Journal of Labelled Compounds and Radiopharmaceuticals—Vol. XXXI, No. 12

AN IMPROVED SYNTHESIS OF α-^{13}C GLYCINE AND HETERONUCLEAR NMR STUDIES OF ITS INCORPORATION INTO THIOREDOXIN

David S. Wishart[1], Brian D. Sykes[1*] and Frederic M. Richards[2]

[1] MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

[2] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, USA 06511

SUMMARY

We present an improved method to easily prepare gram quantities of α-^{13}C glycine beginning from K^{13}CN. The four step synthesis involves the production of an N, N-diphenyl-cyanoformamidine intermediate through the coupling of cyanide to N, N-diphenylcarbodiimide. Subsequent reduction by $LiAlH_4$ and hydrolysis of the resulting amidine produces fully enriched α-^{13}C labelled glycine with a 45-50% yield. This relatively fast and simple synthesis uses only commonly available compounds and requires no special equipment, making the process easy to perform in any well equipped biochemistry laboratory. We further demonstrate that the product may be used, without extensive purification, to specifically label bacterially expressed proteins (E. coli thioredoxin) through standard biosynthetic procedures. We also show that the ^{13}C glycine-labelled protein may be readily analyzed using commonly available heteronuclear NMR techniques. Complete assignments for all 9 glycines of native E. coli thioredoxin are presented.

Key Words: α-^{13}C Glycine, K^{13}CN, NMR, Thioredoxin, Protein, Isotopic labeling

INTRODUCTION

Recent advances in NMR spectroscopy along with improvements in protein biotechnology have amply demonstrated how isotopically enriched amino acids can greatly simplify the spectra of proteins in solution (1, 2, 3). This aspect of spectral simplification is particularly important in cases where there are large numbers of potentially degenerate resonances that are known to exist, such as in denatured or unfolded proteins. Work in our laboratory is currently proceeding on studies of the denatured states of a small protein (108 residues) known as E. coli thioredoxin. We are using ^{13}C NMR spectroscopy of selectively labelled protein samples with the hope that this technique will to allow us to overcome the problems of limited resolution, spectral degeneracy and assignment difficulties which so regularly plague studies of this kind (4, 5).

* Author to whom correspondence should be addressed

0362–4803/92/121019–10$10.00

Received 14 July, 1992
Revised 16 September, 1992

In anticipation of the large numbers of uniquely labelled protein samples required for this project, both the cost and the restricted availability of some commercially obtainable ^{13}C amino acids have had to be addressed. Consideration of many possible alternatives led us to abandon the commercial route and, instead, to adopt an "in house" synthetic strategy based on a single very useful precursor for naturally occurring amino acids: KCN. Although not widely known, it is theoretically possible to prepare either 1-^{13}C (using the Strecker synthesis) or 2-^{13}C labelled glycine (6, 7) using only K^{13}CN as the starting material. Furthermore, with the labelled glycine in hand, it is also possible to prepare a host of other 1-C or 2-C labelled amino acids using, for instance, the phenylhydantoin method of Finkbeiner (8). While other procedures certainly exist for the preparation of ^{13}C labelled glycine (9, 10) along with many other amino acids, few of these methods offer the potential flexibility that KCN does for labeling at either the 1 or 2 positions.

Our initial optimism about using K^{13}CN to prepare α-^{13}C labelled glycine was based on work first reported by Ehrensvard and Stjernholm in 1949 (6). However, in our hands, this presumably simple and high yielding procedure produced no more than trace quantities of the desired product. In this report we wish to describe some very useful improvements we have made to the original Ehrensvard and Stjernholm (E & S) synthesis which have allowed us to consistently prepare gram quantities of α-^{13}C glycine with 45-50% yields (based on K^{13}CN). We also will demonstrate how the labelled product can be biosynthetically incorporated into thioredoxin and how the selectively labelled protein may be easily assigned using standard ^{1}H-^{13}C HMQC techniques.

RESULTS AND DISCUSSION

The fact that small quantities of glycine were occasionally produced in the E & S synthesis suggested that the principles behind the synthesis could prove to be quite tenable. Consequently, the two most problematic steps (steps 1 and 2 -- See Figure 1) were revised.

