Front Loading Natural Product Screening Libraries for $\log P$: Background, Development and Implementation

by **David Camp^a**, **Marc Campitelli^a**, **Anthony R. Carroll^a**, **Rohan A. Davis^a**, and **Ronald J. Quinn^{*a}**

^a) Eskitis Institute, Griffith University, Brisbane, Queensland 4111, Australia

(phone: +61-7-37356006; fax: +61-7-37356001; e-mail: r.quinn@griffith.edu.au)

In the period from January 1981 to December 2010, 1068 small-molecule new chemical entities (NCEs) were introduced of which ~34% were either a natural product or a close analogue. While this metric reflects the impact natural products have played in delivering new chemical starting points (leads) for the pharmaceutical industry, it does not capture the decline this approach has suffered over the last 20 years as the high-throughput screening (HTS) of pure compound libraries has become more popular. An impediment to natural product drug discovery in the HTS paradigm is the lack of a clear strategy that enables front-loading of an extract or fraction's chemical constituents so that they are compliant with lead- and drug-like chemical space. To address this imbalance, an approach based on lipophilicity, as measured by $\text{clog } P$ has been developed that, together with advances being made in isolation and structural elucidation, can afford natural product leads in timelines compatible with pure compound screening.

Keywords: natural product drug discovery, $\text{clog } P$, lipophilicity, extract, high-throughput screening

Introduction. – Small-molecule drug discovery has changed significantly over the past two decades. The previous paradigm in which tens-to-hundreds of compounds were tested in animal models was replaced in the early 1990s by the high-throughput screening (HTS) of megalibraries against modified cell lines or isolated biomolecular targets produced by recombinant proteins [1]. These early libraries were typically comprised of molecules that could be produced in a highly efficient manner using combinatorial chemistry techniques [1].

However, the early period of the HTS era was not without controversy [2]. The vast numbers of compounds that could be efficiently screened was seductive and resulted in unrealistic expectations being placed on combinatorial chemistry and HTS to reverse the lack of new chemical entities (NCEs) emerging from company pipelines. When this did not eventuate, the focus rightly shifted towards increasing the *quality* of both the libraries, assays and supporting infrastructure [2]. For libraries in particular, this readjustment was driven, in part, by a better understanding of the types of molecules that should be screened and the end point that resulted in a drug [3].

A turning point for the pharmaceutical industry occurred in 1997 when Lipinski and co-workers reported that the combinatorial chemistry libraries then being screened were simply not comprised of ‘drug-like’ molecules [4]. They identified four simple physicochemical properties common to the 90th percentile of more than 2000 drugs and candidate drugs achieving phase II clinical status and, from this, formulated the ‘Rule of 5’ (Ro5), so-called because the maximum value for each parameter is a multiple of five: molecular weight (MW) < 500; H-bond donors (HBD) < 5; H-bond acceptors (HBA) < 10 and; the calculated log *P* (clog *P*) < 5. In essence, the Ro5 is a set of guidelines that identifies key properties for orally administered drugs while, at the same time, highlighting potential bioavailability issues if two or more violations occur [4].

While the Ro5 recognizes the end point of a drug discovery program, it is the identification of lead compounds, the chemical starting points for further development, that is considered to be critically important [5]. Much emphasis is now placed on addressing physicochemical properties early in contemporary drug discovery [5] and, as a result, many companies routinely screen libraries comprised

of molecules with profiles below those dictated by the Ro5. The HTS of ‘lead-like’ [6] or ‘reduced complexity’ [7] sets facilitates downstream chemical optimization which typically adds mass and lipophilicity *en route* to the final product [5].

Sometime later, Leeson and Davis proposed that properties remaining consistent over time were actually the most important to be considered [3a]. Their analysis of 864 drug approvals prior to 1983 compared with 329 between 1983-2002 revealed that drug lipophilicity, as estimated by $\text{clog } P$, changed less than other physicochemical properties and concluded $\log P$ was the most important parameter. This was not completely unexpected as the role of $\log P$ in influencing drug potency, pharmacokinetics and toxicity had already been established [8].

Despite the fact natural products have remained a fruitful source of lead molecules for the pharmaceutical industry, their popularity has waned considerably over the past 20 years [9]. One of the reasons often cited for disbanding natural products drug discovery has been that the time and resources required to identify leads, generally via iterative bioassay-guided fractionation following screening of extract or prefractionated libraries, is not competitive with the screening of pure compound libraries [10]. However, unlike the pharmaceutical industry that embraced sweeping reforms and is now reaping the benefits [1] [11], the natural product community has seemingly not learned, or applied, these invaluable lessons to its own work streams.

By way of example, classical bioassay-guided fractionation, a mainstay of natural product drug discovery, has only been able to undertake physicochemical profiling *after* compounds have been isolated and structures elucidated. Until now, there has been no strategy that aligns (*i.e.* front loads) a natural product screening set with lead- and drug-like properties to assist the downstream triage of biological and chemical data to expedite isolation of actives in timelines approaching that of pure compound libraries. Clearly, an approach based on $\log P$, for example, would advance natural product drug discovery within the HTS paradigm as the molecules would automatically comply with the most important of the physicochemical properties and therefore have a higher probability of progressing from hits to leads, even though their structures may not be known at the time of screening. Fortunately for

natural product drug discovery log P lends itself to a chromatographic-based protocol that facilitates construction of a screening library which addresses the lead- and drug-likenesses of its constituents early in the work stream. Herein, we provide details pertaining to the development and implementation of a log P filter for natural product screening extract libraries.

