

## HUNT FOR MOLECULES AT THE ORIGIN OF LIFE: THE QUANTUM CHEMISTRY APPROACH

Y. Ellinger, J.M. Chiaranello, G. Berthier, M. Lattalais, F. Pauzat,  
D. Talbi<sup>1</sup>

**Abstract.** The contribution of Quantum Chemistry to the interdisciplinary field that is the hunt for molecules in space is illustrated in several domains of interest to the search for the origin of life: the critical checking of a reaction scheme for the formation of Adenine, the search for phosphorus containing candidates, the reasons for the observation versus non observation of formamide and glycine.

### 1 Introduction

Beyond all philosophical connotations, the pursuit of the origin of life has never ceased to focus the attention of the scientific community. The first step was taken by Wöhler in 1828 with the synthesis of urea,  $\text{O}=\text{C}(\text{NH}_2)_2$  when treating silver cyanate by ammonium chloride. The most widely held belief at that time was that "vital" compounds, that is, those which were of animate origin, could only be produced by living organisms. This synthesis of an organic compound from purely mineral precursors was the beginning of the end for vitalism theory.

A century later, Haldane (1929) developed the first modern theory of the primary steps of emerging life. It was not until the fifties that experiments were conducted by Miller's group (Miller 1953, Miller & Urey 1959)

---

<sup>1</sup> Laboratoire d'Etude Théorique des Milieux Extrêmes, (LETMEX), USM 205, Muséum National d'Histoire Naturelle, Paris and Laboratoire de Chimie Théorique, Université Pierre et Marie Curie, Paris - FRANCE.

which would change the approach of scientific investigation into the origin of life. Molecules which were believed to represent the major components of the early Earth atmosphere,  $\text{CH}_4$ ,  $\text{NH}_3$ ,  $\text{H}_2$ ,  $\text{H}_2\text{O}$ , were submitted to electric discharges; after a week, 10-15 % of the carbon was in the form of organic materials. About 2 % was in the form of some amino acids. This stunning discovery inspired and still stimulates a multitude of further experiments. In most of them, the energy necessary to initiate the reactions is provided in a "violent" form, as electric discharges for lightning, particle beams for cosmic rays or projectiles simulating volcano eruptions (Ponamperuma 1979, Devienne et al. 2000, Takano 2004). In the meantime, Oro (1961) had found that amino acids could also be made from  $\text{HCN}$  and  $\text{NH}_3$  in aqueous solution. He also found a large quantity of Adenine, a basic component of more elaborate material such as DNA.

Many of the compounds found in the Urey-Miller type of experiments have been identified in meteorites, especially carbonaceous chondrites. Over 70 amino acids, including 8 out of the 20 present in living organisms (Cooper et al. 1992), have been identified to date in the Murchison meteorite (fall in 1969). More importantly, the Murchison and Ivuna meteorites have demonstrated that the Earth may have acquired at least some of its elaborate organic material via meteorite and comet infall. If amino acids and other organic compounds were able to survive in the extreme conditions of outer space and reach the Earth then where did they come from?

More than 120 species have been identified in space by radio observations, a number of them of prebiotic interest. A precursor of the peptide bond,  $\text{HNCO}$ , was identified in the interstellar medium (Snyder et al. 1972, Brown 1981) and more recently in comets Hyakutake and Hale-Bopp (Lis et al. 1997, Bockelee et al. 2000). The next step on the amino acids and proteins route,  $\text{NH}_2\text{CH}=\text{O}$ , i.e. the simplest possible species containing the peptide bond, was identified in space (Rubin et al. 1971) and in the comets mentioned above. If more complex molecules have been discovered recently such as  $\text{CH}_3\text{COOH}$  or  $(\text{CH}_3)_2\text{C}=\text{O}$ , glycine,  $\text{NH}_2\text{CH}_2\text{COOH}$ , is still not identified, nor any phosphorous compound.

