Electronic Supplementary Material (ESI) for Reaction Chemistry & Engineering.

Multi-objective Bayesian optimisation using *q*-Noisy Expected Hypervolume Improvement (*q*NEHVI) for Schotten-Baumann reaction

Jiyizhe Zhang,[a] Naoto Sugisawa,[a,b] Kobi C Felton,[a] Shinichiro Fuse[b] and Alexei A Lapkin*[a,c,d]

[a]Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB3 0AS, UK. E-mail: aal35@cam.ac.uk

[b]Department of Basic Medicinal Sciences, Graduate School of Pharmaceutical Sciences, Nagoya University, Nagoya, 464-8601, Japan.

[c]Innovation Centre in Digital Molecular Technologies, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.

[d]Cambridge Centre for Advanced Research and Education in Singapore (CARES Ltd), #05-05 CREATE Tower, 1 Create Way, 138602, Singapore.

Supplementary Information

Experimental setup

A PTFE T-shape mixer (inner diameter: 0.5 mm) was connected to Teflon tubings (Vapourtec R-Series Tubing Kit) and was immersed in a water bath. The solutions were injected independently into T-shape mixer using TriContinent syringe pumps and reacted at 25 °C.

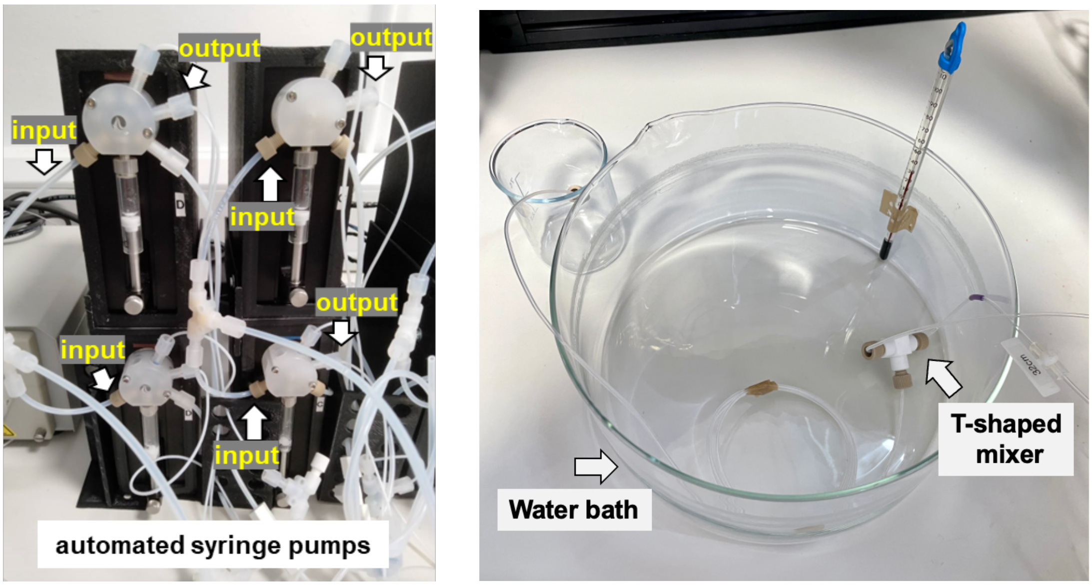


Figure S1. Experimental setup for synthesis of *N*-benzylacetamide (**4**).

