Differential reactivity of carbohydrate hydroxyls in glycosylations. II.[1] The likely role of intramolecular hydrogen bonding on glycosylation reactions. Galactosylation of nucleoside 5′-hydroxyls for the syntheses of novel potential anticancer agents

DENNIS M. WHITFIELD AND STEPHEN P. DOUGLAS

Department of Molecular and Medical Genetics, and Carbohydrate Research Centre, Protein Engineering Network of Centres and Excellence (PENCE), University of Toronto, Toronto, ON M5S 1A8, Canada

TING-HUA TANG

Protein Engineering Network of Centres of Excellence (PENCE) and Department of Chemistry, University of Toronto, Toronto, ON M5S 1A8, Canada

IMRE G. CSIZMADIA

Department of Chemistry, University of Toronto, Toronto, ON M5S 1A8, Canada

HENRIANNA Y.S. PANG

Department of Molecular and Medical Genetics, and Carbohydrate Research Centre, University of Toronto, Toronto, ON M5S 1A8, Canada

FREDERICK L. MOOLTEN

Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA 01730, U.S.A. and Department of Microbiology, Boston University, Boston, MA 02215, U.S.A.

AND

JIRI J. KREPINSKY[2]

Department of Molecular and Medical Genetics, and Carbohydrate Research Centre, Protein Engineering Network of Centres of Excellence (PENCE), University of Toronto, Toronto, ON M5S 1A8, Canada

Received June 25, 1993[3]

DENNIS M. WHITFIELD, STEPHEN P. DOUGLAS, TING-HUA TANG, IMRE G. CSIZMADIA, HENRIANNA Y.S. PANG, FREDERICK L. MOOLTEN, and JIRI J. KREPINSKY. Can. J. Chem. **72**, 2225 (1994).

Contrary to expectations, many primary hydroxy groups are completely unreactive in glycosylation reactions, or give the desired glycosides in very low yields accompanied by products of many side reactions. Hydrogens of such primary hydroxyls are shown to be intramolecularly hydrogen bonded. Intermediates formed by nucleophilic attack by these hydroxyls on activated glycosylating agents may resist hydrogen abstraction. This resistance to proton loss is postulated to be the origin of the observed unreactivity. It is shown that successful glycosylations take place under acidic conditions under which such hydrogen bonds cease to exist. Accordingly, direct galactosylations of the normally unreactive 5′-hydroxyls of nucleosides were accomplished for the first time with a galactose trichloroacetimidate donor in chloroform under silver triflate promotion. It is noted that such galactosylated anticancer nucleosides may have improved biological specificity.

DENNIS M. WHITFIELD, STEPHEN P. DOUGLAS, TING-HUA TANG, IMRE G. CSIZMADIA, HENRIANNA Y.S. PANG, FREDERICK L. MOOLTEN et JIRI J. KREPINSKY. Can. J. Chem. **72**, 2225 (1994).

Contrairement à ce que l'on peut prévoir, plusieurs groupes hydroxyles primaires ne réagissent pas du tout au cours des réactions de glycosylation ou avec seulement de faibles rendements accompagnés de plusieurs sous-produits. On a démontré que les hydrogènes de ces hydroxyles primaires sont impliqués dans des liaisons hydrogènes. Les intermédiaires qui se forment par attaque nucléophile de ces hydroxyles sur des agents activés de glycosylation peuvent résister à la réaction d'enlèvement de l'hydrogène. Cette résistance à la perte d'un proton se produit en milieu acide alors que de telles liaisons hydrogènes cessent d'exister. On a réalisé pour la première fois des galactosylations directes des hydroxyles normalement non réactifs en 5′ des nucléosides en utilisant comme donneur un trichloroacétimidate de galactose dans le chloroforme sous promotion par le triflate d'argent. Il a été suggéré que de tels nucléosides anticancéreux galactosylés pourraient avoir une spécificité biologique améliorée.

[Traduit par la rédaction]

Introduction

Glycosylation of carbohydrate primary hydroxyls usually proceeds smoothly and many examples of glycosylations of OH-6 of hexopyranosides,[4] using Koenigs–Knorr, Helferich, and related procedures, have been described (3). However, we have observed surprisingly unreactive primary hydroxyl groups with certain acceptors. In our experience with difficult glycosylations, these acceptors often contained benzyl or allyl protecting groups (4). Similar observations have been made in other laboratories as well (e.g., P. Sinaÿ, V.V. Bencomo, personal communications). The common-sense explanation is that steric hindrance renders the primary hydroxyl groups inaccessible to the glycosylating agent. We were intrigued by reports of surprising difficulties (5) in the glycosylation of the primary O-5′ H of nucleosides (6, 7). Not only are orthoesters (8) formed when using an ester protective group at C-2 under common Koenigs–Knorr reaction conditions, but even this

[1]For article I in this series, see ref. 1.

[2]Author to whom correspondence may be addressed.

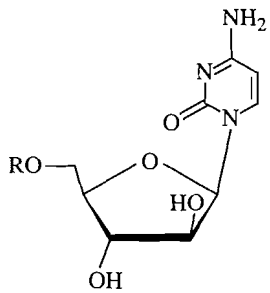
[3]Revision received July 18, 1994.

[4]There has been little interest in the syntheses of oligosaccharides through the OH-5 of pentofuranoses; this is perhaps because the 1–5 linkage is very rare among naturally occurring oligosaccharides.

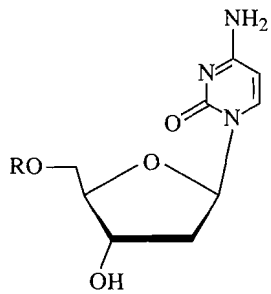
reaction is sluggish and low yielding (9). To invoke steric effects of bulky protective groups resulting in inaccessibility of reactive sites did not seem quite correct in the case of nucleosides since no large protective groups were present in the reported cases. Consequently, we wondered whether intramolecular interactions such as hydrogen bonding between the hydroxyl and a lone pair or a π-electron system may be responsible for this unreactivity.

The success, although very limited, of the Bredereck modification of the Koenigs–Knorr reaction involving 2-deoxynucleosides such as thymidine (**T1**) indicated that these effects may indeed be of importance. In the Bredereck reaction, the primary hydroxyl is protected with a trityl group and glycosylations are performed under acidic conditions (10). Obviously, since the hydrogen atom of the 5′-hydroxy group is not present, it cannot participate in hydrogen bonding. This observation supports our hypothesis that glycosylations should be successful in the absence of the hydrogen bonding. In addition, glycosylations should occur if the hydrogen bond breaks under the reaction conditions.

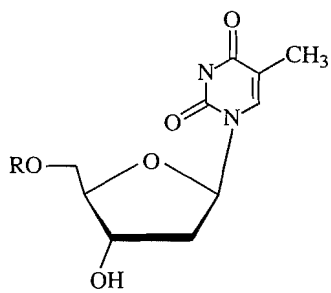
To control effectively the outcome of glycosylations, the factors limiting reactivity should be determined. Because of our long-standing interest in a polymer-supported design (11) for synthesis of oligosaccharides in which such control is essential, we have undertaken a program investigating the origins of the low reactivity of such primary hydroxyls. We are also exploring a novel concept of cancer therapy based on the assumption that, under certain conditions, 5′-galactosylation yields nucleoside



Ar1
Ar2



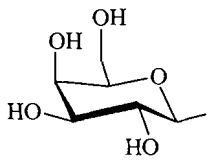
C1
C2



T1
T2

1, R = H

2, R =



Ar, derivatives of arabinocytosine; **C,** derivatives of cytidine; **T,** derivatives of thymidine

anticancer drugs that can be specifically targeted against neoplastic cells (see Appendix I). We have already demonstrated (1) through ab initio mechanical calculations the existence of non-covalent interactions in nucleosides and of the potential for hydrogen bonding between the O5′-H and the nucleoside base in low-polarity solvents (12). To the best of our knowledge, direct glycosylation of nucleoside 5′-hydroxyl groups has not yet been reported. Consequently, we had to develop methodologies, capable of overcoming this hydrogen bonding, for the preparation of 5′-galactosylated nucleosides araC (**Ar2**), 2-deoxycytidine (**C2**), and thymidine (**T2**).

Results

Conformation of acceptor nucleosides

To determine the origins of the unusual unreactivity of primary hydroxyls of aglycons such as **Ar3**, **C3**, and **T3** (intermediates for the synthesis of **Ar2**, **C2**, and **T2**, see Scheme 1), and considering the hypothesis that such hydroxyls may be involved in intramolecular hydrogen bonding, their conformations in dry CD_2Cl_2 (to mimic the glycosylation conditions) were determined by NMR spectrometry. Infrared spectrometry was also used to probe the presence of internal hydrogen bonding in these substances.

