Supporting Information

Natural Language Processing for Automated Workflow and Knowledge Graph Generation in Self-Driving Labs

Bastian Rühle*,[a]

[a] Federal Institute for Materials Research and Testing (BAM), Richard-Willstaetter-Strasse 11, D-12489 Berlin, Germany

*Correspondence: bastian.ruehle@bam.de

Further resources

The github repository accompanying this publication can be found here:

https://github.com/BAMresearch/MAPz_at_BAM/tree/main/Minerva-Workflow-Generator

It contains the python code of the node editor, some example files, a jupyter notebook for visualizing and querying the knowledge graphs, and links to the annotated datasets and the fully trained LLMs.

Generating the (semi-)automatically annotated dataset dataset_llama_cleaned

The template that was used for generating the ICL prompts for automatically creating action graphs with the Llama-3.1-8B-Instruct model can be found in ICL-Template-1. The example procedure in the prompts was taken from the supporting information of a publication[26] and the example output was created by manually annotating it as an action graph. This procedure was mainly chosen due to its style of writing being similar to the procedures in the training set, its reasonable length, and the fact that it covers many different and commonly occurring action tags.

While experimenting with different ICL prompts and inspecting the output generated by the Llama-3.1-8B-Instruct model, it was noticed that the model sometimes used new action tags that did in fact not have a clear correspondence to one of the original 21 ChemicalTagger action tags, or they indicated a specific but important distinction of a more general action defined by these tags. For example, the model often used tags related to adjusting the pH of a solution in its annotations, such as <ACIDIFY>, <ADJUST>, <NEUTRALIZE>, and <BASE>. While these could in principle be represented by an <ADD> action of an acid or base, it was decided to include a new tag <ADJUSTPH>. Similarly, there were many cases in which <MIX> or <REACT> were used as new tags in the annotations created by the Llama-3.1-8B-Instruct model. While these could in principle often be replaced with a simple <STIR> tag, they commonly indicate a non-linear action in which components are mixed or reacted before being added to a reaction mixture, so it was decided to also introduce a new <MIX> tag to keep this distinction. The ICL prompt template 1 for generating action graphs hence included a list of 23 action tags, the original 21 tags from ChemicalTagger plus the new tags <MIX> and <ADJUSTPH>.

It should also be noted that using other instructions and other examples in the ICL prompt will likely give different results. However, besides some basic tests, no further experimentation with ICL prompt engineering was done in this work and the same instruction template was used for all experimental procedures from the training dataset. After removing the example output of the ICL (which was sometimes repeated by the model in the output) and extra line breaks and tab stops (which were also sometimes present in the output) from the output data, the new dataset dataset_llama_raw was obtained. In total, there were 2956 different tags occurring with frequencies varying between 3’134’027 for the most common action <ADD> (a tag the model was supposed to use) to 1 for the least common action <ZnI2> (a new, custom tag that was invented by the Llama-3.1-8B model), which is considerably more than the 23 tags the model was instructed to use. These additional tags occurred in 194’073 examples (12.3%).

Such a large number of tags is certainly too specific to be useful in the automated downstream processing of the generated action graphs. Upon closer inspection of the 47 most frequent (i.e., occurring with frequencies > 1’000), undesired tags, it was found that some of these tags could readily be replaced by an already existing tag with some further specification in its parameter set, e.g., <REFLUX> was replaced by <HEAT> reflux, <DROPWISE> was replaced by <ADD> slowly, and <HYDROGENATE> by <ADD> H2. There were also tags that could be replaced directly by a single, synonymous tag, such as <COMBINE> with <MIX>, <DILUTE> with <ADD>, or <WARM> with <HEAT>. A full list of the substitutions that were made can be found in Table S1.

It was also found that the model used the tag <REPEAT> 8’597 times, which is a tag that is potentially very useful for constructing action graphs, but neither part of the original 21 ChemicalTagger Tags nor of the 23 “allowed” action tags from ICL-Template-1. So instead of replacing this tag, it was added to the list of “allowed” tags in ICL-Template-2, and all procedures

