Journal of Chromatography B, 910 (2012) 95–102

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Test-set reduction in the screening step definition of a chiral separation strategy in polar organic solvents chromatography☆

Hasret Ates, Bart Desmedt, Yvan Vander Heyden*

Department of Analytical Chemistry and Pharmaceutical Technology, Center for Pharmaceutical Research (CePhaR), Vrije Universiteit Brussel-VUB, Laarbeeklaan 103, B-1090 Brussels, Belgium

ARTICLE INFO

Article history:
Received 26 October 2011
Accepted 11 April 2012
Available online 15 April 2012

Keywords:
Enantioseparations
Chiral separation strategies
Polar organic solvents chromatography
Test-set reduction

ABSTRACT

The last decades, many efforts have been made to design and develop chiral separation strategies for different analytical techniques. To ensure that these strategies are broadly applicable rather large test-sets of molecules with very diverse molecules are used. The most enantioselective and complementary separation systems are then used as screening conditions in separation strategies. Potential changes in conditions e.g. implementation of new chiral selectors, requires screening of the entire set to retain the most enantioselective systems. A rational reduction of the test-sets may open new perspectives for developing and updating separation strategies. In the present work, it is investigated whether the screening step of an existing separation strategy in polar organic solvents chromatography can be reconstructed based on reduced test-set results Therefore, the structures of the 58 molecules of the test-set are digitally drawn and their optimal geometrical conformations calculated. From these conformations 3D-molecular descriptors are calculated. The test-set reduction is performed using the Kennard and Stone algorithm: compounds with the most diverse descriptors are selected. The test-sets are gradually reduced with 10% starting from 90% to 30% of the initial size. The results pointed out that with some reduced test-sets the same chromatographic systems are selected. A test-set reduction with 30% (41 remaining compounds) seems possible without losing information on the global enantioselectivity and complementarity of the tested chiral stationary phases.

1. Introduction

Administered drug compounds will only be able to interact with target receptors to cause the intended therapeutic effect when this interaction is stereochemically stable i.e. when compound and receptor fit well. When this is not the case and the compound only fits partially the receptor, the drug shows no or a reduced effect, but it can also express unexpected and even harmful effects from interaction with other receptors. This phenomenon is particularly of importance for chiral drug compounds, consisting of enantiomers, that interact with different chiral media, such as receptors or enzymes. One enantiomer will interact properly with the target resulting in the intended effect, while the interaction of the other enantiomer may cause pharmacological or toxicological problems as mentioned above. Therefore, during the development of chiral lead compounds, the possible differences in safety and/or efficacy of the enantiomers should be taken into account and studied [1].

Testing enantiomers for safety and/or efficacy intentions implies that these entities are available as individual compounds. Besides enantiopure synthesis, pharmaceutical enantiomers can also be obtained by enantioselective chromatography. For analytical purposes many chiral separation strategies have been developed e.g. to assay the impurity of an enantiomer in an enantiopure drug. Because the interactions between analytes and selectors, responsible for enantiorecognition, are not completely understood and mainly based on hypotheses [2–4], it is hardly possible to predict experimental conditions which will ensure the separation of enantiomers. Consequently in the past many efforts were put into the development of generic (screening) strategies for chiral compounds [5–15]. The goal of these strategies is to have a set of experimental conditions that allow separating a broad range of chiral substances into their enantiomers in a limited number of experiments. The importance of these strategies is that they can be used in several stages of drug development, particularly in early high-throughput screening of new lead compounds. However, the development of these strategies is time consuming for various reasons. First of all, the availability of many different chiral selectors and the continuously growing number of chiral stationary phases (CSPs) associated herewith, complicate the choice of selecting the systems with broad and complementary enantiorecognition abilities. The selection of

☆ This paper belongs to the Special Issue Chemometrics in Chromatography, Edited by Pedro Araujo and Bjørn Grung.

* Corresponding author. Tel.: +32 2 477 4734; fax: +32 2 477 4735.
E-mail address: yvanvdh@vub.ac.be (Y.V. Heyden).

1570-0232/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jchromb.2012.04.017

the final experimental conditions happens by selecting first the CSP-mobile phase combination with the broadest enantiorecognition ability toward the test-set. For the selection of the next system, no longer the enantioselectivity of the individual chromatographic system is evaluated but its maximal complementarity toward the already selected. This approach continues until the desired number of systems is included or the highest possible success rate attained. For the development of these strategies, rather large test-sets of molecules are used to ensure the generic character of the conclusions. Screening several CSPs with such large test-sets requires many experiments and thus time.

The goal of this work is to try to speed up the development process of a liquid chromatographic strategy by decreasing the number of tested chiral compounds in a rational way, instead of performing an *ad random* selection of the test molecules. The reference strategy that will be used in this work is the screening step of a strategy in polar organic solvents chromatography [10,11]. Polar organic solvents chromatography is a separation mode where non-aqueous polar organic solvents are used as mobile phase. This mode can be a valuable alternative for the classical normal- and reversed phase modes. It broadened the application field of enantioselective chromatography and has gained much attention [10,11,16–20].

To define the above-mentioned strategy, screening of 58 compounds was done on 8 CSPs with 2 mobile phases consisting of only polar organic solvents with a basic and an acidic additive [10,11]. The experimental setup, i.e. the chromatographic systems, selected based on a reduced test-set will then be 'screened' with all molecules of the original test-set. The decision whether the selected systems based on a reduced test-set may be retained or not in a strategy, depends on the success rate compared to that of the systems selected with the original test-set. Rational test-set reduction is performed by using molecular descriptors and a selection algorithm as will be explained in the next section.