Calibration curve of *N*-benzylacetamide

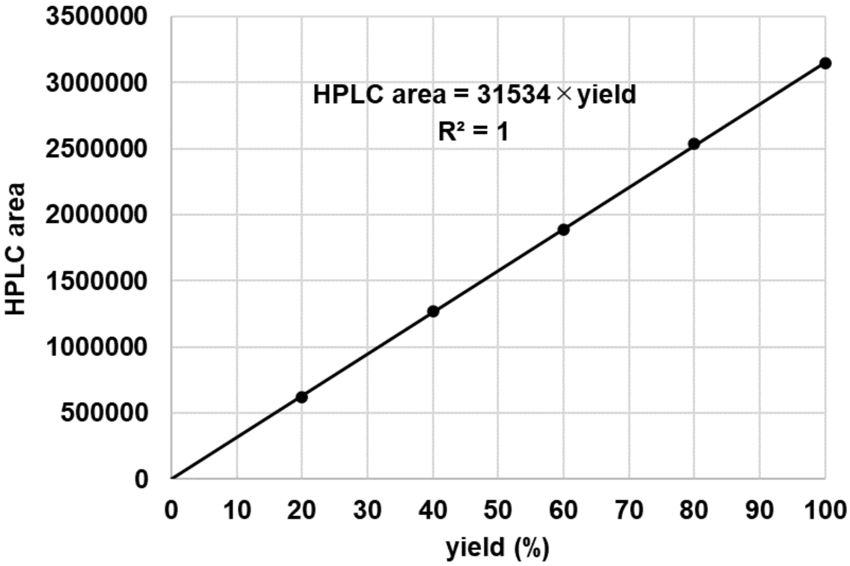


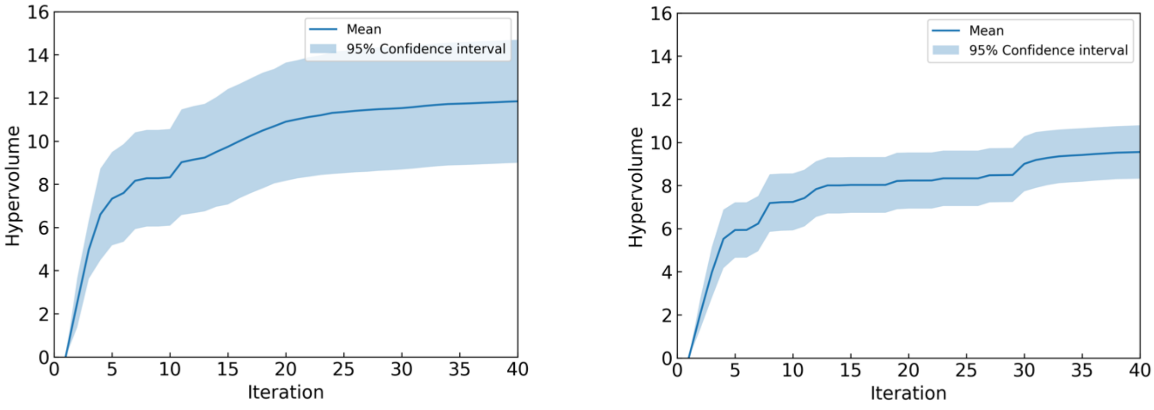
Figure S2. Calibration curve of *N*-benzylacetamide (**4**).

Initial experiments design and reaction simulator

MaxPro criterion[1] was used to generate initial design for the mixed variable space. MaxPro is one of few the optimal design criterions in literature that can incorporate continuous and categorical factors and shows good space-filling properties. [2] Previous studies have recommended 10 training experiments per continuous variable, [3, 4] but experimental applications for two-objective reaction optimization with four continuous variables[5, 6] have showed that 5 experiments per variable were effective enough. Given this information, we set initial experiments to be 20 for our case with two continuous variables and two categorical variables.

In addition, we tested the influence of initial design points on optimization results *in silico*. For comparison, we generated 20 different initial designs with 12 sampling points and 28 sampling points, respectively, and set simulation budget to be 40 steps in total.

Figure S3 (a) and (b) shows hypervolume improvement for initial 12 and 28 sampling points. Compared to hypervolume result in the manuscript Figure 2(a), after the initial design, hypervolume shows relatively larger improvement for 12 initial sampling points compared to 20 and 28 initial points. However, variance of the simulation results for 12 initial sampling points is also large, where 2/20 of the simulation results did not reach true Pareto front, while all simulation results for 20 and 28 initial sampling points did. Therefore, 20 initial sampling point was a reasonable choice to have a good space-filling and at the same time not wasting many experiments on initial design.



(a) (b)

Figure S3. Hypervolume improvement of different initial sampling points generated by MaxPro. (a) 12 initial sampling points. (b) 28 initial sampling points.

