研究论文

N- 苄基 - 对甲苯磺酰胺的连续流合成工艺的贝叶斯优化

杨曦林[1]，苏为科[1]，罗贵华[1]，苏　安[1,2]*

（1. 浙江工业大学，长三角绿色制药协同创新中心，国家化学原料药合成工程技术研究中心，浙江省制药工程重点实验室，浙江杭州 310014；2. 浙江工业大学化学工程学院，绿色化学合成技术国家重点实验室培育基地，浙江省绿色化学合成技术重点实验室，浙江杭州 310014）

摘要：*N*- 苄基 - 对甲苯磺酰胺（**1**）是骨骼肌肌球蛋白Ⅱ的抑制剂，可特异性抑制骨骼肌快速纤维的收缩。为开发 **1** 的连续流合成系统，并运用贝叶斯优化算法对主要的反应参数进行优化。经过 10 轮拉丁超立方采样和 10 轮贝叶斯优化迭代后，得到 2 组最优反应条件。①反应温度为 3.8 ℃，停留时间为 14.6 min，缚酸剂（*N*,*N*- 二异丙基乙胺）用量为 1.5 当量，对甲苯磺酰氯（**2**）浓度为 0.26 mol/L 时，收率 91.4%（以 **2** 计），纯度 99.5%；②反应温度为 7.5 ℃，停留时间为 18.5 min，缚酸剂用量为 1.6 当量，**2** 浓度为 0.37 mol/L 时，收率 92.3%（以 **2** 计），纯度 99.7%。贝叶斯优化能大幅提升搜寻最优条件的效率。优化后的工艺反应时间大幅降低，溶剂单一且经济易得，具有工业化应用价值。

关键词：连续流；贝叶斯优化；*N*- 苄基 - 对甲苯磺酰胺；肌球蛋白Ⅱ抑制剂

中图分类号：TQ460.6　　**文献标志码**：A　　**文章编号**：1001-8255(2024)05-0652-06

DOI：10.16522/j.cnki.cjph.2024.05.008

Bayesian Optimization of a Continuous Flow Synthetic Process for *N*-Benzyl-*p*-toluenesulfonamide

YANG Xilin[1], SU Weike[1], LUO Guihua[1], SU An[1,2]*

(*1. Key Lab. of Pharmaceutical Engineering of Zhejiang Province, National Engineering Research Center for Process Development of Active Pharmaceutical Ingredients, Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou 310014; 2. Key Lab. of Green Chemistry-Synthesis Technology of Zhejiang Province, State Key Lab. Breeding Base of Green Chemistry-Synthesis Technology, College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014*)

ABSTRACT: *N*-Benzyl-*p*-toluenesulfonamide (**1**) is an inhibitor of skeletal muscle myosin II that specifically inhibits contraction of skeletal muscle fast fibers. A continuous flow synthesis system was developed for **1** and the main reaction parameters were optimized with a Bayesian optimization algorithm. After 10 rounds of Latin hypercube sampling and 10 rounds of Bayesian optimization iterations, two sets of optimal reaction conditions were obtained. ① The reaction temperature was 3.8 ℃, the residence time was 14.6 min, the dosage of acid binding agent (*N*,*N*-diisopropylethylamine) was 1.5 equiv. and the concentration of *p*-toluenesulfonyl chloride (**2**) was 0.26 mol/L, which corresponded to a product yield of 91.4% (based on **2**) with a purity of 99.5%. ② The reaction temperature was 7.5 ℃, the residence time was 18.5 min, the dosage of DIEA was 1.6 equiv. and the concentration of **2** was 0.37 mol/L, which corresponded to a product yield of 92.3% with a purity of 99.7%. The Bayesian optimization can greatly improve the efficiency of searching optimal conditions. The reaction time of the optimized process is greatly reduced, and the solvent is single and economical, which is valuable for industrial application.