Results and Discussion. – *Relationship Between log P and Retention Time on a C_{18} HPLC Column.*

Before analyzing extracts, the correlation between the retention time of 20 drugs or bioactives of predominately natural product origin and a diverse set of 115 known natural products on a C_{18} HPLC system adapted to handle dimethyl sulfoxide (DMSO) solutions was determined against log P [12], clog P and clog $D_{7.4}$ [Table 1/ Fig. 1 (drugs and bioactives); Fig. 2 (115 natural products)]. While log P is a key measure of a molecule's lead- or drug-likeness in its *unionized* form, log D is also insightful because many drugs ionize under physiological conditions [13]. The clog $D_{7.4}$ is particularly useful as drugs are carried via the bloodstream which is tightly regulated at pH = 7.4 [13].

Table 1. log P , clog P and clog $D_{7.4}$ data for 20 drugs of predominately natural product origin.

Fig. 1a-c. Relationship between C-18 retention time and (a) log P ; (b) clog P (Chemaxon); (c) clog $D_{7.4}$ (ChemAxon) for 20 drug and bioactive reference compounds

Fig. 2a-c. Relationship between C-18 retention time and (a) log P ; (b) clog P (Chemaxon); (c) clog $D_{7.4}$ (ChemAxon) for 115 isolated natural products

Like others [12-13], we observed a relationship between the retention of small organic molecules in immobilized octadecanyl chains and log P . Thus, most compounds eluting up to 25 minutes on the HPLC system generally had log $P \leq 5$ while those eluting after 25 minutes tended to be higher (Fig. 1a, Fig 2a). A similar trend was observed for both the calculated parameters, clog P and clog $D_{7.4}$ (Fig. 1b-

c, Fig. 2b-c) allowing the predicted values to be used with a high degree of confidence where empirical data was not available. Interestingly, the plot of experimental data, *i.e.* $\log P$, was tighter compared with both the calculated sets using a number of different algorithms and may indicate that the training sets in many commercial packages do not yet contain sufficient scaffolds/combinations of functional groups found in natural products.

Comparison of SPE Sorbents Using Crude Extracts. Having established the desirable upper limit of $\log P$ on our HPLC system, we next investigated the ability of various SPE sorbents to retain highly lipophilic constituents in crude extract and, importantly, elute components with $\log P \leq 5$. Thus, 60 mg equivalents (mge), *i.e.* the extract derived from 60 mg of either a plant or marine invertebrate, in 500 μ l of eluting solvent was loaded onto various SPE sorbents and washed with a further 800 μ l of the same solvent. Four different types of sorbent matrices were tested: C₁₈-bonded silica, polystyrene, cross-linked poly(styrene-divinylbenzene) and cross-linked poly(*N*-vinylpyrrolidone-divinylbenzene). Sixteen plant and marine invertebrates were selected from the Griffith University Nature Bank collection for the initial studies (Table 2). The eluent from the load and wash cycles ($\sim$ 4-5 bed volumes depending on sorbent) was collected and analyzed for retention of lipophilic components using a dual diode array detector and evaporative light scattering detector system (Table 3).

The study of eluting solvents was limited to methanol:water (MeOH:H₂O) mixtures; acetonitrile was not considered for cost and environmental considerations. The *N*-vinylpyrrolidone-divinylbenzene (NVP-DVB) copolymer was the most effective sorbent for all eluting conditions and especially superior when trifluoroacetic acid (TFA/H⁺) was added to the eluent. We have observed that highly lipophilic/nuisance compounds like fatty acids, polyphenolics and sulfated steroids can be major constituents of crude extracts derived from plant and marine invertebrates. These molecules are maintained in the neutral form in the presence of acid which facilitates retention [14].

It was interesting to note that the two styrene-based sorbents did not retain as much of the unwanted lipophilic molecules as the NVP-DVB copolymer. This observation would suggest that screening sets derived from a similar polystyrene sorbent to that used here (*i.e.*, Diaion HP-20SS) [15], may contain

molecules that are not drug-like and typically manifest themselves as frequent hitters in biological assays. However, this is still preferred to screening crude extracts, particularly those derived from hexanes, which we and others [16] have found typically contain fatty acids and other lipophilic material.

Table 2. *Retention of lipophilic material on various SPE adsorbents.*

Refining the Eluting Solvent for the NVP-DVB copolymer. Given the NVP-DVB sorbent retained highly lipophilic components from crude extracts more strongly than other matrices, we next set out to refine the eluting conditions. Eleven reference compounds were selected from the pure natural product and drug sets spanning a $\log P$ range between -7.60 to 7.45, and to this was added two more natural products, eupomatenoid 3, and eupomatenoid 5, with $\log P \sim 5$. (Table 1). The molecules in this subset were used as both mixtures of pure compounds and as internal standards in crude extracts to strike an acceptable balance between recovery and selectivity of compounds with desired $\log P$ profiles.

