The new Automated Mass Spectrometry Deconvolution and Identification System (AMDIS)

Antony N. Davies

ISAS, Institut für Spektrochemie und Angewandte Spektroskopie, Dortmund, Germany.

Introduction

This column has not dealt with mass spectrometric issues since 1991–3.[1] This five year lapse needs to be redressed and the recent release of the NIST'98 mass spectral library seemed a good opportunity.

At a recent meeting at NIST in Washington I had the pleasure of meeting with Gary Mallard and Steve Stein who are amongst the staff responsible for the new NIST'98 release of the NIST/EPA/NIH Mass Spectral Library.[2] The new release has been extensively updated, quality controlled and is far larger than earlier versions. There is now a series of helpful utilities shipped with the package amongst which is a powerful program called the Automated Mass Spectrometry Deconvolution and Identification System (AMDIS) written for analysing Gas Chromatography (GC) or Liquid Chromatography (LC) Electron Ionisation Mass Spectral data sets.

This new program has been developed using US government funds over a two year period to provide a badly needed resource for policing the Chemical Weapons Convention.[3] The aim of the development was to provide better analysis interfaces for GC/MS and LC/MS data sets. An additional module was provided by the Ministry of Defense of Israel which provides additional "match factors" for making chemical identifications.

Six speciality target libraries are installed which consist of selected spectra from the 130,000 in the full NIST/EPA/NIH library:

- NISTCW is 62 chemical-weapons related data sets,
- NISTDRUG is 778 drug analysis spectra,
- NISTEPA is 1106 Environmental Protection Agency (EPA) target compounds,
- NISTFDA 419 for Food and Drug Administration (FDA) analyses,
- NISTTOX 1251 spectra toxicological library,
- NISTFF 993 flavours and fragrances

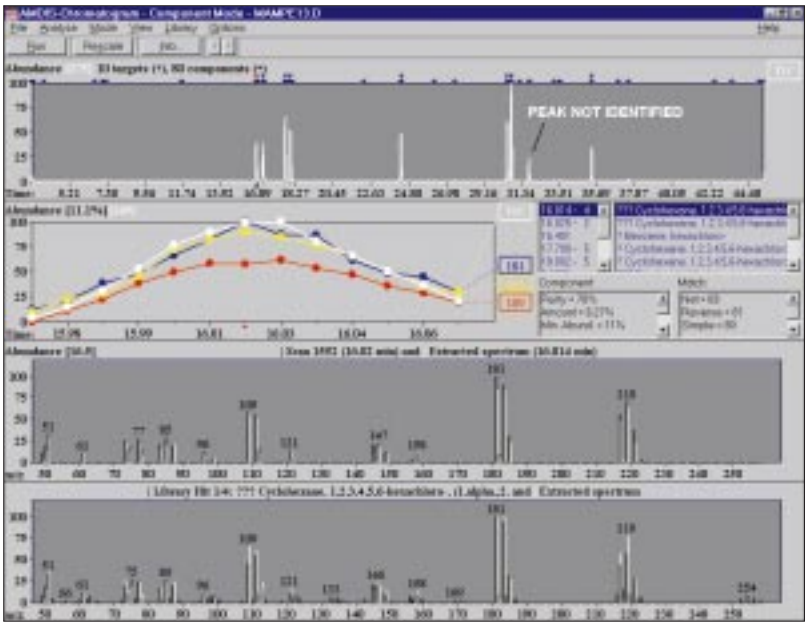

Figure 1. This is the AMDIS user interface. The data set is a GC/MS analysis of a round-robin sample of milk powder contaminated with eight chlor-organic pesticides at a concentration of 4 ng mL^{-1}. The target database is the EPA.

How it works

The program analyses individual single ion chromatograms to identify automatically possible component peaks within the GC/MS data set. All of the masses which are identified as having the same chromatographic characteristics are then put forward as the unknown spectrum for a spectral search within a selected target library (Figure 1).