Two optimisation objectives space-time-yield (STY) and E-factor were calculated as follows:

$$\text{STY (g L}^{-1}\text{ s}^{-1}) = \frac{c_{\text{product}}}{\tau_{\text{res}}}$$

$$\text{E-factor} = \frac{m_{\text{waste}}}{m_{\text{product}}}$$

c_{product} = Desired concentration of **4** (0.15 M) × molecular weight of **4** (149.19 g/mol) × (yield/100) (g/L)

τ_{res} = Residence time (s)

m_{product} = Total mass of product (g)

m_{waste} = Total mass of waste (g) (Total reagents mass − m_{product})

Table S1. Initial 20 experiments using MaxPro.

entry	electrophile	X equiv	solvent	flow rate [mL/min]	Yield of **4** [%]	STY [g L^{-1} s^{-1}]	E-factor
1	Acetic anhydride	1.4	toluene	3.21	85.3	4.08	1.41
2	Acetic anhydride	1.5	EtOAc	5.38	93.2	7.48	1.30
3	Acetic anhydride	1.1	THF	1.65	82.0	2.02	1.15
4	Acetic anhydride	1.2	MeCN	6.00	90.3	8.09	1.06
5	Acetic anhydride	1.3	toluene	4.76	83.3	5.92	1.35
6	Acetic anhydride	1.1	EtOAc	3.52	90.9	4.78	0.94
7	Acetic anhydride	1.3	THF	4.14	92.2	5.70	1.12
8	Acetic anhydride	1.2	MeCN	0.72	78.7	0.84	1.37
9	Acetic anhydride	1.3	toluene	0.10	96.4	0.14	1.03
10	Acetic anhydride	1.4	EtOAc	1.34	98.2	1.96	1.09
11	Acetyl chloride	1.2	THF	2.58	73.1	2.81	1.29
12	Acetyl chloride	1.1	MeCN	2.89	76.6	3.30	1.08
13	Acetyl chloride	1.4	toluene	5.69	83.7	7.11	1.19
14	Acetyl chloride	1.2	EtOAc	4.45	90.9	6.04	0.84
15	Acetyl chloride	1.4	THF	0.41	43.3	0.27	3.22
16	Acetyl chloride	1.5	MeCN	1.96	82.9	2.43	1.30
17	Acetyl chloride	1.1	toluene	1.03	87.3	1.34	0.82
18	Acetyl chloride	1.0	EtOAc	2.27	91.9	3.11	0.65
19	Acetyl chloride	1.0	THF	5.07	79.8	6.04	0.90
20	Acetyl chloride	1.3	MeCN	3.83	89.6	5.12	0.95

Based on the initial 20 experiments, a reaction simulator was created to mimic the performance of reaction systems for algorithm comparison. The simulator was based on a plug flow reactor model to describe the relative concentration relationship between species. Influence of mixing and mass transfer were simplified by defining a mixing index that lowered the concentrations. We chose kinetic parameters and mixing index that gave similar predictions compared with the 20 yield experimental data:

$$\frac{dc_i}{d\tau} = R_j$$

$$R_j = k_j \prod ((\phi c_i)^{n_{i,j}})$$

where c is species concentration, R is reaction rate, τ is residence time, k is kinetic parameter, ϕ is mixing index and n is reaction order.

Multi-objective Bayesian optimisation results

Table S2. Optimisation results using *q*NEHVI.