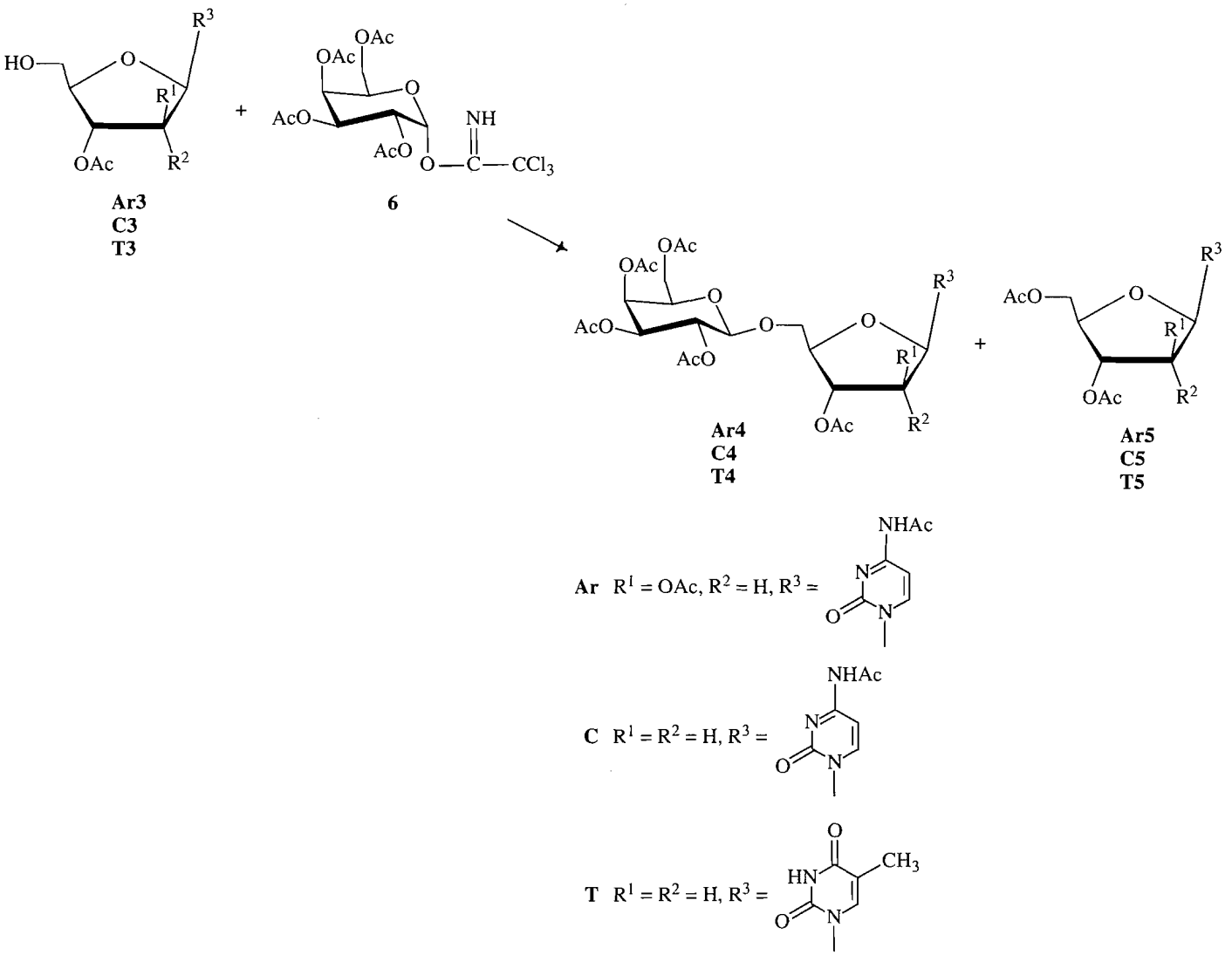
*Deoxycytidine derivative **C3***

A. Nuclear magnetic resonance spectrometry

The data for deoxycytidine derivative **C3** in CD_2Cl_2 solution (cf. also ref. 1) have shown that **C3** exists in one overall predominant conformation. It is the *syn* conformation (cf. Fig. 1 for definition (13)) about the glycosidic linkage, as indicated by a large NOE between the anomeric proton H-1′ and the base H-6 and the *gauche-gauche* ($\gamma = 60°$) conformation for the hydroxymethylene group (based on the small values of the coupling constants $J_{4',5'}$ and $J_{4',5''}$, 2.7 and 2.9 Hz, see Table 3). Such a general conformation allows for a hydrogen bond between O-5′ H and the base carbonyl O-2. In the more acidic deuterated chloroform, the NOE between the H-1′ and H-6 diminished (8.1% → 5.1%), consequently the distance may have increased between O-5′ H and the base carbonyl O-2.

B. Infrared spectrometry

Stretching vibrations of a free OH group are normally absorbing in the frequency range 3610–3645 cm^{-1}, and the intra- and intermolecular hydrogen bonds give rise to an absorption in the frequency range 3200–3600 cm^{-1}. The two hydrogen bondings can be distinguished since the intermolecular hydro-



SCHEME 1. Galactosylation of nucleosidic O-5′H.

gen bonding diminishes with increased dilution of the sample, and consequently the absorption band disappears, while in the case of the intramolecular hydrogen bonding the absorption band remains unchanged. Intermolecular hydrogen bonding is not expected to occur at concentrations below 0.5 molar (14). If an NH group is present in the molecule, N–H stretching vibrations interfere since they give rise to a broad absorption between 3000 and 3500 cm^{-1}, thus overlapping with the absorption due to the OH vibrations.

To mimic the glycosylation conditions, the IR measurements have to be done in CH_2Cl_2, in which **C3** is not very soluble. At the highest concentration possible (1.3 mM), OH stretching vibrations in **C3** are found as a shoulder at 3620 cm^{-1} (free OH) of 3404 cm^{-1} (superimposed OH intramolecular bonding and NH). These absorptions did not change on further dilution until the sample became too dilute to record the spectrum.

Arabinocytidine derivative ***Ar3***

A. Nuclear magnetic resonance spectrometry

The arabinosylcytosine derivative **Ar3** exists in an *anti-syn* equilibrium with the *anti* conformation predominating, since the presence of the O-2′ acetate disfavours the *syn* conformation (15); cf. Fig. 1*a* for the definition of *syn* and *anti*, and the variations of the angle χ within the definition. The small NOE (cf. Fig. 2) between H-1′ and H-6 indicates that for **Ar3** the *anti* conformation, $\chi = 180°$,[5] most likely predominates over the *syn* conformation (16). The conformation around the hydroxymethylene moiety is indicated by the small (3.7 Hz) and medium (5.1 Hz) H-4′–H-5′ and H-4′–H-5″ couplings (see Table 1), nearly identical before and after D_2O exchange (cf. Fig. 3). Such couplings are consistent with an equilibrium between the *gauche-gauche* ($\gamma = 60°$) and the *gauche-trans* ($\gamma = 180°$) rotamers about C-4′–C-5′ (17). Further evidence of such an equilibrium is obtained from NOE measurements after irradiation of H-5′ or H-5″. In both cases a moderate intensity NOE was observed to H-3′ as well as to H-4′. In the *gauche-gauche* conformation only H-5″ (proR) is close enough to H-3′ to expect an observable NOE, while in the *gauche-trans* conformation H-5″ comes close to H-4′.

Another hydrogen bond is possible between O-5′ H and O-2′ in the *gauche-gauche* conformation. Such a hydrogen bond would be best accommodated in the C-2′ *endo* conformation of

[5]The *anti* conformation has been recognized as the predominant conformation for most nucleosides (ref. 43); the parent compound Ara C, **Ar1**, adopts the *anti* conformation in the crystalline state (44). Cf. also ref. 18.

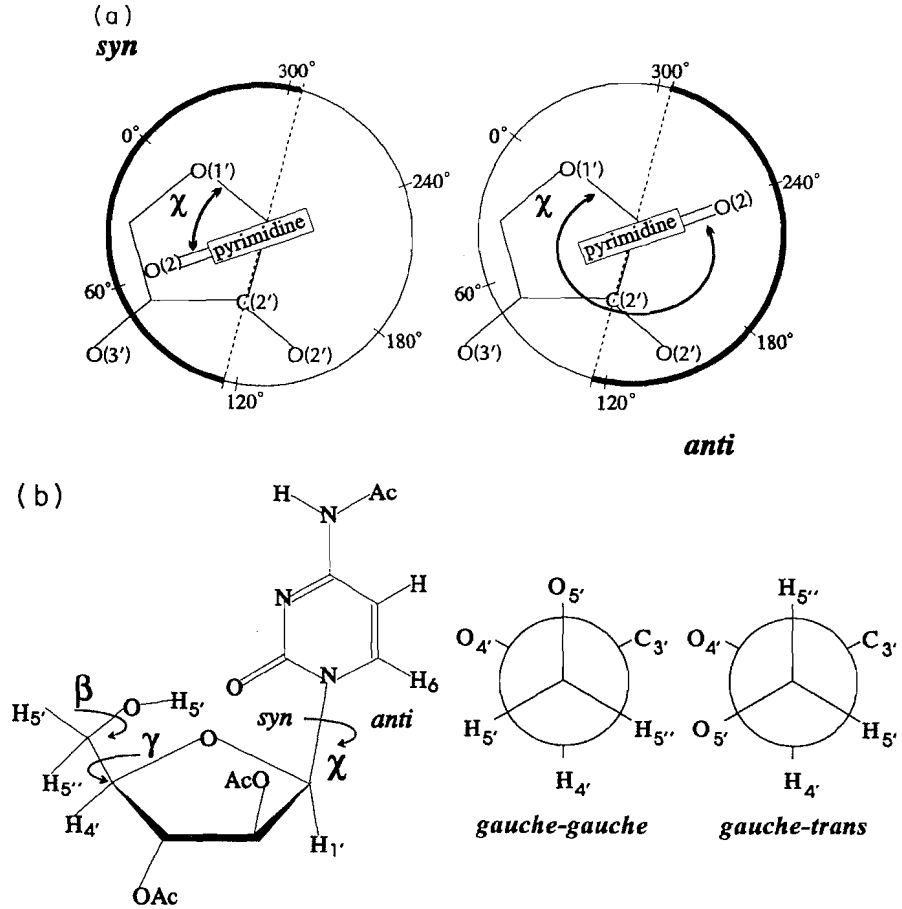


FIG. 1. Definitions of conformational parameters of nucleosides. Part (*a*) is reproduced from Saenger (13), with permission.

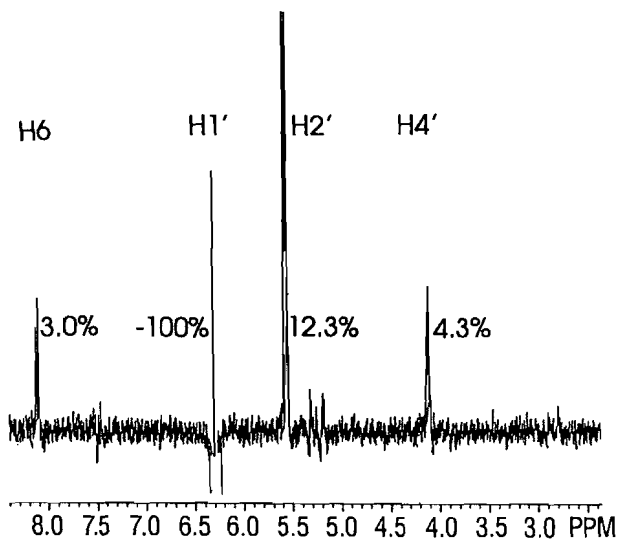


FIG. 2. Partial 500 MHz ^{1}H NMR spectrum of **Ar3** in CD_2Cl_2 solution. Difference 500 MHz ^{1}H NMR nuclear Overhauser effect spectrum of **Ar3** in CD_2Cl_2 solution after irradiation of anomeric H-1′ showing a small NOE to H-6.

the furanoside ring. This ring conformation is deduced from the small values of $J_{3',4'}$ (see Table 1). However, the small values of $J_{1',2'}$ and $J_{2',3'}$ are consistent with a mixture of conformers C-2′ *endo* and C-3′ *endo* (18).

B. Infrared spectrometry

The infrared spectrum of **Ar3** at a concentration of 7.2 mmol (the maximal concentration of freely soluble **Ar3**) displays two broad absorption bands for OH and NH stretching vibrations at 3482 cm^{-1} that did not change on dilution. No free OH absorption could be identified in the spectrum (cf. also Wechter, ref. 19). The absence of the absorption for free OH from the spectra of **Ar3** may seem somewhat puzzling since it appears to suggest that the hydrogen-bonded (O-5′ H...O-2) conformation predominates, while the NMR spectrometry would at first glance indicate that only a certain percentage of **ArC** in the conformational equilibrium is in a hydrogen-bonded *syn* conformation. However, hydrogen bonding from the O-5′ H can be directed to other acceptors than O-2 (18).

*Thymidine derivative **T3***

The thymidine derivative **T3** was not investigated in detail since key resonances overlapped. However, the small $J_{4',5'}$ and $J_{4',5''}$ coupling constants suggest that this derivative resembles the intramolecularly hydrogen-bonded conformation of **C3**.

