Research Article

Magnetic Resonance in Chemistry

Received: 8 February 2013 Revised: 26 May 2013 Accepted: 29 May 2013 Published online in Wiley Online Library: 23 June 2013

(wileyonlinelibrary.com) DOI 10.1002/mrc.3980

Systematic assignment of NMR spectra of 5-substituted-4-thiopyrimidine nucleosides

Xiaohui Zhang,[a]* Jian Wang[a] and Yao-Zhong Xu[b]**

Unambiguous characterization of 5-substituted-4-thiopyrimidine nucleosides (ribonucleosides and 2'-deoxynucleosides) was performed using NMR spectroscopy. Assignments of all proton and carbon signals of 5-bromo-4-thiouridine and related nucleosides were systematically carried out and firmly established by COSY and HMQC techniques. The NMR data of various 4-thiopyrimidine nucleosides are compared, and the key contributing factors discussed. The approach presented here is applicable to other modified nucleosides and nucleotides, as well as nucleobases. Copyright © 2013 John Wiley & Sons, Ltd.

Supporting information may be found in the online version of this article.

Keywords: NMR; ^{1}H; ^{13}C; 5-substituted pyrimidine; 4-thiothymidine; 4-thiouridine; 5-bromo-4-thiouridine; nucleoside; DNA; RNA

Introduction

Nucleic acids (DNA and RNA) are fundamental biomolecules, playing crucial roles in all forms of life. DNA is the genetic material for all living species except some viruses, whereas the primary biological function of RNA lies with the transmission (transcription and translation) of genetic information. There are only four deoxynucleosides used for the storage of genetic information. However, DNA bases and nucleosides are susceptible to damage by chemical, physical and biological actions.[1] Therefore, a great number of base-modified deoxynucleosides have been detected and documented.[2–4] As to RNA, whereas four ribonucleosides are used for the transcription of the genetic information, the number of other naturally occurring ribonucleosides is huge (over 100) because of the wider range of biological functions (see review[5]).

Nucleosides can be modified on either the sugar or the base. Both types of modified nucleosides are of biological relevance and medicinal uses. For instance, sugar-modified nucleosides, such as azidothymidine (AZT), acyclovir (Zovirax), famciclovir (Famvir) and 4′-thionucleosides[6], are primarily used as antiviral agents.[2] The alteration in the sugar moiety is designed to prevent further DNA elongation, thus inhibiting viral production. However, as genetic information is encoded in the bases in DNA and transmitted by the bases in RNA, base-modified nucleosides are certainly of great importance in various biological and medicinal research fields including the study of cancer and have been subjected to extensive studies (see review).[7]

Currently, the incidence of cancer remains extremely high and one in three people will have the disease in their lifetime. The fundamental cause of cancer can be ascribed to DNA damage and subsequent mutations.[1] Any agent or treatment that can lead to DNA damage could also be used to reduce and destroy cancerous cells. This, in fact, is the principle underlying chemotherapy and radiotherapy. However, more often than not, such anti-cancer drugs or treatments are toxic and indiscriminating and thus unsatisfactory. Obviously, better drugs and treatments are urgently required. For the development of anti-cancer drugs, we reported that 4-thiothymidine (**2b**, see Scheme 1 below), an analogue of the naturally occurring nucleoside thymidine (**1b**), can be readily incorporated into DNA of proliferative cells (such as cancerous cells) and activated by UVA light to kill cells.[8] To systemically exploit other 4-thionucleosides as potential anti-cancer drugs, we synthesized 5-bromo-4-thio-2′-deoxyuridine (**2e**)[9] and, more recently, other 5-halo-4-thio-2′-deoxyuridines (**2c, 2d** and **2f**) and their ribonucleosides (**4a-f**). As many of these 5-substituted-4-thionucleosides (IUPAC name: 5-substituted 4-thiooxo-3,4-dihydropyrimidin-2-one nucleosides) are synthesized for the first time, a full characterization of these compounds is essential. Among all analytical methods, nuclear magnetic resonance (NMR) spectroscopy is often the primary tool because it can uncover structural details. NMR is also a solution technique and therefore more likely to reflect the stereochemistry *in vivo*. X-ray crystallography could also accomplish this aim if a diffraction quality crystalline form of the nucleoside is obtainable.

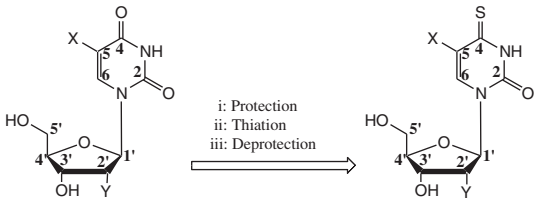
NMR spectroscopy has been extensively used for studies of naturally occurring nucleosides.[10] However, the data on modified nucleosides in literature are patchy and sometimes contradictory because of the rarity of these modified nucleosides.[11] There are few papers systematically examining nucleosides by use of NMR.[12] This has prompted us to carry out a systematic NMR examination of nucleosides, in particular, base-modified nucleosides. Previously, we reported an unambiguous approach to determining structures of purine-modified

* *Correspondence to: Xiaohui Zhang, College of Environment and Chemical Engineering, Dalian University, Dalian 116622, China. E-mail: zhangxiaohui@dlu.edu.cn*

** *Correspondence to: Yao-Zhong Xu, Department of Life, Health and Chemical Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK. E-mail: yao.xu@open.ac.uk*

a *College of Environment and Chemical Engineering, Dalian University, Dalian 116622, China*

b *Department of Life, Health and Chemical Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK*