Key Words: continuous flow; Bayesian optimization; *N*-benzyl-*p*-methylbenzenesulfonamide; myosin Ⅱ inhibitor

N- 苄基 - 对甲苯磺酰胺（**1**）是一种钙离子刺激的肌球蛋白 S1ATP 酶抑制剂，可通过非竞争性抑制肌球蛋白Ⅱ结构域上 ATP 酶的活性，发挥抑制肌肉的作用[1]。目前，**1** 已成为研究肌球蛋白药理活性的特异性工具药，也可开发作为肌球蛋白功能相关疾病（如肥厚型心肌病、格里塞利综合征、耳聋等）治疗药物的先导物[2]。

目前 **1** 的合成路线主要有以下 3 种，如图 1 所示。

路线一：以对甲苯磺酰氯（**2**）和苄胺（**3**）为起始物，在三乙胺作用下经缩合反应得目标产物 **1**，收率 99％（以 **3** 计）[3]。该路线的缺点是反应时间长。

路线二：以 **2** 和 **3** 为起始物，在混合溶剂体系中，于室温经缩合反应得目标产物 **1**，收率为 94％（以 **3** 计）[4]。该方法虽然将反应时间缩短至 8 min，但所需试剂种类繁多，且价格昂贵。

路线三：以对甲苯磺酰胺（**4**）和苄醇（**5**）为起始物，在叔丁基钾作用下，经胺化反应得目标产物 **1**，收率为 88％（以 **5** 计）[5]。该路线反应温度高、时间长，因此能耗较高。

近年来，连续流动合成因更安全的工艺、更好的质量和传热、更精确的反应参数控制、更少的试剂使用，以及有效抑制副反应等优势，受到越来越多的青睐[6—11]。此外，自我优化算法中的贝叶斯优化算法，可降低优化所需的试验次数、减少资源消耗和废物产生，同时节约时间和经济成本，帮助化学家探索广阔的化学空间。

本研究探索以 **2** 和 **3** 为起始物，经 Hinsberg 反应生成目标产物 **1** 的连续流合成（图 2），运用贝叶斯优化 4 个变量（温度、时间、浓度、试剂当量），以提高反应收率。[已确定缚酸剂为 *N*,*N*- 二异丙基乙胺（DIEA），溶剂为二氯甲烷。]

收稿日期：2023-12-13

基金项目：浙江省自然科学基金华东医药企业创新发展联合基金资助项目（LHDMZ23B060001）、浙江省科技计划项目（2022C01179）

作者简介：杨曦林（1998—），女，硕士研究生，专业方向：药物化学。

通信作者：苏　安（1988—），男，副研究员，从事智能医药研发与制造的研究。

E-mail：ansu@zjut.edu.cn

1　贝叶斯优化模型

贝叶斯优化（Bayesian optimization，BO）模型一般是由初始采样、代理模型和采集函数 3 个核心部分构成。本研究初始采样方法采用拉丁超立方抽样（Latin hypercube sampling，LHS），与蒙特卡洛模拟采取的随机抽样相比，避免了当样本量比较小时取样不均匀且带有很大随机性的问题[12]。LHS 先对数据进行分层，再在各个分层后的样本空间内随机抽样，最后打乱顺序。LHS 比传统试验设计（Deign of Experiments，DoE）方法所需的样本更少，从而降低了计算和时间成本。

高斯过程（Gaussian process，GP）是常用的贝叶斯概率代理模型。目前，GP 过程已被广泛应用在回归、分类以及推断黑箱函数中。GP 可以根据历史数据，预估未知数据的方差和平均值。GP 由均值函数 μ(X) 和协方差函数 $\sum(x, x')$ 共同决定，如式①所示。

$$\sum_0(x,x')=\alpha_0\frac{2^{1-\nu}}{\Gamma(\nu)}(\sqrt{2\nu}\,\|x-x'\|)^{\nu}K_{\nu}(\sqrt{2\nu}\,\|x-x'\|) \quad ①$$

式中：K_V 表示修正的贝塞尔函数，$\|x-x'\|$ 表示两点间的距离，α_0 表示方差，Γ 表示伽马函数，V 表示非负参数。

采集函数是通过使采集函数最大化来选择下一个最有“潜力”的评估点，合适的采集函数可以有效保证总损失最小。常用的采集函数包括：概率提升法、期望提升法、置信上限法。DAULTON 等[13]在进行多目标优化时，针对观测值易被噪声影响的问题，提出了一种新的采集函数 qNEHVI，如式②所示。事实证明，qNEHVI 在观察噪声方面更具鲁棒性，故本研究采用该采集函数进行优化。

$$\alpha_{qNEHVI}(x_{cand})=\frac{1}{N}\sum_{t=1}^{N}\mathrm{HVI}(\tilde{f}_t(\chi_{cand})|P_t) \quad ②$$