Thus it is very likely that the deoxycytidine and thymidine derivatives having a free 5′-hydroxyl exist predominantly in an intramolecularly hydrogen-bonded conformation, O-5′ H...O-2. This hydrogen-bonded conformation may also be a significant component of the conformational equilibrium of the arabinosylcytosine derivatives, in addition to other hydrogen bondings.

TABLE 1. 500-MHz ^{1}H NMR data for arabinosylcytidine derivatives **Ar3**, **Ar5**, **Ar4**, and **Ar2**. Chemical shifts are in ppm and coupling constants are in Hz

Resonance	**Ar3**; CD_2Cl_2[a]	**Ar5**; $CDCl_3$	**Ar4**; $CDCl_3$	**Ar2**; D_2O
H5 ($J_{5,6}$)	7.49 (7.5)	7.49 (7.5)	7.51 (7.6)	6.06 (7.5)
H6	8.12	7.92	8.09	7.84
H1′ ($J_{1',2'}$)	6.30 (4.2)	6.36 (3.6)	6.40 (4.9)	6.22 (4.9)
H2′ ($J_{2',3'}$)	5.55 (2.5)	5.55 (<0.5)	5.58 (3.0)	4.42 (4.3)
H3′ ($J_{3',4'}$)	5.20 (3.3)	5.08 (2.6)	5.21 (3.0)	4.17 (4.4)
H4′ ($J_{4',5'}$)	4.12 (3.7)	4.25 (6.6)	4.11 (2.7)	4.16 (1.2)
H5′ ($J_{4',5''}$)	3.94 (5.1)	4.44 (4.4)	3.79 (4.1)	4.25 (3.4)
H5″ ($J_{5',5''}$)	3.87 (−12.2)	4.42 (−11.9)	4.24 (−10.9)	3.94 (−11.3)
NH	9.56	9.60	9.39	—
OH5′ ($J_{OH,5'}$) ($J_{OH,5''}$)	2.82 (4.3) (5.1)	—	—	—
GH1 ($J_{1,2}$)	—	—	4.58 (7.9)	4.49 (7.7)
GH2 ($J_{2,3}$)	—	—	5.22 (10.5)	3.57 (9.7)
GH3 ($J_{3,4}$)	—	—	5.06 (3.4)	3.67 (3.5)
GH4 ($J_{4,5}$)	—	—	5.43 (<0.5)	3.94 (<0.5)
GH5 ($J_{5,6}$)	—	—	3.95 (6.5)	3.69 (4.1)
GH6 ($J_{5,6'}$)	—	—	4.22 (6.9)	3.81 (7.9)
GH6′ ($J_{6,6'}$)	—	—	4.15 (−11.2)	3.77 (−11.7)
CH_3CO	2.24, 2.13, 1.91	2.28, 2.15, 2.14, 1.98	2.27, 2.20, 2.12, 2.06, 2.06, 2.00, 1.98	—

[a]Comparable to partial NMR data in ref. 16.

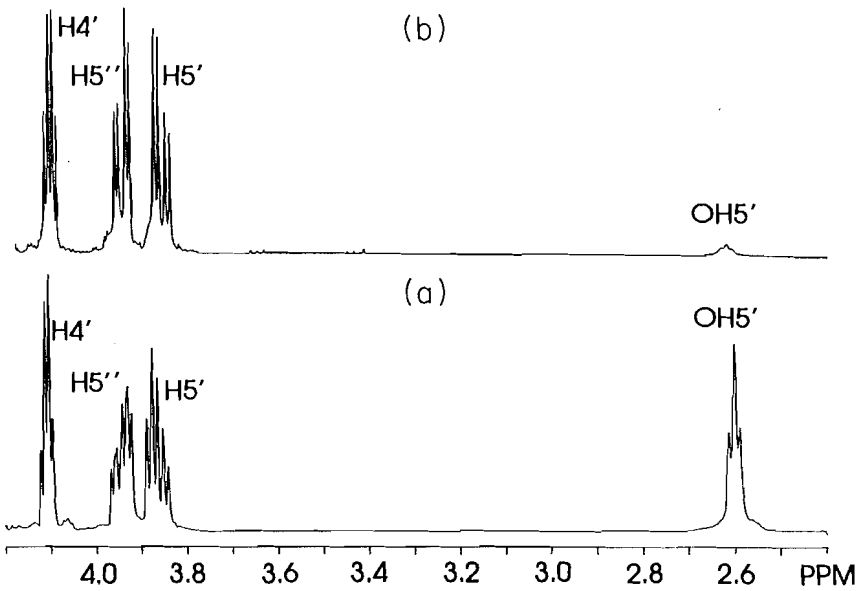


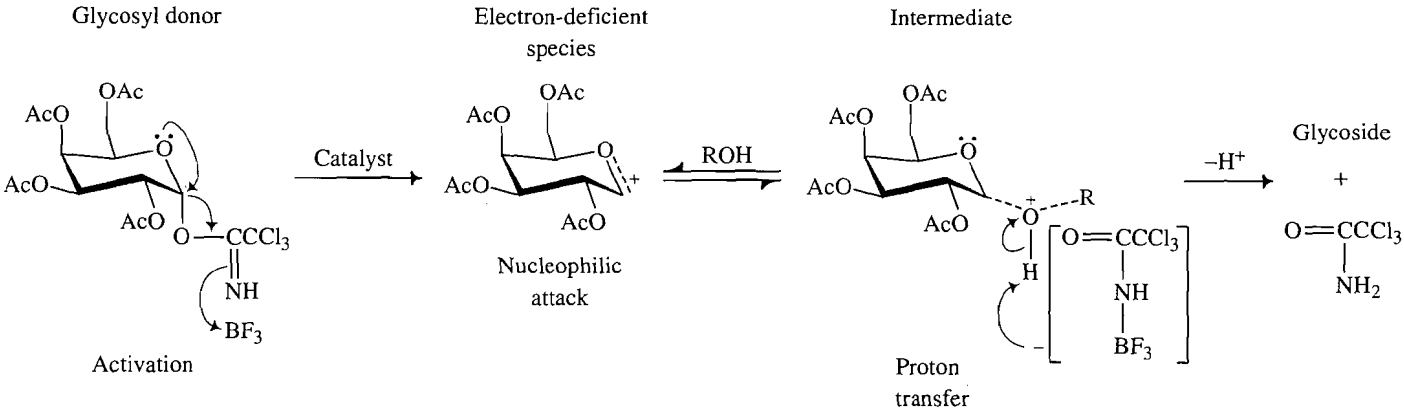
FIG. 3. Partial 500 MHz ^{1}H NMR spectrum of **Ar3** in CD_2Cl_2 solution. Figure 3(*a*) (lower tracing): a portion of the spectrum exhibiting signals of the H-4′, H-5′, H-5″, and OH-5′ protons; Figure 3(*b*) (upper tracing): the same area of the spectrum after exchange with D_2O showing that the signal for OH-5′ disappeared and one coupling was removed from both H-5′ and H-5″ resonances.

Synthesis of 5′-galactosylated nucleosides

The known acetylated nucleosides **Ar3** (19), **C3** (20), and **T3**,[6] having a free 5′-hydroxyl, were prepared as follows: N-acetylation of the free amino group (when present), temporary regioselective protection of the 5′-hydroxyl with the dimethoxytrityl (DMT) group, acetylation of the secondary hydroxyls, and removal of the DMT group. Attempts to selectively react the 5′-hydroxyl with DMT chloride without prior N-acetylation led to considerable substitution with DMT at the amino group. Even mild conditions for the removal of the DMT group with 1% dichloroacetic acid in CH_2Cl_2 led to some unavoidable nucleoside glycosidic bond cleavage necessitating a chromatographic purification.

These derivatives with a free 5′-hydroxyl did not give the desired galactosylated nucleosides with acetobromogalactose under standard Koenigs–Knorr conditions. However, they could all be galactosylated with 2,3,4,6 tetra-*O*-acetyl α-D-galactopyranosyl trichloracetimidate **6** (21) in CH_2Cl_2 promoted by

[6]Also commercially available (e.g., from Sigma, St. Louis, Mo.).



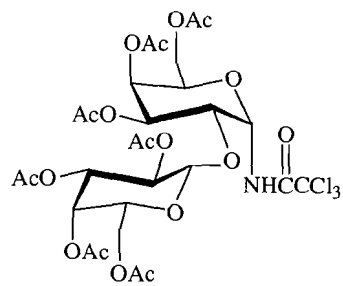
SCHEME 2. Simplified mechanism of the glycosylation reaction showing three principal stages of the reaction: activation, nucleophilic attack, and hydrogen loss. A protonated intermediate is formed in the second step of this reaction sequence. This intermediate must lose the proton to give the desired glycoside. In most cases the deprotonation is facile; it may become difficult if the hydrogen present in the intermediate is stabilized by an intramolecular hydrogen bond.

$BF_3{\cdot}Et_2O$ (22) to give the desired galactosides **Ar4**, **C4**, and **T4**. In all cases galactosylation was accompanied by significant acetate transfer from O-2 of the galactosyl imidate to the 5′-hydroxyl, giving the known nucleoside peracetates **Ar5** (23), **C5** (24), and **T5** (25) in ratios 1.25:1 (**Ar4**:**Ar5**), 0.23:1 (**C4**:**C5**), and 0.36:1 (**T4**:**T5**). No other nucleoside-containing products were observed. Imidate **6** and $BF_3{\cdot}Et_2O$ were used in equimolar quantities; when $BF_3{\cdot}Et_2O$ was used in catalytic amounts, acetate transfer predominated, and the overall yields were low. Variations in the temperature or the rate of addition of the $BF_3{\cdot}Et_2O$ had little effect and $BF_3{\cdot}Et_2O$ had to be added to a mixture of the imidate and the acceptor (26). Some nucleoside (base–sugar) cleavage occurred when reaction times exceeded 1 h.

Acetate transfer is not uncommon and has often been reported to compete with glycosylation reactions (27). To suppress this transfer, other Lewis acids were examined. $ZnBr_2$, although reported to eliminate acetate transfer (28), gave only complex mixtures. Triethylsilyl triflate (TESOTf) (29) gave results similar to those obtained with $BF_3{\cdot}Et_2O$. Silver triflate, recently described as a promoter of glycosylation reactions with trichloroacetimidates (30) in CH_2Cl_2, *significantly suppressed* acetate transfer, raising the ratio of glycosylation to transfer from 0.23:1 to 1.67:1 (**C4**:**C5**).

Further suppression of the acetate transfer reaction was achieved by changing the solvent from CH_2Cl_2 to the more acidic $CHCl_3$; using $BF_3{\cdot}Et_2O$ improved the ratio from 0.23:1 to 0.5:1 (**C4**:**C5**) in favour of the glycoside, and using silver triflate improved the ratio even further to 3.33:1 (**C4**:**C5**). The ratio for **Ar4**:**Ar5** improved to 20:1 from 1.25:1 using AgOTf in $CHCl_3$.

