Synthetic routes of the fundamental building blocks of life: Computational study of the reaction free energy

Kelson Mota T. Oliveira[a] (PQ), Elaine Harada[b] (PQ)

[a] *Laboratório de Química Teórica e Computacional, LQTC/DQ-ICE.* [b]*Instituto de Computação, IComp. Av Rodrigo Otávio 3000, Setor Norte. Manaus-AM. Universidade Federal do Amazonas*
kelsonmota@ufam.edu.br, elaineharada@ufam.edu.br

Keywords: Abiogenesis, nucleic acids, DNA sequences, AM1

INTRODUCTION

In recent years, theoretical and experimental studies have presented efforts to understand the origin of the building blocks that gave rise to life, generating a lot of discussion about abiogenesis theme[1-3]. The most widely widespread idea, known as "the RNA World"[4-5], maintains that life arose spontaneously through five main stages: (1) prebiotic synthesis of nucleotides; (2) prebiotic formation of polynucleotides from the nucleotides; (3) emergence of special RNA molecules catalyzing their own replication; (4) evolution of the primordial replicases (?) towards more efficient ones; (5) emergence and evolution of other and better catalytic RNA molecules[6-7]. All stages are target of much speculation and discussion as original conditions cannot be replicated faithfully in the laboratory, and there is no guarantee that it will be performed in the future[8]. This work, from a semi-empirical theoretical approach, investigates some aspects of the two first stages by means of thermochemical data from the prebiotic nucleotide synthesis reactions, and short sequences of DNA and RNA. The objective was to verify that various abiotic reactions identified in the literature[9-10] as natural precursor in the formation of life blocks can occur spontaneously in abiotic environment. Gibbs free energy was used as calculating criterion , since the thermodynamic parameters calculated by semi-empirical methods such as AM1 and PM3 has shown a good correlation with experimental data[11].

METHODS

The theoretical quantum chemical calculations were performed in an INTEL Quadcore™ PC (8 GB RAM), on the Debian LINUX (5.0 version) by Gaussian 03 Program (Revision E.01). In the optimization geometry and energy routines, the AM1 and PM3 methods were used. Molecular geometries were fully optimized by the force gradient method using Berny's algorithm, and potential energy surfaces were characterized using standard analytical harmonic vibrational analysis to confirm that the stationary points corresponded to minima of the potential energy surfaces (no imaginary frequencies or negative eigenvalues were found). Thermochemical parameters ΔH_r and ΔS_r for each compound were obtained from vibrational analysis (T=298.15K), allowing the calculation of ΔG_r. The methodology consisted of modeling of reaction systems in two distinct stages: 1. Abiotic synthesis reactions of nucleic acids; 2. Formation reactions of short sequences of nucleotides. For this purpose the optimized geometries were obtained as well as the calculation of the energies, frequencies and thermochemical data of each participant's reaction compound. The nucleotide sequences used are for a single helix conformation, as shown Table 1.

Table 1. Conformational data from a single helix sequence.

Single Helix β	
α	-39.17
β	-151,44
γ	30,89
χ	-95,44
δ	156,56
ε	159,19
ζ	-98,97
Sugar form	3' Endo

RESULTS AND DISCUSSION

1. Reactions from nucleic acid synthesis

The estimation of the free energies of formation of each compound (Table 2) allowed the calculation of ΔG_r of each reaction (Table 3). The more traditional reaction routes were chosen (Figure 1). In the case of adenine and cytosine, two main routes have been calculated. The abiotic synthesis of thymine and guanine is subject of discussion, and has not been investigated in this work [12].

Table 2. Thermochemical Data of compounds necessary for the abiotic synthesis reactions of nucleic acids, nucleotides and sequences.

