Automatic Spectra Interpretation, Structure Generation, and Ranking

Patrick Fontana and Ernö Pretsch*

Laboratory for Organic Chemistry, Swiss Federal Institute of Technology, ETH Hönggerberg, CH-8093 Zürich, Switzerland

Received October 19, 2001

A new rule-based spectra interpretation system is described. It directly processes spectral files generated by spectrometers. Its internal information representation and the simultaneous use of several spectroscopic methods allow for the achievement of both high reliability and good performance. For small and medium sized molecules, the new program is capable of automatically reducing the solution space to under 3%.

INTRODUCTION

Research into the automatic generation of all isomers of an organic compound on the basis of its molecular formula and various pieces of structure information began in the late 1960s in several pioneering groups.[1−3] In the meantime, the field has matured and several structure generating programs have become commercially available[4−7] (for reviews, see refs 8 and 9). In contrast to the generation of isomers, automatic spectra interpretation, i.e., the derivation of structural information from spectra, is still problematic. Since it largely relies on empirical knowledge, by principle no rule-based automatic spectra interpretation of an unknown compound can be fully dependable. This also holds in the case where the interpretation system directly uses a spectroscopic database[10−12] instead of explicit interpretation rules.

Recently, we described a new interpretation program, SpecInt,[13] which achieved a high reliability by (1) simultaneously making use of different spectra, (2) restricting the predictions to very conservative ones, and (3) applying several internal consistency checks. While the program is able to derive substructures that are present in, or absent from, the target compound (good-list or bad-list, respectively), it was found that especially the absence of small structural fragments is highly efficient in reducing the number of possible isomers.[14] Despite using only a rather few highly reliable rules, for molecular masses less than ca. 250, SpecInt was capable of reducing the number of isomers to less than 3% of the possible ones defined by the molecular formula alone.[13] However, the program cannot be used in a fully automatic operation mode since the input of spectral information and the transfer of its results to a structure generator have to be done manually. In contrast, SpecInt 2, the new version presented in this contribution, can directly read spectra in formats generated by the spectrometer and its output is fully compatible with the commercially available structure generator Assemble 2.1.[5,15] Moreover, its rule base has been extended, and an editor is included that allows the user to add new interpretation rules.

* Corresponding author phone: +411 632 2926; e-mail: pretsch@org.chem.ethz.ch.

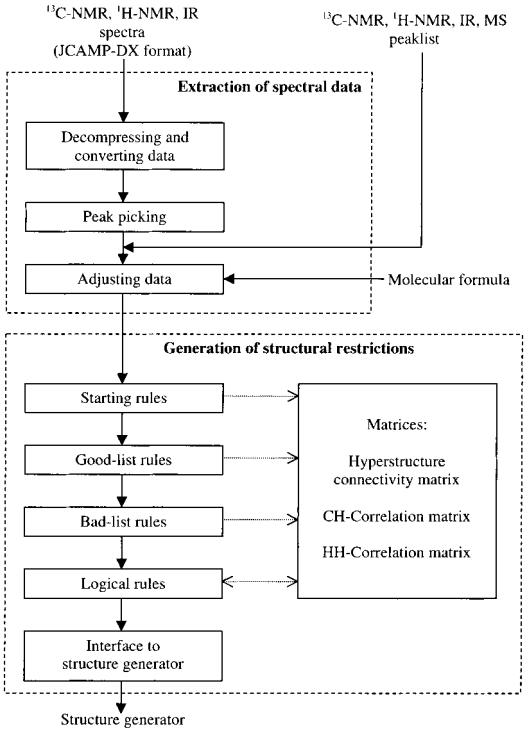

Figure 1. Overview of the program structure of SpecInt 2.

DESCRIPTION OF SPECINT 2

SpecInt 2, the new program for the automatic spectra interpretation of organic compounds, consists of two major parts: (1) modules that extract relevant information from the raw spectral data and (2) modules that generate structure restrictions from the data obtained in step 1 and from the molecular formula (see Figure 1).

