Computers and Chemical Engineering 68 (2014) 151–164

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Optimising chromatography strategies of antibody purification processes by mixed integer fractional programming techniques

Songsong Liu[a], Ana S. Simaria[b], Suzanne S. Farid[b], Lazaros G. Papageorgiou[a,*]

[a] Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK
[b] The Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK

ARTICLE INFO

Article history:
Received 24 January 2014
Received in revised form 28 April 2014
Accepted 3 May 2014
Available online 14 May 2014

Keywords:
Biopharmaceutical manufacturing processes
mAb
Chromatography purification
MINLP
MILFP
Dinkelbach algorithm

ABSTRACT

The strategies employed in chromatography steps play a key role in downstream processes for monoclonal antibody (mAb) manufacture. This work addresses the integrated optimisation of chromatography step sequencing and column sizing in mAb purification processes. Chromatography sequencing decisions include the resin selection at each typical step, while the column sizing decisions include the number of columns, the column diameter and bed height, and number of cycles per batch. A mixed integer nonlinear programming (MINLP) model was developed and then reformulated as a mixed integer linear fractional programming (MILFP) model. A literature approach, the Dinkelbach algorithm, was adopted as the solution method for the MILFP model. Finally, an industrially-relevant case study was investigated for the applicability of the proposed models and approaches.

1. Introduction

The biopharmaceutical industry has been a rapidly growing sector during the past decade. Accounting for 15.6% of the total pharmaceutical market, the global biopharmaceutical market value reached $138 billion in 2011, and is expected to be over $320 billion by 2020 (GBI Research, 2012). Monoclonal antibodies (mAbs) are emerging therapies that have been widely used to treat cancer and autoimmune diseases. Yet mAb manufacturers are faced with increased upstream productivities and purification capacity constraints that can result in purification bottlenecks. Hence it is important to address how to find the most cost-effective purification processes (Low et al., 2007; Langer, 2009; Pujar et al., 2009).

A key decision for mAb purification is the selection of the chromatography sequence. There usually exist multiple chromatography steps in the mAb purification process, and each step has a number of suitable candidate resins/types for selection. The candidate resins often have different characteristics, e.g., yield, price, dynamic binding capacities. Here an importation issue is how to choose the best combination of resins/types for all chromatography steps to be of most benefit to the whole downstream process. Also, at each chromatography step, another key decision is the column sizing strategy, e.g. opting to run a smaller column for several cycles so as to reduce resin costs or a large column for fewer cycles so as to save time and labour costs. Decisions on the chromatography column sizes include the selection, the bed height and diameter of each column and the number of cycles to run and the number of columns to use in parallel.

Recently, many computer-based decisional tools for the bioprocess sector have been developed (Farid et al., 2005; Lim et al., 2006; Chhatre et al., 2006; Farid et al., 2007; Pollock et al., 2013a), and some models have been proposed to assess different solutions for the design and operation of chromatography steps (Joseph et al., 2006; Chhatre et al., 2007; Chan et al., 2008; Stonier et al., 2012; Simaria et al., 2012; Pollock et al., 2013b; Allmendinger et al., 2012, 2013; Bentely and Kawajiri, 2013). Also, mathematical programming models and frameworks have been widely applied to optimisation problems in the pharmaceutical and biopharmaceutical industries, e.g., capacity planning (Papageorgiou et al., 2001; Levis and Papageorgiou, 2004) and production planning (Lakhdar et al., 2005, 2007). At the process level, a number of mathematical programming optimisation models were developed to find the optimal purification sequences, using physicochemical data of protein mixtures and mathematical correlations of the separation techniques. Vasquez-Alvarez et al. (2001) and Vasquez-Alvarez and

* Corresponding author. Tel.: +44 20 76792563; fax: +44 20 73832348.
E-mail address: l.papageorgiou@ucl.ac.uk (L.G. Papageorgiou).

http://dx.doi.org/10.1016/j.compchemeng.2014.05.005

Notation

Indices

bf	bulk fill step
h	harvest step
i	column size
j	column number
k	cycle number
n	digit of the binary representation
r	resin
s	step
t	resin type
ufdf	ultrafiltration step
vf	virus filtration step
vi	virus inactivation step

Sets

BER	set of resins in bind-elute mode
CS	set of chromatography steps = capture, intermediate purification, polishing
FTR	set of resins in flow-through mode
R_s	set of resins suitable to step *s*
R_t	set of resins of the resin type *t*

Parameters

a, *b*, *c*	utilities cost coefficients
aot	annual operating time, day
bcv_r	buffer usage of resin *r*, CV
bpc	buffer price, £/L
brc	bioreactor cost, £
brf	scale-up factor of bioreactor cost
brn	number of bioreactors
brt	bioreaction time, day
brv	bioreactor volume, L
cc_{si}	column cost of size *i* at chromatography step *s*, £
cf	scale-up factor of column cost
cv_{si}	volume of column size *i* at step *s*, L
cy_{sr}	product yield of resin *r* at chromatography step *s*
dbc_r	dynamic binding capacity of resin *r*, g/L
dm_{si}	diameter of column size *i* at chromatography step *s*, L
don	number of operators for downstream processing
dvr_s	diafiltration volume ratio of step *s*
ecv_r	elute volume of resin *r*, CV
el	equipment lifetime, year
fconc	final concentration of product, g/L
fvr_s	flush volume ratio of step *s*
gef	general equipment factor
gu	unit general utilities cost, £/L
h_{si}	height of column size *i* at chromatography step *s*, cm
$i\lambda$	insurance cost ratio to the fixed capital investment
j_s	maximum number of columns at chromatography step *s*, $maxcn_s$
k_s	maximum number of cycles at chromatography step *s*, $maxcyn_s$
l_r	life time of resin *r*, cycle
lang	Lang factor
maxbn	maximum number of batches
maxbbv	maximum buffer volume per batch
$maxcn_s$	maximum number of columns at chromatography step *s*
$maxcv_s$	maximum column volume at chromatography step *s*
$maxcyn_s$	maximum number of cycles at chromatography step *s*
$maxpv_s$	maximum product volume at step *s*
$ma\lambda$	maintenance cost ratio to the fixed capital investment
mepc	media price, £/L
$mi\lambda$	miscellaneous material cost ratio to chemical reagent and consumable costs
$m\lambda$	management cost ratio to direct labour cost
ncy_s	product yield of non-chromatography step *s*
nvr_s	neutralisation volume ratio of step *s*
$oe\lambda$	other equipment cost ratio to the bioreactor volume
of	overpacking factor for resin
q	maximum digit number in the binary representation of number of batches, ($\log_2$ *maxbn*)
$q\lambda$	QCQA cost ratio to direct labour cost
r	interest rate
rpc_r	resin price of resin *r*, £/L
refbrc	reference cost of bioreactor, £
refbrv	reference volume of bioreactor, L
refcc	reference cost of column, £
refdm	reference diameter of column, cm
sfd	duration per shift, h
sfn	number of shifts per day
st	seed train bioreaction time, day
$s\lambda$	supervisors cost ratio to direct labour cost
titre	product titre, g/L
$t\lambda$	tax cost ratio to the fixed capital investment
uon	number of operators per bioreactor in upstream processing
vel_r	linear velocity of flow for resin *r*, cm/h
w	wage of an operator, £/h
wt	DSP window
α	bioreactor working volume ratio
θ	media overfill allowance
μ	chromatography resin utilisation factor
σ	batch success rate

Continuous variables

ABV	annual buffer volume, L
AC	annual cost, £
AP	annual product output, g
AT	annual downstream operating time, day
BAT_s	time for adding buffer per batch at chromatography step *s*, min
BBV	buffer volume added per batch, L
BC	buffer cost, £
BT	downstream processing time per batch, day
BV_s	buffer volume per batch at step *s*, L
CAC	capital cost, £
CC	consumables cost, £
CRC	chemical reagents cost, £
DLC	direct labour cost, £
FCI	fixed capital investment, £
GUC	general utility cost, £
IC	insurance cost, £
LC	labour cost, £
M_0	initial product mass entering downstream processes per batch, g
M_s	product mass per batch remaining after step *s*, g
MAC	maintenance cost, £
MC	management cost, £

MEC media cost, £
MIC miscellaneous material cost, £
OBJ optimisation objective, COG/g, £/g
OIC other indirect costs, £
PLT_s time for loading product per batch at chromatography step s, min
PV_0 initial product volume entering downstream processes per batch, L
PV_s product volume per batch leaving step s, L
QC QCQA cost, £
RV_s resin volume required at chromatography step s, L
SC supervisors cost, £
T_s processing time per batch of step s, min
TC tax cost, £
TCV_s total column volume at chromatography step s, L
UC utilities cost, £
VFR_s volumetric flow rate at chromatography step s, L/min

Binary variables

U_{sr} 1 if resin r is selected for chromatography step s; 0 otherwise
W_{sij} 1 if there are j columns of size i at chromatography step s; 0 otherwise
X_{si} 1 if columns of size i is selected for chromatography step s; 0 otherwise
Y_{sk} 1 if there are k cycles during chromatography step s; 0 otherwise
Z_n 1 if the nth digit of the binary representation of variable BN is equal to 1; 0 otherwise

Integer variables

BN number of completed batches
CN_{si} number of column size i at chromatography step s
CYN_s number of cycles at chromatography step s

Auxiliary variables

$\overline{UM}_{s-1,r} \equiv U_{sr} \cdot M_{s-1}$
$\overline{UV}_{s-1,r} \equiv U_{sr} \cdot PV_{s-1}$
$\overline{UWT}_{srij} \equiv U_{sr} \cdot W_{sij} \cdot PLT_s$
$\overline{UX}_{sri} \equiv U_{sr} \cdot X_{si}$
$\overline{UXY}_{srik} \equiv U_{sr} \cdot X_{si} \cdot Y_{sk}$
$\overline{UYV}_{srk} \equiv U_{sr} \cdot Y_{sk} \cdot TCV_s$
$\overline{YV}_{sk} \equiv Y_{sk} \cdot TCV_s$
$\overline{ZT}_n \equiv Z_n \cdot BT$
$\overline{ZV}_n \equiv Z_n \cdot BBV$
$\overline{ZUYV}_{srkn} \equiv Z_n \cdot U_{sr} \cdot Y_{sk} \cdot TCV_s$

Pinto (2004) developed two mixed integer linear programming (MILP) models for the optimal synthesis of multistep purification processes for a specified purity and recovery. Simeonidis et al. (2005) and Natali et al. (2009), respectively, developed mixed integer nonlinear programming (MINLP) and MILP models to predict the behaviour and design of peptide-fusion tags that alter the purification of proteins. Vasquez-Alvarez and Pinto (2003) extended their previous work by developing an MILP model for the synthesis of protein purification processes that incorporates product losses assuming discrete percentage levels product collection. Then, Polykarpou et al. (2011) incorporated both starting and finishing cut-points for each chromatographic step as optimisation decision variables. Later, this work was extended by developing efficient MILP models with the discretisation approximation (Polykarpou et al., 2012a) and piecewise linearisation approximation (Polykarpou et al., 2012b) to overcome the computational difficulty of MINLP models. These models use the number of chromatography steps, purity and yield as performance metrics, but do not account for the cost of the process. For the optimisation of the chromatography steps in the mAb production, Liu et al. (2013a) developed an MILP model for the optimal column sizing decisions. Then this work was extended to MINLP models for both facility design and facility fit (Liu et al., 2013b).