entry	electrophile	X equiv	solvent	flow rate [mL/min]	Yield of **4** [%]	STY [g L^{-1} s^{-1}]	E-factor
21	Acetic anhydride	1.3	MeCN	6.00	92.4	8.27	1.12
22	Acetic anhydride	1.3	MeCN	5.12	90.8	6.93	1.16
23	Acetyl chloride	1.0	EtOAc	4.74	89.4	6.32	0.69
24	Acetic anhydride	1.3	EtOAc	5.80	90.0	7.78	1.17
25	Acetyl chloride	1.1	toluene	5.75	79.6	6.83	1.00
26	Acetic anhydride	1.5	MeCN	6.00	92.6	8.29	1.32
27	Acetyl chloride	1.0	EtOAc	3.75	89.4	5.00	0.69
28	Acetyl chloride	1.1	EtOAc	6.00	88.8	7.95	0.79
29	Acetyl chloride	1.4	EtOAc	6.00	90.6	8.11	1.02
30	Acetyl chloride	1.2	MeCN	6.00	85.4	7.65	0.96
31	Acetic anhydride	1.5	THF	6.00	91.3	8.17	1.35
32	Acetic anhydride	1.0	THF	6.00	79.6	7.13	1.10
33	Acetic anhydride	1.0	EtOAc	6.00	85.9	7.69	0.95
34	Acetic anhydride	1.0	toluene	6.00	75.8	6.79	1.20
35	Acetyl chloride	1.5	MeCN	6.00	86.6	7.75	1.21
36	Acetyl chloride	1.0	EtOAc	6.00	86.0	7.70	0.76
37	Acetyl chloride	1.1	EtOAc	3.02	90.8	4.09	0.75
38	Acetic anhydride	1.0	MeCN	6.00	82.6	7.39	1.02
39	Acetyl chloride	1.1	THF	6.00	82.0	7.34	0.94

Pareto front data

Based on the optimisation results, the Pareto front consisted of 8 points:

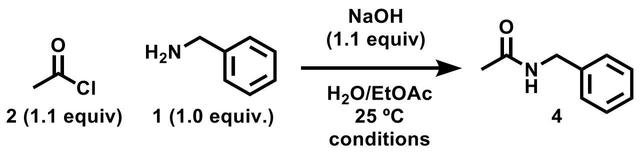
Table S3. Pareto front data.

entry	electrophile	X equiv	solvent	flow rate [mL/min]	Yield of **4** [%]	STY [g L^{-1} s^{-1}]	E-factor
18	Acetyl chloride	1.0	EtOAc	2.27	91.9	3.11	0.65
21	Acetic anhydride	1.3	MeCN	6.00	92.4	8.27	1.12
23	Acetyl chloride	1.0	EtOAc	4.74	89.4	6.32	0.69
26	Acetic anhydride	1.5	MeCN	6.00	92.6	8.29	1.32
27	Acetyl chloride	1.0	EtOAc	3.75	89.4	5.00	0.69

28	Acetyl chloride	1.1	EtOAc	6.00	88.8	7.95	0.79
29	Acetyl chloride	1.4	EtOAc	6.00	90.6	8.11	1.02
36	Acetyl chloride	1.0	EtOAc	6.00	86.0	7.70	0.76

Comparison between flow and batch condition

To further verify the importance of the flow conditions, batch reactions were examined under the same reaction condition with the exception of the reaction time (10 s). The reaction condition included in the Pareto front was used, and three independent experiments were carried out.



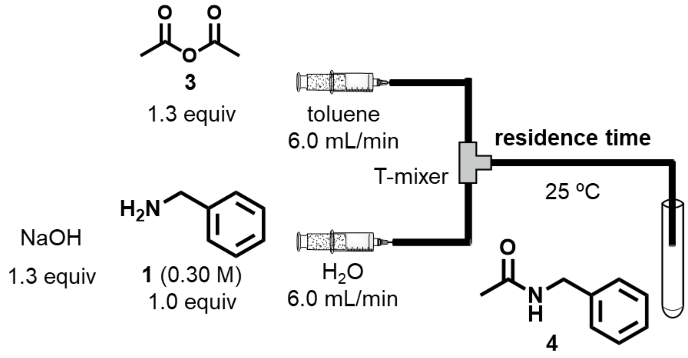
run	yield [%][a]	
	flow condition [6+6 mL/min, 2.5 s]	batch condition [1000 rpm, 10 s]
1	89.5	54.8
2	88.6	53.1
3	88.8	50.1

[a] Yield were determined by HPLC-UV analysis.

Figure S4. Comparison between flow and batch condition.

Examination of the different residence times

Reaction rates were examined using three different residence times with conditions as shown in Figure S5. The residence time was changed by changing the length of reaction tube but kept the same flow rate 6 mL/min (residence time = 2.5 s). As shown in the table, the yield of **4** was almost unchanged when increasing residence time from entries 2.5 s to 22.5 s, indicating both desired and undesired reactions were very fast and finished completely within 2.5 s even for two-phase system.



entry	residence time (s)	HPLC-UV yield of **4** (%)
1	2.5	81.7
2	12.5	81.9
3	22.5	82.2

Figure S5. Examination of the different residence times.

