Journal of Chromatography A 1736 (2024) 465321

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Model-based optimization strategy for intensification in the chromatographic purification of oligonucleotides

Santiago Taguado Menza, Rosella Prestia, Ismaele Fioretti, Mattia Sponchioni *

Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Via Mancinelli 7, Milano, 20131, Italy

ARTICLE INFO

Keywords:
Resource-effective chromatographic optimization
Chromatography
Oligonucleotides
Genetic algorithm
Equilibrium-dispersive model
MCSGP

ABSTRACT

Oligonucleotides (ONs) are acquiring clinical relevance and their demand is expected to grow. However, the ON production capacity is currently limited by high manufacturing costs. Since the purification of the target ON sequence from molecularly similar variants represents a major bottleneck, this work presents a resource-effective strategy for the optimization of their preparative reversed-phase chromatographic purification. First, a model based on the equilibrium-dispersive theory was introduced to describe the chromatographic operation. Considering a deoxyribose nucleic acid with 20 nucleobases as case study, a genetic algorithm was developed to efficiently determine the adsorption isotherm and mass transfer parameters for the target ON and impurities. After the estimation of these parameters, a strategy for the *in-silico* optimization of the operation was established. The product collection window, gradient duration, and resin loading were considered as process variables and their influence on yield and productivity was investigated after setting a purity specification of 99.0%. The optimal process parameters identified through this analysis were experimentally verified, confirming the reliability of the model, calibrated with only 5 experimental runs. In addition, this optimal setpoint was exploited to design the multicolumn countercurrent solvent gradient purification (MCSGP) of this ON mixture, which allowed to boost the yield of the process and to work at cyclic steady state, while respecting the purity constraint. This study confirmed the potential of this *in-silico* optimization strategy in both improving the performance of the traditional single-column operations and in the rapid development of multicolumn processes.

1. Introduction

Oligonucleotides (ONs) are short-chain deoxyribonucleic (DNA) or ribonucleic (RNA) acids that can regulate gene expression [1,2]. Due to their mechanism of action and the possibility of controlling the protein synthesis at a pre-translational level, ONs offer fascinating perspectives as novel therapeutics and their demand is expected to grow significantly in the coming years [3-7].

Despite their potential, major manufacturing limitations are currently limiting the ON production capacity [5]. The current state of the art for ON production is the phosphoramidite-based solid-phase synthesis, which is costly and unsustainable from the waste generation and material burden point of view. Indeed, the American Chemical Society (ACS) Green Chemistry Institute Pharmaceutical Roundtable (GCIPR) has identified the development of greener ON manufacturing practices as an urgent unmet requirement [8]. Many of these practices refer to the ON chromatographic purification, the reference process for isolating the target sequence from impurities due to its high selectivity, scalability, and ease of automation. In fact, the downstream processing of ONs accounts for half of the process mass intensity (PMI, defined as the amount of raw materials required for the production of a unit mass of target compound [9]) of the entire manufacturing [8]. This is attributed to both large volumes of eluent consumed in the purification, as well as to the typical low yield associated to the process [10,11]. Hence, from a process engineering perspective, intensification of the ON chromatographic purification can substantially alleviate material consumption and increase the sustainability of the process while improving the large-scale manufacturing of ONs [8,12].

Nonetheless, this task is hampered by the complex separation of the desired nucleic acid from molecularly similar variants produced during the synthesis (*e.g.* shortmers and longmers). In fact, these impurities often partially co-elute with the product in a center-cut chromatographic separation [13,14], as schematically shown in Fig. 1a. As a result, only a limited portion of the chromatogram can be recovered at the required purity, which is often very stringent for these biopharmaceuticals. On the other hand, the impure side fractions are either

* Corresponding author.
E-mail address: mattia.sponchioni@polimi.it (M. Sponchioni).

https://doi.org/10.1016/j.chroma.2024.465321
Received 3 June 2024; Received in revised form 26 August 2024; Accepted 27 August 2024
Available online 30 August 2024

reprocessed, increasing the process complexity, or discarded, reducing the overall product recovery to less than 40 - 50% [8,15]. Larger yields can only be achieved by enlarging the collection window ($t_3 - t_4$ in Fig. 1**a**), which however increases the amount of impurities included in the product pool, thus penalizing its purity. This trade-off between yield and purity is visualized in the Pareto front shown in Fig. 1**b**, where each point on the front maximizes the purity with respect to the points on its right and the yield compared to all the points on its left [12,15,16]. This trade-off limits the performances of the separation, and is strongly affected by the operating conditions. Therefore, the identification of the critical process parameters and of their role on the chromatographic performance is of paramount importance for intensification of the ON downstream processing. Still, given the complexity of the system, an in depth understanding of the main input-output correlations comes with great experimental effort and material consumption.

To alleviate this shortcoming, in this work we developed a model-based optimization strategy for the chromatographic purification of ONs, highlighting the most critical process parameters and their effect on important process performance metrics like yield and productivity. The investigation is conducted specifically on the reversed-phase liquid chromatographic purification (RPLC) of a 20-nucleotide single-stranded DNA chain. The equilibrium-dispersive theory and the Langmuir isotherm were exploited to describe the molecule transport along the chromatographic column and the adsorption equilibrium, respectively [17]. The characterization of the mass transfer from the liquid to the solid phase, expressed through the lumped rate model, as well as of the isotherm parameters, which represents the major complexity for fully mechanistic models [18-20], was performed with a genetic algorithm. This was developed to minimize the difference in key results, including retention times of the different species and Pareto fronts, between the model and selected experiments by adjusting the model parameters. We demonstrated that by only leveraging five training runs in a central experimental scheme, the model can be accurately tuned to reproduce the experimental evidence, avoiding the extensive and costly experimental effort required by data-driven models [21,22].

Then, the calibrated model was exploited for the *in-silico* optimization of the ON RPLC. In particular, we considered as optimization variables the product collection window, gradient duration, and resin loading. These were iteratively tested in a three-step method to maximize yield and productivity at a minimum acceptable purity of 99.0% [16]. The experimental validation of the *in-silico* optimal process conditions and results confirmed the predictivity of the model. This validation also demonstrated the possibility of increasing yield and productivity compared to the reference process when implementing this integral framework.

Despite the remarkable intensification gained with this approach, Pareto fronts between yield and productivity are invariably obtained with single-column operations. To alleviate this limitation, we demonstrated that the optimal process conditions identified with the *in-silico* optimization proposed herein can be reliably exploited to design the multicolumn countercurrent solvent gradient purification (MCSGP) of the same ON. This process operates two identical columns in alternated interconnected and batch stages with the aim of internally recycling, in a completely automated way, the impure side fractions ($t_2 - t_3$ and $t_4 - t_5$ in Fig. 1**a**). The interested reader is referred to [23-25] for further insights in the MCSGP process. Although MCSGP demonstrated important gains in product recovery and PMI compared to batch operations for monoclonal antibodies [26-28], peptides [29,30], PEGylated proteins [16,31,32] and oligonucleotides [12,33,34], its design is typically performed starting from single-column experiments refined until satisfactory performances are obtained. We demonstrated instead that the optimal process conditions identified through our model-based approach represent a robust setpoint for MCSGP, allowing to reach the steady state soon and complying with the purity constraint set at 99.0%, while boosting the product recovery compared to the single-column process. This can therefore represent a valuable strategy to reduce the experimental effort associated to process development while still ensuring its intensification.

2. Materials and methods

2.1. Model equations

In this work, the equilibrium-dispersive model was adopted to describe the transport of molecules along the column axial coordinate (z) during time (t), as reported in Eq. (1) [17].

$$\frac{\partial C_i}{\partial t} + \frac{(1-\varepsilon)}{\varepsilon}\frac{\partial q_i}{\partial t} = D_{AX,i}\cdot\frac{\partial^2 C_i}{\partial z^2} - v\cdot\frac{\partial C_i}{\partial z} \tag{1}$$

where C_i (g/L) is the concentration of the i^{th} component in the liquid phase, q_i (g/L_{resin}) is the average concentration of the adsorbate in the solid phase, ε (-) is the column porosity, D_{AX} $\left(\frac{cm^2}{min}\right)$ is the axial dispersion coefficient, and v $(\frac{cm}{min})$ the interstitial velocity. To describe the transport of the modifier during the gradient, a similar equation was adopted neglecting its adsorption to the stationary phase (Equation S1).

The model equations are applied during the four characteristic steps of the chromatographic process: loading, wash, elution and strip [35].

