Pseudouridine, a Carbon-Carbon Linked Ribonucleoside in Ribonucleic Acids: Isolation, Structure, and Chemical Characteristics

WALDO E. COHN

From the Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

(Received for publication, November 17, 1959)

Until the invention of chromatographic analytical methods (1), which do not depend upon the precipitation of major components of complex mixtures, knowledge of the constituents of nucleic acids was confined to the principal bases, adenine, guanine, cytosine, uracil, and thymine, and their nucleosides and nucleotides. Although the existence of isomeric forms of the nucleotides was the first discovery attributable to the new methods (3, 4), the presence of 5-methylcytosine (5, 6) (and later 5-hydroxymethylcytosine (7)) in deoxyribonucleic acids constituted the first indication of different bases. At about the same time, unidentified peaks (and spots) began to appear in chromatograms of RNA hydrolysates. Some of these were compounds of one or more of the known nucleotides (in polynucleotide form, for example), but others, more recently recognized, have been identified as methyl derivatives of the common RNA nucleotides occurring in small amounts (8, 9).

All of the more recently discovered bases, as with 5-methylcytosine, occur in the usual N-ribosyl linkage. However, the first “minor” nucleotide from RNA to appear (Fig. 1 of Cohn and Volkin (10), peak labeled ?) was not readily identified with such a linkage. In retrospect it seems likely that the difficulty derives from the fact, demonstrated in this paper, that the base-ribosyl linkage is not C-N but C-C.

The isolation and certain of the physical and chemical properties of this unusual nucleoside have been described (1a, 11-13). Much of the chemical and physical evidence that establishes its structure as 5-ribosyluracil has been summarized previously (1a) and the term pseudouridine has been proposed for a trivial name. In this paper are described certain of the chemical, chromatographic, spectrophotometric, and nuclear magnetic properties of the substance and its derivatives that pertain to the question of structure and natural occurrence.

EXPERIMENTAL
Materials and Methods

The material used as a source for the preparation of large (50 to 500 mg) amounts of pseudouridylate was yeast RNA from the Schwarz Laboratories and the Pabst Laboratories. Both contained 0.2 to 0.3% of the desired substance. Alkaline hydrolysis was carried out in 0.5 N NaOH for 18 hours at 37°. NaOH was removed by adding Dowex 50-H+ or Amberite IR-120(H+) with stirring until the pH fell to approximately 3. To the filtered solution, one volume of alcohol was added and the solution was kept at 0° overnight to precipitate guanylic acid. Sometimes a second precipitation, at 75% alcohol, was used to eliminate part of the adenylic and cytidylic acids. The supernatant solution, somewhat diluted and made alkaline with ammonia, was subjected to ion exchange chromatography.

The ion exchange chromatography of the mixed 2' and 3' nucleotides utilized columns of Dowex 1-formate (200 to 400 mesh, 8% divinylbenzene content) of various sizes, depending upon the amounts of nucleotide material being handled. The sequence of acids and buffers used was essentially that of Cohn and Volkin (10), but the final separation of pseudouridylic acid from uridylic acid was performed by chromatography. A common contaminant of the 2' is separable only by a repetition of the third elution sequence of Fig. 1 (see Section 1c and Fig. 2) or by paper chromatography in NH4 formate-isobutyric acid buffer (15). Most dephosphorylations were performed with acid prostatic phosphatase, generously supplied by Dr. Gerhard Schmidt.

RESULTS

1. Separation of Pseudouridylic Acid by Ion Exchange Chromatography

a. Preparation from Alkaline Hydrolysate of RNA—Pseudouridylic acid is readily isolated from alkaline hydrolysates of RNA by the same procedure that first led to its detection by Cohn and Volkin (10). Pseudouridylic acid obtained in this fashion is relatively free of all other major nucleotides. Any adenylic acids that happen to contaminate the product are automatically removed in the concentration and recovery steps, but any uridine 2'-phosphate picked up in the pseudouridylic acid peak is separable only by a repetition of the third elution sequence of Fig. 1 (see Section 1c and Fig. 2) or by paper chromatography. A common contaminant of the 2's pseudouridylic acid is uridine 5'-phosphate, easily eliminated by hydrolysis with a 5' nucleotidase.

It is possible to combine the first two steps (0.02 M and 0.2 M formic acid), in which case the cytidylic and adenylic acids

1 In preparation for Biochemical Preparations, Vol. 8, in press.
may or may not separate from each other, or to omit them altogether and proceed directly to the formate buffer. In the latter case, pseudouridylic acid is obtained in a large amount of cytidylic acid, from which it separates in the concentration procedure described in Section Ib.