## 2 The Quantum Chemistry alternative

The hunt for molecules mainly relies on radioastronomy. Identification of a molecule in space relies on a perfect match between the observed lines and the spectrum of the same species in the laboratory. It is something difficult

**Table 1.** General possibilities and actual limitations of the quantum chemistry approach.

Observable	Precision
Rotational constants	< 1 %
Dipole Moment	5 %
Vibrational frequencies	< 1 %
Relative intensities	10 %
Isomerization energy	1 kcal/mol

to realize, since experiments can hardly approach the extreme conditions in space; it is even more true for unstable species. Generally indeed, these species are formed together with a number of other molecules that may be isomers or even compounds with completely different chemical formulas. The role of numerical simulation is to propose theoretical spectra that are precise enough to make possible an interpretation of the laboratory spectra and extract the spectrum of the target molecule from the mixture of experimental values for comparison with the spatial observations. Another role of numerical simulation is to test the relative stability of the possible isomers that can be formed together with the feasibility of the corresponding chemical models. It means looking for exothermic or athermic reaction paths with no or weak activation barrier.

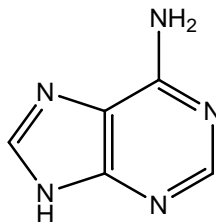
The real challenge in this type of study is to reach the level of accuracy necessary to unambiguously decide whether a reaction is feasible, or which of the possible isomers is the most stable or what precisely are the spectral signatures to be searched. Practically, it is the nature of the problem to be solved that decides the type of simulation to be made. How these numerical simulations can respond to the demands of the hunt for molecules is illustrated in Table 1.

### 3 Results and Discussions

#### 3.1 Testing chemical models

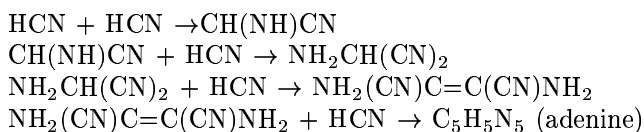
Adenine has the molecular formula  $C_5H_5N_5$ . It can be seen as a pentamer of HCN, a molecule that is quite abundant under primitive conditions in the universe. Recently, Chakrabarti & Chakrabarti (2000) proposed that *oligomerization* of HCN could produce adenine (Figure 1) by gas-phase re-

actions in dense interstellar clouds.



**Fig. 1.** The structure of Adenine

The postulated reaction chain would present four successive steps:

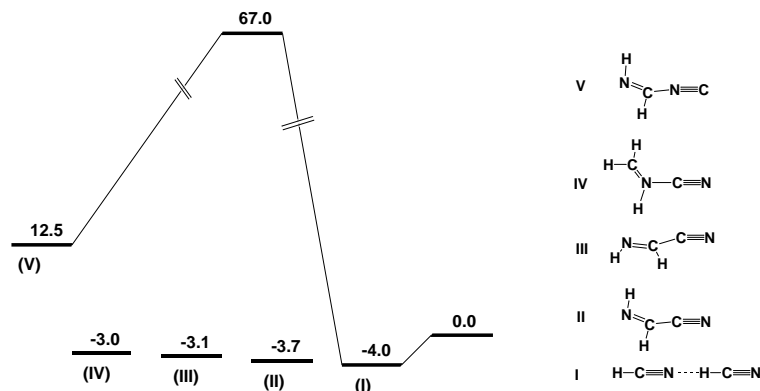


Assuming for each step a temperature independent rate coefficient of  $K = 10^{-10} \text{ cm}^3 \text{ s}^{-1}$ , they found considerable formation of Adenine on relatively short time-scale (188 yr).

Can this series of reactions, extrapolated from chemistry in solution, really occur in the 10-100K conditions of the interstellar medium ?

The first question to be solved is the feasibility of the first reaction. An extensive search for a pathway connecting two isolated HCN molecules to an  $\text{H}_2\text{C}_2\text{N}_2$  chemically bound complex was undertaken. The calculations show (Figure 2) that only two chemically bound isomers (II and III) are lower in energy than two HCN monomers, together with a weakly bound Van der Waals dimer (I). The other two chemically bound isomers IV and V are higher in energy. In fact, none of the dimer structures is strongly bound with respect to  $\text{HCN} + \text{HCN}$ . Even dimer II, which has the structure expected for the product of first reaction is not more stable than the Van der Waals structure.

A search for a pathway connecting the two separated HCN species to isomers I or II failed. The only transition state found connects the hydrogen-bonded dimer to structure V in an endothermic process. From these quantum chemical simulations, it appears that there are no low energy paths to



**Fig. 2.** The lowest energy reaction pathway for formation of a chemically bound HCN dimer obtained at the MP4SDTQ/6-311+G(2df,pd) level of theory (zero point energy correction included) and the five lowest-energy dimers.

any of the chemically-bound dimers II - V. Independently of what particular chemically-bound isomer of HCN is formed, the activation barrier of 71 kcal/mol, equivalent to 36000 K, is far higher than  $kT$  in even warm interstellar regions surrounding new-born stars.

