H$_3^+$ at the interface between astrochemistry and astroparticle physics

BY JOHN H. BLACK*

Department of Earth and Space Sciences, Chalmers University of Technology,
Onsala Space Observatory, 43992 Onsala, Sweden

The H$_3^+$ molecular ion has been used by Oka and collaborators to trace the rate of ionization by cosmic rays in the interstellar medium. More energetic cosmic rays also produce diffuse γ-radiation. Now that several supernova remnants (SNRs) have been identified as γ-ray sources, it is possible to use spectroscopy of molecular ions to search for enhanced ionization rates that would pinpoint the SNRs as the accelerators of cosmic rays. It is proposed that the warm, dilute molecular gas revealed by H$_3^+$ absorption in the central molecular zone of the Galaxy can also be investigated via radio recombination lines of atoms and possibly triatomic hydrogen.

Keywords: astrochemistry; cosmic rays; interstellar matter; gamma radiation

1. Introduction

Oka and co-workers have used infrared absorption spectroscopy of H$_3^+$ to measure the rate of ionization in the molecular gas of the Galactic interstellar medium [1–7]. In the course of this work, they have discovered a new component of warm ($T \sim 250$ K) and dilute (number density $n < 10^5$ cm$^{-3}$) molecular gas in the central molecular zone of our Galaxy [3,5,7]. Most recently, Indriolo & McCall [8] have used a large collection of H$_3^+$ measurements to determine ionization rates and to show that the rate of ionization by cosmic rays varies from place to place. These findings have broad significance. The cosmic ray particles important for interstellar ionization are those with energies of the order of 100 MeV per nucleon or less. Cosmic ray protons of higher energy, more than 1 GeV, are responsible for the production of neutral pions, $\pi^0$, which decay with the emission of γ-rays. The resulting diffuse γ-radiation has been used as a tracer of the overall distribution of interstellar hydrogen, because the intensity depends only upon the total density of interstellar hydrogen nuclei and on the flux of energetic cosmic rays. Other methods of mapping interstellar matter are sensitive to specific components: the H I 21 cm line traces neutral atomic gas, the $J = 1 \rightarrow 0$ rotational line emission of CO traces molecular gas and thermal emission of interstellar dust traces a combination of the distribution of dust and of the starlight that heats it. By comparison with these other tracers of interstellar matter [9–11], recent

*john.black@chalmers.se

One contribution of 21 to a Theo Murphy Meeting Issue ‘Chemistry, astronomy and physics of H$_3^+$’.
analyses of diffuse γ-radiation suggest the existence of a ‘dark molecular gas’, which is not readily detectable through surveys of the most widely used indirect tracer of interstellar H$_2$, the CO $J = 1 \rightarrow 0$ rotational line [12]. Such interpretation of diffuse γ-radiation requires some independent knowledge of the degree of uniformity of the cosmic ray flux throughout the Galactic disc [13]. The use of H$_3^+$ to infer the ionization rate is crucial. Whether the ‘dark gas’ is a dynamically significant component of dark matter in the Galaxy is still controversial.

It has been known for decades from visible and ultraviolet absorption spectroscopy towards relatively nearby stars (distances up to 3 kiloparsecs) that there is a diffuse and translucent molecular gas, outside self-gravitating structures, that is not well traced by CO line emission [14]. However, the amount of such gas may have been underestimated, according to new submillimetre-wave absorption spectroscopy of interstellar OH$^+$, H$_2$O$^+$, HF and other species that flourish in regions where the molecular fraction $f = 2n(H_2)/(2n(H_2) + n(H)) \ll 1$ [15–18]. These sensitive submillimetre-wave observations carried out with the Herschel Space Observatory have made it possible to survey the composition of diffuse and translucent gas over much greater distances—up to 12 kiloparsecs—across the Galaxy than was possible through visible and ultraviolet absorption spectroscopy.

