Chemical Physics Letters 868 (2025) 141978

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Extension of the Lego-brick approach to protonated molecules

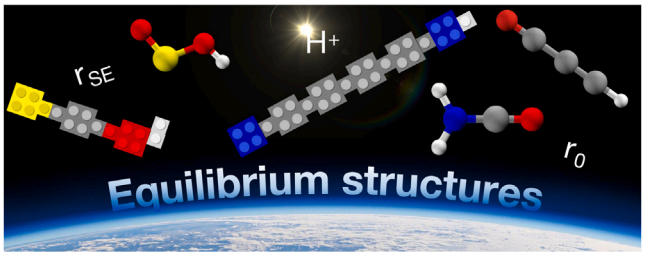
Terri E. Field-Theodore [a,b], Silvia Alessandrini [a], Mattia Melosso [a], Cristina Puzzarini [a,*]

[a] Department of Chemistry "Giacomo Ciamician", University of Bologna, Via F. Selmi 2, Bologna, I-40126, Italy
[b] Scuola Superiore Meridionale, Largo San Marcellino, Naples, 80138, Italy

HIGHLIGHTS

- The Lego-brick approach has been applied to protonated species.
- LETSGO: a new variant of the Lego-brick approach has been introduced and tested.
- Lego-brick / LETSGO reproduce rotational constants with an accuracy of 0.1% / 0.3%.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Protonated molecules
Equilibrium structure
Rotational constants
Lego-brick approach

ABSTRACT

We take the opportunity to extend the applicability of the Lego-brick approach (that is, templating molecular systems from small fragments), by applying it to characterize structural properties of protonated molecules. We additionally present a new variant for it (LETSGO): instead of employing semi-experimental equilibrium geometries of fragments in the templating procedure, experimentally available structures are used. By comparison with experiment (rotational constants), we evaluate the performance of Lego-brick and LETSGO models on a significant range of systems. Our results appear to be a promising extension to techniques for generating equilibrium structures of protonated molecules.

1. Introduction

In the field of astrochemistry, in particular for the identification of molecular species in the interstellar medium (ISM), rotational spectroscopy plays a fundamental role. Radioastronomy exploits rotational transitions as molecular fingerprints to prove the existence of molecules in astrophysical environments [1]. In terms of the knowledge required for a rotational spectrum, the accurate determination of spectroscopic parameters is a fruitful area for collaboration between experimentalists and quantum chemists, with computation often providing supporting and/or complementary information to experimental measurements [2]. In rotational spectroscopy, the leading terms are the rotational constants: their dominance is such that other terms (centrifugal distortion parameters, for example) usually contribute to significantly less than 1%. Rotational constants strongly depend on the molecular geometry [3]; in fact, their equilibrium values can be straightforwardly determined from the equilibrium structure. The inverse problem, which is obtaining structural information from rotational constants, is instead rather complicated. Vibrational effects are one of the major issues when equilibrium geometries are sought.

Focusing on isolated (gas phase) molecules, three approaches immediately come to mind for structural determination: experimental, theoretical, and semi-experimental (SE) [4]. The first considers obtaining geometrical parameters by means of (i) gas electron diffraction (GED) – either ro-vibrationally averaged (r_α) or derived from a thermal average of internuclear distances (r_g), or (ii) rotational spectroscopy – acquiring either an effective structure (r_0) [5] by fitting structural parameters in a least-squares manner from experimental ground-state rotational constants for a set of isotopologues, or a substitution structure (r_s) [6], derived from the same set of data by exploiting the

* Corresponding author.
E-mail address: cristina.puzzarini@unibo.it (C. Puzzarini).

https://doi.org/10.1016/j.cplett.2025.141978
Received 15 December 2024; Accepted 15 February 2025
Available online 26 February 2025

Kraitchman equations [7]. A significant limitation of the experimentally determined structures mentioned above, however, is that they are subject to zero-point vibrational and temperature effects. In principle, information on all vibrationally excited states may be collected in ro-vibrational and rotational spectroscopy studies to derive the experimental equilibrium geometry (r_e). However, this is a viable approach for only very small molecules (up to 3–4 atoms). This, of course, complicates comparisons between the results obtained from different experimental techniques, and hampers the subsequent use of r_α, r_g, r_0, or r_s geometries when calculating molecular properties [4]. Nevertheless, to circumvent these difficulties one may choose to employ quantum chemistry to obtain equilibrium structures. Here, within the Born–Oppenheimer approximation, each structure is associated with a local minimum on a potential energy surface. Moreover, structural parameters at equilibrium are isotopically invariant, enabling a sounded comparison between molecular geometries of various molecules. While optimizing molecular structures using theoretical techniques has today become a routine task, predicting accurate equilibrium geometries requires careful attention to the basis set and electron-correlation treatment [8,9]. Indeed, for small molecules, theory has become increasingly competitive with experiment in this area [8–10]. However, this implies to employ state-of-the-art ab initio methodology, that is, extrapolating coupled-cluster (CC) theory [11] results to the complete basis-set limit to minimize the basis-set truncation error, accounting for electron excitation up to the triple contributions or above to recover to reduce as much as possible the N-electron error, and possibly incorporating very small contributions due to, for example, relativistic effects. Unfortunately, for medium-sized molecular systems and larger, the accurate determination of equilibrium structures, if possible, represents a heavy and expensive computational task. Finally, we turn to the interplay between experiment and theory: the semi-experimental approach. By combining experimental ground-state rotational constants of different isotopologues with theoretical vibrational corrections, we may derive semi-experimental equilibrium structures (r_e^{SE}) from a least-squares fit [12,13]. The success of SE equilibrium structures over the past 30 years indicates how versatile the technique is [4,13]. However, one issue that needs consideration, is the extensive amount of experimental data required for a complete structural characterization: analysis of different isotopologues, possibly accounting for isotopic substitution at each atom. Here, things become exceedingly cumbersome, and thus as yet, this methodology is not easily applied to molecules of significant molecular size and complexity [14–16].

In the case where molecular species have not yet been studied experimentally, among the three approaches discussed above, we turn to the theoretical determination. If the computational study indeed aims to support experimental work in the field of rotational spectroscopy, the target accuracy (that is, on the order of 1 mÅ for bond lengths, 0.1° for angles, and 0.1% for rotational constants [3]) requires the use of state-of-the-art ab initio methodologies, which unfortunately are computationally expensive and time consuming. In recent years, an alternative approach, coined “Lego-brick”, claims to have solved the difficulties inherent in the determination of accurate equilibrium structures for larger systems [17–20]. The basic philosophy of the technique is that one can template a molecule from small fragments (for which an r_e^{SE} is known), much akin to building blocks, and apply linear-regression corrections between different fragments. The “Lego-brick” approach exploits the accurate r_e^{SE} structure of the fragments, scales results with molecular size without a significant loss of accuracy, and avoids highly expensive calculations by combining template molecule (TM) [21] and linear regression (LR) [22] models. For the interested reader, we refer to Refs. [17–19], where the TM and LR approaches are detailed in any aspect, and to Refs. [23,24], where LR data for common bond distances and angles are available for a variety of density functionals and basis sets [22–24]. One striking advantage of the Lego-brick approach is its degree of accuracy in predicting equilibrium geometries and, thus, equilibrium rotational constants, at a fraction of the computational cost compared to ab initio means, owing to the use of density functional theory (DFT). It has yielded quite promising results for polycyclic aromatic hydrocarbons (PAHs) [18], polycyclic aromatic nitrogen heterocycles (PANHs) [18], radical species [19], biomolecules [23,24], and sulfur-containing species bearing the S-S moiety [20], just to name a few. One class of molecules, as yet to be evaluated however, are protonated species.

To date, more than 320 molecules have been discovered in the ISM [25], with ~11% being cations, while only 7 anions have been detected. This is likely related to the inherent difficulties in deriving high quality spectroscopic data for cations, and even more for anions in terrestrial laboratory experiments, while they are expected to be relevant species in astrophysical environments. On the whole, cations play a central role in interstellar chemistry. They are key intermediates in gas-phase ion-molecule reactions, and unique tracers to fully symmetric neutral analogues which lack a permanent dipole moment. Within the interstellar cations subset, 86% are protonated species whose neutral analogues are known to exist in space [1,25]. In essence, any molecule in the ISM may be protonated by the potent Brønsted acid, H_3^+. Protonation, of course, not only causes a redistribution of the electron density in a molecule, but also produces substantial changes in molecular properties, by analogy with the neutral species. The question here is whether the Lego-brick approach can be successfully applied to protonated species. If so, this would enable one to take full advantage of calculating accurate structures of currently unknown cations at a low computational cost. In turn, this would lay the foundation for their spectroscopic characterization.

The first aim of the present study is to test the Lego-brick approach for protonated molecules. To do so, a significant selection of linear, asymmetric-top, and symmetric-top protonated species (that have specifically been experimentally characterized) has been considered. Our test set consists of three subsets. The first includes twelve linear molecules: HC_nNH^+ (with n = 1,3,5,7), $HCCNCH^+$, NC_nNH^+ (with n = 2,4), HOC^+, HCO^+, HC_3O^+, HC_3S^+, and $HCCS^+$ ($^3\Sigma^-$). The second subset, asymmetric-top molecules, contains eight ions: H_2CCCH^+, H_2NCO^+, $HOCS^+$, $HSCO^+$, *cis*- and *trans*-$HOSO^+$, $HOCO^+$, and CH_2CHCNH^+. Finally, the third subset is comprised of five symmetric-top protonated species: CH_3CNH^+, CH_3CO^+, NH_3D^+, H_3S^+, and H_3O^+.

