Density Functional Theory Calculations on the Interstellar Formation of Biomolecules Postprint

Authors: Qingli Liao, Junzhi Wang, Peng Xie, Enwei Liang and Zhao Wang

Date: 2024-01-09T00:00:00+00:00

Abstract

Density functional theory (DFT) is the most versatile electronic structure method used in quantum chemical calculations, and is increasingly applied in astrochemical research. This mini-review provides an overview of the applications of DFT calculations in understanding the chemistry that occurs in star-forming regions. We survey investigations into the formation of biologically relevant compounds such as nucleobases in the interstellar medium, and also cover the formation of both achiral and chiral amino acids, as well as biologically relevant molecules such as sugars, and nitrogen-containing polycyclic aromatic hydrocarbons. Additionally, DFT calculations are used to estimate the potential barriers for chemical reactions in astronomical environments. We conclude by noting several areas that require more research, such as the formation pathways of chiral amino acids, complex sugars, and other biologically important molecules, and the role of environmental factors in the formation of interstellar biomolecules.

Full Text

Preamble

Research in Astronomy and Astrophysics, 23:122001 (12pp), 2023 December

https://doi.org/10.1088/1674-4527/ad013c

Density Functional Theory Calculations on the Interstellar Formation of Biomolecules

Qingli Liao[1], Junzhi Wang[1], Peng Xie[2], Enwei Liang[1], and Zhao Wang[1]
[1] Laboratory for Relativistic Astrophysics, Department of Physics, Guangxi University, Nanning 530004, China; zw@gxu.edu.cn

[2] School of Chemistry and Chemical Engineering, Guangxi University, Nanning 530004, China

Received 2023 July 6; accepted 2023 October 7; published 2023 November 15

Abstract

Density functional theory (DFT) represents the most versatile electronic structure method employed in quantum chemical calculations and is increasingly applied in astrochemical research. This mini-review provides a comprehensive overview of DFT calculations used to understand chemistry occurring in star-forming regions. We survey investigations into the formation of biologically relevant compounds such as nucleobases in the interstellar medium, covering both achiral and chiral amino acids, sugars, and nitrogen-containing polycyclic aromatic hydrocarbons. Additionally, DFT calculations are utilized to estimate potential barriers for chemical reactions in astronomical environments. We conclude by identifying several areas requiring further research, including formation pathways of chiral amino acids, complex sugars, and other biologically important molecules, as well as the role of environmental factors in interstellar biomolecule formation.

Key words: astrochemistry –molecular processes –methods: laboratory: molecular –astrobiology

1.1. Astronomical Detection

A plethora of biologically relevant compounds, including amino acids, nucleobases, and sugar derivatives, have been identified in meteoritic, cometary, and asteroidal samples (Burton et al. 2012). Additionally, numerous biologically relevant molecules have been observed in the interstellar medium (ISM) (Guelin & Cernicharo 2022). These findings strongly suggest an extraterrestrial origin for life' s elemental building blocks and raise intriguing questions about the formation of these molecules in space and their delivery to planets, as depicted in Figure 1 [Figure 1: see original paper]. Addressing these questions is crucial for understanding abiogenesis and the search for life throughout the Universe. To date, over 200 molecular species have been detected in interstellar and circumstellar environments (McGuire 2022). Most of these molecules are organic, reflecting the cosmic abundance of H, C, N, and O. Although the majority are abiotic, reactions among them present a vast array of potential complex organic molecules (COMs) that could be utilized by life.

Despite the long-held belief that space environments are hostile to COM formation and survival, advances in astronomical spectroscopy have provided evidence to the contrary (López-Sepulcre et al. 2019; Li 2020; Li et al. 2021; Lei et al. 2022). For instance, COMs have been detected in various interstellar regions exhibiting high degrees of organic complexity, including formamide (H_2NCHO) and other amides identified in Sagittarius B2 and Orion KL (Rubin et al. 1971;

Suzuki et al. 2018). These molecules contain amino groups analogous to those in peptide chains and are considered precursors of life' s genetic material. Another class of astronomically detected compounds associated with life is nitriles—molecules containing one or more nitrile functional groups that serve as raw materials for RNA formation. Acrylonitrile (HCCHCN) exemplifies such a nitrile, detected in a massive star-forming region (Suzuki et al. 2018).

COM detection presents greater challenges than simpler molecules like CO, CS, and HCN. Snyder et al. (2005) summarized the criteria for COM detection, considering factors such as low abundance, complex energy level structures, and line confusion caused by similar rest frequencies. Among COMs, amino acids are particularly relevant to the origin of life, though reliable detection remains elusive. The simplest amino acid, glycine, has been reported in several sources (Kuan et al. 2003), but subsequent analysis indicated these lines do not actually originate from glycine (Snyder et al. 2005). Significant recent discoveries include peptide-bond-containing molecules such as formamide (Rubin et al. 1971), acetamide (Hollis et al. 2006), N-methylformamide (Belloche et al. 2019), and propionamide (Li et al. 2021). Chiral molecules like propylene oxide (McGuire et al. 2016), sugar-type molecules like ethylene glycol (Hollis et al. 2002), and branched alkyl molecules like isopropyl cyanide (Belloche et al. 2014) have also been detected, marking important milestones in astrochemistry. However, fundamental questions regarding the formation of these COMs in space, their resilience to radiation, and their transport mechanisms to planets remain largely unexplored. Answering these questions holds potential for unveiling life' s origins, discovering extraterrestrial life, and identifying habitable environments for humanity.

1.2. Chemistry in the ISM

Following the detection of biomolecule precursors in space, identifying reactive pathways has become a key research question (Aponte et al. 2017; Barone & Puzzarini 2022; Rimola et al. 2022). Due to molecular complexity, biomolecules are generally difficult to form through direct collisions between closed-shell molecules in the cold ISM environment. Photochemical and radiation chemical processes are therefore considered primary steps in interstellar COM synthesis (Hollenbach & Tielens 1999; Castellanos et al. 2018; Arumainayagam et al. 2019). Laboratory studies of astronomical ice analogs suggest complex chemistry occurs on ice surfaces exposed to radiation in star-forming regions (Materese et al. 2021). Ice mantles concentrate molecules while radiation from young stellar objects can excite them, leading to barrierless reactions, as demonstrated in Figure 2 [Figure 2: see original paper]. However, generating meaningful quantities of biomolecules on ice mantles remains challenging due to short excited-state lifetimes and photoinstability of intermediates and products.

