JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Deuterium Studies Reveal a New Mechanism for the Formose Reaction Involving Hydride Shifts

Chandrakumar Appayee and Ronald Breslow*

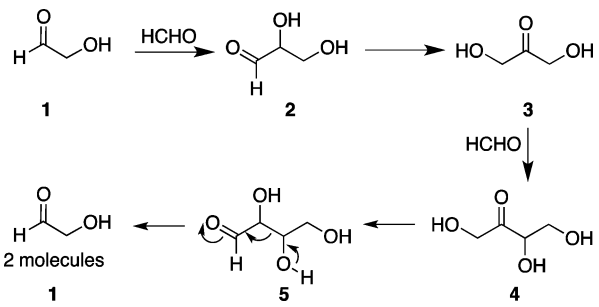
Department of Chemistry, Columbia University, New York, New York 10027, United States

Supporting Information

ABSTRACT: In the formose reaction, formaldehyde is converted to glycolaldehyde, its dimer, under credible prebiotic conditions. Breslow proposed a mechanism for the process in 1959, but recent studies by Benner showed that it was wrong in detail. Our present studies clarify the mechanism, which involves the original Breslow intermediates but some different connecting steps.

The formose reaction. Simple sugars were probably created from formaldehyde on prebiotic earth; most are formally formaldehyde polymers, $(CH_2O)_n$. In 1861, Butlerow first noticed the formation of sugar-like substances when formaldehyde solutions were treated with base,[1] a finding that led to the extensive investigation of this reaction, known as the "formose" reaction. In 1959, Breslow[2] proposed a mechanism for this process based on his and literature evidence[3−5] (Scheme 1). In this mechanism, glycolaldehyde (**1**), the dimer of formaldehyde, is created from formaldehyde by an autocatalytic cycle.

Scheme 1. The Original Proposed Sequence for the Formose Reaction[2]



The reaction is initially extremely slow until the first molecule of **1** is formed by an undetermined mechanism that may well involve radiation. Then the autocatalytic cycle runs rapidly to produce more glycolaldehyde, which can also react further to form trioses, tetroses, pentoses, etc. Some of the proposed intermediates can also be diverted in part from this cycle. The rapid formose process can also be initiated by added glycolaldehyde or by analogous ketols such as hydroxyacetylbenzene.[2] In the discussion, Breslow assumed that the conversions of **2** to **3** and **4** to **5** involved enolization to give an enediol and reprotonation at a different carbon.[2] The conversion of **1** to **2** was proposed to involve formaldehyde addition to 1,2-dihydroxyethene (**1A**) (Figure 1), while the conversion of **3** to **4** was proposed to involve addition of

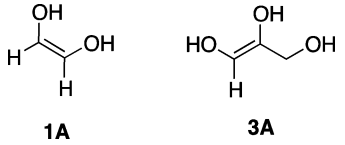


Figure 1. Proposed enediol intermediates in the formose reaction.

formaldehyde to 1,2,3-trihydroxypropene (**3A**). In all of these cases, the enediols were complexed with the Ca^{2+} that is present as a catalyst.

In an important study, Benner[6] examined the reaction more recently in D_2O solvent at 65 °C with high (350 mM) formaldehyde concentrations and found that deuterium is not incorporated into the products until the formaldehyde is largely consumed. He pointed out that his work excluded the Breslow mechanism as originally formulated, with protonation of intermediate enols that would have incorporated deuterium in his intermediates.[6,7] In particular, Benner proposed that glyceraldehyde (**2**) enolizes irreversibly since enediol **3A** would be trapped by formaldehyde addition before it can be protonated to form dihydroxyacetone (**3**). He proposed a mechanism in which **3** is not an intermediate in the process. He did not detect it under his conditions, and if it had been formed from glyceraldehyde by enolization, it would have incorporated deuterium in D_2O solvent.[7] He proposed a mechanism for the formose reaction with some new intermediates, detected by mass spectrometry, that had more than the four carbons of structures **4** and **5**.

