Assessment, from an industrial user perspective, of some major competitor information files on pharmaceutical development products*

A. Mullen, M. Blunck, E. Kalbfleisch, E. Möller and H.G. Rohbeck

Bayer AG, Wuppertal, Germany

Received 7 February 1996
Revised 18 June 1996

Abstract.

This paper discusses in a critical manner the spectrum of commercially available files – for in-house use – on pharmaceutical development products, which has been expanding rapidly over the past few years. Increasing degrees of sophistication in both the file contents and ways of manipulating the data are becoming more evident. A brief overview of competitor information files of this type, currently being used by PDR (Pharmaceutical Documentation Ring) member companies, is included in this appraisal.

The strengths and weaknesses of some of these information resources of importance to pharmaceutical research and development, strategic planning, marketing and other functions within the industrial environment of the health care sector are outlined, using examples. A number of proposals are made for improving the coverage and content of the range of files under discussion, since optimal use of these resources can only be made if certain standards of quality control and indexing are met.

*Paper presented at the International Chemical Information Conference, Nîmes, France, 25 October 1995.
Correspondence to: Dr A. Mullen, Bayer AG, Pharma Research Centre, D-42096, Wuppertal, Germany. Tel: +49 (0)202 36 8495. Fax: +49 (0)202 36 4200. Internet: debay679 @ibmmail.com

1. Introduction

It would perhaps be fitting to start this overview – which will attempt to critically assess some major commercially available competitor information files on pharmaceutical development products – with a quotation from the famous Chinese philosopher Sun-tzu [1], which, although almost 2,500 years old, is still valid today:

> He who knows himself and his enemy always wins.
> He who knows himself but not his enemy sometimes wins and sometimes loses.
> He who does not know himself nor his enemy always loses.

Fig. 1. Sun and Zu in Ping-fa (The Art of War) 4th Century B.C.

To put the quotation (see Fig. 1) into the context of pharmaceutical development [2, 3], it is imperative to be aware of the activities of competitors when working on projects leading to a new ethical drug, in order that one'sproduct will enter the market first and is most likely to be the best of its class. As pharmaceutical development is a lengthy as well as a costly process, obvious mistakes, misdirections and wasted

Growing Impact of Competitor Activities

Prentis *et al.* cited the principal reasons for terminating NCEs*, namely inappropriate pharmacokinetics in man, lack of clinical efficacy, adverse effects in man, commercial reasons, and animal toxicity. No new causes were suggested from the present survey but it was noticeable that economic factors and actions of competitors were frequently mentioned. This may be a consequence of the increasing commercial pressures as total development times have lengthened and effective patent lives therefore decreased, so producing a smaller window of opportunity for the introduction and premium pricing of products. As competition becomes quickly apparent in the form of generic or alternative products there is increasing pressure to proceed with only those products which are likely to be very successful financially by showing a distinct therapeutic advantage over existing treatments.

*New chemical entities
Source: Halliday, Walker and Lumley [4].

Fig. 2. Growing impact of competitor activities.

effort/resources should be avoided at the outset, and therefore *you must constantly reassess your own activities*.

The growing need for the monitoring of competitor activities was stressed in a Centre for Medicines Research (CMR) report [4] on the reasons for terminating pharmaceutical research and development (R&D) projects. Besides unsuitable clinical efficacy, side effects and toxicity, competitor activities were frequently alluded to.

As can be appreciated from the extract depicted in Fig. 2, this report [4] stresses:

(1) increased commercial pressures as total development times have lengthened. Another result is the decreased lifetime of patents, meaning that opportunities are narrowing for the introduction of new products;
(2) increasing pressure to proceed only with highly promising products which exhibit a significant therapeutic advance over state-of-the art treatment, i.e. competitor products.

The Russian Intelligence Service and the French Deuxième Bureau have long recognised the fact that a considerable amount of economic-based competitor information can be gleaned from conventional sources by analysing them in detail. Making relevant sources available within the corporate environment opens a treasure house of data readily available for analysis by experts within the company [5, 6].

Competitor Information System in R&D Area Coverage

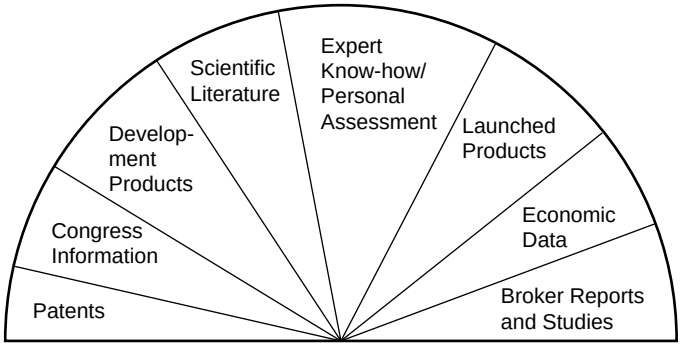

Fig. 3. Spectrum of competitor information sources.

To quote from an article in *Time*, 5 July 1993:

- 'Economic intelligence has nothing to do with the stealing of secrets – it is the analysis of (available) information'; and
- '80% of today's economic intelligence gathering can be done by analysing publicly available sources, such as online and other databases, publications, etc.'

To paraphrase Warr, from a recent article [7] in *Chemistry and Industry* on competitor intelligence:

- competitive intelligence enables a business to develop effective corporate strategies, expedite product development and to anticipate competitor actions and motives.

Consequently, it would appear to make sense to consider assembling a system of this type.

2. A possible 'ideal' in-house competitor information system

An 'ideal' in-house competitor information system – available to the corporate end-user community – should cover the spectrum shown in Fig. 3, i.e. from the inception of a potential drug (patent application) through congress information to development product reports to information about launched and discontinued products. The 'ideal' system should be backed up by access to the current scientific literature (including original publications), as well as providing facilities to input personal assessments of competitor products. The system could be rounded off by incidence and prevalence data sources, current pharmaceutical industry news, as well as information about available broker reports and studies, annual reports, etc.

Competitor Information System in R&D Area Objectives

- Making available a system with competitor information on biologically active substances from research organisations as well as pharmaceutical development compounds from competitor companies
- Basis – commercially available files to reduce the manpower effort needed while maintaining currency
- Integration of the various sources by means of a uniform search form, facilitating text and (sub)structure searching
- Easy-to-use information system for end-users
- Text and (sub)structure queries within standard corporate software environment
- Tool for the generation of specific information as well as overviews of competitor activities
- Standardised analysis program available

Fig. 4. Objectives of competitor information systems in R&D areas.

However, in this paper, the focus will be on selected development and launched product-related files stemming from commercial sources [8, 9].

Why are commercial databases so attractive as the backbone of the 'ideal' competitor information system?

