A Multiobjective Evolutionary Optimization Framework for Protein Purification Process Design

Richard Allmendinger and Suzanne S. Farid

Department of Biochemical Engineering, University College London,
Torrington Place, London WC1E 7JE, UK
{r.allmendinger,s.farid}@ucl.ac.uk

Abstract. Increasing demand in therapeutic drugs has resulted in the need to design cost-effective, flexible and robust manufacturing processes capable of meeting regulatory product purity requirements. To facilitate this design procedure, a framework linking an evolutionary multiobjective algorithm (EMOA) with a biomanufacturing process economics model is presented. The EMOA is tuned to discover sequences of chromatographic purification steps, and equipment sizing strategies adopted at each step, that provide the best trade-off with respect to multiple objectives including cost of goods per gram (COG/g), robustness in COG/g, and impurity removal capabilities. The framework also simulates and optimizes subject to various process uncertainties and design constraints. Experiments on an industrially-relevant case study showed that the EMOA is able to discover purification processes that outperform the industrial standard, and revealed several interesting trade-offs between the objectives.

1 Introduction

The biotech sector is facing increasing pressures to design more cost-efficient, robust and flexible manufacturing processes [1]. Among biotech therapies, monoclonal antibodies (mAbs) represent one of the fastest growing category due to their unique binding specificity to targets. A typical antibody purification process is depicted in Figure 1: in upstream processing (USP) mammalian cells expressing the mAb of interest are cultured in bioreactors, whilst in downstream processing (DSP) the mAb is recovered, purified and cleared from viruses using a variety of operations. Of these steps, chromatography operations are identified as critical steps and can represent a significant proportion of the purification material costs. The design of cost-effective purification processes can help addressing this challenge.

The design stage is further complicated by the fact that regulatory bodies expect biopharmaceutical companies to fully understand their manufacturing process, thus account for uncertainty in the manufacturing process, and be able to establish a purification process that conforms to strict purity requirements. To assist the process of tackling these challenges, presented here is an optimization-based framework linking an evolutionary multiobjective optimization algorithm (EMOA) with a biomanufacturing process economics model. The goal of the EMOA is to discover sequences of chromatographic purification steps, and sizing strategies adopted at each step, that provide the best trade-off with respect to multiple objectives including cost of goods per

T. Bartz-Beielstein et al. (Eds.): PPSN XIII 2014, LNCS 8672, pp. 498–507, 2014.

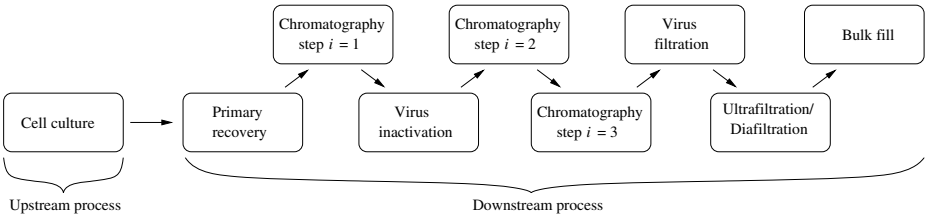


Fig. 1. Typical flowsheet for an antibody manufacturing process

gram (COG/g), robustness in COG/g, and impurity removal capabilities. The objectives are then computed by the process economics model serving as the fitness evaluation tool. Additional complexities accounted for by the framework include simulating and optimizing subject to uncertainty and constraints.

This paper extends our previous work on chromatography design/optimization [2,3], which assumed a fixed sequence of chromatography steps, and focused on tuning (using a single-objective EA) the chromatography column sizing adopted at each step such that the COG/g are minimized only. This extension posed two challenges including (i) the development of a customized EMOA accounting for constraints and variables of different type, and (ii) the extension of the process economics model so as to account for additional design choices and their impact on manufacturing performance.

Chromatography design/optimization can be tackled from several other angles. For example, the exploration of non-Protein A based purification processes was considered by Chen et al. [4]. Tuning of chromatographic operating conditions is another prominent research field [5], and so is resin screening [6]. Unlike our simulation-based work, these studies are based on real physical experiments. A simulation-based approach was also adopted by Liu et al. [7], where mathematical programming is proposed to address chromatography column sizing and sequencing in the context of biopharmaceutical facility design. Stonier et al. [8] proposed a discrete-event simulation for the selection of optimal chromatography column diameters over a range of titres.

