Preparative High-Performance Liquid Chromatography–Mass Spectrometry for the High-Throughput Purification of Combinatorial Libraries

Marcus Bauser*
Medicinal Chemistry VII, Business Group Pharma, BAYER AG, 42096 Wuppertal, Germany

Abstract

Accurate results for the testing of combinatorial libraries necessitates high purity of the library members. Therefore, combinatorial libraries derived from a combinatorial solution or solid-phase synthesis often require the purification of compounds that do not achieve a certain purity threshold. This study describes that preparative high-performance liquid chromatography (HPLC)–mass spectrometry (MS) is the method of choice for the purification of large arrays of diverse compounds. The adoption of this technology to the workflow of a solution phase combinatorial chemistry laboratory producing more than 20,000 compounds per year is described. Furthermore, the setup and logistics are discussed as well as the purity achievable for large libraries. Efficiency, speed, quality, and universality of preparative HPLC–MS are presented in detail for a library of 140 compounds, including data logistics and downstream processes as well.

Introduction

Genomics and proteomics provide a unique source of new targets for many kinds of diseases (1). After a new drugable target is discovered and an appropriate assay is established, high-throughput screening (HTS) and nowadays ultra-HTP are the key technologies for discovering small molecules interacting with this target (2). These screening technologies have inherently changed the drug discovery process during the last decade and have forced chemists to develop strategies for the synthesis of large numbers of compounds in less time. Consequently, combinatorial chemistry became popular by promising thousands of compounds in short time frames. Together with genomics and proteomics this technique represents a technology platform that enables scientists to investigate all kinds of diseases on the molecular level in a fast and efficient manner. Test results can only be as good as the library tested; therefore, strategies in combinatorial chemistry have changed tremendously from synthesizing millions of compounds as mixtures to the parallel synthesis of hundreds and thousands of distinguished well-characterized compounds in high purity. One reason for this change for example was that the deconvolution of large mixtures often failed or caused severe problems.

The lead finding/refinement process can be described as follows. The result of an HTS run mostly consists of a set of active compounds having activity in the submicromolar level. After analytics the activity of these hits is confirmed in more complex second line assays as well as in vivo models. This is the starting point for structural refinement in order to turn the lead compound into a development candidate. In this step further parameters such as toxicology and absorption distribution metabolism excretion parameters have to be optimized. Combinatorial chemistry plays two major roles in this step. First of all, a diverse library for initial screening has to be built up. Together with historical compound collections this is achieved by using high-throughput combinatorial synthesis. In the second part, the structure activity relationship (SAR) from singular hits or small hit sets has to be determined. This is done by synthesizing focused libraries that cover the diversity around the scaffold of the initial hit. These combinatorial libraries can either be prepared on solid support (3) or in solution (4); formats are usually microtiter plates (MTPs) or related footprints. Automation is used when appropriate, typical automated steps are sample distribution and reformatting using liquid handlers. Today, almost all reaction conditions (e.g., reflux, cooling, and inert atmosphere) can be applied even to large arrays of compounds. Library design addressing diversity (5) and biophysical parameters (6) can be applied to virtual libraries for selecting an optimal subset for synthesis. A must is the availability of diverse and exclusive educts to cover a diverse chemical space and produce exclusive unique compounds even by using simple reaction sequences. As mentioned previously, purity is crucial in terms of accurate SAR determination, and not too many tools are available for the purification of libraries. Liquid–liquid extraction is only applicable in cases of high-polar byproducts or excess

* Email: Marcus.Bauser.mb@Bayer-ag.de.

educts (7), and scavenging with polymeric reagents (8) is strongly educt-dependent. Chromatography using silica columns also has disadvantages because the columns are not inert, compounds should not be too polar, and complex sample preparation is necessary. In order to cope with these problems, reversed-phase chromatography is the method of choice for the purification of libraries.