(1) They eliminate the in-house need to assemble and gather data to keep databases current.
(2) The maintenance effort is low.
(3) The files are often supplied in standard formats and can be immediately loaded.
(4) They free manpower, which can be used to upgrade/enhance the information, i.e. the gathering of the raw data is 'outsourced'.

What are the objectives (cf. Fig. 4) of such a competitor information system based on pharmaceutical development and launched products?

(1) It should cover ongoing projects of competitor companies as well as those from academia.
(2) To keep the system user-friendly, standard corporate hard/software should form the basis of the competitor information system.
(3) Text and chemical structure retrieval should both be available.
(4) As already mentioned, commercial files should form the basis to help to reduce the internal manpower effort needed and to ensure that currency is maintained.
(5) The various data sources should be integrated as far as possible to support effective retrieval.
(6) In addition, standard programs should be available to generate specific information on a particular product as well as report outputs, statistical analyses, etc.

The competitor information system must be easy to use in order to encourage its wide acceptance within the corporate environment.

What are the various applications of this system?

(1) Provision of brief current information on specific competitor development products/new chemical entities (NCEs).
(2) Constant assessment of ongoing and planned internal R&D projects.
(3) Making transparent the new R&D approaches made by competitors.
(4) Generation of overviews of the activities of competitor organisations.

Additionally, as a spin-off, it is also a source of data/chemical structures for drug design and molecular modelling.

Before discoursing further about the benefits and otherwise of using commercially available files in-house for such applications, it would perhaps be interesting to take a more pragmatic approach by studying which files, of the currently available fifteen or so, are actually being used by the R&D-based pharmaceutical industry, by disclosing some data from an internal survey recently conducted among 27 Pharmaceutical Documentation Ring (PDR) member companies [10, 11, 12].

3. Activities of the pharmaceutical industry – PDR companies

The PDR is an international corporate-based organisation, which focuses on scientific information and documentation activities as they relate to pharmaceuticals. The present membership is displayed in Fig. 5 – almost all major European-based pharmaceutical companies can be found within the ranks of the PDR. Including ongoing mergers and acquisitions. PDR companies now account for around 65% of the worldwide turnover of the top 50 multinational

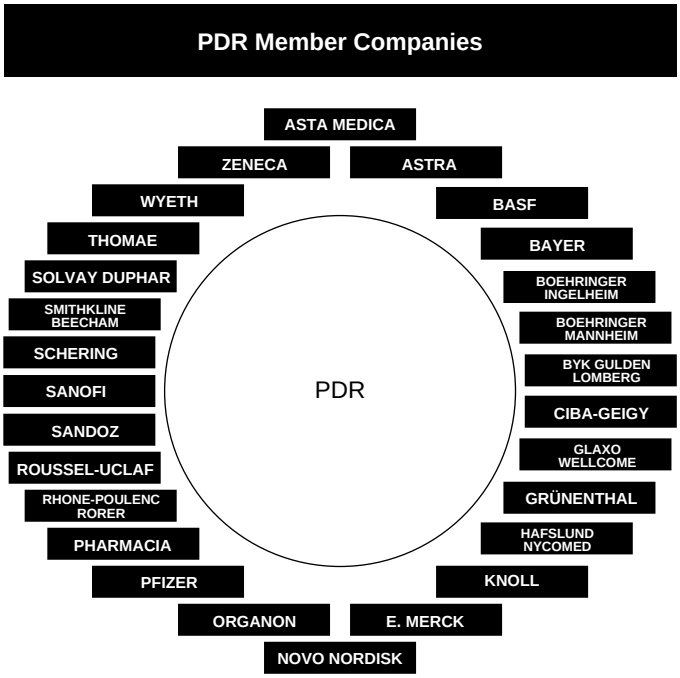

Fig. 5. 'Ring' of PDR companies.

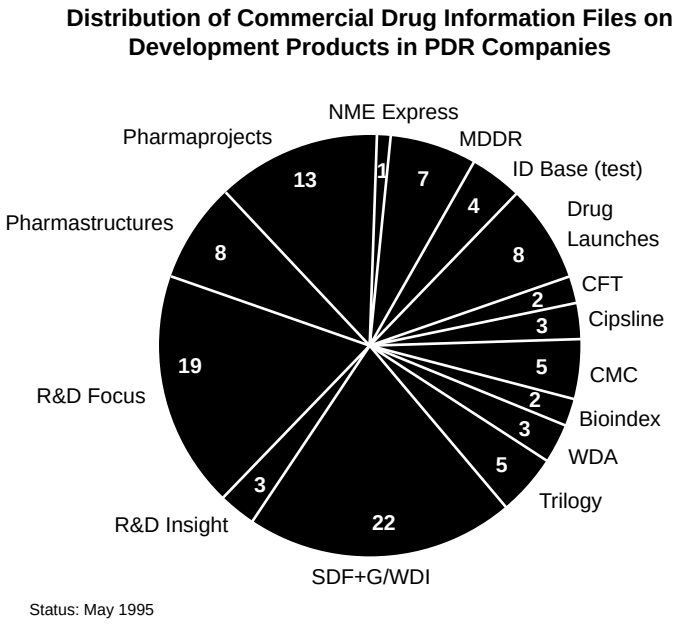

Fig. 6. Distribution of commercial in-house files among PDR companies.

pharmaceutical companies. In its early history, the PDR developed effective documentation systems for the scientific literature, patents and reactions, which were then, in part, handed over to Derwent. Over the last decade and more, the PDR has realigned its spectrum of activities to concentrate more closely on the exchange of experience among its members in non-confidential areas, as well as to provide an interface for the information industry to contact its customers as a group. A number of topics of mutual interest to its members are covered by PDR coordinators.

One PDR topic of special relevance to this presentation is entitled *Drug Information Files on Development Products* (8, 9, 10]. The present spread of commercial development product files used in-house by the PDR companies is shown in Fig. 6.

These files, which stem from eight different producers, are all substance or development product-based, i.e. a record in the database generally corresponds to a compound. While some of these files mainly focus on current awareness, e.g. NME Express and Conference Fast Track (CFT), as several years of data are made available, the user is justified in demanding that existing entries should be upgraded at a later point with, for example, CAS Registry Numbers, in order to make the identification of the same compound in the various sources easier.

However, it is not really possible to classify files into categories such as 'current awareness' or 'archival' files, as some of the files, e.g. Pharmaprojects and R&D Focus, which might be expected to fall into the latter category, experience very frequent updates (daily to weekly) in their online version, i.e. the currency and coverage of the data in these development-product based files can therefore be compared with one another.

As Derwent's [13] SDF (Standard Drug File) was one of the first files of its genre, it is not surprising to find it present in some 22 of the 27 PDR companies which participated in the survey in May 1995, even though this ranking probably does not reflect its relative level of usage within these organisations. Thereafter, follow R&D Focus (from IMS [14]) with nineteen installations, and Pharmaprojects (from PJB [15]) with thirteen. Together with the Pharmastructures file from PJB, these four files account for almost two-thirds of the area of the pie chart in Fig. 6.

