HHS Public Access

Author manuscript

J Chromatogr A. Author manuscript; available in PMC 2018 June 30.

Published in final edited form as:

J Chromatogr A. 2017 June 30; 1504: 46–54. doi:10.1016/j.chroma.2017.04.055.

Sweet spot matching: a thin-layer chromatography-based countercurrent solvent system selection strategy

Yang Liu[a,b], **J. Brent Friesen**[b,c,d,*], **Edyta M. Grzelak**[e], **Qingfei Fan**[a,f], **Ting Tang**[a,g], **Kemal Durić**[a,h], **Birgit U. Jaki**[a,e], **James B. McAlpine**[a,c,e], **Scott G. Franzblau**[e], **Shao-Nong Chen**[a,b,c,e], and **Guido F. Pauli**[a,b,c,d,e,**]

[a]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, USA

[b]UIC/NIH Center for Botanical Dietary Supplements Research, Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, USA

[c]Center for Natural Product Technologies (CENAPT), Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, USA

[d]Physical Sciences Department, Rosary College of Arts and Sciences, Dominican University, River Forest, IL 60305, USA

[e]Institute for Tuberculosis Research, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, USA

[f]Key Laboratory of Tropical Plant Resources and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, P.R. China

[g]Central Laboratory, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, P.R. China

[h]Department of Pharmacognosy, Faculty of Pharmacy, University of Sarajevo, Sarajevo 71000, Bosnia and Herzegovina

Abstract

TLC-based strategies were proposed in 1979 (Hostettmann et al.) and 2005 (Friesen & Pauli; GUESS method) to minimize the number of partitioning experiments required for countercurrent separation (CCS) solvent system selection. As semi-empirical approaches, both proposed that the K values defining the sweet spot of optimal CCS corresponded to a matching R*f* value range from the silica gel TLC plate developed in the organic phase of a biphasic or a corresponding monophasic solvent system. Despite their simplicity, there has been an absence of theoretical support and a deficiency of reported experimental evidence. The present study explores the theory

*Corresponding Author. Tel.: (708) 524-6972; Fax: (708) 524-5990. jbfriesen@dom.edu (J.B. Friesen). **Corresponding Author. Tel.: (312) 355-1949; Fax: (312) 355-2693. gfp@uic.edu (G.F. Pauli).

required to develop correlations between R*f* and K. All theoretical models surmise that the optimal R*f* value range should be centered at 0.5. In order to validate the feasibility of the concept of matching R*f* and K values, 43 natural products and six solvent system families were investigated. Out of 62 correlations, 45 resulted in matched R*f* and K values. Based on this study, practical guidelines for the TLC-based prediction strategy are provided. These approaches will equip CCS users with an updated understanding of how to apply the TLC-based solvent system selection strategy to accelerate a targeted selection of CCS conditions.

Keywords

countercurrent separation; solvent system selection; TLC retention factor; partition coefficient; sweet spot; GUESS

1. Introduction

Countercurrent separation (CCS) includes both centrifugal and gravitational chromatographic methods of stationary phase retention [1]. Centrifugal CCS as a modern liquid-only chromatography technique, was introduced in the 1970s [2]. It can be divided into hydrodynamic mode, i.e., countercurrent chromatography, and hydrostatic mode, i.e., centrifugal partition chromatography [1]. Because of the consistency of the partition coefficient (K value), the solvent system compatibility, and the reproducibility of column behavior, centrifugal CCS provides orthogonal and scalable separation capabilities [3]. Thus, it has been widely used for natural product research [4–11].

An appropriate solvent system (SS) is the key for a high partition efficiency [12]. However, the SS selection process tends to be a bottleneck for CCS applications, and hampers the wider implementation of CCS in chemical research involving preparative separations. Usually, it involves many partitioning experiments to screen for a SS that yields the appropriate K value(s) to deliver the target analyte(s) into the K value sweet spot of optimum resolution, which has been defined as $0.4 < K < 2.5$ in regular elution mode [13,14]. To minimize the tedious bench work required for SS selection, a series of biphasic SS selection strategies have been developed [3]. Among these, in 1979, Hostettmann et al. recommended a TLC-based SS selection strategy as the first step towards providing the CCS practitioner with a more rapid assessment of various SS formulations. In this method, the sweet spot of optimal CCS corresponded to a matching R*f* value range from the silica gel TLC plate developed in the organic phase of the biphasic SS. Inspired by this approach, the Generally Useful Estimate of Solvent Systems (GUESS) TLC-based strategy introduced in 2005 [14] avoided the need for equilibrated SSs altogether by using monophasic, standardized SSs. Employing TLC conditions, that are both polarity matched and commonly used for fraction monitoring, established a semi-empirical method in which TLC R*f* and CCS K values coalesce in a shared sweet spot. TLC-based CCS SS optimization methods are underutilized and require a stronger theoretical base.

Defining a range of practically matching R*f* values that are useful to predict the sweet spot K values, becomes the key in any TLC-based SS selection strategy. Until now, the matching R*f* value range was an empirically defined TLC “sweet spot.” One article [15] set it as the range

$0.4 < Rf < 0.6$; while another [13] suggested this range should be $0.29 < Rf < 0.71$. Both studies assumed that (i) the TLC sweet spot corresponds to a particular K value sweet spot range; and (ii) the resolution on TLC can be translated into the CCS performance. However, while numerous studies have used the GUESSmix selection of commercially available natural products, few applications of the TLC-based SS selection scheme have been reported [16]. Another area of concern is the observation that CCS selectivity, which depends only on liquid-liquid partitioning, cannot always be predicted from the TLC behavior [15]. This raises the question whether the defined matching of Rf and K value ranges are appropriate. To answer this question, both theoretical hypotheses and case studies were developed.

2. Experimental

2.1. Materials

All HPLC grade solvents and NMR solvents were from Sigma-Aldrich (St. Louis, MO, USA) and Cambridge Isotope Laboratories (Tewksbury, MA, USA), respectively. All analytical grade solvents were purchased from Pharmco-AAPER (Crookfield, CT, USA) and redistilled before use. Hexanes is a mixture of C_6H_{14} isomers having much the same chromatographic properties as *n*-hexane. Chemicals used, including the commercially available reference standards, were purchased from Sigma-Aldrich (St. Louis, MO, USA) or purified from their natural source in the UIC/NIH Center for Botanical Dietary Supplements Research.

2.2. Candidate Solvent System Prediction Using Thin-Layer Chromatography

Analytical TLC was performed on Alugram precoated 0.2 mm thick silica gel G/UV254 aluminum plates (12 × 20 cm; Macherey-Nagel, Düren, Germany) at 25 °C. Plates were cut to proper widths before spotting. The analytes were usually dissolved in methanol. Water or ethyl acetate was also used for more or less polar analytes, respectively.

For the bioautography experiment, samples were obtained from actinomycete fermentation extracts (the 100% methanol vacuum liquid chromatography fraction). Subsequent CCS fractions were applied to TLC glass-backed plates pre-coated silica gel G/UV254 (20 × 20 cm; No. 809023), manually or by using an Automatic TLC Sampler ATS4 (CAMAG, Muttenz, Switzerland) with the following application conditions: filling speed 15 μL/s, pre-dosage volume 200 nL, dosage application speed 150 nL/s, rinsing cycle one with methanol-water (9:1, v/v), rinsing vacuum time 4 s, filling vacuum time 1 s.