During molecular cloud collapse to form stars, the temperature of gas surrounding young stellar objects increases passively due to protostellar radiation. When temperatures become sufficiently high to evaporate ice mantles, warm molecular gas is observed at 100-300 K, known as hot cores. These regions exhibit rich organic chemistry, with most observed O-, N-, and S-containing COMs detected in regions like Sagittarius B2 and Orion KL (Herbst & van Dishoeck 2009; Herbst 2014). Additionally, protoplanetary disks—where planets form—have become a focus of abiogenesis research. Although only relatively simple organics have been detected in protoplanetary disks (Oberg et al. 2020), gas temperatures can reach levels comparable to hot cores, increasing from 10 K to several thousand K from the outer layers to the central forming star (Akimkin et al. 2013). In hot cores and protoplanetary disks, molecules at relatively high densities (typically 10^7-10^9 cm^{-3}) can effectively screen ultraviolet radiation and provide the kinetic energy necessary to overcome barriers in photochemical and radiation chemical processes.

Gas-phase COM chemistry in the ISM is less studied experimentally than ice chemistry (Fuente et al. 2019; Herbst 2021; Puzzarini 2022). Debate continues regarding the primary reaction mechanisms leading to their formation. Most COMs with direct biological interest are complex molecules ($ $10 atoms), and the energy barriers for their synthesis from closed-shell species are typically very high, requiring elevated temperatures for activation. Consequently, previous studies have predominantly focused on ion-molecule reactions or neutral-neutral reactions involving highly reactive radicals (Sandford et al. 2020). However, this viewpoint does not fully align with the short lifetimes of ions and radicals or with astronomical observations showing that most detected interstellar molecules are neutral closed-shell species (McGuire 2022).

To address this, researchers have explored catalyzed reactions as an alternative approach for synthesizing complex interstellar molecules. However, the limited abundance of suggested catalysts often renders catalyzed reactions seemingly impossible in the ISM, as they typically require simultaneous three-body collisions (Mendoza et al. 2004; Roy et al. 2007; Rimola et al. 2012; Saladino et al. 2013; Vinogradoff et al. 2015; da Silva & de Araujo 2017; Qi et al. 2018; Potapov et al. 2019; Hanine et al. 2020). Fortunately, this is not always the case, particularly when catalyst concentrations significantly exceed those of reactants, as with neutral atomic hydrogen (H I). Recent discoveries show that H I leads to alternative pathways that substantially reduce energy barriers for complex reactions, significantly enhancing reaction rates (Yang et al. 2023). Consequently, these reactions become thermodynamically feasible in certain ISM regions. This revolutionary finding suggests H I-catalyzed reactions could represent a significant mechanism for interstellar COM formation from closed-shell species.

1.3. Focus of this Review

Identifying intermediates in interstellar COM formation pathways is essential for understanding their formation mechanisms. However, many intermediates are highly reactive and short-lived, making laboratory detection difficult. To address this, chemical networks have been developed to model the nature and abundance of interstellar molecules, incorporating spectroscopic observations and laboratory experiments (Quan et al. 2010; Herbst & Yates 2013; Zhang et al. 2020a, 2020b; Unsleber & Reiher 2020; Zhao et al. 2021). While successful for many simple interstellar molecules, this approach has not fully explained observed COM abundances. A viable solution involves performing quantum chemical calculations (QCCs), where molecular geometry optimization via ab initio electronic structure methods can determine intermediate structures.

This concise review provides an overview of dedicated QCCs employing density functional theory (DFT) to model the synthesis of biologically relevant COMs in the ISM, with primary emphasis on gas-phase chemistry. Previous reviews by Zamirri et al. (2019) covered interstellar COM formation focusing on ice surface reactions, while Sandford et al. (2020) and Jørgensen et al. (2020) provided comprehensive reviews of interstellar biomolecule formation studies, primarily concerning laboratory experiments.

2. Density Functional Theory

QCCs have been utilized to assess electronic contributions to molecular physical and chemical properties (Mata & Suhm 2017; Sumiya et al. 2022). This involves systematic application of approximations to the electronic Schrödinger equation to obtain electron densities and energies related to nuclear positions and electron number. QCCs are classified according to the approximation type used to solve the Schrödinger equation, with the most common approaches being DFT, Hartree-Fock (HF), and their extensions.

The HF method is the simplest wavefunction-based approach that neglects many-electron correlations, limiting its accuracy in describing chemical bond formation. Post-HF methods such as Møller-Plesset second-order perturbation (MP2) (Møller & Plesset 1934) and coupled cluster with single, double, and perturbative triple excitations (CCSD(T)) (Gyevi-Nagy et al. 2020) improve the wavefunction to recover missing electron correlation, providing increased accuracy at reduced computational efficiency.

DFT is now widely regarded as the most versatile electronic structure method in QCCs due to its high computational efficiency (Jones 2015). In Kohn-Sham theory, DFT reduces the many-body problem of interacting electrons in an external field to non-interacting electrons moving in an effective potential that includes exchange and correlation (XC) interactions. Despite early limitations, DFT accuracy for QCCs has improved significantly since the 1990s following refinement of XC interaction approximations.

The Local Density Approximation (LDA) in DFT assumes the XC energy functional depends solely on electron density at each spatial point, leading to underestimation of exchange energy and overestimation of correlation energy. To address this, Generalized Gradient Approximations (GGAs) incorporate higher-order electron density derivatives. Despite widespread use, GGA-level DFT has difficulty accurately describing intermolecular interactions such as van der Waals dispersion, charge transfer excitation, and transition states.

Difficulties in describing exchange interactions can be addressed by incorporating contributions from HF theory, known as hybrid functionals. The popular B3LYP functional exemplifies this approach—a three-parameter, Lee-Yang-Parr functional based on a linear combination of HF exact exchange and GGA. An illustrative application to abiogenesis involves determining the abiotic synthesis pathway of adenine (Ade) from HCN. While experiments since 1960 demonstrated Ade synthesis from HCN and ammonia (NH_3) solutions (Oró 1960), possible reaction pathways remained unknown until Roy et al. (2007) performed B3LYP-level DFT calculations.

