Supplemental Information

This section contains 3 items: A – Synthesis of dκ; B – Incorporation of dXTP by KF, and; C – Analysis of dXTP incorporation by mass spectroscopy.

A. The synthesis of dκ required significant alteration from the published procedure (Figure 1).

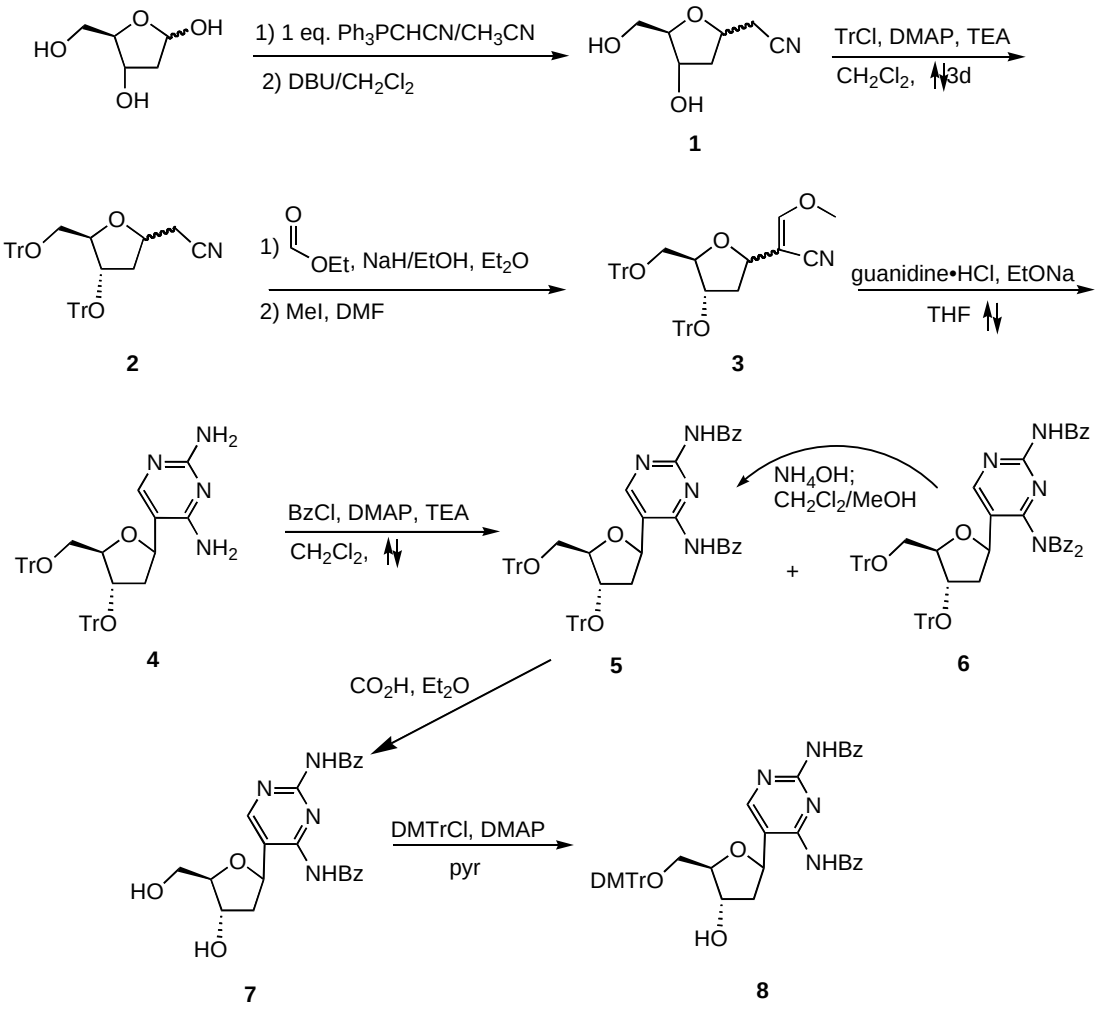


Figure 1. Modified Synthesis of dκ nucleoside.

*2-(2-deoxy-D-ribofuranosyl)-acetonitrile **(1)***. 250 mL dry acetonitrile was added to a 500 mL round bottom flask containing 2'-deoxy-D-ribose (4.3 g) and Ph_3P=CHCN (10.6 g), and the solution was brought to reflux. After 18 hr, the solution was cooled and the solvent removed. The residue was redissolved in 200 mL dry CH_2Cl_2. 5.3 mL DBU was added and the solution stirred an additional 5.5 hr. After quenching with DOWEX H^+ resin, the solvent was removed and the resulting solid dried overnight *en vacuo*. Silica chromatography (9:1 ethyl acetate:methanol) afforded the desired product as a mixture of isomers. The mixture can be carried forward to next step without further purification. ^{1}HNMR (δ): 4.38-4.46 (m, 4H); 4.02 (q, 1H, J=4.2 Hz); 3.93 (m, 1H); 3.53-3.79 (m, 7H); 2.79 (dd, 1H, J1=16.9 Hz, J2=5.2 Hz); 2.72 (dd, 2H, J1=6.2 Hz, J2=2.2 Hz); 2.58-2.62 (m, 2H); 2.43 (m, 1H); 1.67 (br s, 1H); 1.76 (br s, 1H); 1.82-2.06 (m, 5H).

*2-(3,5-O-Ditrityl-2-deoxy-D-ribofuranosyl)-acetonitrile **(2)***. 125 mL of dry dichloromethane was added to **1** (2.2 g) in a 250 mL round bottom flask. Triethylamine (5.7 mL) was added, followed by DMAP (134 mg) and TrCl (8.4 g). After refluxing for four days, the solution was cooled and the solvent removed. Separation by silica chromatography gave the title compound in 28% yield from 2-deoxy-D-ribose. ^{1}HNMR (δ): 7.16-.33 [m, 33H, Ph (+ $CDCl_3$)]; 4.41 (m, 1H); 4.10 (m, 1H); 3.88 (br t, 0.5H); 3.84 (br t, 0.5H); 2.99 (m, 1H); 2.70-2.78 (m, 1.5H); 2.57 (m, 1H); 2.57 (m, 2H). MS(EI): calc. 641.29; found 641.3.