Examination of relationship between solvent and flow rate

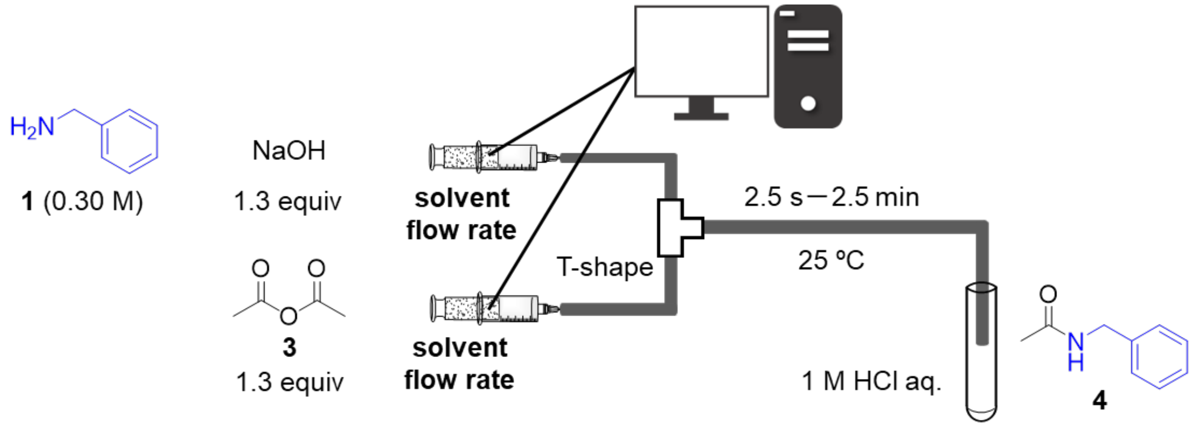


Figure S6. Examination of relationship between solvent and flow rate.

Table S4. Reaction yields with different solvents and flow rates.

entry	solvent	flow rate [mL/min]	yield of **4** [%]
1	**MeCN**	**6.00**	**92.4**
2	**MeCN**	**5.12**	**90.8**
3	MeCN	4.00	91.8
4	MeCN	2.00	92.2
5	MeCN	0.50	76.0
6	MeCN	0.50	76.0
7	MeCN	0.50	77.2
8	MeCN	0.10	97.6
9	MeCN	0.10	97.5
10	MeCN	0.10	97.8
11	THF	6.00	90.8
12	**THF**	**4.14**	**92.2**
13	THF	4.00	93.6
14	THF	2.00	83.9
15	THF	0.50	73.0
16	THF	0.10	95.3
17	THF	0.10	94.4
18	THF	0.10	95.0
19	EtOAc	6.00	89.4
20	**EtOAc**	**5.80**	**90.0**
21	EtOAc	4.00	94.7
22	EtOAc	2.00	96.3

23	EtOAc	0.50	98.0
24	EtOAc	0.10	98.5
25	EtOAc	0.10	99.8
26	EtOAc	0.10	98.6
27	toluene	6.00	81.7
28	**toluene**	**4.76**	**83.3**
29	toluene	4.00	85.5
30	toluene	2.00	91.1
31	toluene	0.50	92.5
32	**toluene**	**0.10**	**96.4**

Data in **bold** are from reaction optimisation results.

Reynolds numbers of four systems are calculated based on characteristic length and physical properties of the system, with single flow rate ranging from 0.00 – 6.00 mL/min. For the two-phase systems (EtOAc and toluene), Reynolds numbers are calculated based on the physical properties of the continuous phase:

$$Re = \frac{du\rho}{\mu}$$

d = inner diameter of T-shape mixer (0.5 mm)

u = fluid velocity (m/s)

ρ = fluid density (kg/m^3)

μ = fluid viscosity (Pa·s)

Table S5. Calculation of the Reynolds number

solvent	density [kg/m^3]	viscosity [mPa·s]	Reynolds number
MeCN	782	0.343	0 – 1162
THF	889	0.454	0 – 998
EtOAc	901	0.416	0 – 552
toluene	867	0.790	0 – 280