Thus, the initial conditions vary depending on the chromatographic

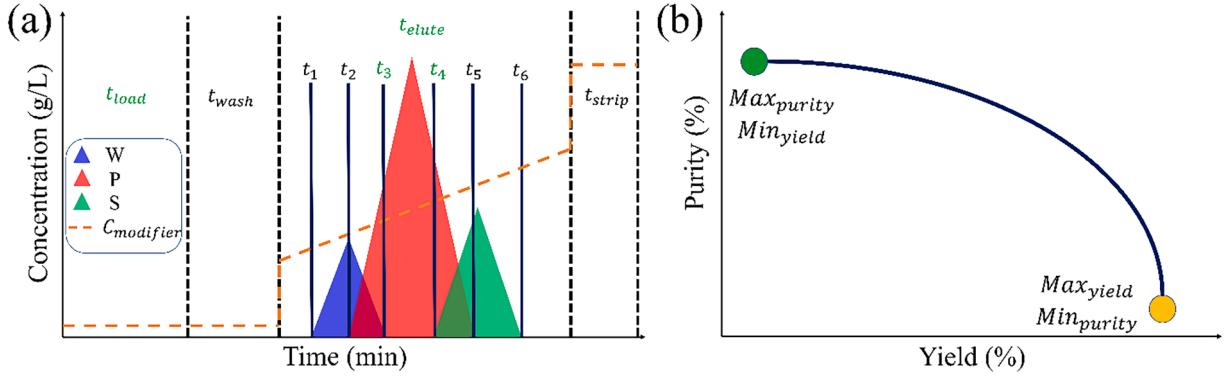

Fig. 1. (a) Schematic representation of a typical ON chromatogram obtained in a linear gradient elution and its different characteristic regions. The similar physicochemical properties of the product (P, in red) and those of weakly adsorbed (W, blue peak) and strongly adsorbed (S, in green) impurities cause their partial co-elution. In this example, P and W co-elute from t_2 to t_3, while P and S co-elute from t_4 to t_5. Therefore, only in the interval from t_3 to t_4, P reaches the required purity specification. (b) Due to the partial co-elution of the product and the impurities in the two characteristic regions $t_2 - t_3$ and $t_4 - t_5$, a Pareto front yield *vs.* purity is typically obtained by changing the product collection window ($t_3 - t_4$). Enlarging this window allows recovering more product, but at a lower pool purity. The point of the maximum yield corresponds to the collection interval $t_2 - t_5$.

step. For the loading phase, *i.e.* $0 < t \leq t_{load}$, we assumed the column is clean and equilibrated to the desired value of modifier volume fraction, in this case 30% v/v. The initial conditions are reported in details for each chromatographic step in the Supporting Information, Equations S2-S13. Moreover, Danckwerts boundary conditions are applied, as reported in Equations S14-S17.

The mass transfer to the solid phase is described using the linear driving force approximation, with the change of the average concentration of the species in the solid phase computed as in Eq. (2).

$$\frac{dq_i}{dt} = K_i \left(q_i^* - q_i\right) \tag{2}$$

where q_i^* (g/L_{resin}) is the concentration of the species *i* in the solid phase when thermodynamic equilibrium with the liquid phase is reached, q_i (g/L_{resin}) is the instantaneous average concentration of the same molecule in the solid, and *K* (min^{-1}) is a rate coefficient lumping together the mass transfer resistances encountered by the solute in its partitioning between the solid and liquid phases. Throughout the investigation, the adsorption isotherm that describes the solid concentration of a species in equilibrium with the liquid phase is the generalized Langmuir isotherm with linear solvent strength (LSS) to account for the effect of the modifier, as reported in Eq. (3).

$$q_i^* = \frac{a_i \cdot \exp\left(-S_{a,i} \cdot \varphi\right) \cdot C_i}{1 + \sum_j b_j \cdot \exp\left(-S_{b,j} \cdot \varphi\right) \cdot C_j} \tag{3}$$

where b_j (L/g) is the adsorption equilibrium constant for the *j-th* specie at $\varphi = 0$, a_i (L/L_{resin}) is the product of b_i and the saturation capacity of the resin at $\varphi = 0$, while $S_{a,i}$ (-) and $S_{b,j}$ (-) are parameters expressing the sensitivity of *a* and *b* for the *i th* specie from the modifier volume fraction, respectively.

The partial differential equations (PDEs) involved in the model were discretized along the longitudinal coordinate. The semi-discretization converts the PDE into a system of ordinary differential equations (ODEs) [35,36]. The integration of this ODE system describing the chromatographic column in space and time is then performed using the MATLAB® ODE15s proprietary solver.

For this semi-discretization, we used the finite volume method (FVM) with van Leer's flux limiters [35-39]. The semi-discretized form of Eq. (1) using this scheme is as follows:

$$\begin{aligned}\frac{dC_{i,j}}{dt} = & \frac{D_{AX}}{h^2}\left(C_{i,j+1} - 2C_{i,j} + C_{i,j-1}\right) \\ & + \frac{v}{h}\left[C_{i,j-1} - C_{i,j} + 0.5\,\phi\left(r_{i,j-1}\right)\left(C_{i,j} - C_{i,j-1}\right) - 0.5\,\phi\left(r_{i,j}\right)\left(C_{i,j+1} - C_{i,j}\right)\right] \\ & - \frac{(1-\varepsilon)}{\varepsilon} \cdot \frac{dq_{i,j}}{dt}\end{aligned} \tag{4}$$

where *h* (cm) is the integration step, determined by dividing L_c by an appropriate number of nodes on the longitudinal coordinate (N_z), *j* represents the finite volume considered, while ϕ and r_j are the van Leer's flux limiter and the gradient sensor, which are calculated as follows:

$$r_j = \frac{C_j - C_{j-1} + \delta}{C_{j+1} - C_j + \delta} \tag{5}$$

$$\phi\left(r_j\right) = \frac{r_j + |r_j|}{1 + |r_j|} \tag{6}$$

Where δ is a constant equal to 10^{-16} to prevent the algorithm from diverging to zero [40]. Finally, the half-cell approximation assuming $C_{i,1} - C_{i,1/2} = C_{i,1/2} - C_{i,0}$ was applied to describe the first finite volume ($j = 1$).

For the integration of the resulting ODE system, the number of time and space steps was held constant at 1500 and 100, respectively, to allow for feasible computation times while minimizing integration errors.

2.2. Model calibration

The 20mer single-stranded DNA sequence investigated in this work was kindly provided by YMC Japan. The analytical characterization of the mixture of ONs and of the fractions collected during the different linear gradient elutions was performed via high-performance liquid chromatography (HPLC) following the method reported in [12] and detailed in the Supporting Information.

From the analytical chromatogram, the purity was computed as the ratio between the area of the product peak appearing at 30.2 min (see **Figure S1** in the supporting information) and the sum of the areas of all the peaks. Based on the retention time of the different impurities, the ON mixture was simplified as constituted by three virtual key components and product. Early weakly adsorbing impurities W1, lumping up all the species eluting before 8.0 min, accounted for those species that do not co-elute with the target product in preparative experiments and represented the 0.33% of the crude material. Late weakly adsorbing impurities W2, eluting from 8.0 to 30.0 min, represented the oligonucleotides that overlapped with the target sequence in preparative experiments and accounted for 8.57% of the mixture. The product P, eluting at 30.2 min, was characterized by 89.70% purity. Finally, the strongly adsorbing impurities S, eluting later than the product, accounted for 1.40% of the mixture.

To calibrate the model and determine the isotherm parameters and mass transport properties, preparative RPLC experiments were carried out using a Contichrom CUBE 30 system (YMC ChromaCon) equipped with a YMC Triart Prep C18-S column (10 μm particle size, 100 mm length x 4.6 mm internal diameter), having a porosity (ε) of 0.6. The equilibration buffer was a 0.2 M sodium acetate ($C_2H_3NaO_2$, purity ≥99%, molecular weight = 82.03 g/mol; Sigma-Aldrich)/acetonitrile (C_2H_3N, purity ≥99.9%, molecular weight = 41.05 g/mol; sourced from Honeywell) 99/1% v/v solution. The elution buffer contained 90% v/v of 0.2 M sodium acetate and 10% v/v acetonitrile. Loading of the crude mixture was invariably performed at 300 cm/h, while different amounts of fed ONs per liter of resin were explored as reported in Table 1. Washing was performed at 30% elution buffer for 2 column volumes (CV) at 150 cm/h. The gradient was conducted by linearly increasing the elution buffer volume fraction from 30% up to 100% at 200 cm/h for different durations, as shown in Table 1. Stripping was performed at 150 cm/h for 2 CV at 100% elution buffer. Finally, the system was re-equilibrated at 30% elution buffer for 3 CV at 400 cm/h. In each experiment, a feed solution at 4.04 g/L of ONs with the composition reported above was used. The separation was performed at 50 °C through an external column thermostat Knauer Azura CT 2.1 equipped with a solvent pre-heating cartridge. The chromatogram was recorded at the outlet of the column with a Knauer BlueShadow 40D UV detector set at 300 nm.

During the gradient, 20 fractions were collected for each experiment, and analyzed via HPLC. The product purity in each fraction was estimated by dividing the area of the peak associated to the product by the sum of the areas of all the peaks in the analytical chromatogram. The yield associated to each fraction was defined according to Eq. (7).

$$Y = \frac{V_f \; C_P}{Q_{load} t_{load} C_{P,in}} \tag{7}$$

Where V_f is the volume of the fraction considered, C_P is the product concentration in the fraction determined via HPLC, Q_{load} and t_{load} are the loading flow rate and duration, respectively, and $C_{P,in}$ is the product concentration in the fed mixture.

For each experiment, different collection windows were obtained by pooling contiguous fractions and the yield and purity of each of them was determined from the HPLC characterization of the pooled fractions to create the purity *versus* yield Pareto fronts.

The experimental data was then used to estimate the adsorption

Table 1
Conditions adopted in the different experiments used to calibrate the model. CV=Column volume.

