Parallel screening of selective and high-affinity displacers for proteins in ion-exchange systems

Kaushal Rege [a], Asif Ladiwala [a], Nihal Tugcu [a],
Curt M. Breneman [b], Steven M. Cramer [a],*

[a] *Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA*
[b] *Department of Chemistry, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA*

Abstract

This paper employs a parallel batch screening technique for the identification of both selective and high-affinity displacers for a model binary mixture of proteins in a cation-exchange system. A variety of molecules were screened as possible displacers for the proteins ribonuclease A (RNAseA) and α-chymotrypsinogen A (α-chyA) on high performance Sepharose SP. The batch screening data for each protein was used to select leads for selective and high-affinity displacers and column experiments were carried out to evaluate the performance of the selected leads. The data from the batch displacements was also employed to generate quantitative structure–efficacy relationship (QSER) models based on a support vector machine regression approach. The resulting models had high correlation coefficients and were able to predict the behaviour of molecules not included in the training set. The descriptors selected in the QSER models for both proteins were examined to provide insights into factors influencing displacer selectivity in ion-exchange systems. The results presented in this paper demonstrate that this parallel batch screening–QSER approach can be employed for the identification of selective and high-affinity displacers for protein mixtures.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Screening; Displacement chromatography; Selectivity; Regression analysis; Mathematical modelling; Support vector machine; Structure–activity relationships; Proteins

1. Introduction

Ion-exchange displacement chromatography with low-molecular-mass displacers has attracted significant attention as a powerful technique for the purification of biomolecules from complex biological mixtures [1–8]. A major challenge for methods development in displacement chromatography of proteins is the identification of efficient displacers. The choice of an appropriate displacer molecule depends not only on the mixture of proteins in the feed but also on the stationary phase chemistry and mobile phase conditions. One of the advantages of low-molecular-mass displacers is their ability to carry out selective displacement chromatography [9] where the bioproduct of interest is displaced while the low-affinity impurities are eluted in the induced salt gradient ahead of the displacement train and the high-affinity

impurities are desorbed in the displacer zone. To date, the relative affinity of various displacers in ion-exchange systems has been evaluated using column experiments and the steric mass action formalism [10]. While this approach can readily predict column performance, it is time intensive. Accordingly, there is a need to develop more rapid methods for the identification of selective and high-affinity low-molecular-mass displacers.

High-throughput screening (HTS) is widely employed to identify novel lead compounds in drug discovery [11–16] and for the identification of ligands for affinity chromatography from combinatorial libraries [17]. Although the earliest reports for the use of combinatorial techniques for affinity ligands involved screening of epitope peptide libraries [18–20], small molecule affinity ligands have been identified for a variety of targets including kallikrein [21], IgG [22], recombinant insulin precursor [23], etc. from focused libraries. Recently, a parallel batch screening technique has been developed for rapid displacer discovery [24]. Using this technique, displacers are screened in parallel and are

* Corresponding author. Tel.: +1-518-276-6198;
fax: +1-518-276-4030.

E-mail address: crames@rpi.edu (S.M. Cramer).

ranked according to the "percent protein displaced" from the stationary phase material.

Property modelling techniques have been generalised into a broader field known as quantitative structure–property analysis (QSPR). Quantitative structure–activity relationship (QSAR) approaches have been designed to assist the development of bioactive compounds [25–31]. Quantitative structure–retention relationship (QSRR) models have also been reported for various modes of chromatography [32–34]. Recently, quantitative structure–efficacy relationship (QSER) models have been developed to predict displacer efficacy in ion-exchange systems [24]. While several successful methods are known, the current work is based on the use of the robust support vector machine (SVM) approach due to its utility in handling large sets of non-orthogonal descriptors and its inherent resistance to over-fitting [35].

In this paper, potential displacers are screened in parallel for ribonuclease A (RNAseA) and α-chymotrypsinogen A (α-chyA) adsorbed together on high-performance (HP) Sepharose SP to enable the identification of both selective and high-affinity displacers for this model protein mixture. A potential selective displacer is one that produces a sufficiently high value of percent protein displaced for one protein while showing lower values for the other protein in the mixture. In contrast, high-affinity displacers show high percent protein displaced values for both proteins in the mixture. In addition, the selected displacer leads are employed in column experiments to validate the utility of this approach.

Percent protein displaced values from the parallel batch screening experiments are also used to generate QSER models using SVM regression. Two-dimensional (2D), three-dimensional (3D) molecular operating environment (MOE, Chemical Computing Group, Montreal, Canada), molecular fragment (FRAG) and transferable atom equivalent (TAE) descriptors [36,37] are calculated from the energy-minimised structures of the displacers. A *training set* of displacers is selected to derive a predictive QSER model (i.e., learning from the database) using SVM regression and bootstrapping techniques [35]. It is shown that the resulting QSER models are able to successfully predict the efficacy of the test set displacers in the database. Finally, interpretation of the resulting models enables the importance of various structural and electronic features of the displacers to be elucidated.

2. Experimental

2.1. Materials

HP Sepharose SP stationary phase material was donated by Amersham Biosciences (Uppsala, Sweden). A Phenomenex Jupiter C_4 10 μm (250 mm × 4.6 mm) column was purchased from Phenomenex (Torrance, CA, USA). A strong cation-exchange (SCX, Waters SP-8HR,

100 mm × 5 mm i.d.) column was obtained from Waters (Milford, MA, USA). Ribonuclease A, α-chymotrypsinogen A, sodium phosphate (dibasic), sodium phosphate (monobasic) and sodium chloride were purchased from Sigma (St. Louis, MO, USA).

Potential displacer molecules: 2,2 dimethyl-1,3-propanediamine, 3,3′-diamino-N-methyldipropylamine, butylamine, diethylenetriamine, hydroxylamine, malonamamidine, malonamide, methylamine, N-methyl-1,3-propanediamine, N,N′-diethyl-1,3-propanediamine, N-2-(aminoethyl)-1,3 propanediamine, pentaethylene hexamine were purchased from Aldrich (Milwaukee WI, USA). Bekanamycin sulfate, N-α-benzoyl-L-arginine ethyl ester (BAEE), neomycin sulfate, paromomycin sulfate, spermidine were purchased from Sigma. 1,2-Diaminocyclohexane, cyclohexylamine, piperazine hydrochloride were purchased from TCI America (Portland, OR, USA). Pentaerythrityl-(trimethylammonium chloride)$_4$ (PETMA4), dipentaerythrityl-(trimethylammonium)$_6$ (DPETMA6) were synthesised in Professor Moore's laboratory at Rensselaer [38].