NMR data

N-Benzylacetamide (4)

^{1}H NMR (400 MHz, CDCl$_3$)

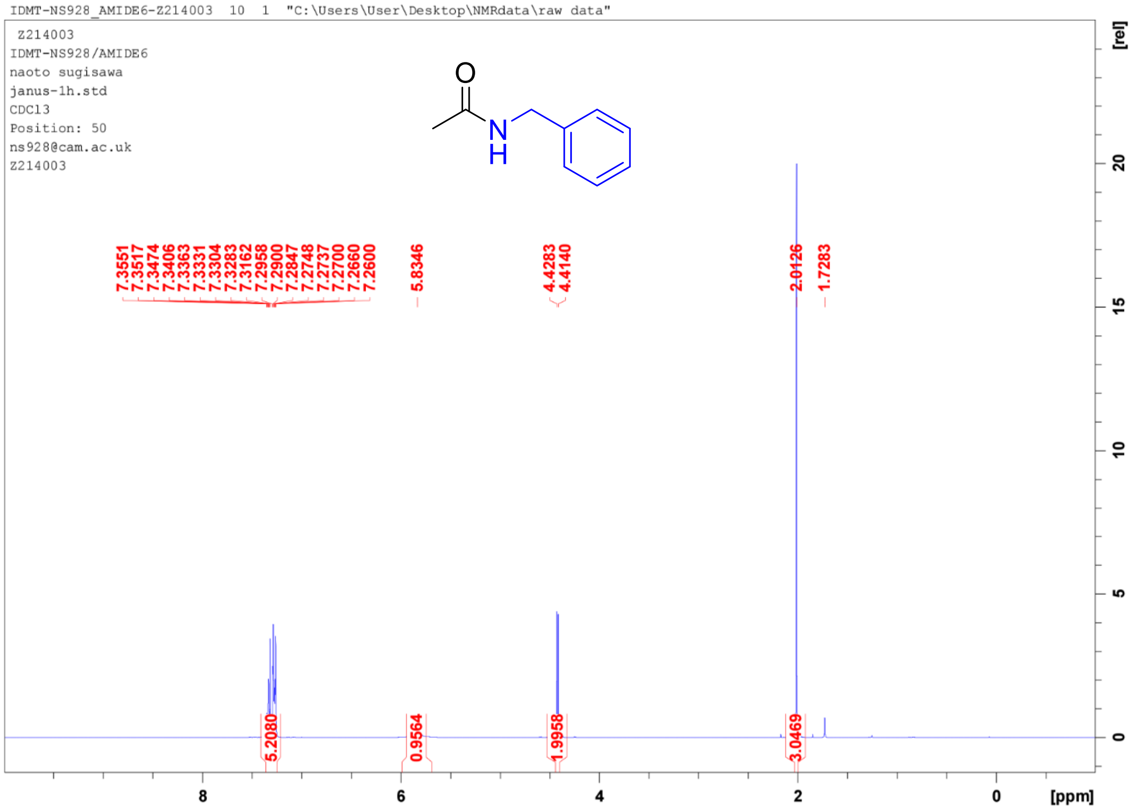


Figure S7. ^{1}H NMR data of *N*-benzylacetamide

References

1 V. R. Joseph, E. Gul and S. Ba, *Biometrika.*, 2015, **102**, 371–380.

2 V. R. Joseph, E. Gul and S. Ba, *J. Qual. Technol.*, 2020, **52**, 343–354.

3 J. L. Loeppky, J. Sacks, W. J. Welch, *Technometrics*, 2009, **51**, 366–376.

4 D. R. Jones, M. Schonlau, W. J. Welch, *J. Glob. Optim.*, 1998, **13**, 455–492.

5 A. M. Schweidtmann, A. D. Clayton, N. Holmes, E. Bradford, R. A. Bourne and A. A. Lapkin, *Chem. Eng. J.*, 2018, **352**, 277–282.

6 M. I. Jeraal, S. Sung and A. A. Lapkin, *Chem. Methods*, 2021, **1**, 71–77.