However, to the best of our knowledge, no work has been done to integrate the optimisation of decisions on chromatography step sequences and columns sizes using mathematical programming techniques. The work in this paper aims to incorporate the chromatography step sequencing decisions into our previous model (Liu et al., 2013a,b), and to develop mathematical programming models and approaches for the optimisation of the downstream mAb purification process.

The rest of this paper is organised as follows: Section 2 describes the problem considered in this work. The mathematical formulation of the proposed MINLP model is presented in Section 3. Then, the model is reformulated as a mixed integer linear fractional programming (MILFP) model and a literature algorithm is adapted for solution, as detailed in Section 4. In Section 5, an industrially-relevant case study of mAb production is presented, followed by the computational results and discussion in Section 6. Finally, the concluding remarks are drawn in Section 7.

2. Problem statement

This work focused on determining the optimal column sizing strategies and chromatography steps sequencing strategies for a mAb purification process. A typical mAb platform process was used in this study, as shown in Fig. 1. In the upstream processing (USP) mammalian cells expressing the mAb of interest are cultured in bioreactors. In the downstream processing (DSP), the mAb is recovered, purified and cleared from viruses by a variety of operations, including a number of chromatography steps. A typical three-step chromatography sequence was considered, including three packed-bed chromatography steps, for product capture, intermediate purification and polishing, respectively (Fig. 1).

For the above three chromatography steps, it is important to determine the resins to be used. There are a number of candidate chromatography resins of different types for selection, with different suitability, yields, dynamic binding capacities, prices, linear velocities and buffer volume usages. In order to ensure that orthogonal separation mechanisms were used in the purification process, it was assumed that each resin type can only be used once, i.e. no two resins of the same type could be used in one sequence. For example, considering two resins of the same type, A and B, suitable for both capture and intermediate purification steps, if resin A has been used in the capture step, then resin B cannot be used at the intermediate purification step.

Furthermore, the chromatography column sizing strategy at each chromatography step was determined, including the bed height, diameter, number of cycles and number of columns to run in parallel at each step. The selected strategy has a direct impact on key metrics related to cost, time and annual product output. For both the bed heights and diameters, some typical ranges with discrete candidate values were available for selection, while the number of cycles and number of columns per step were limited by specific predetermined upper bounds.

In this problem, several scenarios with multiple USP trains feeding a single DSP train were investigated. The cost of goods per gram (COG/g) was considered as the criterion for this problem, equalling the cost of goods (COG), comprising of both capital-related costs and operating costs, divided by the product output.

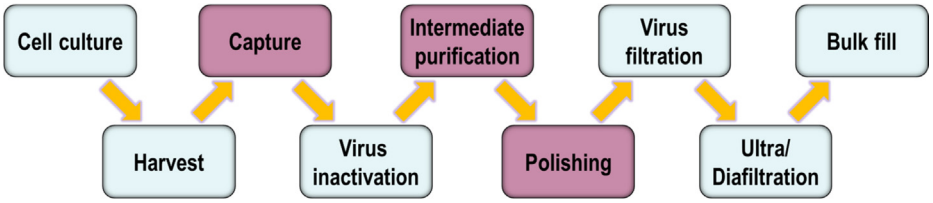

Fig. 1. A typical mAb manufacturing process.

This is a standard approach (Farid et al., 2005), which allows the incorporation of multiple process features into a single metric. COG/g, as an objective function, prefers the solutions with higher production output values.

Overall, the optimisation problem considered in this work is described as follows:

Given are:

- process sequence of a mAb product;
- number of USP trains, bioreactor volume and product titre;
- candidate chromatography resins per step, and their key parameters (e.g. yield, linear velocity, buffer usage, dynamic binding capacity);
- key parameters of non-chromatography steps (e.g. yield, time, buffer usage);
- cost data (e.g., reference equipment costs, labour wage, resin, buffer and media prices); and
- candidate column diameters and heights, and the maximum number of cycles and columns.

Determine:

- chromatography step sequencing strategy (i.e., resin for each step);
- chromatography column sizing strategy (i.e. column diameter and height, the number of cycles, number of columns per step) of the resin used at each chromatography step;
- number of total completed batches;
- annual total processing time;
- annual total production output; and
- annual total cost.

So as to:

minimise COG/g, equalling the total annual cost divided by the total production output.

3. MINLP model

In this section, an MINLP model, extended from our previous work (Liu et al., 2013b), is developed for the integrated chromatography sequencing and column sizing optimisation problem described above. In the mathematical formulation of the developed model, the parameters are represented by lower case letters, while the variables are represented by upper case letters.

3.1. Chromatography Sequence

For the three packed-bed chromatography steps (capture, intermediate purification, polishing) considered, among all candidate suitable resins, only one resin can be used at each step.

$$\sum_{r \in R_s} U_{sr} = 1, \quad \forall_S \in CS \tag{1}$$

Due to the assumption given in the previous section that no two resins of the same type should be used in any sequence, at most one resin of each resin type can be used.

$$\sum_{s \in CS} \sum_{r \in R_s \cap R_t} U_{sr} \leq 1, \quad \forall t \tag{2}$$

3.2. Protein mass

For each batch, the initial product protein mass entering into the downstream processes depends on the titre of the product, *titre*, and the working volume of production bioreactor, which is equal to the working volume ratio, α, multiplied by the single bioreactor volume, *brv*.

$$M_0 = titre \cdot \alpha \cdot brv \tag{3}$$

The product protein mass remaining after step s is equal to the product mass entering into this step, i.e., the mass remaining after the previous step $s-1$, multiplied by the corresponding yield of step s.

$$M_s = ncy_s \cdot M_{s-1}, \quad \forall_S \notin CS \tag{4}$$

$$M_s = \sum_{r \in R_s} (cy_{sr} \cdot U_{sr}) \cdot M_{s-1}, \quad \forall_S \in CS \tag{5}$$

where Eq. (4) is for the non-chromatography steps, while Eq. (5) is for chromatography steps.

The annual product output is the amount of product produced per year by the facility, determined by the product mass remaining after the bulk fill step per batch, M_s, multiplied by the number of completed batches, *BN*, and by the batch success rate, σ.

$$AP = \sigma \cdot BN \cdot M_s, \quad \forall s = bf \tag{6}$$

The number of completed batches, *BN*, is limited by an upper bound given by Eq. (7):

$$BN \leq maxbn \tag{7}$$

where the maximum number of batches $maxbn = brn \cdot \lfloor (aot - st)/brt \rfloor$, in which *aot* is the annual operating time; *st* is the seed train bioreaction time; *brt* is the bioreaction time; and *brn* is the number of bioreactors, i.e., the number of USP trains.

3.3. Resin volume

The total column volume of chromatography step s is given by the number of columns multiplied by the corresponding single column volume.

$$TCV_s = \sum_i cv_{si} \cdot CN_{si}, \quad \forall s \in CS \tag{8}$$

where cv_{si} is the volume (L) of the candidate column size i at chromatography step s, determined by specific diameter dm_{si} and height h_{si} (cm). Thus, if a column size i is selected, both corresponding

diameter and bed height will be known. Here, it is assumed that only one column size can be selected at each chromatography step:

$$\sum_i X_{si} = 1, \quad \forall s \in CS \tag{9}$$

$$CN_{si} \le maxcn_s \cdot X_{si} \quad \forall s \in CS, i \tag{10}$$

where X_{si} is a binary variable to indicate whether column size i is selected at chromatography step s, and $maxcn_s$ is the maximum number of columns at chromatography step s.

The total amount of resin available at each chromatography step must be sufficient to process all protein mass entering into that step. Thus, the number of cycles multiplied by the total column volume should be no less than the required resin volume.

$$CYN_s \cdot TCV_s \ge RV_s, \quad \forall s \in CS \tag{11}$$

where the amount of resin required per batch at chromatography step s, RV_s, is determined by the mass of product to be processed, M_{s-1}, the dynamic binding capacity of the selected resin, dbc_r, and the resin utilisation factor, μ:

$$RV_s = \frac{M_{s-1}}{\mu \cdot \sum_{r \in R_s}(dbc_r \cdot U_{sr})}, \quad \forall s \in CS \tag{12}$$

Also, the number of cycles at each chromatography step cannot exceed an upper bound, $maxcyn_s$:

$$CYN_s \le maxcyn_s, \quad \forall s \in CS \tag{13}$$

3.4. Flow rate

Volumetric flow rate (L/min) at each chromatography step is equal to the velocity for the selected resin, vel_r, multiplied by the selected column diameter, dm_{si}. In order to keep units consistency, conversion factors have been applied:

$$VFR_s = \frac{1}{1000} \cdot \frac{1}{60} \cdot \sum_{r \in R_s}(vel_r \cdot U_{sr}) \cdot \sum_i \pi \cdot \left(\frac{dm_{si}}{2}\right)^2 \cdot X_{si}, \quad \forall s \in CS \tag{14}$$

where vel_r is the linear flow velocity in cm/h given for resin r.