Some fifteen files are present in one or more PDR companies. Except for MDDR (MDL's Drug Data Report) [17, 19] and Drug Launches from IMS [14], the remaining files are present in a maximum of five companies.

How has the PDR database scene changed since 1993? The apparent growth presented in Fig. 6 is partly due to the increase in PDR membership over

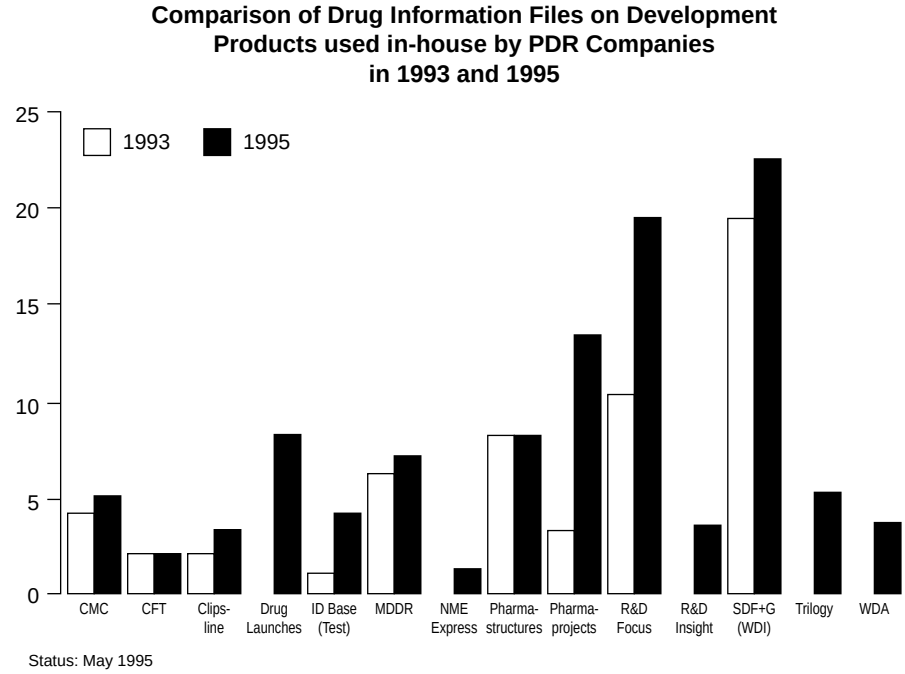

Fig. 7. Comparison of in-house files used by PDR companies in 1993 and 1995.

this period. The most striking increases are – in alphabetical order – for Drug Launches (8), Pharmaprojects (13) and R&D Focus (19). Of course, newer files such as R&D Insight [16], NME Express [17], etc, have still to make their mark.

Fig. 7 also shows that almost half of the PDR companies now have the SDF, R&D Focus and Pharmaprojects files available in-house.

Although Conference Fast Track (CFT) [13] is a highly regarded hard-copy service, it has not made much impact on in-house use as a database – even if it is a component of Derwent's World Drug Alerts (WDA). The CFT file provides a fast coverage of NCEs presented at major conferences (update around three to four weeks after the conclusion of a conference).

4. Sources of information utilised by files on competitor development products

Where does the information originate which is used to assemble these competitor information files? While patent applications are the first signal, who is capable at this stage – without adequate comparative data – of confidently identifying a lead candidate with a very strong likelihood of success? The first official publication is usually made at a scientific conference as a poster or presentation. This is the reason that the PDR persuaded Derwent [13] to launch its 'Conference Fast Track' service to provide rapid information on pharmaceutical development products disclosed at conferences. These disclosures at conferences are often supported by 'soft' facts supplied via the '*Kaffeeklatsch*' network.

Then, a company, particularly from the USA, might publish early information as a press release on new test compounds for public relations purposes and/or to influence its share values. After these steps, publications appear in the scientific literature and the information flow enters external online and other databases. Annual company reports and broker studies with turnover estimates then complete the information chain.

The commercial competitor databases under discussion make use of all of these sources, as well as direct contact with companies.

5. Selection of commercial files for closer appraisal

For this paper, almost half (seven) of the files from the PDR list have been selected; they are shown schematically in Fig. 8. They range from widely to less well-used files, i.e. a good cross-section.

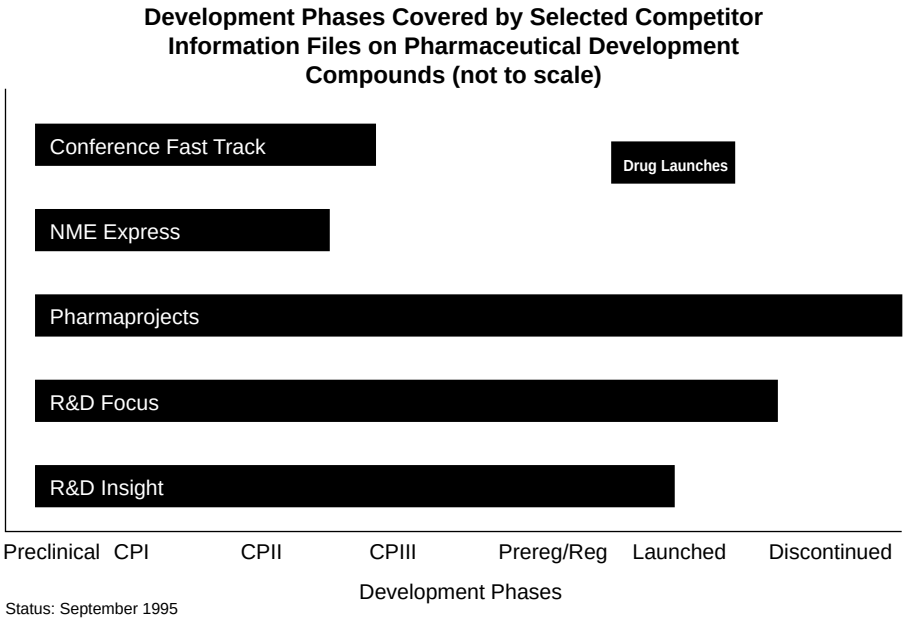

Fig. 8. Pharmaceutical development phases covered by selected files (CPI = Clinical Phase I, CPII = Clinical Phase II, etc).

Fig. 8 shows how the selected files – excluding the SDF/WDI (World Drug Index), which contains virtually no information about development status – cover the various phases in the life of a pharmaceutical product. For sheer extent of coverage, the Pharmaprojects file would appear to be the candidate of choice.

