A Large-scale Approach to Modeling Molecular Biosignatures: The Diatomics

Thomas M. Cross[1] , David M. Benoit[1] , Marco Pignatari[1,2,3,4] , and Brad K. Gibson[1,3]

[1] E.A. Milne Centre for Astrophysics, Department of Physics and Mathematics, University of Hull, HU6 7RX, UK; T.Cross-2019@hull.ac.uk
[2] Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences, Konkoly Thege Miklos ut 15-17, H-1121 Budapest, Hungary
[3] Joint Institute for Nuclear Astrophysics—Center for the Evolution of the Elements, USA

Abstract

This work presents the first steps to modeling synthetic rovibrational spectra for all molecules of astrophysical interest using a new approach implemented in the Prometheus code. The goal is to create a new comprehensive source of first-principles molecular spectra, thus bridging the gap for missing data to help drive future high-resolution studies. Our primary application domain is for molecules identified as signatures of life in planetary atmospheres (biosignatures), but our approach is general and can be applied to other systems. In this work we evaluate the accuracy of our method by studying four diatomic molecules, H_2, O_2, N_2, and CO, all of which have well-known spectra. Prometheus uses the transition-optimised shifted Hermite (TOSH) theory to account for anharmonicity for the fundamental $\nu = 0 \to \nu = 1$ band, along with thermal-profile modeling for the rotational transitions. To this end, we expand TOSH theory to enable the modeling of rotational constants. We show that our simple model achieves results that are a better approximation of the real spectra than those produced through an harmonic approach. We compare our results with high-resolution HITRAN and ExoMol spectral data. We find that modeling accuracy tends to diminish for rovibrational transition away from the band origin, thus highlighting the need for the theory to be further adapted.

1. Introduction

In the past few decades, the search for the origin of life in the universe and the detection of its chemical signatures and primary building blocks has driven remarkable scientific achievements in astronomy and astrobiology (see, e.g., Des Marais et al. 2008). A fundamental step to answer some of those questions is the observation of biologically relevant molecules in the universe. However, apart from extremely well-studied molecules such as water (Viti et al. 1997; Polyansky et al. 2018), ammonia (Yurchenko et al. 2009; Coles et al. 2019), methane (Brown et al. 2013; Yurchenko & Tennyson 2014), and methanol (Falk & Whalley 1961; Beć et al. 2016), the basic data needed to detect most biomolecules are incomplete. Existing data was gathered through meticulous experiments and computations and built from the bottom up to cover about a hundred of molecules and their rovibrational spectra (Tennyson et al. 2020). The molecular data have been carefully curated into well-known databases in the field such as the HIgh-resolution TRANsmission molecular absorption database (HITRAN; Gordon et al. 2017), ExoMol (Tennyson et al. 2020), CDMS (Endres et al. 2016), and JPL (Pearson et al. 2005), for example. Today these databases are fundamental resources for molecular line detection and form the backbone of astrochemistry and astrobiology research using data from present observation missions like the Atacama Large Millimeter/submillimeter Array (ALMA; Harada et al. 2018), the Hubble Space Telescope (HST; Evans et al. 2018; Damiano

et al. 2017), the Transiting Exoplanet Survey Satellite (TESS; Colón et al. 2020) and from future missions like PLanetary Transits and Oscillations of stars (PLATO; Katyal et al. 2019), the James Webb Space Telescope (JWST; Barstow et al. 2015), and Ariel (e.g., Tinetti et al. 2018), to name a few.

Molecular lines are ideal observation targets to study different astrophysical sources. Fundamental information about the properties of the early universe can be collected from integrated galactic spectra at different redshifts (e.g., Muller et al. 2006; Costagliola et al. 2011; Aladro et al. 2015; Zhang et al. 2018), or from a number of local interstellar medium sources (e.g., Bacmann et al. 2012; Rivilla et al. 2016; Shimajiri et al. 2017), from stars (e.g., Yong et al. 2003; Hedrosa et al. 2013), and from both planets (e.g., Greaves et al. 2021; Webster et al. 2015; Tran et al. 2006) and exoplanets (e.g., Swain et al. 2009; Tinetti et al. 2013; Tessenyi et al. 2013; Guilluy et al. 2019).

The evolution of chemistry in the universe has captured the increase of complexity from a metal-free environment (e.g., Galli & Palla 1998) to a metal-, dust- and ice-rich environment (e.g., Morales et al. 1998; Wakelam & Herbst 2008). The stellar production of elements, such as oxygen, silicon, and iron from the first generation of core-collapse supernovae (e.g., Rauscher et al. 2002; Nomoto et al. 2013; Sukhbold et al. 2016; Ritter et al. 2018) and carbon and nitrogen from low-mass stars (e.g., Karakas & Lattanzio 2014; Cristallo et al. 2015; Pignatari et al. 2016), formed the building blocks of biomolecules that are observed today.

Within this same context, the search for molecular fingerprints of life in the atmospheres of exoplanets represents a fundamental goal of astrobiology. A biosignature gas is defined as a gas that is produced by life and accumulates in a planet's atmosphere. An ideal biosignature would be unambiguous with living organisms being its unique source (Greaves et al. 2021). In reality, many

---

[4] NuGrid Collaboration, http://nugridstars.org.

biosignatures can also be produced through abiotic processes and can act as false positives.

For a molecule to be classified as a biosignature under the framework of Seager et al. (2016), the molecule needs to fulfill a series of criteria. Here we will detail what we believe to be the three main points; further classification details can be found in the paper referenced. The first main criterion is molecular stability, identifying compounds that are stable on the order of days as a pure compound at standard conditions and if they are stable to reactions with water. The second main criterion is volatility: the likelihood a molecule would be in gaseous form at standard conditions. Volatility is difficult to assess so Seager et al. (2016) used boiling points instead. A molecule with a boiling below 150 °C was considered sufficiently volatile. The third main criterion focuses on molecular size. Only molecules of up to six non-hydrogen atoms were considered during the search. This criterion was used to limit the number of possible molecules as their number increases exponentially with each additional non-H atom. Moreover, smaller molecules are more likely to be found in gaseous form in planetary atmospheres.

The All Small Molecules (ASM) catalog described by Seager et al. (2016) contains over 14,000 biosignature molecules that comply with the criteria discussed above. For ease throughout this paper we will refer exclusively to the biosignature portion of this catalog as the SBP, an acronym which is simply made from the first letter from each of its author's last names, as ASM also details nonbiogenic molecules.

In order to be able to detect molecules in the SBP catalog, there is a need for the laboratory astrophysics community to characterize the spectral features of each entry. Despite the simple nature of the entries in the catalog, and considering only the rovibrational portion of the spectrum, there are thousands of biosignatures within SBP that have incorrect, incomplete, or completely unknown spectra (Sousa-Silva et al. 2019).

There is therefore a pressing need for vast quantities of spectral signatures to be characterized, as shown by the very recent spectroscopic discovery of phosphine on Venus by Greaves et al. (2021). Not only is this molecule believed to be a biosignature but it was detected in such quantities that it cannot be currently explained abiotically (Bains et al. 2021). One could argue that the spectroscopic analysis which discovered phosphine would have not been possible without the efforts to produce accurate and complete computed line lists, ranging from room temperature (Sousa-Silva et al. 2013) to up to 1500 K (Sousa-Silva et al. 2015). Especially when we consider that suspected inaccuracies found in regions of previous phosphine data might have contributed to past misinterpretations of astronomical spectra (Sousa-Silva et al. 2015; as an example, the PH_3 database was suspected of containing some inaccuracies for the 4.5 μm line intensities, as discussed in Malathy Devi et al. 2014).

Following this discovery, Zapata Trujillo et al. (2021) noted the importance of being able to detect spectroscopically the presence of phosphorous-bearing species. They enumerated a list of phosphorous-bearing species, which could potentially be detected in planetary atmospheres, and compiled all available spectral data. As expected, the data were scarce, therefore they used established computational quantum chemistry methods to produce approximate spectra to fill in the gaps.

Generating an entire set of SBP spectra for a given frequency range is a huge task. Indeed, experimentally it would be hard to

record spectra for the entire catalog without a concerted worldwide effort.

Alternatively, a significant part of the SBP catalog could be explored computationally. This is our current approach for this study, which focuses on the rovibrational spectra of diatomic molecules in SBP. Nonetheless, the envisaged task requires a reliable means of producing good-quality spectral data at reasonably low computational cost so that it can later be easily extended to produce spectral signatures for the rest of the catalog.

The current publicly available version of HITRAN contains detailed spectra for 49 molecules and also atomic oxygen. A further description of HITRAN is provided in Section 3. Out of those 49 molecules, only 26 are actually believed to be biosignatures according to SBP. To break this down even further, Figure 1 compares the number of biogenic molecules for each total number of atoms, up to pentatomics. Rounding the SBP down to exactly 14,000 biosignatures gives HITRAN a completion rate of roughly 0.19%.

The current version of ExoMol (see also further background in Section 3) contains high-quality spectra at a large range of temperatures for 81 molecules; however, only 18 of these molecules are biogenic. We report in Figure 1 a similar breakdown to the one done for HITRAN. Here we see that ExoMol has a completion rate of approximately 0.13%.

Combining both databases gives a maximum total of 26 molecules (excluding SBP entries overlaps). Assuming a constant rate of growth for these databases, without significant change to available technology and resources, it will be difficult to produce spectra for even 1% of SBP, let alone the entire catalog.

The work described here provides a new approach that is meant to complement and support the high-resolution databases mentioned above. We aim to obtain spectra from first principles, but faster than existing approaches (such as vibrational self-consistent field/vibrational configuration inter-action, shown by Clark & Benoit 2020 or discrete variable representation, shown by McKemmish et al. 2019, for example) possibly at the expense of accuracy.

Primarily we are trying to address the gap between approximate, fundamental-only models (such as harmonic, nonadapted transition-optimised shifted Hermite (TOSH) or Rapid Approximate Spectral Calculations for ALL (RAS-CALL)), and labor-intensive, extremely accurate line lists (such as ExoMol and HITRAN). This is done by developing a method which borrows key physics from both approaches to create an intermediate approach. We present here a model that approximates the fundamental ($\nu = 0 \rightarrow \nu = 1$) band *and* its rotational profile, computed at 300 K. The intensities for the spectra use thermal population equations, and as such have the additional ability of modeling rovibrational lines at differing temperatures. The investigation into the effect of differing temperatures on spectra is not considered in this paper.

Additionally our approach is designed to be simple, open source, and computationally cheap, but still more detailed than the aforementioned approximate fundamental-only methods. For example, we do not currently model effects such as external electric or magnetic fields (see, for example, PGOPHER; Western 2017), to preserve simplicity and low computational cost.

Once the catalog is constructed it should, and will, act as a living document, preferably being continually updated with
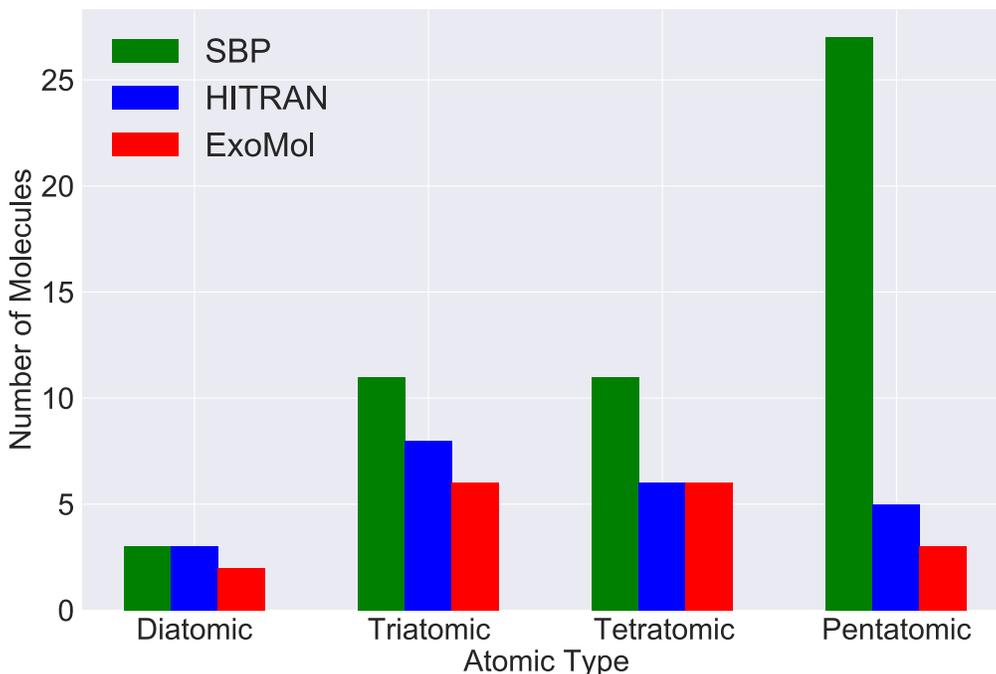
Figure 1. Comparing the availability of data from HITRAN and ExoMol with the molecules needed for the SBP. Only up to pentatomic molecules have been included here.

better results. The rationale for this approach is to provide a wide coverage of the SBP database to help with both detection and atmospheric models with a reasonable accuracy that could be improved with further releases of the database.

In this study we have applied our work to the diatomic biosignature molecules, of which there are three in the SBP catalog. We have also additionally included a fourth diatomic, carbon monoxide (CO), for reasons discussed in Section 3.4.

The paper is organized in the following manner. In Section 2, we introduce the methodology and theory required to produce the synthetic data. The following section, Section 3, presents and then discusses our results. We conclude our findings in Section 4.

2. Methodology

2.1. Requirements

When modeling the rovibrational spectrum, we considered two key parameters: the band origin and the rotational constants. The band origin crucially determines the position of its corresponding rotational lines and thus has a strong influence on the overall shape of the rovibrational band. In our model, we have assumed this aspect to be the most important.