If HCN dimerization to form a chemically-bound dimer is as inefficient as found here, i.e. with a rate coefficient  $cm^3s^{-1}$ :

$$K_{(dimer)}T = 10^{-10} \exp(-36000/T) \quad (3.1)$$

formation of Adenine via reactions of HCN and oligomers (steps 2-4) certainly does not occur in the low-temperature chemistry of interstellar clouds (Smith et al. 2001).

### 3.2 Looking for phosphorous species

The non detection of polyatomic species with phosphorous, that are mandatory for building a DNA world, is nowadays a major difficulty for exobiology models. Only two molecules including phosphorus, PN and PC, have been detected in the interstellar medium. Neither hydrogenated compounds such as HPN, HPC, PNH, PCH (and their corresponding ions) nor

**Table 2.** Relative energies (kcal/mol) of the [C,H,O,P] and [C,3H,O,P] optimized structures including zero-point vibrational energies. The calculations are performed at the B3LYP/6-311G\*\* level of theory.

Isomer	Energy	Isomer	Energy
HPCO	0.0	CH <sub>3</sub> PO	0.0
HOCP	23.5	H <sub>2</sub> C=POH	4.7
cyc-PC(H)O	48.6	PH <sub>2</sub> CH=O	9.8
HCPO	62.2	HPC(H)OH	16.9

PO or PS, which are well characterized in the laboratory, have been observed. On the other hand, complex molecules with a P-C bond have been found in meteorites collected on Earth such as the Murchison meteorite (Cooper et al. 1992). They are phosphonates, derived from phosphonic acid, HPO<sub>3</sub>H<sub>2</sub>, where the hydrogen on the central atom was replaced by hydrocarbon groups.

Taken into account the contradiction between the composition of the interstellar medium and that of meteorites, a systematic investigation of the possible structures and the corresponding properties of the polyatomic systems [C,nH,O,P] with n=(1,3) was undertaken. The relative energies of the four lowest isomers of singlet multiplicity are reported in Table 2 for both series.

The most stable arrangement of the four atoms is HPCO, which is a singlet state. The quasi-linear PCO backbone is well represented by an allenic-like structure HP=C=O similar to that already found for isovalent compounds with low coordinated phosphorus as HP=C=CH<sub>2</sub> (Chentit et al. 1998).

The next isomer on the energy scale, HOCP, has the same PCO connectivity; it is also a quasi-linear singlet state located 23.5 kcal/mol higher in energy. The third singlet state cyc-PC(H)O is a three-membered ring with the hydrogen attached to the carbon. The fourth isomer has a linear structure. Several structures of triplet multiplicity are found between 40 and 70 kcal/mol over the HPCO ground state; however, since the lowest triplet is unstable with respect to dissociation in <sup>3</sup>PH + CO that takes place 38.4 kcal/mol above the ground state, it is not expected that triplet isomers can be formed and a fortiori survive in a spatial environment.

For the six atom system, the most stable structure is CH<sub>3</sub>P=O, The only

**Table 3.** Predicted rotational constants (GHz). The calculations are performed at the B3LYP/6-311G\*\* level of theory.

Candidate	$B_{Predicted}$	$B_{Observed}$	$C_{Predicted}$	$C_{Observed}$
HNCO	11.0910	11.0710	10.9542	10.9106
CH <sub>3</sub> NO	11.4579	11.4593	10.2759	10.2466
HPCO	5.5351		5.4072	
CH <sub>3</sub> PO	8.0275		6.4779	

four stable isomers are within  $\sim 1eV$  above. Contrary to the [C,3H,O,N] family, it is the phosphorous analogue of the nitroso and oxime that are the lowest energy isomers.

The calculation of rotational constants, by quantum chemical methods alone, to an accuracy such that it can be used to support identification of unknown molecules in space, is a difficult task. Various methods of calibration of the raw values are possible to achieve an accuracy better than  $\sim 0.3\%$ , which is usually the lower limit required for identification (see Pauzat & Ellinger, in this volume).

As a consequence of the energy differences between isomers, only the most stable species in each class have a chance of being detected: HP=C=O and CH<sub>3</sub>P=O. After calibration on the known spectra of isovalent HN=C=O (Hocking et al. 1969) and CH<sub>3</sub>N=O (Coffey et al. 1968) one obtains the rotational constants of Table 3.

