



# The hybrid experimental simplex algorithm – An alternative method for ‘sweet spot’ identification in early bioprocess development: Case studies in ion exchange chromatography

Spyridon Konstantinidis<sup>a</sup>, Sunil Chhatre<sup>a</sup>, Ajoy Velayudhan<sup>a</sup>, Eva Heldin<sup>b</sup>, Nigel Titchener-Hooker<sup>a,\*</sup>

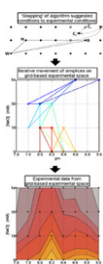
<sup>a</sup> The Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London, Torrington Place, London, WC1E 7JE, United Kingdom

<sup>b</sup> GE Healthcare Life Sciences, BioTechnologies R&D, Björkgatan 30, Uppsala 751 84, Sweden

## HIGHLIGHTS

- ▶ A novel simplex method best suited to coarsely gridded data was developed.
- ▶ This is used for ‘sweet spot’ identification in two bioprocess development studies.
- ▶ The method returns comparable experimental costs to those from DoE methodologies.
- ▶ The method is a viable alternative for ‘sweet spot’ identification in scouting studies.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 27 January 2012

Received in revised form 11 May 2012

Accepted 26 June 2012

Available online 2 July 2012

### Keywords:

Analytical bottleneck

FAB<sup>+</sup>

Green fluorescent protein

High throughput process development

Simplex algorithm

Ion exchange chromatography

## ABSTRACT

The capacity to locate efficiently a subset of experimental conditions necessary for the identification of an operating envelope is a key objective in many studies. We have shown previously how this can be performed by using the simplex algorithm and this paper now extends the approach by augmenting the established simplex method to form a novel hybrid experimental simplex algorithm (HESA) for identifying ‘sweet spots’ during scouting development studies. The paper describes the new algorithm and illustrates its use in two bioprocessing case studies conducted in a 96-well filter plate format. The first investigates the effect of pH and salt concentration on the binding of green fluorescent protein, isolated from *Escherichia coli* homogenate, to a weak anion exchange resin and the second examines the impact of salt concentration, pH and initial feed concentration upon the binding capacities of a FAB<sup>+</sup>, isolated from *E. coli* lysate, to a strong cation exchange resin. Compared with the established algorithm, HESA was better at delivering valuable information regarding the size, shape and location of operating ‘sweet spots’ that could then be further investigated and optimized with follow up studies. To test how favorably these features of HESA compared with conventional DoE (design of experiments) methods, HESA results were also compared with approaches including response surface modeling experimental designs. The results show that HESA can return ‘sweet spots’ that are equivalently or better defined than those obtained from DoE approaches. At the same time the deployment of HESA to identify bioprocess-relevant operating boundaries was accompanied by comparable experimental costs to those of DoE methods. HESA is therefore a viable and valuable alternative route for identifying ‘sweet spots’ during scouting studies in bioprocess development.

© 2012 Elsevier B.V. Open access under [CC BY](http://creativecommons.org/licenses/by/3.0/) license.

## 1. Introduction

Experimental design approaches are well established in many scientific and technological fields as an effective methodology

\* Corresponding author. Tel.: +44 20 7679 3796; fax: +44 7916 3943.

E-mail address: [nigelth@ucl.ac.uk](mailto:nigelth@ucl.ac.uk) (N. Titchener-Hooker).

to determine appropriate operating conditions for complex processes. In the bioprocess industry, the use of such DoE (design of experiments) methods are widely – and appropriately – advocated throughout the development of the various unit operations that are used commercially to produce complex products, while implementing Quality by Design [1]. The DoE approach usually consists of three stages: (1) screening experiments, to determine which of a large set of possible experimental variables are important; (2) scouting experiments on only the important variables, to identify promising regions or ‘sweet spots’ of operating space; and (3) subsequent optimization experiments, within these ‘sweet spots’, to obtain accurate and valid mathematical models which identify and rigorously characterize a design space. Stage 3 is often implemented via the adoption of response surface modeling (RSM) designs.

The derivation of models from RSM designs, such as central composite designs (CCDs), also makes them applicable to stage 2 scouting studies since they can identify ‘sweet spots’ based on their predictions [2]. During these studies however, when the investigated experimental spaces are still large, poorly fitting regression models may be obtained, the predictions of which could yield ill-defined ‘sweet spots’ [3]. This can have a direct impact on the number of follow up optimization studies and their efficiency. Such complexities can be avoided by employing high level factorial designs during scouting studies. These designs are coarse grids of test conditions, which are all evaluated to derive ‘sweet spots’ [4–7]. Hence, when regression models cannot describe experimental data accurately and reliably, favorable operating conditions can be determined from the experimental measurements when they are presented in highly informative contour or response surface plots. Such designs also maximize process understanding since by employing the theoretical knowledge on the investigated system, experimental factors with the greatest interest can be allocated numerous levels and thus assessed in detail. While this approach ensures high levels of information about process behavior, it also leads to increased experimental costs (e.g. increased time and material requirements), especially in those cases where the defined grids contain numerous conditions and necessitating their evaluation by lengthy analytical methods, causing an analytical bottleneck [8]. Early bioprocess development is also characterized by limited amounts of material, especially highly purified material; and often tight timelines and limited personnel. It is therefore worth evaluating which general approaches are best suited to these stringent constraints that compound the analytical bottleneck. Here, we propose and evaluate a variant of an established simplex method as an effective approach to stage 2 scouting experiments intended to identify ‘sweet spots’.

Simplex algorithms, and in particular the variable size simplex algorithm (VSSA) [9], have been widely used for the unconstrained optimization of scalar objective functions. Until comparatively recently, it – along with other direct-search methods that do not use derivatives – was often regarded as a heuristic method, with no guarantees of convergence (e.g. [10]). However, simplex variants have been shown to be convergent [11,12], thereby providing theoretical support to what has always been a remarkably effective method in practice. Simplex algorithms have also been used in the bioprocess area [13,14], but their use in preparative purification (as opposed to analytical chromatography) is more limited [15]. Recent work from our group [3] has shown that the deployment of a simplex method is an efficient approach for carrying out early-stage bioprocess development studies. After the generation of the initial simplex (a polygon with  $k+1$  corners in a  $k$ -dimensional space of input variables), only those experiments suggested by the method are evaluated in an iterative fashion generating useful information and feedback rapidly.

This paper extends our approach by introducing the hybrid experimental simplex algorithm (HESA) as an alternative route for identifying ‘sweet spots’ during scouting studies. HESA is best suited to data generated on a coarse grid of input variables, of the kind commonly generated by deploying high level factorial designs. Standard simplex algorithms cannot deal readily with such constant and coarse grid data. Hence, HESA incorporates modifications which are described in a first section of this paper. It is shown that HESA works well on typical plate data whereas VSSA fails. HESA is then compared to established DoE methods for scouting, in two practical purification problems, and shown to be a viable and useful alternative.

## 2. Materials and methods

### 2.1. Simplex algorithm

The variable size simplex algorithm [9] (VSSA) is an established simplex algorithm that optimizes numerical functions iteratively by using logical rules to traverse a search space rapidly (more so than another established alternative called the fixed-size simplex algorithm [16]) [17]. Central to VSSA is the geometrical shape of a simplex, a polygon with  $k+1$  vertices in a space of  $k$  input variables, and with each vertex being associated with a value for the optimized objective function achieved at that location in the input variable space. Vertices are ranked from best to worst, according to the objective function at every point, and by using simple calculations the algorithm can then define the next simplex. VSSA generates a path of simplices iteratively that lead the search toward better regions of the underlying response surface. The possible movements calculated by VSSA in an iteration are shown in Fig. 1.

### 2.2. Hybrid experimental simplex algorithm

In the proposed methodology (Fig. 2), HESA uses VSSA, with pre-defined settings to calculate reflection, expansion, inside/outside contraction and shrink vertices [18], to calculate the coordinates of vertices which are then replaced with grid conditions by HESA’s approximation procedures. When this replacement does not lead to a *Minimal simplex* and/or degeneracy [19] and/or Next to Worst (N2W) reflections (a type of reflection used by the fixed-size simplex algorithm and HESA), then HESA carries out identical movements to standard VSSA. Otherwise *Actions* and *Rules* that are unique to HESA are used to obtain a new simplex and proceed to the next iteration. A *Minimal simplex* is defined as a simplex for which the differences between its vertex coordinates are equal to the step sizes of neighboring inputs. When a *Minimal simplex* is obtained during a HESA search, then certain VSSA movements (inside contractions and shrinks) become impossible.

#### 2.2.1. Approximation procedures

HESA uses two types of procedures to approximate VSSA vertices by experimental grid conditions (Table 1). These employ the Euclidean distance metric and include tie-breaking rules that take into account the type of movement dictated by VSSA. For example, in two dimensions, during the approximation of a VSSA outside contraction vertex ( $C_r$ ), which is equidistant to four grid conditions, the one that is the furthest away from the reflected  $W$  simplex vertex is selected to substitute the VSSA  $C_r$  vertex. In higher dimension spaces it is possible that even after the first tie-breaker is applied, multiple grid points still remain and a second tie-breaker is applied to select the grid point which is closest to the best simplex vertex. This is based on the weighted centroid principle of Ryan et al. [20] that the true gradient of the surface lies closest to the direction defined by the line connecting worst and best vertices.