The rotational constants have an effect on the spacing of the transitions on each of the rotational branches. As a secondary effect, anharmonicity causes the rotational constant to depend on the vibrational state. Indeed, the bond lengthening in excited vibrational states allows a Q-branch progression (a set of purely vibrational transitions with no rotational transitions; see Equation (28)), since the rotational constants of both starting and final vibrational states are different. An identical rotational constant for all vibrational level (such as predicted by a simple harmonic oscillator approximation, for example) incorrectly suggests that vibrations and rotations are independent. This leads to the Q-branch transitions bunching up at a single position, rather than displaying a typical progression.

Therefore, for an accurate approximation for the entirety of the spectrum, the rotational constant will need to vary with vibrational level (Pawłowski et al. 2002).

The TOSH theory is a modern approach to treating the nuclear wave function that does predict correctly most band origins (Lin et al. 2008). TOSH is effectively a simplification of second-order vibrational perturbation theory (VPT2), and applies limited-order perturbation theory to the nuclear Schrödinger equation. By doing this, it avoids any degeneracy issue that may arise from standard second-order vibrational perturbation theory (Willetts et al. 1990; Vázquez & Stanton 2006), for example.

Fundamentally TOSH works by introducing a shifting parameter (σ) which is used to shift the harmonic basis functions from their equilibrium position. This shift is optimized for the vibrational transition energy expansion, specifically for the fundamental vibrational transition, and as such does an excellent job at correcting the band origins (this is shown by our results in Section 3 and Table 2). An expression for the magnitude of the shift is described in Equations (2)–(10) in Section 2.2. We have discovered, however, that this shift can be applied to the equilibrium distance to recover the anharmonic effects of the first vibrationally excited level (see Section 2.6).

Despite its approximate nature, TOSH is roughly a factor of two cheaper than VPT2, with only some loss in accuracy as a compromise (Lin et al. 2008). This approach also has the potential to be of reasonable accuracy for a large ensemble of molecules (Hanson-Heine 2019) and is straightforward to extend to larger systems (Lin et al. 2008).

2.2. Transition-optimised Shifted Hermite Theory

We will now briefly discuss the necessary theory required to understand the process to produce the Prometheus spectra.

For a diatomic molecule, the Hamiltonian with up to fourth-order terms (quartic potential) and using mass-weighted

displacement coordinate, $Q = \sqrt{\mu}x$, can be expressed as

$$\hat{H} = -\frac{1}{2}\frac{\partial^2}{\partial Q^2} + \frac{1}{2!}\eta_{ii}Q^2 + \frac{1}{3!}\eta_{iii}Q^3 + \frac{1}{4!}\eta_{iiii}Q^4. \qquad (1)$$

Now imagine a shift, σ, along the coordinate Q. If the center of the wave function is shifted by σ, the shape will remain the same but the anharmonic correction can be incorporated into the wave function.

The shifted wave functions for TOSH are now different to the harmonic wave functions, and are described by

$$\psi_n = \left(\frac{\omega^{1/2}}{\pi^{1/2}2^n n!}\right)^{1/2} e^{-\frac{\omega}{2}(Q-\sigma)^2} H_n[(Q - \sigma)\omega^{1/2}], \qquad (2)$$

where $H_n(x)$ naturally refers to the Hermite polynomial, and ω describes the harmonic frequency (defined using TOSH constants in Equation (13); see below).

The energy of this ground vibrational state is

$$E_0 = \langle \psi_0|\hat{H}|\psi_0 \rangle, \qquad (3)$$

$$E_0 = \frac{1}{4}\omega + \frac{1}{2!}\eta_{ii}\left[\frac{1}{2\omega} + \sigma^2\right] + \frac{1}{3!}\eta_{iii}\left[\frac{3\sigma}{2\omega} + \sigma^3\right]$$
$$+ \frac{1}{4!}\eta_{iiii}\left[\frac{3}{4\omega^2} + \frac{3\sigma^2}{\omega} + \sigma^4\right]. \qquad (4)$$

The equation for the energy of the first excited vibrational level is

$$E_1 = \langle \psi_1|\hat{H}|\psi_1 \rangle, \qquad (5)$$

$$E_1 = \frac{3}{4}\omega + \frac{1}{2!}\left[\frac{3}{2\omega} + \sigma^2\right]\eta_{ii} + \frac{1}{3!}\eta_{iii}\left[\frac{9\sigma}{2\omega} + \sigma^3\right]$$
$$+ \frac{1}{4!}\eta_{iiii}\left[\frac{15}{4\omega^2} + \frac{9\sigma^2}{\omega} + \sigma^4\right]. \qquad (6)$$

Therefore the energy difference between the first vibrational state and the ground state is

$$\Delta E^{\mathrm{TOSH}} = \langle \psi_1^*|\hat{H}|\psi_1 \rangle - \langle \psi_0^*|\hat{H}|\psi_0 \rangle, \qquad (7)$$

$$= \omega + \frac{\eta_{iiii}}{8\omega^2} + \frac{\eta_{iii}\sigma}{2\omega} + \frac{\eta_{iiii}\sigma^2}{4\omega}. \qquad (8)$$

Which can now be compared with the energy from the unshifted wave function obtained through second-order perturbation theory, VPT2:

$$\Delta E^{\mathrm{VPT2}} = \omega + \frac{\eta_{iiii}}{8\omega^2} - \frac{5\eta_{iii}^2}{24\omega^4} - \frac{\eta_{iiii}^2}{32\omega^4}. \qquad (9)$$

Within the TOSH theory derivations, σ is assumed to be small, and therefore the σ^2 term can be neglected. By comparing the coefficient of η_{iii} in both the TOSH (8) and VPT2 (9) expressions, a suitable value for the shift can be obtained:

$$\sigma = -\frac{5}{12}\frac{\eta_{iii}}{\omega^3}. \qquad (10)$$

Due to its derivation, this value of the shift parameter is only optimal for the $0 \to 1$ transition. It is also worth noting that the original TOSH paper of Lin et al. (2008) does not match the η_{iiii} terms for the VPT2 energy expression, but remedied this by stating this term is often neglected, in any case. Many intermediate processes for the derivations above have been

assumed and therefore omitted. For a full derivation, please see the Appendix, Section A.1.

One lesser-considered aspect of having a displaced vibrational wave function (mimicking what happens for the exact wave function) is the usage the shift parameter in order to include anharmonic effects in the computation of rotational constants. We explore this possibility in Section 2.6.

2.3. Potential Energy Curve

As shown in Equations (8) and (10), the TOSH approach (and our Prometheus implementation; see Section 2.8) only requires a selection of anharmonic constants for the chosen diatomic. These constants can be obtained either through finite-difference methods or by computing the derivatives of the potential energy curve (PEC) directly, if a functional form is available. In this study, we do not focus on producing PECs, but we use existing ones to derive the required anharmonic constants using a quartic fitting procedure described below.

Prometheus uses the numerical points that make up the PEC to perform a quartic fit (see Equation (11)) and extract the anharmonic constants. We use this methodology for two reasons. Historically quartic force fields have been used as a generic polynomial form to model PECs. In particular, it is used as a tool for analyzing and producing rovibrational spectra for molecules of interest to astrophysical observation (Fortenberry & Lee 2019). Second, the TOSH theory requires only the second-, third- and fourth-order derivatives, without any need for higher-order derivatives.

The quality of the PEC is paramount and will affect the accuracy of the TOSH constants derived from our quartic fit. The code performs a local fit, centered around the equilibrium bond length, rather than a global fit, which would include the potential wall. The local fit is determined via an "inclusion range" determined by starting with the equilibrium bond length and including points at differing percentages of the equilibrium bond length. The ranges tested within the code span from $\pm20\%$ to $\pm7\%$ of the equilibrium bond length, in intervals of 1%.

The fit accuracy has been tested for each range using a simple chi-square test. The range with the lowest chi-square value, or the first range with a sufficiently low chi-square value (arbitrarily defined as 1×10^{-7}) was selected. This was to ensure the range providing the best fit (while still considering the point density of the remaining curve) to the reference data is selected and used to ultimately produce a rovibrational spectrum. This method has been chosen to ensure a consistent selection of inclusion ranges for the PEC of each molecule. It also removes any bias we could have introduced into the analysis by selecting the ranges ourselves.

The quartic fit applied to the PEC is described by the equation

$$V = E_0 + \frac{1}{2!}\zeta_{ii}X^2 + \frac{1}{3!}\zeta_{iii}X^3 + \frac{1}{4!}\zeta_{iiii}X^4, \qquad (11)$$

where $X = r - r_e$ is a displacement coordinate, with r_e describing the equilibrium bond length. Note that E_0 has been included to help Prometheus with the fitting by centering around 0, and has little importance beyond this.

The fit only returns ζ constants, ζ_{ii}, ζ_{iii}, and ζ_{iiii}, which need to be massed weighted for them to become the TOSH constants, as described in the Hamiltonian in Equation (1).

Table 1
Anharmonic Constants Computed by Prometheus (TOSH) for Each Molecule

Constant	H_2	O_2	N_2	CO
η_{ii} ($E_h a_0^{-2} m_e^{-1}$)	4.040×10^{-4}	5.183×10^{-5}	1.156×10^{-4}	9.803×10^{-5}
η_{iii} ($E_h a_0^{-3} m_e^{-3/2}$)	-4.971×10^{-5}	-1.718×10^{-6}	-4.048×10^{-6}	-3.430×10^{-6}
η_{iiii} ($E_h a_0^{-4} m_e^{-2}$)	5.317×10^{-6}	4.789×10^{-8}	1.069×10^{-7}	8.928×10^{-8}

Note. Atomic units are used throughout, where E_h, a_0, and m_e represent Hartrees, Bohr radius, and electron rest mass, respectively. The constants are derived from a quartic fit to the numerical PECs (see Section 2.3 for details).

This was done in the following manner:

$$\eta_{ii} = \frac{\zeta_{ii}}{\mu}, \quad \eta_{iii} = \frac{\zeta_{iii}}{\mu^{3/2}}, \quad \eta_{iiii} = \frac{\zeta_{iiii}}{\mu^2}. \quad (12)$$

Note that fits of order higher than quartic lead to sizable errors and inaccurate results. Additionally, we used spline interpolation to ensure the data spanned the entire range specified.

2.4. Choice of a PEC for Diatomic Anharmonicity Constants Determination

In this study, we obtain the anharmonicity constants by computing the derivatives of a quartic fit to a given PEC. As previously alluded to, these constants are the only data required for Prometheus to operate. The technique used to derive these constants (finite difference, derivatives of PEC, etc.) are inconsequential, but require a suitable description of the interatomic potential around the equilibrium bond length in order to produce reliable spectra. In order to avoid issues originating from a PEC of subspectroscopic quality, we choose to use only data that have been validated through spectroscopic comparisons.

Indeed, while a number of PECs for diatomics are available in the literature, some of the molecules in our selected set have a particularly challenging electronic structure (CO, N_2 and O_2). This in turn affects the quality of the anharmonicity constants, which then impacts the accuracy of the rovibrational lines. As our study focuses on the quality of the rovibrational approach, rather than the quality of the electronic structure approach used for the potential, we chose well-reported modern PECs, obtained through either Rydberg–Klein–Rees (RKR) inversion (see Rees 1947 for further details) or ab initio calculations. All the potentials used within this study are for the ground state of their respective molecule.

The PECs used are those from Schwartz & Le Roy (1987) (the mass-independent clamped nuclei, V_{BO}, potential with adiabatic correction, ΔV_{ad}) for the $^1\Sigma_g^+$ state of H_2. The PEC for $^1\Sigma^+$ for CO is from Meshkov et al. (2018) (specifically, the CO_X_Func_edited Born–Oppenheimer (UBO), potential located within the supplementary material). The PEC for the $^3\Sigma_g^-$ state of O_2 is from Yu et al. (2014) and the potential for the $^1\Sigma_g^+$ state of N_2 is from Edwards et al. (1993).

H_2 is an ab initio potential, CO a semiempirical potential, while N_2 and O_2 are spectroscopic RKR curves. This has been done in part to highlight the flexibility of our model, which can accept any type of PEC, rather than being bound to a single format. Naturally, if the quality of the curve is low, the rovibrational results will reflect this fact, but will still remain a better approximation than a pure harmonic approach.

The necessary anharmonicity or TOSH (quartic) constants, obtained using optimal quartic fitting ranges, for each molecule in this study are given in Table 1.

2.5. Spectroscopic Constants

The TOSH framework allows us to determine key constants, such as the harmonic frequency, ω, the anharmonic fundamental transition, ΔE^{TOSH} or $\nu_{0\to1}$, and the coordinate shift, σ.

The value of ω is calculated using

$$\omega = (\eta_{ii})^{1/2}. \quad (13)$$

The value of ΔE^{TOSH} is obtained from Equation (8), and the shift, σ, is given by Equation (10). The associated errors for each spectroscopic constant is also calculated to allow an assessment of the error on the positions of transitions in the spectra produced by Prometheus (see Section 2.8).

2.6. Rotational Constants

The state-specific rotational constant, B_v, for a diatomic molecule is defined as

$$B_v = \frac{h^2}{8\pi^2 c \mu r_v^2}, \quad (14)$$

where v refers to the vibrational level, μ is the reduced mass, and r_v is a bond length that depends on the vibrational energy level considered. We can see that the bond constant has a reciprocal squared relationship to the bond length, meaning as length increases the constant will decrease. This is one of the fundamental equations used throughout Prometheus to ultimately create the rovibrational spectra.

In the harmonic approximation, r_v is independent of the vibrational energy level and thus $r_v = r_e$, the equilibrium bond length. Consequently, in that approximation $B_v = B_e$, a fixed rotational constant obtained from r_e. More generally, the value of the bond length for the rotational constant for each vibrational state can be obtained from the expectation value of the position, $r_v = \langle \psi_n | r | \psi_n \rangle$. Using this expression for the harmonic model leads to the same conclusions as earlier: $r_v = r_e$, since the harmonic wave function is symmetric around r_e.