The peracetylated galactosylated nucleosides and the products of the acetate transfer reaction (peracetylated nucleosides) were separated by chromatography on silica gel followed by size-exclusion chromatography on Sephadex LH-20. During the silica gel chromatography a major by-product, the peracetylated disaccharide Galβ1,2GalαNHCOCCl$_3$ (**7***a*), was isolated. Its structure was determined by NMR and mass spectrometries; in particular, a strong ^{1}H NMR NOE between the β-Gal H-1 and the H-2 of the α-Gal residue established the β1–2 interglycosidic linkage. The α-NHCOCCl$_3$ anomer **7***a* was accompanied by a variable small amount of the β-anomer **7***b*. Both **7***a* and its anomer arise from the galactosylation of the monosaccharide formed after transfer of the acetate from O-2 of **6**, and imidate rearrangement of **6** (31).



7*a*

The desired glycosylated nucleosides were finally obtained by removing both O- and N-acetate protective groups with DBU-catalysed transesterefication.

Preliminary biological assays

The galactosylated compounds exhibited the increased specificity in toxicity for cell types expressing the lacZ gene while they were nontoxic to other cells. This is the basic requirement for one approach to overcoming the low-selectivity problem associated with cancer chemotherapeutics (see Appendix II).

Discussion

The understanding of glycosylation reactions has been greatly advanced during the last two decades through the efforts of many chemists, and particularly through the pioneering insights into anomeric effects by R.U. Lemieux and his colleagues (32). It is clear that currently most common methods for the formation of glycosidic linkages proceed via nucleophilic substitution at the electron-deficient anomeric carbon as shown in Scheme 2. Details of the mechanism are less clear and apparently vary among different experimental designs. Despite these possible variations, it would be intuitively anticipated that primary hydroxyl groups (e.g., O-5H in pentofuranoses or O-6H in hexopyranoses) should always be reasonably reactive nucleophilic partners in glycosylation reactions. It is not, however, the case for nucleosides.

The results reported in this article suggest that 2′-deoxycytidine derivative **C3** assumes a conformation indicative of a hydrogen bond from the 5′-hydroxyl to the 2-carbonyl group and ab initio calculations and topological bond-path determinations confirm this conclusion (1). The NMR and infrared data demonstrate that analogous derivatives of thymidine, **T3**, and

arabinofuranosyl cytosine, **Ars**, also exist in such hydrogen-bonded conformations, at least to some extent. The O-5′ H of **Ar3** may be engaged in other hydrogen bonding (18).

The hydroxyl groups acting as hydrogen bond donors should have increased electron density at the oxygen (33), and this increase in the electron density should in turn make this oxygen a better (more reactive) nucleophile. Thus such a hydroxyl should be more reactive in a glycosylation reaction than the same hydroxyl would be, if not engaged in hydrogen bonding. Eventually the glycosylating agent becomes bound in a protonated intermediate that is still intramolecularly hydrogen bonded as portrayed in Scheme 2. To obtain the desired glycoside from the reaction, *this hydrogen must be abstracted.* Until this happens, the protonated intermediate remains in equilibrium with other species and therefore can become involved in competing reactions. For instance, the newly formed glycosidic bond can break, thus recreating the electron-deficient species, which may be diverted into other reaction pathways as illustrated in Scheme 3. These pathways, which involve the loss of the 2-O-acetyl and result in acetyl transfer to the aglyconic hydroxyl a well as in 1,2 disaccharides, can effectively compete (cf. the formation of the disaccharides **7***a*, *b*).

It is also important to appreciate that for donors like trichloroacetimidates, once the electron-deficient species is formed, it cannot form the original donor again because the imidate is transformed into an amide after activation. This amide or its complex with a Lewis acid can also act as a nucleophile. Lewis acids such as $BF_3{\cdot}Et_2O$ or silver triflate can also bind to nucleosides; this explains the requirements for larger than equimolar amounts of such acids in successful glycosylations. By such binding the Lewis acids may break up the intramolecular hydrogen bonding: it was shown in the previous communication (1) that complexation at the N-3 of the nucleoside by BF_3 markedly destabilized the hydrogen-bonded conformer. The intramolecular hydrogen bond can be further destabilized by solvent effects; for instance, successful 5′-galactosylations in chloroform (see also footnote 7) are described in this article.

Concluding remarks

We have attempted in this article to explain the origins of the low reactivity of primary hydroxyls acting as acceptors of glycosylating agents. We formulated a hypothesis that such unreactive primary hydroxyls are involved in intramolecular hydrogen bonding. Led by this hypothesis, we devised conditions in which this unreactivity can, at least to some extent, be overcome. Although we developed the argument using nucleosides as examples, it is equally applicable to glycosylations of any substrate, including oligosaccharides, in which the hydroxyl is involved in an intramolecular hydrogen bond. Clearly, more work, using suitably designed experimental strategy, is necessary to provide additional evidence for this hypothesis.

It is also worth noting that the aromatic ring of protecting groups, so commonly used in oligosaccharide synthesis (e.g., benzyls), can be the acceptor in hydrogen bonds as is well known from conformational studies of proteins. π-Type hydrogen-bonded complexes between benzene with acceptors HA or H_nA (including H_2O) have been described both experimentally (34) and theoretically (35). One of the possible ways (both a single or a bifurcated hydrogen bond is possible) to form this type of bonding in $C_6H_6{\cdot}H_2O$ is shown in Fig. 4. Assuming that such a hydrogen bond is sufficiently strong (typically 2–4 kcal mol^{-1}), the reaction mechanism portrayed in Scheme 2 applies to glycosylations of such hydroxyls as well. Examples from this laboratory of such glycosylations will be published in due course.

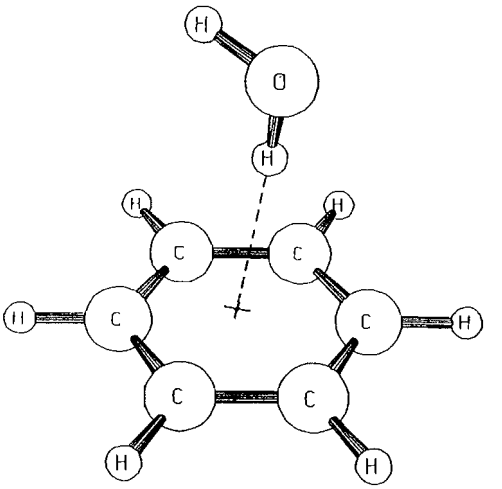


FIG. 4. One of the possible ways to form π-type hydrogen-bonded $C_6H_6{\cdot}H_2O$ complexes. One hydrogen atom of the water molecule is above the centre of the benzene ring and the O—H bond is almost perpendicular to the plane of the benzene ring. One minimum is found using the semiempirical Parametric Method 3 (PM3) method (41), the hydrogen bonding energy ΔE (PM3) = −1.39 kcal mol^{-1}. It has been shown that a similar intramolecular π-type hydrogen bonding in which an aromatic ring is the acceptor is not uncommon (42) and that such bonds are one of the stabilization forces of protein conformations.

Experimental

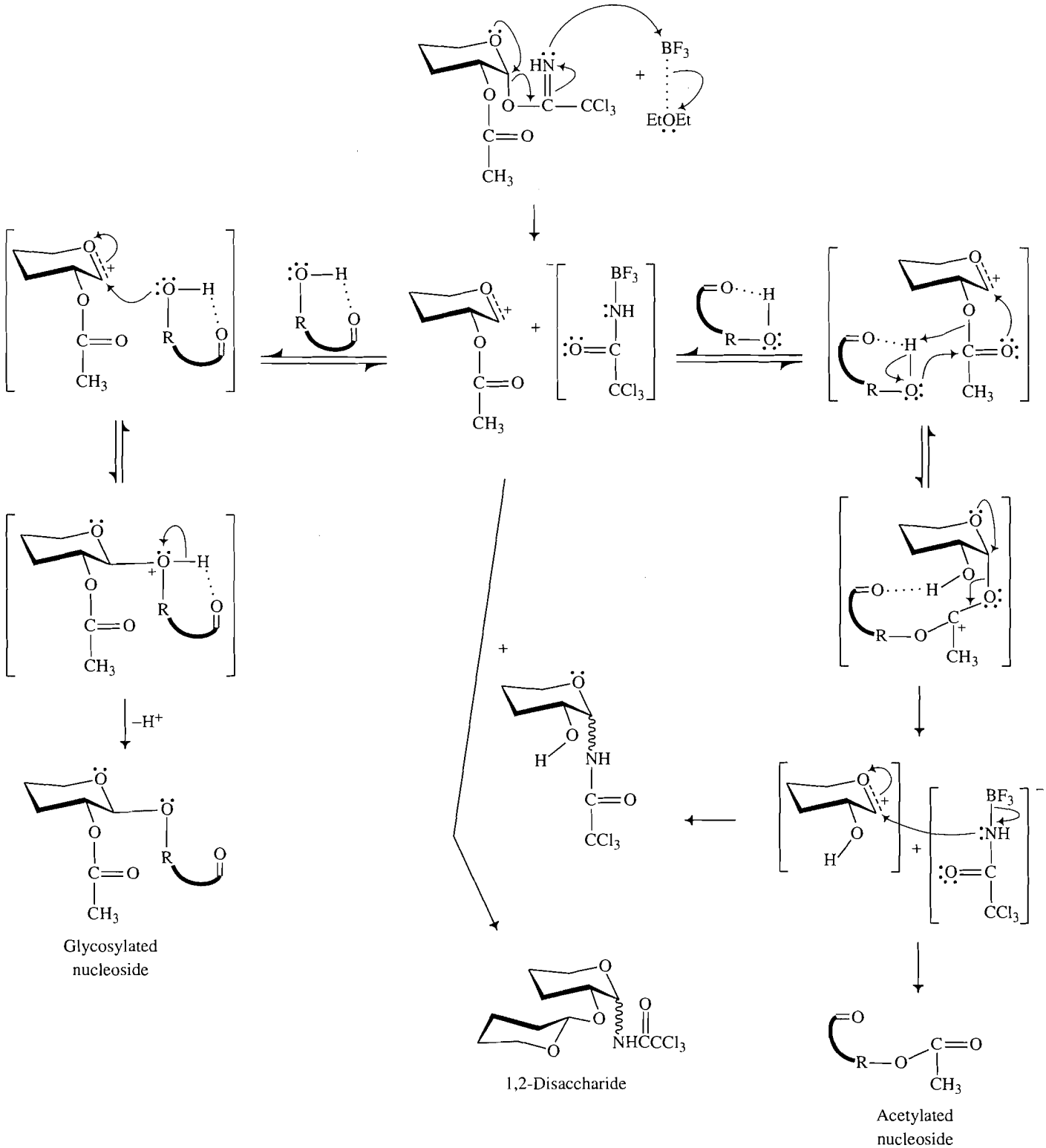
General methods

Thin-layer chromatography (TLC) was performed on silica gel $60F_{254}$ (Merck) plates and visualized by spraying with 50% aqueous sulfuric acid and heating at 200°C. Silica gel (230–400 mesh, Toronto Research Chemicals) was used for flash chromatography. All starting materials were dried overnight in vacuo (10^{-3} Torr; 1 Torr = 133.3 Pa) over KOH or P_2O_5 prior to use, and the solvents were distilled from appropriate drying agents. Solutions were concentrated at 1 Torr pressure in a rotary evaporator.

