

# Tautomerism and Expert Systems in Spectroscopy

## Part II: Assigning the Correct Tautomers Using NMR Spectral Predictors and IR Structure–Spectrum Verification Tools

**Because tautomers are closely related structural isomers, it can be difficult to identify distinguishing traits. Specialized software can be used to accelerate the assignment and validation of proposed tautomers.**

**Michel R.J. Hachey**

In Part I of this series, we saw how different tautomeric assignments can cause problems with respect to spectral databases, and how tautomeric-enabled search algorithms can help (1). The focus here will shift from searching databases for possible duplicates or related tautomeric forms to using software to accelerate the assignment and structure–spectrum validation for proposed tautomers.

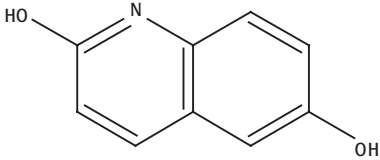
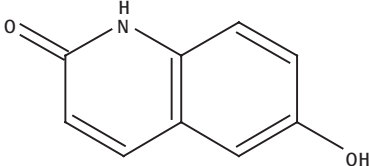
It takes extra effort to ensure that tautomeric forms are assigned accurately. Because tautomers are closely related structural isomers, it can be difficult to identify distinguishing traits. In view of the rather liberal manner in which tautomers often are treated, a brief outline of the policies advocated for spectroscopic databases follows.

To avoid scientific confusion, it is general practice to associate only the predominant tautomeric forms observed at equilibrium with a spectrum. For example, if the spectra of a compound with possible keto–enol tautomerism supports an assignment to the enol form, but gives no characteristic carbonyl bands in the infrared (IR) and lacks the expected C=O carbon chemical shift in a  $^{13}\text{C}$  nuclear magnetic resonance (NMR) spectrum, then it would make little sense attributing these spectra to the keto form. This might appear to be obvious, but some structure databases and libraries of spectra depart from accurate tautomeric representation in favor of simplified and expedient rules for dealing with tautomers. For example, some might choose to standardize all keto–enol pairs to the keto representation regardless of which form actually is observed. Others even go as far as accepting any form as a representative of the whole set of possible tautomers. As long as such conventions are stated clearly at the onset, this is acceptable. However, scientists populating libraries of spectra generally strive to assign the predominant form to the spectrum to which they belong.



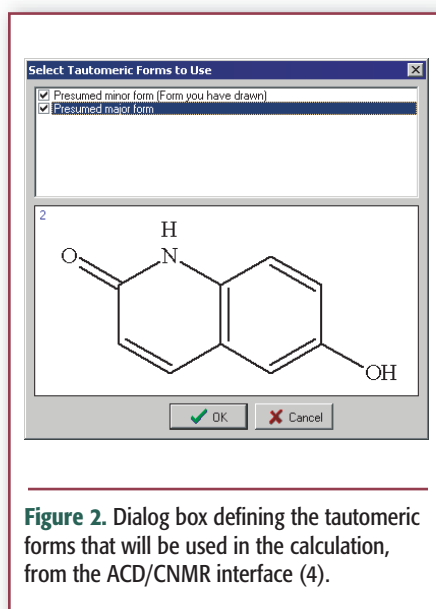
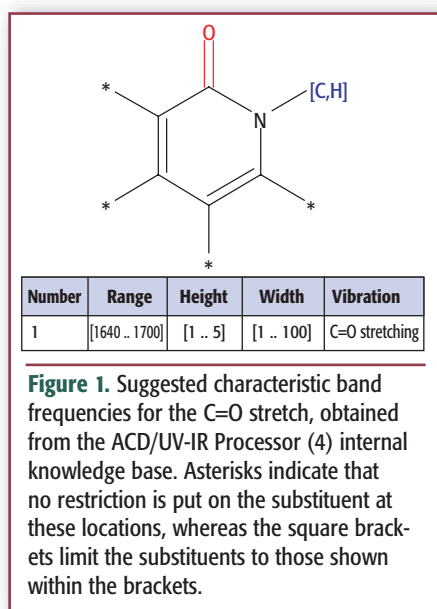
PHOTO COURTESY OF DIGITAL VISION