We have now examined all of the steps of the process using deuterium labels to elucidate their mechanisms. We prepared pure formaldehyde solution from paraformaldehyde. We performed the reaction at pH 12 and 40 °C with catalysis by Ca^{2+}, so the hydroxyaldehydes and hydroxyketones would be bound to the calcium. Intermediates in the reactions would also be bound to calcium. We stopped the reactions after 10−20 min.

The steps in which formaldehyde is added must involve calcium complexes of enediols—**1A** for the conversion of **1** to **2** and **3A** for the conversion of **3** to **4**—and these cannot incorporate deuterium onto carbons unless the first enolizations are reversible. We see that the enols are rapidly captured by 350

Received: October 24, 2013
Published: February 27, 2014

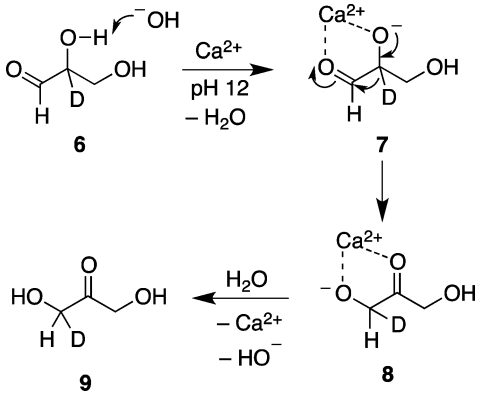
   dx.doi.org/10.1021/ja410886c | J. Am. Chem. Soc. 2014, 136, 3720−3723

mM formaldehyde (Benner's conditions), so they do not reverse to add deuterium from D_2O. However, our studies described below show that deuterium placed in intermediates can be partially washed out in H_2O at lower formaldehyde concentrations through reversible enolization. The isomerizations of **2** to **3** and **4** to **5** would involve deuteration of a carbon in the products if they involved isomerization via the enediols, but we have now shown that they do not use that mechanism. Instead, they use the well-studied[8] hydride shift mechanism that does not incorporate deuterium from the D_2O solvent.

The formation of dihydroxyacetone (3). One of the critical differences between Scheme 1 and a scheme proposed by Benner[7] is whether **3** is formed and utilized during the reaction. Benner reported that the formose reaction with 350 mM formaldehyde shows no detectable **3**, and we confirmed that under his conditions **3** is not present in NMR-detectable amount. However, when we performed the same reaction with only 70 mM formaldehyde and 10 mM glyceraldehyde, we saw **3** in 8% conversion from glyceraldehyde after 10 min. With 35 mM formaldehyde and 10 mM glyceraldehyde, a 4% yield of **3** was found after 20 min. A considerable amount of unreacted formaldehyde (two-thirds at 70 mM formaldehyde and one-third at 35 mM formaldehyde) was still present in the reaction mixture during those reaction times, as well as the products of further reaction by 3. The low percentage of isolated **3** reflects its rapid reaction with formaldehyde, which is apparently even faster under Benner's conditions. We also confirmed the formation of dihydroxyacetone when glycolaldehyde was added instead of glyceraldehyde under the above reaction conditions (see the Supporting Information).

The Delidovich group recently reported the formose reaction at a formaldehyde concentration of 100 mM at various temperatures and observed the formation of **3** under all reaction conditions.[9] Many others have also seen the formation of **3** under formose conditions.[10−14] We now see that deuterium from the D_2O solvent is not incorporated into the intermediates, since **3** and **5** are formed by hydride shift mechanisms (Scheme 2), not by enolization. Although this scheme shows a simple migration of the deuterium, some have studied the possibility that proton tunneling occurs over this short distance.[8] If so, tunneling by deuterium would be less likely.