All compounds were characterized by ^1H (Tables 1, 3, and 5) and ^{13}C NMR (Tables 2, 4, and 6). To the best of our knowledge, the complete NMR spectra of the known compounds **AR3**, **AR5**, **C3**, **C5**, **T3**, and **T5** have not yet been reported. ^1H NMR assignments were made by standard decoupling techniques and all coupling constants were determined by first-order analysis of peak separations where possible. In the case of the fully deprotected compounds in D_2O solutions, the methylene protons were analyzed using inversion–recovery experiments with short delays (0.2−0.4 s) after the 180° pulse. With optimum delays, the short T_1's of the methylene hydrogens allowed the signals for these protons to be observed, whereas all other signals were either nulled or inverted (longer T_1's). ^{13}C NMR spectra of the title compounds were assigned by ^{13}C–^1H correlation, and of protected compounds by comparison to the literature (36, 37). All spectra are consistent with the proposed structures.

H and ^{13}C NMR spectra were recorded at 500 (125, ^{13}C) MHz with a Bruker AM 500 spectrometer or at 300 (75.5, ^{13}C) MHz with a Bruker AM 300 spectrometer, both at the NMR Spectrometry Laboratory of the Carbohydrate Research Centre, University of Toronto. Spectra were obtained either in $CDCl_3$ containing a trace of TMS (0 ppm, ^1H and ^{13}C) as the internal standard or in D_2O (99.98%, Ald-

[7]Another example is silver triflate promoted galactosylation (by trichloroacetimidate, **6**) of a 5-fluorouridine derivative in chloroform at elevated temperature (45).



SCHEME 3. The glycosylation reaction is diverted from yielding the desired glycoside to a formation of by-products. The + and − signs represent electron deficiency or excess rather than true charges. The pyran ring represents properly protected pyranosides, BF_3 is used as an example of a promoter or catalyst.

rich) containing a trace of acetone (2.225 ppm relative to internal DSS, ^1H, and 30.5 ppm, ^{13}C) as the internal standard using the δ(ppm) scale. Spectra in CD_2Cl_2 were referenced to the residual $CHDCl_2$ (5.32 ppm) for ^1H spectra or to the ^{13}C resonance (53.6 ppm). Samples in CD_2Cl_2 had activated powdered 4 Å molecular sieves added directly to the NMR tube to ensure the complete removal of any residual water. For D_2O exchange, 1 drop of D_2O was added to the NMR tube, mixed vigorously, and then left to equilibrate with added molecular sieves before the spectra were obtained. NOE spectra were obtained in the 1D difference mode using an internal reference; the solutions were not de-

TABLE 2. 75.5 MHz ^{13}C NMR data for arabinosylcytidine derivatives **Ar3**, **Ar5**, **Ar4**, and **Ar2**. Chemical Shifts are in ppm

Resonance	**Ar3**; CD_2Cl_2	**Ar5**; $CDCl_3$	**Ar4**; $CDCl_3$	**Ar2**; D_2O
C2	155.1	154.7	154.7	157.0
C4	163.3	163.7	163.3	166.0
C5	96.3	95.9	96.3	95.5
C6	146.0	145.1	145.9	143.3
C1′	85.7	84.6	84.7	82.1
C2′	76.3[a]	75.2[b]	75.4[c]	75.5[d]
C3′	74.7[a]	74.1[b]	74.6[c]	75.7[d]
C4′	83.6	81.3	80.8	86.5
C5′	62.1	62.7	67.6	69.0
G1	—	—	101.3	103.4
G2	—	—	68.6[c]	71.1
G3	—	—	70.8[c]	73.0
G4	—	—	67.0[c]	68.9
G5	—	—	70.9[c]	75.7[d]
G6	—	—	61.2	61.3
COCH_3	171.0, 170.6, 169.0	170.4, 170.3, 170.2, 168.8	171.1, 170.2, 170.1(2), 169.7, 169.5, 168.7	—
NHCO*CH*$_3$	25.2	24.9	24.8	—
OCO*CH*$_3$	21.0, 20.6 20.37	20.5 (2), 20.43, 20.4	20.6(4), 20.5	—

[a]Assignments of these resonances may be interchanged.
[b]Assignments of these resonances may be interchanged.
[c]Assignments of these resonances may be interchanged.
[d]Assignments of these resonances may be interchanged.

TABLE 3. 500 MHz ^1H nmr data for 2′-deoxycytidine derivatives **C3**, **C5**, **C4**, and **C2**. Chemical shifts are in ppm and coupling constants are in Hz

Resonance	**C3**; CD_2Cl_2	**C5**; $CDCl_3$[a]	**C4**; $CDCl_3$	**C2**; D_2O
H5 ($J_{5,6}$)	7.43 (7.5)	7.47 (7.5)	7.48 (7.5)	7.91 (7.5)
H6	8.21	7.99	8.30	6.05
H1′ ($J_{1',2'}$)	6.23 (6.0)	6.25 (5.7)	6.46 (5.8)	6.29 (6.4)
H2′ ($J_{1',2''}$)	2.66 (7.4)	2.81 (7.2)	2.63 (8.3)	2.44 (6.9)
H2″ ($J_{2',3'}$)	2.39 (7.7)	2.09 (1.9)	2.13 (<1.0)	2.32 (4.4)
($J_{2',2''}$)	(−14.1)	(−14.4)	(−14.2)	(−14.2)
H3′ ($J_{2'',3'}$)	5.36 (6.6)	5.22 (4.4)	5.18 (6.3)	4.52 (6.7)
H4′ ($J_{3',4'}$)	4.18 (3.3)	4.37[b]	3.79[b]	4.22 (3.1)
H5′ ($J_{4',5'}$)	3.97 (2.7)	4.37[b]	4.22[b]	4.20 (3.6)
H5″ ($J_{4',5''}$)	3.91 (2.9)	4.37[b]	4.22[b]	3.83 (6.2)
($J_{5',5''}$)	(−11.1)	[b]	[b]	(−12.1)
NH	8.55	9.44	9.20	—
OH5′ ($J_{OH,5'}$)	2.87 (3.3)	—	—	—
($J_{OH,5''}$)	(5.9)			
GH1 ($J_{1,2}$)	—	—	4.52 (7.9)	4.45 (7.8)
GH2 ($J_{2,3}$)	—	—	5.13 (10.6)	3.56 (9.9)
GH3 ($J_{3,4}$)	—	—	5.05 (3.4)	3.66 (3.4)
GH4 ($J_{4,5}$)	—	—	5.44 (<0.5)	3.94 (<0.5)
GH5 ($J_{5,6}$)	—	—	3.94 (6.6)	3.69 (7.8)
GH6 ($J_{5,6'}$)	—	—	4.25 (6.8)	3.81 (4.2)
GH6′ ($J_{6,6'}$)	—	—	4.15 (−11.4)	3.76 (−11.5)
CH_3CO	2.23, 2.11	2.28, 2.11, 2.10	2.22, 2.21, 2.08, 2.06, 2.05, 1.09	—

[a]Comparable to partial NMR data in ref. 21.
[b]Resonances are not 1st order due to chemical shift overlap and so accurate coupling constants could not be determined.

TABLE 4. 75.5 MHz ^{13}C NMR data for 2′-deoxycytidine derivatives **C3**, **C5**, **C4**, and **C2**. Chemical shifts are in ppm

Resonance	**C3**; $CDCl_3$	**C5**; $CDCl_3$	**C4**; $CDCl_3$	**C2**; D_2O
C2	155.4	155.0	155.2	157.6
C4	162.6	162.4	162.7	166.3
C5	97.0	96.4	96.7	96.3
C6	145.3	143.4	145.0	141.9
C1′	86.1	83.0	84.1	85.5
C2′	38.7	39.1	38.6	39.5
C3′	74.9	74.1	68.8[a]	70.9
C4′	88.0	87.4	86.9	86.5
C5′	62.3	63.6	69.1	69.3
G1	—	—	100.8	103.1
G2	—	—	70.4[a]	71.1
G3	—	—	71.0[a]	72.9
G4	—	—	66.8[a]	68.8
G5	—	—	75.2	75.3
G6	—	—	61.2	61.2
*C*OCH$_3$	170.83, 170.75	170.3, 170.2	170.6, 170.5,	—
			170.4, 170.2,	—
			170.0, 169.5	—
NHCO*C*H$_3$	24.9	25.0	25.0	—
OCO*C*H$_3$	21.0	20.9, 20.8	20.9, 20.7,	—
			20.6(2), 20.5	—

[a]Resonance assignments may be interchanged.

TABLE 5. 500 MHz ^1H NMR data for 2′-deoxythymidine derivatives **T4** and **T2**. Chemical shifts are in ppm and coupling constants are in Hz

Resonance	**T4**; $CDCl_3$	**T2**; D_2O
H6	7.67	7.64
H1′ ($J_{1',2'}$)	6.47 (5.7)	6.31 (11.4)
H2′ ($J_{1',2''}$)	2.32 (9.3)	2.40 (11.4)
H2″ ($J_{2',3'}$)	2.18 (<0.5)	2.40 (3.7)
($J_{2',2''}$)	(−14.3)	[a]
H3′ ($J_{2'',3'}$)	5.17 (6.1)	4.55 (6.7)
H4′ ($J_{3',4'}$)	4.17 (<0.5)	4.17 (3.3)
H5′ ($J_{4',5'}$)	4.28 (<0.5)	4.19 (3.3)
H5″ ($J_{4',5''}$)	3.73 (<0.5)	3.84 (4.6)
($J_{5',5''}$)	(−9.8)	(−10.8)
NH	8.53	—
GH1 ($J_{1,2}$)	4.51 (7.9)	4.46 (7.8)
GH2 ($J_{2,3}$)	5.20 (10.6)	3.57 (9.9)
GH3 ($J_{3,4}$)	5.09 (3.5)	3.66 (3.3)
GH4 ($J_{4,5}$)	5.45 (<0.5)	3.93 (<0.5)
GH5 ($J_{5,6}$)	3.97 (6.6)	3.70 (7.9)
GH6 ($J_{5,6'}$)	4.19 (6.8)	3.81 (4.2)
GH6′ ($J_{6,6'}$)	4.14 (−11,3)	3.75 (−11.7)
CH_3C	2.01	1.91
CH_3CO	2.12, 2.11	—
	2.09, 2.05(2)	

[a]The H2′ and H2″ resonances are overlapped and so this coupling constant could not be determined. The remaining coupling constants involving these two resonances were determined from the H1′ and H3′ multiplet patterns.

