Solvent Diversity in Polymorph Screening

MORTEN ALLESØ,[1] FRANS VAN DEN BERG,[2] CLAUS CORNETT,[1] FLEMMING STEEN JØRGENSEN,[3] BENT HALLING-SØRENSEN,[1] HEIDI LOPEZ DE DIEGO,[4] LARS HOVGAARD,[1] JAAKKO AALTONEN,[5] JUKKA RANTANEN[1]

[1]Department of Pharmaceutics and Analytical Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark

[2]Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Frederiksberg C, Denmark

[3]Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Copenhagen, Denmark

[4]Analytical R & D, H. Lundbeck A/S, Valby, Denmark

[5]Division of Pharmaceutical Technology, Faculty of Pharmacy, University of Helsinki, Helsinki, Finland

Received 22 March 2007; revised 29 May 2007; accepted 2 July 2007

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/jps.21153

ABSTRACT: Selecting a diverse set of solvents to be included in polymorph screening assignments can be a challenging task. As an aid to decision making, a database of 218 organic solvents with 24 property descriptors was explored and visualized using multivariate tools. The descriptors included, among others, log*P*, vapor pressure, hydrogen bond formation capabilities, polarity, number of π-bonds and descriptors derived from molecular interaction field calculations (e.g., size/shape parameters and hydrophilic/hydrophobic regions). The data matrix was initially analyzed using principal component analysis (PCA). Results from the PCA showed 57% cumulative variance being explained in the first two principal components (PCs), although relevant information was also found in the third, fourth and fifth component, revealing distinct clusters of solvents. Since five dimensions were not suitable for visual presentation, a nonlinear method, self-organizing maps (SOMs), was applied to the dataset. The constructed SOM displayed features of clusters observed in the first three PCs, however in a more compelling way. Thus, the SOM was chosen as the visually most convenient way to display the diversity of the 218 solvents. In addition, it was demonstrated how safety aspects can be considered by labeling a large fraction of the solvents in the SOM with toxicological information. J Pharm Sci 97:2145–2159, 2008

Keywords: solvent; solid state; polymorphism; polymorph screening; crystallization; toxicity; physicochemical descriptors; molecular interaction fields; principal component analysis (PCA); self-organizing maps (SOMs)

INTRODUCTION

Polymorphism, the ability of a compound to exist in more than one crystalline form,[1] is a phenomenon of both great interest and concern for people working with solid state pharmaceutics. The different forms, that is, the thermodynamically stable and one or more metastable forms, can have markedly different physicochemical properties. This may affect the processability and dissolution profile of the active pharmaceutical ingredient as well as the bioavailability.[2,3] Changes in bioavailability, in particular, can have a significant impact on drug efficacy. Thus, choosing an inappropriate polymorph for development may potentially lead to market withdrawals and

This article contains supplementary material, available at www.interscience.wiley.com/jpages/0022-3549/suppmat.

Correspondence to: Jukka Rantanen (Telephone: 45-35336000; Fax: 45-35336030; E-mail: jtr@farma.ku.dk)

Journal of Pharmaceutical Sciences, Vol. 97, 2145–2159 (2008)

costly modifications to the manufacturing process.[4] This, along with the fact that intellectual property rights can be acquired for each polymorph, underlines the need for performing a thorough screening for polymorphs of the drug candidate at the early stage of preformulation. The overall goal of polymorph screening is to find as many crystal forms of the substance as possible in a cost-effective manner with respect to materials and man-hours consumed. To accommodate this, the crystallizations are often performed in high-throughput well-plate systems[5] although solubility issues may require larger vessels. Recently crystallization on self-assembled monolayers (SAMs)[6] and polymorph farming on chips[7] have been proposed as innovative contributions to this field, offering new insight into the role of critical process variables during the screening operation. Also, as a way of increasing throughput, rational experimental designing is imperative, aiming at covering a large multi factorial parameter space.[8,9] Maximizing solvent diversity is no exception to this and should be incorporated in a study in order to increase the probability of finding all forms.[10] In this respect, it is normal practice to deal with solvent libraries, as recently mentioned by, for example, Florence et al.[11] Two of the main difficulties associated with these libraries are which solvent properties to choose from and how to visualize the data. Identifying solvent descriptors that might affect polymorphic outcome of solvent crystallizations of drug candidates in general can be problematic. This is due to the lack of understanding of nucleation and crystallization phenomena[12,13] as well as solvent–solute interactions which are also dependent upon the physicochemical properties of the particular compound being screened.[14] In addition to this, Towler and Taylor[15] recently demonstrated that there are situations in which polymorphic outcome is independent of the interaction between solvent and solute. Taking these issues into account, relatively simple and general descriptors, which are expected to influence polymorphic outcome, should be included in the library. In this respect, several studies have shown hydrogen bond acceptor/donor (HBA/HBD) capabilities and polarity of solvents to be of particular relevance.[14,16,17] Garti et al.[18] and Khoshkhoo and Anwar,[19] proposed a kinetic mechanism for the effect of solvents in determining a particular polymorph of stearic acid and sulfathiazole, respectively; a mechanism explaining how some solvent molecules may interact with greater affinity to certain parts of a given drug molecule, thereby inhibiting nucleation and crystallization of a particular polymorph. Since drug molecules often posses hydrophilic as well as hydrophobic regions, lipophilic solvents, that is, solvents lacking hydrogen bond donors/acceptors and polar groups, may still interact with certain parts of the solute through van der Waals forces. Hence, hydrophobicity should also be taken into account when defining solvent descriptors for the library.

There is a limited amount of published work providing general recommendations for solvent selection in polymorph screening assignments. In a study by Gu et al.[20] 96 solvents were grouped by means of cluster analysis. The analysis was based on eight experimentally derived solvent parameters obtained from published literature, including hydrogen bond donor/acceptor propensities and polarity. This approach is feasible as long as the library is relatively small in size, whereas aiming at more than 200 solvents would create a lot of missing values due to a lack of information from existing literature.[21] Thus, calculated—yet reliable and interpretable—descriptors are needed to accommodate the issue of library size and completeness. Dealing with these large data matrices ultimately requires analysis by multivariate statistical tools. In this context, principal component analysis (PCA, a linear model)[22,23] and self-organizing maps (SOMs, a nonlinear model)[24] have previously been applied separately to visualize solvent diversity in regards to organic synthesis in general.

For this study a database encompassing 218 organic solvents times 24 solvent descriptors was generated. Identifiers such as CAS number, SMILES-notation, boiling and melting point class and safety related information was added to further enhance the value of the database. Solvent data were analyzed using PCA and SOM as two mathematically different approaches. The aim was to visualize the solvent diversity in a manner that allows for convenient and rapid selection of diverse solvents by researchers involved in screening activities.

METHODS

The Solvent Database

Two hundred eighteen organic solvents were selected spanning a wide range of different chemical classes (alcohols, acids, amines, aromatic

 DOI 10.1002/jps

solvents, alkanes, halides, etc.), ensuring that both polar protic, polar aprotic and nonpolar solvents were represented in the database. Water was excluded from the database, mostly because of its extreme polarity and hydrogen bond formation capabilities. In practice though, because of unique chemical properties and low toxicity, water should always be included when screening for crystal forms. The solvents, including solvent IDs, are listed in Table 1.

Practical identifiers include: CAS number, SMILES notation, boiling point class, melting point class, as well as safety-related identifiers: flammability class,[25] Generally Recognized As Safe (GRAS) solvents[26] and ICH Q3C solvents.[27] A detailed description of all identifiers is available in Supporting Information.

Solvent Descriptors

Twenty-four solvent descriptors were chosen to build the data matrix; these *X*-variables include properties that are relevant in the interaction between solvent and solute and thus polymorphic outcome such as: size/shape, volatility, hydrophilicity/lipophilicity and the balance between these, HBA/HBD capabilities, polarity, and electron distribution (Tab. 2). Acids and bases were treated as neutral molecules with respect to the obtained descriptor values. Surface tension—as a measure of intermolecular forces—was not included due to lack of experimental data. A more detailed explanation of some of the descriptors is formulated below and additional information is available in Supporting Information.