Extraction of Relevant Spectral Information: Handling of Raw Data. SpecInt 2 can process various JCAMP-DX spectra file formats (Versions 4.24 upward)[16−18] or peak lists, including data compressed by algorithms such as SQZ or DIFDUP.[16] If the solvent used is not specified, the program sends a warning, and the user has to check the generated peak list for possible solvent signals.

10.1021/ci0101096 CCC: $22.00
Published on Web 03/22/2002

After decompressing the spectral data, SpecInt 2 performs a peak picking on each full spectrum with algorithms that are optimized for each method. This part is skipped when the spectral information is already stored in a JCAMP-DX peak list (e.g., mass spectra) or the peak list has been entered manually.

For ^{13}C NMR spectra with a sufficiently good signal-to-noise ratio, peak picking is straightforward. If the solvent is specified, SpecInt 2 automatically eliminates the pertinent signals provided that their positions figure in the corresponding file. Since exact peak positions vary with every sample, the chemical shifts of solvents are given with a certain tolerance (typically, $\pm$ 0.1 ppm). All solvent data are stored in a file that can be edited by the user. The multiplicities for each carbon atom are derived directly from the DEPT90 and DEPT135 spectra, if available. Otherwise, the user is prompted to enter them manually. Finally, SpecInt 2 checks if the number of carbon and hydrogen atoms found is consistent with the molecular formula, which must be entered manually since the current version of SpecInt 2 does not include an appropriate routine.[19]

The handling of ^{1}H NMR spectra is somewhat more complex since small signals may appear as parts of multiplets, with amplitudes often not significantly higher than those of spikes. Therefore, a two-step process is used. First, the ranges of the spectrum are determined where signals occur. This is done after smoothing the spectrum with a simple moving-average filter algorithm ($y_n = (y_{n-1} + y_{n+1} + 2 \times y_n)/4$). Depending on the amplitude of the noise (determined on a marginal segment of the spectrum), SpecInt 2 may repeat this smoothing procedure several times. Subsequently, the signals are classified in different groups separated by at least 25 Hz from each other so as to avoid that parts of one and the same multiplet be assigned to different groups. From the original spectrum, SpecInt 2 now executes a separate peak picking on each group of signals. In the next step, each group is searched for simple coupling patterns such as first-order multiplets of up to seven lines, doublets of doublet, and AA′BB′ systems. Since signals of different protons may occur within a group, special precaution is needed to avoid that they are erroneously assigned to the same multiplet. To achieve a high reliability, doublets or triplets are only assigned if a coupling partner can be identified in another group that has the same coupling constant within a threshold of 0.1 Hz. If the solvent signals are present in a group of their own, SpecInt 2 eliminates them, otherwise the user has to remove them manually. The program then determines the integrated intensities of each signal group and assigns the corresponding number of protons using the information on the total number of hydrogen atoms in the molecular formula.

Since IR spectra show strongly varying bandwidths, a two-step peak picking process is performed on them. First, SpecInt 2 identifies all sharp bands by straightforward peak picking (see Figure 2, top). Then, it transforms the spectrum by connecting the minima that are within 500 cm^{-1} of each other to remove all sharp bands. A second peak picking follows, in which the originally hidden broad bands are located (see Figure 2, bottom). Occasionally, both algorithms used in this process may find the same peaks; therefore, in a further step, SpecInt 2 eliminates duplicate signals from the combined peak list. Since broad bands are relevant at higher frequencies only, this procedure is applied to the region above 2000 cm^{-1}.