The following seven files were selected for closer appraisal:

(1) **Conference Fast Track** (diskette in, for example, ChemBase/ISIS format) from Derwent [13], which, as already mentioned, focuses on NCEs presented at recent major scientific conferences. This file is mainly designed for current awareness purposes and given a good coverage of around fifteen selected medicinal chemistry-related conferences per year. The file encompasses around 3,200 records. (Approximate price for single user: £3,400 per year.)

(2) **Drug Launches** (CD-ROM) from IMS [14], which concentrates on drugs introduced to over 50 countries since 1982 and includes around 78,000 records. As the name implies, there are no development compounds in this file. While the coverage of launched compounds is not 100%, different languages (French, Italian, Portuguese, etc) are used in the descriptive text and the updating of existing entries must be improved, it is a useful source when searching for drug names or studying the strategies of companies in different regions. (Approximate price for single user: £2,600 per year [20].)

(3) **NME Express** (diskette in ChemBase/ISIS format) from Prous [17], which reports on new test compounds presented at conferences or reported in the recent literature. This file focuses on current awareness and encompasses around 1,100 records. As shown later (cf. Section 6), it does report on compounds not to be found elsewhere. (Approximate price for single user: £530 per year.)

(4) **Pharmaprojects** (CD-ROM) – probably one of the most systematic, widely-used information sources on competitor development and launched products – is produced by PJB [15] and provides reports on products from the preclinical phase through launched to discontinued projects. Pharmaprojects provides currently the widest coverage of all the files presently available. This file contains around 18,000 records. (Approximate price for single user: £3,500 per year.)

(5) **R&D Focus** from IMS [14], while not covering as extensive an area as Pharmaprojects, often provides more detailed information in the way of reports than the latter. However, its level of standardisation, e.g. mechanism of action terminology, company name and their assignment to corporate groups, etc, is open to improvement. R&D Focus

contains around 8,300 records. Comparison of the coverage of compounds under 'active' development in Pharmaprojects and R&D Focus is discussed in Section 8. (Approximate price for single user: £3,650 per year [20].)

(6) The recently introduced **R&D Insight** (CD-ROM) file from ADIS [16], although it contains fewer records, i.e. development compounds, than either Pharmaprojects or R&D Focus, also covers compounds in the preclinical to launched phases. However, this file also provides some turnover estimates (around 9% of all compounds) from the Broker Lehman Brothers, as well as the most extensive summary of all files hitherto mentioned on the properties of the compound (pharmacokinetics, pharmacodynamics, clinical efficacy, etc). In addition, literature references on about two-thirds of the compounds are provided for further information. Around 45% of the literature references are backed up by high-quality LMS summaries of actual cited publications, often in tabular form, enabling the user to assess rapidly the merits or otherwise of the particular test compound without needing to go to other sources. The presence of more cited references as LMS summaries, and not just 45% thereof, would be a welcome enhancement. The file contains around 5,100 records. (Approximate price for single user: £5,300 per year.)

(7) The last file to be examined is Derwent's [13] **SDF** (Standard Drug File), now renamed WDI (World Drug Index). This file, one of the first of its type, is available in CD-ROM, as well as in other formats. The SDF is really a brief dictionary of biologically active compounds reported in the literature. It contains no information about the stage of development of the substance or the developer, merely the name(s) of the compound, its therapeutic use and pharmacological activity. The level of standardisation of this terminology needs to be improved. In terms of the number of records (around 48,000), it is, next to Drug Launches, the second largest of the seven files. (Approximate price for single user: £9,540 per year.)

As will be shown in later sections of this paper, each of the above files has its merits and its weaknesses, there is currently no *single* source which can be used with confidence for comprehensive coverage of an indication or a company.

5.1. Fields present in the selected commercial files

Table 1 presents the main thrusts of the different selected files – as can be seen, the best overall coverage is provided by Pharmaprojects, followed by R&D Focus.

Besides the compound name field, which, as it is in all files, is not extra listed, the chemical structure and pharmacological activity/therapeutic use fields are present in virtually every file.

Other files, such as Conference Fast Track and Drug Launches, focus on specific areas – test compounds

Table 1
Competitor information system in R&D area focal points of selected files on pharmaceutical development products. N = upcoming new feature; (•) = only partial coverage

Files	Focal Points									
	CAS Reg. No.	Chemical Structure	Activity/ Use	Patent Information	Congress Information	Development Products	Literature	Licensing Information	Launched Products	Discontinued Projects
Conference Fast Track		•	•		•	•				
Drug Launches	N	N	•						•	N
MDDR	(•)	•	•	•		(•)	(•)			
NME Express		•	•	(•)	•	•	•			
Pharmaprojects	•	•	•	•	•	•	•	•	•	•
R&D Focus	•	•	•	•	•	•	•	•	•	(•)
R&D Insight	•	•	•		•	•	•	•	(•)	(•)
SDF (WDI)	•	•	•			(•)	N		•	

Table 2
Normal updating frequency of files for in-house use

Conference Fast Track	ca. monthly
NME Express	2 or 4 weeks
Pharmaprojects	monthly
R&D Focus	monthly
R&D Insight	monthly
SDF/WDI	4–6 months

presented at conferences or launched products respectively. Conference Fast Track and NME Express are the only files not to provide CAS Registry Numbers [18]: a feature which is really a *must* for a competitor information source. As discussed previously, this data should be supplied in subsequent updates of the file.

No doubt, in time, the recently introduced R&D Insight file from ADIS [16] will widen its in-depth coverage to include more launched and discontinued projects.

5.2. Updating frequency of the selected commercial files

The updating of the selected files for inhouse use is generally monthly; only Prous [16] offers a more frequent standard updating modus for NME Express. The SDF or WDI File from Derwent [13] experiences an update every four to six months. (See Table 2.)

5.3. Spectral map analysis of the selected commercial files

To try to summarise some features of these seven selected commercial files, a spectral map analysis (cf. Fig. 7) was carried out – this is a method developed to analyse tabular data and display the relationships between these data using low dimensional projections.

The circles represent the different databases under discussion, the area of a circle corresponding to the combination of the features (squares), i.e. its:

- total size in bytes;
- total cost; and
- number of records (one record normally represents a compound).