In any case, the appearance of pseudouridylic acid in a cytidylic, adenylic, or uridylic acid fraction, or of any of these in the pseudouridylic acid fraction, may be detected with great (approx. 1%) sensitivity by observing the spectrophotometric shift at 900 nm when the pH is raised from below 8 to 12 and 13. This is described in detail in a later section, but we may note here that pseudouridylic acid is the only one of this group that shows a marked enhancement of absorption in the 290 to 295 nm region at pH 12 over that observed at pH 8.

b. Concentration—The pseudouridylic acid, obtained as indicated in formate (or other) salt solution and possibly contaminated with adenylates or cytidylates, is readily redissolved on a smaller anion exchange column if it is first made alkaline (pH ~ 11) with NH₄OH. With an ionization (pK approx. 9.6) in the pyrimidine nucleus in addition to the secondary phosphate ionization, it has a triple affinity for the exchanger. Advantage is taken of this fact and the column is converted, after the absorption step, from the formate (or other) form to a carbonate form by washing with about 20 column volumes of 0.1 mol Na₂CO₃. Nucleosides not possessing the additional acid dissociation in the alkaline region, such as adenylic and cytidyllic acids, are removed in this step. After a water wash pseudouridylic acid is eluted with 0.2 to 0.4 mol NH₄HCO₃ (20 to 40 column volumes); this has a pH of 7 to 8, significantly below the alkaline pK of pseudouridylic acid. From the NH₄HCO₃ solution, ammonium pseudouridylicate may be recovered by boiling to expel NH₄HCO₃ or the free acid may be recovered by removing the salt with a cation exchanger in the acid form, e.g. Dowex 50-H⁺ or Amberlite IR-120(H+). In either case the pseudouridylic acid is obtained in a water solution free of other nucleotides and salts.

c. Separation of Various Forms of Pseudouridylic Acid—Pseudouridylic acid exists in several forms, owing both to the conventional 2' and 3' phosphate isomerism encountered with all RNA nucleotides (the a and b forms) (4) and to a variety of forms assumed by the sugar moiety upon treatment with acid or alkali or after reduction. These forms may be detected and separated, with varying degrees of completeness, by ion exchange chromatography of the type originally used (16) for the separation of the isomeric uridine 2'- and 3'-phosphates as indicated in Fig. 2.

An original separation from an alkaline RNA hydrolysate is shown in Curve I of Fig. 2. The second and third peaks are presumed to be the 2'- and 3'-phosphate isomers by analogy with all the other pairs of RNA nucleotides and from the enzymic evidence summarized in Section VI (but note that the proportions are reversed; in all other cases, the a and b isomers occur in a 40:60 ratio (3, 4)). Both forms are dephosphorylated by prosthetic phosphatase to yield a pseudouridine (designated as the C nucleoside; see Section II) that has the properties of the expected ribofuranosyl derivative. The small peak that often precedes 2', labeled 2'a in Fig. 2, yields a different nucleoside, termed B, that has a different ultraviolet absorption spectrum. From the evidence of Fig. 2 (Curve II) it appears to be formed from 2' by the alkali used in the RNA degradation step. A different nucleoside, designated 3'b because of spectral and nucleoside identity with the 2'a of Fig. 2 (Curves I and II), appears to be formed when 3' pseudouridylate is treated with alkali (Fig. 2, Curve IV). If any of this nucleoside is present in the original alkaline hydrolysate, it appears under the 2' peak and is almost inseparable from it. A small degree of separation of 3'b and 2' can be achieved by paper chromatography in isobutylate buffer. The slightly altered spectral ratios of unpurified 2' with respect to 3', and the fact that all crude 2' preparations, after dephosphorylation, show a small amount of nucleoside B in the pseudouridine (C) that is formed, indicate that such contamination is the usual case.

Treatment of either 2' or 3' pseudouridylic acid, with which in other ribose phosphates leads to interconversion of each isomer to a mixture of both (4, 16, 17), gives a complex mixture of pseudouridylic acid isomers (Fig. 2, Curves III and V). Besides some nucleoside, owing to simple dephosphorylation, and the expected 2' and 3' nucleotides, there are formed nucleotides that appear to be, from the fact that each appears in more than one position, phosphate isomers of at least two additional nucleoside forms. One of the latter appears to be the B form mentioned previously (the nucleoside of the 2'b and 3'b nucleotides). The other new type, when dephosphorylated, yields nucleoside A₅ (described in Section II) that differs from B and from pseudouridine C in spectrum and in chromatographic behavior. (A fourth nucleoside form, A₇, becomes prominent only after prolonged or more drastic acid treatment. It is not a major constituent of the nucleotide chromatograms shown in Fig. 2).

II. Nucleoside Forms

a. Ion Exchange Behavior—The several nucleoside forms and derivatives of pseudouridine (C, B, A₅, A₇, H) are separable by an ion exchange procedure that is based in part upon the degree of complexing with borate exhibited by these forms.
The five nucleosides derived from pseudouridylic acid by acid, alkali, or H2 treatment are separable from them and from each other as shown in Fig. 3 (Curves I and II). (Those pairs, notably As and Ad, and pseudouridine and uridine, that separate poorly by ion exchange may be resolved by paper chromatography in isobutyrate buffer or butanol-H2O solvents. However, neither of these solvents separates the other substances in satisfactory fashion.)

It may be concluded, from the evidence summarized in Figs. 2 and 3, that the five nucleosides (pseudouridines A, As, H, B, and C) differ from each other in borate complexing ability. This is taken as a reflection of differences in sugar structure quite apart from changes in or adjacent to the pyrimidine ring that would be detectable by pK or ultraviolet absorption differences. From the observation that few, if any, of the substances listed in Fig. 3 are well separated by ion exchange in the absence of borate, and from the data on ionization presented in Section 11c, it is concluded that essentially no changes in pK or in number of ionizations take place in going from one form to another and, therefore, that the pyrimidine portion is not directly involved in these transformations.