When employed as a mixture of pure compounds, a good recovery of components with $\log P < 4$ was possible when 70:30 MeOH:H₂O/H⁺ (10 bed volumes) was used as the first solvent, and 90:10 MeOH:H₂O/H⁺ (10 bed volumes) as the second. Good to excellent recoveries were typically obtained for constituents in which the calculated and experimental $\log P < 5$ as shown in Fig. 3. Components in which the calculated and experimental $\log P > 5$, *e.g.* rapamycin ($\log P = 7.45$), were retained using this combination of sorbent and solvent (Fig. 3).

Fig. 3. *Elution profile of natural product drugs: vancomycin (A); brucine (B); metaprolol (C); reserpine (D); mycophenolic acid (E); taxol (F); capsaicin (G); and rapamycin (H). The red spectrum corresponds to a UV trace at 248 nm; black at 274 nm. The peak at 15 minutes is due to decomposition of vancomycin following drying at reduced pressure.*

Eluting conditions for crude extracts were next carried out using the above results from pure compounds as a guide. In the first set of experiments, 60 mge of two crude extracts known to contain high amounts of lipophilic material (*Table 2*) were loaded onto an SPE cartridge packed with 60 mg of an NVP-DVB copolymer. The loaded sorbent was then successively washed with MeOH:H₂O/H⁺ (2.5 ml) in 5% increments from 70% to 100% MeOH. Breakthrough of the lipophilic components was first observed in the 85% MeOH/H⁺ wash. It is important to note that these two crude extracts were chosen specifically for their highly lipophilic character so that the retentive properties of the NVP-DVB sorbent could be better gauged on samples that were not representative of the bulk collection. Minimal (0 to < 5%) breakthrough of constituents with log *P* > 5 was observed for the other 21 crude extracts used in this study containing less lipophilic material when the eluent contained up to 90% MeOH/H⁺. As with the reference compounds, a dual elution profile of 70:30 MeOH:H₂O/H⁺ (10 bed volumes) followed by 90:10 MeOH:H₂O/H⁺ (10 bed volumes) was the best combination of eluting conditions to capture earlier eluting components in good yield.

The preliminary experiments augured well for scale up from 60 mge of extract to 300 mge extract. Thus, an additional six samples (*Table 2*) were added to the set and all 23 biota extracted using the procedures described in the Experimental Part (*vide infra*) and loaded onto SPE cartridges packed with 300, 400 and 600 mg of the NVP-DVB copolymer. We found that 4.5 ml of the load solvent (70:30 MeOH:H₂O/H⁺) was necessary to solubilize the crude extracts. An additional 8 ml of load solvent (*i.e.* 12.5 ml total, ~10 bed volumes) was used to afford the first lead-like enhanced (LLE) extract. The second LLE extract was obtained using 90:10 MeOH:H₂O/H⁺ (12.5 ml). Under these conditions, 400 mg of sorbent proved sufficiently effective at retaining later eluting material and avoiding breakthrough.

The sorbent and dual solvent system was next analyzed using the six extracts spiked with brucine, theophylline and clotrimazole. Each reference compound was recovered in the same quantity when combined with an extract as in pure form. This result prompted a further set of experiments in which crude extracts were spiked with different reference compounds that did not co-elute with innate constituents. In all cases, the spiked extracts afforded an initial LLE extract comprised entirely of early

eluting material, *i.e.* components that elute around the 20 minute mark on the HPLC analysis system. The second LLE extract captured additional quantities of the early eluting components as well as components that eluted up to ~25 minutes (*Fig. 4*). The second extract was also observed to afford minor amounts (< 5%) of the lipophilic (post 25 minute) reference compound, eupomatenoid 5 (26.5 minutes), but none of eupomatenoid 3 (29 minutes). Interestingly, rapamycin (25.5 min) was retained by the sorbent when analyzed as a mixture of pure compounds. The retention of some molecules and not others on the NVP-DVB sorbent, as gauged by the HPLC data around the 25-26 minute mark, suggests that the structure of the individual components and/or nature of other constituents in the crude extract can affect the outcome.

In summary, the first LLE extract is optimized for constituents with $\log P < 5$ but attains this result with a lower yield of desired compounds. The second LLE extract increases yield at the expense of minor amounts of lipophilic material being present.

Fig. 4. Elution profile of: (a) crude extract derived from Botrylloides perspicuum; (b) spiked with brucine (B), eupomatenoid 5 (E5) and eupomatenoid 3 (E3) reference compounds; (c-d) eupomatenoid 3 is retained but a minor amount of eupomatenoid 5 is observed in the second optimized extract.

Before the project moved into full production mode, the number of samples was expanded (198 randomly selected biota) to gauge the protocol's performance against a more diverse set. It was noticed that some crude extracts (typically those derived from marine invertebrates of particular genera) did not fully solubilize in the load solvent resulting in formation of a quantity of precipitate prior to loading on the NVP-DVB sorbent. Thus, instead of solubilizing the crude extract directly in 70:30 MeOH:H₂O/H⁺ (4.5 ml), the MeOH/H⁺ portion (3.15 ml) was first added, with three aspirate and dispense cycles to facilitate solubilization, followed by addition of H₂O/H⁺ (1.35 ml). This alternate load procedure worked well for the majority of cases.