The relative positions of the circle (files) to the squares (features) indicates the affinity of a file to this feature, e.g. the SDF [13] and Drug Launches [14] files – compared to the other databases – have a large

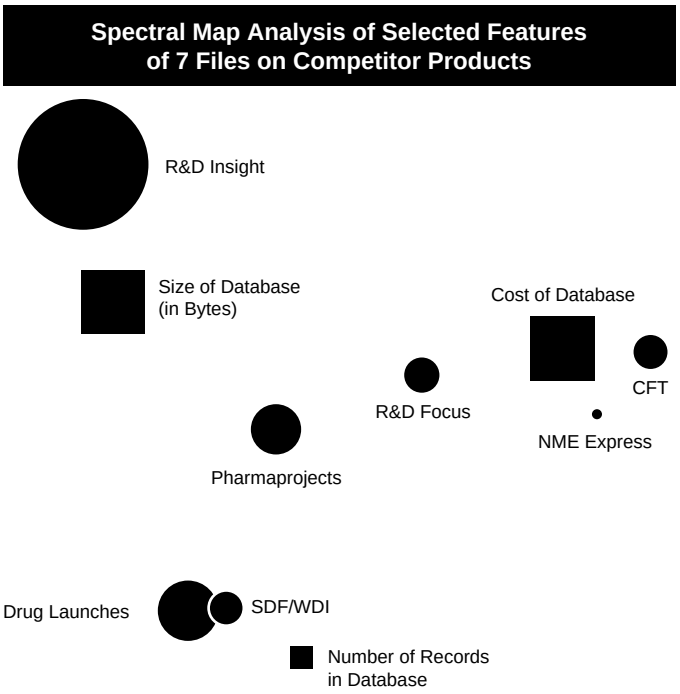

Fig. 9. Spectral map analysis.

number of records in relation to their total size in bytes and total cost.

On the other hand, Conference Fast Track [13], NME Express [17] and R&D Focus [14] are located closer to the cost feature and further away from the other two factors – total size of database in bytes and number of records. Of course, files such as Conference Fast Track are small in terms of the number of records (NCEs from fifteen conferences per year), but of relatively high quality and very up to date – both these factors obviously contribute significantly to their cost.

At the top left of the chart (Fig. 9), R&D Insight's [16] dominating feature is the total size of the database in bytes – almost six times as large as its nearest rival, i.e. a considerable amount of information is provided in the database in relation to the number of records and the total cost.

Pharmaprojects [15] takes up a median position: the relation between size, cost and number of records equals the mean of these relations for the databases under consideration.

Fig. 9 is not meant to pinpoint the 'best buy', as 'quality' cannot be quantified readily, but merely to illustrate some features of these seven files in relation to one another.

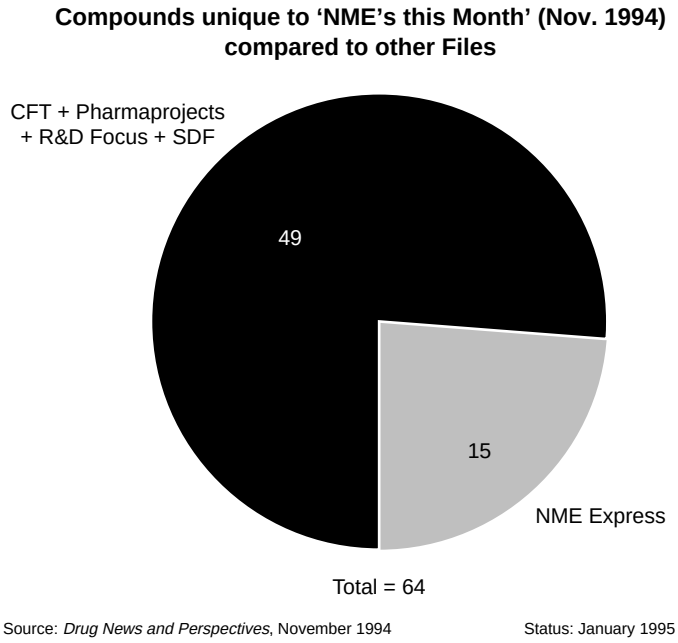

Fig. 10. 'Unique' development compounds in NME Express.

6. Comparison of coverage of NME Express with four of the selected commercial files

The obvious questions crop up when looking at the spectrum of databases which are commercially available: do we really need them all? Is there not a huge degree of overlap?

At the beginning of 1995, the monthly update of Prous' NME Express [17] database (with some 64 compounds from November 1994) was compared with that of four of the files already mentioned. It was found that almost 25% of these 'new molecular entities' – all with test compound numbers – reported in NME Express were not to be found in Conference Fast Track or Pharmaprojects or R&D Focus or the SDF.

The sources used by Prous [17] for NME Express are reported to be the latest journal literature, scientific congresses, company press releases, etc. When the relative coverage of the above-mentioned files was examined, the distribution in Fig. 10 resulted.

Conference Fast Track, Pharmaprojects, R&D Focus and the SDF accounted for 49 of the 64 test substances in NME Express. *Fifteen* compounds were unique to NME Express! The spread of the coverage of the

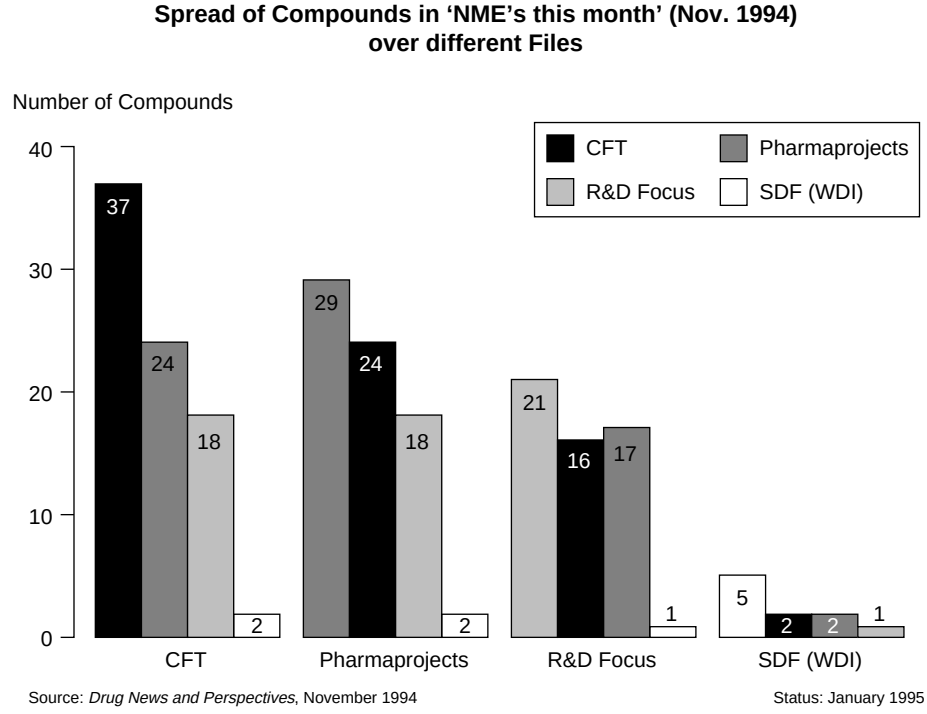

Fig. 11. Coverage of development compounds in NME Express relative to other selected files.