2.2. Equipment

Parallel batch screening experiments were carried out in 1.4 ml glass vials with rubber septum purchased from Fisher. Displacement experiments were carried using a Waters 590 HPLC pump (Waters, Milford, MA, USA) connected to a chromatography column via a Model C10W 10-port valve (Valco, Houston, TX, USA). The column effluent during displacement experiments was monitored using a model 484 UV-Vis absorbance detector (Waters). Fractions of the column effluent were collected using a LKB 2212 Helirac fraction collector (LKB, Bromma, Sweden). Analytical chromatographic experiments were carried out using a model 600 multisolvent delivery system, a 712 WISP auto injector and a 484 UV-Vis absorbance detector controlled by a Millennium chromatography software manager (Waters).

2.3. Procedures

2.3.1. Parallel batch screening

The bulk stationary phase, HP Sepharose SP was washed with deionised water and then the carrier buffer (50 mM phosphate, pH 6.0) and allowed to equilibrate for 2 h. After gravity settling of the stationary phase, the supernatant was removed and 3.0 ml of the remaining stationary phase slurry was equilibrated with 36 ml containing 3 mg/ml of the protein mixture (1.5 mg/ml each of ribonuclease A and α-chymotrypsinogen A) in 50 mM phosphate buffer, pH 6.0 at 20 °C. The proteins were equilibrated for 5 h in order to attain complete equilibrium. After equilibration was complete, the stationary phase was allowed to gravity-settle, the supernatants were removed, and the protein content was determined using linear gradient reversed-phase liquid chromatography (RPLC). The protein adsorbed on the stationary phase was then determined by mass balance.

Twenty-five microliters aliquots of the remaining stationary phase slurry with bound protein were then added to 300 μl containing 10 mM displacer in 50 mM phosphate buffer, pH 6.0 at 20 °C. A separate displacer was employed for each batch experiment to enable parallel screening. A total of 22 different displacers were examined for each protein on each stationary phase material and the experiments were carried out in triplicate. The system was equilibrated for 5 h. After equilibrium was achieved, the supernatant was removed and the protein content was determined via RPLC. The concentration of protein was determined and the "percent protein displaced" by each displacer was calculated by mass balance.

2.3.2. Column displacement chromatography

The column was initially equilibrated with the carrier solution and then sequentially perfused with the feed and displacer solutions. Separate displacement experiments were carried out using 6 mM diethylenetriamine and 8 mM paromomycin as the displacer on a HP Sepharose SP column. Two hundred microliters fractions were collected for subsequent analysis of the proteins and displacers. Following the infusion of displacer, column regeneration was carried out using a 2 M NaCl solution. All displacement experiments were carried out at 0.2 ml/min and the effluent was monitored at 280 nm. Experimental conditions including the feed load, salt concentration, and displacer concentration employed for each separation are given in the figure legends of the corresponding displacement chromatograms.

2.3.3. Linear gradient HPLC analysis

2.3.3.1. Parallel batch screening experiments. Linear gradient RPLC with a Phenomenex Jupiter C_4 10 μm (250 mm × 4.6 mm) column was used to evaluate the amount of α-chyA and RNAseA in the supernatant of the parallel batch screening experiments. A linear gradient of 30–100% (v/v) buffer B was carried out in 5 min [buffer A: 0.1% (v/v) trifluoroacetic acid (TFA) in deionised water; buffer B: 90% (v/v) acetonitrile and 0.1% (v/v) TFA in deionised water]. The flow rate was 1 ml/min and the column effluent was monitored at 280 nm. The injection volumes were varied between 25 and 200 μl; as little as 2 μg of protein could be detected from a single injection.

2.3.3.2. Displacement experiments. Protein and displacer analysis for the displacement of α-chyA and RNAseA using diethylene triamine and paromomycin was carried out using reversed-phase chromatography (Phenomenex C_4 column) under linear gradient conditions. The mobile phase employed for the analyses were deionised water (A) and 90% (v/v) acetonitrile in deionised water adjusted to pH 2.2 with TFA (B). The fractions were diluted 1–3-fold and 10 μl samples were injected. A 20 min gradient from 30 to 100%

(v/v) B at a flow rate of 1 ml/min was employed for the analysis of α-chyA and RNAseA. The column effluent was monitored at 280 nm. Diethylene triamine and paromomycin were analysed by complexation with fluorescamine [39,40]. The fractions were diluted 5–100-fold and 0.28 mg/ml solution of fluorescamine in acetone was added to the fractions with displacer in a 1:3 (v/v) ratio. Excitation at 390 nm and emission at 475 nm were then employed to quantify the amount of displacer in the fractions.

2.3.4. QSER modelling and SVM regression models

In order to construct informative QSER models, electron-density-based TAE quantum mechanics descriptors, molecular operating environment (MOE) descriptors and molecular fragment-type descriptors (FRAG) were employed. Using this hybrid set of descriptors, a SVM sparse regression algorithm was applied in a feature selection mode to determine a subset of relevant molecular property descriptors for each of the training sets involved in the bootstrapping procedure. Subsequently, non-linear SVM models were built based on those relevant descriptors. The overall modelling scheme is shown in Fig. 1.

2.3.5. Implementation

The structures of the displacer molecules were obtained from the supplier's website (http://www.sial.com/). These molecules were drawn in SYBYL 6.5 (Tripos, St. Louis, MO, USA) and energy minimised using the MMFF94 force field. MOE software was used to calculate the MOE molecular descriptors. The laboratory-developed RECON 5.2 package was employed for generating the transferable atom equivalent (TAE) descriptors for the displacers. An SVM program, developed independently in the Department of Mathematics at Rensselaer Polytechnic Institute, was used in the analysis [41].

