Spectroscopic networks

Attila G. Császár *, Tibor Furtenbacher

Laboratory of Molecular Spectroscopy, Institute of Chemistry, Eötvös University, H-1518 Budapest 112, P.O. Box 32, Hungary

A B S T R A C T

Approaches related to graph theory are investigated which allow a better understanding and yield routes for systematic enlargement and improvement of experimental spectroscopic line lists of molecules. The proposed protocols are based on the fact that quantum mechanics builds, in a simple and natural way, large-scale, weighted, undirected graphs, whereby the vertices are discrete energy levels, the edges are transitions, and the weights are transition intensities. A small part of molecular quantum mechanical graphs can be probed experimentally via high-resolution spectroscopic techniques, while the complete graph encompassing the full line list information for a given molecule can be obtained through sophisticated variational nuclear motion computations. Both approaches yield what one may call spectroscopic networks (SNs). It is shown on the example of the $HD^{16}O$ isotopologue of the water molecule that both the measured and the computed one-photon absorption SNs have a scale-free behavior with all of the usual consequences, including appearance of hubs, robustness, error tolerance, and the "small-world" property. For the complete computed "deterministic" network the scale-free property holds if a realistic intensity cut-off is employed during its build-up, thus introducing "stochasticity". The graph-theoretical view of molecular spectra offers several new ideas for improving the accuracy and robustness of the information systems containing high-resolution spectroscopic data.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Complete characterization of high-resolution spectra of a considerable number of atoms and molecules, starting from the microwave and extending to the ultraviolet, is a prerequisite for modeling and understanding of many processes and phenomena in physics, chemistry, and engineering. For example, modelers of the atmospheres of planets and cool stars as well as those investigating combustions in rocket exhausts and turbine engines need detailed line-by-line information which only elaborate spectroscopic measurements can provide. Recent advances in molecular spectroscopy led to a considerable increase in the extent of experimental high-resolution spectroscopic data. Some of these data have been deposited, sometimes in a critically evaluated and annotated form, in databases; see, for example, Refs. [1–9]. Treatment of the rapidly increasing information and the desire to turn information into knowledge requires sophisticated procedures in the generation, accumulation, validation, handling, visualization, and distribution of spectroscopic data. The present study is based on our belief that in search of useful new tools for improving spectroscopic line lists, graph (network) theory and its sophisticated polynomial algorithms offer interesting possibilities which so far have not been explored.

For individual molecules quantum mechanics (QM) offers a simple, natural, and elegant way to build large-scale deterministic, undirected, (weighted) graphs (networks), made up of energy levels as vertices (nodes), allowed transitions between the levels as edges (links), and weights related to transition intensities. The topology of QM networks can be determined experimentally via high-resolution techniques of molecular spectroscopy (see, *e.g.*, Fig. 1). Thus, practical realizations of QM networks can be called spectroscopic networks (SNs) [10–13]. The robust organizing principle of SNs is provided by QM selection rules; different transitions and transition intensities characterize different spectroscopic techniques. It is important to emphasize that even in the experimentally most thoroughly studied cases the observable transitions form just a tiny part of all the transitions allowed [12,13]. The complete line list information about allowed transitions, corresponding to a large SN, can only be determined *via* sophisticated quantum mechanical computations; see, for example, Refs. [14–19].