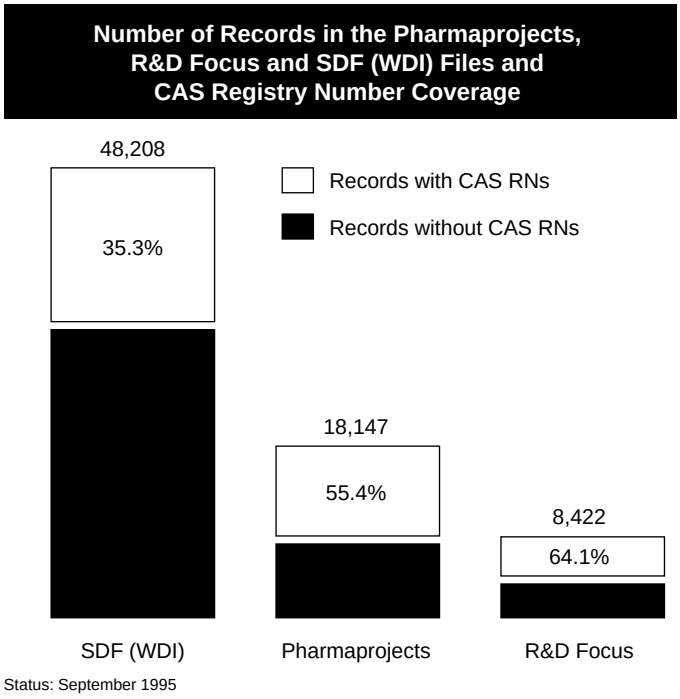

Fig. 12. Coverage of Chemical Abstracts Registry Numbers in three selected files.

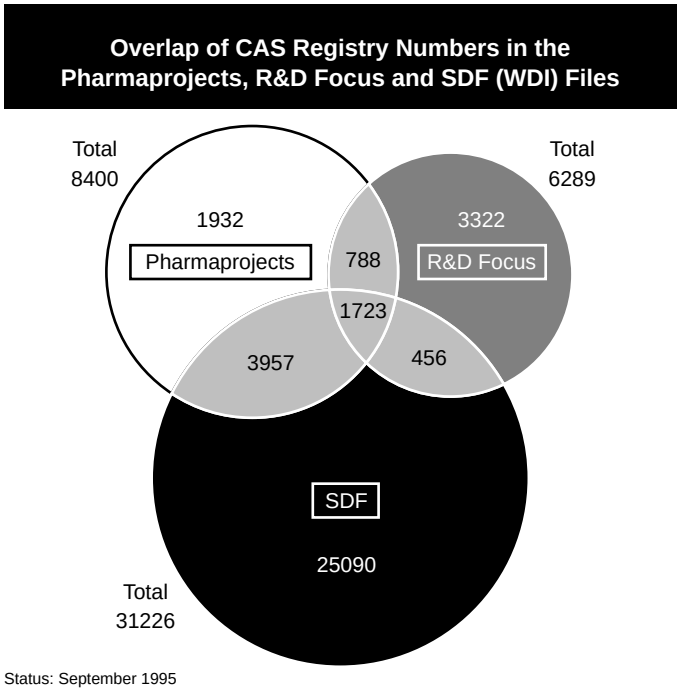

Fig. 13. Overlap of CAS Registry Numbers in three selected files.

different files relative to this NME Express (November 1994 update) can be appreciated from Fig. 12.

Conference Fast Track covered 37 of the mentioned 64 test compounds, followed by Pharmaprojects with 29, R&D Focus with 21 and the SDF with five (in part, due to its four- to six-month updating frequency). Of the 37 compounds found by Conference Fast Track, 24 were also in Pharmaprojects and eighteen in R&D Focus, and so on.

This analysis stresses the complexity of trying to track the development of test compounds comprehensively – the information sources are varied and, so far, all have something unique to offer!

7. Coverage and overlap of CAS Registry Numbers in three of the selected commercial files

The Chemical Abstracts Registry Number [18] is the unique way to identify a development compound in its different guises, i.e. lead compound in a patent application, various test compound numbers, chemical names, trivial names, generic name, etc.

Unfortunately, despite having a defined chemical structure in many cases (e.g. SDF), the CAS Registry Number [18] coverage is seriously lacking (see Fig. 12). Only about 32,000, i.e. about 65% of the Standard Drug File records possess a CAS Registry Number, although almost all exhibit a chemical structure and were reported in the scientific literature – why? Pharmaprojects contains some 8,000 CAS Numbers for its around 18,000 records, while R&D Focus has an approximate 40% share of CAS Numbers. In these last two files, the chemical constitution of the development compound is frequently not known – but not always! In the Conference Fast Track file – in which the chemical structure is known in almost every case – there is not even one CAS Registry Number. This is an area urgently in need of improvement!

Despite the relatively low coverage of CAS Registry Numbers in the different files, it is interesting to look at their overlap in the files – Pharmaprojects, R&D Focus and SDF. From approximately 46,000 CAS Registry Numbers, only about 1,700 are common to the three files (Pharmaprojects, R&D Focus and the SDF) shown in Fig. 13.

R&D Focus has a surprisingly low degree of overlap with Pharmaprojects (about 2,500 CAS Registry

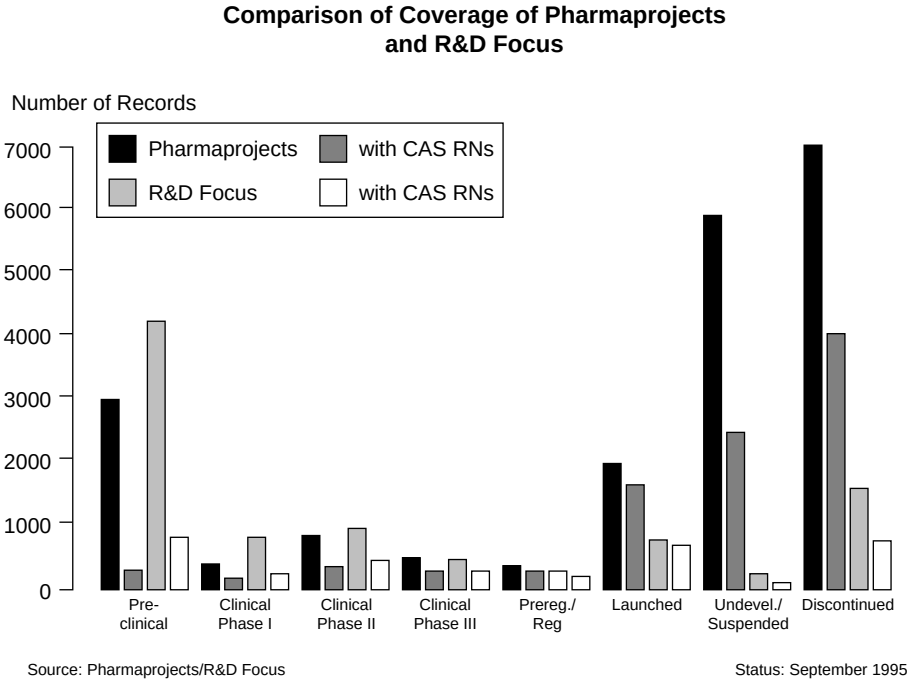

Fig. 14. Comparison of coverage of Pharmaprojects and R&D Focus.