3.5. Product and buffer volume

The initial product volume entering downstream processes for each batch, PV_0, equals the working volume of one bioreactor, i.e., the bioreactor volume multiplied by working volume ratio.

$$PV_0 = \alpha \cdot brv \tag{15}$$

The product volume remaining after harvest step, PV_s for $s = h$, is equal to the initial product volume plus the flush volume, where the flush volume equals the flush volume ratio, fvr_s, multiplied by the initial product volume.

$$PV_s = (fvr_s + 1) \cdot PV_0, \quad \forall s = h \tag{16}$$

While the required buffer material volume used in the harvest step, BV_s for $s = h$, is equal to the flush volume:

$$BV_s = fvr_s \cdot PV_0, \quad \forall s = h \tag{17}$$

The product volume remaining after each chromatography step, PV_s for $s \in CS$, is equal to the eluate volume at this step if the resin is in the bind-elute mode (BER), or the product volume remaining after the previous step if the resin is in the flow-through mode (FTR). Here, the eluate volume is proportional to the product of the number of cycles and the total column volume:

$$PV_s = \sum_{r \in R_s \cap BRE}(ecv_r \cdot U_{sr}) \cdot CYN_s \cdot TCV_s + \sum_{r \in R_s \cap FTR}(U_{sr}) \cdot PV_{s-1}, \quad \forall s \in CS \tag{18}$$

where ecv_r is the eluate volume ratio of resin r. Moreover, the required buffer material volume, BV_s, is also proportional to the product of the number of cycles and the total column volume:

$$BV_s = \sum_{r \in R_s}(bcv_r \cdot U_{sr}) \cdot CYN_s \cdot TCV_s, \quad \forall s \in CS \tag{19}$$

where bcv_r is the buffer volume ratio of resin r.

The product volume remaining after virus inactivation step, PV_s for $s = vi$, is equal to the product volume entering into that step plus the used neutralisation volume. Here, the neutralisation volume equals the neutralisation volume ratio, nvr_s, multiplied by the product volume entering into that step.

$$PV_s = (nvr_s + 1) \cdot PV_{s-1}, \quad \forall s = vi \tag{20}$$

The required buffer volume used in virus inactivation step, BV_s for $s = vi$, is equal to the neutralisation volume:

$$BV_s = nvr_s \cdot PV_{s-1}, \quad \forall s = vi \tag{21}$$

Similar to the harvest step, the product volume remaining after virus filtration step, PV_s for $s = vf$, is the product volume entering into this step plus the flush volume used:

$$PV_s = (fvr_s + 1) \cdot PV_{s-1}, \quad \forall s = vf \tag{22}$$

Meanwhile, the required buffer volume used in virus filtration step, BV_s for $s = vf$, is equal to the flush volume:

$$BV_s = fvr_s \cdot PV_{s-1}, \quad \forall s = vf \tag{23}$$

The product volume remaining after UF/DF step, PV_s for $s = ufdf$, is equal to the remaining product mass divided by filling concentration.

$$PV_s = \frac{M_s}{fconc}, \quad \forall s = ufdf \tag{24}$$

The buffer volume required at UF/DF step, BV_s for $s = ufdf$, is equal to the product volume remaining after that step, multiplied by diafiltration volume ratio, dvr_s.

$$BV_s = dvr_s \cdot PV_s, \quad \forall s = ufdf \tag{25}$$

The total required buffer volume, BBV, for each batch will simply be equal to the summation of buffer volumes, BV_s, used at different steps:

$$BBV = \sum_s BV_s \tag{26}$$

Thus, the annual total buffer volume, ABV, equals the number of completed batch multiplied by the buffer volume per batch:

$$ABV = BN \cdot BBV \tag{27}$$

3.6. Processing time

In each chromatography step, the total processing time per batch (min) is comprised of processing time for both adding buffer and loading product.

$$T_s = PLT_s + BAT_s, \quad \forall s \in CS \tag{28}$$

The processing time for loading product, PLT_s, is given by the total product volume entering into the chromatography step divided by the number of columns, and by the volumetric flow rate.

$$PLT_s = \frac{PV_{s-1}}{VFR_s \cdot \sum_i CN_{si}}, \quad \forall s \in CS \tag{29}$$

It is worth noting that as for each chromatography step s, only one column size i is selected, CN_{si} is positive for only one size i based on Eqs. (9) and (10). Thus, $\sum_i CN_{si}$ in Eq. (29) is the number of the selected columns at the step.

The processing time for adding buffer, PLT_s, is given by the total volume of buffer added in each cycle divided by the volumetric flow rate, then multiplied by the number of cycles.

$$BAT_s = \frac{CYN_s \cdot \sum_{r \in R_s}(bcv_r \cdot U_{sr}) \cdot \sum_i (cv_{si} \cdot X_{si})}{VFR_s}, \quad \forall s \in CS \tag{30}$$

As the non-chromatography steps are not the main concern in this problem, it is assumed that their operating times are constant. The processing time per batch, BT, is the summation of processing times of all steps. In addition, the shift pattern of DSP operators is considered to convert total DSP time into days.

$$BT = \frac{\sum_s T_s}{60 \cdot sfd \cdot sfn} \tag{31}$$

where sfd is the duration of each shift (h), and sfn is the number of shifts per day.

The annual DSP time, AT, is equal to the number of completed batch multiplied by the processing time per batch:

$$AT = BN \cdot BT \tag{32}$$

Also, AT cannot exceed the annual available time for DSP:

$$AT \leq aot - st - brt \tag{33}$$

3.7. Costs

The total cost is categorised into two types: direct cost and indirect cost, which are presented in the following two subsections, respectively.

3.7.1. Direct costs

We firstly consider the direct costs based on the resource utilisation, including labour, materials and utilities.

Labour cost

The labour cost, LC, includes the direct labour cost, DLC, supervisors cost, SC, quality control and quality assurance (QCQA) cost, QC, and management cost, MC.

$$LC = DLC + SC + QC + MC \tag{34}$$

The direct labour cost includes the costs for both USP and DSP. At the USP stage, its direct labour cost is the calculated by the number of operators per bioreactor, uon, hourly wage, w, bioreaction time, brt, number of completed batches, BN, while at the DSP stage, its direct labour cost is the calculated by the number of operators per shift, don, hourly wage, w, duration per shift, sfd, number of shifts per day, sfn, and total processing time, AT:

$$DLC = 24 \cdot uon \cdot w \cdot brt \cdot BN + don \cdot w \cdot sfd \cdot sfn \cdot AT \tag{35}$$

All other three cost terms are assumed to be proportional to the direct labour cost as follows:

$$SC = s\lambda \cdot DLC \tag{36}$$

$$QC = q\lambda \cdot DLC \tag{37}$$

$$MC = m\lambda \cdot DLC \tag{38}$$

Materials cost

The materials cost was split into chemical reagents, consumables and miscellaneous materials. The key materials related to chromatography operations are the resins and buffers, which typically dominate the purification materials cost (e.g. >60% at commercial scales) (Simaria et al., 2012; Pollock et al., 2013a) and so these were considered explicitly under chemical reagents and consumables. The chemical reagents cost also captured the bioreactor media cost. The remaining materials were captured under miscellaneous materials.

Here, the chemical reagents cost, CRC, is assumed to include the cost for buffer, BC, and bioreactor media, MEC:

$$CRC = BC + MEC \tag{39}$$

The buffer cost, BC, is the total buffer required multiplied by the buffer price, bpc.

$$BC = bpc \cdot ABV \tag{40}$$

The bioreactor media cost, MEC, is the total media required for all batches multiplied by the media price, $mepc$.

$$MEC = \theta \cdot mepc \cdot \alpha \cdot brv \cdot BN \tag{41}$$

where θ is the media overfill allowance.

The key consumables cost, CC, in this study is the resin cost that is defined as the total used resin volume multiplied by the resin price, rpc_r. The total resin volume used is equal to the total column volume, multiplied by the number of cycles used (batch number multiplied by cycle number per batch) divided by the resin's lifetime, l_r.

$$CC = \sum_{r \in R_s} \sum_{s \in CS} \frac{of \cdot rpc_r \cdot BN \cdot CYN_s \cdot TCV_s}{l_r} \cdot U_{sr} \tag{42}$$

where of is the overpacking factor.

The miscellaneous material cost, MIC, is proportional to the summation of the chemical reagents cost and consumables cost.

$$MIC = mi\lambda \cdot (CRC + CC) \tag{43}$$

Utilities cost

Utilities cost, UC, includes the costs of steam, compressed air, electricity, cooling water and water for injection (WFI). To simplify the calculation of the utilities cost, according to Simaria et al. (2012), the utilities considered in this work are separated into three parts: the first part is proportional to the total bioreactor volume, $brn \cdot brv$; the second part is related to the total batch sizes, i.e. number of completed batches, BN, multiplied by single bioreactor volume, brv; while the third part involves the annual buffer volume, ABV. Thus, the utilities cost can be expressed as the summation of three terms as follows:

$$UC = a \cdot brn \cdot brv + b \cdot brv \cdot BN + c \cdot ABV \tag{44}$$

where a, b, c are given.

3.7.2. Indirect costs

Next, we consider the indirect costs, including capital cost and other indirect costs, both of which are dependent on the facilities.

Capital Cost

The annualised capital cost, CAC, is calculated by the fixed capital investment, FCI, and the capital recovery factor:

$$CAC = FCI \cdot \frac{r \cdot (1+r)^{el}}{(1+r)^{el} - 1} \tag{45}$$

where r is the interest rate; el is the lifetime of the equipment in years; and FCI is the fixed capital investment, consisting of the capital investment for bioreactors, chromatography columns and other equipment in DSP. As the decision for other equipment is not the

main concern of this work, the capital investment for other equipment is assumed to be proportional to the bioreactor cost. Here, the fixed capital investment is approximated by Lang factor, *lang* (Farid et al., 2005). Also, the general equipment factor, *gef*, is used to take into account the costs of support equipment, spares, utilities equipment and vessels:

$$FCI = lang \cdot (1 + gef) \cdot (brn \cdot brc + \sum_{s \in CS} \sum_{i} CC_{si} \cdot CN_{si} + oe\lambda \cdot brc \cdot brn) \quad (46)$$

For different sizes of chromatography columns and bioreactors, the costs are calculated using the values of reference equipment sizes and costs to scale up the equipment cost, i.e., $cc_{si} = refcc \cdot (dm^{col}/refdm)^{cf}$, $brc = refbrc \cdot (br\nu/refbr\nu)^{brf}$, where *refcc* is the cost of a single chromatography column with a diameter of *refdm*, and *refbrc* is the cost of a single bioreactor with a volume of *refbrv*. Both reference costs are used to scale up the costs of chromatography columns and bioreactors with different sizes, using *cf* and *brf* as the scale-up factors of column cost and bioreactor cost, respectively.