Complex graphs (networks) appear ubiquitously in nature, society, communication, and elsewhere [20–26]. Traditionally, networks have been examined and interpreted via random graph theory, developed and popularized by Erdős and Rényi [27]. Random graphs and thus random networks exhibit a well-defined, characteristic mean value for the probability $P(k)$ characterizing vertices having degree k. In this case most of the nodes have an about average number of links and increasingly less and less have a higher or lower number of connections. However, about 10 years
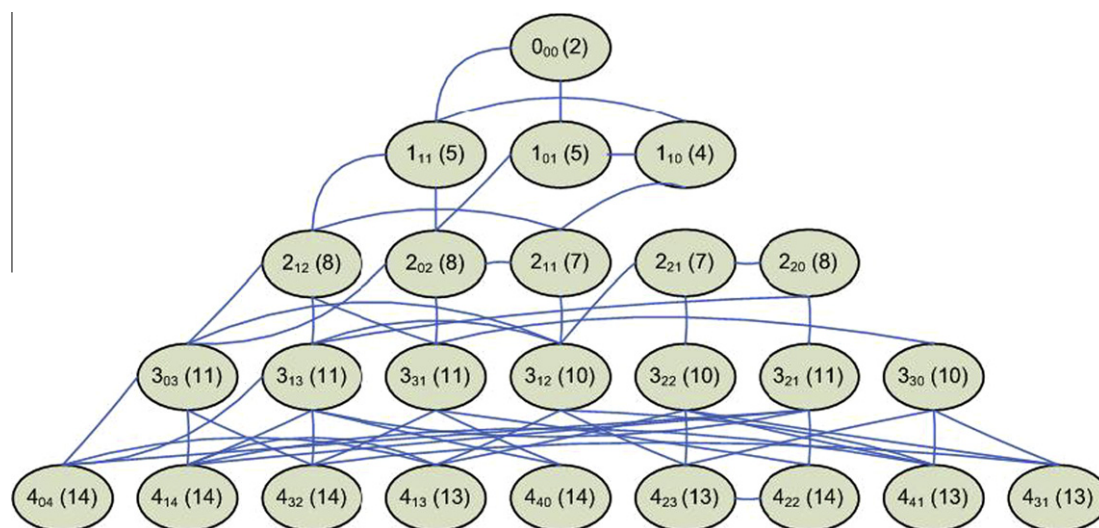
Fig. 1. The experimental purely rotational one-photon absorption spectroscopic network (SN) of the $HD^{16}O$ isotopologue of the water molecule for its ground vibrational state (the single root of the graph is the $J_{K_aK_c} = 0_{00}$ state), up to $J = 4$. The number in parentheses in each oval provides the total number of allowed rotational transitions starting or ending on the given energy level. Missing from the figure are the weights of the links associated with absorption intensities.

ago it was recognized [20] that most networks of practical interest do not follow the laws emerging from the use of random graphs. The size-frequency distributions of the links within most complex natural networks do not exhibit a characteristic mean value, as perhaps first emphasized by Barabási and Albert [20]. Instead, they show what is referred to as scale-free behavior characterized by a monotonically decreasing, polynomial size–frequency probability function. Since that study a plethora of papers appeared developing the "mathematics" and "physics" of networks, both natural and model ones, and finding new occurrences of complex graphs following a scale-free behavior [26].

Viewing energy levels and the related transitions as entities forming a network raises questions like what kind of size-frequency distribution the links in an SN follow and whether this graph-theoretical view could offer new insight and significant consequences during investigations of SNs. In this study we demonstrate for the case of the rotational–vibrational states of the $HD^{16}O$ isotopologue of the water molecule that both the measured and the first-principles computed one-photon absorption SNs are scale free due to the randomness offered by realistic intensity cut-offs for the latter case. This observation has several important consequences for SNs, including appearance of hubs (hubs are those few nodes within a scale-free network which have a very large number of links), robustness and error tolerance (resulting from the large number of cycles within the SN), and "small-world" property (meaning that, on average, a small number of links provide the shortest routes to connect even the most distant nodes). Furthermore, the SN view of line lists offers new design principles and ideas for improving the accuracy and robustness of information systems containing high-resolution spectroscopic data.