Scheme 2. Isomerization of 2-Deuteroglyceraldehyde (6) to 1-Deuterodihydroxyacetone (9) through a 1,2-Hydride Shift under Formose Reaction Conditions



The strongest evidence comes from the reaction of 2-deuteroglyceraldehyde (**6**) under our formose conditions in H_2O. Commercial **6** (100% deuterated at C-2 on the basis of ^1H NMR and mass spectroscopic evidence) was converted to 1-deuterodihydroxyacetone (**9**) with 74% of the deuterium still retained. Some was lost in the enolization of **9**, which is partially reversible under our conditions in competition with enol capture by formaldehyde. With the higher formaldehyde concentration and higher temperature of Benner's conditions, capture of the enol by formaldehyde would be essentially complete, so his intermediate **3** could have an undetectible concentration and would not incorporate deuterium from solvent D_2O to pass on to **4**. In our studies, all of the carbohydrates were converted to 2-nitrophenylhydrazones and analyzed by HPLC, ^1H NMR spectroscopy, and mass spectrometry.

When we simply converted **2** to **3** in D_2O with calcium ion and formaldehyde at pH 12, in 10 min the process yielded **3** with 23% of the molecules carrying a deuterium. That deuterium incorporation must reflect a side reaction (reversal of the enolization of **3** after its formation), but most of the enolized dihydroxyacetone was rapidly trapped by formaldehyde, forming erythrulose (**4**). When **3** was left for the same time in the D_2O medium containing calcium ion at pH 12 but without formaldehyde, **3** was extensively polydeuterated on carbon. In the absence of formaldehyde, the enol cannot be trapped in competition with deuteration.

The deuterium would have been completely lost by enolization of 2-deuteroglyceraldehyde **6**, but not in the hydride shift mechanism, where it migrates to C-1. With compound **6** at pH 12 with calcium ion, the coordinated hydroxyl would be deprotonated to a coordinate hydroxide ion to afford **7**, as shown in Scheme 2, and **7** would not lose deuterium from C-2 but would simply undergo the hydride shift to form **8**. The loss of 26% of the deuterium in the conversion of **6** to **9** could possibly reflect that 26% of the product is formed by the direct conversion of **6** to the enediol and then to dihydroxyacetone **3**, while the other 74% comes from the hydride shift process. More likely, the hydride shift is the only process that occurs, and deuterium could be partially washed out of 1-deuterodihydroxyacetone **9** by enolization and reprotonation. At higher formaldehyde concentrations, reprotonation would be suppressed as a competing reaction.

It is at first a curious fact that we see the reaction of formaldehyde with dihydroxyacetone **3** to be ca. 6-fold faster than is the reaction of formaldehyde with **2** under the same conditions (from the yields of **4** with **2** and **3**). Normally aldehydes are more rapidly enolized than their ketone isomers, but in water the carbonyl group of aldehyde **2** is extensively hydrated; the proton NMR spectrum of **2** in D_2O shows less than 1% of the aldehyde, with the rest a geminal diol. The keto group of **3** is 80% a ketone and 20% a geminal diol in D_2O (as observed by ^{13}C NMR spectroscopy). Thus, the fastest route to the enediol 1,2,3-trihydroxypropene is not by enolization of **2** but instead by its more rapid conversion to **3** by hydride shift followed by the rapid enolization of **3**. Why is the hydride shift mechanism faster than the enolization of **2**? The calcium can form the complex of the aldehyde form of **2**, either trapping the small fraction that is not a hydrate or binding first to the aldehyde hydrate of **2** and then undergoing the hydride shift as soon as the complex loses water to temporarily form the aldehyde complex. The hydride shift is intramolecular, possibly

dx.doi.org/10.1021/ja410886c | J. Am. Chem. Soc. 2014, 136, 3720−3723

even with a tunneling mechanism, while enolization waits for attack by an external hydroxide ion.

We compared the reactions of glyceraldehyde with calcium ion at pH 12 and 40 °C with and without formaldehyde. Without formaldehyde, after 10 min, almost all of the glyceraldehyde was consumed, while with added 70 mM formaldehyde, 36% of unreacted **2** was observed. How did formaldehyde protect glyceraldehyde from a side reaction, and what was that side reaction? The clue comes from the report[15] that a 1:1 mixture of **2** and **3** under even milder conditions (23 °C) than ours has been found to undergo rapid conversion to fructose and its stereoisomers when the enolate of **3** adds to the carbonyl of **2**. This side reaction is a direct analogue of the process by which fructose is formed biochemically.[16] When formaldehyde is present, it captures the enolate, so the glyceraldehyde is spared.