Table I. Illustration of the Advantages and Disadvantages of Preparative HPLC Setup as a UV or Mass Triggered System

Preparative HPLC–UV	Preparative HPLC–MS
Advantageous	
Rugged	One injection, one fraction
Cheap	Identification online
Sharp peak shape	Deck capacity not critical
	Logistics easy to handle
Disadvantageous	
Each peak is a fraction	Some samples might be lost
Fraction collector capacity critical	Price
Complex logistics	Peak broadening

Preparative high-performance liquid chromatography (HPLC) has been used for years in peptide (9) and combinatorial chemistry (10). Columns, buffers, pumps, injectors, fraction collectors, and software are commercially available from many companies providing ready-to-use systems. Fraction collection is achieved using UV detection. Consequently, every UV-active compound can be collected, and each UV peak above a certain threshold is a fraction. At the time these instruments were introduced this was regarded as an advantage and not as a problem, but this turns into a problem when many samples have to be purified because this requires complex logistics for product identification and "cherry picking". In order to speed up this procedure Kassel et al. (11) developed online detection and fractionation of target molecules by using preparative HPLC–mass spectrometry (MS) in the laboratories of Combichem (San Diego, CA) in which this approach is successfully used for fast and efficient lead optimization (12). The main advantage of preparative HPLC–MS is that only the target molecule is selected by detecting the ion current of the desired target molecule. This means that only product-containing fractions are identified and collected, resulting in advantages such as "online" identification, easy-to-handle logistics, and noncritical fraction collector capacity. One disadvantageous is that if for any reason an existing product is not detected by MS, the sample is lost. Also, because of splitting and the nature of the MS detector, peak broadening is sometimes observed. This can lead to impure samples especially when fast chromatography is applied. However, there are still advantages of UV-based systems, such as ruggedness, price, and the optimal detector behavior in terms of peak shape. Additionally, these systems are more or less easy to handle.

This makes UV-based systems the systems of choice for small libraries and complex separation problems (see Table I). The biggest advantage of

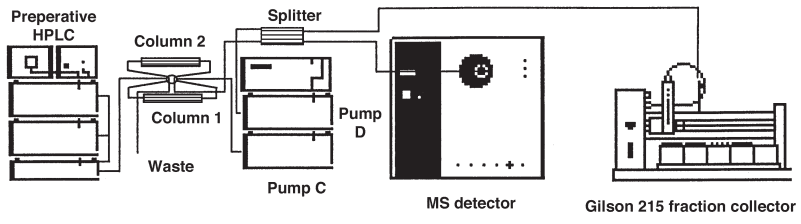

Figure 1. Setup of a preparative HPLC–MS system.

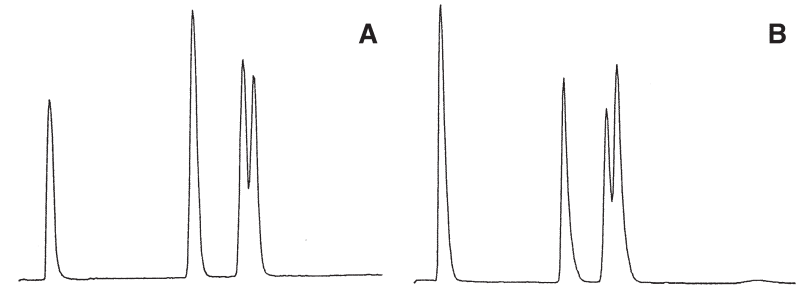

Figure 2. Separation of three compounds in (A) 10- and (B) 50-µmol amounts injected in DMSO.