The chromatographic trace at the top if the figure has a number of small markings along the upper border (Figure 2.) The "▼" shapes indicate components located and the "**T**" above these components indicates that one or more substances in the target library fit the component mass spectrum. By clicking on either the component or target symbol it is possible to see the

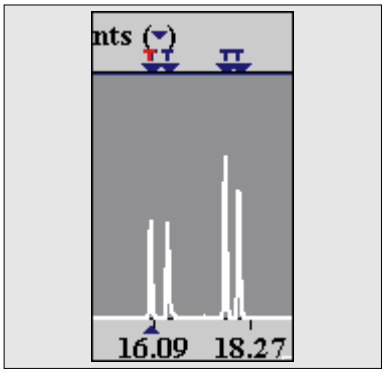

Figure 2. A cut out of Figure 1 showing the four chromatography peaks around 17 minutes and the symbols indicating that AMDIS has located components "▼" and has hits from the target database "T".

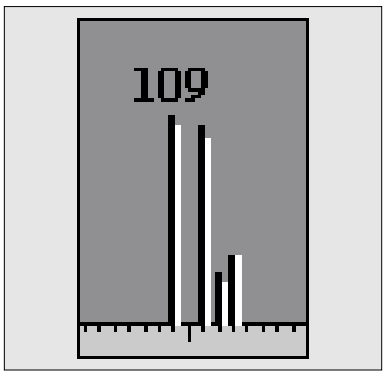

Figure 3. A cut out of Figure 1 showing the original scan data in black and the extracted component spectral data based on the AMDIS analysis.

```
AMDIS GC/MS Analysis Report - Data:D:\DATA\MAMPE13.FIN Page 001
AMDIS GC/MS Analysis Report
Organization: ISAS Dortmund
Division: Organic Analysis
Data: MAMPE13.D
Library: D:\Programme\AMDIS32\LIB\Nistepa.msl
Number of Identifications: 10
RT(min)  Hits     Chemical Name
16.014           4      ??? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
                                ??? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
                                ??? Cyclohexane, 1,2,3,4,5,6-hexachloro-???
                                Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
16.025           2      ??? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
                                ??? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
16.401                  ? Benzene, hexachloro-
17.700           5      ? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
                                ? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
                                ??? Cyclohexane, 1,2,3,4,5,6-hexachloro-???
                                Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
                                ??? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
18.002           5      ? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
                                ?? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
                                ??? Cyclohexane, 1,2,3,4,5,6-hexachloro-???
                                Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
                                ??? Cyclohexane, 1,2,3,4,5,6-hexachloro-, (1.alpha.,2.
24.391                  ??? Di-n-butyl phthalate
30.541                  ??? Dieldrin
30.778                  Benzene, 1,1'-(dichloroethenylidene)bis[4-chloro-
30.789                  Benzene, 1,1'-(dichloroethenylidene)bis[4-chloro-
35.419           2      ??? Benzene, 1,1'-(2,2-dichloroethylidene)bis[4-chloro
                                ??? Benzene, 1-chloro-2-[2,2,2-trichloro-1-(4-chloroph
```

Figure 5. A section taken from the results report printout. This clearly shows the problems of identifying various hexachlorocyclohexane isomers by mass spectral data alone!

relevant mass spectral data sets. The active symbol then turns red.

The mixture of black and white peaks are the scan data and the extracted spectrum (white) and secondly (lower trace) the target database hit and the extracted spectrum (white) (Figure 3).

The complete result report can be viewed within another component of the package (Figure 4) and printed out in Analysis Reports complete with relevant QA/QC data (Figure 5).

The test data set shown in Figure 1 is from a GC/MS round-robin experiment into the analysis of contaminated milk powder. Eight organo-chlorine pesticides were in the samples at concentration around 4 µg µL^{-1}. Using the EPA target library and the standard settings 10 different components were found, however, for the chromatography peaks at 16.0 and 30.7 minutes two different components were located within each peak but were identified as the same compound.