Experimental values of octanol/water partition coefficient ($\log P$) and vapor pressure (VP at 25°C) in mmHg for 183 and 210 solvents, respectively, were obtained through Syracuse Research Corporation.[28] Values of $\log P$ and VP at 25°C for the remaining solvents were acquired from the CAS Registry *via* the search engine SciFinder Scholar (American Chemical Society, version 2006), originally calculated using Advanced Chemistry Development (ACD/Labs, version 8.14) for Solaris.

Because of a limited amount of information available from the literature, a set of descriptors, which could be easily calculated for all solvents, were created. These descriptors include, for example, number of freely rotatable bonds (FRB), hydrogen bond acceptors (HBAs) and donors (HBDs) which were obtained from the CAS Registry *via* the search engine SciFinder Scholar (American Chemical Society, version 2006), originally calculated using Advanced Chemistry Development (ACD/Labs, version 8.14) for Solaris. A polarity descriptor (Pol) was created by counting the number of polar atoms such as: oxygen, nitrogen, sulfur, and phosphor per solvent molecule. In addition to Pol, electron distribution (termed bonding index; BI) was also determined by counting the total number of π-bonds in each solvent molecule, a double and a triple bond contributing with one and two π-bonds, respectively. The number of fluorine atoms per solvent molecule gave the so-called fluorine index (FI) which was included because of the fluorine atom's unique ability to act as a HBA.[29] To compensate for molecular size all values of FRB, HBA, HBD, Pol, BI, and FI were divided separately by the volume ($Å^3$) and surface area ($Å^2$) of the solvent in question (volume and surface area being *VolSurf* descriptors; see below). This way a molecule such as glycerol, having three HBA/HBD atoms and three polar oxygen atoms as well as a small size, will be correctly weighted as a solvent having high hydrogen bond capabilities while also being polar. A similar approach to scaling has been used successfully in a previous study regarding prediction of contact angles of pharmaceutical solids.[30] Finally, aromaticity (Ar) was expressed as the number of aromatic rings in a solvent molecule and is a crude measure of the ability to participate in aromatic π-complexes with solutes—as well as solvent molecules—due to the unique electronic configuration of the aromatic ring.[31] Since aromatic rings are planar, this descriptor is also an indicator of the *flatness* of a particular solvent.

A set of *VolSurf* descriptors was calculated to obtain additional information on the balance between the hydrophilicity and hydrophobicity within each solvent molecule and also information on the size and shape of the different solvents. This is performed by the generation and processing of molecular interaction fields, which is an approach most commonly used in Quantitative Structure-Activity Relationship (QSAR) and Absorption, Distribution, Metabolism, and Excretion (ADME) studies.[32] The procedure encompasses:[33] (1) calculating 3D molecular interaction fields between a target (the solvent) and a selected probe captured in a virtual grid, and (2) extracting the physicochemical information present in the 3D maps and converting it into numerical descriptors.

Conversion of molecular solvent structures from SMILES codes to 3D coordinates was performed in Concord as integrated into the Sybyl software package (Tripos Associates Inc.,

Table 1. Solvent List

No.	Solvent Name	No.	Solvent Name	No.	Solvent Name
1	1,1,1-Trichloroethane	74	2-Phenylethanol	147	Formic acid
2	1,1,2,2-Tetrachloroethane	75	3-Methylpyridine	148	Glycerol
3	1,1,2-Trichloroethane	76	3-Pentanone	149	Heptane
4	1,1,2-Trichloroethene	77	4-Heptanone	150	Hexadecane
5	1,1-Dichloroethane	78	4-Methylpyridine	151	Hexafluorobenzene
6	1,1-Dichloroethene	79	5-Nonanone	152	Hexane
7	1,2,4-Trimethylbenzene	80	Acetic acid	153	Hexanoic acid
8	1,2-Dibromoethane	81	Acetone	154	Iodobenzene
9	1,2-Dichlorobenzene	82	Acetonitrile	155	Iodoethane
10	1,2-Dichloroethane	83	Acetophenone	156	Iodomethane
11	1,2-Diethoxyethane	84	Allyl alcohol	157	Isoamyl alcohol
12	1,2-Dimethoxyethane	85	α,α,α-Trifluorotoluene	158	Isobutyl acetate
13	1,2-Ethanediamine	86	Aniline	159	Isobutyl alcohol
14	1,2-Ethanediol	87	Anisole	160	Isopropyl acetate
15	1,2-Propanediol	88	Benzaldehyde	161	Isopropyl alcohol
16	1,4-Dioxane	89	Benzene	162	Isopropylacetone
17	1,5-Pentanediol	90	Benzonitrile	163	*m*-Cresol
18	1-Bromo-2-methylpropane	91	Benzyl alcohol	164	Mesitylene
19	1-Bromooctane	92	Benzyl chloride	165	Methanol
20	1-Bromopentane	93	Bromobenzene	166	Methyl acetate
21	1-Bromopropane	94	Bromoethane	167	Methyl benzoate
22	1-Butanol	95	Bromoform	168	Methyl formate
23	1-Chlorobutane	96	Butanenitrile	169	Methyl isopropyl ketone
24	1-Chlorohexane	97	Butanoic acid	170	Methyl propionate
25	1-Chloropentane	98	Butanoic acid, methyl ester	171	Methylcyclohexane
26	1-Chloropropane	99	Butyl acetate	172	Morpholine
27	1-Decanol	100	Butylamine	173	*m*-Xylene
28	1-Fluorooctane	101	Butylbenzene	174	*N*,*N*-Dimethylacetamide
29	1-Heptanol	102	Butyraldehyde	175	*N*,*N*-Dimethylformamide
30	1-Hexanol	103	Carbon disulfide	176	Nitrobenzene
31	1-Hexene	104	Carbon tetrachloride	177	Nitroethane
32	1-Hexyne	105	Chlorobenzene	178	Nitromethane
33	1-Iodobutane	106	Chloroform	179	*N*-Methylaniline
34	1-Iodohexadecane	107	*cis*-1,2-Dimethylcyclohexane	180	*N*-Methylformamide
35	1-Iodopentane	108	*cis*-Decalin	181	Nonane
36	1-Iodopropane	109	Cumene	182	*o*-Cresol
37	1-Methyl-2-pyrrolidone	110	Cyclohexane	183	Octane
38	1-Methylnaphthalene	111	Cyclohexanol	184	*o*-Xylene
39	1-Nitropropane	112	Cyclohexanone	185	*p*-Cymene
40	1-Nonanol	113	Cyclopentane	186	Pentadecane
41	1-Octanol	114	Cyclopentanol	187	Pentanal
42	1-Pentanamine	115	Cyclopentanone	188	Pentane
43	1-Pentanol	116	Decalin	189	Pentanoic acid
44	1-Pentene	117	Decane	190	Pentyl acetate
45	1-Propanamine	118	Dibromomethane	191	Perfluorodecalin
46	1-Propanol	119	Dibutyl ether	192	Perfluoroheptane
47	1-Propyl acetate	120	Dichloromethane	193	Perfluorohexane
48	2,2,2-Trifluoroethanol	121	Diethyl carbonate	194	Perfluorooctane
49	2,2,4-Trimethylpentane	122	Diethyl ether	195	Perfluorotoluene
50	2,4-Dimethylpentane	123	Diethyl sulfide	196	Propanal
51	2,4-Dimethylpyridine	124	Diethylamine	197	Propanenitrile
52	2,6-Dimethylpyridine	125	Diethylene glycol dimethyl ether	198	Propanoic acid
53	2-Bromopropane	126	Diiodomethane	199	*p*-Xylene
54	2-Butanol	127	Diisopropyl ether	200	Pyridine
55	2-Butanone	128	Diisopropylamine	201	Quinoline
56	2-Butoxyethanol	129	Diisopropylethylamine	202	sec-Butylbenzene
57	2-Chlorobutane	130	Dimethoxymethane	203	Sulfolane
58	2-Chlorotoluene	131	Dimethyl disulfide	204	*tert*-Butanol
59	2-Ethoxyethanol	132	Dimethyl sulfoxide	205	*tert*-Butyl methyl ether
60	2-Furaldehyde	133	Diphenylether	206	*tert*-Butylbenzene
61	2-Heptanone	134	Dipropyl ether	207	Tetrachloroethene
62	2-Hexanone	135	Dipropylamine	208	Tetrahydrofuran
63	2-Methoxyethanol	136	Dodecane	209	Tetralin
64	2-Methylheptane	137	E-1,2-Dichloroethene	210	Thiophene
65	2-Methylpentane	138	E-2-Pentene	211	Thiophenol
66	2-Methylpyridine	139	Ethanethiol	212	Toluene
67	2-Methyltetrahydrofuran	140	Ethanol	213	*trans*-Decalin
68	2-Nitropropane	141	Ethoxybenzene	214	Tributyl phosphate
69	2-Nitrotoluene	142	Ethyl acetate	215	Triethylamine
70	2-Octanol	143	Ethyl formate	216	Trifluoroacetic acid
71	2-Octanone	144	Ethylbenzene	217	Undecane
72	2-Pentanol	145	Fluorobenzene	218	Z-1,2-Dichloroethene
73	2-Pentanone	146	Formamide		

