bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

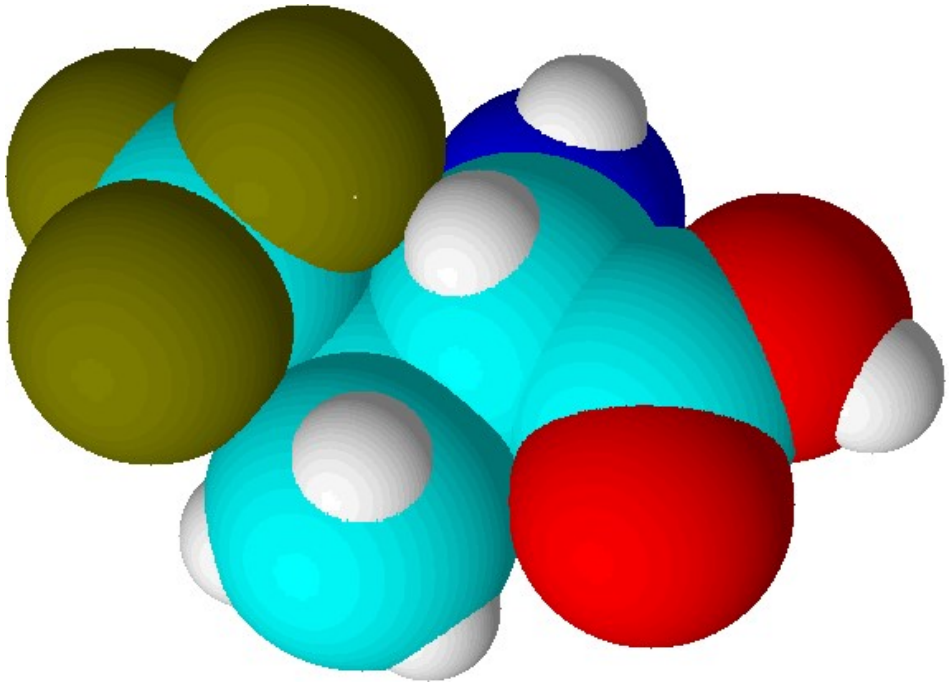
Notes on Synthesis of perdeutero-5-^{13}C,5,5,5-trifluoroisoleucine VI.

D. G. Naugler*, R. S. Prosser
Department of Chemistry, University of Toronto, UTM, 3359 Mississauga Rd. North Mississauga, ON, Canada, L5L 1C6

Abstract: The $^{13}CF_3$ group is a promising label for heteronuclear (^{19}F,^{13}C) NMR studies of proteins. Desirable locations for this NMR spin label include the branched chain amino acid methyl groups. It is known that replacement of CH_3 by CF_3 at such locations preserves protein structure and function and enhances stability. In particular, $^{13}CF_3$ may be introduced at the δ position of isoleucine and incorporated biosynthetically in highly deuterated proteins. This paper reports our work in synthesis and purification of 5,5,5-trifluoroisoleucine, its perdeutero and 5-^{13}C versions and of 2-^{13}C-trifluoroacetate and its utility as a precursor for introduction of the $^{13}CF_3$ group into proteins.

Key words: **^{13}C, ^{19}F, heteronuclear, paramagnetic, NMR, spin label, methyl TROSY, unnatural amino acid, fluorous, lipophilic, protein expression**

- Correspondence: david.naugler@utoronto.ca



bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Fluorine NMR and labeling strategies in proteins

Fluorine NMR spectroscopy is a powerful method for the study of both structure and dynamics of proteins and their interactions with other proteins or ligands ([1],[2],[3],[4]). Because of the ability of the ^{19}F lone-pair electrons to participate in non-bonded interactions with the local environment, ^{19}F chemical shifts are sensitive to changes in van der Waals contacts, electrostatic fields and hydrogen bonding. As such, ^{19}F chemical shifts (or changes in shifts) are often indicative of conformational changes ([5],[6],[7,8]), binding ([9]), and protein folding or unfolding events ([10],[11],[12]). Fluorinated probes are also frequently used to assess solvent exposure, via chemical shift changes or relaxation effects resulting from: 1) substituting H_2O for ^2H_2O ([13],2,[14]), 2) paramagnetic additives such as Gd^{3+}:EDTA ([15]) or 3) heteronuclear nuclear Overhauser effects ([16]). In membranous systems, analogous paramagnetic effects are also observed upon addition of nitroxide spin-labels ([17]) or dissolved oxygen ([18],[19]), facilitating the study of topology and immersion depth via fluorinated probes. Finally, associated ^{19}F spin-spin and spin-lattice relaxation rates are useful for studying conformational dynamics over a wide range of timescales, due to the significant chemical shift dispersion, shift anisotropy, and large hetereonuclear dipolar relaxation terms ([20], [21], [22], [23]).

Fluorine labeling of proteins is achieved in a number of ways. Via biosynthetic means, monofluorinated versions of tyrosine, phenylalanine, and tryptophan may be substituted for their nonfluorinated equivalents, often with little effect on overall expression yields (2,4). Fluorinated versions of methionine (7,[24]), proline ([25]), leucine ([26]), and isoleucine ([27]) have also been successfully incorporated into proteins. An alternative approach to ^{19}F labeling of proteins is to make use of a thiol specific fluorinated probe, which frequently consists of a terminal trifluoromethyl group (4). In this way, a fluorine tag may be placed at virtually any site in the protein, using successive single cysteine mutations of the protein under study. Thus, the majority of ^{19}F labels used in protein NMR fall under the category of isotopic fluoroaromatic or trifluoromethyl species. For very large proteins, spectral overlap may become problematic for biosynthetically labeled proteins.