2.3.6. TAE descriptor generation

The transferable atom equivalent/RECONstruction (TAE/RECON) [36,37] method consists of a rapid charge density reconstruction algorithm that utilises atomic charge density fragments that have been pre-computed from ab initio wave functions. In principle, a library of atomic charge density components (TAEs) can be used to construct molecular electron densities in a form that allows for rapid retrieval of the molecular surface properties needed to generate descriptors. For each molecule, the RECON program reads its molecular structure information, and then reconstructs the electronic properties of the molecular surface from the atomic fragments. The distributions of several electronic properties on molecular surfaces may then be quantified to give a large variety of numerical QSER descriptors. The CPU and disk resources required for TAE reconstruction are minimal—the electronic property distributions of 22 displacers may be computed in about 15 s on a single-headed 1.7 GHz Linux workstation.
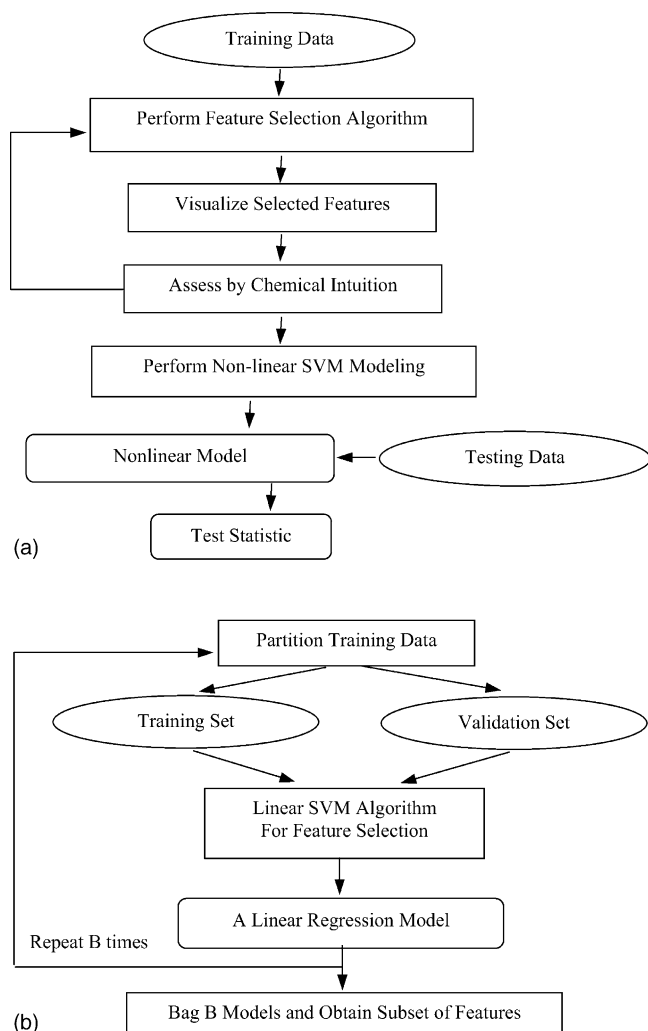
Fig. 1. (a) Computational chemical property design and model validation and (b) feature selection flowchart.

2.3.7. MOE descriptors

The MOE program provides a combination of several types of traditional molecular property descriptors, including connectivity-based topological 2D descriptors, physico-chemical property descriptors, shape-independent 3D molecular features and some pharmacophoric descriptors. These descriptors were calculated for the displacers via the QuaSAR-descriptors module in the MOE package. Prior to MOE descriptor calculation, the displacer structures must be appropriately charged as at the pH of the experiments. For this, the ACD/pK_a DB package (Advanced Chemistry Development, Toronto, Canada) was employed to compute the pK_a's of the charge centers on the displacer molecule. These pK_a values were then used to assign the charges on the displacer molecules at the pH of the experiments.

2.3.8. Support vector machine modelling

The support vector machine (SVM) method proposed by Vapnik [35] is based on statistical learning theory. The SVM

method has a number of interesting properties, including an effective avoidance of overfitting, which improves its ability to build models using large numbers of molecular property descriptors with relatively few experimental results in the training set. Although SVM was originally developed for pattern recognition, it was later extended to solve the regression problem [35]. In the present work, we focus on support vector regression for creating QSER models of displacer efficacy.

To summarise the operation of an SVM modelling procedure, it is important to consider some fundamental principles of SVMs: With a given set of training data, the objective is to find a function $f(x)$, called the ε insensitive loss function, that has less than ε deviation from the experimental protein retention data for all cases in the training set. In other words, those predicted % protein displaced values within ε distance of the actual response are not penalised for being erroneous. Only those prediction points beyond ε of the real response values are considered to contain modelling errors and are included in the "loss function". This technique helps to control the complexity of the model and tends to minimise the risk of overfitting. In QSER studies, the magnitude of ε will be roughly equivalent to the experimental error in the measurement of % protein displaced values.

In typical QSPR studies, many more descriptors are initially available than the number of molecules in the dataset, and usually include some redundant or irrelevant variables. In order to identify the relevant descriptors for a particular problem, variable selection techniques are always employed to choose those informative descriptors and eliminate irrelevant descriptors from consideration. The application of this type of feature selection serves to improve the computational signal-to-noise ratio in the resulting models. In this study, we applied a feature selection approach based on the sparse SVM regression [41]. Within this technique, a series of linear SVM models (usually 20–40) that exhibit good generalisation are constructed. In each linear l_1-norm SVM or bootstrap, the optimal weight vector will have relatively few non-zero weights with the degree of sparsity depending on the SVM model parameters. Those features with non-zero weights then become potential attributes to be used in the non-linear SVM. The method exploits the fact that linear SVM with l_1-norm regularisation inherently performs feature selection as a side effect of minimising capacity in the SVM model. Finally, the union of all obtained subsets of features produces the final descriptor set that can be used to construct the non-linear SVM predictive model. In order to get more robust and general predictive results, multiple QSER models based on the same feature set are built. Thus, instead of using a single model, which is heavily and easily affected by chance correlations, the average of all model predictions are used as our final prediction results. This kind of de-biasing technique is referred to as "bagging" in the statistical analysis field [42].

