PERSPECTIVES: ASTRONOMY

The Complex Story of H_2

William Klemperer

Molecular hydrogen (H_2) is the most abundant species in the universe. Although it is also the simplest molecule, its electronic states frequently show surprising complexity. Evidence is emerging that this complexity may play a key role in astronomical processes involving H_2, with important implications for chemical evolution in space.

Atomic hydrogen (H), whose hyperfine transition at 1420 MHz initiated atomic and molecular radioastronomy, pervades the diffuse gaseous matter in galaxies, but in dense interstellar regions, including the centers of galaxies, hydrogen is molecular. Unfortunately, the terrestrially accessible absorption spectrum of H_2 is weak. H_2 can only be observed from Earth's surface if it forms a strongly heated gas, whose vibrational emission is readily seen (*1*). Electronic spectra of H_2 can be observed by extraterrestrial observatories (*2*, *3*) (see the figures).

The electronic states of H_2 can be exceptionally simple. Its ground state, $^1\Sigma_g^+$, is the pedagogical model for the chemical bond, treated conveniently by valence bond or molecular orbital theory. The potential energy function of this state is a model for the Born-Oppenheimer approximation, according to which nuclear and electronic motion can be treated separately because of the vastly different time scales of their motion. The excited triplet state, $^3\Sigma_u$, is essentially repulsive, and optical transitions between the singlet and triplet state are highly forbidden.

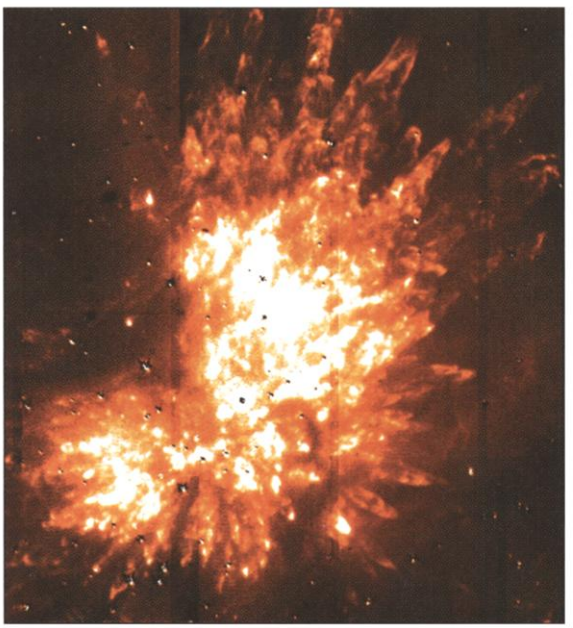
In these two states, both atoms are in the (energetically isolated) ground state, H(1s). If one of the two atoms is excited, however, a very large number of atomic states become accessible. The electronic states of H_2 then become exceptionally complex, with multiple minima and such strong entanglement between nuclear and electronic motion that the Born-Oppenheimer approximation is violated and the concept of a potential energy curve is not useful.

The characteristic features of the electronic states of H_2 and its optical response are seen most prominently in illuminated regions such as diffuse interstellar clouds or in photon-dominated regions near bright stars. In these regions, atomic and molecular hydrogen interconvert by photodissociation of H_2 and surface recombination (*4*) of two H atoms. The galactic radiation field has a sharp upper limit at 13.6 eV because absorption through photoionization of atomic hydrogen leaves very little radiation with higher energy. The dissociation energy of the H_2 ground state is 4.5 eV, and the lowest energy of optically allowed states of H_2 relative to 2H is 10.2 eV. Hence, photon energies >14.7 eV would be required for direct photodissociation. This exceeds the observed upper limit, and photodissociation of H_2 must therefore proceed through a highly nonlinear, indirect route.

The photodissociation occurs through a two-step process. Energy is first absorbed through allowed transitions between the ground state and two excited states, B and C. These transitions have sharply defined energies between 11 and 13 eV. Thus, only selected frequencies of radiation are absorbed, which are attenuated exponentially, thereby shielding the inner part of the cloud by the outermost layer. The equilibrium internuclear separation in B and C is much larger than in the ground state. During absorption, the nuclei do not move, but subsequently they move apart. The H_2 molecule dissociates upon reemission from the bound B and C vibrational states to the dissociated region of the ground state when the nuclei have moved far enough apart. The strong nonlinearity of the system is a result of the sharp line absorption. The system is self-shielding for large volumes of H_2.