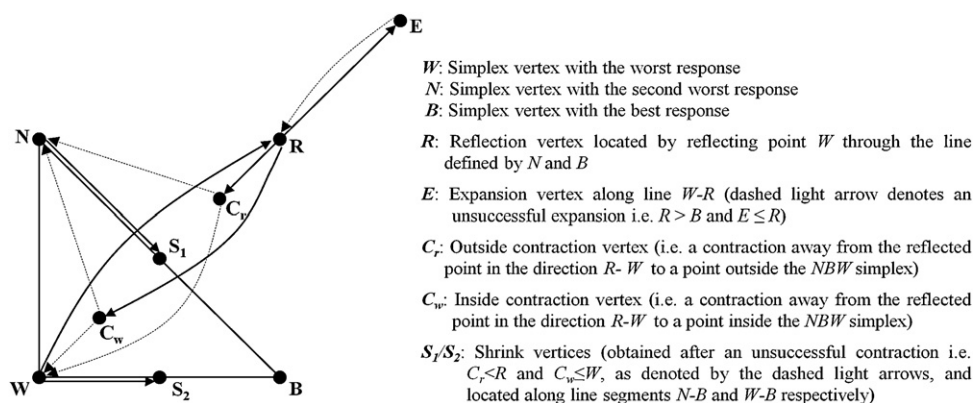


Fig. 1. Example of movements in an iteration of the variable size simplex algorithm.

During the approximation of VSSA shrink vertices when at least one of these vertices can be substituted by multiple equidistant grid conditions, all possible simplices are generated by substituting these VSSA shrink vertices by any of the relative candidate grid conditions and a non-degenerated simplex is then chosen. This approximation procedure also ensures that the VSSA vertices and their approximations all lie on the same half-plane (e.g. a set of three points defining a plane yields three sets of half-planes). If multiple simplices were to be formed after the screens described above, the average distances of the  $k$  new vertices (i.e. grid conditions) of each simplex from the common best simplex vertex would be calculated before selecting the ones resulting in the smallest average distance.

If the two groups of approximation procedures still suggest more than one alternative grid point or simplex respectively, then one is chosen randomly for the next iteration. This random selection uses the auto-corrective nature of the algorithm to discard sub-optimal conditions in later iterations and so correct the search direction [21,22].

### 2.2.2. Actions and Rules

In *Actions* the grid conditions that have been obtained and evaluated throughout an iteration are used to obtain a new simplex and proceed to the next iteration, with the exception of *Action 6* which is a shrink step to deal with cases wherein the approximation of a VSSA inside or outside contraction vertex has led to a degenerated simplex. The *Rules* on the contrary are themselves new approximation procedures which result in the identification of a new set of grid conditions. Different *Actions* are applied depending on the specific sequences of movements occurring in an iteration of HESA (Table 2) and they employ, in their majority,  $N2W$  reflections, where the second least favorable vertex ( $N$ ) is reflected instead of the conventional worst one ( $W$ ), in order to avoid oscillation that would otherwise cause the search to enter a loop that merely switches between the same points repeatedly. *Actions* 3, 4 and 7 handle and apply  $N2W$  reflections in the same manner as used in the fixed-size simplex algorithm [16]. The latter two are applicable only when the

**Table 1**  
Description of procedures for the approximation of the variable size simplex algorithm (VSSA) vertices.

Reflection, expansion, inside/outside contraction	Shrink
1. Calculate VSSA vertex	1. Calculate the $k$ VSSA shrink vertices
2. Exclude grid conditions occupied by the vertices of the current simplex	2. Exclude conditions occupied by the vertices of the current simplex
3. Find conditions with minimum distance from the VSSA vertex	3. Find conditions with minimum distance from each VSSA shrink vertex
4. If VSSA movement is reflection, expansion or outside contraction	4. If each VSSA shrink vertex can be replaced by one condition.
4.1 Find the conditions furthest away from the reflected vertex	4.1 Replace the calculated vertices by their respective selected conditions
5. If VSSA movement is inside contraction	5. If at least one VSSA shrink vertex can be replaced by more than one condition
5.1 Find the conditions closest to the reflected vertex.	5.1 Replace each of the $k$ VSSA shrink vertices by each of their respective selected conditions to generate all possible simplices
6. Find the condition closest to the best vertex of the current simplex	5.2. Exclude the degenerate simplices
	5.3 Exclude the simplices for which the direction of the shrink is reversed
	5.4 Find simplices with minimum average distance of the $k$ new vertices, from the best vertex

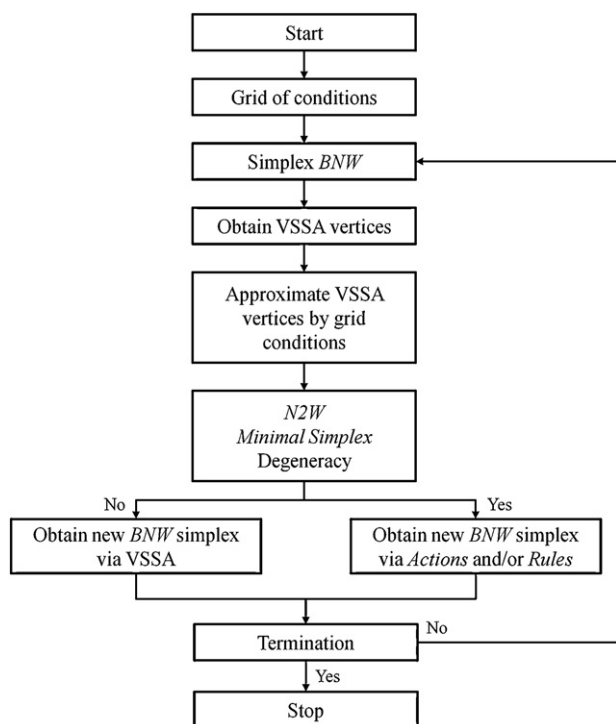


Fig. 2. Proposed methodology for the application of the hybrid experimental simplex algorithm based on the variable size simplex algorithm (VSSA).

**Table 2**  
Description of Actions.

Action	Sequence of movements	Description of Action
1	N2W expansion followed by: (a) Unsuccessful expansion or (b) Violation of experimental bounds or (c) Degeneracy	Substitute <i>N</i> simplex vertex by <i>R</i>
2	Successful N2W expansion	Substitute <i>N</i> simplex vertex by <i>E</i>
3	(a) <i>Minimal simplex</i> and N2W $R \geq W$ or (b) N2W $R > W$	Substitute <i>N</i> simplex vertex by <i>R</i>
4	<i>Minimal simplex</i> and N2W $R < W$	Substitute <i>N</i> simplex vertex by <i>R</i> and carry out a N2W reflection in next iteration
5	(a) <i>Minimal simplex</i> and unsuccessful outside contraction or (b) Unsuccessful outside contraction followed by degeneracy in shrink	Substitute <i>W</i> simplex vertex by <i>C<sub>r</sub></i> and carry out a N2W reflection in next iteration
6	Degeneracy in outside/inside contraction	Perform a shrink
7	<i>Minimal simplex</i> and $R \leq W$	Substitute <i>W</i> simplex vertex by <i>R</i> and carry out a N2W reflection in next iteration
8	Unsuccessful inside contraction followed by degeneracy in shrink	Substitute <i>W</i> simplex vertex by <i>C<sub>w</sub></i> and carry out a N2W reflection in next iteration

simplex in the beginning of an iteration is a *Minimal simplex* while the former is applicable for any type of simplex. Actions 5 and 8 use the Deming and Parker procedure [23] instead of the original VSSA shrink as used previously [24,25] and they are employed to deal with unsuccessful inside or outside contractions that are followed by shrink steps which lead to degenerated simplices. Finally, in HESA, N2W reflections are allowed to undergo expansions to a vertex along the *N*–*R* line instead of the normal *W*–*R* line. These ‘N2W expansions’ are handled by Actions 1 and 2.

Different Rules are also applied depending on the specific sequences of movements that occur in an iteration of HESA (Table 3). Rule 1 is applied when degeneracy occurs after the approximation of a VSSA *R* vertex. Rule 3 is applied when the

approximation of the shrink vertices, carried out due to Action 6, leads to degenerated simplices alone. Rule 3 may instruct HESA to perform a N2W reflection in the following iteration depending on the comparison between the newly obtained *C<sub>r</sub>* or *C<sub>w</sub>* vertex and the *W* simplex vertex. Rule 4 is a shrink step which is applied when the simplex at the beginning of an iteration is not a *Minimal simplex* and a N2W reflection is either unsuccessful or the VSSA N2W *R* vertex is out of the experimental bounds. When this rule does not identify a suitable set of grid conditions, then HESA is terminated. If the simplex at the beginning of the iteration is a *Minimal simplex*, then Rule 4 is replaced by Action 4 or Rule 2 respectively. Rule 2 identifies an alternative simplex vertex for reflection in order to obtain a VSSA *R* vertex within the experimental bounds while the simplex obtained after the approximation of this reflection vertex is not degenerated and not obtained in previous iterations. In the event of a *Minimal simplex* with an orientation where only reflection of the most favorable vertex gives a new VSSA *R* vertex in the experimental space, Rule 2 allows its reflection. As a result the simplex search either moves toward a new optimum or it returns back to the reflected best vertex.