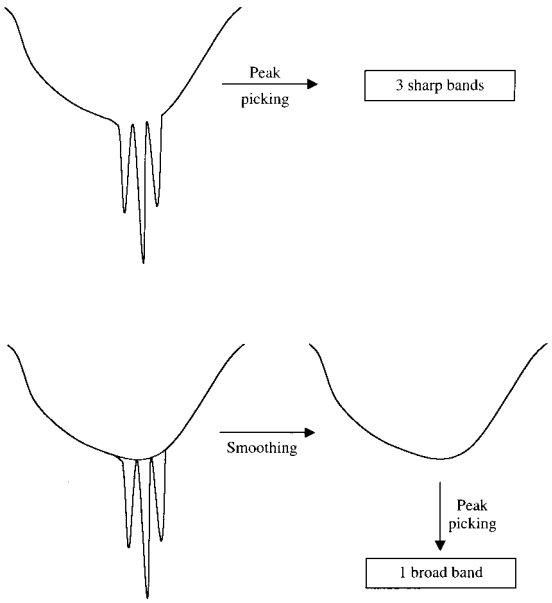

Figure 2. Effect of peak picking algorithms for regular (top) and broad (bottom) IR bands.

Table 1. Three-Digit Representation for Connectivities between Two Basic Fragments, A and B[a]

presence of bond	bond type triple	double	single
no	0	0	0
possible	1	1	1
yes	2	2	2

[a] Example: 011 signifies the absence of a triple bond, while a single and a double bond are possible.

Generating Structural Restrictions. The second part of SpecInt 2 is a revised and extended version of SpecInt.[13] Since the basic structure of the program remains the same, only a brief summary is given here. The knowledge base of SpecInt 2 consists of a set of interpretation rules relating spectral information with the presence (good-list) or absence (bad-list) of a given substructure. To achieve the highest possible reliability, SpecInt 2 does not use any probabilistic rules. The program first derives the possible basic fragments, which contain the non-hydrogen atoms together with the hydrogen atoms directly attached to them. All rules, i.e., the starting rules, the substructure-related rules, and the logical rules, operate on three matrices: the hyperstructure connectivity matrix defining the connectivities between basic fragments,[20] the CH-correlation matrix, and the HH-correlation matrix. For a future version of SpecInt, it is planned to directly enter 2D NMR information into these correlation matrices. The connectivities are represented by a three-digit code describing the bonding state (no, possible, or yes) for the three bond types (triple, double, or single bond) as shown in Table 1. No aromatic bonds are used since Assemble 2 identifies aromatic rings by keeping track of different resonance structures.[5] While in the first version, the spectral information had to be entered by the user, SpecInt 2 directly reads the data generated from the analysis of spectra.

The *Starting Rules* set up minimal and maximal limits for the number of all possible basic fragments by using the information from spectra and molecular formula. The revised

Scheme 1. Starting Rule for Basic Fragment NH_2

If ((nhm - nhc) / 2) < nnm **Then** NH_2^{max} = (nhm - nhc) / 2
If ((nhm - nhc) / 2) ≥ nnm **Then** NH_2^{max} = nnm
If (mxh - (nhm - nhc)) < nnm **Then** NH_2^{min} = nnm - (mxh - (nhm - nhc))
If (mxh - (nhm - nhc)) ≥ nnm **Then** NH_2^{min} = 0

NH_2^{max}: maximal number of NH_2 groups in the molecule
NH_2^{min}: minimal number of NH_2 groups in the molecule
nhm: number of H atoms according to molecular formula
nhc: number of H atoms from ^{13}C-NMR spectra
nnm: number of N atoms according to molecular formula
mxh: maximal imaginary number of H atoms (if all heteroatoms and all isochronous C atoms carry (valence - 1) H atoms, e.g., for N with valence 3: only NH_2)

rules of SpecInt 2 additionally derive constraints for the number of individual basic fragments. As an example, the starting rule for the basic fragment NH_2 is given in Scheme 1. This rule derives the minimal and maximal number of NH_2 groups from the total number of N- and H-atoms and the number of H-atoms bound to carbon atoms. Although these constraints would implicitly be found also by a structure generator, to derive them at this stage has the advantage of reducing the complexity of the matrices. As a consequence, it becomes more likely to find new good-list and bad-list entries during the interpretation process.