If we now consider the TOSH model, the shifted TOSH position, x, can be described using the harmonic position, u, and the shift constant, σ:

$$x = u + \sigma. \quad (15)$$

As shown earlier, we can take the expectation value of the position, but this time implementing TOSH:

$$\langle \psi_n^T | x | \psi_n^T \rangle = \langle \psi_n^H(u) | u + \sigma | \psi_n^H(u) \rangle, \quad (16)$$

where ψ_n^T and ψ_n^H represent the TOSH (Equation (2)) and harmonic wave functions, respectively:

$$\langle \psi_n^T | x | \psi_n^T \rangle = \langle \psi_n^H(u) | u | \psi_n^H(u) \rangle + \langle \psi_n^H(u) | \sigma | \psi_n^H(u) \rangle. \quad (17)$$

The first expectation value is the position expectation position for a harmonic wave function (i.e., zero) and thus we can see that the expectation of the position for TOSH is simply

$$\langle \psi_n^T | x | \psi_n^T \rangle = \sigma. \quad (18)$$

The full derivation, for both required levels, can be found in the Appendix, Section A.2.

Thus the TOSH model leads to $r_\nu = r_e + \sigma$. As is the case for the harmonic model, the TOSH approach only produces a single static value for all rotational constants (see Section A.2).

We cannot accurately reproduce rotational constants that exhibit rotation–vibration interaction, using solely either a TOSH or a harmonic model: a hybrid theory is required. The approach we use in Prometheus approximates the ground-state rotational constant using the harmonic value. The first excited vibrational level uses the modified TOSH bond length derived above, where sigma is combined with the equilibrium bond length. The vibrationally averaged bond distances are thus expressed as follows:

$$r_0 = r_e, \quad (19)$$

$$r_1 = r_e + \sigma, \quad (20)$$

where σ is not mass weighted to harmonise units.

The harmonic approach effectively costs nothing (using $r_0 = r_e$), so can be exploited to provide the ground-state rotational constant with little effort. Our approach does not necessarily provide a strong quantitative agreement but is qualitatively correct: bond length gets larger with increased vibration, hence the rotational constant changes (gets smaller).

2.7. Rovibrational Spectra

In order to generate a rovibrational line list and corresponding spectrum, we compute the maximum allowed quantum rotational number, J, for a given input temperature. In the present study, the temperature was set to 300 K to match the other data sets. The spectra from each of the experimental databases are all set at 300 K, also.

Rotation symmetry also influences the spectrum the molecule will produce, and therefore conditional arguments within the intensity calculations have been created to account for this. If a diatomic molecule has a center of symmetry, hence meaning it is homonuclear, it is Raman active. This occurs instead of IR activity due to having no dipole moment; instead stretching and contraction of the bond leads to changes in the polarizability. This type of molecular response implies different transition selection rules leading to the O, Q, and S branches, rather than the typical P and R branches in infrared.

Nuclear spin also has an effect on the spectra produced. This will be discussed further in Section 3. The code is automated to account for additional effects by using the initial inputs of the mass of a molecule.

We use a Boltzmann distribution to obtain the relative intensities for each transition. This statistical law states that, for a system of N total molecules, only a fraction, $\frac{N_J}{N}$, will occupy

particular energy level, E_J, a fraction. This can be written as

$$\frac{N_J}{N} = \frac{g_J e^{\left(\frac{-E_J}{kT}\right)}}{f}, \quad (21)$$

where the degeneracy, g_J, the energy of a given level, E_J, and the partition function, f, are described as

$$g_J = (2J + 1), \quad (22)$$

$$E_J = B_0 J(J + 1), \quad (23)$$

$$f = \sum g_J e^{\left(\frac{-E_J}{kT}\right)}. \quad (24)$$

Hence, the full equation now becomes

$$\frac{N_J}{N} = \frac{(2J + 1) e^{\left(\frac{-B J(J+1)}{kT}\right)}}{f}. \quad (25)$$

From here the code calculates the transition positions using the appropriate equations depending on the rotational branches considered. The full derivations for these equations have been included in the Appendix, Section A.3. They are summarized as follows:

$$\nu_O = \nu_{0\to1} + B_1 J'(J' + 1) - B_0(J' + 2)(J' + 3), \quad (26)$$

$$\nu_P = \nu_{0\to1} - (B_1 + B_0)(J' + 1) + (B_1 - B_0)(J' + 1)^2, \quad (27)$$

$$\nu_Q = \nu_{0\to1} - J'(J' + 1)(B_1 - B_0), \quad (28)$$

$$\nu_R = \nu_{0\to1} + (B_1 + B_0)(J'' + 1) + (B_1 - B_0)(J'' + 1)^2, \quad (29)$$

$$\nu_S = \nu_{0\to1} + B_1(J'' + 2)(J'' + 3) - B_0 J''(J'' + 2), \quad (30)$$

where $\nu_{0\to1}$ represents the band origin, and B_0 and B_1 represent the rotational constants for their appropriate vibrational levels. As previously mentioned, for our code we use the solution calculated by TOSH (Equation (8)) as the band origin.

Here is a reminder of the labeling for the line series: $\Delta J = -2$ (O branch), $\Delta J = -1$ (P branch), $\Delta J = 0$ (Q branch), $\Delta J = +1$ (R branch), and $\Delta J = +2$ (S branch).

Finally, the code combines the intensity calculations with the transition positions to produce the synthetic spectra.

2.8. Prometheus

Prometheus is the name given to our code, which is written in Python. It is maintained by the authors and is available on the Milne Centre Github (https://github.com/Milne-Centre/Prometheus). The version of Prometheus used to obtain the results of this paper is archived at 10.5281/zenodo.5494420.

Our code is designed to be an approximate method of simulating spectra, potentially trading off accuracy for simplicity and speed. The main idea behind this code is to offer a new complementary approach to modeling the vast amount of molecules which have astrophysical and astrobiological importance.

Our analysis may have applications beyond the SBP catalog, as we have done ourselves within this work by modeling CO.

Prometheus primarily implements the TOSH theory detailed in Section 2.2 and makes use of the shift parameter, σ, to calculate band origins and provide anharmonic corrections to spectroscopic constants. To our knowledge this is the first time it has been used to provide anharmonic corrections to rotational

constants, rather than purely being used for anharmonically corrected vibrational frequency calculations.

Prometheus then produces a stick spectrum (without any line-broadening effects) using the constants it has determined.

3. Results and Discussion

The spectra Prometheus produces can be compared to different sources to evaluate the accuracy of our method. This does not necessarily mean that if the transitions line up well with another spectrum that Prometheus is correct, rather that it has a better agreement with one of the mainstream methods.

When producing spectra, our program can either use purely TOSH, purely harmonic, or a mixture of both sets of constants (such as the hybrid approach that we use within this study; detailed further in Section 2.6). Additionally, Prometheus is capable of using any other experimental/literature values for the spectroscopic constants. A caveat, however, is that although the experimental/literature results may be more accurate, they are still bound to our code's capabilities and do not include all the effects that ExoMol or HITRAN have, for example.

Within this paper we will only be considering the data from HITRAN and ExoMol, and not RASCALL. This is because, as previously mentioned, the spectra produced by RASCALL 1.0 focus mainly on approximate band centers (Sousa-Silva et al. 2019) rather than rotational structure.

The comparison data we use to evaluate the performance of Prometheus are briefly described below:

1. *Literature*. We used the spectroscopic constants published in papers that are well known and commonly cited in the spectroscopic community. We have included these results to highlight the theoretical "best" that our code could produce without a TOSH approximation. For the discussion of each diatomic, the appropriate paper has been stated and annotated on the results.
2. *HITRAN2016: data from HITRANs 2016 release*. HITRAN is an acronym for high-resolution transmission molecular absorption database and is a compilation of spectroscopic parameters that a variety of computer codes use to predict and simulate the transmission and emission of light in the atmosphere (Gordon et al. 2017). Data were obtained from the https://hitran.org website, using the line list data access interface, and directly read into our code. These data are currently acting as the co-optimal result along with ExoMol. We normalized the intensities of the lines to match the relative intensities produced by Prometheus. Future work will potentially look at the intensities of the rovibrational spectrum—whether that be regarding opacity functions or using units of atmospheric concentration rather than the current relative intensities.
3. *ExoMol*. A database which provides high-accuracy and complete line lists for application in hot astrophysical environments (Tennyson et al. 2020). The ExoMol data structure can be used to generate lifetimes, cooling functions and partition functions. More details can be found in Tennyson et al. (2016) and Chubb et al. (2021). The website http://www.exomol.com/ is the source of the ExoMol data used for comparisons. Like the HITRAN database, we have normalized the intensities of the lines to allow comparisons in intensities.
4. *Harmonic*. This is simply using Prometheus constants, calculated from PECs, without any anharmonic

corrections. For example, the band origin is the harmonic frequency, ω, instead of using Equation (7) for the origin. These data are meant to highlight how harmonic methods, while simple and easy to do, are typically the weakest in accuracy and therefore there is a need for anharmonic corrections.

The comparison spectra that are available (ExoMol, or HITRAN, or both) have been inverted on the figures below. This is merely an aesthetic choice to prevent the figures from becoming difficult to interpret, by having up to four sets of data layered upon each other.

We will briefly outline the color scheme used throughout the spectra. Prometheus is designated as red, and the spectra produced by literature constants via Prometheus as blue. The HITRAN 2016 spectra is denoted by the color black whereas the ExoMol spectra is the color purple. Finally the harmonic spectra is green.

The computed error on the line positions for the spectra generated by Prometheus are typically less than 0.1 wavenumbers, except H_2, which is less than four wavenumbers. We did not include error bars on each line position in the figures for clarity.

3.1. Molecular Hydrogen

The first biosignature molecule is molecular hydrogen. H_2 is the most abundant molecule in the universe by orders of magnitude (Wakelam et al. 2017). It is difficult to observe directly in the interstellar medium due to a lack of permanent electric dipole moment and most transitions are of quadrupolar nature (and extremely weak) (Herzberg 1949; Roueff et al. 2019), or in emission for selected environments (see Habart et al. 2005 for a review). Indirect techniques, such as the amount of dust present using the gas-to-dust ratio (Jo et al. 2017), also allow some degree of detection. See Seager et al. (2020) for further discussion of possible H_2 presence and detection on exoplanets. From a biology standpoint, many microorganisms produce hydrogen as primary product through metabolism under anaerobic conditions (Schlegel 1974).

For H_2, data from HITRAN2016 (Gordon et al. 2017) were available (specifically, the line list comes from Komasa et al. 2011 and Wolniewicz et al. 1998). The spectroscopic constants were obtained from Campargue et al. (2012). Additionally, spectra were available from ExoMol (Roueff et al. 2019).

In Table 2, Campargue et al. (2012) determined the band origin to be $4161.17 \, \text{cm}^{-1}$. The harmonic approach gives $4411.45 \, \text{cm}^{-1}$, which deviates from the literature by $250.28 \, \text{cm}^{-1}$. TOSH, with a result of $4174.71 \, \text{cm}^{-1}$, manages to produce a band origin with a difference of only $13.54 \, \text{cm}^{-1}$, an order of magnitude less than the harmonic approach.

The experimental ground-state rotational constant, B_0, was determined by Campargue et al. (2012) to be $59.3329 \, \text{cm}^{-1}$. The harmonic method approximates B_0 better, with a value of $60.9513 \, \text{cm}^{-1}$, than TOSH, which appears to drastically overcorrect, leading to a value of $54.2350 \, \text{cm}^{-1}$.

However, since the predicted rotational constants has no vibrational dependence for TOSH or the harmonic model, the reverse is true for B_1. Campargue et al. (2012) determined $56.3732 \, \text{cm}^{-1}$, with TOSH and harmonic now differing by $2.1382 \, \text{cm}^{-1}$ and $4.5781 \, \text{cm}^{-1}$, respectively.
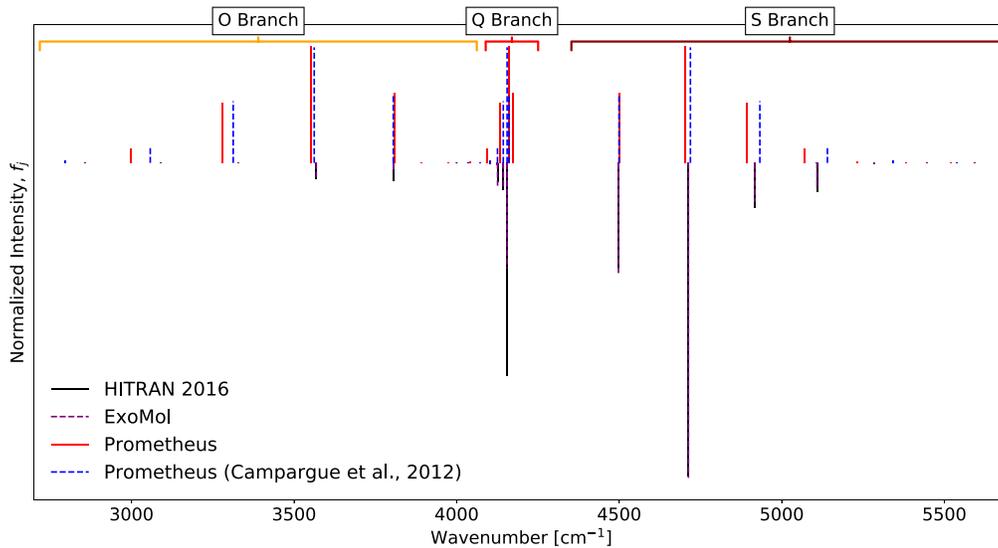
Figure 2. A comparison of the H_2 spectra produced by Prometheus, ExoMol (Roueff et al. 2019), and Prometheus using the spectroscopic constants from Campargue et al. (2012). The intensities for all sets of data have been normalized and therefore are relative values. Note that HITRAN and ExoMol transitions overlap with each other.