2. Theoretical background

2.1. Molecular descriptors

A molecular descriptor transforms chemical information contained in a molecule into a useful number through a logical and mathematical procedure or is the result of a standardized experiment [16]. As a result, two main classes of molecular descriptors can be distinguished: experimental descriptors derived from experimental measurements and theoretical descriptors derived from a symbolic representation of the chemical entity. The representation of the molecules allows defining subclasses within the theoretical descriptors. The most simple representation is the molecular formula. The descriptors derived from it are called zero-dimensional descriptors (0D). One-dimensional descriptors (1D) are derived from substructure-list representations. These can be considered as a one-dimensional representation of a molecule and consists of a list of structural fragments. Two-dimensional descriptors (2D) describe, for instance, the connectivity of atoms in the molecule in terms of the presence and nature of the chemical bonds. In fact the 2D-descriptors are derived from the topological representation of the molecule. The three-dimensional descriptors (3D) are calculated from the geometrical structure of the molecule, which gives an overall representation of the spatial configuration of the entity. Finally, four-dimensional descriptors (4D) are derived from a representation of the molecules through their stereo-electronic properties. These descriptors are related to those molecular properties that arise from their electronic distribution and are characterized by a scalar field associated with the 3D molecular geometry.

With these descriptors, a numerical characterization of the molecules can be made which enables a mathematical treatment and comparison of these chemical entities. The descriptor classes contain different subclasses which will not be discussed here. More information can be found in Ref. [21].

2.2. Selection algorithm

The Kennard and Stone (KS) selection algorithm [22] is used to compose the reduced test-sets. It selects sequentially the molecules in such a way that they are uniformly spread over the entire space of the original test-set. First, the Euclidean distance between all pairs of points (i.e. the molecules described by the descriptors) is determined. The molecules that are the furthest away from each other are selected as the first two objects for the reduced test-set. Then, the shortest distance of the remaining molecules to the two selected ones is determined and the molecule that is the most distant (where this shortest distance is maximal) is selected as third. This procedure is repeated until the desired number of molecules for the reduced test-set is selected. Based on the results of the reduced test-set compounds, the column or system sequence for screening is determined. The success rates for this sequence are then considered taking into account all compounds.

3. Materials and methods

3.1. Screened compounds

The 58 racemates that have been screened are: **acebutolol**, **alprenolol**, **atenolol**, **atropine**, **betaxolol**, chlorthalidone, **ephedrine**, fenoprofen, ibuprofen, ketoprofen, **labetalol**, mandelic acid, **nadolol**, naproxen, naringenin, **oxazepam**, **pindolol**, praziquantel, promethazine, sulpiride, suprofen, **tetramisole** and warfarin (all from Sigma–Aldrich, Steinheim, Germany), acenocoumarol and **dimethindene** (from Novartis, Basel, Switzerland), nimodipine, nisoldipine and nitrendipine (Bayer, Leverkusen, Germany), leucovorin and **oxprenolol** (Cynamid Benelux, Brussels, Belgium), **propranolol** and **verapamil** (Fluka, Neu-Ulm, Switzerland), **ambucetamide** (Janssen Pharmaceutica, Beerse, Belgium), **bopindolol** (Sandoz, Holskirchen, Germany), **carvedilol** (Boehringer, Mannheim, Germany), **esmolol** (Du Pont de Nemours, Saconnex, Switzerland), flurbiprofen (ICN Biomedicals, Ohio, USA), **mebeverine** (Duphar, Amsterdam, The Netherlands), **metoprolol** (Astra Hassle AB, Lund, Sweden), nicardipine (UCB, Brussels, Belgium), sotalol (Merck, Darmstadt, Germany), terbutaline (Astra Draco, Lund, Sweden), **bupranolol**, **carazolol**, salbutamol, salmeterol, **bisoprolol**, **methadone**, **carbinoxamine**, **chlorphenamine**, hexobarbital, **isothipendyl**, **mepindolol**, meptazinol, **mianserin**, **propiomazine**, **procyclidine** and **tertatolol** were gifts from different sources. The test compounds were chosen based on availability in the lab and the differences or similarities in molecular and pharmacological properties.

A closer look to the test-set shows a larger representation of basic compounds (in **bold**). Their fraction represents that of the occurring fraction within the commercialized drugs (with a majority of basic compounds).

3.2. Chromatographic systems

The chromatographic systems that are screened consist of following CSPs (see also Table 1): (1) amylose tris(3,5-dimethylphenylcarbamate) 150 mm × 4.6 mm, 5 μm; (2) cellulose tris(3,5-dimethylphenylcarbamate) 150 mm × 4.6 mm, 5 μm; (3) amylose tris([S]-α-methylbenzylcarbamate) 150 mm × 4.6 mm, 5 μm; (4) cellulose tris (4-methylbenzoate) 150 mm × 4.6 mm, 5 μm; (5) cellulose tris(3-chloro-4-methylphenylcarbamate)

Table 1
Chiral stationary phases used in the initial screening experiments.

