DNA–polypeptide conjugate and does not add to the electromotive force on the conjugate during electrophoresis. However, when glutamine is deamidated to form glutamic acid, it acquires a negative

charge in aqueous buffer near neutral pH. This increased charge increases the mobility of the DNA–polypeptide conjugate toward the anode during electrophoresis. The magnitude of the conjugate’s negative charge increases in a stepwise fashion as deamidation progresses. Therefore, DNA–polypeptide conjugates containing one or more deamidated glutamic acid residues should elute earlier than a conjugate having all of the native glutamine residues. The peak distribution of the samples represents quantitatively how many glutamine residues are deamidated in a protein polymer. Note that the final peak (which should be a conjugate having only the native glutamine residues) is always much sharper than other peaks. Because deamidation can happen at any glutamine residue in the protein (with a total of 48 glutamine residues in this protein polymer), polypeptides containing the same number of glutamic acid residues may have slightly different electrophoretic mobilities depending on the locations of the glutamic acid along the backbone. These small mobility differences can cause

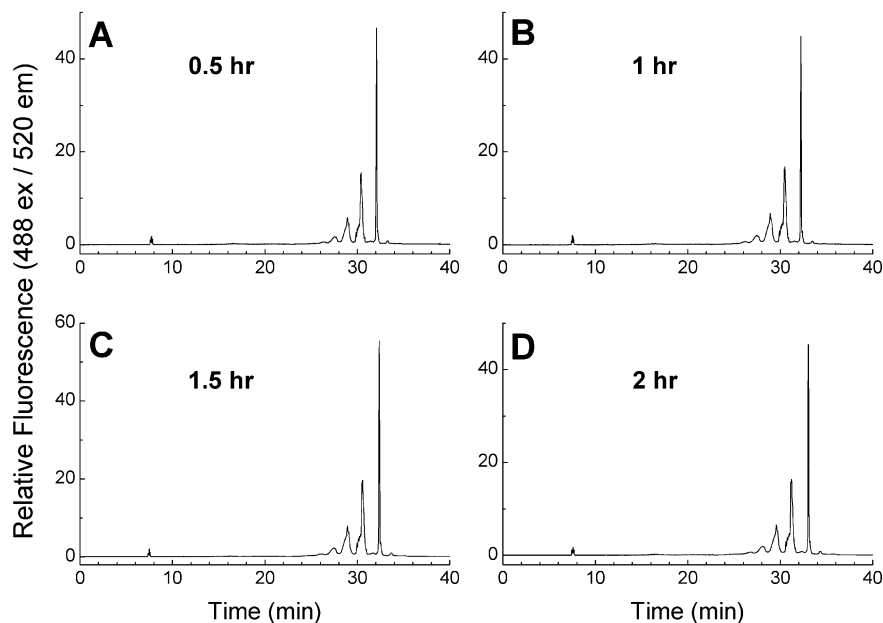


Figure 4. Capillary electropherograms of protein–DNA conjugates, with protein samples cleaved with CNBr for short times: 0.5 (A), 1.0 (B), 1.5 (C), 2.0 h (D). Electrophoresis conditions are the same as in Figure 3.

increased peak broadness and may lead to the inability to resolve individual peaks in the middle region (from about 10–25 min) of the electropherogram.

To investigate deamidation effects during the initial stages of the CNBr cleavage reaction, four aliquots were taken at intervals of 0.5, 1, 1.5, and 2 h. The cleavage reaction was quenched by drying the solution under vacuum, and the volatiles were removed immediately by rotary evaporation. The cleavage products were then resuspended in 500 μ L of native protein purification buffer (50 mM NaH_2PO_4 , 500 mM NaCl, pH 8.0). To separate the target protein (i.e., a protein polymer without a His-tag) [(GAGQGSA)₄₈G] from the residual, uncleaved fusion protein (with His-tag) completely, each protein mixture was mixed with 0.5 mL of Ni-chelating resin (Probond, Invitrogen, CA) and centrifuged at 1000 g for 5 min. Whereas the uncleaved fusion protein binds to the resin strongly due to its histidine residues, the target protein will not bind to the resin because it has no histidine residues. Thus, virtually pure protein could be obtained by collecting the column flow-through solution. The purity and identity of each target protein was confirmed again by MALDI-TOF. After conjugation to DNA using the method above, the conjugate molecules were analyzed by CE. As shown in Figure 4, the peak pattern does not change significantly between 0.5 and 2 h, with almost the same peak pattern as after 2.5 h seen in Figure 3A. This implies that the initial polydispersity does not arise mainly during the short CNBr reaction but may happen during the protein purification, DNA conjugation, or CE operation.