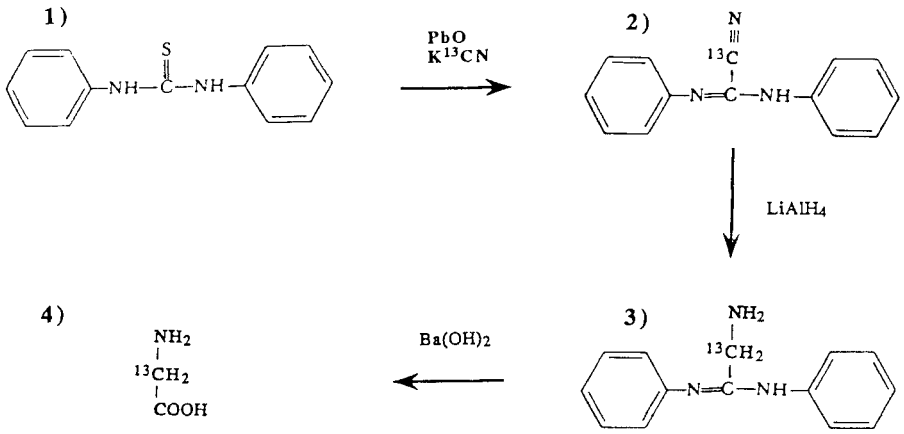

Figure 1. Summary of the Ehrensvard and Stjernholm synthesis.
1) N, N-diphenylthiourea 2) N, N-diphenylcyanoformamidine
3) Glycine-N, N-diphenylamidine 4) Glycine

Step 1 often resulted in the creation of a dirty grey sludge of almost uncharacterizable products. Particularly troublesome was the large quantity of dense, black lead sulfide which often combined with the resulting oils to yield an insoluble "plastic".

To avoid this problem we turned to carbodiimides instead of thiocarbanilides as coupling agents for the cyanide ions. Work by Khorana (12) and others had previously demonstrated a high level of reactivity of nitriles and other electronegative constituents to a wide variety of carbodiimides. This suggested that reacting KCN with a pure carbodiimide such as diphenylcarbodiimide under moderately basic conditions could produce the desired product (N, N - diphenylcyanoformamidine) in high yields. As a result, a two step reaction based on, first, preparing the carbodiimide from diphenylthiourea and then reacting the carbodiimide with cyanide was developed and implemented as a replacement for the first step of the E & S synthesis. The resulting yields of N, N-diphenylcyanoformamidine were 90-95%.

Several attempts were made to eliminate the need to prepare diphenylcarbodiimide (which is not commercially available) by using commercially available carbodiimides such as N, N-dicyclohexylcarbodiimide or 1-cyclohexyl-3-(morpholinyl-(4)-ethyl) carbodiimide in the nitrile coupling reaction. Interestingly, none of the cyanoformamidines prepared from these aliphatic carbodiimides were efficiently hydrolyzed to glycine. Evidently the aromaticity of the N, N-diphenylcarbodiimide is critical to the success of both the reduction and subsequent hydrolysis steps in the E & S synthesis.

In addition to the above changes it was also found that a revision of the $LiAlH_4$ reduction procedure (step 2) was required in order to prevent a pernicious white "gum" from developing after the initial H_2O quenching step. This sticky aluminum hydroxide precipitate often prevented any kind of efficient extraction or purification of the desired product and no doubt contributed to the very low yields in our early attempts using the old E & S protocol. A solution to this was eventually found through the use of a multi-step quenching protocol.

The improved E & S synthesis is depicted in Figure 2 and is described in more detail in the Experimental Section of this paper. This new protocol is somewhat more complex than Ehrensvard and Stjernholm's original proposal although it does offer several advantages. Specifically it is a much "cleaner" synthesis wherein almost all of the products are found to be translucent oils that are easily extracted and purified. Furthermore, the yields are high for nearly all products (80 - 90%) with the notable exception of the final $Ba(OH)_2$ hydrolysis step. As mentioned previously, yields for the final product, glycine, are generally between 45 - 50% (based on $K^{13}CN$). The 75 MHz ^{13}C spectrum of a sample of α-^{13}C glycine prepared from this new method is shown in Figure 3. Attempts to improve the yield of the final hydrolytic step with other bases (NaOH, KOH, etc.) have so far been unsuccessful.