Good-List and Bad-List Rules. SpecInt 2 uses the same rules (totalling 37) for predicting the presence of fragments as SpecInt. However, the rules (a total of 49) predicting the absence of fragments have been redesigned and enhanced. Now, all rules are accommodated in a separate file that can be checked, edited, or extended by the user.

Logical Rules. After executing all good-list and bad-list rules, SpecInt 2 applies a set of logical rules which mainly draw conclusions from the matrices updated by the preceding good-list and bad-list rules. In the majority of cases, the hyperstructure connectivity matrix is considered in the first place because it is modified most by the good-list and bad-list rules. Due to its three-digit representation of bonds (see Table 1), it is very well suited for applying the logical rules. Since these rules can draw new conclusions after each change in the hyperstructure connectivity matrix, they are executed in a recursive way as long as such changes occur. Besides drawing conclusions, these rules perform consistency checks on the three matrices to achieve the highest possible reliability.

SpecInt 2 contains mainly two kinds of logical rules: self-related and neighbor-related rules. The former are simple ones trying to constrain the bonds of each basic fragment with the help of its free valences and the remaining double bond equivalents as determined by the molecular formula and the identified good-list fragments. The latter rules are more complicated because they also consider all possible neighbors of the basic fragments of the hyperstructure matrix. As an example, one such rule for a basic fragment, A, having three free valences and two possible neighbors, B and C, is explained in Table 2. The initial bonds between A and B or C (first two columns) lead to conclusions shown in columns 3 and 4. For example, A having three free valences cannot form a double bond to B if a single bond to C is forbidden (first row) and *vice versa* (second row). A triple bond must be present between A and B, if a single bond between them is forbidden and only a double bond between A and C is allowed (line 3) or a double bond A═B is forbidden and only a single bond A−C is allowed (line 4). The same conclusion can be drawn if single and double bonds between A and B and a triple bond between A and C are forbidden (line 5). The remaining examples of Table 2 show how

Table 2. Example of a Logical Rule for a Basic Fragment, A, with Three Free Valencies and Two Possible Neighbors, B and C[a]

initial connectivity between		resulting connectivity between		
A and B	A and C	A and B	A and C	remarks
xxx	xx0	x0x	xx0	no A═B and A−C
x0x	xxx	x0x	xx0	no A═B and A−C
110	010	200	000	A≡B
101	001	200	000	A≡B
100	0xx	200	000	A≡B
011	001	020	002	A═B and A−C
010	0x1	020	002	A═B and A−C
xxx	xx0	x0x	xx0	no A═B and A−C
x0x	xxx	x0x	xx0	no A═B and A−C

[a] In the three-digit code, x stands for either 0 or 1.

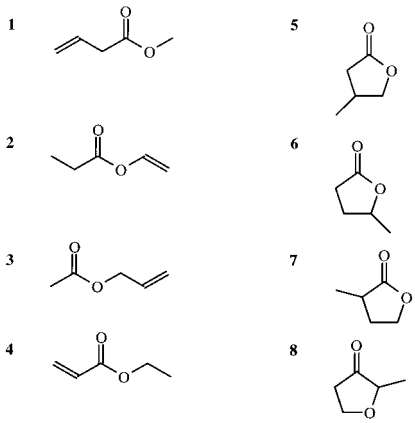

Figure 3. Structures of eight isomers of $C_5H_8O_2$ processed by SpecInt 2.

further combinations of possible and forbidden bonds can be used to constrain the number of possibilities. SpecInt 2 contains similar sets of rules for fragments having one, two, or four free valences.

Interface to the Structure Generator. The results of SpecInt 2 consist of constraints and of good-list and bad-list fragments, which figure as structural restrictions for the target structure. They are compiled in a separate ASCII file directly readable by the structure generator Assemble 2.1. From there, the results can be viewed and, if necessary, edited in a graphical environment. Subsequently, all possible structures are generated by Assemble 2.1 and ranked on the basis of the estimated spectra.[5] The output of SpecInt 2 in ASCII format also allows the results to be used otherwise.