TABLE 6. 75.5 MHz ^{13}C NMR data for 2′-deoxythymidine derivatives **T4** and **T2**. Chemical shifts are in ppm

Resonance	**T4**; $CDCl_3$	**T2**; D_2O
C2	150.4	152.0
C4	163.6	166.9
C5	111.4	111.8
C6	135.9	137.7
C1′	84.6[a]	85.4[a]
C2′	37.0	38.6
C3′	68.8[b]	68.9
C4′	83.4[a]	85.6[a]
C5′	70.1	69.7
G1	101.3	103.5
G2	70.3[b]	73.1
G3	71.1[b]	71.18
G4	66.9[b]	71.13
G5	75.4[b]	75.5
G6	61.0	61.3
CH_3C	12.5	11.9
*C*HCH_3	170.7, 170.3,	—
	169.92(2), 169.85	
OCO*C*H$_3$	20.9, 20.63,	—
	20.58, 20.4,	
	20.0	

[a]Assignments for C4′ and C1′ may be interchanged.
[b]Assignments for these resonances may be interchanged.

gassed.[8] Inversion–recovery experiments used the standard Bruker software for T_1 determinations.

UV spectra were recorded with a Cary 20 UV–VIS spectrophotometer. Infrared spectra of compounds in saturated CH_2Cl_2 solutions were recorded with a Nicolet FTIR 5DX instrument. These solutions were subjected to stepwise dilutions to halve the previous concentration, and recorded again.

[8]A referee has suggested that traces of oxygen in solutions may significantly influence the quantitation of the NOE measurements. Although the specimens were not purged free of air, a possible experimental error would not exceed 15% as can be seen from the comparison of T_1's in $CDCl_3$ for **Ar3** before and after purging (data after purging are in parentheses): NH, 0.70 (0.71); H5, 1.55 (1.79); H6, 1.20 (1.32); H1′, 1.17 (1.46); H2′, 1.31 (1.46); H3′, 1.35 (1.53); H4′, 1.14 (1.24); H5′ + H5″, 0.41 (0.46); O5′H, 0.84 (0.82).

TABLE 7. Summary of EI mass spectral data for nucleoside intermediates. Relative intensities (%) of the fragment ions are shown in parentheses

Compound	M^b	Sugar	Hexose	Base (BH; B + 2H)	Others
Ar3	369(1.97)	217(38)	—	154(50) 112(80);	193(100); B + 41 (ref. 21)
Ar4	699(0.8)	547(4)	331(37)	154(50)	81(100)
					193(48); B + 41
					352(62); M − Hex[a]
C3	311(3.4)	159(15)	—	112(30)	69(100)
				154(60)	
C4	641(0.2)	489(0.2)	331(8.6)	152(15)	81(100)
				153(10)	294(2.6); M − Hex[a]
				154(10)	
T4	614(0.5)	489(2.8)	331(29)	126(10)	81(100)
				127(20)	
A5	411(4.5)	259(45.7)	—	154(40)	193(100); B + 41
C5	353(5)	201(8)	—	154(95)	81(100)

[a]Hex = hexose.
[b]For molecular formulae, see Experimental.

All mass spectra were recorded with a VG Analytical ZAB-SE mass spectrometer at the Mass Spectrometry Laboratory of the Carbohydrate Research Centre, University of Toronto. For Fast Atom Bombardment (FAB) the samples dissolved in methanol were loaded onto a target that contained 1–2 μL of glycerol and thioglycerol (3:1) and were bombarded by neutral xenon atoms (8 keV and 1 mA anode current) generated by an Ion-Tech Saddle Field ion gun. The spectra were recorded using the multichannel analyzing mode with a VG 11-250 data system. For low-resolution FAB-MS, the resolution was set at 1000 and the instrument was calibrated with CsI. For high-resolution FAB-MS measurements, i.e., 10 000 resolution (10% valley definition), 2% polyethylene glycol (avg. mol. wt. 300 or 600) was added to the matrix as internal calibrant and a narrow mass range was scanned using the accelerating voltage.

Electron-impact mass spectra (EI) were recorded using the direct insertion probe, and the source conditions were 200°C, 70 eV, and the electron trap current 100 μA. Resolution was approximately 1000 for low-resolution spectra and 10 000 for high-resolution spectra. Perfluoroalkane used as internal calibrant was introduced into the ion source via the septum inlet and the high-resolution spectra were acquired by scanning the magnet.

All derivatives of sufficient volatility, i.e., protected nucleosides, were characterized by electron ionization mass spectrometry, EIMS, and the title galactosylated nucleosides by fast-atom-bombardment mass spectrometry, FAB-MS. All new compounds purified by chromatography gave satisfactory elemental composition as determined by high-resolution MS spectra. The mass spectra of the peracetylated nucleosides exhibit facile elimination of ketene or acetic acid from the original molecules or fragment ions but otherwise the fragment ions are very similar to those reported for the trimethylsilyl derivatives (38, 39). Ions characteristic of the carbohydrate and base moieties are summarized in Table 7. In the case of the compounds **Ar3** and **C3**, the free hydroxyl group was also blocked by the trimethylsilyl group to confirm the presence of the free hydroxyl group at the 5′-position by observing the *m/z* 103 ion in the derivatized compounds (35, 36).

N^4-*Acetyl-1-[2′,3′-di-*O*-acetyl-*β*-D-arabinofuranosyl] cytosine (**Ar3**)*

1-[β-D-Arabinosyl]cytosine (**Ar1**, 3.0 g, 12.3 mmol) was dissolved in dry methanol (300 mL) and acetic anhydride (3.0 mL) and refluxed under argon. Additional portions of acetic anhydride were added after 1 h and, again, after 2 h (40). After heating for a total of 3 h, the mixture was concentrated to about 10 mL, cooled on ice, and diethyl ether (100 mL) was added. The resulting white precipitate was isolated by filtration with suction. The dried solid was dissolved in dry pyridine (60 mL) containing 4-dimethylaminopyridine (20 mg) under argon, dimethoxytrityl chloride (6.27 g; 18.5 mmol) was added, and the mixture was stirred at room temperature for 24 h. Another portion (0.9 g) of dimethoxytrityl chloride was added and stirring was continued for another 8 h. At this time TLC (1% CH_3OH:CH_2Cl_2) indicated that the reaction was complete, and the flask was cooled in an ice-bath, acetic anhydride (18 mL) was added dropwise by a syringe, and the mixture was stirred overnight in the ice-bath. After warming up to room temperature, the volatiles were removed by evaporation in high vacuum, and the residue was subjected to chromatography on a column of silica gel eluted sequentially with 500 mL CH_2Cl_2, 1% CH_3OH:CH_2Cl_2, 3% CH_3OH:CH_2Cl_2, and 5% CH_3OH:CH_2Cl_2, to yield slightly impure solid peracetylated 5′-dimethoxytritylated compound (6.0 g). After drying, the solid was dissolved in CH_2Cl_2 (150 mL) and a 1% solution of $CHCl_2COOH$ in CH_2Cl_2 (1.5 mL; 18.3 mL) was added dropwise under argon. After stirring was continued for another 5 min, solid $NaHCO_3$ (1.6 g; 20 mmol) was added and the reaction mixture was stirred for another 10 min. The solids were filtered off, rinsed with $CHCl_3$, and the combined filtrates were evaporated to dryness. The residue was at once subjected to chromatography on a column of silica gel and eluted sequentially with 1% CH_3OH:CH_2Cl_2, 3% CH_3OH:CH_2Cl_2, 5% CH_3OH:CH_2Cl_2, and 10% CH_3OH:CH_2Cl_2 to yield slightly impure **Ar3** (3.5 g; 77%). After repeated chromatography on silica gel with 5% CH_3OH:CH_2Cl_2, a pure specimen of **Ar3** ($C_{15}H_{19}N_3O_8$) was obtained: $[\alpha]_D$ +96.3 (*c* 0.30, CH_3OH:CH_2Cl_2 1:15); UV: 304, 245 nm (CH_2Cl_2) (lit. (16) UV 299, 249 nm); IR: 3482, 3295, 1754, 1727, 1674, 1655 (sh), 1628 cm^{-1} (CH_2Cl_2) (lit. (19) IR 3400, 3280, 3240, 1750, 1710, 1655 cm^{-1}).

N^4-*Acetyl-1-[3′-*O*-acetyl-2′-deoxy-*β*-D-ribofuranosyl] cytosine (**C3**)*

Compound **C3** ($C_{15}H_{19}N_3O_7$) was prepared in a manner analogous to that described for **Ar3**; yield 40%; $[\alpha]_D$ +54.9 (c 0.36, CH_3OH:CH_2Cl_2 1:15); UV: 302, 245 nm (CH_2Cl_2) (lit. (20) UV 296, 247 nm); IR: 3620, 3404, 1739, 1727, 1662, 1632 cm^{-1} (CH_2Cl_2).

N^4-*Acetyl-1-[2′,3′-di-*O*-acetyl-5′-(2,3,4,6-tetra-*O*-acetyl-*β*-D-galactopyranosyl)-*β*-D-arabinofuranosyl] cytosine (**Ar4**)*

Procedure A

To a solution of **Ar3** (300 mg, 0.81 mmol) in CH_2Cl_2 (20 mL) containing activated powdered 4 Å molecular sieves (500 mg) was added a solution of trichloroacetimidate **6** (520 mg; 1.22 mmol) in CH_2Cl_2 (36.6 mL) at 5°C, followed by $BF_3{\cdot}Et_2O$ (150 μL; 1.22 mmol). After stirring for 40 min, solid NH_4HCO_3 (100 mg) was added and stirring was continued for another 10 min. The solids were filtered off, rinsed with CH_2Cl_2, and the combined filtrates were evaporated to dryness. The residue was subjected to flash chromatography on silica gel, which on elution with 3% CH_3OH:CH_2Cl_2, followed by 5% CH_3OH:CH_2Cl_2, yielded fractions containing **6**, decomposition products of **6**, disaccharide **7** (see below), and a mixture of galactosylated

nucleoside **Ar4** and peracetylated nucleoside **Ar5** (total 296 mg). This mixture was further chromatographed on Sephadex LH-20 in CH_3OH:$CHCl_3$ (4:3) to give **Ar4** (174 mg, 20%) and a fraction containing **Ar5** contaminated with **Ar4**. The latter fraction, after a repeated chromatography on LH-20, yielded an additional small amount of **Ar4** and pure N^4-*acetyl-1-[2',3',5'-tri*-O-*acetyl*-β-D-arabinofuranosyl] cystosine (**Ar5**) ($C_{17}H_{21}N_3O_{17}$, $[\alpha]_D$ +118.2 (*c* 4.25, CH_2Cl_2) (lit. (footnote 6) $[\alpha]_D$ +87); UV: 304, 245 nm (CH_2Cl_2)). The fractions containing **Ar4** were further purified by flash chromatography on silica gel in 4% CH_3OH:$CHCl_3$ to yield pure **Ar4** ($C_{29}H_{37}N_3O_{17}$; 120 mg, 14%), $[\alpha]_D$ +52.0 (*c* 0.59, CH_2Cl_2); UV: 303, 244 nm (CH_2Cl_2). FAB-MS showed *m/z*: 700 (MH^+), 658 ($MH^+ - CH_2CO$), 370 (MH^+ − hexose − H_2O), and 331 (hexose). The observed Exact Mass value of MH^+ is 700.2215 (calcd. 700.2201).