To test the hypothesis that the individual peaks represent conjugate molecules with different degrees of glutamine deamidation, the electrophoretic mobilities of the different conjugate peaks in Figure 3 were compared. In particular, it is useful to consider the ratio of mobilities of adjacent peaks.

The electrophoretic mobility of a DNA–polypeptide conjugate can be thought of in terms of a ratio of total charge q to hydrodynamic friction f . The total charge can be broken

down into components q_{DNA} and q_{P} , representing respectively the charge of the DNA and polypeptide portions of the conjugate molecule:

$$\mu = \frac{q}{f} = \frac{q_{\text{DNA}} + q_{\text{P}}}{f} \quad (1)$$

For the case of a glutamine-containing polypeptide conjugated to monodisperse DNA, the charge q_{P} will vary depending on the number and possibly the location of deamidated glutamine residues. The charge on the DNA, q_{DNA} , is constant. The deamidation of a glutamine to form glutamic acid should not drastically alter the conformation of an unstructured polypeptide (CE analysis was done in 7 M urea solution), and thus, the friction f can also be considered essentially constant. Hence, the ratio of mobility of two conjugate peaks (1) and (2) is given approximately by

$$\frac{\mu_1}{\mu_2} = \frac{q_{\text{DNA}} + q_{\text{P1}}}{q_{\text{DNA}} + q_{\text{P2}}} \quad (2)$$

where q_{P1} and q_{P2} are the charges of the polypeptides in conjugate peaks (1) and (2), respectively.

The electrophoretic mobilities were compared for the final four peaks in each of the electropherograms in Figure 3. These peaks have been designated “0” (final peak), “1” (second from last), “2” (third from last), and “3” (fourth from last). These peak assignments are also illustrated in Figure 5A. The ratio of mobilities μ_N/μ_0 for peaks 1–3 (relative to peak 0) is presented in Table 1. This ratio increases with a relatively constant increment of about 0.06 between the three peaks, meaning that the data can be fit to a form like

$$\frac{\mu_N}{\mu_0} = \frac{q_{\text{DNA}} + q_0 + cN}{q_{\text{DNA}} + q_0} \quad (3)$$

where q_0 is the charge of the polypeptide in conjugate peak

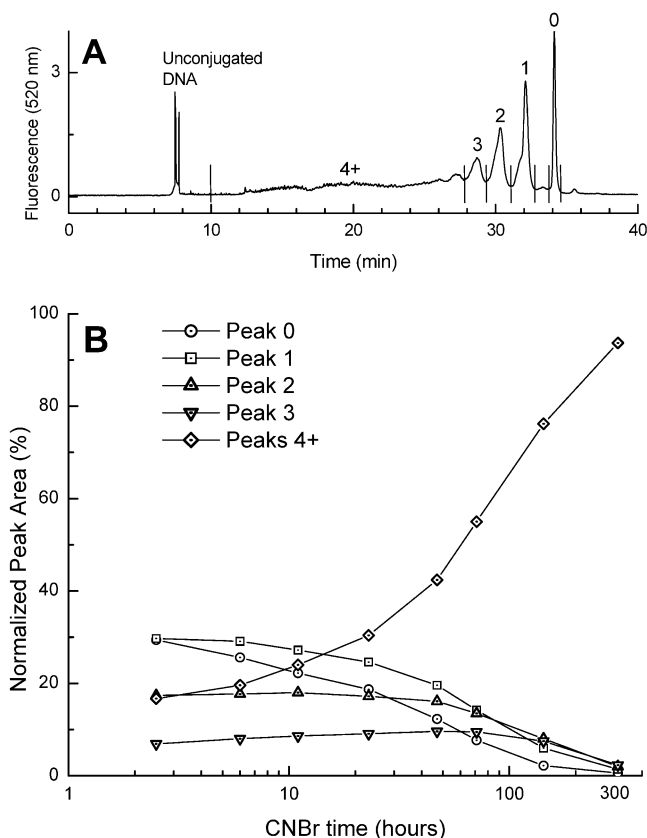


Figure 5. (A) Electropherogram obtained for protein–DNA conjugates, where the protein was exposed to a 71 h CNBr cleavage reaction, illustrating the division of the electropherogram for peak area analysis. The peaks labeled 0, 1, 2, 3, and 4+ are believed to represent DNA conjugated to proteins with 0, 1, 2, 3, and 4 or more glutamine residues deamidated. The vertical lines represent the division of regions for integration. (B) Normalized area of each peak marked in panel A, as a function of CNBr cleavage time. Unconjugated DNA was not included in the area analysis.