All spotted plates were dried before developing. The TLC developing solution was the organic phase of an equilibrated biphasic SS or the equivalent organic-only SS based on HEMWat and ChMWat families (SSE and SSC respectively) [13]. Then, the spotted TLC plate was developed in an appropriate SS in a TLC chamber or in a CAMAG Automated Multiple Development Chamber AMD 2 (CAMAG, Muttenz, Switzerland). The eluted TLC plates were observed under UV at 254 nm and 365 nm. If the analyte had no UV absorption, the general-purpose reagent *p*-anisaldehyde-sulfuric acid-acetic acid 1:1:48 (v/v) was sprayed on the TLC plates, the plates were drained, and then heated on a CAMAG TLC Plate Heater III. TLC experiments were carried out in duplicate.

2.3. Evaluation of Candidate Solvent System

2.3.1. Partitioning Experiments—An equilibrium between all interacting components of a biphasic SS was performed before separating the two phases. Equal volumes of upper and lower phases were mixed with a certain amount of a given analyte in a vial. After partitioning the two phases, the equilibrium concentrations in both phases were examined by UV or ^{1}H NMR.

2.3.2. Small-Volume CCS Experiments—Small-volume CCS was carried out using a Tauto 20 with a 16 mL coil volume (Tauto Biotech., Shanghai, China) equipped with a CherryOne system (Wrightwood Technologies, Chicago, IL, USA) [17], featuring a 500 μL sample loop, pump, phase metering apparatus, and UV detector. After mixing with equal volumes of upper and lower phases, the sample was loaded into the sample loop. The column was first filled with stationary phase. The column was rotated at 800 rpm, and the lower (mobile) phase was pumped into the column until dynamic equilibrium was attained. Subsequently, the sample was injected into the column through a six-port valve and elution started. Where indicated, the elution-extrusion mode was applied. The eluate was collected with a Foxy Jr. fraction collector (Teledyne Isco, Lincoln, NE, USA).

3. Theoretical Discussion

Normal-phase TLC uses silica gel as a stationary phase and various organic solvents as a mobile phase. The migration rate of an analyte is affected by both its solubility in the mobile phase and its affinity for the silica gel. In a normal-phase CCS, an aqueous alcohol phase approximates the polarity and selectivity characteristics of the TLC silica gel stationary phase. Normal-phase CCS holds the aqueous phase in the column while the mobile phase carries the analytes which partition between the two liquid phases by a series of mixing and settling stages. In both TLC and CCS, the eluotropic series of solvent “eluting strengths” is generally followed. The similarities of action in TLC and CCS are the theoretical foundation of the TLC-based CCS SS prediction strategy.

At the same time, differences between the two methods exist. One basic distinction between TLC and CCS is the nature of the mobile and stationary phase interactions. The rate of analyte migration in TLC is largely influenced by low surface area adsorption and desorption interactions. In contrast, the rate of analyte elution in CCS is governed almost completely by high surface area liquid-liquid partitioning, although the phase interactions with the tubing (CCC) or cell (CPC) walls may also be a factor. Another difference between CCS and TLC is that the nature of the silica gel stationary phase is rather static, whereas in CCS the composition of the stationary phase may be highly variable depending on the composition of the chosen SS.

In the GUESS method [13] the mobile phase in both chromatography systems has approximately the same composition. Indeed, it was proposed that a simplified, polarity matched organic phase composition could be substituted for the equilibrated organic phase in TLC [13]. The rationale for this simplification is that equilibration requires mixing the solvents of the biphasic SS, agitation, settling, and physically separating the two layers in order to prepare the TLC mobile phase. The simplified TLC eluting solution in the GUESS

method was found to give approximately the same results as those with equilibrated organic phases. The aim of the GUESS method was not to calculate K values based on TLC R*f* values, but simply to determine which SS may provide a good starting point for the SS selection process where K values are calculated based on observed partitioning. There has always been some concern that the simplified approach may not be adequate for SSs where one or more organic solvents is/are found in significant quantities in both phases of the equilibrated SS. This potential limitation led to the proposal that e.g., ethyl acetate containing SSs such as hexanes-ethyl acetate 3:7, v/v (SSE 7) corresponded to three HEMWat CCS SSs: 3:7:5:5, 3:7:4:6, and 3:7:3:7 (v/v).

3.1. The Correlations between R*f*, K, and a Solvent System

In order to devise a system of comparing R*f* and K values, the nature of these values should be understood. R*f* values in TLC result from the experimental measurement of two distances. R*f* values are rarely recorded for analytes in the literature due to their relatively variable nature and the limitations of the most widely used simplistic experimental conditions. In contrast, in CCS, the K value is a partition coefficient, a physicochemical constant that can be measured from the appropriate CCS experimental values and by partitioning experiments [17].

A practical consideration is that lipophilic analytes in normal-phase TLC receive higher R*f* values (near 1), while the same analytes in normal-phase CCS receive low K values (near 0). This can be compensated for by using the reversed-phase K values for CCS, which are simply (and remarkably) the reciprocal of the normal-phase K values [10]. This strict, mathematical reversibility cannot be achieved with TLC even though both normal-phase and reversed-phase media are available.

A much more serious consideration is the scale used for each value. TLC R*f* values are calculated as ratios between zero and unity. On the other hand, K values are calculated as ratios between zero and infinity. There are at least two methods of reconciling these values so that their scales have the same range. (i) The CCS elution scale may be placed in the zero to unity range by defining a partition retention factor (P*f*) as the concentration of the analyte in the stationary phase, divided by the sum of the concentrations in the mobile and stationary phases ($\mathrm{P}f=\frac{\mathrm{K}}{\mathrm{K}+1}$) [13]. (ii) Another approach is to compare R*f* values with logK values. This is also a reciprocal scale, as logK and $\log(\frac{1}{\mathrm{K}})$ are equidistant from unity. In this case, however, the resulting logK values are between negative and positive infinity with log0 being undefined.

For the second method, the possible range of R*f* and logK values are not the same, therefore, a factor must be introduced to correlate R*f* and logK. A linear correlation (*m*) between R*f* and logK, where R*f* ∝ logK, may be created by proposing that any three SSs (a, b, and c) with respective TLC R*f* and CCS K values will create a straight line with a slope that follows Eq. 1. A model behavior of natural mixtures in both chromatographic systems is shown in Supplementary Information S1.

$$m=\frac{\mathrm{R}f\mathrm{b}-\mathrm{R}f\mathrm{a}}{\log K\mathrm{b}-\log K\mathrm{a}}=\frac{\mathrm{R}f\mathrm{c}-\mathrm{R}f\mathrm{b}}{\log K\mathrm{c}-\log K\mathrm{b}} \quad \text{Eq. 1}$$

where the assumptions are that $\mathrm{R}f_c > \mathrm{R}f_b > \mathrm{R}f_a$ and $\log\mathrm{K}_c > \log\mathrm{K}_b > \log\mathrm{K}_a$.