that contained “repeat” (10’657 entries total, i.e., including those that were already annotated with <REPEAT> tags by the model) were again included in the list of procedures for re-annotating.
Furthermore, the tags <SONICATE> (4’183 occurrences plus 638 occurrences for <SONICATION>) and <CENTRIFUGE> (340 occurrences plus 2 occurrences each for <CENTRIFUGALIZE> and <CENTRIFUGATE> and 1 occurrence for <CENTRIFUGAL SEPARATION>) were also added to the list of “allowed” tags at this stage and included in ICL-Template-2. While these could be represented adequately by <APPARATUSACTION> sonicate and <SEPARATE>, respectively, it is anticipated that they will prove useful in generating action graphs, especially for materials science applications. In total, the list of allowed tags now comprised 26 tags, i.e., the 21 ChemicalTagger tags plus <MIX>, <ADJUSTPH>, <REPEAT>, <SONICATE>, and <CENTRIFUGE>, which is still a reasonable amount and a good balance between being overly specific and too general.
Next, all other procedures that still contained tags that were neither in the list of the 26 allowed tags nor in the list of the 47 manually substituted tags were added to the collection of procedures for re-annotation. For these, the ICL prompt was adjusted (ICL-Template-2), using a slightly different wording in the instruction part and a different procedure with manually annotated output in the example part of the ICL prompt and the Llama-3.1-8B-Instruct model was run again on these experimental procedures. The new example for the ICL prompt was chosen because it had several additional action tags that occurred frequently in other procedures as well, namely <GRADIENT>, <FLOW>, <EVACUATE>, <BACKFILL>, <MOBILE PHASE A>, <MOBILE PHASE B>, and <GUARD COLUMN>.
The resulting dataset (dataset_llama_raw_1) was still using over 1000 different tags, i.e., still many more than the 26 allowed tags. Since the Llama-3.1-8B-Instruct model kept using undesired tags in the output, a large portion of the dataset was cleaned up manually by replacing tags that had a clear synonym from the list of the 26 desired tags. In some cases, these substitutions were rather straightforward, such as replacing <CHILL> with <COOL> or <BOIL> with <HEAT>, in some cases the actions were just spelling variations of the same action, e.g., , <LYOPHALLIZE>, <LYOPHILISATION>, <LYOPHILISE>, <LYOPHILIZATION>, <LYOPHILLIZE>, <LYOPHILYZE>, <LYOPHOLISE>, <LYOPHOLIZE>, <LYOPHOLYZE>, <LYPHILIZATION>, <LYPHOLISE>, <LYPHOLIZE>, <LYPHOLYZE>, <FREEZE DRY>, <FREEZEDRY>, and <FREEZE-DRY>, which were all replaced by <DRY>, and in yet other cases, the action did have a correspondence in the context it was used, but this was less obvious, e.g., <CHASE> was used for describing the action of changing the eluent in column chromatography. In some cases, the use of the tag was very context-dependent, e.g., <REDUCE> was used for describing a chemical reduction as well as the action of reducing the volume of a solution through evaporation, and the tag <SUBSTITUTE> could in principle be replaced by a <REMOVE> action followed by an <ADD> action, but the replacement would have to take the parameters given after these tags into account. In those cases, no substitutions were made, and the re-annotation was left to the large language model. A list of all substitutions can be found in Table S2.
Since the Llama-3.1-8B-Instruct model was clearly struggling with these annotations, the more powerful Llama-3.1-70B-Instruct model (that has 70 billion instead of 8 billion parameters) was used for re-annotating these procedures, albeit with reduced accuracy (8 bit instead of 16 bit to keep resource consumption in a reasonable frame). The ICL prompt was also adjusted again, using a variation of ICL-Template-1 that included an example of an <APPARATUSACTION>, as well as the “history” of the interaction of the “user” with the “agent”. This new ICL prompt hence consisted of the modified instruction part, the initial response of the Llama-3.1-8B-Instruct, and a new instruction that listed the undesired tags that were used by the model in the output and instructed the model again to use only the desired tags (ICL-Template-3). The resulting dataset still contained a few hundred instances in which tags other than the 26 desired tags were used and 4 instances that still used tags that were also not in the lists of substitutions that were manually compiled before. These were cleaned up manually.
Finally, after combining the output of all 3 annotation steps, there were still some instances in which the output was syntactically incorrect (e.g., ending with an open angle bracket <), cases in which parts of the example procedure or example output were repeated, or cases in which the model simply kept repeating the same 2-3 tags over and over again. These were filtered out by compiling a list of procedures in which the output contained over 40 tags total or over 15 times the same tag, or that contained the exact same chemicals and amounts that were used in the example instructions. These were re-annotated iteratively using a combination of manual corrections and re-annotations using Llama-70B-instruct with ICL Templates 1, 2, and 3, or slight variations thereof.
Finally, the dataset was post-processed by removing all text including and following “Procedure “ or “Output” up to the next action tag (in some cases the model repeated the procedure or sometimes even completely hallucinated a procedure and appended it along with its annotation to the output), and applying the same clean-up that was also used for the ChemicalTagger Dataset, i.e., removing extra line-breaks, leading and trailing spaces, and grouping actions in cases where the same action tags occurred consecutively. This gave the final dataset (dataset_llama_cleaned) that was used for training the surrogate models.

Templates for ICL prompts

ICL-Template-1
Given a chemical synthesis procedure, summarize the procedure in structured output. Use exclusively these markup tags in the structured output, no other tags: <ADD>, <YIELD>, <DISSOLVE>, <STIR>, <WASH>, <DRY>, <CONCENTRATE>, <PURIFY>, <REMOVE>, <FILTER>, <HEAT>, <EXTRACT>, <COOL>, <SYNTHESIZE>, <WAIT>, <PARTITION>, <DEGASS>, <QUENCH>,

<RECOVER>, <APPARATUSACTION>, <PRECIPITATE>, <MIX>, <ADJUSTPH>. Give only the output, no explanation. Here is an example:
Procedure: In a flame-dried 100 mL round bottom flask, a mixture of 2.0 mL of furfurylamine (22 mmol) and 3 mL of triethylamine were stirred in 45 mL of dry dichloromethane under nitrogen at 0 °C. Then, 4.4 g of 1-adamantane carbonylchloride (22 mmol) in 5 mL of dry dichloromethane was added slowly, and the solution was allowed to warm to room temperature. After stirring for 1 h at room temperature, the solution was washed with 40 mL of an aqueous ammonium chloride solution (saturated) and 40 mL of an aqueous potassium carbonate solution (5%), the organic layer was separated, dried over MgSO4, filtered, and evaporated to dryness in vacuo. The crude product was recrystallized from heptane/EtOAc = 1:1 (v/v) to yield the product as off-white needles (3.15 g, 55%)
Output: <ADD> furfurylamine 2.0 mL 22 mmol triethylamine 3 mL dichloromethane 45 mL <COOL> 0 °C <MIX> 1-adamantane carbonylchloride 22 mmol 4.4 g dry dichloromethane 5 mL <ADD> mixture slowly <HEAT> room temperature <STIR> 1 h <WASH> aqueous ammonium chloride solution (saturated) 40 mL aqueous potassium carbonate solution (5%) 40 mL <EXTRACT> organic layer <DRY> MgSO4 <FILTER> <REMOVE> in vacuo <PURIFY> heptane/EtOAc = 1:1 (v/v) <YIELD> off-white needles 3.15 g 55 %
Procedure: ***<<experimental procedure from dataset inserted here>>***

