Chemical evolution: The mechanism of the formation of adenine under prebiotic conditions

Debjani Roy[†], Katayoun Najafian[‡], and Paul von Ragué Schleyer[†§]

[†]Center for Computational Chemistry, Department of Chemistry, University of Georgia, Athens, GA 30602-2525; and [‡]Department of Chemistry, Acadia University, Wolfville, NS, Canada B4P 2R6

Communicated by Norman L. Allinger, University of Georgia, Athens, GA, September 6, 2007 (received for review May 10, 2007)

Fundamental building blocks of life have been detected extraterrestrially, even in interstellar space, and are known to form nonenzymatically. Thus, the HCN pentamer, adenine (a base present in DNA and RNA), was first isolated in abiogenic experiments from an aqueous solution of ammonia and HCN in 1960. Although many variations of the reaction conditions giving adenine have been reported since then, the mechanistic details remain unexplored. Our predictions are based on extensive computations of sequences of reaction steps along several possible mechanistic routes. H_2O- or NH_3-catalyzed pathways are more favorable than uncatalyzed neutral or anionic alternatives, and they may well have been the major source of adenine on primitive earth. Our report provides a more detailed understanding of some of the chemical processes involved in chemical evolution, and a partial answer to the fundamental question of molecular biogenesis. Our investigation should trigger similar explorations of the detailed mechanisms of the abiotic formation of the remaining nucleic acid bases and other biologically relevant molecules.

reaction mechanism | DNA | RNA | density functional theory | gas phase model

How did life on earth begin? The presence of biomolecules was a prerequisite, but the origin of even the simplest of these remains a fascinating but unsolved puzzle (1, 2). Numerous experiments have demonstrated that amino acids, nucleotides, carbohydrates, and other essential compounds form under simulated primitive earth conditions from simple starting materials, hydrocarbons, HCN, cyano compounds, aldehydes, and ketones (3, 4). HCN, a high-energy prebiotic precursor, is produced in appreciable amounts, for example, by the action of electric discharges on simulated primitive atmospheres (5). The HCN pentamer adenine (a constituent of DNA, RNA, and many coenzymes) is one of the most abundant biochemical molecules. The abiotic synthesis of adenine from a solution of HCN and ammonia was first reported by Oró and colleagues in 1960 (6–9). Although many experimentalists have contributed to our understanding of possible prebiotic processes, low yields and mechanistic complexities rule out detailed investigations (7, 10–15). The key riddle remains: how do five HCN molecules combine to form adenine under prebiotic conditions? We now propose a detailed, thermochemically favorable mechanism based on density functional theory computations.

In Oró's first experiment (1960), adenine was formed in 0.5% yield by heating solutions of ammonium cyanide (>1.0 M) at 70°C for several days. Since then, the abiotic synthesis of adenine from the polymerization of HCN under various conditions has been achieved many times (7, 10–15). The highest yield of adenine (20%) resulted from the sealed-tube reaction of HCN with liquid ammonia (16). In 1978, Ferris *et al.* (17) detected 0.04% adenine from 0.1 M NH_4CN kept in the dark at room temperature for 4–12 months. In a particularly striking experiment, Levy *et al.* (18) detected 0.035–0.04% adenine from dilute solution of ammoniacal HCN, frozen for 25 years at −20°C and −78°C (simulated prebiotic synthetic process in the Jovian moon Europa).

Experimental findings have suggested the outlines of possible prebiotic pathways for the formation of adenine in prebiotic earth conditions (6, 10–15) (Fig. 1). However, the mechanistic details have not been explored (apart from an *ab initio* mechanistic study of the HCN dimerization and adenine protonation) (19, 20). Experimental investigations would be very difficult because adenine is not formed cleanly, yields are small, and many steps are involved. However, clues are provided by four putative intermediates detected in the product mixtures (Fig. 1): formamidine, 2,3-diaminomaleonitrile (DAMN), 4-amino-5-cyanoimidazole (4-aminoimidazole-5-carbonitrile; AICN) (10, 11, 13–15), and 4-aminoimidazole-5-carboxamidine (9, 11).

The pentamerization of HCN to give adenine is very exothermic overall ($\Delta G^{298} = -53.7$ kcal/mol; Fig. 2). Likewise, each successive step, HCN dimerization and the sequential HCN additions to give trimer, tetramer, and pentamer, also is quite exothermic. But a favorable thermochemistry does not ensure that reactions actually will proceed, because they may be prevented by insurmountable kinetic impediments.