3.2. K and R*f* Value Sweet Spots

"Sweet spot" is a term used in bat and racket sports to denote the area on a hitting implement that makes the optimal contact with the ball. Sweet spot was, therefore, used to describe the region of optimum separation in CCS [13]. There is evidence that the great majority of natural products isolated with CCS have a K value in the range of 0.4 to 2.5 in the solvent system used in their isolation. The information in Fig. 1 was gathered from reports of natural product separations where the K values of the analytes were calculated based on partitioning or CCS experiments [1].

There are practical reasons, why the sweet spot paradigm is widely practiced in CCS [1,9]. Analytes with low partition constants ($0 \leq \mathrm{K} < 0.4$) are held in the column a short time, therefore, they undergo an insufficient number of mixing and settling stages that are required for adequate chromatographic separation. In addition, there is a practical lower limit to peak width for rapidly eluting compounds. Analytes with longer retention times ($2.5 < \mathrm{K} \leq \infty$), on the other hand, are held in the column during many mixing and settling stages so that they are better resolved. However, the practical limitation of high K value compounds is the increased time and solvent consumption associated with eluting high K value and the concomitant peak broadening. Indeed, several methods have been devised to hasten the elution of highly retained compounds such as gradient, elution-extrusion, dual, and cocurrrent modes in CCS. It is interesting to note that experimental duration also plays an important role in the SS selection process, which is in itself prolonged and labor intensive. Finding the best balance between both continues to be an analytical challenge.

There is an attraction to specifying K=1 as the epicenter of CCS that goes beyond simple practical considerations. At unity, the concentration of the analyte is the same in both phases and it will elute as a narrow peak at one column volume in both normal-phase and reversed-phase modes. This minimizes potential error in the selection of experimental and instrument conditions, which are known to pose binary challenges that can lead to complete experiment failure [18] and are likely to contribute to the frequent impression by novice practitioners that CCS is a capricious technique.

In order to ascertain the probability of discovering a particular analyte in the sweet spot of a SS, the question was asked, "How often is an analyte found to be in the sweet spot of a (randomly chosen) solvent system?" To answer this question, K values from 19 GUESSmix compounds (excluding 0 and ∞ markers) in 57 solvent systems were plotted in frequency bins with reciprocal intervals centered at unity (Fig. 2) [19]. The resulting convex curve shows a minimum in the middle of the sweet spot. Out of 1053 data points, 198, or 18%, are in the sweet spot. This shows the general rarity of finding a specific analyte in the sweet spot of a specific SS. Each analyte was found, on the average, in the sweet spot of 10 out of 57 SSs. Combined with Fig. 1, this graphic illustrates why SS selection is a challenging, and at

the same time an important process. While Fig. 1 shows the distribution of useful K values, Fig. 2 shows the distribution of all the K values of a set of compounds in a SS panel. This is in sharp contrast to gradient methods in solid support liquid chromatography that tend to place a large number of analytes within a solvent composition range that facilitates their purification.

The GUESS method proposes a correlation between the TLC R_f value in a certain region (R_f sweet spot) with the region of optimal separation (K sweet spot) in CCS. The optimum separation region in CCS has been proposed: $0.4 \leq K \leq 2.5$ [13] or $0.5 \leq K \leq 2$ [12]. Interestingly, both sweet spot boundaries, described by different groups, are represented in terms of reciprocal values ($2/5 \leq K \leq 5/2$ or $½ \leq K \leq 2$ respectively) which illustrates the importance of K = 1 in CCS. The TLC vs. CCS sweet spot correlation strategy (ii) may be used to illustrate the challenges in correlating R_f and K values. The potential correlation between R_f and K or R_f and logK is proposed in Supplementary Information S2. Because log0 is undefined, the previously proposed two-dimensional sweet spots $0.4 \leq K \leq 2.5$ and $0.29 \leq R_f \leq 0.71$ may be used to calculate the logK working interval that corresponds this 2D sweet spot. The resulting calculations define the logK working interval to be $-0.95 \leq \log K \leq 0.95$ which corresponds to $0.11 \leq K \leq 8.9$ (Fig. 3A). This is a rather narrow window considering the full range of possible K values of zero to infinity.

The logK working interval may be expanded by adjusting the R_f sweet spot. For example, the K sweet spot remains $0.4 \leq K \leq 2.5$, and the logK working interval is set at $-2.0 \leq \log K \leq 2.0$. The resulting R_f sweet spot (on the y-axis) narrows to $0.4 \leq R_f \leq 0.6$ (Fig. 3B). The results of larger logK working intervals may be found in Supplementary Information S2 and S3. These data show that, as the K working interval widens, the slope of the line tends towards 0, and the R_f sweet spot becomes extremely narrow and nearly equal to 0.5. This demonstrates the challenge of defining the two-dimensional $R_f \times$ K sweet spot in a universal way. Proposing a one-to-one correlation with adjusted R_f or K values is very likely an oversimplification. In order to universalize such a correlation, a correlation factor would need to be created to account for differences in SS chemistry and chromatographic methodology (e.g., following the basic form of $R_f = a \log K + b$).

The following example with curcuminoids illustrates the challenges of matching both polarity and selectivity between TLC and CCS. Initially, HEMWat and ChMWat organic phases were used in the GUESS method for the separation of curcumin (**1**), demethoxycurcumin (**2**), and bisdemethoxycurcumin (**3**). Preliminary results indicated that HEMWat (6:4:5:5, v/v) and ChMWat (10:4:6, v/v) were good SS candidates (Fig. 4A). After the partitioning experiments, it was discovered that HEMWat (6:4:5:5, v/v) provided reasonable K values for the curcuminoids (0.94 for **1**, 1.02 for **2**, and 1.07 for **3**), but with separation factors less than 1.2. However, ChMWat (10:4:6, v/v) partitioning experiments showed that all K values were in fact lower than 0.02. When hexanes was added into the ChMWat system to make the HChMWat SSs 3:7:5:5, 3:7:6:4, and 3:7:7:3 (Table 1 and Supplementary Information S4), the curcuminoids were more tightly grouped in the TLC sweet spot for the organic phase of HChMWat 3:7:7:3. When the CCS experiment was then performed (Supplementary Information S4), the three analytes showed baseline separation with experimentally determined K values of 0.67 (**1**), 1.09 (**2**), and 1.76 (**3**). Notably, even

though the three curcuminoids do not appear to be well separated in the GUESS method TLC, they were resolved in the corresponding CCS experiment. The outcome of this experiment suggests that the TLC method is possibly a poor indicator of CCS selectivity (not polarity). However, if the analytes of interest can be placed in the sweet spot of the SS, the selectivity may also be satisfactory.

4. Results and Discussion

The ability to match *both* polarity *and* selectivity remains a challenge for correlating TLC and CCS. Thus, it was useful to validate the feasibility of the proposed R*f* sweet spot with an expanded range of analytes and SSs. Two basic strategies were utilized: (i) “normal prediction,” where the R*f* value sweet spot (TLC) was used to predict the K value sweet spot (CCS), and (ii) “reversed prediction” for an analyte that had a K value in the sweet spot range for a particular biphasic SS, the organic phase of that SS was used to develop the analyte in TLC.

4.1. Validation of R*f* Sweet Spots

In order to explore the correlation between R*f* and K value sweet spots, a sufficient diversity of both analytes and SSs must be considered. Hexanes-ethyl acetate-methanol-water (HEMWat), chloroform-methanol-water (ChMWat), and ethyl acetate-*n*-butanol-water (EBuWat) SS families were studied [3]. These three SS families were used initially along with 29 commercially available natural products to validate the proposed R*f* sweet spot in the “normal prediction” strategy.