The relation of the cis-glycol grouping of a ribofuranosyl moiety, as in uridine (and pseudouridine C), to position in the borate elution scheme is shown in Fig. 3 (Curves III and IV). Fig. 3 (Curve IV) shows the behavior of pseudouridine C once the cis-glycol grouping is destroyed by periodate oxidation; Peak I is presumably the expected dialdehyde, whereas Y and Z appear to be related products, possibly polymers, that have not been identified. Further degradation of Peak I of Fig. 3 by reduction and hydrolysis, as described later (Section IVb), yields the substance, Peak II, in Fig. 3 (Curve III) that occupies the same position as 5-CH2OH uracil and that, like it, shows no ability to complex with borate.

It is of interest that hydrogenated pseudouridine (H in Fig. 2, Curve II), with a spectrum practically identical with nucleosides B and As, appears between them and close to the bases. This is in keeping with the acyclic nature of the sugar that is postulated for the hydrogenated nucleoside (Section IVc).

b. Acid-catalyzed Interconversions—Treatment with strong acids (n HCl) at elevated temperatures (100°) produces extensive interconversion among the nucleosides. Since at least two nucleotides of each of the B and As nucleosides may be obtained from acid- (or alkali-) treated 2' or 3' pseudouridylates, it is concluded that the ribose moiety remains intact; none of these substances is a simple pyrimidine base or has suffered a loss of carbon.

The rates of appearance of the B, As, and Af forms from the pseudouridine C form are indicated in Fig. 4. Although Af (and 5-hydroxymethyluracil, or substances that closely resemble it) steadily accumulates throughout the hydrolysis period examined, it has been demonstrated (Table I) that an equilibrium or steady state exists among them; each, when heated in acid, gives rise to the other three, or at least to substances that are chromatographically and spectrophotometrically indistinguishable from them. This fact supports the previous conclusion that all four forms contain the same number of carbon (and nitrogen) atoms, and that their differences are attributable to changes in the ribose portion of the molecule.

c. Acid Dissociation Constants—Taking the spectrophotometric changes with pH as a reflection of ionizations within the chromophore (19, 20), it may be concluded that the principal
The ultraviolet absorption spectra of the four pseudouridines, the hydrogenated product, and the three methylated derivatives (the presumed 1- and 3-methyl, and 1,3-dimethyl), together with uridine and the three related pyrimidine bases for comparison, are given in Fig. 7. For each compound are given the spectra at pH 7, 12, and 14 (in NaOH, involving a 10% dilution with 10 N alkali that is not corrected for in the tracings); these pH values were chosen to give the various ionic species in as pure form as possible.

In other tests in which methylation was interrupted before completion, spots intermediate between pseudouridine and dimethyl pseudouridine were obtained. Each of these substances had only one of the two spectral shifts noted in the unmethylated material (their spectra are included in Fig. 7). From this fact and their intermediate chromatographic position, it is concluded that these represent the 1-methyl and 3-methyl derivatives of pseudouridine.

### III. Ultraviolet Absorption Spectra

The ultraviolet absorption spectra of the four pseudouridines, the hydrogenated product, and the three methylated derivatives (the presumed 1- and 3-methyl, and 1,3-dimethyl), together with uridine and the three related pyrimidine bases for comparison, as determined on a Cary automatic recording spectrophotometer, are given in Fig. 7. For each compound are given the spectra at pH 7, 12, and 14 (in NaOH, involving a 10% dilution with 10 N alkali that is not corrected for in the tracings); these pH values were chosen to give the various ionic species in as pure form as possible.

Certain differences in relative heights and locations of maxima (see Table II) that bear upon the structures are apparent in these curves. *(a)* Pseudouridine C, of all the substances recorded, most closely resembles 5-hydroxy methyluracil. *(b)* Derived pseudouridines B, A, and H resemble each other and

![Ultraviolet Absorption Spectra](image-url)

**Fig. 4. Acid catalyzed interconversions of pseudouridines.** Distribution of forms A, A, B, and C with time of heating at 100° in HCl, starting with pseudouridine C. At right side are shown distributions after treatment with 5 at H2SO4 for 1 hr, and after "short" and "long" treatments of the dried pseudouridine with anhydrous HF (by Dr. D. Lipkin at Washington University). About half of the material no longer appearing as A, B, or C nucleosides after long-term acid treatment is recovered in definite but unidentified peaks by ion exchange chromatography (carried out as in Fig. 5). The positions of these peaks, to the right of uridine, indicates either a polymeric or a more acidic structure without loss of the ultraviolet absorption characteristics of the pseudouridines.