The application of multiobjective optimization to chromatography design/ optimization has become popular only recently. For example, Nfor et al. [9] used EMO to tune operating parameters (e.g. column loading, flowrate and gradient length) of a single chromatography step so as to improve recovery yield, purity, and productivity. The focus in this paper is on optimizing "high-level" criteria relating to all chromatography steps (e.g. impurity removal capabilities) or the complete manufacturing process (e.g. COG/g and its robustness). Moreover, uncertainty is associated with global operating parameters (e.g. product titre and initial impurity levels) as well as with chromatography specific parameters (e.g. step yields and step specific removal capabilities).

2 Constrained Multiobjective Purification Process Design

The framework proposed is based on the following closed-loop: an EMOA creates solutions $\mathbf{x}$ (i.e. a sequence of chromatography steps and column sizing strategies), which are then decoded, embedded into a feasible manufacturing process (see Figure 1),

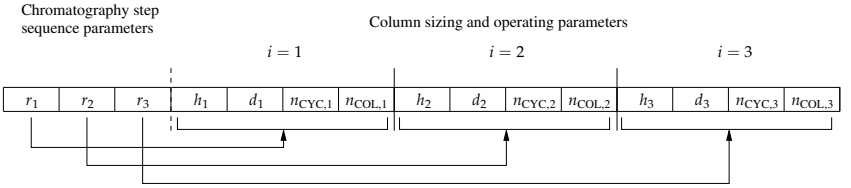


Fig. 2. Representation of a candidate solution for $k = 3$ chromatography steps. Each step $i = 1, ..., k$ is defined by a resin r_i and a column sizing strategy, which is composed of the bed height h_i and diameter d_i of columns, number of cycles $n_{\mathrm{CYC},i}$ each column is used, and the number of columns $n_{\mathrm{COL},i}$ operating in parallel.

and evaluated by a biomanufacturing process economics model; manufacturing uncertainties are accounted for using Monte Carlo (MC) trials. Objective values pertaining to $\mathbf{x}$ are recorded and fed back to the EMOA to be considered in the generation of future solutions. The decision variables, constraints, objective functions, and uncertain factors the problem is subject to are explained in more detail in the following.

Decision Variables: Figure 2 shows the string encoding developed to represent a purification process or solution $\mathbf{x}$. Assuming a fixed number of chromatographic steps k, the task is to define, for each step $i = 1, ..., k$, the resin $r_i \in \{\mathrm{resin}_1, ..., \mathrm{resin}_q\}$ and column sizing strategy, which is composed of the bed height h_i and diameter d_i of a column, number of cycles each column is used for $n_{\mathrm{CYC},i}$, and the number of columns operating in parallel $n_{\mathrm{COL},i}$. Therefore, the problem is subject to $l = k + 4 \cdot k$ variables in total. The choice of the resin r_i used dictates several chromatographic operation and cost parameters considered by the biomanufacturing process economics model, such as the step yield, resin price, and impurity removal capabilities. On the other hand, the sizing strategy adopted at each chromatographic step i defines the total volume of resin V_i available, and the processing time T_i that the chromatography step take; T_i and V_i are calculated as follows [10]:

$$V_i = \pi \cdot d_i^2/4 \cdot h_i \cdot n_{\mathrm{CYC},i} \cdot n_{\mathrm{COL},i} \tag{1}$$

$$T_i = n_{\mathrm{CYC},i} \cdot h_i \cdot (CV_{\mathrm{BUFF},i} + CV_{\mathrm{LOAD},i}/n_{\mathrm{COL},i}) \cdot u_i, \tag{2}$$

where $CV_{\mathrm{BUFF},i}$ and $CV_{\mathrm{LOAD},i}$ are the number of column volumes of buffer and product load per cycle, and u_i is the linear velocity of resin r_i.

Constraints: The problem is subject to three types of constraints:

1. *Chromatography sequence constraints* are defined on the variables $r_i, i = 1, ..., k$ and ensure that a purification process consists of non-identical, feasible and orthogonal (i.e. different typed) chromatography steps, or more formally

$$g_1: \quad r_i \neq r_j, \quad i, j = 1, ..., k, \quad i \neq j, \tag{3}$$

$$g_2: \quad r_i^i = 1, \quad i = 1, ..., k, \tag{4}$$

$$g_3: \quad r_i^T \neq r_j^T, \quad i, j = 1, ..., k, \tag{5}$$

where r_i^T denotes the resin type of r_i, and r_i^i is a boolean variable indicating whether resin r_i is permitted to be used at position i.