Applications for HTS and Drug Discovery. The scaled-up optimization protocol employed 300 mg of plant or marine invertebrate, and 400 mg of NVP-DVB copolymer to ultimately yield a set of 70000 natural product LLE extracts from 35000 biota resident in Nature Bank. The subsequent natural product library has since been used in two distinct ways (*Fig. 5*).

Fig. 5. *Workflow for production of the LLE extracts and subsequent screening.*

First, LLE extracts derived from 35000 plant and marine organisms were screened in duplicate to give a campaign of 140000 data points (*i.e.* $2 \times$ LLE extracts per biota sample = 70000 LLE extracts screened in duplicate) and the ensuing actives analyzed via a process coined ‘small scale bioprofiling’ [17]. Small scale bioprofiling provided information that could be used to determine if the active compound in an extract was likely to have conventional lead-like properties, to semi-quantify the potency of the bioactive compound and to predict the likely uniqueness of the bioactive component [17]. For each screening campaign, up to 200 active extracts were processed using a rapid automated C₁₈ HPLC chromatography protocol. The separated fractions (22 per extract) were then screened and the spectroscopic properties (NMR, MS and nitrogen content) of the active fractions analyzed. Screening of the active fractions allowed the partially purified fraction to be compared with the crude extract. Screening of LLE extracts in tandem with small scale bioprofiling led to an improvement in the delivery of lead-like compounds with 64% of screens yielding lead-like bioactive compounds [17].

More recently, a subset of 18453 LLE extracts from the 70:30 MeOH:H₂O/H⁺ subset have been prefractionated to afford 202983 screening points (*i.e.* 11 fractions per extract) with each well of the screening set containing 1-5 compounds [18]. The prefractionated LLE extracts have been screened against two parasitic diseases: malaria [19] and human African trypanosomiasis [20]; and two bacterial targets [21] to yield both new [19a-d] [20a-d] and novel [20e] structures. Compounds with nanomolar activity against specific strains of both parasites were ultimately isolated following screening [19d]

[19g] [20a] [20b] while one lead from the antibacterial assays has progressed into preclinical trials [21c]. Importantly, the set of isolated compounds from all four screening campaigns was enriched in lead- and drug-like molecules compared to the total entries of all reported natural products from the *Dictionary of Natural Products* [22] using the same Ro5 criteria [10] [18].

Acknowledgement. – We thank AstraZeneca for financial support to develop the optimized extraction procedure, and Waters Australia for allowing the bulk purchase of Oasis HLB. We also thank James Ebdon, Anthony Boyle, Caterina Ekenas and Stephen Toms for invaluable technical assistance.

Experimental Part

General. Twenty reference compounds within a clog P range between -7.60 and 7.45 were purchased from Sigma-Aldrich (*Table 1*). Eupomatenoid 3 and eupomatenoid 5 were isolated in-house and shown to be identical to previously reported material.[23] HPLC grade methanol (MeOH), dichloromethane (DCM) and *n*-hexane were purchased from Lab-Scan. Dimethyl sulfoxide (DMSO, 99.9%) and trifluoroacetic acid (TFA/H⁺, 99.5%) were purchased from Fluka and used without further purification. Water was purified with a Milli-Q system from Millipore. Dried plant and marine invertebrate samples were obtained from Griffith University's Nature Bank (*Table 2*). Alltech Davisil octadecanysilica (C₁₈-bonded silica); Supelco Diaion HP-20 (polystyrene); Supelco Amberlite XAD-16 [cross-linked poly(styrene-divinylbenzene) copolymer]; Waters Oasis HLB [cross-linked poly(*N*-vinylpyrrolidone-divinylbenzene) copolymer] and Alltech polyamide gel (PAG) were obtained direct from the vendor.

The calculated log P and calculated log $D_{7.4}$ for reference compounds and isolated natural products were performed with ChemAxon's Instant JChem version 3.0.4. The HPLC method developed by Lombardo *et al.*[12] was used to determine the experimental log P of the 115 natural products. Reference log P values for drugs and bioactive compounds were obtained from Sangster Research Laboratories except where stated (*Table 1*).

Stock Solutions. The 12 reference compounds used in protocol development were prepared at varying concentrations (mg/ml) in MeOH in an effort to obtain reasonably similar UV peak heights via a diode array detector: brucine (1), capsaicin (1), clotrimazole (5), metaprolol (1), mycophenolic acid (1), rapamycin (1), reserpine (1), paclitaxel (1), theophylline (1), vancomycin (15), eupomatenoid 3 (1), and eupomatenoid 5 (1). All stock solutions were stored in the dark at 4 °C following preparation. To facilitate fine tuning of the elution conditions through SPE, stock solutions (100 µl) were selected in an *ad hoc* manner, combined and dried. The resulting mixture was reconstituted with 100 µl DMSO prior to HPLC analyses (*vide infra*). A stock solution of brucine (0.1 mg/ml), dipyrimadole (0.1 mg/ml) and capsaicin (0.5 mg/ml) was prepared as an internal standard to monitor the production of optimized extracts for HTS.