**Table I**

<table>
<thead>
<tr>
<th>Starting material</th>
<th>Distribution of species (% of total)*</th>
<th>Recovery (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>73</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>6</td>
<td>74</td>
</tr>
<tr>
<td>B</td>
<td>16</td>
<td>33</td>
</tr>
<tr>
<td>C</td>
<td>14</td>
<td>32</td>
</tr>
</tbody>
</table>

*The nucleosides produced were separated and identified by ion exchange (Fig. 3) and spectrophotometric (Fig. 7) properties. In the case of the A-A pair, separation of the mixture from the column was achieved by paper chromatography.

also resemble uracil and thymine; H approaches thymine more closely, but the trend of all three towards thymine from the 5-hydroxymethyluracil type of spectrum shown by pseudouridine C possibly indicates a change in the side chain from the one type (—CH2OH) towards the other (—CH3). *(c)* Although As at first glance resembles 3-methyl pseudouridine, the spectral shift between pH 12 and 14 relates it to the unsubstituted pyrimidines and we may conclude that a change in the side chain that exaggerates the difference between —CH3 and —CH2OH has occurred. *(d)* The two derivatives that are apparently singly methylated have spectra of the type that would be predicted for singly methylated uracil compounds (see comparison data in Table II) and specific structures have been assigned on this basis. *(e)* The completely methylated pseudouridine shows no major shift of spectrum with pH and resembles dimethyluracil in this respect (Table II). A minor shift in the longer wave lengths at pH 14 (also noted in uridine and seemingly present in the other ribose derivatives shown in Fig. 7) may be ascribed to ionizations of the ribose hydroxyl at this high pH (21, 22).
Fig. 5. Variation of spectrophotometric absorbancy ratio, \( A_{260}/A_{250} \), with pH. The curves are for uracil (Ura), thymine (Thy), 5-hydroxymethyluracil ("5") and uridine (Urd). The points are for pseudouridines C (△), B (○), A\(_y\) (□), A\(_y\) (○), and for hydrogenated pseudouridine (H). The values indicated by the vertical arrows represent the midpoints (pK’s) of the curves in the region of 9.6. The presence of two pK’s for all substances, one near 9.6 and one above 13, and the similarities of these pK’s are clearly indicated.

IV. Chemical Properties

a. Elemental Analysis, Extinction Coefficient, and Acid Degradations—Analysis of various preparations of the nucleotide as the free acid, the cyclohexylamine salt, or the brucine salt (the salts were prepared by Dr. D. G. Doherty) indicate a C:N:P ratio in the nucleotide of 9:2:1. Analysis of pseudouridine C after recrystallization from alcohol (m.p. 220-221°) gave a C:N ratio of 9:2.

\[ \text{CH}_{12}\text{N}_{2}\text{O}_6 \]

Calculated: C 44.25 H 4.95 N 11.45

Found: C 43.75 H 5.25 N 11.30

The hydrogenated nucleoside (H) sinters at 189° and melts at 195-196°. The extinction coefficient of pseudouridylic acid, based on N or on P, or of pseudouridine, based on N, is 8600 at 260 m\(\mu\), pH 2 to 7. This is in agreement with the value found by Davis and Allen (11).

Degradation by HClO\(_4\) yields no ultraviolet absorbing material. Degradation by formic acid gives a variety of ultraviolet absorbing products, separable by ion exchange and paper chromatography, and all showing similar spectral properties. Their chromatographic behavior is considerably changed and indicates the formation of polymers or of acidic groups.

b. Periodate Oxidation—Pseudouridine C consumes periodate. The rate of the oxidation at 10-4 M, observed spectrophotometrically at 230 m\(\mu\) (23), and the extent of oxidation at any concentration are the same as for uridine. Ion exchange chromatography of the product gives a variety of substances of which one (Peak I in Fig. 3, Curve IV) appears to be the expected dialdehyde and the others (Y and Z), polymers or oxidation products of this. Reduction of Peak I with NaBH\(_4\) to form alcoholic groups in place of the aldehydes gave a small yield of a substance that behaved almost identically with 5-hydroxymethyluracil on ion exchange (Peak II in Fig. 3, Curve III) but that ran considerably slower in butanol-water paper chromatography (Column III, Fig. 8). This substance was again oxidized by periodate (1 mole was consumed on the basis of a pseudouridine extinction coefficient of 8600 at 260 m\(\mu\)) and the product, after reduction by NaBH\(_4\), behaved like 5-hydroxymethyluracil in two chromatographic systems (Column III, Fig. 8) and had spectra identical with those of 5-hydroxymethyluracil.

The course of the reaction, ignoring the large yield of side products and following only those of interest, is formulated as shown in Fig. 9. The exact course of the reactions following the production of the dialdehyde and the reasons for the low
FIG. 7. Ultraviolet absorption spectra of various pseudouridines and of uracil compounds of related structure. The spectra were automatically recorded, each group on a single sample that was not removed from the instrument when the pH was altered from 7 to 12, and from 12 to 14. The curves at pH 14 are therefore at a 9% greater dilution than those at 7 or 12, owing to the addition of 0.1 volume of 10 N NaOH to each sample. NaOH was used in the reference cell for the pH 14 curves, water for the others.