*3-Methoxy-2-(3,5-O-ditrityl-2-deoxy-D-ribofuranosyl)-acrylonitrile **(3)***. 100 mL dry ether was added to **2** (5.8 g) in a 250 mL round bottom flask. 350 µL of absolute ethanol was added, followed by solid NaH (5.4 g). After stirring at room temperature for one hour, ethyl formate

(7.25 mL) was added. The white suspension stirred at room temperature until no starting material was detectable by TLC (2:1 hexanes:ethyl acetate; about two hours). The mixture was then filtered through a frit to remove excess NaH, washed with THF, and the solvent removed. The resulting residue was redissolved in dry DMF (100 mL) and MeI added dropwise (2.8 mL). After stirring for 11 h, the reaction was quenched with water and partitioned between water and ether. The combined organic fractions were washed with water and dried over Na_2SO_4. After evaporation of the solvent, the product was purified by silica chromatography to obtain the title compound (4.12 g, 67%) as a mixture of four isomers (α and ß anomers; cis and trans olefin; the integration of the four olefinic protons is diagnostic of the isomer distribution). ^{1}HNMR (δ): 7.13-7.56 [m, 112H, Ph (+ $CDCl_3$)]; 6.83 (s, 0.7H); 6.79 (s, 1.0H); 6.75 (s, 0.8H); 6.72 (s, 0.4H); 5.06 (m, 1.1H); 4.62 (m, 1.1H); 4.11-4.26 (m, 10.3H); 3.93 (s, 0.5H); 3.77 (s, 1.2H); 3.70, 3.71 (2s, 5.4H); 3.65 (s, 1.2H); 3.13 (m, 3.7H); 2.96 (m, 2.0H); 2.65-2.80 (m, 0.6H); 1.80-2.05 (m, 1.4H); 1.64 (m, 2.3H).

*2,4-Diamino-5-(3,5-O-ditrityl-2-deoxy-ß-D-ribofuranosyl)-pyrimidine **(4)***. **3** (3.98 g) in 50 mL dry THF was transferred via canula into a solution of guanidinium hydrochloride (205 mg) and solid sodium ethoxide (258 mg) in 50 mL THF in a 250 mL round bottom flask. The flask containing **3** was rinsed 1 x 25mL THF to ensure complete transfer. The solution was brought to reflux and stirred seven days. After cooling to room temperature, the reaction was quenched with aqueous NH_4Cl and partitioned between ethyl acetate and water. The combined organic layers were washed with water and dried over Na_2SO_4 prior to solvent removal. The title compound crystallized in 29% yield as a single anomer from 9:1 ethyl acetate:methanol. ^{1}HNMR (δ): 7.68 (s, 1H); 7.46 (m, 6H, Ph); 7.26-7.30 (m, 24H, Ph); 4.95 (dd, 1H, H1, J1=11.6 Hz,

J2=4.7 Hz); 4.63 (br s, 2H, NH2); 4.55 (d, 1H, H3, J=6.0 Hz); 4.19(d, 1H, H4, J=2.0 Hz); 3.15-3.28 (m, 2H, H5, 5'); 2.02 (m, 1H, H2); 1.24 (m, 1H, H2'). MS (ESI): calc. 710.33; found 711($M+H^+$).

*N^2,N^4-Dibenzoyl-2,4-diamino-5-(3,5-O-dityrityl-2-deoxy-ß-D-ribofuranosyl)-pyrimidine **(5)***. 40 mL dry dichloromethane was added to DMAP (cat.) and **4** (121 mg) in a 100 mL round bottom flask. Triethylamine (59 µL) followed by benzoyl chloride (42 µL) was added, and the solution brought to reflux. After 29 h, an additional 42 µL of benzoyl chloride was added the reaction refluxed an additional 18 h. After cooling to room temperature, the reaction was quenched with aqueous NH_4Cl and partitioned between water and dichloromethane. The combined organic layers were washed with dilute HCl and brine. After drying over Na_2SO_4, the solvent was removed and the resulting residue dried *en vacuo* overnight. Silica chromatography (1:1 hexanes:ethyl acetate) afforded compound **6**, which was taken up in CH_2Cl_2/MeOH (20 mL). 3 mL of concentrated NH_4OH was added to the solution and the mixture stirred for 45 minutes. After TLC (1:1 hexanes:ethyl acetate) showed complete consumption of **6**, solvent was removed and the residue taken up in ethyl acetate and washed with dilute HCl, followed by brine. The title compound was purified by silica chromatography (1:1 hexanes:ethyl acetate) in 20% yield. The monobenzoylated product was isolated in 66% yield (based on moles of nucleoside) and recycled through the reaction sequence. ^{1}HNMR (δ): 10.32 (s, 1H, NH); 8.85 (s, 1H, NH); 8.22 (s, 1H, Ar); 7.91 (d, 2H, J=5.7 Hz); 7.98 (d, 2H, J=5.5 Hz); 7.36-7.55 (m, 13H); 7.11-7.22 (m, 27H); 5.27(dd, 1H, J1=10.7 Hz, J2=4.6 Hz); 4.18 (m, 1H); 4.11 (m, 1H); 3.08 (dd, 1H, J1=10.4 Hz, J2=2.5 Hz); 2.80 (m, 1H); 1.79 (m, 1H); 1.71 (m, 1H). MS(ESI): calc. 918.38; found 919

($M+H^+$), 925 ($M+Li^+$), 941 ($M+Na^+$), 957 ($M+K^+$). MS (ESI, monobenzoylated): calc. 814.35, found 815 ($M+H^+$), 837 ($M+Na^+$).

*N^2,N^4-Dibenzoyl-2,4-diamino-5-(2-deoxy-ß-D-ribofuranosyl)-pyrimidine **(7)***. **5** (53 mg) was dissolved in 3:2 ether:formic acid (10 mL) and stirred at room temperature. After ~30 minutes, the solvent was removed and the product purified by silica chromatography (5:1 ethyl acetate:methanol). The title compound was obtained in 63% yield and its identity confirmed by comparison to the published ^{1}HNMR spectrum[*].

*N^2,N^4-Dibenzoyl-2,4-diamino-5-(5-O-dimethoxytrityl-2-deoxy-ß-D-ribofuranosyl)-pyrimidine **(8)***. **7** (15.4 mg) and DMAP (2.8 mg) were dissolved in dry pyridine (5 mL) to which DMTrCl (18.0 mg) was added. The solution was stirred two days, during which an additional 18 mg was added to the reaction three times. Once TLC (5:1 ethyl acetate:methanol) showed complete consumption of starting material, the reaction was quenched by the addition of water, and the mixture partitioned between water and ethyl acetate. The combined organic layers were washed with brine, dried over Na_2SO_4 and evaporated to dryness. The residue was then purified by silica chromatography (49:1 $CHCl_3$:MeOH). The fraction containing the desired product was then subjected to a second round of chromatography to yield the title compound cleanly (14.4 mg, 55%). Its identity was confirmed by comparison to the published ^{1}HNMR spectrum[*].

