Supplementary Information (SI) for Digital Discovery.
This journal is © The Royal Society of Chemistry 2026

Supplementary Information (SI) for: Spectroscopy-Assisted Bayesian Optimization for Efficient Inclusion Body Refolding

Florian Gisperg[a,b], Robert Klausser[a,b], Matthias Kierein[a], Eva Prada Brichtova[a,b], Mohamed Elshazsly[a,b], Julian Kopp[a,b], and Oliver Spadiut[a,b]

[a] Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, Vienna, Austria
[b] Christian Doppler Laboratory for Inclusion Body Processing 4.0, TU Wien, Vienna, Austria

Contents

1 Target protein sequence

The following is the amino acid sequence of the target single chain fragment (scFvM), as expressed with a 6x His-tag.

```
MQEQLVESGGGLVQPGGSLRLSCAASGFDFSSHWIYWVRQAPGKGLEWVSTIYTGSDSTYYATWAKGRFTISK
DNSKNTVYLQMNSLRAEDTAVYYCARDLGGSSSTSYISDLWGQGTLVTVSSGGGGSGGGGSGGGGSELVLTQS
PATLSLSPGERATLSCTLSSAHKTYSIAWYQQKPGQAPRYLIQLKSDGSYTKGTGVPARFSGSSSGADRTLTI
SSLEPEDFAVYYCSTDYATGYYVFGQGTKVEIKRHHHHHH
```

2 In-house scFvM standard generation

The internal scFvM standard used to quantify was produced from solubilized and refolded IBs. For each production run a 1 g aliquot of wet IBs was solubilized in 10 mL of solubilization buffer (8 Molar urea, 50 mM glycine, pH 10). After the pellet was dissolved, 1 M DTT stock was added to reach the final desired DTT concentration in the solubilizate (10 mM for the first production runs, adapted to 6.5 mM for later runs based on the optimizations). After centrifugation at

20,000 rcf, at 4 °C, for 20 min, the clarified solubilized protein was refolded by rapid dilution into a pre-cooled refolding buffer. For the initial standard production a dilution factor of 20 was used, with a refolding buffer consisting of 4 M urea, 50 mM glycine, and no GSSG, based on a standard protocol used in our lab. For later standard production runs, the dilution factor and refolding buffer composition was adjusted based on the latest optimization results.

After refolding at 10 °C over night, the pH was adjusted by rapid 1:2 dilution into a pre-cooled 100 mM citrate buffer with pH 4, resulting in a final pH of ≈ 6. As some precipitate was usually formed in this step, the pH adjusted solution was once again clarified by centrifugation (20,000 rcf, at 4 °C, for 20 min) after which the supernatant was purified.

Purification of the refolded protein was carried out using an Äkta pure system (Cytiva, Marlborough, Massachusetts, USA) equiped with a fraction collector. Two 1 mL Hi-Trap™ SP Sepharose FF columns were connected in series and equilibrated with ≈ 10 column volumes (CV) buffer A (20 mM phosphate buffer, pH 6.5). The full volume of refolded scFvM was loaded at 0.5 mL min^{-1} and subsequently washed with buffer A until the UV signal was back at the baseline. Subsequently the protein was eluted with a gradient of 0 % – 40 % buffer B (20 mM phosphate buffer, pH 6.5, 1 M NaCl) over 30 CV, followed by regenerating the column with 100 % buffer B over ≈ 10 CV. The chromatogram of an exemplary standard production run is shown in figure S1 with the pooled standard fraction highlighted in yellow. The pooled fractions were desalted by exchanging the buffer to buffer A using a PD-10 desalting column (Cytiva, Marlborough, Massachusetts, USA) using gravitational flow. Subsequently, the concentration of the purified and desalted scFvM standard was determined by absorbance at 280 nm with a NanoDrop One™ spectrometer (Thermo Fisher, Waltham, Massachusetts, US) setting the molecular weight to 27,630 Da and the extinction coefficient to 58,580 $\text{M}^{-1}\,\text{cm}^{-1}$.

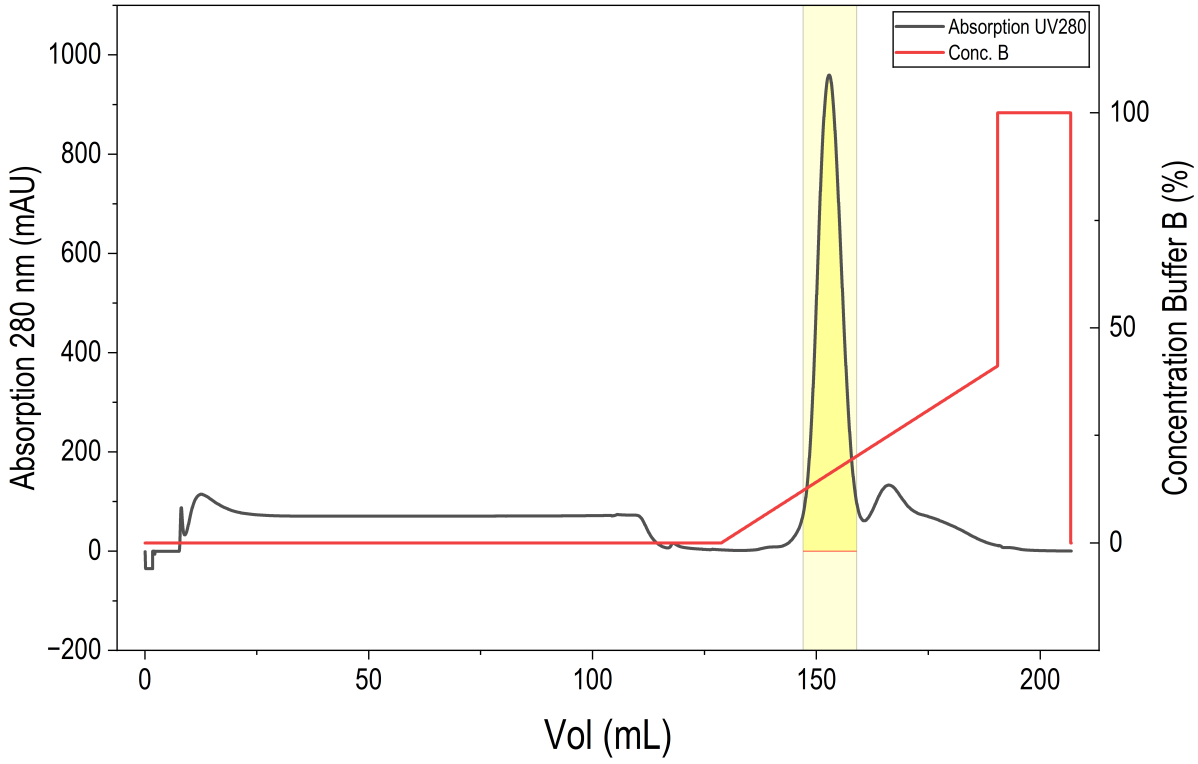

Figure S1: Chromatogram of an exemplary purification run for the generation of scFvM standard. Preparative cation exchange (CEX) was performed at pH 6.5 on Hi-Trap™ SP Sepharose FF resin using 20 mM phosphate buffer and eluting with a sodium chloride salt gradient (1 M NaCl in buffer B. The concentration of buffer B is shown as a red line, scaling on the right y-axis. The UV absorbtion signal at 280 nm is shown as a grey line, scaling on the left y-axis. The pooled fraction collected for use as a protein standard in this study is highlighted in yellow.