	Gradient Duration [CV] (30-100%B)		
	4.5	6.5	8.5
Load 7.5 g/L_{resin}	/	E1	/
Load 15.0 g/L_{resin}	E4	E2 (Center Point)	E5
Load 22.5 g/L_{resin}	/	E3	/

model parameters described in Eq. (3) for each compound as well as the axial dispersion and lumped mass transfer coefficients shown in Eqs. (2) and (4). The parameter estimation was carried out using the genetic algorithm (GA) function in MATLAB [41]. The fitness function F developed in this work is reported in Eq. (8).

$$F = \sum_{exp=1}^{5}\left(\sum_{j}^{NC}\frac{\left(\mu_{exp,j}-\mu_{mod,j}\right)^2}{\mu_{exp,j}^2} * \boldsymbol{P} + \sum_{j}^{NC}\frac{\left(C_{max,exp,j}-C_{max,mod,j}\right)^2}{C_{max,exp,j}^2} + \sum_{j}^{NC}\frac{\left(\sigma_{exp,j}-\sigma_{mod,j}\right)^2}{\sigma_{exp,j}^2} + \sum_{j}^{NC}\frac{\left(HETP_{exp,j}-HETP_{mod,j}\right)^2}{HETP_{exp}^2,j} + \sum_{k}^{NP}\frac{\left(purity_{exp,k}-purity_{mod,k}\right)^2}{purity_{exp,k}^2}\right) \quad (8)$$

F sums up the squared relative errors on mean retention time (μ), concentration at peak maximum (C_{max}), variance (σ), and height equivalent to a theoretical plate (HETP) for each component *j* (with NC total components) and experiment *exp*, as well as the pool purities for *k* experimental collection windows.

To ensure the retention times are accurately predicted, they are penalized within the fitness function by a parameter P [41-43]. The penalization factor is estimated after a preliminary single-component parameter estimation.

For the GA optimization process, the population size, generation size, and absolute error criteria were set to 400, 100, and 10^{-6}, respectively, to ensure high accuracy and computational time trade-off [44,45]. Finally, the accuracy of the model was measured as follows:

$$Accuracy = \left(1 - \frac{|model\ prediction - experiment\ result|}{experiment\ result}\right)\cdot 100\% \quad (9)$$

2.3. Optimization algorithm

The *in-silico* procedure aims at optimizing the loading time, gradient duration, and collection window to maximize the yield and productivity of the ON RPLC whilst being subjected to strict purity constraints [16]. With reference to the typical chromatogram for a center-cut purification shown in Fig. 1a, the productivity, yield and purity for a collection window defined by the interval $t_3 - t_4$ were calculated as in Eqs. (10)–(12) and utilized throughout the optimization procedure.

$$\mathrm{Pr} = \frac{C_{P,in}\cdot Q_{load}\cdot t_{load}\cdot Y}{V_C\cdot\left(t_{load}+t_{wash}+t_{elution}+t_{strip}\right)} \quad (10)$$

$$Y = \frac{Q_{elution}\cdot\int_{t_3}^{t_4} C_P dt}{Q_{load}t_{load}C_{P,in}} \quad (11)$$

$$\mathrm{Purity} = \frac{\int_{t_3}^{t_4} C_P dt}{\sum_i \int_{t_3}^{t_4} C_i dt} \quad (12)$$

Based on these definitions, four characteristic process variables are iteratively adjusted: t_3 (beginning of collection window), t_4 (end of collection window), $t_{elution}$, and t_{load}. The remaining process parameters are kept constant.

The optimization strategy is divided in 3 steps.

During the first step, $t_{elution}$ and t_{load} are fixed at 19.5 min (*i.e.* 6.5 CV) and 7.5 min (*i.e.* 15.0 g/L_{resin}) respectively, to reproduce the conditions of the center point in Table 1. On the other hand, diverse collection windows (t_3, t_4) are analyzed. In particular, different collection starting times in the interval $t_2 \leq t_3 < t_5$ were considered. For each value of t_3, different collection end times in the range $t_3 < t_4 \leq t_5$ were investigated. Each pair (t_3, t_4) constitutes a possible product collection window, for which purity, yield, and productivity were recorded. Among all the possible combinations, the optimal collection window (t_3, t_4) was the one leading to the maximum yield, and correspondingly, to the maximum productivity, while respecting the purity constraint (*i.e.* purity ≥ 99.0%).

The second step repeats the procedure mentioned so far for different elution times ($t_{elution}$) from 13.8 min (*i.e.* 4.5 CV) to 25.5 min (*i.e.* 8.5 CV) while holding t_{load} constant. In this way, the optimal collection window and $t_{elution}$ maximizing yield and productivity at the required purity specification can be identified for a specific t_{load}.

Finally, in the third step, also t_{load} is changed from 3.75 min (*i.e.* 7.5 g/L_{resin}) to 11.25 min (*i.e.* 22.5 g/L_{resin}). The ultimate result in this third step is the optimal collection window, $t_{elution}$, and t_{load} that maximized productivity or yield.

2.4. Model validation and MCSGP

The optimal conditions in terms of loading, gradient duration and extension of the collection window defined *in-silico* were tested experimentally using the same equipment and method reported in Section 2.2. This served as a validation of the model developed, as the conditions implemented were different from those used for parameter calibration. 22 fractions were collected during the linear gradient elution and analyzed via HPLC to define purity and yield of the product as well as the productivity according to Eqs. (10)–(12). These were compared with the model results and the deviations for each parameter were quantified in terms of prediction accuracy as shown in Eq. (9).

The process conditions leading to the maximum productivity in the single-column process were also transferred to MCSGP. The process is conducted on the same system and using the same buffers as reported in Section 2.2 for the batch operation. The design is performed using the ChromIQ software (version 8.0) and requires the specification of 5 characteristic times, from t_1 to t_5 in Fig. 1a. These represent the start of the gradient, start of the W/P overlap internal recycling, start and stop of product collection and stop of P/S internal recycling, respectively [13, 14]. Often, these times are determined from single-column experiments [34,46]. Here, t_3 and t_4 directly came from the model-based optimization, t_1 was set at the beginning of the gradient, while t_2 and t_5 were taken as the times when the product started and ended to elute based on the model simulation. Therefore, we set t_1=19.5 min, t_2=28.1 min, t_3=36.9 min, t_4=40.9 min, and t_5=46.8 min. The MCSGP was conducted for 5 consecutive cycles and every cycle a product pool was collected in the interval t_3 - t_4, combining the elutions from each of the two columns. These pools were analyzed via HPLC to determine purity, yield and productivity of the process cycle after cycle.

3. Results

3.1. Model development and calibration

The equilibrium-dispersive model (Eq. (1)) was adopted to describe the transport of the different components and of the modifier in the chromatographic column. The accumulation of each molecule in the solid phase is computed through the linear driving force approximation, as the product between a lumped rate constant and a driving force,

which is the distance between the average species concentration in the stationary phase from the equilibrium conditions (Eq. (2)). These were accounted for using the Langmuir isotherm, whose parameters are function of the local volume fraction of modifier φ through the linear solvent strength model (Eq. (3)). The obtained system of partial differential equations was numerically solved with the FVM.

The reliability of the model was first verified by reproducing literature results. As an example, we qualitatively compared the model results obtained at different concentrations of the loaded mixture with those proposed by Qamar et al. [35] in **Figure S2** (see Supporting Information section). The numerical algorithm displays good agreement with the analytical solutions provided by the authors at both 2 g/L and 10 g/L injections. Therefore, it was considered sufficiently robust and reliable.

After having verified its reliability, the model was calibrated through a statistical approach. This exploits a genetic algorithm to minimize the difference between experimental and model results in the form of the fitness function reported in Eq. (8). To minimize the number of experimental runs, and hence the resources to be dedicated to this model calibration stage, while ensuring highly accurate parameter estimations, a central scheme was adopted [47-49]. In particular, we investigated 3 different loadings, *i.e.* 7.5, 15.0 and 22.5 g/L_{resin}, and 3 different gradient durations, *i.e.* 4.5, 6.5 and 8.5 CV, as shown in Table 1.

With the aim of restricting the parameter domain, we preliminarily applied the genetic algorithm considering only the target product in the simulatons and limiting the number of experiments to E1, E2, and E3. The search boundaries and parameter estimations of the target product are displayed in **Table S1**, while **Figure S3** shows the comparison between model and experimental results with the optimal set of parameters after this preliminary screening. It is possible to observe that the retention time of the product is invariably overestimated by the model. The limitation detected in capturing the retention time for the single-component fitting was addressed by placing a penalty of $P = 1000$ on the contribution from the retention time within the fitness function.

With this expedient and the refined lower and upper boundaries for the different parameters from the single-component investigation, the GA optimization was extended to the multi-component system, comprising 3 virtual key components (*i.e.* W1, W2, and S), the product, and the modifier. For this system, the model requires 19 parameters: 4 parameters of the Langmuir isotherm for each of the 4 components (16 parameters), 1 axial dispersion coefficient for all components, 1 mass transfer rate constant for all components, and 1 axial dispersion coefficient for the modifier.

The lower boundaries, upper boundaries, and the parameter estimates obtained with this approach are listed in Table 2.

Using these parameters, the multi-component chromatograms and Pareto curves were reproduced for each run and compared to the experimental results in Fig. 2.