Other indirect costs

Other indirect costs considered here include the annual maintenance cost, *MAC*, insurance cost, *IC*, and local tax costs, *TC*, which are all assumed to be dependent on the fixed capital investment, respectively. Thus, these costs can be expressed by following equations:

$$MAC = ma\lambda \cdot FCI \quad (47)$$

$$IC = i\lambda \cdot FCI \quad (48)$$

$$TC = t\lambda \cdot FCI \quad (49)$$

The general utilities cost, *GUC*, is equal to the unit general utilities cost, *gu*, multiplied by the total bioreactor volume, which is equal to the number of bioreactors multiplied by the single bioreactor volume:

$$GUC = gu \cdot brn \cdot br\nu \quad (50)$$

Thus, we have the other indirect costs, *OIC*:

$$OIC = MAC + IC + TC + GUC \quad (51)$$

According to the above equations, the annual total cost, *AC*, is calculated:

$$AC = LC + CRC + CC + MIC + UC + CAC + OIC \quad (52)$$

3.8. Objective function

In the work, the objective is to minimise COG/g, which equals the annual total cost divided by the annual production output. Thus, we have the following objective function:

$$OBJ = \frac{AC}{AP} \quad (53)$$

Overall, the considered problem is formulated as an MINLP model with Eqs. (1)–(52) as constraints and Eq. (53) as the objective function, in which Eqs. (5), (6), (11), (12), (14), (18), (19), (27), (29), (30), (32), (42) and (53) are nonlinear equations.

4. MILFP model

In general, MINLP models are difficult to solve using standard commercial solvers so as to identify global optimal solutions or even feasible solutions. In this section, we reformulate the proposed MINLP model to overcome the computational difficulties. In the proposed MINLP model, all nonlinearities involve the products or fractions comprising of discrete variables. Only the nonlinearity of the objective function is the fraction of two continuous variables. Thus, the proposed MINLP model can be reformulated as an MILFP model, in which all the constraints are linear, and the objective function is a fraction of two linear functions. Thus, in this section, we present the MILFP model based on the proposed MINLP model using exact linearisation techniques (Floudas, 1995; Harjunkoski et al., 1998; Sherali and Adams, 1999).

4.1. Linearisation methods

To reformulate the proposed MINLP model as an MILFP model, we need to linearise all the nonlinear constraints. If an integer variable is involved in the nonlinearity, it should be expressed by a number of auxiliary binary variables at first. An arbitrary integer variable *I* can be expressed by its decimal representation as follows:

$$I = \sum_{l} l \cdot E_l \quad (54)$$

$$\sum_{l} E_l = 1 \quad (55)$$

where E_l is a binary variable to indicate whether the value of variable *I* is equal to *l*. Alternatively, the integer variable *I* can be expressed by its binary representation:

$$I = \sum_{m} 2^{m-1} \cdot F_m \quad (56)$$

where F_m is a binary variable to indicate whether the *m*th digit of the binary representation of the variable is equal to 1.

Thus, an integer variable is expressed by a set of binary variables in both the above two ways. Then, the bilinear term in the form of the product of an integer variable and a nonnegative continuous/integer variable is transformed into a series of multiplications of a binary variable and a nonnegative continuous/integer variable. Here, in order to linearise a bilinear term with a binary variable and a nonnegative continuous variable, we introduce an auxiliary nonnegative variable, a big-*M* parameter and additional auxiliary constraints. If the decimal representation is used to express an integer variable (Eqs. (54) and (55)), the following two constraints should be added to express the product of binary variable E_l and nonnegative continuous variable, *C*:

$$\overline{CE}_l \le M \cdot E_l, \quad \forall l \quad (57)$$

$$\sum_{l} \overline{CE}_l = C \quad (58)$$

where $\overline{CE}_l$ is an auxiliary variable, $\overline{CE}_l \equiv C \cdot E_l$, and *M* is a large number, an upper bound of variable *C*.

If the binary representation is used to express an integer variable (Eq. (56)), the following three constraints should be added to express the product of binary variable F_m and nonnegative continuous variable, *C* (Glover, 1975):

$$\overline{CF}_m \le M \cdot F_m, \quad \forall m \quad (59)$$

$$\overline{CF}_m \le C, \quad \forall m \quad (60)$$

$$\overline{CF}_m \ge C - M(1 - F_m), \quad \forall m \quad (61)$$

where, $\overline{CF}_m$ is an auxiliary variable, $\overline{CF}_m \equiv C \cdot F_m$.

Comparing the above two representations used to linearise a bilinear term involving an integer variable, if the maximum possible value of integer variable *I* is δ, then, $|k| = \delta$, and $|l| = (\log_2 \delta)$. Thus, the binary representation of the integer variable requires fewer binary variables. For the number of auxiliary constraints, the binary representation method requires $3 \times |l| = 3 \times (\log_2 \delta)$ constraints, while the decimal representation requires $|k| + 2 = \delta + 2$ constraints. Thus, when δ is large ($\delta > 10$), the binary representation is preferred, which saves both the number of constraints and

the number of variables. When δ is small ($\delta \leq 10$), the decimal representation is better to be used, which requires a smaller number of constraints. In the proposed model, the integer variables involved in the bilinear terms include CN_{si}, CYN_s and BN. As the maximum number of columns and cycles are small ($\leq$10), we use the decimal representation for CN_{si} and CYN_s, while BN is expressed using its binary representation, as its value can be very large (>100).

4.2. Model reformulation

According to the above discussion, firstly, we express CN_{si} and CYN_s by their decimal representations as follows:

$$CN_{si} = \sum_{j=1}^{j_s} j \cdot W_{sij}, \quad \forall s \in CS, i \tag{62}$$

$$\sum_{j=1}^{j_s} W_{sij} = X_{si}, \quad \forall s \in CS, i \tag{63}$$

$$CYN_s = \sum_{k=1}^{k_s} k \cdot Y_{sk}, \quad \forall s \in CS \tag{64}$$

$$\sum_{k=1}^{k_s} Y_{sk} = 1, \quad \forall s \in CS \tag{65}$$

where W_{sij} indicates whether there are j columns of size i at chromatography step s, while Y_{sk} indicates whether there are k cycles during chromatography step s. j_s and k_s are the maximum values of CN_{si} and CYN_s for each chromatography step, respectively, i.e., $j_s = maxcn_s$, and $k_s = maxcyn_s$.

Then, BN can be expressed by its binary representation:

$$BN = \sum_{n=1}^{q} 2^{n-1} \cdot Z_n \tag{66}$$

where Z_n indicates whether the nth digit of the binary representation of variable BN is equal to 1, and $q = (\log_2 maxbn)$.

Eq. (5) can be linearised using the following constraints by introducing an auxiliary variable, $\overline{UM}_{s-1,r} \equiv U_{sr} \cdot M_{s-1}$:

$$\overline{UM}_{s-1,r} \leq titre \cdot \alpha \cdot brv \cdot U_{sr}, \forall s \in CS, r \in R_s \tag{67}$$

$$\sum_{r \in R_s} \overline{UM}_{s-1,r} = M_{s-1,r}, \quad \forall s \in CS \tag{68}$$

$$M_s = \sum_{r \in R_s} (ncy_{sr} \cdot \overline{UM}_{s-1,r}), \quad \forall s \in CS \tag{69}$$

Based on Eq. (66), Eq. (6) is equivalent to the following equations:

$$AP = \sum_{n=1}^{q} \sigma \cdot 2^{n-1} \cdot \overline{ZM}_{sn}, \quad \forall s = bf \tag{70}$$

$$\overline{ZM}_{sn} \leq titre \cdot \alpha \cdot brv \cdot Z_n, \quad \forall s = bf, n = 1, \ldots, q \tag{71}$$

$$\overline{ZM}_{sn} \leq M_s, \quad \forall s = bf, n = 1, \ldots, q \tag{72}$$

$$\overline{ZM}_{sn} \geq M_s - titre \cdot \alpha \cdot brv \cdot (1 - Z_n), \quad \forall s = bf, n = 1, \ldots, q \tag{73}$$

where the auxiliary variable $\overline{ZM}_{sn} \equiv Z_n \cdot M_s$.

Based on Eq. (64), Eq. (11) is equivalent to the following equations:

$$\sum_{k=1}^{k_s} k \cdot \overline{YV}_{sk} \geq RV_s, \quad \forall s \in CS \tag{74}$$

$$\overline{YV}_{sk} \leq maxcn_s \cdot \max cv_s \cdot Y_{sk}, \quad \forall s \in CS, k = 1, \ldots, k_s \tag{75}$$

$$\sum_{k=1}^{k_s} \overline{YV}_{sk} = TCV_s, \quad \forall s \in CS \tag{76}$$

where $\overline{YV}_{sk} \equiv Y_{sk} \cdot TCV_s$, and $maxcv_s$ is the maximum column volume at chromatography step s.