The desalted standard set to 1 g L^{-1} was checked for oligomerization by analytical size exclusion chromatography using a TSKgel SuperSW2000 column (Tosoh, Tokio, Japan) and isochratic elution (20 mM phosphate buffer, 250 mM NaCl; 0.3 mL min^{-1} flowrate). The single observed peak of the monomer is shown in Figure S2A. In Figure S2B the chromatographic profile of the standard compared to a refolded sample is shown for the HIC-HPLC protocol described in the main manuscript. Solubilized, and unfolded scFvM did not lead to any peak eluting off the HIC column, and was therefore assumed to be irreversibly adsorbed to the pre-column due to its high hydrophobicity. Only the main peak eluted at around 6.9 min (Figure S2B) was to quantify the native scFvM.

The purified standard was also confirmed to be the correct target protein by mass spectrometry. The target protein sequence was found as the top hit with 100 % sequence coverage (Figure S3). The full results of all identified peptides within the standard can be viewed in the data repository (10.5281/zenodo.18270843) with the sample "Fraction 1" being the desalted scFvM standard, the sample "Fraction 2" being the desalted fraction peak eluting immediately after the product from the preparative CEX column, and the other four samples being from a storage study unrelated to this manuscript.

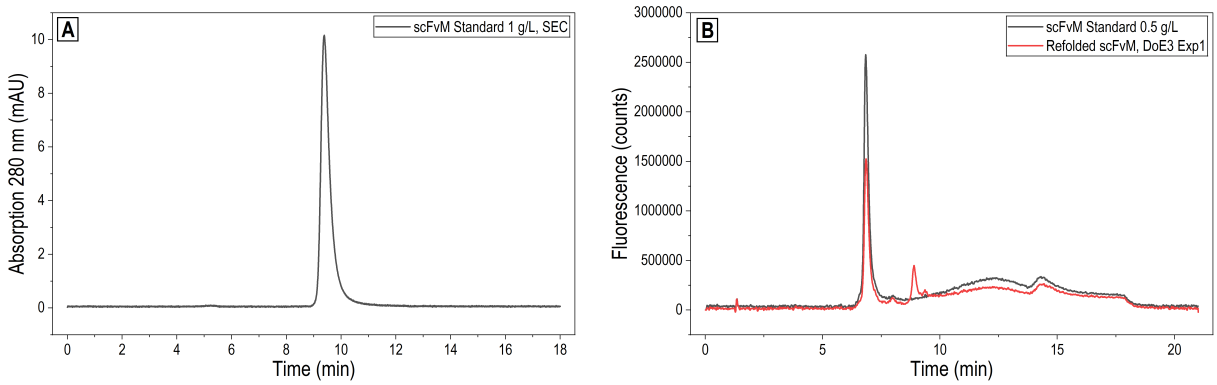

Figure S2: Exemplary HPLC chromatograms of the in-house scFvM standard used in this study. A: Size exclusion chromatogram of a 1 g L^{-1} scFvM standard, showing the lack of oligomers within the standard. B: Hydrophobic interaction chromatogram of the method used for quantification of the product monomer, comparing the in-house standard (grey) with a refolding sample of the DoE optimization (red).

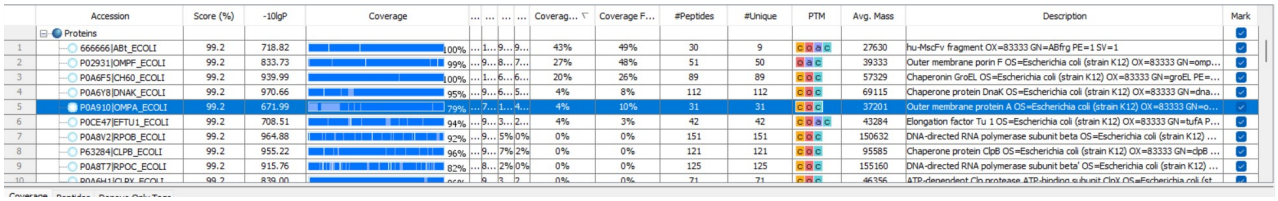

Figure S3: Screenshot of the protein Top-List for MS analysis. The target protein scFvM was found as the top-scoring protein with 100 % sequence coverage.

3 Cost estimation of refolding buffer components

Table S1: Estimation of per liter cost of refolding buffer components. As a cost basis we used the specific grades of chemicals we used in this study. In an industrial production setting, the cost of bulk urea would likely be even cheaper than assumed here.

Buffer component	Cost €L^{-1}	Carl Roth article number
Glycine	0.092	3187.7
Urea	5.09	2317.2
GSSG	58.92	6378.2

4 Detailed DoE results

While the most relevant outcomes of the DoE optimization were covered in the main manuscript, the individual contour plots of each model were omitted to aid readability. In this section, the contour plots, observed versus predicted plots, individual model coefficients and variable importance scores for all three DoEs are shown in Figures S4–S6.

5 Endpoint stability of AEW time profiles

Full emission spectra were acquired throughout each BO refolding experiment, but the optimization objectives were deliberately computed from the initial spectrum and the ∼17 h endpoint spectrum. This endpoint was selected as a practical refolding time based on prior experience