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

However, a doubly (^{13}C, ^{19}F) labeled amino acid should provide greater resolution in two-dimensional (^{19}F,^{13}C) NMR spectra, with further possibilities of assignment without mutational analysis. Moreover, such two-dimensional NMR schemes should benefit from the relatively large one bond (^{13}C,^{19}F) coupling which is between 265 and 285 Hz for both fluoroaromatics and trifluoromethyl groups. Finally, the possibilities for studying dynamics from a ^{13}C,^{19}F pair encompasses a much greater range, since various zero, single, and double quantum coherences in addition to Zeeman and two-spin longitudinal order, may be separately evolved and studied ([28],[29]). In this paper, we present a method for the preparation of perdeuterated isoleucine, in which the terminal trifluoromethyl group consists of a ^{13}C-^{19}F pair. The motivation for this work is to develop a useful doubly labeled species for subsequent two-dimensional NMR studies of proteins, whose isoleucine residues have been fluorinated.

Advantages of a trifluoromethyl group

The trifluoromethyl group is expected to be a useful probe of molecular structure and dynamics, particularly in the hydrophobic core of proteins, at the interface between protein complexes, and in the membrane or detergent interior in studies of integral membrane proteins. Expressed within proteins, the CF_3 group offers additional benefits of sensitivity and relatively long transverse relaxation times. However, the inherent slow rotational tumbling associated with large proteins or protein complexes, and membrane proteins, results in line broadening and reduced sensitivity. ^{19}F spin labels also suffer extensively from dipolar relaxation with nearby proton spins of the protein (2), which may be largely avoided by extensive deuteration. Furthermore, in situations where ^{13}C,^{19}F two-dimensional NMR schemes are employed, the use of transverse relaxation optimized spectroscopy (TROSY) techniques ([30],[31]) may be considered. The TROSY effect in methyl groups, results from interference between intra-methyl dipolar interactions (31). As such, the effect is independent of field, to the extent that chemical shift anisotropy does not contribute to relaxation. Since the geometry of the trifluoromethyl group is similar to that of a CH_3 group, while the gyromagnetic ratio of the ^{19}F nucleus is 0.83 times that of ^{1}H, the methyl TROSY effect would be expected to

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

be preserved in appropriate (^{19}F,^{13}C) two-dimensional schemes. In particular, in the rigid limit, the maximum peak intensities in the (^{1}H,^{13}C) HMQC are predicted to depend on terms which are derived from relaxation via reorientation of either the CF or FF intramolecular bonds [*ie* transverse rates proportional to $S_{axis}^2\gamma_F^2\gamma_C^2\hbar^2\tau_C / r_{FC}^6$ and, to a lesser extent, $S_{axis}^2\gamma_F^4\hbar^2\tau_C / r_{FF}^6$ (31)]. In this case, S_{axis} and τ_c represent the local order parameter of the methyl rotor and the global correlation time associated with rotational tumbling of the protein. Finally, γ_C and γ_F represent the magnetogyric ratios of ^{13}C and ^{19}F, respectively while r_{CF} and r_{FF} designate the intramolecular bond lengths. In the (^{19}F,^{13}C) HMQC, where we estimate the CF and FF intramolecular bond lengths in a trifluoromethyl group to be 1.33 Å and 2.12 Å respectively, the above transverse rates are predicted to be more than three times smaller than those for the CH_3 groups, in the absence of external dipolar relaxation or relaxation due to chemical shift anisotropy.

The trifluoromethyl probe in isoleucine

In light of these anticipated advantages for one- and two-dimensional NMR, we have developed a protocol for the synthesis and purification perdeuterated 5,5,5-trifluoroisoleucine, in which the carbon nucleus of the trifluoromethyl group is ^{13}C enriched. The incorporation, using a cell-free protein expression technique, and eventual protein NMR studies are ongoing and will be reported elsewhere (M. Kainosho, private communication ([32])). We also describe herein a synthesis strategy for 2-^{13}C-trifluoroacetate and purification of the ammonium salt, in an effort to produce perdeutero 5-^{13}C-5,5,5-trifluoroisoleucine. This report is intended to communicate some of the subtleties involved in these efforts.

Isoleucine has some additional features that make it attractive for (^{19}F, ^{13}C) double labeling. At neutral pH, isoleucine is the most hydrophobic of the aliphatic side chain amino acids and it is often highly conserved at positions involved in hydrophobic interactions. Consequently, isoleucine often functions at the hydrophobic binding cleft.

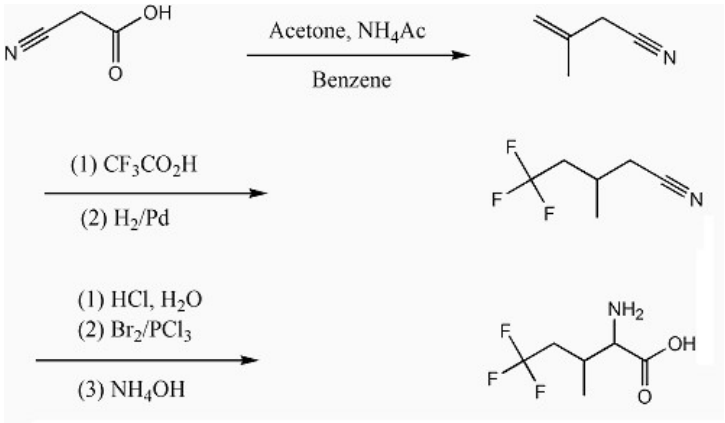
bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For example, isoleucine residues 737 and 898 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH_2-terminal domain, which in turn mediates transcriptional activation ([33]). Calmodulin provides another example. ^{1}H NMR studies of amide proton exchange rates of Ca^{2+}-saturated calmodulin and a Ca^{2+}-saturated calmodulin-mastoparan complex showed a reduction in solvent accessibility of Ile27 upon mastoparan binding ([34]). Genomic analysis of membrane protein families informs us of amino acid abundances and conserved motifs. The most abundant amino acids in transmembrane regions are leucine, isoleucine, valine, phenylalanine, alanine, glycine, serine, and threonine. Taken together, these amino acids account for 75% of the amino acids in transmembrane regions. In contrast, isoleucine (10% abundance versus 4% conserved), valine (8% versus 4%), methionine (4% versus 1%) and threonine (7% versus 4%) are less prevalent in conserved positions ([35]). Large hydrophobic (Phe, Leu, Ile, Val) residues show a clear preference for the protein surfaces facing the lipids for β-barrels, but in α-helical proteins, no such preference is seen, with these residues equally distributed between the interior and the surface of the protein ([36]).