Reaction mechanisms involving several intermediate and transition states (which are difficult to detect and identify experimentally) can be studied effectively computationally. These allow selection among various possibilities. Our investigations have identified a plausible detailed step-by-step mechanism for the formation of adenine based on density functional theory computations (21) (described in *Methods*).

We focus on the last, crucial stage, the reaction of the HCN tetramer, AICN, with a fifth HCN to give adenine. This is the key, rate-limiting step in the general adenine formation pathway starting from HCN (see Fig. 1). Although DAMN and AICN are both $C_4H_4N_4$ isomers, AICN was selected as the adenine precursor because of its much lower energy (19.3 kcal/mol according to our computations) and greater structural similarity to adenine (Fig. 2). Like adenine, AICN has a C–C–C–N sequence, whereas DAMN has a C–C–C–C backbone not present in adenine (Fig. 1). DAMN may form experimentally as a kinetically controlled side product. Although DAMN cannot be involved directly in the pathway, it may serve as a "material reservoir." Indeed, many details of photoisomerization of DAMN to AICN in water are established (13, 15). In contrast, the nonphotolytic conversion of DAMN to AICN or to adenine has not been observed (11). Here, we describe our investigation of the mechanism for adenine formation from AICN under nonphotochemical conditions (reactions in the dark); consideration of DAMN as adenine precursor will be reported subsequently.

Author contributions: D.R. and P.v.R.S. designed research; D.R., K.N., and P.v.R.S. performed research; K.N. analyzed data; and D.R. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: AICN, 4-amino-5-cyanoimidazole; DAMN, 2,3-diaminomaleonitrile; PCM, polarizable continuum model; ZPVE, zero-point vibrational energy.

[§]To whom correspondence should be addressed. E-mail: schleyer@chem.uga.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0708434104/DC1.

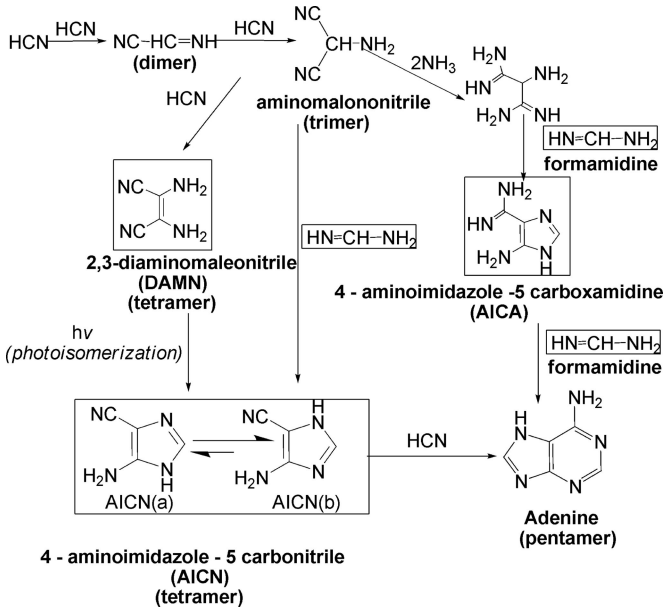

Fig. 1. Proposed steps for the formation of adenine in aqueous ammonium cyanide solution (7, 10–15). Experimentally detected putative intermediates in the abiotic formation of adenine are enclosed in boxes. Two tautomers of AICN can exist; AICN(b) is the more stable. Note that a photoisomerization step is proposed for the formation of AICN from DAMN. DAMN has not been demonstrated to be an adenine intermediate in a nonphotolytic reaction.