The formation of erythrulose (4). The aldol addition of formaldehyde to **3** forms **4**, a racemic mixture of D- and L-erythrulose. Starting with **3,** formaldehyde, and calcium ion at pH 12, we observed the formation of erythrulose in 5% yield after 20 min along with higher sugars and recovered **3** (5%). When glycolaldehyde or glyceraldehyde was used instead of dihydroxyacetone under the above reaction conditions, the yield of erythrulose was 3% or 11%, respectively.

When 2-deuteroglyceraldehyde **6** was used under formose reaction conditions in water, we observed 33% deuterium retention in the product erythrulose hydrazones that could come in part from 1-deuterodihydroxyacetone **9** (74% of deuterium retained), the product formed from **6** through the 1,2-hydride shift, and later from some reversible enolization of **4**.

Tautomerization of erythrulose to aldotetroses. When erythrulose was treated with $Ca(OH)_2$ at 50 °C for 60 min, we observed the formation of aldotetroses **5** (69% of the new products formed) and glycolaldehyde (1%) along with unreacted erythrulose (84%) in the crude reaction mixture, which were identified as 2-nitrophenylhydrazones. Again, when we performed this reaction in D_2O we saw no incorporation of deuterium onto the carbons of **5**, showing that it uses the hydride shift mechanism. The formation of glycolaldehyde in this reaction further confirms the retroaldol cleavage of aldotetrose (below) as described in the mechanism.

Retroaldol cleavage of aldotetroses 5 to glycolaldehyde. To explore another important step in the proposed mechanism of Scheme 1, the retroaldol cleavage of aldotetroses **5** to form two molecules of glycolaldehyde **1**, we treated an equimolar 4 mM mixture of erythrose and threose with 0.6 mM $Ca(OH)_2$ in water at 65 °C for 60 min. After converting the reaction mixture into the corresponding 2-nitrophenylhydrazones, we analyzed the solution by ^{1}H NMR spectroscopy and HPLC and observed the formation of glycolaldehyde (9% of the new products), erythrulose (47%), and unreacted erythrose and threose (45%). Of course, this reaction does not produce two molecules of glycolaldehyde **1** directly; rather, it produces one molecule of **1** and one molecule of the enediol **1A**. This could be protonated to make **1**, but it would have deuterium from the D_2O solvent unless it reacted with formaldehyde to form glyceraldehyde **2**. This would simply move the reaction further along the normal pathway.

If every molecule of **1** were doubled on every cycle, the only product would be glycolaldehyde, but in real life the full cycle would not run perfectly for every molecule of **1**. Other products would be isolated, such as **2**−**5** or compounds branching from them. The observed retroaldol reaction is just the critical part of the minimal pathway for an autocatalytic cycle to amplify the concentration of glycolaldehyde or its enediol **1A**. Such amplification is needed to explain how a tiny amount of added **1** can generate significant amounts of its derived products.

Benner's mechanistic proposals. Benner[6,7] was struck by the fact that no deuterium was incorporated by any of the intermediates in the process when he carried the reaction out in D_2O with very high concentrations of formaldehyde at 65 °C. He therefore proposed that the isomerizations of **2** to **3** and **4** to **5** did not occur, since he assumed that they must have involved enolizations followed by protonations. He proposed that the enols that could have converted **2** to **3** and **4** to **5** were instead caught by formaldehyde, leading eventually to a proposed pentose instead of compound **5**.

We have found no evidence for such a compound at our lower formaldehyde concentrations or in our studies with the higher formaldehyde concentrations used by Benner. It was not found in the studies by Delidovich et al.,[9] in which all of the intermediates of Scheme 1 were identified. Most seriously, Benner reported that he could not find his proposed pentose unless he included borate in the catalyst system.[7] The formose reaction does not require borate to carry out its autocatalytic cyclic process.[1,3,5] Now we have shown that the isomerizations involve hydride shifts rather than enolizations, so their failure to incorporate deuterium in D_2O is explained without the need for the intermediates Benner proposed.