RESULTS AND DISCUSSION

SpecInt 2, the new spectra interpretation program is capable of generating structural restrictions solely on the basis of ^{13}C NMR spectra. However, the combination of information from different spectra leads to more comprehensive results and higher reliability. The performance of SpecInt 2, therefore, has been investigated on more than 100 structures by simultaneously using ^{13}C NMR, ^{1}H NMR, and IR spectra. The correct structure was always found among the solutions calculated applying the automatically generated restrictions, i.e., SpecInt 2 did not fail in any of the cases. Here, we describe tests with eight constitutional isomers (see Figure 3) having the molecular formula $C_5H_8O_2$ and the same number of C, CH, CH_2, and CH_3 groups. Their ^{13}C NMR

Table 3. Results Obtained with SpecInt 2 for the Eight Isomers in Figure 3

no.	substructures generated by SpecInt 2[a] good-list	bad-list	remaining structures; reduction factor in %	ranking position of the correct structure (details see Table 4) ^{13}C NMR	^{1}H NMR
1	CH_3-O	C≡C, CH_3-C, CH═O, *CH_2⟨H22⟩ -O⟨H00⟩*, *CH_3-C⟨H00⟩*	32; 97.3	1	1
2	CH_3-CH_2	C≡C, CH═O, CH_3-O, CH_3-CH_2-O, *CH_3-CH⟨H11⟩*, *CH_3-C⟨H00⟩*	26; 97.8	1	1
3	CH_3-C	C≡C, CH_3-CH, CH═O, CH_3-O	31; 97.3	1	1
4	CH_3-CH_2	C≡C, CH═O, CH_3-O, *CH_3-CH⟨H11⟩*, *CH_3-C⟨H00⟩*	37; 96.8	1	1
5	CH_3-CH	C≡C, CH═CH, CH═O, CH_3-O, *CH_3-C⟨H00⟩*, *CH⟨H11⟩ =C⟨H00⟩*, *CH⟨H11⟩ -O⟨H00⟩*	55; 95.3	1	3
6	CH_3-CH	C≡C, CH_2═C, CH═CH, CH═O, CH_2-O, *CH_3-C⟨H00⟩*	13; 98.9	1	1
7	CH_3-CH	C≡C, CH═CH, CH═O, CH_3-O, *CH_3-C⟨H00⟩*, *CH⟨H11⟩ =C⟨H00⟩*, *CH⟨H11⟩ -O⟨H00⟩*	55; 95.3	1	1
8	CH_3-CH, *$C-CH_2$-CH_2*	C≡C, CH_2═C, CH═CH, CH═O, CH_3-O, C(═O)-O, *CH_3-C⟨H00⟩*, C⟨H00⟩ -O⟨H00⟩, *CH⟨H11⟩* -CH_2⟨H22⟩	4; 99.7	1	1

[a] In roman type, entries found by interpretation rules, in italics, those found by logical rules. The number of attached H atoms is specified by the atom tags within angle brackets, e.g., CH_2⟨H22⟩ signifies a C atom bearing exactly two hydrogens.

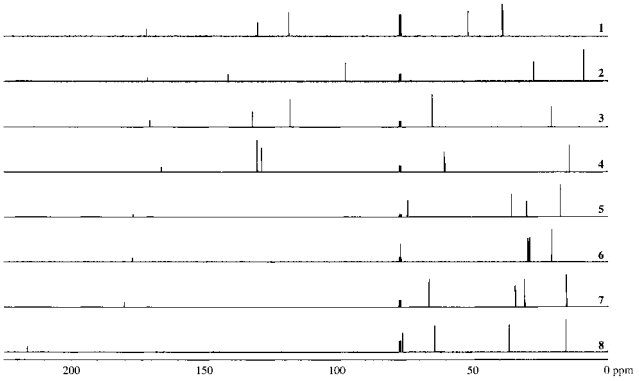

Figure 4. ^{13}C NMR spectra of the eight isomers shown in Figure 3.