Crude Extracts. Dried and milled biota samples (300 mg) were placed in an SPE cartridge fitted with a 20 µm frit and a cap applied. Plant biota were extracted at 1 ml/min with *n*-hexane (7 ml, discarded), followed sequentially by DCM (7 ml) and HPLC MeOH (13 ml). The extracts were collected and dried in the same test tube using a combined Martin Christ Beta-RVC centrifugal evaporator/Alpha 2-4 refrigeration unit. Dried extracts were reconstituted in MeOH (4 ml) and loaded onto a cartridge filled with 900 mg PAG using a Gilson XL4 to remove unwanted polyphenolic compounds [24]. The cartridge was subsequently washed with MeOH (2 × 4 ml) at 5 ml/min and the combined MeOH washes dried under vacuum to afford the crude plant extract. Marine invertebrates were extracted in a similar manner with *n*-hexane (7 ml, discarded), 80:20 DCM:MeOH (7 ml) and MeOH (13 ml). The vast majority of marine organisms do not contain polyphenolic compounds thereby obviating the need to filter extracts through PAG.

Comparison of Solid-Phase Extraction Sorbents. 120 mg of sorbent (Table 3) was pre-weighed into empty SPE cartridges fitted with a 20 µm frit. A second 20 µm frit was placed on top of the sorbent and a cap applied to facilitate automated SPE using the Gilson XL4. The sorbents were conditioned with MeOH and equilibrated with the eluting solvent immediately prior to use. TFA was added to MeOH:H₂O mixtures to make a final concentration of 1% during refinement of the eluting solvents.

Each biota specimen was extracted in batches of eight. The resulting dried, crude extracts were reconstituted in MeOH (5 ml) and combined. Eight aliquots (5 ml each) were then removed from the mix to maintain extract uniformity for the comparative study. One fifth (*i.e.* 1 ml) of each aliquot was tested on each sorbent and the remaining 1/5 retained as an untreated reference. Each batch of biota extracts thus afforded sufficient material to compare eight different eluting conditions on each of the four sorbents.

Sixteen extracts (eight DCM extracts derived from plants and eight DCM:MeOH extracts from marine invertebrates) were initially chosen for this study (*Table 2*). Eight of the extracts yielded lead- and drug-like molecules while the other eight afforded non drug-like, highly lipophilic compounds. A further seven biota (*Table 2*) were selected (one for its high level of lipophilic material; the other six from randomly selected plant and marine organisms) after the comparative SPE investigation had concluded in an effort to fine tune the eluting solvent and also undertake preliminary scale-up testing on a smaller subset. Finally, an additional 198 biota were analyzed to increase biological and chemical diversity, and further refine the eluent conditions for retention of high log *P* components prior to production of the natural product extract screening library.

HPLC Analysis. All extracts in the screening set were prepared as DMSO solutions to facilitate HTS. To maintain consistency, all extracts from the comparative SPE sorbent investigation were, likewise, prepared as DMSO solutions, while aliquots of the reference compounds were dried and reconstituted in DMSO prior to analysis. A Waters 600 preparative HPLC system consisting of a controller, pump, degasser, Waters 996 Diode Array Detector and Alltech Varex MKIII Evaporative Light Scattering Detector, and two Betasil C₁₈ (5 µm, 10 mm × 100 mm) column connected in series was used for all analyses. The first column was employed to retain reference compounds or extracts injected as DMSO solutions (50, 75 or 100 µl) while concomitantly removing residual DMSO. A constant flow of MeOH (0.8 ml/min) was pumped into one inlet of a metal T-piece while H₂O (10 ml/min) was pumped into the other inlet via a second pump. The outlet flow was eluted directly onto the first C₁₈ column. Once the DMSO was washed from the first column (5.5 min), a switching valve was activated, connecting the

first column in series to the second. The coupled columns were then eluted with a linear gradient from H₂O containing 1% TFA (H₂O/H⁺) to MeOH containing 1% TFA (MeOH/H⁺) over 17 min at 4 ml/min followed by isocratic elution with MeOH/H⁺ for a further 10 min at 4 ml/min.

Preparation of the Natural Product Extract Screening Library. Dried crude extracts derived from 300 mg of biota were solubilized in 70:30 MeOH:H₂O/H⁺ (4 ml), loaded onto an Oasis HLB SPE cartridge (400 mg) and eluted with a further 8.5 ml of the same solvent into a test tube using a Gilson XL4. A solution of 90:10 MeOH:H₂O/H⁺ (4 ml) was then added to the extract test tube, loaded onto the same SPE cartridge and eluted with a further 8.5 ml of the same solvent into a second tube. Both fractions were dried and reconstituted in DMSO (1200 µl) to afford extracts with a concentration of 250 µg/ul (*i.e.* each extract derived from 300 mg of biota). The resulting solution was transferred to a Thermo-Abgene microtube with a 2D barcode affixed to the base using a Perkin Elmer MultiProbe II liquid handler.

