Fresenius Z Anal Chem (1989) 335:884–886

Fresenius Zeitschrift für Analytische Chemie

Infrared spectra for a multi discipline spectroscopy system*

Antony N. Davies

ISAS, Institut für Spektrochemie, Postfach 10 13 52, D-4600 Dortmund 1, Federal Republic of Germany

Infrarotspektren für ein multidisziplinäres Spektroskopie-System

Summary. The infrared spectra quality requirements of the "Spektrendatenbanken-Verbundsystem" (SDVS) are reported with special reference to Expert System needs. Some of the potential benefits of the system are given. A "Call for Spectra" is made and the guidelines for submitting spectra to the database outlined.

Introduction

Automation in spectrochemical analysis has produced a tidal wave of data requiring processing. In modern analytical laboratories samples no longer have to wait in long queues to be analysed, however, the bottle neck has moved to the analysis of the resulting spectra [1]. Automated spectral comparison with library data in a similarity or identity search give adequate results where the analysis is of a relatively simple nature or of virtually pure compounds [2–13].

A good multicomponent analysis (MCA) programme which has been correctly parameterised can give quite adequate results for complex mixtures (beware, some of the "latest" commercial MCA software supplied by spectrometer manufacturers is flawed in it's basic mathematics and can give misleading results). However, analysis of research results, where it would be hoped that the substance or mixture under analysis is not known to the analyst, negates both of the classic methods of automated spectral analysis. Something more that the analysis with known standards is required.

One way forward would be the some form of substructure analysis where the elements of an unknown sample could be identified from similar features observed and assigned in well characterised reference data [1]. Combining these observations would then lead to a identification of the probable unknown. The regular combination of different analytical spectroscopic techniques to produce a unified analytical "engine" is not at present common practice despite the fact that a simple IR, NMR, and MS based substructure analysis combination should be far more powerful than the best algorithm working only on data from one of these fields [14].

Inference engines based on library similarity searches, spectral data interpretation followed by substructure generation and analysis ought to be able to converge to a plausible solution. When this convergence is not possible the system should be capable of suggesting further experimental work to fill gaps in the knowledge base or root out conflicting anomalies in the data already held. Inversely, if chemical property data were included the knowledge base then the inference engine would be able to predict possible chemical behaviour in the unknown or even in hypothetical structures and environments. This could easily be extended such that when given a set of required chemical properties the inference engine would suggest compounds or mixtures and their appropriate environments in which could be synthesised to produce the required results.

However, this is flying before we have learnt to walk. The capability of any such system is limited to the quality of the knowledge base reference data at it's disposal. Although any reasonable inference engine should be able to cope with outliers it is unreasonable to except "expert system" results from data which no "expert" human would regard as acceptable. From this point of view the abundance of poor quality spectral data measured on dispersive instruments and later "digitised" which is available as "digital" data is only a hinderance, or at best a start point from which to rapidly improve.

The government of the Federal Republic of Germany recognised the potential of a combined spectroscopic database system working on data collected from many spectroscopic fields as a potential national asset and instigated a project to build a system, the Spektren-Datenbank Verbundsystem or SDVS, such as is described below (Fig. 1.). The Institute for Spectrochemistry and Applied Spectroscopy in Dortmund is responsible for, amongst other things, optical spectra collection and evaluation.

Spectral quality

As recognised by many authors "rigorous attention to quality during the initial development of the database would pay off in future applications" [15]. To this end several committees were formed to provide guidelines as to what would be acceptable quality for submitted spectra.

The results obtained from these committees were as expected somewhat dependant on their particular interpretation of their remit [16]. The Mass Spectrometry Committee provided a set guidelines along the lines of "no worse than this will be accepted". The Optical Spectroscopy Committee took a different approach and provided a set of rules for top

* Dedicated to Prof. Dr. G. Tölg on the occasion of his 60th birthday

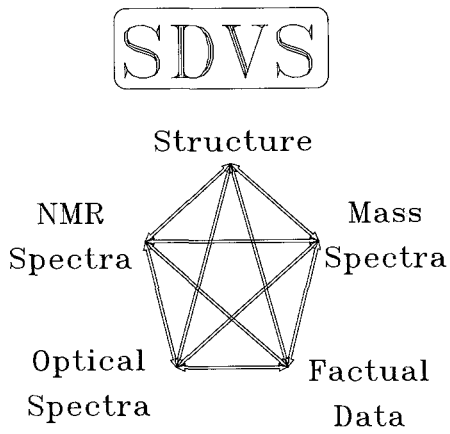

Fig. 1. The concept behind SDVS

quality spectra. The guidelines of the Optical Spectroscopy Subcommittee were as follows:

1. All optical spectra must be stored in full without previous data manipulation (e.g. smoothing, baseline correction) apart from conversion to absorption units.