	Selector	Commercial name
CSP1	Amylose tris(3,5-dimethylphenylcarbamate)	Chiralpak AD-RH
CSP2	Cellulose tris(3,5-dimethylphenylcarbamate)	Chiralcel OD-RH
CSP3	Amylose tris([S]-α-methylbenzylcarbamate)	Chiralpak AS-RH
CSP4	Cellulose tris (4-methylbenzoate)	Chiralcel OJ-RH
CSP5	Cellulose tris(3-chloro-4-methylphenylcarbamate)	LuxCellulose-2
CSP6	Amylose tris(5-chloro-2-methylphenylcarbamate)	LuxAmylose-2
CSP7	Cellulose tris(4-chloro-3-methylphenylcarbamate)	LuxCellulose-4
CSP8	Cellulose tris(3,5-dichlorophenylcarbamate)	Sepapak-5

250 mm × 4.6 mm, 5 μm; (6) amylose tris(5-chloro-2-methylphenylcarbamate) 250 mm × 4.6 mm, 5 μm; (7) cellulose tris(4-chloro-3-methylphenylcarbamate) 250 mm × 4.6 mm, 5 μm and (8) cellulose tris(3,5-dichlorophenylcarbamate) 250 mm × 4.6 mm, 5 μm. The screening mobile phases are acetonitrile/di-ethylamine/trifluoroacetic acid and methanol/diethylamine/trifluoroacetic acid both in 100/0.1/0.1 (v/v/v) ratios. Acetonitrile (ACN) and methanol (MeOH) are qualified as HPLC grade (Fisher Scientific – Loughborough, Leicestershire, UK). Trifluoroacetic acid (TFA) and diethylamine (DEA) (Sigma–Aldrich, Steinheim, Germany) are added to the mobile phases as acidic and basic additive, respectively. These additives are used to improve enantioselectivity and minimize peak broadening that arises from unwanted interactions between the polar analytes and the stationary phases. Initially eight mobile phases have been tested [23]. Acetonitrile or methanol were the main solvents but some mobile phases also contain 5% (v/v) of an alcohol (ethanol, 2-propanol, butanol or methanol) as organic modifiers [10,11]. Despite the use of all these mobile phases, only ACN/DEA/TFA and MeOH/DEA/TFA are selected because of their superior results (more enantioseparations).

The chromatographic experiments are performed on a Merck-Hitachi HPLC system (Tokyo, Japan) with an L-7100 pump, an autosampler L-7200 with a 100 μl loop, an L-7400 UV detector, a D-7000 interface and an L-7360 column oven. Data-management is done with the D-7000 HPLC System Manager software (Merck-Hitachi, 1994–2001, version 4.1). The sample injection volume was 5 μl and all experiments are run under isocratic elution conditions with a flow rate of 0.5 ml min^{-1}. The column temperature during the screenings is set at 20 °C, while the detection of the compounds occurred at 220 nm.

3.3. Data processing

In chiral separations, as generally in chromatography, the quality of the separation of enantiomers is expressed by the resolution (Rs). The Rs is calculated according to the equation used in the United States Pharmacopeia [24]:

$$\mathrm{Rs} = \frac{2(t_{r2} - t_{r1})}{w_1 + w_2}$$

with t_{r1} and t_{r2} the retention times (in min) of the first and second eluting peak, and w_1 and w_2 the baseline widths (in min) (tangent method) of the corresponding peaks, respectively. When Rs ≥ 1.50 a baseline separation is achieved. When 0 < Rs < 1.50, the compounds are only partially separated and for Rs = 0, no enantioselectivity at all is observed. In this work, a compound is called 'separated' from the moment it has an Rs > 0 (because enantioselectivity is observed and normally analytical conditions are defined from baseline separation).

3.4. Molecular structure optimization

The structures of the 58 compounds are digitally drawn and optimized with the software package Hyperchem™ 6.03 Professional (Hypercube, Gainesville, FL, USA). The geometrical optimization of the molecules is performed using the molecular mechanics force field method (MM+) using the Polak-Ribière conjugate gradient algorithm to calculate a geometry with a minimum potential energy. The termination conditions of the calculations are set at an RMS gradient of 0.1 kcal/(Å mol). In the geometrical optimization of the molecular structures, so-called vacuum conditions are applied (in this situations the solvent, in which a molecule would occur, is ignored). The optimized molecular structures are then imported as a data matrix in Dragon® [25] to calculate the molecular descriptors.

3.5. Calculation of molecular descriptors

The molecular descriptors of the compounds are calculated using the Dragon® 5.0 Professional software [25]. The software provides the possibility to calculate 1664 descriptors, divided over 20 descriptor classes. The 0D-descriptors are categorized under the constitutional descriptors. The calculated 1D-descriptors are situated in the subclasses called functional group counts and atom-centered fragments. The 2D-descriptors consist of topological descriptors, walk and path counts, connectivity indices, information indices, 2D-autocorrelations, edge adjacency indices, Burden eigenvalues, topological charge indices and eigenvalue-based indices. The 3D-descriptors are found in the subclasses called Randic molecular profiles, geometrical descriptors, radial distribution function descriptors (RDF descriptors), 3D-MoRSE descriptors, WHIM descriptors and GETAWAY descriptors, while a final class is called 'others' and contains charge descriptors and molecular properties (including farmacological indices). More detailed information about the descriptors can be found in [21].

3.6. Selection of the reduced test-sets

The molecules which will be part of the reduced test-sets are selected by using the Kennard and Stone selection algorithm as stated above. These calculations are performed with Matlab™, version 7.1 (The Mathworks, Natick, MA, USA). For every compound, the same descriptors are calculated. It concerns 721 molecular descriptors per compound (41 Randic molecular profiles, 74 geometrical descriptors, 150 RDF descriptors, 160 3D-MoRSE descriptors, 90 WHIM descriptors and 197 GETAWAY descriptors). The Kennard and Stone algorithm then selects the most diverse molecules, i.e. those situated furthest from each other of from the previously selected in the 721-dimensional space. For instance for a test-set reduced with 10% it means that the 90% most diverse compounds are selected.