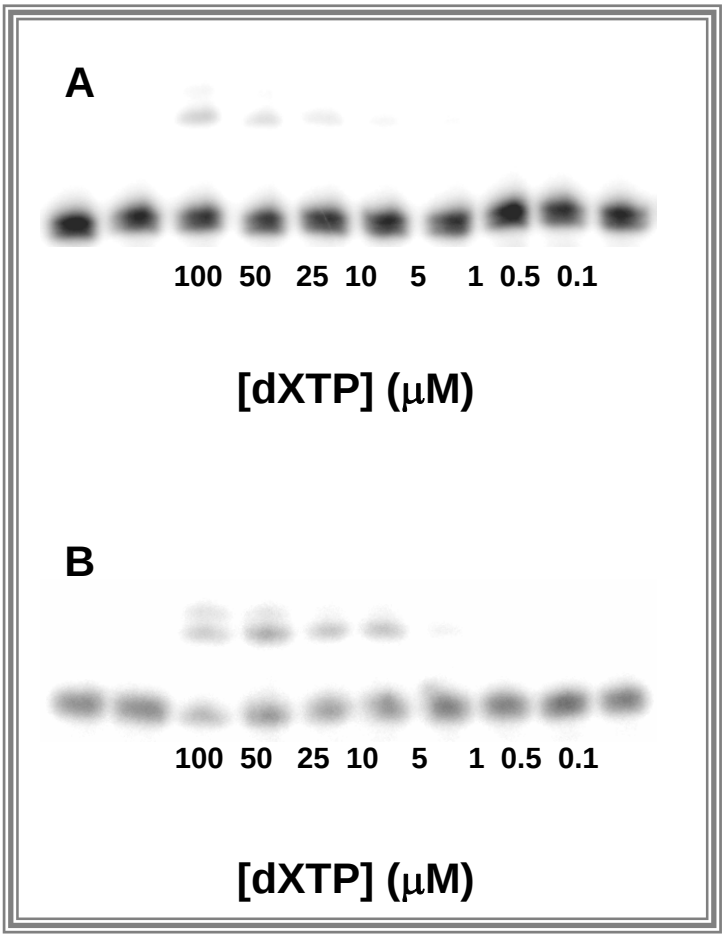
N^2,N^4-Dibenzoyl-2,4-diamino-5-(5-O-dimethoxytrityl-3-O-diisopropylcyanoethylphosphoramidite-2-deoxy-ß-D-ribofuranosyl)-pyrimidine. **8** was co-evaporated with toluene three times, followed by vacuum drying overnight. Molecular sieves ere

[*] Piccirilli, J. A. *The Evolution of Biological Catalysis*; Harvard University: Cambridge, MA, 1989.

added to the reaction flask, and the nucleoside (1.0 eq) was then dissolved in freshly distilled dichloromethane containing Hunig's base (2.0 eq).

Diisopropylcyanoethylchlorophosphoramidite (Aldrich) was then added, and the solution allowed to stir at room temperature for approximately three hours. The reaction was monitored by silica TLC (45:45:10 EtOAc:CH_2Cl_2:TEA), and often had to be stopped prior to complete conversion of starting material to avoid formation of the H-phosphonate at longer times. The reaction was quenched with several drops of dry MeOH and evaporated to dryness. The residue was immediately subjected to flash chromatography (45:45:10 EtOAc: CH_2Cl_2:TEA) to provide a pure phosphoramidite, which was dissolved in dry CH_3CN (0.1M final) and used immediately on an ABI394 DNA synthesizer.

B. Primer extension tests with DNAκ and dXTP.



A) 10μM primer/template DNA; B) 1μM primer/template DNA. In both panels, the lanes contained (left to right): no enzyme control, no dXTP control, and +dXTP in the concentration indicated on the Figure. All reactions contained 0.1U/μL KF.

[dXTP] (μM)	10μM Template	1μM Template
100	9%	47%
50	6%	41%
25	4%	30%
10	2%	27%
5	1%	6%
1	0.1%	2%
0.5	<0.1%	1%
0.1	<0.1%	0.9%
Control	<0.1%	0.9%

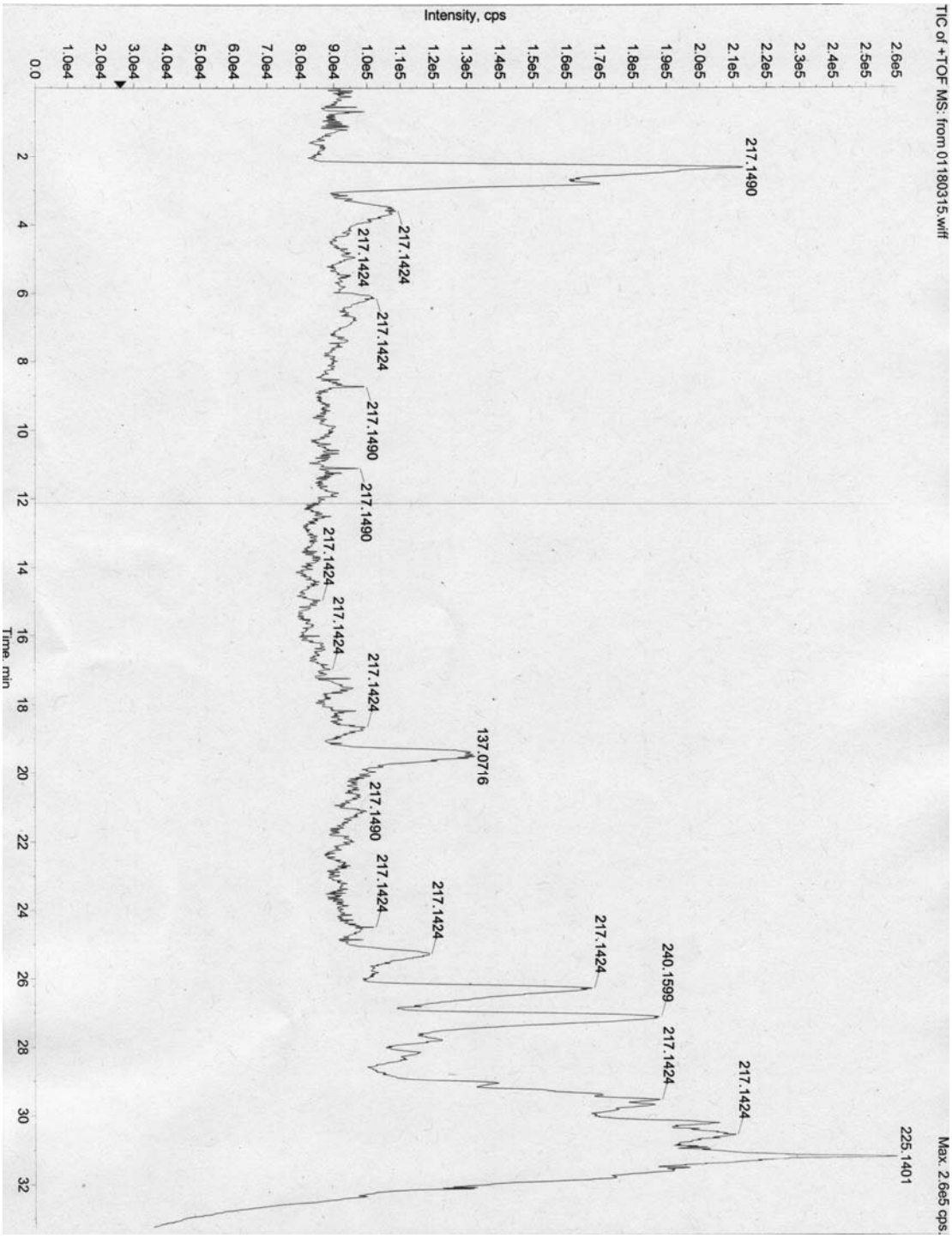
Percent of primer extended, as determined by phosphorimagery and calculated using ImageQuant software.