Meta-hybrid GGA functionals, such as the Minnesota family, may offer greater accuracy than B3LYP (Hohenstein et al. 2008; Riley et al. 2010; Peverati & Truhlar 2011; Ferrighi et al. 2012). This family includes M06-L, M06, M06-2X, and M06-HF functionals, each with varying exact exchange contributions. Their expansion terms depend on electron density, density gradient, and density second derivatives, demonstrating success for systems with dispersion forces and correcting prominent standard DFT deficiencies.

DFT accuracy depends on XC functional choice and electronic wavefunction representation (the basis set), which can comprise atomic orbitals or plane waves—the former being common in quantum chemistry. Prevalent DFT functional/basis set combinations for astrochemistry include B3LYP/6-31G*, B3LYP/6-311G(d,p), B3LYP/aug-cc-pVTZ, M06-2X/6-31G, M06-2X/6-311G(d,p), and M06-2X/aug-cc-pVTZ, along with their derivatives.

Deficiencies exist in the popular B3LYP/6-31G* model, including missing London dispersion and basis set superposition error, as discussed by Kruse et al. (2012) regarding DFT molecular thermochemistry calculations. Additionally, Mardirossian & Head-Gordon (2017) reviewed different DFT levels, benchmarking approximately 200 density functionals on a molecular database of nearly 5000 data points for non-covalent interactions, isomerization energies, thermochemistry, and barrier heights.

Although DFT is a valuable tool, it possesses certain limitations (Cohen et al. 2012). A significant challenge involves accurately describing reaction barriers and energetics. For certain systems, DFT often struggles to predict activation energies precisely, especially for reactions with complex transition states or substantial molecular rearrangements. This limitation arises from inherent exchange-correlation functional approximations, leading to energy barrier errors. Additionally, handling solvent effects can be challenging, though this is less relevant in the low-density ISM environment. Accurate solvent modeling may require alternative methods such as implicit solvent models, explicit solvent molecules, or hybrid quantum mechanics/molecular mechanics approaches (Senn & Thiel 2009; Acevedo & Jorgensen 2010). Moreover, DFT encounters difficulties with large systems. Despite being more computationally efficient than most wavefunction-based methods, DFT calculations become demanding for larger molecular systems, such as mixtures containing more than ten molecules. In such cases, molecular dynamics simulations may be more suitable than QCCs

(Wang 2019, 2020, 2023). Nevertheless, by carefully considering limitations and implementing appropriate strategies, DFT remains a powerful tool for studying interstellar chemical reactions (Jones 2015).

Accurate computation of electron density and free energy has made DFT valuable for exploring reactive free-energy surfaces. Through DFT structural optimization, one can predict intermediates and products constituting reaction pathways. This process involves updating nuclear positions in colliding molecules toward the potential energy minimum until optimal geometries are found at stationary points on the system’ s potential energy surface (PES), where net interatomic forces on each atom are acceptably close to zero (Althorpe & Clary 2003). To confirm intermediates and transition states, vibrational frequency calculations are typically performed. Additionally, intrinsic reaction coordinate (IRC) calculations can ensure transition states reasonably connect reactants and products.

The following section surveys DFT studies on interstellar synthesis of biologically relevant COMs, organized by biomolecule category: nucleobases, amino acids, sugars, and other species.

3. Nucleobases

Nucleobases are nitrogen-containing compounds constituting the basic components of nucleic acids, comprising five primary nucleobases: adenine (Ade, $C_5H_5N_5$), cytosine (Cyt, $C_4H_5N_3O$), guanine (Gua, $C_5H_5N_5O$), thymine (Thy, $C_5H_6N_2O_2$), and uracil (Ura, $C_4H_4N_2O_2$). The three pyrimidine bases (Cyt, Thy, Ura) contain single-ring structures, while the two purine bases (Ade, Gua) have fused-ring configurations, as illustrated in Figure 3 [Figure 3: see original paper]. These five primary nucleobases serve as fundamental genetic code units, with Ade, Gua, Cyt, and Thy present in deoxyribonucleic acid (DNA) and Ade, Gua, Cyt, and Ura present in ribonucleic acid (RNA). While direct ISM observations of primary nucleobases remain elusive, various adenine precursors have been identified, including aminoacetonitrile (H_2NCH_2CN) (Belloche et al. 2008), carbodiimide (HNCNH) (McGuire et al. 2012), cyanamide (NH_2CN) (Turner et al. 1975), cyanomethanimine (HNCHCN) (Zeng et al. 2019), and glycolonitrile ($HOCH_2CN$) (Zaleski et al. 2013). Additionally, hydroxylamine (NH_2OH) is a known precursor for pyrimidines and purines (Rivilla et al. 2020), while ethanimine (CH_3CHNH) may play a role in amino acid synthesis (Loomis et al. 2013).

3.1. Pyrimidine Bases

Most studies on interstellar pyrimidine nucleobase formation focus on routes from pyrimidine ($C_4H_4N_2$) due to its direct chemical relevance to nucleobase structures. In astronomical contexts, pyrimidine is known as a nucleobase precursor in interstellar ice analogs. Using laboratory simulations, Nuevo et al. (2009) observed uracil formation in residues generated from UV photolysis of pyrimidine in pure water ices. To explore reaction mechanisms, B3LYP-level DFT was employed to aid identification of key intermediates (Bera et al. 2010). Oba et al. (2019) reported pyrimidine and purine nucleobases in interstellar ice analogs composed of H_2O, CO, NH_3, and CH_3OH after UV irradiation followed by thermal processing. Laboratory experiments demonstrate that UV photoprocessing of pyrimidine and purine in simple astrophysical ices can result in formation of all five primary nucleobases (Materese et al. 2018).