Here, the good accuracy of the model calibrated through the genetic algorithm in reproducing both the chromatogram and Pareto front for each of the experimental run is evident. In particular, the best fit is displayed at the intermediate loading of 15.0 g/L_{resin}, where the yield-purity tradeoff in place for this center-cut purification can be reliably captured by the model for the different gradient durations tested. On the other hand, the Pareto front tends to be underestimated when higher loadings are considered. This might be due to the displacement of the weak impurities operated by the product, which increases the measured purity when increasing the loading. Indeed, this peculiar phenomenon was already reported for the same model ON mixture, while at the same time being difficult to be captured by the model [12]. The quantification of model accuracy for the different peak features, calculated from Eq. (9), is reported in **Table S2**. From this, we can conclude that the GA optimization manages to reliably predict the mean retention time, with an accuracy that for all the experiments is >95%. Similar great reliability was found in the prediction of the Pareto fronts, whose accuracy is always >96%. More imprecision is observed instead in the reproduction of the peak maxima and in the variance of the peaks, mainly at large loadings.

3.2. In-silico optimization

After having validated the model accuracy, an *in-silico* optimization of the ON RPLC was carried out. The procedure was articulated on three levels, aimed at maximizing yield and productivity at a given purity constraint of 99.0% by changing: i) the collection window, ii) $t_{elution}$, and iii) t_{load}.

During the first step of the batch process optimization, the gradient duration and loading were fixed to 6.5 CV and 15.0 g/L_{resin}, respectively, while different collection windows $t_3 - t_4$ were investigated in the elution time interval between t_2 and t_5. Among all the tested combinations, only the product pools characterized by >99.0% purity were retained. Fig. 3a graphically displays the iterative procedure. In particular, for each t_3 considered, corresponding to a single line in the plot, increasing values of t_4 as second border for the product collection window were investigated. By increasing this parameter, as it can be deduced from the example chromatogram in Fig. 1a, more and more product is initially included in the pool, which causes the purity to increase and eventually surpass the purity specification. However, if t_4 is further increased, more and more strongly adsorbing impurities are included in the pool, whose purity starts dropping. As such, for a given t_3, the optimal t_4 is obtained when the purity decreases below P_{spec} this second time, as this is associated to the maximum possible recovery of the product at acceptable purity. Following this approach, from Fig. 3b it is possible to observe that the purity constraint is respected only for $31.4 < t_3 < 32.4$ min. Among all the values of t_3 (and the corresponding optimal t_4) satisfying these requirements, an optimal collection window maximizing the yield can be identified. This was then chosen as the optimal product recovery interval at fixed loading and gradient duration. It is worth highlighting that, being all the characteristic times of the process fixed, maximizing the yield ensures that also a maximum productivity is obtained (see Eq. (10)). These maximum productivity and yield, together with the optimal collection window are reported in **Table S3**, which confirms that the constraint on the minimum acceptable purity is satisfied. At the same time, with these loading and gradient duration, only 21% of the fed product could be recovered at this strict purity requirement, confirming the necessity of process optimization.

In the second step of the optimization procedure, the investigation of

Table 2
Lower boundaries, upper boundaries, and parameter estimates for the optimization run using population and generation size of 400 and 100, respectively.

	Lower Boundary	Upper Boundary	W1	W2	P	S1	φ
$D_{AX}\ \left(\frac{cm^2}{min}\right)$	1.00E-04	1.00E-02	3.50E-03	3.50E-03	3.50E-03	3.50E-03	3.00E-03
$a_i\ \left(\frac{L}{L_{resin}}\right)$	5000	25,000	15,280	8820	7540	19,200	–
$S_{a,i}\ (-)$	5.00	15.00	14.66	11.67	9.93	9.17	–
$b_i\ \left(\frac{L}{g}\right)$	500	3000	2140	1820	2150	1350	–
$S_{b,i}\ (-)$	5.00	30.00	10.69	22.44	24.46	20.09	–
$K\ (min^{-1})$	0.80	2.50	0.99	0.99	0.99	0.99	–

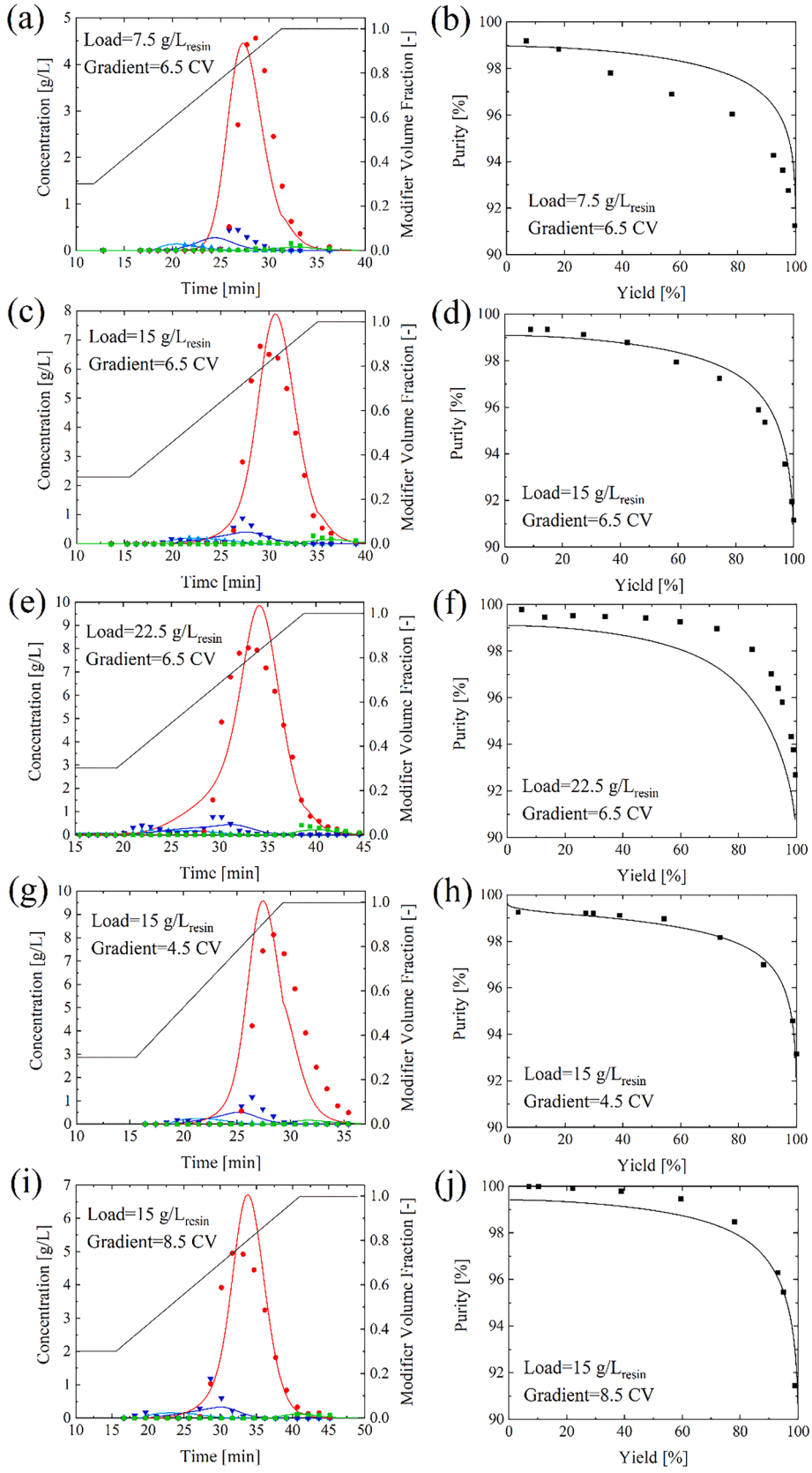

Fig. 2. Comparison of experimental multi-component chromatograms and purity-yield Pareto fronts (symbols) against model results (lines) using the parameters estimated from the genetic algorithm and reported in Table 2. (a-b) Load = 7.5 g/L$_{resin}$, gradient duration = 6.5 CV; (c-d) Load = 15.0 g/L$_{resin}$, gradient duration = 6.5 CV; (e-f) Load = 22.5 g/L$_{resin}$, gradient duration = 6.5 CV; (g-h) Load = 15.0 g/L$_{resin}$, gradient duration = 4.5 CV; (i-j) Load = 15.0 g/L$_{resin}$, gradient duration = 8.5 CV. W1 is displayed in light blue, W2 in dark blue, P in red and S in green.