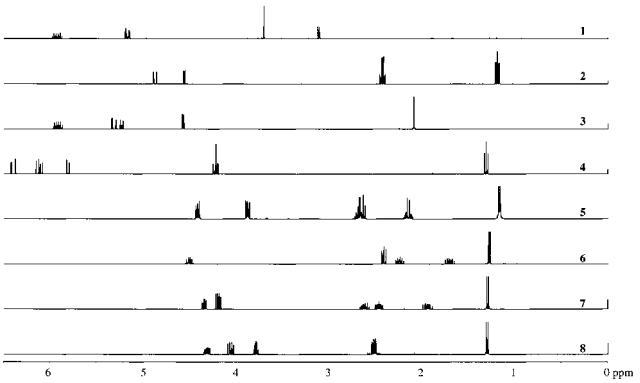

Figure 5. ^{1}H NMR spectra of the eight isomers shown in Figure 3.

and ^{1}H NMR spectra are represented in Figures 4 and 5. Their IR spectra were either full spectra or peak lists. All data were entered in JCAMP-DX format and processed fully automatically. The output of SpecInt 2 was directly forwarded to Assemble 2.1 without any editing.

The results are summarized in Table 3. The second column shows the fragments which were found to be present in the different structures (good-list); in roman letters are those found by the interpretation rules, and in italics those generated by the logical rules. In some cases, the number of attached hydrogen atoms is specified by the atom tags shown within angle brackets. The third column gives the fragments that were found to be absent (bad-list). The same atom constraints were obtained for all eight test examples, i.e., no molecular symmetry, exactly 1 C, 1 CH, 2 CH_2, 1 CH_3, and no OH groups.

With the molecular formula ($C_5H_8O_2$) alone, the structure generator Assemble 2.1 generated a total of 1168 isomers. With the help of the structural restrictions of SpecInt 2, the solution space was drastically reduced in each case (Table 3, column 4). The mean reduction factor was 97.3%. With Assemble 2.1, chemically improbable substructures such as small, unsaturated, condensed rings can be excluded.[5] When using this feature with our example, the total number of isomers was reduced to 1090, and the mean reduction factor was 97.5%. The last two columns of Table 3 show the position at which the correct solution was ranked by Assemble 2.1 on the basis of comparing estimated and measured NMR chemical shifts. Except in one case (compound **5**, ranking by ^{1}H NMR), the correct solutions figure as the best ones. More detailed information is given in Table 4. The noncyclic compounds **1**−**4** perform very well with both NMR methods. The mean and maximal deviations between estimated and measured ^{13}C NMR shifts are rather large for compounds **5**−**8**. It is known that the approach used here performs less good with five-membered rings than with other compounds.[21] Interestingly however, despite the presence of diastereotopic protons, the mean deviations for ^{1}H NMR were found to be rather small (= 0.21 ppm). Based on this detailed information, the number of solutions that are within a given threshold can be determined. Despite using rather wide limits for the mean and the maximal deviations (^{13}C NMR: 10.0 and 20.0 ppm, ^{1}H NMR: 0.4 and 0.8 ppm, respectively), only in three cases (compounds **5**−**7**), a few alternative structures were found with the ^{13}C NMR ranking. On the other hand, their number varies more strongly for the ^{1}H NMR ranking (Table 4, bottom part).