Table 2
A Comparison of the Spectroscopic Constants between Prometheus (TOSH), the Literature and the Harmonic Method

		Band Origin $\nu_{0\to1}$ (cm^{-1})		Rotational Constant B_0 (cm^{-1})		Rotational Constant B_1 (cm^{-1})	
		Calculated	ΔLit.	Calculated	ΔLit.	Calculated	ΔLit.
H_2	Campargue et al. (2012)	4161.17	⋯	59.3329	⋯	56.3732	⋯
	Harmonic	4411.45 ± 0.13	+ 250.28	**60.9513**	+ 1.6184	60.9513	+ 4.5781
	TOSH	**4174.71** ± 2.53	+ 13.54	54.2350 ± 0.0015	−5.0979	**54.2350** ± 0.0015	−2.1382
O_2	Yu et al. (2014)	1556.39	⋯	1.4377	⋯	1.4219	⋯
	Harmonic	1580.00	+ 23.61	**1.4456**	+ 0.0079	1.4456	+ 0.0237
	TOSH	**1556.43** ± 0.07	+ 0.04	1.4257	−0.0120	**1.4257**	+ 0.0038
N_2	Bendtsen & Rasmussen (2000)	2329.91	⋯	1.9896	⋯	1.9722	⋯
	Harmonic	2359.56 ± 0.02	+ 29.65	**1.9982**	+ 0.0086	1.9982	+ 0.0260
	TOSH	**2329.85** ± 0.14	−0.06	1.9753	−0.0143	**1.9753**	+ 0.0031
CO	Coxon & Hajigeorgiou (2004)	2143.27	⋯	1.9225	⋯	1.9050	⋯
	Harmonic	2173.07 ± 0.02	+ 29.80	**1.9316**	+ 0.0091	1.9316	+ 0.0266
	TOSH	**2143.15** ± 0.08	−0.12	1.9080	−0.0145	**1.9080**	+ 0.0030

Note. ΔLit. represents the difference between the calculated value and the appropriate literature. The highlighted values represent the values which Prometheus uses to create spectra, with the band origin and upper rotational coming from TOSH theory and the ground level rotational constant coming from the harmonic approximation. The band origins are calculated to two decimal places. The rotational constants have been calculated to four decimal places. Errors have been included where available (some literature sources did not provide them) and other errors have been omitted if they were of a lower order than the rounding criterion of the data.

Figures 2 and 3 show the resulting simulated spectra from Prometheus compared to ExoMol, data by Campargue et al. (2012), and the harmonic method.

Two major effects need to be considered when the rovibrational spectrum of H_2 is analyzed. First, this homonuclear molecule exhibits a center of symmetry. As a consequence of this, it is Raman active and IR inactive with O, Q and S branches present in its spectra (Campargue et al. 2012). Prometheus will model a molecule in the correct region by considering symmetry of the diatomic using the initial mass inputs.

As also mentioned by Campargue et al. (2012), a molecule composed of nuclei with nonzero nuclear spin (such as H_2) will exhibit spectral lines that show an alternation in intensity. Unfortunately Prometheus does not currently model this. Instead the intensities have the typical Boltzmann distribution regardless of the nuclear spin of the molecule's nuclei components.

A key aspect of Figure 2, is how well all three band origins align. Using spectroscopic constants of Campargue et al. (2012), Prometheus can quite accurately reproduce the ExoMol positions for the transitions. Prometheus TOSH approach however appears to slightly overestimate the band origin by about 20 cm^{-1}. This is likely due to the strong quantum nature of the hydrogen nuclei, and thus this molecule displays strong anharmonic effects that are less well modeled by the TOSH approach. Regardless, Figure 2 shows all three approaches are in relatively good agreement with one another.

Regarding the Q branch, Prometheus displays a larger spread of transitions than ExoMol and Campargue et al. (2012). The Prometheus central transitions potentially would roughly line up provided the band origin was corrected. Once again Campargue et al. (2012) adequately models the ExoMol data. It does falter slightly with the distribution of the branch, as the Campargue et al. (2012) Q branch appears to model a greater spread in the transitions than ExoMol.
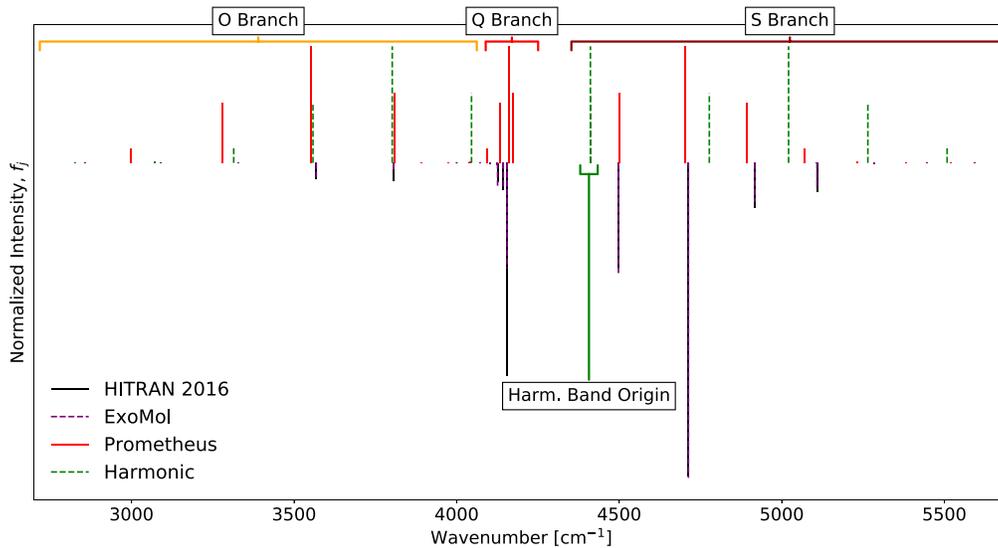
Figure 3. A comparison of the H_2 spectra produced by Prometheus, ExoMol (Roueff et al. 2019) and the Harmonic method. The intensities for all sets of data have been normalized and therefore are relative values. Note that HITRAN and ExoMol transitions overlap with each other.

The O branch, indicated by the labels on Figures 2 and 3, shows an interesting result. Campargue et al. (2012) appears to decrease slightly in accuracy the higher the transition, whereas Prometheus does the same but the decrease in accuracy is larger. This deterioration is to such a point it becomes difficult to match the ExoMol's transitions to Prometheus.

For the S branch (also indicated by the label on Figures 2 and 3), Prometheus seems to initially fare quite well. The Prometheus transition positions appear to slightly increase in difference to the ExoMol data with each higher transition. Possibly if the origins had aligned better Prometheus would be able to continue to quite effectively model this branch at higher rotational transitions. As before, Campargue et al. (2012) gives a satisfactory level of comparison to ExoMol, until the higher transitions are reached at which point it begins to falter. Some difficulty occurs with the comparisons at higher transitions as the relative intensities for the ExoMol data are very low. The ExoMol S branch seems to have a far greater weighting than the O branch, something which is not modeled in any other spectra.

The difficulties in describing the higher rotational transitions can be partly explained by the fact that hydrogen is a light molecule and thus distortion constants are required to accurately model anything beyond $J = 0$, 1. For example, it was found by Jennings & Brault (1983) that to accurately fit up to $J = 5$, four rotational constants (B, D, H, and L) were required. This presumably explains Prometheus' inaccuracies for the hydrogen spectra, which are not as pronounced in the other heavier molecules considered within these simulations.

Finally, we compare the Prometheus spectrum with the harmonic model in Figure 3. As discussed earlier, it can be seen that Prometheus slightly overestimates the origin, whereas the harmonic method substantially overestimates it. Due to this, it is difficult to comment on much of the harmonic spectrum, as it barely matches ExoMol.

We can see that despite Prometheus also having some difficulties with modeling H_2, particularly at higher transitions and in the O branch, it still performs better than the harmonic approach. Certainly, this shows anharmonic corrections are required to produce a qualitatively correct result.

3.2. Molecular Oxygen

The second biosignature is molecular oxygen. Oxygen is the third most abundant element in the universe, yet molecular oxygen is one of the most elusive molecules (Wang et al. 2020). This is often cited to be due to O_2's high chemical reactivity and lack of electric dipole moment (Luspay-Kuti et al. 2018). Even now, a comprehensive picture of oxygen chemistry in interstellar environments is still missing (Wang et al. 2020; Wakelam et al. 2010; Agundez & Cernicharo 2006).

Oxygen is crucial to our understanding of most life on Earth, with oxygenic photosynthesis being the dominant producing metabolism on our planet (Domagal-Goldman et al. 2014). Concerning exoplanet observations, oxygen is a potential biosignature since it can be produced through photosynthesis processes. However, it is also a possible false positive since it can be formed in larger quantities through abiotic processes such as runaway greenhouse effects (Meadows 2017).

Oxygen's potential to act as a false-positive biosignature is rooted in the diversity of exoplanets. For example, on Earth there are no abiotic processes that would produce it in large abundance (Meadows et al. 2018). It has been shown for a very different star and planetary system that O_2 could be generated in a planetary environment without life (Wordsworth & Pierrehumbert 2014).

The literature spectroscopic constants for O_2 were obtained from Yu et al. (2014), and our values are compared to data from the HITRAN 2016 release (Gamache et al. 1998; Yu et al. 2014; Gordon et al. 2017) only, as no ExoMol data were available. It is worth noting that the HITRAN data contain not Raman but magnetic dipole and electric quadrupole IR transitions, which may lead to some discrepancies between the data sets.

Like H_2, O_2 has a center of symmetry, which causes it to be Raman active but IR inactive. This once again means that the O, Q, and S branches are present in the spectra. In addition, the symmetry also means the effects of nuclear spin will be observed, but these effects differ to that of the other homonuclear diatomic molecules.
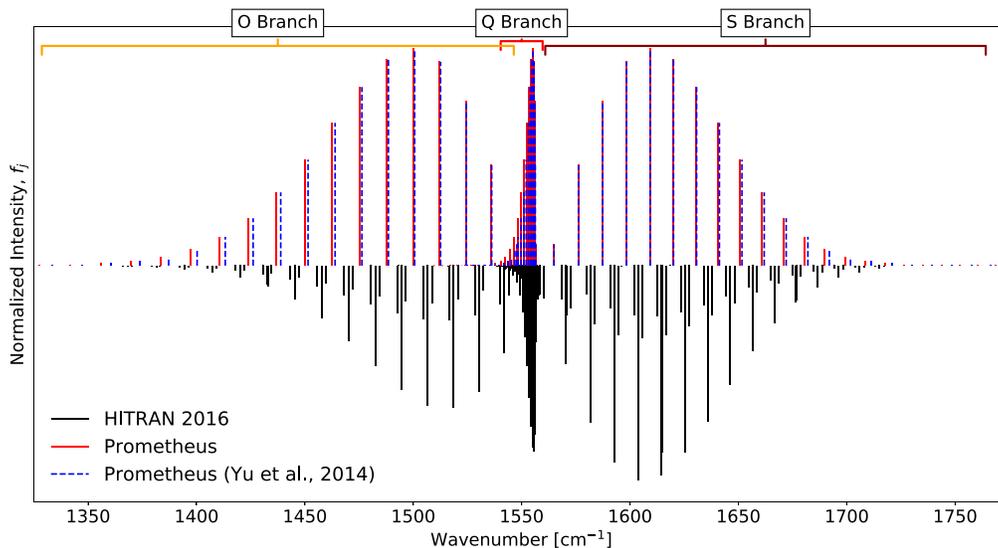
Figure 4. A comparison of the O_2 spectra produced by Prometheus, Prometheus using the spectroscopic constants from Yu et al. (2014) and HITRAN 2016 data (Gordon et al. 2017; Gamache et al. 1998; Yu et al. 2014). The intensities for all sets of data have been normalized and therefore are relative values.
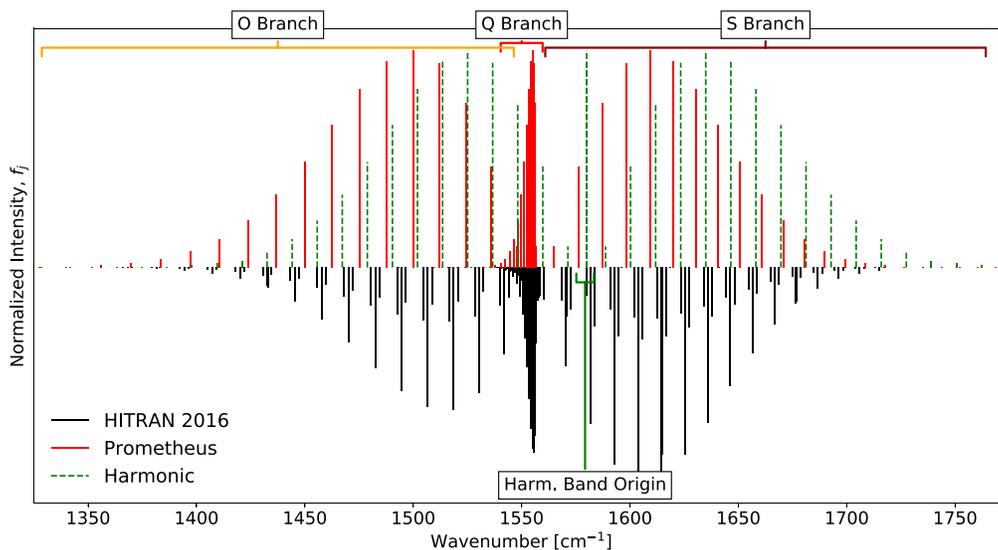

Figure 5. A comparison of the O_2 spectra produced by Prometheus, the harmonic method, and HITRAN 2016 data (Gamache et al. 1998; Yu et al. 2014; Gordon et al. 2017). The intensities for all sets of data have been normalized and therefore are relative values.