![Ultraviolet absorption spectra](image)

Substance*  | \( \lambda_{	ext{max}} \) at pH 7 (Å) | \( \Delta \) | \( \lambda_{	ext{max}} \) at pH 12 (Å) | \( \Delta \) | \( \lambda_{	ext{max}} \) at pH 14 (Å) | \( A_{260}/A_{320} \) at pH 7 | \( A_{260}/A_{320} \) at pH 12 | \( A_{260}/A_{320} \) at pH 14 | \( \lambda_{	ext{max}}/\lambda_{	ext{max}}' \) at pH 14
--- | --- | --- | --- | --- | --- | --- | --- | --- | --- | ---
Uracil | 258 | 0 | 283 | 25 | 275 | 0.17 | 1.42 | 1.39 | 0.74 | 1.00
5-CH2OH-uracil | 261 | 25 | 286 | 25 | 278 | 0.32 | 1.80 | 1.72 | 0.89 | 1.00
Thymine | 265 | 26 | 291 | 10 | 281 | 0.53 | 1.31 | 1.50 | 0.68 | 1.00
3-CH3-uracil† | 259 | 24 | 283 | 0 | 283 | 0.14 | 3.50 | 3.50 | 1.47 | 1.00
ψ uridine C | 262 | 24 | 286 | 25 | 278 | 0.40 | 1.96 | 1.93 | 0.96 | 1.00
ψ uridine A₅ | 262 | 24 | 286 | 25 | 281 | 0.36 | 2.50 | 2.04 | 1.08 | 1.00
ψ uridine A₇ | 263 | 22 | 285 | 7 | 278 | 0.50 | 1.37 | 1.64 | 0.74 | 1.00
ψ uridine B | 263 | 24 | 287 | 9 | 278 | 0.48 | 1.46 | 1.73 | 0.75 | 1.00
ψ uridine-H₂ | 264 | 25 | 287 | 7 | 282 | 0.57 | 1.53 | 1.56 | 0.74 | 1.00
3-CH₂ψ uridine† | 264 | 19 | 283 | 2 | 285 | 0.92 | 2.23 | 2.25 | 1.27 | 1.00
Uridine | 262 | 0 | 262 | 0 | 263 | 0.35 | 0.29 | 0.41 | 0.76 | 1.00
5-CH2OH-uridine | 263 | 0 | 262 | 0 | 262 | 0.35 | 0.50 | 0.73 | 1.00 | 1.00
Thymidine | 263 | 0 | 267 | 0 | 267 | 0.35 | 0.66 | 0.75 | 1.00 | 1.00
1-CH₃-uracil† | 268 | -3 | 265 | 0 | 265 | 0.69 | 0.40 | 0.40 | 0.72 | 1.00
1,3-diCH₃-uracil† | 266 | 0 | 266 | 0 | 266 | 0.62 | 0.62 | 0.62 | 1.00 | 1.00
1-CH₃ψ uridine† | 265 | +4 | 269 | +3 | 272 | 0.66 | 0.80 | 1.05 | 0.78 | 1.00
1,3-diCH₃ψ uridine | 269 | +1 | 270 | +2 | 272 | 0.94 | 0.94 | 1.00 | 0.90 | 1.00

*ψ = pseudo.
† Data from Fox and Shugar (21).
‡ Structure tentative.

FIG. 8. Ultraviolet absorption print of descending paper chromatogram of pseudouridines, partially degraded pseudouridines (see Figs. 3 and 9), and known substances, run approx. 16 hours in butanol-water (86:14) on Whatman 3MM paper. Columns II and III are substances obtained from the periodate oxidation sequence (Fig. 9); II also appears in Fig. 3 (Peak II).
Fig. 9. Proposed scheme of degradation of pseudouridine C to 5-hydroxymethyluracil. Substances I and II are the same as those indicated in Figs. 3 and 8.

Fig. 10. Spectra of colors produced in the orcinol reaction, after 1 hour of heating, by pseudouridine and other compounds. The final concentrations were: ribose, 0.083 mM; pseudouridine, 0.1 mM; uridine, 0.10 mM; ribosylthymine, 0.10 mM.

Fig. 11. Nuclear magnetic resonance spectra, at 40 mc frequency, of pseudouridine C, 5-methyluridine (ribosylthymine), thymine, and 6-methyluracil, normalized to the HDO peak (artifact). Room temperature. (Spectra obtained in Cambridge, England, by Drs. N. Sheppard and D. M. Brown on material prepared by W. E. Cohn.)

yield of Substance II (Fig. 9) and of the final product are under investigation. In this connection, it may be noted that small yields of substances that appear to be 5-hydroxymethyluracil are found in acid hydrolysates of long duration.

The B nucleoside also consumes 1 mole of periodate, but each A form consumes 2 moles, producing one mole of acid but no formaldehyde. This is consistent with a pyranose structure.

c. Hydrogenation—The hydrogenation of pseudouridylic acid (24) with the uptake of 1 mole of H₂ per mole of nucleotide gave the variety of nucleotides shown in Fig. 2, Curve VI. Each of these was dephosphorylated to give a nucleoside (designated as H) that was also formed by the direct hydrogenation of pseudouridine C. The spectrum of the II nucleoside, or of the three nucleotides of Fig. 2, Curve VI that contain it, resembles the spectrum of B (or A₁) but not that of pseudouridine C, or A₂ (see Section III). The retention of the ultraviolet spectrum indicates that there is no hydrogenation in the pyrimidine ring.