The validity of the Lego-brick approach rests on the agreement with experimental rotational constants, given their high accuracy (usually determined with an accuracy better than 1 part in 10^8) and strong correlation with molecular structure. However, in order to make a meaningful comparison, vibrational effects must be accounted for in the determination of theoretical rotational constants, which otherwise refer to equilibrium structures, while the experimental parameters correspond to the vibrational ground state. Secondly, we propose here a new variant of Lego-brick that may be a viable alternative when semi-experimental structures (used for templating) are not available, but experimental effective structures (such as r_0 and r_s) are. We call this “LETSGO” – Leveraging ExperimenTal Structures for Geometry Optimization. While effective structures are often unable to describe the equilibrium geometry of a molecule well, they may be the only reliable geometric data available. This is, for example, the case of flexible molecules which possess large-amplitude motions (LAMs). Needless to say, the presence of LAMs hampers the reliable calculation of vibrational corrections and straightforward use of the SE approach (see Ref. [20] and references therein). In such cases, it would be desirable to have a valid alternative. Thus, alongside Lego-brick, we test the performance of LETSGO by applying it to the set of systems listed above, and compare results with experiment. In the next section, we present the methodology for both Lego-brick and LETSGO approaches, followed by our results and discussion. Before moving to summarize our findings and give some concluding remarks, we apply the Lego-brick approach to four larger systems (protonated forms of benzene, benzonitrile, quinoline and isoquinoline) whose experimental characterization is still lacking.

2. Methodology

As already noted, the Lego-brick model applies both the TM and LR approaches to the problem of calculating structural parameters. In this context, our interest lies in the structural characterization of protonated species of interstellar interest. For a concise overview, we refer to Refs. [17,19,21,24]. However, for completeness, we briefly present the main points that merit discussion.

Within the TM scheme, one improves structural determinations of low level, usually obtained from DFT (r_e^{DFT}), for a medium- to large-sized molecular system by considering it as formed by smaller fragments, for which accurate r_e^{SE} structures are available. Comparable to building blocks, one can thus mix and match individual fragments to "build" the desired target species [19]. TM-corrected equilibrium parameters (r_e^{TM}) for the target molecule (T) (in this case, our protonated species, $r_e^{TM,T}$) may then be written as

$$r_e{}^{TM,T} = r_e{}^{DFT,T} + \Delta r_e^{(F)} \tag{1}$$

where $r_e^{DFT,T}$ is the target species geometry evaluated at the low (DFT) level, and

$$\Delta r_e^{(F)} = r_e{}^{SE,F} - r_e{}^{DFT,F} \tag{2}$$

Here, F denotes the fragment, and $r_e{}^{SE,F}$ corresponds to its $r_e{}^{SE}$ structure. The fragment species, in this instance, represents the neutral analogue of our protonated target molecule. Incorporating the LR approach, that is, correcting interfragment bond lengths between structural motifs, $\Delta r_e{}^{(F)}$ is replaced by

$$\Delta r_e{}^{LR,T} = a \times r_e{}^{DFT,T} + b \tag{3}$$

where a and b are linear regression parameters, determined by a least-squares fit for a significant set of molecules containing the bond under consideration, and available for different DFT levels [23,24]. In this respect, we are testing the LR functionality with respect to C-, N-, or O- protonation.

The modification of Lego-brick to obtain LETSGO is straightforward. All that is required is to use an experimental effective structure (r_0 or r_s) as a template fragment. The well-known limitations of such geometries lies in their inability to provide an accurate description of molecular structures at equilibrium [10,21]. However, as addressed in the Introduction, they may be the only experimental information available. It is thus important to understand if LETSGO may be a valid alternative to improve low-level results.

Based on previous experience with Lego-brick calculations, $r_e{}^{DFT,T}$ (and $r_e{}^{DFT,F}$) equilibrium geometries have been obtained with the rev-DSDPBE86 [26] double-hybrid functional, incorporating Grimme's D3BJ [27,28] dispersion correction, and the jun-cc-pVTZ basis set [29]. For simplicity, we abbreviate rev-DSDPBEP86-D3BJ/jun-cc-pVTZ to revDSD/junTZ. As mentioned, for a meaningful comparison with experiment, equilibrium rotational constants need to be corrected for vibrational effects. The corresponding vibrational corrections were obtained at the fc-MP2/cc-pVTZ level of theory (fc representing the frozen-core approximation and MP2 standing for Møller-Plesset second-order theory [30]) within vibrational perturbation theory to second order (VPT2) [31]. According to the latter, ground-state rotational constants can be seen to comprise of two contributions:

$$B_0^i = B_e^i - \frac{1}{2}\sum_r \alpha_i^r = B_e^i + \Delta B_{vib}^i \tag{4}$$

where α_i^r denotes the vibration–rotation interaction constants, i is the inertial axis (a, b, or c), and the sum runs over all r vibrational modes in the molecule. From a computational point of view, evaluation of the vibration–rotation interaction constants requires anharmonic force field calculations. Owing to the fact the ΔB_{vib} term usually accounts for 0.5% or less of the corresponding rotational constant, as pointed out in Ref. [3], a low level of theory may be employed for its computation (with an uncertainty smaller than 0.05% affecting the final B_0 value). All MP2 calculations were performed in CFOUR [32], while DFT computations were carried out in the Gaussian suite of programs [33].

Herein for brevity, "TM+LR" refers to the application of the TM approach in conjunction with LR corrections for C-, N-, or O- protonation. "TM" thus corresponds to the model that solely uses the TM approach. As mentioned above, templating with SE equilibrium geometries or experimental effective structures leads to classification of Lego-brick and LETSGO procedures, respectively.

3. Results and discussion

To assess the performance of Lego-brick and LETSGO for protonated molecules, we have divided the discussion into three parts: results for linear (Table 1), asymmetric-top (Table 2), and symmetric-top (Table 3) molecules. All tables list the ground-state rotational constants as computed using Lego-brick, LETSGO, and revDSD/junTZ schemes together with the vibrational corrections employed. The performance of each approach is analyzed using relative percentage errors (noted in square brackets) with respect to experimental data. In Figs. 1–3, we compare equilibrium structures generated using revDSD/junTZ, Lego-brick and LETSGO methodologies for the linear, asymmetric and symmetric species studied, respectively. The LR-corrected value for protonation is also reported. One potential limitation of the LR method here, is that corrections for C–H, N–H, and O–H are based on the neutral bond length. The validity of thus employing them for structures of protonated species is unclear.

3.1. Rotational constants

We begin our discussion with results for the linear subset of molecules (Table 1 and Fig. 1). If only closed-shell species are considered (thus excluding $HCCS^+$), the average relative deviation of revDSD/junTZ is −0.54%. This is somewhat in line with reported literature [34]: while the underestimation of rotational constants with respect to experiment is related to the typical overestimation of bond lengths at the revDSD/junTZ level, the absolute magnitude is a bit larger than the expected ~0.4% value [18,34]. For Lego-brick, the average (unsigned) relative deviation is 0.08% and 0.09% for TM and TM+LR, respectively. For LETSGO, the average relative deviation is 0.23% for both TM and TM+LR. For Lego-Brick, only in four cases does the LR correction lead to an improvement of results, which in general is very small. Interestingly by comparison, incorporating the LR correction for LETSGO improves the agreement with experiment. For the most part, the Lego-brick approach appears to be suitable and effectively applicable to protonated species, providing an accuracy which is in line with previous studies. Particularly intriguing is LETSGO's ability to significantly reduce revDSD/junTZ discrepancies without any additional computational cost.

Let us now discuss Table 1 results in more detail. In the first row of this table, we report the series HC_nNH^+, with n = 1, 3, 5 and 7. Regarding Lego-brick results, the mean unsigned deviation (MUD) is 0.05% for TM and 0.06% for TM+LR. When employing the LR correction, the N–H bond length becomes shorter by about 0.002 Å, and this affords either a small improvement in the results, or increased the error. The LETSGO approach yields a MUD of 0.13% for TM and 0.12% for TM+LR. However, it is noted that the relative discrepancies of both TM and TM+LR for $HCNH^+$ are −0.36% and −0.31%, respectively, while the other members of the series (HC_nNH^+, where n = 3, 5 and 7) show deviations smaller than 0.1%. The performance of revDSD/junTZ is in accord with its expected accuracy: the average deviation is −0.48% and the greatest discrepancy with respect to experiment is found for the smallest member of the series ($HCNH^+$, −0.58%). In our test set, we consider $HCCNCH^+$, an isomer of HC_3NH^+. Both Lego-brick and LETSGO perform very well. Specifically, when the Lego-brick approach is exploited, the deviation is 0.01% for TM and 0.02% TM+LR. Moving to LETSGO, agreement with experiment is similar to Lego-brick:

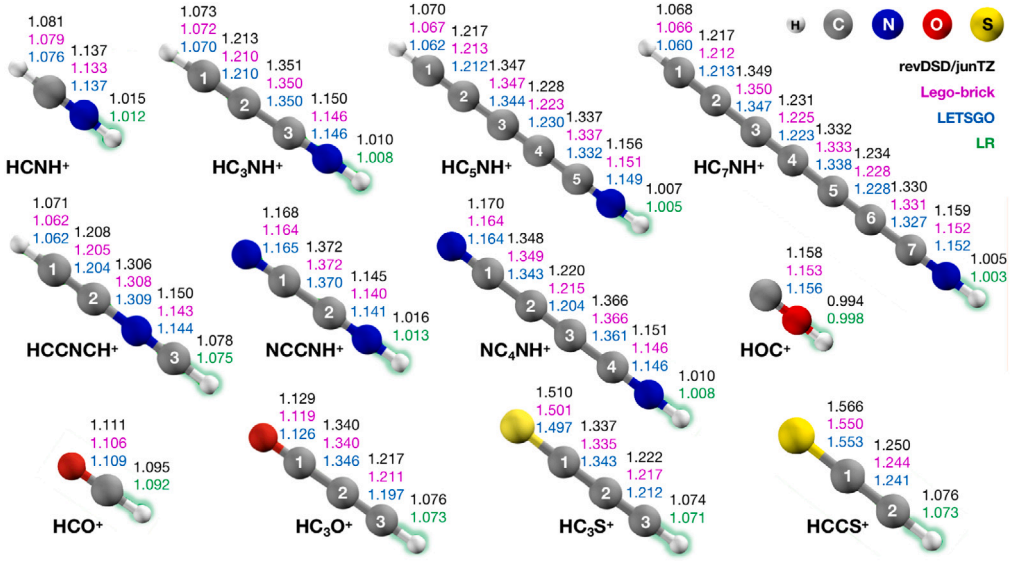


Fig. 1. Linear protonated species set: Equilibrium structures obtained with revDSD/junTZ, Lego-brick and LETSGO approaches. Bond lengths reported in units of angstroms and bond angles in degrees.