Based on the equivalence of $\overline{UM}_{s-1,r}$ and $U_{sr} \cdot M_{s-1}$ defined by Eqs. (68) and (69), Eq. (12) can be rewritten as the following equation:

$$RV_s = \sum_{r \in R_s} \frac{\overline{UM}_{s-1,r}}{dbc_r \cdot \mu}, \quad \forall s \in CS \tag{77}$$

By introducing an auxiliary variable $\overline{UX}_{sri} \equiv U_{sr} \cdot X_{si}$, Eq. (14) can be replaced by the following equations:

$$VFR_s = \frac{1}{1000} \cdot \frac{1}{60} \cdot \sum_{r \in R_s} \sum_i vel_r \cdot \pi \cdot \left(\frac{dm_{si}}{2}\right)^2 \cdot \overline{UX}_{sri}, \quad \forall s \in CS \tag{78}$$

$$\sum_{r \in R_s} \overline{UX}_{sri} = X_{si}, \quad \forall s \in CS, i \tag{79}$$

$$\sum_i \overline{UX}_{sri} = U_{sr}, \quad \forall s \in CS, r \in R_s \tag{80}$$

We define auxiliary variables $\overline{UYV}_{srk} \equiv U_{sr} \cdot \overline{YV}_{sk} \equiv U_{sr} \cdot Y_{sk} \cdot TCV_s$, and $\overline{UV}_{s-1,r} \equiv U_{sr} \cdot PV_{s-1}$, by the following constraints:

$$\overline{UYV}_{srk} \leq maxcn_s \cdot maxcv_s \cdot U_{sr}, \quad \forall s \in CS, r \in R_s, k = 1, \ldots, k_s \tag{81}$$

$$\sum_{r \in R_s} \overline{UYV}_{srk} = \overline{YV}_{sk}, \quad \forall s \in CS, k = 1, \ldots k_s \tag{82}$$

$$\overline{UV}_{s-1,r} \leq maxpv_{s-1} \cdot U_{sr}, \quad \forall s \in CS, r \in R_s \tag{83}$$

$$\sum_{r \in R_s} \overline{UV}_{s-1,r} = PV_{s-1}, \quad \forall s \in CS \tag{84}$$

Then, based on Eq. (64), Eqs. (18) and (19) can be rewritten as:

$$PV_s = \sum_{r \in R_s \cap BER} \sum_{k=1}^{k_s} ecv_r \cdot k \cdot \overline{UYV}_{srk} + \sum_{r \in R_s \cap FTR} \cdot \overline{UV}_{s-1,r}, \quad \forall s \in CS \tag{85}$$

$$BV_s = \sum_{r \in R_s} \sum_{k=1}^{k_s} bcv_r \cdot k \cdot \overline{UYV}_{srt}, \quad \forall s \in CS \tag{86}$$

Replacing BN by Eq. (66), and introducing an auxiliary variable $\overline{ZV}_n = Z_n \cdot BBV$, Eq. (27) can be expressed as:

$$ABV = \sum_{n=1}^{q} 2^{n-1} \cdot \overline{ZV}_n \tag{87}$$

$$\overline{ZV}_n \leq maxbbv \cdot Z_n, \quad \forall n = 1, \ldots, q \tag{88}$$

$$\overline{ZV}_n \leq BBV, \quad \forall n = 1, \ldots, q \tag{89}$$

$$\overline{ZV}_n \geq BBV - maxbbv \cdot (1 - Z_n), \quad \forall n = 1, \ldots, q \tag{90}$$

According to a newly introduced auxiliary variable $\overline{UWT}_{srij} \equiv U_{sr} \cdot W_{sij} \cdot PLT_s$ and Eq. (62), we have the following constraints to replace Eq. (29):

$$\frac{1}{1000} \cdot \frac{1}{60} \cdot \sum_{r \in R_s} \sum_i \sum_{j=1}^{j_s} vel_r \cdot \pi \cdot \left(\frac{dm_{si}}{2}\right)^2 \cdot j \cdot \overline{UWT}_{srij} = PV_{s-1}, \quad \forall s \in CS \tag{91}$$

$$\overline{UWT}_{srij} \leq brt \cdot W_{sij}, \quad \forall s \in CS, r \in R_s, i, j = 1, \ldots, j_s \tag{92}$$

$$\overline{UWT}_{srij} \leq brt \cdot U_{sr}, \quad \forall s \in CS, r \in R_s, i, j = 1, \ldots, j_s \tag{93}$$

$$\sum_i \sum_{j=1}^{j_s} \sum_{r \in R_s} \overline{UWT}_{srij} = PLT_s, \quad \forall s \in CS \tag{94}$$

Here, we introduce another auxiliary variable $\overline{UXY}_{srik} \equiv \overline{UX}_{sri} \cdot Y_{sk} \equiv U_{sr} \cdot X_{si} \cdot Y_{sk}$, and the following constraints are equivalent to Eq. (30) with the existence of Eq. (64):

$$BAT_s = \sum_{r \in R_s} \sum_i \sum_{k=1}^{k_s} \frac{bcv_r \cdot cv_{si} \cdot k \cdot \overline{UXY}_{srik}}{(1/1000) \cdot (1/60) \cdot vel_s \cdot \pi (dm_{si}/2)^2}, \quad \forall s \in CS \tag{95}$$

$$\sum_{r \in R_s} \sum_i \overline{UXY}_{srik} = Y_{sk}, \quad \forall s \in CS, k = 1, \ldots, k_s \tag{96}$$

$$\sum_{k=1}^{k_s} \overline{UXY}_{srik} = \overline{UX}_{sri}, \quad \forall s \in CS, r \in R_s, i \tag{97}$$

Let $\overline{ZT}_n \equiv Z_n \cdot BT$, due to Eq. (66), Eq. (32) can be rewritten as:

$$AT = \sum_{n=1}^{q} 2^{n-1} \cdot \overline{ZT}_n \tag{98}$$

$$\overline{ZT}_n \le (aot - st - brt) \cdot Z_n, \quad \forall n = 1, \ldots, q \tag{99}$$

$$\overline{ZT}_n \le BT, \quad \forall n = 1, \ldots, q \tag{100}$$

$$\overline{ZT}_n \ge BT - (aot - st - brt) \cdot (1 - Z_n), \quad \forall n = 1, \ldots, q \tag{101}$$

Finally, we introduce auxiliary variables $\overline{ZUYV}_{srkn} \equiv Z_n \cdot \overline{UYV}_{srk} \equiv U_{sr} \cdot \overline{YV}_{sk} \equiv U_{sr} \cdot Y_{sk} \cdot TCV_s$. Thus, due to Eq. (66), Eq. (42) can be reformulated as:

$$CC = \sum_{s \in CS} \sum_{r \in R_s} \sum_{n=1}^{q} \sum_{k=1}^{k_s} \frac{of \cdot rpc_r \cdot 2^{n-1} \cdot k \cdot \overline{ZUYV}_{srkn}}{l_r} \tag{102}$$

$$\overline{ZUYV}_{srkn} \le \max tcv_s \cdot Z_n, \quad \forall s \in CS, r \in R_s, k = 1, \ldots, k_s, n = 1, \ldots, q \tag{103}$$

$$\overline{ZUYV}_{srkn} \le \overline{UYV}_{srk}, \quad \forall s \in CS, r \in R_s, k = 1, \ldots, k_s, n = 1, \ldots, q \tag{104}$$

$$\overline{ZUYV}_{srkn} \le \overline{UYV}_{sk} - maxtcv_s \cdot (1 - Z_n), \quad \forall s \in CS, r \in R_s, k = 1, \ldots, k_s, n = 1, \ldots, q \tag{105}$$

After the above reformulation, the original MINLP model is reformulated with Eqs. (1)–(4), (7)–(10), (13), (15)–(17), (20)–(26), (28), (31), (33)–(41), (43)–(52), (62)–(105) as constraints with Eq. (53) as the objective function. In this model, all constraints are linear and the objective function is a fraction of two continuous variables. Thus, it is an MILFP model.

4.3. Dinkelbach algorithm

Now, we apply the Dinkelbach algorithm to the above developed MILFP model. The Dinkelbach algorithm, developed by Dinkelbach (1967), is an application of the classical Newton method to solve convex nonlinear fractional programming (NFP) models by solving a sequence of nonlinear programming (NLP) models successively. Recently, several works have implemented the Dinkelbach algorithm to solve MILFP problems (Bradley and Arntzen, 1999; Pochet and Warichet, 2008; You et al., 2009; Billionnet, 2010; Espinoza et al., 2010; Trinh et al., 2012; Yue and You, 2013).

For an optimisation problem $\max\{\psi(x)/\phi(x) \mid x \in \Omega\}$, where $\psi(x)$ and $\phi(x)$ are two linear functions, and Ω is the variable x' feasible region, the Dinkelbach algorithm procedure is described as below:

Step 1: Initialise f;
Step 2: Solve the MILP problem $\max\{\psi(x) - f \cdot \phi(x) | x \in \Omega\}$, and the optimal solution is denoted as x^*;
Step 3: If $|\psi(x^*) - f \cdot \phi(x^*)| \le \varepsilon\}$, stop and x^* is the optimal solution; otherwise, update $= (\psi(x^*))/(\phi(x^*))$, and go to Step 2.

It has been proven that the Dinkelbach algorithm can find the global optimal solution of the MILFP model for the maximisation problem (You et al., 2009). Similarly, it can be applied to the minimisation optimisation model with the guarantee of the optimality.

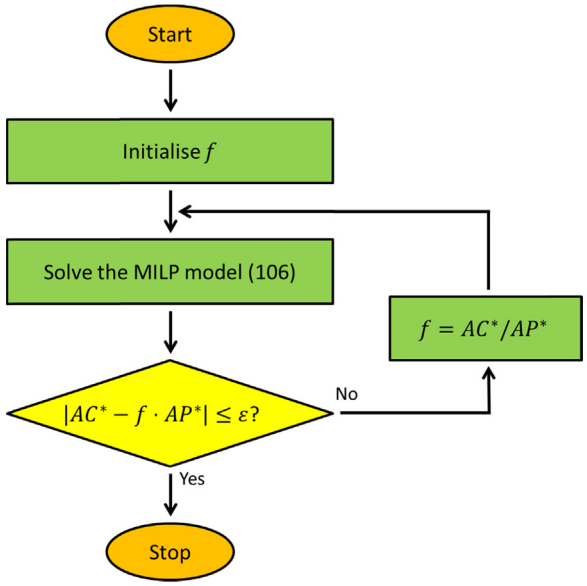

Fig. 2. Procedure of the Dinkelbach algorithm.

For the proposed MILFP model, the MILP model implemented in the Dinkelbach algorithm is given below:

$$\begin{aligned} \min \quad & AC - f \cdot AP \\ s.t. \quad & \text{Eqs.(1)–(4), (7)–(10), (13), (15)–(17), (20)–(26), (28), (31)–(41), (43)–(52), (62)–(105)} \end{aligned} \tag{106}$$

Thus, as shown in Fig. 2, the procedure of the Dinkelbach algorithm for the problem discussed in this work can be implemented as follows:

Step 1: Initialise f;
Step 2: Solve the MILP model (106), and the obtained values of AC and AP in the solutions are denoted as AC^* and AP^*, respectively;
Step 3: If $|AC^* - f \cdot AP^*| \le \varepsilon$, stop and the optimal COG/g $= AC^*/AP^*$; otherwise, update $f = AC^*/AP^*$, then go to Step 2.