A peak was identified at 24.4 minutes as a plasticiser contaminant and the Endrin peak at 31.84 minutes was located but as Endrin was missing from the target library and as no similar compounds were found no hit was registered.

Figure 6 is a report constructed by exporting text via the clipboard into a word processing package. Information in the AMDIS reports view relating to the first hits and their synonyms is combined with the chromatographic peak retention time along with the correct pesticide identification provided by the user.

AMDIS had succeeded in identifying a component spectra at 31.78 minutes but unable to identify it. The next step was to take the unknown component spectrum extracted by AMDIS and, having added it to a search list, transfer it to the full NIST'98 database. This was singularly unsuccessful with Endrin only appearing at rank 12 in the hit list. However, using the hybrid search option combining the standard similarity search with a neutral loss search promoted Endrin to third in the hit list. This unspectacular performance on Endrin is apparently (according to our mass spectroscopists) down to the poor nature of the mass spectrum recorded here with little isotopic information essential to achieving good database search results.

To give AMDIS a better chance of finding all our substances, a new target database was created combining the EPA and FDA libraries and the Endrin spectrum exported from the NIST'98 database.

This proved the old adage that if "you know beforehand what you are looking for you are sure to find it" in mass spectrometry and as can be seen in Figure 7 the missing peak is now

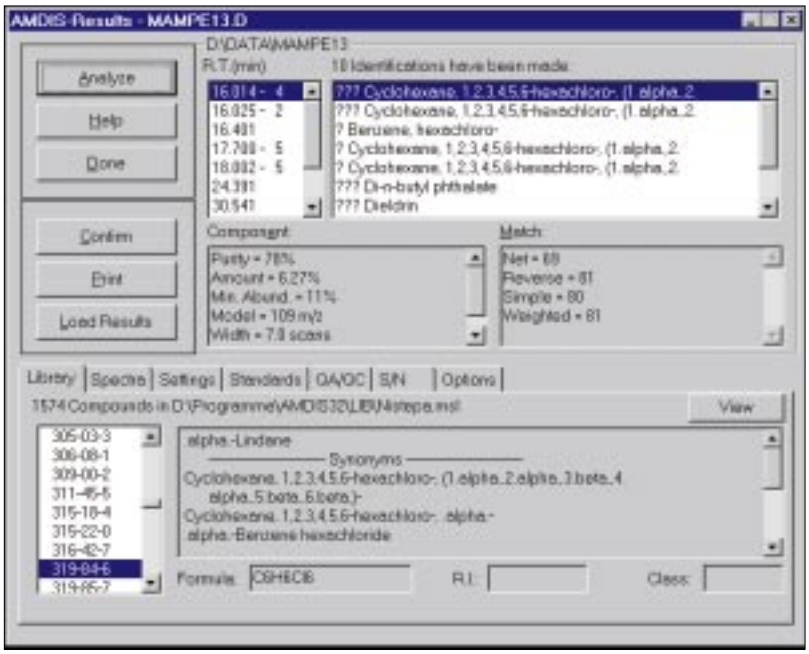

Figure 4. The AMDIS results interface. The text windows can be copied using windows clipboard which helps in making up user report sheets such as shown in Figure 6.

PEAK 1 (16.014) *alpha-HCH*
.alpha.-Lindane
---------- Synonyms ----------
Cyclohexane, 1,2,3,4,5,6-hexachloro-,
(1.alpha.,2.alpha.,3.beta.,4.alpha.,5.beta.,6.beta.)-
Cyclohexane, 1,2,3,4,5,6-hexachloro-, .alpha.-
.alpha.-Benzene hexachloride
PEAK 2 (16.0254)
.alpha.-Lindane
---------- Synonyms ----------
Cyclohexane, 1,2,3,4,5,6-hexachloro-,
(1.alpha.,2.alpha.,3.beta.,4.alpha.,5.beta.,6.beta.)-
Cyclohexane, 1,2,3,4,5,6-hexachloro-, .alpha.-
.alpha.-Benzene hexachloride
PEAK 3 (16.401) *HCB*
Benzene, hexachloro-
---------- Synonyms ----------
Amatin; Anticarie; Bunt-cure; Bunt-no-more; Co-op Hexa;
Hexachlorobenzene; HCB; Julin's Carbon chloride;
No Bunt;