Table 1
Set of linear protonated species: computed ground-state rotational constants (MHz). Relative errors (%) (listed in square brackets) with respect to experimental data[a].

Molecule	$HCNH^+$	HC_3NH^+	HC_5NH^+	HC_7NH^+
Expt.[b]	37 055.7482(3)	4328.9970(5)	1295.8159(3)	553.938802(160)
Lego-brick				
(TM)	37 069.804 [0.04]	4326.580 [−0.06]	1294.982 [−0.06]	554.373 [0.08]
(TM+LR)	37 088.561 [0.09]	4326.670 [−0.05]	1295.040 [−0.06]	554.386 [0.08]
LETSGO				
(TM)	36 922.903 [−0.36]	4326.711 [−0.05]	1296.145 [0.03]	554.450 [0.09]
(TM+LR)	36 941.529 [−0.31]	4327.164 [−0.04]	1296.204 [0.03]	554.436 [0.09]
revDSD/junTZ	36 842.381 [−0.58]	4308.438 [−0.47]	1289.854 [−0.46]	551.687 [−0.41]
ΔB_{vib}[c]	−223.480	−2.418	1.020	0.744
Molecule	$HCCNCH^+$	$NCCNH^+$	NC_4NH^+	HOC^+
Expt.[b]	4664.431891(692)	4438.0116(11)	1293.90840(60)	44 743.9141(35)
Lego-brick				
(TM)	4664.691 [0.01]	4435.814 [−0.05]	1293.384 [−0.04]	44 815.067 [0.16]
(TM+LR)	4665.431 [0.02]	4436.339 [−0.04]	1293.435 [−0.04]	44 596.740 [−0.33]
LETSGO				
(TM)	4661.906 [−0.05]	4436.172 [−0.04]	1304.427 [0.81]	44 645.936 [−0.22]
(TM+LR)	4662.547 [−0.04]	4436.653 [−0.03]	1305.478 [0.82]	44 765.533 [0.38]
revDSD/junTZ	4646.906 [−0.38]	4414.589 [−0.53]	1289.323 [−0.36]	44 539.617 [−0.46]
ΔB_{vib}[c]	−3.196	−1.472	1.065	−21.600
Molecule	HCO^+	HC_3O^+	HC_3S^+	$HCCS^+$ $[^3\Sigma^-]$
Expt.[b]	44 594.42895(27)	4460.590(1)	2735.46311(23)	6021.89878(55)
Lego-brick				
(TM)	44 513.954 [−0.18]	4466.338 [0.13]	2733.445 [−0.07]	6126.033 [1.73]
(TM+LR)	44 548.842 [−0.10]	4467.058 [0.15]	2733.767 [−0.06]	6127.143 [1.75]
LETSGO				
(TM)	44 345.081 [−0.56]	4468.601 [0.18]	2732.767 [−0.10]	6124.207 [1.70]
(TM+LR)	44 379.735 [−0.48]	4469.217 [0.19]	2733.088 [−0.09]	6125.316 [1.72]
revDSD/junTZ	44 217.353 [−0.85]	4433.081 [−0.62]	2712.853 [−0.83]	6031.082 [0.15]
ΔB_{vib}[c]	−241.526	−1.395	−1.788	0.601

[a] A full table referencing SE equilibrium (Lego-brick) and experimental effective (LETSGO) geometries is given in the Supplementary Information. The r_e^{SE} structures of HC_7N, C_3O and C_3S are available in the Supplementary Information.

[b] Values in parentheses denote 1σ error and apply to last digits.

[c] fc-MP2/cc-pVTZ vibrational corrections.

−0.05% and −0.04% for TM and TM+LR, respectively. Agreement is worst for the revDSD/junTZ level, for which the relative deviation is −0.38%.

Another interesting comparison is the performance of Lego-brick and LETSGO for NC_nNH^+, with n = 2 and 4. For $NCCNH^+$, Lego-brick and LETSGO schemes are of comparable impressive quality. Either employing Botschwina's $r_e{}^{SE}$ structure [35] or Maki's r_s structure [36], TM relative errors are similar: −0.05% and −0.04%, respectively. Including the LR correction to N–H makes only a very small improvement in the rotational constants: −0.04% for Lego-brick and −0.03% for LETSGO. Moving on to the largest ion of the series, NC_4NH^+, the performance of Lego-brick remains excellent: −0.04% for both TM

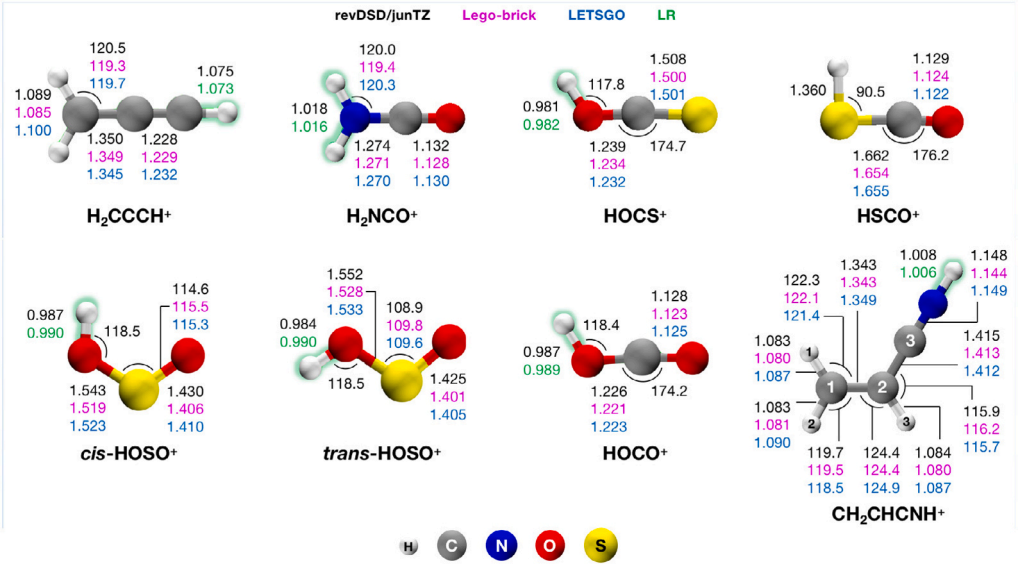


Fig. 2. Asymmetric-top protonated species set: Equilibrium structures obtained with revDSD/junTZ, Lego-brick and LETSGO approaches. Bond lengths reported in units of angstroms and bond angles in degrees.

and TM+LR. Conversely, for LETSGO, deviations are large (0.81% and 0.82% for TM and TM+LR, respectively). This may be attributed to the fact that the experimental structure available for NC_4N is the so-called $r_\alpha{}^0$ [37]. Here, bond lengths are obtained from the refinement of the electron-diffraction data, which also consider the thermal average of the instantaneous internuclear distances. Of our test set, this is the only case which we use a $r_\alpha{}^0$ structure. From our results, this type of structure, compared with r_0 or r_s, is inadequate for templating bond distances.