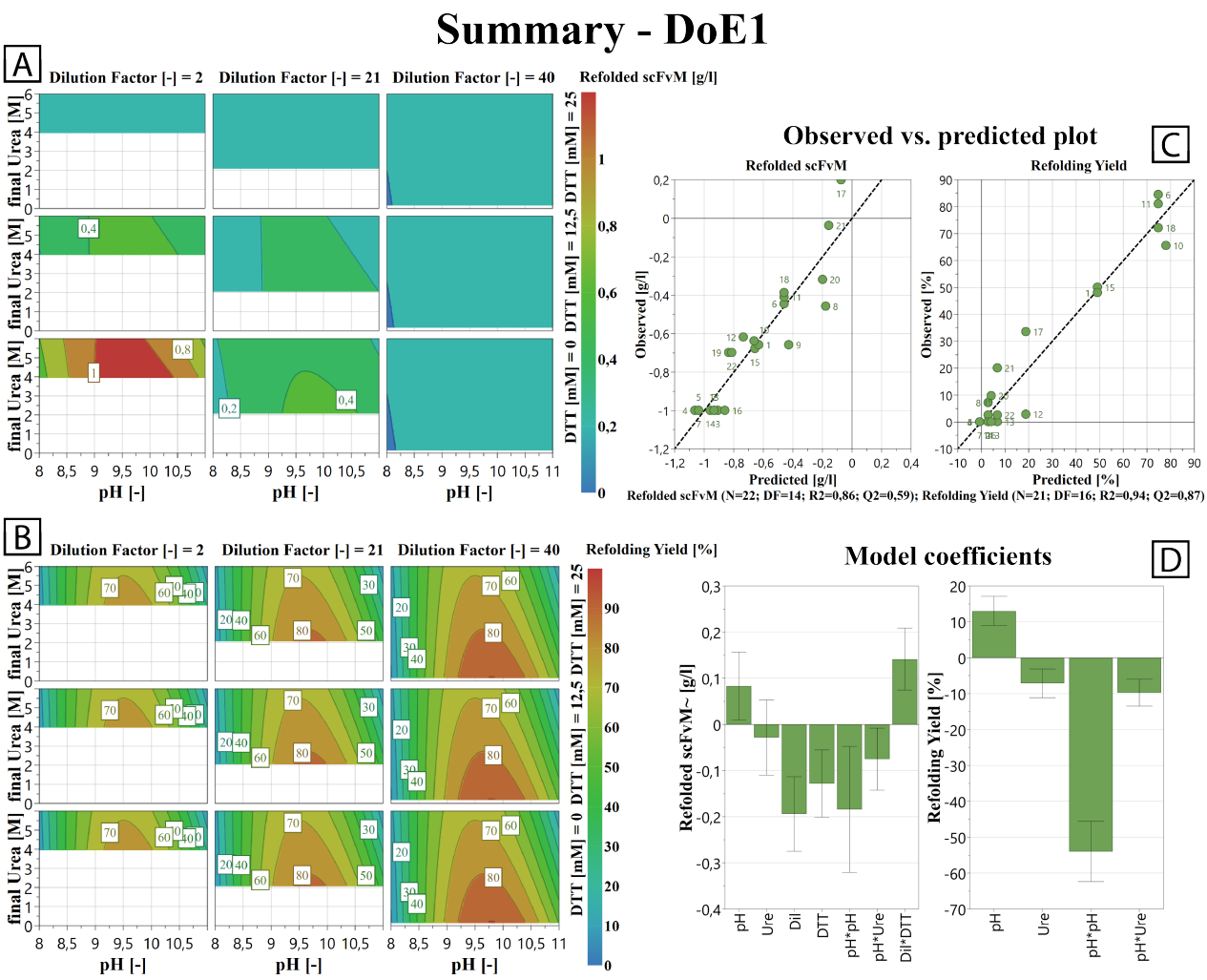

Figure S4: Detailed results of the initial screening DoE1. A: Contour plot of the correctly folded scFvM concentration after refolding. The the dependency on the final urea concentration and pH are shown as x and y axes respectively. Low medium and high levels of both the dilution factor and DTT concentrations are used as faceting variables, splitting the graph into a 3 x 3 panel grid. The response is color mapped consistently across all nine combinations of the faceting variables. B: Corresponding contour plot of the refolding yield. Despite the dilution factor and DTT concentration not being significant model terms, the faceting variables and axes were chosen analogous to the contour plot in A, to enable a direct comparison. C: Observed versus predicted plots for both response models. Negative $g\,L^{-1}$ values for the refolded scFvM are a result of the negative logarithmic transformation employed to achieve a better model fit. Below the plots, the total number of experiments (N), the residual degrees of freedom (DF), the coefficient of determination (R^2) and the cross validated coefficient of determination (Q^2) of the models are given. D: Bar charts of the model coefficients including 95 % confidence intervals shown as error bars.