ICL-Template-2
Given a chemical synthesis procedure, summarize the procedure in structured output. Use exclusively the following markup tags in the structured output, no other tags: <ADD>, <YIELD>, <DISSOLVE>, <STIR>, <WASH>, <DRY>, <CONCENTRATE>, <PURIFY>, <REMOVE>, <FILTER>, <HEAT>, <EXTRACT>, <COOL>, <SYNTHESIZE>, <WAIT>, <PARTITION>, <DEGASS>, <QUENCH>, <RECOVER>, <APPARATUSACTION>, <PRECIPITATE>, <MIX>, <ADJUSTPH>, <REPEAT>, <SONICATE>, <CENTRIFUGE>. Give only the output, no explanation. Here is an example:
Procedure: 2-chloro-4-(cyclopropyl(4-methoxybenzyl)amino)imidazo[2,1-f][1,2,4]triazine-7-carbonitrile (Intermediate 9) (60 mg, 0.169 mmol), 3-amino-4-fluoro-5-(3-(4-methylpiperazin-1-yl)azetidin-1-yl)benzonitrile (48 mg, 0.166 mmol), DPPF (6.44 mg, 0.012 mmol), Cs2CO3 (92 mg, 0.282 mmol), Xantphos (9.60 mg, 0.017 mmol), Palladium(II)Acetate (11.17 mg, 0.050 mmol) and 1,4-dioxane (2 ml) were combined in a microwave vial. The vial was evacuated and backfilled with Nitrogen 3 x. The reaction stirred at 100 °C for 3 hr. The reaction mixture was cooled to 25 °C, diluted with EtOAc, washed with brine and dried (Na2SO4). The solvents were removed and DCE (1 mL)/TFA (0.5 mL) was added. The reaction stirred 3 h and solvents were removed. The crude material was purified via preparative LC/MS with the following conditions: Column: Waters XBridge C18, 19 x 200 mm, 5-um particles; Guard Column: Waters XBridge C18, 19 x 10 mm, 5-um particles; Mobile Phase A: water with 20-mM ammonium acetate; Mobile Phase B: 95: 5 acetonitrile: water with 20-mM ammonium acetate; Gradient: 50-100 % B over 20 minutes, then a 5-minute hold at 100 % B; Flow: 20 mL/min. Fractions containing the desired product were combined and dried via centrifugal evaporation to afford 2-((5-cyano-2-fluoro-3-(3-(4-methylpiperazin-1-yl)azetidin-1-yl)phenyl)amino)-4-(cyclopropylamino)imidazo[2,1-f][1,2,4]triazine-7-carbonitrile (16.9 mg). MS: (ESI) m/z 48
Output: <ADD> 2-chloro-4-(cyclopropyl(4-methoxybenzyl)amino)imidazo[2,1-f][1,2,4]triazine-7-carbonitrile (Intermediate 9) 60 mg 0.169 mmol 3-amino-4-fluoro-5-(3-(4-methylpiperazin-1-yl)azetidin-1-yl)benzonitrile 48 mg 0.166 mmol DPPF 6.44 mg 0.012 mmol Cs2CO3 92 mg 0.282 mmol Xantphos 9.60 mg 0.017 mmol Palladium(II)Acetate 11.17 mg 0.050 mmol 1,4-dioxane 2 mL <MIX> <APPARATUSACTION> vacuum <ADD> nitrogen <REPEAT> 3x <HEAT> 100 °C <WAIT> 3 h <COOL> 25 °C <ADD> EtOAc <WASH> brine <DRY> Na2SO4 <REMOVE> <ADD> DCE 1 mL TFA 0.5 mL <STIR> 3 h <REMOVE> <PURIFY> preparative LC/MS; Column Waters XBridge C18 19 x 200 mm 5-um particles; Guard Column Waters XBridge C18 19 x 10 mm 5-um particles; Mobile Phase A water with 20-mM ammonium acetate; Mobile Phase B 95: 5 acetonitrile: water with 20-mM ammonium acetate; Gradient 50-100 % B over 20 minutes then a 5-minute hold at 100 % B; Flow 20 mL/min <MIX> <CENTRIFUGE> <YIELD> 16.9 mg
Procedure: ***<<experimental procedure from dataset inserted here>>***

ICL-Template-3
<<role: user>>
Given a chemical synthesis procedure, summarize the procedure in structured output. Use exclusively the following markup tags in the structured output, no other tags: <ADD>, <YIELD>, <DISSOLVE>, <STIR>, <WASH>, <DRY>, <CONCENTRATE>, <PURIFY>, <REMOVE>, <FILTER>, <HEAT>, <EXTRACT>, <COOL>, <SYNTHESIZE>, <WAIT>, <PARTITION>, <DEGASS>, <QUENCH>, <RECOVER>, <APPARATUSACTION>, <PRECIPITATE>, <MIX>, <ADJUSTPH>, <REPEAT>, <SONICATE>, <CENTRIFUGE>. Give only the output, no explanation. Here is an example:
Procedure: In a flame-dried 100 mL round bottom flask, a mixture of 2.0 mL of furfurylamine (22 mmol) and 3 mL of triethylamine were stirred in 45 mL of dry dichloromethane under nitrogen at 0 °C. Then, 4.4 g of 1-adamantane carbonylchloride (22 mmol) in 5 mL of dry dichloromethane was added slowly, and the solution was allowed to warm to room temperature. After stirring for 1 h at room temperature, the solution was washed with 40 mL of an aqueous ammonium chloride solution (saturated) and 40 mL of an aqueous potassium carbonate solution (5%), the organic layer was separated, dried over MgSO4, filtered, and evaporated to dryness in vacuo. The crude product was recrystallized from heptane/EtOAc = 1:1 (v/v) to yield the product as off-white needles (3.15 g, 55%) and analyzed with 1H-NMR.