It should be noted that in the Dinkelbach algorithm, although a higher optimality gap of each MILP may require more iterations to terminate the algorithm, the optimality gap does not affect the global optimality of the final solution obtained, if it is less than 100%. The optimal objective value in the last iteration is very close to zero. Therefore, during last iteration, the upper bound of the objective value is always positive, while its lower bound is always negative. The difference between the above two bounds should always be larger than 100%, until the objective value reaches below the tolerance, ε. Therefore, the optimality gap for the solution of each MILP is set below 100% and the global optimality is then guaranteed. More discussion on the optimality gaps will be presented in Section 6.

5. A Case Study

In this section, an industrially-relevant example is considered as a case study, based on the example of a biopharmaceutical company using a platform process for mAb purification described by Simaria et al. (2012). The proposed models and approaches are applied to this case study to find the optimal chromatography step sequencing and column sizing strategies for the design of a new purification suite in an existing facility.

In this example, we consider one product, whose process flowsheet is shown in Fig. 1. There are 6 candidate resins of 5 types, including Affinity Chromatography (AFF), Cation Exchange

Table 1
Characteristics of the resins in the case study.[a]

Resin	Type	Mode	Binding capacity (g/L)	Eluate volume (CV)	Buffer volume (CV)	Linear velocity (cm/h)	Resin lifetime (cycle)	Resin price (£/L)	Yield		
									Capture	Intermediate purification	Polishing
R1	AFF	Bind-elute	50	2.3	37	150	100	9200	91%	95%	–
R2	AFF	Bind-elute	30	2.3	37	300	100	6400	91%	95%	–
R3	CEX	Bind-elute	40	1.4	26	300	100	400	86%	92%	92%
R4	AEX	Flow-through	100	0	10	300	100	700	–	95%	95%
R5	MM	Flow-through	150	0	10	375	100	3500	–	90%	90%
R6	HIC	Bind-elute	27.5	1.4	26	175	100	2500	–	89%	89%

[a] Simaria et al. (2012).

Chromatography (CEX), Anion Exchange Chromatography (AEX), Mixed-Mode Chromatography (MM) and Hydrophobic Interaction Chromatography (HIC). The details of the resins' characteristics are given in Table 1, from which it can be seen that each candidate has different suitability and performs differently in terms of yield when used in different positions (capture, intermediate purification and polishing). The resin utilisation factor (μ) is 0.95, and the overpacking factor (*of*) is 1.1.

The candidate values of the chromatography equipment sizing decision variables are shown in Table 2. As there are 11 possible bed heights and 7 possible diameters, single column volume has 77 possible sizes. The number of cycles can be up to 10, while at most 4 parallel columns are allowed at each chromatography step.

The main details of the non-chromatographic DSP unit operations in the process flowsheet used in the model to perform mass balance and cost calculations are presented in Table 3. The annual operating time considered in this case study is 340 days. The batch success rate is 90%. In the USP stage, there are 3 operators for each bioreactor. In the DSP stage, there are 15 operators per shift, with a shift pattern of 8 h/day, 7 days/week. The wage for the operators is £20/h. The media and buffer prices are £32/L and £1/L, respectively. The overfill allowance of media, θ, is set at 1.2. More cost data in this case study is shown in Table 4.

Table 2
Candidate values of the column sizing decisions in the case study.[a]

Decision	Candidate value
Bed height (cm)	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
Diameter (cm)	50, 70, 100, 120, 160, 180, 200
Number of cycles	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Number of columns	1, 2, 3, 4

[a] Simaria et al. (2012)

Table 3
Characteristics of non-chromatographic unit operations in the case study.[a]

Unit operation parameter	Value	Unit operation parameter	Value
Cell culture		*Virus filtration*	
Bioreaction time (days)	15	Yield (%)	95
Seed train bioreaction time (days)	29	Flush volume ratio	0.3
Bioreactor working volume ratio (%)	75	Processing time (h)	4
Harvest		*Ultra/diafiltration*	
Yield (%)	95	Yield (%)	90
Flush volume ratio	0.1	Processing time (h)	4
Processing time (h)	4	Final concentration (g/L)	75
Virus inactivation		Diafiltration volume	7
Yield (%)	90	*Bulk fill*	
Neutralisation volume ratio	1.75	Yield (%)	98
Processing time (h)	1.5	Filling time (h)	6

[a] Simaria et al. (2012).

Table 4
Cost data in the case study.[a]

Parameter	Value	Parameter	Value
a (£/L)	14.145	$q\lambda$	1
b (£/L)	4.234	$m\lambda$	1
c (£/L)	0.071	$mi\lambda$	0.1
gef	0.7	$oe\lambda$	0.8
gu (£/L)	90	$ma\lambda$	0.05
lang	6	$i\lambda$	0.005
$s\lambda$	0.2	$t\lambda$	0.01

[a] Simaria et al. (2012).

To calculate the fixed capital investment on the bioreactors and chromatography columns, we use the reference equipment sizes and costs, and scale-up factors as given in Table 5.

The equipment lifetime is 10 years, and a 10% interest rate is used. Here, three scenarios of the case study will be investigated, with different configurations of USP and DSP trains, i.e., 1USP:1DSP, 2USP:1DSP, and 4USP:1DSP. It is assumed that the total bioreactor volume is given (25,000 L) and is kept the same for all scenarios. Thus, in the three investigated scenarios, the volumes of each bioreactor are 25,000 L, 12,500 L, and 6250 L, respectively.

The proposed optimisation model and approach were implemented in GAMS 24.0 (Brooke et al., 2012) on a 64-bit Windows 7 based machine with 3.20 GHz six-core Intel Xeon processor W3670 and 12.0 GB RAM. Three MINLP solvers, including BARON, COUENNE and SCIP, were used to solve the MINLP model. CPLEX was used as the MILP solver to implement the Dinkelbach algorithm. The CPU time for each run was limited to 1 h.

6. Results and discussion

Firstly, we solved the proposed MINLP model for all three scenarios. From Table 6, all three solvers can find feasible solutions, but cannot converge within the CPU time limit. Comparing the three solvers, SCIP solver always gives solutions that are equal or better than those from the other two solvers. Then, we investigated the computational performance of the Dinkelbach algorithm for the proposed MILFP model. As discussed above, the optimality gap of each MILP model should be set to less than 100%. Here, we used 6 different gaps for each scenario, including 0%, 5%, 10%, 20%, 50% and 80%. All these gaps find the same optimal solutions for all scenarios, whose optimal objective values are given in Table 6. Comparing the

Table 5
Reference sizes, costs, scale-up factors in the case study.[a]

	Reference size	Reference cost (£)	Scale-up factor
Bioreactor	Volume = 2000 L	612,000	0.6
Chromatography column	Diameter = 100 cm	170,000	0.8

[a] Simaria et al. (2012).

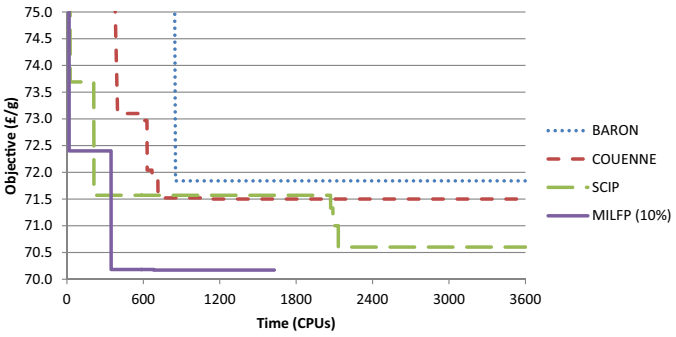

Fig. 3. Convergence curves of MINLP and MILFP models for the 1USP:1DSP scenario.

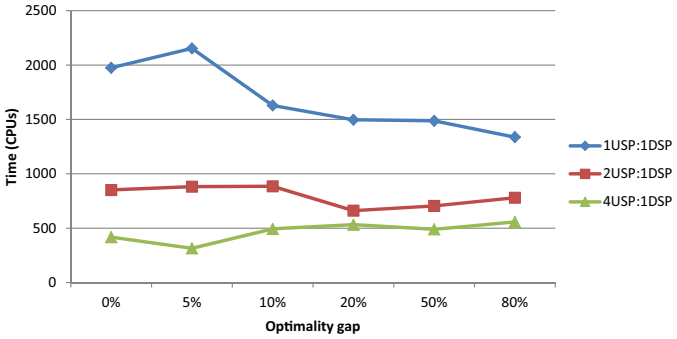

Fig. 4. CPU times of the Dinkelbach algorithm with different optimality gaps for different USP:DSP scenarios.

average CPU time for implementing the Dinkelbach algorithm with the aforementioned optimality gaps to the CPU times of the MINLP models, the Dinkelbach algorithm has a much better computational performance than the proposed MINLP model. In the last two scenarios, the CPU time is even improved by one order of magnitude. Furthermore, the 1USP:1DSP scenario was used as an example to investigate the convergence curves of the proposed MINLP models and MILFP model, as presented in Fig. 3, in which a 10% optimality gap is used for the MILP models in the Dinkelbach algorithm. The Dinkelbach algorithm takes two iterations and 18 CPU seconds (CPUs) to obtain a good feasible solution, whose objective value is 72.38, while the SCIP MINLP solver takes 26 CPUs to find a good feasible solution (objective value is 73.69), the fastest among all three MINLP solvers. Also, the Dinkelbach algorithm takes about 680 CPUs to obtain the global optimal solution, and for the MINLP model, BARON, COUENNE and SCIP solvers take about 850, 1070 and 2130 CPUs, respectively, to find their current best solutions. Thus, the proposed MILFP model is more efficient to find good feasible solutions and converge to the optimal solution.

As shown in Fig. 4, the impact of the optimality gap on the computational time varies by the scenario investigated. For the 1USP:1DSP scenario, a higher gap generally leads to a shorter CPU time, except when the gap = 5%, which leads to a longer CPU time than all other gaps. For the 2USP:1DSP scenario, small gaps ($\leq$10%) have similar performance and take longer times than larger gaps (>20%). For the 4USP:1DSP scenario, except for the gaps $\leq$5%, other higher gaps have similar performance with CPU times around 500 s. From Fig. 4, although a specific optimality gap with a dominant advantage cannot be found, we can still observe that there is consistently good performance in all scenarios.