C. Mass Spectroscopy Analysis for Incorporation of dXTP.

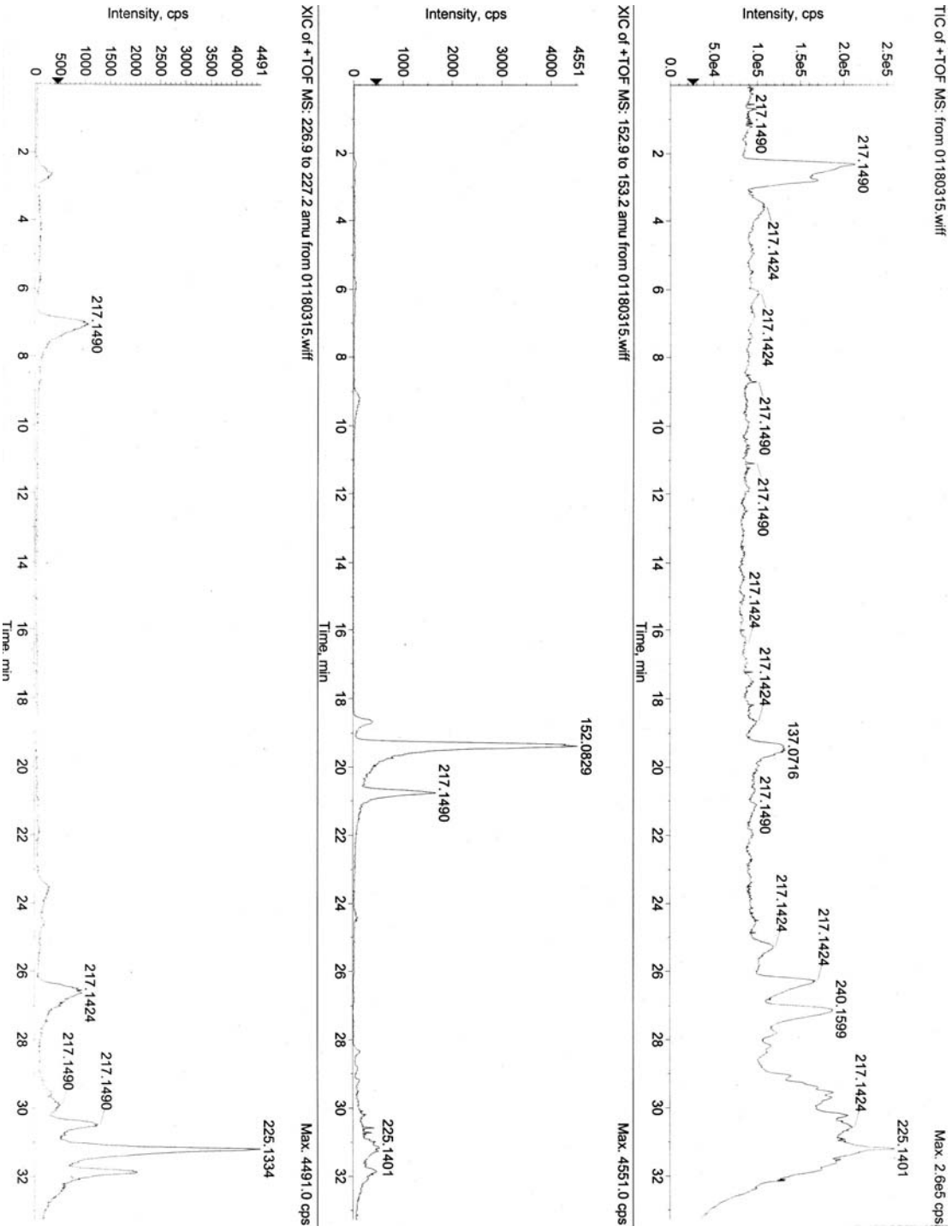
The following are representative data for the LC-MS analysis. For each sample listed in Table 2, the spectrum of ion counts versus time for a specific mass was extracted from the total ion count versus time data. This gives a chromatogram that shows the elution peak of the nucleoside corresponding to the specified mass. The mass spectrum for the time period of the elution peak is then extracted. The masses are calibrated using the internal PEG standards, and the mass of the unnatural nucleoside thus determined. From this mass, a molecular formula can be deduced, and the identity of the nucleoside inferred. Although more than one molecular formula may result in the detected mass, generally only a very small number of chemically reasonable formulae are possible.

The following graphs show the complete analysis for one sample (line 4 in Table 2).