Unlike the other diatomic molecules in this study, O_2 has a zero nuclear spin, which has a different effect on the spectra than the alternating intensities effect. Instead, every rotational level with an even value is absent from the spectra. This presents itself as a spectra with seemingly large gaps between the rotational transitions (Yu et al. 2014; see Figure 4).

O_2 has a triplet sigma state and exhibits a triplet structure for each of the remaining transition lines, due to the splitting of the rotational levels. This is another effect that Prometheus does not currently model and needs to be taken into consideration when evaluating spectra. However, the weighting of the relative intensities due to triplet splitting have been included into Prometheus, to allow for ease of comparison.

When considering the band origins (see Table 2), TOSH produces a better approximation to the literature than the harmonic method, despite the points mentioned previously. Yu et al. (2014) reports the band origin to be at 1556.39 cm^{-1}, and

this is extremely well estimated by TOSH, with a result of 1556.43 cm^{-1}. The harmonic approximation, on the other hand, displays a significant displacement, roughly 24 cm^{-1}. We also note again that the harmonic approximation provides a better approximation of B_0, with TOSH doing the same for B_1.

In Figure 5, we can quite clearly see the shift in band origin has a significant effect on the harmonic approximation's ability to mimic the literature.

Prometheus, by using our hybrid methodology, models the Q branch well (in good agreement with Yu et al. 2014). This is something the purely harmonic method (or a purely TOSH approach) repeatedly fails at due to the fixed rotational constants between levels.

3.3. Molecular Nitrogen

The third diatomic biosignature is molecular nitrogen. This is a key species in cosmology and has been observed in various
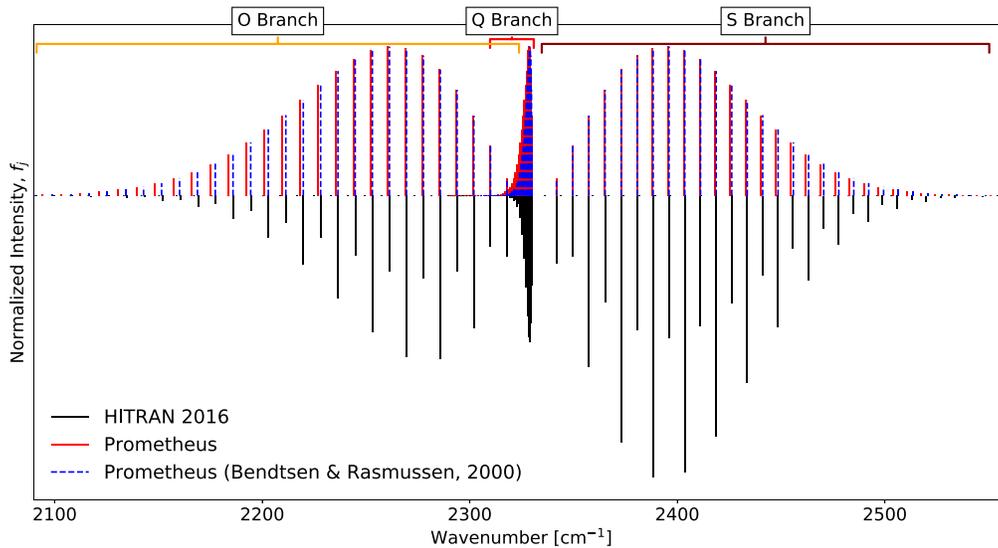
Figure 6. A comparison of the N_2 spectra produced by Prometheus, Prometheus using the spectroscopic constants from Bendtsen & Rasmussen (2000), and HITRAN 2016 (Gordon et al. 2017; Li & Roy 2007; Roy et al. 2006) data. The intensities for all sets of data have been normalized and therefore are relative values.

galactic environments (Vangioni et al. 2018), although direct observation of N_2 is often difficult as it lacks strong pure rotational or vibrational lines (Li et al. 2013).

Within our own solar system we find N_2 present in various planetary and satellite atmospheres; it accounts for 3.5% (Oyama et al. 1980) of the Venusian atmosphere, 78.1% (Cox 2002) of the terrestrial atmosphere, 2.8% (Franz et al. 2017) of the Martian atmosphere, and, perhaps most significantly, 98.4% (Strobel & Shemansky 1982) of Titan's atmosphere.

From a biogenic perspective, N_2 is an essential ingredient for the building blocks for life as we know it (Sproß et al. 2018) since it is required, along with carbon and phosphorus, for the formation of nucleic acids and proteins (Lammer et al. 2019).

The spectroscopic constants were obtained from Bendtsen & Rasmussen (2000). Our N_2 results are compared to the harmonic approximation, literature optimal results and the data from the HITRAN 2016 release (Roy et al. 2006; Li & Roy 2007; Gordon et al. 2017). The available ExoMol data was not suitable for our study as it provides the Western et al. (2018) data, which focus on a different region of the spectrum ($4500–11,000\,cm^{-1}$), rather than the ground state.

Finally, N_2, like O_2 and H_2, has a center of symmetry and thus the spectrum is Raman active but IR inactive. As N_2 is composed of nuclei of nonzero nuclear spin, the HITRAN spectrum shows the characteristic alternation in intensities (again, not modeled here by our Prometheus approach).

By looking at the results of Table 2, we see that N_2 rotational constants follow the typical pattern of the previous molecules within this study. Bendtsen & Rasmussen (2000) determined the first excited level rotational constant to be $1.9722\,cm^{-1}$ and the ground rotational constants to be $1.9896\,cm^{-1}$. The TOSH approach, with a fixed value of $1.9753\,cm^{-1}$, overestimates the ground state but provides a good approximation for the first excited level. Whereas the harmonic approximation, with a constant value of $1.9982\,cm^{-1}$, achieves the opposite.

We can see in Table 2 that for N_2, TOSH ($2329.85\,cm^{-1}$) once again provides a superior approximation of the band origin than the harmonic approximation ($2359.56\,cm^{-1}$), when comparing to the literature value ($2329.91\,cm^{-1}$). In Figure 7,

we can draw comparisons between HITRAN, Prometheus, and the Bendtsen & Rasmussen (2000) constants used in a Prometheus spectrum. As expected from using the TOSH band origin, Prometheus provides an excellent approximation of both Bendtsen & Rasmussen (2000) and HITRAN.

For the lower rotational transitions of both the O and S branches, Prometheus does a satisfactory job of modeling the positions. In most cases it lines up incredibly well with not only Bendtsen & Rasmussen (2000) but also HITRAN data. As was the case for the other molecules, this accuracy is lower for higher J transitions.

Prometheus does an excellent job at approximating the Q branch for N_2 and the relative spread of the transitions of this branch is comparable to the HITRAN data.

For each increasing J transition the position is incrementally overestimated, but not to the point where the Prometheus spectrum is incorrectly matching lower number transitions to higher HITRAN ones. With this is mind, Prometheus is a very good approximation of both Bendtsen & Rasmussen (2000) and HITRAN.

A comparison with the harmonic spectrum is shown in Figure 7. The harmonic band origin is displaced from the literature by roughly 30 wavenumbers (see Table 2). This effectively renders any comparisons of the transition positions to the other spectra pointless.

As was the case for H_2 the relative intensities of N_2 alternate in weighting, an effect that Prometheus does not currently model, once again causing the difficulties in drawing comparisons. The general intensities of the harmonic and the Bendtsen & Rasmussen (2000) spectra have the same problem, which is expected, as it is produced via our Prometheus methodology.

Comparing the relative intensities of Prometheus and harmonic to HITRAN does raise interesting points. The "peak" for the O branch for HITRAN appears to occur 10 to 20 wavenumbers away from Prometheus. Potentially the alternating intensities are artificially causing the "peak" to occur at higher wavenumbers. The S branch does not appear to be as exaggerated. As discussed earlier, this discrepancy is not a
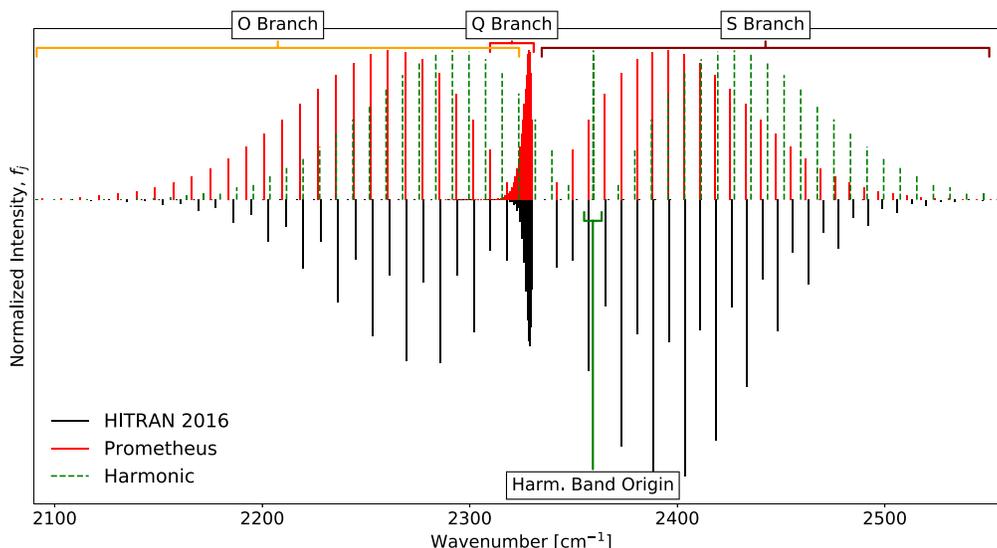
Figure 7. A comparison of the N_2 spectra produced by Prometheus, the harmonic method, and HITRAN 2016 (Roy et al. 2006; Li & Roy 2007; Gordon et al. 2017) data. The intensities for all sets of data have been normalized and therefore are relative values.

major concern as intensities calculations are the not the focus of our present study.

We may conclude that, for N_2, Prometheus is producing a better approximate spectrum than the harmonic method. Like the other molecules seen in previous sections, Prometheus' predictive power is reduced at higher rotational transitions. Presumably the uncorrected harmonic value for ground rotational constant and the TOSH approximate values for the rotational constant of the first excited level is having an effect on the later transitions' positions.

3.4. Carbon Monoxide

The last molecule considered in this work is CO. Unlike the previous molecules, CO has a permanent dipole moment and therefore readily absorbs, even at the low temperatures found in molecular clouds (Whitworth & Jaffa 2018). More relevantly, CO has already been detected in an exoplanet's atmosphere (HD 189733; de Kok et al. 2013).

Within a biogenic context, CO is often classified as an antibiosignature. For example, for an inhabited planet, it could be difficult for CO to accumulate in the atmosphere as it acts as an energy source for some microbes on Earth (Wang et al. 2016). Anthropologically, it can be formed instead when any organic substance is combusted incompletely and therefore can be an indication of not just life but intelligent life (Horner 2000).

The literature spectroscopic constants were obtained from Coxon & Hajigeorgiou (2004). Our results for CO are compared to the harmonic approximation, optimal results, and data from both the HITRAN 2016 release (Coxon & Hajigeorgiou 2004; Li et al. 2015; Gordon et al. 2017; Devi et al. 2018) and ExoMol's Li2015 line list (Li et al. 2015).

CO does not have a center symmetry and therefore it is IR active and Raman inactive. It is easily modeled by Prometheus to an accuracy akin to the HITRAN data at the lower transitions. Unlike the other biogenic diatomic molecules in this analysis, CO solely demonstrates P and R branches, with no Q-branch transitions.

Coxon & Hajigeorgiou (2004), as shown in Table 2, provided $1.9225\,cm^{-1}$ for B_0 and $1.9050\,cm^{-1}$ for B_1. As

before, the harmonic approximation for B_0 is better ($1.9316\,cm^{-1}$) than TOSH's $1.9080\,cm^{-1}$. For the upper rotational constant, B_1, TOSH provides a rotational constant closer to literature.

The band origin of the Prometheus spectrum (using TOSH), is $2143.15\,cm^{-1}$, and easily lies within a single wavenumber of Coxon & Hajigeorgiou (2004) (at $2143.27\,cm^{-1}$). Figure 8 shows that, even when compared to the HITRAN/ExoMol spectra, Prometheus' band origin still lies within a few wavenumbers.

The spectrum produced via the harmonic method (Figure 9) gives a band origin of $2173.07\,cm^{-1}$, which has a difference of $29.80\,cm^{-1}$ to Coxon & Hajigeorgiou (2004). It is also displaced from ExoMol/HITRAN by approximately 30 to 40 wavenumbers. Such a displacement naturally renders the harmonic CO spectrum considerably inaccurate when considering the positions of individual J transitions.

In general the distributions of the intensities between HITRAN and Prometheus are similar. On closer inspection, only the fundamental transitions at lower rotational energy levels, J, are modeled with high accuracy.

Indeed, most of the lower-target transitions lie within a couple of wavenumbers of Prometheus transitions, but this starts to falter for the higher values of J. The error for the higher transitions is then exaggerated due to the slight position differences of the band origin.

The relative intensities appear to be in a good agreement, although it is worth mentioning that ExoMol is exhibiting some increased intensities for certain higher-level transitions. This is not shown by either HITRAN2016 or Prometheus data.