Let's not forget that tautomers actually are mixtures of different isomeric components. A significant number of different forms might coexist at equilibrium, which complicates the spectral interpretation. While some tautomeric mixtures can be separated by chromatographic techniques (2), this is not universally applicable. And even if possible,

Tautomers	<sup>13</sup> C NMR Match Factor*	<sup>1</sup> H NMR Match Factor*
Quinoline-2,6-diol 	0.65	0.93
6-Hydroxyquinolin-2(1H)-one 	0.88	1.00

\* The match factor was evaluated automatically using the verification command in ACD/NMR Manager (4), and is based upon predictions with ACD/HNMR and ACD/CNMR predictors version 7.07 (4).

**Table I.** Predicted match factor between an experimental spectrum and two possible tautomeric forms.



the tautomeric composition within the column might not be representative of that in the original sample. In short, the existence of conditions, normal or otherwise, that permit the analyst to isolate one form or another cannot be taken for granted. Ideally, each spectrum in a spectral library should be permitted to have more than one chemical structure attached and should contain some indication about the components' ratio or preferred forms. In practice, however, it is not uncommon for the chemical representation to be restricted to the one component in

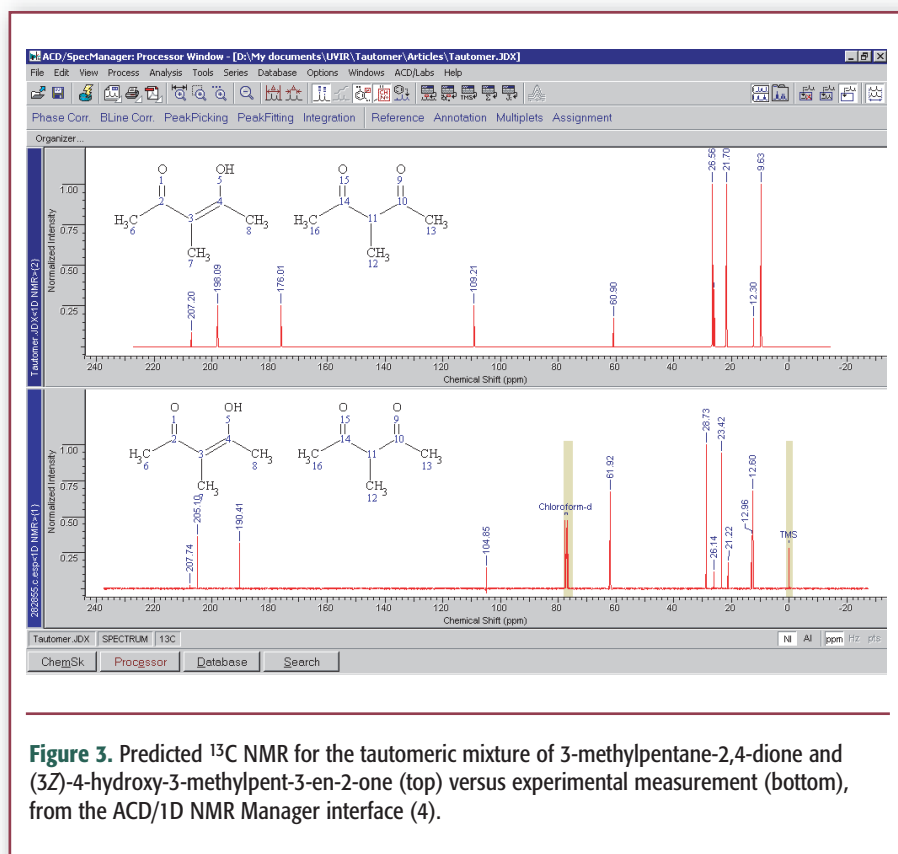
the mixture with the greatest spectroscopic impact.