 DOI 10.1002/jps

Table 2. Descriptor List

No.	Abbrev.	Descriptor Name	No.	Abbrev.	Descriptor Name
1	MW	Molecular weight (g mol^{-1}) [32.04; 462.1]	13	BI/area	Bonding index ($Å^{-2}$) [0; 0.0210]
2	Log*P*	Partition coefficient (octanol/water) [-2.04; 9.47]	14	Ar	Aromaticity [0; 2]
3	VP	Vapor pressure (mmHg at 25°C) [0.000057; 635]	15	FI/vol	Fluorine index ($Å^{-3}$) [0; 0.0346]
4	FRB/vol	Freely rotatable bonds ($Å^{-3}$) [0; 0.0221]	16	FI/area	Fluorine index ($Å^{-2}$) [0; 0.0505]
5	FRB/area	Freely rotatable bonds ($Å^{-2}$) [0; 0.0258]	17	V	Molecular volume ($Å^3$) [123.88; 831.12]
6	HBA/vol	Hydrogen bond acceptor capability ($Å^{-3}$) [0; 0.0162]	18	S	Molecular surface area ($Å^2$) [117.97; 598.40]
7	HBD/vol	Hydrogen bond donor capability ($Å^{-3}$) [0; 0.0214]	19	R	Rugosity ($Å^3Å^{-2}$) [1.04; 1.47]
8	HBA/area	Hydrogen bond acceptor capability ($Å^{-2}$) [0; 0.0188]	20	W3	Hydrophilic regions (at -1.0 kcal mol^{-1} interaction) [0.00; 636.00]
9	HBD/area	Hydrogen bond donor capability ($Å^{-2}$) [0; 0.0251]	21	W6	Hydrophilic regions (at -4.0 kcal mol^{-1} interaction) [0.00; 59.25]
10	Pol/vol	Polarity ($Å^{-3}$) [0; 0.0162]	22	D3	Hydrophobic regions (at -0.6 kcal mol^{-1} interaction) [0.00; 115.00]
11	Pol/area	Polarity ($Å^{-2}$) [0; 0.0188]	23	D6	Hydrophobic regions (at -1.2 kcal mol^{-1} interaction) [0.00; 27.88]
12	BI/vol	Bonding index ($Å^{-3}$) [0; 0.0162]	24	HL2	Hydrophilic-hydrophobic balance (hydrophilic and hydrophobic regions measured at -4.0 and -0.8 kcal mol^{-1} interaction, respectively) [0.00; 118.50]

The range of a particular descriptor is shown in square brackets.

St. Louis, www.tripos.com, version 7.2). Prior to generation of the 3D molecular interaction fields, all solvent molecules were energy minimized by 100 iterations in Sybyl using the Merck Molecular Force Field (MMFF). 3D molecular interaction fields for all solvent molecules were generated in GRID (Molecular Discovery Ltd., United Kingdom, www.moldiscovery.com, version 22) using the water probe and hydrophobic (DRY) probe and subsequently converted into numerical descriptors using *VolSurf* (Molecular Discovery Ltd, version 4.1.4.1). Eight descriptors were selected for further analysis. These included molecular volume (V), surface area (S), rugosity (R), hydrophilic regions at -1.0 and -4.0 kcal/mol interaction (W3 and W6, respectively), hydrophobic regions at -0.6 and -1.2 kcal/mol interaction (D3 and D6, respectively), and hydrophilic-hydrophobic balance (HL2; see Tab. 2).

Data Analysis

PCA and SOMs are categorized as unsupervised multivariate tools, meaning that no *a priori* information exists to which the original data is to be correlated. Both methods have their pros and cons. The loading vectors (see below) calculated in PCA allow a convenient way of understanding the clustering of solvents with respect to the original property descriptors. In contrast to PCA, the SOM approach promises a better visual interpretation of the clustering in large datasets, due to the fact that the solvents are distributed onto nodes-vectors of a predefined organized map (see below). Unfortunately, the relationship between the observed clusters in the SOM and the original variables is not that clear-cut and becomes very challenging to deal with, especially when dealing with many property descriptors.

Thus, a good understanding of the dataset prior to SOM analysis, for example, gained from PCA, is compulsory. These considerations formed the rationale for using both methods and to investigate complementarities.

Principal Component Analysis (PCA)

PCA is a bilinear modeling method that approximates the original data table (the *X*-matrix of size 218×24) through decomposition into a set of mutually orthogonal components termed principal components (PCs) and a residual matrix, **E**.[34] Each PC is the product of orthogonal objects-score vectors, $\boldsymbol{t_i}$ (length 218) and orthonormal variables-loading vectors $\boldsymbol{p_i}$ (length 24).

$$\mathbf{X} = \mathbf{t}_1 \cdot \mathbf{p'}_1 + \mathbf{t}_2 \cdot \mathbf{p'}_2 + \cdots + \mathbf{E} \qquad (1)$$

During computation the loading vectors $\mathbf{p}_i$ are normalized to length one to avoid any scaling ambiguity. The first PC corresponds to the direction of maximum explained variance with each successive PC accounting for as much as the remaining variance as possible. Plotting two score vectors against each other will give the position of objects or samples (i.e., the rows of the *X*-matrix) in that respective PC direction, while a plot of loading vectors describes the relationship between the original variables (i.e., columns of the *X*-matrix) and the PC direction in question. Combined, the plots will provide information on how the samples behave mutually.

In the current study all X-variables were centered and scaled to unit variance prior to PCA to avoid suppression of descriptors with low numeric values. For easy interpretation, score and loading plots for a given set of PCs were combined into a so-called biplot by multiplying the normalized score vectors with a constant only after modeling, for strictly visual purposes. Correlation plot and PCA were calculated and visualized in MatLab R2006b (The MathWorks, Natick, version 7.3.0.267).