### 2.2.3. Stopping criteria

After identifying a promising region of the search space, inside contractions and shrinks take place and eventually a *Minimal simplex* is derived. Then HESA switches to the fixed-size simplex algorithm and the search encircles the optimum. It is during this switch-over that HESA results to the identification of a well resolved ‘sweet spot’ containing the experimental optimum that could then be studied further. In two dimensional spaces this encirclement generates close packing simplices, but in higher dimensions the simplices do not close pack complicating a termination decision [26]. Hence, in HESA when the search enters a state that encircles a condition, Rule 5 (Table 3) is used to construct a *Minimal simplex* that has not been obtained previously but with a rearranged orientation in a final effort to study the favorable area. Rule 5 is similar to the Gustavsson and Sundkvist rotation method [27] which can improve the performance of VSSA [28]. Similar to Rule 4, if this rule does not identify a suitable set of grid conditions, it terminates HESA. User-defined termination criteria can also be

**Table 3**  
Description of Rules.

Rule	Sequence of movements	Description of Rule
1	Degeneracy in reflection	Find grid condition closest to the VSSA <i>R</i> vertex, furthest away from the <i>Reflected simplex vertex</i> and closest to the <i>B</i> simplex vertex and is such that the obtained simplex is not degenerated.
2	<i>Minimal simplex</i> and VSSA <i>R</i> vertex is out of experimental bounds	1. Reflect simplex vertices from <i>W</i> to <i>N</i> (or <i>B</i> ) and identify the first simplex vertex that once reflected its VSSA <i>R</i> vertex falls within the bounds ( $Min_k - step\ size_k, Max_k + step\ size_k$ ) and for which the approximation of the VSSA <i>R</i> vertex by a grid condition yields a simplex with a minimum volume, which is not degenerated and not previously obtained. If no such simplex is found use Rule 5; else: 2. Evaluate the approximated <i>R</i> vertex; 3. If $R > Reflected\ simplex\ vertex$ rearrange simplex to reflect the identified simplex vertex in the next iteration; else substitute the <i>Reflected simplex vertex</i> by <i>R</i> and perform a N2W reflection in the next iteration
3	Action 6 leads to degeneracy	1. Find grid condition closest to the VSSA <i>C<sub>r</sub></i> or <i>C<sub>w</sub></i> vertex, furthest or closest to the <i>W</i> simplex vertex respectively and closest to the <i>B</i> simplex vertex and is such that the obtained simplex is not degenerated; 2. Evaluate the approximated <i>C<sub>r</sub></i> or <i>C<sub>w</sub></i> vertex and replace the <i>W</i> simplex vertex by <i>C<sub>r</sub></i> or <i>C<sub>w</sub></i> ; 3. If $C_r$ or $C_w \leq W$ perform a N2W reflection in the next iteration; else proceed normally
4	(a) N2W $R \leq W$ or (b) N2W <i>R</i> vertex is out of experimental bounds	Perform Shrink by finding grid conditions within an area inscribed by the simplex vertices extended in either direction by the step size of each of the <i>k</i> factors which have the smallest average distance from the VSSA Shrink vertices and are such that the obtained simplex has the smallest volume, is not degenerated and does not have reversed orientation
5	(a) <i>k</i> + 1 consecutive repeated simplices or (b) Rule 2 invocation	Carry out as Rule 4 with the difference that the grid conditions are sought in an area inscribed by the simplex vertices without the extension



used. For example, it is possible to stop a simplex very quickly, if it is not making useful progress, and save time and material to start another simplex elsewhere in the space of input variables.

To enable objective function maximization, HESA was encoded in Matlab 7.8 (The MathWorks® Incorporated, MA, U.S.A.) on a 2.53 GHz PC with 4 GB of RAM running Windows 7 (Microsoft Corporation, WA, U.S.A.). By deploying HESA, experimental data are generated on-the-fly, i.e. selection of experiments for evaluation and their evaluation is carried out iteratively until HESA terminates.

### 2.3. Case studies

Two case studies are used to demonstrate the deployment of HESA and its returned results. In the first study, HESA is used to identify loading conditions for an anion exchange chromatographic step that minimize the content of the target analyte (green fluorescent protein) in the flowthrough fraction from a two dimensional grid of 32 conditions. In the second study, HESA is used to maximize the binding capacity of an antibody fragment (FAB') to a cation exchange chromatography resin using a three dimensional grid with 72 conditions. In the first case study HESA is compared with a version of VSSA using only low level modifications (VSSA-LLM) in which the approximations of the VSSA vertices are based on simple roundings and neither *Actions* nor *Rules* are applied. In both case studies the ability of HESA to locate and define an accurate 'sweet spot' is compared to approaches employing high level factorial, central composite face centered (CCF) and 2 level factorial designs. To understand more about how HESA compares against DoE approaches, and in particular RSM, in terms of their associated experimental costs, HESA searches were carried out from different starting points (initial simplices) in both case studies. Regression analysis and evaluation of experimental designs were carried with Matlab (The MathWorks® Incorporated) and MINITAB® (Minitab Inc., PA, U.S.A.). Experimentation details for these two case studies are given below.

#### 2.3.1. GFP binding case study

**2.3.1.1. Production of green fluorescent protein (GFP) material.** The target analyte was recombinant GFP expressed in *Escherichia coli*. The starting material was obtained from an *E. coli* cell paste, dissolved in 5 mM Tris pH 7.5 and then disrupted by high pressure homogenization for four passes at 400 bar. The homogenate was spun at  $35,000 \times g$  for 40 min in an Avanti® J-E centrifuge (Beckman Coulter Inc., CA, U.S.A.). The resultant supernatant was diafiltered in 5 mM Tris pH 7.5 and concentrated using Vivaspin™ 20 MWCO 10,000 tubes (GE Healthcare, Uppsala, Sweden).

**2.3.1.2. High throughput anion exchange system.** Two factors (pH at eight levels [7.6–9.0 with a step size of 0.2 pH units] and sodium chloride in 50 mM Tris at four concentrations [0–54 mM with a step size of 18 mM]) – were used to generate the grid of 32 conditions in order to evaluate their impact on the binding of GFP to Capto DEAE (GE Healthcare) using a PreDicator™ 96-well filter plate (GE Healthcare) that was pre-filled with 2 µL of resin. A Tecan Freedom Evo® 200 station equipped with standard fixed soft Teflon® coated tips and controlled by Freedom EVOware® version 2.1 software (Tecan Group Ltd, Männedorf, Switzerland) was used to conduct all liquid handling steps involved in the preparation and the analysis of the samples in the study. The handling of the plate was performed according to the manufacturer's instructions unless stated otherwise. The combination of the two factors required 32 buffer solutions to equilibrate the filter plate. After equilibration 32 feed solutions were applied, prepared directly in the filter plate by adding 100 µL aliquots of the starting GFP material and 200 µL from a second set of 32 buffer solutions to obtain feed solutions with the desired pH and salt concentration. The final GFP concentration

in each feed aliquot was 0.45 mg mL<sup>-1</sup> and the total protein concentration was 6.5 mg mL<sup>-1</sup>. These were applied in triplicate to fill all 96 wells of the filter plate and so generate triplicates of all samples. The plate was agitated on an orbital shaker at 1100 rpm for 1 h and the filtrate (i.e. flowthrough) was collected in a 96-well collection plate. The 32 equilibration and loading solution compositions were all generated using a GE Healthcare in-house Microsoft Excel® based user interface (Microsoft Corporation) capable of generating worklists that are recognized by Freedom EVOware®.

**2.3.1.3. GFP quantification by HT protein express.** GFP content in the flowthrough fraction was quantified using an HT Protein Express assay kit operated on a LabChip® 90 station (Caliper Life Sciences Ltd., Halton, U.K.) according to the instructions of the manufacturer (Caliper Life Sciences Ltd.). Samples were prepared in a 96-well Thermowell® PCR plate (Corning Inc. Lifesciences) and the GFP concentration was calculated by monitoring the density of the band at 26 kDa in the electropherograms. The GFP measurements from the triplicated 32 flowthrough samples yielded a CV% of 6.8%.

#### 2.3.2. FAB' binding case study

**2.3.2.1. Production of FAB' material for chromatography study.** *E. coli* strain W3110 containing plasmid pTTOD A33 IGS2 (producing an antibody fragment (FAB') under the control of the tac promoter) was provided by UCB (Slough, UK). This was used to prepare a working cell bank as described by Bowering et al. [29] and the fermentation followed the protocol presented by Tustian et al. [30]. Upon completion, the fermentation broth was spun at 6300 rpm for 60 min using an Avanti® J-E centrifuge (Beckman Coulter Inc.). The supernatant was discarded and the cell pellet was resuspended to a final concentration of 7.1% (w/v) in periplasmic extraction buffer (100 mM Tris, pH 7.4, 10 mM EDTA) overnight at 60 °C while undergoing agitation at 200 rpm in an ISF-1-W shaker (Kühner AG, Birsfelden, Switzerland). The completion of this lysis step was followed by a second centrifugation step at the same conditions to pellet the spheroplasts, after which the FAB'-rich supernatant was collected. This was then diafiltered in water to a final concentration of 1.10 mg mL<sup>-1</sup> using a Pellicon® XL 50 membrane with a cut-off limit of 10 kDa (Millipore, Massachusetts, U.S.A.) connected to a Labscale TFF System with a reservoir volume of 500 mL (Millipore, Massachusetts, U.S.A.).