These values should be accurate within the 0.3% limit necessary to initiate observations in space or in the laboratory. Before such an investigation can be considered, another quantity has to be taken into account, i.e., the dipole moment of each molecule since the intensity of the radio spectra scales with the square of it. Unfortunately, the value of the dipole moment of HP=C=O ( $\mu = 0.8$  Debye), may not be large enough to allow detection in space, contrary to CH<sub>3</sub>P=O ( $\mu = 3.5$  Debye).

### 3.3 *The elementary bricks with a peptide bond*

The problem here is not the calculation of the spectral signatures of simple prebiotic species (most of them are well established in the laboratory), but to understand why some are observed and some are not.

The simplest species containing the peptide bond, i.e. the N-C=O link-

**Table 4.** Relative energies (kcal/mol) of the [C,3H,O,N] and [C,3H,S,N] optimized structures including zero-point vibrational energies. The calculations are performed at the CCSD(T)/cc-pVQZ level of theory. Left entry: B3LYP/6-311G\*\* optimized geometries; Right entry: MP2/6-311G\*\* optimized geometries.

Isomer	Energy	Isomer	Energy
NH <sub>2</sub> CH=O	0.0/0.0	NH <sub>2</sub> CH=S	0.0/0.0
HNC(H)OH	10.9/10.8	HNC(H)OH	9.0/8.9
H <sub>2</sub> C=NOH	54.7/55.2	H <sub>2</sub> C=NSH	22.6/22.4
CH <sub>3</sub> NO	61.6/61.5	CH <sub>3</sub> NS	38.0/37.9
H <sub>2</sub> C=N(H)O	62.1/62.4	H <sub>2</sub> C=N(H)S	30.3/30.4
<i>cyc</i> -H <sub>2</sub> CONH	70.1/69.9	<i>cyc</i> -H <sub>2</sub> CSNH	31.1/31.0

age is formamide, NH<sub>2</sub>CH=O. In the violent conditions of Miller experiments, all the isomers that can be formed with the six-atom system [C, 3H, O, N] are a-priori possible. Five of them have a linear backbone, the last one is a cyclic molecule. Theoretical calculations are able to give the geometries and energies, which makes it possible to compare their relative stabilities.

The numerical simulations show (Chiaromello et al. 2005), without any ambiguity, that the most stable system is formamide. The next isomer on the energy scale, HN=CH-OH, has the same N-C-O connectivity; it is the corresponding enolic tautomer, located  $\sim 11$  kcal/mol higher in energy. The third isomer, H<sub>2</sub>C=N-OH is formaldoxyme; it is the first structure with a C=N double bond. The next two isomers are nitrogen oxydes: H<sub>2</sub>C=NH-O, and CH<sub>3</sub>N=O, i.e. the nitron and the corresponding nitroso compound. The last isomer is a three membered ring.

Replacement of oxygen by sulphur leads to the same six stable isomers, namely, five open-chain structures and one cyclic molecule. The results are presented in Table 4 where all isomers lie now within a 2 eV energy range at the CCSD(T)/cc-pVQZ level.

The most stable system is thioformamide NH<sub>2</sub>CH=S, i.e., the molecule containing the peptide bond. As for oxygen compounds, the next isomer on the energy scale, HN=CH-SH, is the corresponding enolic form, located  $\sim 10$  kcal/mol higher in energy. The third isomer on the energy scale, a thiooxime H<sub>2</sub>C=N-SH,  $\sim 22$  kcal/mol above thio-formamide, is the first isomer with C-N-S connectivity with a double bond at nitrogen. Just above



**Table 5.** Relative energies (kcal/mol) of the neutral, protonated and ionized forms of isomers with  $C_2H_5O_2N$  chemical formula. The calculations are performed at the B3LYP/6-311G\*\* level of theory.

Neutral form	$\Delta E$	Protonated form	$\Delta E$
$HOCH_2C(NH_2)O$	12.1	$HOCH_2C(NH_2)O^+$	22.1
$H_2N-CH_2-COOH$	10.0	$CH_3OC(NH_3^+)O$	16.0
$CH_3OC(NH_2)O$	4.5	$CH_3NH_2^+COOH$	10.6
$CH_3NHCOOH$	0.0	$NH_3^+CH_2COOH$	0.0

( $\sim 8$  kcal/mol) is the N-sulfide. The last isomer with this connectivity, analogous to nitroso compounds, is the highest in energy. With sulphur in place of oxygen, the cyclic molecule,  $\sim 31$  kcal/mol above thioformamide, is no longer the less stable isomer in contrast with the [C,3H,O,N] series.