Self-Organizing Maps (SOMs)

SOMs are a nonlinear approach used for unsupervised pattern recognition that seeks to map multi dimensional data while preserving as much of the topology of the original data as possible. SOM fitting is a so-called natural algorithm that belongs to the class of artificial neural networks. In short, a SOM is constructed by the following procedure:[35,36] (1) Select and fix the size $\mathrm{A} \times \mathrm{B}$ of the map, expressed as an array of nodes. The nodes of the map are represented by vectors of the same length as the object or sample vectors. (2) Select the map structure/connectivity, the two most commonly used being "rectangular" where the nodes are connected over the horizontal and vertical paths and "hexagonal" where the connection is over the diagonal path. (3) Initialize the map by filling in random numbers in the weight vectors of all nodes, hence giving them a random orientation on the map. (4a—Training) Randomly select a sample and calculate a specified distance measure (e.g., Euclidean or Mahalanobis) between the weight-vector on each node and the sample vector. The particular sample is associated with the best-matching node, that is, having the smallest distance. (4b—Training) The weight-vector of this winning node is adjusted according to Kohonen's learning rule which includes a specified learning rate value:

$$\begin{aligned} \mathbf{w}_n(i) &= \mathbf{w}_n(i-1) + \alpha\beta(\mathbf{x}_i - \mathbf{w}_n(i-1)) \\ &= (1-\alpha\beta)\mathbf{w}_n(i-1) + \alpha\beta\mathbf{x}_i \end{aligned} \qquad (2)$$

Hence the winning weight-vector in the present iteration i, $\mathbf{w}_n$(i), is made more similar (but not equal to) the sample $\mathbf{x}_i$ used for training using learning rate $\alpha < 1$ and neighborhood weight $\beta = 1$. Similarly, the neighboring nodes are adjusted, albeit to a lesser extent than the wining node ($\beta < 1$), by a specified neighborhood function (e.g., Linear, Mexican-hat, Gaussian bell). This way a small area of the map will represent properties dictated by the present training-sample. (4c—Training) Steps 4a and 4b are repeated until all samples have been used in the first training cycle, thus completing one iteration or epoch. To achieve the final SOM model several hundreds to thousands of iterations (4a–4c) are performed. The aim is to gradually adapt the map to represent the topology of the original data set (e.g., clusters of similar samples are positioned in the same area on the node-map) by slowly adjusting the weight-vectors on the nodes towards the objects (e.g., the samples within a cluster).

As is evident from the description above numerous parameters are involved in the algorithm, with map size and number of epochs being some of the critical ones with respect to obtaining the optimal representation of the data.[36] Unfortunately, a closed form or analytical expression for the SOM is not available—rather, the method is based on the learning strategy. This, in combination with

 DOI 10.1002/jps

the random initialization of the weight-vectors and the algorithmic approach taken in natural learning methods, makes interpretation sometimes difficult, especially since no obvious quality criterion is available for the evaluation of different SOM-runs/restarts.

The SOM used in the current study was developed using the Neural Network toolbox (The MathWorks, Natick, version 5.0.1) for MatLab. A hexagonal map of size 10×10 was chosen with default settings for all parameters. All solvent descriptors were scaled to unit variance. The map was randomly initialized and trained using an adaptive learning rate starting at 0.9 going to 0.02 (going from the so-called ordering phase to the tuning phase) and locally adjusted by a simple linear neighborhood function of diameter one. The data was iterated 5000 times. Additionally, 15 consecutive runs with random starting values were performed. As *ad-hoc* evaluation criterion the following energy value is determined for each run:

$$E(\mathrm{SOM}) = \sum_{i=1}^{I} ||\mathbf{w}_n - \mathbf{x}_i|| \qquad (3)$$

where $\mathbf{w}_n$ is again the winning node (after training) for sample $\mathbf{x}_i$ and $||\,||$ is the Euclidean norm. The map having the lowest energy (lowest error) was chosen as the best solution within the 15 runs/restarts computed.

RESULTS AND DISCUSSION

Overview of Solvent Descriptors

Figure 1 shows the correlation coefficients between the 24 descriptor variables computed over the 218 solvents in the *X*-matrix. A high degree

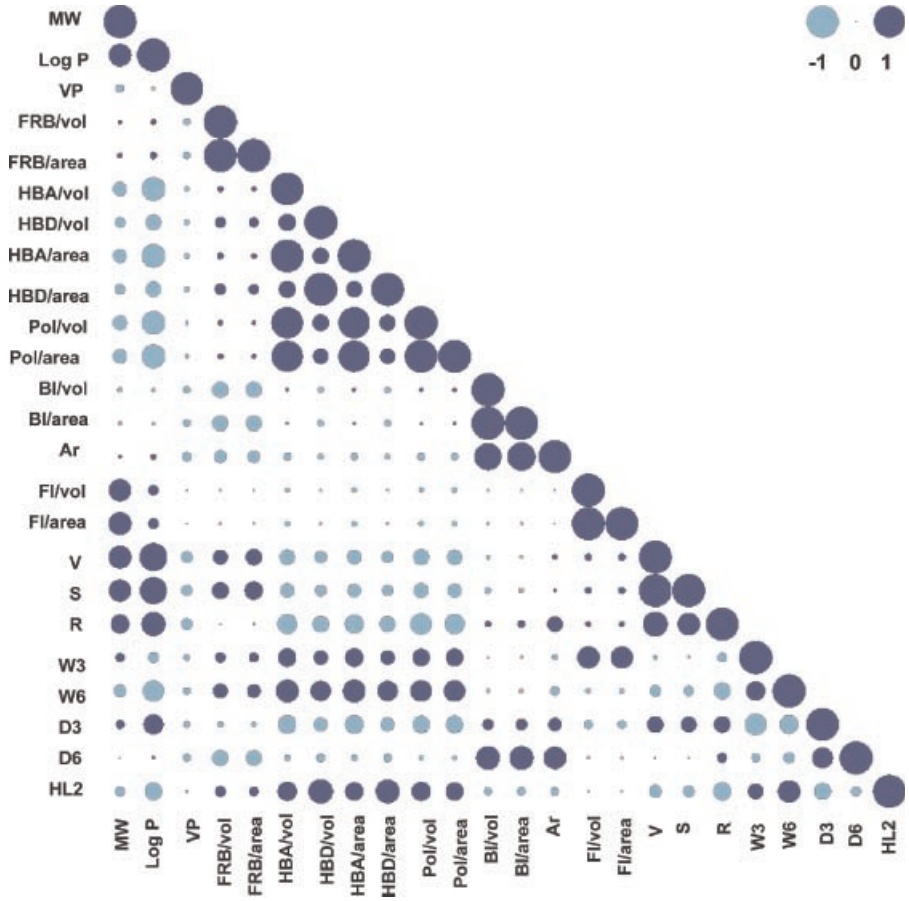

Figure 1. Correlation plot of solvent descriptors. Dot size indicates correlation strength, color indicates sign.

of correlation between volume (V) and surface area (S) is observed, which is also reflected in descriptors being weighted according to these two properties. This correlation is reasonable since all 218 solvent molecules are of small size (i.e., MW < 500 g/mol). Hydrogen bond capabilities (HBA/HBD) and polarity (Pol) descriptors are positively correlated to hydrophilic *VolSurf* descriptors such as W3, W6, and HL2, the latter describing the balance between hydrophilicity and lipophilicity. HBA/HBD, Pol, W3, W6, and HL2 correlate negatively with measures of lipophilicity, log*P* and to a certain extent MW and D3. The mentioned correlations are expected since the hydrophilicity and lipophilicity of a solvent are largely determined by its ability to form hydrogen bonds as well as the presence of polarized molecular bonds. Furthermore, the low energy hydrophobic *VolSurf* descriptor, D6, mainly contains information related to the π-bond composition (BI) and aromaticity (Ar) of the solvents. As an interesting note, W3's good correlation with FI might indicate the ability of the water probe to interact favorably with fluorine at −1.0 kcal/mol (Tab. 2), either through van der Waals forces or weak hydrogen bonds.[29,33] In contrast, this type of correlation is not seen for W6 (e.g., hydrophilic regions measured at −4.0 kcal/mol interaction). There is no simple relationship between VP and the other 23 descriptors, although a low degree of negative correlation, in particular for size/shape descriptors, is observed. It is well known that VP varies with the strength of intermolecular forces in a given solvent, as dictated by, for example, solvent–solvent hydrogen bonding and dipole–dipole interactions. The HBA/HBD descriptors included in this study are crude, merely stating the number of donors and acceptors and do not take into account the strength of the hydrogen bonds and dipole–dipole interactions. Thus—in contrast to what may be expected—no marked negative correlation is observed between hydrogen bonding descriptors (HBA/HBD) and VP.