Output: <ADD> furfurylamine 2.0 mL 22 mmol triethylamine 3 mL dichloromethane 45 mL <COOL> 0 °C <MIX> 1-adamantane carbonylchloride 22 mmol 4.4 g dry dichloromethane 5 mL <ADD> slowly <HEAT> room temperature <STIR> 1 h <WASH> aqueous ammonium chloride solution (saturated) 40 mL aqueous potassium carbonate solution (5%) 40 mL <REMOVE> organic layer <DRY> MgSO4 <FILTER> <REMOVE> in vacuo <PURIFY> heptane/EtOAc = 1:1 (v/v) <YIELD> off-white needles 3.15 g 55% <APPARATUSACTION> analyze 1H-NMR
Procedure: ***<<experimental procedure from dataset inserted here>>***
<<role: assistant>>
<<output generated by Llama-3.1-8B-Instruct inserted here>>
<<role: user>>
This output contains extra tags, such as: ***<<list of undesired tags inserted here>>***. Please use only the following tags: <ADD>, <YIELD>, <DISSOLVE>, <STIR>, <WASH>, <DRY>, <CONCENTRATE>, <PURIFY>, <REMOVE>, <FILTER>, <HEAT>, <EXTRACT>, <COOL>, <SYNTHESIZE>, <WAIT>, <PARTITION>, <DEGASS>, <QUENCH>, <RECOVER>, <APPARATUSACTION>, <PRECIPITATE>, <MIX>, <ADJUSTPH>, <REPEAT>, <SONICATE>, <CENTRIFUGE>. Give only the output, no explanation.

Table S1: Tags (gray background) and their substitutions (white background) applied to dataset_llama_raw before the first re-annotation.

<COMBINE>	<MIX>	<SONICATE>	<APPARATUSACTION> sonicate
<SEPARATE>	<REMOVE>	<AZEOTROPE>	<PURIFY>
<ELUTE>	<WASH>	<ANALYZE>	<APPARATUSACTION> analyze
<TRITURATE>	<MIX>	<DISTILL>	<PURIFY>
<DILUTE>	<ADD>	<PURGE>	<ADD>
<EVAPORATE>	<REMOVE>	<CHROMATOGRAPH>	<APPARATUSACTION> chromatograph
<DROPWISE>	<ADD> dropwise	<LOAD>	<ADD>
<RINSE>	<WASH>	<MS>	<APPARATUSACTION> MS
<NEUTRALIZE>	<ADJUSTPH>	<REFLUX>	<HEAT> reflux
<COLLECT>	<RECOVER>	<CRYSTALLIZE>	<PURIFY>
<ACIDIFY>	<ADJUSTPH>	<MONITOR>	<APPARATUSACTION> monitor
<LYOPHILIZE>	<DRY>	<HYDROGENATE>	<ADD> H2
<WARM>	<HEAT>	<MICROWAVE>	<HEAT> microwave
<REACT>	<MIX>		

Table S2: Tags (gray background) and their substitutions (white background) applied to dataset_llama_raw_1 before further re-annotation (including the substitutions already mentioned in Table S1).

<FLOW>	; flow	<DEPOSIT>	<ADD>
<GRADIENT>	; gradient	<DISPENSE>	<ADD>
< 2,4 >	2,4	<EXPOSE>	<ADD>
<DILUTE>	<ADD>	<FILL>	<ADD>
<PURGE>	<ADD>	<INJECT>	<ADD>
<LOAD>	<ADD>	<INJECTION>	<ADD>
<FLUSH>	<ADD>	<LAYER>	<ADD>
<BUBBLE>	<ADD>	<PREADSORB>	<ADD>
<ADSORB>	<ADD>	<RECHARGE>	<ADD>
<TRANSFER>	<ADD>	<RE-DILUTE>	<ADD>
<POUR>	<ADD>	<REFILL>	<ADD>
<TREAT>	<ADD>	<SATURATE>	<ADD>
<ADDSOLVENT>	<ADD>	<SEED>	<ADD>
<ADMINISTER>	<ADD>	<SERIALDILUTION>	<ADD>
<APPLY>	<ADD>	<AERATE>	<ADD> air
<BACKFILL>	<ADD>	<AIR>	<ADD> air
<COAT>	<ADD>	<VENT>	<ADD> air