The computational performance of the Dinkelbach algorithm was analysed. Considering optimality gaps of 0% and 80%, the value of f found by each iteration is shown in Fig. 5. When optimality gap = 0%, the Dinkelbach algorithm takes 3 or 4 iterations to complete, while more iterations (5 or 6) are required when the optimality gap = 80%. Investigating other gaps, we can find that when the gap $\leq$ 5%, the convergence requires 3 to 4 iterations, and when the gap $\geq$ 10%, 5 to 6 iterations are needed to terminate the algorithm. There is significant trade-off between the CPU time per iteration and the number of iterations. Although the CPU time per iteration decreases when the optimality gap is increased, the number of iterations increases as well. From Fig. 5, the value of f decreases significantly from iteration 1 to 2, and then it decreases slightly iteration by iteration until convergence.

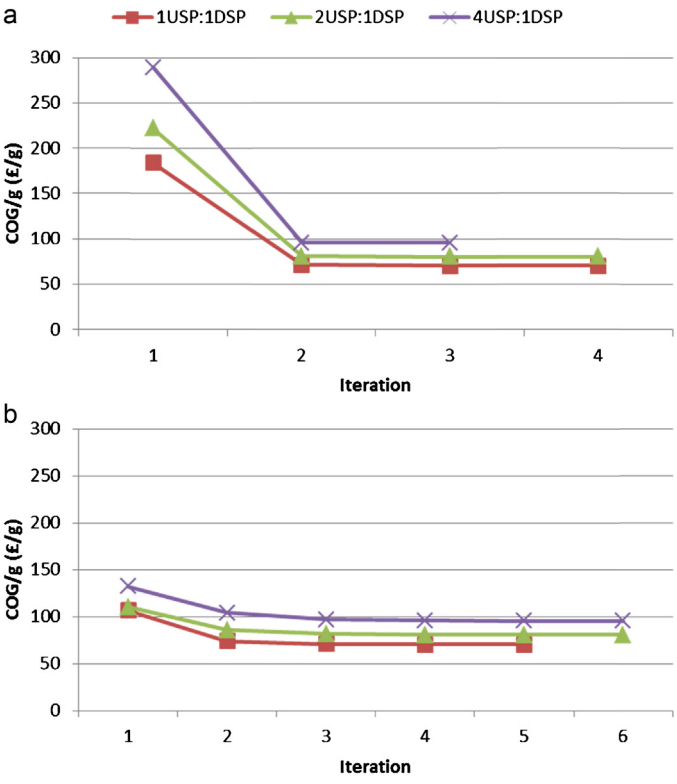

Fig. 5. Iterations of the Dinkelbach algorithm for different USP:DSP scenarios when optimality gap = (a) 0% and (b) 80%.

Table 6
Computational performances of MINLP and MILFP models for different USP:DSP scenarios.

		1USP:1DSP		2USP:1DSP		4USP:1DSP	
		OBJ	CPUs	OBJ	CPUs	OBJ	CPUs
MINLP	BARON	71.8	3600[a]	80.9	3600[a]	108.4	3600[a]
	COUENNE	71.5	3600[a]	82.9	3600[a]	102.5	3600[a]
	SCIP	70.6	3600[a]	80.9	3600[a]	95.5	3600[a]
MILFP (Dinkelbach algorithm)		70.2	1680[b]	80.5	794[b]	95.4	468[b]

[a] Terminated by CPU time limit.
[b] Average time with all optimality gaps investigated.

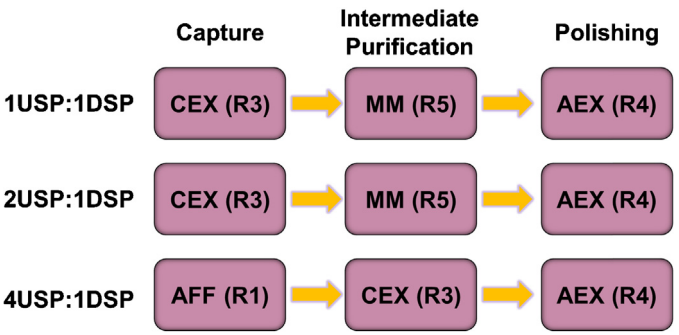

Fig. 6. Optimal chromatography sequencing decisions for different USP:DSP scenarios.

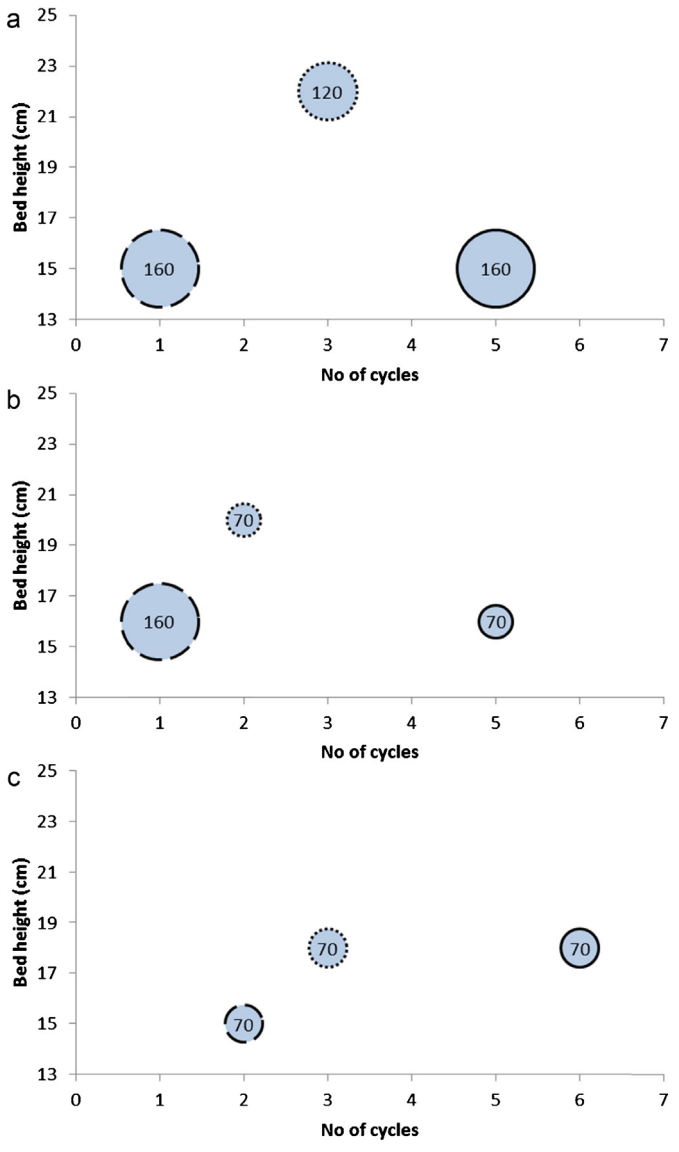

Fig. 7. Optimal column sizing decisions: bed height (*x*-axis), number of cycles (*y*-axis), diameter (proportional to the bubble size, indicated at the centres of the circles), at (a) capture step; (b) intermediate purification step; (c) polishing step, for 1USP:1DSP (solid circle lines), 2USP:1DSP (dotted circle lines) and 4USP:1DSP (dash circle lines) scenarios.

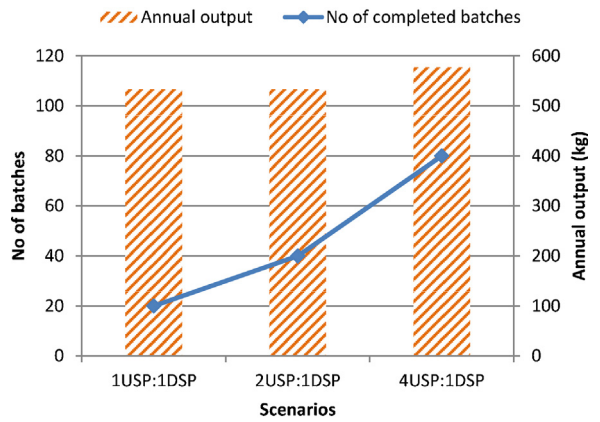

Fig. 8. Optimal annual output production and number of completed batches for different USP:DSP scenarios.

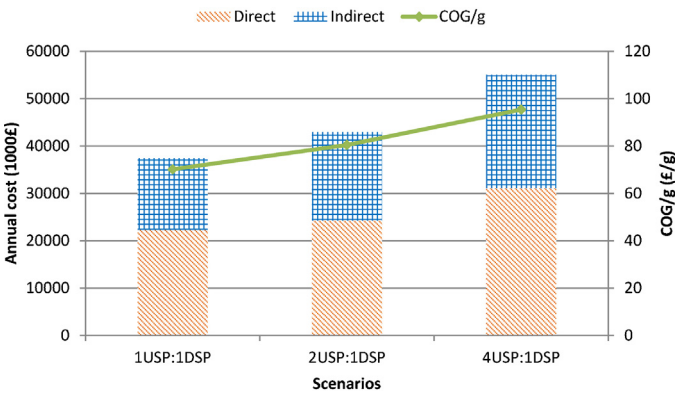

Fig. 9. Annual total cost and COG/g for different USP:DSP scenarios.

Next, we investigated the optimal solutions found by the MILFP model. Fig. 6 shows that for all scenarios, except for the scenario with 4USP:1DSP, the same resin sequence is selected, i.e., CEX (R3) for capture, MM (R5) for intermediate purification, and AEX (R4) for polishing. For the scenario with 4USP:1DSP, AFF (R1) is selected for capture, while CEX (R3) is for intermediate purification and AEX (R4) is for polishing. Thus, the USP:DSP configuration is important to be investigated, and has an effect on the chromatography step sequencing decisions.

The optimal chromatography column sizing decisions at each step are presented in Fig. 7, including the number of cycles per batch (x-axis), the column bed height (y-axis), the column diameter (proportional to the bubble size, number in cm at the circle centre). It should be noted that for all scenarios, only one column is used at each chromatography step. Because both 1USP:1DSP and 2USP:1DSP scenarios have the same chromatography sequence, it is fair to compare the column sizing decisions of the aforementioned two scenarios. As shown in Fig. 7, the number of cycles or column diameter decrease with the increasing number of bioreactors that lead to decreasing bioreactor size and batch size. Also, for the above two scenarios, it can be observed that the selected column diameters in the subsequent chromatography steps usually become smaller or not larger, because less protein mass enters into the subsequent chromatography steps, resulting from the yield at each manufacturing step.