4.1.1. GUESSmix—The GUESSmix, as originally formulated, includes 22 natural products with a broad polarity range [13]. The initial purpose of the GUESSmix was to investigate the correlations between R*f*, K, and SSs. Accordingly, the GUESS method was established and applied to HEMWat and ChMWat SS families. For the present work, a database with the three-dimension correlation of R*f*, K, and SS was available from that study. The results showed that in the HEMWat family, these three factors exhibited significant correlation for most of the analytes. However, for the ChMWat family, the same potential correlations were not as obvious. According to the theoretical discussion (Fig. 3 and Supplementary Information S3), the R*f* sweet spot in the HEMWat family may be differently correlated to the K value sweet spot than in ChMWat SSs.

The GUESS method database was mined to identify three R*f* values closest to 0.5 for 15 analytes (Fig. 5A and Supplementary Information S5). Applying the normal prediction strategy, the corresponding K values in their respective SSs were extracted (Fig. 5B). According to the GUESS method, some TLC SSs may correspond to two or three CCS SSs. Therefore, all corresponding K values were obtained for the R*f* value closest to 0.5, e.g., the TLC SS for reserpine, SSE7, corresponded to HEMWat 3:7:5:5, 3:7:4:6, and 3:7:3:7. Three of the 15 analytes had R*f* values in chloroform-based SSs in the TLC sweet spot. Cholesterol had TLC sweet spot values in both HEMWat and ChMWat families. R*f* values of 0.48 in the hexanes-ethyl acetate SSs and 0.42 in chloroform-methanol base SSs corresponded to K values of 1.03 in HEMWat 6:4:6:4, 1.63 in HEMWat 6:4:5:5, and 10.9 in ChMWat 10:2:8. To display the results in Fig. 5B, the K values were converted to their P*f* values. The $0.4 < K$

< 2.5 interval corresponds to 0.29 < P*f* < 0.71. The HEMWat correlation results show that 19 times out of 23 experiments the GUESS method predicted a suitable SS for the targeted separation of the analyte. Interestingly, all the misses were on the high side, which arguably still represents a viable SS for targeted separation (see section 3.2) or at least a reasonable starting point. Chloroform-methanol based TLC SSs predicted one match and two misses with their corresponding ChMWat CCS SSs. Therefore, the feasibility of this TLC and CCS correlation scheme needs to be further investigated (Supplementary Information S6).

The ethyl acetate-*n*-butanol-water (EBuWat) SS family was investigated using the reversed prediction strategy. Although K values of polar GUESSmix components [14] have been investigated using the EBuWat family (Supplementary Information S7), the TLC R*f* information was absent. The polar GUESSmix components: caffeine, chlorogenic acid, nicotine acid, tryptophan, arbutin, and salicin had K values in the sweet spot for the EBuWat family SSs. The results, shown in Fig. 6, indicate that three of the eight SS and analyte pairs had both R*f* and K values in the accepted sweet spots. SS pairs for chlorogenic acid and arbutin had either R*f* or K values outside of the sweet spot.

The investigation of the GUESSmix compounds in three popular SS families showed that the GUESS method is most effective for compounds in HEMWat SSs. Generally, HEMWat covers a range of compounds with lower polarity than ChMWat and EBuWat.

4.1.2. Other Natural Products and Solvent Systems—According to the theoretical discussion, the proposed R*f* sweet spot may be applied to a broad range of SS polarities. In order to test different SSs with the GUESS method, we employed the “reversed prediction” strategy which involved determining Rf values from compounds in SSs that had already be assigned K values as a result of partition and/or CCS studies. As a result, hexanes-chloroform-methanol-water (HChMWat), ethyl acetate-acetonitrile-water (EAcWat), and hexanes-ethyl acetate-ethanol-water (HEEtWat) were added to the previously investigated HEMWat, ChMWat, and EBuWat SS families. A cohort of 25 natural products representing seven natural product classes (alkaloids, lactones, curcuminoids, lignans, chalcones, flavonoids, and steroids) were tested in 28 SSs representing six SS families (Fig. 6 and Supplementary Information S8). TLC plates were eluted with the organic phase of the prequilibrated biphasic SS in which the K value of analytes had already been determined..

The alkaloids 3′-*O*-methylpyridoxine, 4′-*O*-methylpyridoxine (ginkgotoxin), and 5′-*O*-methylpyridoxine are positional isomers. Ginkgotoxin was used to perform the prediction by screening in both HEMWat and ChMWat SS families [4]. ChMWat (10:5:5, v/v) placed all three isomers in the both the TLC and CCS sweet spots and afforded a baseline separation [16]. This is an example similar to the case of curcuminoids where placing the analytes of interest into the TLC sweet spot also gave a reasonable CCS separation.

Ligustilide has been successfully purified in HEMWat 9:1:9:1 [20]. The organic phase of this SS gives ligustilide an R*f* value of 0.45 in silica gel TLC, indicating a good predictive match.

Curcumin, demethoxycurcumin, and bisdemethoxycurcumin were tested in a variety of HChMWat SSs [21,22]. These systems also showed matches between R*f* and K sweet spots for these compounds.

Schisandrin and gomisin were tested in HEEtWat SSs, and again gave matches between R*f* and K sweet spots. This was the first time that this particular SS family has been used to separate these natural products.

The chalcone, licochalcone A was screened using the HEMWat family, and the organic phase of HEMWat 4:6:5:5 was found to give a 0.43 R*f* value. Its K value in this SS was 3.00. Using the predicted SS yielded licochalcone A in a purity of > 95% when evaluated by quantitative NMR. This demonstrates the predictive power of the TLC/CCS sweet spot approach for this compound.

The (iso-)flavonoids, daidzein, kaempferol, and quercetin were screened using HEMWat SSs as well. According to its polarity estimate, the rutin separation was approached with EBuWat. The selected SS gave an R*f* values close to 0.5 on the high side and a K value above 7.00. Higher K values were deemed superior to lower K values for effective separations of flavonoids [23]. The flavonoids, isoliquiritin, isoliquiritin apioside, and liquiritin performed well in EAcWat SSs [5]. It is noteworthy that flavonoids are a class of compounds which has shown to be particularly amenable to CCS [24].

Finally, a group of eight commercially available steroids were screened with HEMWat SSs. For all but one, the K values correlated to R*f* values near 0.5 were matches with K > 1. This further supported the utility of the sweet spot validation.

In summary, six SS families, i.e., HEMWat, HEEtWat, HChMWat, ChMWat, EAcWat, and EBuWat, and 25 natural products were investigated for the R*f* sweet spot validation. Of 28 R*f* and K pairs, TLC can predict a usable CCS SS for 22 of them. Notably, in all 6 “misses,” the sweet spot R*f* values correlated to K values above the upper limit of the accepted sweet spot and, therefore, still represented a viable CCS SS. In terms of polarity, the great majority of these analytes would be considered as having polarities in the moderately lipophilic range, which may also be a factor in their applicability to the GUESS method.