<ARGON>	<ADD> Argon
<BOC2O>	<ADD> BOC2O
<CHARCOAL>	<ADD> charcoal
<CHARCOALISE>	<ADD> charcoal
<CHARCOALIZE>	<ADD> charcoal
<ADDCARBONDISULFIDE>	<ADD> carbon disulfide
<HYDROGENATE>	<ADD> hydrogen
<HYDROGEN>	<ADD> hydrogen
<ADDMETHANOL>	<ADD> methanol
<N2>	<ADD> N2
<OXYGEN>	<ADD> oxygen
<ADDSILICA>	<ADD> silica
<ADDSILICA GEL>	<ADD> silica gel
<DROPWISE>	<ADD> slowly
<DROP>	<ADD> slowly
<DRIP>	<ADD> slowly
<GRADUALADD>	<ADD> slowly
<GRADUALLYADD>	<ADD> slowly
<SLOWADD>	<ADD> slowly
<WATER>	<ADD> water
<ACIDIFY>	<ADJUSTPH>
<ACID>	<ADJUSTPH>
<BASE>	<ADJUSTPH>
<pH>	<ADJUSTPH>
<PH>	<ADJUSTPH>
<READJUSTPH>	<ADJUSTPH>
<pH 3>	<ADJUSTPH> 3
<NEUTRALIZE>	<ADJUSTPH> 7.0
<ACIDICMEDIUM>	<ADJUSTPH> acidic
<MADEACIDIC>	<ADJUSTPH> acidic
<MAKEACIDIC>	<ADJUSTPH> acidic
<MAKEDACIDIC>	<ADJUSTPH> acidic
<BASEIFY>	<ADJUSTPH> basic
<BASICIFY>	<ADJUSTPH> basic
<NEUTRALISE>	<ADJUSTPH> neutralize
<NEUTRALIZATION TITRATION>	<ADJUSTPH> neutralize
<APPARATUSEACTION>	<APPARATUSACTION>
<EXAMINE>	<APPARATUSACTION>
<ANALYZE>	<APPARATUSACTION> analyze
<ANALYSE>	<APPARATUSACTION> analyze
<ANALYSIS>	<APPARATUSACTION> analyze
<CHARACTERISE>	<APPARATUSACTION> analyze
<CHARACTERIZE>	<APPARATUSACTION> analyze
<DETECT>	<APPARATUSACTION> analyze
<READ>	<APPARATUSACTION> analyze
<SCINTILLATIONCOUNT>	<APPARATUSACTION> analyze
<CHIRAL RESOLUTION>	<APPARATUSACTION> chiral resolution
<CLASSIFY>	<APPARATUSACTION> classify
<CUT>	<APPARATUSACTION> cut
<EXTRUDE>	<APPARATUSACTION> extrude
<EXTRUSION>	<APPARATUSACTION> extrude
<SHEAR>	<APPARATUSACTION> extrude
<CRUSH>	<APPARATUSACTION> grind
<GRIND>	<APPARATUSACTION> grind
<GROUND>	<APPARATUSACTION> grind
<MILL>	<APPARATUSACTION> grind
<PULVERIZE>	<APPARATUSACTION> grind
<SIZE REDUCE>	<APPARATUSACTION> grind
<WETGRANULATE>	<APPARATUSACTION> grind
<WETPULVERIZATION>	<APPARATUSACTION> grind
<HIGHVACUUM>	<APPARATUSACTION> high vacuum
<HPLC>	<APPARATUSACTION> HPLC
<ILLUMINATE>	<APPARATUSACTION> illuminate
<ILLUMINATION>	<APPARATUSACTION> illuminate
<IRRADIATE>	<APPARATUSACTION> irradiate
<JETMILL>	<APPARATUSACTION> jetmill
<MEASURE>	<APPARATUSACTION> measure
<Fitzmilled>	<APPARATUSACTION> mill
<MONITOR>	<APPARATUSACTION> monitor
<MPLC>	<APPARATUSACTION> MPLC
<NMR>	<APPARATUSACTION> NMR
<PREPARATIVEHPLC>	<APPARATUSACTION> preparative HPLC
<COMPRESSION>	<APPARATUSACTION> press
<PRESS>	<APPARATUSACTION> press
<SPHERONISE>	<APPARATUSACTION> spheronize
<SPHERONIZE>	<APPARATUSACTION> spheronize
<STERILIZE>	<APPARATUSACTION> sterilize
<PRESSTABLET>	<APPARATUSACTION> tablet
<TABLET>	<APPARATUSACTION> tablet

<TEST>	<APPARATUSACTION> test
<TLC>	<APPARATUSACTION> TLC
<EVACUATE>	<APPARATUSACTION> vacuum
<VACUO>	<APPARATUSACTION> vacuum
<VACUUM>	<APPARATUSACTION> vacuum
<vacuum>	<APPARATUSACTION> vacuum
<SPIN>	<CENTRIFUGE>
<SPINNING>	<CENTRIFUGE>
<COCONCENTRATE>	<CONCENTRATE>
<CO-CONCENTRATE>	<CONCENTRATE>
<PARTIAL CONCENTRATE>	<CONCENTRATE>
<RECONCENTRATE>	<CONCENTRATE>
<RE-CONCENTRATE>	<CONCENTRATE>
<FREEZE>	<COOL>
<CHILL>	<COOL>
<CHILLED>	<COOL>
<COLD>	<COOL>
<RECOOL>	<COOL>
<REFREEZE>	<COOL>
<REFRIGERATE>	<COOL>
<0 °C>	<COOL> 0 °C
<ICE BATH>	<COOL> icebath
<ICEBATH>	<COOL> icebath
<ICEWATERBATH>	<COOL> icebath
<SUSPEND>	<DISSOLVE>
<DISOLVE>	<DISSOLVE>
<DISPERS>	<DISSOLVE>
<Dissolve>	<DISSOLVE>
<PARTIALDISSOLVE>	<DISSOLVE>
<RECONSTITUTE>	<DISSOLVE>
<RE-CONSTITUTE>	<DISSOLVE>
<REDDISSOLVE>	<DISSOLVE>
<REDESOLVE>	<DISSOLVE>
<REDISOLVE>	<DISSOLVE>
<REDISSOLVE>	<DISSOLVE>
<RE-DISSOLVE>	<DISSOLVE>
<REDOSSOLVE>	<DISSOLVE>
<RESUSPEND>	<DISSOLVE>
<RE-SUSPEND>	<DISSOLVE>
<AIR DRY>	<DRY>
<AIRDRY>	<DRY>
<AZEOTROPIC DRY>	<DRY>
<AZEOTROPICALLY DRY>	<DRY>
<AZOTROPIC DRYING>	<DRY>
<DESICCATE>	<DRY>
<DRI>	<DRY>
<DRIE>	<DRY>
<DRIED>	<DRY>
<DRIEDRY>	<DRY>
<DRIEDRYNESS>	<DRY>
<DRIER>	<DRY>
<DRIEVACUO>	<DRY>
<DRIEVACUUM>	<DRY>
<DRIY>	<DRY>
<EXSICCATE>	<DRY>
<FREEZE DRY>	<DRY>
<FREEZEDRY>	<DRY>
<FREEZE-DRY>	<DRY>
<LIOPHILIZE>	<DRY>
<LIOPHYLIZE>	<DRY>
<LIPHOLIZE>	<DRY>
<LYOPHALLIZE>	<DRY>
<LYOPHILISATION>	<DRY>
<LYOPHILISE>	<DRY>
<LYOPHILIZATION>	<DRY>
<LYOPHILLIZE>	<DRY>
<LYOPHILYZE>	<DRY>
<LYOPHOLISE>	<DRY>
<LYOPHOLIZE>	<DRY>
<LYOPHOLYZE>	<DRY>
<LYPHILIZATION>	<DRY>
<LYPHOLISE>	<DRY>
<LYPHOLIZE>	<DRY>
<LYPHOLYZE>	<DRY>
<OVEN DRY>	<DRY>
<PARTIAL DRY>	<DRY>
<SPRAYDRY>	<DRY>
<SUBLIMATE>	<DRY>
<VACUUMDRY>	<DRY>
<BACK EXTRACT>	<EXTRACT>
<BACKEXTRACT>	<EXTRACT>
<BACK-EXTRACT>	<EXTRACT>
<REEXTRACT>	<EXTRACT>
<RE-EXTRACT>	<EXTRACT>
<SOXHLET>	<EXTRACT>
<ADSORPTIVEFILTER>	<FILTER>
<NATURALFILTER>	<FILTER>
<PASS>	<FILTER>
<SIEVE>	<FILTER>
<SIZE>	<FILTER>
<SYRINGEFILTER>	<FILTER>
<CHARCOALFILTERATION>	<FILTER> charcoal