With a simple and efficient method now available for the production of α-^{13}C labelled glycine from $K^{13}CN$, various combinations of labelled glycine can now be produced using KCN as the sole precursor, for instance, $^+H_3{}^{15}N^{13}CH_2COO^-$ from $K^{13}C^{15}N$ or $^+H_3{}^{15}N^{13}CH_2COO^-$ from $KC^{15}N$. In addition, 1-^{13}C glycine as well as

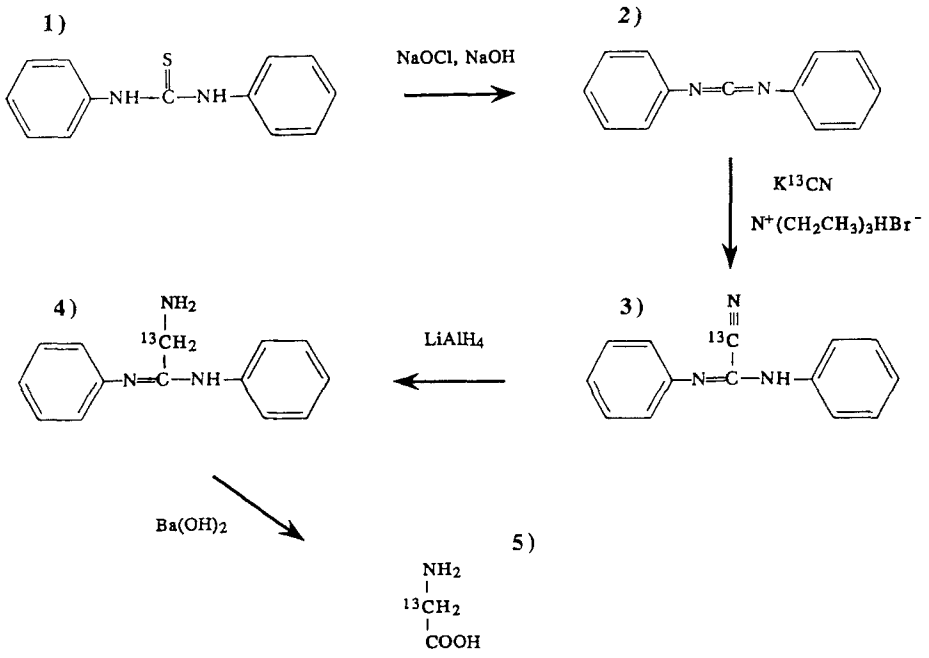

Figure 2. Summary of the carbodiimide-based synthesis developed for this work. 1) N, N-diphenylthiourea 2) N, N-diphenylcarbodiimide 3) N, N-diphenyl-cyanoformamidine 4) Glycine-N, N-diphenylamidine 5) Glycine

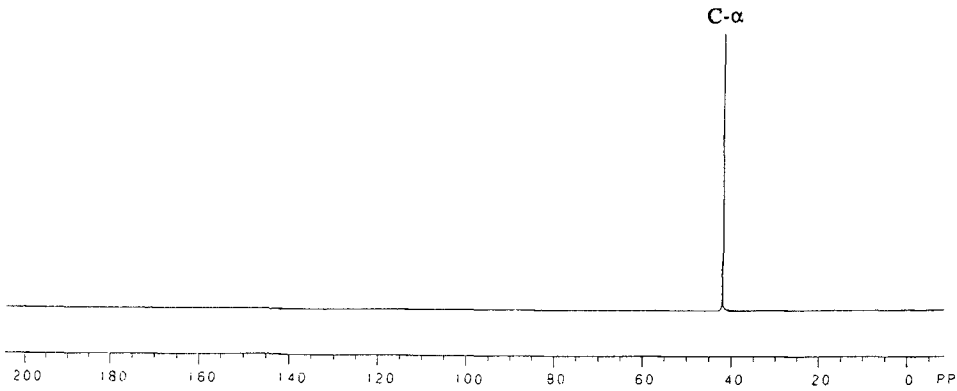

Figure 3. 75.4 MHz ^{13}C NMR spectrum of a 20 mM solution of α-^{13}C glycine.