2. Methodological details

The scale-free property of networks means that the probability that a randomly selected node has exactly k links is $P(k) \propto k^{-\gamma}$, where γ is called the scaling index [26]. The following dynamical features are the usually assumed requirements of a scale-free network: (a) evolutionary growth with more or less random generation of new nodes, (b) highly interactive self-organization, and (c) preferential connectivity of new nodes to old ones. There are additional signatures beyond power-law degree distribution which

characterize scale-free networks, an important one is the small-world property, these networks can be characterized by a small diameter. The diameter of a network corresponds to the average length of the shortest paths defined by the links connecting any two nodes of the network. Scale-free networks, as opposed to random (Erdős–Rényi) networks [27] are further characterized by: (a) a relatively few highly connected nodes (hubs) and (b) a robust connectivity structure hard to fragment by random removal of nodes.

The molecule $HD^{16}O$, hereafter called HDO, was chosen for this investigation as it was the subject of a large number of experimental high-resolution spectroscopic studies validated recently [13] via a robust inversion protocol called MARVEL [11] (standing for Measured Active Rotational-Vibrational Energy Levels), a high-quality first-principles line list, including energy levels, transitions, and one-photon absorption intensities, is available for it [18], and it contains a single SN as all of its nuclei are different. The experimental dataset of Ref. [13] contains 8 819 nodes and 54 740 links (of which 36 690 are unique), forming a multiedge graph. The size of a first-principles SN, both in the number of nodes and links, depends heavily on the chosen cut-off of the absorption intensities. All the results reported here for the computed SN correspond to a temperature of 300 K and most to an absorption intensity cut-off of 10^{-30} cm molecule^{-1}. With this cut-off, the SN contains 62 140 nodes and 738 869 links. With still realistic cut-offs of 10^{-26} and 10^{-28} cm molecule^{-1}, the first-principles SN contains 9 196(76 107) and 23 313(231 713) nodes(links), respectively. Note that the computed network of HDO employed here contains altogether 163 491 and 697 444 828 nodes and links, respectively. From the numbers presented one can conclude that it is easy to find a few relatively strong transitions for almost all nodes in the SN but the great majority of the transitions is extremely weak.

3. Results and discussion

Fig. 1 shows a small part of the purely rotational experimental SN of HDO, without the weights and up to $J = 4$, where J is the rotational quantum number [28]. The connectivity is defined here by well-known selection rules of quantum mechanics for one-photon absorption spectroscopy. These state for HDO that if $\Delta J = 0(\pm 1)$ than $p = \pm 1(0)$, where p is the parity describing the energy level.
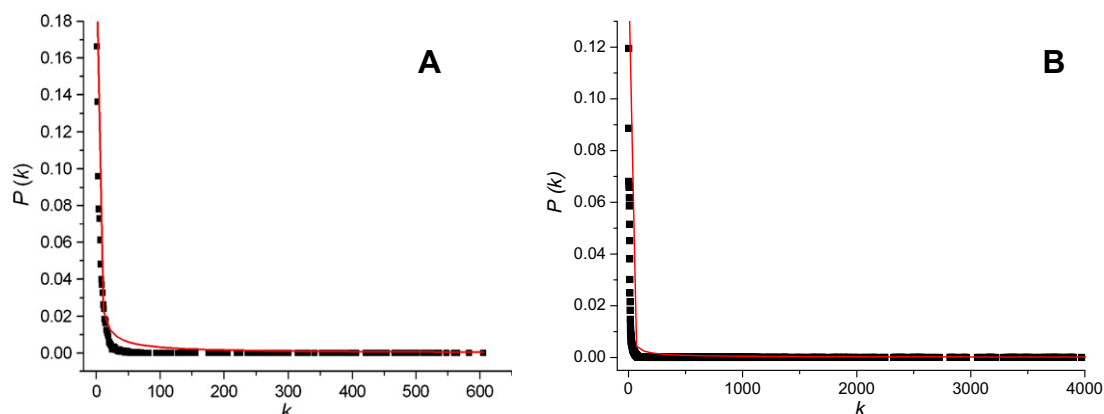
Fig. 2. Size–frequency [k–$P(k)$] plots for measured (Panel A) and first-principles (Panel B) transitions for HDO, with an absorption intensity cut-off of 10^{-30} cm molecule^{-1} in the latter case. The best polynomial fit, see text, to the data points is given by the continuous line.