1: Y=H
3: Y=OH

2: Y=H
4: Y=OH

Cpd	a	b	c	d	e	f
X=	H	CH_3	F	Cl	Br	I

Scheme 1. Transformation of 5-substituted nucleosides to their 4-thio-anaologues. The starting materials **1** (**Y** = H) and **3** (**Y** = OH) were converted to 4-thioanalogues **2** (**Y** = H) and **4** (**Y** = OH) *via* a three reaction process: protection of the OH groups of the sugar moiety, replacement of the oxygen atom at 4-position with a sulfur atom, then removal of the protecting groups on the sugar. **X** at 5-position represents an H atom, CH_3 (methyl group) or any of the halogens.

deoxynucleosides.[13] Recently, we extended our work to some 4-thiopyrimidine deoxynucleosides.[14] With the availability of all 5-halogenated-4-thionucleosides (ribo and deoxyribo), we were able to carry out a systematic and comparable NMR study of both types of nucleosides. Here, we take 5-bromo-4-thiopyrimidine ribonucleoside (**4e**) as an example to illustrate a general approach to the NMR study of pyrimidine-modified nucleosides. Together with our early work on purine-modified nucleosides,[13] we can offer a complete ^{1}H and ^{13}C NMR assignments for both purine-modified and pyrimidine-modified nucleosides and clarify some inconsistent NMR assignments in the literature.

Results and Discussion

The standard nucleosides fall into two types: purine and pyrimidine nucleosides. In our previous paper,[13] we studied NMR of purine nucleosides extensively. In this paper, our focus will be on pyrimidine nucleosides, namely, thymidine (**1b**) (of DNA) and uridine (**3a**) (of RNA) and their base-modified analogues, although the principles discussed here should be applicable to cytidine, another pyrimidine nucleoside. Scheme 1 shows the chemical structures of these compounds and the transformation of 5-substituted pyrimidine nucleosides (**1** and **3**) to their 4-thioanalogues (**2** and **4**).

Chemical synthesis

5-substituted-4-thiouridines (**4a-f**) and their deoxy-analogues (**2a-f**) were prepared by adapting methods reported in the literature.[15–17] A general synthetic approach consists of three stages: (i) first, protecting the hydroxyl groups on the sugar by acylation; (ii) then replacing the oxygen atom at 4-position with a sulfur atom using suitable thiating reagents (such as P_2S_5[18]); and (iii) finally, removing the protecting groups with milder deprotecting reagents (such as NH_3) to produce the target compounds.

NMR peak assignment

Sugar protons

The first step in NMR structural elucidation is to assign all the NMR signals (peaks) to their corresponding atoms (e.g. H or C). Proton assignment is always the initial step as it is relatively easy to acquire ^{1}H NMR spectra because of the abundance of NMR-sensitive protons in most organic molecules (including pyrimidine nucleosides). Here, we used 5-bromo-4-thiouridine, **4e**, to exemplify an unambiguous route to assigning all the sugar protons.

Although 5-bromo-4-thiouridine is a doubly modified nucleoside, however, their modifications take place at 4- and 5-positions of the base and are thus expected to have little effects on the sugar protons in terms of their chemical shifts. ^{1}H NMR spectrum of **4e** is shown in the supplementary material.

By comparison with the reported data for un-modified nucleosides,[13,19] we can be confident that the chemical shifts of all the sugar protons would be below 6 ppm (Table 2). The protons of the sugar OHs can be readily identified by the means of D_2O exchanges. Thus, we can tentatively assign the peaks below 6 ppm to their corresponding sugar protons. Our above assignments are further supported by the COSY spectrum as shown Fig. 1.

Figure 1 shows the sugar part of the ^{1}H-^{1}H Correlation Spectroscopy (COSY) spectrum of **4e** (5-bromo-4-thiouridine). COSY is commonly used to identify nuclei (such as protons in the case of ^{1}H-^{1}H COSY) that exhibit a scalar (J) coupling. The presence of off-diagonal peaks (cross-peaks) in the spectrum directly correlates the coupled partners (the protons in the case of ^{1}H-^{1}H COSY). It has long been established[19] that 1′-H has the highest chemical shift value among all sugar protons because of the deshielding effect of the electronegative N atom (at the glycosidic position) in the base and the O atom (at 4′-position) in the sugar. Thus, the peaks at δ 5.67 ppm can be confidently assigned to 1′-H. Subsequently, we can start from the assigned 1′-H on the diagonal line (indicated by a dotted line in Fig. 1) to trace its cross-peak with 2′-H signal, then following the arrows to identify 2′-H (δ 4.07) on the diagonal line. From 2′-H signal on the diagonal line, we can easily find its cross-peak with 2′-OH (δ 5.48). Using the same arrow approach, all other sugar protons can be identified (namely, 3′-H, 3′-OH, 4′-H, 5′-OH, 5′-H_a and 5′-H_b). The ^{1}H-^{1}H COSY spectrum also clearly show that these two chemically equivalent 5′-H_a and 5′-H_b atoms are not magnetically equivalent as evidenced by the presence of two sets of signals for these two 5′-H atoms.

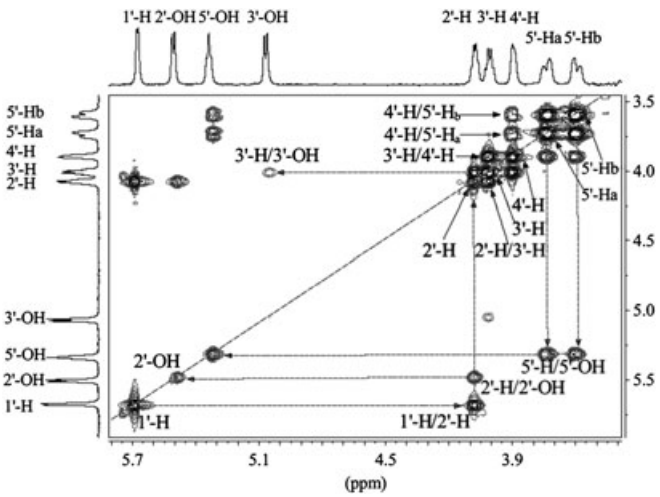


Figure 1. The sugar section of H-H COSY spectrum of 4e (5-bromo-4-thiouridine). Peaks on the diagonal line are the NMR signals of the protons on the sugar. The off-diagonal peaks (cross-peaks) directly correlate the coupled protons. The arrows indicate routes to the coupled partners.