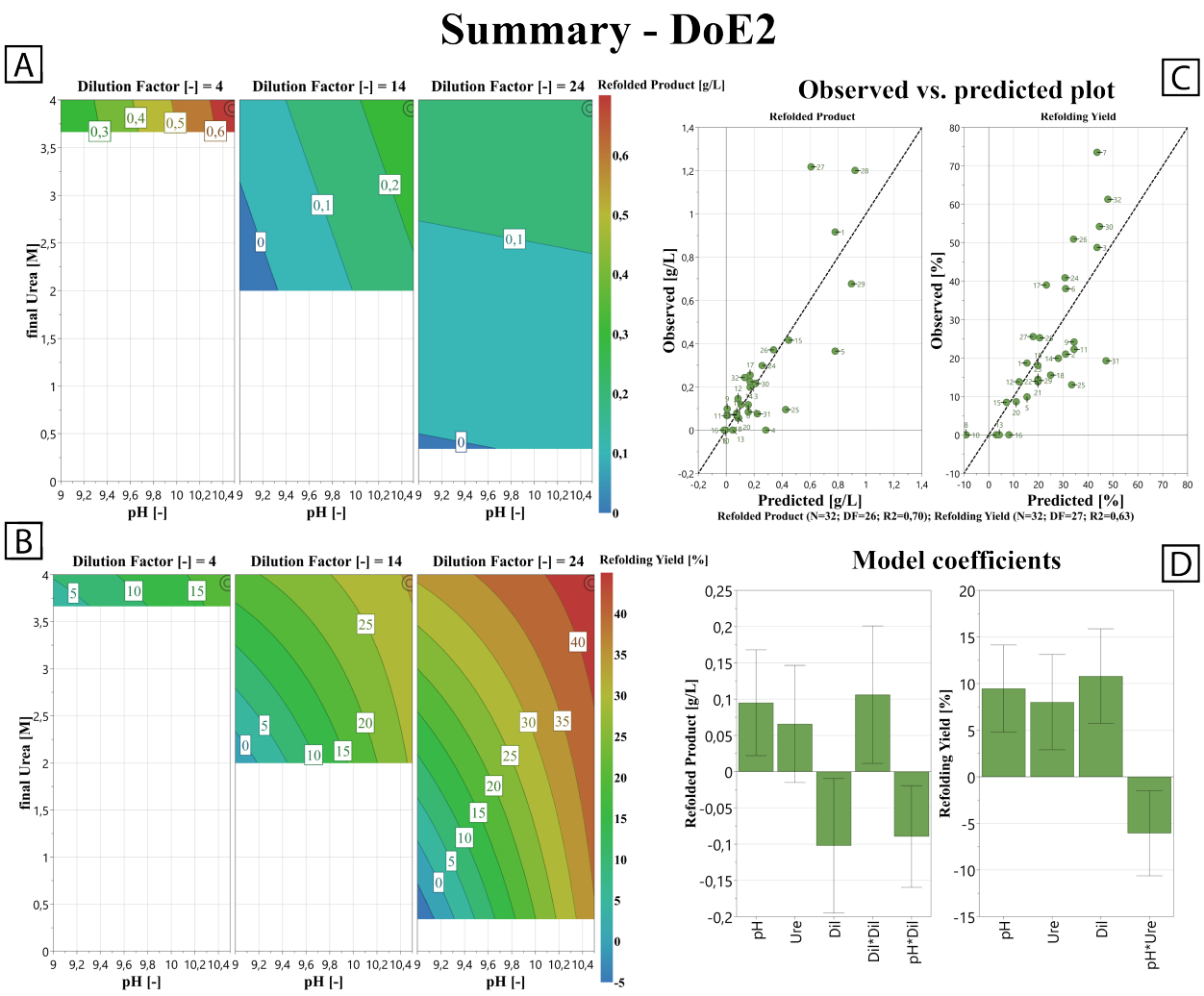

Figure S5: Detailed results of the second, optimization, DoE2. A: Contour plot of the correctly folded scFvM concentration after refolding. The the dependency on the final urea concentration and pH are shown as x and y axes respectively. Low medium and high levels of the dilution factor were used as a faceting variable. Dilution factor DF = 4 was used as the low value, because the physical constraints (a maximum final urea concentration of 4 M) would leave no visible design space at DF = 2. The response is color mapped consistently across all three panels. B: Corresponding contour plot of the refolding yield. Factors and panels are scaled equally to the contour plot for the refolded scFvM concentration. C: Observed versus predicted plots for both response models. Below the plots, the total number of experiments (N), the residual degrees of freedom (DF), the coefficient of determination (R^2) and the cross validated coefficient of determination (Q^2) of the models are given. D: Bar charts of the model coefficients including 95 % confidence intervals shown as error bars.

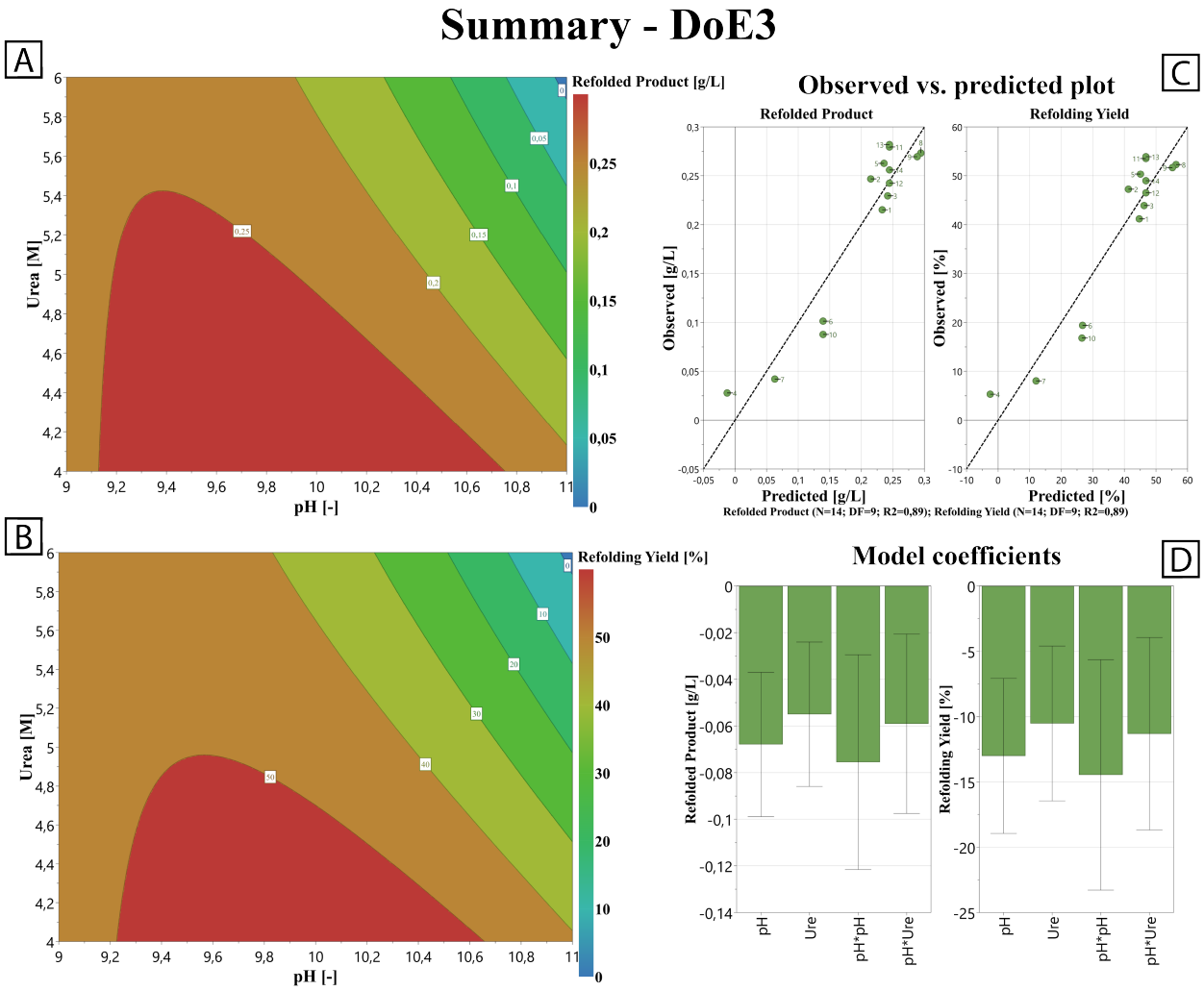

Figure S6: Detailed results of the final optimization DoE3. A: Contour plot of the correctly folded scFvM concentration after refolding. The the dependency on the final urea concentration and pH are shown as x and y axes respectively. B: Corresponding contour plot of the refolding yield. Factors are scaled equally to the contour plot for the refolded scFvM concentration. Due to the lack of variation of the dilution factor and DTT concentration, the models for refolded scFvM concentration and refolding yield are essentially linear transformations of each other. The subtle difference in the contour plots stems from a slightly different scaling of the color scale. C: Observed versus predicted plots for both response models. Below the plots, the total number of experiments (N), the residual degrees of freedom (DF), the coefficient of determination (R^2) and the cross validated coefficient of determination (Q^2) of the models are given. D: Bar charts of the model coefficients including 95 % confidence intervals shown as error bars.

with scFvM and defined as the maximum tolerable refolding time. Furthermore, it was practical from an experimental workflow point of view.