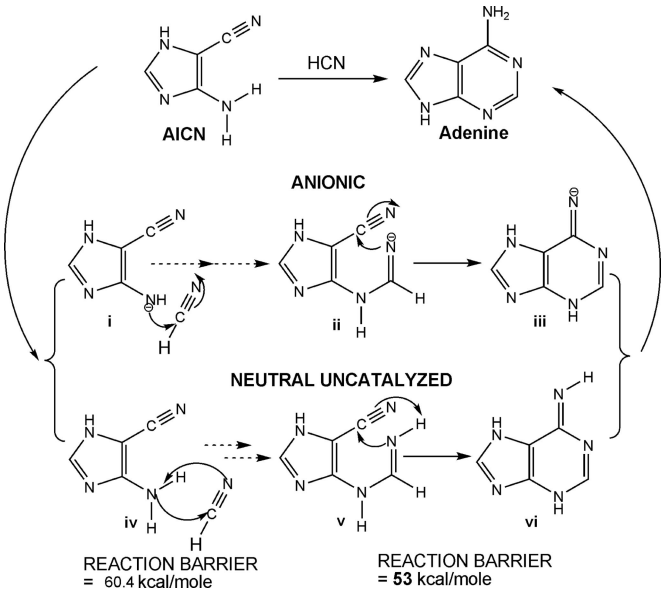

Fig. 3. Anionic mechanisms are unfeasible in isolation. On optimization, both **ii** and **iii** revert back to **i** (reactants). Free radical and neutral *uncatalyzed* mechanisms are also not viable because of the very large reaction barriers for the two steps shown.

An anionic mechanism for the abiotic formation of adenine seems plausible from a physical organic chemist's point of view. HCN is a weak acid and adenine forms in ammoniacal solution, but mechanisms involving unsolvated anionic species proved not to be feasible computationally. Ring closure (Fig. 3 **ii** → **iii**) is precluded because the putative anionic intermediates **ii** and **iii** are not minima, and **iii** reverts to the ring-opened isomers on optimization. Mechanisms involving free radical intermediates also were unpromising. In contrast, optimizations of the reactant, product, and transition states for the neutral closed-shell mechanism were successful.

We first point out explicitly that two AICN tautomeric forms, AICN(a) and AICN(b) (Fig. 1), can exist and that AICN(b) is favored strongly at equilibrium (by $\Delta G^{298} = 3.73$ kcal/mol in the gas phase and 1.73 kcal/mol with bulk solvation, according to our computations). [Bulk solvation (22) is described in *Methods*.] The pathway to adenine from the less stable AICN(a) is precluded by the second high reaction barrier associated with the six-membered ring closure step. The reaction profile is depicted in Fig. 4.

We report the favorable mechanism of adenine formation starting from the thermodynamically more stable AICN(b) tautomer. However, the first step from AICN(b), the addition of HCN to the NH_2 group, cannot proceed directly through a four-center transition state: the reaction barrier (60.4 kcal/mol in the gas phase relative to the energy of the 1:1 HCN–AICN complex; Figs. 3 and 5) is much too high. Moreover, a prohibitive reaction barrier (53 kcal/mol) also precludes a subsequent step in the neutral uncatalyzed pathway (involving the concerted six-membered ring closure and H transfer; Fig. 3). Therefore, the neutral, uncatalyzed mechanism for adenine formation also is unlikely. The involvement of additional molecules serving as catalysts is required to lower the high barriers.

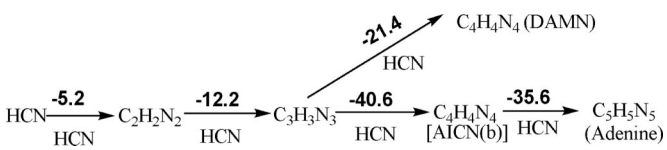

Fig. 2. Thermochemistry of pentamerization of HCN. The relative energies in gas phase are in kilocalories per mole computed at B3LYP/6–311+G**+ZPVE. Entropy is unfavorable but is not included in each step. Overall energy for pentamerization of adenine (5 HCN → $C_5H_5N_5$) is −93.8 kcal/mol ($\Delta G^{298} = -53.7$ kcal/mol). Note that the last crucial step for formation of pentamer (adenine) from tetramer [AICN(b)] is highly exothermic.

The unfavorable four-center transition structure (TS0) associated with the first step from AICN(b) is shown in Fig. 5. The catalytic participation of a water molecule in the six-center alternative (TS1) reduces the barrier considerably (ammonia or similar molecules can function similarly). The importance of water-assisted proton transfer is well known in keto–enol tautomerization (23, 24). The water bridge connects the donor and acceptor sites and stabilizes the transition structure. The classical proton transfer barrier is lowered substantially. The energetically most favorable six-center cyclic transition structure leading to the HCN–AICN(b) addition product (TS1) (Fig. 5) is very different from TS0 for the neutral uncatalyzed mechanism. The water molecule in TS1 transfers its hydrogen-bonded proton to the HCN nitrogen concertedly with the formation of the new bond between the AICN amine nitrogen and the electron-deficient HCN carbon.