3. Results and discussion

3.1. Parallel batch screening

Parallel batch displacement experiments were carried out with the two proteins ribonuclease A (RNAseA) and α-chymotrypsinogen A (α-chyA) as described in Section 2. These two proteins exhibit similar retention times on HP Sepharose SP under linear gradient conditions at pH 6.0 [43]. Table 1 shows the % protein displaced data for both proteins for the 22 displacers screened. The data is arranged in increasing order of % α-chyA displaced. As seen in the table, displacers with low charge such as methylamine, malonamamidine, etc. showed negligible values of % protein displaced for both proteins. As the molecular weight and basicity of the molecules increased, the % protein displaced values of both proteins also increased. Interestingly, towards the middle of the table, the values of % protein displaced for RNAseA are higher than those for α-chyA. For example, N,N′-diethyl-1,3-propanediamine propanediamine (% α-chyA displaced: 28, % RNAseA displaced: 59.57), diethylene triamine (% α-chyA displaced: 28.9, % RNAse A displaced: 61.8) and 2,2-dimethyl-1,3-propanediamine (% α-chyA displaced: 34.78, % RNAseA displaced: 59.30) all showed higher values of % RNAseA displaced as compared to % α-chyA displaced. These results indicate that these displacers are possible lead compounds for selective displacement of RNAseA from a mixture of the two

proteins. Of the three possible leads, diethylene triamine was investigated for its utility to act as a selective displacer of RNAseA under column conditions as described below.

Displacers located towards the end of Table 1 were able to efficiently displace both proteins under these batch conditions indicating that they are good candidates for "high-affinity" displacers. For example, the aminoglycosides, paromomycin (% α-chyA displaced: 88.51, % RNAseA displaced: 66.32) and neomycin (% α-chyA displaced: 87.25, % RNAseA displaced: 71.75) showed high % protein displaced values for both proteins and therefore stand out as possible high-affinity displacers that may be used for the displacement of both proteins. The fact that these high-affinity compounds displace more of the α-chyA than RNAseA is due primarily to the multicomponent adsorption behaviour of this binary protein mixture (data not shown). Paromomycin was used as a possible high-affinity displacer in the subsequent column experiment.

Fig. 2 shows the structures of the potential selective and high-affinity displacer lead molecules. Column displacements were carried out using the equilibrium displacer concentration obtained from the batch experiments. Experiments were first carried out with the selective displacer lead compound diethylene triamine. As seen in Fig. 3, this column experiment resulted in the selective displacement of RNAseA with α-chyA remaining behind the displacer zone. The high-affinity displacer lead, paromomycin, was examined for its ability to displace both proteins. As seen in Fig. 4, paromomycin was indeed able to successfully displace both RNAseA and α-chyA under the column displacement conditions. These results indicate that % protein displaced data from parallel batch screening experiments can be used as a qualitative guide for the selection of column conditions for both selective and high-affinity displacements.

Table 1
Parallel batch screening data for % RNAseA and % α-chyA displaced on HP Sepharose SP

Displacer	% α-chyA displaced	% RNAseA displaced
BAEE	1.05	3.78
Malonamamidine	1.76	3.16
Malonamide	2.05	5.59
Methylamine	3.79	8.66
Butylamine	3.23	6.24
Hydroxylamine	4.73	11.38
Piperazine	7.82	30.39
Cyclohexylamine	18.37	4.25
N,N′-Diethyl-1,3-propanediamine[a]	28.00	59.57
Diethylene triamine[a]	28.93	61.80
2,2-Dimethylpropanediamine[a]	34.78	59.30
1,2-Diaminocyclohexane	41.99	53.40
N-Methyl-1,3-propanediamine	54.07	69.16
N-(2-Aminoethylamine)-1,3-propanediamine	63.91	61.73
Spermidine	75.31	77.77
PETMA4	83.45	83.16
Pentaethylenehexamine	81.14	59.49
3,3′-Diamino-N-methyldipropylamine	87.33	59.82
DPETMA6	82.96	67.81
Bekanamycin	82.04	57.13
Neomycin[b]	87.25	71.75
Paromomycin[b]	88.51	66.32

[a] Potential selective displacers for RNAseA.
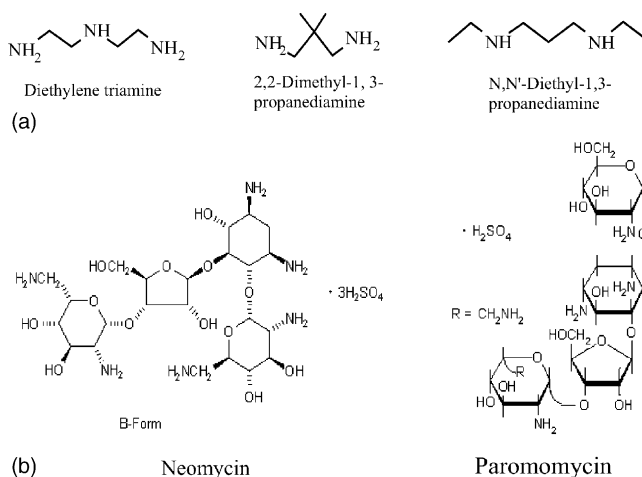[b] High-affinity displacer leads for RNAseA and α-chyA.

Fig. 2. Molecular structures of lead displacer candidates selected from parallel batch screening data: (a) selective displacers and (b) high-affinity displacers.
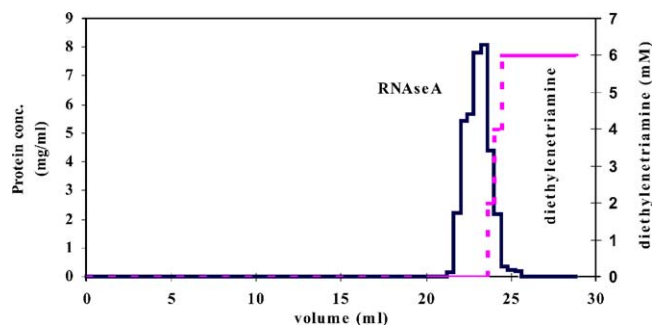
Fig. 3. Selective displacement of RNAseA on HP Sepharose SP using 6 mM diethylenetriamine. Column: 100 mm × 5 mm i.d. HP Sepharose SP; carrier: 50 mM phosphate, pH 6.0; protein: 35 mg of α-chyA and RNAseA; flow rate: 0.2 ml/min.