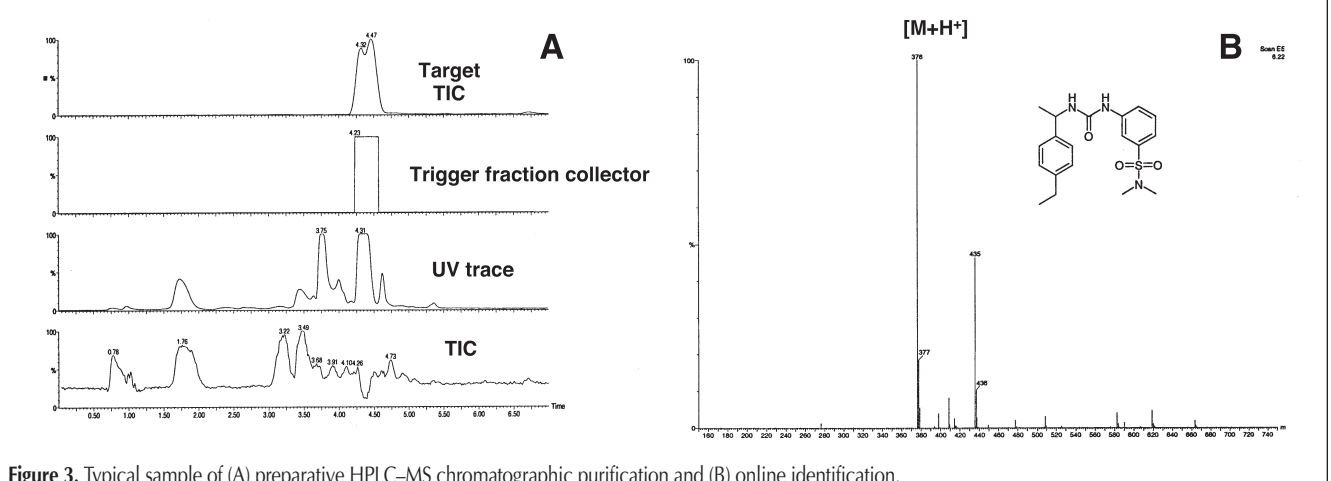

Figure 3. Typical sample of (A) preparative HPLC–MS chromatographic purification and (B) online identification.

preparative HPLC–MS is the one-injection, one-fraction purification procedure that solves many logistical problems normally encountered with large libraries.

In addition to the setup of a standard preparative UV system, an MS and at least one more pump is needed. These systems are commercially available from different suppliers (Gilson Abimed, Langenfeld, Germany; PerkinElmer, Weiterstadt, Germany; Shimadzu, Langenfeld, Germany; and Waters, Eschborn, Germany) and all have their pros and cons. Nevertheless, the setup is almost the same, which is described schematically in Figure 1. Differences include the software platforms used and some of the hardware components.

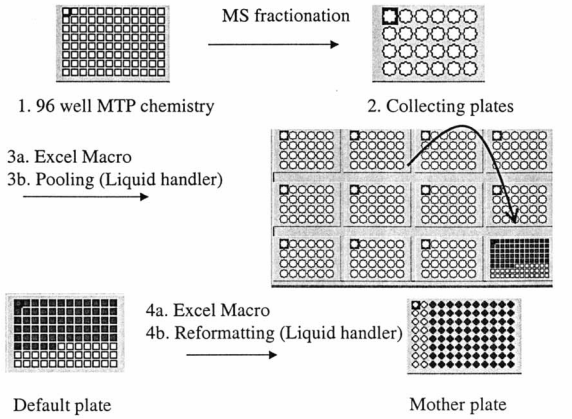

Figure 4. Downstream process encountered with preparative HPLC–MS schematically.

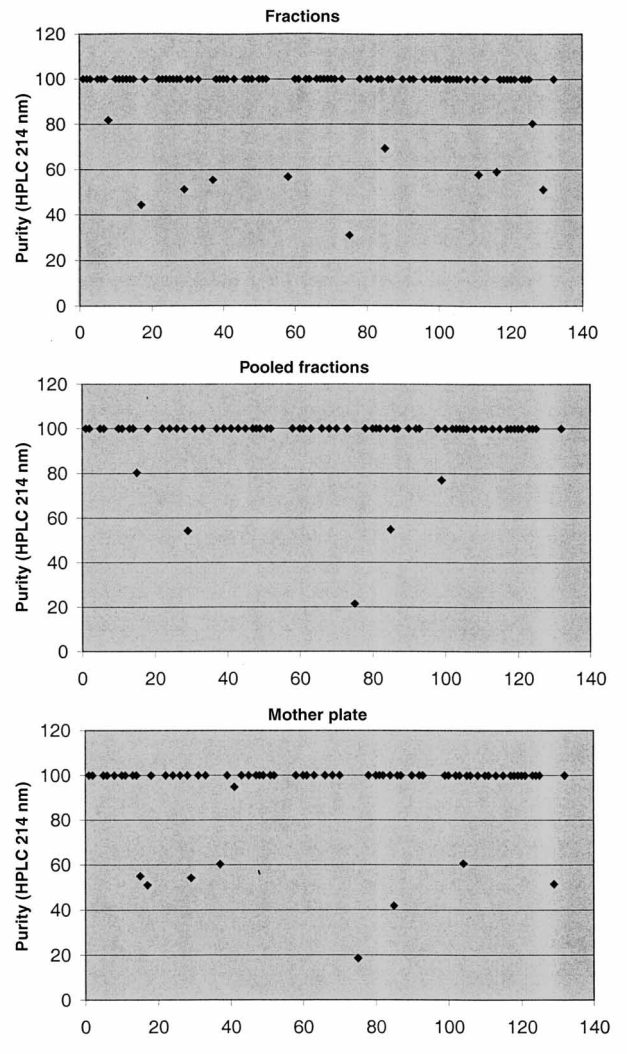

Figure 5. Purity check of a subset of 140 compounds on each process stage.