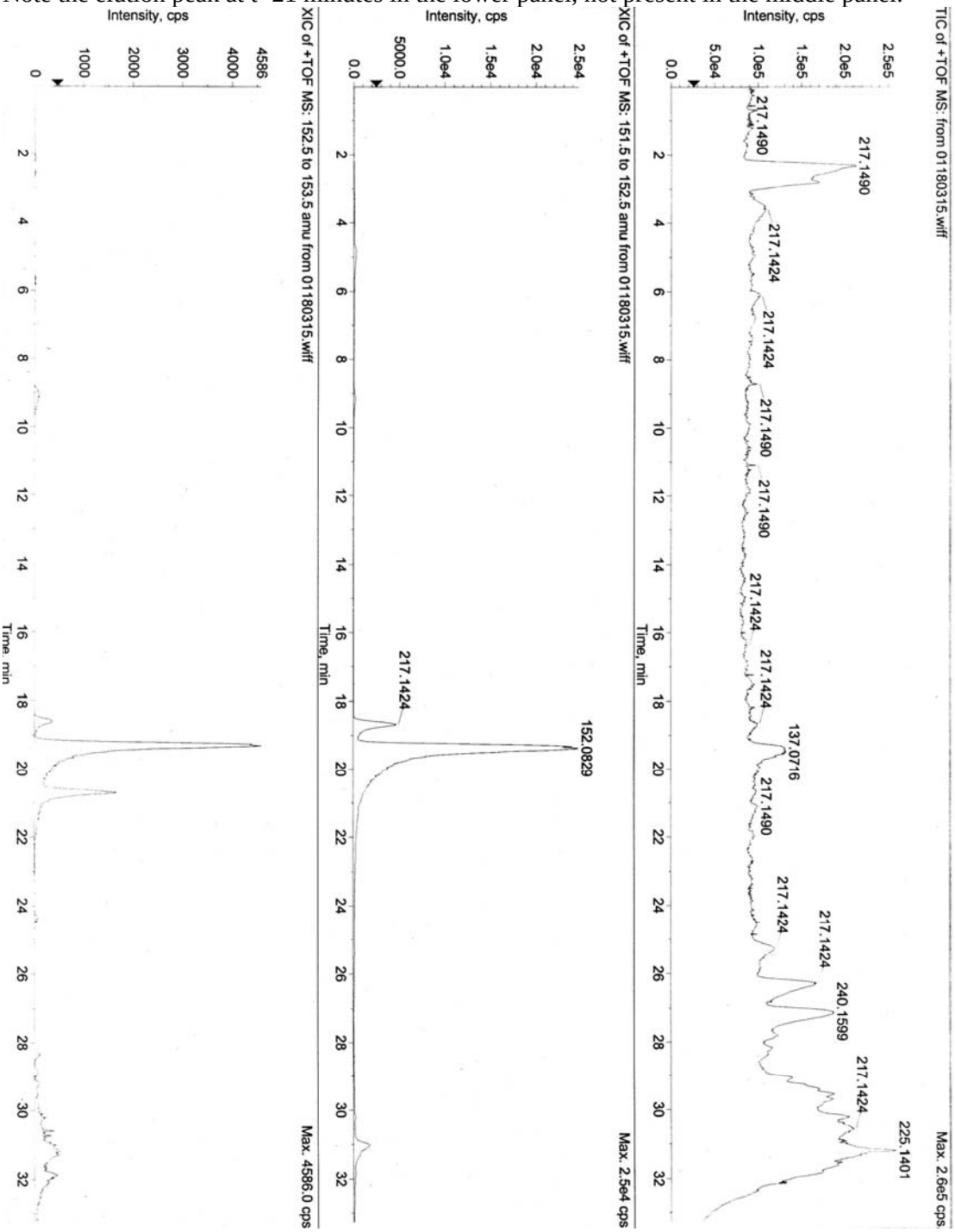
Total ion count vs. time. For each peak the most abundant mass is listed, and is almost always 217, one of the PEG standards.



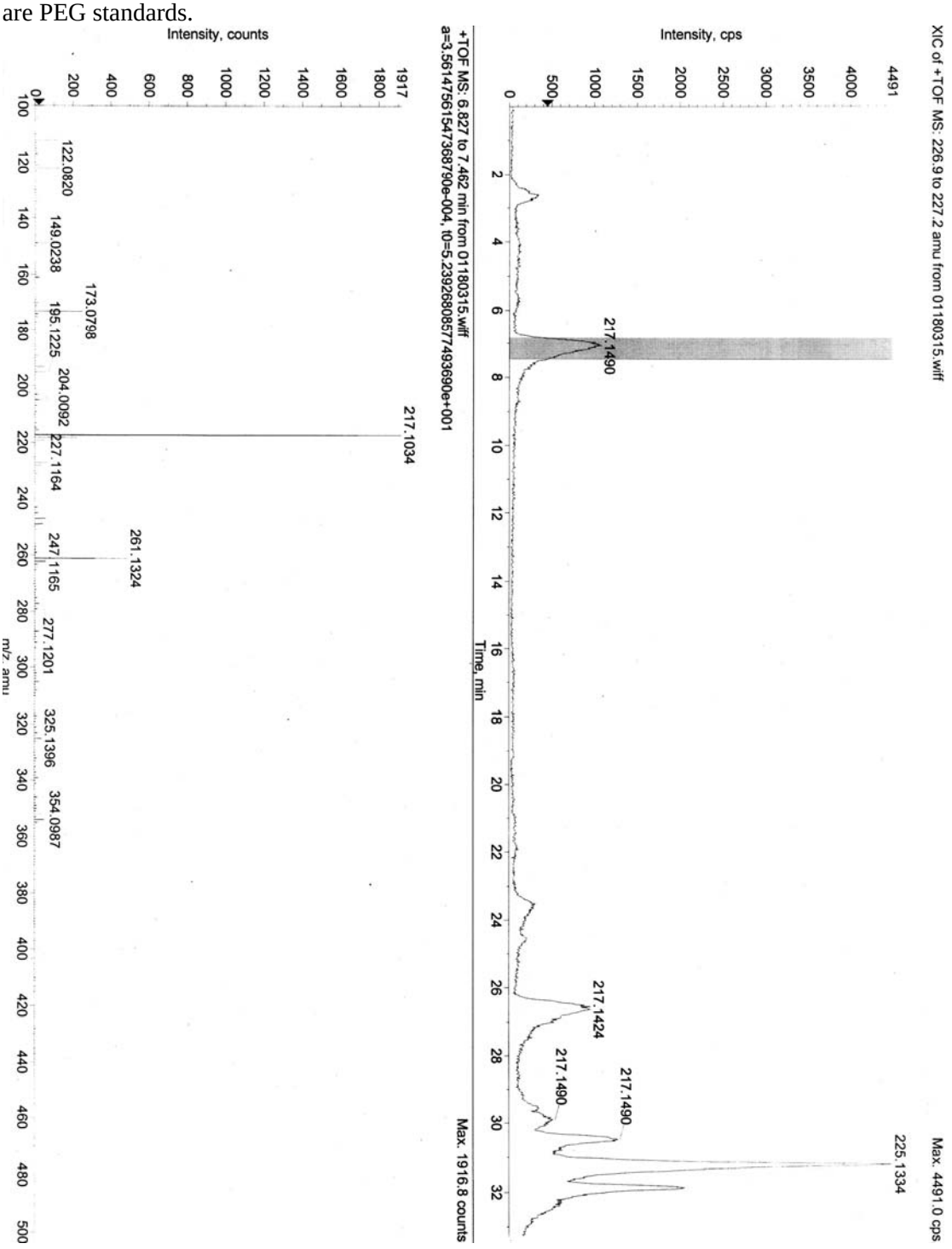
Upper panel: total ion chromatogram shown previously.
Middle panel: extracted spectrum of counts vs time for dX (mass of free base=153), showing the elution peak at approximately 21 minutes
Lower panel: extracted spectrum of counts vs time for dκ (mass=227), showing the elution peak at approximately 7 minutes.