3.5. Comparisons of Constants

As shown in Table 2, the harmonic approximation consistently describes the ground level rotational constant (B_0) well but poorly estimates the upper rotational constant (B_1). This makes realistically modeling any Q branch impossible since it relies on a difference between the ground and excited state rotational constant (see Equation (28) and Section 2.7).
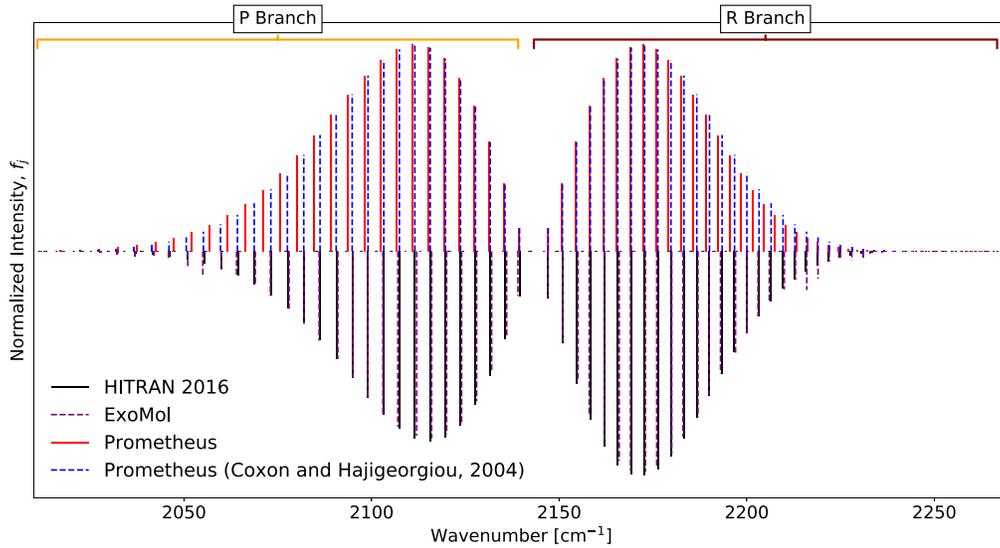
Figure 8. A comparison of the CO spectra produced by Prometheus, Prometheus using the spectroscopic constants from Coxon & Hajigeorgiou (2004), ExoMol (Li et al. 2015), and HITRAN 2016 data (Coxon & Hajigeorgiou 2004; Li et al. 2015; Gordon et al. 2017; Devi et al. 2018). The intensities for all sets of data have been normalized and therefore are relative values.
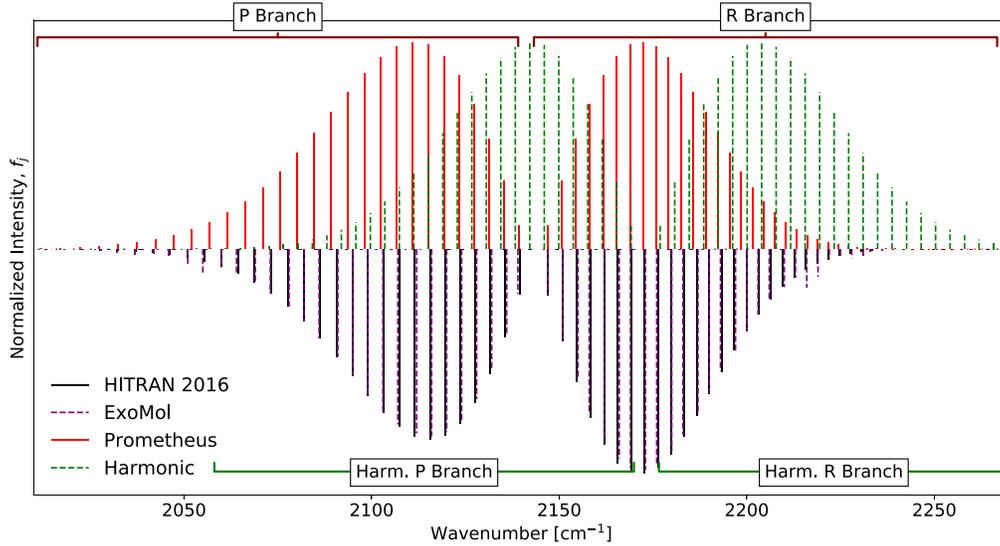

Figure 9. A comparison of the CO spectra produced by Prometheus, the harmonic method, ExoMol (Li et al. 2015), and HITRAN 2016 data (Coxon & Hajigeorgiou 2004; Li et al. 2015; Gordon et al. 2017; Devi et al. 2018). The intensities for all sets of data have been normalized and therefore are relative values.

The TOSH-corrected rotational constants, rather than offering the predicted "vibrationally averaged" value between the first and ground states, actually model the first excited state well (see Table 2). Unfortunately, the TOSH rotational constant overestimates the ground-state value, with the harmonic approach typically offering a better approximation. Other studies have also shown (e.g., Table 1 in Endres 1967) that the uncorrected harmonic approximation is a good approximation for the ground state, whereas for the higher levels this is not the case.

Both the ground and first-excited-level rotational results have been included in Table 2 and discussed within the previous results sections. This has been done to highlight that using the current modified theory, where we amalgamate the TOSH and harmonic results, is the best option to match reference spectra. The results used for Prometheus are highlighted in the table for each case.

Typically Prometheus is able to replicate the band origin within a single wavenumber using TOSH, with the exception of H_2. Prometheus has also shown it is capable of estimating the band origins to a greater accuracy than that of the harmonic method, often with order of magnitude improvements.

While TOSH can approximate the upper rotational constant better than the harmonic model approximates the lower one, both still only approximate the literature values.

This has the further effect of causing incorrect spacing of transitions in the branches, which has been shown for the high-J rotational transitions in all spectra produced. We can verify this conjecture by reviewing Equations (26)–(30). Each of these position equations have a quadratic dependence on the quantum number, J, therefore any variation between the rotational constants (such as the differences between the literature and Prometheus) will become more apparent at the higher transitions.

It is also worth noting that we have not used any distortion constants or additional rotation constants in our calculations. We did this to keep the method as simple as possible. The effects of this exclusion are most notable for H_2.

4. Conclusions

As shown and discussed in Section 3, the TOSH method produces full fundamental rovibrational spectra at higher accuracy than the harmonic approximation for a slight increase in computational complexity. The novelty here is the suitable modeling of the rovibrational transition, rather than a simple vibrational modeling of the fundamental. This has never been attempted within the TOSH framework before.

Provided the anharmonicity constants are of high quality (essentially meaning the PEC they are obtained from is of spectroscopic quality), Prometheus can seemingly replicate a spectrum that is comparable to HITRAN at the lower transitions. Despite the breakdown in accuracy at higher transitions, Prometheus still produces a satisfactory approximation at extremely low computational cost (essentially only the cost of determining the necessary quartic constants: η_{ii}, η_{iiii}, and η_{iiii}).

One major caveat is that the current theory for the σ shift does not accommodate for two differing vibrational constants for the upper and lower states. Instead the lower value must be calculated via a harmonic methodology, which in turn leads to similar inaccuracies in the transition positions.

A second issue is the Prometheus spectra do not currently model specific phenomena such as alteration in line intensity due to nuclear spin and the splitting of rotational levels. This arguably is not a crucial element, as the code currently focuses on line position rather than intensities. In the case of effects due to nuclear spin, the intensities variation is more of a cosmetic issue. The absent lines due to zero spin is arguably a more important effect, and as such has been modeled by Prometheus.

A recurring note for improvement throughout this paper has pertained to the shift variable, σ. It is a useful tool to use to account for anharmonic corrections of the rotational constant, however in its current form it is only suitable for the fundamental transition. To improve the spectra, we need to modify the theory to obtain a shift that varies with vibrational level instead of being optimized for the fundamental, $0 \rightarrow 1$, vibrational transition.

Another aspect of further work would revolve around creating a more robust means of evaluating the additional effects arising from the symmetry of a molecule. This would include considering the phenomena arising due to properties such as spin and triplet structure.

Finally, the next stage is to adapt the theory to triatomic molecules. Once triatomics have been successfully modeled we can then progress to modeling larger polyatomic molecules.

Appendix
Further Derivations

A.1. Deriving the Energy

We will go through how to derive the corresponding energy for each vibrational level as referenced by Equations (3) and (5). We will use the following notation from Section 2.6 within the following Appendix sections. As a reminder, the relation between the anharmonic position, x, and the harmonic position, u, can be described as

$$x - \sigma = u. \tag{A1}$$

We will now demonstrate deriving the ground-state energy. To begin we must use perturbation theory. As the anharmonic Hamiltonian (Equation (1)) and wave functions are known (Equation (2)) we can express it as

$$E_0 = \langle \psi_0^* | \hat{H} | \psi_0 \rangle, \tag{A2}$$

$$= \int_{-\infty}^{\infty} \psi_0^* \left(-\frac{1}{2} \frac{\partial^2}{\partial x^2} + \frac{1}{2!} \eta_{ii} x^2 + \frac{1}{3!} \eta_{iii} x^3 + \frac{1}{4!} \eta_{iiii} x^4 \right) \psi_0 dx. \tag{A3}$$

For ease of comprehension, we will break the following derivations down into four components and solve individually, then sum back together. We will now describe the separated parts and label for ease:

$$A = \int_{-\infty}^{\infty} \psi_0^* \left(-\frac{1}{2} \frac{\delta^2}{\delta x^2} \right) \psi_0 \, dx, \tag{A4}$$

$$B = \int_{-\infty}^{\infty} \psi_0 \left(\frac{1}{2!} \eta_{ii} x^2 \right) \psi_0 \, dx, \tag{A5}$$

$$C = \int_{-\infty}^{\infty} \psi_0 \left(\frac{1}{3!} \eta_{iii} x^3 \right) \psi_0 \, dx, \tag{A6}$$

$$D = \int_{-\infty}^{\infty} \psi_0 \left(\frac{1}{4!} \eta_{iiii} x^4 \right) \psi_0 \, dx. \tag{A7}$$

Let us begin with A first. We need to remember that when an operator involves a differentiation, it does not commute. This means we now need to first differentiate ψ_0 twice, but leave ψ_0 unaffected:

$$A = \int_{-\infty}^{\infty} \psi_0^* \left(-\frac{1}{2} \frac{\partial^2 \psi_0}{\partial Q^2} \right) dx. \tag{A8}$$

Now to insert the equations for the wave functions:

$$= -\frac{1}{2} \left(\frac{\omega}{\pi} \right)^{1/2} \int_{-\infty}^{\infty} e^{-\frac{\omega}{2}(x-\sigma)^2} \frac{\partial^2}{\partial x^2} e^{-\frac{\omega}{2}(x-\sigma)^2} \, dx. \tag{A9}$$

Recalling $u = x - \sigma$, therefore, $dx = du$. Substitute this back in:

$$= -\frac{1}{2} \left(\frac{\omega}{\pi} \right)^{1/2} \int_{-\infty}^{\infty} e^{-\frac{\omega}{2}(u)^2} \frac{\partial^2}{\partial x^2} e^{-\frac{\omega}{2}(u)^2} \, du. \tag{A10}$$

Doubly differentiating within the equation we get

$$\frac{\partial}{\partial x}e^{-\frac{\omega}{2}(u)^2} = -\omega u e^{-\frac{\omega}{2}(u)^2}, \tag{A11}$$

$$\frac{\partial}{\partial x} - \omega u e^{-\frac{\omega}{2}(u)^2} = u^2\omega^2 e^{-\frac{\omega}{2}(u)^2} - \omega e^{-\frac{\omega}{2}(u)^2}, \tag{A12}$$

$$\therefore = -\frac{1}{2}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty} e^{-\frac{\omega}{2}(u)^2}(u^2\omega^2 e^{-\frac{\omega}{2}(u)^2} - \omega e^{-\frac{\omega}{2}(u)^2})du, \tag{A13}$$

$$= -\frac{1}{2}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}(u^2\omega^2 e^{\omega(u)^2} - \omega e^{-\omega(u)^2})du, \tag{A14}$$

$$= -\frac{1}{2}\left(\frac{\omega}{\pi}\right)^{1/2}\left[\omega^2\frac{1}{2\omega}\left(\frac{\pi}{\omega}\right)^{1/2} - 2\omega\left(\frac{\pi}{4\omega}\right)^{1/2}\right], \tag{A15}$$

$$= \frac{1}{4}\omega. \tag{A16}$$

Now onto part B, same as before we will introduce the wave functions then use substitution to solve:

$$B = \int_{-\infty}^{\infty}\psi_0^*\left(\frac{1}{2!}\eta_{ii}x^2\right)\psi_0\,dx, \tag{A17}$$

$$= \frac{1}{2!}\eta_{ii}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}x^2 e^{-\omega(x-\sigma)^2}\,dx, \tag{A18}$$

$$= \frac{1}{2!}\eta_{ii}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}(u+\sigma)^2 e^{-\omega u^2}\,du, \tag{A19}$$

$$= \frac{1}{2!}\eta_{ii}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}(u^2+2u\sigma+\sigma^2)e^{-\omega u^2}\,du. \tag{A20}$$

Once again we can break this down into pieces:

$$\int_{-\infty}^{\infty}(u^2)e^{-\omega u^2}\,du = \frac{1}{2}\left(\frac{\pi}{\omega^3}\right), \tag{A21}$$

$$2\sigma\int_{-\infty}^{\infty}(u)e^{-\omega u^2}\,du = 0, \tag{A22}$$

$$\sigma^2\int_{-\infty}^{\infty}e^{-\omega u^2}\,du = \sigma^2\left(\frac{\pi}{\omega}\right). \tag{A23}$$