Experimental

HPLC–MS setup

The setup (shown in Figure 1) consisted of a preparative pump (Waters LC 4000) generating a gradient flow. This flow was pumped through a switching valve (Rheodyne, L.P., Rohnert Park, CA) that allowed two column switchings, thus achieving optimal throughput by saving time for column conditioning. After the column, the eluent ran through a UV detector (Waters 2487) and then into a special splitter device (LC packings). This splitter device diluted the sample by a factor of 1000 using the flow of pump D (Waters 215) to achieve a concentration suitable for MS characterization. The remaining preparative flow was pumped through a valve at the fraction collector (Gilson 215), which was only triggered when the target mass was determined to be above a certain threshold. Pump C (Waters 515) was used for equilibrating one column during the chromatography on the second column. For running such a system, a few more parameters have to be checked by comparing with a conventional UV-triggered system. The HPLC parameters flow rate, pressure, external pulses, and run time have to be optimized. MS parameters also have to be edited (e.g., mass or molecular composition of the target molecule, possible ion adducts and their intensity thresholds, lens voltages, and scan rates). The MS parameters have to be carefully evaluated before each run in comparison with a standard compound in order to ensure that no compound gets lost because of ion intensity reasons.

HPLC conditions

Two chromatography modes were applied depending on the chemistry project (a semipreparative setup with the goal of a 10-µmol product and a preparative setup with the goal of 50 µmol). Typical results of the chromatography and experimental conditions for both setups are outlined in Figure 2.

Because a gradient acetonitrile and water was used, all attempts to use methanol instead of acetonitrile failed mainly because of back pressure and peak shape. Different reversed-phase materials were tested, and depending on the physicochemical properties of the library, hydrophilic-endcapped or standard-endcapped C18 material was used. The particle size was found to be optimal between 5 and 7 µm, being that it was the best compromise between the mass loading and theoretical plate height. The 8- × 75-mm columns for the purification of 10 µmol were run with 12 mL/min and a 5-min 10% to 90% acetonitrile gradient. The 20- × 50-mm columns for the purification of 100 µmol were run with 40 mL/min and a 5-min 10% to 90% acetonitrile gradient. The column life time was determined as 500 to 1000 injections depending on the mixtures injected. Formic acid (0.05%) was added to the gradient system in order to obtain proper chro-

matography and good ionization of the target molecules in electrospray ionization (ESI) +/– mode. The lens voltages were optimized to smooth ionization in order to prevent fragmentation. This was a prerequisite because the molecular mass was the target mass for the fractionation. Fragmentation was also the reason for preferring ESI over atmospheric pressure chemical ionization (APCI). When using APCI, generally higher fragmentation was observed. A real sample is shown in Figure 3.

The UV trace as well as the total ion current (TIC) are shown. Based on the TIC of the target molecule, the fraction collector was triggered. This demonstrated very well the advantages of preparative HPLC–MS over preparative HPLC–UV. This example could lead to at least 8 fractions by collecting every UV-active component of the reaction mixture.

Logistics and data handling

The whole downstream process is schematically outlined in Figure 4. As already mentioned, sample logistics can be very user-friendly with this technology. In a combinatorial synthesis the molecules are defined by the applied chemistry and the educts used. This information is usually stored within a suitable database having all information for the purification process available. The information for the sample carrier containing the filtered reaction mixtures can be tracked via this database, and the necessary information for the preparative HPLC–MS can also be extracted. Usually, the molecular composition is sufficient. During the purification process the samples are fractionated into an MTP format of 24 or 48 wells, respectively (normally one but sometimes two fractions are obtained from one injection). These fraction collection plates are evaporated and the fractions redissolved in dimethylsulfoxide (DMSO). Meanwhile, the MS data is automatically processed and allows for precise sample tracking in the collecting plates. This information is transferred into an Excel VBA-Macro, which transforms the data into a run file for a liquid handler. The liquid handler pools the samples back into a 96-well format, and the final results are written into the entire database. Either 96- or 384-well MTP copies are formed, and the corresponding data can then be extracted for screening purposes. It is important to notice that data flow and data management can be a severe bottleneck in case they are not automated and capable of carrying all the information and processing the data in every stage. In summary, the data logistic framework has to store structures and synthesis information as well as MS data. Besides data storage the system has to provide run files for liquid handlers and track this information.