The next group we consider contains the systems HOC^+, HCO^+, HC_3O^+, HC_3S^+ and $HCCS^+$. Herein we discuss some interesting comparisons: first, HOC^+ and HCO^+ to gain insight on the performance of the TM approach on isomerization; second, HCO^+ and HC_3O^+ to investigate (once again) the lengthening of the carbon chain; third, HC_3O^+ and HC_3S^+ to examine the effect of third-row elements; finally, HC_3S^+ and $HCCS^+$ to study the differences between closed- and open-shell species. Concerning the comparison between HOC^+ and HCO^+, while TM Lego-brick results yield similar agreement, the deviation from experiment is of the opposite sign (0.16% and −0.18%, respectively). With respect to the O–H LR correction, lengthening the bond distance worsens agreement with experiment. Similarly for HOC^+, TM-LETSGO results show a discrepancy with experiment of −0.22%, which becomes 0.38% by applying the LR correction. For HCO^+, LETSGO is unable to recover the large revDSD/junTZ deviation (−0.85%), thus leading to TM and TM+LR relative errors of −0.56% and −0.48%, respectively. Moving from HCO^+ to HC_3O^+, we note an improvement in both TM and TM+LR LETSGO results. As evident from Table 1, specifically comparing HC_3O^+ and HC_3S^+, both Lego-brick and LETSGO perform well, highlighting the methods are not affected by substituting a second-row atom (O) with its third-row analogue (S). For the latter species, agreement with experiment is very good, with TM+LR Lego-brick and LETSGO deviations down to −0.06% and −0.09%, respectively. Analysis of the data in Table 1 reveals that revDSD/junTZ deviates noticeably from experiment for the last three cations discussed: by −0.85% for HCO^+, −0.62% for HC_3O^+, and −0.83% for HC_3S^+. The final comparison to be discussed is HC_3S^+ and $HCCS^+$, the latter molecule being the only open-shell species of the entire test set. At the revDSD/junTZ level, agreement with experiment deviates by 0.15%. This is the only case by which revDSD/junTZ overestimates the rotational constant. Comparison of Lego-brick and LETSGO data for $HCCS^+$ reveals large discrepancies with experiment, on the order of 1.70% to 1.75%. This is not surprising, and somewhat expected. For closed-shell molecules, the revDSD functional is known to systematically predict equilibrium bond lengths that are too long [19]. Incidentally, this is why the Lego-brick approach is so successful. Conversely for open-shell molecules, the revDSD functional predicts bond lengths that are too short, as pointed out in a recent study on radicals [19]. Another aspect that warrants discussion here is the vibrational correction. As expected, MP2 theory is inadequate in describing the electronic structure of open-shell species. This fact is reflected in the computed ΔB^i_{vib}. At the fc-MP2/cc-pVTZ level, the vibrational correction to the rotational constant is 0.60 MHz. Upon correlating all electrons (all-MP2/cc-pCVTZ), ΔB^i_{vib} remains nearly unchanged (0.47 MHz), however when moving to coupled-cluster theory (CC), we note not only a change in sign, but also an increase in an order of magnitude: ΔB^i_{vib} = −5.11 MHz at fc-CCSD/cc-pVTZ, −4.62 MHz at fc-CCSD(T)/cc-pVTZ, −4.05 MHz at fc-CCSD(T)/cc-pVQZ and −4.20 MHz at all-CCSD(T)/cc-pCVQZ [38]. Certainly CC yields improvement in the treatment of the vibrational correction, however it has a small effect on the overall deviation from experiment. When the best ΔB_{vib} is employed to TM-Lego-brick results, the discrepancy is 1.65%. We note the inadequacy of MP2 theory is limited to radicals. This is illustrated by the results of HC_3S^+. Vibrational corrections at both the fc-MP2/cc-pVTZ and fc-CCSD(T)/cc-pVTZ levels of theory are of comparable quality, −1.79 MHz and −1.48 MHz, respectively.

Moving on to our series of asymmetric-top protonated species (Table 2 and Fig. 2), our data set consists of eight molecules: H_2CCCH^+, H_2NCO^+, the O- and S-protonated forms of OCS, *cis*- and *trans*-$HOSO^+$, $HOCO^+$, and CH_2CHCNH^+. With the exception of the first (H_2CCCH^+) and last species (CH_2CHCNH^+) of this list, we note a change in molecular point group upon templating the neutral molecule to the protonated form. This modification predominantly affects the moment of inertia along the *a* axis, and thus the A_0 rotational constant. Indeed, for CH_2CHCNH^+, whose C_s symmetry remains unchanged upon protonation, A_0 agrees very well with experiment (the relative deviation is 0.06%). Conversely, for H_2CCCH^+ (same C_{2v} symmetry of the unprotonated form), it is noted that the experimental A_0 rotational constant is poorly determined (statistical uncertainty of 247 MHz). For our test set of asymmetric-top molecules, revDSD/junTZ provides a MUD of 1.27%. This large average discrepancy can be traced to the poor performance of the approach for *cis*- and *trans*-$HOSO^+$. The MUD for these two cases is 3.06%, with unsigned deviations greater than 4% on both A_0 rotational constants. If one were to exclude both *cis*- and *trans*-$HOSO^+$ from the statistical analysis, the performance of revDSD/junTZ (0.67%)

Table 2
Computed ground-state rotational constants for asymmetric-top molecules (MHz). Relative errors (%) (listed in square brackets) with respect to experimental data[a].

Molecule	H_2CCCH^+			H_2NCO^+		
Rot. constants	A_0	B_0	C_0	A_0	B_0	C_0
Expt.[b]	281 856.(247)	9675.841(1)	9342.877(1)	319 782(103)	10 278.6846(26)	9948.9034(23)
Lego-brick						
(TM)	284 090.480 [0.79]	9662.698 [−0.14]	9329.296 [−0.15]	314 526.502 [−1.64]	10 282.537 [0.04]	9949.437 [0.01]
(TM+LR)	284 090.480 [0.79]	9664.889 [−0.11]	9331.400 [−0.12]	315 907.388 [−1.21]	10 284.089 [0.05]	9952.246 [0.03]
LETSGO						
(TM)	272 670.392 [−3.26]	9667.354 [−0.09]	9320.974 [−0.23]	320 289.747 [0.16]	10 251.760 [−0.26]	9926.094 [−0.23]
(TM+LR)	272 972.790 [−3.26]	9669.648 [−0.06]	9323.115 [−0.21]	321 759.582 [0.62]	10 253.431 [−0.25]	9929.045 [−0.20]
revDSD/junTZ	282 814.090 [0.34]	9648.641 [−0.28]	9314.779 [−0.30]	318 327.127 [−0.48]	10 208.695 [−0.68]	9883.695 [−0.66]
ΔB_{vib}[c]	−1934.526	5.950	−12.069	−4017.520	−28.150	−37.975
Molecule	$HOCS^+$			$HSCO^+$		
Rot. constants	A_0	B_0	C_0	A_0	B_0	C_0
Expt.[b]	779 000.1(41)	5750.3771(30)	5702.9444(32)	279 431.995(10)	5696.74654(86)	5575.97796(66)
Lego-brick						
(TM)	776 197.021 [−0.36]	5750.273 [0.00]	5703.185 [0.00]	277 532.800 [−0.68]	5698.785 [0.04]	5578.488 [0.05]
(TM+LR)	774 615.478 [−0.56]	5750.983 [0.01]	5703.794 [0.01]	–	–	–
LETSGO						
(TM)	776 297.736 [−0.35]	5751.160 [0.01]	5704.064 [0.02]	277 532.258 [−0.68]	5699.211 [0.04]	5578.897 [0.05]
(TM+LR)	774 615.478 [−0.56]	5750.983 [0.01]	5703.794 [0.01]	–	–	–
revDSD/junTZ	775 669.353 [−0.43]	5695.065 [−0.96]	5648.776 [−0.95]	277 534.598 [−0.68]	5642.959 [−0.94]	5524.867 [−0.92]
ΔB_{vib}[c]	19 513.818	−14.368	−17.872	−1725.362	−11.246	−17.129
Molecule	*cis*-$HOSO^+$			*trans*-$HOSO^+$		
Rot. constants	A_0	B_0	C_0	A_0	B_0	C_0
Expt.[b]	44 183.605(41)	9899.77223(67)	8070.09973(58)	44 766.215(35)	9961.72064(93)	8131.9572(13)
Lego-brick						
(TM)	44 510.599 [0.74]	9910.649 [0.11]	8088.802 [0.23]	45 169.697 [0.90]	9944.955 [−0.17]	8134.77 [0.03]
(TM+LR)	44 484.669 [0.68]	9910.617 [0.11]	8087.905 [0.22]	45 168.912 [0.90]	9943.447 [−0.18]	8133.738 [0.02]
LETSGO						
(TM)	43 551.132 [−1.43]	9955.269 [0.56]	8085.315 [0.19]	44 710.106 [−0.13]	9909.112 [−0.53]	8095.489 [−0.45]
(TM+LR)	43 525.029 [−1.49]	9955.355 [0.56]	8084.451 [0.18]	44 709.362 [−0.13]	9907.614 [−0.54]	8094.465 [−0.46]
revDSD/junTZ	42 280.255 [−4.31]	9691.305 [−2.11]	7866.359 [−2.52]	42 761.167 [−4.48]	9737.27 [−2.25]	7914.533 [−2.67]
ΔB_{vib}[c]	217.453	−56.131	−47.221	263.456	−60.229	−47.408
Molecule	$HOCO^+$			$H_2CCHCNH^+$		
Rot. constants	A_0	B_0	C_0	A_0	B_0	C_0
Expt.[b]	789 944.610(21)	10 773.68964(34)	10 610.3413(23)	46 199.5(167)	4791.13798(92)	4334.79 183(92)
Lego-brick						
(TM)	780 720.417 [−1.17]	10 786.718 [0.12]	10 623.190 [0.12]	46 227.521 [0.06]	4788.049 [−0.06]	4332.475 [−0.05]
(TM+LR)	777 063.288 [−1.63]	10 785.652 [0.11]	10 621.430 [0.10]	46 259.330 [0.13]	4788.944 [−0.05]	4333.766 [−0.02]
LETSGO						
(TM)	780 600.255 [−1.18]	10 749.3863 [−0.23]	10 586.874 [−0.22]	45 547.309 [−1.41]	4780.714 [−0.22]	4320.375 [−0.33]
(TM+LR)	776 944.002 [−1.65]	10 748.327 [−0.24]	10 585.125 [−0.24]	45 549.404 [−1.41]	4781.190 [−0.21]	4320.784 [−0.32]
revDSD/junTZ	780 421.573 [−1.21]	10 694.375 [−0.74]	10 533.3519 [−0.73]	46 511.767 [0.68]	4761.817 [−0.61]	4313.689 [−0.49]
ΔB_{vib}[c]	20 994.534	−28.922	−40.637	−66.436	−17.223	−20.918