Visualizing Solvent Diversity by Principal Component Analysis

The first two PCs, PC1 and PC2, explain 37% and 20% of the total variance, respectively. PC1 describes variance mainly related to hydrophilic and lipophilic properties; solvents with high PC1 scores being highly hydrophilic, meaning high loadings of HBA/HBD, Pol, HL2 and W6 (Fig. 2a). Examples of solvents in this category are glycerol (no. 148) as the very extreme, 1,2-ethanediamine (no. 13), 1,2-ethanediol (no. 14), 1,2-propanediol (no. 15), formamide (no. 146), and formic acid (no. 147), the latter two having lower PC2 scores due to the presence of a double bond and no FRBs (i.e., higher BI and lower FRB). The more lipophilic solvents are positioned on the opposite side of the biplot, that is, negative PC1 scores, with a homologous series of higher order alkanes and perfluoro compounds located in the 2nd quadrant (no. 34, 150, 186, 136, 217, 117, and 191–194) and most of the aromatic solvents in the 3rd quadrant as expected from the loadings of log*P*, V, S, R, MW, Ar, and D6. PC2 may also be viewed upon as describing the rigidity of the solvents, since FRB correlate negatively with Ar and BI in this component. Besides being expected from a physicochemical viewpoint, the relationship between log*P* and Pol, HBA/HBD observed in PC1 (Fig. 2a) is also in accordance with results from a comparable study by Carlson et al.,[22] where log*P* correlated negatively in PC1 with experimental measures of solvent polarity such as dielectric constant and dipole moment. Solvents having low PC1 and PC2 scores, that is, located close to the center of the PC coordinate system, are not that well described in this two component PCA model. These primarily include ethers and polar aprotic solvents such as ketones. To enhance the variance within these groups one should perform a PCA on the respective solvent cluster exclusively (not shown). Their position in the PC1 and PC2 subspace are justified though, since they roughly possess in-between numeric values of the solvent properties explained in this two-dimensional subspace.

Variance related to differences in VP, in particular, and to some extent FI, is not described in Figure 2a. To highlight these uncorrelated descriptors it is necessary to look into the higher order PCs. As seen in Figure 2b, PC3 (explaining 13% of the total variance) contains information related to VP, FI and additional information on W3, BI, Ar and MW, thus providing complementary interpretation and further clustering/grouping.

As an example hexafluorobenzene (no. 151) and perfluorotoluene (no. 195) are now located closer to its perfluoro counterparts (no. 191–194) which is in contrast to Figure 2a, where no. 151 and 195 were placed closer to the aromatic compounds. This illustrates some of the difficulties of doing unsupervised exploratory data analysis; due to a lack of response, in this case a polymorphic

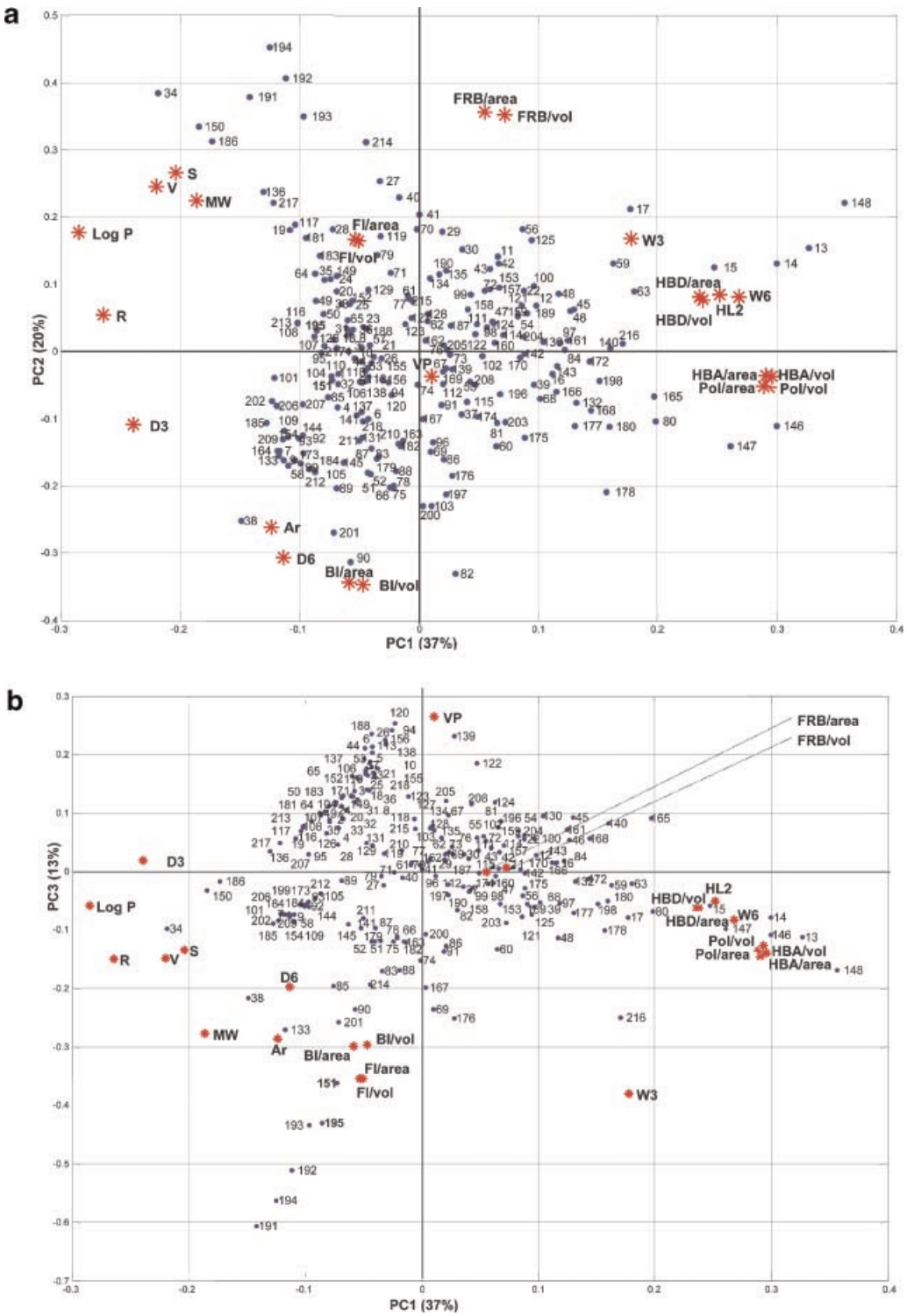

Figure 2. (a) The positions of solvents (labeled as dots) in the PC1 and PC2 subspace along with loading values (labeled as stars) for each of the original variables. (b) The positions of solvents (labeled as dots) in the PC1 and PC3 subspace along with loading values (labeled as stars) for each of the original variables.

outcome for all 218 solvents, it is not possible to unambiguously determine which type of clustering is the correct one. Based on the explained variance and loadings, five PCs (explaining cumulative 87% variance) were deemed relevant, all combinations showing distinct groupings or isolations (results not shown), although not as conclusive as the interpretation of PC1, 2 and 3. The presence of uncorrelated descriptors in the dataset clearly results in only a modest fraction of the original information being explained in the first two PCs (57%). This was also the case in the study by Carlson et al.,[22] mentioned previously, in which merely 51% of the original information based on 8 solvent properties was explained in the final two component model. Going to the higher order PCs, for example, PC3, would most likely have prompted a different clustering of the solvents, as seen from the results of this study. Considering the aim of this study, that is, to visualize the information contained in the database, only two (linear) dimensions are allowed. Obviously, the PCA approach does not fulfill this, and instead of discarding the higher order PCs, the data analysis should be supported using additional multivariate tools. In this study the raw data was modeled onto a nonlinear SOM to elaborate on the multi dimensional space.