Cole et al. (2015) utilized an ion trap mass spectrometer with B3LYP-level DFT calculations to reveal connections between pyrimidine nucleobase anions and cyanate (OCN^-). Similarly, Gupta et al. (2013) employed the same theoretical method to propose two exothermic cytosine formation pathways beginning from propynylidyne (CCCH) and cyanoacetylene (HCCCN). Choe (2021) used DFT calculations to investigate synthesis pathways toward Cyt, Ura, and Thy on icy grain mantles through reactions between HCCCN, protonated isocyanic acid (H_2NCO^+), and NH_3, H_2O, or CH_3OH. Furthermore, Choe (2020) proposed a cytosine formation route from urea reacting with protonated cyanoacetylene (CAH^+) and H_2O.

Pyrimidine has been found unstable under UV radiation (Peeters et al. 2005), prompting consideration of alternative formation routes. Formamide (H_2NCHO) has been proposed as a reactant (Rotelli et al. 2016; Saladino et al. 2016), with dehydration products from its polymerization shown to lead to nucleobase formation (Nguyen et al. 2015; Jeilani et al. 2016). B3LYP/6-311++G(d,p) DFT studies suggested that reactions of free radicals such as CCCNH or CCCO may lead to pyrimidine base formation (Wang & Bowie 2012).

Using M06/6-31+G(d,p)/6-311++G(d,p)-level DFT calculations, Lu et al. (2021) investigated gas-phase reactions between partially dehydrogenated formamide and vinyl cyanide (H_2CCHCN) for synthesizing Cyt, Thy, and Ura via 1H-pyrimidin-2-one ($C_4H_4N_2O$), which acts as a direct nucleobase precursor. Their results indicate H migration is the rate-determining step in pyrimidine base synthesis. However, highly dehydrogenated free radicals may be unstable in interstellar molecular clouds due to high chemical reactivity and consequent short lifetimes. Moreover, Komatsu & Suzuki (2022) employed an automated path search approach based on UB3LYP and UM06-2X-level DFT to study cytosine synthesis from NH_2CCHO.

3.2. Purine Bases

Purine base (Ade and Gua) formation is more complex than pyrimidine base formation. Laboratory experiments have employed photochemical reactions of purine in interstellar ice analogs to synthesize Ade, Gua, and their derivatives (Materese et al. 2017, 2018), with 2-aminopurine and isoguanine hypothesized as key intermediates. B3LYP/cc-pVDZ DFT calculations modeled this reaction (Bera et al. 2017), suggesting a multistep mechanism involving purine cations, hydroxyl and amino radicals, water, and ammonia, with ice surface effects mimicked using a conductor solvent model.

The HCN pentamerization pathway has been extensively studied. Gupta et al. (2011) used B3LYP/6-31G** DFT to investigate Ade formation from HCN, CN, and HCCN, estimating rate coefficients up to 8.71×10^{-9} cm^3 s^{-1}. Jung & Choe (2013) applied B3LYP/6-311G(2d,d,p) DFT to simulate Ade synthesis via HCN oligomerization, reporting high reaction barriers (a few hundred kJ/mol). Cole et al. (2015) combined experiments with B3LYP/6-311++G(d,p) DFT to study purine base formation in reverse, by measuring dissociation products of Ade and Gua, revealing carbodiimide (HNCNH), cyanamide (NH_2CN), and isocyanic acid (HNCO). Merz et al. (2014) proposed an Ade synthesis pathway from CCCNH and HNCNH (and its isomer H_2NCN) using MP2 calculations.

Wang et al. (2013) proposed a formamide-to-adenine synthesis pathway using B3LYP/6-311G(d,p) DFT. Choe (2018) utilized B3LYP/6-311G(2d,d,p) DFT to suggest guanine formation pathways from 4-aminoimidazole-5-carbonitrile (AICN), with H_2O considered as a catalyst substantially decreasing activation barriers. Jeilani et al. (2018) further studied free radical routes to Ade and Gua formation using UB3LYP/6-311G(d,p) DFT, identifying two key intermediates: 5-aminoimidazole-4-carboxamide (AICA) and 5-(formylamino)imidazole-4-carboxamide (fAICA). B3LYP results were comparable to M06 data.

4. Amino Acids

Alpha-amino acids, which constitute proteins in the genetic code, have been detected in meteorites and comets, making them particularly interesting for astrobiology. All alpha-amino acids in the genetic code except glycine (Gly, $C_2H_5NO_2$) are chiral, existing as two mirror-image stereoisomers called enantiomers. Figure 4 [Figure 4: see original paper] shows alanine (Ala, $C_3H_7NO_2$) as an example, with its two enantiomers D-Ala and L-Ala possessing a typical chiral center represented by a central carbon atom bonded to four distinct groups. The origin of biological homochirality remains unresolved, particularly regarding chemical processes responsible for chiral characteristics of alpha-amino acids in astrochemical contexts. Previous investigations into interstellar amino acid synthesis have primarily focused on glycine due to its relative simplicity, with proposed mechanisms often involving Strecker-type synthesis and radical-radical or radical-neutral interactions.

4.1. Glycine

Bernstein et al. (2002) conducted experiments exploring glycine and other amino acid formation when cryogenic water ice containing small amounts of CH_3OH, NH_3, and HCN was irradiated with UV light. Woon (2002) evaluated these pathways using MP2-level QCCs with aug-cc-pVDZ basis sets, finding that isotopic substitution experiments could identify CH_3OH as the carbon source in amino acid carboxylic acid groups. Subsequently, Holtom et al. (2005) performed combined experimental and theoretical studies on glycine and its isomer formation in interstellar ices using B3LYP/6-311G(d,p) DFT, revealing that H atoms with sufficient kinetic energy could overcome entrance barriers to add to CO_2, yielding a trans-hydroxycarbonyl radical that could recombine with other radicals to form glycine and its isomer.

Similar studies were undertaken by Nhlabatsi et al. (2016) using DFT approaches, Rimola et al. (2010, 2012) using cluster approaches with B3LYP/6-31+G(d,p) DFT, and Bossa et al. (2010), Oba et al. (2015), Sato et al. (2018), Joshi & Lee (2022), Thripati (2022), and Krasnokutski et al. (2020) using DFT calculations and experiments, all proposing glycine formation pathways on interstellar ices/grains.