2. From the field of optical spectroscopy, the collection of infrared spectra should take priority. Not only condensed phase transmission spectra but also ATR and low resolution gas phase spectra should be accepted. The wavenumber range in the infrared should be at least 4000 cm^{-1} to 400 cm^{-1} although partial spectra should also be accepted. The wavenumber scale should be linear without breaks and the ordinate scale linear in absorbance to at least 4 significant figures. The strongest band should have an absorbance of between 1 and 2A. If the remaining bands are more than a factor of 4 weaker then a second spectrum should be supplied from a sample of greater thickness or higher concentration.

3. It is envisaged that eventually Raman, UV/Vis transmission and fluorescence, and NIR spectra will be included.

4. Initially a "critical mass" of good quality spectra should be obtained from existing libraries as a nucleus for future development. This would enable meaningful work to be carried out immediately the database came on-line. This core data would then be replaced with higher quality spectra as they became available.

5. Old spectra of rare or "exotic" substances measured on dispersive instruments or digitised from hardcopy spectra will be tolerated where better data is not available.

These guidelines whilst being simple in form present significant problems for the spectroscopist specifically in the area of data manipulation. All interferometrically based spectrometers manipulate the data to some extent. Certain spectrometers more than others by introducing some degree of smoothing automatically. There are also no guidelines laid down relating to sample purity nor any quantitative method for determining the quality of a given spectrum. As the aim of the project is to operate an active regularly updated online library it is necessary to have a numerical quality index of some nature to indicate when a newly received spectra is of a higher standard to that of a currently held spectra of the same sample structure prepared by the same method.

Much has been published in the last few years on spectral quality and database requirements [17–21]. Recently Griffiths published a multi-page algorithm to be completed

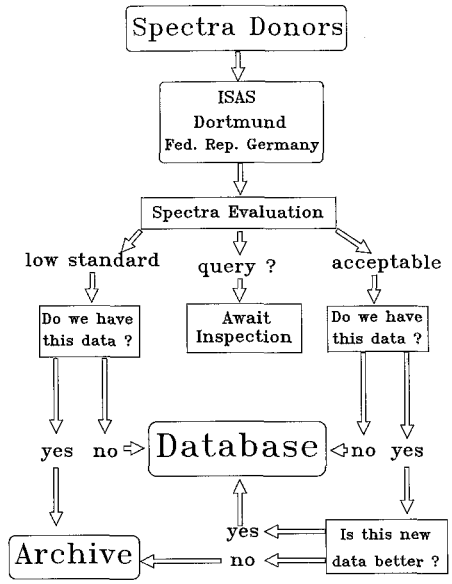

Fig. 2. Schematic diagram of the data flow through ISAS within SDVS

whenever an infrared spectrum was to be submitted to the EPA database that he is collecting [22]. From our standpoint the use of question and answer sessions would only be practical if it were carried out at the time a spectrum is being measured. If a database is to be founded on spectra already recorded then the use of this form-filling approach is of questionable value. There is also the danger that when different laboratories are presented with the same set of questions different quality indices will result for what are basically identical spectra.

For old spectra it is our intention to analyse the data numerically taking as much information as possible from the JCAMP-DX header zone and the rest from automated study of the spectra [23]. We will be using a combination of noise, background trend, and bandshape analysis alongside basic spectral interpretation related to the submitted structure data. Previous spectral manipulation will be looked for and the work will be carried out both in the absorbance as well as the fourier domain.

Spectra which are obviously of inferior quality to those already in the database will be archived but not included. Spectra whose quality is dubious but which are from samples or sampling methods from which we currently do not have a spectra in the database will be accepted pending the acquisition of better spectra. Spectra which throw doubt on the structure submitted or vice versa will be side-lined to await human intervention. Figure 2 shows a data handling chart for spectra passing through ISAS within the SDVS.

It is hoped that once the spectra-measuring public have become used to submitting spectra and the corresponding data to the database then the question and answer session approach to spectral quality classification might become more feasible. We envisage issuing such a system linked into a software package for converting data into JCAMP-DX format for submission.

Structural data will be expected to be in JCAMP-CS format [24] and all textual information on the respective structures and spectra will be available in a parallel textual database to that doing the spectral searching.