Visualizing Solvent Diversity by Self-Organizing Maps

Generally, the clusters apparent from the SOM (Fig. 3) are in agreement with the PC1/PC2 biplot (Fig. 2a). Solvents having high HBA/HBD, Pol and low log$P$ (i.e., hydrophilic) are located in the upper left corner of the map (no. 148, 146, etc.). Highly fluorinated solvents (i.e., solvents no. 191–195, 151) are better separated compared to Figure 2a, most of these being in the lower left corner of Figure 3, opposite to the higher order alkanes (no. 34, 150, 186, 136, 217), the latter located in the upper right corner. Due to the nonlinear nature of the SOM approach, the distances between each node is not uniform throughout the map, resulting in some neighboring nodes being closer to each other than others (Fig. 3). This is denoted as the topology of the map and provides

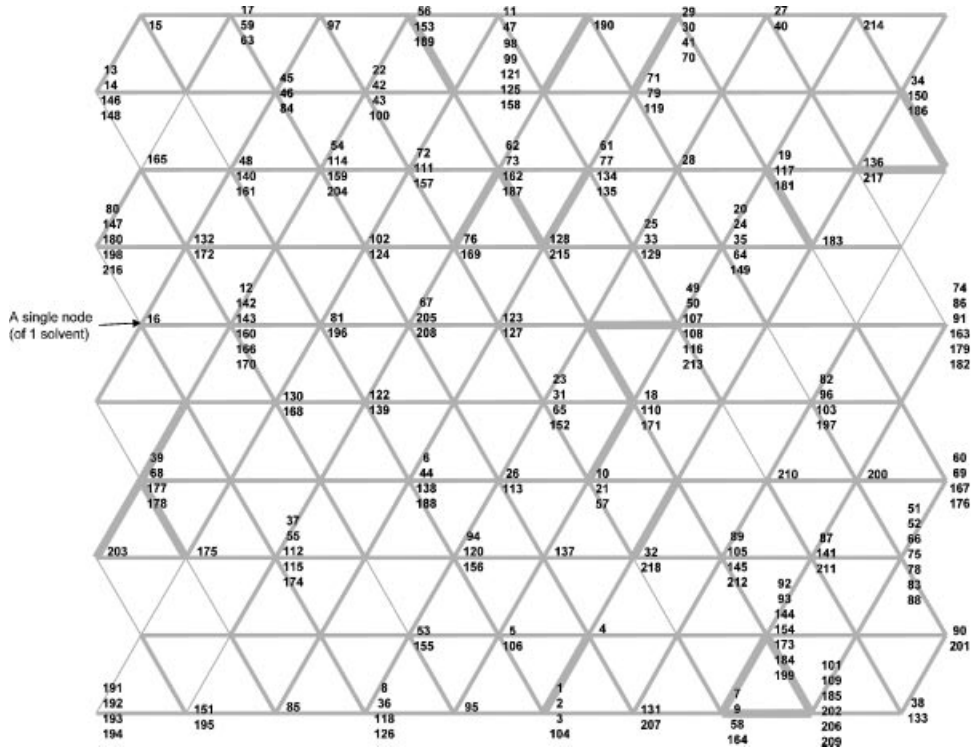

Figure 3. Locations of solvents on the 10×10 SOM map of nodes with solvent IDs situated to the right of each node. The Euclidean distances between neighboring node weight-vectors, representing the similarity or connectivity, are indicated by the thickness of connector lines, the thickest lines indicating short distances ($D < 1$) and thinnest lines large distances ($D \geq 4$).

 DOI 10.1002/jps

additional information on the clusters. For instance the previously mentioned perfluoro weight-vectors, in the lower left corner, are very far (i.e., thin lines) from its neighboring nodes above (compared to the average node distances of the map), meaning that perfluoro solvents have been correctly modeled as being very diverse. The clear separation of fluorinated solvents bears some resemblance to that as seen in PC3 (Fig. 2b) and for this reason the SOM might be viewed upon as a weighted representation of all the systematic variance present in the original data, that is, PC1, PC2, PC3 and possibly some of the higher PCs. Finally, some nodes have short relative distances between the weight-vectors, a good example being the three nodes containing aromatic compounds in the lower right corner of the map (e.g., solvents no. 7, 9, . . ., 92, 93, . . ., 101, 109). Despite being located in separate nodes these solvents are to some extent similar in physicochemical properties, at least when based on the descriptors included in the present study. Hence, it might be more rational to regard these nodes as being one common region inside the map when investigating solvent diversity using the SOM.

Safety Aspects Inside the Self-Organizing Map

When performing crystallization studies an effort should be made to ensure that the selected solvents are toxicologically safe to the extent that the experimental conditions allow for, an assertion that has been widely recognized.[14,27,37] This is especially relevant for primary manufacturing where the crystallizations are up-scaled.[38] In the current study, the inclusion of toxicological identifiers, such as GRAS and solvent classes from ICH Q3C, allow the user to take safety aspects into consideration when selecting the desired set of solvents to include in a polymorph screening. As is apparent from Figure 4, the vast majority of the least toxic solvents, that is, Class 3 (green colored) and GRAS, are located in the upper left quadrant of the map, for example, 1-propanamine (no. 45), acetic acid (no. 80) and ethyl formate (no. 143). Highly toxic Class 1 solvents (red colored)

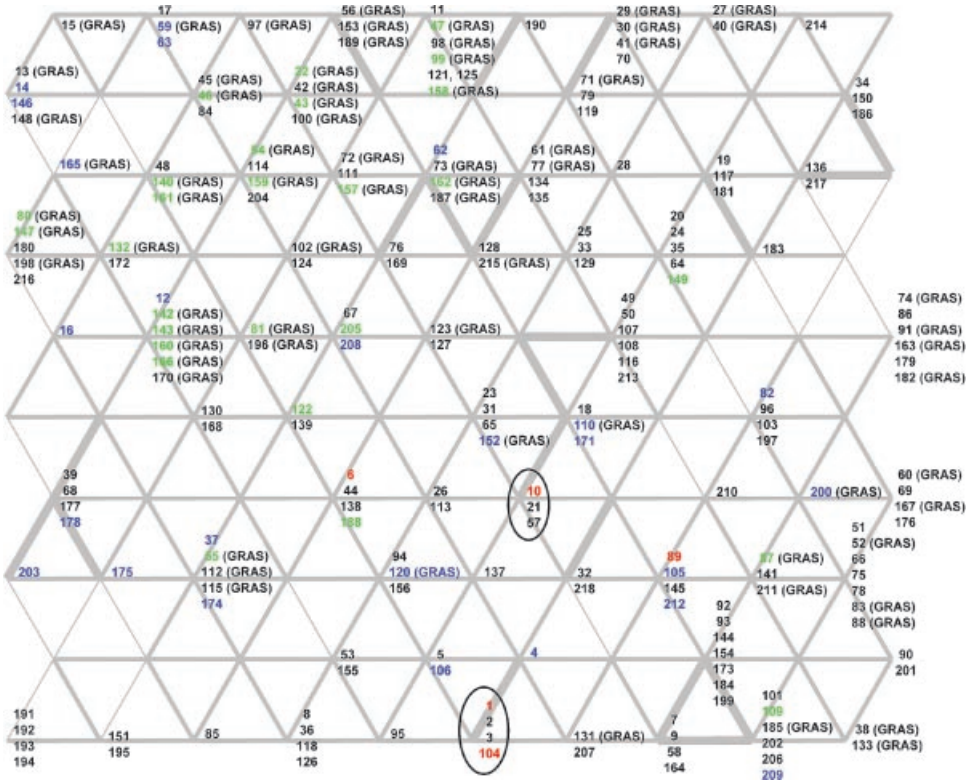

Figure 4. Safety labeled solvents in the SOM. (GRAS = Generally Recognized As Safe), (ICH Q3C: "Red = Class 1: To be avoided," "Blue = Class 2: To be limited," "Green = Class 3: Low toxic potential").

are confined to the lower middle part of the map among them the chlorinated compounds and benzene (no. 1, 6, 10, 104, and 89, respectively). The confinement of toxic solvents presents a problem in situations where the researcher wishes to maximize physicochemical diversity of the solvents while simultaneously minimizing toxicity. This is the case of two nodes, highlighted in Figure 4, that only contain information on Class 1 solvents. In crystallization experiments where toxicity is an issue this situation can be circumvented by choosing the less toxic solvent(s) from one of the neighboring nodes of that region, if such a region exists. Hence, the ICH Class 1 solvents 1,1,1-trichlorethane (no. 1) and carbontetrachloride (no. 104) may be replaced with the ICH Class 2 solvent 1,1,2-trichloroethene (no. 4), closely located in the neighboring node, albeit with slightly different physicochemical properties (i.e., a double bond). An identical approach can be applied to the Class 1 solvent 1,2-dichlorethane (no. 10). One should bear in mind that other solvents of this region, for which no toxicological information exist, might also be potentially toxic, hence care should be taken when selecting, handling and mixing[39] solvents from the lower middle region.