These selection rules dictate that if J–K_c is even or odd then the maximum number of pure rotational transitions from a given rovibrational energy level with asymmetric-top label [28] $J_{K_a K_c}$ is $3J + 2$ or $3J + 1$. As clear from Fig. 1, where these maximum numbers as well as the measured transitions are indicated, measurements miss several purely rotational transitions even for low J values. As J increases, the number of transitions not measured increases extremely quickly. Note that while many of the transitions are not directly measured, their identification in experimental spectra becomes straightforward if inverted rovibrational energies obtained, for example, via the MARVEL protocol [11] are available and the transition intensity is high enough. Note also that since connectivity within quantum mechanical networks is determined by selection rules, different spectroscopic techniques built upon different physical principles, *e.g.*, absorption infrared vs. vibrational Raman spectroscopy, result in SNs with drastically different topologies.

Panel A of Fig. 2 depicts the *measured* size–frequency [k–$P(k)$] plot for HDO, based on a recent study [13] of its rotational-vibrational spectrum. Apart from the very low and very high k part, one can find a linear relationship on the corresponding log–log plot indicating a power-law behavior over several orders of magnitude with a scaling exponent, $\gamma = 0.93$, close to 1. This value is quite different from the usually obtained $\gamma = 2$–3 values for large, complex, scale-free networks [26]. The first-principles SN, with an absorption intensity cut-off of 10^{-30} cm molecule^{-1}, shows a very similar $P(k)$–k fall-off, of course with much larger k values, and a slightly smaller $\gamma = 0.87$. These graphs are clear indicators of the scale-free property of SNs.

Complex scale-free networks have usually been studied via principally *stochastic* models. Note in this respect that SNs are principally deterministic and stochasticity is introduced, whether experimentally or in computations, via realistic intensity cut-offs. Thus, SNs offer an interesting class of networks for future studies also from this respect.

Scale-free networks contain hubs, *i.e.*, nodes with extremely large number of connections. As expected, the most important hubs in an absorption spectrum are on the ground vibrational state. For the measured SN they are as follows: $J_{K_a K_c} = 4_{22}$, 4_{23}, and 3_{13}, with 605, 583, and 565 links, respectively. In the computed SN, with a cut-off value of 10^{-30} cm molecule^{-1}, the nodes with the largest number of connections are $6_{34}(4042)$, $7_{35}(3970)$, and $6_{24}(3897)$, where the number of links is given in parentheses.

The interconnectedness of a particular spectroscopic network can be described efficiently by a single number, the diameter, defined above. The diameter computed statistically for the measured SN of HDO is only about 7, a value slightly larger but similar to the diameter of the first-principles SN. As the absorption intensity cut-off is decreased, the diameter of the computed SN seems to stay around this value though becomes smaller. Thus, SNs clearly have an intrinsic small-world property, similarly to most other complex networks studied in nature, society, communication, and elsewhere.

Selection rules allow only a limited number of links between the nodes of the SN. As the spectroscopic network becomes larger, either via new measurements or by a decrease in the intensity cut-off, the number of links increases substantially but not the number of nodes. The number of cycles within the network also increases drastically. Thus, SNs appear to be extremely robust. After random removal of nodes the relative size of the largest remaining network compared to the full size of the remaining network remains very close to 1 even if up to 70% of the nodes are removed (see Fig. 3). The network only fragments when about 85% of the links are randomly removed. This extreme error tolerance is another characteristic property of SNs.