Synthesis outline

The scheme shown below shows the route used to make 5,5,5-trifluoroisoleucine as a racemic mixture of diastereomers ([37],27). This scheme is modified for perdeuteration by substitution with cyanoacetic acid-d_3, acetone-d_6, CF_3CO_2D, ammonium-d_4 acetate-d_3 and D_2/Pd. Modification of this scheme for the $^{13}CF_3$ amino acid requires electrochemical trifluoromethylation using $^{13}CF_3CO_2^-$.

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Experimental

Melting points are uncorrected. ^{1}H, ^{19}F and ^{13}C NMR spectra were recorded on Varian Gemini 200 MHz and Unity Inova 600 MHz spectrometers. ^{2}H NMR spectra were recorded on the Varian Unity Inova 600 MHz spectrometer. UV spectra were recorded using a Biochrom Libra S22 UV/Visible spectrophotometer. IR spectra were recorded using a Nicolet Avatar 360 E.S.P. FT-IR spectrometer. GC/MS was done using a Shimadzu GCMS-QP5050 system. LC/MS was performed using a system comprised of a Waters/Alliance 2690 liquid chromatography module and a Waters Micromass ZQ ESI mass spectrometer. Hydrogenation was performed with a Parr high-pressure hydrogenation apparatus. Electrochemical trifluoromethylation was performed with a Sargent Welch analyzer with (one spinning) platinum mesh electrodes.

Synthesis of methallylcyanide and methallylcyanide-d_7

Synthesis of methallylcyanide and methallylcyanide-d_7 relies on 'active hydrogen' chemistry, *i.e.* exchangeable protons/deuterons. Wang et al. (27) cite a synthesis reported by Marson *et al.* ([38]).Cyanoacetic acid (e.g. 50.0 g, 0.59 mol), acetone (34.2 g, 0.59 mol), and ammonium acetate (4.0 g, 0.05 mol) (or their perdeutero counterparts) in dry benzene (100 mL) were refluxed with a Dean-Stark trap. Reflux was performed until one

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

equivalent of water (or D_2O) was collected in the trap. Due to the cost of perdeuteroammonium acetate, a minimal amount of this catalyst was used. The quantity was observed not to be critical to yield. One gram of ammonium-d_4 acetate-d_3 will suffice for the synthesis of methallylcyanide-d_7. The Dean-Stark unit was replaced with a distillation head and the fraction collected between 110 °C and 115 °C.

Compton et al. ([39]) reports the existence of both 3-methyl-2-butenenitrile and 3-methyl-3-butenenitrile in a sample of methallylcyanide. This is consistent with an equilibrium arising from 'active hydrogen' chemistry as evidenced by the ^2H NMR spectrum in Figure 1, below. The predicted boiling points for these isomers are within 2 °C, so they cannot be separated by fractional distillation. Accordingly, this product is properly called methallylcyanide (37) rather than 3-methyl-but-3-enenitrile (27). Appreciation of the active hydrogen nature of this product is required in order to make methallylcyanide-d_7.

Synthesis of methallylcyanide-d_7 requires the preparation of cyanoacetic acid-d_3, first and then its condensation with acetone-d_6, catalyzed by ammonium-d_4 acetate-d_3. 50 grams of cyanoacetic acid was dissolved with warming in a minimum quantity of D_2O, then left overnight to affect exchange. Excess H_2O/D_2O was subsequently removed with a rotary-evaporator, and the wet crystals further dried overnight in a vacuum desiccator. Because not all H_2O/D_2O can be removed by this procedure, subsequent steps require less D_2O to affect dissolution. The proton/deuteron exchange procedure was executed three times and the product submitted for mass spectral analysis. Three more H/D exchange steps were performed. The resulting cyanoacetic acid-d_3 was condensed with a molar excess of actetone-d_6 in the presence of ammonium-d_4 acetate-d_3 and the water of condensation was analyzed by ^1H NMR. The trace aqueous solubility of benzene is known [40]. Comparison of integrated peak intensities showed that the water of condensation contained 0.6% H_2O and 99.4% D_2O. Given that all H/D sites are exchangeable, it is concluded that methallylcyanide-d_7 obtained after decarboxylation and distillation of the condensation product was 99.4% isotopically pure. This isotopic purity is consistent with that of the reagents used. See the ^2H NMR spectrum in Figure 1, below.

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Electrochemical trifluoromethylation of methallylcyanide-d_7

Muller (37) chose to perform electrochemical trifluoromethylation of methallylcyanide in aqueous 90% methanol between platinum electrodes whereas we chose a later system ([41]) of CH_3CN and H_2O (8:1). An electrochemical solvent system of acetonitrile and water necessitates further use of acetonitrile and D_2O for methallylcyanide-d_7.