Selecting Solvents Using the Self-Organizing Map

Giving absolute guidelines for selecting solvents is not an easy task, and it is not the purpose of this paper to provide them either; needless to say, the selection must to a large extent depend upon the physicochemical properties of the drug-candidate in question (see below). To remain within the scope of the current study, the diversity of solvents included in previously published studies on polymorph screening will be explored, as a way of connecting the findings of this study to real-life decisions. Therefore five independent studies were identified, performing polymorph screening of the following drugs: ritonavir[40] (study 1; 17 solvents), piroxicam[41] (study 2; 9 solvents), formoterol fumarate[42] (study 3; 11 solvents), and carbamazepine (studies 4[38] and 5;[11] 11 solvents and 67 solvents, respectively), with studies 1–3 using conventional crystallization technology and studies 4–5 using high-throughput systems. All solvents included in each of the five studies, water excluded, are highlighted on the SOM (Fig. 5). Clearly, there is a preference for 10 specific solvents, with 1,4-dioxane (no. 16), 2-butanone (no. 55), acetone (no. 81), acetonitrile (no. 82), chloroform (no. 106), ethanol (no. 140), isopropyl alcohol (no. 161), methanol (no. 165), *N*,*N*-dimethylformamide (no. 175) and toluene (no. 212) appearing in at least three out of the five studies. The rationale for this selection may be due to the low toxicity of these solvents (compare with Fig. 4) and also because of the simple fact that they are considered quite standard pharmaceutical solvents, hence being readily available in the pharmaceutical industry. The latter is in accordance with a solvent table proposed by Guillory[43] who lists 15 solvents (including water) as often being used in the preparation of polymorphs. Seven out of the ten solvents identified here are included in that list. Studies 1–4 include solvents from a relatively small area of the map compared to the study by Florence et al.[11] (study 5) which covers many (but not all) of the regions in the SOM. Although, based on a different set of solvent descriptors, the large diversity observed in study 5 is expected, since Florence and coauthors deliberately selected widely diverse solvents by PCA when setting up their high-throughput crystallization experiments. This diversity is clearly repeated in the SOM computed in our work (Fig. 5).

It should be noted that the desired span of solvent diversity in screening studies can be restricted by the solubility of the solute in the respective solvents, as was the case in study 3 where crystallization of formoterol fumarate was unfeasible in 2 out of 11 solvents. From a practical viewpoint the viscosity of the selected solvents should also be considered. A highly viscous solvent, such as glycerol (no. 148) for instance, may not be suitable for high-throughput automated systems using cannulas for liquid handling.[5] Related, filtering of suspensions and subsequent washing of the generated crystals can be difficult to carry out when using viscous solvents.[43] Other issues to consider for people assigned to solvent selection are: degradation of solute, which can be favored in some solvents, and the degree of acidity/basicity of the solute. For instance, deprotonation of an acidic solvent changes the hydrogen bond formation capabilities and the hydrophilicity of the solvent, thus affecting the interaction with the solute considerably. In these cases it would be more relevant to work with log*D* of the solvents at certain pH levels and additionally calculate new values of HBA/HBD and *VolSurf* descriptors as a way to optimize the database to different chemical environments. This is out of scope for the current study, mostly because the chemical environment is also dictated by the physicochemical properties of

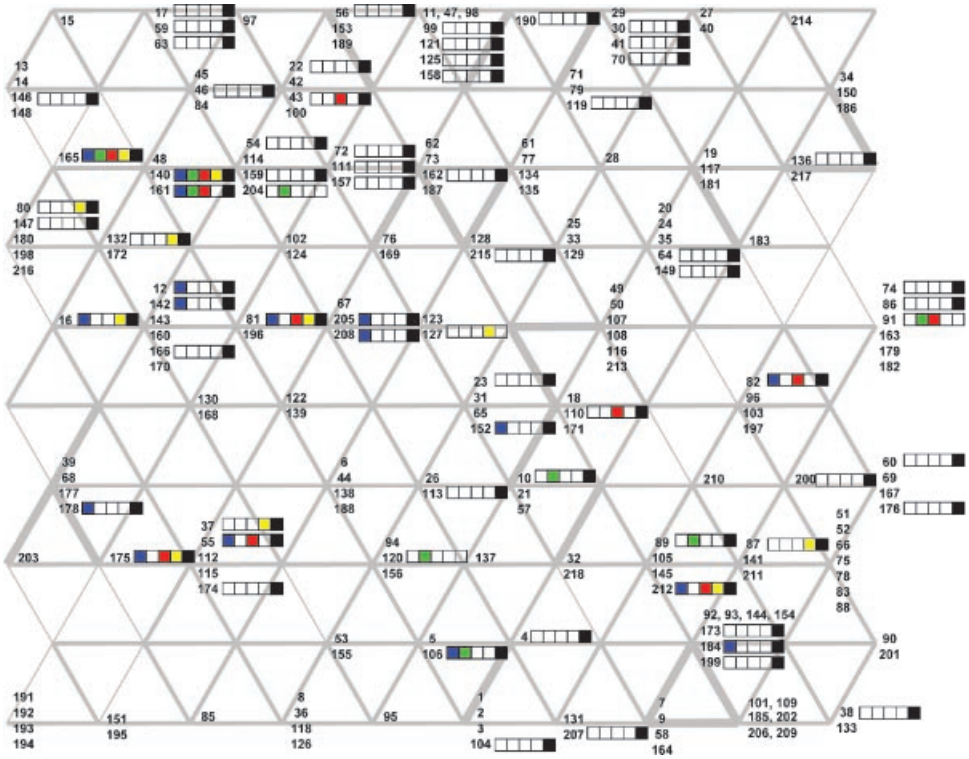

Figure 5. Selected solvents in previous polymorph screening studies. Blue = Miller et al.,[40] green = Vrecer et al.,[41] red = Jarring et al.,[42] yellow = Hilfiker et al.,[38] black = Florence et al.[11] Xylene, which was used in the study by Florence and coauthors, is highlighted as the ortho, meta, and para forms for use in this database.

the particular solute being screened. Nonetheless, solvent diversity—in general—can quickly and efficiently be maximized by use of the SOM presented in Figures 3–5.

CONCLUSIONS

Selecting a diverse enough set of solvents for use in polymorph screening can be a challenging task, and since time constraints are an important factor in the pharmaceutical industry, rapid decisions are often required. In this study a database of 218 organic solvents times 24 property descriptors was explored using PCA. The first two PCs, explaining 57% of the total variance, predominantly contained information related to lipophilicity, hydrophilicity, hydrogen bond formation capabilities, polarity, aromaticity, number of π-bonds and FRBs. Additional insight into the role of these and the remaining descriptors was found in the third PC, in particular, as well as in the fourth and fifth. However, five (linear) dimensions are undesirable with respect to a visual presentation of the diversity. Therefore, a (nonlinear) SOM was applied to the solvent data. The SOM displays features of clusters that were observed in the first three PCs of the PCA, while doing so in a more compelling way. The price for going from linear PCA to nonlinear SOM is a loss of information on the role of individual descriptors/variables in the solution (as found from the loading vectors) and a complex abstract mapping of the geometric distance between different samples/solvents. Hence, the linear and nonlinear modeling methods are clearly complementary in this study.

Although the selected 218 solvents in this study do indeed represent just a minor fraction of the vast chemical space,[44] the constructed SOM may be used as an assisting decision-making tool when selecting solvents to include in polymorph screening.

REFERENCES

1. McCrone WC. 1965. Polymorphism. In: Fox D, Labes MM, Weissberger A, editors. Physics and chemistry of the organic solid state. New York: Interscience Publishers. pp 725–767.