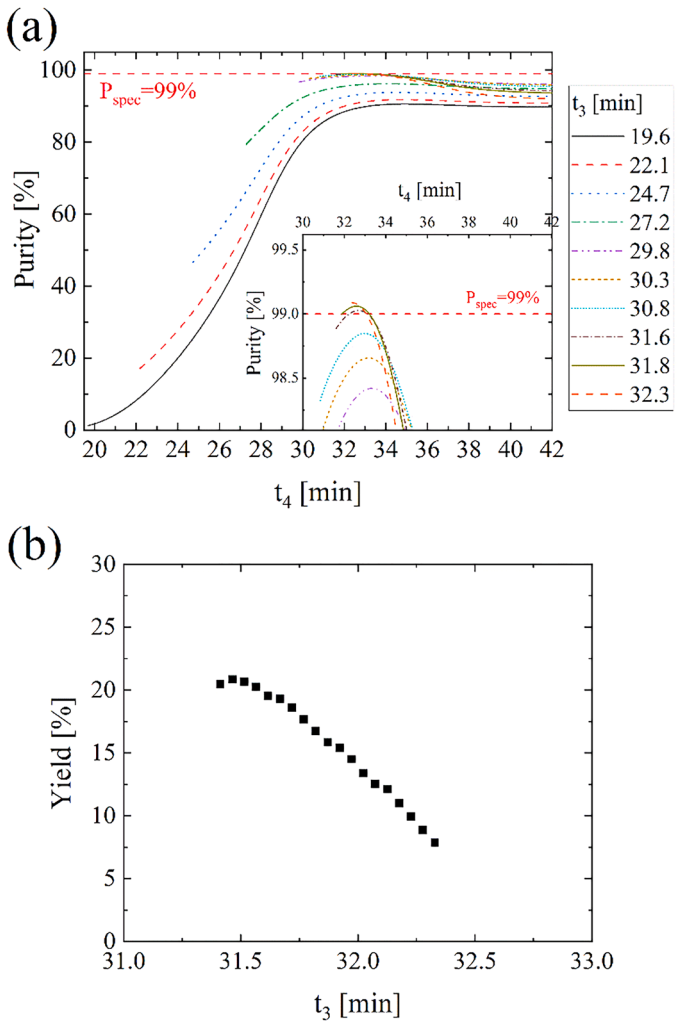

Fig. 3. First stage of the optimization strategy. (a) Effect of t_4 on purity (%) for selected values of t_3 to illustrate the procedure. The dashed line represents P_{spec}=99.0%. In the inset, magnification of the results in proximity of the purity constraint. For a given curve, the optimal value of t_4 is considered as the largest one at the intersection with the purity specification. (b) Evolution of the process yield for the different collection windows satisfying the constraint on the product purity. The results reported in this figure refer to a loading of 15.0 g/L_{resin} and gradient duration 6.5 CV.

the optimal collection window is repeated for different gradient durations. **Figure S4** depicts the productivity and yield associated to the optimal collection window satisfying P_{spec} at each tested elution time. It is worth observing that this constraint on purity can only be met for elution durations larger than 6.1 CV. For steeper gradients, W2 and S are not separated enough from P to identify a collection window with purity >99.0%. On the other side, in the highlighted design space, both yield and productivity monotonously increase with the elution time. This is not surprising as shallower gradients allow for a better separation of the different components, so that more product can be collected in the pool at the required purity threshold. Based on this, the longer gradient duration is considered the optimal one for this specific process. Still, it is worth noting the reduction in the slope of the productivity when proceeding to larger elution times. In fact, despite the larger amount of product recovered, the increase in the total process duration associated to longer elutions negatively impacts this performance indicator. Indeed, we expect an inversion in the trend for even larger gradient durations, whose negative effect on productivity would eventually prevail the improvement introduced by the increase in the yield, thus not justifying long operations.

The last step of the optimization repeats the first and second steps whilst varying the loading. Fig. 4a shows the evolution of yield and productivity with the gradient duration and loading. In particular, each line refers to a value of tested loading time and reports yield and productivity at variable gradient duration.

This analysis reflects that both yield and productivity monotonically increase with longer gradient durations at all tested loadings (from 7.5 to 22.5 g/L_{resin}), as already observed during the second step of the optimization. Therefore, the best elution time in this process is the largest one (8.5 CV) in our design space. Based on this evidence, Fig. 4b reports yield and productivity at this largest gradient duration for the different loads. By increasing this process parameter, in a first stage from 7.5 to 11.0 g/L_{resin} both yield and productivity improve. Therefore, it is not useful for this process to work at loadings below 11.0 g/L_{resin}. On the other hand, above this threshold, a tradeoff between yield and productivity can be observed, with the former increasing at lower loadings and the latter growing in the opposite direction. This behavior can be explained considering that increasing the mass of ONs fed to the column corresponds to augmenting the area below their elution peaks. This results in a stronger overlap between the product and main impurities, which reduces the amount of product collectable at P_{spec}. On the other hand, increasing the ON loaded to the column has a positive effect on the

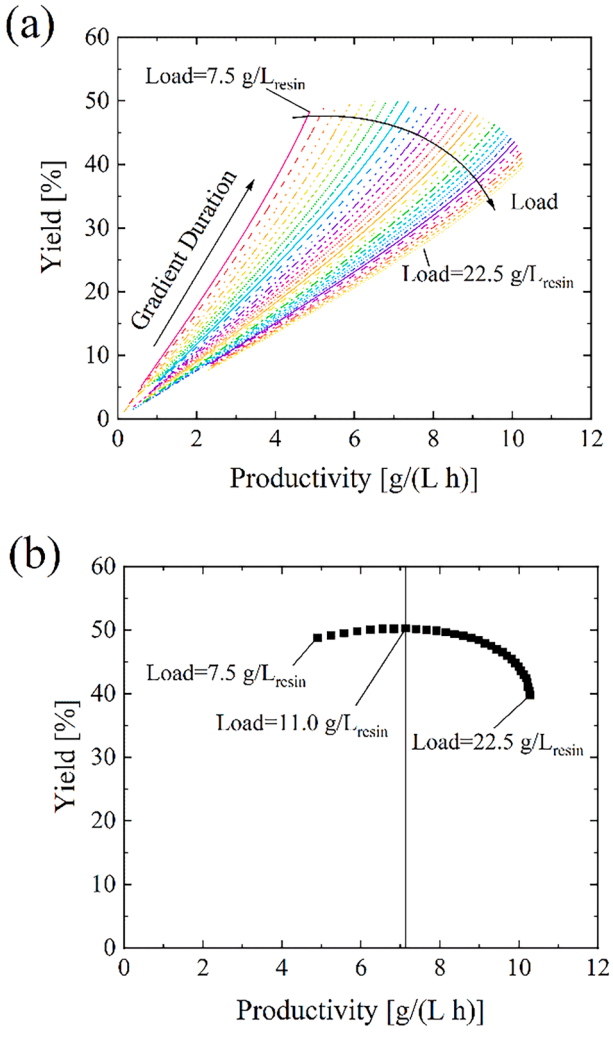

Fig. 4. Third optimization step. (a) Effect of loading, from 7.5 to 22.5 g/L_{resin}, and gradient duration, from 4.5 to 8.5 CV, on productivity and yield. (b) Optimal points at the longest gradient duration of 8.5 CV and different loadings from 7.5 to 22.5 g/L_{resin}, revealing a yield-productivity tradeoff starting from a load of 11.0 g/L_{resin}.

productivity. Therefore, the optimization strategy is in agreement with the common sense and the physics of the system.

We can then conclude that for this process there is no absolute optimum, but the best process parameters depend on the desired performance. If the yield is to be privileged, which is the case for difficult-to-manufacture products, as this would reduce the pressure on the upstream processing, a loading of 11.0 g/L_{resin} should be preferred. In this case, up to 50.3% yield can be achieved. Larger values would instead violate the constraint on 99.0% purity. On the other hand, working at 22.5 g/L_{resin} maximizes the productivity to 10.28 g/(L h). The optimal set of process parameters and performances for these two scenarios are summarized in Table 3.

Comparing the results with the center point of the experimental scheme shown in Table 1, the optimization procedure allowed to increase the yield from 20.9 to 50.3% (+140% gain) when this is the parameter of interest, and the productivity from 4.41 to 10.28 g/(L h) (+133% gain) when focusing on the latter.

This is an outstanding result, considering that only 5 experiments were required to calibrate the model, which in turn was used to simulate more than 3200 chromatograms to define the optimal process conditions. It is easy to understand the major saving in experimental effort and material consumption introduced by the optimization strategy developed in this work, which allowed to achieve a remarkable gain in the process performances.

3.3. Model validation

The process conditions identified through the model-based optimization leading to the maximum yield or productivity were experimentally tested, with the aim of validating the predictions. In particular, the setpoints on load, gradient duration, t_3 and t_4 are those reported in Table 3. For both experimental runs, 22 fractions were collected during the elution phase and analyzed via HPLC to define the concentration of each component. The comparison between the experimental and model results in terms of preparative chromatogram and Pareto fronts are then shown in Fig. 5.

For both runs, it is possible to appreciate the accurate prediction of the mean retention time for all the components as well as of the peak height for the product, the most abundant species. In addition, the Pareto fronts reliably depict the tradeoff between purity and yield for this single-column process. This provides a good indication that the partial co-elution of W2 and P as well as of P and S is properly described by the model. This confirms the potential of the approach investigated in this work and the relevance of the model-based optimization.

Table 3
Comparison of model predictions and experimental results for productivity, purity and yield in the two validation runs maximizing yield (load = 11.0 g/L_{resin}, gradient duration = 8.5 CV) and productivity (load = 22.5 g/L_{resin}, gradient duration = 8.5 CV).