Compounds	ΔH (Kcal/mol)	ΔS (Cal/mol)	ΔG^* (Kcal/mol)
C-Diaminomaleonitrile	136.575669	87.49	110.4905255
T-Diaminofumaronitrile	137.0281037	87.35	110.9847012
Aminoimidazol Carbonitrile	143.6157037	81.11	119.4327572
HCN	44.00288373	47.89	29.72448023
HCN dimer	82.05885519	63.89	63.00885909
Aminomaleonitrile	110.2911576	76.81	87.3896598
Formamidine	55.92557373	61.35	37.63228233
Adenine	163.2818671	86.24	137.5694111
Thymine	17.07517461	84.59	-8.14533389
Citosine	69.92092926	78.88	46.40285726
Guanine	129.5707748	89.54	102.8744238
Uracil	6.44703774	79.56	-17.27377626
Deoxyribose	-101.4005959	91.61	-128.7141174
H_3PO_4	-252.5621073	77.99	-275.8148258
$NH_2CONHCHCHCN$	73.8830274	90.05	47.0346199
PO_4	-57.79178847	74.71	-80.06776757
dAdenine	120.3156299	129.89	81.58892635
dCitosine	28.44314577	125.81	-9.06710573
dGuanine	86.9979864	133.73	47.1263869
dTIMINA	-25.51518411	131.41	-64.69507561
HCCCN	107.2251437	57.67	90.03083324
HOCHCHCN	38.33019333	71.48	17.01843133
NH_2CONH_2	-1.44390051	65.95	-21.10689301
Thiocyanate	13.1588847	47.63	-1.0419998
H_2CO_3	-122.1824721	64.00	-141.2640721
H_2	3.09613434	30.76	-6.07525781
O_2	5.75803176	46.38	-8.07016524
H_2O	-44.05496706	45.09	-57.49855056
NH_4OH	-28.76192085	78.46	-52.15476985
NH_3	16.7168664	45.84	3.0365518
AT	-117.1605096	222.92	-183.6241076
CT	-208.5400	212.75	-271.9772667
ATC	-40.72853	284.05	-107.1921346
ATCG	-231.2606529	398.96	-350.2105769
ATCGA	-69.66929775	485.23	-214.3406223
AUCG	-312.8915462	426.92	-440.1777442
$dAPO_2$	107.5150534	121.98	71.14671636
$dTPO_2$	-43.41678939	125.04	-80.69746539
$dCPO_2$	3.79267044	118.24	-31.46058556
$dGPO_2$	70.75740009	120.5	34.83032509
dPO_4	-301.1972699	117.78	-336.3142713

Table 3. Energias livre (in Kcal/mol) das rotas de síntese de ácidos nucleicos. Free energy (in Kcal/mol) of routes of synthesis of nucleic acids.

Reaction	ΔG_1	ΔG_2	ΔG_3	ΔG_4	ΔG_5	ΔG_6
R 01	3.54	-5.33	-6.62	0.49	8.45	-11.58
R 02	3.54	-5.33	-6.62	8.49	-16.46	
R 03	-15.51	51.12	56.87			
R 04	-68.22	56.87				
R 05	56.87	-0.82				

Results showed that all the nucleic base synthesis routes are energetically unfavorable and may not spontaneously occur in the way it has been proposed. Reactions 01 and 02 had more than one intermediate step thermodynamically unfavorable.

The second step of the reaction 03 presents a large ΔG_r, which makes it impossible to produce cytosine. There is a similar argument to reaction 04. Reaction 05, as proposed by Luisi[15] depends on the spontaneous formation of cytosine by other mechanisms, especially thiocyante and urea routes, which proved to be energetically unfavorable. In this case, with the formation of cytosine, uracil spontaneous generation is possible, a basic component for RNA.