The inclusion of one specific water molecule decreases the reaction barrier drastically, from 60.4 kcal/mol (without water) to 38.0 kcal/mol (with a single H_2O) (Fig. 4). Consequently, we also explored the effect of more than one solvent molecule in the mechanism. A second H_2O does not participate in the proton relay effectively, but can form relatively strong hydrogen bonds stabilizing the reactant, product, and transition state complexes. The energetically most favorable transition structure with two water molecules (TS2) also is depicted in Fig. 5. However, inclusion of two explicit H_2O molecules decreases the reaction barrier by only an additional 0.4 kcal/mol (to 37.6 kcal/mol). It does not seem likely that additional explicit H_2O molecules would have much of a further effect. However, a complete solvation shell does have a significant influence (see below). Almost all of the reported abiotic

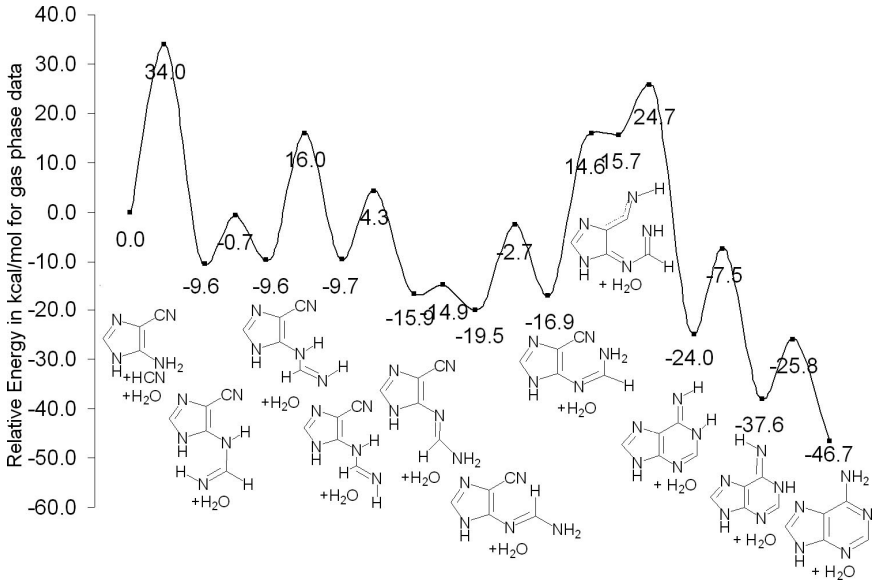

Fig. 4. Gas-phase reaction profile [B3LYP/6-31G*] for adenine formation from the less stable AICN(a) isomer, when one explicit H_2O molecule is included as catalyst. This pathway is precluded by a second high reaction barrier.

syntheses of adenine were carried out with HCN dissolved in water–ammonia solutions. Besides maintaining the pH of the medium, ammonia might also participate mechanistically. Indeed, an explicit NH_3 molecule is as good a catalyst as an explicit H_2O molecule. The reaction barrier for the first step, the NH_3-catalyzed addition of HCN to AICN(b) (with one explicit NH_3), is 37.1 kcal/mol, as compared with 38.0 kcal/mol with an explicit H_2O. Although computations based on species in isolation (gas phase) may simulate extraterrestrial conditions, they do not suffice to predict prebiotic processes on the primitive earth, where reactions