Putting these components back together:

$$=\frac{1}{2!}\eta_{ii}\left(\frac{\omega}{\pi}\right)^{1/2}\left(\frac{1}{2\omega}\left(\frac{\pi}{\omega}\right)^{1/2} + \sigma^2\left(\frac{\pi}{\omega}\right)^{1/2}\right), \tag{A24}$$

$$=\frac{1}{2!}\eta_{ii}\left[\frac{1}{2\omega} + \sigma^2\right]. \tag{A25}$$

Part C, using the same method as before:

$$C = \int_{-\infty}^{\infty}\psi_0^*\left(\frac{1}{3!}\eta_{iii}x^3\right)\psi_0\,dx, \tag{A26}$$

$$= \frac{1}{3!}\eta_{iii}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}x^3 e^{-\omega(Q-\sigma)^2}\,dx, \tag{A27}$$

$$= \frac{1}{3!}\eta_{iii}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}(u+\sigma)^3 e^{-\omega(u)^2}\,du, \tag{A28}$$

$$= \frac{1}{3!}\eta_{iii}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}(u^3+3\sigma u^2+\sigma^2 u+\sigma^3)^3 e^{-\omega(u)^2}\,du. \tag{A29}$$

Breaking it down into parts again for ease of solving:

$$\int_{-\infty}^{\infty}u^3 e^{-\omega(u)^2}\,du = 0, \tag{A30}$$

$$3\sigma\int_{-\infty}^{\infty}u^2 e^{-\omega(u)^2}\,du = 3\sigma\left(\frac{\pi}{4\omega^3}\right)^{1/2}, \tag{A31}$$

$$\sigma^2\int_{-\infty}^{\infty}u e^{-\omega(u)^2}\,du = 0, \tag{A32}$$

$$\sigma^3\int_{-\infty}^{\infty}e^{-\omega(u)^2}\,du = \sigma^3\left(\frac{\pi}{\omega}\right)^{1/2}. \tag{A33}$$

Substituting these components back into the main equation:

$$= \frac{1}{3!}\eta_{iii}\left(\frac{\omega}{\pi}\right)^{1/2}\left[\frac{3\sigma}{2\omega}\left(\frac{\pi}{\omega}\right)^{1/2} + \sigma^3\left(\frac{\pi}{\omega}\right)^{1/2}\right], \tag{A34}$$

$$= \frac{1}{3!}\eta_{iii}\left[\frac{3\sigma}{2\omega} + \sigma^3\right]. \tag{A35}$$

Finally, onto section D of the original equation:

$$D = \int_{-\infty}^{\infty}\psi_0^*\left(\frac{1}{4!}\eta_{iiii}x^4\right)\psi_0\,dx, \tag{A36}$$

$$= \frac{1}{4!}\eta_{iiii}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}x^4 e^{-\omega(Q-\sigma)^2}\,dx, \tag{A37}$$

$$= \frac{1}{4!}\eta_{iiii}\left(\frac{\omega}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}(u+\sigma)^4 e^{-\omega(u)^2}\,du, \tag{A38}$$

$$= \frac{1}{4!}\eta_{iiii}\left(\frac{\omega}{\pi}\right)^{1/2}$$
$$\times \int_{-\infty}^{\infty}(u^4+4\sigma u^3+6\sigma^2 u^2+4\sigma^3 u+u^4)e^{-\omega u^2}\,du. \tag{A39}$$

Break down into parts:

$$\int_{-\infty}^{\infty}u^4 e^{-\omega u^2}du = \frac{6}{8\omega^2}\left(\frac{\pi}{\omega}\right)^{1/2}, \tag{A40}$$

$$4\sigma\int_{-\infty}^{\infty}u^3 e^{-\omega u^2}du = 0, \tag{A41}$$

$$6\sigma^2\int_{-\infty}^{\infty}u^2 e^{-\omega u^2}du = \frac{6\sigma^2}{2\omega}\left(\frac{\pi}{\omega}\right)^{1/2}, \tag{A42}$$

$$4\sigma^3\int_{-\infty}^{\infty}u\,du = 0, \tag{A43}$$

$$\sigma^4\int_{-\infty}^{\infty}e^{-\omega u^2}du = \sigma^4\left(\frac{\pi}{\omega}\right)^{1/2}. \tag{A44}$$

Summing the pieces back together:

$$= \frac{1}{4!}\eta_{iiii}\left(\frac{\omega}{\pi}\right)^{1/2}\left[\frac{6}{8\omega^2}\left(\frac{\pi}{\omega}\right)^{1/2} + \frac{6\sigma^2}{2\omega}\left(\frac{\pi}{\omega}\right)^{1/2} + \sigma^4\left(\frac{\pi}{\omega}\right)^{1/2}\right], \tag{A45}$$

$$= \frac{1}{4!}\eta_{iiii}\left[\frac{3}{4\omega^2} + \frac{3\sigma^2}{\omega} + \sigma^4\right]. \tag{A46}$$

If we now take all the parts, A, B, C, and D, we can now have the full equation for the energy of the ground state, which

matches the expression in Equation (6):

$$E_0 = \frac{1}{4}\omega + \frac{1}{2!}\eta_{ii}\left[\frac{1}{2\omega} + \sigma^2\right] + \frac{1}{3!}\eta_{iii}\left[\frac{3\sigma}{2\omega} + \sigma^3\right]$$
$$+ \frac{1}{4!}\eta_{iiii}\left[\frac{3}{4\omega^2} + \frac{3\sigma^2}{\omega} + \sigma^4\right]. \tag{A47}$$

The method described here can be applied to the first excited state to achieve the answer shown in Equation (6).

A.2. Position Expectation Values

Let us first start with the expectation value for the harmonic position, u:

$$\langle u \rangle = \int_{-\infty}^{\infty} u|\psi|^2 du. \tag{A48}$$

We can use a single simple identity to calculate the expectation value, as we know the wave function can be normalized:

$$\int_{-t}^{t} o(u)du = 0; \quad |\psi|^2 = 1, \tag{A49}$$

$$\therefore \langle u \rangle = 0. \tag{A50}$$

Now to do the anharmonic position expectation values. Although the anharmonic wave functions are also normalized, we cannot use only the identities described previously, as the equations are more complex. Therefore, in this case we need to define the wave functions. Using Equation (2), we see that the for the ground state is

$$\psi_0 = \left(\frac{\omega}{\pi}\right)^{1/4} e^{-\frac{\omega}{2}(x-\sigma)^2}. \tag{A51}$$

Additionally, for the first excited vibrational state:

$$\psi_1 = \left(\frac{4\omega^3}{\pi}\right)^{1/4}(x-\sigma)e^{-\frac{\omega}{2}(x-\sigma)^2}. \tag{A52}$$

Now to derive the expectation value for the shifted anharmonic position, $\langle x \rangle$, with respect to the anharmonic wave functions. We can now derive the expectation value of the position, x, for the anharmonic ground and first vibrational level:

$$\langle \psi_0|x|\psi_0 \rangle = \int_{-\infty}^{\infty} |\psi_0|^2 x \, dx, \tag{A53}$$

$$= \left(\frac{\omega}{\pi}\right)^{1/2} \int_{-\infty}^{\infty} xe^{-\omega(x-\sigma)^2} dx. \tag{A54}$$

Substituting $x = u + \sigma$:

$$= \left(\frac{\omega}{\pi}\right)^{1/2} \int_{-\infty}^{\infty} (u + \sigma)e^{-\omega(u)^2} du. \tag{A55}$$

Separating into two components (A and B) for easier integration:

$$A = \int_{-\infty}^{\infty} (u)e^{-\omega(u)^2} du. \tag{A56}$$

Using the identity for an odd function (where $o(u) = -o(-u)$ for all u), and knowing the wave function can be normalized:

$$\int_{-t}^{t} o(u)du = 0, \tag{A57}$$

$$\therefore A = \int_{-\infty}^{\infty} (u)e^{-\omega(u)^2} du = 0. \tag{A58}$$

Now for part B:

$$B = \sigma \int_{-\infty}^{\infty} e^{-\omega(u)^2} du. \tag{A59}$$

An even function $e(u)$ satisfies $e(u) = e(-u)$ for all u. There, for any t,

$$\int_{-t}^{t} e(u)du = 2\int_0^t e(u)du, \tag{A60}$$

$$\therefore B = \sigma \int_{-\infty}^{\infty} e^{-\omega(u)^2} du = 2\sigma \int_0^{\infty} e^{-\omega(u)^2} du. \tag{A61}$$

Then using a secondary integration identity:

$$\int_0^{\infty} e^{-\omega u^2} du = \left(\frac{\pi}{4\omega}\right)^{1/2}, \tag{A62}$$

$$\therefore B = 2\sigma\left(\frac{\pi}{4\omega}\right)^{1/2}. \tag{A63}$$

Put A and B back together into the original expression:

$$= \left(\frac{\omega}{\pi}\right)^{1/2}\left[0 + 2\sigma\left(\frac{\pi}{4\omega}\right)^{1/2}\right], \tag{A64}$$

$$= \sigma. \tag{A65}$$

Now onto the first vibrational level:

$$\langle \psi_1|x|\psi_1 \rangle = \int_{-\infty}^{\infty} |\psi_1|^2 x \, dx, \tag{A66}$$

$$= \left(\frac{4\omega^3}{\pi}\right)^{1/2} \int_{-\infty}^{\infty} (x-\sigma)^2 xe^{-\omega(x-\sigma)^2} dx. \tag{A67}$$

Substituting $x = u + \sigma$:

$$= \left(\frac{4\omega^3}{\pi}\right)^{1/2} \int_{-\infty}^{\infty} u^2(u+\sigma)e^{-\omega(u)^2} du, \tag{A68}$$

$$= \left(\frac{4\omega^3}{\pi}\right)^{1/2} \int_{-\infty}^{\infty} (u^3 + \sigma u^2)e^{-\omega(u)^2} du. \tag{A69}$$

Once again separating into two components (A and B) for easier integration:

$$A = \int_{-\infty}^{\infty} (u^3)e^{-\omega(u)^2} du. \tag{A70}$$

Using the identity for an odd function (where $o(u) = -o(-u)$ for all u):

$$\int_{-t}^{t} o(u)du = 0, \tag{A71}$$

$$\therefore A = \int_{-\infty}^{\infty} (u^3)e^{-\omega(u)^2} du = 0, \tag{A72}$$

$$B = \sigma \int_{-\infty}^{\infty} (u^2)e^{-\alpha(u)^2} du. \tag{A73}$$

Using the integration identity:

$$\int_{-\infty}^{\infty} u^2 e^{-\omega u^2} du = \frac{1}{2}\left(\frac{\pi}{\omega^3}\right)^{1/2}, \tag{A74}$$

$$\therefore B = \frac{\sigma}{2}\left(\frac{\pi}{\omega^3}\right)^{1/2}. \tag{A75}$$

Putting A and B back together into the original expression:

$$= \left(\frac{4\omega^3}{\pi}\right)^{1/2}\left[0 + \frac{\sigma}{2}\left(\frac{\pi}{\omega^3}\right)^{1/2}\right], \quad (A76)$$

$$= \sigma. \quad (A77)$$

Clearly the outcome is independent of the vibrational level.

A.3. Branches

This next section will detail the working out required to produce Equations (26), (27), (28), (29) and (30). The general analytical expression for a spectrum of a diatomic molecule is the following:

$$\varepsilon_{(J,v)} = \varepsilon_J + \varepsilon_v, \quad (A78)$$

$$= \bar{\omega}_v(v + 1/2) + B_v J(J + 1) - D_e J^2(J + 1)^2, \quad (A79)$$

where

$$\bar{\omega}_v = \bar{\omega}_e - \bar{\omega}_e x_e(v + 1/2), \quad (A80)$$

$$\therefore \varepsilon_{(J,v)} = \bar{\omega}_e(v + 1/2) - \bar{\omega}_e x_e(v + 1/2)^2 \\ + B_v J(J + 1) - D_e J^2(J + 1)^2 + ... \quad (A81)$$

The frequency $\nu_{0\to1}$ is usually called the band origin or band center, which is represented by the following equation:

$$\nu_{0\to1} = \bar{\omega}_e - 2\bar{\omega}_e x_e. \quad (A82)$$

D_e represents the centrifugal distortion constant. For current theory we can ignore the small centrifugal distortions from D_e, etc., and the equation can be rewritten simply as

$$\varepsilon_{(J,v)} = \bar{\omega}_e(v + 1/2) - \bar{\omega}_e x_e(v + 1/2)^2 + B_v J(J + 1). \quad (A83)$$

For the next part we will be restricting the discussion to solely fundamental vibrational transitions, $v = 0 \to v = 1$. We also take the respective B values as B_0 and B_1 with $B_0 > B_1$. This transition can generally be described as (in cm^{-1}):

$$\Delta\varepsilon = \varepsilon_{J',v=1} - \varepsilon_{J'',v=0}, \quad (A84)$$

$$= \nu_{0\to1} + B_1 J'(J' + 1) - B_0 J''(J'' + 1). \quad (A85)$$

The transitions for the P branch ($\Delta J = -1$, $J'' = J' + 1$) is given by

$$\Delta\varepsilon = \bar{\nu}_P, \quad (A86)$$

$$= \nu_{0\to1} - (B_1 + B_0)(J' + 1) + (B_1 - B_0)(J' + 1)^2, \quad (A87)$$

where $J' = 0, 1, 2, 3,...$. The transitions for the Q branch ($\Delta J = 0$, $J'' = J'$) is given by

$$\Delta\varepsilon = \bar{\nu}_Q, \quad (A88)$$

$$= \nu_{0\to1} - J'(J' + 1)(B_1 - B_0), \quad (A89)$$

where $J' = 0, 1, 2, 3,...$.