4. Results and discussion

4.1. Chromatographic screening

First all 58 compounds are screened on eight CSPs with eight mobile phases. With the results of these chromatographic experiments, the most enantioselective and complementary systems are selected as screening step in a generic chiral separation strategy in polar organic solvent chromatography [10]. The column and mobile phase sequence with the highest separation success rate is: CSP5 > CSP2 > CSP8, first screened with ACN/DEA/TFA then with MeOH/DEA/TFA. The selection of the first system is based on the highest success rate or the broadest enantioselectivity. Then, the

most complementary system is sequenced as second. This is continued until from all screened systems the most complementary are selected or the success rate is not increasing anymore. This selection rules out all mobile phases with 5% alcohol because in comparison to ACN/DEA/TFA or MEOH/DEA/TFA they express a lower separation performance [23]. The screening with the ACN/DEA/TFA mobile phase results in 40/58 (69%) partial and baseline separated compounds, and with the MeOH-based mobile phase in 21/58 (36%) separations. Combined, this gives a success rate of 44/58 (76%). These three phases will be considered as reference throughout the manuscript since no other combination of CSPs performs better applying the entire data set for CSP selection.

4.2. Test-set reductions

As mentioned above, the compounds in a reduced test-set are selected using the KS algorithm. The test-set with the predetermined size is then applied to select the chromatographic systems. Reductions of the initial test-set are made per step of 10% and this up to 70% reduction. Because no difference is observed in the selected molecules using either the 3D or all available descriptors, only the results, based on describing the molecules with 3D descriptors are finally discussed.

For every reduced set the results for the remaining compounds are first re-analyzed on the reference chromatographic systems as an exploratory analysis. This means that the selected compounds will be treated as a new test-set subjected to a screening on the eight CSPs and on the two mobile phases, ACN/DEA/TFA and MeOH/DEA/TFA. Secondly, it is investigated which CSPs would be selected by each of these reduced test-sets, and whether the obtained success rate is conserved in the new selections, in comparison with the results obtained in the selection based on the entire test-set.

4.2.1. Exploratory analysis

After a reduction of 10%, the compounds carbinoxamine, esmolol, flurbiprofen, ketoprofen, metoprolol and salbutamol are excluded (Table 2). These molecules contain the most similar structural information compared to selected compounds and are therefore not selected by the KS algorithm. When the remaining 52 compounds are screened at the earlier established conditions, ACN/DEA/TFA gives 36/52 (69%) separations, while screening with MeOH/DEA/TFA gives 21/52 (40%) separations. The sequential screening of both mobile phases on the three most enantioselective columns CSP5, CSP2 and CSP8, gives cumulatively 40/52 (77%) separated compounds. The increased global success rate of 77% relative to the individual mobile phases on these CSPs, indicates a certain degree of complementarity between both mobile phases.

To obtain a test-set reduction of 20%, next to the previously mentioned compounds, alprenolol, betaxolol, ibuprofen, mepindolol, oxprenolol and sotalol are excluded. Results obtained with this test-set are, 21/46 (67%), 19/46 (41%) and 35/46 (76%) for the above-mentioned mobile phases (Table 3). A 30% reduced test-set gives success rates of 29/41 (71%) for the ACN-based, 16/41 (39%) for the MeOH-based and 32/41 (78%) for both mobile phases. The additionally excluded compounds are fenoprofen, mianserin, naproxen, pindolol and terbutaline. Similar success rates are seen with a 40% reduced test-set. In this set, 24/35 (69%), 15/35 (43%) and 27/35 (77%) compounds are separated by the ACN-based, the MeOH-based and both mobile phases, respectively. For this test-set we had the additional exclusion of bisoprolol, bupranolol, carazolol, nitrendipine, tertatolol and warfarin.

For the larger reductions, the percentages of the success rates of the ACN-based mobile phases remain very similar to those of the smaller reduction as illustrated in Table 3. Another deduction that can be made from the results is the increasing success rate of the

Table 2
Cumulative success rates for the different reduced test-sets on the reference column sequence CSP5 > CSP2 > CSP8.

Reduction	Number of remaining molecules	Mobile phase		Combined results	Excluded molecules
		ACN/DEA/TFA	MeOH/DEA/TFA	ACN/DEA/TFA + MeOH/DEA/TFA	
Cumulative success rates					
0%	58	40/58 (69%)	21/58 (36%)	44/58 (76%)	–
10%	52	36/52 (69%)	21/52 (40%)	40/52 (77%)	Carbinoxamine, esmolol, flurbiprofen, ketoprofen, metoprolol, salbutamol
20%	46	31/46 (67%)	19/46 (41%)	35/46 (76%)	Previous ones + alprenolol, betaxolol, ibuprofen, mepindolol, oxprenolol, sotalol
30%	41	29/41 (71%)	16/41 (39%)	32/41 (78%)	Previous ones + fenoprofen, mianserine, naproxen, pindolol, terbutaline
40%	35	24/35 (69%)	15/35 (43%)	27/35 (77%)	Previous ones + bisoprolol, bupranolol, carazolol, nitrendipine, tertatolol, warfarine
50%	29	20/29 (69%)	14/29 (48%)	22/29 (76%)	Previous ones + atenolol, dimethindene, ephedrine, isothipendyl, meptazinol, nadolol
60%	23	15/23 (65%)	11/23 (48%)	17/23 (74%)	Previous ones + atropine, labetalol, nitrendipine, praziquantel, promethazine, sulpiride, suprofen
70%	17	12/17 (71%)	8/17 (47%)	13/17 (76%)	Previous ones + acenocoumarol, ambucetamide, bopindolol, chlorphenamine, naringenin, nisoldipine

Table 3
Column sequence selected after reducing the test-set.