Because of hydrogen exchange chemistry of methallylcyanide-d_7, special care needs to be applied to the adaptation of this reaction to a perdeuterated substrate. H_2O needs to be replaced by D_2O. Both MeOD and CH_3CN methyls ought to be resistant to H/D exchange under the conditions of the reaction but CH_3CN is cheaper than MeOD and so is a better choice. Muller (37) chose to use a 73% molar excess of CF_3CO_2H relative to methallylcyanide and to use a slight coulombic excess. Dmowski (41) chose to use a 20 times molar excess of TFA relative to substrate and 1.5 Faradays per mole of TFA. Because both methallylcyanide-d_7 and 2-^{13}C-trifluoracetate are very expensive these should be used in 1:1 molar ratio and with 50% excess current. Both authors achieved mild basic condition with 10% Na.

Muller's workup (37) was followed, as described below. For methallylcyanide-d_7, H_2O should be replaced by D_2O, H_2 by D_2 and MeOH by MeOD or some other exchange resistant solvent such as CH_3CN. An iron (steel) cathode would minimize loss of valuable substrate by electrolytic hydrogenation.

The mixture was poured into 700 mL of water, the dense oil collected, and the aqueous layer extracted with two 40 mL portions of dichloromethane. The combined organic layers from three identical runs were distilled to remove the solvent and then steam distilled. The nonaqueous layer was a mixture of 3-methyl-5,5,5-trifluoropentanonitrile, several isomeric 3-methyl-5,5,5-trifluoropentenonitriles, methallyl cyanide, and unidentified by-products. It was diluted with methanol and hydrogenated at low pressure over 5% Pd/C. Distillation afforded 66 g of nearly pure 3-methyl-5,5,5-trifluoropentanonitrile, b.p. 166-171 °C.

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For the perdeutero route, once the various isomeric perdeutero- 3-methyl-5,5,5-trifluoropentenonitriles are reduced by D_2, exchange is no longer possible. Subsequent steps closely follow the protocol outlined by Muller as does our exposition. Quantities were adjusted proportionately.

The distillation product, 3-methyl-5,5,5-trifluoropentanonitrile was stirred with sufficient concentrated aqueous hydrochloric acid to bring most of the organic material into solution, for several days, then diluted with water and refluxed overnight. The separated organic layer was isolated, dried over Na_2SO_4, and distilled at 6 torr, giving 58.3 g of 3-methyl-5,5,5-trifluoropentanoic acid, b.p. 76-81 °C.

58.3 g (0.343 mol) of 3-methyl-5,5,5-trifluoropentanoic acid, 19.1 mL of bromine and 0.6 mL of phosphorus trichloride were refluxed with a trap to absorb gaseous hydrogen bromide until the color of bromine had disappeared. On distilling at 6 torr, about 6 g of 3-methyl-5,5,5-trifluoropentanoic acid were recovered. 60 g of distillate boiling at 100-106 °C consisted mainly of nearly equal amounts of the two diastereomers of 2-bromo-3-methyl-5,5,5-trifluoropentanoic acid . This mixture was used for the preparation of 2-amino-3-methyl-5,5,5-trifluoropentanoic acid without further purification. 64.4 g (0.259 mol) of this nearly pure 2-bromo-3-methyl-5,5,5-trifluoropentanoic acid and 225 mL of concentrated aqueous ammonia were cautiously mixed and stored in a closed flask at 44-48 °C for 5 days. The stopper was removed and the mixture gently heated with a water bath to drive off excess ammonia and reduce the volume to about 90 mL.

In our hands we observed the formation of polymeric materials and so we could not crystallize the product from our reaction mixture. Accordingly we developed several purification methods discussed in the next section.

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Purification of a reaction mixture containing 2-amino-3-methyl-5,5,5-trifluoropentanoic acid

Two factors are likely to alter the physical properties of 5,5,5-trifluoroisoleucine relative to native isoleucine. The inductive effect of the CF_3 group ([42]) will make the amino acid and amino groups more acidic and the greater hydrophobicity of the CF_3 group will enhance the hydrolytic stability of its polymers ([43]).

In our hands, reaction of a mixture of diastereomers of 2-bromo-3-methyl-5,5,5-trifluoropentanoic acid with aqueous ammonia produced a product mixture that contained a significant quantity of polymeric material. Some of this material was readily soluble in diethyl ether and $CDCl_3$. That fraction that was soluble in organic solvent could be hydrolyzed by dissolution in TFA followed by gradual addition of water. We found that the reaction mixture could be stabilized by formation of the TFA salt.

We first chose a method of chemical purification that was appropriate for the partial fluorous character of the amino acid ([44],[45],[46]). The use of a C_8F_{17} BOC derivative (46) proved to be problematic because the only rational method for its hydrolysis was the use of TFA and we already had made the TFA salt. Exploitation of the C_8F_{17} Cbz derivative (46) proved to be more fruitful. A product mixture was dissolved in THF and a minimal quantity of water and reacted with a small molar excess of N-[4-(1H,1H,2H,2H-perfluorodecyl)benzyloxycarbonyloxy]succinimide. The resulting C_8F_{17} Cbz derivative was purified by solid phase extraction on fluorous silica (44). This derivative was subjected to atmospheric pressure hydrogenolysis over 5%Pd/C and the subsequent product mixture subjected to fluorous SPE. Lyophilization yielded a soluble product that was white. Incorporation of this product in calmodulin (M. Kainosho: private communication) using a cell free protein expression system proved that the product mixture contained ≤25% 5,5,5-trifluoro-L-isoleucine and also proved the efficacy of the chemical purification.

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

During lyophilization, much of the product was lost due to sublimation. Sublimation has recently been revisited as a means for purification of amino acids ([47]). We found that our chemically purified product mixture could be sublimed at 150 °C and 6 mm Hg, but that fractional sublimation would require better vacuum and temperature control.

Finally, we exploited ion exchange chromatography using cellulose phosphate to separate monomeric and polymeric fractions. Elution with distilled H_2O yielded a microcrystalline fraction while elution with dilute aqueous ammonia yielded a waxy fraction identifiable as polymeric material. Refinement of ion exchange column purification will be developed using ion exchange TLC on cellulose phosphate paper.