Table 1. Electrophoretic Mobilities of the Three Penultimate Peaks in the Electropherograms in Figure 3, Relative to the Final Peak^a

peak (N)	1	2	3
μ_N/μ_0	1.060 ± 0.002	1.118 ± 0.003	1.176 ± 0.008
Increment	0.060	0.059	0.057

^a See Figure 5A for illustration of peak assignment.

0, and c is a constant. In this case, the observed increment between peaks is about 0.06, and so

$$\frac{\mu_N}{\mu_0} \approx 1 + 0.06N \quad (4)$$

Since the DNA used was a 22-base oligonucleotide with an Oregon Green label (which also carries a negative charge), the formal charge of the DNA portion of the conjugate molecules, q_{DNA} , would be -23 at the buffer conditions of pH 8.4. For the native protein with no deamidation, the formal charge would be -1 , due to the C-terminus of the polypeptide. Thus, for the first few deamidations, one might expect mobility increments of $1/24$, $1/25$, and $1/26$, or about 4%, for the first few peaks. The observed increment of 6% per peak is thus of the correct order of magnitude for successive deamidations. The difference between the observed 6% and the expected 4% might be due to different

degrees of ion condensation and charge screening for the densely charged DNA backbone versus the sparsely charged polypeptide.

A similar procedure also allows us to estimate the maximum extent of deamidation observed in the electropherograms. In Figure 3F–H, with long cleavage times, the broad peaks for highly deamidated polypeptides begin to elute around 12 min. The ratio of mobility μ_N/μ_0 for the most highly deamidated species in these figures is about 2.8. Based on eq 4, using the observed spacing of about 6% for the first few deamidations, this mobility ratio corresponds to about 30 deamidations. However, the peak spacing may change at higher charge densities, and thus, the maximum extent of deamidation here is difficult to determine exactly.

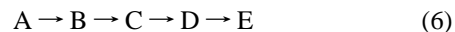
Peak areas in the electropherograms can be used to quantitate the relative concentrations of each species. Each electropherogram was divided into five regions, representing DNA conjugated to protein with 0, 1, 2, 3, and 4 or more glutamines deamidated; this division is illustrated in Figure 5A. Areas were computed for each of these regions, giving an estimate of the relative abundance of each species. Higher degrees of deamidation (4 or more) are lumped together because these peaks are not well-resolved and thus are difficult to quantify reliably.

It is customary in capillary electrophoresis to normalize peak areas by the migration time, to correct for the different spatial velocity of analytes that elute at different times. In this case, since we are lumping together the areas of all of the poorly resolved peaks between the “free” DNA and the final 4 well-resolved peaks, we calculate more formally the normalized peak area between times t_1 and t_2 as

$$\text{normalized area} = \int_{t_1}^{t_2} \frac{s(t)}{t} dt \quad (5)$$

where $s(t)$ is the fluorescence signal at time t . Prior to integration, a simple baseline correction was performed by subtracting the average value of the signal during a five-minute interval prior to elution of the free DNA peak, during which time the baseline was typically flat. Numerical integration of the baseline-corrected data was performed using MATLAB software (The MathWorks, Inc, Natick, MA).

Results of the peak area analysis are shown in Figure 5B. The relative areas of peaks 0 and 1 decline steadily as the cyanogen bromide cleavage progresses. Peaks 2 and 3 increase slightly during the initial stages of the reaction and then also decrease at longer reaction times. The area in the 4+ region increases steadily throughout the cleavage reaction. This behavior is consistent with a process of the form



in which the majority of the material starts out in one state (i.e., no deamidation), and progresses irreversibly through a series of intermediate states (i.e., 1, 2, and 3 deamidations), and ends up in some final state (i.e., highly deamidated). Assuming first-order kinetics, a rate constant for the first step of this process was determined by plotting the natural logarithm of the normalized area for peak 0. The resulting plot was linear from 2.5 to 143 h ($R^2 = 0.998$), and a first-

order rate constant of 0.018 h^{-1} was determined for the deamidation of a single glutamine residue during the CNBr cleavage. These results argue in favor of shorter CNBr cleavage times, or enzymatic methods where applicable, when deamidation of glutamine must be avoided.