Table 1. ^{1}H NMR of 5-substituted-2'-deoxyuridines and their 4-thioanalogues (in DMSO-d_6)

Cpd	1'-H	2'-H	3'-H	4'-H	5'-H	3'-OH	5'-OH	6-H	**NH**	5-H (CH_3)
1a	6.12	2.47	4.21	3.76	3.54	5.24	5.01	7.84	**11.28**	5.67
1b	6.15	2.49	4.22	3.74	3.55	5.21	5.01	7.68	**11.25**	(1.79)
1c	6.12	2.10	4.25	3.78	3.62	5.25	5.15	8.21	**11.82**	
1d[20]	6.08	2.15	4.21	3.77	3.57	5.22	5.14	8.28	**11.80**	
1e[20]	6.07	2.19	4.21	3.77	3.57	5.22	5.14	8.36	**11.76**	
1f[20]	6.10	2.12	4.24	3.79	3.58	5.24	5.14	8.40	**11.66**	
2a	6.07	2.11	4.22	3.81	3.56	5.26	5.03	7.78	**12.68**	6.30
2b	6.10	2.14	4.24	3.79	3.59	5.27	5.10	7.89	**12.69**	(1.96)
2c	6.05	2.18	4.25	3.82	3.60	5.28	5.21	8.26	**13.06**	
2d	6.06	2.21	4.27	3.84	3.58	5.21	5.21	8.43	**13.07**	
2e	6.02	2.21	4.22	4.09	3.61	5.26	5.21	8.52	**13.08**	
2f	6.01	2.17	4.23	3.81	3.59	5.25	5.18	8.55	**12.99**	

The above assignment of NMR peaks of 5-bromo-4-thiouridine (**4e**) is in good agreement with those of its related 5-bromo-uridine (**3e**) (Table 2) and 5-bromo-4-thiodeoxyuridine (**2e**) (Table 1). The excellent agreement offers a solid cross-check of our sequential approach to assigning nucleoside sugar protons. Because the sugar moieties are not altered in base-modified nucleosides, this method would thus be applicable to the sugar protons in other base-modified nucleosides. Therefore, the same approach has been used to assign the sugar protons of all 5-substituted-4-thiopyrimidne deoxynucleosides (**2a-f**) and ribonucleosides (**4a-f**) examined in this paper. The chemical shifts are summarized in Tables 1 and 2.

Pyrimidine protons

Pyrimidine bases have few protons, namely, 3-NH and 5-H and 6-H. When the 5-position is substituted for by a bromo group, for instance, in the case of **4e**, there are only two protons for assignment, namely the imino proton at the 3-N-position (exchangeable) and the 6-H proton (non-exchangeable). The exchangeable proton, 3-NH, can be readily identified by using D_2O exchange. Furthermore, because of the strong deshielding effect of the N atom (in the form of two amide groups), the imino proton always appears at lower field (with higher δ value). Thus, the exchangeable peak at around 13 ppm can be undoubtedly assigned as the 3-NH. It is worth noting that when the oxygen at the 4-position is replaced by a sulfur atom, the 3-NH shifts further downfield from 11.82 (in 5-bromouridine, **3e**) to 13.02 (in 5-bromo-4-thiouridine **4e**). This also holds true for other 5-modified 4-thio-nucleosides (Tables 1 and 2). However, it is also worth pointing out that the signals of exchangeable protons are often broad, and their chemical shifts can vary somewhat depending upon the solvents and conditions used.

Pyrimidine-modified nucleoside **4e** has one non-exchangeable pyrimidine proton appearing 8.65 (Table 2); the peak is readily assigned to the proton at the 6-position. Other pyrimidine-modified nucleosides (**2a-f** and **4a-f**) have also been examined, and their proton chemical shifts are listed in Tables 1 and 2, from which the following common features can be summarized:

a) All the sugar protons appear between δ 2.0 to 6.0 ppm.
b) 1'-H has the highest δ value (around 6 ppm) among all sugar protons and is in the form of a doublet for ribonucleosides and two doublets for deoxynucleosides. The remaining sugar protons always give multiplets.
c) 6-H always appears as a singlet and shows the highest δ value among all non-exchangeable protons derived from the nucleosides.
d) 2'-OH (doublet, existing only in ribose) and 3'-OH (doublet) and 5'-OH (triplet) protons are exchangeable as well as the

Table 2. ^{1}H NMR of 5-substituted uridines and their 4-thioanalogues (in DMSO-d_6)