We explored the use of analytical HPLC-MS first using an acetonitrile gradient in 0.1% aqueous TFA on a C_{18} column. 2-amino-3-methyl-5,5,5-trifluoropentanoic acid has a molecular weight of 185. We identified two separated peaks corresponding to [M+] and [M+H+] on an ESI-MS instrument. We conclude that these are the diastereomers and that differential inductive effects due to CF_3 cause differing acid/base properties of the diastereomers. These peaks were followed at longer time by peaks due to polymers. We wished to explore the possibility of preparative scale chromatography where TFA would be counterproductive. An isocratic method was developed using 5% acetonitrile in 0.2% aqueous formic acid on an analytical C_{18} column.

Ammonium 2-^{13}C-trifluoroacetate

In order to embed the $^{13}CF_3$ group into a synthetic scheme for 5,5,5-trifluoroisoleucine following the synthetic scheme above, a route to 2-^{13}C-trifluoroacetate is required. Trifluoroacetic acid has been made via electrochemical fluorination ([48]). Since this electrolysis would entail use of anhydrous hydrogen fluoride within a customized Teflon reaction vessel, a refrigeration unit, a high current power supply and a process control system, we chose to explore an alternate route involving halogen exchange with 2-^{13}C-tribromoacetic acid. We devised a new route to 2-^{13}C-tribromoacetic acid starting with 2-^{13}C-ethanol, discussed below.

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Halogen exchange reaction design considerations

Preliminary experiments were performed following the halogen exchange reaction originated by ([49]). Mass spectral analysis of an early natural abundance test reaction showed that trifluoroacetic acid was formed by reaction of $AgBF_4$ with CBr_3CO_2H in DCM. However, ^{13}C NMR multiplet analysis of a test reaction with $^{13}CBr_3CO_2H$ showed a conversion to $^{13}CF_3CO_2H$ of only 18% after stirring at for 10 days at room temperature. A longer reaction time in glassware was found to be counterproductive because of failure of containment. $AgBF_4$ releases highly aggressive BF_3 during the course of the reaction. The reaction between BF_3 and silica gel ([50],[51]) is useful at the purification stage, however, reaction with ground glass joints may give rise to leakage.

A procedure for making 2-^{13}C-trifluoroacetic acid

The following procedure was designed to avoid the necessity of handling hydrogen fluoride, either as a solvent, reagent, or product. Synopsis: 2-^{13}C-ethanol is converted to the tribromoacetaldehyde, 2-^{13}C-bromal (hydrate) using a molar excess of bromine, Br_2. One equivalent of water converts the product mixture to bromal hydrate. Reaction with excess nitric acid at a temperature less than 50 °C converted bromal hydrate to 2-^{13}C-bromoacetic acid. This product is isolated and converted to 2-^{13}C-trifluoroacetic acid using $AgBF_4$ under pressure in dichloromethane. The 2-^{13}C-trifluoroacetate is extracted into ammonia. Impure ammonium 2-^{13}C-trifluoroacetate can be enhanced in purity by sublimation at 85 °C and with a vacuum less than 10 microns Hg. The yields were very low.

Conversion of 2-^{13}C-ethanol to 2-^{13}C-bromal (hydrate)

The reaction proceeds according to:

2-^{13}C-ethanol + $4Br_2$ → 2-^{13}C-bromal + 5HBr

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In a closed system, the above is an equilibrium reaction. In order to drive the reaction to completion, product HBr gas must be permitted to escape. This loss of mass results in a considerable reduction in the volume of the reaction mixture. We have tried sulfur and I_2 as catalysts. TFA is probably a better catalyst for this reaction. The oxidation potential of Br_2 is not sufficient to carry oxidation beyond the aldehyde. The aldehyde is required for tribromination, because each bromination step proceeds via the enol. In order to avoid loss of volatiles, 2-^{13}C-ethanol and excess Br_2 were combined at liquid nitrogen temperature and warmed very slowly to reflux temperature. When the reaction has been driven to completion, ^{13}C NMR shows the presence of only 2-^{13}C-bromal (~40 ppm) and its hydrate (~12 ppm). Prior to the next step, it may desirable to isolate 2-^{13}C-bromal via distillation, and its hydrate by crystallization but it is not essential. One equivalent of H_2O is added to convert all to hydrate.

Conversion of 2-^{13}C-bromal hydrate to 2-^{13}C-tribromoacetic acid

$$^{13}CBr_3CH(OH)_2 + HNO_3 \rightarrow {}^{13}CBr_3CO_2H + \tfrac{1}{2}N_2O_4 + H_2O$$

The more stable hydrate of tribromoacetaldehyde is the species oxidized. Completion of this reaction can be determined by ^{13}C NMR (~34 ppm). At completion of reaction, excess nitric acid is removed under vacuum. Purity can be determined by melting point, 128 °C.

Conversion of 2-^{13}C-tribromoacetic acid to 2-^{13}C-trifluoroacetic acid

$$^{13}CBr_3CO_2H + 3AgBF_4 + 3(C_2H_5)_2O \rightarrow {}^{13}CF_3CO_2H + 3AgBr + 3[BF_3{\cdot}(C_2H_5)_2O]$$

This reaction is conducted in dichloromethane. This three step reaction is exceedingly slow. For practical synthesis, it is necessary to conduct this reaction in a sealed pressure reaction vessel at above 75 °C. The completion of this reaction can be determined by ^{13}C NMR (quartet at 116.6 ppm is dominant) or ^{19}F NMR (doublet at -76.55 ppm is

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dominant). Under standard conditions BF_3 is a gas, whereas $BF_3{\cdot}(C_2H_5)_2O$ is liquid. A strong Lewis acid, BF_3 forms a complex with diethyl ether. This reduces the pressure. At the completion of the reaction, the DCM reaction mixture is eluted through silica gel. SiO_2 reacts with BF_3. 2-^{13}C-trifluoroacetate is extracted into aqueous ammonia and excess removed under vacuum. Ammonium 2-^{13}C-trifluoroacetate was converted into a fine powder by lyophilization in order to aid in sublimation. Ammonium 2-^{13}C-trifluoroacetate was sublimed to improve purity by sublimation at 85 °C and less than 10 microns Hg vacuum. Because initial purity was poor, yield was poor. Gram scale quantities of material could be processed in this way.