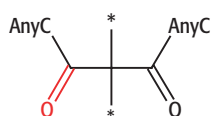
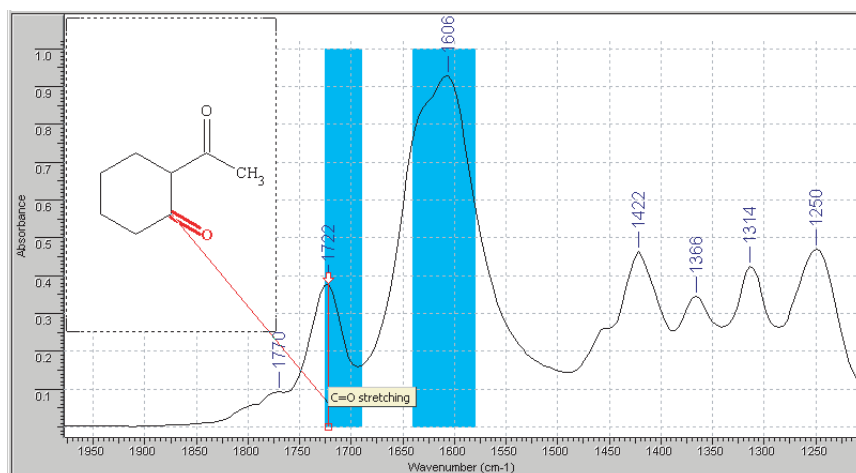
Spectroscopic expert systems are valuable in assigning tautomer mixtures correctly and rapidly. Expert systems are defined as computer programs that aid scientists in making better decisions about the dominant mixture forms through the application of a knowledge-based or rule-based approach. Two tautomerically enabled expert systems for interpreting tautomeric data will be discussed here — NMR predictors and IR structure-spectrum verification tools.

## Tautomerically Enabled NMR Predictors

NMR is a very powerful technique for detecting tautomers and identifying their forms at equilibrium. But even experts can suffer from lapses in attention, as shown by the case of quinoline-2,6-diol found in the Aldrich/ACD FT-NMR spectral library (3). This assignment most likely is wrong according to NMR spectral predictions. Table I shows match factors between the experimental (3) and predicted  $^{13}\text{C}$  and  $^1\text{H}$  NMR spectra for both forms. The match factors were evaluated automatically using the verification command in ACD/1D NMR Manager (4) and are based upon predictions with ACD/HNMR and ACD/CNMR predictors version 7.07 (4). The matches overwhelmingly favor the reassignment to the keto form, 6-hydroxyquinolin-2(1H)-one, instead of the enolic diol form (3).

Under some conditions, tautomeric equilibria are quite rapid relative to the NMR time scale, so that the resulting NMR spectrum might be indistinct with





Number	Range	Height	Width	Vibration	Comment	Result
1	[1580 .. 1640]	[1 .. 5]	[1 .. 100]	C=O stretching	Enol form	Passed
2	[1690 .. 1725]	[1 .. 5]	[1 .. 100]	C=O stretching	Corrected; keto form	Passed

**Figure 4.** Autoassignment of 2-acetylcyclohexanone showing the C=O stretch regions corresponding to the keto and enol tautomers of the 1,3-diketone substructure, (ACD/NIST IR database [6]). Screen captures from ACD/UV-IR Manager (4) software.

shifts between one form and the other, or with both species present. Complementary confirmation with an IR interpretation therefore is advised. To determine if the reassignment to 6-hydroxyquinolin-2(1*H*)-one is correct, the corresponding condensed-phase IR spectrum was reanalyzed. For the keto form, the IR spectrum should show a C=O stretch frequency between 1640 and 1700  $\text{cm}^{-1}$  for the 2-pyridone substructure as shown in Figure 1. Indeed, a fairly intense and sharp band is observed experimentally at 1660  $\text{cm}^{-1}$  (5), which supports the proposed reassignment based upon the NMR match factors. In this case, the original error likely occurred because of a legacy assignment.