Cpd	1'-H	2'-H	3'-H	4'-H	5'-H	2'-OH	3'-OH	5'-OH	6-H	**NH**	5-H (CH_3)
3a	5.79	4.03	3.97	3.85	3.56-3.63	5.40	5.10	5.10	7.896	**11.3**	5.66
3b	6.10	4.31	4.27	4.11	3.84-3.88	5.82	5.82	5.13	7.73		(1.77)
3c[23]	5.91	4.37	4.36	4.15	3.88-4.00	N/A	N/A	N/A	8.17	**N/A**	
3d[20]	5.73	4.05	3.99	3.87	3.54-3.74	5.43	5.08	5.28	8.41	**11.85**	
3e[20]	5.70	4.01	3.96	3.84	3.59-3.67	5.40	5.04	5.25	8.45	**11.78**	
3f[20]	5.72	4.01	3.97	3.87	3.53-3.73	5.41	5.07	5.26	8.48	**11.68**	
4a	5.72	4.03	3.96	3.86	3.54-3.66	5.44	5.10	5.10	7.82	**12.65**	6.31
4b	5.75	4.06	3.99	3.87	3.56-3.70	5.44	5.08	5.19	7.96	**12.71**	(1.96)
4c	5.68	4.05	4.00	3.88	3.58-3.74	5.49	5.08	5.31	8.37	**13.01**	
4d	5.68	4.08	4.01	3.86	3.58-3.76	5.52	5.08	5.34	8.57	**13.12**	
4e	5.67	4.07	4.01	3.81	3.49-3.80	5.50	5.05	5.34	8.65	**13.02**	
4f	5.68	4.06	3.89	3.89	3.58-3.72	5.40	5.10	5.30	8.68	**13.00**	

3c: in CD_3COCD_3.

imino proton (3-NH). The latter has the highest δ value in the whole molecule of each nucleoside. These exchangeable protons are distinguishable and easily singled out by D_2O exchange.

^{13}C peaks of the sugar and pyrimidine

A ^{13}C NMR spectrum of 5-bromo-4-thiouridine was obtained (shown in supplementary material). The ^{13}C NMR signals are initially assigned in the 1D spectrum and then verified by using a Heteronuclear Multiple-Quantum Correlation (HMQC) technique. HMQC is used to correlate directly bonded nuclei (in this case, carbon-proton nuclei) and can offer structural information on the bonded atoms. This approach has the same principle as that of COSY used to assign ^{1}H signals. Because each of the sugar carbons has at least one proton attached, the ^{13}C peaks of **4e**, for example, can be readily assigned to their carbon atoms from the known protons by HMQC as demonstrated in Fig. 2 below.

Assignment of sugar carbons

In the previous section, we have unambiguously assigned ^{1}H signals to their corresponding protons in the sugar. Now, we can easily trace from the assigned ^{1}H protons to their coupled ^{13}C partners. For instance, the ^{1}H signal located at δ 5.67 ppm has been confirmed to be the 1′-H of **4e** (Table 2). We can find a cross-peak from which we can trace to its coupled ^{13}C partner, that is, 1′-C at δ 89.34 ppm (Fig. 2). In a similar vein, we can allocate the signals for 2′-C, 3′-C, 4′-C and 5′-C. It is also reassuring that there is no cross-peak for the sugar OHs as the protons in the OH groups are not directly linked to any carbon.

Assignment of pyrimidine carbons

The pyrimidine has two types of carbons, ones with proton attached (e.g. 6-C in **4e**) and the others without (e.g. 2-C, 4-C and 5-C in **4e**). The former can be readily identified by using the same HMQC technique. Taking **4e** again as an example, as the peak of singlet (δ 8.65 ppm) has been previously assigned as 6-H (cf. Table 2), the 6-H peak can be traced in the HMQC spectrum (Fig. 2) to identify a cross-peak leading (*via* a broken line) to its coupled carbon (δ 137.37 ppm).

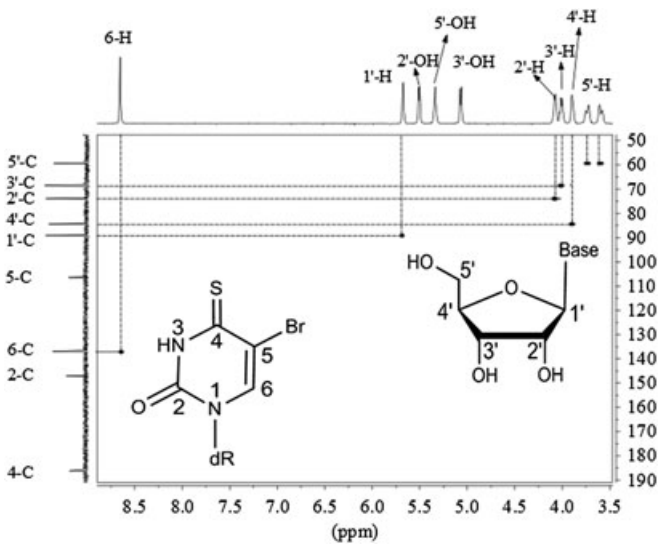


Figure 2. H-C COSY spectra of 4e (5-bromo-4-thiouridine). ^{1}H NMR is shown at the top. ^{13}C NMR is shown in the left side. Inserts are chemical structures of the sugar and base.

To assign the remaining pyrimidine carbons (i.e. 2-C, 4-C and 5-C) is more challenging as these carbon atoms do not bear any proton; thus, HMQC is no use in this instance. HMBC (heteronuclear multiple bond correlation) is a possible means; however, the only non-exchangeable proton in the base is at 6-position and is of three bond distance to both 2-C and 4-C. Clearly, HMBC is unlikely to provide a conclusive answer to the question of distinguishing between 2-C and 4-C. On the other hand, the chemical shift of any atom (^{13}C in this case) is influenced by its atomic surroundings. 5-Bromo-4-thiouridine (**4e**) is a nucleoside modified at 4-C and 5-C positions. By comparing NMR data of **4e** with those of its related compounds, it is possible to reveal how the modifications at 4-C and 5-C positions influence the chemical shifts of the pyrimidine carbons. Therefore, such a comparison could offer a useful route to the assignment of the ^{13}C NMR signals from these three carbons.