Manipulation of ^{13}C haloacetates

The halogen exchange reaction between 2-^{13}C-tribromoacetic acid and $AgBF_4$ proceeds in a stepwise fashion and hence yields a mixture of haloacetates. The target compound, 2-^{13}C-trifluoroacetic acid is too volatile and so the haloacetate mixture is best manipulated as a salt. The ammonium salts have volatilities that were exploited for purification by high vacuum sublimation and the progress of purification was monitored by ^{19}F and ^{13}C NMR. Whereas high vacuum sublimation has only one theoretical plate, the greater load capacity compared to chromatography affords it an advantage for the first stages of purification. Repeated stages of high vacuum sublimation yielded a product that gave three spots on silica gel TLC. Neutral alumina TLC gave streaks. One spot showed an R_f=0.78 on silica gel TLC that was identical to the R_f of natural abundance ammonium trifluoroacetate, eluted with 10% aqueous ammonia in methanol. On the basis of this observation, the ammonium ^{13}C haloacetate mixture was subjected to preparative TLC under the same conditions. ^{19}F and ^{13}C NMR of the product from the preparative TLC target band showed that the mixture consisted of ammonium salts of 2-^{13}C-trifluoroacetate, 2-^{13}C-bromodifluoroacetate and 2-^{13}C-dibromofluoroacetate (See Table I).

It is interesting to note that CF_3I has been enriched to 86% in ^{13}C by selective multiphoton dissociation of $^{12}CF_3I$ at pressures less than 1.0 torr ([52]). A more practical

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

method would rely on enrichment in the condensed phase. An isotope effect is often observed on melting points ([53]). In recent work, boron has been enriched to 93.21% in ^{10}B and 99.01% in ^{11}B by zone refining ([54]). It is known that isotope effects on heat capacity and crystal transition temperature can be detected by differential scanning calorimetry ([55]). Zone refining of low melting ammonium trifluoroacetate (M.P. 123 °C) would be more energy efficient than that of high melting boron (M.P. 2079 °C). In future work, we plan to develop an analytical HPLC method and to use preparative HPLC for the purification of ammonium 2-^{13}C-trifluoroacetate. Infrared difference spectroscopy may be able to resolve the carbon isotope effect for ammonium 2-^{13}C-trifluoroacetate in the condensed phase. Differential scanning calorimetry of ammonium 2-^{13}C-trifluoroacetate will provide the thermodynamic information needed to plan and develop a zone refining method for the extraction of ammonium 2-^{13}C-trifluoroacetate from an inexpensive natural abundance melt.

Another promising route to ^{13}C enriched TFA that we plan to explore is that of fluorodeoxygenation ([56]), starting with glycine. Recently a new and better synthesis of arylsulfur trifluorides has been reported ([57]) for reagents that may provide a convenient route to fluorodeoxygenation of carboxylic acids.

Conclusion

We have presented a *critical scientific narrative* of an unfolding technology. The advantage of the $^{13}CF_3$ NMR spin label in protein NMR have been explained. *Progress in incorporation* of this spin label in a perdeuterated amino acid and in an important protein *is reported*. We *comment* on some of the synthetic subtleties encountered.

Whereas incorporation of monofluorinated aromatic amino acids in proteins using in vivo or in vitro expression systems is now routine, incorporation of trifluoromethyl analogs of branched chain amino acids is at present an emerging technology. A biochemical hurdle is the specificity of an aminoacyl tRNA synthetase. Wang et al. (27) determined that the specificity constant, k_{cat}/K_M of 5,5,5-trifluoroisoleucine is only 1/134 that of isoleucine

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for E. coli isoleucyl tRNA synthetase. Considerations of enzyme kinetics and competitive inhibition dictate that the background concentration of isoleucine needs to be effectively zero in order for $tRNA^{Ile}$ to be charged with 5,5,5-trifluoroisoleucine by isoleucyl tRNA synthetase. In addition to the catalytic domain, where the amino acid and $tRNA^{aa}$ are specifically ligated, aminoacyl tRNA synthetase also have an editing domain for the hydrolysis of mischarged tRNA. In the case of isoleucyl tRNA synthetase, the editing domain has been evolved most specifically for the hydrolysis of val· $tRNA^{Ile}$. The successful incorporation of 5,5,5-trifluoroisoleucine in mDHFR, mIL II (27) and in calmodulin (M. Kainosho: private communication) implies that 5TFI· $tRNA^{Ile}$ is too large for the editing site. Practical questions remain. Is an expression system based upon E. coli more advantageous than one based upon a eukaryotic organism? Is a cell free lysate system preferable to in vivo expression system?

This paper *covers many areas* of synthetic chemistry, organic, fluoro, isotopic, and biochemical. We have *identified* ammonium 2-^{13}C-trifluoroacetate as an important synthon for introduction of the $^{13}CF_3$ group into amino acids. Completed synthesis of methallylcyanide-d_7 and conceptual synthesis of 5-^{13}C-5,5,5-trifluoroisoleucine-d_7, provide an important building block for the exploitation of $^{13}CF_3$ in protein NMR.