Procedure B

Alcohol **Ar3** (500 mg, 1.37 mmol), trichloracetimidate **6** (945 mg, 1.92 mmol), and silver triflate (458 mg, 1.8 mmol) were dried together at high vacuum overnight. Chloroform (5 mL), freshly distilled from CaH_2, was added to this mixture of solids under an atmosphere of argon at room temperature. The reaction mixture was stirred for 24 h and then diisopropylethylamine (5 drops) was added, the solids were filtered off and rinsed with dichloromethane (2 × 23 mL), and the combined filtrate plus washings were washed with cold 0.05 M HCl (3 × 50 mL) and cold saturated aqueous bicarbonate (1 × 50 mL). The organic layer was dried with $MgSO_4$, filtered, and evaporated to dryness. The residue was purified by sequential chromatography on silica gel and Sephadex LH20 (cf. procedure A) to yield 394 mg, 41% of **Ar4**.

N-*[2-O-(2,3,4,6-Tetra*-O-*acetyl*-β-D-*galactopyranosyl)-3,4,6-tri*-O-*acetyl*-α-D-*galactopyranosyl] trichloroacetamide* (*7a*)

The fractions containing **7***a*, *b* (cf. the preparation of **Ar4**) were subjected to another chromatography on silica gel using ethyl acetate:hexane (3:2) as eluent, yielding **7***a* contaminated with *N*-[2-*O*-(2,3,4,6 tetra-*O*-acetyl-β-D-galactopyranosyl)-3,4,6-tri-*O*-acetyl-β-D-galactopyranosyl] trichloroacetamide (**7***b*). ^{1}H NMR: 7.33 (d, $J_{1,NH}$ 5.0, NH), 5.77 (t, $J_{1,2}$ 5.4, $J_{1,NH}$ 5.0, αGal H1), 5.38 (m, *J* 1.9, αGal H4), 5.35 (brd, $J_{3,4}$ 3.4, $J_{4,5} < 0.5$, βGal H4), 5.11 (dd, $J_{1,2}$ 7.5, $J_{2,3}$ 9.6, βGal H2), 5.11 (dd, $J_{2,3}$ 9.6, $J_{3,4}$ 3.6, αGal H3), 4.98 ($J_{2,3}$ 9.6, $J_{3,4}$ 3.4, βGal H3), 4.62 (d, $J_{1,2}$ 7.5, βGal H1), 4.18 (dd, $J_{1,2}$ 5.4, $J_{2,3}$ 9.6, αGal H2), 4.16 (m, αGal H6 and H6'), 4.13 (m, αGal H5), 4.11 (dd, $J_{6,6'}$ −11.3, $J_{5,6}$ 6.6, βGal H6), 4.08 (dd, $J_{6,6'}$ −11.3, $J_{5,6'}$ 6.2, βGal H6'), 3.93 (brdd, $J_{5,6}$ 6.6, $J_{5,6'}$ 6.2, βGal H5), 2.16, 2.12, 2.08, 2.06, 2.04, 2.03, 1.97 (7 × s, CH_3CO); ^{13}C NMR: 170.4 (2), 170.0 (2) 169.9, 169.5, 169.0 (CH_3COO), 162.5 CON, 101.7 βGal C1, 92.2 CCl_3, 78.0 αGal C1, 72.7, 71.0, 70.4, 68.9, 68.7, 67.9, 66.9, 66.7 (αGal and βGal C2, C3, C4 and C5), 61.1 and 60.8 (αGal and βGal C6), 20.62(4), 20.58(2), 20.5 (CH_3CO). FAB-MS exhibited fragments at *m/z* 780 (MH^+), 619 (peracetylated hexose − hexose), and 331 (peracetylated hexose). Exact Mass: calcd. for $C_{28}H_{37}NO_{18}Cl_3$: 780.1076; found: 780.1074.

N^4-*Acetyl-1-[2'-deoxy-2',3'-di*-O-*acetyl-5'-(2,3,4,6-tetra*-O-*acetyl*-β-D-*glactopyranosyl)*-β-D-*ribofuranosyl] cytosine* (***C4***)

Compound **C4** ($C_{27}H_{35}N_3O_{15}$, $[\alpha]_D$ +6.93 (*c* 0.37, CH_2Cl_2); UV: 303, 243 nm (CH_2Cl_2)) was prepared as described for **Ar4**. Yield 9% (procedure A) and 35% (procedure B). FAB-MS exhibited ions at *m/z* 642 (MH^+), 600 ($MH - CH_2CO$)$^+$, and 331 (peracetylated hexose). Exact Mass calcd. for $C_{27}H_{35}N_3O_{15}$: 642.2146; found: 642.2128. Compound **C4** was accompanied by N^4-*acetyl-1-[2'-deoxy-3',5'-di*-O-*acetyl*-β-D-ribofuranosyl] cytosine (**C5**), ($C_{15}H_{19}N_3O_7$, $[\alpha]_D$ +71.4 (*c* 12.4, CH_2Cl_2); UV: 304, 244 nm (CH_2Cl_2), (lit. (21) UV 307, 239 nm). The compounds **C4** and **C5** were separated by chromatography on silica gel and Sephadex LH20.

*1-[3'-*O-*Acetyl-2' deoxy-5'-(2,3,4,6-tetra*-O-*acetyl*-β-D-*galactopyranosyl)*-β-D-*ribofuranosyl] thymine* (***T4***)

Compound **T4** ($C_{26}H_{34}N_2O_{15}$, $[\alpha]_D$ −13.2 (*c* 0.67, CH_2Cl_2); UV: 262 nm (CH_2Cl_2)) was prepared as described for **Ar4**. Yield 24% (procedure A). FAB-MS exhibited ions at *m/z* 637(MNa^+), 615 (MH^+), 573 ($MH - CH_2CO$)$^+$, and 331 (peracetylated hexose). Exact Mass calcd. for $C_{26}H_{34}N_2O_{15}$: 615.2037; found: 615.2040. Compound **T4** was accompanied by *1-[2'-deoxy-3',5'-di*-O-*acetyl*-β-D-ribofuranosyl] thymine (**T5**).

*1-[5'-*β-D-*Galactopyranosyl*-β-D-*arabinofuranosyl] cytosine* (***Ar2***)

Peracetylated galactosylated nucleoside **Ar4** (120 mg) was dissolved in CH_3OH (20 mL) under argon, DBU (5 drops) was added, and the mixture was stirred overnight. After an addition of toluene (15 mL), the reaction mixture was concentrated using a rotary evaporator until a precipitate appeared. The solid was filtered off, rinsed with toluene and CH_2Cl_2, and then redissolved in CH_3OH and subjected to chromatography on LH-20 using $CHCl_3$:CH_3OH 4:3. Pure **Ar2** (62 mg, 89%) ($C_{15}H_{22}N_3O_{10}$, $[\alpha]_D$ +93.0 (*c* 0.13, CH_3OH); UV: 273, 230 nm (CH_3OH)) was obtained. FAB-MS exhibited ions at *m/z* 406 (MH^+), 244 (MH^+ − hexose), and 112 (base + 2H). Exact Mass calcd. for $C_{15}H_{22}N_3O_{10}$: 406.1462; found: 406.1469.

*1-[2'-Deoxy-5'-*β-D-*galactopyranosyl*-β-D-*ribofuranosyl] cytosine* (***C2***)

Compound **C2**; yield 71% ($C_{15}H_{22}N_3O_9$, $[\alpha]_D$ +51.3 (*c* 0.37, CH_3OH); UV: 273, 230 nm (CH_3OH)) was prepared as described for **Ar2**. FAB-MS exhibited ions at *m/z* 390 (MH^+), 228(MH^+ − hexose), and 112 (base + 2H). Exact Mass calcd. for $C_{15}H_{22}N_3O_9$: 390.1512; found: 390.1502.

*1-[2'-Deoxy-5'-*β-D-*galactopyranosyl*-β-D-*ribofuranosyl] thymine* (***T2***)

Compound **T2** ($C_{16}H_{24}N_2O_{10}$, $[\alpha]_D$ +11.2 (*c* 0.42, CH_3OH:H_2O 3:2); UV: 264, 202 nm (CH_3OH)) was prepared as described for **Ar2**. Yield 96%; FAB-MS exhibited ions at *m/z* 427 (MNa^+), 405 (MH^+), 243 (MH^+ − hexose), and 127 (base + 2H). Exact Mass calcd. for $C_{16}H_{24}N_2O_{10}$: 405.1509; found: 405.1544.

Acknowledgements

This work was supported through the Protein Engineering Network of Centres of Excellence (Canada), the Natural Sciences and Engineering Research Council of Canada, and the Department of Veteran Affairs (USA). The Carbohydrate Research Centre is a Faculty of Medicine (University of Toronto) facility partially supported by a maintenance grant from the Medical Research Council of Canada (MT-6499). Our thanks are due to Mr. Richard Crombie for obtaining the mass spectra. The skilful technical assistance of Mr. Raymond Ngan and Mr. Minh D. To is appreciated with thanks.