[a] A full table referencing SE equilibrium (Lego-brick) and experimental effective (LETSGO) geometries is given in the Supplementary Information.
[b] Values in parentheses denote 1σ error and apply to last digits.
[c] fc-MP2/cc-pVTZ vibrational corrections.

is in line with results from our linear subset. Turning to the TM Lego-brick approach, if we disregard A_0 rotational constants for the subset of species whose point group change upon protonation (and that of H_2CCCH^+ whose experimental value is poorly determined), we obtain a MUD of 0.08%. Albeit, if we consider all rotational constants, the MUD increases to 0.32%. The incorporation of the LR correction has a small effect. Statistically considering the rotational constants A_0, B_0, and C_0 in TM+LR Lego-brick results, the MUD increases to 0.39%. Alternatively, discounting the A_0 rotational constants (as mentioned above), we note the MUD remains unchanged. On average, the application of TM Lego-brick leads to agreement with experiment better than 0.1% (0.08%), when only B_0 and C_0 rotational constants are considered in the statistical analysis. Given all eight species of our asymmetric-top test set are planar, for rotational spectroscopy applications, B_0 and C_0 are most often the rotational constants of interest. With respect to LETSGO results, the MUDs (statistically including all rotational constants) are 0.51% and 0.70% for TM and TM+LR, respectively. However, these deviations decrease upon excluding the problematic A_0 values: 0.30% for TM, and 0.45% for TM+LR. Discarding the problematic A_0 values, the MUD of revDSD/junTZ is 0.97%, which decreases (without any additional cost) to 0.30% by TM LETSGO, and further to 0.08% by TM Lego-brick.

Among the eight ions in our asymmetric-top test set, $HOSO^+$ yields the largest deviation from experiment at the revDSD/junTZ level. We note this outcome is not however a failure of this level of theory. For both *cis*- and *trans* isomers of $HOSO^+$, equilibrium geometries and rotational constants have been evaluated at the coupled-cluster level [39]. Specifically, if we apply fc-MP2/cc-pVTZ vibrational corrections (reported in Table 2) to the fc-CCSD(T)/cc-pVTZ equilibrium rotational constants of *cis*-$HOSO^+$ [39], we obtain deviations from experiment on the order of −4.21% for A_0, −2.03% for B_0 and −2.44% for C_0. Notably, these discrepancies reduce to 0.44% for A_0, −0.05% for B_0, and 0.04%

Table 3
Computed ground-state rotational constants for symmetric-top molecules (MHz). Relative errors (%) (listed in square brackets) with respect to experimental data[a].

Molecule	CH_3CNH^+	CH_3CO^+	NH_3D^+	SH_3^+	H_3O^+
Rot. constant	B_0	B_0	B_0	B_0	B_0
Expt.[b]	8590.5589(10)	9134.47211(20)	131 412.1315(130)	146 737.6663(13)	331 411.766(107)
Lego-brick					
(TM)	8594.776 [0.05]	9132.590 [−0.02]	131 268.635 [−0.11]	146 788.081 [0.03]	332 416.058 [0.30]
(TM+LR)	8596.115 [0.06]	9135.177 [0.01]	131 052.052 [−0.27]	–	329 508.111 [−0.57]
LETSGO					
(TM)	8584.797 [−0.07]	91 144.593[−0.22]	131 060.667 [−0.27]	143 188.675 [−2.42]	333 753.168 [0.71]
(TM+LR)	8587.266 [−0.04]	9117.145 [−0.19]	131 898.638 [0.37]	–	330 426.601 [−0.30]
revDSD/junTZ	8553.738 [−0.43]	9082.291 [-0.57]	130 660.147 [−0.57]	146 049.942 [−0.47]	329 769.185 [−0.50]
ΔB_{vib}[c]	−26.476	−26.977	−1974.076	−1629.101	81.148

[a] A full table referencing SE equilibrium (Lego-brick) and experimental effective (LETSGO) geometries is given in the Supplementary Information. The r_e^{SE} structure of H_2S is available in the Supplementary Information.
[b] Values in parentheses denote 1σ error and apply to last digits.
[c] fc-MP2/cc-pVTZ vibrational corrections.

for C_0 when considering fc-MP2/cc-pVTZ vibrational corrections to all-CCSD(T)/cc-pwCVQZ equilibrium rotational constants. A similar trend is obtained for *trans*-$HOSO^+$. A substantial part of the improvement can be attributed to the shortening of the two S-O bond distances (by about 21–23 mÅ), when going from fc-CCSD(T)/cc-pVTZ to all-CCSD(T)/cc-pwCVQZ. Interestingly, as evident in Fig. 2, Lego-brick and LETSGO are able to recover the large core-correlation effects at no computational cost. These effects are predominantly related to the sulfur atom. By comparison, the average deviation of revDSD/junTZ for $HOCO^+$ is −0.89%, and for each rotational constant, the agreement with experiment is about one third of the corresponding value for protonated sulfur dioxide.

The last set of protonated molecules to be discussed is that of the symmetric-top species (Table 3 and Fig. 3): CH_3CNH^+, CH_3CO^+, NH_3D^+, SH_3^+, and H_3O^+. For these, only B_0 has been considered, owing to it being the only determinable rotational constant from analysis of the rotational spectrum (if perturbations are excluded). At the revDSD/junTZ level, the mean relative deviation across the series is −0.51%. These predictions are in accordance with results of both our linear and asymmetric-top molecules; relative to experiment, rotational constants are underestimated due to the typical overestimation of bond lengths of the revDSD functional. Turning to the TM Lego-brick approach, the MUD is 0.10%, which increases to 0.23% when the LR correction is applied. This increase in error is attributable to NH_3D^+ and H_3O^+. Indeed, for these two species, the TM+LR structure is simply a LR-corrected geometry because all the bond distances are equivalent: application of the LR correction substitutes the TM model. Since the LR-corrected bond lengths are longer than the corresponding TM-corrected values, the larger MUD is not all surprising. Given there is no LR correction for S-H available in the literature, for SH_3^+, we report only the TM result. Regarding the performance of LETSGO, if SH_3^+ is excluded, the TM MUD is 0.32%. If we were to consider SH_3^+ in the statistical analysis, we yield a MUD of 0.74%. Strikingly, the relative deviation from experiment for SH_3^+ is −2.42%. This is largely due to the difference in r_0 and r_e^{SE} geometries; the former possesses an S–H bond distance that is longer (by 11 mÅ) and HSH angle that is smaller (by about 1°). Incorporating the LR correction to LETSGO results reduces the MUD to 0.23%. Akin to TM+LR Lego-brick results, the TM+LR LETSGO structures for NH_3D^+ and H_3O^+ are simply LR-corrected geometries.

As already discussed for our asymmetric-top test set, in some cases, we note a change of point group when moving from the templating neutral molecule to the protonated species. This is indeed the case for SH_3^+ and H_3O^+ (H_2S and H_2O possess C_{2v} symmetry and, upon protonation, their symmetry becomes C_{3v}), and NH_3D^+ (the C_{3v} symmetry of NH_3 changes to T_d, at equilibrium, upon protonation). However, as previously mentioned, the change in the symmetry predominantly affects the rotational constant A_0. Focusing specifically on TM Lego-brick results, if we exclude H_3O^+ from the statistical analysis, the MUD is halved, reducing to 0.05%. A closer inspection of Table 3 highlights the TM Lego-brick relative error of H_3O^+ is 0.30%. This is larger than the typical accuracy of 0.1% (or better). However, this is not surprising, owing to the umbrella-inversion motion of H_3O^+, which affects both theory and experiment. VPT2 calculations lead to vibrational corrections that strongly depend on the level of theory (5.25 MHz with fc-CCSD(T)/cc-pVTZ, 106.79 MHz with all-CCSD(T)/cc-pCVTZ, and 238.83 MHz with fc-CCSD(T)/cc-pVQZ). Experimentally, the 0^- and 0^+ inversion splitting of the vibrational ground state is relevant, with the B_0 values of the two states differing significantly.

3.2. Equilibrium structures

From the inspection of Figs. 1–3, it is apparent that the bond lengths of revDSD/junTZ, TM Lego-brick and TM LETSGO equilibrium structures agree to two decimal places, variations lying in the third decimal place and being usually on the order of 1 to 4 mÅ. Between the three structures the angle parameters are also very similar, the differences being usually within 0.5°. These small differences highlight the great sensitivity of rotational constants to molecular geometries. For example, comparing revDSD/junTZ and TM Lego-brick structures for $HCNH^+$, we note that the overall length of the molecule (that is, the sum of the bond lengths) decreases by only 6 mÅ from the former to the latter. This, however, leads to an increase of 227.4 MHz in the computed rotational constant. Another example we may consider is HCO^+. In this particular case, we observe a 5 mÅ reduction in the total length of TM Lego-brick, relative to revDSD/junTZ. This small change in the geometry results in the rotational constant enlarging by 296.6 MHz.

As previously mentioned, the revDSD/junTZ level of theory inadequately describes S-O bonds (as in the case of $HOSO^+$), predicting bond lengths too long (by about 21–23 mÅ). It is apparent in Fig. 2 that both Lego-brick and LETSGO are able to recover this overestimation quite well. Another ion showing a large variation in predicted geometries is $HCCS^+$. Given its open-shell nature and the effect of protonation, it is hardly surprising this cation poses a challenge for Lego-brick and LETSGO approaches.