Fig. 8 shows the optimal annual output and the number of batches. According to Fig. 8, when the number of bioreactors is increased, more batches are processed, increasing from 20 batches

with one bioreactor to 80 batches with four bioreactors. It should be noted that although there are more batches with more bioreactors, batch size becomes smaller. Also, as the first two scenarios select the same chromatography sequence, their annual productions are the same, while the 4USP:1DSP scenario has a higher annual production, due to the higher yields of the resins selected relative to the other two scenarios. An extra 8% of product is obtained in this scenario.

The breakdown of the total cost for all scenarios is shown in Fig. 9. With increasing number of bioreactors, there is a higher direct cost, as more batches are produced, and a higher indirect cost mainly due to the purchase of more equipment. Compared to the 1USP:1DSP scenario, the 2USP:1DSP scenario has a higher COG and the same production output, therefore its COG/g becomes higher, as shown in Fig. 9. The 4USP:1DSP scenario has both higher COG and production output than the other two scenarios. Because the increase in the COG is relatively higher than that in the production output, its COG/g is still the highest among all three scenarios.

7. Concluding remarks

In this paper, we addressed the integrated optimisation of chromatography sequencing and column sizing, extending from our previous work (Liu et al., 2013b). At first, an MINLP model is developed to minimise COG/g. Then, to overcome the computational expense of the proposed MINLP model, an MILFP model is developed as the reformulation of the MINLP model. The literature Dinkelbach algorithm is used as the solution approach of the MILFP model. An industrially-relevant example is investigated as the case study. The computational results prove the computational advantage of the proposed MILFP model and the Dinkelbach algorithm. The case study also demonstrated the applicability of the proposed models and approaches. Future work will investigate the incorporation of uncertainty issues and multi-criteria decision-making into the current work.

Acknowledgements

Funding from the UK Engineering & Physical Sciences Research Council (EPSRC) for the EPSRC Centre for Innovative Manufacturing in Emergent Macromolecular Therapies hosted by University College London is gratefully acknowledged. Financial support from the consortium of industrial and governmental users is also acknowledged.

References

Allmendinger R, Simaria AS, Farid SS. Efficient discovery of chromatography equipment sizing strategies for antibody purification processes using evolutionary computing. Lect Notes Comput Sci 2012;7492:468–77.
Allmendinger R, Simaria AS, Turner R, Farid SS. Closed-loop optimization of chromatography column sizing strategies in biopharmaceutical manufacture. J Chem Technol Biotechnol 2013., http://dx.doi.org/10.1002/jctb.4267 [in press].
Bentely J, Kawajiri Y. Prediction-correction method for optimization of simulated moving bed chromatography. AIChE J 2013;59(3):736–46.
Billionnet A. Optimal selection of forest patches using integer and fractional programming. Operat Res An Int J 2010;10:1–26.
Bradley JR, Arntzen BC. The simultaneous planning of production, capacity, and inventory in seasonal demand environments. Operat Res 1999;47:795–806.
Brooke A, Kendrick D, Meeraus A, Raman R. GAMS – a user's guide. Washington, DC: GAMS Development Corporation; 2012.
Chan S, Tichener-Hooker N, Gracewell DG, Sorensen E. A systematic approach for modeling chromatographic processes – application to protein purification. AIChE J 2008;54:965–77.
Chhatre S, Jones C, Francis R, O'Donovan K, Titchener-Hooker NJ, Newcombe A, et al. The integrated simulation and assessment of the impacts of process change in biotherapeutic antibody production. Biotechnol Prog 2006;23:1612–20.
Chhatre S, Thillaivinayagalingam P, Francis R, Titchener-Hooker NJ, Newcombe A, Keshavarz-Moore E. Decision-support software for the industrial-scale chromatographic purification of antibodies. Biotechnol Prog 2007;23:888–94.
Espinoza D, Fukasawa R, Goycoolea M. Lifting, tilting and fractional programming revisited. Operat Res Lett 2010;38:559–63.
Farid SS, Washbrook J, Titchener-Hooker NJ. Decision-support tool for assessing biomanufacturing strategies under uncertainty: stainless steel versus disposable equipment for clinical trial material preparation. Biotechnol Prog 2005;21:486–97.
Farid SS, Washbrook J, Titchener-Hooker NJ. Modelling biopharmaceutical manufacture: design and implementation of SimBiopharma. Comput Chem Eng 2007;31:1141–58.
Floudas CA. Nonlinear and mixed-integer opitmization: fundamentals and applications. Oxford: Oxford Univeristy Press; 1995.
GBI Research. Biopharmaceutical manufacturing in India, China and South Korea – regulatory framework, infrastructure support and discovery funding create an environment conducive to growth; 2012.
Dinkelbach W. On nonlinear fractional programming. Manage Sci 1967;13:492–8.
Glover F. Improved linear integer programming formulations of nonlinear integer problems. Manage Sci 1975;22:455–60.
Harjunkoski I, Westerlund T, Pörn R, Skrifvars H. Different transformations for soving non-convex trim-loss probems by MINLP. Eur J Operat Res 1998;3:594–603.
Joseph JR, Sinclair A, Tichener-Hooker NJ, Zhou Y. A framework for assessing the solutions in chromatographic process design and operation for large-scale manufacture. J Chem Technol Biotechnol 2006;81:1009–20.
Lakhdar K, Savery J, Papageorgiou LG, Farid SS. Multiobjective long-term planning of biopharmaceutical manufacturing facilities. Biotechnol Prog 2007;23:1383–93.
Lakhdar K, Zhou Y, Savery J, Titchener-Hooker NJ, Papageorgiou LG. Medium term planning of biopharmaceutical manufacture using mathematical programming. Biotechnol Prog 2005;21:1478–89.
Langer E. Downstream factors that will continue to constrain manufacturing through 2013. BioProcessing J 2009;8:22–6.
Levis AA, Papageorgiou LG. A hierarchical solution approach for multi-site capacity planning under uncertainty in the pharmaceutical industry. Comput Chem Eng 2004;28:707–25.
Lim AC, Washbrook J, Titchener-Hooker NJ, Farid SS. A computer-aided approach to compare the production economics of fed-batch and perfusion culture under uncertainty. Biotechnol Bioeng 2006;93:687–97.
Liu S, Simaria AS, Farid SS, Papageorgiou LG. Mixed integer optimisation of antibody purification processes. In: Kraslawski A, Turunen I, editors. Proceedings of the 23rd European Symposium on Computers Aided Process Engineering, Computer Aided Chemical Engineering, vol. 32. Amsterdam: Elsevier; 2013a. p. 157–62.
Liu S, Simaria AS, Farid SS, Papageorgiou LG. Designing cost-effective biopharmaceutical facilities using mixed-integer optimization. Biotechnol Prog 2013b;29:1472–83.
Low D, O'Leary R, Pujar NS. Future of antibody purification. J Chromatogr B 2007;848:48–63.
Natali JM, Pinto JM, Papageorgiou LG. Efficient MILP formulations for the simultaneous optimal peptide tag design and downstream processing synthesis. AIChE J 2009;55:2303–17.
Papageorgiou LG, Rotstein GE, Shah N. Strategic supply chain optimization for the pharmaceutical industries. Ind Eng Chem Res 2001;40:275–86.
Pochet Y, Warichet F. A tighter continuous time formulation for the cyclic scheduling of a mixed plant. Comput Chem Eng 2008;32:2723–44.
Pollock J, Ho SV, Farid SS. Fed-batch and perfusion culture processes: operational, economic and environmental feasibility under uncertainty. Biotechnol Bioeng 2013a;110:206–19.
Pollock J, Bolton G, Coffman J, Ho SV, Bracewell DG, Farid SS. Optimising the design and operation of semi-continuous affinity chromatography for clinical and commercial manufacture. J Chromatogr A 2013b;1284:17–27.
Polykarpou EM, Dalby PA, Papageorgiou LG. Optimal synthesis of chromatographic trains for downstream protein processing. Biotechnol Prog 2011;27:1653–60.
Polykarpou EM, Dalby PA, Papageorgiou LG. A novel efficient optimisation system for purification process synthesis. Biochem Eng J 2012a;67:186–93.
Polykarpou EM, Dalby PA, Papageorgiou LG. An MILP formulation for the synthesis of protein purification processes. Chem Eng Res Des 2012b;90:1262–70.
Pujar NS, Low D, O'Leary R. Antibody purification: drivers of change. In: Gottschalk U, editor. Process scale purification of antibodies. New Jersey: John Wiley & Sons, Inc.; 2009. p. 407–26.
Sherali HD, Adams WP. A reformulation-linearization technique for solving discrete and continuous nonconvex problems. Dordrecht: Kluwer Academic Publishers; 1999.
Simaria AS, Turner R, Farid SS. A multi-level meta-heuristic algorithm for the optimisation of antibody purification processes. Biochem Eng J 2012;69:144–54.
Simeonidis E, Pinto JM, Lienqueo ME, Tsoka S, Papageorgiou LG. MINLP models for the synthesis of optimal peptide tags and downstream protein processing. Biotechnol Prog 2005;21:875–84.
Stonier A, Simaria AS, Smith M, Farid SS. Decisional tool to assess current and future process robustness in an antibody purification facility. Biotechnol Prog 2012;28:1019–28.
Trinh K, Ferland J, Dinh T. A stochastic optimization method for solving the machine—part cell formation problem. In: Huang DS, et al., editors. Avanced intelligent computing. Lecture notes in computer science, vol. 6838. Heidelberg: Springer-Verlag; 2012. p. 162–9.
Vasquez-Alvarez E, Lienqueo ME, Pinto JM. Optimal synthesis of protein purification processes. Biotechnol Prog 2001;17:685–96.

Vasquez-Alvarez E, Pinto JM. A mixed integer linear programing model for the optimal synthesis of protein purification processes with product loss. Chem Biochem Eng Quart 2003;17:77–84.
Vasquez-Alvarez E, Pinto JM. Efficient MILP formulations for the optimal synthesis of chromatographic protein purification processes. J Biotechnol 2004;110:295–311.
You F, Castro PM, Grossmann IE. Dinkelbach's algorithm as an efficient method to solve a class of MINLP models for large-scale cyclic scheduling problems. Comput Chem Eng 2009;33:1879–89.
Yue D, You F. Sustainable scheduling of batch processes under economic and environmental criteria with MINLP models and algorithms. Comput Chem Eng 2013;54:44–59.