First, we prepared 5-fluoro-4-thiouridine (**4c**) and examined its ^{19}F and ^{13}C spectra. Only a single peak at 151.7 ppm is observed in its ^{19}F NMR spectrum (see the supplementary material) and is not informative. ^{13}C NMR appears more useful. The pyrimidine ^{13}C peaks are tentatively assigned and summarized in Table 3. Fluorine is an NMR sensitive atom and can couple with ^{13}C signals through bonding. The ^{13}C signal at δ 147.74 has the largest coupling constant (J = 217 Hz), so this carbon must be the carbon directly bonded with the fluorine atom and thus can be unambiguously assigned as 5-C. Both neighboring carbon atoms (i.e. 4-C and 6-C) have weak couplings with the fluorine atom, but they can be easily distinguished. This is because 6-C has a proton attached with it, and its ^{13}C signal (at δ 122.13) can be readily confirmed using ^{1}H-^{13}C HMQC technique as discussed above. Thus, the other fluorine-coupled peak (at δ 180.67) with a lower coupling constant (J = 30Hz) can be confidently assigned as 4-C. The ^{13}C signal located at δ 146.85 and not coupled with fluorine is certainly from 2-C, the remaining pyrimidine carbon atom.

The chemical structure of 5-bromo-4-thiouridine (**4e**) is very similar to that of 5-fluoro-4-thiouridine (**4c**). The only difference lies with the substituent at 5-position; thus, this difference would be reflected primarily in the chemical shifts of 5-C and less in the shifts of 2-C and 4-C atoms. Therefore, the little-shifted ^{13}C signals at 147.37 and 186.45 should be assigned to 2-C and 4-C of **4e**, whereas the substantially shifted ^{13}C signals (at 106.70) can confidently be assigned to 5-C (Table 3). This initial assignment is also consistent with the data from its deoxy analogue, namely, 5-bromo-4-thio-2′-deoxynucleoside (**2e**) (also see Table 4 and Fig. 3).

Our approach involves the following steps: (i) use HMQC to assign 6-C, (ii) use ^{19}F-^{13}C coupling to identify 5-C (larger coupling constant) and 4-C (smaller coupling constant), and (iii) recognize the fact that 2-C is away from and little affected by the modifications. Using this approach, all ^{13}C peaks from a number of 5-substituted-4-thiopyrimidine nucleosides have been unambiguously assigned and are listed in Tables 4 and 5, from which the following conclusions can be drawn:

a) All of the sugar carbons have δ values lower than 100 ppm. 1′-C has a higher value than most of other sugar carbons. 2′-C (in deoxynucleoside) has the lowest δ value and its signals are located near 40 ppm, sometimes buried within the signals of DMSO-d_6 when used as the solvent.

Table 3. Comparison of ^{13}C NMR chemical shifts of the carbons in bases

	2-C	4-C	5-C	6-C
4c	146.85	180.67 $^{2}J = 30$ Hz	147.74 $^{1}J = 217$Hz	122.13 $^{2}J = 41$Hz
4e	147.37	186.45	106.70	137.37
2e	147.12	186.33	106.57	137.18

Table 4. ^{13}C NMR of 5-substituted-2'-deoxyuridines and their 4-thioanalogues (in DMSO-d_6)

Cpd	1'-C	2'-C	3'-C	4'-C	5'-C	2-C	4-C	5-C	6-C	CH_3
1a	84.11	39.63	70.43	87.40	61.29	150.45	163.14	101.76	140.53	
1b	83.73	39.40	70.43	87.24	61.33	150.46	163.75	109.36	136.12	12.26
1c[21]	84.5	42.05	70.1	87.5	61.0	149.0	157.1	141.1	124.9	
							156.9	138.8	124.5	
							$^{2}J = 20$Hz	$^{2}J = 230$Hz	$^{1}J = 40$Hz	
1d[20]	84.73	40.09	70.02	87.45	60.89	150.40	160.09	107.27	137.39	
1e[20]	84.86	40.13	69.96	87.57	60.79	149.74	159.17	95.66	140.27	
1f[20]	84.63	40.18.	69.99	87.50	60.80	150.09	160.48	69.25	145.03	
2a	85.00	N/A	70.10	87.70	61.00	147.70	190.00	112.60	135.90	
2b	84.70	39.90	70.10	87.70	61.00	147.80	190.70	117.70	133.50	17.00
2c	85.36	40.06	69.64	87.75	60.63	146.63	180.60	147.73	122.12	
							180.30	145.56	121.71	
							$^{2}J = 30$Hz	$^{1}J = 213$Hz	$^{2}J = 41$Hz	
2d	85.65	40.32	69.46	87.79	60.44	146.89	185.32	116.77	134.67	
2e	85.61	40.31	69.43	87.77	60.39	147.12	186.33	106.57	137.18	
2f	85.39	39.50	69.50	87.74	60.41	147.63	189.31	82.99	140.15	

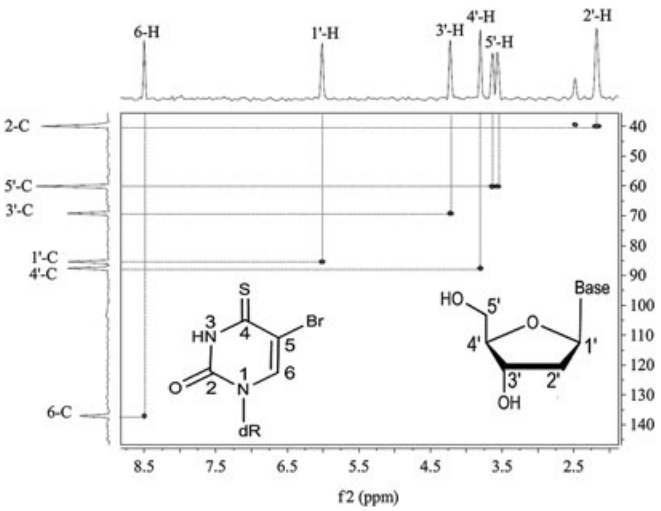


Figure 3. H-C COSY spectra of 2e (5-bromo-4-thiodeoxyuridine). ^{1}H NMR is shown at the top. ^{13}C NMR is shown in the left side. Inserts are chemical structures of the sugar and base. The chemical shift for 4'-C (δ: 87.77) is clearly higher than that 1'-C (δ: 85.61).