Reduction	Number of Molecules	CSP Sequence	Cumulative success rates		
			Mobile phase		Combined results
			ACN/DEA/TFA	MeOH/DEA/TFA	ACN/DEA/TFA + MeOH/DEA/TFA
0%	58	CSP5 > CSP2 > CSP8	40/58 (69%)	21/58 (36%)	44/58 (76%)
10%	52	CSP5 > CSP2 > CSP8	36/52 (69%)	21/52 (40%)	40/52 (77%)
20%	46	CSP5 > CSP2 > CSP8	31/46 (67%)	19/46 (41%)	35/46 (76%)
30%	41	CSP5 > CSP2 > CSP8	29/41 (71%)	16/41 (39%)	32/41 (78%)
40%	35	CSP5 > CSP3 > CSP2	23/35 (66%)	15/35 (43%)	28/35 (80%)
		Sequence applied on 58 molecules	*36/58 (62%)*	*21/58 (36%)*	*42/58 (72%)*
		CSP5 > CSP2 > CSP8	24/35 (69%)	15/35 (43%)	27/35 (77%)
		Sequence applied on 58 molecules	*40/58 (69%)*	*21/58 (36%)*	*44/58 (76%)*
		CSP5 > CSP2 > CSP1	23/35 (66%)	15/35 (43%)	26/35 (74%)
		Sequence applied on 58 molecules	*37/58 (64%)*	*23/58 (40%)*	*42/58 (72%)*
50%	29	CSP5 > CSP3 > CSP2	19/29 (66%)	14/29 (48%)	24/29 (83%)
		Sequence applied on 58 molecules	*36/58 (62%)*	*21/58 (36%)*	*42/58 (72%)*
		CSP5 > CSP2 > CSP8	20/29 (69%)	14/29 (48%)	22/29 (76%)
		Sequence applied on 58 molecules	*40/58 (69%)*	*21/58 (36%)*	*44/58 (76%)*
		CSP5 > CSP2 > CSP1	19/29 (66%)	14/29 (48%)	22/29 (76%)
		Sequence applied on 58 molecules	*37/58 (64%)*	*23/58 (40%)*	*42/58 (72%)*
60%	23	CSP5 > CSP3 > CSP2	15/23 (65%)	12/23 (52%)	19/23 (83%)
		Sequence applied on 58 molecules	*36/58 (62%)*	*21/58 (36%)*	*42/58 (72%)*
		CSP5 > CSP2 > CSP1	15/23 (65%)	12/23 (52%)	18/23 (78%)
		Sequence applied on 58 molecules	*37/58 (64%)*	*23/58 (40%)*	*42/58 (72%)*
		CSP5 > CSP2 > CSP8	15/23 (65%)	11/23 (48%)	17/23 (74%)
		Sequence applied on 58 molecules	*40/58 (69%)*	*21/58 (36%)*	*44/58 (76%)*
70%	17	CSP5 > CSP3 > CSP2	12/17 (71%)	8/17 (47%)	14/17 (82%)
		Sequence applied on 58 molecules	*36/58 (62%)*	*21/58 (36%)*	*42/58 (72%)*
		CSP5 > CSP2 > CSP8	12/17 (71%)	8/17 (47%)	13/17 (76%)
		Sequence applied on 58 molecules	*40/58 (69%)*	*21/58 (36%)*	*44/58 (76%)*

The italic text refers to the results obtained after application of the mentioned CSP sequence on the initial test set.

MeOH mobile phase with a further decreasing number of components. The improved success rate for MeOH with smaller test-sets, is caused by the fact that the separated compounds are slightly more selected. This increased percentages are also the consequence of the 'statistics of small numbers', i.e. for the three smallest sets a difference of one component means already a percentage change of more than 3%.

The cumulative success rates are for both mobile phases comparable for all test-sets. The above indicates that the selected subsets are representative to evaluate enantioselectivity for the initial test-set, i.e. in each set similar fractions of separated and not-separated compounds are selected. This observation could not be predicted a priori but was hoped for. The KS algorithm selects compounds described by structural properties that are quantified (i.e. the receptors). By applying the KS algorithm for subset selection in our study, one expects that the structural properties and the observed enantioselectivities would be related, else the subsets would not be representative. The obtained results demonstrate that this is the case.

During the exploratory analysis, it might also be interesting to have a closer look to the selected and the excluded compounds. From a pharmaceutical point of view, compounds belonging to the same pharmacological family have a similar pharmacophore (active site of the molecule) and thus should be similar in their interactions with the target receptors. This could imply that the compilation of a test-set containing only one compound of each pharmacological group should be sufficient to define a 'standard test-set'. However, the experimental results of chiral separations are not always that unambiguous. It is not unusual to separate a compound from a certain pharmacological group into its enantiomers while another compound from that group, that only differs in a small substituent, cannot be separated under the same experimental conditions. This is, for example, the case with nisoldipine and nitrendipine, two calcium channel blockers with very similar structures, where the former compound is separated on CSP5 with ACN/DEA/TFA as well as with MeOH/DEA/TFA, while the latter is not [10]. Another similar example concerns alprenolol and oxprenolol, two β-blockers for which separations are obtained for oxprenolol on CSP5 and on CSP8 with ACN/DEA/TFA but not for alprenolol.