2. Datta S, Grant DJ. 2004. Crystal structures of drugs: Advances in determination, prediction and engineering. Nat Rev Drug Discov 3:42–57.
3. Vippagunta SR, Brittain HG, Grant DJ. 2001. Crystalline solids. Adv Drug Deliv Rev 48:3–26.
4. Chemburkar SR, Bauer J, Deming K, Spiwek H, Patel K, Morris J, Henry R, Spanton S, Dziki W, Porter W, Quick J, Bauer P, Donaubauer J, Narayanan BA, Soldani M, Riley D, McFarland K. 2000. Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. Org Process Res Dev 4:413–417.
5. Storey R, Docherty R, Higginson P, Dallman C, Gilmore C, Barr G, Dong W. 2004. Automation of solid form screening procedures in the pharmaceutical industry—how to avoid the bottlenecks. Crystallogr Rev 10:45–56.
6. Lee AY, Lee IS, Dette SS, Boerner J, Myerson AS. 2005. Crystallization on confined engineered surfaces: A method to control crystal size and generate different polymorphs. J Am Chem Soc 127:14982–14983.
7. Lee T, Hung ST, Kuo CS. 2006. Polymorph farming of acetaminophen and sulfathiazole on a chip. Pharm Res 23:2542–2555.
8. Rawlings JB, Sink CW, Miller SM. 2002. Control of crystallization processes. In: Myerson AS, editor. Handbook of industrial crystallization. 2nd edition. Boston: Butterworth-Heinemann. pp 201–230.
9. Morissette SL, Almarsson O, Peterson ML, Remenar JF, Read MJ, Lemmo AV, Ellis S, Cima MJ, Gardner CR. 2004. High-throughput crystallization: Polymorphs, salts, co-crystals and solvates of pharmaceutical solids. Adv Drug Deliv Rev 56: 275–300.
10. Hilfiker R, De Paul SM, Szelagiewicz M. 2006. Approaches to polymorphism screening. In: Hilfiker R, editor. Polymorphism: In the pharmaceutical industry. Weinheim: Wiley-VCH. pp 287–308.
11. Florence AJ, Johnston A, Price SL, Nowell H, Kennedy AR, Shankland N. 2006. An automated parallel crystallisation search for predicted crystal structures and packing motifs of carbamazepine. J Pharm Sci 95:1918–1930.
12. Davey RJ. 2004. Crystallization: How come you look so good? Nature 428:374–375.
13. Dunitz JD, Bernstein J. 1995. Disappearing polymorphs. Acc Chem Res 28:193–200.
14. Mirmehrabi M, Rohani S. 2005. An approach to solvent screening for crystallization of polymorphic pharmaceuticals and fine chemicals. J Pharm Sci 94:1560–1576.
15. Towler CS, Taylor LS. 2007. Spectroscopic characterization of intermolecular interactions in solution and their influence on crystallization outcome. Cryst Growth Des 7:633–638.
16. Cross WI, Blagden N, Davey RJ, Pritchard RG, Neumann MA, Roberts RJ, Rowe RC. 2003. A whole output strategy for polymorph screening: Combining crystal structure prediction, graph set analysis, and targeted crystallization experiments in the case of diflunisal. Cryst Growth Des 3:151–158.
17. Gu CH, Young V, Jr., Grant DJ. 2001. Polymorph screening: Influence of solvents on the rate of solvent-mediated polymorphic transformation. J Pharm Sci 90:1878–1890.
18. Garti N, Wellner E, Sarig S. 1980. Stearic-acid polymorphs in correlation with crystallization conditions and solvents. Cryst Res Technol 15:1303–1310.
19. Khoshkhoo S, Anwar J. 1993. Crystallization of polymorphs—the effect of solvent. J Phys D: Appl Phys 26:B90–B93.
20. Gu CH, Li H, Gandhi RB, Raghavan K. 2004. Grouping solvents by statistical analysis of solvent property parameters: Implication to polymorph screening. Int J Pharm 283:117–125.
21. Marcus Y. 1998. The properties of solvents. West Sussex: John Wiley & Sons.
22. Carlson R, Lundstedt T, Albano C. 1985. Screening of suitable solvents in organic synthesis—strategies for solvent selection. Acta Chem Scand B 39:79–91.
23. Chastrette M, Rajzmann M, Chanon M, Purcell KF. 1985. Approach to a general classification of solvents using a multivariate statistical treatment of quantitative solvent parameters. J Am Chem Soc 107:1–11.
24. Gramatica P, Navas N, Todeschini R. 1999. Classification of organic solvents and modelling of their physico-chemical properties by chemometric methods using different sets of molecular descriptors. TrAC, Trends Anal Chem 18:461–471.
25. Occupational Safety and Health Administration. 2006. Flammable and combustible liquids—1910.106. http://www.osha.gov/SLTC/smallbusiness/ sec8.html.
26. FDA. 2006. EAFUS: A Food Additive Database. http://www.cfsan.fda.gov/~dms/eafus.html.
27. (a) ICH. 1997. Q3C Guideline. Impurities: Residual Solvents. (b) ICH. 2003. Q3C Guideline. Tables and list.
28. Syracuse Research Corporation. 2006. Interactive PhysProp Database Demo. http://www.syrres.com/esc/physdemo.htm.
29. Kollman PA, Allen LC. 1972. The theory of the hydrogen bond. Chem Rev 72:283–303.
30. Suihko E, Forbes RT, Korhonen O, Ketolainen J, Paronen P, Gynther J, Poso A. 2005. Prediction of contact angle for pharmaceutical solids from their molecular structure. J Pharm Sci 94:745–758.
31. Hunter CA, Lawson KR, Perkins J, Urch CJ. 2001. Aromatic interactions. J Chem Soc, Perkin Trans 2651–2669.

32. Cruciani G, Pastor M, Guba W. 2000. VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. Eur J Pharm Sci 11:S29–S39.
33. Cruciani G, Crivori P, Carrupt PA, Testa B. 2000. Molecular fields in quantitative structure-permeation relationships: The VolSurf approach. J Mol Struct (Theochem) 503:17–30.
34. Wold S, Esbensen K, Geladi P. 1987. Principal component analysis. Chemom Intell Lab Syst 2: 37–52.
35. Kohonen T. 1997. Self-organizing maps. 2nd edition. Heidelberg: Springer.
36. Vandeginste BGM, Massart DL, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J. 1998. Handbook of chemometrics and qualimetrics—Part B. Amsterdam: Elsevier.
37. Morissette SL, Soukasene S, Levinson D, Cima MJ, Almarsson O. 2003. Elucidation of crystal form diversity of the HIV protease inhibitor ritonavir by high-throughput crystallization. Proc Natl Acad Sci USA 100:2180–2184.
38. Hilfiker R, Berghausen J, Blatter F, Burkhard A, De Paul SM, Freiermuth B, Geoffroy A, Hofmeier U, Marcolli C, Siebenhaar B, Szelagiewicz M, Vit A, von Raumer M. 2003. Polymorphism—integrated approach from high-throughput screening to crystallization optimization. J Therm Anal Calorim 73: 429–440.
39. White RF, Proctor SP. 1997. Solvents and neurotoxicity. Lancet 349:1239–1243.
40. Miller JM, Collman BM, Greene LR, Grant DJW, Blackburn AC. 2005. Identifying the stable polymorph early in the drug discovery-development process. Pharmaceut Dev Technol 10:291–297.
41. Vrecer F, Vrbinc M, Meden A. 2003. Characterization of piroxicam crystal modifications. Int J Pharm 256:3–15.
42. Jarring K, Larsson T, Stensland B, Ymen I. 2006. Thermodynamic stability and crystal structures for polymorphs and solvates of formoterol fumarate. J Pharm Sci 95:1144–1161.
43. Guillory JK. 1999. Generation of polymorphs. In: Brittain HG, editor. Polymorphism in pharmaceutical solids. New York: Marcel Dekker, Inc.: pp 183–226.
44. Lipinski C, Hopkins A. 2004. Navigating chemical space for biology and medicine. Nature 432:855–861.