The shift in spectrum toward that of thymine suggests that the reaction is analogous to that of allyl and benzyl ethers, i.e. —C — C — OR + H₂ → —C — C — H + ROH, and of 5-CH₂OH uracil → 5-CH₃ uracil (25). This behavior is consistent with the formulation of pseudouridine as 5-ribosyluracil and of the reduced product, H, as 5-ribityluracil. It consumes 3 moles of periodate and liberates 2 moles of acid and 1 mole of formaldehyde, as predicted from this structure.

d. Orcinol Reaction—Inasmuch as pseudouridine appears to be a pyrimidine nucleoside with a ribosyl linkage more stable than the N₁-Cl linkage of uridine, a positive reaction in the conventional orcinol procedure (25a) is not to be expected. However, uridine (and even cytidine, to a lesser degree) gives a partial reaction on prolonged treatment with the orcinol reagent, owing to progressive hydrolysis of the ribosyl linkage. A quantitative reaction can be obtained from uridine in the normal 20-minute reaction time if the ribosyl linkage is weakened by destruction of the double bond by hydrogenation (24), ultraviolet irradiation, or bromination, especially if followed by alkali (26) to split the pyrimidine ring. These treatments have been ineffective on pseudouridine.

Prolonged treatment of pseudouridine with orcinol gives only a trace of the blue-green color (λmax = 665 mp) given by ribose; the principal color is a red-brown (λmax = 515 to 535 mp, depending on the length of heating) that appears in about 20 minutes and deepens progressively up to about 2 hours (Fig. 10). This color reaction is independent of the form of nucleoside or nucleotide of pseudouridine, and from the behavior of the isomeric forms in acid (Section IIb) it is suggested that the reactive form is A₁. The extinction coefficient at 555 mp is high enough (7500) to make this a reaction suitably sensitive for analytical purposes. It should be noted that reduced pseudouridine does not give any color at all.

An absorption maximum at 535 to 550 mp in the orcinol reaction is given by neuraminic acid and certain keto sugars (27), by unsubstituted ketopentoses (28), and by malonaldehyde and related substances when coupled with thiobarbituric acid (29, 30) or similar substances. The nature of the chromophores giving the red color is still not certain, but it is considered that cyclizations occur to give pyrrole or furfural structures (31). In the present case, we consider C₄ and its attached O atom to be of critical importance in the chromophore because of...
the fact that when this is reduced by hydrogenation to \(-\text{CH}_2\) (Section IVc), no color is formed in the orcinol reaction.

V. Nuclear Magnetic Resonance Spectra

From the nuclear magnetic resonance spectra, much information on the nature of the various structures has been obtained. In Fig. 11, a comparison of 5-methyl- and 6-methyluracil locates the positions (at 40 mc) of the C_4 and C_5 protons of uracil at \(-115\) and \(-30\) cycles per second, respectively, whereas pseudouridine shows only the C_4 peak. Comparison of 5-methyluracil (thymine) and its 1-ribofuranosyl derivative (ribothymidine) indicates the position of the stable protons attached to C_1', C_2', C_3', C_4', and C_5' atoms. With the exception of C_1', each peak is reflected in the pseudouridine trace.

The seeming absence of a C_1' peak in pseudouridine C, the presumed ribofuranosyl form, gave rise to the surmise that it was hidden behind the HDO artifact. It is known that signals from protons attached as in C-CH normally lie in a higher (more positive) position than those attached N-CH (as in uridine). When the spectrum of pseudouridine was observed at progressively higher temperatures (Fig. 12), a process that moves the HDO peak positively relative to the other peaks, both members of the missing C_1'H peak were observed. Also to be noted are the several C_5', C_6', and C_7 proton peaks and the C_6'H doublet, all of which are identical with those observed in the spectra of uridine and uridylic acid.

A direct comparison of uridine and 1-methyluracil is shown in Fig. 13. The mirror-image doublets of the C_4 and C_5 protons are clearly seen in the 1-CH_3-uracil trace. Uridine adds to these a doublet for the C_1'H (lying within the C_5'H doublet) and substitutes the ribosyl complex for the CH_3. The similarity (or identity) between the ribosyl moieties of uridine and pseudouridine, with the exception of the C_1'H, is clearly shown (here the HDO peak is shifted to the left by cooling in order to display the C_1'H peaks).

The spectra obtained on specimens of the B, mixed A, and purified pseudouridine are shown in the top portion of Fig. 13 and indicate certain structural relationships and differences among these substances (a spectrum of AF could not be obtained because of technical difficulties). All show similar C_5'H and C_6'H peaks. In B the latter is slightly split (as are those of thymidine and ribothymidine), suggesting an influence upon the C_5'H of B similar to those that exist in the latter substances with methyl groups at C_5. The peaks for the C_5', C_6', and C_7 protons (the latter well separated) of B and the small spin coupling and more positive position noted in the C_6'H of B suggest an equatorial conformation for it or for the C_6'H proton.

Comparing the spectra of the mixed A nucleosides with that of relatively pure AS, the broadening in the C_5'H peak of the former may indicate a tendency towards splitting, as in the case in nucleoside B. Since no such tendency is noted in the C_5'H peak of AS, we may tentatively ascribe the slight splitting in the AS spectrum to the C_5'H of an AF impurity and conclude, in accordance with the ultraviolet absorption evidence, that AF is related to B (and to H_5-pseudouridine) and has a more thymine-like
structure with $C_1'$ being in a more positive environment. The smaller spin coupling observed between $C_2'$ and $C_3'$ protons of $A_3$ indicates an equatorial conformation for the $C_3'$-H. The change in relative height of the $C_1'$-H peaks (just to the right of the HDO peak) may indicate that this is totally absent in pure $A_3$ (or that there exists an $A_3$ without $C_1'$ protons).