The procedure described here was well-suited to the analysis of the relatively simple and well-characterized polypeptide sequence (GAGQGSA)₄₈G. Notably, this sequence lacks lysine, which allows for the straightforward conjugation of the oligonucleotide to the amino terminus of the polypeptide with sulfo-SMCC. For sequences that do contain lysine (which would include most natural proteins), different conjugation chemistry may be necessary to provide unique attachment of the oligonucleotide, for example by reaction of a free cysteine. Other than analysis time, the method does not appear to have an obvious size limit, as different lengths of the (GAGQSA)_nG sequence ranging from 85 to 673 amino acids have been analyzed successfully by this technique. Electrophoresis analysis time for neutral polypeptides increases roughly linearly with polypeptide length. The size of the DNA oligonucleotide can be chosen to speed analysis time for larger polypeptides, although resolution may decrease somewhat if longer oligonucleotides are used. The method should also be applicable to sequences containing charged side chains, although resolution of deamidation products may be more difficult in polypeptides that are already rich in glutamic acid or aspartic acid. Finally, sequences that contain multiple types of reactive side chains, for example, both glutamine and asparagine, could lead to some uncertainty in assigning peaks.

Conclusions

We have presented a new method for analyzing the extent of glutamine deamidation in a long, repetitive protein polymer by a free-solution electrophoretic analysis. By conjugating a monodisperse, fluorescently labeled DNA oligomer to polydisperse protein polymers, protein polymers differing in degree of deamidation were successfully separated by capillary electrophoresis. Since the target protein polymer, with the sequence [(GAGQGSA)₄₈G] in its native form, is nearly electrostatically neutral, it does not migrate in an electric field, and consequently cannot be detected by electrophoresis. However, conjugation to a DNA oligomer imparts the neutral polypeptide with an electrophoretic mobility, thus allowing DNA + protein polymer conjugates to be separated and detected by capillary electrophoresis. Glutamine residues in the protein polymer sequence were deamidated during the harsh cyanogen bromide cleavage reaction used to remove the polyhistidine tag included for protein purification purposes. As the CNBr reaction progresses, the protein polymer becomes more polydisperse due to glutamine deamidation, which results in a distribution of peaks upon analysis by capillary electrophoresis. By comparing the electrophoretic mobilities of the conjugate molecules, low degrees of glutamine deamidation in a protein polymer were determined quantitatively, and a rate constant of 0.018 h^{-1} was measured for the first deamidation event. More

extensive deamidation can be clearly seen qualitatively, although quantitative analysis is more difficult in this case. We believe that this analytical technique can provide a method for the high-resolution characterization of important specialty polymers and proteins, where accurate charge characterization is crucial to the protein's application. The method should be applicable to natural proteins as well, if appropriate methods for end-on conjugation to DNA are available and CE analysis is carried out under denaturing conditions.

Acknowledgment. We acknowledge support from the U.S. Department of Energy, Office of Biological and Environmental Research Grant DE-FG02-99ER6278 and National Institutes of Health (Grant No. NHGRI R01 HG002918-01).

Abbreviations

ssDNA, single-stranded DNA
 dsDNA, double-stranded DNA
 DEAE, diethylaminoethyl
 PCR, polymerase chain reaction
 CE, capillary electrophoresis
 CNBr, cyanogen bromide