Gas-phase reactions leading to glycine have been less studied than surface chemistry. Pilling et al. (2011) used B3LYP DFT and MP2 calculations to propose glycine formation routes from carboxylic acids and carboxyl radicals (COOH) in both gas and solid phases, suggesting the most favorable gas-phase reactions involved acetic acid (CH_3COOH) with NH^+ or NH_2OH, or NH_2CH_2 with $COOH^+$. Singh et al. (2013) reported that simple molecules NH_2, CH_2, CH, and CO could form glycine in both gas and grain phases based on B3LYP/6-31G(d,p)-level DFT calculations.

Kayanuma et al. (2017) used B3LYP/6-31G* DFT to propose glycine synthesis pathways from aminoacetonitrile (NH_2CH_2CN) and hydantoin (2,4-imidazolidinedione, $C_3H_4N_2O_2$) through the Bücherer-Bergs reaction and hydrolysis in both gas and solid phases. Additionally, B3LYP/6-311++G(2df,p)-level DFT calculations have been used to study glycine peptide chain formation via unimolecular reactions (Comte et al. 2023).

4.2. Chiral Alpha-amino Acids

Chiral amino acid formation in the ISM is less understood than glycine formation, despite homochirality being a common property of biological amino acids. Most research has focused on relatively simple amino acids such as alanine (Ala) and serine (Ser). Elsila et al. (2007) used isotopic labeling to observe multiple reaction pathways for glycine and serine formation in interstellar ice analogs, suggesting amino acid formation is not narrowly dependent on ice composition.

Shivani et al. (2017) proposed through DFT calculations that reactions between CH_3CN, CH_3CNH_2, and HCOOH could lead to alanine formation. The same group explored serine formation in the ISM through detected interstellar molecules CH, CO, and OH via radical-radical and radical-neutral interactions in the gas phase using B3LYP/6-311G+(2df,2p) DFT (Shivani et al. 2014). Results were calibrated with higher-level M06 and M06-2X functionals, suggesting proposed reactions with low potential barriers could occur in the ISM.

Rani & Vikas (2018) studied stereoinversion pathways in L-threonine, an amino acid with two stereocenters, using M06-2X/aug-cc-pVTZ DFT. They observed that simultaneous intramolecular proton and hydrogen atom transfer drives stereoinversion, with significant reaction rates in the 500-1000 K temperature range. In a recent study, Liao et al. (2023) employed M06-2X/6-31+G(d,p) DFT to examine interstellar formation of Ala, Ser, and isovaline, focusing on key chirality-determining steps in synthesis pathways.

5. Sugars and Other Biology-relevant COMs

The presence of sugars such as ribose and xylose in meteorites (Laneuville et al. 2018; Smith et al. 2020) has prompted investigation into their formation in star-forming regions. Woods et al. (2013) used DFT calculations to examine formyl radical dimerization and its implications for glycolaldehyde ($C_2H_4O_2$) formation, the simplest sugar-related molecule. Results suggest the dimerization transition state is located within the formyl radical and that the reaction is energetically favorable under low-temperature conditions, consequently leading to glycolaldehyde formation. This finding has implications for monosaccharide, disaccharide, and polysaccharide formation, all crucial for astrochemistry and astrobiology.

Thripati & Ramabhadran (2017) investigated the role of metal ions and hydrogen bonds in the formose reaction—the first step in sugar synthesis—using B3LYP/aug-cc-pVTZ DFT. Simulations revealed metal ions serve as catalysts while hydrogen bonds stabilize reaction intermediates. Vazart et al. (2018) and Skouteris et al. (2018) studied gas-phase formation routes for glycolaldehyde, acetic acid, and formic acid (the "ethanol tree") through combined infrared spectroscopy and B2PLYP/m-aug-cc-pVTZ DFT calculations. They observed radiative association as the principal mechanism for glycolaldehyde and acetic acid formation, while gas-phase elimination is the primary route for formic acid.

Jeilani & Nguyen (2020) investigated the autocatalytic formose reaction and its potential role in RNA nucleoside formation using B3LYP/6-311G(d,p) DFT. Results indicated the autocatalytic formose reaction can form several RNA nucleosides, including cytosine and uracil. Woon (2021) used DFT calculations to investigate glycolonitrile ($HOCH_2CN$) formation from C^+ reactions with HCN and HNC on icy grain mantles, demonstrating these reactions contribute to molecule formation, with C^+ + HCN being a major contributor in interstellar regions.

Ahmad et al. (2020) presented a theoretical approach to elucidate $C_2H_4O_2$ isomer formation in the ISM through reactions of interstellar formaldehyde molecules using DFT molecular dynamics simulations. Results revealed formaldehyde reactions are exothermic and occur concertedly to produce three isomers: glycolaldehyde, methyl formate, and acetic acid.

DFT calculations have also investigated other biologically relevant COM formations in the ISM. Xie et al. (2022) proposed a new ISM pathway leading to amine-group or peptide-bond-bearing biorelevant molecules via single-photon ionization-induced Michael/cyclization reactions of acrylonitrile-alcohol heterodimer complexes in the gas phase, combining experiments and DFT calculations. Ahmadvand et al. (2014) investigated the cyclopropenone (c-H_2C_3O) formation potential energy surface using B3LYP DFT, finding the spin-allowed pathway more energetically favorable than the spin-forbidden pathway under interstellar conditions.

Nitrogen-substituted polycyclic aromatic hydrocarbons (NPAHs) exhibit strong infrared emission features and are important for astronomical observations addressing abiogenesis, as they could serve as biomolecule building blocks (Kovács et al. 2020; Meng et al. 2021, 2023). Parker and coworkers demonstrated NPAH formation from ISM precursor molecules through combined DFT calculations and experiments (Parker et al. 2015a, 2015b, 2015c). Furthermore, DFT-fitted empirical reactive potentials in molecular dynamics simulations investigated interstellar PAH formation (Qi et al. 2018; Hanine et al. 2020). Finally, Yang et al. (2022) used B3LYP/6-311++G** and 6-31+G** DFT calculations to examine PAH interactions with organic molecules for PAH-organic cluster formation. Results suggested clusters form through hydrogen bonding, with formation being more favorable when PAHs are more convex and highly functionalized.