2. The *demand constraint* ensures that the annual amount of product manufactured P is sufficient to satisfy the annual demand D, or $g_4 : P \geq D$.
3. *Resin requirement constraints* act on the column sizing variables and ensure that, at each step $i = 1, ..., k$, there is sufficient resin volume V_i available to process the mass of product M_i coming in from the previous unit operation. Formally, this constraint can be defined as

$$g_5 : V_i \geq \frac{M_i}{r_{i,DBC} \cdot \kappa} \quad i = 1, ..., k, \tag{6}$$

where V_i is computed according to Equation (2), $r_{i,DBC}$ is the DBC of the resin used at step i, and $0 < \kappa \leq 1$ the maximum capacity utilization factor.

Manufacturing Uncertainties: Several uncertain factors arising in the manufacturing process are captured by the framework: (i) product titre, (ii) chromatography step yields, (iii) DBC, (iv) eluate volumes, (v) HCP log reduction, and (vi) initial HCP level. While uncertainties in (i) and (vi) are due to fluctuations arising in USP, the other factors are associated with the resins r_i available for selection and sensitivity of operating conditions. Uncertainties are modeled by associating each factor with a probability distribution (reflecting real-world variability) from which values are drawn at random during Monte Carlo (MC) trials; the way the data resulting from the MC trials is processed by the EMOA will be detailed in the next section.

Performance Metrics: Three objectives are considered to drive the search for cost-efficient and reliable purification process yielding highly pure products:

1. The *cost of goods per gram* $COG/g = C/P$, where C is the sum of annual direct costs (e.g. consumables and labor) and indirect costs (e.g. capital charge and facility-related costs) and P the annual product output, represent the costs for manufacturing a single gram of product and are to be *minimized*.
2. The *robustness in* COG/g, η, is defined here as the ratio

$$\eta = \frac{\sigma_{\text{COG/g}}}{\mu_{\text{COG/g}}} = \frac{\sqrt{\frac{1}{N}\sum_{j=1}^{N}(COG/g_j - \mu_{\text{COG/g}})^2}}{\mu_{\text{COG/g}}}, \tag{7}$$

where N is the number of MC trials performed for a specific process so far, COG/g_j the COG/g value at MC trial j, and $\mu_{\text{COG/g}} = \frac{1}{N}\sum_{j=1}^{N} COG/g_j$. The smaller the value of η, the less variation there is in COG/g in the presence of uncertainty. Hence, the objective is to *minimize* η.
3. The *probability of meeting purity requirements* $p(\textit{meeting required purity})$ is the probability that a purification process reduces the HCP impurity level in a product below a certain limit HCP^*. This probability is to be *maximized* and computed here by

$$p(\text{meeting required purity}) = \frac{1}{N}\sum_{i=1}^{N}\delta_i, \text{ where } \delta_i = \begin{cases} 1 & \text{if } HCPFinal_i < HCP^* \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where N is the number of MC trials performed, and $HCPFinal_i$ the final HCP level at MC trial i. The final HCP level is calculated here by $HCPFinal_i = HCP_{\text{initial}} / 10^{\sum_{i=1}^{k} r_{i,HCP}}$, where $r_{i,HCP}$ is the HCP log reduction of resin r_i.

All objectives are obtained by running the process economics model, which is based on mass balance and cost equations as defined in [10]. Note, whilst COG/g is a standard metric, the metrics, $p(\text{meeting required purity})$ and η, have not been considered in the literature so far.

3 Experimental Setup

This section describes the case study, EMOA and its parameter settings as used in the subsequent experimental analysis.

Case Study Setup: The case study was adopted from [3] and focuses on a single-product mAb manufacturing facility that employs a process flowsheet as shown in Figure 1 with $k = 3$ chromatography steps. Assumed is an annual demand of $D = 400$kg, a product titre of 3g/L, and a desired final HCP level of $HCP^* = 100$ng/mg (which is typical of final product specification limits for recombinant proteins). Two initial HCP levels are investigated, $HCP_{\text{initial}} = \{10^5, 10^6\}$ng/mg.