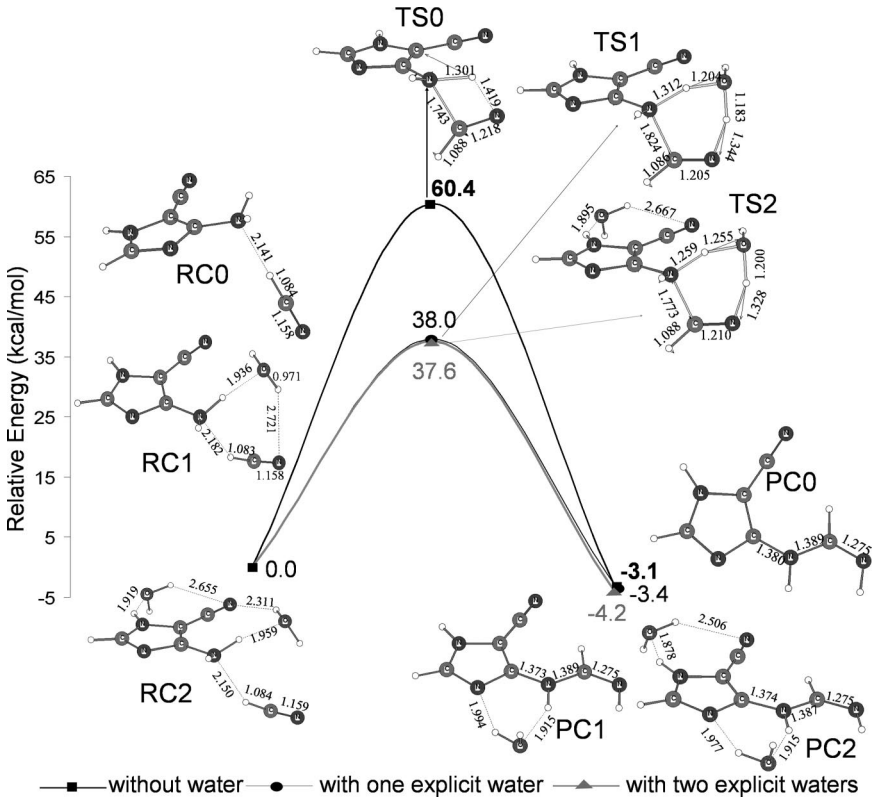

Fig. 5. Gas-phase potential energy profiles (in kilocalories per mole) for the first key addition step of HCN to AICN(b) (B3LYP/6-311+G** + ZPVE). The uncatalyzed reaction, shown by the dashed line at the top, has a prohibitively high 60.4 kcal/mol barrier. Optimized geometries (in angstroms) are given for reactant complexes (RC), transition structures (TS), and product complexes (PC) having one or two explicit catalytic H_2O molecules. Note that these are active participants and reduce the barrier to 38.0 and 37.6 kcal/mol, respectively (entropy was not considered).