By concentrating on interstellar matter at the molecular level, we sometimes lose contact with larger issues. The interstellar medium is a dynamic and, in places, quite violent system, where the overall balance of energy, pressure and ionization is controlled in part by supernovae explosions and by cosmic rays. There remains an underlying mystery about the origin of the cosmic rays. The energy spectra and nuclear composition of cosmic rays are measured, but the effects of propagation through the magnetic fields and gas of the Galaxy (and through the heliosphere at the lower energies) have effectively erased the paths from their sources. Diffusive shock acceleration in expanding supernova remnants (SNRs) is generally thought to be the main mechanism responsible for the Galactic cosmic rays, but detailed observational tests are only now becoming possible. The Large Area Telescope on the Fermi Gamma-ray Space Telescope has recently made it possible to identify localized sources of energetic ($h\nu > 1$ GeV) γ-rays with well-known SNRs. In most of these cases, the SNR is closely associated with a large concentration of the interstellar molecular gas (see §3). The γ-ray spectra are best fit by models in which the dominant radiation processes are hadronic; that is, interactions between energetic cosmic-ray nuclei and ambient hydrogen yield pions, which subsequently decay with a characteristic spectrum of γ-rays. If this is the correct interpretation, then there is strong evidence that SNRs are indeed sources of cosmic rays. Moreover, the total energy in relativistic particles can be estimated for specific sources. Molecular astrophysics can help us to confirm and extend this work. Observations of the molecular gas associated with SNRs help characterize the conditions in the γ-ray sources. An abundance of molecular ions such as H$_3^+$ is sensitive to the rate of ionization by particles with energy $E < 100$ MeV, while the diffuse γ-radiation is caused by particles of higher energy, $E \geq 1$ GeV. As shown below, the forbidden vibration–rotation spectrum of H$_3^+$ may provide an even more direct probe of ionization rate and may become detectable near the sources where the ionization rate is orders of magnitude larger than the Galactic average. Thus, it will become possible to place constraints on the acceleration of cosmic rays at $E < 100$ MeV and on their ability to escape the regions where they are accelerated.

Phil. Trans. R. Soc. A (2012)
The interstellar chemistry of $H_3^+$ is admirably simple, which makes it an excellent diagnostic probe of superthermal ionization. In the neutral interstellar medium, virtually all H-ionizing starlight is confined to the photoionized nebulae that surround hot stars, which are the main thermal sources of ultraviolet radiation. Thus, interstellar molecular hydrogen is ionized mainly by cosmic ray protons and in some localized regions by X-rays that can penetrate rather thick interstellar clouds. The range of a 10 MeV proton is $3.2 \times 10^{22}$ hydrogen atoms cm$^{-2}$, which corresponds to a fairly thick interstellar cloud of visual extinction $A_V = 20$ magnitudes. More energetic protons travel even farther against ionization losses.

The initial ionization of $H_2$

$$H_2 + CR \rightarrow H_2^+ + e^- \quad (2.1)$$

is quickly followed in a fully molecular region by formation of $H_3^+$

$$H_2^+ + H_2 \rightarrow H_3^+ + H, \quad (2.2)$$

and in a partly atomic region, the $H_2^+$ can also be removed by

$$H_2^+ + H \rightarrow H_2 + H^+ \quad (2.3)$$

and

$$H_2^+ + e^- \rightarrow H + H. \quad (2.4)$$

In most of the literature on interstellar $H_3^+$, the fate of its predecessor, $H_2^+$, is left unexamined. The lifetime of $H_2^+$ against the above reactions (equations (2.2)–(2.4)) is so short that it should have a negligible abundance in steady state. A low steady-state abundance does not render $H_2^+$ completely unobservable. The ionization of $H_2$—whether by photoionization, electron impact or proton impact—leaves the product $H_2^+$ in a broad distribution of excited vibrational states. The vibrationally excited $H_2^+$ can decay towards the ground state by quadrupole vibration–rotation transitions, with lifetimes of the order of $10^6$ s or less [19]. While this seems a rather slow process, radiative relaxation is at least as rapid as reactive or inelastic collisions with $H_2$ and $H$, wherever the density is of the order of 1000 cm$^{-3}$ or less. Even at higher densities, quadrupole radiation will occur, but its emissivity will be reduced by collisions. Thus, during its short lifetime, interstellar $H_2^+$ will produce a characteristic infrared emission spectrum. In the low-density limit, the intensity of this spectrum depends only upon the rate of ionization of $H_2$, the relative strengths of the various vibration–rotation lines being fixed by the array of quadrupole transition probabilities. The initial distribution of vibrational states of $H_2^+$, following ionization of $H_2$, is fairly well known both experimentally and theoretically [20–22], except for the last two bound states, $v = 18$ and $v = 19$, which support four and two rotational levels, respectively. Although $H_2^+$ will be vibrationally hot at birth, its initial rotational distribution is likely to reflect that of its parent $H_2$.