6. Astronomical Implications

Using DFT, one can calculate free energies of reactants, intermediates, transition states, or products assuming non-interacting particles at a given thermodynamic temperature (McQuarrie & Simon 1999), such as the average temperature of a specific star-forming region phase. These calculations determine potential barriers and assess proposed reaction feasibility in given astronomical environments. Two common barrier estimation strategies exist, each yielding different values. One approach assumes energy does not leave the system via radiation during atomic rearrangements, while the other considers energy dissipation after each pathway step. Under the former assumption, the barrier is defined as the free energy difference between the highest-energy state and the reactant complex, while in the latter it is defined as the energy needed to overcome the most costly step. Although reality lies between these idealized scenarios, the former approach is typically used for gas-phase interstellar reactions because atomic rearrangement is usually faster than vibrational cascades.

The potential barrier is crucial for determining reaction feasibility in the ISM. As temperature increases, molecules collide more frequently and with greater kinetic energy. The proportion of collisions surpassing the barrier increases with temperature following the Arrhenius relation. The relationship between reaction rate constant k, reactant half-life $t_1/_2$, temperature, and barrier ΔG is described by transition state theory (Eyring 1935):

$$k = \kappa \frac{k_B T}{h} e^{-\Delta G^{\ddagger}/RT}$$

$$t_{1/2} = \frac{\ln 2}{k}$$

where is the transmission coefficient, k is the Boltzmann constant, T is absolute temperature, h is Planck's constant, and R is the molar gas constant. Figure 5 [Figure 5: see original paper] shows that to achieve a half-life shorter than 10^5 years—a typical astrochemical timescale highlighted by the green horizontal line in the lower panel—the reaction barrier must be lower than approximately 35 kcal mol^{-1} at 300 K. This limit roughly corresponds to a rate coefficient greater than 10^{-13} cm^3 s^{-1}, as highlighted by the horizontal line in the upper panel. Note that the considered reactions are two-body collision processes or catalyzed binary reactions where the catalyst concentration greatly exceeds those of the two reactants (so-called pseudo-binary reactions).

According to Figure 5, a potential barrier lower than approximately 35 kcal mol^{-1} signifies thermal feasibility in formation regions of young stellar objects, which are often associated with warm ambient gas at average temperatures up to 300 K, commonly referred to as hot cores or corinos. These regions typically exhibit rich organic chemistry and contain initial materials required for organic compound formation (Herbst & van Dishoeck 2009). A common mistake to avoid is attempting to determine activation temperature by direct conversion of 35 kcal mol^{-1} to its temperature equivalent (yielding approximately 17,613 K), which overlooks that molecules react via collisions with Boltzmann-distributed energies.

Furthermore, protoplanetary disks are particularly interesting for abiogenesis studies since planetary chemical compositions are believed to be inherited from the disk itself. Disk temperatures gradually increase from 10 K to several thousand K from outer layers to the central forming star (Akimkin et al. 2013). Table 1 summarizes DFT-calculated potential barriers for previously proposed reactions synthesizing biologically relevant species in the ISM. For complex biomolecule synthesis from closed-shell species, potential barriers ΔG are typically greater than 50 kcal mol^{-1}. However, reactions involving at least one transient species (atoms, radicals, ions) have much lower barriers and are feasible at star-forming region temperatures. Alternatively, catalyzed reactions have been proposed for interstellar COM formation. Catalysts can facilitate difficult reactions between closed-shell species, enabling them to proceed under laboratory conditions. Various free radicals, ions, acids, and substrates have been suggested as catalysts for interstellar COM synthesis (Mendoza et al. 2004; Roy et al. 2007; Rimola et al. 2012; Jeilani et al. 2013; Saladino et al. 2013; Vinogradoff et al. 2015; Jeilani et al. 2016; da Silva & de Araujo 2017; Qi et al. 2018; Potapov et al. 2019; Hanine et al. 2020). Notably, neutral atomic hydrogen (H I) has recently been proposed as an efficient catalyst for interstellar adenine and guanine synthesis (Yang et al. 2023). As the most abundant free radical in the Universe, H I likely acts as a general catalyst for interstellar organic formation.

7. Conclusion and Perspective

This mini-review delves into DFT applications in astrochemistry, particularly for studying complex interstellar biomolecule formation. DFT calculations have proven valuable for elucidating formation mechanisms of biologically relevant compounds, either guiding experimental investigations or predicting reaction pathways. Despite these advancements, numerous areas remain unexplored. From our perspective, future research should encompass: exploration of complex purine base formation; elucidation of chiral amino acid synthesis pathways; comprehension of intricate sugar and biomolecule production; examination of metal ion and hydrogen bonding influences in ISM chemical reactions; refinement of computational models; and assessment of temperature and environmental variable impacts on interstellar biomolecule formation. Furthermore, integration of machine learning techniques offers intriguing possibilities.

Acknowledgments

The authors acknowledge financial support from the National Natural Science Foundation of China (NSFC, grant Nos. 11964002 and 22168002) and the Natural Science Foundation of Guangxi Province (2020GXNSFAA159119).

References

Acevedo, O., & Jorgensen, W. L. 2010, ACR, 43, 142
Ahmad, A., Shivani, Misra, A., & Tandon, P. 2020, RAA, 20, 14
Ahmadvand, S., Zaari, R. R., & Varganov, S. A. 2014, ApJ, 795, 173
Akimkin, V., Zhukovska, S., Wiebe, D., et al. 2013, ApJ, 766, 8
Althorpe, S. C., & Clary, D. C. 2003, ARPC, 54, 493
Aponte, J. C., Elsila, J. E., Glavin, D. P., et al. 2017, ESC, 1, 1
Arumainayagam, C. R., Garrod, R. T., Boyer, M. C., et al. 2019, ChSoRv, 48, 2293
Barone, V., & Puzzarini, C. 2022, FrASS, 8, 814384
Belloche, A., Menten, K. M., Comito, C., et al. 2008, A&A, 482, 179
Belloche, A., Garrod, R. T., Muller, H. S. P., et al. 2019, A&A, 628, A10
Belloche, A., Garrod, R. T., Muller, H. S. P., et al. 2014, Sci, 345, 1584
Benallou, A. 2019, JTUS, 13, 105
Bera, P. P., Nuevo, M., Milam, S. N., et al. 2010, JChPh, 133, 104303
Bera, P. P., Stein, T., Head-Gordon, M., et al. 2017, AsBio, 17, 771
Bernstein, M. P., Dworkin, J. P., Sandford, S. A., et al. 2002, Natur, 416, 401
Bossa, J. B., Borget, F., Duvernay, F., et al. 2010, JPOC, 23, 333
Burton, A. S., Stern, J. C., Elsila, J. E., et al. 2012, ChSocRev, 41, 5459
Castellanos, P., Candian, A., Zhen, J., et al. 2018, A&A, 616, A166
Choe, J. C. 2018, CPL, 708, 71
Choe, J. C. 2020, ApJ, 898, 13
Choe, J. C. 2021, ApJ, 914, 136
Cohen, A. J., Mori-Sanchez, P., & Yang, W. 2012, ChRv, 112, 289