To verify that this endpoint was representative, we re-analysed the AEW time profiles from all 25 BO experiments. For each run, the recorded time point closest to 17 h was identified and compared with the final recorded AEW value. The nearest recorded 17 h time point ranged from 16.981 h – 17.028 h. Final recorded times ranged from 17.877 h – 23.263 h, reflecting minor differences in run duration after the planned overnight endpoint. Across all experiments, the change between AEW at ∼17 h and the final recorded AEW had a median of −0.009 nm and a maximum absolute value of 0.181 nm. Similarly, the median AEW change during the final recorded hour was 0.005 nm, with a maximum absolute change of 0.181 nm. These changes were small compared with the total refolding-induced AEW shifts used for optimization, supporting the use of the 17 h endpoint for the BO objectives.

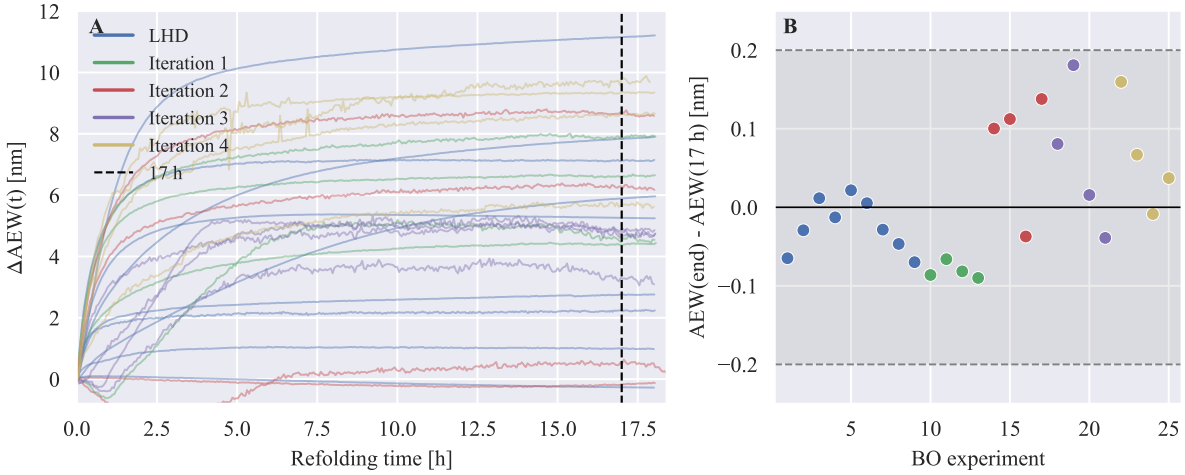

Figure S7: Endpoint stability of AEW during BO refolding experiments. (A) Time profiles of the refolding-induced AEW shift, $\Delta\mathrm{AEW}(t) = \mathrm{AEW}(0) - \mathrm{AEW}(t)$, for all 25 BO experiments. The dashed vertical line indicates the planned 17 h endpoint used for objective calculation. (B) Difference between the final recorded AEW value and the AEW value closest to 17 h. The shaded region marks ±0.2 nm.

The full time-resolved spectra may also contain kinetic information, such as initial folding rates, time-to-plateau estimates, or integrated AEW shifts. These kinetic descriptors were not used as BO objectives in the present study because the objective was to optimize an endpoint refolding process and because the available HPLC validation was likewise endpoint-based. Future work could incorporate kinetic objectives.

6 GSSG omission experiment at the BO-optimized condition

To test whether the high GSSG concentration identified by BO was relevant under the final operating point, the BO-optimized process was repeated without GSSG while keeping the remaining setpoints unchanged. Omitting GSSG approximately halved the refolding yield compared with the Pareto-optimal BO condition and changed the time-resolved fluorescence response (Figure S8). The increase in ΔAEW was delayed and slower without GSSG, indicating delayed productive refolding. A possible explanation is that residual DTT carried over from the solubilized protein must first be oxidized before disulfide-bond formation can proceed efficiently. This effect is expected to be strongest at low dilution factors, where reducing-agent carry-over from the solubilizate is highest.

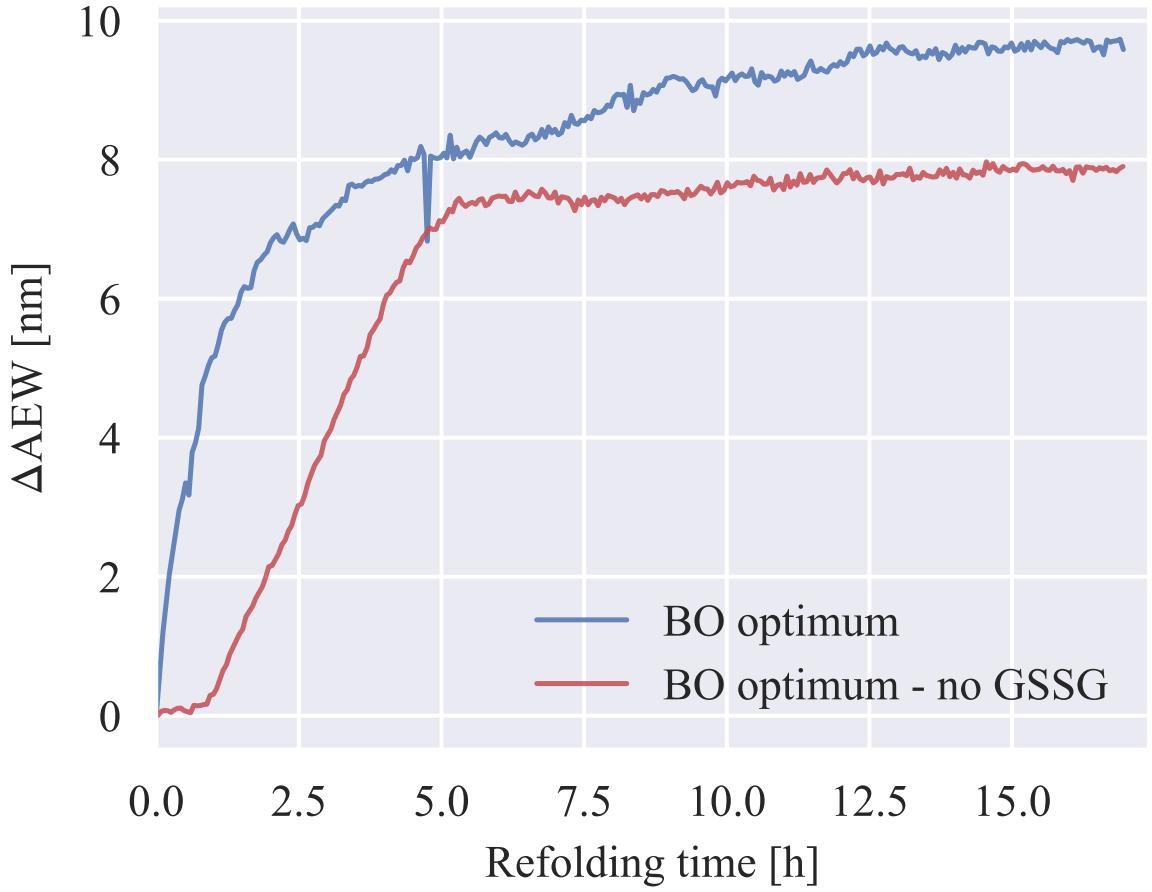

Figure S8: Comparison of time-resolved refolding-induced AEW shifts for the BO-optimized Pareto-optimal condition and the same process setpoints without GSSG. Removing GSSG delayed the ΔAEW increase and slowed the fluorescence response.