3.1.1. QSER models

The aim of this study was to generate models that can predict the displacer efficacy for α-chyA and RNAseA as well as aid in identifying the structural components that contribute to the efficacy and selectivity of the displacers in cation-exchange systems. To this end, experimental batch screening data (two responses) were combined with 189 MOE descriptors, 208 FRAG descriptors and 147 TAE descriptors. Descriptors having the same values for all 22 molecules were eliminated from the dataset since these provide no real information in the model. Furthermore, descriptors showing a variance of greater than four times a standard deviation (i.e., 4σ) were also removed from the dataset so as to reduce outliers and to enable interpretation of the models. The final dataset consisted of the two responses and 273 descriptors and was used to generate two independent QSER models. This dataset was subjected to SVM feature selection, as described above, to give two independent feature sets, one corresponding to each response. Finally, the dataset was divided into training and test sets for the purpose of model building. It was decided to keep 10% of the molecules (i.e., 2 molecules) in the test set and the rest (i.e., 20 molecules) in the training set. Methylamine (Mol 12) and paromomycin sulfate (Mol 18) were arbitrarily chosen to be in the ex-
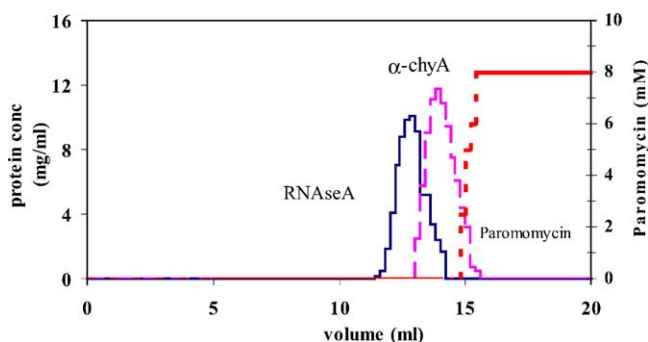
ternal test set. Once established, the QSER models were tested for their predictive ability for the external test set of displacers.

Fig. 5a and b shows the correlation between the experimental and predicted results for % α-chyA and % RNAseA displaced on HP Sepharose SP. The open circles represent the "bagged" predictions for the training set molecules and the dark squares represent "bagged" predictions for the test set molecules. The error bars in these figures represent the standard deviation in the predicted % protein displaced values. The cross-validated r^2 for these models were 0.9506 and 0.9785, respectively, which indicated that the predicted values for the training set were in good agreement with the experimental data. Importantly, the % protein displaced results for the molecules not included in the training of the model (i.e., the external test set) demonstrate the predictive power of these models (Fig. 6a and b). Even though the training set used to generate these models contained molecules with a high level of structural diversity (e.g., linear, cyclic,
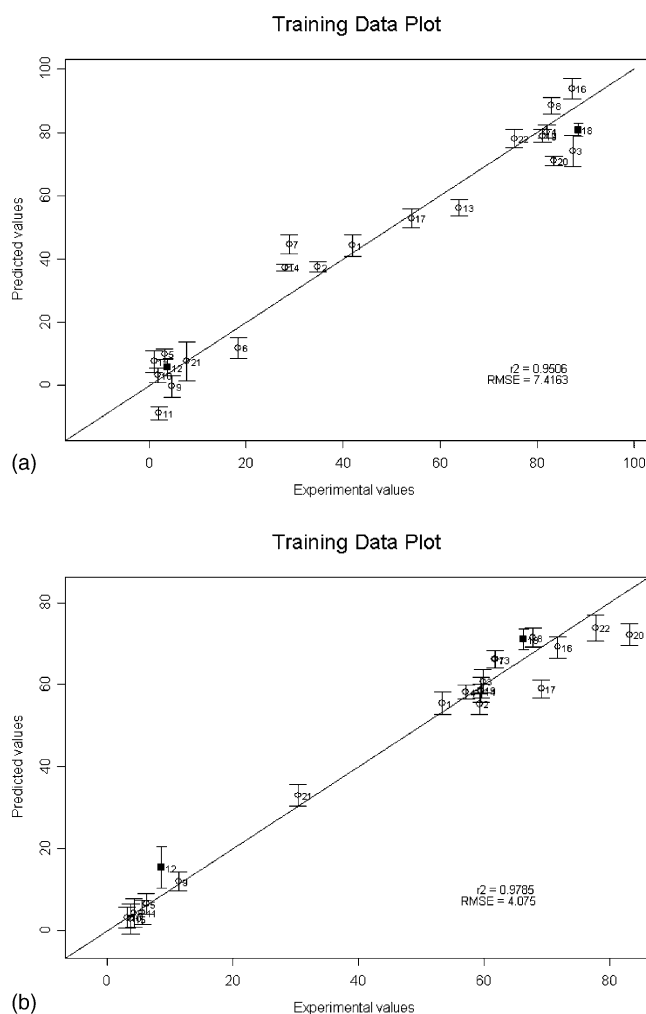

(a)

(b)

Fig. 5. QSER models based on a support vector machine (SVM) regression approach for % protein displaced on HP Sepharose SP for (a) α-chyA and (b) RNAseA.