Most importantly, the structure elucidation system consisting of SpecInt 2 and Assemble 2.1 has been working without making any erroneous statements. The correct structures were always found among the best ones in the combined ranking based on both ^{13}C- and ^{1}H NMR information. By applying the structural restrictions, the solution space is reduced to a few percent of the number of isomers defined by the

Table 4. Ranking Results for the Eight Isomers of Figure 3 Based on ^{13}C- or ^{1}H NMR Spectra

NMR spectra	no.	ranking position	deviation for correct structure in ppm		no. of incorrect structures below threshold[a]
			mean	maximal	
^{13}C	**1**	1	1.90	2.70	0
	2	1	1.10	3.50	0
	3	1	3.30	5.20	0
	4	1	0.50	1.30	0
	5	1	5.80	9.30	3
	6	1	3.60	6.10	2
	7	1	5.10	11.00	2
	8	1	8.70	16.30	0
^{1}H	**1**	1	0.06	0.19	5
	2	1	0.07	0.15	2
	3	1	0.08	0.18	3
	4	1	0.02	0.07	1
	5	3	0.21	0.40	9
	6	1	0.15	0.36	0
	7	1	0.11	0.45	10
	8	1	0.18	0.66	1

[a] Threshold for ^{13}C NMR: mean deviation = 10.00 ppm, maximal deviation = 20.00 ppm, for ^{1}H NMR: mean deviation = 0.40 ppm, maximal deviation = 0.80 ppm.

molecular formula alone. The same results were found with a large set of test compounds, which will be described elsewhere. Although in the above example, ranking would be easy even when including all possible isomers, in more complex cases, a reduction of the solution space is absolutely necessary.

Preliminary investigations show that the mean reduction factor of ca. 98% remains constant up to a total number of possible isomers of 10 000 000. Since the algorithms used for NMR spectra estimations are fast (ca. 1 ms/estimated shift),[22−24] ranking is feasible with up to about 100 000 generated structures. Thus, the combined procedure can be applied to molecules having about 5 000 000 isomers. To larger molecules, it is not applicable at present. However, since the program internally applies a CH- and an HH-correlation matrix, it may easily be extended to use corresponding two-dimensional NMR spectra.

The performance of the combination of SpecInt 2 with a structure generator may be further improved since, at present, Assemble 2.1 does not make full use of the available information. To describe the neighborhood of an atom in a fragment, Assemble uses atom tags as a unique feature, which allows for representation of partial overlaps between nonoverlapping fragments and significantly contribute to the speed of processing. SpecInt 2, for its part, also provides this kind of information. However, in the case of overlapping good-list fragments as well as of bad-list fragments in general, Assemble 2.1 ignores the corresponding atom tags. This loss of information is especially severe with bad-list fragments, as it is not possible to delete their atom tags. For example, $CH\langle H11\rangle = C\langle H00\rangle$ as a bad-list item must not be replaced by CH═C since this would also demand the absence of the fragments CH═CH, CH_2═C, and CH_2═CH. As shown in Table 3, the logical rules of SpecInt 2 generate such bad-list fragments, but the actual version of Assemble does not use them further.

CONCLUSIONS

This paper shows that SpecInt 2, a new spectra interpretation program, is capable of generating structural information with high reliability. In combination with the structure generator, Assemble 2.1, and its ranking modules, structure problems of small and intermediate sized molecules can be automatically solved. The key to this success is that SpecInt 2 only uses highly reliable interpretation rules and that the fast NMR shift estimation programs of Assemble 2.1 allow a ranking of up to ca. 100 000 structures.

ACKNOWLEDGMENT

The authors thank Dr. D. Wegmann for careful reading of the manuscript.