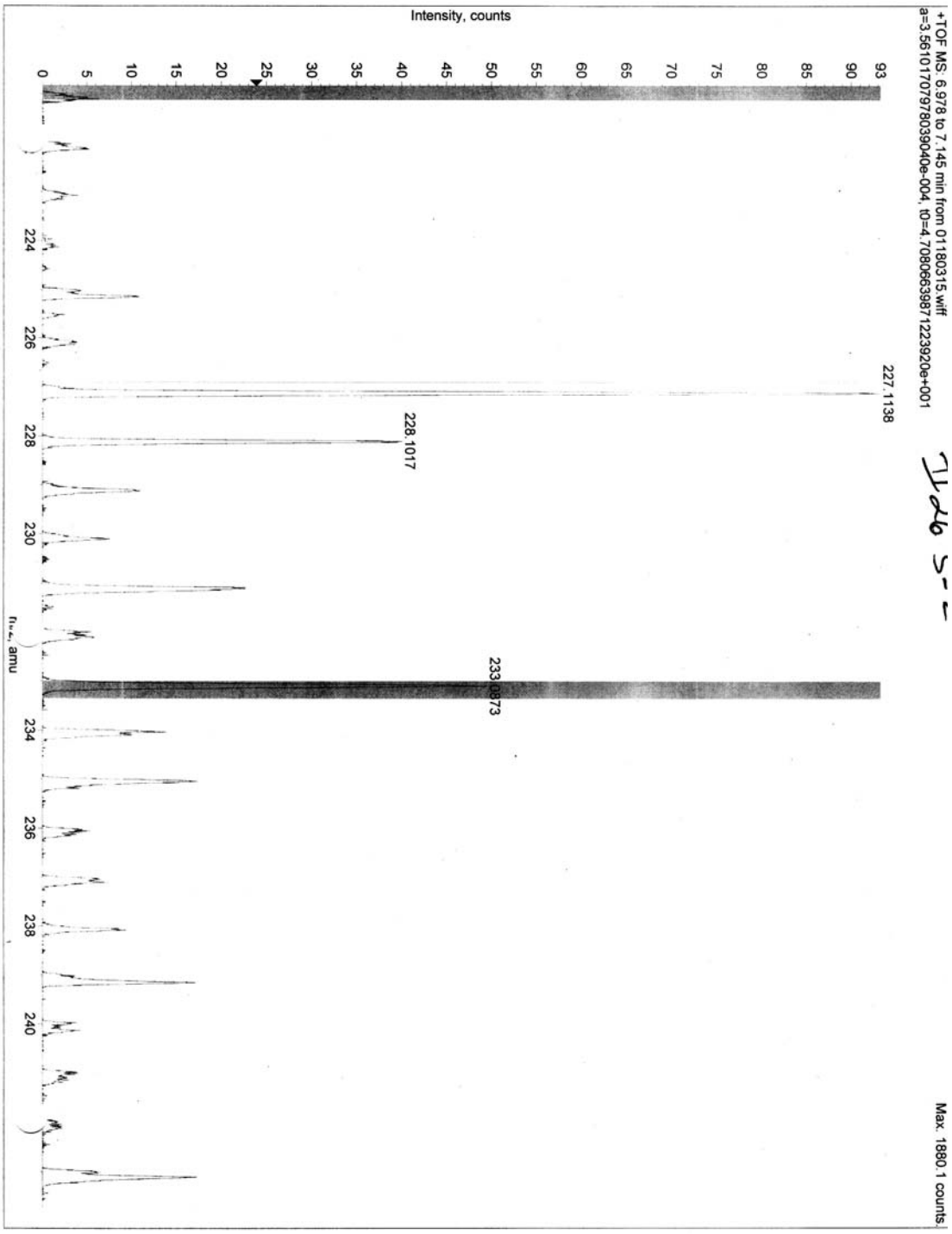
Comparison of extracted spectra for dG) and dX. Because these two nucleosides have similar masses and some overlap in the LC elution profile, it is necessary to separate them computationally.
Upper panel: total ion chromatogram
Middle panel: extracted spectrum for dG (mass of free base=152)
Lower panel: extracted spectrum for dX (mass of free base=153)
Note the elution peak at t=21 minutes in the lower panel, not present in the middle panel.