It is remarkable that the two isomers with C-N-O and C-N-S connectivities, that is those corresponding to the atomic sequence of peptidic structures, are by far the most stable. It is also worth noting that only the most stable, isomer has been identified in the formamide family. At the same time, this observation provides a test of the numerical simulations. The theoretical rotational constants for  $NH_2CH=O$  are:  $B_{Theory} = 11.3321$ ;  $C_{Theory} = 9.8246$  when compared with  $B_{Observed} = 11.3738$ ;  $C_{Observed} = 9.8337$  show the desired accuracy ( $\Delta B \sim 0.3\%$ ;  $\Delta C \sim 0.1\%$ ). The quality of the simulations gives confidence in the predictions of  $B_{Theory} = 6.0342$ ;  $C_{Theory} = 5.4921$  for the still to be detected  $NH_2CH=S$  molecule.

### 3.4 The glycine problem

On the one hand, complex amino acids are found in the carbonaceous chondrites collected on Earth; on the other hand, no amino acid has ever been observed in the interstellar medium, even the simplest representative, glycine:  $NH_2CH_2COOH$ . A possible answer to this surprising fact may be tentatively searched in the relative stability of glycine compared with the other molecules that can be formed from the same set of atoms.

A systematic investigation of the molecules of prebiotic interest with chemical formula  $C_2H_5O_2N$  led to the relative energies reported in Table 5. From this study, it can be seen that the amino acid structure, glycine, is not the most stable compound that can be formed; it is 10 kcal/mol (5000 K) higher in energy. Such a difference is presumably the reason why the

simplest of the amino acids cannot be detected: it is not formed, at least, as a neutral compound.

Since interstellar chemistry is known to rely on ion-molecule more than on neutral-neutral reactions, a parallel investigation was carried out on the protonated forms. In this case, the ion derived from glycine is the most stable structure (Table 5). It is 10 kcal/mol below  $\text{CH}_3\text{NH}_2^+\text{COOH}$ , namely the protonated derivative of the molecule that was the most stable in the neutral form.

The present study suggests that the search for glycine could be preferably reoriented to the search for its protonated ion.

#### 4 Conclusion

Different applications of Quantum Chemistry that are pertinent to the hunt for molecules in space have been addressed in this presentation. The focus on molecules at the origin of life is an illustration of the versatility of numerical simulations using wave functions or electronic density based methods.

The Adenine formation starting from HCN is an example where numerical simulations demonstrate that extrapolation of a well-known chemical scheme in solution to the interstellar medium leads to erroneous conclusions. It is clear that Adenine cannot be formed by neutral-neutral reactions, and has to be searched in close relation with solid chemistry.

The second example is a prospective study aimed at finding a tracer of molecular phosphorus in space. The basic idea is to look at the simplest molecules built from the most widespread atoms that could incorporate the P-C bond and, at the same time, be accessible to radio observations; HPCO and  $\text{CH}_3\text{PO}$  are found promising candidates whose rotational constants have been determined within an estimated 0.3 % error bar.