The higher energy electronic states of H_2 have many consequences beyond their different internuclear separations. The optical response of H_2 depends on its excited electronic states, which are complex (*5*). The density of electronic energy levels for two H atoms bound to each other where one is excited is very high. The interaction between the two atoms can be quite long range, varying as $1/R^3$, where R is the internuclear separation. A second long-range interaction is that of the ionic states, H^+H^-. The ionic curve and that for the H(1s)H(n = 4) state (where n is the principal quantum number of H) cross at very large distances. The crossing of the ionic state with the n = 3 state occurs at 20 Å and that with the n = 2 state at 4 Å. The character of the electronic wave function thus varies with internuclear distance and can change character sharply in the neighborhood of a crossing. Multiple curve crossings can result in a complex mixing of electronic and nuclear motion.

The first H_2 state recognized as having multiple minima is still labeled EF because



Molecular hydrogen abounds. This image of the Kleinman Low nebula in Orion, taken with the Subaru Telescope in Japan, is formed by infrared emission from the vibrational transition of molecular hydrogen from the first excited to the ground state.

The Orion Nebula. In this lower resolution image from the Subaru Telescope, the Kleinman Low can be seen at the top right.

The author is in the Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. E-mail: klemperer@chemistry.harvard.edu

it was initially thought to consist of two distinct electronic states, E and F, with internuclear distances of 1.01 and 2.31 Å, respectively. Theoretical analysis (*6*) showed the existence of a single adiabatic curve. The idea of multiple isomers on a single electronic potential surface is accepted as standard for polyatomic molecules but was thought to be a rarity for diatomic molecules, which have a single nuclear geometry coordinate. Several electronic states of $^1\Sigma_g^+$ symmetry with multiple minima are now known (*4*). Theory and experiment have also shown this phenomenon for a second electronic symmetry (*7*). The potential energy curve in the latter system has minima at 1.05 and 5.7 Å.

The complexity of the electronic energy levels of H_2 is also seen in their electric and magnetic susceptibilities. The surprising paramagnetism observed in one state (*4*) is explained by the mixing in of another nearby excited state. A molecular Stark effect has also been reported for excited states of H_2 (*8*). The apparent electric dipole moment of these states results from the frequent close proximity of gerade and ungerade states, that is, states whose wave functions are respectively symmetric and antisymmetric with respect to inversion.

The reactivity of H_2 depends on which electronic states are accessible. Among the simplest processes are electron and proton transfer. The rich chemistry of organic species that has been observed by radioastronomers (*9*) occurs primarily in massive dense dark molecular clouds, which cannot be penetrated by starlight. Here, molecular processing is initiated by cosmic ray ionization of H_2 and He. The latter is 1000 times more abundant than CO.

The secondary reactions of the ion products, H_2^+ and He^+, with the prevalent neutral H_2 exhibit electronic complexity. Collisions result in facile conversion of H_2^+ to H_3^+ (*10*). The highly exothermic reactions $He^+ + H_2 \rightarrow HeH^+ + H$, $He + H_2^+$, or $He + H + H^+$ essentially do not occur, whereas the much less exothermic reaction $H_2^+ + He \rightarrow HeH^+ + H$ occurs at collision frequency (*11*). It is the lack of H_2 reactivity with He^+ that permits the rich interstellar organic chemistry observed by radioastronomers. The latter has its origin in the copious production of carbon ions by the reaction $He^+ + CO \rightarrow C^+ + O + He$. C^+ adds efficiently to existing organic species, thereby increasing their carbon chain length and producing an interesting organic-rich molecular universe.