References

- [1] R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. S. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, U. Schopfer, G. S. Sittampalam, *Nat. Rev. Drug Discov.* **2011**, *10*, 188.
- [2] R. Macarron, *Drug Discov. Today* **2006**, *11*, 277.
- [3] a) P. D. Leeson, A. M. Davis, *J. Med. Chem.* **2004**, *47*, 6338; b) P. D. Leeson, A. M. Davis, J. Steele, *Drug Discov. Today: Technol.* **2004**, *1*, 189.
- [4] C. A. Lipinski, F. Lombardo, B. Dominy, P. J. Feeney, *Adv. Drug Del. Rev.* **1997**, *23*, 3.
- [5] P. D. Leeson, B. Springthorpe, *Nat. Rev. Drug Discov.* **2007**, *6*, 881.
- [6] a) S. J. Teague, A. M. Davis, P. D. Leeson, T. Oprea, *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 3743; b) T. I. Oprea, A. M. Davis, S. J. Teague, P. D. Leeson, *J. Chem. Inf. Comp. Sc.* **2001**, *41*, 1308.
- [7] M. M. Hann, T. I. Oprea, *Curr. Opin. Chem. Biol.* **2004**, *8*, 255.
- [8] a) C. Hansch, *Acc. Chem. Res.* **1969**, *2*, 232; b) M. S. Tute, in *Adv. Drug Res.*, Vol. 6 (Eds.: N. J. Harper, A. B. Simmonds), Academic, New York, USA, **1971**, pp. 1; c) H. van de Waterbeemd, D. A. Smith, K. Beaumont, D. K. Walker, *J. Med. Chem.* **2001**, *44*, 1313; d) M. T. D. Cronin, *Curr. Comp. Aided Drug Des.* **2006**, *2*, 405.
- [9] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2012**, *75*, 311.
- [10] D. Camp, R. A. Davis, E. A. Evans-Illidge, R. J. Quinn, *Future Med. Chem.* **2012**, *4*, 1067.
- [11] B. Munos, *Nat. Rev. Drug Discov.* **2009**, *8*, 959.
- [12] F. Lombardo, M. Y. Shalaeva, K. A. Tupper, F. Gao, M. H. Abraham, *J. Med. Chem.* **2000**, *43*, 2922.
- [13] X. Li, M. A. Cooper, *Anal. Chem.* **2012**, *84*, 2609.

- [14] a) D. Puig, D. Barceló, *Chromatographia* **1995**, *40*, 435; b) R. Wissiack, E. Rosenberg, M. Grasserbauer, *J. Chromatogr. A* **2000**, *896*, 159; c) A. Stafiej, K. Pyrzynska, F. Regan, *J. Sep. Sci.* **2007**, *30*, 985.
- [15] T. S. Bugni, B. Richards, L. Bhoite, D. Cimborá, M. K. Harper, C. M. Ireland, *J. Nat. Prod.* **2008**, *71*, 1095.
- [16] a) G. R. Eldridge, H. C. Veroort, C. M. Lee, P. A. Cremin, C. T. Williams, S. M. Hart, M. G. Goering, M. O'Neil-Johnson, L. Zeng, *Anal. Chem.* **2002**, *74*, 3963; b) T. A. Johnson, J. Sohn, W. D. Inman, S. A. Estee, S. T. Loveridge, H. C. Vervoort, K. Tenney, J. Liu, K. K.-H. Ang, J. Ratnam, W. M. Bray, N. C. Gassner, Y. Y. Shen, R. S. Lokey, J. H. McKerrow, K. Boundy-Mills, A. Nukanto, A. Kanti, H. Julistiono, L. B. S. Kardono, L. F. Bjeldanes, P. Crews, *J. Nat. Prod.* **2011**, *74*, 2545.
- [17] V. M. Avery, D. Camp, A. R. Carroll, I. D. Jenkins, R. J. Quinn, in *Comprehensive Natural Products II. Chemistry and Biology, Vol. 3* (Eds.: L. N. Mander, H.-W. Lui), Elsevier, Oxford, UK, **2010**, pp. 177.
- [18] D. Camp, R. A. Davis, M. Campitelli, J. Ebdon, R. J. Quinn, *J. Nat. Prod.* **2012**, *75*, 72.
- [19] a) M. S. Buchanan, R. A. Davis, S. Duffy, V. M. Avery, R. J. Quinn, *J. Nat. Prod.* **2009**, *72*, 1541; b) D. Mueller, R. A. Davis, S. Duffy, V. M. Avery, D. Camp, R. J. Quinn, *J. Nat. Prod.* **2009**, *72*, 1538; c) R. A. Davis, S. Duffy, V. M. Avery, D. Camp, J. N. A. Hooper, R. J. Quinn, *Tetrahedron Lett.* **2010**, *51*, 583; d) X. Yang, R. A. Davis, M. S. Buchanan, S. Duffy, V. M. Avery, D. Camp, R. J. Quinn, *J. Nat. Prod.* **2010**, *73*, 985; e) M. Xu, K. T. Andrews, G. W. Birrell, T. L. Tran, D. Camp, R. A. Davis, R. J. Quinn, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 846; f) X. Yang, Y. Feng, S. Duffy, V. M. Avery, D. Camp, R. J. Quinn, R. A. Davis, *Planta Med.* **2011**, *77*, 1644; g) R. A. Davis, M. S. Buchanan, S. Duffy, V. M. Avery, S. A. Charman, W. N. Charman, K. L. White, D. M. Shackelford, M. D. Edstein, K. T. Andrews, D. Camp, R. J. Quinn, *J. Med. Chem.* **2012**, *55*, 5851.
- [20] a) Y. Feng, R. A. Davis, M. L. Sykes, V. M. Avery, A. R. Carroll, D. Camp, R. J. Quinn, *Tetrahedron Lett.* **2010**, *51*, 2477; b) Y. Feng, R. A. Davis, M. L. Sykes, V. M. Avery, D. Camp, R. J. Quinn, *J. Nat. Prod.* **2010**, *73*, 716; c) R. A. Davis, O. Demirkiran, M. L. Sykes, V. M. Avery, L. Suraweera, G. A. Fechner, R. J. Quinn, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 4057; d) R. A. Davis, M. L. Sykes, V. M. Avery, D. Camp, R. J. Quinn, *Bioorg. Med. Chem.* **2011**, *19*, 6615; e) Y. Feng, R. A. Davis, M. L. Sykes, V. M. Avery, D. Camp, R. J. Quinn, *Tetrahedron Lett.* **2010**, *51*, 4847.
- [21] a) S. Yin, M. L. Sykes, R. A. Davis, T. Schelper, V. M. Avery, D. Camp, R. J. Quinn, *Planta Med.* **2010**, *76*, 1; b) F. Martin, T. Grkovic, M. L. Sykes, T. Shelper, V. M. Avery, D. Camp, R. J. Quinn, R. A. Davis, *J. Nat. Prod.* **2011**, *74*, 2425; c) S. Yin, R. A. Davis, T. Shelper, M. L. Sykes, V. M. Avery, M. Eloffsson, C. Sundinc, R. J. Quinn, *Org. Biomol. Chem.* **2011**, *9*, 6755.
- [22] *Dictionary of Natural Products on DVD*, Taylor and Francis Group / CRC Press, London, UK, **2011**.
- [23] B. F. Bowden, E. Ritchie, W. C. Taylor, *Aust. J. Chem.* **1972**, *25*, 2659.
- [24] Y. Tu, C. Jeffries, H. Ruan, C. Nelson, D. Smithson, A. A. Shelat, K. M. Brown, X.-C. Li, J. P. Hester, T. Smillie, I. A. Khan, L. Walker, R. K. Guy, B. Yan, *J. Nat. Prod.* **2010**, *73*, 751.