式中：HVI 表示超体积改进，P_t 表示帕累托边界，χ_{cand} 表示候选采样，N 表示样本的数量。

2　优化过程

贝叶斯优化的优化流程如图 3 所示。

首先，采用 LHS 在受试试验范围（温度 0～25 ℃；停留时间 0 ～ 30 min；DIEA 当量 1 ～ 2.5；化合物 **2** 浓度 0.01 ～ 1 mol/L）内抽取 10 组条件，采用图 4 所示进行试验，并将结果作为初始训练数据。在每一轮迭代中，经由高斯代理模型根据已知的试验点，

计算每一点处函数值的均值和方差。再由采集函数推荐下一组试验条件，经试验得到新的数据进行新的迭代，直至达到所设定的最大迭代次数（10 次）。贝叶斯优化的结果如图 5 所示，体现了目标值与试验次数的关系。此处目标值指的是转化率 × 选择性的数值，即试验后稀释分析，从液相谱图中计算得到，记为产率。补充说明：试验部分经后处理得到的数值记为收率，以此区分。

分析上述贝叶斯优化结果所对应的反应条件，如表 1 所示。

最高收率发生在第 7 组试验，因此最优条件为 3.8 ℃、停留时间 14.6 min、DIEA 当量为 1.53 且

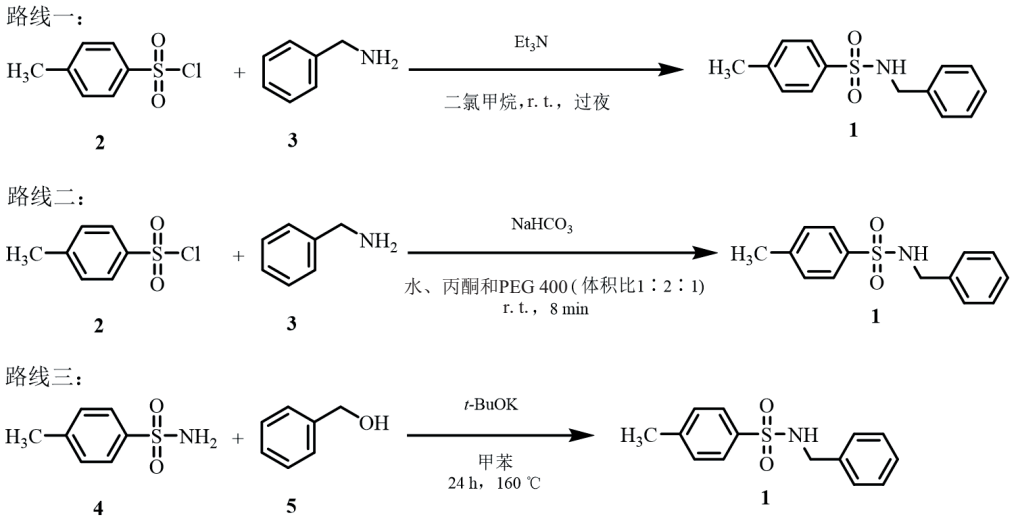


图 1　1 的合成路线

Fig.1　Synthetic Routes of **1**

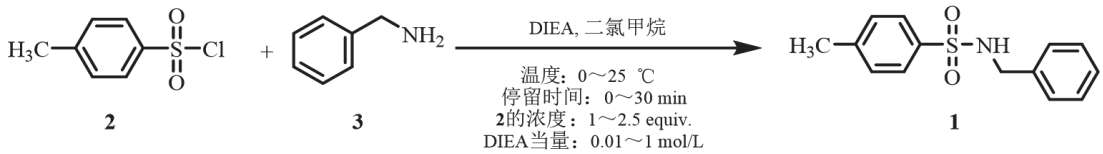


图 2　连续流合成 1 的路线

Fig.2　The Synthetic Route of **1** in the Continuous Flow

表 1　贝叶斯优化中每轮试验的反应条件

Tab.1　Reaction Conditions of Each Round Experiment in Bayesian Optimization

编号	温度 /℃	停留时间 /min	DIEA 当量	化合物 **2** 浓度 /（mol·L^{-1}）	编号	温度 /℃	停留时间 /min	DIEA 当量	化合物 **2** 浓度 /（mol·L^{-1}）
1	24.0	13.5	1.08	0.45	11	13.0	21.2	1.65	0.459
2	21.0	19.7	2.30	0.75	12	11.0	27.6	2.26	0.433
3	16.0	22.6	1.80	0.55	13	8.0	23.3	1.63	0.274
4	10.0	28.6	1.38	0.65	14	18.0	22.2	1.38	0.511
5	6.0	10.5	1.22	0.16	15	10.0	30.0	1.54	0.352
6	8.8	25.5	2.13	0.36	16	12.0	18.6	1.33	0.382
7	3.8	14.6	1.53	0.26	17	7.5	18.5	1.60	0.376
8	1.0	7.5	1.68	0.85	18	2.0	21.1	1.22	0.342
9	11.0	4.5	2.43	0.06	19	7.0	20.6	1.94	0.252
10	19.0	1.5	1.98	0.95	20	11.0	26.7	1.05	0.599

