This paper describes the syntheses of representative members of a new class of monodisperse, semisynthetic polymers—polymers obtained by perfunctionalization of amine groups on a protein. Monodisperse polymers are centrally important in understanding the relationships between molecular and physical properties of macromolecules. They may be useful where well-defined macromolecular structures are necessary (e.g., polyvalent biological agents), or where physical properties must be tightly controlled (e.g., in molecular weight standards, or as subjects of biological or spectroscopic study). Truly monodisperse polymers are, however, difficult to prepare using the methods of conventional polymer synthesis. We hypothesized that by starting with polymeric backbones that were already monodisperse—proteins—and by modifying their side chains chemically, we might be able to synthesize monodisperse polymers with precisely tailored sequence and side chain functionality.

Proteins have a limited number of functional groups (e.g., Lys-NH$_2$, Asp-COO$_2$H, and Glu-COO$_2$H, Cys-CH$_2$SH, Ser/Thr-OH, Tyr-and side chain functionality. We performed the reactions under pH-controlled conditions using 10 mM SDS present to render all lysine residues accessible to derivatization. We arbitrarily set the concentration of protein at 100$\mu\text{M}$ in all experiments. The Supporting Information summarizes the complete experimental details.

In initial studies, we used acetic anhydride to acetylate the amino groups of three different proteins with activated carboxylic acid groups. We investigated the nucleophilic reaction of lysine side chains of three different proteins with acetic anhydride previously. This procedure provides a practical route, in principle, to a range of monodisperse polyamide-based polymers, having controlled sequence and functional, nonbiological side chains. These polymers may, of course, lose the native structure and function of the protein from which they were derived. In this work, however, we consider proteins only as readily available, truly monodisperse backbones for other, functional polymers.

We synthesized monodisperse polymers with precisely tailored sequence modifying their side chains chemically, we might be able to synthesize monodisperse polymers with precisely tailored sequence and side chain functionality.

Figure 1. CE electropherograms (monitored at $\lambda = 214$ nm; 1/t is directly proportional to electrophoretic mobility) of native, partially acetylated (ladder), and peracetylated (a) ubiquitin, (b) bovine carbonic anhydrase (BCA), and (c) lysozyme. (c) indicates mesityl oxide added to measure electroosmotic flow; (b) indicates impurities in the starting protein. For analysis of lysozyme (a positively charged protein), we used a capillary with positively charged walls: using this capillary, the most negatively charged species eluted first (i.e., at the smallest value of electrophoretic mobility). The peracetylated derivatives of each protein correspond to the most negatively charged peak of the charge ladder. In (c), the last two peaks of the charge ladder are more closely spaced than the other peaks due to the difference in $pK_a$ and thus in the effective charge, between the Lys-NH$_2$ and N-terminal NH$_2$ groups. The $y$-axis in (a-c) has arbitrary units and is not meant to reflect absolute absorbance.

N-terminal amine), lysozyme (EC 3.2.1.17, 14.3 kDa, 6 lysines, 1 N-terminal amine), and bovine carbonic anhydrase II (BCA) (EC 4.2.1.1, 29 kDa, 18 lysines, post-translationally acetylated N-terminal amine). We have studied the partial acetylation of the lysine $\varepsilon$-NH$_2$ residues of these proteins with acetic anhydride previously.

We allowed solutions of ubiquitin, lysozyme, or BCA (in the presence of SDS) to react with acetic anhydride (20 mol equiv per amine group of the protein) and analyzed the resulting mixture of proteins by capillary electrophoresis (CE). We determined the extent of modification (i.e., the number of functionalized amine groups) by comparison of the electropherogram of the perfunctionalized species with a charge ladder derived from that protein. We observed (Figure 1) peracylated derivatives of these proteins as the only products in $\sim$93%, $\sim$90%, and $\sim$90% purity for ubiquitin, lysozyme, and BCA, respectively. Mass spectrometric (MS) analysis of peracetylated ubiquitin indicated that this compound was indeed perfunctionalized on the amines, but that a species...
containing one additional acetyl group was also present. Ubiquitin has a single tyrosine residue, and we hypothesized that the phenolic hydroxyl group of this amino acid might be acetylated. To remove these ester groups selectively, we incubated perfunctionalized ubiquitin (100 μM) in 0.1 M LiOH for 1 h at 4 °C. Dialysis of the hydrolyzed protein solution against water and reanalysis by mass spectrometry gave a single peak, consistent in mass with one acylation per amine in ubiquitin. We subjected all other derivatized proteins to the same de-esterification procedure prior to CE and MS analysis. With ubiquitin and BCA, we achieved >93% recovery of soluble protein after analysis of solutions by UV (λ = 280 nm) before and after reaction with acetic anhydride (see Supporting Information). To demonstrate that this procedure for modification can yield useful quantities of derivatized protein, we converted 1 g of lysozyme to the peracetylated species (90% purity).

To establish conditions that consistently perfunctionalize a protein regardless of the size of or number of reactive residues, we explored the role of protein denaturation on the reaction of the lysine and N-terminal amine groups with electrophiles. We subjected ubiquitin and BCA, both in their native form, and denatured in SDS, to acetic anhydride (20 mol equiv per amine group in the protein) in 0.1 M HEPES buffer (pH = 8.2) and analyzed the products by CE, after de-esterification and dialysis. There was little difference in the reactivity of ubiquitin in native and denatured forms with acetic anhydride. With BCA (Figure 1b), however, we found essentially a single product (>90% purity) corresponding to 18 acetylated lysines when we carried out the reaction in SDS-containing solution, while we found a mixture of partially reacted products when we carried out the reaction in SDS before reaction of the protein with acylating agents; SDS can be removed by dialysis after the modification step. For smaller proteins, SDS may not be necessary.

An efficient route to truly monodisperse polymers with tailored chemical groups may provide access to materials with defined structure (e.g., rigid rodlike molecules that display liquid crystalline behavior or supramolecular organization), with useful mechanical properties (e.g., by analogy with silk), or with applications as adhesives (e.g., by analogy with mussel proteins). Monodisperse polymeric materials may also be useful in fundamental studies in biochemistry, biophysics, and polymer science, or as scaffolds for the polyanion display of ligands and receptors.

Acknowledgment. This work was supported by the NIH (grants GM51559, GM30367, and RR10888). I.G. acknowledges support from an NSF predoctoral fellowship. V.M.K. acknowledges support from NDSEG for a predoctoral research fellowship.

Supporting Information Available: Additional experimental details and mass spectra data (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

Table 1. Electrophiles Used To Perfunctionalize Ubiquitin

<table>
<thead>
<tr>
<th>compound ID</th>
<th>modifying reagent</th>
<th>relative purity (%)</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(CH₃CO)₂O</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CH₂OH(CH₂O)₅CH₂COO⁻(NHS)</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CO(CH₃)₂COOH</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CF₃CH₂COO⁻(NHS)</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CH₃CO⁻(NHS)</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ICH₃CO⁻(NHS)</td>
<td>84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a **NHS** = N-hydroxysuccinimide. b As determined by CE. c See ref 8. d See Supporting Information.

References

(7) A charge ladder is a collection of protein derivatives differing in the number of modified lysine (and N-terminal) amino groups and, thus, in charge. CE separates these derivatives into distinct groups of peaks. See ref 6.
(8) We estimated purity by integration of the major (peracetylated) peak relative to the integrated area of all observable deviations from baseline, assuming equal molar absorptivity coefficients for all species detected; these deviations may arise from partially functionalized products, decomposition, side reactions, aggregated polymeric species, or interaction of the polymers with the walls of the capillary.