b) All of the pyrimidine carbons have δ values higher than 90 ppm and higher than the sugar carbons except in 5-iodo analogues where the δ values are affected by the heavy atom effect (discussed below).

c) The carbon at 4-position has the highest δ value among all the carbons. The sulfur atom at 4-position shifts δ for 4-C to even higher values. In all cases, 4-C of 4-thiouridines has a higher δ value than that of 4-oxy-uridine analogues.

d) The following order 4-C> 2-C> 6-C> 5-C has been noted except in the case of fluorine-substituted nucleosides when the strong electronegativity of fluorine shifts its bonded C-5 toward a higher δ value.

Further discussion

Comparison between ribonucleosides and deoxyribonucleosides

Tables 1,2 and 4,5 list ^{1}H and ^{13}C chemical shifts of base-modified deoxynucleosides and ribonucleosides. There is little difference in the chemical shifts of the bases between these two types of nucleosides. However, interesting differences were noted in chemical shifts of the sugar part. As expected, the major difference is related to the 2'-position. 2'-H and 2'-C in all of the deoxynucleosides (Tables 1 and 4) appear at higher fields (i.e. with lower δ values) than those in the ribonucleosides (Tables 2 and 5). This can be attributed to the presence of the electronegative oxygen atom of 2'-OH in the latter. The orders of chemical shifts for the sugar protons are the same for both types of nucleosides, namely, 1'-H> 3'-H> 4'-H >5'-H, indicating the deshielding effects from the presence of 2'-OH on the sugar protons (except 2'-H) is minimal.

However, the deshielding effects derived from the presence of 2'-OH on the sugar carbons (except 2'-C) are more complex. ^{13}C signals from 1'-C and 4'-C often appear very closely; however, the presence (or lack) of 2'-OH in ribonucleosides (or in deoxyribonucleosides) would seriously affect the chemical shifts of 1'-C and 4'-C and possibly change their order. This is why the data from one-dimensional ^{13}C NMR in the

Table 5. ^{13}C NMR of 5-substituted uridines and their 4-thioanalogues (in DMSO-d_6)

	1'-C	2'-C	3'-C	4'-C	5'-C	2-C	4-C	5-C	6-C	CH_3
3a	87.71	73.46	69.80	84.77	60.80	150.66	163.01	101.66	140.64	
3b	89.18	74.25	70.24	85.01	61.60	152.59	168.19	112.31	138.21	12.8
3c[24]	88.77	74.31	69.94	85.22	60.89	149.80	157.68	141.38	125.53	
							157.48	139.56	125.25	
							^2J = 25Hz	^1J = 227Hz	^2J = 35Hz	
3d[20]	88.50	73.93	69.24	84.68	60.14	149.74	158.99	107.12	137.89	
3e[20]	88.49	73.94	69.25	84.67	60.11	149.97	159.14	95.69	140.35	
3f[20]	88.27	73.95	69.31	84.73	60.18	150.35	160.48	96.17	145.11	
4a	88.5	74.0	69.6	84.7	60.5	148.0	190.2	112.6	136.0	
4b	88.24	73.72	69.43	84.81	60.43	147.99	190.72	117.56	133.45	16.82
4c	89.08	74.06	68.90	84.63	59.87	146.85	180.67	147.74	122.13	
							180.37	145.57	121.72	
							^2J = 30Hz	^1J = 217Hz	^2J = 41Hz	
4d	89.36	74.13	68.72	84.57	59.63	147.07	185.36	116.81	134.71	
4e	89.34	74.18	68.77	84.61	59.65	147.37	186.45	106.70	137.37	
4f	89.01	74.10	68.86	84.60	59.66	147.73	189.28	89.01	142.05	

literature[20,21] could not unambiguously distinguish the ^{13}C signals between 1'-C and 4'-C and sometimes led to contradictory assignments. Thus, we used the ^{1}H-^{13}C HMQC technique to clarify this inconsistency in the literature. Figure 2 clearly shows that 1'-C has a higher δ value than 4'-C; hence, the order of the chemical shifts for the ribo-sugar carbons is **1'-C**> **4'-C**> 2'> 3'-C> 5'C. For the deoxyribonucleosides, the respective order is **4'-C**> **1**'-C> 3'-C> 5'-C> 2'-C (Fig. 3). Clearly, the OH at 2'-position (in ribo-sugar) has more influence on its neighboring 1'-C than on the more distant 4'-C.

The increased δ value of 1'-C in the ribonucleosides can be ascribed to the strong electronegativity of the oxygen atom of 2'-OH. It is interesting to note that the effects of the presence of 2'-oxygen atom are substantially different on the two neighboring carbons (1'-C and 3'-C) (Tables 4 and 5). The influence on 1'-C is noticeably high, whereas the influence on 3'-C is very small and at the scale similar to the distantly located 5'-C. This might indicate the orientation of 2'-OH points towards to 1'-C and away from 3'-C to avoid the potential electron repulsion from 3'-OH.

