Supporting Information: Schedule Optimization for Chemical Library Synthesis

Qianxiang Ai,[†,§] Fanwang Meng,[†,§] Runzhong Wang,[†] J. Cullen Klein,[‡] Alexander G. Godfrey,[‡] and Connor W. Coley[*,†,¶]

[†] *Department of Chemical Engineering, MIT, Cambridge, MA, 02139*

[‡] *National Center for Advancing Translational Sciences, Rockville, MD, 20850*

[¶] *Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 02139*

§ *authors contributed equally*

E-mail: ccoley@mit.edu

S1 Supporting figures

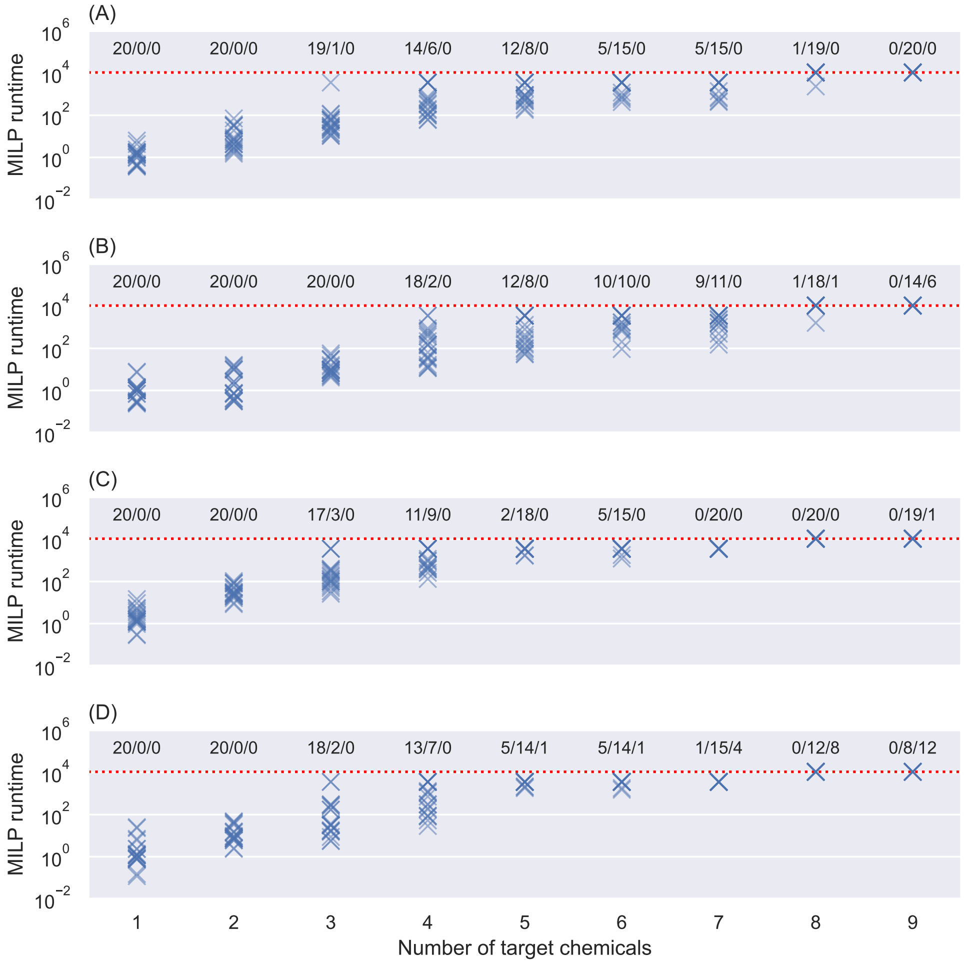

Figure S1: Runtime distributions for finding optimal schedules. Red dashed lines indicate the solver time limit of three hours. Each column (strip plot) represents the results of a library group scheduled on LAB-1 and LAB-2, totaling 20 scheduling instances. (A) FDA libraries scheduled without work shift constraints; (B) FDA libraries scheduled with work shift constraints; (C) VS libraries scheduled without work shift constraints; (D) VS libraries scheduled with work shift constraints. The columns are annotated with triples of the format "x/y/z", where "x", "y", and "z" represent the number of "Optimal", "Suboptimal", and "No solution" instances, respectively, as defined in Figure 3 caption.