Numbers), while R&D Focus has more unique CAS Registry Numbers (3,300) than Pharmaprojects (1,900). This diagram (Fig 11) merely serves to re-stress the fact that, currently, one will probably need to use both of these sources.

Of course, there may be different versions of a basic chemical structure in the different files, i.e. an acid and its salts, for example, which will mean that different CAS Registry Numbers are employed.

There still remains the urgent need to assign more comprehensively CAS Registry Numbers to the records in all the databases discussed.

8. Brief comparison of the coverage of the Pharmaprojects and R&D Focus files

The next chart (Fig. 14) presents the spread of compounds in the different phases in the Pharmaprojects [15] and R&D Focus [14] files, as well as the CAS Registry Number content of the files per phase. In general, for the active phases, the R&D Focus coverage of CAS Registry Numbers is slightly better. The diagram also stresses the very extensive coverage of launched, suspended and discontinued products by Pharmaprojects.

The apparent dominance of R&D Focus in the preclinical phase (for reasons to be explained in Table 3) is possibly misleading. It is surprising in Fig. 14 that so many launched compounds – where the chemical

Table 3
Comparison of currency of entries in the preclinical phase of Pharmaprojects and R&D Focus

	Pharmaprojects		R&D Focus	
1995 entries	931		578	
1994 entries	1,014	1,945	1,440	2,018
1993 entries	429		776	
1992 entries	242		414	
1991 entries	53		34	
1990 entries	24	763	394	2,026
1989 entries	15		38	
>1989 entries	∅		62	
Others*	47		3	
Total entries	2,755		4,047	

*no dates specified or input errors
Status: 8/1995.

Table 4
Coverage of selected renin inhibitors in Pharmaprojects, R&D Focus and R&D Insight files

Compound	Company	Development Status		
Name		Pharmaprojects	R&D Focus	R&D Insight
A-64662	Abbott	No development reported	Suspended	CP II
CL-331049	AHP	Preclinical	–	Preclinical
FK-906	Fujisawa	CP II	CPII	CPII
GR-70982	Glaxo	No development reported	–	Preclinical
S-2864	Hoechst	No development reported	Preclinical	No development reported
JTP-4761	Japan Tobacco	Preclinical	–	–
CP-80794	Pfizer	Discontinued	CP I	CP II
Ro-42-5892	Roche	No development reported	Discontinued	CP II
SC-56526	Searle	Preclinical	–	Preclinical
U-71038	Upjohn	No development reported (CPII)	CP I	Discontinued
PD-132002	Warner-Lambert	Discontinued	Preclinical	Preclinical
YM-26365	Yamanouchi	–	–	Preclinical

Status: 7/1995 Pharmaprojects/R&D Focus/R&D Insight.

structure is well documented – do not exhibit a CAS Registry Number.

As we have seen, a first naive comparison of the mere numbers of development products in the preclinical phases elicits the impression that R&D Focus with its around 4,000 entries (products) is much superior to Pharmaprojects with its around 2,800 entries.

However, if we use the history field to compare when records (products) last experienced a significant update/advance, we see that, in the case of Pharmaprojects, about 760 stem from the period 1989–93; the corresponding figure for R&D Focus being about 2,000. Both files therefore exhibit roughly the same number of relatively recent entries – around 2,000 – in the preclinical phase. Therefore, an overview of the current R&D activities of a company will probably be more accurately reflected by Pharmaprojects, or an appropriate search strategy must be used with R&D Focus to concentrate on recent entries/updates.

9. Coverage of some renin inhibitor development products by the Pharmaprojects, R&D Focus and R&D Insight files

Table 4 presents a random selection of twelve renin inhibitors from twelve different companies (A–Y) in the Pharmaprojects [15], R&D Focus [14] and R&D Insight [16] files.

The spread of the development status for a specific compound can be appreciated from Table 4. Pharmaprojects suggests that half the projects are either discontinued or show little signs of activity for eighteen months or more. This status is present only twice in the other two files. The development compound YM-26365 from Yamanouchi was only to be found in the R&D Insight file, while JTP-4761 from the Japan Tobacco Company was solely in the Pharmaprojects file.

Five test compounds, categorised as '*discontinued/no development reported*' in Pharmaprojects, are listed as still active in R&D Insight.

In only *one* of the twelve cases examined do the entries in the three selected files concur with one another. What is one to believe?

10. Some aspects of quality control in the selected files

10.1. Errors in spelling of 'Ciprofloxacin'

A major Bayer antiinfective product 'Ciprofloxacin' (including brand names) was taken and, using fuzzy searching, the number of misspellings in some of the

Table 5
Misspellings of 'Ciprofloxacin' present in selected files

		No. of Records
Conference Fast Track	5%	140
NME Express	0%	4
Pharmaprojects	0%	47
R&D Focus	9%	56
SDF/WDI	0%	11

selected files was examined (see Table 5). R&D Focus had a 9% error rate, followed by Conference Fast Track with 5%. The obvious question arises: why are not more sophisticated and effective spelling checker programmes being used? In this case, NME Express, Pharmaprojects and the SDF exhibited zero errors over all fields. As Pharmaprojects [15] contains more data than NME Express [17] or the SDF [13], this clearly speaks for their quality assurance (QA) checks.

10.2. Systematic treatment of the term 'Calcium Antagonist'

If the systematic treatment of a relatively simple, well-established concept such as 'Calcium Antagonist' is examined in the therapeutic use/mechanism of action fields of the selected files, four terminology variations (cf. Table 6) are to be found. In addition, the term is present:

- five times in the 'Mechanism of Action' field;
- three times in the 'Therapeutic Use' field; and
- once in a hybrid field (Therapeutic Use/Mechanism of Action).

There are four different coding systems employed:

- sometimes in Mechanism of Action field; and
- sometimes in Therapeutic Use field.

The SDF [13] and R&D Focus [14] files have two spelling variants of this term within the same records and, even then, their frequency of occurrence is not even 1:1 in these records!

How is the poor unsuspecting end-user to cope with this situation? Particularly, as the calcium antagonist concept is an established pharmacological activity, i.e. one of the simpler examples. A more clean-cut system is needed instead of the present diversity of terms, codes and their assignments.

With these last few diagrams, the need for better quality control and more standardised indexing, as well as a uniform approach to the terminology used over the different files, has been stressed.