Because of this, it is difficult to compile any generically representative screening set. The use of descriptors and a selection algorithm makes it at least possible to ensure a diverse selection from a larger set, based on mathematical information derived from structural properties. The compounds that are first eliminated are carbinoxamine, esmolol, flurbiprofen, ketoprofen, metoprolol and salbutamol. The compounds that are always selected by the algorithm and thus are present in every test-set are acebutolol, carvedilol, chlorthalidone, hexobarbital, leucovorin, mandelic acid, mebeverine, methadone, nicardipine, nimodipine, oxazepam, procyclidine, propranolol, propiomazine, salmeterol, tetramisole and verapamil. These compounds can be considered – based on the descriptors – as the most diverse from a structural point of view. A comparison between these two groups illustrates the above discussed: for example esmolol and metoprolol, two β-blockers, are excluded already from the first selection, while acebutolol, carvedilol and propranolol, another three β-blockers, remained after every reduction. The exclusion of a number of β-blockers indicates their structural and descriptional similarity with selected compounds. The fact that three β-blockers remained in the smallest test-set indicates that their descriptors are sufficiently different and that they are not only mainly reflecting the common structural features of β-blockers.

4.2.2. Column selection from reduced test-sets

The main objective of this work is to study whether the selected chromatographic systems remain the same, when only the analytes of the reduced test-sets are screened. Therefore, for every reduced test-set, column selection is performed and the results are compared to the initial selection.

Table 4
The most frequently selected column sequences considered for the larger test-sets.

Reduction	Number of molecules	CSP sequence	Cumulative success rates		
			Mobile phase		Combined results
			ACN/DEA/TFA	MeOH/DEA/TFA	ACN/DEA/TFA + MeOH/DEA/TFA
0%	58	CSP5 > CSP2 > CSP8	40/58 (69%)	21/58 (36%)	44/58 (76%)
		CSP5 > CSP3 > CSP2	36/58 (62%)	21/58 (36%)	42/58 (72%)
		CSP5 > CSP2 > CSP1	37/58 (64%)	23/58 (40%)	42/58 (72%)
10%	52	CSP5 > CSP2 > CSP8	36/52 (69%)	21/52 (40%)	40/52 (77%)
		CSP5 > CSP3 > CSP2	33/52 (63%)	21/52 (40%)	39/52 (75%)
		CSP5 > CSP2 > CSP1	33/52 (63%)	22/52 (42%)	38/52 (73%)
20%	46	CSP5 > CSP2 > CSP8	31/46 (67%)	19/46 (41%)	35/46 (76%)
		CSP5 > CSP3 > CSP2	28/46 (61%)	18/46 (39%)	34/46 (74%)
		CSP5 > CSP2 > CSP1	27/46 (59%)	20/46 (43%)	32/46 (70%)
30%	41	CSP5 > CSP2 > CSP8	29/41 (71%)	16/41 (39%)	32/41 (78%)
		CSP5 > CSP3 > CSP2	26/41 (63%)	16/41 (39%)	31/41 (76%)
		CSP5 > CSP2 > CSP1	26/41 (63%)	17/41 (41%)	30/41 (73%)
40%	35	CSP5 > CSP3 > CSP2	23/35 (66%)	15/35 (43%)	28/35 (80%)
		CSP5 > CSP2 > CSP8	24/35 (69%)	15/35 (43%)	27/35 (77%)
		CSP5 > CSP2 > CSP1	23/35 (66%)	15/35 (43%)	26/35 (74%)

The column sequence that in absolute numbers gives the best results for test-set reductions going from 10% till 30% is the same as originally proposed for the complete set i.e. CSP5 > CSP2 > CSP8. The success rates for these test-sets are those mentioned in Table 3. Because of the equal column sequence, the extrapolation of the results to the complete test-set is identical to the results of the screening with 58 compounds.

For a test-set reduction of 40%, three column-sequence combinations are interesting because they give similar results and thus no real distinction in performance between these sequences can be made (see Table 3). The first possibility is CSP5 > CSP3 > CSP2. The success rates obtained for this sequence with the 35 compounds test-set are 23/35 with ACN/DEA/TFA, 15/35 with MeOH/DEA/TFA and 28/35 with both mobile phases. When screening the complete test-set with this sequence, the results are 36/58, 21/58 and 42/58 respectively. The initially selected sequence CSP5 > CSP2 > CSP8 can be considered as a second best: 23/35 separations with the ACN-based mobile phase, 15/35 with the MeOH-based mobile phase and 27/35 with both mobile phases. Finally a third interesting column sequence is CSP5 > CSP2 > CSP1 with 23/35, 15/35 and 26/35 separations respectively. For the reduction of 50%, when only half of the initial compounds are selected, the same sequences are observed. The success rate obtained for the sequence CSP5 > CSP3 > CSP2 with ACN/DEA/TFA is 19/29, with MeOH/DEA/TFA 14/29 and for both mobile phases 24/29. When screening the complete test-set with this column sequence, the results are 36/58 for ACN/DEA/TFA, 21/58 for MeOH/DEA/TFA and 42/58 for the two mobile phases. A second possibility is the screening with CSP5 > CSP2 > CSP8. Screening of the 29 compounds on these columns with the ACN-based mobile phase gives 20 separations, with the MeOH-based 14 and cumulatively for both mobile phases 22/29 compounds are separated. When the complete test-set is screened on this sequence of CSPs the outcome is 40/58 for ACN/DEA/TFA, 21/58 for MeOH/DEA/TFA and 44/58 considering both mobile phases. A third interesting column sequence is again CSP5 > CSP2 > CSP1. The same success rates as with the first column sequence are obtained for the individual mobile phases but the complementarity of the mobile phases in this case is different on the latter sequence. Both mobile phases only separate 22/29 compounds (Table 3).