7 Quantitative proxy-to-HPLC validation metrics

The spectroscopy-derived objectives were intended as rapid, standard-free optimization signals for BO, rather than as calibrated replacements for HPLC quantification. To make this distinction explicit, we quantified proxy-HPLC accuracy using two complementary datasets. First, an independent validation dataset ($n = 18$), not used for BO training, was used to evaluate the relationship between ΔAEW and HPLC-derived refolding yield. Second, the 25 experiments from the BO campaign were used for a post hoc chromatographic evaluation of the proxy-guided search. The HPLC values from these experiments were measured for validation and were not used by the BO algorithm to propose candidates.

Prediction errors were estimated using leave-one-out cross-validation (LOOCV) with a simple linear calibration between the respective proxy and HPLC response. The results are summarized in Table S2. For the independent validation dataset, ΔAEW correlated with HPLC-derived refolding yield with $R^2 = 0.712$, and LOOCV gave a mean absolute error (MAE) of 12.83 percentage points and a root-mean-square error (RMSE) of 16.50 percentage points. For the BO campaign, ΔAEW showed similar agreement with HPLC-derived yield ($R^2 = 0.724$; MAE 11.37 percentage points; RMSE 15.26 percentage points). The volumetric proxy P_{proxy} correlated with HPLC-derived titer with $R^2 = 0.675$, corresponding to an LOOCV MAE of 197.92 $\mathrm{mg\,L^{-1}}$ and RMSE of 285.51 $\mathrm{mg\,L^{-1}}$.

Table S2: Quantitative proxy-to-HPLC validation metrics. Prediction errors were calculated by leave-one-out cross-validation using a linear calibration between the indicated proxy and HPLC response.

Dataset	Proxy / HPLC response	n	Pearson r	R^2	LOOCV MAE	LOOCV RMSE
Independent validation dataset	ΔAEW / yield	18	0.844	0.712	12.83 pp	16.50 pp
BO campaign HPLC evaluation	ΔAEW / yield	25	0.851	0.724	11.37 pp	15.26 pp
BO campaign HPLC evaluation	P_{proxy} / titer	25	0.822	0.675	197.92 mg L^{-1}	285.51 mg L^{-1}

These error magnitudes show that the spectroscopy-derived objectives provide useful directional information for experimental allocation, but they are not sufficiently accurate to replace chromatographic quantification when absolute yield or titer values are required. This is particularly important for P_{proxy}, which is calculated as

$$P_{\mathrm{proxy}} = \Delta\mathrm{AEW}\,\frac{C_{\mathrm{sol}}}{DF}.$$

Consequently, uncertainty in ΔAEW, SDS-PAGE-derived solubilized protein concentration (C_{sol}), and the dilution factor (DF) propagates multiplicatively. Therefore, accuracy is lower for the the concentration proxy. Nonetheless, both ΔAEW and P_{proxy} provided directional information robust enough to guide optimization successfully.

Error propagation in the volumetric titer proxy

Because P_{proxy} is a product and quotient of measured or set quantities, uncertainty is expected to propagate approximately through relative errors:

$$\left(\frac{\sigma_{P_{\mathrm{proxy}}}}{P_{\mathrm{proxy}}}\right)^2 \approx \left(\frac{\sigma_{\Delta\mathrm{AEW}}}{\Delta\mathrm{AEW}}\right)^2 + \left(\frac{\sigma_{C_{\mathrm{sol}}}}{C_{\mathrm{sol}}}\right)^2 + \left(\frac{\sigma_{DF}}{DF}\right)^2.$$

The first two contributions arise from fluorescence spectroscopy and SDS-PAGE quantification, respectively. The dilution factor is an experimentally set value and is therefore expected to contribute less uncertainty than the measured quantities, but it still enters the proxy inversely. This multiplicative structure explains why P_{proxy} is more sensitive to measurement variation than ΔAEW alone and why extreme proxy values can overestimate the chromatographic titer. During BO, this propagated uncertainty was not supplied as point-specific heteroscedastic noise to qNEHVI. Each objective was modeled with a learnable homoscedastic GP noise term, and the propagated-error discussion is used here to clarify the expected limitations of the proxy.

8 Final validation and replicate statistics

The final validation experiments were performed as independent process triplicates. Each replicate started from a separate IB aliquot and included independent solubilization, refolding, sample preparation, and HIC-HPLC analysis. The raw replicate values are shown in Table S3. Table S4 summarizes the primary HIC-HPLC validation results with 95% confidence intervals.

Table S3: Raw independent process replicate results for the final DoE and BO validation experiments. Titer and yield are based on the primary HIC-HPLC quantification.

Operating point	Replicate	Solubilized scFvM / $g\,L^{-1}$	DF	Titer / $g\,L^{-1}$	Yield / %
DoE final	A	6.262	11.39	0.340	61.77
DoE final	B	7.442	11.39	0.380	58.08
DoE final	C	6.827	11.39	0.385	64.21
BO final	A	6.680	3.13	1.232	57.74
BO final	B	7.039	3.13	1.311	58.30
BO final	C	6.976	3.13	1.341	60.17

Table S4: Independent triplicate validation statistics for the final DoE and BO operating points. Titer and yield are based on the primary HIC-HPLC quantification. The 95% confidence intervals were calculated from the triplicate process replicates using the t-distribution with $n-1$ degrees of freedom.