We are grateful for the comments of reviewers, each of whom described this manuscript as unusual. 5,5,5-TFI is an unnatural AA, more so than monofluorinated amino acids. Either as uniform, or site specific isotopic labels, deuterium and ^{13}C amino acids are readily synthesized and incorporated. Fluorine is the 13th most abundant isotope in the earth's crust, yet even after 3.5b years of biology only about a dozen fluorinated natural products have been evolved, attributed to fluorine's chemistry as a "superhalogen" [58]. Organofluorine compounds as polymers or as drugs have proven useful in material science and pharmacology. The target spin $^{13}CF_3$ label should prove useful in multidimensional heteronuclear NMR structure dynamics studies of proteins. Synthesis of 5,5,5-TFI has been proven by its incorporation in the calcium binding protein calmodulin. Methallylcyanide-d_7 has been produced with military grade deuterium

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

isotope purity, 99.4%. Trace quantities of 2-^{13}C-trifluoroacetate have been characterized by ^{13}C-^{19}F NMR coupling. This contribution may open a door to the future.

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table I

	$\delta(^{19}F)$ ppm*	$\delta(2\text{-}^{13}C)$ ppm	$^{1}J_{CF}$ Hz
$^{13}CF_3CO_2^-$	-76.68	116.6	291.58
$^{13}CBrF_2CO_2^-$	-60.18	112.76	318.39
$^{13}CBr_2FCO_2^-$	-56.51	89.21	327.75

*relative to TFA, -76.55 ppm

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Illustrations

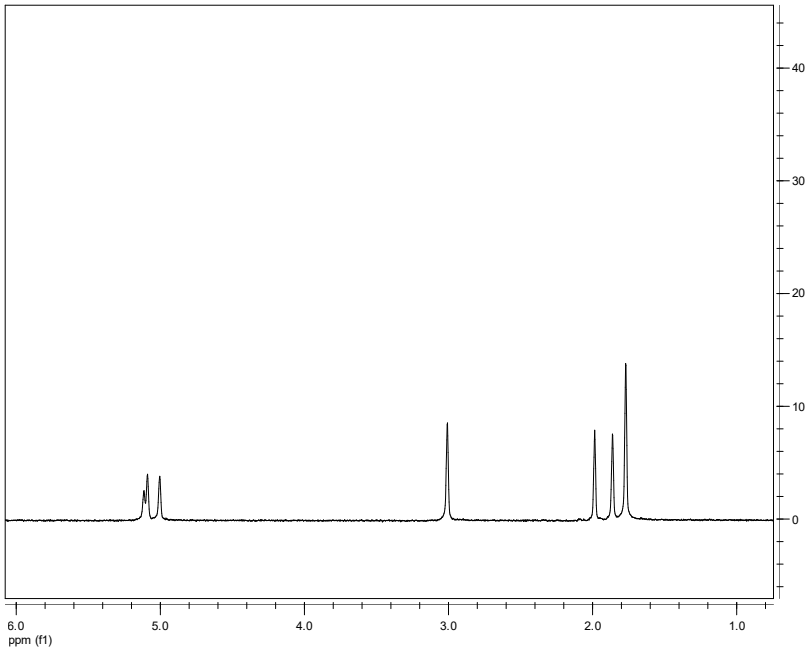


Figure 1. Deuterium NMR spectrum of methyallylcyanide-d_7, an equilibrium mixture of 3-methyl-2-butenenitrile-d_7 and 3-methyl-3-butenenitrile-d_7. The predicted boiling points of these two components are within 2 °C, thus these isomers cannot be separated by fractional distillation.

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[1]. M. A. Danielson, and J. J. Falke. Annu. Rev. Biophys. Biomol. Struct. **25,** 163-195 (1996).

[2]. G. T. Gerig. Prog. NMR Spect. **26**, 293-370 (1994).

[3]. Y. G. Gakh, A. A. Gakh, and A. M. Gronenborn. Magn. Reson. Chem. **38**, 551-558 (2000).

[4]. C. Frieden, S. D. Hoeltzli, and J. G. Bann. Methods Enzymol. **380**, 400-415 (2004).

[5]. L. A. Luck and J. J. Falke. Biochemistry **30**, 4248-4256 (1991).

[6]. L. A. Luck and J. J. Falke. Biochemistry **30**, 4257-4261 (1991).

[7]. B. Salopek-Sondi, M. D. Vaughan, M. C. Skeels, J. F. Honek, and L. A. Luck. J. Biomol. Struct. and Dynamics. **21**, 235-246 (2003).

[8]. H. Li and C. Frieden. Biochemistry **44**, 2369-2377 (2005)

[9]. J. G. Bann, J. Pinkner, S. J. Hultgren, and C. Frieden. Proc. Natl. Acad. Sci. USA **99**, 709–714 (2002).

[10]. C. Frieden. Biochemistry **42**, 12440-12446 (2003).

[11]. Q. Shu and C. Frieden. Biochemistry **43**, 1432-1439 (2004).

[12]. Q. Shu and C. Frieden. J. Mol. Biol. **345**, 599–610 (2005).

[13]. W. E. Hull and B. D. Sykes. Biochemistry **15**, 1535-1546 (1976).

[14]. G. S. Rule, E. A. Pratt, V. Simplaceanu, and C. Ho. Biochemistry **26**, 549-556 (1987).

[15]. M. A. Danielson, H. P. Biemann, D. E. Koshland Jr., and J. J. Falke. Biochemistry **33**, 6100-6109 (1994).

[16]. D. P. Cistola and K. B. Hall. J. Biomol. Nucl. Magn. Reson. **5**, 415-419 (1995).

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[17]. H. T. N. Truong, E. A. Pratt, and C. Ho. Biochemistry **30**, 3893-3898 (1991).

[18]. R. S. Prosser, P. A. Luchette and P. W. Westerman. Proc Natl Acad Sci U S A **97,** 9967-71 (2000).

[19] P. A. Luchette, R.S. Prosser and C.R. Sanders. J Am Chem Soc **124,** 1778-81 (2002).