Selecting the column sequence based on 23 compounds (60% test-set reduction) gives again three interesting possibilities (the same CSP combinations but different preferences). The sequence with the highest success rates is still CSP5 > CSP3 > CSP2; it gives 15/23 separations with ACN, 12/23 with MeOH and 19/23 cumulatively (see Table 4). When the 58 compounds are tested, the success rates are 36/58 for ACN, 21/58 for MeOH and 42/58 for both mobile phases. Secondly CSP5 > CSP2 > CSP1, which is selected as third sequence for 40% and 50% reductions, tends to be a successful combination with again 15/23 separations for ACN, 12/23 for MeOH and 18/23 for the two mobile phases. Performing screening with the complete test-set on this sequence, results in 36/58, 23/58 and 40/58 separations for ACN, MeOH and both mobile phases respectively. Based on the experiments with the 23 compounds, the original sequence CSP5 > CSP2 > CSP8 is the third possibility. Here again 15/23 compounds are separated by ACN, 11/23 by MeOH and 17/23 cumulatively.

For a reduction rate of 70% the first selected sequence is again identical to the previous selections: CSP5 > CSP3 > CSP2. In this case the results are 12/17, 8/17 and 14/17 for the different mobile phases (see Table 3). The results for the 58 compounds on this sequence is as reported above. The original sequence CSP5 > CSP2 > CSP8 comes out as second best sequence with 12/17, 8/17 and finally 13/17 separations.

Because the sequences CSP5 > CSP3 > CSP2 and CSP5 > CSP2 > CSP1 are preferentially selected by the reduced test-sets, it was worth looking to the results of larger test-sets on these sequences. There again it seems that there is no big difference in enantioselectivity between the different system sequences (Table 4). The initially selected sequence CSP5 > CSP2 > CSP8 is slightly more successful.

From the above we conclude that a test-set reduction with 30% (i.e. till 41 compounds) results in a test-set providing similar information as the entire set and allowing a similar selection of chiral systems. But it must be remarked that the difference with the preferred sequences obtained after 40% reduction or more, is small. When chiral screening strategies are defined, the first chromatographic system selected is usually the one with the most separations, i.e. with the broadest enantioselectivity. For the larger test-sets, with reductions from 10% to 30%, the column selection sequence is always the same, and equal to the best sequence that was initially determined. When smaller test-sets are screened, different CSP sequences show a similar success rate. From a test-set reduction with 40% or i.e. 35 remaining compounds, the conclusions to be drawn become less straightforward due to the lower number of remaining compounds as already discussed for Table 3. Due to the low number of compounds, CSPs with a slightly lower preference from the entire test-set, show similar or better success rates as the initially selected, generally more preferred CSPs.

Table 5
Complementarity of the different sequences.

Reduction	Number of molecules	CSP sequence	Cumulative success rates			
			Mobile phase		Combined results	Differently separated compounds
			ACN/DEA/TFA	MeOH/DEA/TFA	ACN/DEA/TFA + MeOH/DEA/TFA	(+) Separated (−) Not separated
50%	29	CSP5 > CSP2 > CSP8	20/29 (69%)	14/29 (48%)	22/29 (76%)	Naringenin (−) Promethazine (+) Propiomazine (−)
		CSP5 > CSP3 > CSP2	19/29 (66%)	14/29 (48%)	24/29 (83%)	Naringenin (+) Promethazine (+) Propiomazine (+)
		CSP5 > CSP2 > CSP1	19/29 (66%)	14/29 (48%)	22/29 (76%)	Naringenin (+) Promethazine (−) Propiomazine (−)
60%	23	CSP5 > CSP3 > CSP2	15/23 (65%)	12/23 (52%)	19/23 (83%)	Propiomazine (+) Naringenin (+)
		CSP5 > CSP2 > CSP1	15/23 (65%)	12/23 (52%)	18/23 (78%)	Propiomazine (−) Naringenin (+)
		CSP5 > CSP2 > CSP8	15/23 (65%)	11/23 (48%)	17/23 (74%)	Naringenin (−) Propiomazine (−)
70%	17	CSP5 > CSP3 > CSP2	12/17 (71%)	8/17 (47%)	14/17 (82%)	Propiomazine (+)
		CSP5 > CSP2 > CSP8	12/17 (71%)	8/17 (47%)	13/17 (76%)	Propiomazine (−)

This results in the changed preferences from the screening set of columns. The complementarity of these different sequences toward the separated compounds of the reduced test-set is similar: the separated compounds are all the same except for some analytes. For the 50% reduced test-set, the differences in absolute number of separations depends on the compounds naringenin, promethazine and propiomazine (see Table 5). On the systems CSP5 > CSP3 > CSP2 the three compounds are separated while on CSP5 > CSP2 > CSP1 only naringenin is separated and on CSP5 > CSP2 > CSP8 only promethazine. Obviously a similar situation is seen for the 60% reduction: but now only for propiomazine and naringenin; promethazine is not selected by the algorithm and thus cannot be taken into account during the comparison. Finally, when the 70% reduced test-set is examined, naringenin is excluded which makes the differences in number of separated compounds only depending on propiomazine.

It should be noticed that in the above the complementarity of series of sequences is compared. The complementarity differences between individual systems will, of course, be larger. The different series, as discussed from Table 5, result in the separation of about the same compounds from the reduced test-sets. Also for the initial test-set the different series separate about the same compounds. The unseparated compounds for the three sequences CSP5 > CSP3 > CSP2, CSP5 > CSP2 > CSP1 and CSP5 > CSP2 > CSP8 are chlorphenamine, fenoprofen, ibuprofen, ketoprofen, labetalol, mandelic acid, meptazinol, naproxen, nitrendipine, procyclidine and verapamil. The first and the second sequence could not separate bupranolol and ephedrine, two compounds that are separated on the third sequence (which is the initially preferred sequence). Flurbiprofen is not separated by the first and the third sequences. Propiomazine is not separated by the second and the third sequences. Finally, bisoprolol is not separated by the first sequence, promethazine not by the second and naringenin not by the third.