Conclusion. This new mechanism is consistent with the evidence obtained by Benner and others and our new evidence using specifically deuterated glyceraldehyde or D_2O studies. It is interesting that it involves the same four intermediates originally proposed by Breslow 55 years ago, but the details of the transformations are different. This seems to be the mechanism with the fewest intermediates that explains the autocatalysis in the formose reaction, but other mechanisms are in principle also possible. However, they must account for all of the findings in this and other work.

While the formose reaction was an initial inspiration for chemists to imitate possible prebiotic reactions, it forms racemic and diastereomeric mixtures while the prebiotic world probably needed single enantiomers such as in our current D sugars. Elsewhere we described how a process related to the formose reaction but with L-amino acids as catalysts, forming enamines at low pH, can afford glyceraldehyde with an excess of the D enantiomer.[17−19] Furthermore, we showed how modest D excesses can be amplified to high D/L ratios by simple aqueous processes.[17−19] These more recent studies, inspired by the formose reaction, are better models of the likely prebiotic syntheses of carbohydrates larger than glycolaldehyde.

■ ASSOCIATED CONTENT

Supporting Information

Experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

rb33@columbia.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support of this work by NASA is gratefully acknowledged.

■ REFERENCES

(1) Butlerow, A. *C. R. Acad. Sci.* **1861**, *53*, 145.
(2) Breslow, R. *Tetrahedron Lett.* **1959**, *1* (21), 22.
(3) Langenbeck, W. *Angew. Chem.* **1954**, *66*, 151.
(4) Langenbeck, W. *Tetrahedron* **1958**, *3*, 185.
(5) Pfeil, E.; Schroth, G. *Chem. Ber.* **1952**, *85*, 293.
(6) Ricardo, A.; Frye, F.; Carrigan, M. A.; Tipton, J. D.; Powell, D. H.; Benner, S. A. *J. Org. Chem.* **2006**, *71*, 9503.
(7) Kim, H. J.; Ricardo, A.; Illangkoon, H. I.; Kim, M. J.; Carrigan, M. A.; Frye, F.; Benner, S. A. *J. Am. Chem. Soc.* **2011**, *133*, 9457.
(8) Nicoll, R.; Hillier, I.; Truhlar, D. *J. Am. Chem. Soc.* **2001**, *123*, 1459.
(9) Delidovich, I. V.; Simonov, A. N.; Pestunova, O. P.; Parmon, V. N. *Kinet. Catal.* **2009**, *50*, 297.
(10) Henry, L. *C. R. Acad. Sci.* **1895**, *121*, 210.
(11) Pfeil, E.; Ruckert, H. *Justus Liebigs Ann. Chem.* **1961**, *641*, 121.
(12) Weiss, A. H.; LaPierre, R. B.; Shapira, J. *J. Catal.* **1970**, *16*, 332.
(13) Harsch, G.; Harsch, M.; Bauer, H.; Voelter, W. J. *J. Chem. Soc. Pak* **1979**, *1*, 95.
(14) Iqbal, Z.; Novalin, S. *Emirates J. Food Agric.* **2011**, *23*, 338.
(15) Harsch, G.; Bauer, H.; Voelter, W. *Liebigs Ann. Chem.* **1984**, 623.
(16) Rose, I. A. *J. Am. Chem. Soc.* **1958**, *80*, 5835.
(17) Breslow, R.; Levine, M.; Cheng, Z. *Origins Life Evol. Biospheres* **2010**, *40*, 11.
(18) Breslow, R. *Tetrahedron Lett.* **2011**, *52*, 2028.
(19) Breslow, R.; Ramalingam, V.; Appayee, C. *Origins Life Evol. Biospheres* **2013**, DOI: 10.1007/s11084-013-9347-0.