Although it may be difficult to imagine such a reactive molecular ion being spectroscopically measurable through forbidden electric quadrupole transitions, similarly ‘impossible’ things happen all the time in space. The transition at 21 cm wavelength between the hyperfine-structure sublevels of the ground state of the $H_2^+$.

*Phil. Trans. R. Soc. A (2012)*
Astrochemistry and astroparticle physics

H atom has a radiative lifetime of $1.1 \times 10^7$ years, much longer than the collision time in much of the interstellar medium, yet enough atoms perform this unlikely transition that the 21 cm line is easily observed.

There may be interesting chemical consequences of hot, transient $\text{H}_2^+$. The reaction

$$\text{H}_2^+(v) + \text{He} \rightarrow \text{HeH}^+ + \text{H}$$

is endoergic by 0.81 eV for the ground state ($v = 0$), but is fairly rapid for higher states ($v > 3$). Like radiative association, most other proposed sources of $\text{HeH}^+$ in space are very slow processes [23]. In regions where $\text{H}_2$ is subjected to rapid ionization, $\text{HeH}^+$ may be an interesting by-product. This ion has the advantage of a permanent dipole moment and strongly allowed rotational and vibration–rotation transitions. A bit more speculative is a possible source of $\text{NeH}^+$ by

$$\text{H}_2^+ + \text{Ne} \rightarrow \text{NeH}^+ + \text{H}.$$  \hfill (2.6)

$\text{NeH}^+$ has not yet been observed in space, but its spectrum is known. The lowest rotational transition $J = 1–0$ lies at 1.039 THz, and the $v = 1–0$ band at low $J$ is centred around 2600–2800 cm$^{-1}$, close to low-lying vibration–rotation transitions of $\text{H}_3^+$.

The astronomical vibration–rotation spectrum of $\text{H}_2^+$ can be predicted. It will be detectable only in situations where a high rate of ionization persists over a large column density of $\text{H}_2$. There are two obvious kinds of targets for detecting $\text{H}_2^+$. One is a molecular cloud that is directly associated with a source of cosmic rays, where the flux of cosmic-ray nuclei might be orders of magnitude higher than the average flux in the Galactic disc. The second is the molecular torus that is thought to surround and to provide fuel for the X-ray-emitting accretion disc orbiting a supermassive black hole in an active galactic nucleus.

3. Searching for the sources of cosmic rays

Cosmic rays of interest here wiggle around in the Galactic magnetic field and thus do not follow straight-line paths from their sources to us. Direct evidence of the sources of Galactic cosmic rays is lacking. Although cosmic rays are observed with energies up to $10^{20}$ eV per nucleon, even the comparatively modest ones at 10 MeV to 1 TeV require extraordinary natural accelerators. It has long been thought that most cosmic-ray nuclei are formed by a process of diffusive shock acceleration, primarily in SNRs that expand into the surrounding interstellar medium at a high velocity, compressing gas and magnetic field lines. Several recent observations of $\gamma$-rays from SNRs strongly suggest that these remnants are indeed the sources of cosmic rays. All of these SNRs are known to be associated with concentrations of neutral gas, and their $\gamma$-ray spectra are best fit by models in which the dominant emission mechanism is the decay of $\pi^0$ mesons, which in turn were produced by hadronic interactions: that is, by interaction of cosmic-ray protons with interstellar atoms and molecules. The general Galactic background of cosmic rays is sufficient to produce observable diffuse $\gamma$-radiation from the interstellar medium [24]. This diffuse emission has been mapped with increasing sensitivity and improving angular resolution over the years. Most recently, the comparison of the distribution of diffuse $\gamma$-rays with other tracers of interstellar matter has been used to infer a significant component of the so-called dark molecular gas.
that is not very emissive in the CO $J = 1 \rightarrow 0$ rotational line [9–12]. This is probably the same partly atomic, partly molecular gas that some of us have labelled ‘diffuse’ and ‘translucent’ for decades, but the new data now afford an improved accounting of it over a large part of the Galaxy. The Fermi Gamma-ray Space Telescope and several ground-based arrays sensitive to TeV photons have recently achieved such improvements in sensitivity and resolution that it is now possible to recognize the localized, supernova-related sources above the diffuse background of $\gamma$-rays [25–38].