The results of Table 4 showcase the great accuracy one can obtain with the Lego-brick approach. The reference structures employed to carry out our benchmark are SE equilibrium geometries and/or very accurate theoretical determinations (that are known to have an accuracy of about 1 mÅ or better). On average, TM(+LR) structural parameters agree with reference values to within 1–2 mÅ for bond lengths, and 0.1–0.2° for valence angles. Overall, we note very few exceptions. For *cis*- and *trans*-$HOSO^+$, the HOS angle deviates by 1.5° for the former, and 2.9° for the latter. Indeed, these angles are neither TM nor LR corrected, but rather the optimized revDSD/junTZ parameters. As we

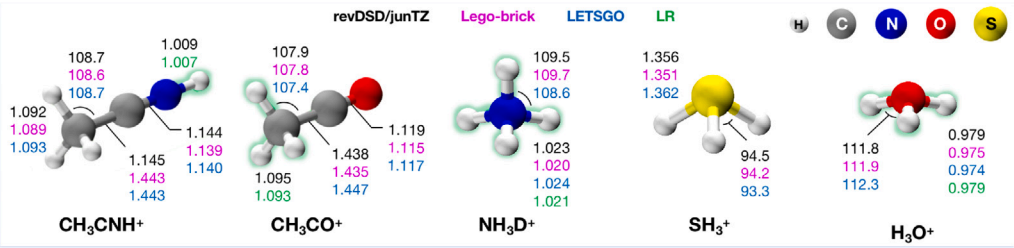


Fig. 3. Symmetric-top protonated species set: Equilibrium structures obtained with revDSD/junTZ, Lego-brick and LETSGO approaches. Bond lengths reported in units of angstroms and bond angles in degrees.

Table 4
Comparison between TM(+LR), semi-experimental (SE) and computed (theo) equilibrium structures.[a].

Molecule	Param.	TM(+LR)	theo	SE	Molecule	Param.	TM(+LR)	theo
H_3S^+[b]	S–H	1.352	1.350	1.3500(1)	*cis*-$HOSO^+$[g]	H–O1	0.987 (0.990)	0.983
	HSH	94.2	94.3	94.15(1)		O1–S	1.519	1.519
HCO^+[c]	C–O	1.106	1.107	1.1091(1)		S–O2	1.406	1.408
	C–H	1.095 (1.092)	1.093	1.0919(1)		HO1S	118.5	120.0
$HCNH^+$[d]	H–C	1.079	–	1.0779(2)		O1SO2	115.5	115.2
	C–N	1.133	–	1.1340(1)	*trans*-$HOSO^+$[g]	H–O1	0.984 (0.990)	0.981
	N–H	1.015 (1.012)	–	1.0123(3)		O1–S	1.528	1.529
$HNCCN^+$[e]	H–N	1.016 (1.013)	1.014	1.0133(1)		S–O2	1.401	1.404
	N–C	1.140	1.141	1.1406(1)		HO1S	118.5	115.6
	C–C	1.372	1.374	1.3724(1)		O1SO2	109.8	109.6
	C–N	1.164	1.163	1.1634(1)	H_2CCCH^+[h]	H1–C1	1.085	1.086
$HCCS^+$[f]	H–C2	1.076 (1.073)	1.074	–		C1–C2	1.349	1.347
	C2–C1	1.244	1.257	–		C2–C3	1.229	1.228
	C1–S	1.550	1.561	–		C3–H2	1.075 (1.073)	1.073
$HOCS^+$[f]	H–O	0.981 (0.982)	0.978	–		H1C2C3	119.3	120.4
	O–C	1.234	1.236	–	CH_2CHCNH^+[i]	H1–C1	1.080	1.081
	C–S	1.500	1.498	–		H2–C1	1.081	1.080
	HOC	117.8	117.6	–		C1–C2	1.343	1.341
	OCS	174.7	174.8	–		H3–C2	1.080	1.081
$HSCO^+$[f]	H–S	1.360	1.355	–		C2–C3	1.413	1.416
	C–S	1.654	1.655	–		C3–N	1.144	1.145
	C–O	1.124	1.124	–		N–H	1.008 (1.006)	1.006
	HSC	90.5	90.3	–		H1C1C2	122.1	122.2
	SCO	176.2	176.3	–		H2C1C2	119.5	119.7
						H3C2C3	124.4	124.7
						C1C2C3	119.4	119.4

[a] Bond distances in angstroms and angles in degrees. For TM(+LR) structures, the TM structural parameters are reported, where applies the LR-corrected values is given in parentheses.
[b] r_e^{SE} (SE) structure from Ref. [21], theoretical (theo) structure (all-CCSD(T)/cc-pwCVQZ) from Table 1 of Ref. [40].
[c] r_e^{SE} (SE) structure from Ref. [21], theoretical (theo) structure (CCSD(T)/CBS+CV) from Ref. [38].
[d] r_e^{SE} (SE) structure from Ref. [41].
[e] r_e^{SE} (SE) structure from Ref. [21], theoretical (theo) structure (CCSD(T)/CBS+CV+fT+pQ) from Table 1 of Ref. [42].
[f] Theoretical (theo) structure (CCSD(T)/CBS+CV+fT+fQ) from Ref. [38].
[g] Theoretical (theo) structure (CCSD(T)/CBS+CV+fT+fQ) from Ref. [39].
[h] Theoretical (theo) structure (all-CCSD(T)/cc-pwCVQZ) from Ref. [43].
[i] Theoretical (theo) structure (all-CCSD(T)/cc-pwCVQZ) from Ref. [44].

have already discussed, revDSD/junTZ does not perform well for this system. Another large discrepancy, that is not surprising however, is the open-shell species $HCCS^+$. Here, we note TM(+LR) C–C and C–S bond distances deviate from the reference (CCSD(T)/CBS+CV+fT+fQ) values by approximately 11–13 mÅ.

3.3. Application to large systems

Finally, this work deals with the accurate prediction of equilibrium geometries (and resultant rotational constants) for a few protonated systems containing one or two aromatic rings. In this section, we restrict ourselves to the application of Lego-brick. Its accuracy is thus far very impressive, instilling confidence in results for species where there is considerably less experimental data available for comparison.

The species considered in this section are protonated benzene ($C_6H_7^+$), protonated benzonitrile ($C_6H_5CNH^+$), protonated quinoline (Quinoline-H^+), and protonated isoquinoline (Isoquinoline-H^+). There is limited spectroscopic information on $C_6H_7^+$ and $C_6H_5CNH^+$: rotational constants have been calculated at CCSD(T*)-F12a/cc-pVTZ-F12 with an empirical correction for expected bond length overestimation for the former species [45], while B3LYP-D3/aug-cc-pVTZ [46] calculations have been performed on the latter. No gas-phase structural data are available for protonated quinoline and protonated isoquinoline.

The rotational constants derived from TM and TM+LR Lego-brick results are given in Table 5, where a comparison to revDSD/junTZ is also provided. In analogy to our benchmark study, the templating molecule is the non-protonated one: benzene for $C_6H_7^+$, benzonitrile for $C_6H_5CNH^+$, quinoline for Quinoline-H^+, and isoquinoline for Isoquinoline-H^+. While the r_e^{SE} structure is available for benzene and benzonitrile, the TM structure from Ref. [18] has been employed for quinoline and isoquinoline. The resulting structural parameters are reported in Fig. 4. Due to the molecular size and the related computational cost of fc-MP2/cc-pVTZ anharmonic calculations, vibrational corrections have been obtained (within VPT2) at the B3LYP-D3BJ/jun-cc-pVDZ (hereafter B3/junDZ) level, which is analogous to that used in

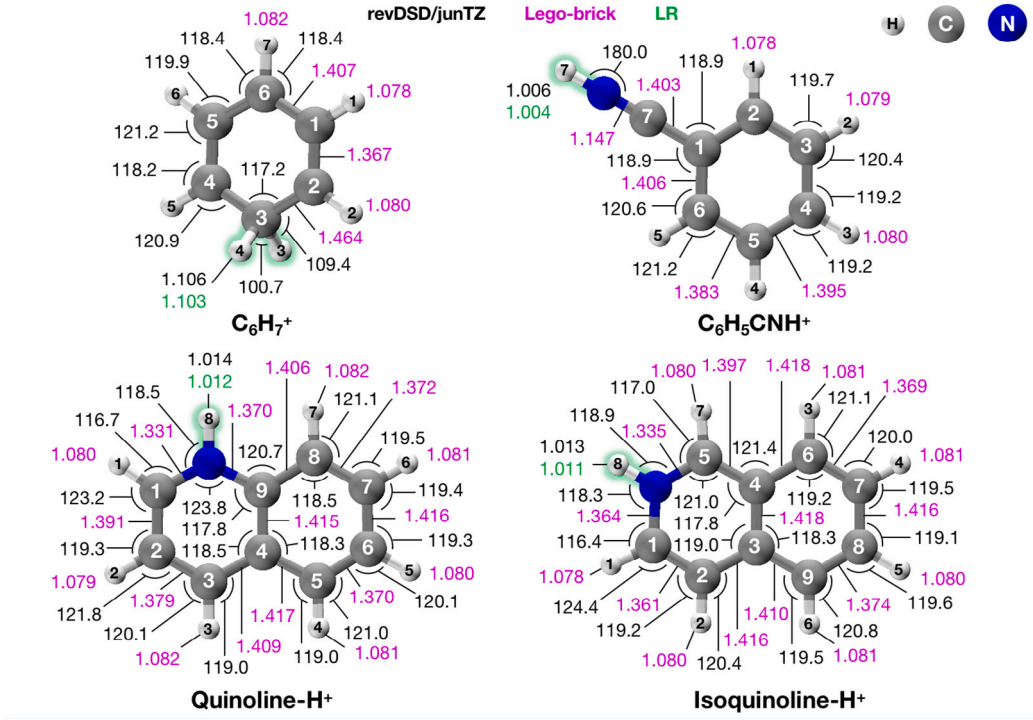


Fig. 4. Application subset: Equilibrium structures obtained with the Lego-brick approach. Bond lengths reported in units of angstroms.