Therefore for the R branch ($\Delta J = +1$, $J' = J'' + 1$) the transition is given as

$$\Delta\varepsilon = \bar{\nu}_R, \quad (A90)$$

$$= \nu_{0\to1} + (B_1 + B_0)(J'' + 1) + (B_1 - B_0)(J'' + 1)^2, \quad (A91)$$

where $J'' = 0, 1, 2, 3, ...$

Taking the general rovibrational equation:

$$\varepsilon(J, v) = \bar{\omega}_e(v + 1/2) - \bar{\omega} x_e(v + 1/2)^2 + B_v J(J + 1). \quad (A92)$$

Let us now apply this for the ground state with quantum number J'' and the upper vibrational state with quantum number J' as per standard notation:

$$\varepsilon_{(J'',v=0)} = \frac{1}{2}\bar{\omega}_e - \frac{1}{4}\bar{\omega} x_e + B_0 J''(J'' + 1), \quad (A93)$$

$$\varepsilon_{(J',v=1)} = \frac{3}{2}\bar{\omega}_e - \frac{9}{4}\bar{\omega} x_e + B_1 J'(J' + 1), \quad (A94)$$

$$\therefore \Delta\varepsilon = \nu_{0\to1} + B_1 J'(J' + 1) - B_0 J''(J'' + 1). \quad (A95)$$

Now consider the S branch where $\Delta J = +2$, $J' = J'' + 2$. By plugging in these formulas:

$$\Delta\varepsilon = \nu_S, \quad (A96)$$

$$= \nu_{0\to1} + B_1(J'' + 2)(J'' + 3) - B_0 J''(J'' + 2), \quad (A97)$$

$$= \nu_{0\to1} + B_1((J'')^2 + 5J'' + 6) - B_0((J'')^2 + J''). \quad (A98)$$

In the cases where $B_0 = B_1 = B$, the expression simplifies to

$$\nu_S = \nu_{0\to1} + B(4J'' + 6). \quad (A99)$$

Finally, onto the O branch where $\Delta J = -2$, $J'' = J' + 2$. By plugging in these formulas:

$$\Delta\varepsilon = \nu_O, \quad (A100)$$

$$= \nu_{0\to1} + B_1 J'(J' + 1) - B_0(J' + 2)(J' + 3), \quad (A101)$$

$$= \nu_{0\to1} + B_1((J')^2 + J') - B_0((J')^2 + 5J' + 6). \quad (A102)$$

In the cases where $B_0 = B_1 = B$, the expression simplifies to

$$\nu_O = \nu_{0\to1} - B(4J'' + 6). \quad (A103)$$

ORCID iDs

Thomas M. Cross ⓘ https://orcid.org/0000-0001-6621-8182
David M. Benoit ⓘ https://orcid.org/0000-0002-7773-6863
Marco Pignatari ⓘ https://orcid.org/0000-0002-9048-6010
Brad K. Gibson ⓘ https://orcid.org/0000-0003-4446-3130

References

Agundez, M., & Cernicharo, J. 2006, ApJ, 650, 374
Aladro, R., Martín, S., Riquelme, D., et al. 2015, A&A, 579, A101
Bacmann, A., Taquet, V., Faure, A., Kahane, C., & Ceccarelli, C. 2012, A&A, 541, L12
Bains, W., Petkowski, J. J., Seager, S., et al. 2021, AsBio, 21, 1277
Barstow, J. K., Aigrain, S., Irwin, P. G. J., Kendrew, S., & Fletcher, L. N. 2015, MNRAS, 448, 2546
Beć, K., Futami, Y., Wójcik, M., & Ozaki, Y. 2016, PCCP, 18, 13666
Bendtsen, J., & Rasmussen, F. 2000, JRSp, 31, 433
Brown, L., Sung, K., Benner, D., et al. 2013, JQSRT, 130, 201
Campargue, A., Kassi, S., Pachucki, K., & Komasa, J. 2012, PCCP, 14, 802
Chubb, K. L., Rocchetto, M., Yurchenko, S. N., et al. 2021, A&A, 646, A21
Clark, V. H., & Benoit, D. M. 2020, in IAU Symp. 350, Laboratory Astrophysics: From Observations to Interpretation, ed. F. Salama & H. Linnartz (Cambridge: Cambridge Univ. Press), 468
Coles, P. A., Yurchenko, S. N., & Tennyson, J. 2019, MNRAS, 490, 4638
Colón, K. D., Kreidberg, L., Welbanks, L., et al. 2020, AJ, 160, 280
Costagliola, F., Aalto, S., Rodriguez, M. I., et al. 2011, A&A, 528, A30
Cox, A. N. (ed.) 2002, Allen's Astrophysical Quantities (New York: Springer)
Coxon, J. A., & Hajigeorgiou, P. G. 2004, JChPh, 121, 2992
Cristallo, S., Straniero, O., Piersanti, L., & Gobrecht, D. 2015, ApJS, 219, 40
Damiano, M., Morello, G., Tsiaras, A., Zingales, T., & Tinetti, G. 2017, AJ, 154, 39
de Kok, R., Brogi, M., Snellen, I., et al. 2013, A&A, 554, A82

Des Marais, D. J., Nuth, J. A. I., Allamandola, L. J., et al. 2008, AsBio, 8, 715
Devi, V. M., Benner, D. C., Sung, K., et al. 2018, JQSRT, 218, 203
Domagal-Goldman, S. D., Segura, A., Claire, M. W., Robinson, T. D., & Meadows, V. S. 2014, ApJ, 792, 90
Edwards, S., Roncin, J. Y., Launay, F., & Rostas, F. 1993, JMoSp, 162, 257
Endres, C. P., Schlemmer, S., Schilke, P., Stutzki, J., & Müller, H. S. P. 2016, JMoSp, 327, 95
Endres, P. F. 1967, JChPh, 47, 798
Evans, T. M., Sing, D. K., Goyal, J. M., et al. 2018, AJ, 156, 283
Falk, M., & Whalley, E. 1961, JChPh, 34, 1554
Fortenberry, R. C., & Lee, T. J. 2019, Annual Reports in Computational Chemistry, Vol. 15 (1st ed.; Amsterdam: Elsevier) , 173
Franz, H. B., Trainer, M. G., Malespin, C. A., et al. 2017, P&SS, 138, 44
Galli, D., & Palla, F. 1998, A&A, 335, 403
Gamache, R., Goldman, A., & Rothman, L. 1998, JQSRT, 59, 495
Gordon, I. E., Rothman, L. S., Hill, C., et al. 2017, JQSRT, 203, 3
Greaves, J. S., Richards, A. M. S., Bains, W., et al. 2021, NatAs, 5, 655
Guilluy, G., Sozzetti, A., Brogi, M., et al. 2019, A&A, 625, A107
Habart, E., Walmsley, M., Verstraete, L., et al. 2005, SSRv, 119, 71
Hanson-Heine, M. W. D. 2019, JPCA, 123, 9800
Harada, N., Sakamoto, K., Martín, S., et al. 2018, ApJ, 855, 49
Hedrosa, R. P., Abia, C., Busso, M., et al. 2013, ApJL, 768, L11
Herzberg, G. 1949, Natur, 163, 170
Horner, J. M. 2000, Rev. Environ. Health, 15, 289
Jennings, D. E., & Brault, J. W. 1983, JMoSp, 102, 265
Jo, Y.-S., Seon, K.-I., Min, K.-W., Edelstein, J., & Han, W. 2017, ApJS, 231, 21
Karakas, A. I., & Lattanzio, J. C. 2014, PASA, 31, e030
Katyal, N., Nikolaou, A., Godolt, M., et al. 2019, ApJ, 875, 31
Komasa, J., Piszczatowski, K., Lach, G., et al. 2011, JChTC, 7, 3105
Lammer, H., Sproß, L., Grenfell, J. L., et al. 2019, AsBio, 19, 927
Li, G., Gordon, I. E., Rothman, L. S., et al. 2015, ApJS, 216, 15
Li, H., & Roy, R. L. 2007, JChPh, 126, 224301
Li, X., Heays, A. N., Visser, R., et al. 2013, A&A, 555, A14
Lin, C. Y., Gilbert, A. T. B., & Gill, P. M. W. 2008, Theor. Chem. Acc., 120, 23
Luspay-Kuti, A., Mousis, O., Lunine, J. I., et al. 2018, SSRv, 214, 115
Malathy Devi, V., Kleiner, I., Sams, R. L., et al. 2014, JMoSp, 298, 11
McKemmish, L. K., Masseron, T., Hoeijmakers, H. J., et al. 2019, MNRAS, 488, 2836
Meadows, V. S. 2017, AsBio, 17, 1022
Meadows, V. S., Reinhard, C. T., Arney, G. N., et al. 2018, AsBio, 18, 630
Meshkov, V. V., Stolyarov, A. V., Ermilov, A. Y., et al. 2018, JQSRT, 217, 262
Morales, R., Espinosa, G., Borrell, Y. J., et al. 1998, Biotecnol. Apl, 15, 149
Muller, S., Guélin, M., Dumke, M., Lucas, R., & Combes, F. 2006, A&A, 458, 417
Nomoto, K., Kobayashi, C., & Tominaga, N. 2013, ARA&A, 51, 457
Oyama, V. I., Carle, G. C., Woeller, F., et al. 1980, JGR, 85, 7891
Pawłowski, F., Jørgensen, P., Olsen, J., et al. 2002, JChPh, 116, 6482
Pearson, J. C., Drouin, B. J., & Pickett, H. M. 2005, in IAU Symp. 231, Astrochemistry: Recent Successes and Current Challenges, ed. D. C. Lis, G. A. Blake, & E. Herbst (Cambridge: Cambridge Univ. Press), 270
Pignatari, M., Herwig, F., Hirschi, R., et al. 2016, ApJS, 225, 24

Polyansky, O. L., Kyuberis, A. A., Zobov, N. F., et al. 2018, MNRAS, 480, 2597
Rauscher, T., Heger, A., Hoffman, R. D., & Woosley, S. E. 2002, ApJ, 576, 323
Rees, A. L. 1947, PPS, 59, 998
Ritter, C., Herwig, F., Jones, S., et al. 2018, MNRAS, 480, 538
Rivilla, V. M., Fontani, F., Beltrán, M. T., et al. 2016, ApJ, 826, 161
Roueff, E., Abgrall, H., Czachorowski, P., et al. 2019, A&A, 630, A58
Roy, R. L., Huang, Y., & Jary, C. 2006, JChPh, 125, 164310
Schlegel, H. G. 1974, Tell, 26, 11
Schwartz, C., & Le Roy, R. J. 1987, JMoSp, 121, 420
Seager, S., Bains, W., & Petkowski, J. J. 2016, AsBio, 16, 465
Seager, S., Huang, J., Petkowski, J. J., & Pajusalu, M. 2020, NatAs, 4, 802
Shimajiri, Y., André, P., Braine, J., et al. 2017, A&A, 604, A74
Sousa-Silva, C., Al-Refaie, A. F., Tennyson, J., & Yurchenko, S. N. 2015, MNRAS, 446, 2337
Sousa-Silva, C., Petkowski, J. J., & Seager, S. 2019, PCCP, 21, 18970
Sousa-Silva, C., Yurchenko, S. N., & Tennyson, J. 2013, JMoSp, 288, 28
Sproß, L., Lammer, H., Lee Grenfell, J., et al. 2018, EPSC, 12, 2018
Strobel, D. F., & Shemansky, D. E. 1982, JGR, 87, 1361
Sukhbold, T., Ertl, T., Woosley, S. E., Brown, J. M., & Janka, H. T. 2016, ApJ, 821, 38
Swain, M. R., Tinetti, G., Vasisht, G., et al. 2009, ApJ, 704, 1616
Tennyson, J., Yurchenko, S. N., Al-Refaie, A. F., et al. 2016, JMoSp, 327, 73
Tennyson, J., Yurchenko, S. N., Al-Refaie, A. F., et al. 2020, JQSRT, 255, 107228
Tessenyi, M., Tinetti, G., Savini, G., & Pascale, E. 2013, Icar, 226, 1654
Tinetti, G., Drossart, P., Eccleston, P., et al. 2018, ExA, 46, 135
Tinetti, G., Encrenaz, T., & Coustenis, A. 2013, A&ARv, 21, 63
Tran, H., Flaud, P. M., Fouchet, T., Gabard, T., & Hartmann, J. M. 2006, JQSRT, 101, 306
Vangioni, E., Dvorkin, I., Olive, K. A., et al. 2018, MNRAS, 477, 56
Vázquez, J., & Stanton, J. F. 2006, MolPh, 104, 377
Viti, S., Tennyson, J., & Polyansky, O. L. 1997, MNRAS, 287, 79
Wakelam, V., Bron, E., Cazaux, S., et al. 2017, MolAs, 9, 1
Wakelam, V., & Herbst, E. 2008, ApJ, 680, 371
Wakelam, V., Smith, I. W. M., Herbst, E., et al. 2010, SSRv, 156, 13
Wang, J., Li, D., Goldsmith, P. F., et al. 2020, ApJ, 889, 129
Wang, Y., Tian, F., Li, T., & Hu, Y. 2016, Icar, 266, 15
Webster, C. R., Mahaffy, P. R., Atreya, S. K., et al. 2015, Sci, 347, 415
Western, C. M. 2017, JQSRT, 186, 221
Western, C. M., Carter-Blatchford, L., Crozet, P., et al. 2018, JQSRT, 219, 127
Whitworth, A. P., & Jaffa, S. E. 2018, A&A, 611, A20
Willetts, A., Handy, N. C., Green, W. H., & Jayatilaka, D. 1990, JChPh, 94, 5608
Wolniewicz, L., Simbotin, I., & Dalgarno, A. 1998, ApJS, 115, 293
Wordsworth, R., & Pierrehumbert, R. 2014, ApJL, 785, L20
Yong, D., Lambert, D. L., & Ivans, I. I. 2003, ApJ, 599, 1357
Yu, S., Drouin, B. J., & Miller, C. E. 2014, JChPh, 141, 174302
Yurchenko, S. N., Barber, R. J., Yachmenev, A., et al. 2009, JPCA, 113, 11845
Yurchenko, S. N., & Tennyson, J. 2014, MNRAS, 440, 1649
Zapata Trujillo, J. C., Syme, A.-M., Rowell, K. N., et al. 2021, FrASS, 8, 43
Zhang, Z.-Y., Romano, D., Ivison, R. J., Papadopoulos, P. P., & Matteucci, F. 2018, Natur, 558, 260