References and Notes

- Panitch, A.; Yamaoka, T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1999**, *32*, 1701–1703.
- Welsh, E. R.; Tirrell, D. A. *Biomacromolecules* **2000**, *1*, 23–30.
- Betre, H.; Setton, L. A.; Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2002**, *3*, 910–916.
- Di Zio, K.; Tirrell, D. A. *Macromolecules* **2003**, *36*, 1553–1558.
- Fahnestock, S. R.; Irwin, S. L. *Appl. Microbiol. Biotechnol.* **1997**, *47*, 23–32.
- Winkler, S.; Szela, S.; Avtges, P.; Valluzzi, R.; Kirschner, D. A.; Kaplan, D. *Int. J. Biol. Macromol.* **1999**, *24*, 265–270.
- Zhou, Y.; Wu, S.; Conticello, V. P. *Biomacromolecules* **2001**, *2*, 111–125.
- Arcidiacono, S.; Mello, C.; Kaplan, D.; Cheley, S.; Bayley, H. *Appl. Microbiol. Biotechnol.* **1998**, *49*, 31–38.
- Yu, S. M.; Conticello, V.; Kayser, C.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Nature* **1997**, *389*, 167–170.
- Yu, S. M.; Soto, C. M.; Tirrell, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 6552–6559.
- Nath, N.; Chilkoti, A. *Anal. Chem.* **2003**, *75*, 709–715.
- Nagarsekar, A.; Crissman, J.; Crissman, M.; Ferrari, F.; Cappello, J.; Ghandehari, H. *Biomacromolecules* **2003**, *4*.
- Nagarsekar, A.; Crissman, J.; Crissman, M.; Ferrari, F.; Cappello, J.; Ghandehari, H. *J. Biomed. Mater. Res.* **2002**, *62*, 195–203.
- Kostal, J.; Mulchandani, A.; Chen, W. *Macromolecules* **2001**, *34*, 2257–2261.
- Feeney, K. A.; Tatham, A. S.; Gilbert, S. M.; Fido, R. J.; Halford, N. G.; Shewry, P. R. *Biochim. Biophys. Acta* **2001**, *1546*, 346–355.
- Prince, J. T.; Mcgrath, K. P.; Digirolamo, C. M.; Kaplan, D. L. *Biochemistry* **1995**, *34*, 10879–10885.
- McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727–733.
- Fukushima, Y. *Biopolymers* **1998**, *45*, 269–279.
- McMillan, R. A.; Lee, T. A. T.; Conticello, V. P. *Macromolecules* **1999**, *32*, 3643–3648.
- Creighton, T. E. *Proteins, structures and molecular properties*, 2nd ed.; W. H. Freeman and Company: New York, 1993.
- Scotchler, J. W.; Robinson, A. B. *Anal. Biochem.* **1974**, *59*, 319–322.
- McKerrow, J. H.; Robinson, A. B. *Anal. Biochem.* **1971**, *42*, 565–568.
- Robinson, N. E. *Proc. Natl. Acad. Sci.* **2002**, *99*, 5283–5288.
- Robinson, N. E.; Robinson, A. B. *Proc. Natl. Acad. Sci.* **2001**, *98*, 12409–12413.

- (25) Robinson, N. E.; Robinson, A. B. *Proc. Natl. Acad. Sci.* **2001**, *98*, 944–949.
- (26) Joshi, A. B.; Kirsch, L. E. *J. Pharm. Sci.* **2002**, *91*, 2332–2345.
- (27) Robinson, N. E.; Robinson, A. B.; Merrifield, R. B. *J. Peptide Res.* **2001**, *57*, 483–493.
- (28) Robinson, N. E.; Robinson, A. B. *Proc. Natl. Acad. Sci.* **2001**, *98*, 4367–4372.
- (29) Vreeland, W. N.; Desruisseaux, C.; Karger, A. E.; Drouin, G.; Slater, G. W.; Barron, A. E. *Anal. Chem.* **2001**, *73*, 1795–1803.
- (30) Vreeland, W. N.; Slater, G. W.; Barron, A. E. *Bioconjugate Chem.* **2002**, *13*, 663–670.
- (31) McCormick, L. C.; Slater, G. W.; Karger, A. E.; Vreeland, W. N.; Barron, A. E.; Desruisseaux, C.; Drouin, G. *J. Chromatogr. A* **2001**, *924*, 43–52.
- (32) Won, J.-I.; Barron, A. E. *Macromolecules* **2002**, *35*, 8281–8287.
- (33) Tournier, E. J. M.; Wallach, J.; Blond, P. *Anal. Chim. Acta* **1998**, *361*, 33–44.
- (34) Gross, E. *Methods Enzymol.* **1967**, *11*, 238–255.
- (35) Fabrizio, E. F.; Nadim, A.; Sterling, J. D. *Anal. Chem.* **2003**, *75*, 5012–5021.

BM034442P