		Model	Experiment	Accuracy [%]
Max Yield	**Load [g/L_{resin}]**	11.0		–
	$t_{elution}$ [CV]	8.5		–
	t_3 [min]	31.7		–
	t_4 [min]	35.7		–
	Productivity [g/(L h)]	7.13	5.64	73.6
	Purity [%]	99.0	99.3	99.7
	Yield [%]	50.3	40.3	75.1
Max Productivity	**Load [g/L_{resin}]**	22.5		–
	$t_{elution}$ [CV]	8.5		–
	t_3 [min]	36.9		–
	t_4 [min]	40.9		–
	Productivity [g/(L h)]	10.28	8.94	85.0
	Purity [%]	99.0	99.1	99.9
	Yield [%]	39.8	34.8	85.6

Nonetheless, at load = 11.0 g/L_{resin}, the concentration at peak maximum for W2 is underestimated by the model, which predicts a broader peak for this pseudo-component. Since the peak front elutes before the product, this is reflected in an overestimation of the maximum purity achievable at high yields, as highlighted in the Pareto front. On the other side, at high loading (*i.e.* 22.5 g/L_{resin}), the model accurately predicts the mean retention time of the different components, but overestimates the peak front. Indeed, the significant overlap shown by the model results for P and W2 in the front leads to the underestimation of the maximum purity achievable at high yield, as reflected in the Pareto front.

For both experiments, the process performances in terms of purity, yield and productivity were finally determined by pooling the appropriate number of fractions to reproduce the optimal collection window (t_3, t_4) predicted by the model in each condition. These are compared with the model outputs in Table 3. In all cases, the accuracy of the model is >74%, which considering the experimental variability confirms the significance of the process optimization. At the same time, it is worth highlighting that the implemented collection window shows a remarkable agreement in terms of product purity with the result from the model and, in particular, it allows respecting the constraint on the purity specification. This is particularly important considering that meeting a certain purity requirement discriminates the release of the entire drug lot. These experimental runs also validate the drastic improvement in yield and productivity compared to the central point of Table 1, testifying the potential of this model-based optimization.

3.4. MCSGP

In addition to the maximization of yield or productivity for a single-column operation, the approach proposed in this work can be appealing for the rapid definition of an optimal setpoint for MCSGP. In particular, this twin-column operation requires, in addition to the typical process parameters like load and gradient duration, a collection window for product recovery (t_3-t_4 in Fig. 1a), as well as time intervals for internal recycling of W/P overlaps (t_2-t_3 in Fig. 1a) and P/S overlaps (t_4-t_5 in Fig. 1a). By applying this internal recycling, the MCSGP significantly improves the yield compared to a single-column operation at comparable purity of the product pool. For this reason, the most suitable process conditions to be transferred to MCSGP are those maximizing the productivity, letting the internal recycling of the impure side fractions compensate for the poor product recovery of this operation.

Therefore, with reference to the batch chromatogram in Fig. 5c, we designed the MCSGP with the same startup load of 22.5 g/L_{resin}, gradient duration of 8.5 CV and collection window from 36.9 to 40.9 min. On the other side, the product-containing impure fractions eluting from 28.1 to 36.9 min and from 40.9 to 46.8 min were recycled from the upstream to the downstream column during the interconnection steps of the process. The detailed MCSGP phases are shown in Fig. 6a. The MCSGP was conducted for 5 cycles and the UV signal recorded at the outlet of each of the two columns is shown in Fig. 6b.

The product pool was collected every cycle and analyzed by HPLC to determine its composition and concentration. The results are shown in Table 4. It is possible to observe that the system reaches the cyclic steady state very soon. In fact, already from the third cycle the composition of the pool as well as the process performance do not change anymore. This can be visually assessed also from the UV chromatogram in Fig. 6b, where the peak profile is preserved very similar starting from 125 min, which marked the beginning of cycle 3. In addition, the transfer of the optimal collection window defined by the model ensured that the purity specification was respected in all the cycles of the MCSGP. This confirmed the strength of this approach in the resource effective design of a continuous operation providing a product pool in specification. Finally, although the productivity is comparable to the one obtained in the single-column experiment, the yield could be drastically improved from 34.8 to >96.1%. The MCSGP designed through the model developed in this work allows then major savings in valuable product, while

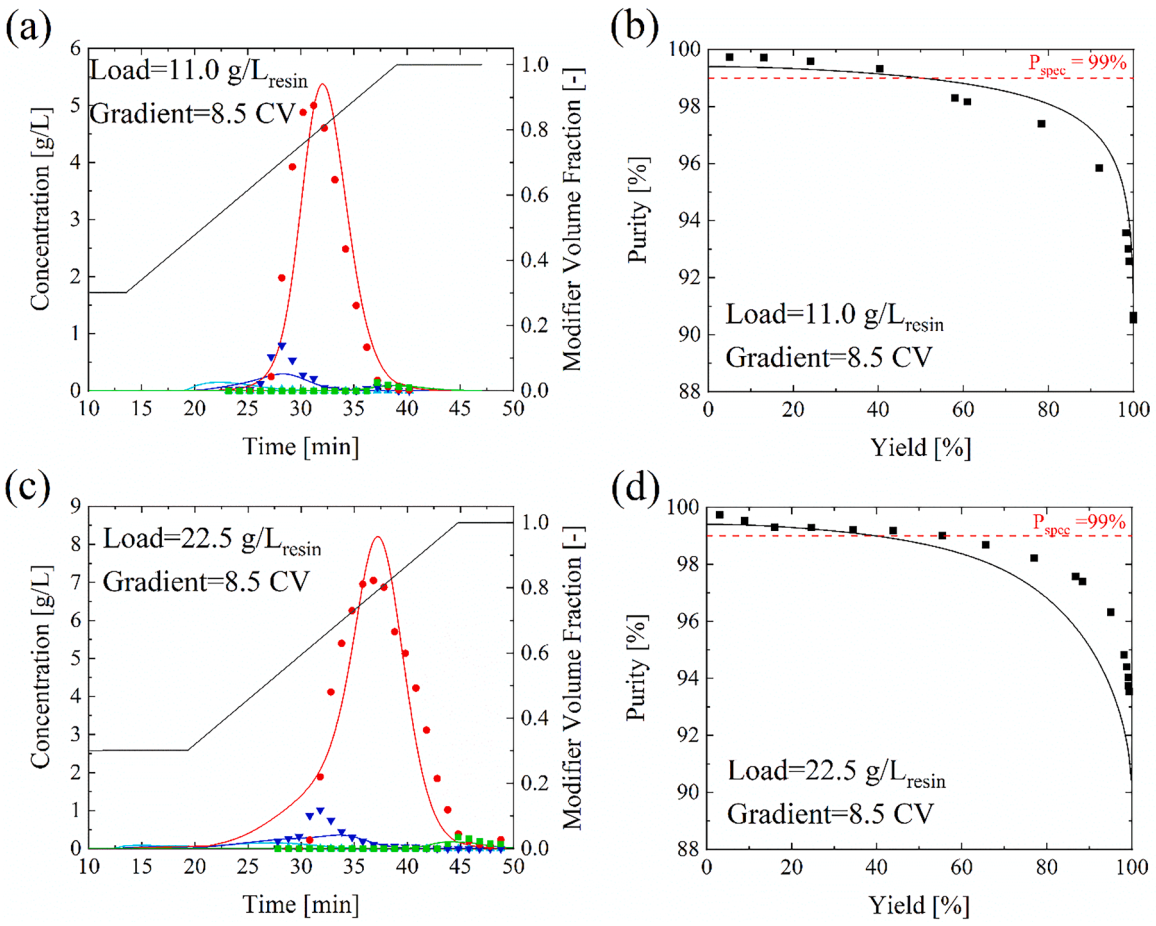

Fig. 5. Comparison between experimental results (symbols) and model predictions (lines) in terms of preparative chromatogram and Pareto fronts for the runs maximizing the yield, with load = 11.0 g/L$_{resin}$ and gradient duration = 8.5 CV (a-b) and the productivity, with load = 22.5 g/L$_{resin}$ and gradient duration = 8.5 CV (c-d). W1 is displayed in light blue, W2 in dark blue, P in red and S in green.

providing the required purity.

4. Conclusions

The chromatographic purification of oligonucleotides currently impacts 43% of the overall process mass intensity [8]. This is attributed to the large consumption of buffers as well as to the limited product recoveries. At the same time, process optimization can be extremely time and resource intensive. To overcome this limitation and permit process intensification in the purification of oligonucleotides, we developed a model-based approach to optimize the performances of the single-column chromatographic purification of a 20mer DNA sequence. The equilibrium-dispersive model describing the molecule transport along the column was considered and the model parameters were calibrated through a genetic algorithm by exploiting only 5 experimental runs in a central scheme. After having validated the reliability of the model, we were able to exploit it in an *ad hoc* optimization strategy, investigating the role of the collection window, loading and gradient duration on the yield and productivity of the process, imposing a minimum acceptable purity of 99.0%. While an increase in the elution duration improves both yield and productivity, we found out that these are bound to a tradeoff driven by the column loading. At 11.0 g/L$_{resin}$, the yield could be maximized to 50%, but the productivity was modest (*i.e.* 7.13 g/L/h). On the other side, the latter could be maximized (*i.e.* 10.28 g/L/h) by increasing the load to 22.5 g/L$_{resin}$, which in turn led to a decrease in the maximum achievable yield to <40%. Still, this is a remarkable improvement compared to the central point of the scheme used to calibrate the model, suggesting the potential of this optimization strategy in improving the performances of the operation.