A total of $q = 10$ resins, comprising around 125 different sequences, are available for tuning the sequence of chromatography steps. For the characteristics of these resins, and technical details of the manufacturing process and resource cost assumptions please refer to [2]. Table 1 lists the uncertain parameters and their common levels of uncertainty in the context of triangular probability distributions; i.e. a variation of $x\%$ corresponds to the distribution $\text{Tr}(x \cdot (100 - x)/100,\ x,\ x \cdot (100 + x)/100)$. The value range of column sizing parameters is $15\text{cm} \leq h_i \leq 25\text{cm}$ (11 values), $50\text{cm} \leq d_i \leq 200\text{cm}$ (10 values), $n_{\text{CYC},i} \in \{1, ..., 10\}$cm, $n_{\text{COL},i} \in \{1, ..., 4\}$, $i = 1, 2, 3$; i.e. in total the search space comprises $(11 \cdot 10 \cdot 10 \cdot 4)^3 \cdot 125 \approx 10.5 \cdot 10^{12}$ different purification processes. The industrial platform employs a fixed and commonly used chroamtography step sequence, *PrA L→CEX L→AEX*, in combination with the sizing strategy (which is set based on empirical rules) $n_{\text{COL},i} = 1, h_i = 20\text{cm}, n_{\text{CYC},i} = 5, i = 1, 2, 3$, with d_i being adjusted such that the resulting total resin volume V_i (Equation (2)) satisfies the resin requirement constraint (Equation (6)).

Evolutionary Multiobjective Optimization: The focus in this work is to understand how EMO can be tuned to tackle the purification process design problem rather than comparing different EMOAs. Hence, to guide the search, the popular NSGA-II [11] is extended with methods for coping with the model uncertainties and constraints, which are explained below. The algorithm uses binary tournament selection, uniform crossover, and a mutation operator that selects a random value from the range of possible values.

Constraint-Handling Strategies: The *chromatography sequence constraints* (Equations (3) to (5)) are addressed by programming the sequence-related variables $r_i, i = 1, ..., k$ as a single variable S representing all feasible sequences. For population initialization, a sequence is selected at random from S. Crossover and mutation are applied directly on the variables r_i but resulting infeasible offspring are repaired by selecting a sequence from S that differs in as few steps $i = 1, ..., k$ as possible from the

Table 1. Probability distributions associated with uncertain factors

Uncertain factor	*Variation* (%)
Product titre	13.3
Chrom. step yields	5
DBC	10
Eluate volume	10
HCP logs	20
Initial HCP level	20

Table 2. Default parameter settings of the EMOA

Parameter	*Setting*
Population size μ	50
Per-variable mutation probability	$1/l$
Crossover probability	0.6
Number of generations G	50
Monte Carlo trials N	100

original sequence. Ties between equally close sequences are broken at random. The *demand constraint* is circumvented by setting up USP such that there is a slight product surplus. To cope with the *resin requirement constraint* (Equation (6)), a 'repairing' strategy is employed that iteratively increases the values of the column-sizing related variables (associated with a particular chromatography step i), one variable at the time, until sufficient resin is available (i.e. until Equation (6) is satisfied) or until the maximum value of a variable is reached, in which case the value of another variable is increased. The sequence in which variables are modified affects the performance of the optimizer as indicated in [3] for the single-objective case. The default sequence adopted is $d_i \rightarrow n_{\mathrm{CYC},i} \rightarrow h_i \rightarrow n_{\mathrm{COL},i}$, which performed best in [3], but alternative sequences will be considered in the experimental study.

Uncertainty-Handling Strategy: Model uncertainties are accounted for by exposing a manufacturing process to N MC trials with values of uncertain factors being drawn at each trial from the probability distribution associated with the factors. The objective values of a solution were then the averages of the different performance metrics across the N trials, and these averages were updated if the same solution is evaluated multiple times during the search.

The experimental study presents a sensitivity analysis of the performance metrics, and investigates the robustness of the EMOA using the proposed framework. The default settings of the EMOA are given in Table 2. To allow for fair comparison of processes discovered by the EMOA, all processes present in the final population are evaluated using 1000 MC trials. Any results shown are average results across 30 independent algorithm runs.