Further correlations are not presently possible inasmuch as the positions occupied by the various protons may shift relative to each other as their immediate chemical environment is changed. Thus the two $C_2'$ protons in deoxythymidine or deoxouridine appear at $+143$ cycles per second, whereas the single one of ribothymidine lies in the region of $+20$. In the $A_3$ trace, it is considered that the $C_2'$-H and $C_4'$-H complex groups are reversed in order; although this movement of $C_2'$-H of $A_3$ in the positive direction may be ascribed to a similarity to thymidine in structure, there is as yet insufficient evidence for this interpretation. However, it is clear that (a) all forms contain a $C_6$H but no $C_5$H, (b) the $C_1'$-H is in a more positive environment than in uridine, (c) the rest of the ribose moiety of pseudouridine $C$ is essentially identical with that of uridine, and (d) the differences between the various forms of pseudouridine involve conformational or chemical changes in the ribose component only.

VI. Enzyme Reactions

Pseudouridylate is released from RNA by ribonuclease, both as the free $3'$ nucleotide (b, according to Davis and Allen) and in the polynucleotide fraction, indicating that it appears adjacent to both purine and pyrimidine nucleotides in the chain and that it is a pyrimidine $3'$ nucleotide as far as susceptibility to ribonuclease is concerned.

All pseudouridylates are dephosphorylated by prostatic phosphatase; the $2'$, $3'$, and $5'$ acids yield pseudouridine $C$, the $2'$- and $3'$- acids yield nucleoside $B$, and others yield nucleoside $A_3$.

Hydrolysis by total snake venom yields pseudouridine $C$ from RNA. Snake venom diesterase liberates a nucleotide (from the RNA of Escherichia coli infected with bacteriophage T7) that, in terms of spectrum and chromatographic behavior, appears to be pseudouridine $5'$-phosphate; this is dephosphorylated by venom $5'$-nucleotidase (whereas $2'$-, $3'$-, $3'$- $A$, and $3'$- $B$, and three other nucleotide forms derived from $2'$ and $3'$ are not) to pseudouridine $C$. A barley nucleotidase that hydrolyzes only adenosine $3'$-phosphate and not adenosine $2'$-phosphate acts solely upon $3'$-pseudouridylate (not upon $2'$-, $2'$- $A$, or the derived forms) to yield pseudouridine $C$. From these facts, we conclude that the third pseudouridylate acid in Fig. 2 (Curve I) is the $3'$-phosphate, that the second is therefore the $2'$, that a $5'$-phosphate exists, that the mode of production of all three indicates the presence of the three hydroxyls of a normal nucleotide, and that RNA contains pseudouridylate in a $3'$ or $2'$-$3'$ linkage, with $3'$-$5'$ indicated by the ribonuclease evidence.

Three nucleosidases (from yeast (32), E. coli (33), and Lactobacillus arabinosus (34)) that hydrolyzed uridine to uracil and ribose failed to give a significant amount of material reacting as ribose in the orcinol reaction when incubated under comparable conditions with pseudouridylate.

DISCUSSION

The existence of a $C-C$ link between a pyrimidine base and the ribosyl radical requires a reconsideration of methods of analysis and detection that have been built around the properties of N-C nucleosides. Pseudouridine does not yield uracil or significant quantities of 5-hydroxymethyluracil upon strong hydrolytic treatment to yield bases from nucleic acids (e.g. HClO$_4$ or formic acid hydrolysis tends to destroy it or convert it into a variety of as yet unidentified products). It does not give a conventional reaction with orcinol either before or after bromination, hydroxylamine, or other treatment that will weaken even the strongest N-C linked pyrimidine nucleosides. Upon hydrolysis to nucleoside or nucleotide from the parent nucleic acid, it is not easily separable from uridine or uridylic acid, although its detection in a mixture is made more certain by virtue of the bathochromic shift in alkali, a property not shared by other nucleotides.

(However, this property, usually associated with the bases, may itself be misleading because of the possibility of base contamination in one or another chromatographic system.)

The absence of substitution on either nitrogen is shown quite clearly by ultraviolet absorption spectra that are very similar to 5-hydroxymethyluracil and reflect the two ionizations and pK's characteristic of uracil and its derivatives, and by the ability to form two monomethylated and one double methylated derivatives with consistent spectral changes. The appearance of 5-hydroxymethyluracil in one or another degradation supports the assignment of uracil as the base component, as does the $C_4$N$_2$ empirical formula and the absence of an amino group.

The proof that ribose is the sugar and that it is in the furanosyl form is also indirect. The presence of $2'$ and $3'$ pseudouridylates on alkaline hydrolysis, the appearance of only the $3'$ on ribonuclease hydrolysis (11), the appearance of a $5'$ pseudouridylate from diesterase hydrolysates, and of pseudouridine in total snake venom hydrolysates indicate the presence of the usual $2'$, $3'$, and $5'$ hydroxy groups (although the 40:60 ratio of $2'$ and $3'$ found for other nucleotides in alkaline hydrolysates is reversed here, indicating some new influence upon the cyclic intermediate) and a $3'$-$5'$ diester linkage. The nucleoside forms a complex with borate and reacts with periodate to exactly the same degree and at the same rate as uridine, consistent with the cis-glycol configuration of a furanosyl structure.