Tables

Table 1. $\log P$, $\text{clog } P$ and $\text{clog } D_{7.4}$ data for 20 drugs of predominately natural product origin.

Drug/Bioactive/NP	CAS number	Retention time	$\log P^a$	$\text{clog } P^b$	$\text{clog } D_{7.4}^b$
vancomycin ^c	1404-90-6	14	-0.84 ^d	-7.60	-4.86
theophylline ^c	58-55-9	15	-0.02	-0.81	-0.89
brucine ^c	357-57-3	17	0.98	0.61	-0.85
quinine	130-95-0	19	2.10	2.51	0.86
penicillin G	61-33-6	20	3.44	1.08	-2.29
metoprolol ^c	83-43-2	20	2.42	1.76	-0.47
clozapine	5786-21-0	20	2.99	3.52	1.10
tetracycline	60-54-8	20	-1.07	-0.20	-3.42
captopril	62571-86-2	21	0.18	0.73	-2.42
triamcinolone	124-94-7	21	1.16	0.24	0.24
reserpine ^c	50-55-5	21	3.32 ^d	3.53	3.14
dipyridamole ^e	58-32-2	21	2.23	1.81	1.81
verapamil	152-11-4	22	2.31	5.04	2.79
mycophenolic acid ^c	24280-93-1	22	2.76	3.53	0.17
erythromycin	114-07-8	23	2.54	2.60	1.57
clotrimazole ^c	23593-75-1	23	3.50	5.84	5.79
paclitaxel ^c	33069-62-4	23	4.50	3.54	3.54
capsaicin ^c	404-86-4	24	3.04	3.75	3.75
lovastatin	75330-75-5	25	4.04	3.90	3.90
rapamycin ^c	53123-88-9	25	n.a.	7.45	7.45
eupomatenoid 5 ^f	41744-28-9	26	6.15	4.88	4.87
eupomatenoid 3 ^f	41744-27-8	29	6.45	4.96	4.96

^a Data obtained from Sangster Research Laboratories, Montreal, Canada (<http://logkow.cisti.nrc.ca/logkow/index.jsp>) unless otherwise stated.

^b $\text{clog } P$ and $\text{clog } D_{7.4}$ values were calculated using Instant JChem (ChemAxon).

^c Reference compounds from original set of 20 drugs/bioactives/natural products used to develop protocols.

^d Data obtained from SRC, Syracuse, New York
(<http://www.syrres.com/what-we-do/databaseforms.aspx?id=386>).

^e Used for QC purposes.

^f Additional natural products added to the reference set used to develop protocols.