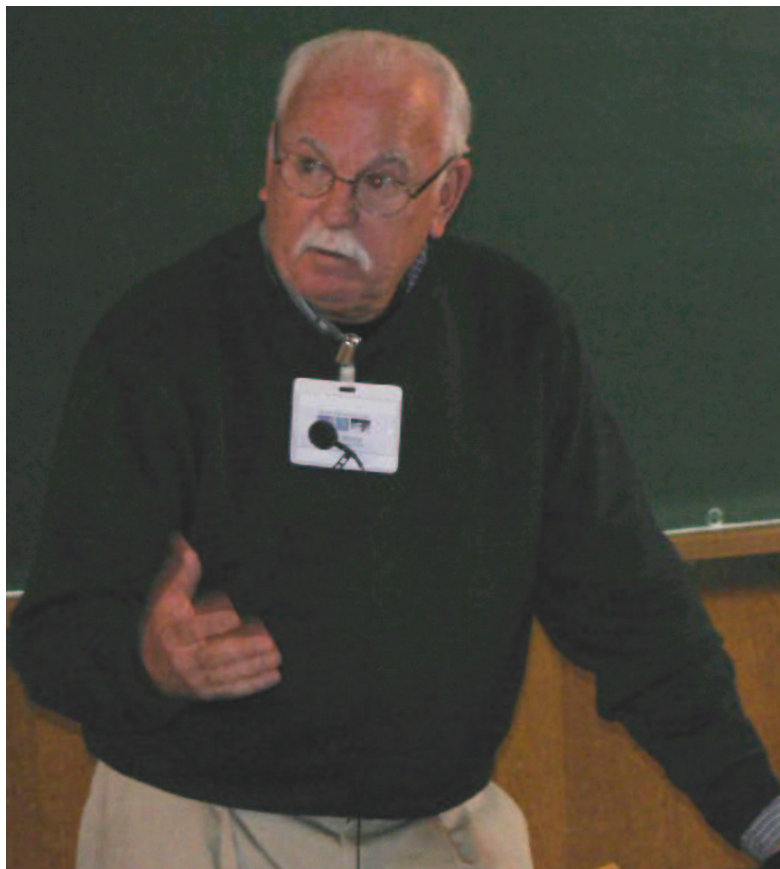
The last two examples are concerned with the observation/non observation of prebiotic species. The numerical simulations show that formamide,  $\text{H}_2\text{N-HC=O}$ , which is the only observed isomer of  $\text{CH}_3\text{ON}$  chemical formula is the lowest energy structure. The same argument of relative stability provides a rationalization of the non observation of Glycine,  $\text{CH}_3\text{NH}_2\text{COOH}$  in the interstellar medium. It also suggests that the corresponding protonated ion should be a much better candidate.

*The hunt for molecules is not over.*

This research was supported by CNRS national programs PNP, PCMI and GDR Exobiology. The calculations presented in this contribution were financed by national supercomputing facilities IDRIS and CINES.

## References

- Bockelee D., Lis D. C., Wink J. E., Despois D., Crovisier J, et al. (2000). *Astron. Astrophys.*, 353, 1101.
- Brown, *Astrophys.J.* 1981, 248, L119.
- Chakrabarti, S. & Chakrabarti, S. K. 2000, *Astron. Astrophys.*, 354, L6.
- Chentit, M., Sidorenkova, H., Ellinger, Y. & Geoffroy, M. 1998, *J. Phys. Chem.*, A102,10469.
- Chiaramello, J.M., Talbi, D., Berthier, G. & Ellinger, Y. 2005, *Int. J. Astrobiology*, 4, 125.
- Coffey, D., Britt, C.O. & Boggs, J.E. 1968, *J. Chem. Phys.*, 49, 591.
- Cooper, G.W., Onwo, W.M. & Cronin, J.R. 1992, *Geochim. Cosmochim. Acta*, 56, 4109.
- Devienne, F.M., Barnab., C., Couderc, M. & Ourisson, G. 2000, *C.R. Acad. Sci.*, IIc 3, 345.
- Dimur, C., Pauzat, F., Ellinger, Y. & Berthier, G. 2001, *Spectrochimica Acta*, 57A , 859.
- Haldane, J.B. 1929, *Rationalist Annual*, The origin of Life,148.
- Hocking, W.H., Gerry, M.C.L. & Winnewisser, M. 1969, *Can. J. Phys.*, 53, 1975.
- Lis, D.C., Keene J., Young, K., Phillips, T.G., Bockelee, D., Crovisier, J., Schilke, P., Goldsmith P.F. & Bergin, E.A. 1997, *Icarus*, 130, 355.
- Miller, S. 1953, *J. Amer. Chem. Soc.*, 77, 2351.
- Miller S., & Urey, H.C. 1959, *Science*, 130, 245.
- Oro, J. 1961, *Nature*, 191, 1193
- Ponnamperuma, C. 1979, *Chem. Br.*, 15, 560.
- Rubin R.H., Swenson G.W.Jr., Benson R.C., Tigelaar H.L. & Flygare W.H. (1971). *Astrophys. J.*, 169, L39.
- Smith, I. W. M., Talbi, D. & Herbst, E. 2001, *Astron. Astrophys.*, 369, 611.
- Snyder, L.E., Buhl, B. & Edrich, C.H. 1972, *Astrophys.J.*, 177, 625.
- Takano Y., Ohashi A., Kaneko T., & Kobayashi K. (2004). *Applied Phys. Letters*, 84, 90.



Sam Gulkis