Cole, C. A., Wang, Z. C., Snow, T. P., et al. 2015, ApJ, 812, 77
Comte, D., Lavy, L., Bertier, P., et al. 2023, JPCA, 127, 775
da Silva, J. B. P., & de Araujo, A. P. M. 2017, ESC, 1, 376
Elsila, J. E., Dworkin, J. P., Bernstein, M. P., et al. 2007, ApJ, 660, 911
Eyring, H. 1935, JChPh, 3, 107
Ferrighi, L., Pan, Y. X., Gronbeck, H., et al. 2012, JPCC, 116, 7374
Fuente, A., Navarro, D. G., Caselli, P., et al. 2019, A&A, 624, A105
Guelin, M., & Cernicharo, J. 2022, FrASS, 9, 787567
Gupta, V. P., Tandon, P., & Mishra, P. 2013, AdSpR, 51, 797
Gupta, V. P., Tandon, P., Rawat, P., et al. 2011, A&A, 528, 6
Gyevi-Nagy, L., Kallay, M., & Nagy, P. R. 2020, JCTC, 16, 366
Hanine, M., Meng, Z. S., Lu, S. R., et al. 2020, ApJ, 900, 188
Herbst, E. 2014, PCCP, 16, 3344
Herbst, E. 2021, FrASS, 8, 776942
Herbst, E., & van Dishoeck, E. F. 2009, ARA&A, 47, 427
Herbst, E., & Yates, J. T. J. 2013, ChRv, 113, 8707
Hohenstein, E. G., Chill, S. T., & Sherrill, C. D. 2008, JCTC, 4, 1996
Hollenbach, D. J., & Tielens, A. G. G. M. 1999, RvMoPh, 71, 173
Hollis, J. M., Lovas, F. J., Jewell, P. R., et al. 2002, ApJL, 571, L59
Hollis, J. M., Lovas, F. J., Remijan, A. J., et al. 2006, ApJL, 643, L25
Holtom, P. D., Bennett, C. J., Osamura, Y., et al. 2005, ApJ, 626, 940
Jeilani, Y. A., Nguyen, H. T., Newallo, D., et al. 2013, PCCP, 15, 21084
Jeilani, Y. A., & Nguyen, M. T. 2020, JPCB, 124, 11324
Jeilani, Y. A., Ross, B., Aweis, N., et al. 2018, JPCA, 122, 2992
Jeilani, Y. A., Williams, P. N., Walton, S., et al. 2016, PCCP, 18, 20177
Jones, R. O. 2015, RvMP, 87, 897
Jørgensen, J. K., Belloche, A., & Garrod, R. T. 2020, ARA&A, 58, 727
Joshi, P. R., & Lee, Y. P. 2022, CmCh, 5, 62
Jung, S. H., & Choe, J. C. 2013, AsBio, 13, 465
Kayanuma, M., Kidachi, K., Shoji, M., et al. 2017, CPL, 687, 178
Komatsu, Y., & Suzuki, T. 2022, ESC, 6, 2491
Kovács, P., Zhu, X. S., Carrete, J., et al. 2020, ApJ, 902, 100
Krasnokutski, S. A., Jäger, C., & Henning, T. 2020, ApJ, 889, 67
Krishnan, Y., Vincent, A., & Paranjothy, M. 2017, JChS, 129, 1571
Kruse, H., Goerigk, L., & Grimme, S. 2012, JOC, 77, 10824
Kuan, Y. J., Charnley, S. B., Huang, H. C., et al. 2003, ApJ, 593, 848
Laneuville, M., Kameya, M., & Cleaves, H. J., II 2018, AsBio, 18, 897
Lei, J. C., Li, M., Duan, C. G., et al. 2022, JPCA, 126, 3549
Li, A. 2020, NatAs, 4, 339
Li, J., Wang, J. Z., Lu, X., et al. 2021, ApJ, 919, 4
Liao, Q., Xie, P., & Wang, Z. 2023, PCCP, 25, 28829
Loomis, R. A., Zaleski, D. P., Steber, A. L., et al. 2013, ApJL, 765, L9
López-Sepulcre, A., Balucani, N., Ceccarelli, C., et al. 2019, ESC, 3, 2122
Lu, S. R., Meng, Z. S., Xie, P., et al. 2021, A&A, 656, 7
Mardirossian, N., & Head-Gordon, M. 2017, MolPh, 115, 2315
Mata, R. A., & Suhm, M. A. 2017, ACIE, 56, 11011