Modification at 4-position

The replacement of the oxygen at the 4-position with a sulfur atom gives rise to a thiocarbonyl group, which is evident from the shifts of the imino proton (3-NH) and thiocarbonyl carbon (4-C) respectively. Exchangeable ^{1}H signals in the range of δ 12.68–13.10 ppm, attributable to the 3-NH protons, support the structures of thiolated compounds **4(a-f)**. The appearance of only one carbon signal in the δ_C 185.30–190.70 ppm region (characteristic for a thiocarbonyl group) confirms the presence of the thiocarbonyl moieties in compounds **4(a-f)**. It is interesting to note that the ^{1}H chemical shifts of the imino proton (3-NH) in all the thionucleosides **4(a-f)** are substantially higher (at around 13 ppm) than those of the oxy-nucleosides **3(a-f)** that resonate at δ 11.2–11.8 ppm (Tables 1 and 2). This difference offers a valuable NMR window to detect the imino proton of thionucleosides because in general, there are no signals from un-modified nucleosides appearing at such a low field. In addition, these 3-NH proton signals are exchangeable and readily identifiable by D_2O exchange experiments. Therefore, these would also be a good marker in NMR studies of 4-thionucleosides and their corresponding DNA and RNA.

Modification at 5-position

The presence of a substituent at the 5-position affects chemical shifts of carbon atoms in the base, in particular the bonded carbon (i.e. 5-C) (Tables 4 and 5). As expected, a fluorine atom has the strongest effect on the 5-C signals. Figure 4 plots ^{13}C chemical shifts of halogenated 5-C atoms against the electronegativity of the 5-substituents. The greater the electronegativity of the attached halogen atom, the lower the electron density around the 5-C and the further downfield the chemical shift of the 5-C. In all cases, when the H atom at 5-position is replaced by a fluorine atom, the 5-C has the highest value, shifting from around 100 to 140 ppm.

This is due to the extremely high electronegativity of fluorine. Because F, Cl, Br and I atoms are each more electronegative than H atom, it would be anticipated that, when the H atom at 5-C is replaced by these substituents, their ^{13}C resonances should have higher δ values. This prediction holds true

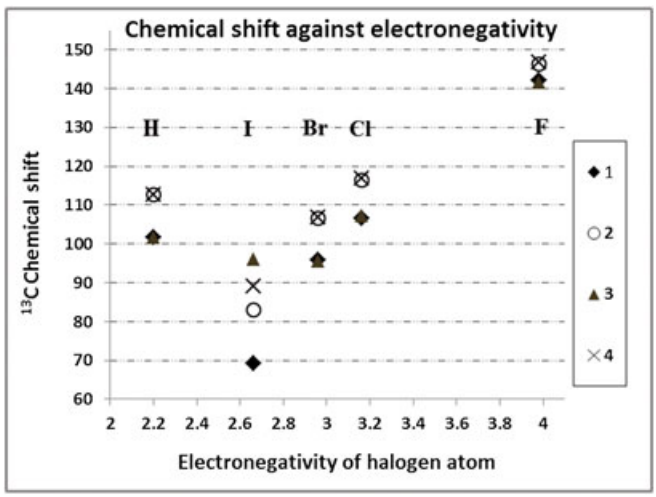


Figure 4. ^{13}C chemical shift of halogenated 5-C plotted against electronegativity of the halogen atom. The electronegativity is 3.98 (F), 3.16 (Cl), 2.96 (Br) or 2.66 (I), respectively, and each of them is higher than that of H atom (2.2). However, the ^{13}C chemical shifts of the 5-carbons bearing bromo (Br) or iodo (I) atom are lower than those of the carbons bearing H atom in all the nucleosides (**1–4**).

for F and Cl but not for Br and I. The ^{13}C resonances of 5-C in 5-bromonucleosides (**1e**, **2e**, **3e** and **4e**) and 5-iodo-nucleosides (**1f**, **2f**, **3f** and **4f**) in fact have lower δ values relative to that of the un-modified uracil nucleosides. Clearly, the electron-withdrawing effect alone is not enough to explain these observations. This unusual effect could be explained by the 'heavy atom effect'[22] that is when a carbon atom is attached to a heavy halogen atom (such as Br or I), the diamagnetic interactions arising from the numerous electrons of bromo or iodine atom increase the shielding effect of the substituted carbon atom so that the NMR resonances shift upfield.

Conclusions

A series of 5-substituted-4-thio-uridines and their 2'-deoxy analogues have been successfully prepared from their respective parent nucleosides. Their ^{1}H and ^{13}C NMR have been systemically investigated. Practical methods are established to unambiguously assign all pyrimidine carbons and to distinguish between 1'-C and 4'-C signals in both types of nucleosides. These assignments can provide useful references for 4-thiopyrimidine nucleosides and other base-modified nucleosides. The finding that the ^{1}H chemical shifts of the imino proton in these 4-thionucleosides are unusually high would offer an excellent marker in NMR studies of nucleic acids containing these thio-bases.

Experimental

NMR instruments

500 MHz from Bruker (AV-500, FT NMR) and 400 MHz from Bruker (AV-400, FT NMR). The COSY spectra (DMSO-d_6) were obtained in the magnitude mode with 1024 points in the F2 dimension and 256 increments in the F1 dimension. Each increment FID was obtained with 12 scans with a relaxation delay of 2 s. HMQC spectra (DMSO-d_6) were obtained in the magnitude mode with 1024 points in the F2 dimension and 256 increments in the F1 dimension. Each increment FID was obtained with 16 scans with a relaxation delay of 2 s.

Materials and methods

All chemicals and solvents, unless stated otherwise, were from either Aldrich or Sigma. All chemicals and solvents were used directly without further purification. Nucleosides on TLC were identified using p-anisaldehyde/ethanol/H_2SO_4 (5:90:5) or EtOAc: Petro Ether (8: 2) solution that converted the nucleosides into black spots on heating.