化合物 **2** 浓度 0.260 mol/L，对应产率为 98.2%。另外，第 17 组试验条件，也获得了接近的收率为 98.0%，可认为在误差范围内，都可算作为最优条件。相较而言，第 17 组反应温度（7.5 ℃）比第 7 组（3.8 ℃）更接近室温，能减少制冷所需的能源消耗，但 **2** 浓度（0.376 mol/L）更高，因此这 2 组条件各有优点。此外，第 11 组和第 19 组也得到了≥ 96.3%的产率，可作为备用方案。优化后的工艺，通过连续流合成，经贝叶斯优化所需的反应条件，反应时间由过夜缩短至小于 20 min，溶剂用量大幅降低，从而大大提高了生产效率和降低了生产成本。值得注意的是，在本研究中，贝叶斯优化只针对连续变量（如温度、停留时间等）。今后的研究可以考虑将溶剂类型、缚酸剂类型等离散变量纳入到贝叶斯优化的范围，使得优化体系更加完善，满足医药制造业多样化的优化需求。

将上述受试的 4 个变量两两作图，分析贝叶斯优化过程的采样空间（图 6），体现出贝叶斯优化的一个重要特点，即平衡探索和开发这 2 个决策。其中，由采集函数主导的迭代循环不仅进一步开发出 LHS 阶段较佳反应条件（第 6、7 组）附近的取样空间，而且还在 LHS 阶段未能涉足的采样空间内进行探索（如远离所有 LHS 取样点的第 11、12、15、16 组），尽管这样可能导致较低的收率（如第 12、15 组），但也有可能接近甚至超越现有最优解（如第 11、16 组），最大程度地避免局部最优解。