5. Conclusion

In this work, the possibility to reconstitute the chromatographic screening sequence of a chiral separation strategy by using reduced test-sets is investigated. The test-set molecules are described by information included in molecular descriptors and a rational reduction is performed by using the Kennard and Stone selection algorithm. The reductions investigated reach from 10% to 70% reduction of the initial test-set. For every reduction, up to 30% (i.e. 41 remaining compounds), the initially found chromatographic sequence is reselected. When going to higher reduction rates, also other sequences are defined as equally or more interesting. The less molecules to test, the more systems achieve comparable results. Therefore it seems better to introduce a cut off value for the size of the test-set. According to the examined results a test-set with about 41 compounds (70% of the initial test-set) would be a good compromise between minimal size and possible decision making about system complementarity.

Summarized, our initial test-set can be reduced maximally with 30% without losing information on the enantioselectivity and complementarity of the tested chiral stationary phases. On the other hand, it might also be interesting to evaluate in the future how representative, both the initial test-set and the 30% reduced one, are for the population of drug molecules.

Another topic that might be studied in the future is whether taking into account the used solvent (ACN or MeOH) will affect the molecular conformations, their descriptor values and the subset selections. However, for simplicity reasons, the solvent initially was ignored. Another reason is that we also did not know whether a subset selection based on the descriptors applied (which do not distinguish between enantiomers) would be correlating with the enantioselectivity on the systems considered. Taking into account the influence of the subset on the molecular conformation, introduces a number of additional problems to be solved. A first one is that conformation of each molecule is to determine twice–once for every solvent. Secondly the reduced test-sets will be different for the different solvents. This creates a problem when cumulative success rates are to be considered. It should be carefully considered how test-set reduction is to be done when conformations of molecules is once done in ACN and once in MeOH solvents. Probably a 'smart' combination of these selections made from both solvent systems is needed to determine realistic cumulative success rates using ACN and MeOH based mobile phases in that situation. However, such study was actually considered outside the scope of this paper, but might be examined in future.

References

[1] http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q6A/Step4/Q6Astep4.pdf (accessed August 2011).
[2] W.H. Pirkle, T.C. Pochapsky, Chem. Rev. 89 (1989) 347.
[3] T.D. Booth, D. Wahnon, I.W. Wainer, Chirality 9 (1997) 96.
[4] M. Lämmerhofer, J. Chromatogr. A 1217 (2010) 814.
[5] L. Peng, S. Jayapalan, B. Chankvetadze, T. Farkas, J. Chromatogr. A 1217 (2010) 6942.
[6] L. Zhou, C. Welch, C. Lee, X. Gong, V. Antonucci, Z. Ge, J. Pharm. Biomed. Anal. 49 (2009) 964.
[7] T. Zhang, D. Nguyen, P. Franco, J. Chromatogr. A 1191 (2008) 214.
[8] S. Morante-Zarcero, I. del Hierro, M. Fajardo, I. Sierra, Anal. Chim. Acta 618 (2008) 102.
[9] A.A. Younes, D. Mangelings, Y. Vander Heyden, J. Pharm. Biomed. Anal. 55 (2011) 414.
[10] H. Ates, D. Mangelings, Y. Vander Heyden, J. Chromatogr. B 875 (2008) 57.
[11] N. Matthijs, M. Maftouh, Y. Vander Heyden, J. Sep. Sci. 29 (2006) 1353.
[12] M. Maftouh, C. Granier-Loyaux, E. Chavana, J. Marini, A. Pradines, Y. Vander Heyden, C. Picard, J. Chromatogr. A 1088 (2005) 67.
[13] N. Matthijs, C. Perrin, M. Maftouh, D.L. Massart, Y. Vander Heyden, J. Chromatogr. A 1041 (2004) 119.
[14] C. Perrin, V.A. Vu, N. Matthijs, M. Maftouh, D.L. Massart, Y. Vander Heyden, J. Chromatogr. A 947 (2002) 69.
[15] N. Matthijs, M. Maftouh, Y. Vander Heyden, J. Chromatogr. A 1111 (2006) 48.
[16] K.S.S Dossou, P. Chiap, B. Chankvetadze, A.-C. Servais, M. Fillet, J. Crommen, J. Chromatogr. A 1216 (2009) 7450.
[17] K.S.S. Dossou, P. Chiap, B. Chankvetadze, A.-C. Servais, M. Fillet, J. Crommen, J. Sep. Sci. 33 (2010) 1699.
[18] K.S.S. Dossou, P. Chiap, A.-C. Servais, M. Fillet, J. Crommen, J. Pharm. Biomed. Anal. 54 (2010) 687.
[19] K.S.S. Dossou, P. Chiap, A.-C. Servais, M. Fillet, J. Crommen, J. Sep. Sci. 34 (2011) 617.
[20] K.S.S. Dossou, P.A. Edorh, P. Chiap, B. Chankvetadze, A.-C. Servais, M. Fillet, J. Crommen, J. Pharm. Biomed. Anal. 54 (2011) 687.
[21] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, WILEY-VCH, Weinheim, Germany, 2000.
[22] R. Kennard, L. Stone, Technometrics 11 (1969) 137.
[23] H. Ates, A. A. Younes, D. Mangelings and Y. Vander Heyden, Enantioselectivity of polysaccharide-based chiral selector in polar oragnic solvents chromatography: implementation of chlorine-containing selectors in a separation strategy (to be submitted).
[24] The United States Pharmacopeia, USP 29, NF 26, 2006, United States Pharmacopeial Convention, Rockville, MD, pp. 2649.
[25] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, TALETE srl, Dragon-Software for molecular descriptor calculations-Version 5.4, 2006.