1. T.H. Tang, D.M. Whitfield, S.P. Douglas, J.J. Krepinsky, and I.G. Csizmadia. Can. J. Chem. **70**, 2434 (1992).
2. F.L. Moolten. CRC Crit. Rev. Immunol. **10**, 203 (1990).
3. H. Paulsen. Angew. Chem. Int. Ed. Engl. **21**, 155 (1982); **29**, 823 (1990).
4. S.P. Douglas, D.M. Whitfield, and J.J. Krepinsky. Abstracts, Fuji '90 Post-Symposium: Recent progress of synthetic methods in carbohydrates and their applications to synthetic chemistry. Susono City, Shizuoka, Japan. 1990; The Fuji Training Institute, Shizuoka.
5. F.W. Lichtenthaler, Y. Sanemitsu, and T. Nohara. Angew. Chem. Int. Ed. Engl. **17**, 772 (1978).
6. N.D. Chkanikov and M.N. Preobrazhenskaya. (*a*) Bioorg. Khim. **6**, 67 (1980); *b*) J. Carbohydr. Nucleosides Nucleotides, **8**, 391 (1981).
7. N.B. Hanna, R.K. Robins, and G.R. Revankar. Carbohydr. Res. **165**, 267 (1987).
8. N.K. Kochetkov and A.F. Bochkov. *In* Recent developments in the

chemistry of natural carbon compounds. *Edited by* A. Bognár, V. Bruckner, and Cs. Szántay. Akademiai Kiadó, Budapest. 1971.
9. N.D. Chkanikov, V.N. Tolkachev, and M.N. Preobrazhenskaya. Bioorg. Khim. **4**, 1620 (1978).
10. H. Bredereck, A. Wagner, G. Faber, H. Ott, and J. Rauther. Chem. Ber. **92**, 1135 (1959).
11. S.P. Douglas, D.M. Whitfield, and J.J. Krepinsky. J. Am. Chem. Soc. **113**, 5095 (1991).
12. E. Bosch, M. Moreno, and J.M. Leuch. Can. J. Chem. **70**, 1640 (1992).
13. W. Saenger. Angew. Chem. **85**, 680 (1973); K.H. Scheit. Nucleotide analogs. Wiley–Interscience, New York. 1980. p. 3.
14. L.P. Kuhn. J. Am. Chem. Soc. **74**, 2492 (1952).
15. J.G. Dalton, A.L. George, F.E. Hruska, T.N. McGaig, K.K. Ogilvie, J. Feeling, and D.J. Wood. Biochim. Biophys. Acta, **478**, 261 (1978).
16. A. Saran, B. Pullman, and D. Perahia. Biochim. Biophys. Acta, **349**, 189 (1974).
17. F.E. Hruska, D.J. Wood, T.N. McCaig, A.A. Smith, and A. Holy. Can. J. Chem. **52**, 497 (1974).
18. G.I. Birnbaum and D. Shugar. Top. Nucleic Acid Struct. Part 3. *Edited by* S. Neidle. Macmillan Press, London. 1987. p. 1.
19. W.J. Wechter. J. Med. Chem. **10**, 762 (1967).
20. A.M. Michelson and A.R. Todd. J. Chem. Soc. 34 (1954).
21. P. Zimmermann, R. Bommer, T. Bär, and R.R. Schmidt. J. Carbohydr. Chem. **7**, 435 (1988).
22. R.R. Schmidt. Angew. Chem. Int. Ed. Engl. **25**, 212 (1986).
23. A.P. Martinez, W.W. Lee, and L. Goodman. J. Med. Chem. **9**, 268 (1966).
24. M.J. Robins, M. MacCoss, S.R. Naik, and G. Ramani. J. Am. Chem. Soc. **98**, 7381 (1976).
25. A.M. Michelson and A.R. Todd. J. Chem. Soc. 816 (1955).
26. R. Bommer, W. Kinzy, and R.R. Schmidt. Liebigs Ann. Chem. 425 (1991).
27. (*a*) P.J. Garegg, P. Konradsson, I. Kvarnström, T. Norberg, S.C.T. Svensson, and B. Wigilius. Acta Chem. Scand. B, **39**, 569 (1985); (*b*) T. Ziegler, P. Kováć, and C.P.J. Glaudemans. Ann. Chem. 613 (1990); (*c*) N.I. Uvarova, G.I. Oshitok, and G.B. Elyakov. Carbohydr. Res. **27**, 79 (1973); (*d*) A. Ya. Khorlin, V.A. Nesmeyanov, and S.E. Zurabyan. Carbohydr. Res. **43**, 69 (1975); (*e*) R.U. Lemieux. Chem. Can. **16**, 14 (1964); (*f*) J. Banoub and D.R. Bundle. Can. J. Chem. **57**, 2091 (1979); (*g*) G. Wulff, and G. Röhle. Angew. Chem. Int. Ed. Engl. **13**, 157 (1974).
28. F.J. Urban, B.S. Moore, and R. Breitenbach. Tetrahedron Lett. **31**, 4421 (1990).
29. Y.D. Vankar, P.S. Vankar, M. Behrendt, and R.R. Schmidt. Tetrahedron, **47**, 9985 (1991).
30. S.P. Douglas, D.M. Whitfield, and J.J. Krepinsky. J. Carbohydr. Chem. **12**, 131 (1993).
31. (*a*) S. Nilsson, H. Lönn, and T. Norberg. Glycoconjugate J. **8**, 9 (1991); (*b*) C.W. Yue, J. Thierry, and P. Potier. Tetrahedron Lett. **34**, 323 (1993).
32. A.J. Kirby. The anomeric effect and related stereoelectronic effects at oxygen. Springer, Berlin and New York. 1983.
33. (*a*) K. Kefurt, J. Staněk, Jr., Z. Kefurtová, and J. Jarý. Collect. Czech. Chem. Commun. **70**, 1640 (1992); (*b*) G.J.F. Chittenden and J.G. Buchanan. Carbohydr. Res. **11**, 379 (1969); (*c*) G.J.F. Chittenden. Carbohydr. Res. **43**, 366 (1975); (*d*) D.M. Whitfield, J.P. Carver, and J.J. Krepinsky. J. Carbohydr. Chem. **4**, 369 (1985).
34. (*a*) A. Engdahl and B. Nelander. J. Phys. Chem. **89**, 2860 (1985); (*b*) J. Wanna, J.A. Menapace, and E.R. Bernstein. J. Phys. Chem. **91**, 2533 (1987); (*c*) F.A. Baiocchi, J.H. Williams, and J. Klemperer. J. Phys. Chem. **87**, 2079 (1983); (*d*) L. Andrews, G.L. Johnson, and S.R. Davis. J. Phys. Chem. **89**, 1706 (1985); (*e*) S. Suzuki, P.G. Green, R.E. Bumgarner, S. Dasgupta, W.A. Goddard III, and G.A. Blake. Science, **257**, 942 (1992); (*f*) J.L. Atwood, F. Hamada, K.D. Robinson, G.W. Orr, and R.L. Vincent. Nature, **349**, 683 (1991).
35. (*a*) J.L. Bredas and G.B. Street. J. Chem. Phys. **90**, 7291 (1989); (*b*) T.-H. Tang, W.-J. Hu, D.-Y. Yan, and Y.-P. Cui. J. Mol. Struct. (THEOCHEM), **207**, 319 (1990); (*c*) B.V. Cheney, M.W. Schulz, J. Cheney, and W.G. Richards. J. Am. Chem. Soc. **110**, 4195 (1988).
36. E. Breitmeier and W. Voelter. Monogr. Mod. Chem. **5**, 270 (1978).
37. D.B. Davies. Prog. NMR Spectrosc. **12**, 135 (1978).
38. J.A. McCloskey. Methods Enzymol. **93**, 825 (1990), and references therein.
39. H. Pang, K.H. Schram, D.L. Smith, S.P. Gupta, L.B. Townsend, and J.A. McCloskey. J. Org. Chem. **47**, 3823 (1982).
40. K.A. Watanabe and J.J. Fox. Angew. Chem. Int. Ed. Engl. **5**, 579 (1966).
41. (*a*) J.J.P. Stewart. J. Comput. Chem. **10**, 209 (1989); **10**, 221 (1989); (*b*) J. Comput. Chem. **12**, 320 (1991).
42. E.A. Gallo and S.H. Gellman. Tetrahedron Lett. **33**, 7485 (1992); (*b*) S.K. Burley and G.A. Petsko. Adv. Protein Chem. **39**, 125 (1988).
43. D.B. Davies, P. Rajani, and H.J. Sadikot. J. Chem. Soc. Perkin Trans. 2, 279 (1985).
44. J.S. Sherfinski and R.E. Marsh. Acta Crystallogr. **52**, 497 (1974).
45. J.J. Krepinsky, D.M. Whitfield, S.P. Douglas, N. Lupescu, D. Pulleyblank, F.L. Moolten. Methods Enzymol. **247**, 144 (1994).

Appendix I

Many cancer chemotherapeutic agents are capable of eradicating cancer cells. Practical clinical problems arise from the low selectivity of these agents: they kill normal healthy cells along with the cancerous ones, albeit at a slower rate. A significant increase in selectivity would dramatically enhance the curative potential of such drugs. For instance, if the cytotoxicity of β-D-arabinofuranosylcytosine, Ara C (**Ar1**), clinically used for treatment of leukemias, can be abolished by glycosylation of the drug, and if a suitable glycosidase is present only in the neoplastic cells, then this glycosylation would bring about the desired increase in the selectivity. One of us has been investigating such a scenario for future cancer chemotherapies (cf. ref. 2 for a review). The glycosidase employed is β-galactosidase, the product of the *Escherichia coli* β-galactosidase (lacZ) gene, inserted artificially into cancer cells. Galactosylation of Ara C then masks the toxicity of the nucleoside until the masking β-galactoside is removed inside such a cell, and Ara C is activated. Galactosylation of an Ara C antagonist, i.e., 2-deoxy-β-D-ribofuranosylcytosine (**C1**), would result in analogous deactivation of this nucleoside. To permit direct testing of this hypothesis, galactosylated nucleosides 5′-β-galactosyl β-D-arabinofuranosylcytosine (**Ar2**) and 5′-β-galactosyl 2-deoxy-β-D-ribofuranosylcytosine (**C2**) were required.

Appendix II

Galactosylation presumably turns the active compounds into proactive ones: e.g., toxic Ara C (**Ar1**) is turned into nontoxic Gal-Ara C (**Ar2**), which should become toxic again after the hydrolysis of the galactose–arabinose linkage using β-galactosidase. Two cell lines were used: (1) mouse fibroblasts *psi-2* capable of packaging retroviral genetic

information into infectious viral particles that can transfer the information to new cells, and its subline *psi-2 BAG* α containing retroviral information that includes β-galactosidase (lacZ) gene from *Escherichia coli*; and (2) murine sarcoma cell line *K3T3* from which a subline *K3T3 lacZ+* was created by exposing *K3T3* cells to *lacZ*-bearing virus released from *psi-2 BAG* α. As expected, Ara C (**Ar1**), added to the growth medium, was equally toxic to all the above cell lines and sublines. When Gal-Ara C (**Ar2**) was used, it was toxic to the sublines expressing lacZ gene, i.e., to *psi-2 BAG* α and *K3T3 lacZ+* cells, albeit about two orders of magnitude less than Ara C (**Ar1**), while non-toxic to the lines without lacZ gene (*psi-2* and *K3T3*).

Galactosylated compounds **C2** and **T2** were also used to examine rescue effects of deoxycytidine and thymidine. For instance, when cells are grown in the presence of Ara C (**Ar1**), they can be saved from death by an addition into the medium of deoxycytidine that acts as an antidote to the toxicity of **Ar1**. When such an experiment has been performed with *K3T3* and *K3T3 lacZ+* cells and galactosylated deoxycytidine (**C2**), only the line containing lacZ gene showed a slight rescue effect. A similar small rescue effect was observed when *psi-2 BAG* α cells (but not *psi-2* cells) were exposed to a "HAT" (hypoxanthin–aminopterin–thymidine) regimen in which galactosylated thymidine (**T2**) was substituted for thymidine as an antidote to aminopterin toxicity.

The low level of activity of galactosylated compounds in lacZ-containing cell lines may reflect, at least in part, two possible causes: first, the galactosylated compound **Ar2** is inefficiently taken up by the cells or, second, the rate of release of toxic **Ar1** in the cells is slow. The latter may happen since the enzyme is originally targeted for the Galβ1-4Gal linkage and so it may be less active for the Galβ1-5Ara linkage. Studies to clarify these issues and remedy the situation are underway.