实验部分

以第 7 组的反应条件进行试验。

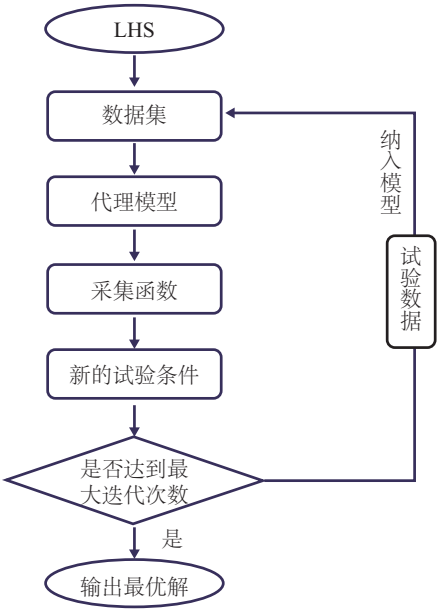


图 3　贝叶斯优化流程图

Fig.3　The Flowchart of Bayesian Optimization

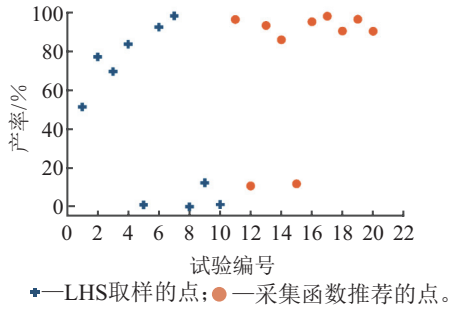


图 5　贝叶斯优化结果

Fig.5　The Result of Bayesian Optimization

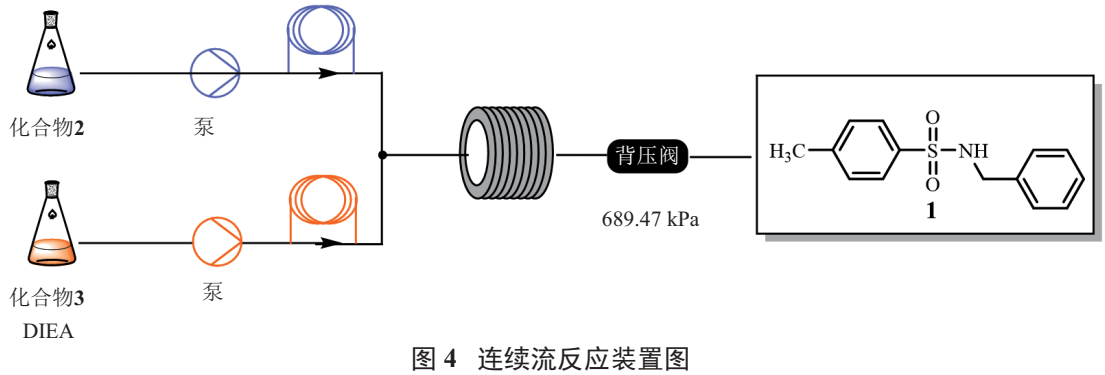
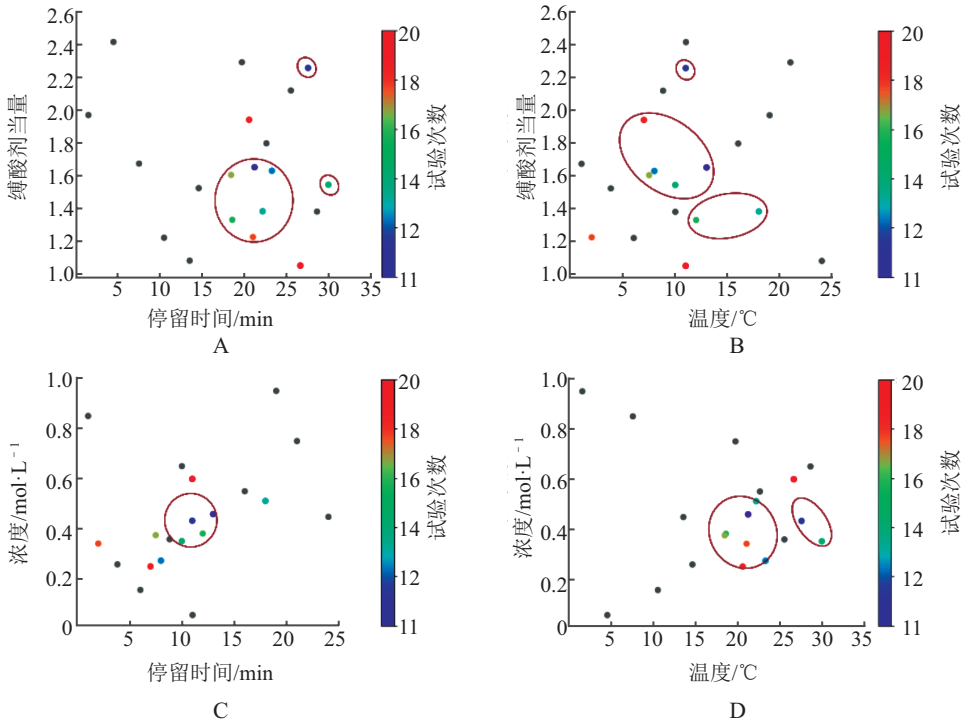


图 4　连续流反应装置图

Fig.4　Schematic Diagram of the Continuous Flow System



注：灰色点代表初始 LHS 取样点；彩色点代表由采集函数推荐的取样点（色条显示迭代次数），其中红色圈内的是采集函数推荐的与初始 LHS 取样点相距较远的取样点。

图 6　贝叶斯优化过程中的取样空间分析

Fig.6　The Analysis of Sampling Space during Bayesian Optimization

将 **2**（49.6 g，260 mmol，上海麦克林生化科技有限公司）溶于 DCM（500 mL）中；另将 **3**［30.6 g，286 mmol，梯希爱（上海）化成工业发展有限公司］和 DIEA（51.4 g，398 mmol）溶于 DCM（500 mL）中。两液分别以 0.27 mL/min 的流速泵入反应体系中。设置反应温度为 3.8 ℃，背压阀调至 689.48 kPa（100 psi），停留时间为 14.6 min。待反应液收集完全后，加等量的饱和氯化钠溶液（1 000 mL），静置分层，用 DCM（300 mL×3）萃取，有机相减压蒸除溶剂，常温减压干燥 12 h 后，得白色固体 **1**（62.4 g，91.4%）。mp 111.8 ～ 112.9 ℃（文献[14]：113.0 ～ 117.0 ℃）。纯度 99.5%［HPLC 归一化法：色谱柱 Welch Ultimate XB-C_{18} 柱（4.6 mm×250 mm，3 μm）；流动相 甲醇和水（体积比 60 ∶ 40）；检测波长 254 nm；柱温 30 ℃；流速 1 mL/min；进样量 10 μL。保留时间为 20.0 min］。^{1}H NMR（600 MHz，DMSO）δ: 8.10（t, *J*=6.3 Hz, 1H），7.71（d, *J*=8.1 Hz, 2H），7.38（d, *J*=8.0 Hz, 2H），7.27 ～ 7.31（m, 2H），7.22 ～ 7.25（m, 3H），3.95（d, *J*=6.4 Hz, 2H），2.39（s, 3H）。