**2.3.2.2. High throughput cation exchange system.** Three factors (pH at six levels [5.0–7.5 with a 0.5 unit step size], sodium phosphate concentration at four levels [15–60 mM with a 15 mM step size] and impure lysate FAB' concentration at three levels [0.11–0.33 mg mL<sup>-1</sup> with a 0.11 mg mL<sup>-1</sup> step size]) – were used to generate a grid of 72 conditions in order to evaluate their impact upon the binding of FAB' to the cation exchange resin Capto S (GE Healthcare) using two PreDicator™ 96-well filter plates (GE Healthcare) that were pre-filled with 2 µL of resin. A Tecan Freedom Evo® 100 station controlled by Freedom EVOware® version 2.1 software (Tecan Group Ltd) was used to conduct all liquid handling steps employing BioRobotix tips (Molecular BioProducts Inc., California, U.S.A.). The combination of the three factors required 24 buffer solutions to equilibrate the filter plate and 72 feed solutions, each with the desired pH and buffer salt composition. The equilibration buffer solutions were prepared in a 24-well plate (E&K Scientific, California, U.S.A.) using stock phosphate solutions. An additional 24 salt solutions were prepared from stock phosphate solutions and these were applied along with the concentrated FAB' lysate solution to each well of the filter plate at the appropriate volumes, resulting in the required FAB' concentration, pH and salt concentration and a final supernatant volume of 300 µL. The handling of the plates and the samples was performed as described in the anion exchange study, with the exception that the incubation period was 8 h. All

experiments were performed in duplicate at room temperature. After incubating the resin with the 72 feed solutions and removing the supernatant by centrifugation, the concentration of FAB' in the filtrate (i.e. flowthrough) was measured by HPLC as described below.

**2.3.2.3. FAB' quantification by high performance liquid chromatography.** A 1 mL HiTrap™ Protein G HP Column (GE Healthcare) was operated on an Agilent 1200 series LC system (Agilent Technologies) at 2 mL min<sup>-1</sup>. The injection volume was 100 µL and the absorbance of the eluted FAB' was monitored at 220 nm. The loading and elution buffers were 20 mM sodium phosphate at pH 7.40 and pH 2.50 respectively and a step elution was used. FAB' concentrations in the samples were calculated using a calibration curve prepared from a stock solution of pure FAB' with concentrations in the range of 0.1–0.5 mg mL<sup>-1</sup>. Calibration samples were prepared in duplicate and the coefficient of determination of the standard curve was found to be close to unity. The FAB' measurements from the duplicated 72 flowthrough samples yielded a CV% of ~6.9%. Based on the measured FAB' flowthrough concentrations, i.e. [FAB'] (mg mL<sup>-1</sup>) as well as the initial FAB' concentration ([FAB']<sub>i</sub>) and the resin and sample volumes (*V*<sub>resin</sub> and *V*<sub>liquid</sub> respectively in mL), binding capacities (*Q*<sub>FAB'</sub> in mg mL<sup>-1</sup>), were calculated by using Eq. (1).

$$Q_{\text{FAB}'} = \frac{[\text{FAB}']_i - [\text{FAB}']}{V_{\text{resin}}} \times V_{\text{liquid}} \quad (1)$$

### 3. Results and discussion

#### 3.1. GFP case study

The [GFP] measurements from the flowthrough fraction are shown in Fig. 3A. The condition defined by pH 8.2, [NaCl]=0 mM gave the smallest GFP content in the flowthrough (i.e. the global optimum) while the condition at pH 8.8, [NaCl]=0 mM gave the second smallest GFP content (i.e. a local optimum). Based on the raw experimental data, a 'sweet spot' containing the optimum was defined and which could be further studied in an optimization follow up study employing a RSM design such as CCF (Fig. 3B).

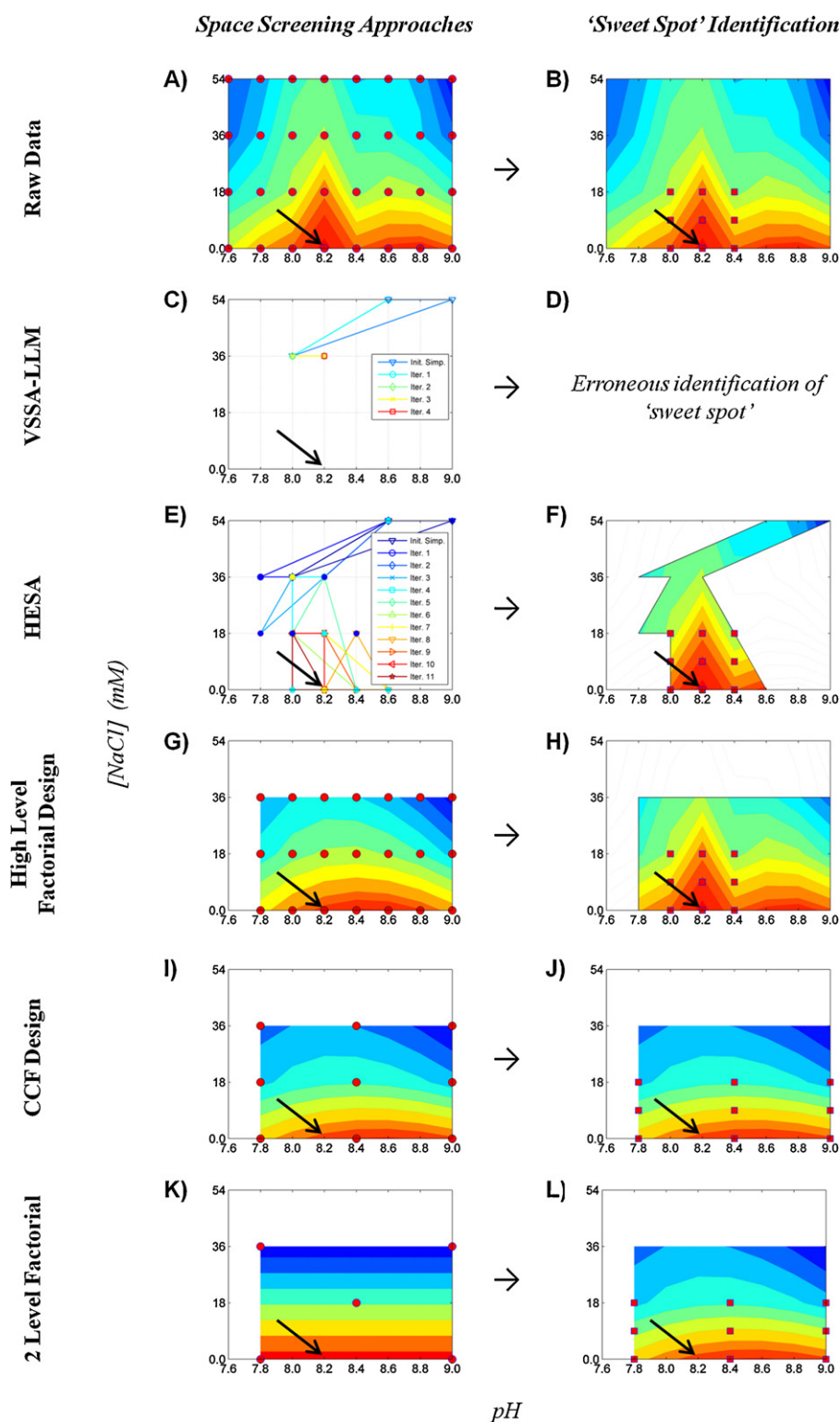
##### 3.1.1. HESA and VSSA-LLM deployments

The averaged GFP measurements from the 32 triplicated experiments were used to define the objective function which was then minimized by HESA and VSSA-LLM. Both simplex searches were initialized from the same initial simplex which was purposely located in an unfavorable area of the search space and away from the true optimum so as to allow the better description of the HESA movements. VSSA-LLM degenerated quickly from the second iteration (a simplex was obtained with all of its three vertices lying on a line) and by the end of the VSSA-LLM search (four iterations – five triplicated experiments) a simplex was obtained with all of its vertices lying on the same grid condition (pH 8.2 and [NaCl]=36 mM) (Fig. 3C). This condition was well away from the global and local optima. Thus VSSA-LLM displayed critical non-convergence and led to an entirely erroneous identification of a 'sweet spot' (Fig. 3D). By contrast, HESA converged to the global optimum after selecting 17 triplicated experiments (i.e. ~53% from the total available number of the 32 triplicated experiments) (Fig. 3E and Table 4). The global optimum was selected for the first time in the eighth iteration and in iterations 9–11 it was encircled to yield an accurate and well defined 'sweet spot' that could then be further assessed with follow up optimization studies (Fig. 3F).

##### 3.1.2. HESA, high level factorial, CCF, and 2 level factorial 'sweet spot' identification

The locations of the experimental space that were altogether less favorable (i.e. leading to high [GFP] in the flowthrough) were removed from the original experimental space and the remaining subsection was used to form and evaluate a high level factorial design (Fig. 3G), a CCF design (Fig. 3I) and a 2 level factorial design plus a center point (Fig. 3K) which all spanned the same factor ranges. The results from the evaluations of these three designs are detailed in Table 5. By comparing the predicted response surfaces, yielded by the three designs (Fig. 3G, I and K), to the raw experimental data (Fig. 3A), discrepancies were observed. The high level factorial and CCF designs returned predicted response surfaces which compared realistically to the one from the raw experimental data. However, they did not distinguish between the two optima and at the same time they overestimated the effect of the [NaCl] factor. This was also verified from Table 5 since for the high level factorial design a %Q<sup>2</sup> of 74.12 was obtained whereas when the CCF design model was challenged to predict the [GFP] of the remaining conditions of the space that were not part of the design (i.e. external validation of the model) a %R<sup>2</sup><sub>prediction</sub> of 56.82 was returned. Moreover, the residuals of the models from each design displayed a 'fanning out' behavior with increasing predicted [GFP] measurements. Hence, in this case study linear additive regression models encountered difficulties in capturing the trends in the experimental data accurately and reliably. In the case of the high level factorial design, since all the conditions in the experimental space are evaluated with such an approach, a 'sweet spot' would be defined based on the experimental data instead of relying on the moderately ill-fitting regression model. Consequently, the same 'sweet spot' to the one obtained by the deployment of HESA would be defined for further investigation and optimization (Fig. 3H). In the case of the CCF design, conservative follow up studies could lead to the definition of a new design, such as a CCF, spanning wider ranges of both factors (Fig. 3J) than those implied by the predicted [GFP] measurements. In such a case, due to the candidate 'sweet spot' covering a wide range for the pH factor, for which the [GFP] measurements underwent strongly non linear changes (shown from the raw experimental data (Fig. 3A), such follow up studies would be very likely to be repeated until a 'sweet spot' was defined that could be accurately modeled by a valid regression model as is the goal of these optimization follow up studies. In contrast, the deployments of HESA and of the high level factorial design, both returned the best candidate 'sweet spot' for such further investigation.