1-^{13}C, 2-^{13}C glycine can be made using the standard Strecker synthesis (13). The fact that $K^{13}CN$, $KC^{15}N$ or $K^{13}C^{15}N$ can now be used to prepare all possible permutations of ^{13}C or ^{15}N labelled glycine, combined with the fact that glycine can be used as a precursor to prepare most naturally occurring amino acids (8), means that selective isotopic enrichment

of amino acids could become both simpler and easier than commonly imagined. We are hopeful that others may find this potentially generalizable synthetic strategy useful in their own work.

In addition to our interest in developing a general and efficient strategy for preparing labelled amino acids, we have also been interested in determining how well these labelled amino acids could be incorporated into proteins. To this end, we conducted two parallel cell growth experiments, one using commercially available, unlabelled glycine in our fermentation medium and the other using crude α-^{13}C glycine prepared from simply lyophilizing the supernatant derived from the $Ba(OH)_2$ hydrolysis. Despite the concerns of barium contamination or the presence of other toxic substances, we observed no reduction in wet cell weight or in the expected yield of the labelled protein. This result seems to indicate that α-^{13}C glycine prepared in this manner has no adverse effect on either cell growth or biosynthetic incorporation efficiency. Extensive purification of the amino acid, therefore, appears to be unnecessary. Figure 4 shows the 75 MHz ^{13}C NMR spectrum of a sample of E. coli thioredoxin prepared from one of these labeling experiments. There are nine well-defined peaks arising from the nine glycine residues known to exist in this protein. There is little evidence of isotopic scrambling. Figure 5 illustrates a 500 MHz ^{1}H-^{13}C HMQC spectrum collected on the same sample with all nine resonances appropriately labelled. Table 1, below, provides the chemical shift data for each of the glycine residues. Proton chemical shifts used to identify the ^{13}C resonances were taken from (1).

Table 1

^{1}H and ^{13}C glycine chemical shifts for E. coli thioredoxin at pH 5.0, 25 °C.

Residue	^{1}H shift (ppm)	α-^{13}C shift (ppm)
Gly 21	3.99, 3.99	42.2
Gly 33	4.04, 4.32	46.6
Gly 51	3.73, 4.40	43.3
Gly 65	3.65, 3.98	43.9
Gly 71	3.75, 3.80	44.5
Gly 74	3.76, 4.25	42.4
Gly 84	3.75, 4.37	44.1
Gly 92	3.63, 4.37	41.5
Gly 97	3.78, 3.92	44.9

The results from these studies clearly indicate that α-^{13}C glycine prepared from the method described here can be biosynthetically incorporated into proteins without the need for extensive purification.

To conclude, we have succeeded in developing an improved method to synthesize α-^{13}C glycine using K^{13}CN as a precursor. The success and simplicity of the improved protocol indicates that this synthesis can readily be performed using standard laboratory

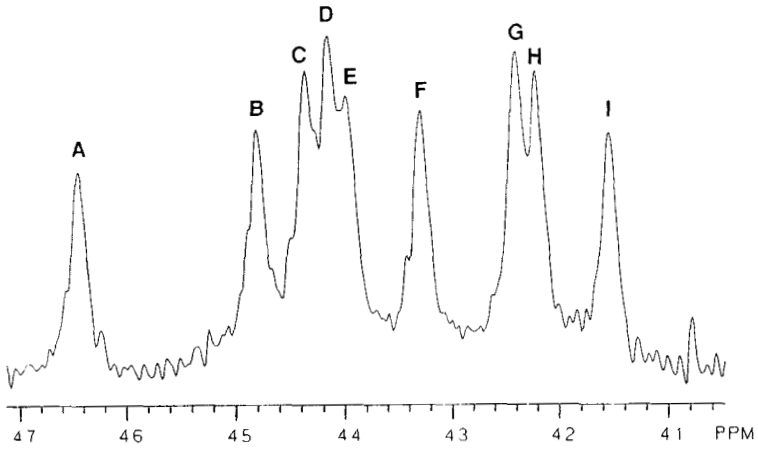

Figure 4. 75.4 MHz ^{13}C NMR spectrum of a 0.4 mM solution of α-^{13}C glycine labelled thioredoxin.