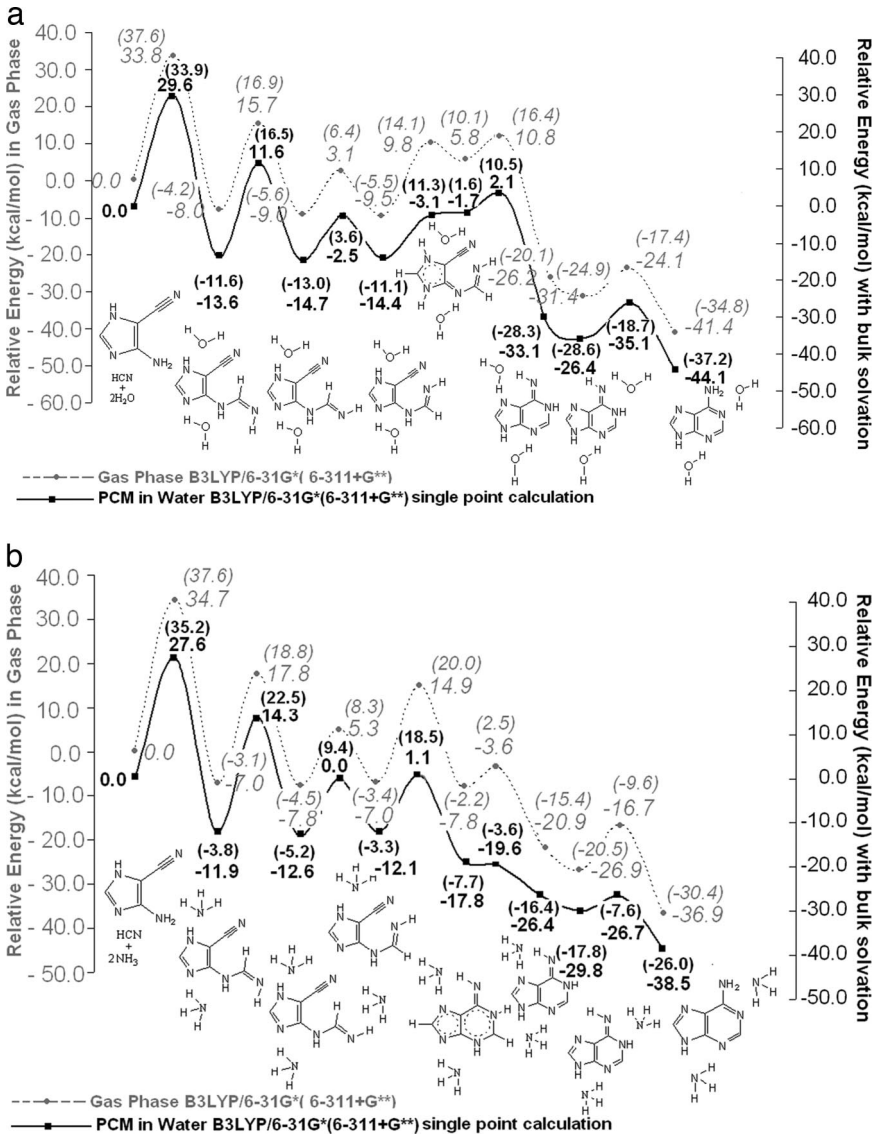

Fig. 6. Reaction profiles for the formation of adenine starting from AICN(b) and HCN in the gas phase and with simulated bulk water solvation by means of explicit solvent-catalyzed mechanisms (two solvent molecules). (*a*) Reaction profile with water as explicit catalytic molecule (gas phase vs. simulated bulk water solvation). (*b*) Reaction profile with ammonia as explicit catalytic molecule (gas phase vs. simulated bulk water solvation). The two H_2O or NH_3 molecules facilitate a ''proton relay'' by forming an H-bonded ''circuit'' for the proton transfer in a six-membered transition state. All species shown are stable minima. (Dotted lines depict partial bonds in complexes or transition states.) The comparisons with the gas-phase profiles show the large extent to which simulated water solvation reduces the barrier electrostatically. The first step is rate-determining in all cases. The basis set dependency of the barrier heights is shown by the comparison data at 6-31G* and at 6-311+G** (in parentheses).

might have taken place in solution. In addition to the explicit solvent modeling (see above), we simulated the bulk solvation effect by employing the polarizable continuum model (PCM) (22), which considers the solvent as a macroscopic continuum of dielectric constant. Bulk solvation stabilizes structures involving greater charge separation (e.g., transition states) preferentially. [This is a simple alternative to explicit solvation modeling (involving a box of water molecules) by using computationally demanding *ab initio* molecular dynamics methods (25).] Indeed, bulk solvation reduces the reaction barriers for the rate-determining step to 33.9 kcal/mol (from 37.6 kcal/mol) and to 35.2 kcal/mol (from 37.6 kcal/mol) for specific H_2O/NH_3-catalyzed mechanisms, respectively. These barriers are low enough to be consistent with the experimental observations as well as conjectures regarding the abiotic genesis of adenine. The subsequent steps from the HCN–AICN adduct are depicted in Fig. 6*a* (for catalysis by two H_2O molecules) and Fig. 6*b* (for catalysis by two NH_3 molecules). The effects of bulk solvation are included in both plots. Except for the *syn–anti* hydrogen transfer and the C–N bond rotation in the second and third steps, all of the stages require the catalytic participation of at least one H_2O or NH_3 molecule. The proton relays across six- and five-membered rings are interesting mechanistic features, because two H_2O or NH_3 molecules are needed to complete the H-bonded "circuit." One molecule participates in the proton transfer directly as a catalyst, while the second H_2O or NH_3 assists by hydrogen bonding. The detailed step-by-step mechanism for the formation of [adenine·$(H_2O)_2$] starting from AICN(b) + HCN + 2 H_2O is depicted in Fig. 6*a*. This reaction profile includes the relative energies of each stationary point (with respect to AICN + HCN + 2 H_2O 1:1:2 reactant complex) without and with bulk solvation. The first step is rate-determining (as stated in our earlier discussion). The subsequent steps have lower reaction barriers.

All of the mechanistic features of the NH_3- (Fig. 6*b*) and the H_2O-catalyzed (Fig. 6*a*) pathways are not alike, because ammonia and water have different Lewis base properties. Although the first three steps, that is, addition of HCN to AICN catalyzed by ammonia, *syn–anti* hydrogen transfer, and rotation around C–N, are similar, concerted six-membered ring closure and 1,4 H transfer take place in the NH_3-catalyzed mechanism before 1,3 H transfer.