Finally, the deployment of a 2 level factorial design led to the least realistic description of the experimental data (Fig. 3K) since it wrongly indicated that [GFP] was affected solely by the [NaCl] factor (Table 5). While this design correctly led to the conclusion that low [NaCl] led to a decrease of GFP content in the flowthrough, the indication that [GFP] was not affected at all by the pH factor could lead to detrimental results if subsequent studies further studied only the [NaCl] factor. If on the contrary follow up studies were carried out by keeping the [NaCl] factor to its lowest setting, as indicated by the 2 level factorial design, while varying only the pH factor, despite the indications provided by the 2 level factorial design showing that it did not affect [GFP], then the absolute optimal operating condition would be identified. However, such studies would be of relatively low utility since they would provide limited insights regarding the robustness of the performance of the chromatographic separation in face of simultaneous variations in both, pH and [NaCl], factors. This also is one of the primary purposes of such follow up studies. A more cautious and correct strategy would be to augment the 2 level factorial design to the CCF design (Fig. 3I) which would then lead to the earlier obtained 'sweet spot' (Fig. 3L).



**Fig. 3.** GFP case study: (A) response surface of raw [GFP] data; (B) 'sweet spot' identification based on raw [GFP] data; (C) simplices obtained at the end of each iteration from the variable size simplex algorithm using the low level modifications (VSSA-LLM); (D) 'sweet spot' identification based on VSSA-LLM deployment; (E) simplices obtained at the end of each iteration from the hybrid experimental simplex algorithm (HESA); (F) 'sweet spot' identification based on HESA deployment; (G) response surface of predicted [GFP] data based on a high level factorial design; (H) 'sweet spot' identification based on high level factorial design deployment; (I) response surface of predicted [GFP] data based on a central composite face centered (CCF) design; (J) 'sweet spot' identification based on CCF design deployment; (K) response surface of predicted [GFP] data based on a 2 level factorial design plus center point; (L) 'sweet spot' identification based on a 2 level factorial design plus center point deployment. Circles denote experiments employed by the deployment of each approach. Squares denote proposed experiments for follow up studies based on a CCF design deployment. The arrow denotes the location of the experimental optimum (i.e. conditions leading to the lowest concentration of GFP in the flowthrough fraction).

**Table 4**  
Description of the movements of the hybrid experimental simplex algorithm (HESA) for the GFP case study. Where different the VSSA calculated vertices are shown in brackets. The light arrows show the sequence of the movements carried out during an iteration. A bold arrow represents the end of the iteration. The last column contains the coordinates of the vertices in the simplex formed at the end of each iteration ordered from best to worst ( $B, N, W$ ).

Iteration number		pH	[NaCl] (mM)		pH	[NaCl] (mM)		pH	[NaCl](mM)		pH	[NaCl] (mM)
0	Start	8.0	36									
		8.6	54									
		9.0	54									
1	Reflection	7.6	36	→	Outside contraction (successful)	7.8 (7.95)	36 (40.5)	→	8.0	36		
									8.6	54		
									7.8	36		
2	Reflection	8.8	54	→	Inside contraction (successful)	8.2 (8.05)	36 (40.5)	→	8.2	36		
									8.0	36		
									8.6	54		
3	Reflection	7.6	18	→	Outside contraction (successful)	7.8 (7.85)	18 (27)	→	8.2	36		
									8.0	36		
									7.8	18		
4	Reflection	8.4	54	→	Inside contraction (successful)	8.0 (7.95)	18 (27)	→	8.0	18		
									8.2	36		
									8.0	36		
5	Reflection	8.2	18	→	Expansion (successful)	8.4 (8.3)	0 (9)	→	8.4	0		
									8.0	18		
									8.2	36		
6	Reflection	<i>Boundary Violation</i>		→	Inside contraction (successful)	8.2	18 (22.5)	→	8.4	0		
									8.2	18		
									8.0	18		
7	Reflection	(8.2)	(–18)	→	Expansion	<i>Boundary Violation</i>		→	8.6	0		
		8.6	0						8.4	0		
									8.2	18		
8	Reflection	<i>Boundary Violation</i>		→	Inside contraction (unsuccessful)	(8.9)	(–9)	→	8.2	0		
						8.4 (8.35)	18 (9)		8.4 (8.5)	18 (9)		
									8.6	0		
									8.4	18		
9	Reflection	(8.8)	(–18)	→	Inside contraction	Degeneracy		→	8.2	0		
		<i>Boundary Violation</i>							8.2 (8.3)	18 (9)		
									8.4	0		
									8.2	18		
10	Reflection	(8.4)	(–18)	→	Inside contraction	8.4	0 (9)	→	8.0 (8.3)	18 (0)		
		<i>Boundary Violation</i>				<i>Minimal simplex</i>						
									8.2	0		
									8.2	18		
									8.0	18		
11	Next to worst reflection	(8.4)	(–18)	→	Stop				8.2	0		
		8.0	0						8.0	0		
									8.0	18		



**Table 5**

Details of experimental designs and regression analysis for the investigation of the [GFP] data.

	High level factorial design		CCF design		2 Level factorial design plus center point	
Runs	63 (21 triplicated)		27 (9 triplicated)		15 (5 triplicated)	
Model terms	Coefficient value <sup>a</sup>	p-Value	Coefficient value <sup>a</sup>	p-Value	Coefficient value <sup>a</sup>	p-Value
Constant	0.17 (69)	<0.0001	0.1914	<0.0001	0.1910 (0.1905)	<0.0001
pH	0.0039	0.1831 <sup>b</sup>	0.0012	0.7098 <sup>b</sup>	(0.0013)	(0.7584)
[NaCl]	0.0314	<0.0001	0.0328	<0.0001	0.0359 (0.0359)	<0.001
pH × [NaCl]	0.0092	0.0119	0.0078	0.0638 <sup>c</sup>	(0.0078)	(0.0866)
pH <sup>2</sup>	0.0225	<0.0001	0.0125	0.0383	NA	NA
[NaCl] <sup>2</sup>	−0.0091	0.0293	−0.0139	0.0231	NA	NA
Center point	NA	NA	NA	NA	(0.0028)	(0.7685)
ANOVA						
Model						
DF	5		5		1	
SS	0.0493		0.0222		0.0155	
MS	0.0098		0.0044		0.0155	
F-ratio	42.1341		23.0521		71.4890	
p-Value	<0.0001		<0.0001		<0.0001	
Residual error						
DF	57		21		13	
SS	0.0133		0.0040		0.0028	
MS	0.0002		0.0002		0.0002	
Lack of fit						
DF	15		3		3	
SS	0.0046		0.0004		0.0007	
MS	0.0003		0.0001		0.0002	
F-ratio	1.46		0.6321		1.2675	
p-Value	0.1673		0.6038		0.3376	
Pure error						
DF	42		18		10	
SS	0.0088		0.0037		0.0020	
MS	0.0002		0.0002		0.0002	
Summary of fit						
%R <sup>2</sup>	78.71		84.59		84.61	
%Q <sup>2</sup>	74.12		74.12		79.03	
External validation						
Runs	NA		36 (12 triplicated)		48 (16 triplicated)	
Residual error						
SS	NA		0.0138		0.0258	
MS	NA		0.0004		0.0005	
Summary of fit						
%R <sup>2</sup> <sub>prediction</sub>	NA		56.82		38.77	

<sup>a</sup> Coefficient values are based on coded factors.<sup>b</sup> Retained to support model hierarchy.<sup>c</sup> Retained based on an observed deterioration of goodness of fit upon removal; for the 2 level factorial design model, terms in brackets indicate coefficient values and p-values prior to model reduction.