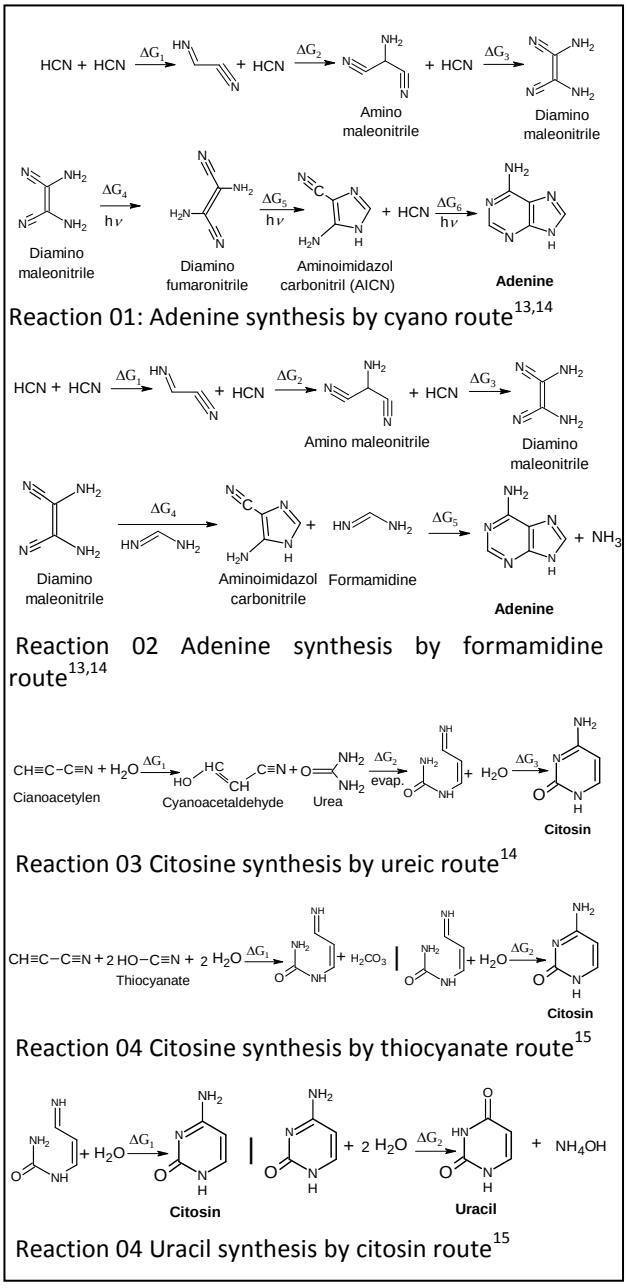

Figure 1. Abiotic routes of synthesis of Adenine, Cytosine, Uracil nucleic acids.

2. Reactions from formation of short sequences of DNA and RNA

The abiotic mechanism of simple sequences of nucleotides formation has never been adequately

and sufficiently demonstrated. There are at least three possible suggestions of formation: (a) reaction from the individual components present in the environment; (b) nucleosides reaction in environments rich in phosphorus; (c) reaction from pre-existing sequences containing a phosphodiester bridge.

In case (a), free energy was calculated with the thermochemical data of all necessary components to form a single sequence(Table 4). Analysis of the resulting values presents a strong correlation with the size of the sequence. The longer the more positive the value of ΔG_r (Figure 2). Extrapolation to a complete sequence (single helix) with $200x10^6$ nucleotides, as contained in the human genetic code, without any particular order, will result in a ΔG_f of $3.19x10^{10}$ Kcal/mol.

Table 4. Free energy (in kcal/mol) of formation of short sequences, from the individual components present in the environment (nitrogen basis, deoxyribose and phosphoric acid).

COMPONENTS	→ SEQUENCE	ΔG_r
A + T + 2DEOXY + 1 H_3PO_4	→ AT + 2H_2O	105.18
A + T + C+ 3DEOXY + 2 H_3PO_4	→ ATC + 4H_2O	424.74
A + T + C + G + 4DEOXY + 3 H_3PO_4	→ ATCG + 6H_2O	368.37
A + T + C + G + A+ 5DEOXY + 4 H_3PO_4	→ ATCGA + 8H_2O	656.18
A + U + C + G + 4DEOXY + 3 H_3PO_4	→ AUCG + 6H_2O	287.52

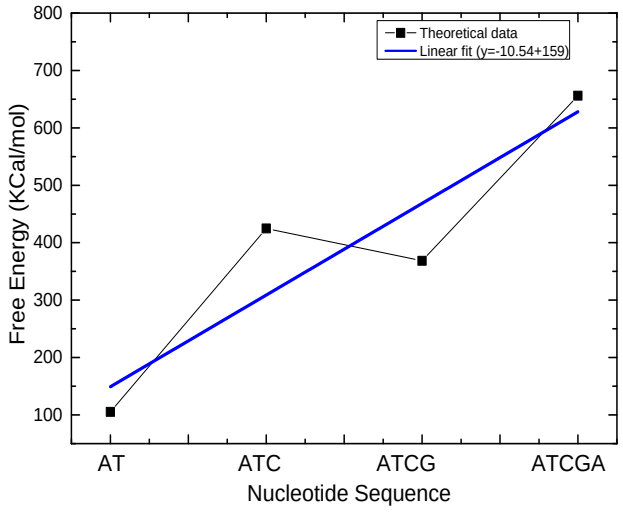

Figure 2. ΔG_r formation of random nucleotide sequences according to the size of the sequence. In blue, the linear fit.

For case (b), some longer sequences were shown to be energetically favorable to spontaneous formation, if complete nucleosides are available in the environment in the presence of anion of phosphoric acid (Table 5). However, the formation deoxynucleotides (dA, dC, dG and dT) proved to be energetically unfavorable.

Table 5. Free energies (in kcal / mol) of d-nucleosides complete reactions in the presence of phosphorus.