Fig. 4. Displacement of α-chyA and RNAseA on HP Sepharose SP using 8 mM paromomycin. Column: 100 mm × 5 mm i.d. HP Sepharose SP; carrier: 50 mM phosphate, pH 6.0; protein: 35 mg of α-chyA and RNAseA; flow rate: 0.2 ml/min.
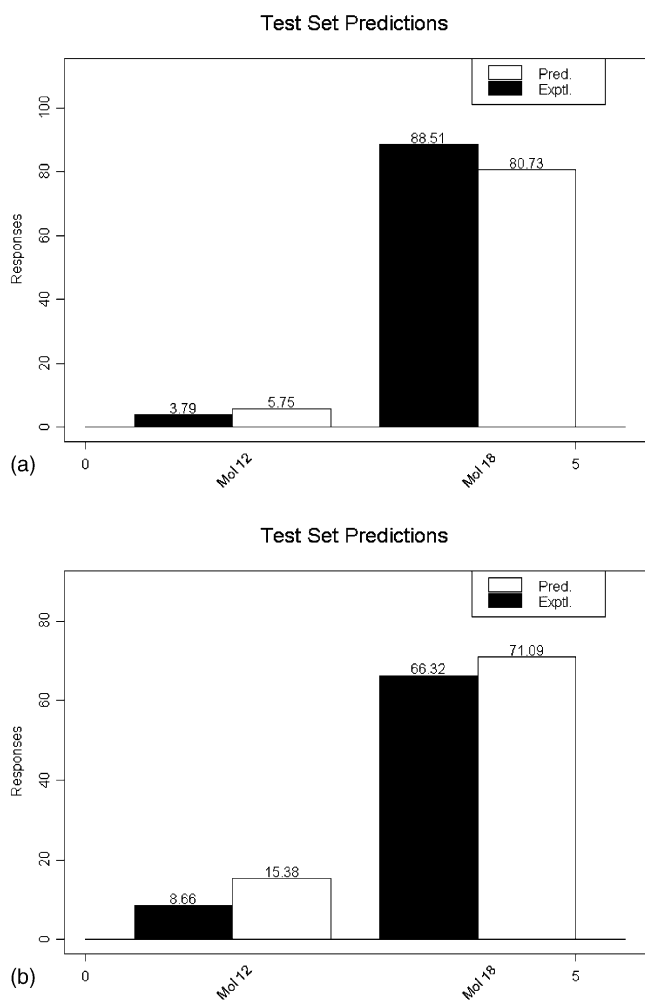
Fig. 6. Predictions of the QSER models for % protein displaced on HP Sepharose SP for the external test set of molecules comprising of methylamine (Mol 12) and paromomycin (Mol 18) for: (a) α-chyA and (b) RNAseA.
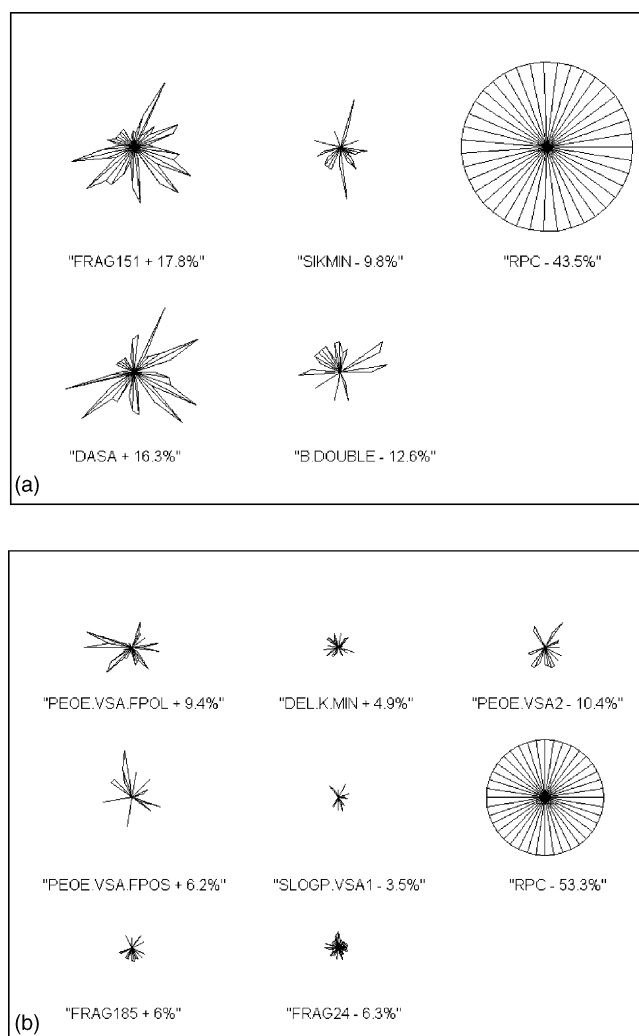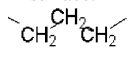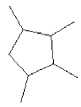

Fig. 7. Star plots of important model descriptors selected in the QSER models based on the SVM regression approach for % protein displaced on HP Sepharose SP for (a) α-chyA and (b) RNAseA.

dendritic, and aromatic), the models were able to produce good predictions.

Models were then examined to determine the physico-chemical phenomena influencing displacer efficacy and selectivity in cation-exchange systems. In order to facilitate interpretation, it was necessary to determine which descriptors were consistently important when different combinations of training and validation molecules were used. As indicated in Fig. 1, each different set of training molecules is called a "fold" and the model created using this set is used to make predictions on the validation molecules left out of the training set for that particular "fold". Accordingly, star plots were generated to evaluate the relative importance of each of these descriptors selected throughout each of 40 bootstrap "folds" used for creating the composite model set. A star plot presentation of these results consists of a set of radial graphics that represent a multivariate data matrix—in this case, the weight matrix. In these plots, each star corresponds to a specific descriptor in the weight matrix generated by

all 40 linear SVM regression bootstrap folds and the length of each ray represents the weight of this descriptor in one of the bootstrap iterations. For each star plot, the selected descriptors are ranked according to the sum of their radii for all bootstraps, placing the most significant positively weighted descriptor in the upper left hand corner and the most significant negative contributor on the lower right. The order proceeds from left to right in a columnar fashion. The relative contribution of a descriptor to the aggregate model is quantified by the sum of the weights of that descriptor in all bootstraps, relative to the sum of the weights of all descriptors in the model. The star plots for the α-chyA and RNAseA displacement models are shown in Fig. 7a and b.

The relevant QSER descriptors were found to include shape, size, surface property, molecular fragment, and electron-density derived descriptors. The definitions of the important descriptors are given in Table 2. As seen in the table, many descriptors have direct physical/chemical significance and are thus easy to interpret. For descriptors that

Table 2
List and definition of important descriptors used for displacer efficacy modelling

Descriptor	Type	Definition
B.DOUBLE	MOE	Number of double bonds. Aromatic bonds are not considered to be double bonds.
DASA	MOE	Absolute value of the difference between the water accessible surface area of all atoms with positive and negative partial charges.
DEL.K.MIN	TAE	Minimum rate of change of the K electronic kinetic energy density normal to and away from the molecular surface.
FRAG151	FRAG	
FRAG185	FRAG	
FRAG24	FRAG	
PEOE.VSA.FPOL	MOE	Fractional polar van der Waals surface area.
PEOE.VSA.FPOS	MOE	Fractional positive van der Waals surface area.
PEOE.VSA2	MOE	Sum of VDW surface area where partial charge of atoms is in the range (0.10, 0.15).
RPC	MOE	Relative positive partial charge: the largest positive partial charge on an atom in the molecule divided by the sum of the positive partial charge on the molecule.
SIKMIN	TAE	The minimum of the surface integral of the K electronic kinetic energy density (SIK) distribution.
SLOGP.VSA1	MOE	Sum of VDW surface area such that the contribution to $\log P$(o/w) of atom(s) is in the range (-0.4, -0.2).

are more difficult to interpret (e.g., PEOE.VSA2), molecular surface visualisation was employed to gain insight into the physico-chemical information provided by these descriptors in the model.