PEAK 4 (17.700) *beta-HCH*
.delta.-Lindane
---------- Synonyms ----------
Cyclohexane, 1,2,3,4,5,6-hexachloro,
(1.alpha.,2.alpha.,3.alpha.,4.beta.,5.alpha.,6.beta.)-
Cyclohexane, 1,2,3,4,5,6-hexachloro-, .delta.-
.delta.-(Aeeee)-1,2,3,4,5,6-Hexachlorocyclohexane
.delta.-Benzene hexachloride; .delta.-BHC
.delta.-Hexachlorocyclohexane
.delta.-HCH; .delta.,1,2,3,4,5,6-Hexachlorocyclohexane
Cyclohexane, .delta.,1,2,3,4,5,6-hexachloro-
PEAK 5 (18.002) *gamma-HCH*
.delta.-Lindane
---------- Synonyms ----------
Cyclohexane, 1,2,3,4,5,6-hexachloro-,
(1.alpha.,2.alpha.,3.alpha.,4.beta.,5.alpha.,6.beta.)-
Cyclohexane, 1,2,3,4,5,6-hexachloro-, .delta.-
.delta.-(Aeeee)-1,2,3,4,5,6-Hexachlorocyclohexane
.delta.-Benzene hexachloride; .delta.-BHC
.delta.-Hexachlorocyclohexane
.delta.-HCH; .delta.,1,2,3,4,5,6-Hexachlorocyclohexane
Cyclohexane, .delta.,1,2,3,4,5,6-hexachloro-

PEAK 6 (24.391) *plasticiser contamination*
Dibutyl phthalate
---------- Synonyms ----------
1,2-Benzenedicarboxylic acid, dibutyl ester
Phthalic acid, dibutyl ester
n-Butyl phthalate; Butyl phthalate; Celluflex DPB
Elaol; Genoplast B; Hexaplas M/B; Palatinol C
PEAK 7 (30.541) *dieldrin*
Dieldrin
---------- Synonyms ----------
2,7:3,6-Dimethanonaphth[2,3-b]oxirene, 3,4,5,6,9,9-
hexachloro-
1a,2,2a,3,6,6a,7,7a-octahydro-,
(1a.alpha.,2.beta.,2a.alpha.,3.
Beta.,6.beta.,6a.alpha.,7.beta.,7a.alpha.)-
1,4:5,8-Dimethanonaphthalene, 1,2,3,4,10,10-hexachloro-6,7-
epoxy-1,4,4a,5,6,7,8,8a-octahydro-, endo,exo-
exo-Dieldrin; Aldrin epoxide; Alvit 55; Dieldrex
Dielmoth; Dildrin; Dorytox
PEAK 8 (30.778) *p,p'-DDE*
p,p'-DDE
---------- Synonyms ----------
Benzene, 1,1'-(dichloroethenylidene)bis[4-chloro-
Ethylene, 1,1-dichloro-2,2-bis(p-chlorophenyl)-
p,p'-(Dichlorodiphenyl)dichloroethylene
DDE; 1,1-Dichloro-2,2-Bis(p-chlorophenyl)ethylene
1,1-Dichloro-2,2-di(p-chlorophenyl)ethylene
2,2-Bis(4-chlorophenyl)-1,1-dichloroethylene
4,4'-DDE ; 2,2-Bis(p-chlorophenyl)-1,1-dichloroethylene
PEAK 9 (30.789)
p,p'-DDE
---------- Synonyms ----------
Benzene, 1,1'-(dichloroethenylidene)bis[4-chloro-
Ethylene, 1,1-dichloro-2,2-bis(p-chlorophenyl)-
p,p'-(Dichlorodiphenyl)dichloroethylene; DDE
1,1-Dichloro-2,2-Bis(p-chlorophenyl)ethylene
1,1-Dichloro-2,2-di(p-chlorophenyl)ethylene
2,2-Bis(4-chlorophenyl)-1,1-dichloroethylene
4,4'-DDE; 2,2-Bis(p-chlorophenyl)-1,1-dichloroethylene
PEAK 10 (35.419) *p,p'-DDT*
1,1-Dichloro-2,2-bis(p-chlorophenyl)ethane
---------- Synonyms ----------
Benzene, 1,1'-(2,2-dichloroethylidene)bis[4-chloro-
Ethane, 1,1-dichloro-2,2-bis(p-chlorophenyl)-
p,p'-(Dichlorodiphenyl)dichloroethane
p,p'-DDD; p,p'-TDE; Dichlorodiphenyldichloroethane
Dilene; DDD; ME1700