4.2. Metabolomic Applications of the GUESS method

In natural product research, analytes in extracts may be either known compounds or unknown entities. Targeted isolation of important known compounds from metabolic mixtures has been widely practiced with CCS for decades. For the known analytes, their spectral and/or chromatographic properties can be exploited to measure K values by partitioning experiments as well as the GUESS method. Indeed, analytical methods such as HPLC-UV, fluorimetry, UV-vis, GC-MS, and ^{1}H NMR have been used to quantitate compounds in partitioning experiments [3]. The GUESS “normal prediction” method usually relies on the availability of reference standards to screen SSs. In addition, the GUESS method can also be performed with mixtures by tracking easily distinguishable TLC bands [25].

The chemical approach, which attempts the targeted isolation of a particular analyte, is clearly insufficient for bioactivity-guided fractionation studies, because the most easily identifiable analytes are likely not to be the bioactive ingredient of interest. In fact, restricting the search for bioactives to easily identifiable known compounds, defeats the actual purpose of the search. Moreover, using chemical methods to pursue biological targets creates incoherency between the methodology and the desired outcome. To overcome this disadvantage, a TLC-based bioautography was introduced [26,27]. The use of TLC in both the GUESS method and bioautography to visualize bioactivity creates seamless transition from chemical based fractionation to biological targets. Bioautography linked with CCS provides for a bioactivity-guided fractionation methodology in which bioactivity is the main driving force for separation strategies. The fact that CCS is essentially a loss-free separation technique further enhances its suitability for bioactivity-guided fractionation.

4.2.1. Chemical analysis of a crude natural product extract—An example of the usefulness of the GUESS method for targeted isolation of analytes may be found in our recent work with *Schisandra chinensis* metabolites. Fig. 7A shows the initial GUESS method exploration of a crude sample. A dark band occupies the R*f* sweet spot for TLC plates developed in the upper phase of the equilibrated HEMWat SSs 6:4:6:4, 5:5:5:5, and 4:6:4:6, as seen by UV inspection at 254 nm. The subsequent HEMWat (5:5:5:5, v/v) CCS of the sample shows a likely major compound in the K value sweet spot shown on the TLC fractogram along with other analytes, which are closely-related by polarity (Fig. 7B). Interestingly, the CCS reveals a greater sample complexity than simple TLC analysis provided, especially concerning analytes with lower polarity than the targeted region.

4.2.2. Bioactivity-guided fractionation of bacterial extracts—The GUESS method may be extended to bioactivity evaluation by the bioautography, a method of determining bioactivity on a TLC plate. Such bioautography can be used to directly screen the potential hits or leads. Bioactive analytes on the developed TLC slide will inhibit bacterial growth to reveal inhibition zones. The GUESS method may be directly applied with bioautography in order to provide information regarding the optimal CCS SS needed for further chromatographic purification of the active principle(s). Typically, more SS families have to be employed to achieve an adequately wide search for suitable SSs. Combined with bioautography, the GUESS strategy can be applied without a predetermined reference or target compound, establishing a chemically untargeted and biologically fully targeted approach. Thus, GUESS-based bioautography exhibits two major advantages: (i) the target prediction can be achieved without chemical reference; (ii) CCS can be performed directly and avoids multiple partitioning experiments and *in vitro* assays. The TLCs shown in Fig. 8A, exhibit a crude Streptomycetes' extract sample TLC accompanied by the TLCs of both the upper and lower phases of the crude sample in the same SS. In the "crude" TLC, one part of inhibition zone occupied the R*f* sweet spot for the TLC plate developed in the upper phase of the SS. In a partitioning experiment (Fig. 8A "LP" and "UP" lanes), the majority of activity around the R*f* sweet spot presented in both LP and UP. According to the intensities, the corresponding K value(s) of the bioactive analyte(s) was(were) estimated to be close to unity. While the majority of activity were present in the organic lower phase (far left), the range of R*f* values for bioactive analytes suggested that the SS may be suitable for CCS. The

TLC monitoring of the CCS fractionation in Fig. 8B demonstrates the affinity of the activity for the organic phase, along with a reasonable separation of active components into discrete fractions. While further results from this ongoing anti-tuberculosis drug discovery project will be reported in due course, the results demonstrate the utility and feasibility of GUESS-guided bioautography.

5. Conclusions

As a result of this study, practical TLC prediction zones can be recommended such as those shown in the Supplementary Information S10. The expanded investigation of the GUESS method showed that the R*f* sweet spot can be used in most of SS families. Generally, matching R*f* and K value sweet spots was suitable for analytes with a broad polarity range, but more work needs to be done with polar analytes. To address this problem, more polar SS families, e.g., *ter*AcWat, H*ter*AcWat, and EAcWat, should be tested. Future work should also consider pH as a variable, based on the common pH-dependence of the solubility of polar and ionizable constituents.

The collaborative nature of this investigation by multiple authors, who perform CCS for different reasons, showed that the GUESS method is versatile and readily applied in a variety of research environments. Furthermore, the present study did not only contribute to redefinition of the R*f* sweet spot for the CCS, but also exhibited a novel methodology, i.e., CCS and TLC combined with bioautography. The TLC-based CCS SS prediction can be performed without reference materials, thereby enhancing the discovery nature of bioactivity-guided fractionation. GUESS-guided bioautography can recognize bioactivity and predict the SS for CCS, in the same step.

Overall, 43 natural products (Supplementary Information S9) and six SS families were involved in this study, providing coverage for a substantial range of natural product polarity and CCS application space. The results showed that out of 62 correlations, 45 (73%) had both R*f* and K values in their respective sweet spots. This confirms the fitness of the GUESS method as a semi-empirical methodology with good predictive potential for the combinations of analytes and solvent systems investigated. While the determination of the K value(s) of each target analyte prior to a preparative CCS run is desirable especially in cases of the targeted isolation of known compounds, a TLC-based SS prediction methodology provides a straightforward technique to greatly accelerate CCS SS selection and to foster the usability of CCS for novice practitioners. In conclusion, the GUESS concept is capable of the rapid approximation (“GUESSing”) of SSs and CCS experimental conditions for any separation problem that can be examined by TLC.

Acknowledgments

The authors gratefully acknowledge support for their research involving CCS methodology through grants P50 AT000155, R01 DE21040, U41 AT008706 and R44 AT004534), as well as research concerning bioautography through grant R21 AI093919, all from the National Institutes of Health. We are also grateful for support from the team at Cherry Instruments - Wrightwood Technologies (Chicago, IL) in particular by Mr. Samuel Pro and Mr. Warren Friedel. Finally, the authors wish to express their gratitude to Drs. Rasika Phansalkar, Charlotte Simmler, Joo-Won Nam and Geping Cai for their collegial support and helpful discussions.