As seen from the star plots, descriptors associated with positive charges on the displacer molecules were found to exist in both models. DASA represents the difference between the positively and the negatively charged surface area of the displacer molecule. Since all our displacers had a net positive charge at the pH of the batch screening experiments (i.e., pH 6.0), this descriptor is associated with the relative abundance of positive charges on the displacer molecules. This descriptor had a positive weight in the α-chyA model, which indicates that net positive charge is desirable, as would be expected in cation-exchange systems. The presence of a combination of PEOE.VSA.FPOS and PEOE.VSA.FPOL in the QSER model for % RNAseA displaced indicates that a positive and polar surface area on the molecules is required for effecting high values of % RNAseA displaced. The positive contributions of these two descriptors in the model are in agreement with the nature of cation-exchange chromatography. PEOE.VSA2 is a histogram bin-type descriptor associated with molecular surface area with a positive PEOE partial charge value in the range (0.10–0.15). From molecular surface visualisation we observed that this range of partial charge values was typically found on hydrogen atoms associated with uncharged amines (Fig. 8). The negative value associated with this PEOE.VSA2 descriptor indicates that amine groups on the candidate molecules that were still uncharged at pH 6 had a negative influence on displacer efficacy.

As seen in Fig. 7, RPC is the most dominant descriptor in both QSER models. RPC is associated with the relative positive partial charge (based on the MMFF94 force field) on the molecule and is mathematically defined as the ratio of

the largest positive partial charge to the sum of the positive partial charge on the molecule. It is thus a measure of the distribution of the positive partial charge, with a higher value corresponding to concentrated partial charge and a lower value indicating that the partial charge is more evenly distributed over the molecular surface. This descriptor is seen to have a high negative contribution in both models, suggesting that lower values of RPC are favourable for higher displacer efficacy. This indicates that, in addition to the amount of charge, the distribution of charge on the molecule also plays an important role in determining displacer efficacy.

It can be seen from Fig. 7a that FRAG151 is important for the displacement of α-chyA, but was not observed to be important for RNAseA. FRAG151 describes a three-methylene spacing in a candidate molecule. This result is consistent with previous observations in our laboratory that displacers such as spermidine and spermine with three and four nitrogen atoms, respectively, containing a three-methylene spacing between nitrogen atoms are more efficacious than diethylene triamine and triethylene tetramine which only
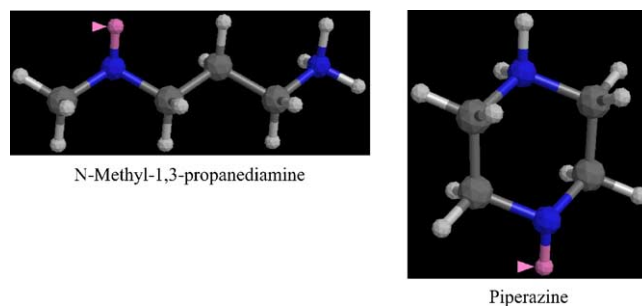

Fig. 8. Molecular surface visualisation enables us to identify atoms/groups associated with the PEOE.VSA2 descriptor having PEOE partial charges in the range (0.10, 0.15). Atoms that contribute to PEOE.VSA2 are indicated with arrows.

have two-methylene spacing between the nitrogen atoms [10,24]. It is interesting to note that while molecules with two-methylene spacing such as diethylene triamine were able to displace RNAseA, they were not able to effectively displace α-chyA (Table 1). In contrast, molecules with three-methylene spacing such as spermidine were indeed able to effectively displace α-chyA, in qualitative agreement with the star plots results. These results with spermidine and diethylene triamine also make sense when considering the results described for the PEOE.VSA2 descriptor, which indicated that uncharged amine groups on candidate molecules (e.g., diethylenetriamine) had a negative influence on displacer efficacy.

4. Conclusions

In this paper, we demonstrate that the parallel batch displacement protocol can be extended to a mixture of proteins adsorbed on the stationary phase to enable the identification of high-affinity and selective displacers. Twenty-two different molecules ranging from simple linear polyamines to complex aminoglycosides were screened for two proteins, α-chymotrypsinogen A and ribonuclease A, on HP Sepharose SP. The two proteins were adsorbed on the resin and the respective % protein displaced values were used to rank the displacers. Three displacers structurally related to each other (diethylene triamine, 2,2-dimethyl-1,3-propanediamine, N,N'-diethyl-1,3-propanediamine) with two charges at the employed pH, were found to have higher selectivity for the displacement of RNAseA as compared to α-chyA. In contrast, high % protein displaced values were observed for both proteins when aminoglycosides such as neomycin and paromomycin were employed. The information from the batch screens was validated using column displacement experiments in which the selective and high-affinity displacer "leads" resulted in selective displacement of RNAseA and displacement of both proteins, respectively. These experimental results indicate that the parallel batch screening technique enables the rapid identification of displacer selectivity and affinity, and that the results can be employed directly in a column setting.

The results from the parallel batch screening experiments were then used to generate predictive QSER models using SVM regression. The models showed high regression coefficients ($r^2 > 0.95$) and the predicted % protein displaced values were in good agreement with the experimental data. This is particularly significant in view of the fact that the two molecules used as the external test set were quite diverse (i.e., methylamine and paromomycin). The star plots from the QSER models revealed the influence of molecular characteristics, such as charge and charge distribution, on displacer efficacy. In addition, more specific characteristics were also revealed in the models such as the importance of a three-methylene spacing between amine groups for enhancing displacer affinity. The results indicated that a three-methylene spacing was advantageous for increasing the displacement of α-chyA. This was corroborated by the experimental data, which showed that while molecules with three-methylene spacing such as spermidine were able to effectively displace α-chyA, displacers with two-methylene spacing such as diethylene triamine were only able to effectively displace RNAseA.