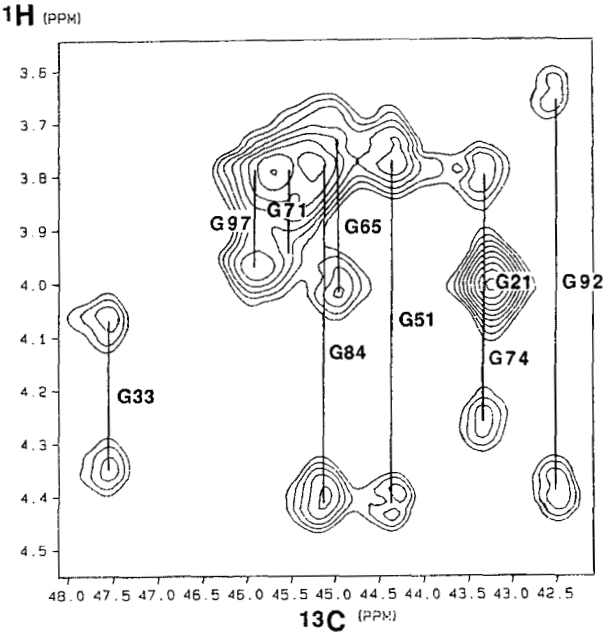

Figure 5. 500 MHz ^{1}H-^{13}C HMQC spectrum of α-^{13}C glycine labelled thioredoxin.

equipment. We believe that this improved synthesis, in combination with the standard Strecker synthesis, could allow isotopically labelled cyanide to serve as an excellent general precursor for a large variety of amino acid syntheses. This work represents a first step in what we hope will be a successful approach in combining novel chemical strategies and ^{13}C NMR spectroscopy to study the denatured states of proteins.

EXPERIMENTAL SECTION

Diphenylthiourea (thiocarbanilide) and barium hydroxide octahydrate were purchased from Aldrich. $K^{13}CN$ was a generous gift from the Stable Isotope Division of Los Alamos Laboratories.

N, N - Diphenylcarbodiimide

17.6 grams (0.075 mol) of diphenylthiourea (thiocarbanilide) were dissolved in 200 mL of benzene in a 2L flask. A solution of basic aqueous NaOCl was prepared by dissolving 3 g of NaOH (0.075 mol) into 220 mL (0.16 mol) of commercial grade Clorox and 200 mL of water. The Clorox solution was slowly added (over a period of 10 minutes) to the thiocarbanilide/benzene solution, thereby producing a two phase system. The mixture was stirred vigorously at room temperature for one hour. Upon completion of the reaction the aqueous phase was removed and discarded using a 1000 mL separator funnel. The remaining organic phase was filtered to eliminate particulate matter and unreacted material. This organic component was rotary evaporated to remove the remaining benzene so as to yield a pale yellow oil. This oil was further purified by passing it over a short (4 cm x 20 cm) column containing 100 - 130 g of silica. The desired product was eluted and collected with 250 - 300mL of a 1:1 mixture of chloroform and hexane. The diphenylcarbodiimide was found to be the only mobile component under these particular elution conditions. The resulting filtrate was rotary evaporated at room temperature to yield between 8 - 9 g (yield = 55 - 60%) of a clear viscous oil. This was identified through IR and NMR as N, N - diphenylcarbodiimide. TLC of the product on fluorescent coated silica gel plates (Merck, 60F) using 1:1 chloroform/hexane as the solvent indicated an R_f of 0.80. The product can be stored under refrigerated conditions for about two weeks.