<WARM>	<HEAT>
<BAKE>	<HEAT>
<CALCINE>	<HEAT>
<COBOIL>	<HEAT>
<HEAT-GUN>	<HEAT>
<MELT>	<HEAT>
<REHEAT>	<HEAT>
<SINTER>	<HEAT>
<THAW>	<HEAT>
<AZEOTROPICBOIL>	<HEAT> boil
<BOIL>	<HEAT> boil
<MICROWAVE>	<HEAT> microwave
<REFLUX>	<HEAT> reflux
<AZEOTROPICREFLUX>	<HEAT> reflux
<COMBINE>	<MIX>
<TRITURATE>	<MIX>
<REACT>	<MIX>
<HOMOGENISE>	<MIX>
<HOMOGENIZE>	<MIX>
<POOL>	<MIX>
<RESLURRY>	<MIX>
<RE-SLURRY>	<MIX>
<RESURRY>	<MIX>
<SLURRY>	<MIX>
<Triturate>	<MIX>
<TRITURATION>	<MIX>
<VORTEX>	<MIX>
<WETMIX>	<MIX>
<SEPARATE>	<PARTITION>
<LIQUIDSEPARATION>	<PARTITION>
<PHASESEPARATE>	<PARTITION>
<PHASESEPARATION>	<PARTITION>
<REPARTITION>	<PARTITION>
<SOLUTION SEPARATION>	<PARTITION>
<CRYSTALLIZE>	<PRECIPITATE>
<CO-PRECIPITATE>	<PRECIPITATE>
<CRISTALISE>	<PRECIPITATE>
<CRYSTALIZE>	<PRECIPITATE>
<GRANULATE>	<PRECIPITATE>
<PRECIPTATE>	<PRECIPITATE>
<QUICKCRYSTALLIZE>	<PRECIPITATE>
<RECIPITATE>	<PRECIPITATE>
<REPRECIPITATE>	<PRECIPITATE>
<RE-PRECIPITATE>	<PRECIPITATE>
<SCRATCH>	<PRECIPITATE>
<SOLIDIFY>	<PRECIPITATE>
<DISTILL>	<PURIFY>
<LYOPHILIZE>	<PURIFY>
<AZEOTROPE>	<PURIFY>
<CHROMATOGRAPH>	<PURIFY>
<COLUMN>	<PURIFY>
<AZEOTROP>	<PURIFY>
<AZEOTROPIC DISTILL>	<PURIFY>
<AZEOTROPIC DISTILLATION>	<PURIFY>
<AZEOTROPIC>	<PURIFY>
<AZEOTROPICDISTILL>	<PURIFY>
<AZEOTROPICDISTILLATION>	<PURIFY>
<AZEOTROPY>	<PURIFY>
<AZETROPE>	<PURIFY>
<AZOTROPE>	<PURIFY>
<BULB-TO-BULB DISTILL>	<PURIFY>
<BULB-TO-BULB DISTILLATION>	<PURIFY>
<BULBTOBULBDISTILLATE>	<PURIFY>
<BULBTOBULBDISTILLATION>	<PURIFY>
<CHIRAL PREPARATIVE CHROMATOGRAPHY>	<PURIFY>
<CHIRALCHROMATOGRAPHY>	<PURIFY>
<CHROMATOGRAPHY>	<PURIFY>
<CO DISTILL>	<PURIFY>
<CODISTILL>	<PURIFY>
<CO-DISTILL>	<PURIFY>
<CODISTILLATE>	<PURIFY>
<CO-DISTILLATE>	<PURIFY>
<COLUMNCHROMATOGRAPH>	<PURIFY>
<COLUMNCHROMATOGRAPHY>	<PURIFY>
<COOLDISTILL>	<PURIFY>
<CRISTALIZE>	<PURIFY>
<CRYSTAL>	<PURIFY>
<CRYSTALISE>	<PURIFY>
<CRYSTALLISE>	<PURIFY>
<DIALYSE>	<PURIFY>
<DIALYZE>	<PURIFY>
<DISTIL>	<PURIFY>
<DISTILLATE>	<PURIFY>
<DISTILLATION>	<PURIFY>
<FLASH COLUMN CHROMATOGRAPHY>	<PURIFY>
<FLASH>	<PURIFY>
<FLASHCHROMATOGRAPH>	<PURIFY>
<FLASHCHROMATOGRAPHY>	<PURIFY>
<FLASHCOLUMN>	<PURIFY>
<FLASHCOLUMNCHROMATOGRAPHY>	<PURIFY>
<FLASHDISTILLATION>	<PURIFY>
<FLASHVACUUMDISTILLATION>	<PURIFY>
<FRACDIST>	<PURIFY>