Preparation of 5-substituted- 4-thiopyrimidine nucleosides: *Chemical synthesis of 5-bromo-4-thiouridine is described below as a typical example.*

2', 3', 5'-tri-O-acetyl-5-bromouridine

To a solution of 5-bromouridine (**1e**) (0.87 g, 2.70 mmol) in dry pyridine (15 ml) at 0 °C, dry $(CH_3CO)_2O$ (3.0 ml, 32 mmol) was added slowly. The reaction was stirred for 5 h and monitored by the formation of fluorescent spots on TLC (5% MeOH in CH_2Cl_2). After the reaction was complete, the solvent was removed under reduced pressure. A white solid crystallized from ethanol (95%) was collected by filtration and dried on an oil vacuum pump to give the title compound (1.16 g, 95.9%), mp 52–53 °C.

2', 3', 5'-tri-O-acetyl-5-bromo-4-thiouridine

2', 3', 5'-tri-O-acetyl-5-bromouridine (0.90 g, 2.00 mmol) was dissolved in 1, 4-dioxane (30 ml), and P_2S_5 (0.85 g, 3.84 mmol) was added. The mixture was refluxed at 106 °C for 1.5 h (monitored by TLC). The solvent was removed under reduced pressure, and the residue was treated several times with CH_2Cl_2. The residue was purified on a silica gel column and eluted with petroleum: ethyl acetate (3:2, v/v) to give the title compound (0.51 g, 55.6%).

5-bromo-4-thiouridine (4e)

2', 3', 5'-tri-O-acetyl-5-bromo-4-thiouridine (0.63 g, 1.95 mmol) was suspended in absolute MeOH (120 ml) and was saturated with dry ammonia gas by stirring at room temperature for 4.5 h. The resulting solution was evaporated under reduced pressure at 35 °C, and the residue was purified on a silica gel column packed in CH_2Cl_2: MeOH = 9:1. Column was eluted with CH_2Cl_2: MeOH = 9.5:0.5; the solvent was then removed *in vacuo* to provide a solid residue. The residue was recrystallized from H_2O to give the title compound (0.33 g, 71%).

Acknowledgements

The authors are most grateful to Mr Yong-Ting Xu at Liaoning Normal University, China, for his excellent technical assistance with NMR. Dr Xu is most grateful to Dr Elaine Moore for critical reading of the draft. The work is financially supported by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, China.

References

[1] T. Lindahl. *Nature* **1993**, *362*, 709–715.
[2] E. Ichikawa, K. Kato. *J. Synth. Org. Chem. Jpn.* **2001**, *59*, 331–345.
[3] S. Porcher, S. Pitsch. *Helv. Chim. Acta* **2005**, *88*, 2683–2704.
[4] M. K. Lakshman. *Curr. Org. Synth.* **2005**, *2*, 83–112.
[5] P. A. Limbach, P. F. Crain, J. A. McCloskey. *Nucleic Acids Res.* **1994**, *22*, 2183–2196.
[6] M. N. Prichard, E. R. Kern. *Viruses-Basel* **2010**, *2*, 1968–1983.
[7] L. C. J. Gillet, O. D. Scharer. *Chem. Rev.* **2006**, *106*, 253–276.
[8] A. Massey, Y. Z. Xu, P. Karran. *Curr. Biol.* **2001**, *11*, 1142–1146.
[9] Y. Z. Xu, X. H. Zhang, H. C. Wu, A. Massey, P. Karran. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 995–997.
[10] R. Marek, V. Sklenar, in Annual Reports on NMR Spectroscopy, vol. 54, **2005**, pp. 201–242.
[11] P. F. Agris, S. C. Brown. *Methods Enzymol.* **1995**, *261*, 270–299.
[12] B. Ancian, in Annual Reports on Nmr Spectroscopy, vol. *69* (Ed: G. A. Webb), Elsevier Academic Press Inc., San Diego, **2010**, pp. 39–143.
[13] R. Narukulla, D. E. G. Shuker, V. Ramesh, Y. Z. Xu. *Magn. Reson. Chem.* **2008**, *46*, 1–8.
[14] X. H. Zhang, Y. Z. Xu. *Molecules* **2011**, *16*, 5655–5664.
[15] J. Asakura, M. J. Robins. *J. Org. Chem.* **1990**, *55*, 4928–4933.
[16] H. Schott, D. Goltz, T. C. Schott, C. Jauch, R. A. Schwendener. *Bioorg. Med. Chem.* **2011**, *19*, 3520–3526.
[17] V. Kumar, S. V. Malhotra. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5640–5642.
[18] J. J. Fox, D. Vanpraag, I. Wempen, I. L. Doerr, L. Cheong, J. E. Knoll, M. L. Eidinoff, A. Bendich, G. B. Brown. *J. Am. Chem. Soc.* **1959**, *81*, 178–187.
[19] C. B. Reese. *Tetrahedron* **1978**, *34*, 3143–3179.
[20] V. Kumar, J. Yap, A. Muroyama, S. V. Malhotra. *Synthesis-Stuttgart* **2009**, 3957–3962.
[21] F. Yang, Y. G. Zhu, B. A. Yu. *Chem. Commun.* **2012**, *48*, 7097–7099.
[22] S. Standara, K. Malinakova, R. Marek, J. Marek, M. Hocek, J. Vaara, M. Straka. *Phys. Chem. Chem. Phys.* **2010**, *12*, 5126–5139.
[23] G. W. M. Visser, R. E. Herder, P. Noordhuis, O. Zwaagstra, J. D. M. Herscheid, F. J. J. Dekanter. *J. Chem. Soc., Perkin Trans. 1* **1988**, 2547–2554.
[24] Q. Wu, A. H. Xia, X. F. Lin. *J. Mol. Catal. B: Enzym.* **2008**, *54*, 76–82.