The developments in $\gamma$-ray astronomy suggest that there is a role for low-energy molecular astrophysics, too, in the search for the origins of cosmic rays. The production rates of interstellar H$_2^+$ and H$_3^+$ are directly related to the cosmic-ray flux. Indriolo & McCall [8] have derived ionization rates from a large sample of H$_3^+$ measurements and have shown that the ionization rate varies from place to place in the Galaxy. In at least one case, the line of sight near IC 443, there is a higher-than-average ionization rate that can plausibly be attributed to a SNR [39]. The growing body of data on H$_3^+$ in the central molecular zone (CMZ) of the Galaxy suggests a high rate of ionization there [2,3,5–7].

In order to understand the possible molecular signatures of greatly enhanced ionization rates near cosmic-ray sources, Black and co-workers have been developing models of highly irradiated molecular gas [40,41]. The models refer specifically to the SNRs that have been detected as $\gamma$-ray sources with $\gamma$-ray spectra that are well fit by hadronic processes. There are two distinct aspects of the models. One is to extrapolate the inferred cosmic-ray spectra to lower energies, 10–100 MeV, at which ionization is most efficient over hydrogen columns greater than $10^{22}$ cm$^{-2}$. This makes it possible to estimate the ionization rates in the same gas that radiates the observed $\gamma$-rays. Unfortunately, the estimated rate is extremely sensitive to the description of the low-energy spectrum, which probably has either a low-energy cut-off or a change of slope somewhere in the ionizing range of the particle spectrum. For example, the $\gamma$-ray emission from the W51C SNR requires that the product of total cosmic-ray energy and average interstellar density be $W_p n_H \approx 7.7 \times 10^{51}$ ergs cm$^{-3}$, where the total energy in energetic protons is integrated from a lower limit of 10 MeV. The resulting ionization rate is very sensitive to the energy at which the particle spectrum changes slope: for W51C, the inferred ionization rate is approximately 0.03 times the Galactic average value if the slope changes at 1 GeV, but four times the Galactic average value if the slope changes at 30 MeV. Much larger enhancements are deduced from several of the other SNRs, the most extreme being W49B, for which the interstellar ionization rate is at least 1800 times the Galactic average (slope change at 1 GeV) and could be enhanced by a factor of $4 \times 10^5$ (slope change at 30 MeV). Indriolo et al. [39] have inferred the H$_2$ ionization rates of $\zeta \approx 2 \times 10^{-15}$ s$^{-1}$ for two sightlines near the IC 443 SNR. Ceccarelli et al. [42] estimate a similar value of $\zeta$ from emission line measurements of HCO$^+$ and DCO$^+$ near W51C. Both these results are in accord with our analysis of the $\gamma$-ray emission when the low-energy limit is 10 MeV and the slope of the particle spectrum changes at 30 MeV or less. These cases represent only a factor of 10 enhancement over the Galactic average ionization rate. If these parameters apply to the more extreme cases of W49B and 3C391, then cosmic-ray ionization rates $\zeta > 10^{-13}$ s$^{-1}$ might be found near these sources. At this level, detection of H$_3^+$ and other rare ions might become possible. The crucial parameter for the production
Figure 1. Predicted emission spectrum of H$_2^+$ in the mid-infrared. This model has a molecular density $n$(H$_2$) = 1000 cm$^{-3}$, a kinetic temperature $T$ = 150 K and a cosmic ray exposure $\eta$ = 10$^9$ s$^{-1}$ cm$^{-2}$. The spectrum has been convolved to a resolution of 0.001 μm.

of transient molecular ions is the ion production rate integrated over path length, which we call the cosmic-ray exposure

$$\eta = \int \zeta n(H_2) \, d\ell,$$

which approaches $\eta \approx \zeta \int n(H_2) \, d\ell$ if the column density is not so large that $\zeta$ is severely diminished by ionization losses. If the extreme $\gamma$-radiating sources have hydrogen column densities as large as $\int n(H_2) \, d\ell = N(H_2) \sim 10^{22}$ cm$^{-2}$, then the cosmic-ray exposure could approach $\eta \sim 10^9$ s$^{-1}$ cm$^{-2}$.