<FRACTIONAL DISTILL>	<PURIFY>
<KUGELDISTILLATION>	<PURIFY>
<KUGELROHR>	<PURIFY>
<KUGELROHRDISTILL>	<PURIFY>
<KUGELROHRDISTILLATION>	<PURIFY>
<PAPERCHROMATOGRAPHY>	<PURIFY>
<PARTIAL PURIFICATION>	<PURIFY>
<RADIALCHROMATOGRAPHY>	<PURIFY>
<RECHROMATOGRAPHY>	<PURIFY>
<RECRISTALIZE>	<PURIFY>
<RECRYSTAL>	<PURIFY>
<RECRYSTALISE>	<PURIFY>
<RECRYSTALIZE>	<PURIFY>
<RECRYSTALLISE>	<PURIFY>
<RE-CRYSTALLISE>	<PURIFY>
<RECRYSTALLIZE>	<PURIFY>
<RE-CRYSTALLIZE>	<PURIFY>
<REDISTILL>	<PURIFY>
<REDISTILLATE>	<PURIFY>
<RE-PURIFY>	<PURIFY>
<SHORT-PATH DISTILLATION>	<PURIFY>
<STEAM DISTILLATION>	<PURIFY>
<STEAMDISTILLATE>	<PURIFY>
<STEAMDISTILLATION>	<PURIFY>
<SUB>	<PURIFY>
<SUBLIMATION>	<PURIFY>
<SUBLIME>	<PURIFY>
<SUBLINE>	<PURIFY>
<VACUUMACTION>	<PURIFY>
<VACUUMDISTILL>	<PURIFY>
<VACUUMDISTILLATE>	<PURIFY>
<VACUUMDISTILLATION>	<PURIFY>
<VACUUMFRACTIONALDISTILLATE>	<PURIFY>
<VIGREUXDISTILLATION>	<PURIFY>
<WIDMERDISTILLATION>	<PURIFY>
<WIPED-FILM DISTILLATION>	<PURIFY>
<COLLECT>	<RECOVER>
<ISOLATE>	<RECOVER>
<COEVAPORATE>	<REMOVE>
<AZEOTROPIC EVAPORATE>	<REMOVE>
<AZEOTROPIC EVAPORATION>	<REMOVE>
<AZEOTROPICALLY EVAPORATE>	<REMOVE>
<CO-evaporate>	<REMOVE>
<FAST EVAPORATE>	<REMOVE>
<REEVAPORATE>	<REMOVE>
<RE-EVAPORATE>	<REMOVE>
<ROTARYEVAPORATE>	<REMOVE>
<ROTAVAP>	<REMOVE>
<ROTAVAPOR>	<REMOVE>
<STRIP>	<REMOVE>
<EVAPORATE>	<REMOVE>
<DECANT>	<REMOVE>
<CO-EVAPORATE>	<REMOVE>
<DISCARD>	<REMOVE>
<DISPOSE>	<REMOVE>
<DRAIN>	<REMOVE>
<REMOVED>	<REMOVE>
<Sonicate>	<SONICATE>
<AGITATE>	<STIR>
<SHAKE>	<STIR> shaking
<SYNTHESISE>	<SYNTHESIZE>
<OR>	<SYNTHESIZE> alternative
<INCUBATE>	<WAIT>
<STAND>	<WAIT>
<STORAGE>	<WAIT>
<STORE>	<WAIT>
<OVERNIGHT>	<WAIT> overnight
<WEEKEND>	<WAIT> weekend
<ELUTE>	<WASH>
<RINSE>	<WASH>
<BACKWASH>	<WASH>
<CHASE>	<WASH>
<ELUATE>	<WASH>
<ELUENT>	<WASH>
<RUN>	<WASH>
<SELECT>	<YIELD>
<TOTAL YIELD>	<YIELD>
<MS>	*[delete tag and parameters]*
<USE>	*[delete tag and parameters]*
<SEAL>	*[delete tag and parameters]*
<13C NMR>	*[delete tag and parameters]*
<1H NMR>	*[delete tag and parameters]*
<1H-NMR>	*[delete tag and parameters]*
<CALCULATE>	*[delete tag and parameters]*
<COUNT>	*[delete tag and parameters]*
<M.p.>	*[delete tag and parameters]*
<M.P>	*[delete tag and parameters]*
<mp>	*[delete tag and parameters]*
<OPEN>	*[delete tag and parameters]*
<OPTIONAL>	*[delete tag and parameters]*

<QUANTIFY>	*[delete tag and parameters]*	<MP>	*[delete tag and parameters]*
<RECORD>	*[delete tag and parameters]*	<MS>	*[delete tag and parameters]*
<WEIGH>	*[delete tag and parameters]*	<USE>	*[delete tag and parameters]*
<WET>	*[delete tag and parameters]*		

Built-in substitution heuristics and error corrections:

The node editor will automatically apply certain heuristics and error correction measures when building the node graph from the action graph to handle imprecise experimental descriptions or missing parameters from the synthesis protocol and inform the user of the measures taken through a warning message.

For example, if a stirring step is described as “slow” in the natural language input, a value of “100 rpm” will be used, while for “vigorous” or “quick” or “fast” stirring, a speed of “600 rpm” is assumed. Similarly, a time duration given as “overnight” will be changed to “16 h”, and if a temperature is specified as “room temperature”, it will be set to “25 °C”. For some process steps, default values will be used if certain required values are missing. For example, if a heating/stirring step does not specify a time, temperature and/or speed, 5 min, 25 °C, and 300 rpm will be used as the default for the missing value(s), while for a centrifugation step, 15 min at 8000 rpm and 25 °C are the default, and a sonication step without a specified time will be done for 10 min. In more severe cases, such as a missing or imprecise amount of a chemical, a warning will be generated and the node will be created, but the corresponding field of the node will be left empty rather than trying to “guess” a default value. This gives the user the opportunity to correct this warning manually at the node graph level. If no value is given, and the node graph is compiled to a python script and attempted to run, the Minerva-OS backend will produce an error when it comes to this step.

These substitutions are currently hard coded into the node generation step, and again somewhat specific to our use case of nano and advanced material synthesis. Other hardware platforms will most probably require different default values and heuristics.

Table S3: Example output from each model (Column1: BigBirdPegasus_ChemicalTagger; Column2: LED-Base-16384_ChemicalTagger; Column3: BigBirdPegasus_Llama; Column4: LED-Base-16384_Llama) for exemplary procedures from the domains of materials science (examples 1-4), patents (example 5), organic chemistry (example 6), and inorganic chemistry (example 7). Line breaks, indentations, and bold text formatting were added for clarity.