The results presented in this paper demonstrate that parallel batch screening can be used in concert with advanced QSER modelling for the a priori design of both selective and high-affinity displacers. Future work will examine the efficacy of this multi-protein screening and modelling technique for more complex mixtures of industrial relevance.

Acknowledgements

References

[1] J.A. Gerstner, J. Morris, T. Hunt, R. Hamilton, N.B. Afeyan, J. Chromatogr. A 695 (1995) 195.
[2] G. Jayaraman, Y.F. Li, J.A. Moore, S.M. Cramer, J. Chromatogr. A 702 (1995) 143.
[3] A. Kundu, S. Vunnum, G. Jayaraman, S.M. Cramer, Biotechnol. Bioeng. 48 (1995) 452.
[4] A. Kundu, S. Vunnum, S.M. Cramer, J. Chromatogr. A 707 (1995) 57.
[5] E. Luellau, U. Stockar, S. Vogt, R. Freitag, J. Chromatogr. A 796 (1998) 165.
[6] A. Shukla, S.M. Cramer, in: S. Ahuja (Ed.), Handbook of Bioseparations, vol.1, Academic Press, 2000.
[7] N. Tugcu, R.R. Deshmukh, Y.S. Sanghvi, J.A. Moore, S.M. Cramer, J. Chromatogr. A 923 (2001) 65.
[8] N. Tugcu, R.R. Deshmukh, Y.S. Sanghvi, S.M. Cramer, React. Functl. Polym. 1–3 (2003) 37.
[9] A. Kundu, K.A. Barnthouse, S.M. Cramer, Biotechnol. Bioeng. 56 (1997) 119.
[10] A.A. Shukla, S.S. Bae, J.A. Moore, S.M. Cramer, J. Chromatogr. A 827 (1998) 295.
[11] M.A. Gallop, R.W. Barret, W.J. Dower, S.P.A. Fodor, M. Gordon, J. Med. Chem. 37 (1994) 1233.
[12] M. Gordon, R.W. Barret, W.J. Dower, S.P.A. Fodor, M.A. Gallop, J. Med. Chem. 37 (1994) 1385.
[13] W.P. Janzen, Lab. Robotics Autom. 8 (1996) 261.
[14] B.A. Kenny, M. Bushfield, D.J. Parry-Smith, S. Fogarty, J.M. Treherne, Prog. Drug Res. 51 (1998) 246.
[15] C.P. Woodbury Jr., D.L. Venton, J. Chromatogr. B 712 (1999) 113.
[16] L.V. Hijfte, G. Marciniak, N. Froloff, J. Chromatogr. B 725 (1999) 3.
[17] C.R. Lowe, Curr. Opin. Chem. Biol. 5 (2001) 248.
[18] J.K. Scott, G.P. Smith, Science 249 (1990) 249.
[19] R.W. Barrett, S.E. Cwirla, M.S. Ackerman, A.M. Olson, E.A. Peters, W.J. Dower, Anal. Biochem. 204 (1992) 357.
[20] K.R. Oldenburg, D. Loganathan, I.J. Goldstein, P. G Schultz, M.A. Gallop, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 5393.

[21] N.P. Burton, C.R. Lowe, J. Mol. Recog. 5 (1992) 55.

[22] S.-F. Teng, K. Sproule, A. Hussain, C.R. Lowe, J. Mol. Recog. 12 (1999) 67.

[23] K. Sproule, P. Morrill, J.C. Pearson, S.J. Burton, K.R. Hejnaes, H. Valore, S. Ludvigsen, C.R. Lowe, J. Chromatogr. B 740 (2000) 17.

[24] C.B. Mazza, K. Rege, N. Sukumar, C.M. Breneman, J.S. Dordick, S.M. Cramer, Biotechnol. Bioeng. 80 (2002) 60.

[25] A.J. Hopfinger, J. Am. Chem. Soc. 102 (1980) 7196.

[26] J.K. Seydel, D. Trettin, H.P. Cordes, J. Med. Chem. 23 (1980) 607.

[27] M.I. Skvortsova, I.I. Baskin, O.L. Slovokhotova, V.A. Palyulin, N.S. Zefirov, J. Chem. Info. Comp. Sci. 33 (1993) 630.

[28] C. Hansch, A. Leo, Exploring QSAR, American Chemical Society, Washington, DC, 1995.

[29] D.C. Horwell, W. Howson, M. Higginbottom, D. Naylor, G.S. Ratcliffe, S. Williams, J. Med. Chem. 38 (1995) 4454.

[30] Z. Mazerska, E. Augustin, J. Dziegielewski, M.W. Cholody, J. Konopa, Anti-Cancer Drug Design 11 (1996) 73.

[31] W. Karcher, S. Karabunarliev, J. Chem. Info. Comp. Sci. 36 (1996) 672.

[32] R.J. Kaliszan, J. Chromatogr. A 656 (1993) 417.

[33] P.W. Carr, Microchem. J. 48 (1993) 4.

[34] C.B. Mazza, N. Sukumar, C.M. Breneman, S.M. Cramer, Anal. Chem. 73 (2001) 5457.

[35] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.

[36] C.M. Breneman, T.R. Thompson, M. Rhem, M. Dung, Comput. Chem. 19 (1995) 161.

[37] C.M. Breneman, N. Sukumar, RECON2000 Program, Rensselaer Polytechnic Institute, Troy, NY, 2000.

[38] A.A. Shukla, S.S. Bae, J.A. Moore, K.A. Barnthouse, S.M. Cramer, Ind. Eng. Chem. Res. 370 (1998) 4090.

[39] S. Udenfied, S. Steinb, P. Bohlen, W. Dairman, Science 178 (1972) 871.

[40] S. De Bernardo, M. Wiegele, V. Toomw, K. Manhart, W. Leimgruber, Arch. Biochem. Biophys. 163 (1974) 390.

[41] J. Bi, K. Bennett, M. Embrechts, C.M. Breneman, M. Song, J. Mach. Learn. Res. 3 (2003) 1229.

[42] L. Breiman, Machine Learning 24 (1996) 123.

[43] V. Natarajan, S.M. Cramer, Sep. Sci. Technol. 35 (2000) 1719.