4 Experimental Study

Sensitivity Analysis to Identify Global Drivers of Cost and Purity: Figure 3 uses the idea of tornado plots to show the impact of several uncertain factors on COG/g and the final HCP level *HCPFinal*. The boxplots have been created based on 10000 randomly generated, feasible and unique purification processes. For each process, plotted is the overall maximal effect on the two metrics of the best and worst case setting of the uncertain factors. It can be observed from the plots that generally the impact of model parameters depends on the objective being optimized, and increasing the number of parameters affects performance more significantly. There seems to be a symmetric

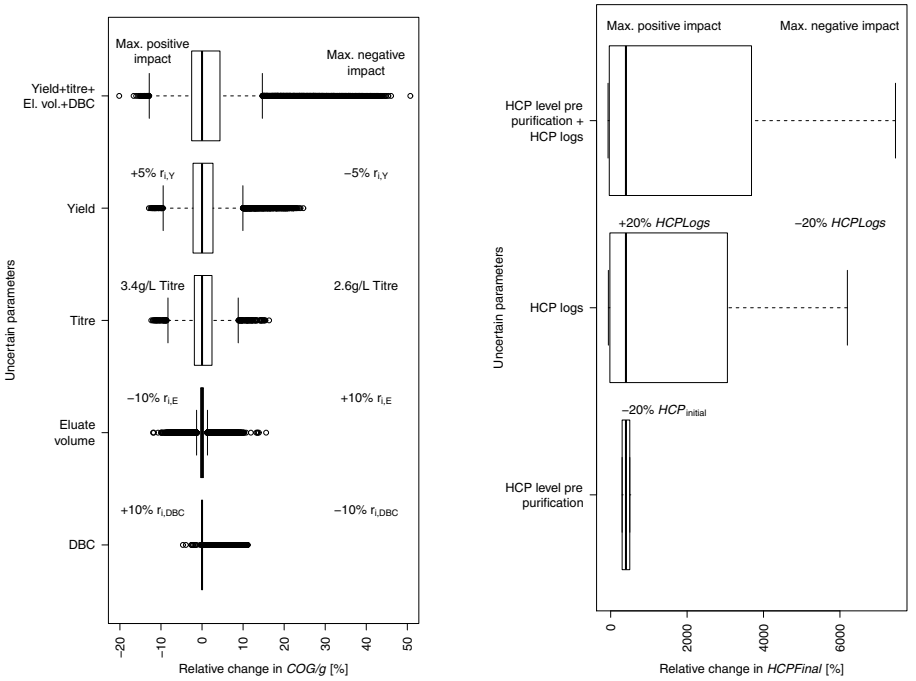


Fig. 3. Tornado diagrams illustrating the overall maximal effects of best (left of zero) and worst (right of zero) case settings of uncertain factors — DBC, elution volume, titre, and yield (bottom four boxplots in (a)) and HCP log reductions and initial HCP level (bottom two boxplots in (b)), and all four, respectively, two parameters at once (top boxplot) — on (a) COG/g and (b) final HCP levels *HCPFinal*

positive and negative impact of the uncertain factors on the objective COG/g (Figure 3(a)). The step yield and titre are most sensitive to uncertainty as they have a direct impact on the mass of product manufactured (and thus the denominator of the metric COG/g). Uncertainties in initial HCP levels and HCP log reductions are the only factors that affect the final HCP level *HCPFinal* (Figure 3(b)). The negative effect on *HCPFinal* is significantly greater than the positive effect because many of the (randomly generated) purification processes are able to reduce the HCP level down to $HCPFinal \approx 0$ (though this might be associated with high COG/g), leaving limited scope for further improvements. Note, a reduction in *HCPFinal* translates into an increase in $p(\text{meeting required purity})$.

Tuning an EMOA to Cope with Uncertainty and Constraints: This section gives a taste of how the discovery process of optimal purification processes can be affected by the choice of algorithm parameter settings. Figure 4 uses the concept of (median) attainment surfaces [12] to visualize the typical convergence behavior of the EMOA (top left plot) and the performance impact on the EMOA by two algorithm settings,