First results of our model of the coupled ion chemistry and excitation have been published by Becker et al. [40] for $\eta = 10^9$ s$^{-1}$ cm$^{-2}$ over a low average density, $n$(H$_2$) = 100 cm$^{-3}$. Similar models have been computed for a higher density, $n$(H$_2$) = 1000 cm$^{-3}$, and the predicted infrared spectrum of H$_2^+$ is displayed in figure 1. The excitation model of H$_2^+$ has been computed through use of a slightly modified version of the RADEX program [43]. It includes a state-specific formation distribution function for all bound vibration–rotation levels of the ground electronic state of H$_2^+$, assuming that the rotational distribution of parent H$_2$ is preserved in the product ion. It includes destruction by the reactions with H$_2$, H and e$^-$ mentioned earlier. It includes all quadrupole vibration–rotation transitions and it takes into account estimates of inelastic collisions with H$_2$ and e$^-$. The rate equations (incorporating the formation and destruction processes
explicitly) that describe the level populations are solved simultaneously with a simplified form of the radiative transfer and line formation. The resulting column density of H$_2^+$ is $4.6 \times 10^{14}$ cm$^{-2}$, compared with $N$(H$_2$) = $10^{22}$ cm$^{-2}$. The kinetic temperature in the model is 150 K, and the rotational distribution of nascent H$_2^+$ is characterized by a somewhat higher rotational temperature of 300 K. The character of the emitted spectrum is completely dominated by the non-thermal vibrational distribution. If the computed populations of the 40 most populous levels are forced to fit a Boltzmann distribution, the result is 14 000 K, but this is a fictitious temperature. Again, the crucial elements in the line intensities are the high rate of ionization of H$_2$ and the broad distribution of product ions over excited vibrational states. As seen in figure 1, the strongest lines appear at wavelengths of 5–6 μm.

The computed chemistry and excitation of H$_2^+$ have been used to compute the closely related abundance and excitation of HeH$^+$ and H$_3^+$ for the same conditions. Parts of the combined spectra for all three ions are shown in figure 2. This preliminary model includes only a small subset of the H$_3^+$ spectrum, with rotational quantum numbers limited to $J \leq 5$.

4. Rydberg atoms and molecules

As mentioned already, the observations of metastable H$_3^+$ absorbing over long sightlines towards the Galactic Centre have been taken as evidence of warm (200–300 K), dilute molecular gas throughout the CMZ of our Galaxy. The inferred conditions in this gas are favourable for the production of radio recombination lines—that is, transitions between adjacent, highly excited Rydberg states of atoms like H and C, which are populated by radiative recombination of the ions

$$C^+ + e^- \rightarrow C(n^*) + \gamma. \quad (4.1)$$

The subsequent radiative cascade from a highly excited Rydberg state of principal quantum number $n^*$ to $n^* - 1$ and lower states yields several photons before the atom has relaxed to the ground state or has suffered some other kind of interaction. Astronomical observations at decametre wavelengths have revealed absorption lines of interstellar carbon atoms in states as high as $n = 1009$ in $\Delta n = 4$ transitions that flank the stronger $n \approx 630 \Delta n = 1$ lines near 26 MHz frequency [44]. At $n = 1009$, the binding energy of the state is only $-0.089$ cm$^{-1}$ relative to C$^+\left(2p^2P_{1/2}\right)$, and the radius of the classical electron orbit is a startling 54 μm. Carbon is favoured because it remains photoionized by ultraviolet starlight through much of the diffuse, neutral interstellar medium, so that C$^+$ is often the most abundant positive ion. We can imagine that there are regions where the superthermal ionization rate of H and H$_2$ by cosmic rays (or nearby X-ray sources) is high enough that H$^+$ leads to H$(n^*)$ in Rydberg states. Indeed, this effect has been seen in recombination lines towards the Galactic Centre. Although the C lines appear in absorption at $n > 300$ [45], diffuse emission lines of both H and C ($n = 270$ to 273) have been measured at frequencies near 327 MHz [46,47]. The angular resolution of these observations is too coarse to permit strong constraints on some of the physical parameters of the gas. Fortunately, it appears that related gas can be detected via recombination lines at much higher radio frequencies with the Atacama Large Millimetre/submillimetre Array, which is now going
Astrochemistry and astroparticle physics