The graph-theoretical interpretation of spectroscopic line lists offers several new ideas and algorithms to improve the accuracy and robustness of experimental SNs. First, as we learn from the scale-free property of SNs, despite the fact that SNs can be very large there are only relatively few energy levels whose accuracy principally determines the accuracy of the full experimental SN. Therefore, improving the accuracy of hubs by decreasing the uncertainties of energy levels qualifying to be hubs via new and accurate measurements is vital. For example, all microwave (MW) and millimeterwave (MMW) measurements with this aim
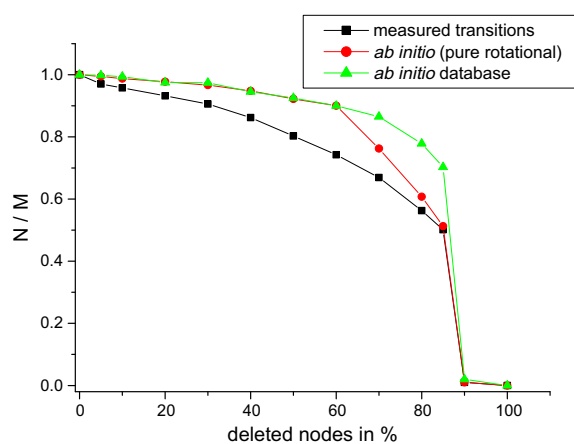

Fig. 3. Fragmentation dynamics of spectroscopic networks following random removal of nodes, where N/M means the number of nodes within the largest remaining network (N) compared to the maximum number of nodes (M) in the SN.
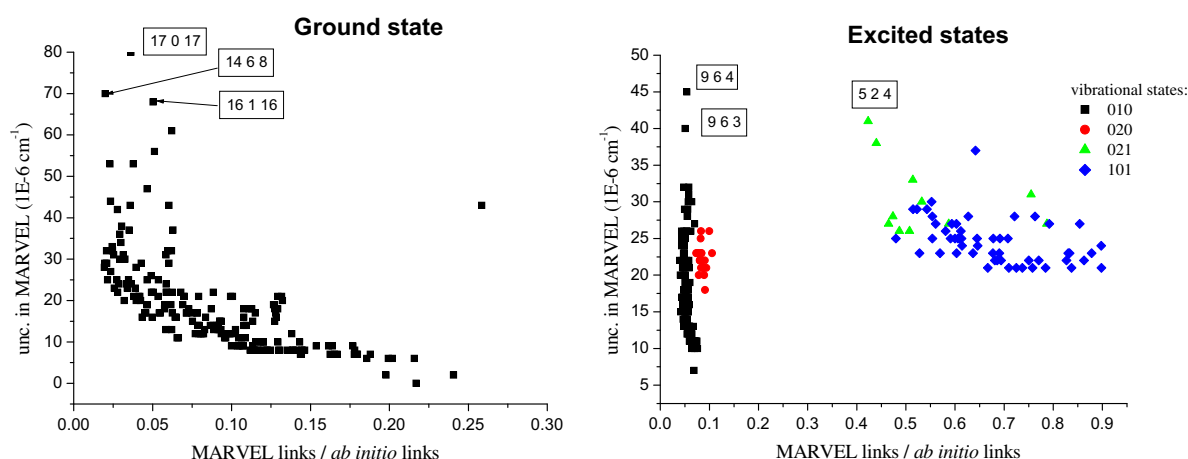
Fig. 4. The least well determined rotational–vibrational energy levels among the first 200 hubs for the ground vibrational state (Panel A) and among the first 200 hubs for the next four vibrational states (Panel B) of HDO determined by plotting the measured MARVEL uncertainties against the relative number of MARVEL links compared to the first-principles links with an intensity cut-off of 10^{-30} cm molecule^{-1}.

would be highly beneficial for improving the overall accuracy of the experimental SN. Second, spectroscopic measurements often result in disjoined subgraphs, whose vertices have no path to any of the roots of the experimental SN. Previously [11], we named these disjoint components of the SN floating spectroscopic networks (FSN) or orphans (if the component contains only two vertices). Unification of graph components helps to improve the robustness of the network and to allow attachment of correct energy values to the vertices of the disjoint components. Investigation of the minimum weight spanning forest should tell us how to connect FSNs and orphans to components containing one of the roots in the most simple and efficient way. Note that since each link has a unique weight (unique intensity), there will be a unique minimum weight spanning tree for each component. Third, comparison of the experimental and the computed SNs results in practical information about the least well determined hubs present in the experimental SN, and about design recommendations for future experiments. As shown in the next paragraph and Fig. 4, detailed comparison of measured and computed hubs helps to determine the "weakest nodes" within the experimental SN and to design new experiments which help, with a minimum amount of effort, the determination of a more accurate and more robust SN.