COMPONENTS → NUCLEOSIDES/SEQUENCES		ΔG_r
d + A	→ dA + H_2O	15.13
d + T	→ dT + H_2O	14.66
d + C	→ dC + H_2O	15.74
d + G	→ dG + H_2O	15.46
dC + dT + PO_4	→ CT	-118.15
dA + dT + PO_4	→ AT	-120.45
dA + dT + dC + 2PO_4	→ ATC	45.12
dA + dT + dC + dG + 3PO_4	→ ATCG	-164.95
dA + dT + dC + dG + dA+ 4PO_4	→ ATCGA	-30.61

The latter case (c) investigated the formation of new sequences from other smaller d-nucleosides sequences containing a phosphodiester bridge (see Figure 3). The results showed that some sequences may have negative ΔG_r, but the data are not sufficient to set as a rule. Once the first sequence is not energetically favored (107.89 Kcal/mol), the next cannot be formed therefrom (Table 6).

Table 6. Formation free energy (in kcal/mol) of new sequences from pre-existing sequences and d-nucleosides with phosphodiester bridge.

SEQUENCE PREEXISTING → RESULTING SEQUENCE		ΔG_r
AT + $dCPO_2$	→ ATC	107.89
ATC + $dGPO_2$	→ ATCG	-277.85
ATCG + $dAPO_2$	→ ATCGA	64.72

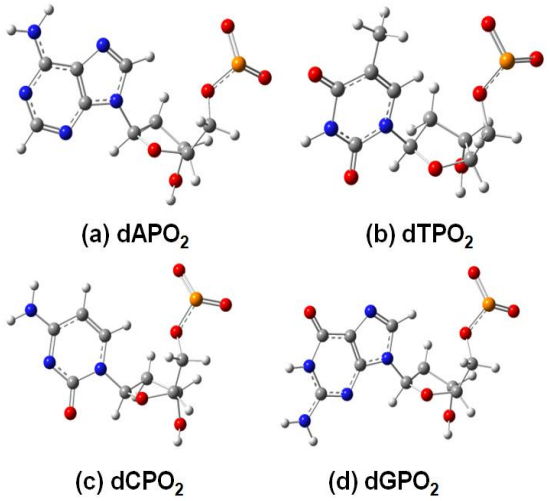

Figure 3. Deoxyribonucleosides containing a phosphodiester bridge.

CONCLUSIONS

None of the investigated abiotic reactions of nucleic acids synthesis showed positive ΔG_r in all their stages. Calculated values do not contemplate a spontaneous occurrence of these reactions in an abiotic way. The formation of sequences of nucleotides, considering the presence of all components in an abiotic reaction environment also resulted in positive values. In this case, even with all components present, there is no formation of nucleotides sequences. In the presence of complete deoxynucleotides there is the possibility of spontaneous formation of larger sequences. However, the formation of these deoxynucleotides remains not spontaneous , which prevents the generation of longer sequences. Finally, reactions with complete nucleosides which exhibit a phosphodiester bridge, do not guarantee that larger DNA sequences can be formed continuously. Therefore, it is suggested that other reactions and proposals should be exhaustively investigated before giving a final word on the appearance of bases of life from an abiotic environment.

ACKNOWLEDGMENTS

The authors are grateful for the support given from the FAPEAM and PPGQ/UFAM.

[1] B.S. Palmer, Int. J. of Astrobio., 12, 39 (2013).
[2] L. Li, C.Francklyn, C.W. Carter Jr, J. Biol. Chem., 288, 26856 (2013).
[3] W. Ma, C. Yu, W. Zhang, Biosystems 90, 28 (2007).
[4] W. Gilbert, Nature 319, 618 (1986).
[5] L.E. Orgel, Crit. Rev. Biochem. Mol. Biol. 39, 99 (2004).
[6] G.F. Joyce, L.E. Orgel. The RNA World. Cold Spring Harbor Laboratory Press, NY, 49–77 (1999)
[7] G.F. Joyce, Nature, 418, 214 (2002)
[8] J.L. Bada, Earth and Planet. Sc. Letters 226, 1 (2004).
[9] S.A. Miller, Science. 117, 528 (1953)
[10] S.A. Miller, A. Lazcano (78-109); J. Ferris (113-135); L. Orgel (140-154) . In: Life's Origin: The Beginnings of Biological Evolution. J. William Schopf (Ed). University of California Press, 78 (2002).
[11] X. Ma, H.H. Schobert. J. Phys. Chem. A 104, 1064(2000).
[12] S.G. Trevino, N. Zhang, M.P. Elenko, A. Lupták, J.W. Szostak. PNAS, 108, 13492 (2011)
[13] J.P. Ferris, A.R. Hill, R. Liu, L.E. Orgel. Nature, 381, 59 (1996).
[14] S.A. Miller, A. Lazcano. Formation of the Building Blocks of Life. W. Schopf, ed. (2002)
[15] P.L. Luisi. The Emergence of Life: From Chemical Origins to Synthetic Biology. Cambridge University Press (2006).