Tautomeric-specific tools were built directly into some NMR predictor modules (4). For example, users trying to run a prediction on a minor form or condition-dependent tautomeric form will receive an automatic prompt to select a tautomeric form. This automatic prompting could be salutary if the potential tautomerism had been missed, such as in the case of quinoline-2,6-diol.

Figure 2 shows the tautomer selection dialog box currently highlighting 6-hydroxyquinolin-2(1*H*)-one, which is predicted correctly as being the major form by the algorithm, in agreement with the spectral analyses. Note that quinoline-2,6-diol was used as the input structure, but the ACD/Tautomer algorithm (1) correctly suggests the preferred form. The same tautomer selection options used in the database-searching module are available in the NMR predictors.

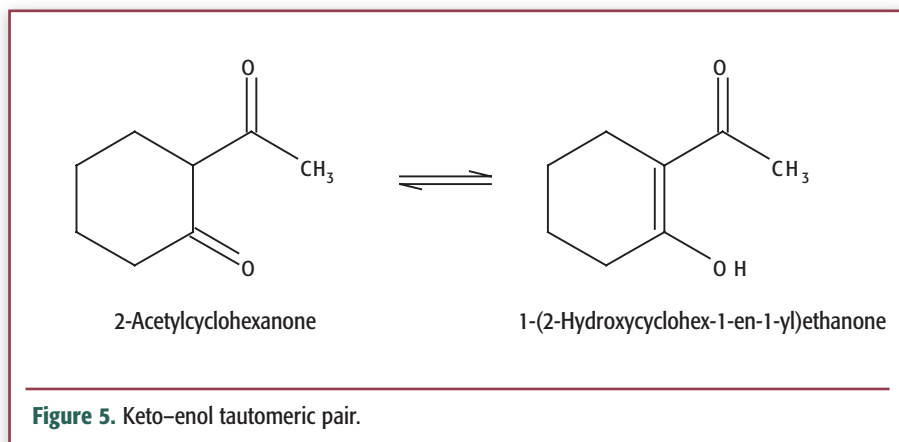
Significantly, selecting “All Forms” or “Most Stable Forms” will cause the NMR predictors to calculate the spectrum of the corresponding tautomeric mixture. Taking another example from the Aldrich/ACD Library of FT-NMR spectra, the tautomeric equilibrium between 3-methylpentane-2,4-dione and (3*Z*)-4-hydroxy-3-methylpent-3-en-2-one is reported to be approximately 85% and 15%, respectively (3). Drawing the 3-methylpentane-2,4-dione form and clicking “Calculate” causes a tautomeric form selector dialog box to appear. Selecting “Most Major

Forms” now yields a predicted NMR spectrum for the mixture of both forms as shown in Figure 3. This predicted spectrum is consistent with the experimentally observed spectrum after adjusting the composition ratio to match the experimental one.

Next, we will describe an IR verification routine that could have helped check the respective structure–spectrum correspondence semiautomatically among different possible tautomeric forms.

### Tautomerically Enabled IR Structure–Spectrum Verification

The discussion so far has emphasized that tautomers depend upon two things: the structural environment of their functional groups and the overall physicochemical environment of the molecule. Interestingly, characteristic IR bands are subject to these same factors. Therefore, many charts and books on IR characteristic frequencies contain information about chemical groups that relate to the tautomeric space. Until



recently, software tools had not been built with the ability to describe the structural environment around vibrational groups in sufficient detail to be very useful for tautomerism studies.

Advanced fragment objects were developed that allow the representation of groups of atoms responsible for the vibration within a narrowly defined chemical environment via a descriptive and flexible chemical-environment shell (6). We saw previously an example of

such an object in Figure 1, where the 2-pyridone substructure provided a typical C=O stretching frequency region between 1640 and 1700  $\text{cm}^{-1}$ . These sophisticated structure representations can describe almost any structural environment, including those that distinguish tautomers. An IR knowledge base of more than 700 highly characteristic IR fragments, including those of some tautomers, was built that narrowly defines correlating frequency regions (4,



6). This knowledge base was used to provide an automatic IR verification expert tool as described next.