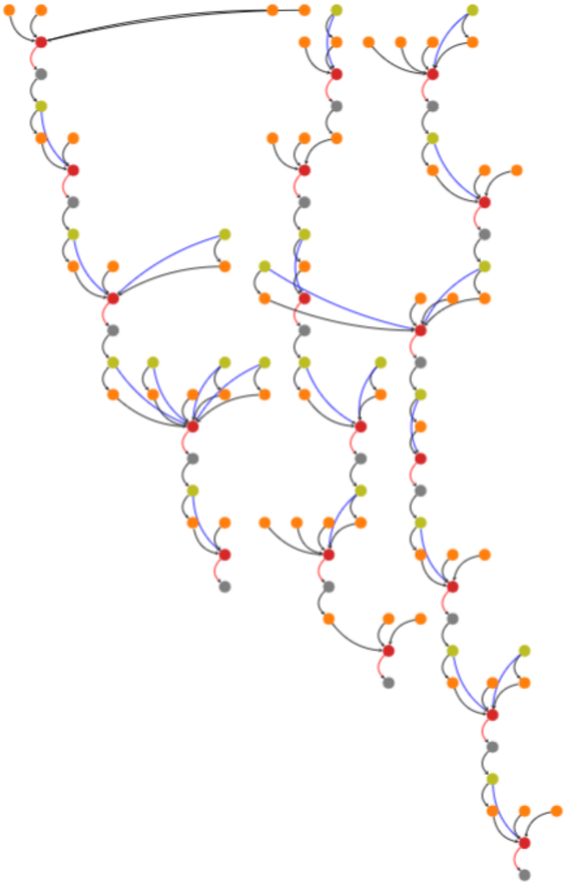

Figure S2: Operation network constructed for FDA.03.09. The orange, red, gray, and yellow markers represent `LiquidTransfer`, `Heating`, `ConcentrationAndPurification`, and `Makesolution` operations, respectively. The blue (red) edges indicate minimum (maximum) lag time constraints. This is a screenshot taken from the visualization interface.

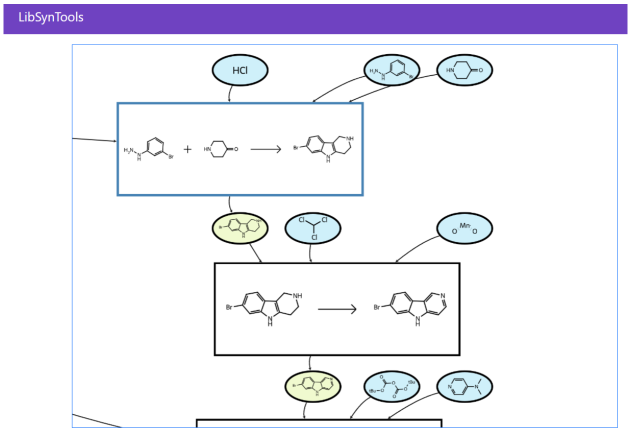

Figure S3: Part of the reaction network for FDA.03.09 from a screenshot taken from the visualization interface. Molecules with blue and green backgrounds represent starting (include solvents and reagents) and intermediate molecules, respectively.