Operating point	n	Titer / $g\,L^{-1}$	95% CI / $g\,L^{-1}$	Yield / %	95% CI / %
DoE final	3	0.368 ± 0.025	0.307 – 0.429	61.35 ± 3.08	53.69 – 69.01
BO final	3	1.295 ± 0.056	1.155 – 1.434	58.74 ± 1.27	55.58 – 61.90

Welch two-sample t-tests comparisons of the independent triplicates showed that the BO condition produced a substantially higher refolded scFvM titer than the final DoE condition (difference $0.927\,g\,L^{-1}$; 95% CI $0.808\,g\,L^{-1}$ – $1.046\,g\,L^{-1}$; $p = 2.24 \times 10^{-4}$). In contrast, the refolding yield was not significantly different between the two validated operating points (difference -2.61 percentage points; 95% CI -9.21 to 3.98 percentage points; $p = 0.278$). Thus, within the resolution of the triplicate process validation, the BO condition increased titer while maintaining comparable refolding yield.

9 Software and Modeling Details of the Bayesian Optimization Implementation

Bayesian optimization (BO) was implemented in Python using `BoTorch` and `GPyTorch`, which provide Gaussian process (GP) surrogate modeling and acquisition-based experimental design. The complete implementation (data preprocessing, GP training, acquisition optimization, and constraint handling) is available in a public code repository.

Framework and Environment

Table S5 summarizes the software environment used for all BO computations.

Table S5: Framework and environment information used for Bayesian optimization.

Component	Version / Details
Python	3.13.7
BoTorch	0.12.0
GPyTorch	1.13
PyTorch	2.5.1
Operating system	Windows

Gaussian Process Surrogate Model

The two optimization objectives (ΔAEW and P_{proxy}) were modeled independently using separate single-task Gaussian process models. Each GP used a Matérn 5/2 kernel with automatic relevance determination (ARD) to capture nonlinear and anisotropic dependencies across the five-dimensional design space.

Input parameters were linearly scaled to the unit hypercube based on their predefined physical bounds, i.e., each parameter was mapped to $[0, 1]$. Objective values were standardized to zero mean and unit variance prior to model fitting.

Kernel hyperparameters were learned by maximizing the exact marginal log-likelihood using gradient-based optimization with the Adam optimizer for a fixed number of iterations. A Gaussian likelihood (`gpytorch.likelihoods.GaussianLikelihood`) with a learnable noise parameter was used; the noise was initialized to 0.05 and optimized jointly with the GP hyperparameters. The noise parameter is constrained to be positive by GPyTorch's default constraint. No additional minimum noise floor was applied.

The learnable likelihood noise was estimated from the full training dataset during marginal-likelihood optimization. The GP covariance structure allows the model to distinguish smooth latent-function variation across nearby design points from residual observation noise. In this implementation, the fitted noise parameter should therefore be interpreted as an objective-level residual noise estimate that includes spectroscopic measurement variation, uncertainty in SDS-PAGE-derived solubilized protein concentration, dilution and pipetting variability, proxy uncertainty, and unmodeled process variability.

For multi-objective optimization, the independently trained single-task GPs were combined using BoTorch's `ModelListGP`, enabling joint posterior sampling for the acquisition function while retaining independent likelihoods/noise models per objective.

Table S6: Gaussian process model configuration.

Parameter	Value
Kernel type	Matérn 5/2 (ARD)
Likelihood	Gaussian (learnable noise; init 0.05; positive constraint)
Input scaling	Linear scaling to $[0, 1]^d$
Output standardization	Zero mean, unit variance
Optimizer	Adam
Training iterations	1000

Model Validation

To assess surrogate prediction accuracy and guard against overfitting in the low-data regime, leave-one-out cross-validation (LOOCV) was performed for each objective after every model update. Predictive performance was evaluated using the coefficient of determination (R^2) and range-normalized root-mean-square error (nRMSE). LOOCV was used here to assess point-prediction performance.

Bayesian Optimization Strategy

The optimization problem was formulated as a multi-objective task aiming to simultaneously maximize the spectral shift (ΔAEW) and the volumetric productivity proxy (P_{proxy}). Batch Bayesian optimization was conducted using the q-noisy expected hypervolume improvement (qNEHVI) acquisition function.

An initial design consisting of nine experiments was generated using a Latin hypercube design. Subsequently, four BO iterations were performed, each proposing a batch of four new

Table S7: Averaged LOOCV validation results (nRMSE) by iteration. Training size denotes the number of points used in each LOOCV fold (i.e., total data minus the held-out point).

Iteration	Training size	ΔAEW nRMSE	P_{proxy} nRMSE
0 (LHS)	8	0.261	0.367
1	12	0.278	0.180
2	16	0.254	0.157
3	20	0.198	0.113
4	24	0.240	0.121

experimental conditions ($q = 4$). After each iteration, both GP models were retrained using all accumulated data.

The initial Latin hypercube design was selected as a pragmatic space-filling initialization strategy for the five-dimensional continuous design space. The choice of initial design size and the allocation between initial exploration and subsequent BO samples is an important practical decision in BO, but there is no generally accepted optimum because the appropriate balance depends on the dimensionality, smoothness, noise level, and practical constraints of the specific experimental workflow. Published experimental BO studies therefore use different initial designs and initial/BO sample allocations depending on the application [1–3]. As a pragmatic lower-bound heuristic, at least $d + 1$ initial samples can be used for a d-dimensional design space. The nine-point design used here corresponds to 1.8 initial samples per design variable and was used as a practical compromise between broad initial coverage and model-guided BO iterations. A systematic comparison of alternative initial designs and initial/BO sample allocations was outside the scope of the present experimental demonstration.

Hypervolume-based optimization used a fixed reference point of $[0, 0]$ in objective space.

Table S8: Bayesian optimization overview parameters.

Parameter	Value
Initial design	Latin hypercube
Initial samples	9
Acquisition function	qNEHVI
Batch size (q)	4
Number of BO iterations	4
Total experiments	25
Reference point (hypervolume)	[0, 0]

Acquisition Optimization Settings

The qNEHVI acquisition function was optimized using multi-start gradient-based optimization with Monte Carlo sampling. Table S9 summarizes the parameters used for acquisition optimization.

As stated above for the initial design, testing the sensitivity of the full experimental optimization trajectory to different initial designs or random seeds was outside the scope of this study, because it would have required several independent wet-lab BO campaigns. At the same time, the acquisition optimization itself was not based on a single local optimizer start. qNEHVI was estimated using Sobol quasi-Monte Carlo sampling with 2048 MC samples, and BoTorch's `optimize_acqf` was run with 2048 raw samples and 200 optimization restarts for each qNEHVI step. This improves the numerical robustness of acquisition maximization and reduces sensitivity to local optimizer initialization.

Table S9: Acquisition optimization settings used for qNEHVI.