Results

By applying straightforward chemistry many compounds can be produced by liquid-phase combinatorial chemistry just by adding reactants A+B or by performing multicomponent reactions such as A+B+C+D. Preparative HPLC–MS can widen the bottleneck of the purification and characterization of these libraries. We have purified more than 20,000 compounds by using this logistically interlocked approach. Purity checks indicated at all times very pure samples, and the SAR turning out of those arrays were highly consistent compared with the resynthesized compounds. Purity was checked intensively for a subset of 140 compounds on each process stage. Figure 5 shows the results of this purity check. The first graph showed the purity of the fractions after evaporation from acetonitrile–water and the redissolving in DMSO before pooling. The purity was measured by analyzing the fractions with HPLC–MS and UV detection at 214 nm. Some of the raw fractions showed impurities that were caused by the very high dilution of some fractions. After pooling the fractions, only five samples out of 140 were less than 90% pure. The pooled samples were again evaporated and redissolved in DMSO before reformatting into the mother plate for screening purposes. We observed some thermal degradation, but overall only 8 out of 140 had to be discharged.

Conclusion

Preparative HPLC–MS is now routinely used for high-throughput purification for targeted as well as random libraries and has a major impact on the purity of our compound collections. Moreover, impure HTS hits are also purified by using this technology, leading to a fast evaluation of early hits by sorting out false positives. Preparative HPLC–MS is not yet as simple to use as its analytical counterpart, but for the final goal of the adoption of this technology it is intensively being worked on.

Acknowledgments

The author acknowledges the contributions of colleagues U. Rosentreter, H. Stakemeier, W. Bender, and other coworkers (namely, A. Grawinger, L. Jansen, C. Oschwald, and C. Wilkop).

References

1. S.R. Wiley. Genomics in the real world. *Curr. Pharmaceut. Design* **4:** 417–22 (1998).
2. T. Mander. Beyond uHTS, ridiculously HTS? *Drug Disc. Today* **5:** 223–25 (2000).
3. R.G. Franzen. Recent advances in the preparation of heterocycles on solid support: A review of the literature. *J. Combinat. Chem.* **2:** 195–214 (2000).
4. C.M. Sun. Recent advances in liquid phase combinatorial chemistry. *Combinat. Chem. High Throughput Screening* **2:** 299–318 (1999).
5. P. Willett. Chemoinformatics—similarity and diversity in chemical libraries. *Curr. Opin. Biotechnol.* **11:** 85–88 (2000).
6. F. Darvas, G. Dorman, and A. Papp. Diversity measures for enhancing ADME admissibility of combinatorial libraries. *J. Chem. Info. Comput. Sci.* **40:** 314–22 (2000).
7. S.X. Peng, C. Henson, M.J. Strojnowski, A. Golebiowski, and S.R. Klopfenstein. Automated high-throughput liquid–liquid extraction for initial purification of combinatorial libraries. *Anal. Chem.* **72:** 261–66 (2000).
8. H.N. Weller. Purification of combinatorial libraries. *Molecul. Divers.* **4:** 47–52 (1998).
9. J.S. Hamada, A.M. Spanier, J.M. Bland, and M. Diack. Preparative separation of value added peptides from rice bran proteins by high-

performance liquid chromatography. *J. Chromatogr. A* **827:** 319–27 (1998).
10. L. Schultz, C.D. Garr, L.M. Cameron, and J. Bukowski. High throughput purification of combinatorial libraries. *Bioorgan. Medicin. Chem. Lett.* **8:** 2409–14 (1998).
11. L. Zeng, L. Burton, K. Yung, B. Sushan, and D.B. Kassel. Automated analytical/preparative high-performance liquid chromatography mass spectrometry system for rapid characterization and purification of compound libraries. *J. Chromatogr. A* **794:** 3–13 (1998).
12. J.P. Williams and K. Lavrador. A solution-phase combinatorial synthesis of selective dopamine D-4 ligands. *Combinat. Chem. High Throughput Screening* **3:** 43–50 (2000).

Manuscript accepted February 25, 2002.