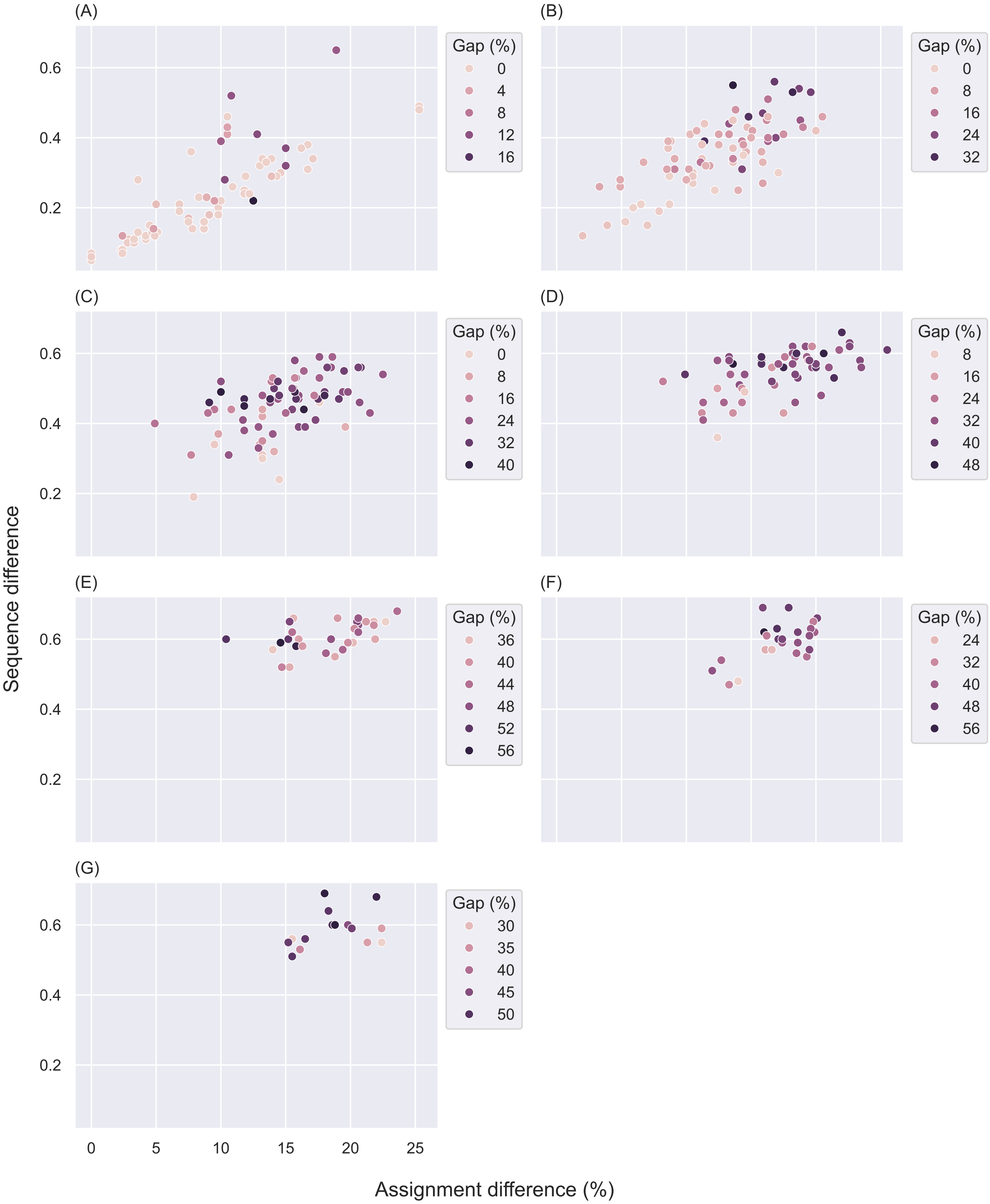

Figure S4: Schedule difference between optimal and baseline schedules. Scheduling instances are grouped by the number of target chemicals in (A) - (G) for number of target chemicals from 1 to 7.

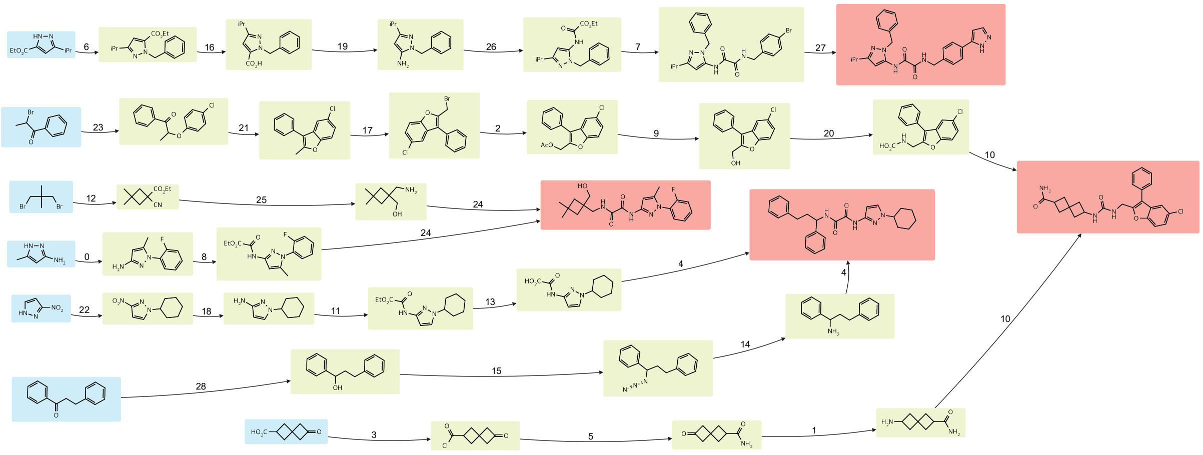

Figure S5: Reaction network for VS.04.07 generated by ASKCOS and SPARROW. Molecules with blue, green, and red backgrounds represent starting, intermediate, and target molecules, respectively. Reagents and solvents of these reactions, which also exist in the reaction network, are excluded in this figure for clarity purposes.

S2 Additional information about case studies

S2.1 Operation types and processing time estimates

Table S1 shows the list of operation types and the methods to compute their time estimates. A multiplier randomly sampled from $[0.7, 1.3]$ is applied to the final processing time when estimating the operation on a specific functional module such that different operation-to-module assignments can have different makespans. The maximum relative variance of 30% is used as a rough estimation as no empirical data is available. The random multipliers are used and only used in generating processing times. Given an operation graph, the scheduling input (including processing times p_{im}) is consistent except heater capacities (K_m where m are indices of heaters) regardless which scheduling algorithm or which LAB is used in scheduling. We define `Purification` operations to be executed in batches, thus estimating their processing times involves pre-defined batch size (e.g. the volume of the mixture) and batch processing time.