Parameter	Value	Description
MC_SAMPLES	2048	Monte Carlo samples for acquisition evaluation
NUM_RESTARTS	200	Optimization restarts for acquisition maximization
RAW_SAMPLES	2048	Raw samples for restart initialization
Sequential optimization	True	Sequential candidate generation (`sequential=True`)
Batch size (q)	4	Candidates proposed per iteration
Random seed	0	`torch.manual_seed(0)`
maxiter	200	Maximum iterations per restart
batch_limit	5	Batch limit for optimizer

Physical Constraints and Candidate Post-processing

To ensure physically feasible buffer compositions, a constraint linking the final urea concentration and the dilution factor was enforced. Candidate points violating this constraint were corrected using an iterative post-processing strategy that alternated between decreasing the target urea concentration (0.1 M steps) and increasing the dilution factor (0.5 increments) until feasibility was restored or a predefined maximum number of adjustment steps was reached. These corrections were small relative to the tested factor ranges and ensured that all proposed experiments could be executed without violating material or formulation constraints.

Three BO candidates were affected by feasibility handling and lay at or near the active urea/dilution boundary (Table S10). These were iteration 1 candidate 3, iteration 3 candidate 4, and iteration 4 candidate 2. At the lower dilution-factor bound of DF = 2.00, the corresponding feasibility boundary is a final urea concentration of 4.00 M. Thus, one adjusted condition lay exactly on the boundary, while the other two were slightly inside the feasible region.

Table S10: BO candidates affected by urea/dilution feasibility handling.

Global row	BO iteration	Candidate	Dilution factor	Final urea / M
12	1	3	2.00	4.00
21	3	4	2.00	4.15
23	4	2	2.00	4.11

Because post-hoc repair moves a candidate from the originally proposed acquisition point to an executable point, the acquisition value of the repaired point is not necessarily identical to that of the unrepaired proposal. This can introduce ranking bias relative to direct constrained acquisition optimization or pre-acquisition repair strategies. Directly embedding the constraint into the acquisition calculation directly generally represents a more elegant approach and will be implemented in the future. Nevertheless, all experimentally executed conditions were feasible, and solutions near active physical constraints are common in constrained optimization when the productive operating region lies close to the boundary.

Reproducibility and Availability

The complete implementation, including configuration files, GP training routines, acquisition optimization, and constraint handling, is available in a public repository. In addition, the full dataset used for BO (all experimental conditions and objective values) is provided as a separate data file.

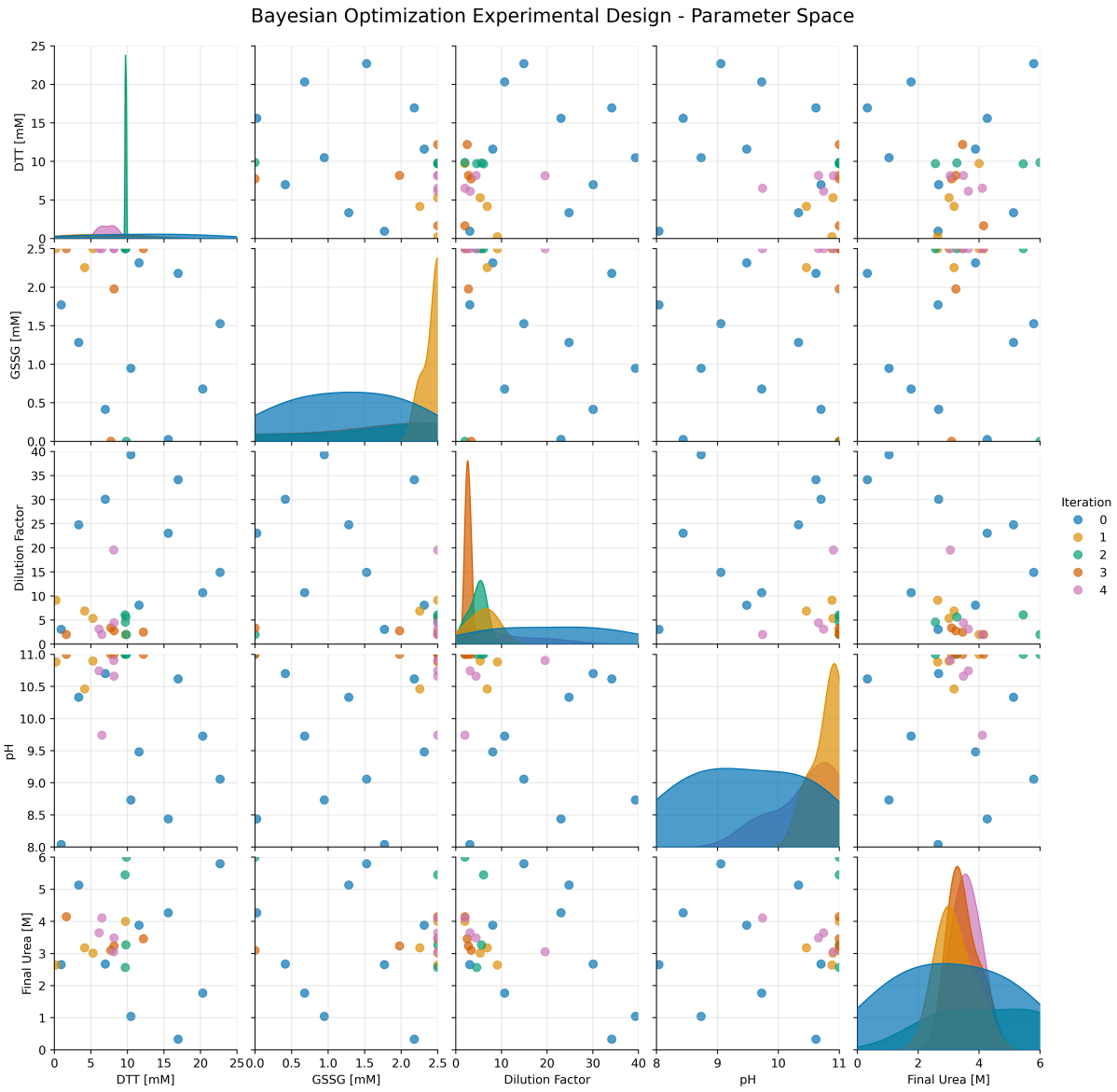

Figure S9: Pairwise distribution of experimental conditions evaluated during Bayesian optimization.

References

[1] M. Kondo, H. D. P. Wathsala, K. Ishikawa, D. Yamashita, T. Miyazaki, Y. Ohno, H. Sasai, T. Washio and S. Takizawa, *Molecules*, 2023, **28**, 5180.

[2] J. Bader, H. Narayanan, P. Arosio and J.-C. Leroux, *European Journal of Pharmaceutics and Biopharmaceutics*, 2023, **182**, 103–114.

[3] S. S. Rosa, D. Nunes, L. Antunes, D. M. F. Prazeres, M. P. C. Marques and A. M. Azevedo, *Biotechnology and Bioengineering*, 2022, **119**, 3127–3139.