Table 2. *Taxonomy of biota samples used in this investigation.*

Kingdom	Phylum	Family	Species
Plantae ^a	Spermatophyta	Rubiaceae	<i>Gardenia</i> sp.
Plantae ^a	Spermatophyta	Sapindaceae	<i>Dodonaea</i> sp.
Plantae ^{a,b}	Spermatophyta	Sapindaceae	<i>Dodonaea</i> sp.
Plantae ^a	Spermatophyta	Cupressaceae	<i>Callitris</i> sp.
Plantae ^a	Spermatophyta	Dilleniaceae	<i>Dillenia</i> sp.
Plantae ^a	Spermatophyta	Cupressaceae	<i>Callitris</i> sp.
Plantae ^a	Spermatophyta	Lauraceae	<i>Beilschmiedia</i> sp.
Plantae ^a	Spermatophyta	Sapindaceae	<i>Cupaniopsis</i> sp.
Plantae ^{c,d}	Spermatophyta	Euphorbiaceae	<i>Mallotus</i> sp.
Plantae ^{c,d}	Spermatophyta	Sapindaceae	<i>Dodonaea</i> sp.
Plantae ^{c,d}	Spermatophyta	Rutaceae	<i>Geijera</i> sp.
Animalia ^a	Cnidaria	Tubiporidae	<i>Pachyclavularia violacea</i>
Animalia ^a	Porifera	Irciniidae	<i>Ircinia</i> sp.
Animalia ^a	Porifera	Thorectidae	<i>Carteriospongia foliascens</i>
Animalia ^a	Porifera	Irciniidae	<i>Ircinia</i> sp.
Animalia ^a	Cnidaria	Alcyoniidae	<i>Sinularia</i> sp.
Animalia ^a	Porifera	Agelasidae	<i>Agelas axifera</i>
Animalia ^a	Porifera	Dictyonellidae	<i>Acanthella</i> sp.
Animalia ^a	Porifera	Crellidae	<i>Crella</i> sp.
Animalia ^{b,c}	Porifera	Thorectidae	<i>Fascaplysinopsis</i> sp.
Animalia ^{c,d}	Porifera	Dysideidae	<i>Lamellodysidea herbacea</i>
Animalia ^{c,d}	Porifera	Ianthellidae	<i>Ianthella flabelliformis</i>
Animalia ^{c,d}	Chordata	Styelidae	<i>Botrylloides perspicuum</i>

^a Original 16 extracts used in SPE studies.

^b Additional biota extracts (one plant, one animal) containing lipophilic material used as referents for development of protocols to remove high log P components.

^c Additional 7 extracts added to set.

^d Extracts (6) used for protocol development, spiking experiments and initial scale-up.

Table 3. *Retention of lipophilic material on various SPE adsorbents.*

Adsorbent	Matrix	Eluent	% retention of lipophilics ^b
		% MeOH:H ₂ O ^a	(~ 4-5 bed volumes eluent)
C ₁₈ -bonded silica	silica	100:0 ^c	50
		80:20	75
		70:30	91
		80:20 (H ⁺) ^d	88
		70:30 (H ⁺)	94
Diaion HP-20	polystyrene	100:0 ^c	66
		80:20	52
		70:30	73
		80:20 (H ⁺)	87
		70:30 (H ⁺)	82
Amberlite XAD-16	polystyrene	100:0 ^c	72
		divinylbenzene	57
		copolymer	65
		80:20 (H ⁺)	82
		70:30 (H ⁺)	79
Oasis HLB	<i>N</i> -vinylpyrrolidone-	100:0 ^c	78
		divinylbenzene	94
		copolymer	100
		80:20 (H ⁺)	100
		70:30 (H ⁺)	100

^a Extracts were loaded onto SPE cartridges using 500 µl of solvent. A further 800 µl of the same solvent was then used to wash the sorbent.

^b Mean value from $n = 3$, *i.e.* 16 biota analyzed in triplicate for each combination of sorbent and eluent.

^c Extracts in MeOH (1 ml) were eluted through the sorbent in a one-step procedure.

^d H⁺ refers to 1% TFA in the eluent.

Reprinted (adapted) with permission from D. Camp, R. A. Davis, M. Campitelli, J. Ebdon, R. J. Quinn, *J. Nat. Prod.* 2012, 75, 72 Copyright © 2012 American Chemical Society.

Figures

Fig. 1a-c. Relationship between C-18 retention time and (a) $\log P$; (b) $\text{clog } P$; (c) $\text{clog } D_{7.4}$ for 20 drug and bioactive reference compounds of mainly natural product origin

Fig. 2a-c. Relationship between C-18 retention time and (a) $\log P$; (b) $\text{clog } P$; (c) $\text{clog } D_{7.4}$ for 115 isolated natural products

Fig. 3. Elution profile of natural product drugs: vancomycin (A); brucine (B); metaprolol (C); reserpine (D); mycophenolic acid (E); taxol (F); capsaicin (G); and rapamycin (H). The red spectrum corresponds to a UV trace at 248 nm; black at 274 nm. The peak at 15 minutes is due to decomposition of vancomycin following drying at reduced pressure.

Fig. 4. Elution profile of: (a) crude extract derived from *Botrylloides perspicuum*; (b) spiked with brucine (B), eupomatenoid 5 (E5) and eupomatenoid 3 (E3) reference compounds; (c-d) eupomatenoid 3 is retained but a minor amount of eupomatenoid 5 is observed in the second optimized extract.

Fig. 5. Workflow for production of the LLE extracts and subsequent screening.

Graphical Illustration