The yield of the process can be further boosted at comparable purity and productivity by MCSGP. We demonstrated that the process parameters maximizing the productivity identified by our model-based strategy provided excellent results in the design of an MCSGP operation reaching the steady state already at the third cycle and respecting the purity specification imposed, while increasing the product recovery from <40% to >96%. So overall, the model-based optimization developed in this work can serve as both a way for improving the existing processes based on single-column operations, improving either the yield or the productivity based on the most suitable target, or be exploited for a resource-effective design of innovative processes, both in the direction of process intensification.

It is finally worth highlighting that the approach reported herein is agnostic to the type of molecule to be purified, provided that its adsorption can be described by a Langmuir isotherm. Hence, this optimization strategy can be extended to the intensification of the chromatographic purification of many other molecules, where product-related impurities cannot be baseline separated.

Supporting Information: Electronic supplementary information is available at the publisher's website and reports analytical chromatogram for the crude mixture used in the experiments, the reproduction of a literature case, the chromatogram fitting for the single-component case, the estimated parameters, and the results of the first and second optimization steps.

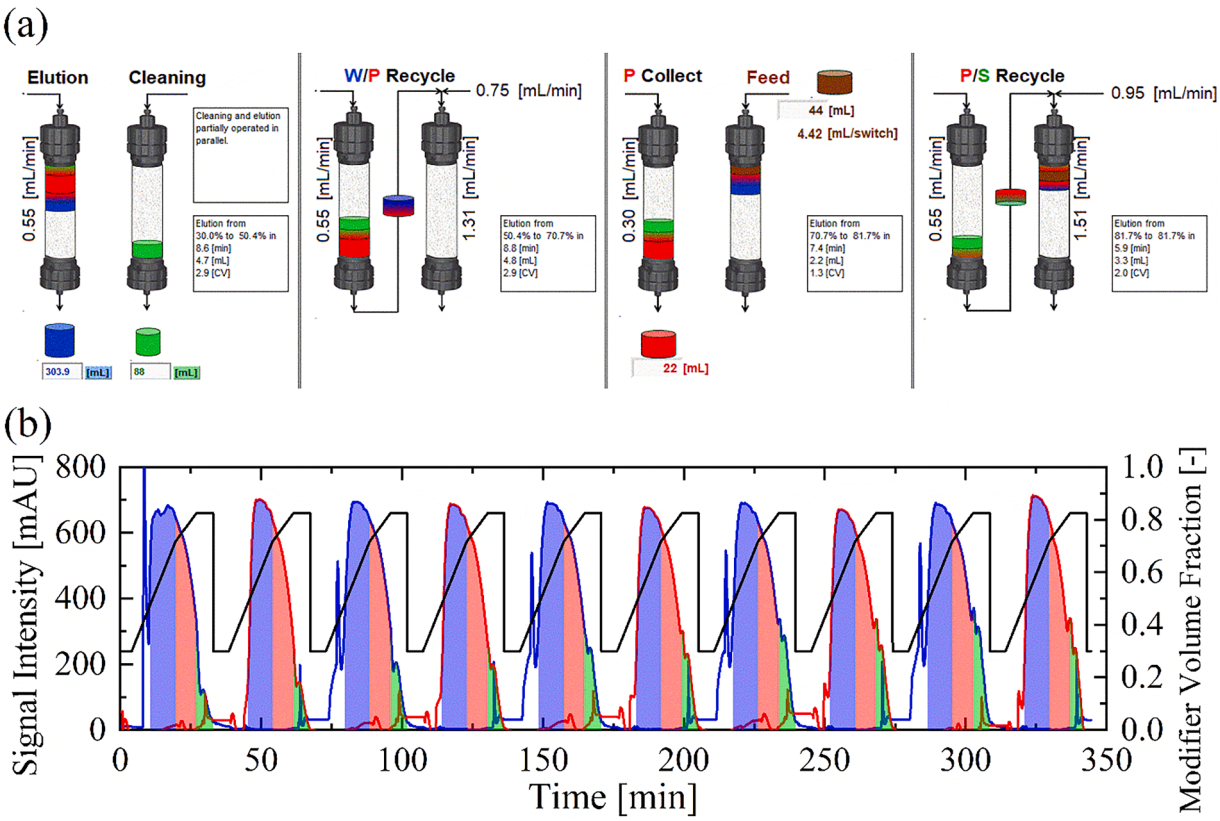

Fig. 6. (a) MCSGP process designed from the single-column operation maximizing the productivity. (b) UV signal during the 5 cycles of the MCSGP recorded by the detector at the outlet of the left column (blue line) and at the outlet of the right column (red line). The black line shows the volume fraction of the modifier, the blue region the W/P internal recycling, the red area the product collection window and the green region the P/S internal recycling.

Table 4
Characterization of the product pools collected during the 5 cycles of MCSGP compared to the one obtained in the single-column experiment maximizing the productivity.

Cycle #	W1 [g/L]	W2 [g/L]	P [g/L]	S [g/L]	Purity [%]	Yield [%]	Productivity [g/(L h)]
1	0.00	0.03	4.95	0.00	99.4	68.7	5.64
2	0.00	0.04	6.58	0.01	99.2	93.5	7.68
3	0.00	0.06	6.92	0.01	99.0	96.8	7.95
4	0.00	0.06	6.84	0.01	99.0	96.1	7.89
5	0.00	0.06	6.88	0.01	99.0	96.6	7.93
Batch reference	0.00	0.05	5.29	0.00	99.1	34.8	8.94

CRediT authorship contribution statement

Santiago Taguado Menza: Writing – original draft, Investigation, Data curation. **Rosella Prestia:** Validation. **Ismaele Fioretti:** Writing – review & editing, Validation. **Mattia Sponchioni:** Writing – review & editing, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors are grateful to Thomas Müller-Späth from YMC ChromaCon for the discussion and the revision of this work, as well as to YMC JP for the provision of the oligonucleotide mixture.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2024.465321.

References

[1] M. Catani, C. De Luca, J. Alcântara Medeiros Garcia, N. Manfredini, D. Perrone, E. Marchesi, R. Weldon, T. Müller-Späth, A. Cavazzini, M. Morbidelli, M. Sponchioni, Oligonucleotides: current trends and innovative applications in the synthesis, characterization, and purification, Biotechnol. J. 15 (8) (2020) 1900226.
[2] N. Dias, C.A. Stein, Oligonucleotides: basic concepts and mechanisms, Mol. Cancer Ther. 1 (5) (2002) 347–355.
[3] E.C. Smith, R. Zain, Therapeutic oligonucleotides: state of the art, Annu. Rev. Pharmacol. Toxicol. 59 (1) (2019) 605–630.