Comparison of the measured and first-principles (*ab initio*) SNs show that there are no energy levels which contain more than 200 links and have not been involved in measured transitions. Thus, for HDO all the important hubs are available (*cf.* Fig. 2B) and assignment of measured transitions can be considered to be rather complete in this sense, which probably does not hold for most other molecules. Then, determination of the perhaps most useful new experiments for improving the accuracy of the presently available experimental SN of HDO was done as follows. We selected the first 200 hubs of the ground vibrational state (GS) and the first 200 non-GS hubs. Then, for each of the 400 hubs we formed the ratio of the MARVEL and first-principles number of links characterizing these hubs (an intensity cut-off of 10^{-30} cm molecule^{-1} at 300 K was used for the first-principles list). Next, we plotted the MARVEL uncertainties of the energy levels, based on the experimental SN, against the ratio just obtained. These plots for the ground and four excited vibrational states are given in Fig. 4. On this figure one can identify the "weakest nodes" among the hubs selected. Next, we selected the first three least well determined nodes for both the ground and the excited vibrational states. It turns out that two of the "weakest nodes" of the excited states correspond to the first excited bending state (0 1 0). Then, we checked the first-principles

line list for those transitions which involve these energy levels but are not part of the experimental SN. These transitions should be the best candidates for new experiments designed to improve the experimental SN of HDO. All those transitions were left out for which the assignment in the two databases is different. All the rovibrational transitions selected start from the ground vibrational state. For the ground vibrational state, the selected predicted transitions, in cm^{-1}, are $\{17_{0,17} - 18_{1,18} = 234.4991(2)$, $14_{6,8} - 14_{7,7} = 182.55552(8)$, $15_{1,15} - 16_{1,16} = 209.72935(9)\}$ with corresponding absorption intensities of $\{1.2 \times 10^{-21}, 1.8 \times 10^{-22}, 7.6 \times 10^{-22}\}$ cm molecule^{-1}. For the (0 1 0) vibrational state, the two selected transitions are $\{9_{6,4} - 8_{6,3} = 1608.54197(5)$, $9_{6,3} - 8_{6,2} = 1608.54316(5)\}$ with corresponding absorption intensities of $\{4.4 \times 10^{-22}, 4.4 \times 10^{-22}\}$ cm molecule^{-1}. For the (0 2 1) vibrational state, the selected transition is $5_{2,4} - 4_{2,3} = 6539.05647(4)$ with a corresponding absorption intensity of 1.1×10^{-22} cm molecule^{-1}. These recommended transitions clearly identify regions of the absorption spectrum of HDO whose accurate measurement and thorough investigation would be most beneficial (beyond those already investigated).

The above arguments clearly suggest that a considerable number of new experimental MW and MMW measurements are required in order to further improve the quality of the experimental SN of HD^{16}O. It is likely that the same conclusion can be reached for most other molecules for which determination of a complete line list is an important goal. Unfortunately, there are less and less high-precision MW and MMW studies appearing in the literature for stable molecules but hopefully this can be remedied in the future given the pronounced need for such data.

4. Conclusions

It is suggested that to make maximum use of spectroscopic line list information available for a large number of molecules, whether obtained experimentally or first principles, weighted undirected graphs should be constructed from them which we call spectroscopic networks (SN). The vertices of the graphs are the energy levels, the edges are the spectroscopically allowed transitions, and the weights are the transition intensities. Experiments yield relatively small multiedge graphs, with a considerable number of parallel edges, while first-principles computations result in very large simple graphs. Of course, different spectroscopic techniques yield SNs with drastically different topologies. As demonstrated in this

study, many of the popular notions of interdisciplinary scientific, social, and communication network investigations, like scale-free and "small world" properties, hubs, network dynamics, self-organization, robustness, and attack/error tolerance, can be used to characterize SNs.