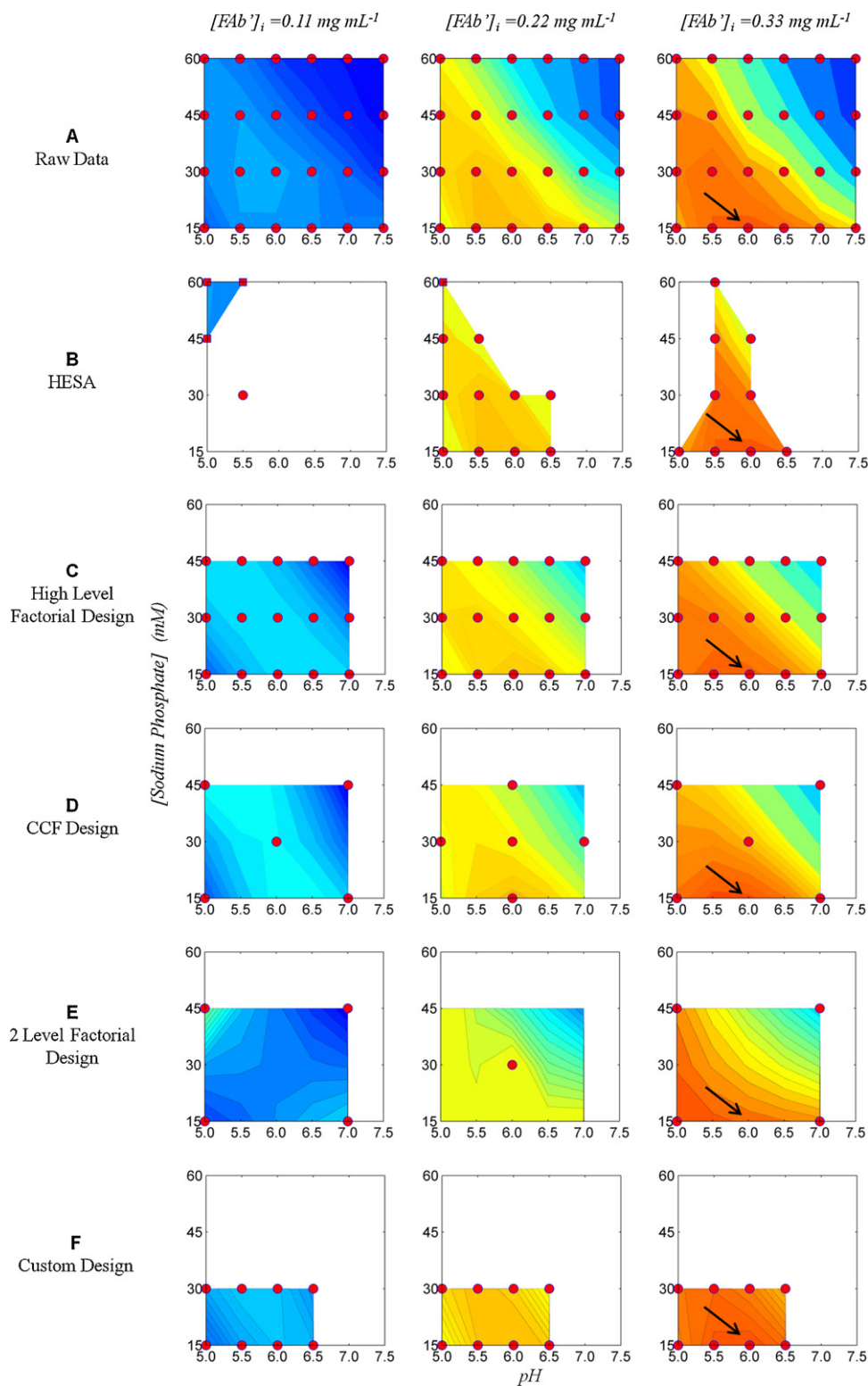
### 3.2. FAB' case study

#### 3.2.1. HESA, high level factorial, CCF, and 2 level factorial 'sweet spot' identification

The raw  $Q_{FAB'}$  measurements are provided in Fig. 4A. These measurements have a global optimum at the grid condition pH 6.0, [sodium phosphate] = 15 mM and  $[FAB']_i = 0.33 \text{ mg mL}^{-1}$ . The  $[FAB']$  measurements from the 72 duplicated experiments were used to calculate  $Q_{FAB'}$  values (Eq. (1)). These values were averaged and used to define the objective function which was maximized by HESA. A HESA search was purposely initialized from a simplex with its vertices lying in an unfavorable area of the search space and after selection of 24 duplicated experiments (i.e. 33% from the total available number of the 72 duplicated experiments) it converged to the global optimum and encircled it (Fig. 4B). Similar to the first case study the HESA results could then be used to identify an accurate and well defined 'sweet spot', enclosing the experimental optimum, for investigation in follow up studies.

In this case study, a smaller experimental space was also defined by removing from the original space those conditions that were altogether unfavorable (i.e. low  $Q_{FAB'}$  values) in order to form and

evaluate a high level factorial design (Fig. 4C), a CCF design (Fig. 4D) and a 2 level factorial design plus a center point (Fig. 4E) which all spanned the same factor ranges. Details from the evaluations of these designs are shown in Table 6. The models from each of these designs displayed a significant lack of fit. This is easily understood in the case of the 2 level factorial design since without the ability to evaluate the quadratic terms of the three factors (pH, [sodium phosphate], and  $[FAB']_i$ ) it indicated that follow up studies should investigate a 'sweet spot' spanning a pH range that was also extended toward low values (i.e. lower than pH 5.0) (Fig. 4E) and it thus led to the least realistic description of the experimental data. By comparing the response surface from the raw experimental data (Fig. 4A) to the predicted response surfaces that were based on the high level factorial and CCF design models (Fig. 4C and D), it was observed that the predicted  $Q_{FAB'}$  values matched closely the measured ones despite the significant lack of fit. The latter was therefore further investigated by observing the residuals of the two models. For both models, the residuals were found to exhibit a quadratic trend when they were plotted against the predicted  $Q_{FAB'}$  values. This trend was stronger for medium to low predicted  $Q_{FAB'}$  values and weaker for medium to high predicted  $Q_{FAB'}$  values. This was



**Fig. 4.** FAB' case study: (A) response surface of raw binding capacity,  $Q_{FAB'}$ , data; (B) segments of the response surface revealed by the deployment of the hybrid experimental simplex algorithm (HESA); (C)–(F) response surfaces of predicted  $Q_{FAB'}$  data obtained from the high level factorial, central composite face centered (CCF), 2 level factorial plus center point, and custom designs respectively. Circles denote experiments employed by the deployment of each approach. Squares denote experiments forming the starting point (initial simplex) for the deployment of HESA. The arrow indicates the experimental optimum (i.e. conditions leading to the highest  $Q_{FAB'}$  value).

further supported when a custom design was used to fit a regression model using an even smaller subset of the available experiments by removing additional conditions that led to poor  $Q_{FAB'}$  values from the original experimental space (Fig. 4F). Such a model returned excellent statistics with an insignificant lack of fit (Table 6) and

its residuals were found to be homoscedastic. Using the high level factorial design experiments (Fig. 4C), a model was fitted containing additional higher order terms, but the trend in the residuals remained and so did the significant lack of fit error. Data transformations (e.g.  $\log_{10}$  of response) did not counter the problem

**Table 6**  
Details of experimental designs and regression analysis for the investigation of the  $Q_{FAB}$  data.

	High level factorial design		CCF design		2 Level factorial plus center point		Custom design	
Runs	90 (45 duplicated)		30 (15 duplicated)		18 (9 duplicated)		24 (12 duplicated)	
Model terms	Coefficient value <sup>a</sup>	p-Value	Coefficient value <sup>a</sup>	p-Value	Coefficient value <sup>a</sup>	p-Value	Coefficient value <sup>a</sup>	p-Value
Ct	27.8467	<0.0001	27.1978	<0.0001	20.0571	<0.0001	29.6277	<0.0001
pH	−3.9279	<0.0001	−3.7238	<0.0001	−3.2817	0.0007	0.0591	0.6687 <sup>b</sup>
[sod. phosph.]	−3.1325	<0.0001	−3.0293	<0.0001	−2.885	0.0017	−0.1694	0.1055 <sup>b</sup>
[FAB'] <sub>i</sub>	10.4308	<0.0001	10.3163	<0.0001	10.0726	<0.0001	12.0459	<0.0001
pH × [sod. phosph.]	−4.3306	<0.0001	−3.8945	<0.0001	−3.8945	0.0002	−1.0598	<0.0001
pH × [FAB'] <sub>i</sub>	−2.4221	<0.0001	−2.3716	0.0006	−2.3716	0.0058	−0.5743	0.0015
[sod. phosph.] × [FAB'] <sub>i</sub>	−2.1827	<0.0001	−2.6066	0.0002	−2.6066	0.0033	−0.5642	<0.0001
pH <sup>2</sup>	−3.2692	<0.0001	−4.3669	0.0003	NA	NA	−1.8799	<0.0001
[sod. phosph.] <sup>2</sup>	−1.2507	0.0025	–	–	NA	NA	–	–
[FAB'] <sub>i</sub> <sup>2</sup>	−3.4571	<0.0001	−2.7133	0.0129	NA	NA	−4.2300	<0.0001
Center point	NA	NA	NA	NA	8.9546	0.0013	NA	NA
ANOVA								
Model								
DF	9		8		7		8	
SS	9178.9244		3270.1632		2513.0637		4915.1782	
MS	1019.8805		408.7704		359.0091		614.3973	
F-ratio	319.0817		73.1976		48.7201		1224.7518	
p-Value	<0.0001		<0.0001		<0.0001		<0.0001	
Residual error								
DF	80		21		10		39	
SS	255.7039		117.2740		73.6881		19.5644	
MS	3.1963		5.5845		7.3688		0.5017	
Lack of fit								
DF	35		6		1		15	
SS	212.4731		99.7198		65.3602		8.1539	
MS	6.0707		16.6200		65.3602		0.5436	
F-ratio	6.3191		14.2018		70.6347		1.1434	
p-Value	<0.0001		<0.0001		<0.0001		0.3739	
Pure error								
DF	45		15		9		24	
SS	43.2308		17.5541		8.3279		11.4105	
MS	0.9607		1.1703		0.9253		0.4754	
Summary of fit								
%R <sup>2</sup>	97.29		96.54		97.15		99.60	
%Q <sup>2</sup>	96.42		91.95		83.96		99.34	
External validation								
Runs	NA		60 (30 duplicated)		72 (36 duplicated)		NA	
Residual error								
SS	NA		217.5863		1322.2860		NA	
MS	NA		3.6264		18.3651		NA	
Summary of fit								
%R <sup>2</sup> <sub>prediction</sub>	NA		96.39		80.43		NA	

<sup>a</sup> Coefficient values are based on coded factors.