Table 5
Computed ground-state rotational constants for the "application" subset of molecules (MHz)[a].

Molecule	$C_6H_7^+$			$C_6H_5CNH^+$		
Rot. constants	A_0	B_0	C_0	A_0	B_0	C_0
Lego-brick						
(TM)	5443.973	5297.062	2727.236	5544.196	1505.176	1183.642
(TM+LR)	5445.413	5297.508	2727.479	5544.196	1505.250	1183.688
revDSD/junTZ	5419.571	5289.534	2719.081	5515.275	5289.534	1182.273
ΔB_{vib}[b]	−35.574	−36.165	−18.370	−46.318	−4.976	−5.334
Molecule	Quinoline-H^+			Isoquinoline-H^+		
Rot. constants	A_0	B_0	C_0	A_0	B_0	C_0
Lego-brick						
(TM)	3095.012	1246.919	889.007	3135.851	1215.081	875.921
(TM+LR)	3095.192	1246.919	889.022	3135.898	1215.115	875.941
revDSD/junTZ	3085.062	1243.380	886.388	3111.576	1218.670	875.876
ΔB_{vib}[b]	−23.938	−8.102	−5.912	−24.594	−7.822	−5.807

[a] A full table referencing the equilibrium geometries of benzene, benzonitrile, quinoline and isoquinoline is given in the Supplementary Information.
[b] B3/junDZ vibrational corrections.

Ref. [18]. In accordance with this latter work on PAHs and PANHs [18], bond angles were kept fixed at the corresponding revDSD/junTZ values.

As noted from Fig. 4, incorporation of the LR correction shortens the C–H^+/N–H^+ bond length, thus increasing the rotational constants. Such an increase is extremely small and ranges from +0.001% to +0.009%, which in absolute terms means variations from +15 kHz to ~450 kHz. Only for A_0 of $C_6H_7^+$, a more relevant modification (from TM to TM+LR) is noted: +0.03%, corresponding to +1.44 MHz. However, this was somewhat expected because protonation of benzene changes its symmetry point group. Based on our benchmark study, TM and TM+LR constants are expected to have an accuracy of about 0.1%, employing B3/junDZ vibrational corrections does not affect the overall result [18]. Therefore, it is hoped these results will support laboratory spectroscopy experiments in a near future. Among the "application" subset, $C_6H_7^+$ is perhaps the most interesting case because it can be considered as another proxy of benzene; indeed, protonation generates a dipole moment of about 0.7 D.

4. Conclusions

In this work, we have extended the applicability of Lego-brick to a new area of chemistry: determining accurate equilibrium molecular structures of protonated molecules. Furthermore, we have introduced a new variant, LETSGO, which further extends its applicability. As in previous works, the accuracy has been tested for a significant set of species and based on the comparison of rotational constants. For the systems considered, the Lego-brick mean absolute error from experiment was found to be 0.08% for TM and 0.11% for TM+LR. Preliminary investigation of the newly proposed LETSGO approach also produced encouraging results: the mean absolute error of the TM and TM+LR schemes are 0.28% and 0.34%, respectively. From the results presented,

it can be concluded that linear regression corrections to the molecular geometry (with respect to C-, N-, or O- protonation) does not represent an improvement in the calculated rotational constants of protonated systems. The mean absolute error of revDSD/junTZ calculations with respect to experiment is a striking 0.89%. The conclusion that can be drawn is that the Lego-brick approach can be successfully applied to protonated species using the corresponding unprotonated molecules for the TM scheme. The new variant, LETSGO, is a promising method because it allows for improving the accuracy of the revDSD/junTZ level without any additional cost.

More than two decades ago, theoretical spectroscopists predicted calculations would progress to the point where they can be accurate enough to guide the detection of difficult molecular species (such as small anions or cations) from ab initio predictions [47]. To date, only a few interstellar cations have been astronomically detected based solely on theoretical calculations. To the best of our knowledge, they include N_2H^+ [48,49], HCS^+ [50], $HCCS^+$ [51], l-C_3H^+ [52], HC_3O^+ [53], and HC_5NH^+ [54]. Indeed, the accuracy required by astronomical searches can be obtained only experimentally. However, in case of linear species, which are characterized by thinned simple rotational spectra, accurate computational predictions turned out to be effective when experiment was lacking [51,53,54]. The Lego-brick calculations of this and previous works are convincing evidence of the impressive accuracy that is obtainable at less cost and greater convenience of a CCSD(T) computations. We hope the theoretical strategy presented herein will be useful in proposing new interstellar molecules and predicting their equilibrium structures, in guiding future laboratory experiments and assisting in line assignments. Indeed, Lego-brick turns out to be a less expensive computational approach than the empirical scaling procedure which is often used in the field of rotational spectroscopy (see, e.g., Refs. [51,53]). In fact, this latter approach requires the accurate structural determination (with high-level ab initio methods) of the reference (experimentally characterized) and target (new) species together with the computation of their anharmonic force fields (for vibrational corrections to rotational constants).

CRediT authorship contribution statement

Terri E. Field-Theodore: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation. **Silvia Alessandrini:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mattia Melosso:** Investigation, Conceptualization. **Cristina Puzzarini:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Peter R. Taylor is acknowledged for helpful discussions. This research has been supported by computing resources from the Department of Chemistry "Giacomo Ciamician" (University of Bologna). This work has been supported by MUR (PRIN Grant Numbers 202082CE3T and P2022ZFNBL) and by the University of Bologna (RFO funds). The COST Action CA21101 "COSY - Confined molecular systems: from a new generation of materials to the stars" is acknowledged.

MM thanks the European Union – Next Generation EU under the Italian National Recovery and Resilience Plan (PNRR M4C2, Investment 1.4 – Call for tender n. 3138 dated 16/12/2021—CN00000013 National Centre for HPC, Big Data and Quantum Computing (HPC) – CUP J33C22001170001

Appendix A. Supplementary data

Supplementary material collects: (i) the semi-experimental equilibrium and experimental effective r_0 structures determined in this work; (ii) replicas of Tables 1, 2, 3, and 5 with references to literature.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cplett.2025.141978.

Data availability

Data will be made available on request.

References

[1] B.A. McGuire, 2021 census of interstellar, circumstellar, extragalactic, protoplanetary disk, and exoplanetary molecules, Astrophys. J. Suppl. Ser. 259 (2) (2022) 30.
[2] C. Puzzarini, S. Alessandrini, L. Bizzocchi, M. Melosso, V.M. Rivilla, From the laboratory to the interstellar medium: A strategy to search for exotic molecules in space, Front. Astron. Space Sci. 10 (2002) 1211784.
[3] C. Puzzarini, J.F. Stanton, Connections between the accuracy of rotational constants and equilibrium molecular structures, Phys. Chem. Chem. Phys. 25 (2023) 1421–1429.
[4] J. Demaison, Experimental, semi-experimental and ab initio equilibrium structures, Mol. Phys. 105 (23–24) (2007) 3109.
[5] W. Gordy, R.L. Cook, in: A. Weissberger (Ed.), Microwave Molecular Spectra, 3rd ed., Wiley, New York, 1984.
[6] C.C. Costain, Determination of molecular structures from ground state rotational constants, J. Chem. Phys. 29 (4) (1958) 864.
[7] J. Kraitchman, Determination of molecular structure from microwave spectroscopic data, Am. J. Phys. 21 (1) (1953) 17.
[8] T. Helgaker, P. Jørgensen, J. Olsen, Electronic-Structure Theory, Wiley, Chichester, 2000.
[9] C. Puzzarini, Accurate molecular structures of small- and medium-sized molecules, Int. J. Quantum Chem. 116 (21) (2016) 1513.
[10] C. Puzzarini, J.F. Stanton, J. Gauss, Quantum-chemical calculation of spectroscopic parameters for rotational spectroscopy, Int. Rev. Phys. Chem. 29 (2) (2010) 273–367.
[11] I. Shavitt, R.J. Bartlett, Many-body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory, Cambridge University Press, 2009.
[12] P. Pulay, W. Meyer, J.E. Boggs, Cubic force constants and equilibrium geometry of methane from Hartree-Fock and correlated wavefunctions, J. Chem. Phys. 68 (11) (1978) 5077.
[13] J. Demaison, J.E. Boggs, A.G. Császár (Eds.), Equilibrium Molecular Structures: From Spectroscopy to Quantum Chemistry, CRC Press, Taylor & Francis Group, Boca Raton, FL, US, 2011.
[14] N. Vogt, J. Demaison, E.J. Cocinero, P. Écija, A. Lesarri, H.D. Rudolph, J. Vogt, The equilibrium molecular structures of 2-deoxyribose and fructose by the semiexperimental mixed estimation method and coupled-cluster computations, Phys. Chem. Chem. Phys. 18 (2016) 15555–15563.
[15] D.A. Obenchain, L. Spada, S. Alessandrini, S. Rampino, S. Herbers, N. Tasinato, M. Mendolicchio, P. Kraus, J. Gauss, C. Puzzarini, J.-U. Grabow, V. Barone, Unveiling the sulfur–sulfur bridge: Accurate structural and energetic characterization of a homochalcogen intermolecular bond, Angew. Chem. Intern. Ed. 57 (48) (2018) 15822–15826.
[16] L. Uribe, S. Di Grande, M. Mendolicchio, N. Tasinato, V. Barone, Accurate structure and spectroscopic properties of azulene and its derivatives by means of pisa composite schemes and vibrational perturbation theory to second order, J. Phys. Chem. A 128 (49) (2024) 10474–10488.
[17] A. Melli, F. Tonolo, V. Barone, C. Puzzarini, Extending the applicability of the semi-experimental approach by means of "Template molecule" and "Linear regression" models on top of DFT computations, J. Phys. Chem. A 125 (45) (2021) 9904.
[18] H. Ye, S. Alessandrini, M. Melosso, C. Puzzarini, Exploiting the "Lego brick" approach to predict accurate molecular structures of PAHs and PANHs, Phys. Chem. Chem. Phys. 24 (38) (2022) 23254.
[19] S. Alessandrini, M. Melosso, L. Bizzocchi, V. Barone, C. Puzzarini, The semiexperimental approach at work: Equilibrium structure of radical species, J. Phys. Chem. A. 128 (29) (2024) 5833.
[20] S. Alessandrini, H. Ye, M. Biczysko, C. Puzzarini, Describing the disulfide bond: From the density functional theory and back through the "Lego brick" approach, J. Phys. Chem. A 128 (43) (2024) 9383.
[21] M. Piccardo, E. Penocchio, C. Puzzarini, M. Biczysko, V. Barone, Semi-experimental equilibrium structure determinations by employing B3LYP/SNSD anharmonic force fields: Validation and application to semirigid organic molecules, J. Phys. Chem. A 119 (10) (2015) 2058.