The network-theoretical view advocated here for complex SNs offers distinct advantages toward the complete characterization of high-resolution molecular spectra. The established scale-free property of the overall network structure leads to the concept of hubs and thus straightforwardly to the design of new spectroscopic experiments which, with a minimum amount of effort, help to determine a more accurate and more robust SN by preferentially measuring, with high accuracy, transitions in which hubs are involved. For example, detailed comparison of measured and first-principles hubs helps to determine the "weakest nodes" among these energy levels in an existing experimental SN which, in turn, leads to the identification of transitions which should preferentially be investigated in new experiments designed specifically for their determination.

Driven by the need of scientific and engineering applications, complex spectroscopic networks, perhaps as part of active databases [10–13], are expected to become an intrinsic part of the description of the high-resolution spectra of molecules.

Acknowledgments

References

[1] L.S. Rothman, J. Quant. Spectrosc. Rad. Transfer 111 (2010) 1565–1567.
[2] L.S. Rothman, I.E. Gordon, A. Barbe, D. Chris Benner, P.F. Bernath, M. Birk, V. Boudon, L.R. Brown, A. Campargue, J.-P. Champion, K. Chance, L.H. Coudert, V. Dana, V.M. Devi, S. Fally, J.-M. Flaud, R.R. Gamache, A. Goldman, D. Jacquemart, I. Kleiner, N. Lacome, W.J. Lafferty, J.-Y. Mandin, S.T. Massie, S.N. Mikhailenko, C.E. Miller, N. Moazzen-Ahmadi, O.V. Naumenko, A.V. Nikitin, J. Orphal, V.I. Perevalov, A. Perrin, A. Predoi-Cross, C.P. Rinsland, M. Rotger, M. Simeckova, M.A.H. Smith, K. Sung, S.A. Tashkun, J. Tennyson, R.A. Toth, A.C. Vandaele, J. Vander Auwera, J. Quant. Spectrosc. Rad. Transfer 110 (2009) 533–572.
[3] L.S. Rothman, I.E. Gordon, R.J. Barber, H. Dothe, R.R. Gamache, A. Goldman, V.I. Perevalov, S.A. Tashkun, J. Tennyson, J. Quant. Spectrosc. Rad. Transfer 111 (2010) 2139–2150.
[4] N. Jacquinet-Husson, N.A. Scott, A. Chédin, K. Garceran, R. Armante, A.A. Chursin, A. Barbe, M. Birk, L.R. Brown, C. Camy-Peyret, C. Claveau, C. Clerbaux, P.F. Coheur, V. Dana, L. Daumont, M.R. Debacker-Barilly, J.M. Flaud, A. Goldman, A. Hamdouni, M. Hess, A. Nikitin, D. Newnham, A. Perrin, V.I. Perevalov, L. Régalia-Jarlot, A. Rublev, F. Schreier, I. Schult, K.M. Smith, S.A. Tashkun, J.L. Teffo, R.A. Toth, Vl.G. Tyuterev, J. Vander Auwera, P. Varanasi, G. Wagner, J. Quant. Spectrosc. Rad. Transfer 95 (2005) 429–467.
[5] K.P. Dere, E. Landi, P.R. Young, G. Del Zanna, M. Landini, H.E. Mason, Astron. Astrophys. 498 (2009) 915–929.