<sup>b</sup> Retained to support model hierarchy.

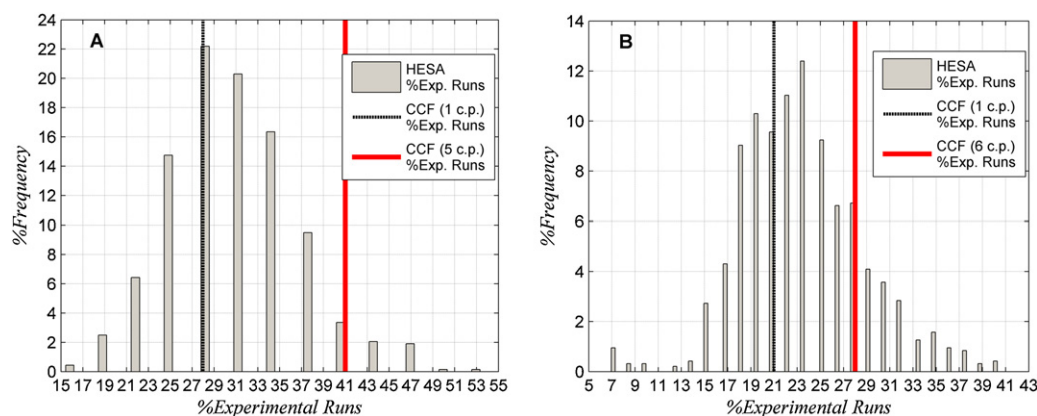
either. Hence, in the case of the high level factorial design the 'sweet spot' could be identified based on the experimental data, leading thus to the same 'sweet spot' obtained via the deployment of HESA. In the case of the CCF design, since a higher order model would most likely not result to an insignificant lack of fit, either external validation runs could be employed, which would increase the experimental costs, or a 'sweet spot' could be identified in a conservative manner and thus be greater in size than the one obtained from HESA and the high level factorial design deployments, but still correctly focusing on the experimental optimum.

### 3.3. Comparison of HESA and DoE approaches

To assess the extent to which HESA can be considered as a viable alternative approach for carrying out scouting studies during early bioprocess development, it is compared against DoE approaches in two aspects: (1) Ability to return accurate and well resolved 'sweet spots'; (2) Involved experimental burden.

#### 3.3.1. 'Sweet spot' identification

Both case studies support the earlier observation that when investigating large experimental spaces, as is the case for bioprocess development scouting studies, ill-fitting regression models may be obtained. Instead, when the space is narrowed down, valid regression models can be derived with greater ease (e.g. the regression model based on the custom design in the second case study). Based on the CCF designs from the two case studies, 'sweet spots' could be obtained that also contained the experimental optima. The 'sweet spot' obtained in the second case study would be better defined than the one obtained in the first case study. However, for both cases, and particularly for the first one, the obtained 'sweet spots' could be less well defined than those obtained by deploying HESA and the high level factorial designs. The latter approach will always be capable of yielding such accurate and well defined 'sweet spots' regardless of the shape of the underlying response surface, being accompanied however by high experimental costs. This is not the case for HESA or for other DoE approaches such as RSM and 2 level factorial designs. HESA may be affected by the presence of multiple optima, similar to its parent algorithm VSSA, while DOE



**Fig. 5.** Comparison of hybrid experimental simplex algorithm (HESA) experimental burden from 1000 random searches to experimental burden from response surface modeling employing central composite designs (CCDs) for: (A) case study 1; and (B) case study 2.

approaches can also be affected by the presence of multiple optima and in general they may not deal well with non linear responses even if they contain only one optimum. In the first case study, when 1000 random HESA searches (i.e. starting from different, randomly defined, initial simplices) were carried out, 64% and 36% of them resulted to the definition of a 'sweet spot' encircling the global and local optima respectively. In the second case study, where only one optimum was present in the experimental space, more than 98% of 1000 random HESA searches resulted in the correct derivation of a 'sweet spot' encircling the optimum. Hence, the ability of HESA to return correctly located 'sweet spots' is affected by the presence of multiple optima. However, it needs to be underlined that HESA will always locate and characterize an optimal condition, even if it is not a global optimum, as opposed to the deployment of a DoE approach. This is an attractive feature of HESA since if the associated performance of the investigated system is found to be acceptable, then the identified 'sweet spot' can still be further assessed and optimized in follow up studies. The fact that in the first case study the 'sweet spot' yielded by the CCF design included both global and local optima is not a generic feature of RSM and DoE methodologies. Instead, it is equally likely that in other cases of responses with multiple and well separated optima or non linear trends, a DoE approach could lead to an entirely erroneously identified 'sweet spot'. Hence, the adoption of HESA in scouting studies appears to occupy the middle ground between approaches adopting a high level factorial design and other conventional DoE methodologies, such as RSM, in terms of its ability to yield 'sweet spots' for further investigation. Finally, employing 2 level factorial designs for identifying process relevant 'sweet spots' was shown to be in both case studies the least successful approach. This was expected as such designs are not meant to be used to obtain predictions of responses. Instead they are used to screen experimental factors.

### 3.3.2. Involved experimental burden

To compare HESA and DoE approaches, and in particular RSM designs, in terms of their associated experimental burden, the number of replicated experiments selected by HESA from the 1000 random searches mentioned earlier, were compared against RSM approaches employing a central composite design with one or multiple center points for each case study (Fig. 5). The number of replicated experimental runs from each approach were normalized against the total number of available experimental runs and thus expressed in terms of percentages. For example, the HESA search that selected 17 triplicated conditions in the first case study yielded an experimental burden of  $100 \times (17 \times 3) / (32 \times 3) \approx 53\%$  whereas the deployment of a CCD with one center point yielded an experimental burden of  $100 \times (9 \times 3) / (32 \times 3) \approx 28\%$ .

Based on Fig. 5A, in the first case study  $\sim 45\%$  of the HESA searches resulted in a number of experimental runs that was approximately equal or lower than that required by a CCD with one center point. An additional  $\sim 50\%$  of the HESA searches were accompanied with an experimental burden that was between those from the CC designs with one and five center points whereas the remaining  $\sim 5\%$  of the searches required a greater number of experimental runs than the latter design. For the second case study, Fig. 5B showed that the equivalent percentages were  $\sim 38\%$ ,  $\sim 46\%$ , and  $\sim 16\%$  respectively. According to these results,  $\sim 40\%$  and  $\sim 80\%$  of the HESA searches required a number of runs that was equal or lower than those required by CC designs with one or multiple center points respectively in both case studies. Hence, both HESA and RSM approaches are more rapid than adopting a high level factorial design approach for identifying 'sweet spots' in scouting studies during early bioprocess development. Between them the involved experimental costs appear to be comparable. The latter conclusion is also supported by the fact that HESA application does not require a downtime for 'off-line' data analysis by an end-user between successive experimental studies. This is because HESA does not require a mathematical model; instead it only uses directly measurable experimental data. The same does not apply for RSM approaches. The duration of such a downtime period is not insignificant as can be understood by the elaborated data analyses in the two case studies.

### 3.4. Implications of HESA

As mentioned earlier the hybrid experimental simplex algorithm does not guarantee the identification of a 'sweet spot' enclosing the global optimum. Like any other simplex method, it may find local optima. The likelihood of finding the global optimum can be improved by deploying multiple searches. This will also inevitably increase the associated experimental costs but not in a multiplicative manner due to the finite number of experiments within a space investigated by HESA. For example, for each case study the 1000 random HESA searches were split in half and combined to simulate 500 paired searches starting from two different and randomly defined initial simplices. For the first case study 85% of these paired searches encircled the global optimum, whereas for the second case study all paired searches encircled the single optimum. The experimental burden was increased on average by  $\sim 15\%$  and  $\sim 10\%$  respectively. This increase can be minimized by initializing a search from an initial simplex with good size, location, and orientation, possibly based on initial experiments or previous experience. Such optimally started HESA searches are faster than their random alternatives and display improved convergence. The



derivation of rules of thumb for initializing HESA and the investigation of strategies for deploying multiple searches is the topic of ongoing research in our group.

Finally, HESA is also affected by the presence of experimental error. Such effects were countered in the described application examples by running all HESA suggested experiments at least in duplicate. This is in agreement with good laboratory practice regardless of the way that a set of experiments is to be performed. In cases where the experimental error is expected to be overly great and extensive replication is required, then a DoE approach may be deployed instead since such approaches can usually deal with experimental errors without requiring extensive replication.

#### 4. Conclusions

Scouting studies in early bioprocess development aim to identify a process viable 'sweet spot' for further investigation in follow up studies. Such 'sweet spots' are routinely identified by deploying high level factorial or RSM designs. An alternative methodology, the hybrid experimental simplex algorithm, was introduced in this paper, and it was demonstrated in two case studies which sought to optimize the binding conditions for GFP to a weak anion exchange resin and FAB' to a cation exchange resin. The obtained results suggest that HESA can readily deal with experimental spaces that cannot be directly searched via the application of the standard variable size simplex algorithm even when the latter is lightly modified. At the same time deployment of HESA was found to be comparable to the deployment of RSM designs regarding its ability to yield well defined and accurate 'sweet spots' and its associated experimental costs. These results suggest that HESA can be considered as a viable alternative approach for carrying out scouting studies in early bioprocess development which aim to identify process relevant 'sweet spots' for further investigation and optimization by traditional DoE methodologies.