以第 17 组的反应条件进行试验。

将 **2**（71.7 g，376 mmol，上海麦克林生化科技有限公司）溶于 DCM（500 mL）中；另将 **3**［44.3 g，412 mmol，梯希爱（上海）化成工业发展有限公司］和 DIEA（77.8 g，602 mmol）溶于 DCM（500 mL）中。两液分别以 0.27 mL/min 的流速泵入反应体系中。设置反应温度为 7.5 ℃，背压阀调节至 100 psi，停留时间为 18.5 min。其后按上述方法操作，得白色固体 **1**（91.0 g，92.3%）。mp 111.6 ～ 112.8 ℃。纯度 99.7%（HPLC 条件同上）。

参考文献：

[1]　CHEUNG A, DANTZIG J A, HOLLINGWORTH S, *et al*. A small-molecule inhibitor of skeletal muscle myosin Ⅱ［J］. *Nat Cell Biol*, 2002, **4**(1): 83-88.

[2]　BOND L M, TUMBARELLO D A, KENDRICK-JONES J, *et*

al. Small-molecule inhibitors of myosin proteins [J]. *Future Med Chem*, 2013, **5**(1): 41-52.

[3] BROM J, MARUANI A, MICOUIN L, *et al*. Light-driven reductive cleavage of sulfonamides promoted by thiourea organophotosensitizers [J]. *J Org Chem*, 2023, **88**(9): 5923-5935.

[4] GIOIELLO A, ROSATELLI E, TEOFRASTI M, *et al*. Building a sulfonamide library by eco-friendly flow synthesis [J]. *ACS Comb Sci*, 2013, **15**(5): 235-239.

[5] KOZLOV A S, AFANASYEV O I, CHUSOV D. Borrowing hydrogen amination: whether a catalyst is required? [J]. *J Catal*, 2022, **413**: 1070-1076.

[6] HONE C A, KAPPE C O. Towards the standardization of flow chemistry protocols for organic reactions [J]. *Chem Methods*, 2021, **1**(11): 454-467.

[7] GUIDI M, SEEBERGER P H, GILMORE K. How to approach flow chemistry [J]. Chem Soc Rev, 2020, **49**(24): 8910-8932.

[8] GÉRARDY R, EMMANUEL N, TOUPY T, *et al*. Continuous flow organic chemistry: successes and pitfalls at the interface with current societal challenges [J]. *Eur J Org Chem*, 2018, **2018**(20-21): 2301-2351.

[9] BRITTON J, RASTON C L. Multi-step continuous-flow synthesis [J]. *Chem Soc Rev*, 2017, **46**(5): 1250-1271.

[10] MOVSISYAN M, DELBEKE E I P, BERTON J K E T, *et al*. Taming hazardous chemistry by continuous flow technology [J]. *Chem Soc Rev*, 2016, **45**(18): 4892-4928.

[11] PLUTSCHACK M B, PIEBER B, GILMORE K, *et al*. The hitchhiker's guide to flow chemistry [J]. *Chem Rev*, 2017, **117**(18): 11796-11893.

[12] OLSSON A, SANDBERG G, DAHLBLOM O. On Latin hypercube sampling for structural reliability analysis [J]. *Structural Safety*, 2003, **25**(1): 47-68.

[13] DAULTON S, BALANDAT M, BAKSHY E. Parallel Bayesian optimization of multiple noisy objectives with expected hypervolume improvement [J]. *Adv Neural Inf Process Syst*, 2021.

[14] TCI. Product details page of *N*-Benzyl-p-toluenesulfonamide [EB/OL]. [2024-03-01]. https://www.tcichemicals.com/US/en/p/B3082.