[4] S.M. Hammond, A. Aartsma-Rus, S. Alves, S.E. Borgos, R.A.M. Buijsen, R.W. J. Collin, G. Covello, M.A. Denti, L.R. Desviat, L. Echevarría, C. Foged, G. Gaina, V. Arechavala-Gomeza, Delivery of oligonucleotide-based therapeutics: challenges and opportunities, EMBo Mol. Med. 13 (4) (2021) e13243.
[5] R.G. Ingle, W.-J. Fang, An overview of the stability and delivery challenges of commercial nucleic acid therapeutics, Pharmaceutics. 15 (4) (2023) 1158.
[6] T.C. Roberts, R. Langer, M.J.A. Wood, Advances in oligonucleotide drug delivery, Nature Rev. Drug Discov. 19 (10) (2020) 673–694.
[7] J. Kim, C. Hu, C. Moufawad El Achkar, L.E. Black, J. Douville, A. Larson, M. K. Pendergast, et al., Patient-customized oligonucleotide therapy for a rare genetic disease, New Engl. J. Med. 381 (17) (2019) 1644–1652.
[8] A.I. Benjamin, F.D. Antia, S.B. Brueggemeier, L.J. Dioraziov, S.G. Koenig, M. E. Kopach, H. Lee, M. Olbrich, A.L. Watson, Sustainability challenges and opportunities in oligonucleotide manufacturing, J. Org. Chem. 86 (1) (2021) 49–61.
[9] S.R. Madabhushi, J. Gavin, S. Xu, C. Cutler, R. Chmielowski, W. Rayfield, et al., Quantitative assessment of environmental impact of biologics manufacturing using process mass intensity analysis, Biotechnol. Prog. 34 (2018) 1566–1573.
[10] I. Cedillo, B. Jarvis, T. Pavone, Designing commercial-scale oligonucleotide synthesis, Pharmaceutical Technol. 44 (2) (2020) 55.
[11] L.K. Shekhawat, A.S. Rathore, An overview of mechanistic modeling of liquid chromatography, Prepar. Biochem. Biotechnol. 49 (6) (2019) 623–638.
[12] I. Fioretti, T. Müller-Späth, R. Weldon, S. Vogg, M. Morbidelli, M. Sponchioni, Continuous countercurrent chromatographic twin-column purification of oligonucleotides: the role of the displacement effect, Biotechnol. Bioeng. 119 (7) (2022) 1861–1872.
[13] G. Guiochon, Preparative liquid chromatography, J. Chromatogr. A 965 (1–2) (2002) 129–161.
[14] M. Zoubaïr, E. Fallah, G. Guiochon, Prediction of a protein band profile in preparative reversed-phase gradient elution chromatography, Biotechnol. Bioeng. 39 (8) (1992) 877–885.
[15] F. Steinebach, N. Ulmer, L. Decker, L. Aumann, M. Morbidelli, Experimental design of a twin-column countercurrent gradient purification process, J. Chromatogr. A 1492 (2017) 19–26.
[16] T.K. Kim, C. Botti, J. Angelo, X. Xu, S. Ghose, Z.J. Li, M. Morbidelli, M. Sponchioni, Experimental design of the multicolumn countercurrent solvent gradient purification (MCSGP) unit for the separation of pegylated protein, Ind. Eng. Chem. Res. 60 (29) (2021) 10764–10776.
[17] G. Carta, A. Jungbauer, Protein Chromatography: Process Development and Scale-Up, Wiley-VCH, Weinheim, 2010.
[18] A. Osberghaus, S. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, J. Hubbuch, Optimizing a chromatographic three component separation: a comparison of mechanistic and empiric modeling approaches, J. Chromatogr. A 1237 (2012) 86–95.
[19] H. Narayanan, T. Seidler, M. Luna, M. Sokolov, M. Morbidelli, A. Buttè, Hybrid models for the simulation and prediction of chromatographic processes for protein capture, J. Chromatogr. A 1650 (2021) 462248 article no.
[20] K. Kobl, L. Nicoud, E. Nicoud, A. Watson, J. Andrews, E.A. Wilkinson, M. Shahid, C. McKay, B.I. Andrews, B.A. Omer, O. Narducci, E. Masson, S.H. Davies, T. Vandermeersch, Oligonucleotide purification by ion exchange chromatography: a step-by-step guide to process understanding, modeling, and simulation, Org. Process. Res. Dev. 28 (7) (2024) 2569–2589.
[21] M. Sturm, S. Quinten, C.G. Huber, O. Kohlbacher, A statistical learning approach to the modeling of chromatographic retention of oligonucleotides incorporating sequence and secondary structure data, Nucleic Acids Res. 35 (12) (2007) 4195–4202.
[22] M. Enmark, J. Häggström, J. Samuelsson, T. Fornstedt, Building machine-learning-based models for retention time and resolution predictions in ion pair chromatography of oligonucleotides, J. Chromatogr. A 1671 (2022) 462999 article no.
[23] L. Aumann, M. Morbidelli, A continuous multicolumn countercurrent solvent gradient purification (MCSGP) process, Biotechnol. Bioeng. 98 (2007) 1043–1055.
[24] C. De Luca, S. Felletti, G. Lievore, T. Chenet, M. Morbidelli, M. Sponchioni, A. Cavazzini, M. Catani, Modern trends in downstream processing of biotherapeutics through continuous chromatography: the potential of multicolumn countercurrent solvent gradient purification, Trends Analytical Chem. 132 (2020) 116051.
[25] H. Narayanan, M. Sponchioni, M. Morbidelli, Integration and digitalization in the manufacturing of therapeutic proteins, Chem. Eng. Sci. 248 (2022) 117159.
[26] T. Müller-Späth, L. Aumann, L. Melter, G. Ströhlein, M. Morbidelli, Chromatographic separation of three monoclonal antibody variants using multicolumn countercurrent solvent gradient purification (MCSGP), Biotechnol. Bioeng. 100 (6) (2008) 1166–1177.
[27] T. Müller-Späth, L. Aumann, G. Ströhlein, H. Kornmann, P. Valax, L. Delegrange, E. Charbaut, G. Baer, A. Lamproye, M. Jöhnck, M. Schulte, M. Morbidelli, Two step capture and purification of IgG2 using multicolumn countercurrent solvent gradient purification (MCSGP), Biotechnol. Bioeng. 107 (6) (2010) 974–984.
[28] T. Müller-Späth, M. Krättli, L. Aumann, G. Ströhlein, M. Morbidelli, Increasing the activity of monoclonal antibody therapeutics by continuous chromatography (MCSGP), Biotechnol. Bioeng. 107 (4) (2010) 652–662.
[29] C. De Luca, G. Lievore, D. Bozza, A. Buratti, A. Cavazzini, A. Ricci, M. Macis, W. Cabri, S. Felletti, M. Catani, Downstream processing of therapeutic peptides by means of preparative liquid chromatography, Molecules 26 (15) (2021) 4688 article no.
[30] C. De Luca, S. Felletti, D. Bozza, G. Lievore, M. Morbidelli, M. Sponchioni, A. Cavazzini, M. Catani, W. Cabri, M. Macis, A. Ricci, Process intensification for the purification of peptidomimetics: the case of icatibant through multicolumn countercurrent solvent gradient purification (MCSGP), Ind. Eng. Chem. Res. 60 (18) (2021) 6826–6834.
[31] T.K. Kim, B. Sechi, J.J. Romero Conde, J. Angelo, X. Xu, S. Ghose, M. Morbidelli, M. Sponchioni, Design and economic investigation of a Multicolumn Countercurrent Solvent Gradient Purification unit for the separation of an industrially relevant PEGylated protein, J. Chromatogr. A 1681 (2022) 463487 article no.
[32] T.K. Kim, A.A. Bham, I. Fioretti, J. Angelo, X. Xu, S. Ghose, M. Morbidelli, M. Sponchioni, Role of the gradient slope during the product internal recycling for the multicolumn countercurrent solvent gradient purification of PEGylated proteins, J. Chromatogr. A 1692 (2023) 463868 article no.
[33] R. Weldon, J. Lill, M. Olbrich, P. Schmidt, T. Müller-Späth, Purification of a GalNAc-cluster-conjugated oligonucleotide by reversed-phase twin-column continuous chromatography, J. Chromatogr. A 1663 (2022) 462734 article no.
[34] I. Fioretti, T. Müller-Späth, L. Aumann, M. Sponchioni, UV-based dynamic control improves the robustness of multicolumn countercurrent solvent gradient purification of oligonucleotides, Biotechnol. J. 19 (7) (2024) 2400170 article no.
[35] S. Qamar, J.N. Abbasi, S. Javeed, M. Shah, F.U. Khan, A. Seidel-Morgenstern, Analytical solutions and moment analysis of chromatographic models for rectangular pulse injections, J. Chromatogr. A 1315 (2013) 92–106.
[36] M.J. den Uijl, P.J. Schoenmakers, B.W.J. Pirok, M.R. van Bommel, Recent applications of retention modelling in liquid chromatography, J. Sep. Sci. 44 (1) (2021) 88–114.
[37] B. Medi, M. Amanullah, Application of a finite-volume method in the simulation of chromatographic systems: effects of flux limiters, Ind. Eng. Chem. Res. 50 (3) (2011) 1739–1748.
[38] S. Javeed, S. Qamar, A. Seidel-Morgenstern, G. Warnecke, Efficient and accurate numerical simulation of nonlinear chromatographic processes, Comput. Chem. Eng. 35 (11) (2011) 2294–2305.
[39] S. Perveen, A. Khan, A. Iqbal, S. Qamar, Simulations of liquid chromatography using two-dimensional non-equilibrium lumped kinetic model with bi-Langmuir isotherm, Chem. Eng. Res. Des. 181 (2022) 14–26.
[40] E. Rossi, M. Paloni, G. Storti and R. Rota, "Modeling dual reflux pressure swing adsorption processes: numerical solution based on the finite volume method," vol. 203, p. 173–185, 2019.
[41] W. Paszkowicz, Properties of a genetic algorithm equipped with a dynamic penalty function, Comput. Mater. Sci. 45 (1) (2009) 77–83.
[42] Ö. Yeniay, Penalty function methods for constrained optimization with genetic algorithms, Math. Comput. Appl. 10 (1) (2005) 45–56.
[43] M. Gen, R. Cheng, Survey of penalty techniques in genetic algorithms, in: Proceedings of the IEEE Conference on Evolutionary Computation, 1996, pp. 804–809.
[44] T. Alvarez-Segura, S. López-Ureña, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Multi-scale optimisation *vs.* genetic algorithms in the gradient separation of diuretics by reversed-phase liquid chromatography, J. Chromatogr. A 1609 (2020) 460427.
[45] C.A. Coello Coello, Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art, Comput. Methods Appl. Mech. Eng. 191 (11–12) (2002) 1245–1287.
[46] M.H. Ardakani, C. Nosengo, S. Felletti, M. Catani, A. Cavazzini, C. De Luca, H. Rezadoost, Enhancing the purification of crocin-I from saffron through the combination of multicolumn countercurrent chromatography and green solvents, Anal. Bioanal. Chem. 416 (2024) 2553–2564.
[47] F. Rischaway, D. Saleh, T. Hahn, S. Oelmeier, J. Spitz, S. Kluters, Good modeling practice for industrial chromatography: mechanistic modeling of ion exchange chromatography of a bispecific antibody, Comput. Chem. Eng. 130 (3) (2019) 106532 article no.
[48] K. Behere, S. Yoon, Chromatography bioseparation technologies and in-silico modelings for continuous production of biotherapeutics, J. Chromatogr. A 1627 (2020) 461376 article no.
[49] S.W. Benner, J.P. Welsh, M.A. Rauscher, J.M. Pollard, Prediction of lab and manufacturing scale chromatography performance using mini-columns and mechanistic modeling, J. Chromatogr. A 1593 (2019) 54–62.