[6] H.S.P. Müller, F. Schlöder, J. Stutzki, G. Winnewisser, J. Mol. Struct. 742 (2005) 215–227.
[7] H.S.P. Müller, S. Thorwirth, D.A. Roth, G. Winnewisser, Astron. Astrophys. 370 (2001) L49–L52.
[8] H.M. Pickett, R.L. Poynter, E.A. Cohen, M.L. Delitsky, J.C. Pearson, H.S.P. Müller, J. Quant. Spectrosc. Rad. Transfer 60 (1998) 883–890.
[9] S.A. Tashkun, V.I. Perevalov, J.-L. Teffo, A.D. Bykov, N.N. Lavrentieva, J. Quant. Spectrosc. Rad. Transfer 82 (2003) 165–196.
[10] A.G. Császár, G. Czakó, T. Furtenbacher, E. Mátyus, Annu. Rep. Comp. Chem. 3 (2007) 155–176.
[11] T. Furtenbacher, A.G. Császár, J. Tennyson, J. Mol. Spectrosc. 245 (2007) 115–125.
[12] J. Tennyson, P.F. Bernath, L.R. Brown, A. Campargue, M.R. Carleer, A.G. Császár, R.R. Gamache, J.T. Hodges, A. Jenouvrier, O.V. Naumenko, O.L. Polyansky, L.S. Rothman, R.A. Toth, A.C. Vandaele, N.F. Zobov, L. Daumont, A.Z. Fazliev, T. Furtenbacher, I.F. Gordon, S.N. Mikhailenko, S.V. Shirin, J. Quant. Spectrosc. Rad. Transfer 110 (2009) 573–596.
[13] J. Tennyson, P.F. Bernath, L.R. Brown, A. Campargue, M.R. Carleer, A.G. Császár, L. Daumont, R.R. Gamache, J.T. Hodges, A. Jenouvrier, O.V. Naumenko, O.L. Polyansky, L.S. Rothman, R.A. Toth, A.C. Vandaele, N.F. Zobov, A.Z. Fazliev, T. Furtenbacher, I.F. Gordon, S.-M. Hu, S.N. Mikhailenko, B. Voronin, J. Quant. Spectrosc. Rad. Transfer 111 (2010) 2160–2184.
[14] O.L. Polyansky, A.G. Császár, S.V. Shirin, N.F. Zobov, P. Barletta, J. Tennyson, D.W. Schwenke, P.J. Knowles, Science 299 (2003) 539–542.
[15] R.J. Barber, J. Tennyson, G.J. Harris, R.N. Tolchenov, Mon. Not. Roy. Astron. Soc. 368 (2006) 1087–1094.
[16] T. Sochi, J. Tennyson, Mon. Not. Roy. Astron. Soc. 405 (2010) 2345–2350.
[17] S.V. Shirin, N.F. Zobov, R.I. Ovsyannikov, O.L. Polyansky, J. Tennyson, J. Chem. Phys. 128 (2008) 224306.
[18] B.A. Voronin, J. Tennyson, R.N. Tolchenov, A.A. Lugovskoy, S.N. Yurchenko, Mon. Not. Roy. Astron. Soc. 402 (2010) 492–496.
[19] S.N. Yurchenko, R.J. Barber, A. Yachmenev, W. Thiel, P. Jensen, J. Tennyson, J. Phys. Chem. A 113 (2009) 11845–11855.
[20] A.-L. Barabási, R. Albert, Science 286 (1988) 509–512.
[21] R. Albert, H. Jeong, A.-L. Barabási, Nature 406 (2000) 378–382.
[22] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, Nature 407 (2000) 651–654.
[23] H. Jeong, S. Mason, A.-L. Barabási, Z.N. Oltvai, Nature 411 (2001) 41–42.
[24] R. Albert, H. Jeong, A.-L. Barabási, Nature 401 (1999) 130–131.
[25] A.-L. Barabási, E. Ravasz, T. Vicsek, Physica A 299 (2001) 559–564.
[26] R. Albert, A.-L. Barabási, Rev. Mod. Phys. 74 (2002) 47–97.
[27] P. Erdős, A. Rényi, Publ. Math. 6 (1959) 290–297.
[28] H.W. Kroto, Molecular Rotation Spectra, Wiley, New York, 1975.