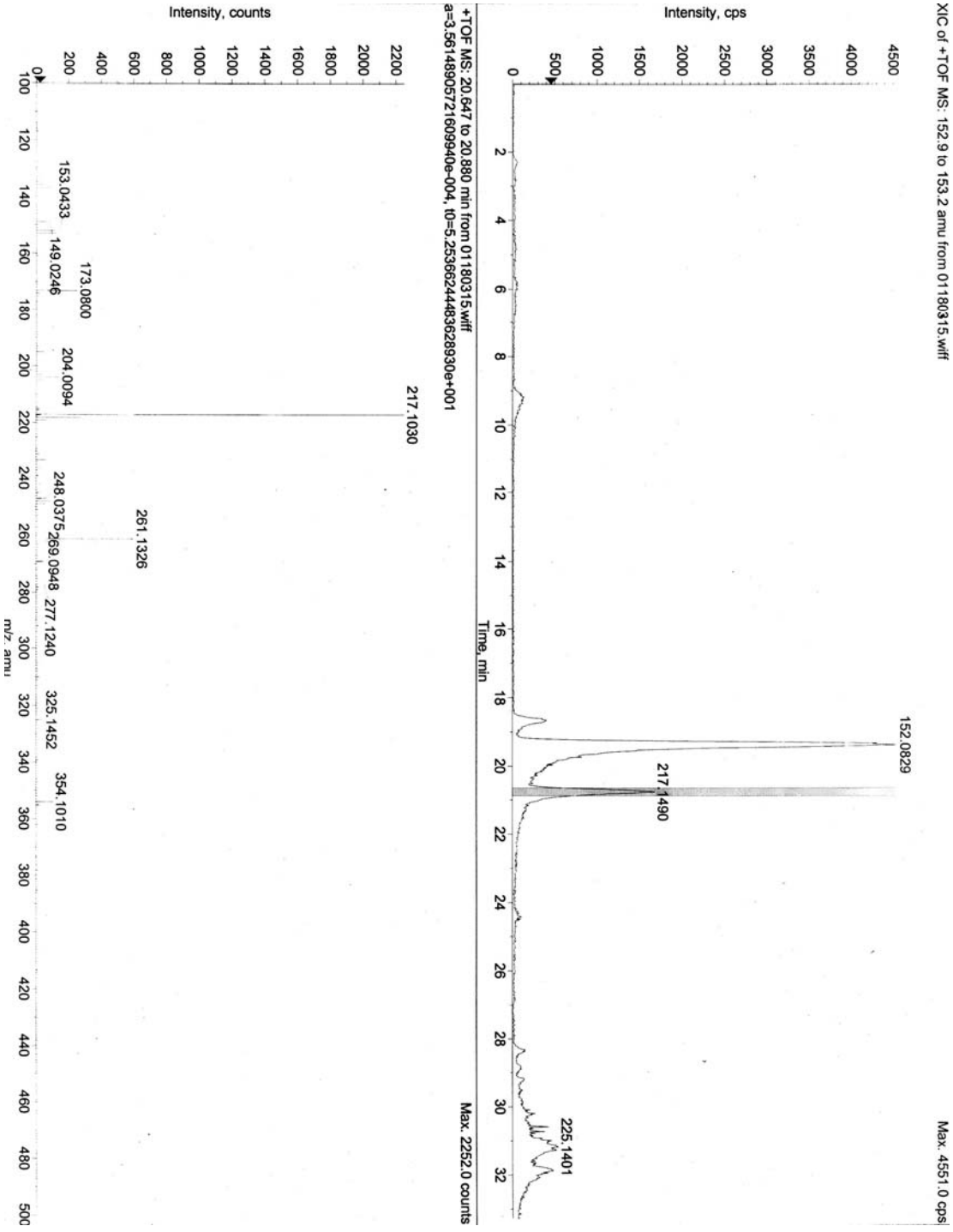
Detection of dκ and calculation of exact mass.
Upper panel: extracted spectrum of dκ (mass=227) with the elution peak highlighted.
Lower panel: mass spectrum (counts vs. mass) of elution peak highlighted above. The major peaks seen are PEG standards.



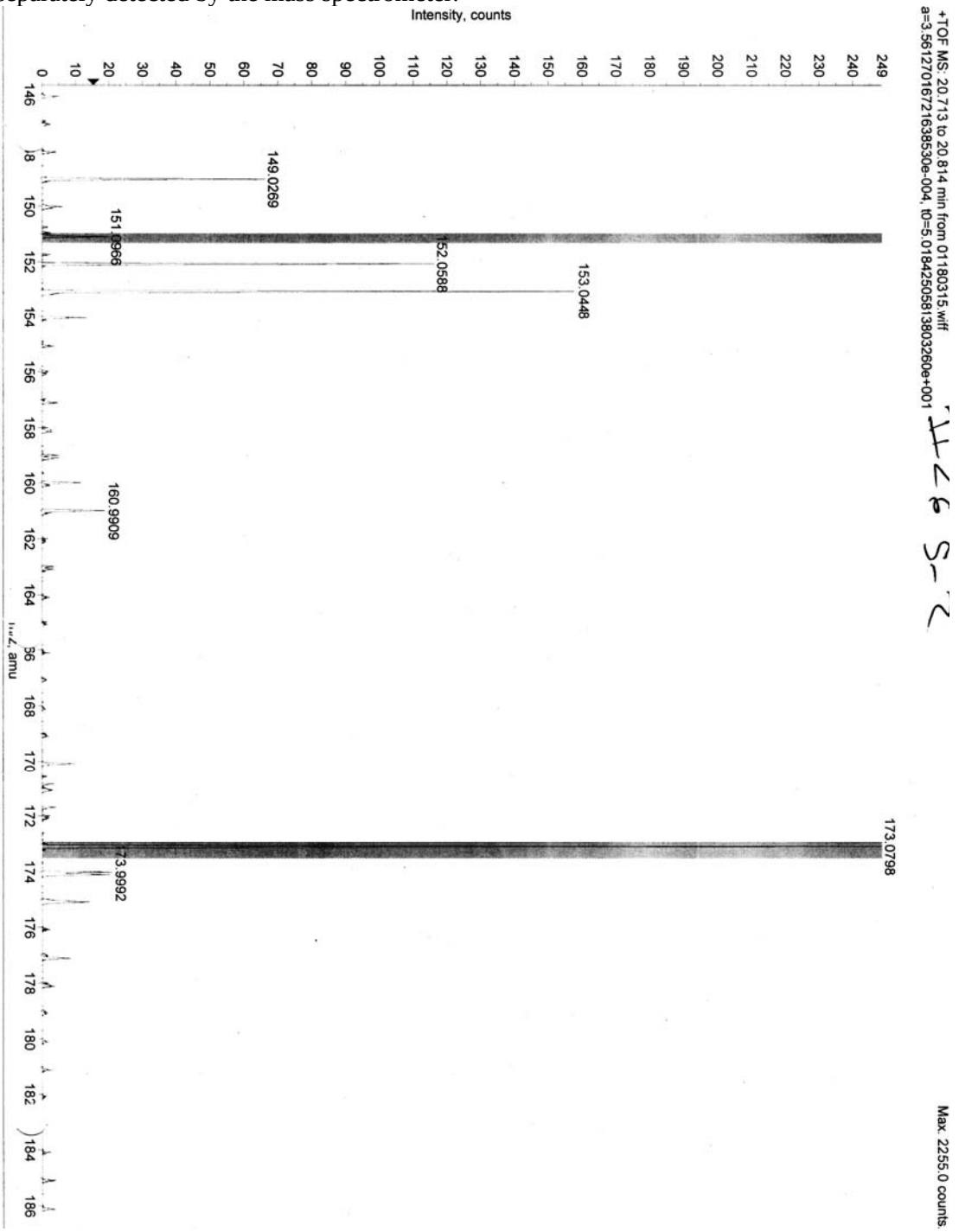
Exact mass calculation for the peak at 227 amu. The spectrum is recalibrated using the PEG standards, two of which are highlighted in this graph. This gives a mass of 227.1138.