However, it is the nuclear magnetic resonance behavior that most clearly indicates all the essential features of pseudouridine. In the pyrimidine portion, the presence of a proton at $C_1'$ and the absence of one at $C_5$ is clearly revealed. An identity between the ribose moieties of pseudouridine, uridine, and ribothymidine is shown by the identity of the respective spectra in the $C_2'$, $C_4'$, $C_5'$, and $C_6'$ proton region, in contrast to the large divergence noted in the deoxyribosyl derivatives and in acid-treated pseudouridines. The ability of nuclear magnetic resonance to distinguish axial from equatorial conformations again supports the ribofuranosyl hypothesis for pseudouridine. Attachment of the methyl group (or of any methylene group, except for $C_5'$) is also clearly shown.

The presence of a

$$\text{OR}$$

$\text{C} - \text{C} - \text{C}$

grouping serves to explain the uptake of one mole of hydrogen without saturation of the double bond (which in most pyrimidines results in loss of the characteristic ultraviolet absorption (18)).
since allyl and benzyl ethers (and 5-CH2OH uracil (20)) tend to undergo hydrolysis (in our case, rupture of the furanose bridge) to form $-C\equiv C\equiv CH - + HO-OR$ rather than to saturate the double bond.

The exact nature of the changes that occur in acid to form the B, A$_R$, and A$_F$ forms of the nucleoside are not yet to be specified. From the spectral changes alone, one might suppose that the B, A$_R$, and hydrogenated forms all possess a $-CH_2-C$ or $-CHOH-C$ structure at C$_7$ and C$_9$ rather than

$$\begin{align*}
\text{O} & \quad \text{C} \\
\text{C} & \quad \text{CH}
\end{align*}$$

(that is, the furanose bridge is ruptured, perhaps followed by dehydration between C$_1$ and C$_2$). On the basis of the ultraviolet and nuclear resonance spectra, it is tempting to postulate a $-CO-CH_2-$ dehydrate form for A$_F$. However, B consumes but 1 mole of periodate verse 2 moles for A$_R$ and A$_F$ and no formaldehyde is liberated by any, which is difficult to reconcile with an acyclic structure. The several forms may represent conformational changes only, without permanent ring rupture. From spectra and interconvertibility considerations, it seems certain that these forms are all relatively simple rearrangements of the sugar moiety, observable in this case because of the stability of the C$_7$-C$_9$ linkage and possibly because of the influence of the neighboring C$_7$-C$_9$ double bond and the $\beta$ carbonyl group at C$_6$.

Perhaps it is because of the directing influences of these groups that the pseudouridine C form reappears on acid treatment of the B, A$_R$, and A$_F$ forms, for one might expect a trans configuration owing to repulsion between C$_7$ and C$_9$ hydroxyl groups. Yet it must be recalled that only the pseudouridine C form is obtained in a way that lends confidence to considering it as a single structure; B, A$_R$, and A$_F$ could be mixtures having identical chromatographic properties. For this reason, conjectures as to their structures must be reserved.

Regardless of the remaining uncertainties, however, we are here faced with the natural occurrence of a new type of nucleoside structure, one that does not lend itself to the established routes of biosynthesis or of detection and analysis. In one way it follows in the pattern noted for the methylated purine nucleosides, of being relatively enriched in the so-called "soluble" or "supernatant" RNA's (35, 36), and it is reported to appear in urine (37). These observations do not offer any clue as to biosynthesis or function. With regard to the former, it may be noted that although C-C glycosyls are not unknown in Nature, they are not of common occurrence and have not heretofore been detected in nucleic acids. With regard to the latter, this substance may be unique for (alone among the commoner nucleic acid bases) only uracil can be so bonded and yet, while within a polynucleotide chain, can present a $-CO-HN-$ group in the proper position for interchain hydrogen bonding. This unique feature, however, does not necessarily exclude from consideration the possibility that other nucleosides of this type exist in nucleic acids.

The incorporation of uracil into the pseudouridine fraction of a uracil-requiring mutant has been observed by J. D. Smith (personal communication).

M. Uziel, personal communication.

W. S. Adams (personal communication) and W. E. Cohn, unpublished results.

**SUMMARY**

A nucleoside isomeric with uridine, and therefore termed pseudouridine, that appears to be 5-ribosyluracil exists in ribonucleic acids. It has the double bond. A nucleoside isomeric with uridine, and therefore termed pseudouridine, that appears to be 5-ribosyluracil exists in ribonucleic acids. It has the double bond.

REFERENCES

3. **Cohn, W. E., Federation Proc.,** 16, 166 (1957); 17, 203 (1958).
7. **Cohn, W. E., J. Am. Chem. Soc.,** 73, 1539 (1951); also (16).
20. **Shugar, D., and Fox, J. J., Biochim. et Biophys. Acta, 9, 199 (1959).**
22. **Fox, J. J., and Shugar, D., Biochim. et Biophys. Acta, 9, 309 (1959).**
Pseudouridine, a Carbon-Carbon Linked Ribonucleoside in Ribonucleic Acids: Isolation, Structure, and Chemical Characteristics
Waldo E. Cohn


Access the most updated version of this article at http://www.jbc.org/content/235/5/1488.citation

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/235/5/1488.citation.full.html#ref-list-1