Abbreviations

CCS	countercurrent separation
TLC	thin-layer chromatography
K	partition coefficient
R*f*	retention factor for TLC
SS	solvent system
P*f*	partitioning retention factor
ChMWat	chloroform-methanol-water solvent system
HEMWat	hexanes-ethyl acetate-methanol-water solvent system
SSE	TLC solvent systems based on ethyl acetate
SSC	TLC solvent systems based on chloroform
GUESS	generally useful estimate of solvent systems
Sf	stationary phase volume retention ratio

References

1. Friesen JB, McAlpine JB, Chen SN, Pauli GF. Countercurrent separation of natural products: an update. J Nat Prod. 2015; 78:1765–1796. [PubMed: 26177360]
2. Ito Y, Bowman RL. Countercurrent chromatography: liquid-liquid partition chromatography without solid support. Science. 1970; 167:281–283. [PubMed: 5409709]
3. Liu Y, Friesen JB, McAlpine JB, Pauli GF. Solvent system selection strategies in countercurrent separation. Planta Med. 2015; 81:1582–1591. [PubMed: 26393937]
4. Liu Y, Chen SN, McAlpine JB, Klein LL, Friesen JB, Lankin DC, Pauli GF. Quantification of a botanical negative marker without an identical standard: ginkgotoxin in *Ginkgo biloba*. J Nat Prod. 2014; 77:611–617. [PubMed: 24432981]
5. Simmler C, Nikolic D, Lankin DC, Yu Y, Friesen JB, van Breemen RB, Lecomte A, Le Quemener C, Audo G, Pauli GF. Orthogonal analysis underscores the relevance of primary and secondary metabolites in licorice. J Nat Prod. 2014; 77:1806–1816. [PubMed: 25080313]
6. Ramos Alvarenga RF, Friesen JB, Nikolic D, Simmler C, Napolitano JG, van Breemen R, Lankin DC, McAlpine JB, Pauli GF, Chen S-N. K-targeted metabolomic analysis extends chemical subtraction to DESIGNER extracts: selective depletion of extracts of hops (*Humulus lupulus*). J Nat Prod. 2014; 77:2595–2604. [PubMed: 25437744]
7. Qiu F, Friesen JB, McAlpine JB, Pauli GF. Design of countercurrent separation of *Ginkgo biloba* terpene lactones by nuclear magnetic resonance. J Chromatogr A. 2012; 1242:26–34. [PubMed: 22579361]
8. McAlpine JB, Friesen JB, Pauli GF. Separation of natural products by countercurrent chromatography. Methods Mol Biol. 2012; 864:221–254. [PubMed: 22367899]
9. Pauli GF, Pro SM, Friesen JB. Countercurrent separation of natural products. J Nat Prod. 2008; 71:1489–1508. [PubMed: 18666799]
10. Friesen JB, Pauli GF. Performance characteristics of countercurrent separation in analysis of natural products of agricultural significance. J Agric Food Chem. 2008; 56:19–28. [PubMed: 18069794]

11. Inui T, Wang Y, Deng S, Smith DC, Franzblau SG, Pauli GF. Counter-current chromatography based analysis of synergy in an anti-tuberculosis ethnobotanical. J Chromatogr A. 2007; 1151:211–215. [PubMed: 17316661]
12. Ito Y. Golden rules and pitfalls in selecting optimum conditions for high-speed counter-current chromatography. J Chromatogr A. 2005; 1065:145–168. [PubMed: 15782961]
13. Friesen JB, Pauli GF. G.U.E.S.S.—a generally useful estimate of solvent systems for CCC. J Liq Chromatogr Relat Technol. 2005; 28:2777–2806.
14. Friesen JB, Pauli GF. Rational development of solvent system families in counter-current chromatography. J Chromatogr A. 2007; 1151:51–59. [PubMed: 17320092]
15. Hostettmann K, Hostettmann-Kaldas M, Sticher O. Application of droplet counter-current chromatography to the isolation of natural products. J Chromatogr A. 1979; 186:529–534.
16. Liu Y, Friesen JB, Klein LL, McAlpine JB, Lankin DC, Pauli GF, Chen SN. The generally useful estimate of solvent systems (GUESS) method enables the rapid purification of methylpyridoxine regioisomers by countercurrent chromatography. J Chromatogr A. 2015; 1426:248–251. [PubMed: 26680272]
17. Pauli GF, Pro SM, Chadwick LR, Burdick T, Pro L, Friedl W, Novak N, Maltby J, Qiu F, Friesen JB. Real-time volumetric phase monitoring: advancing chemical analysis by countercurrent separation. Anal Chem. 2015; 87:7418–7425. [PubMed: 26152934]
18. Friesen JB, Pauli GF. Binary concepts and standardization in counter-current separation technology. J Chromatogr A. 2009; 1216:4237–4244. [PubMed: 19203761]
19. Friesen JB, Ahmed S, Pauli GF. Qualitative and quantitative evaluation of solvent systems for countercurrent separation. J Chromatogr A. 2015; 1377:55–63. [PubMed: 25542704]
20. Schinkovitz A, Pro SM, Main M, Chen SN, Jaki BU, Lankin DC, Pauli GF. Dynamic nature of the ligustilide complex. J Nat Prod. 2008; 71:1604–1611. [PubMed: 18781813]
21. Inoue K, Nomura C, Ito S, Nagatsu A, Hino T, Oka H. Purification of curcumin, demethoxycurcumin, and bisdemethoxycurcumin by high-speed countercurrent chromatography. J Agric Food Chem. 2008; 56:9328–9336. [PubMed: 18826228]
22. Liu Y, Friesen JB, McAlpine JB, Chen SN, Pauli GF. The generally useful estimate of solvent systems method enables the rapid separation of curcuminoids by countercurrent separation. Planta Med. 2015; 81:PJ4.
23. Liu Y, Garzon J, Friesen JB, Zhang Y, McAlpine JB, Lankin DC, Chen SN, Pauli GF. Countercurrent assisted quantitative recovery of metabolites from plant-associated natural deep eutectic solvents. Fitoterapia. 2016; 112:30–37. [PubMed: 27118320]
24. Neves Costa, Fd, Leitão, GG. Strategies of solvent system selection for the isolation of flavonoids by countercurrent chromatography. J Sep Sci. 2010; 33:336–347. [PubMed: 20169552]
25. Fan Q, Zhang H, Hu H, Xu Y, Song Q. Quick method for separating target compounds from the bark of Maqian (*Zanthoxylum myriacanthum* var. pubescens) by high-performance countercurrent chromatography. J Sep Sci. 2016; 39:4049–4052. [PubMed: 27557568]
26. Grzelak EM, Hwang C, Cai G, Nam JW, Choules MP, Gao W, Lankin DC, McAlpine JB, Mulugeta SG, Napolitano JG, Suh JW, Yang SH, Cheng J, Lee H, Kim JY, Cho SH, Pauli GF, Franzblau SG, Jaki BU. Bioautography with TLC-MS/NMR for rapid discovery of anti-tuberculosis lead compounds from natural sources. ACS Infect Dis. 2016; 2:294–301. [PubMed: 27478868]
27. Grzelak EM, Liu Y, Nam JW, Friesen JB, Chen SN, Lankin DC, McAlpine JB, Suh JW, Yang SH, Cheng J, Lee H, Kim JY, Cho SH, Pauli GF, Franzblau SG, Jaki BU. TLC-bioautography linked with guess: a truely targeted active compound isolation process. Planta Med. 2015; 81:PJ5.

Highlights

- TLC method of solvent system selection for countercurrent separation is revisited.
- A broader approach to the GUESS method is implemented.
- The theory supporting TLC and CCS sweet spot matching is presented.
- Application of the GUESS method to natural product separation is demonstrated.