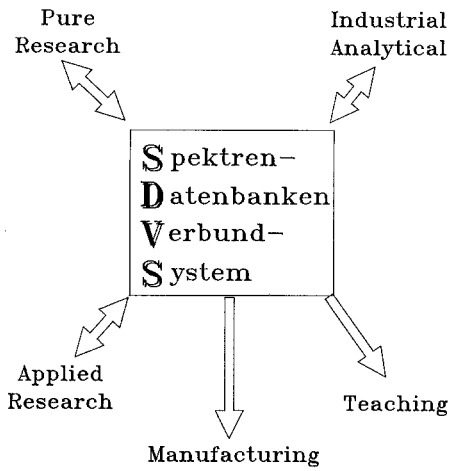

Fig. 3. Prospective users of the SDVS

Call for spectra

The Combined Database System will be available online and will be of use not only as a research or analytical spectroscopists aid but also will be of use for teaching purposes. It will provide for smaller companies and institutes a much broader factual resource for their work than they would previously have had access to and hopefully prevent expensive errors caused use of limited knowledge bases (Fig. 3).

All data held on the system will retain its origin information be it a chemical manufacturer or a university researcher enabling a new avenue of contact to be opened up. Users of the system being able to find out at a glance who produces or who is working on what, all the textual information from the different structured compounds being retained in the text database even after the IR spectra have been superceded by ones of better quality.

Anyone interested in submitting spectra to the database, or obtaining more information about this project, should contact the author at the above address. We are looking initially for high quality fourier transform infrared spectra of pure compounds submitted with the structure of the compound in JCAMP standard format.

Guidelines for contributed spectra

1. All Infrared spectra should be recorded interferometrically. They should be original, unreduced, and not manipulated in any way.
2. The spectra should be submitted in JCAMP-DX format with the relevant structure information in JCAMP-CS format.
3. The sample purity should be stated with at least one other measurement as secondary verification.
4. The spectra should comply as closely as possible to the Coblentz Society Class II specifications.
5. Data can be submitted on whatever magnetic medium is suitable for the donor.

Acknowledgements. The author would like to thank the Bundesministerium für Forschung und Technologie for financial support and encouragement (project 08 G32 21), and the Ministerium für Wissenschaft und Forschung des Landes Nordrhein-Westfalen.

References

1. Bremser W, Neudert R (1987) ESN European Spectroscopy News 75:10
2. Olson ML (1981) SPIE 289:236
3. Kawata S, Noda T, Miniami S (1987) Appl Spectros 41:1176
4. Hangac G, Wiebolt RC, Lam RB, Isenhour TL (1982) Appl Spectros 36:40
5. Azarraga LV, Williams RR, DEHaseth JA (1981) Appl Spectros 35:466
6. Owens PM, Isenhour TL (1983) Anal Chem 55:1548
7. Fields RE, White RL (1987) Anal Chem 59:2709
8. Wang Ching Po, Isenhour TL (1987) Appl Spectros 14:185
9. Qian Cheng, Wilkins CL, (1988) J Chem Inf Comput Sci 28:53
10. Yu Jung-Pin, Friedrich HB (1987) Appl Spectros 41:869
11. Warren Jr. FV, Delaney MF (1983) Appl Spectros 37:172
12. Rasmussen GT, Isenhour TL (1979) Appl Spectros 33:371
13. Delaney MF, Hallowell Jr. JR, Warren Jr. FV (1985) J Chem Inf Comput Sci 25:27
14. Hippe Z, Duliban J, Licbarska R, Koziol J, Mazur M (1986) Anal Chem (Warsaw) 31:607
15. Lowry SR, Huppler DA, Anderson CR (1985) J Chem Inf Sci 25:235
16. Final Report (1989) SDVS Project, Subproject: Test and development of the pilot systems. Pötzscher, Berlin, 27/1/89 Draft
17. Derendjaev BG (1986) Appl Spectros 40:707
18. Jones RN (1987) ESN European Spectroscopy News 75:28
19. Griffiths PR et al. (1983) Appl Spectros 37:458
20. No Author (1975) Anal Chem 47:945A
21. Becker ED (1978) Pure Appl Chem 50:231
22. Griffiths PR, Wilkins CL (1988) Appl Spectros 42:538
23. McDonald RSM, Wilks Jr. PA (1988) Appl Spectros 42:151
24. Gasteiger J, Hendriks BMP, Hoever P, Jochum C, Somberg H, (1988) Preprint, Revision 3.6, June 21st 1988

Received June 15, 1989