Figure 2. (a) Predicted emission spectra of H$_3^+$ (red), H$_2^+$ (black) and HeH$^+$ (blue) in the infrared. The same conditions as in figure 1 apply here, with kinetic temperature of 150 K. H$_3^+$ adds stronger lines at 3–4.5 μm wavelength and dominates the far-infrared spectrum at 50–150 μm wavelength. The contribution of HeH$^+$ is imperceptible on these scales. (b) The intensity per unit frequency displayed as a function of wavelength at low resolution exaggerates the peak intensities of lines at wavelengths $\lambda > 100\,\mu$m in relation to those at shorter wavelengths. (Online version in colour.)

into operation in Chile. Further observation and analysis will show whether the gas detected through Rydberg atoms is closely related to the gas that produces H$_3^+$ absorption throughout the CMZ.

Radio recombination line observations of Rydberg atoms in astronomy date back to 1964 and are quite extensive. The carbon and sulphur lines, in particular, trace components of the interstellar medium where most of the hydrogen remains neutral. This raises the question whether it might be possible to detect
corresponding lines between highly excited Rydberg states in molecules like H₃. Herzberg [48] proved spectroscopically that the H₃ molecule possesses stable Rydberg states, even though it is unstable in its ground state. It is expected that dissociative recombination of H₃⁺ will remain more rapid than radiative recombination, which would otherwise populate the highly excited states. The nd Rydberg series that converges on the ground level of ortho-H₃⁺ has been studied up to n ≈ 100 by field ionization [49]. Indeed, in subsequent experiments, the natural lifetimes of H₃ Rydberg states have been measured and have been found to decrease with increasing n, with the implication that radiative transitions are unimportant compared with predissociation in these states [50]. In the near future, we will make some quantitative estimates of the H₃ recombination line intensities, based upon the known H₃⁺ abundance and estimated electron densities in the Galactic Centre region, but they will probably be quite weak. In any case, the high-n recombination lines of H₃ will have a constant offset in fractional frequency, Δν/ν = 3.63 × 10⁻⁴, from the corresponding H lines. This corresponds to a constant Doppler shift of −108.85 km s⁻¹ in the line profiles.

There is another class of molecular Rydberg states that occur at large internuclear separations in the ionic core. This includes the ‘heavy Rydberg states’ of H₂ that correlate with H⁻–H⁺ ion pairs [51–53]. Although these states can be populated efficiently in the laboratory by laser excitation from the H₂ EF 1Σ⁺ ground state, they might be excited in space only in a small fraction of the exceedingly unlikely collisions between H⁻ and H⁺. It will be interesting to try to imagine other processes that could excite Rydberg states of simple molecules in space.

5. Conclusions

It is a paradox of nature that giant molecular clouds, the largest low-energy structures in the Galaxy, are recognizable γ-ray sources because they are such big targets for cosmic rays. It is interesting that low-energy molecular astrophysics provides clues about one of the central problems in high-energy particle astrophysics, namely the origin of cosmic rays.

Support of the Swedish National Space Board is gratefully acknowledged.

References

17 Gerin, M. et al. 2010 Interstellar OH$^+$, H$_2$O$^+$ and H$_3$O$^+$ along the sight-line to G10.6-0.4. Astron. Astrophys. 518, L110. (doi:10.1051/0004-6361/201015476)
20 Flannery, M. R., Tai, H. & Albritton, D. L. 1977 Cross sections for the photoionization of H$_2$(X$^1\Sigma^+$, $v_i=0–14$) with the formation of H$_2^+$(X$^2\Sigma^+$, $v_f=0–18$), and vibrational overlaps and $R^a$-centroids for the associated vibrational transitions. Atomic Data Nucl. Data Tables 20, 563–585. (doi:10.1016/0092-640X(77)90039-0)