REFERENCES AND NOTES

(1) Munk, M. E.; Sodano, C. S.; McLean, R. L.; Haskell, T. H. Actinobolin. I. Structure of actinobolamine. *J. Am. Chem. Soc.* **1967**, *89*, 4158−4165.
(2) Sasaki, S.; Abe, H.; Ouki, T.; Sakamoto, M.; Ochiai, S. *Anal. Chem.* **1968**, *40*, 2220−2223.
(3) Carhart, R. E.; Smith, D. H.; Brown, H.; Djerassi, C. Application of artificial intelligence for chemical inference. XVII. An approach to computer-assisted elucidation of molecular structure. *J. Am. Chem. Soc.* **1975**, *97*, 5755−5762.
(4) Funatsu, K.; Miyabayashi, N.; Sasaki, S. Further development of the structure generation in the automated structure elucidation system CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18−28.
(5) Badertscher, M.; Korytko, A.; Schulz, K. P.; M. Madison; Munk, M. E.; Portmann, P.; Junghans, M.; Fontana, P.; Pretsch, E. Assemble 2.0: A structure generator. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 73−79.
(6) Elyashberg, M. E.; Martirosian, E. R.; Karasev, Y. Z.; Thiele, H.; Somberg, H. X-PERT: A user-friendly expert system for molecular structure elucidation by spectral methods. *Anal. Chim. Acta* **1997**, *337*, 265−286.
(7) Kerber, A.; Laue, R.; Moser, D. Ein Strukturgenerator für molekulare Graphen. *Anal. Chim. Acta* **1990**, *235*, 221−228.
(8) Gray, N. A. B. *Computer-Assisted Structure Elucidation*; Wiley: New York, 1986.
(9) Munk, M. E. Computer-based structure determination: Then and now. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 997−1009.
(10) Dubois, J. E.; Carabedian, M.; Dagane, I. Computer-aided elucidation of structures by carbon-13 NMR. *Anal. Chim. Acta* **1984**, *158*, 217−233.
(11) Shelley, C. A.; Munk, M. E. Computer prediction of substructures from carbon-13 nuclear magnetic resonance spectra. *Anal. Chem.* **1982**, *54*, 516−521.
(12) Will, M.; Fachinger, W.; Richert, J. R. Fully automated structure elucidation − a spectroscopist's dream comes true. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 221−227.
(13) Schriber, H.; Pretsch, E. A rule-based system to automatically derive goodlist and badlist entries for structure generators from spectra. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 884−891.
(14) Schriber, H.; Pretsch, E. General characteristics of good-list and bad-list entries for structure generators from spectra. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 879−883.
(15) *Assemble 2.1;* Upstream Solutions GmbH, Scientific Software Engineering: Bergstrasse 144, CH-8032 Zürich, Switzerland. Free actual version treating up to 15 non-hydrogen atoms can be downloaded from www.upstream.ch.
(16) McDonald, R. S.; Jr., P. A. W. JCAMP-DX: A standard form for exchange of infrared spectra in computer readable form. *Appl. Spectrosc.* **1988**, *42*, 151−162.
(17) Davies, A. N.; Lampen, P. JCAMP-DX for NMR. *Appl. Spectrosc.* **1993**, *47*, 1093−1099.
(18) Lampen, P.; Hillig, H.; Davies, A. N.; Linscheid, M. JCAMP-DX for mass spectrometry. *Appl. Spectrosc.* **1994**, *48*, 1545−1552.
(19) Fürst, A.; Clerc, J. T.; Pretsch, E. A computer program for the computation of the molecular formula. *Chemom. Intell. Lab. Syst.* **1989**, *5*, 329−334.
(20) Christie, B. D.; Munk, M. E. Structure generation by reduction: A new strategy for computer-assisted structure elucidation. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 87−93.
(21) Fürst, A.; Pretsch, E.; Robien, W. A comprehensive parameter set for the prediction of the ^{13}C NMR chemical shifts of sp^3-hybridized carbon atoms in organic compounds. *Anal. Chim. Acta* **1990**, *233*, 213−222.

(22) Fürst, A.; Pretsch, E. A computer program for the prediction of ^{13}C NMR chemical shifts of organic compounds. *Anal. Chim. Acta* **1990**, *229*, 17−25.
(23) Bürgin Schaller, R.; Pretsch, E. A computer program for the automatic estimation of ^{1}H NMR chemical shifts. *Anal. Chim. Acta* **1994**, *290*, 295−302.
(24) Bürgin Schaller, R.; Munk, M. E.; Pretsch, E. Spectra estimation for computer-aided structure determination. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 239−243.

CI0101096