Materese, C. K., Gerakines, P. A., & Hudson, R. L. 2021, AcChR, 54, 280
Materese, C. K., Nuevo, M., Mcdowell, B. L., et al. 2018, ApJ, 864, 44
Materese, C. K., Nuevo, M., & Sandford, S. A. 2017, AsBio, 17, 761
McGuire, B. A. 2022, ApJS, 259, 30
McGuire, B. A., Carroll, P. B., Loomis, R. A., et al. 2016, Sci, 352, 1449
McGuire, B. A., Ryan, A., Loomis, R. A., et al. 2012, ApJL, 758, L33
McQuarrie, D. A., & Simon, J. D. 1999, Molecular Thermodynamics (Mill Valley, CA: Univ. Science Books)
Mendoza, C., Ruette, F., Martorell, G., et al. 2004, ApJL, 601, L59
Meng, Z. S., Zhang, Y., Liang, E. W., et al. 2023, MNRASL, 525, L29
Meng, Z. S., Zhu, X. S., Kovács, P., et al. 2021, ApJ, 922, 101
Merz, K. M., Jr, Aguiar, E. C., & da Silva, J. B. P. 2014, JPCA, 118, 3637
Møller, C., & Plesset, M. S. 1934, PhRv, 46, 618
Nguyen, H. T., Jeilani, Y. A., Hung, H. M., et al. 2015, JPCA, 119, 8871
Nhlabatsi, Z. P., Bhasi, P., & Sitha, S. 2016, PCCP, 18, 20109
Nuevo, M., Milam, S. N., Sandford, S. A., et al. 2009, AsBio, 9, 683
Oba, Y., Takano, Y., Naraoka, H., et al. 2019, NatCo, 10, 4413
Oba, Y., Watanabe, N., Osamura, Y., et al. 2015, CPL, 634, 53
Oberg, N., Kamp, I., Cazaux, S., & Rab, C. 2020, A&A, 638, 16
Oró, J. 1960, BBRC, 2, 407
Parker, D. S. N., Kaiser, R. I., Kostko, O., et al. 2015a, ApJ, 803, 53
Parker, D. S. N., Kaiser, R. I., Kostko, O., et al. 2015b, PCCP, 17, 32000
Parker, D. S. N., Yang, T., Dangi, B. B., et al. 2015c, ApJ, 815, 115
Peeters, Z., Botta, O., Charnley, S. B., et al. 2005, A&A, 433, 583
Peverati, R., & Truhlar, D. G. 2011, JPCL, 2, 2810
Pilling, S., Baptista, L., Boechat-Roberty, H. M., et al. 2011, AsBio, 11, 883
Potapov, A., Theulé, P., Jäger, C., et al. 2019, ApJL, 878, L20
Puzzarini, C. 2022, FrASS, 8, 811342
Qi, H. N., Picaud, S., Devel, M., et al. 2018, ApJ, 867, 133
Quan, D. H., Herbst, E., Osamura, Y., et al. 2010, ApJ, 725, 2101
Rani, N., & Vikas 2018, JPCA, 122, 7572
Riley, K. E., Pitonak, M., Jurecka, P., et al. 2010, ChRv, 110, 5023
Rimola, A., Balucani, N., Ceccarelli, C., et al. 2022, IJMS, 23, 4252
Rimola, A., Sodupe, M., & Ugliengo, P. 2010, PCCP, 12, 5285
Rimola, A., Sodupe, M., & Ugliengo, P. 2012, ApJ, 754, 24
Rivilla, V. M., Martin-Pintado, J., Jimenez-Serra, I., et al. 2020, ApJL, 899, L28
Rotelli, L., Trigo-Rodríguez, J. M., Moyano-Cambero, C. E., et al. 2016, NatSR, 6, 38888
Roy, D., Najafian, K., & Schleyer, P. R. 2007, PNAS, 104, 17272
Rubin, R. H., Swenson, G. W. J., Benson, R. C., et al. 1971, ApJL, 169, L39
Saladino, R., Botta, G., Delfino, M., et al. 2013, ChEurJ, 19, 16916
Saladino, R., Carota, E., Botta, G., et al. 2016, OLEB, 46, 515
Sandford, S. A., Nuevo, M., Bera, P. P., et al. 2020, ChRv, 120, 4616
Sato, A., Kitazawa, Y., Ochi, T., et al. 2018, MolAs, 10, 11
Senn, H. M., & Thiel, W. 2009, ACIE, 48, 1198

Shivani, P. P., Misra, A., & Tandon, P. 2017, EPJD, 71, 215
Shivani, S. A., Gupta, V., Misra, A., et al. 2014, A&A, 563, 8
Singh, A., Shivani, Misra, A., et al. 2013, RAA, 13, 912
Skouteris, D., Balucani, N., Ceccarelli, C., et al. 2018, ApJ, 854, 135
Smith, H. B., Drew, A., Malloy, J. F., et al. 2020, AsBio, 21, 177
Snyder, L. E., Lovas, F. J., Hollis, J. M., et al. 2005, ApJ, 619, 914
Sumiya, Y., Harabuchi, Y., Nagata, Y., et al. 2022, JACS, 2, 1181
Suzuki, T., Ohishi, M., Saito, M., et al. 2018, ApJS, 237, 3
Thripati, S. 2022, OrBioCh, 20, 4189
Thripati, S., & Ramabhadran, R. O. 2017, JPCA, 121, 8659
Turner, B. E., Liszt, H. S., Kaifu, N., et al. 1975, ApJL, 201, L149
Unsleber, J. P., & Reiher, M. 2020, ARPC, 71, 121
Vazart, F., Skouteris, D., Balucani, N., et al. 2018, ICCSA, 10961, 730
Vinogradoff, V., Duvernay, F., Fray, N., et al. 2015, ApJL, 809, L18
Wang, J., Gu, J. D., Nguyen, M. T., et al. 2013, JPCB, 117, 14039
Wang, T. F., & Bowie, J. H. 2012, OrBioCh, 10, 652
Wang, Z. 2019, JPCC, 123, 15166
Wang, Z. 2020, JPCC, 124, 3851
Wang, Z. 2023, JPCB, 127, 6227
Woods, P. M., Slater, B., Raza, Z., et al. 2013, ApJ, 777, 90
Woon, D. E. 2002, ApJL, 571, L177
Woon, D. E. 2021, ApJ, 906, 20
Xie, M., Sun, X. N., Li, W. X., et al. 2022, JPCL, 13, 8207
Yang, S. M., Xie, P., Liang, E. W., et al. 2023, RAA, 23, 055019
Yang, Y. Y., Hu, X. Y., Zhang, C. C., et al. 2022, A&A, 663, 13
Zaleski, D. P., Seifert, N. A., Steber, A. L., et al. 2013, ApJL, 765, L10
Zamirri, L., Ugliengo, P., Ceccarelli, C., et al. 2019, ESC, 3, 1499
Zeng, S., Quénard, D., Jiménez-Serra, I., et al. 2019, MNRASL, 484, L43
Zhang, X., Quan, D. H., Chang, Q., et al. 2020a, MNRAS, 497, 609
Zhang, X., Quan, D. H., & Esimbek, J. 2020b, ChJCP, 33, 107
Zhao, G. M., Quan, D. H., Zhang, X., et al. 2021, APJS, 257, 26

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.