N, N - Diphenylcyanoformamidine

3.6 g (0.02 mol) of triethylammonium bromide* was dissolved in 150 mL of methanol in an externally cooled (ice bath at 0 °C) round bottom flask (500 mL). To this solution 1.32 g (0.02 mol) of $K^{13}CN$ was slowly added and permitted to completely dissolve (about one hour). After thorough dissolution of the $K^{13}CN$, a solution containing 4.2 g (0.021 mol) of N, N - diphenylcarbodiimide in 40 mL of THF was slowly added over a period of 10 - 15 minutes. The material was allowed to stir for an additional 30 minutes after which the product was concentrated by rotary evaporation at room temperature. Addition of ether to the resulting oil precipitated the KBr salts, which were removed by filtration. In general, no further purification of the product was undertaken before proceeding to the next step. Yields were typically between 4.3 - 4.5 g or about 90 - 95% of the expected result. TLC in a 1:1:1 mixture of hexanes, ether and chloroform yields an R_f of about 0.60.

* The triethylammonium bromide used in this part of the synthesis was prepared as follows: To a solution of triethylamine (60 mL) and water (30 mL) was added a total of 65 mL of concentrated (9 M) HBr. This was stirred for approximately 30 minutes to mix both

layers. The mixture was then rotary evaporated to dryness. To this material was added a small amount of ethanol to create a dark brown slurry. This slurry was filtered through a sintered glass filter and the filtrate collected. The addition of a sufficient quantity of ether to the filtrate precipitated the triethylammonium bromide crystals. These appeared as large, white needles. The crystals were washed several times with ether before being collected and dried.

Glycine - N, N - diphenylamidine

Using a 1L three-necked flask equipped with a drying tube, a 150 mL drip tube and a nitrogen inlet affixed with a glass tube "bubbler", 4 g (0.01 mol) of $LiAlH_4$ was dissolved in 150 mL of anhydrous ether over a period of 30 minutes. A solution of 4.5 g (0.02 mol) of the unpurified N, N - diphenylcyanoformamidine in 100 mL of ether was added through the drip tube over a period of about two hours. During the entire reduction procedure dry nitrogen was continually bubbled through the mixture. After addition of the N, N - diphenylcyanoformamidine the mixture was stirred for another 30 minutes to complete the reduction. Afterwards the reaction vessel was cooled in an external ice bath. To the cooled flask, water was added to destroy the remaining $LiAlH_4$. First a small quantity of H_2O (4 mL) was added drop-wise through the drip tube. This was followed by a slow addition of 4 mL of 15% NaOH. Finally 12 more mL of water was added to finish off the oxidation. This stepwise destruction of $LiAlH_4$ prevented the formation of a sticky white "gum" and as a result kept the precipitate highly flocculent and easy to work with. After the addition of the water, the precipitate was removed by filtration and the remaining solution was extracted three times with ether (70 mL each). The organic phase was concentrated on a rotary evaporator to yield a bright yellow oil. For this step yields were typically 3.3 -3.8 g (75 -85%). TLC of the product was best performed in methanol which was charged by the addition of a few drops of ammonium hydroxide.

Glycine

To 3.5 g (0.016 mol) of crude glycine - N, N - diphenylamidine was added 20 mL of water in a 250 mL round bottom flask equipped with a reflux condenser. The apparatus was heated to 100 °C in an oil bath for about 15 minutes after which 3.15 g (0.01 mol) of $Ba(OH)_2 \cdot 8H_2O$ was added to the amidine/water mixture. The material was allowed to reflux at 130 °C for 1 hour. After the hydrolysis the resulting aniline by-products were extracted from the dark brown mixture with three washes of ether (50 mL) each. The barium ions in the remaining aqueous phase, which also contains the glycine, were removed by adding 5.5 mL of 10% sulfuric acid to create an insoluble $BaSO_4$ precipitate. This precipitate was centrifuged off, boiled with 50 mL H_2O to extract additional glycine and centrifuged again. The combined aqueous extracts were rotary evaporated to dryness. To the resulting material a small amount of water was added (2 mL) and an additional 50 mL of ethanol. Glycine crystals precipitated from this solution on cooling to 4 °C overnight. Typical yields from this procedure were between 0.60 - 0.70 g or about 45 - 50% relative to $K^{13}CN$. Note that for preparations that were to be used for the selective labeling of proteins, the final crystallization step was usually left out.

Confirmation of the identity and purity of the product was done by TLC using the standard solvent mixture of butanol, acetic acid and water (4:1:1) with detection by ninhydrin. 500 MHz ^{1}H and 75 MHz ^{13}C NMR of the product also aided in this confirmation.