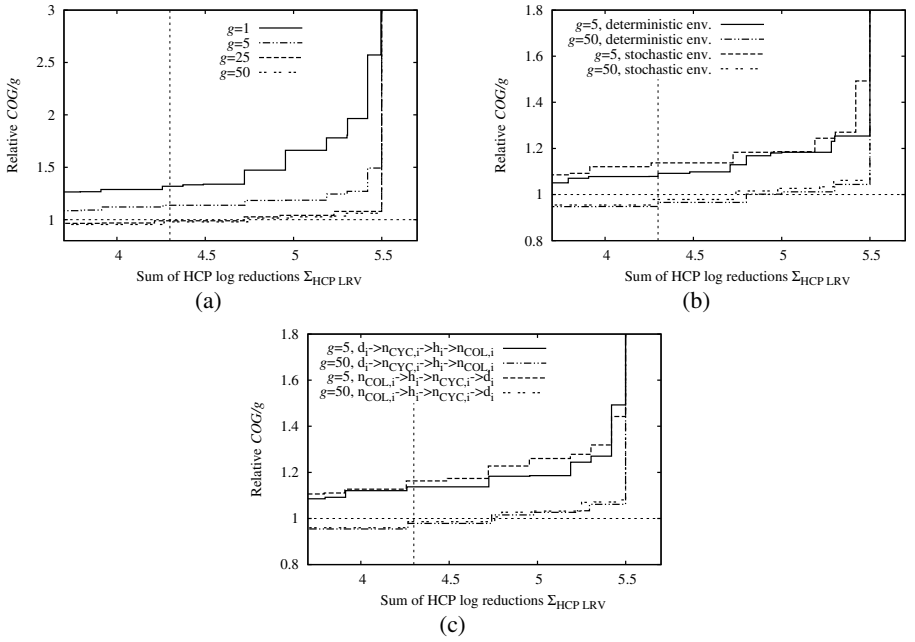
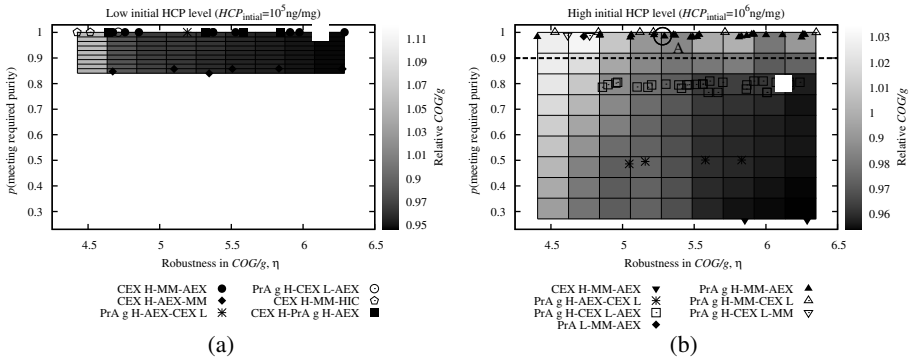


Fig. 4. Median attainment surfaces (MASs) obtained by an EMOA optimizing COG/g and the sum of HCP log reductions $\sum_{\text{HCP LRV}} = \sum_{i=1}^{k} r_{i,HCP}$. The plots show (a) the MASs at different generations g with uncertainty, (b) MASs for different g obtained in a deterministic and stochastic environment, and (c) the MASs for different g and repairing strategies with uncertainty. The performance of the industrial platform is indicated by the dashed horizontal and vertical lines.

the number of MC trials N (top right plot) and the constraint-handling strategy (bottom plot). In all three plots, the EMOA minimized the COG/g and the sum of HCP log reductions $\sum_{i=1}^{k} r_{i,HCP}$.

From Figure 4(a) it can be seen that the EMOA needs to be run for around $g \approx 25$ generations to match and outperform the industrial platform. Comparing the convergence speed and final solution quality obtained by the EMOA with and without uncertainty (Figure 4(b)), it is apparent that uncertainty harms both aspects significantly. Figure 4(c) shows that the constraint-handling strategy adopted is crucial too. In fact, repairing according to the scheme $d_i \rightarrow n_{\text{CYC},i} \rightarrow h_i \rightarrow n_{\text{COL},i}$ yields best results as increasing the column diameter d_i first is often sufficient to satisfy the resin requirement constraint without sacrificing processing time significantly.

EMO Applied to all Three Objectives Subject to Uncertainty: Figure 5 uses heatmaps to visualize the trade-off between all three objectives for two HCP levels $HCP_{\text{initial}} = 10^5$ng/mg (Figure 5(a)) and 10^6ng/mg (Figure 5(b)). Several trade-offs can be observed from the plots: (i) the range of the metric $p(\text{meeting } \textit{required purity})$ increases with the initial HCP level, (ii) the COG/g increases as $p(\text{meeting } \textit{required purity})$ increases and/or η decreases, and (iii) an improvement in the robustness η is achieved by adopting smaller column dimensions (supporting figure not shown here). The heatmaps can also be exploited to make design decisions. For example, assume that the goal of a



(a) (b)