Our computed reaction profile with low energy barriers reveals the feasibility of adenine formation from AICN under abiotic conditions. But are the computed reaction barriers reliable? The B3LYP density functional theory method we used is known to give relatively accurate structures and spectroscopic properties of first-row molecules, but activation barriers may be more problematical (26, 27). Our computed reaction barriers at B3LYP/6–31G* are ≈3–4 kcal/mol lower than at B3LYP/6-311+G**. The key rate-determining step was recomputed at the MP2/6-311+G** and the gas-phase reaction barrier with zero-point vibrational energy (ZPVE) correction is 43.9 kcal/mol, which is higher than that at B3LYP/6-311+G**. It has been reported that although MP2 in general gives accurate geometries for reaction complexes, it also overestimates barrier heights (28, 29). Thus for calibration, we performed higher-level CCSD(T)/cc-pVTZ single-point computations on the simpler HCN + NH_3 + H_2O system on MP2 geometries optimized at 6-311+G** level. This models the key AICN(b) + HCN + H_2O rate-determining step, because it also involves the addition of a NH bond to HCN. (The geometries of the reactants, transition structures, products, and barriers, computed at various levels of theory, are given in supporting information (SI) Fig. 7). The CCSD(T)/aug-cc-pVTZ single-point reaction barrier is only 3 kcal/mol higher than the B3LYP/6–311+G** result.

Conclusion

Our detailed computational investigations of step-by-step formation pathways from AICN(b) show how adenine can arise abiotically. Although the formation of adenine by the pentamerization of HCN is very exothermic, this process is quite unlikely in isolation (gas phase). Not only must five HCN molecules come together, but also the reaction barriers are very high. The intimate participation of an additional molecule, such as H_2O or NH_3 (or perhaps HCN) is needed to lower the barriers considerably to realistic energies. Moreover, an aqueous medium facilitates the reaction, because both specific and bulk solvation lowers the barrier of the rate-determining step further. The reaction energetics are mostly governed by the enthalpy change (see Fig. 2 legend) especially at the low temperatures of some of the experiments.

Finding a viable, thermodynamically feasible, step-by-step mechanism that can account for the formation of adenine did not prove to be easy. Our approach has a model character to it. The "first model" is without water or ammonia, the second model is with explicit participation of ammonia (along with water) in gas phase, and the third model adds bulk solvation. Because there are no quantitative experimental data to match, a highly refined study, for example, by using the approach described by Jorgensen and his coworkers (30) to model the medium effects, is not called for at this initial stage. The computed activation energy is reasonable considering the long time scale of chemical evolution toward complex organic systems on primitive earth. An alternate pathway based on the experimental isolation of 2- and 8-cyanoadenine or adenine 8-carboxamide as adenine precursor suggests a further complex mechanism involving hexamer and heptamers of HCN (31). The feasibility of that alternate pathway requires further investigation. The subject is of continuing interest, and detailed calculations on reactions of this kind are needed before a full picture emerges. Our investigation should trigger similar explorations of the detailed mechanisms of the abiotic formation of the remaining nucleic acid bases and other biologically relevant molecules.

Methods

The computations were performed at the B3LYP density functional theory level with the 6-31G* and the 6-311+G** basis sets. In addition, CCSD(T)/aug-ccpVTZ *ab initio* computations simulated the first HCN + RNH_2 addition step with the highest barrier. All computed harmonic frequencies of fully optimized minima were real, whereas transition structures (states) had a single imaginary frequency. Intrinsic reaction coordinate analyses of the minimum-energy pathways confirmed that the transition states led to the reactants and products shown in the figures. All of the relative energies are corrected by the ZPVE and correspond to the classical reaction barriers (i.e., without proton tunneling). The reaction barriers are defined as the difference in sum of electronic and zero-point energies (ε_0 + ε_{ZPE}) of the reactant complex and transition state. ΔG is defined as difference between the sum of electronic and thermal free energies (ε_0 + G_{corr}) of reactant and product.

Instead of modeling bulk solvation by an explicit shell involving many solvent molecules surrounding the solute (25), we used the PCM implemented in the Gaussian 98 program (21). The PCM "bulk solvent medium" is simulated as a continuum of dielectric constant ε. This surrounds a solute cavity, which is defined by the union of a series of interlocking spheres centered on the atoms. Our computed single-point PCM bulk solvent simulations used the optimized equilibrium geometries. We have added an estimated ≈2.5 kcal/mol ZPVE correction to the single-point PCM energy. The derivation of this corrected energy value is discussed in detail in the *SI Dataset*. More sophisticated solvation treatments are not called for in the absence of experimental data.