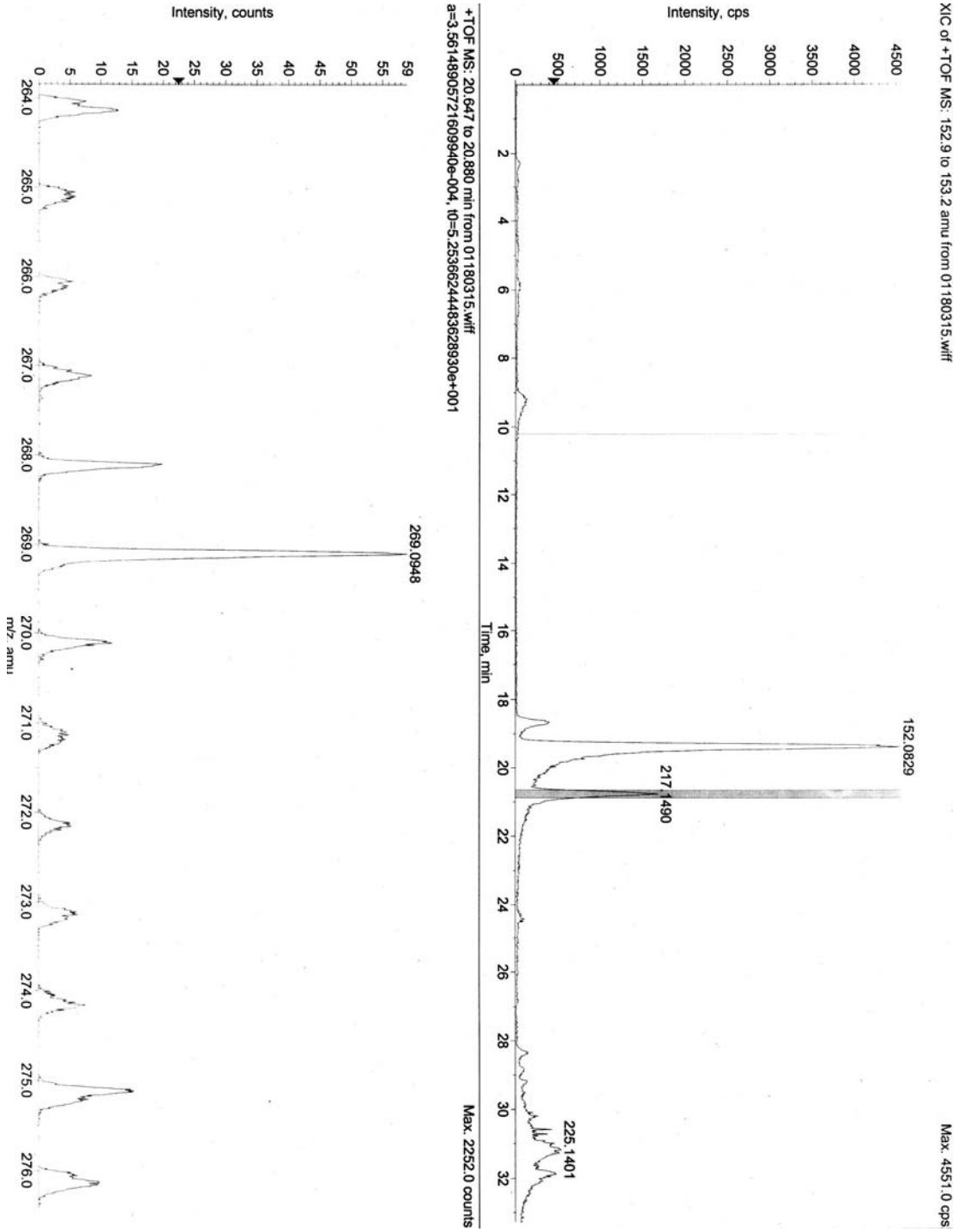
Detection of dX and calculation of exact mass.
Upper panel: extracted spectrum of dX [mass of free base is used (153)] with the elution peak highlighted.
Lower panel: mass spectrum (counts vs. mass) of elution peak highlighted above. The major peaks seen are PEG standards.



Exact mass calculation for the peak at 153 amu. The spectrum is recalibrated using the PEG standards, two of which are highlighted in this graph. This gives a highly accurate mass at 153.0448. Note the peak from dG at 152.0588. Despite overlap in the LC elution profile of these two nucleosides, they are easily separately detected by the mass spectrometer.



Confirmation of identity of dX.
Upper panel: extracted spectrum of dX (mass of free base=153) with the elution peak highlighted.
Lower panel: mass spectrum (counts vs. mass) of elution peak highlighted above. Although the free base is the major mass detected, the peak at 269 (full nucleoside) helps to confirm the identity of the nucleoside.



Molecular formulae calculations:

Using a tolerance of 5 millimass units (all of our detected masses were within 4 mmu, and most within 1mmu, so this represents a significantly outer limit for mass), we can calculate all possible molecular formulae for the detected masses, as well as each formula's DBE (double bond equivalent).

mass=153.0448 +/- 0.005 amu	DBE
H11N1O8	-4
C3H3N7O	6
C4H9O6	0.5
C5H5N4O2	5.5
C7H7NO3	5
C10H5N2	9.5

mass=227.1138 +/- 0.005 amu	DBE
C2H13N9O4	1
C3H9N13	6
C3H19N2O9	-4.5
C4H15N6O5	0.5
C5H11N10O	5.5
C6H17N3O6	0
C7H13N7O2	5
C8H19O7	-0.5
C9H15N4O3	4.5
C11H17NO4	4
C12H13N5	9
C14H15N2O	8.5

Because the ionization method is ESI, we can exclude those formulae that give an integer DBE (which would result from a radical ion).

That leaves only 3 possibilities for mass 153 and 6 for 227. Depending on preference, further refinement can be done by inspection or using a computerized screen. Either way, by excluding those formulae that do not fall within reasonable parameters for one's set of nucleosides (e.g. restrict the number of nitrogens and oxygens to be in a certain range for the library studied, or exclude any formula for a nucleoside with fewer than 5 carbons), the number of possibilities for each mass falls to only a small number.

For the above data set:
Mass =153. 0448: setting the number of nitrogens and oxygens between 2 and 6 leaves only one formula:$C_5H_5N_4O_2$, which corresponds to xanthine.
Mass=227.1138: setting N and O as above, plus C>5 (as this mass is high enough to correspond to a nucleoside rather than a free base) leaves only one formula as well, $C_9H_{15}N_4O_3$, which corresponds to dκ.

Even without the exclusions based on numbers of certain elements, which would change based on the nucleotides studied, the small number of possibilities that remain after the DBE analysis make manual inspection quite easy.