Figure 6. The retention times reference substance identification and synonyms for the proposed best hits in the target EPA library.

Figure 7. The hits results improve when the Endrin spectrum expected around 31.7 minutes is actually represented in the target database!

correctly identified. Even though AMDIS identifies our "mystery peak" it shows it is not happy by writing "???Endrin" as the hit identification. In AMDIS speak it doesn't come much lower down the hit rankings and still get shown!

Conclusions

This has been a short article on one of the features which have appeared in the NIST'98 new release of the NIST/EPA/NIH Mass Spectral Library. Here we fed some very tough sample data into the program which was not set up to identify most of the compounds we were interested in. A later test using phenoxy alkanoic acid pesticides from spiked drinking water samples was even less successful until we expanded our target library after which all of our compounds were located and identified by AMDIS even where co-elution was taking place.

The best testers of new software systems are the people who work in the labs and who have to use them on a regular basis. The constant comment coming back from the people testing this installation was "why couldn't we have had this system last year when ....."! The 32 bit system running under Windows NT4 was completely stable and quick, the 16 bit version running under Windows 3.11 failed to read only one of the data sets we presented.

We are still only starting to play with this tool but I think the number of features which are being discovered daily and acclaimed by our staff places this as one of their firm favourites.

Readers may be interested in a new spectrum viewer/plugin that will shortly be available under:

http://www.labcontrol.com/ViewP.htm

although at the time of going to press the page was still under construction!

We would be grateful to receive any experiences you may have had using spectroscopic data standards especially with web software. E-mail to davies@isas-dortmund.de

References

1. Kurt Varmuza and Antony N. Davies, "A Software Tool for Recognising Hidden Relationships between Mass Spectra and Chemical Structures", *Spectroscopy International* **3(4),** 14–17 (1991). and A.N. Davies, "Mass Spectrometric Data Systems", *Spectroscopy Europe* **5(1),** 34–38 (1993).
2. NIST'98 Release of the NIST/EPA/NIH Mass Spectral Library is significantly expanded over previous versions with 129,136 spectra and almost complete coverage with chemical structures. For more general information please contact: Standard Reference Data Program, National Institute of Standards and Technology, Bldg 820, Rm 113, Gaithersburg, MD 20899-0001, USA. Te:l +1-301-975-2208, Fax: +1-301-926-0416, E-mail: srdata@nist.gov. The scientific contact for the database is: Stephen E. Stein, Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, MD 20899-0001, USA. Tel: +1-301-975-2505, Fax: +1-301-975-3670, E-mail: steve.stein@nist.gov or W. Gary Mallard, A111/221 Chemical Reference Data and Modelling Group NIST, Gaithersburg, MD 20899, USA.

Tel: +1-301-975-2564, Fax: +1-301-975-3670, E-mail: gary.mallard@nist.gov.

3. The Chemical Weapons Convention Website is at http://www.opcw.nl/ The website is run by the Technical Secretariat of the Organisation for the Prohibition of Chemical Weapons and provides not only information on the Chemical Weapons Convention but also on other issues relating to toxic chemicals and chemical protection.