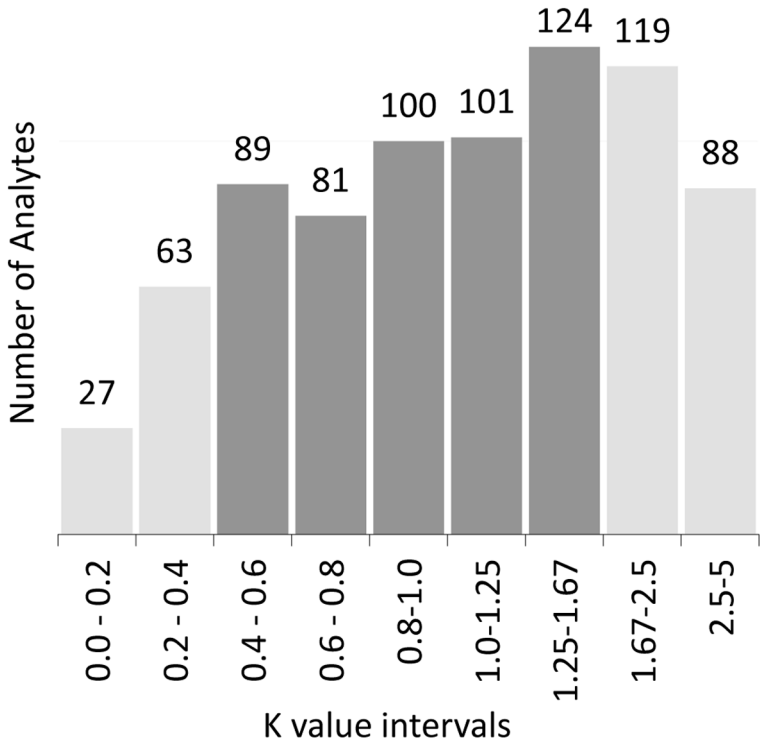

Figure 1.
K value distribution of natural products isolated or purified by CCS [1,9].

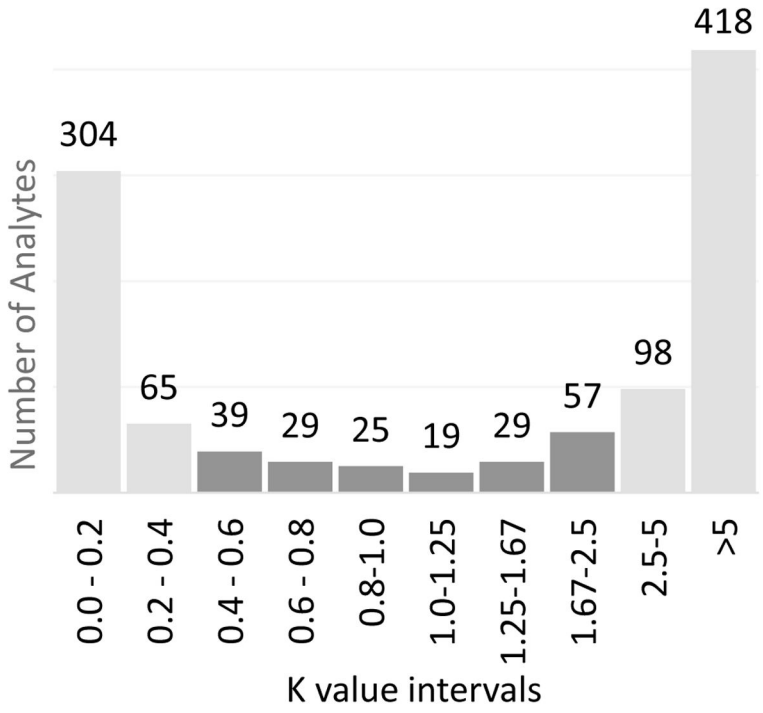

Figure 2.

K value distribution of 19 GUESSmix compounds (excluding 0 and ∞ markers) in 57 solvent systems (1083 data points).

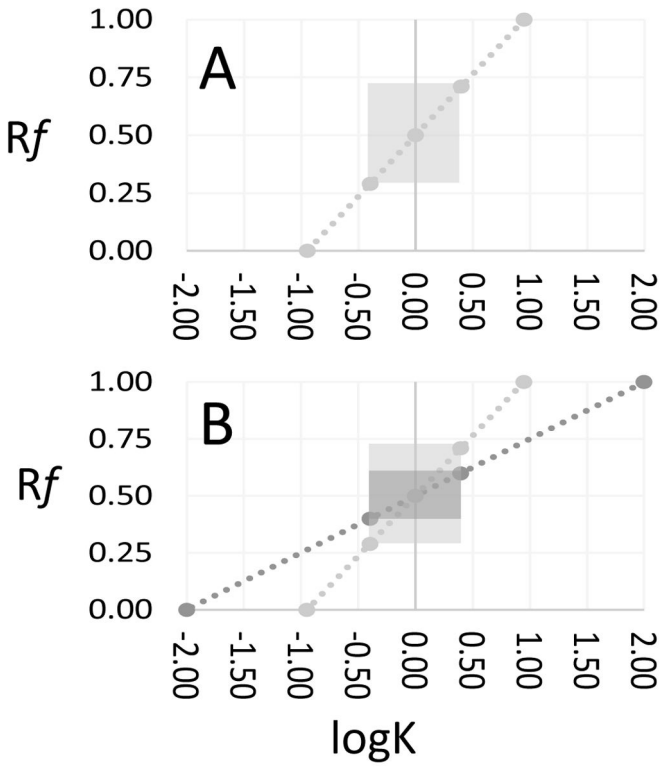

Figure 3.

Models correlating R*f* and K sweet spots by comparison of R*f* and logK values with a K sweet spot $0.4 \leq K \leq 2.5$. (A) logK working interval $-1.0 \leq \log K \leq 1.0$. (B) logK working interval $-2.0 \leq \log K \leq 2.0$. The results of larger logK working intervals are found in Supplementary Information S3.

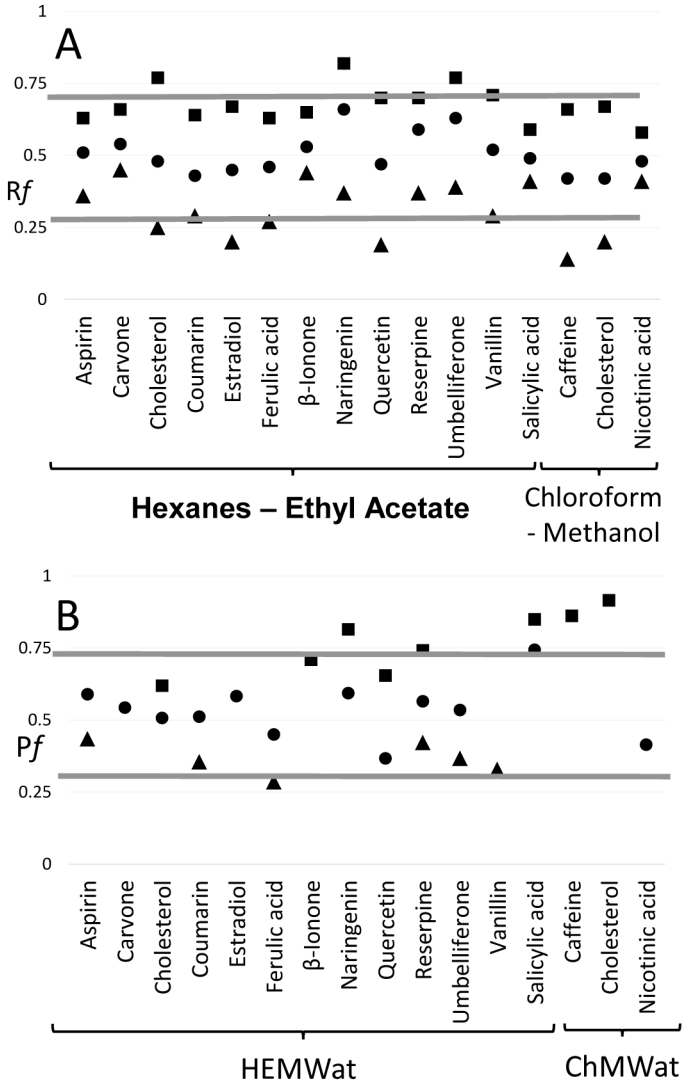

Figure 4.