We note the concept of functional module is used in the current formulation as an abstraction of hardware units, thus two functional modules for the same operation type can be mapped to two (sets of) hardware units that are different instruments. For example, the two "Workup station" modules in LAB-1 can be mapped to one containing a rotary evaporator and another containing a rapid vacuum evaporator.

Table S1: Operation types and their processing time estimates.

Operation type	Process time estimate
`Purification`	batch processing time × (⌈mixture quantity / batch size⌉)
`Concentration`	randomly sampled from a uniform distribution
`Dissolution`	randomly sampled from a uniform distribution
`TransferLiquid`	liquid volume / transferring rate
`MakeSolution`	`TransferLiquid` + `Dissolution`
`Heating`	randomly sampled from a uniform distribution
`ConcentrationAndPurification`	`Purification` + `Concentration`

S2.2 Temperature bins

Table S2: Temperature bins for compatibility among `Heating` operations on the same heating module.

Condition	**Min (°C)**	**Max (°C)**
Very cold	$-\infty$	-5
Cold	-5	20
Room	20	35
Mild	35	85
Hot	85	150
Very hot	150	$+\infty$

S2.3 Chemical libraries

Table S3: Chemical library groups in case studies. Each row represents a group of ten test libraries sharing the same number of target chemicals (shown in the second column).

Group name	Targets	Reactions	Operations	Solids	Liquids
FDA.01	1	[4, 8]	[23, 64]	[10, 27]	[6, 19]
FDA.02	2	[5, 15]	[41, 98]	[17, 41]	[12, 34]
FDA.03	3	[11, 21]	[74, 128]	[29, 53]	[20, 48]
FDA.04	4	[13, 25]	[97, 179]	[31, 72]	[29, 59]
FDA.05	5	[17, 29]	[124, 196]	[50, 75]	[39, 70]
FDA.06	6	[23, 34]	[150, 233]	[56, 94]	[54, 75]
FDA.07	7	[22, 37]	[168, 263]	[69, 110]	[51, 83]
FDA.08	8	[34, 44]	[239, 299]	[94, 121]	[74, 98]
FDA.09	9	[35, 49]	[238, 366]	[93, 148]	[82, 117]
VS.01	1	[3, 12]	[20, 79]	[7, 32]	[8, 26]
VS.02	2	[11, 18]	[63, 116]	[20, 44]	[27, 44]
VS.03	3	[13, 28]	[80, 159]	[22, 62]	[39, 59]
VS.04	4	[19, 34]	[124, 200]	[42, 79]	[50, 79]
VS.05	5	[31, 47]	[189, 279]	[66, 111]	[74, 98]
VS.06	6	[34, 50]	[214, 296]	[75, 113]	[74, 109]
VS.07	7	[39, 57]	[239, 358]	[87, 137]	[89, 127]
VS.08	8	[41, 61]	[264, 388]	[96, 147]	[102, 143]
VS.09	9	[51, 63]	[313, 402]	[109, 155]	[129, 151]

S3 Precedence relation types

Different concepts related to precedence relations among operations were involved in our formulation, and we defined a vocabulary in hopes of minimizing ambiguity. We illustrate these concepts with the following example:

A mixture of water, sugar, and salt is heated at 40 °C for 1 hour. 1 mL of vinegar is then added to the mixture. The temporal gap between the end of heating and the start of vinegar should be at most 3 min.

Let $\mathcal{O}$ be its operation graph, $\mathcal{S}$ be a feasible schedule where water, sugar, and salt were added sequentially, we have:

- **Required precedence relation** in $\mathcal{O}$: A relation that is required by the procedure. For example, the addition of water/sugar/salt is *required* to precede the heating operation as it is explicitly stated. Note the addition order of water, sugar, and salt is not stated in the procedure, so there is no required precedence relation among additions of these components – they can be added sequentially or concurrently.

- **Implied precedence relation** in $\mathcal{S}$: A derived relation from a given schedule $\mathcal{S}$. For example, in $\mathcal{S}$ water addition precedes sugar addition, even this is not required by $\mathcal{O}$.

- **Ordinary precedence constraint** in linear programming: A precedence constraint between two operations without minimum/maximum lag time. For example, the constraint that water addition should precede mixture heating, which is different from the the constraint that mixture heating should precede vinegar addition with a maximum lag of 3 min.