Fig. 5. (Pareto) optimal purification processes discovered by an EMOA optimizing three objectives, COG/g, $p(\text{meeting required purity})$, and η, for two different HCP levels $HCP_{\text{initial}} = 10^5$ng/mg (a) and 10^6ng/mg (b). The performance of the industrial platform is indicated by the big white square in each plot. The response surface was generated by interpolating the processes' objective values using the Kriging function, Krig(), from the *fields* package of the statistical software R.

manufacturer is to establish a process with $p(\text{meeting } \textit{required purity}) > 0.9\%$. Whilst in this case there is no incentive to deviate from the industrial platfrom from the perspective of COG/g and $p(\text{meeting required } \textit{purity})$ for a low initial HCP level, a different sequence is needed for a high initial HCP level. For instance, the sequence $PrAg\,H \rightarrow MM \rightarrow AEX$, as indicated by the letter A in Figure 5(b), meets the purity requirements without increasing the COG/g significantly.

5 Summary and Conclusion

Presented was a framework for designing cost-efficient and robust chromatographic purification process that yield pure biopharmaceuticals. The framework comprised a process economics model and an EMOA, which optimized the sequence of chromatography steps and column sizing strategies with respect to multiple objectives and subject to uncertainty. Validating the framework on an industrially-relevant case study revealed that the performance impact of an uncertain factor depends on the objective being optimized. This knowledge can be used e.g. to diagnose which process parameters need a tighter control. Furthermore, the framework was able to discover purification processes that outperform the industrial platform, and revealed interesting trade-offs between objectives that can facilitate the design of purification process. Future research will look at extending the framework to cover additional design choices and investigate more efficient uncertainty-handling strategies.

References

1. Kelley, B.: Industrialization of mab production technology: The bioprocessing industry at a crossroads. mAbs 1, 443–452 (2009)
2. Simaria, A.S., Turner, R., Farid, S.S.: A multi-level meta-heuristic algorithm for the optimisation of antibody purification processes. Biochemical Engineering Journal 69, 144–154 (2012)

3. Allmendinger, R., Simaria, A.S., Turner, R., Farid, S.S.: Closed-loop optimization of chromatography column sizing strategies in biopharmaceutical manufacture. Journal of Chemical Technology and Biotechnology (2013), doi:10.1002/jctb.4267
4. Chen, J., Tetrault, J., Zhang, Y., Wasserman, A., Conley, G., DiLeo, M., Haimes, E., Nixon, A.E., Ley, A.: The distinctive separation attributes of mixed-mode resins and their application in monoclonal antibody downstream purification process. Journal of Chromatography A 1217(2), 216–224 (2010)
5. Treier, K., Berg, A., Diederich, P., Lang, K., Osberghaus, A., Dismer, F., Hubbuch, J.: Examination of a genetic algorithm for the application in high-throughput downstream process development. Biotechnology Journal 7, 1203–1215 (2012)
6. Rathore, A.S.: Resin screening to optimize chromatographic separations. LC–GC 19(6), 616–622 (2001)
7. Liu, S., Simaria, A.S., Farid, S.S., Papageorgiou, L.G.: Designing cost-effective biopharmaceutical facilities using mixed-integer optimization. Biotechnology Progress 29(6), 1472–1483 (2013)
8. Stonier, A., Smith, M., Hutchinson, N., Farid, S.S.: Dynamic simulation framework for design of lean biopharmaceutical manufacturing operations. Computer Aided Chemical Engineering 26, 1069–1073 (2009)
9. Nfor, B.K., Zuluaga, D.S., Verheijen, P.J.T., Verhaert, P.D.E.M., van der Wielen, L.A.M., Ottens, M.: Model-based rational strategy for chromatographic resin selection. Biotechnology Progress 27(6), 1629–1643 (2001)
10. Farid, S.S., Washbrook, J., Titchener-Hooker, N.J.: Modelling biopharmaceutical manufacture: Design and implementation of SimBiopharma. Computers & Chemical Engineering 31, 1141–1158 (2007)
11. Deb, K., Agrawal, S., Pratap, A., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2), 182–197 (2002)
12. Fonseca, C.M., Fleming, P.J.: On the performance assessment and comparison of stochastic multiobjective optimizers. In: Ebeling, W., Rechenberg, I., Voigt, H.-M., Schwefel, H.-P. (eds.) PPSN IV. LNCS, vol. 1141, pp. 584–593. Springer, Heidelberg (1996)