A. The three R*f* values closest to 0.5 for each of 15 GUESSmix analytes screened in both hexanes-ethyl acetate and chloroform-methanol based TLC SSs [13]. The values for data points and SSs are given in Supplementary Information S5. B. K values (expressed in P*f*) for the GUESSmix compounds in A in the SS(s) predicted by the R*f* values closest to 0.5. Some TLC SSs predict more than one CCS SS. The $0.4 < K < 2.5$ interval corresponds to $0.29 < Pf < 0.71$ [13]. The K values for data points and SSs are given in Supplementary Information S5.

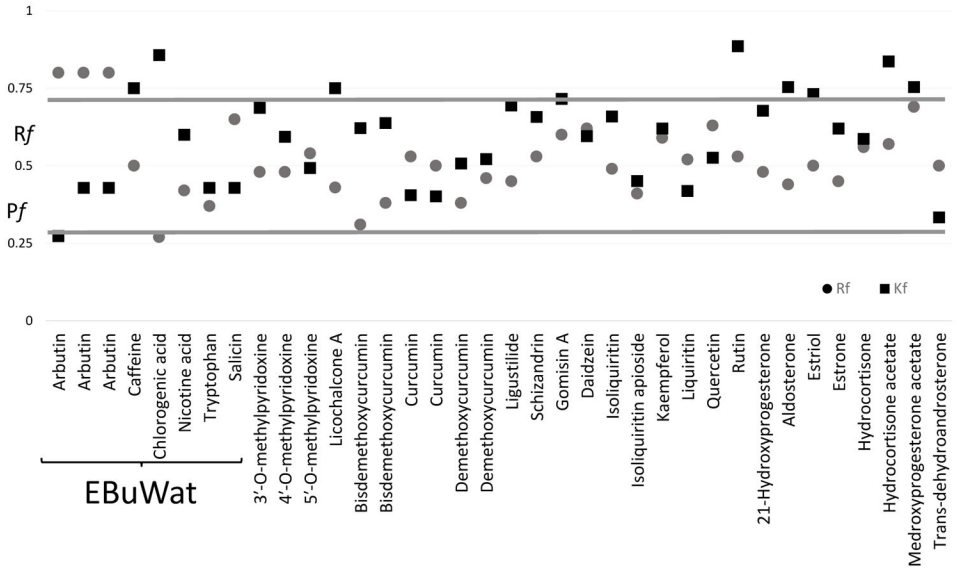

Figure 5.

R*f* and K values (expressed in P*f*) for selected GUESSmix compounds in the SS(s) predicted by the GUESS method. The $0.4 < K < 2.5$ interval corresponds to $0.29 < Pf < 0.71$. The values of the data points and SSs are given in Supplementary Information S7 & S8.

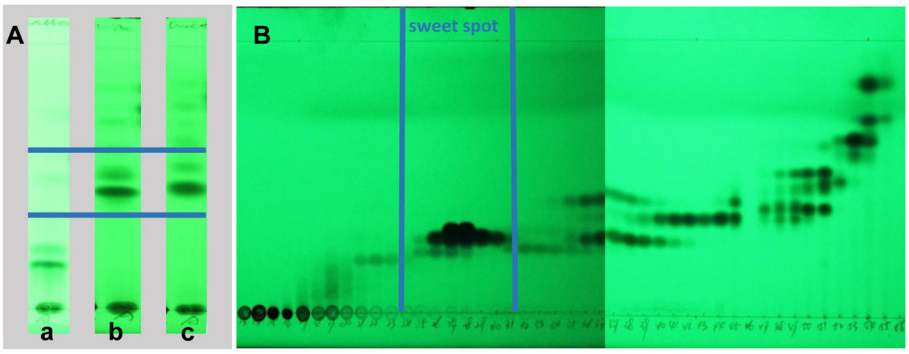

Figure 6.

CCS results of a natural chemical matrix using TLC-based predicted SS. (A) represents a screening procedure, in which the SS candidates (v/v) are HEMWat 6:4:6:4 (a), 5:5:5:5 (b), and 4:6:4:6 (c). (B) TLC monitoring results of a CCS run in HEMWat 5:5:5:5. Analytes with a K value of ~1 reside in fraction 28 (F28), representing one column volume. The test sample was 150 mg of an open HP-20 column methanol fraction from *Schisandra chinesis* extract. The TLC SS was the HEMWat (5:5:5:5, v/v) organic phase.

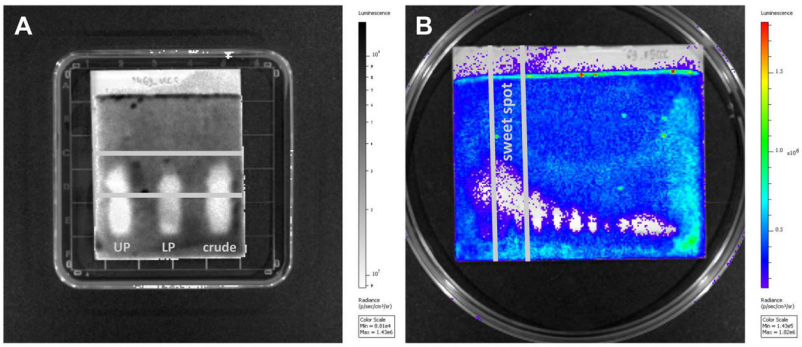

Figure 7.

Exemplary bioautography application for TLC-based CCS SS prediction using the *Mycobacterium tuberculosis* (*M.tb*) inhibition zone (A), and leading directly to the resolution demonstrated in the CCS fractogram (B). “UP” designates the upper phase of the partitioning experiment; “LP” the lower phase of the partitioning experiment; and “crude” the crude sample.

Table 1

TLC values of curcumin, demethoxycurcumin and bisdemethoxycurcumin developed in the organic phase of selected HEMWat (hexanes-ethyl acetate-methanol-water), ChMWat (chloroform-methanol-water) and HChMWat (hexanes-chloroform-methanol-water) SSs.

analytes	HEMWat 4:6:5:5	ChMWat 10:4:6	HChMWat 3:7:5:5	HChMWat 3:7:6:4	HChMWat 3:7:7:3
curcumin	0.45	0.75	0.61	0.42	0.50
demethoxycurcumin	0.45	0.46	0.30	0.26	0.46
bisdemethoxycurcumins	0.45	0.26	0.15	0.18	0.38