During the review of this article, Glaser *et al*. (32) proposed and analyzed mechanisms of pyrimidine-ring formation of isomeric monocyclic HCN pentamers. Pyrimidine-ring formation from a monocyclic HCN pentamer, either by proton-catalyzed or by uncatalyzed cyclization, was found not to be viable thermochemically. However, they propose that photoactivation of either of these cyclization paths might lead eventually to the imino form of adenine.

We thank Prof. T. T. Tidwell for calling our attention to this problem and for his encouragement and the referees who made constructive suggestions, which led to improvements in the article. This work was supported by National Science Foundation Grants CHE-0209857 and CHE-0716718 and the University of Georgia.

1. Miller SL, Orgel LE, eds (1974) *The Origins of Life on Earth* (Prentice-Hall, Englewood Cliffs, NJ).
2. Schopf JW, ed (1983) *Earth's Earliest Biosphere: Its Origin and Evolution* (Princeton Univ Press, Princeton).
3. Eschenmoser A, Loewenthal E (1992) *Chem Soc Rev* 21:1–16.
4. Orgel LE (2004) *Crit Rev Biochem Mol Biol* 39:99–123.
5. Abelson PH (1966) *Proc Natl Acad Sci USA* 55:1365–1372.
6. Oró, J (1960) *Biochem Biophys Res Commun* 2:407–412.
7. Oró, J (1961) *Nature* 191:1193–1194.
8. Oró, J., Kimball AP (1961) *Arch Biochem Biophys* 94:217–227.
9. Oró, J., Kimball AP (1962) *Arch Biochem Biophys* 96:293–313.
10. Ferris JP, Sanchez RA, Orgel LE (1968) *J Mol Biol* 33:693–704.
11. Shuman RF, Shearin WE, Tull RJ (1979) *J Org Chem* 44:4532–4536.
12. Voet AB, Schwartz AW (1983) *Bioorg Chem* 12:8–17.
13. Ferris JP, Orgel LE (1966) *J Am Chem Soc* 88:3829–3831.
14. Ferris JP, Orgel LE (1965) *J Am Chem Soc* 87:4976–4977.
15. Ferris JP, Orgel LE (1966) *J Am Chem Soc* 88:1074–1074.
16. Wakamatsu H, Yamada Y, Saito T, Kumashiro I, Takenishi T (1966) *J Org Chem* 31:2035–2036.
17. Ferris JP, Joshi PC, Edelson EH, Lawless JG (1978) *J Mol Evol* 11:293–311.

18. Levy M, Miller SL, Brinton K, Bada JL (2000) *Icarus* 145:609–613.
19. Kikuchi O, Watanabe T, Satoh Y, Inadomi Y (2000) *J Mol Struct* 507: 53–62.
20. Tureček F, Chen X (2005) *J Am Soc Mass Spectrom* 16:1713–1726.
21. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery AJ, Jr, Stratmann RE, Burant JC, *et al.* (1998) GAUSSIAN 98 (Gaussian, Pittsburgh), Revision A.9.
22. Miertus S, Scrocco E, Tomasi J (1981) *Chem Phys* 55:117–129.
23. Yamabe S, Tsuchida N, Miyajima K (2004) *J Phys Chem A* 108:2750–2757.
24. Kiruba GSM, Wong MW (2003) *J Org Chem* 68:2874–2881.
25. Aida M, Yamataka H, Dupuis M (1998) *Chem Phys Lett* 292:474–480.
26. Barone V (1994) *Chem Phys Lett* 226:392–398.
27. Martell JM, Goddard JD, Eriksson LA (1997) *J Phys Chem A* 101:1927–1934.
28. Wiberg KB, Ochterski JW (1997) *J Comput Chem* 18:108–114.
29. Gonzales-Garcia N, Gonzales-Lafont À, Lluch JM (2005) *J Comput Chem* 26:569–583.
30. Jorgensen WL, Alexandrova AN (2007) *J Phys Chem B* 111:720–730.
31. Borquez E, Cleaves HJ, Lazcano A, Miller SL (2005) *Origins Life* 35:79–90.
32. Glaser R, Hodgen B, Farrelly D, McKee E (2007) *Astrobiology* 7:455–470.