[20]. W. E. Hull and B. D. Sykes. J. Chem. Phys. **63**, 867-880 (1975).

[21]. W. E. Hull and B. D. Sykes. J. Mol. Biol. **98**, 121-153 (1975).

[22]. B. D. Sykes, W. E. Hull, and G. H. Snyder. Biophys. J. **21**, 137-146 (1978)

[23]. S. Rozovsky, G. Jogl, L. Tong, and A. E. McDermott. J. Mol. Biol. **310**, 271-280 (2001)

[24]. M. D. Vaughn, P. Cleve, V. Robinson, H. S. Duewel, and J. F. Honek, J. Am. Chem. Soc. **121**, 8475 (1999).

[25]. C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber, and L. Moroder. Angew. Chem. Intl. Ed. **40**, 923-925 (2001).

[26]. J. Feeney, J. E. McCormick, C. J. Bauer, B. Birdsall, C. M. Moody, B. A. Starkmann, D. W. Young, P. Francis, R. H. Havlin, W. D. Arnold, and E. Oldfield. J. Am. Chem. Soc. **118**, 8700-8706 (1996).

[27]. P. Wang, Y. Tang, and D. A. Tirrell. J. Am. Chem. Soc. **125,** 6900-6906 (2003).

[28]. A. G. Palmer, III. Chem. Rev. **104**, 3623-3640 (2004).

[29]. L. E. Kay. J. Magn. Reson **173,** 193-207 (2005).

[30]. K. Pervushin, R. Riek, G. Wider and K. Wuthrich. Proc Natl Acad Sci U S A 94, 12366-12371 (1997).

[31]. V. Tugarinov, P.M. Hwang, J. E. Ollerenshaw and L.E. Kay. J Am Chem Soc **125,** 10420-10428 (2003).

[32]. T. Torizawa, M. Shimizu, M. Taoka, H. Miyano and M. Kainosho. J Biomol NMR **30,** 311-25 (2004).

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[33]. B. He, J. A. Kemppainen, J. J. Voegel, H. Gronemeyer and E. M. Wilson. J. Biol. Chem. **274,** 37219-37225 (1999).

[34]. S. Ohki, S. Tsuda, S. Joko, M. Yazawa, K. Yagi and K. Hikichi. J. Biochem. (Tokyo) **109,** 234-237 (1991).

[35]. Y. Liu, D. M. Engelman, and M. Gerstein. Genome Biol. **3,** 0054 (2002).

[36]. M. B. Ulmschneider, and M.S. Sansom. Biochim. Biophys. Acta. **1512,** 1-14 (2001).

[37]. N. Muller. J. Fluorine Chem. **36,** 163-70 (1987).

[38]. C. M. Marson, U. Grabowska, T. Walsgrove, D. S. Eggleston, and P. W. Baures. J Org Chem **59,** 284-290 (1994).

[39]. D. A. C. Compton, W. F. Murphy and H. H. Mantsch. Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy **37,** 453-455 (1981).

[40] CRC Handbook of Chemistry and Physics, 87th Edition, 2006-2007, David R. Lide. Editor, pp. 3-32, 8-96.

[41]. W. Dmowski and A. Biernacki. J. Fluorine Chem. **78,** 193-194 (1996).

[42]. V. P. Kukhar. J. Fluorine Chem. **69,** 199-205 (1994).

[43]. T. Nihei, S. Kurata, Y. Kondo, K. Umemoto, N. Yoshino and T. Teranaka. J. Dent. Res. **81,** 482-486 (2002).

[44]. D. P. Curran. Synlett 1488-1496 (2001).

[45]. D. P. Curran, M. Amatore, D. Guthrie, M. Campbell, E. Go and Z. Luo. J. Org. Chem. **68,** 4643-4647 (2003).

bioRxiv preprint doi: https://doi.org/10.1101/140681; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[46]. Z. Luo, J. Williams, R. W. Read and D. P. Curran. J. Org. Chem. **66,** 4261-4266 (2001).

[47]. D. P. Glavin and J. L. Bada. Anal Chem **70,** 3119-3122 (1998).

[48]. D. J. Wasser, P. S. Johnson, F. W. Klink, F. Kucera, and C. C. Liu. J. Fluorine Chem. **35,** 557-569 (1987).

[49]. A. J. Bloodworth, K. J. Bowyer, and J.C. Mitchell. Tetrahedron Lett. **28,** 5347-5350 (1987).

[50]. K. H. Rhee and M. R. Basila. J. Catal. **10,** 243 (1968).

[51]. T. Yoshidome, Z. Y. Dang, and B. A. Morrow. Anal. Sci. **19,** 429-435 (2003).

[52]. S. Bittenson and P. L. Houston. J. Chem. Phys. **67,** 4819-4824 (1977).

[53]. W. A. Vanhook. Zeit. fur Natur. Section A, Phys. Sci. **50,** 337-346 (1995).

[54]. N. Nogi, T. Hirano, K. Honda, S. Tanaka, and T. Noda. J. Surf. Anal. **4,** 280-283 (1998).

[55]. H. Shigematsu, H. Nakadaira, T. Matsui, A. Tobo, and K. Ohoyama. J. Nucl. Sci. Technol. **39,** 395-398 (2002).

[56]. G. V. Shustov, S. N. Denisenko, I. I. Chervin, and R. G. Kostyanovskii. Bul. Acad. Sci. USSR, Division of Chemical Science **37,** 1422-1427 (1988).

[57]. V. E. Pashinnik, E. G. Martyniuk, M. R. Tabachuk, Y. G. Shermolovich, and L. M. Yagupolskii. Synth. Comm. **33,** 2505-2509 (2003).

[58] D. B. Harper and D. O'Hagan, Nat. Prod. Rep., 1994,11, 123-133