References and Notes

1. T. R. Geballe, *ASP Conf. Ser.* **41**, 147 (1993).
2. L. Spitzer, W. D. Cochran, A Hirshfeld, *Astrophys. J. Suppl. Ser.* **28**, 373 (1974).
3. D. C. Morton, *Astrophys. J.* **197**, 85 (1975).
4. O. Biham, I. Furman, V. Pirronello, G. Vidale, *Astrophys. J.* **553**, 595 (2001).
5. J. Rychlewski, *Adv. Quantum. Chem.* **31**, 173 (1999).
6. E. R. Davidson, *J. Chem. Phys.* **33**, 1577 (1960).
7. A. de Lange, W. Hogervorst, W. Ubachs, L. Wolneiwicz, *Phys. Rev. Lett.* **86**, 2988 (2001).
8. J. Stark, *Wien and Harms Handbuch d. Experimentalphysik*, vol. XXI (Akademische Verlagsgesellschaft, Leipzig, 1927), p. 442.
9. E. Herbst, *Annu. Rev. Phys. Chem.* **46**, 27 (1995).
10. B. J. McCall, T. R. Geballe, K. H. Hinkle, T. Oka, *Astrophys. J.* **522**, 338 (1999).
11. B. H. Mahan, *Acc. Chem. Res.* **8**, 55 (1970).

PERSPECTIVES: BIOMEDICINE

Reconstructing Myotonic Dystrophy

Stephen J. Tapscott and Charles A. Thornton

Myotonic dystrophy (DM) is an inherited disease characterized by progressive muscle weakness, hyperexcitability of the muscle membrane (myotonia), ocular cataracts, and cardiac arrhythmias. Some families with DM have a mutation in a gene on chromosome 19q13.3 (the *DM1* locus), whereas others have a mutation in a gene on chromosome 3q21 (the *DM2* locus). In 1992, the DM1 mutation was identified as an unusually large CTG trinucleotide repeat in the *DMPK* gene. However, as the mutation is in the noncoding region of *DMPK*, it is not clear how this expanded repeat actually causes DM1. Many investigators have been hoping that identification of the DM2 mutation would reveal a common pathogenic mechanism underlying both forms of the disease. Now, these hopes have been realized with the report by Liquori *et al.* (*1*) on page 864 of this issue. These investigators reveal that the DM2 mutation is a huge expansion of a tetranucleotide repeat (CCTG) in a noncoding region of the *ZNF9* gene. This finding adds strong support to the growing suspicion that the mutant RNA contributes to the constellation of features that characterize DM.

In many diseases caused by expansions of a trinucleotide repeat, such as Huntington's disease, a CAG repeat is located in a region of the gene that is translated into protein. Consequently, the protein contains an expanded stretch of polyglutamines that confers a toxic property on the protein, leading to the pathology of the disease. In contrast, the CTG repeat at the *DM1* locus is located in the 3′-untranslated region of the *DMPK* gene, a region that is transcribed into RNA but never translated into protein. Because the CTG expansion at the *DM1* locus does not alter the protein sequence encoded by *DMPK*, the mechanism of pathogenesis in DM1 must be different from that in the polyglutamine encoding CAG-repeat diseases.

Three different theories have been proposed to explain how repeat expansions in the *DMPK* gene result in DM1. The expanded CUG repeat in RNA transcribed from the mutant gene has important effects on the metabolism of this and other transcripts (*2*). Most dramatically, the mutant RNA accumulates in discrete foci in the cell nucleus rather than being transported to the cytoplasm, where translation of mRNA into protein normally takes place (see the figure) (*3, 4*). Therefore, DM could be caused by a decrease in the amount of DMPK protein (because the RNA from the mutant allele is in the nucleus and is not available for translation into protein) or by the abnormal RNA itself, which may interact with and disrupt the activity of nuclear RNA binding proteins. A third possibility is that DM is caused by a decrease in the amount of the SIX5 transcription factor, because the CTG expansion at the *DM1* locus suppresses the expression of the adjacent *SIX5* gene.

There is experimental support for each of these theories. Disruption of the *Dmpk* gene in mice causes a cardiac arrhythmia similar to that seen in DM (*5–7*), but does not cause other features of the disease such as myotonia or cataracts. Myotonia and muscle disease can be induced in mice with a transgene that is transcribed into an RNA with a large CUG repeat in its untranslated region (*8*). It is the mutant RNA that promotes muscle disease in these mice, because animals do not develop disease if the transgene containing the expanded repeat is not transcribed into RNA. Because expression of the transgene in this experiment was restricted to skeletal muscle, it remains unclear whether expression of the mutant RNA in other cell types could produce additional features of the disease. Finally, disruption of *Six5* in mice causes ocular cataracts (*9–12*). From these

S. J. Tapscott is in the Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. E-mail: stapscot@fhcrc.org. C. A. Thornton is in the Department of Neurology, School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642, USA. E-mail: charles_thornton@urmc.rochester.edu