[22] E. Penocchio, M. Piccardo, V. Barone, Semiexperimental equilibrium structures for building blocks of organic and biological molecules: The B2PLYP route, J. Chem. Theory Comput. 11 (10) (2015) 4689.

[23] G. Ceselin, N. Tasinato, V. Barone, Accurate biomolecular structures by the nano-LEGO approach: Pick the bricks and build your geometry, J. Chem. Theory Comput. 17 (11) (2021) 7290.

[24] V. Barone, G. Ceselin, F. Lazzari, N. Tasinato, Toward spectroscopic accuracy for the structures of large molecules at DFT cost: Refinement and extension of the nano-LEGO approach, J. Phys. Chem. A 127 (24) (2023) 5183.

[25] D.E. Woon, The astrochymist: An internet resource for astrochemists and interested bystanders, 2004, https://www.astrochymist.org/.

[26] G. Santra, N. Sylvetsky, J.M.L. Martin, Minimally empirical double-hybrid functionals trained against the GMTKN55 database: revDSD-PBEP86-d4, revDOD-PBE-d4, and DOD-SCAN-D4, J. Phys. Chem. A 123 (24) (2019) 5129–5143.

[27] L. Goerigk, S. Grimme, Efficient and accurate double-hybrid-meta-GGA density functionals-evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions, J. Chem. Theory Comput. 7 (2) (2011) 291–309.

[28] S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory, J. Comp. Chem. 32 (7) (2011) 1456.

[29] E. Papajak, J. Zheng, X. Xu, H.R. Leverentz, D.G. Truhlar, Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions, J. Chem. Theory Comput. 7 (10) (2011) 3027.

[30] C. Møller, M.S. Plesset, Note on an approximation treatment for many-electron systems, Phys. Rev. 46 (1934) 618.

[31] I.M. Mills, Vibration-rotation structure in asymmetric- and symmetric-top molecules, in: K.N. Rao, C.W. Matthews (Eds.), Molecular Spectroscopy: Modern Research, Academic Press Inc., New York, 1972, p. 115.

[32] J.F. Stanton, J. Gauss, M.E. Harding, P.G. Szalay with contributions from, A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, J. Bomble, L. Cheng, O. Christiansen, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W.J. Lauderdale, D.A. Matthews, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vüzquez, F. Wang, J.D. Watts, A.T. the integral packages MOLECULE, J. Almlöf, P.R. Taylor, PROPS P.R. Taylor, A.T. Helgaker, H.J.A. Jensen, P. Jørgensen, J. Olsen, E.A.V. Mitin, C. van Wüllen, CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package, (For the current version, see http://www.cfour.de).

[33] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian 16 revision c.01, 2016, Gaussian Inc. Wallingford CT.

[34] V. Barone, G. Ceselin, M. Fusè, N. Tasinato, Accuracy meets interpretability for computational spectroscopy by means of hybrid and double-hybrid functionals, Front. Chem. 8 (2020) 584203.

[35] P. Botschwina, Accurate equilibrium structures for small polyatomic molecules, radicals and carbenes, Mol. Phys. 103 (10) (2005) 1441.

[36] A.G. Maki, High-resolution infrared spectrum of cyanogen, J. Mol. Spectrosc. 269 (2) (2011) 166.

[37] K.W. Brown, J.W. Nibler, K. Hedberg, L. Hedberg, Structure of dicyanoacetylene by electron diffraction and coherent rotational Raman spectroscopy, J. Phys. Chem. 93 (15) (1989) 5679.

[38] S. Alessandrini, J. Gauss, C. Puzzarini, Accuracy of rotational parameters predicted by high-level quantum-chemical calculations: Case study of sulfur-containing molecules of astrochemical interest, J. Chem. Theory Comput. 14 (10) (2018) 5360–5371.

[39] C. Puzzarini, Accurate thermochemistry and spectroscopy of the oxygen-protonated sulfur dioxide isomers, Phys. Chem. Chem. Phys. 13 (48) (2011) 21319–21327.

[40] C. Puzzarini, Ab initio anharmonic force field and equilibrium structure of the sulfonium ion, J. Mol. Spectr. 242 (1) (2007) 70–75.

[41] P. Botschwina, A. Heyl, M. Horn, J. Flugge, Calculated spectroscopic constants and the equilibrium geometry of $HCNH^+$, J. Mol. Spectr. 163 (1) (1994) 127–137.

[42] C. Puzzarini, G. Cazzoli, Equilibrium structure of protonated cyanogen, $HNCCN^+$, J. Mol. Spectr. 256 (1) (2009) 53–56.

[43] W.G.D.P. Silva, J. Cernicharo, S. Schlemmer, N. Marcelino, J.C. Loison, M. Agúndez, D. Gupta, V. Wakelam, S. Thorwirth, C. Cabezas, B. Tercero, J.L. Doménech, R. Fuentetaja, W.J. Kim, P. de Vicente, O. Asvany, Discovery of H_2CCCH^+ in TMC-1, Astron. Astrophys. 676 (2023) L1.

[44] O. Martinez Jr, V. Lattanzi, S. Thorwirth, M.C. McCarthy, Detection of protonated vinyl cyanide, CH_2CHCNH^+, a prototypical branched nitrile cation, J. Chem. Phys. 138 (9) (1970) 094316.

[45] P. Botschwina, R. Oswald, Explicitly correlated coupled cluster calculations for the benzenium ion ($C_6H_7^+$) and its complexes with Ne and Ar, J. Phys. Chem. A 115 (46) (2011) 13664.

[46] K. Chatterjee, O. Dopfer, Infrared signatures of protonated benzonitrile, J. Phys. Chem. A 865 (2) (2018) 114.

[47] T.J. Lee, C.E. Dateo, Accurate spectroscopic characterization of $^{12}C^{14}N^-$, $^{13}C^{14}N^-$, and $^{12}C^{15}N^-$, Spectrochim. Acta A Mol. Biomol. Spectrosc. 55 (3) (1999) 739–747.

[48] P. Thaddeus, B.E. Turner, Confirmation of interstellar $N2H^+$, Astrophys. J. 201 (1975) L25.

[49] S. Green, J.A. Montgomery Jr., P. Thaddeus, Tentative identification of U93.174 as the molecular ion N_2H^+, Astrophys. J. 193 (1974) L89.

[50] P. Thaddeus, M. Guelin, R.A. Linke, Three new "Nonterrestrial" molecules, Astrophys. J. 246 (1981) L41.

[51] C. Cabezas, M. Agúndez, N. Marcelino, B. Tercero, Y. Endo, R. Fuentetaja, J.R. Pardo, Discovery of the elusive thioketenylium, $HCCS^+$, in TMC-1, Astron. Astrophys. 657 (2022) L4.

[52] J. Pety, P. Gratier, V. Guzmán, E. Roueff, M. Gerin, J.R. Goicoechea, S. Bardeau, A. Sievers, F. Le Petit, J. Le Bourlot, A. Belloche, D. Talbi, The IRAM-30 m line survey of the horsehead PDR. II. First detection of the l-C_3H^+ hydrocarbon cation, Astron. Astrophys. 548 (2012) A68.

[53] J. Cernicharo, N. Marcelino, M. Agúndez, Y. Endo, C. Cabezas, C. Bermúdez, B. Tercero, P. de Vicente, Discovery of HC_3O^+ in space: The chemistry of O-bearing species in TMC-1, Astron. Astrophys. 642 (2020) L17.

[54] N. Marcelino, M. Agúndez, B. Tercero, C. Cabezas, C. Bermúdez, J.D. Gallego, P. de Vicente, J. Cernicharo, Tentative detection of HC_5NH^+ in TMC-1, Astron. Astrophys. 643 (2020) L6.