#### Nomenclature

[FAB']	(mg mL <sup>-1</sup> )	Concentration of FAB'
[FAB'] <sub>i</sub>	(mg mL <sup>-1</sup> )	Initial concentration of FAB'
[GFP]	(mg mL <sup>-1</sup> )	Concentration of green fluorescent protein
[NaCl]	(mM)	Concentration of sodium chloride
[sodium phosphate]	(mM)	Concentration of sodium phosphate
<i>B</i>	(–)	Vertex of a simplex with the most favorable objective function value
<i>C<sub>r</sub></i>	(–)	Outside contraction vertex or movement
<i>C<sub>w</sub></i>	(–)	Inside contraction vertex or movement
<i>E</i>	(–)	Expansion vertex or movement
HESA	(–)	Hybrid experimental simplex algorithm
<i>k</i>	(–)	Number of factors in experimental space
<i>Max<sub>k</sub></i>	(–)	Maximum value of factor <i>k</i>
<i>Minimal simplex</i>	(–)	A simplex of minimal size
<i>Min<sub>k</sub></i>	(–)	Minimum value of factor <i>k</i>
<i>N</i>	(–)	Vertex of a simplex with the second less favorable objective function value
<i>N<sub>1</sub></i>	(–)	Vertex of a simplex with the second most favorable objective function value for three factor experimental spaces
<i>N<sub>2</sub></i>	(–)	Vertex of a simplex with the second less favorable objective function value for three factor experimental spaces
<i>N2W</i>	(–)	Next to worst
<i>R</i>	(–)	Reflection vertex or movement
<i>S<sub>1</sub></i>	(–)	Shrink vertex across <i>N–B</i> line
<i>S<sub>2</sub></i>	(–)	Shrink vertex across <i>W–B</i> line
<i>step size<sub>k</sub></i>	(–)	Absolute difference between two adjacent levels of factor <i>k</i>
<i>V<sub>liquid</sub></i>	(mL)	Volume of liquid solution

<i>V<sub>resin</sub></i>	(mL)	Volume of chromatographic resin
VSSA	(–)	Variable size simplex algorithm
VSSA-LLM	(–)	Variable size simplex algorithm with low level modifications
<i>Q<sub>FAB'</sub></i>	(mg mL <sup>-1</sup> )	Binding capacity for FAB'
<i>W</i>	(–)	Vertex of a simplex with the least favorable objective function value

#### Acknowledgements

The support of the Engineering and Physical Sciences Research Council (EPSRC) Center for Innovative Manufacturing in Emergent Macromolecular Therapies is gratefully acknowledged. The Center is part of The Advanced Center for Biochemical Engineering, Department of Biochemical Engineering, University College London, with collaboration from a range of academic partners and biopharmaceutical companies (<http://www.ucl.ac.uk/biochemeng/industry>). The support for Dr. A. Velayudhan as an EPSRC Manufacturing Fellow is gratefully appreciated. The contribution of GE Healthcare to the EngD of S. Konstantinidis is also duly acknowledged.

#### References

- [1] P. McKenzie, S. Kiang, J. Tom, A.E. Rubin, M. Futran, Can pharmaceutical process development become high tech? *AIChE J.* 52 (2006) 3990–3994.
- [2] R.S. Islam, D. Tisi, M.S. Levy, G.J. Lye, Framework for the rapid optimization of soluble protein expression in *Escherichia coli* combining microscale experiments and statistical experimental design, *Biotechnol. Prog.* 23 (2007) 785–793.
- [3] S. Chhatre, S. Konstantinidis, Y. Ji, S. Edwards-Parton, Y. Zhou, N.J. Titchener-Hooker, The simplex algorithm for the rapid identification of operating conditions during early bioprocess development: Case studies in FAB' precipitation and multimodal chromatography, *Biotechnol. Bioeng.* 108 (2011) 2162–2170.
- [4] J.L. Coffman, J.F. Kramarczyk, B.D. Kelley, High-throughput screening of chromatographic separations: I. Method development and column modeling, *Biotechnol. Bioeng.* 100 (2008) 605–618.
- [5] S.D. Doig, S.C.R. Pickering, G.J. Lye, J.M. Woodley, The use of microscale processing technologies for quantification of biocatalytic Baeyer–Villiger oxidation kinetics, *Biotechnol. Bioeng.* 80 (2002) 42–49.
- [6] B.D. Kelley, M. Switzer, P. Bastek, J.F. Kramarczyk, K. Molnar, T. Yu, J.L. Coffman, High-throughput screening of chromatographic separations: IV. Ion-exchange, *Biotechnol. Bioeng.* 100 (2008) 950–963.
- [7] K. Rege, M. Pepsin, B. Falcon, L. Steele, M. Heng, High-throughput process development for recombinant protein purification, *Biotechnol. Bioeng.* 93 (2006) 618–630.
- [8] M. Bensch, P.S. Wierling, E. von Lieres, J. Hubbuch, High throughput screening of chromatographic phases for rapid process development, *Chem. Eng. Technol.* 28 (2005) 1274–1284.
- [9] J.A. Nelder, R. Mead, A simplex method for function minimization, *Computer J.* 7 (1965) 308–313.
- [10] K.I.M. McKinnon, Convergence of the Nelder–Mead method, *SIAM J. Optim.* 9 (1998) 148–158.
- [11] P. Tseng, Fortified-descent simplicial search method: a general approach, *SIAM J. Optim.* 10 (1999) 269–288.
- [12] C.J. Price, I.D. Coope, D. Byatt, A convergent variant of the Nelder Mead algorithm, *J. Optim. Theory Appl.* 113 (2002) 5–19.
- [13] B. Buttr-Indra, W. Kasinrerkb, S. Tayapiwatanab, Sequential simplex optimization of recombinant biotinylated surviving production by *Escherichia coli* using mineral supplementation, *Biochem. Eng. J.* 46 (2009) 115–120.
- [14] J.A. Lawitts, J.D. Biggers, Optimization of mouse embryo culture media using simplex methods, *J. Reprod. Fertil.* 91 (1991) 543–556.
- [15] A. Velayudhan, M. Teeters, H. Shen, D. Bezila, P. Alred, High-throughput and automated screening methods to facilitate purification process development of antibodies, oral presentation, in: Society of Biological Engineering (SBE) International Conference on Accelerating BioPharmaceutical Development, San Diego, March 2007, 2007.
- [16] W. Spendley, G.R. Hext, F.R. Himsworth, Sequential application of simplex designs in optimisation and evolutionary operation, *Technometrics* 4 (1962) 441–461.
- [17] E. Morgan, K.W. Burton, G. Nickless, Optimization using the modified simplex method, *Chemom. Intell. Lab. Syst. J.* 7 (1990) 209–222.
- [18] J.C. Lagarias, J.A. Reeds, M.H. Wright, P.E. Wright, Convergence properties of the Nelder–Mead simplex method in low dimensions, *SIAM J. Optim.* 9 (1998) 112–147.
- [19] P.F.A. van der Wiel, R. Maassen, G. Kateman, The symmetry-controlled simplex optimization procedure, *Anal. Chim. Acta* 153 (1983) 83–92.
- [20] P.B. Ryan, R.L. Barr, H.D. Todd, Simplex techniques for nonlinear optimization, *Anal. Chem.* 52 (1980) 1460–1467.

- [21] D.E. Long, Simplex optimization of the response from chemical systems, *Anal. Chim. Acta* 46 (1969) 193–206.
- [22] S.L. Morgan, S.N. Deming, Simplex optimization of analytical chemical methods, *Anal. Chem.* 46 (1974) 1170–1181.
- [23] S.N. Deming, L.R. Parker Jr., M.B. Denton, A review of simplex optimization in analytical chemistry, *Crit. Rev. Anal. Chem.* 7 (1978) 187–202.
- [24] E.R. Åberg, A.G.T. Gustavsson, Design and evaluation of modified simplex methods, *Anal. Chim. Acta* 144 (1982) 39–53.
- [25] K.W.C. Burton, G. Nickless, Optimisation via simplex Part I. Background, definitions and a simple application, *Chemom. Intell. Lab. Syst.* 1 (1987) 135–149.
- [26] F.H. Walters, L.R. Parker Jr., S.L. Morgan, S.N. Deming, 6: General considerations, in: *Sequential simplex optimization: a technique for improving quality and productivity in research, development, and manufacturing*, CRC Press, Florida, 1999, pp. 184–189 (licensed electronic reprint by MultiSimplex AB).
- [27] A. Gustavsson, J-E. Sundkvist, Design and optimization of modified simplex methods, *Anal. Chim. Acta* 167 (1985) 1–10.
- [28] D.D. Burgess, Rotation in simplex optimization, *Anal. Chim. Acta* 181 (1986) 97–106.
- [29] L.C. Bowering, D.G. Bracewell, E. Kesharvarz-Moore, M. Hoare, A.N.C. Weir, Comparison of techniques for monitoring antibody fragment production in *E. coli* fermentation cultures, *Biotechnol. Progr.* 18 (2002) 1431–1438.
- [30] A.D. Tustian, H. Salte, N.A. Willoughby, I. Hassan, M.H. Rose, F. Baganz, M. Hoare, N.J. Titchener-Hooker, Adapted ultra scale-down approach for predicting the centrifugal separation behavior of high cell density cultures, *Biotechnol. Progr.* 23 (2007) 1404–1410.