Table 6
Treatment of the term 'Calcium Antagonist'

File	Field		Term
(Producer)	Mechanism of Action	Therapeutic Use	
Conference Fast Track (Derwent)	+	–	CALCIUM-ANTAGONIST
NME Express	+		CALCIUM-ANTAGONIST
(Prous)			Code = 31500
Pharmaprojects	+	–	CALCIUM CHANNEL
(PJB)			ANTAGONIST
			Code = CHA-CA-AN
R&D Focus	+		CALCIUM-ANTAGONIST
		+	CALCIUM-ANTAGONISTS
(IMS)			Code = C8
R&D Insight	+		CALCIUM-CHANNEL
			ANTAGONISTS
(ADIS)		+	
		(Codes)	
SDF	+		CALCIUM-ANTAGONIST
(Derwent)		+	CALCIUM-ANTAGONISTS

Status: September 1995.

11. End-user wishes

After having examined the main features of these competitor tracking files, a number of aspects came to our attention, which are compiled below.

(1) The currency issue has already been mentioned and will not be additionally stressed here; however, how often are company names updated, e.g. with takeovers, mergers, etc, or 'old' compounds removed from their active status or compounds from recent conferences included in R&D Focus/ Pharmaprojects?
(2) The reliability of the data (including quality control) is high on the agenda, including the accuracy of the chemical structures.
(3) There is a genuine lack of chemical structures, even when specific chemical nomenclature is available.
(4) A better CAS Registry Number [18] coverage is essential in order to readily gain further information from other sources, as well as to identify the same development compound in the different files. The present CAS Registry Number coverage is (as already shown) sadly lacking; in some files, it is not even present!
(5) The use of many of these files for competitor analysis via statistical techniques is often doomed to failure due to the lack of a systematic approach by the producer. The Pharmaprojects file [15] is one of the very few exceptions to this rule.
(6) The poor end-user is confronted by a plethora of classification systems (if even present!). The development phases, company names, therapeutic use of a development compound, its pharmacological mechanism of action, etc, are all treated in a variety of ways. There are some soundly based classification systems being used. However, due to the different approaches by the various database producers (even varying in their different files!), this results in confusion and frustration for any end-user let loose on these files.

A system for therapeutic use and mechanism of action terminology must be standardised over all files. While each producer could still maintain its own special system to 'enhance' its product, a standard system would facilitate multiple use of these files by end-users and professional searchers alike.

12. Unfulfilled needs

Finally, what are the unfulfilled needs?

Firstly, take account of end-users by keeping things simple, including, as already stressed, the introduction of a uniform system for company names, therapeutic uses, mechanisms of action terminology, etc.

The PDR [12], as a body consisting of some 27 major R&D-based pharmaceutical companies, is more than willing to discuss remedies to this situation with *all* interested parties.

Secondly, more value-added information is needed and not just a mere documentation of events, whereby, as shown, even that is a difficult matter. What is a significant advance and what is merely run-of-the-mill? Provide facilities to make a ready, but well founded, comparison of products feasible. Other important enhancements could be:

- inclusion of incidence and prevalence data, discussion of marketing aspects;
- turnover estimates/projections (regularly updated), etc.

13. Concluding remarks

The need for high quality, reliable, coherently indexed information on pharmaceutical development products is still very high. The competition among database producers to service this market is very evident. Those commercial providers that are most likely to succeed and to continue to be successful are those who are acutely aware of the actual requirements of their users and potential users in the pharmaceutical R&D community. To date, no one file covers these requirements and users are obliged to utilise a range of files to cover their needs.

References and notes

[1] W.C. Hou, L.K. Sheang and B. W. Hidajat, *Sun Tzu – War and Management: Application to Strategic Management and Thinking* (Addison-Wesley, Singapore, 1991).
[2] A. Archer, A. Mullen and W.R. Pickering, The pharmaceutical industry. In: R.T. Bottle and J.F.B. Rowland (eds), *Information Sources in Chemistry* (Bowker-Saur, London, 1993), pp. 251–278.
[3] B.H. Desai and D. Bawden, Competitor intelligence in the pharmaceutical industry: the role of the information professional, *Journal of Information Science* 19(5) (1993) 327–338.
[4] R.G. Halliday, S.R. Walker and C.E. Lumley, *Journal of Pharmaceutical Medicine* 2 (1992) 139–154.
[5] A. Mullen, M. Blunck, T. Busch and E. Möller, *Trend and Competitor Analysis in the Pharmaceutical Area* (196th National ACS Meeting, 25–30 September 1988).
[6] M. Blunck, W.D. Busse, G. Meister, E. Möller, A. Mullen and L.A.A. van Rooijen, Online patent searching – rapid

answer to critical questions, *Naturwissenschaften* 76(3) (1993) 96–98.
[7] W. Warr, *Chemistry and Industry* 15 (1995) 617.
[8] A. Mullen, M. Blunck, T. Busch and E. Möller, *Drug Information Files for Inhouse Use* (Derwent Literature Subscriber Meeting, 1990), pp. 45–59. [Private publication.]
[9] T. Busch, M. Blunck, A. Mullen and E. Möller, Comparison of scope and application of currently available drug information files. In: *Proceedings of the Montreux 1990 International Chemical Information Conference* (Springer-Verlag, Berlin, 1990), pp. 159–176.
[10] A. Mullen, *Aims and Activities of the PDR – Pharma Documentation Ring* (Report of the Derwent Literature Subscriber Meeting, 1992), pp. 41–56. [Private publication.]
[11] A. Mullen, Information transfer: new age – new ways. In: *Proceedings of the Third European Conference of Medical Libraries, Montpellier, 23–26 September 1992*, edited by S. Bakker and M.C. Cleland (Kluwer Academic Publishers, Dordrecht, Boston, 1993), pp. 343–346.
[12] Y. Dubosc and A. Mullen, Overview of the activities of the PDR (Pharma Documentation Ring), *Drug News and Perspectives* 7(9) (1994) 551–555.
[13] Producer: Derwent Information Ltd, Derwent House, 14 Great Queen Street, London WC2B 5DT, UK.
[14] Producer: IMS Global Services, 7 Harewood Avenue, London NW1 6JB, UK.
[15] Producer: PJB Publications Ltd, 18/20 Hill Rise, Richmond, Surrey TW10 6UA, UK.
[16] Producer: ADIS International Ltd, Chowley Oak Lane, Tattenhall, Chester CH3 9GA, UK.
[17] Producer: Prous Science Publishers, Apartado de Correos 540, 08080 Barcelona, Spain.
[18] Producer: CAS, 2540 Olentangy River Road, PO Box 3012, Columbus, Ohio 43210-0012, USA.
[19] Producer: MDL Information Systems Inc, 14600 Catalina Street, San Leandro, CA 94577, USA.
[20] The basic price covers one to five networked users. Exchange rate: £1.00 = $1.50 or DM 2.30 (approximately).