When the “Verify” command was applied to the IR spectrum of anthralin in Part I of this article series (1), a peak consistent with an aromatic carbonyl was identified and a pass value was returned, confirming the postulated keto–diol assignment. By attaching the proposed chemical structure to the spectrum and then invoking an IR analysis mode, the program is designed to find by itself all relevant vibrational groups, including possible tautomeric fragments, present in the IR knowledge base (6). Further, the software verifies the suggested characteristic band frequencies against the labeled peaks in the spectrum, thus providing a pass or fail grade that can be reviewed critically by analysts.

Figure 4 shows an IR spectrum assigned to 2-acetylcyclohexanone (7). An interactive lasso tool was used to select a carbonyl group in the attached structure to highlight the characteristic vibrational regions expected in this structural environment (blue bars). Note that two distinct regions for C=O stretches are suggested because the program detected a 1,3-diketone substructure, which is expected to lead to two distinct enol and keto tautomeric forms as shown in Figure 5. The band at  $1722\text{ cm}^{-1}$  is consistent with C=O stretch frequencies for carbonyl groups with saturated carbon substituents in the keto form and should be assigned to 2-acetylcyclohexanone. In the same way, the band at  $1606\text{ cm}^{-1}$  is consistent with the C=O stretch in the enolic tautomer, 1-(2-hydroxycyclohex-1-en-1-yl)ethanone. The corresponding verification result table is given at the bottom of Figure 4 and shows “pass” results for both. These two carbonyl stretch frequencies are consistent with the expectation that vinylic ketones occur at lower wavenumbers than in a saturated environment.

In the original assignment in the ACD/NIST IR database, the gas phase IR spectrum was attributed to the keto form, 2-acetylcyclohexanone; but note that the enolic carbonyl band is considerably more intense. A similar situation is observed for the neat solution (5). Is

this yet another misassignment in the literature? The corresponding NMR experimental spectrum was found (3), and the predicted match factors were obtained. Match factors of 0.21 ( $^{13}\text{C}$  NMR) and 0.57 ( $^1\text{H}$  NMR) were found for the keto form, while the enol form provided matches of 0.90 ( $^{13}\text{C}$  NMR) and 0.86 ( $^1\text{H}$  NMR), confirming our misassignment suspicion.

## Discussion and Summary

It is noteworthy that the spectral databases from which the tautomeric misassignments were extracted are some of the tautomerically better-reviewed databases. Other databases probably will fare worse or no better. The problems identified came from legacy assignments, and predate the tools discussed herein. We saw that NMR and IR expert systems offer useful insights for assigning and validating the structure–spectrum relationship of tautomers.

It is important to understand that current tautomeric expert systems generally do not make final decisions with respect to tautomers, but rather they are designed to highlight the tautomeric species that the scientist should consider before making a final decision. The final onus always is on the scientist.

## References

1. M.R.J. Hachey, *Spectroscopy* **19**(4), 14–19 (2004).
2. P.J. Skrdla, V. Antonucci, and C. Lindemann, *J. Chrom. Science* **39**, 431–440 (2001).
3. Aldrich/ACD Library of FT-NMR Spectra and reference therein.
4. Advanced Chemistry Development, Inc., (ACD/Labs), ver. 7.0, Toronto ON, Canada, [www.acdlabs.com](http://www.acdlabs.com) (accessed Oct. 2003).
5. C.J. Pouchert, *The Aldrich Library of FT-IR Spectra*, 2nd ed., (Sigma-Aldrich Co., 1997) Vol. **1**, p.703; Vol. **3**, p. 3913.
6. ACD/UV-IR Manager IR Verification tool and Knowledge Base, ver. 7.0, Toronto ON, Canada, 2003.
7. ACD/NIST IR Database: structure searchable form of the NIST/EPA Gas Phase Infrared Database using data licensed from the National Institute of Standard and Technology, Gaithersburg, MD (1998). ■