Cell Growth and Protein Preparation

The E. coli strain DL39G (1) was transfected with the high expression, thermally inducible plasmid pDL59 (1) containing the E. coli thioredoxin gene. Cells were grown in a defined medium containing all 20 amino acids along with other essential salts and carbon sources (13). Protein expression was induced using a thermal shock treatment during the late log phase of cell growth. Thioredoxin was isolated from the contents of lysed cells through a three step protocol involving ammonium sulfate precipitation, size exclusion chromatography on a Sephadex G75 (superfine) column and finally ion exchange chromatography using a short DE52 column (1,14). Purity of the final product was in excess of 95% as judged by SDS polyacrylamide gel electrophoresis and ^{1}H NMR.

NMR Methods

All ^{13}C spectra were collected on a Nicolet NT 300 MHz wide bore spectrometer (^{13}C frequency = 75.4 MHz) equipped with a 12 mm ^{13}C probe. Samples were dissolved in 3.5 mL of 99.9% D_2O. The uncorrected pH of most samples ranged from 4.8 - 5.0. Protons were decoupled throughout the acquisition, pulse and delay periods using a broadband amplifier. Acquisition times were typically 1.02 seconds with relaxation delays being 0.5 ms (for protein work) or 500 ms (for amino acid work). The ^{13}C 90° pulsewidth was 16.0 μs. ^{1}H-^{13}C heteronuclear spectroscopy was performed on a Varian VXR 500 MHz spectrometer equipped with a 5 mm inverse detection probe. Sample sizes for these experiments were typically 0.60 mL. For the HMQC experiment a total of 1024 complex points were collected along the t_2 domain (^{1}H) and 96 increments along the t_1 domain (^{13}C). Data in both dimensions were zero-filled to create a 2K x 1K data set. The ^{13}C 90° pulsewidth was 32.2 μs and the refocusing delay was set to 3.5 ms. The relaxation delay was 2.3 seconds. The sweepwidth for protons was 5998.8 Hz while the sweepwidth for carbon was 3000.0 Hz.

ACKNOWLEDGEMENTS

The authors wish to thank David LeMaster for his helpful advice and for his gift of the DL39G/pDL59 E. coli strain. We also wish to thank Jose Delfino for his help and insight into the chemistry of this often problematic synthesis. Support from the Los Alamos Stable Isotope Resource (USPHS Grant RR02231 and US D.O.E.) along with financial support by the MRC Group in Protein Structure and Function (Canada) and by the National Institutes of Health (NIH Grant GM-22778) is gratefully acknowledged.

REFERENCES

1. LeMaster, D.M., Richards, F.M., Biochemistry 24: 7263 (1988)
2. Henry, G.D., Weiner, J.H., Sykes, B.D., Biochemistry 26: 3619 (1987)
3. Ikura, M., Kay, L.E., Bax, A., Biochemistry 29: 4659 (1990)
4. Howarth, O.W., Lian, L.Y., Biochemistry 23: 3515 (1984)
5. Benz, F.W., Roberts, G.C.K., J. Mol. Biol. 91: 345 (1975)
6. Ehrensvard, G., Stjernholm, R., Acta Chem. Scand. 3: 971 (1949)
7. Teplan, P., Banfi, J., Otvos, L., Acta Chim. Acad. Sci. Hung. 34: 109 (1962)
8. Finkbeiner, H., J. Org. Chem. 30: 3414 (1965)
9. Ott, D.G., Synthesis with Stable Isotopes, Wiley, New York (1981)
10. Kaneko, J., Izumi, Y., Chibata, I., Itoh, T., Synthetic Production and Utilization of Amino Acids, Wiley, New York (1974)
11. Greenstein, J.P., Winitz, M., Chemistry of the Amino Acids (Vols. 1-3), Wiley, New York (1961)
12. Khorana, H.G., Chem. Revs. 53: 145 (1953)
13. LeMaster, D.M., Richards, F.M., Biochemistry 24: 7263 (1985)
14. Wishart, D.S., Ph.D. Thesis, Yale University, New Haven (1991)