Example 1: Unfunctionalized mesoporous silica nanoparticles were synthesized according to a published procedure.5 In brief, 200 mg of CTAB and 1200 µL of sodium hydroxide solution (1.000 M) were dissolved in 100 mL of water under stirring. The solution was heated at 80 °C for 30 minutes, followed by the addition of 1050 µL of TEOS under vigorous stirring. Stirring was continued for 2 h at 80 °C, and the solution was allowed to cool to room temperature. The nanoparticles were collected by centrifugation (10 min at 7197 rcf), washed 2x with water (2x 90 mL), 2x with ethanol (2x 90 mL) and 2x with toluene (2x 90 mL), redispersed in 20 mL of dry toluene and directly used for further functionalization.			
<SYNTHESIZE> silica **<DISSOLVE>** CTAB 200 mg sodium hydroxide 1200 uL 1.000 M water 100 mL **<STIR>** **<HEAT>** 80 C 30 minutes **<ADD>** TEOS 1050 uL **<STIR>** **<WAIT>** 2 h 80 C **<COOL>** **<RECOVER>** 10 min 7197 rcf **<WASH>** water 90 mL ethanol 90 mL toluene 90 mL **<ADD>** toluene 20 m	**<SYNTHESIZE>** silica **<DISSOLVE>** CTAB 200 mg sodium hydroxide 1200 uL 1.000 M water 100 mL **<STIR>** **<HEAT>** 80 °C 30 minutes **<ADD>** TEOS 1050 uL **<STIR>** **<WAIT>** 2 h 80 °C **<COOL>** **<RECOVER>** 10 min 7197 rcf **<WASH>** water 90 mL ethanol 90 mL toluene 90 mL **<ADD>** toluene 20 mL	**<ADD>** CTAB 200 mg sodium hydroxide solution 1200 uL 1.000 M water 100 mL **<MIX>** **<HEAT>** 80 C 30 min **<ADD>** TEOS 1050 uL **<STIR>** 2 h **<COOL>** room temperature **<WASH>** water 2x 90 mL water 2x 90 mL ethanol 2x 90 mL ethanol 2x 90 mL toluene 2x 90 mL **<RECOVER>** dry toluene 20 mL	**<ADD>** CTAB 200 mg sodium hydroxide solution 1200 uL 1.000 M water 100 mL **<MIX>** **<HEAT>** 80 °C 30 min **<ADD>** TEOS 1050 uL **<STIR>** 2 h **<COOL>** room temperature **<WASH>** water 2x 90 mL ethanol 2x 90 mL toluene 2x 90 mL **<MIX>** toluene 20 mL **<SYNTHESIZE>**

Example 2: ZrOCl2·8 H2O (0.028 g, 0.086 mmol) was dissolved in DMF (5 mL) by ultrasonic treatment in a glass bottle (Schott Duran, Borosilicate 3.3, ISO4796, 100 mL) with a PBT cap equipped with a Teflon seal. Then acetic acid (980 µL, 17 mmol) and finally 4,4'-(anthracene-9,10-diyl)dibenzoic acid (PAP diacid; 0.036 g, 0.086 mmol) were added. The PAP diacid was dispersed through sonication. The bottle was placed in an oven preheated to 120 °C and kept at this temperature for 24 h. After removing the bottle from the oven and cooling the suspension to room temperature, the solid was isolated through centrifugation (Eppendorf centrifuge 5430, 7830 rpm, 7197 rcf, 5 min). It was washed by suspending it in DMF (30 mL, anhydrous) and subsequent centrifugation. Suspending and centrifugation were repeated two more times. Afterwards the obtained solid was suspended in unstabilized CH2Cl2 (30 mL) and the suspension was centrifuged. The isolated yellow powder was dried under reduced pressure (2.4·10-3 bar). The material was extracted for 48 h under nitrogen atmosphere in a Soxhlet device (5 mL) with unstabilized CH2Cl2 and with molecular sieve in the distillation flask. The remaining powdery solid in the extraction thimble was dried under reduced pressure (2.4·10-3 bar) and stored under nitrogen atmosphere.

<DISSOLVE> 3.3 , ZrOCl28 H2O 0.028 g 0.086 mmol DMF 5 mL Borosilicate PBT **<APPARATUSACTION>** Teflon **<ADD>** acetic acid 980 uL 17 mmol 4,4'-(anthracene-9,10-diyl)dibenzoic acid 0.036 g 0.086 mmol PAP diacid **<APPARATUSACTION>** **<WAIT>** 24 h **<REMOVE>** **<COOL>** **<YIELD>** 5430 , 7830 rpm 7197 rcf 5 min **<WASH>** **<DRY>** 2.410-3 **<EXTRACT>** 48 h nitrogen CH2Cl2 **<DRY>** 2.410-3	**<DISSOLVE>** 3.3 ISO4796 , ZrOCl2·8 H2O 0.028 g 0.086 mmol DMF 5 mL Borosilicate PBT **<APPARATUSACTION>** Teflon **<ADD>** acetic acid 980 uL 17 mmol 4,4'-(anthracene-9,10-diyl)dibenzoic acid 0.036 g 0.086 mmol Diacid **<APPARATUSACTION>** **<WAIT>** 24 h **<REMOVE>** **<COOL>** **<YIELD>** 5430 , 7830 rpm 7197 rcf 5 min **<WASH>** **<DRY>** 2.4·10-3 **<EXTRACT>** 48 h Nitrogen CH2Cl2 **<DRY>** 2.4·10-3	**<ADD>** ZrOCl28 H2O 0.028 g 0.086 mmol **<DISSOLVE>** DMF 5 mL **<SONICATE>** **<ADD>** acetic acid 980 uL 17 mmol 4,4'-(anthracene-9,10-diyl)dibenzoic acid 0.036 g 0.086 mmol **<MIX>** **<HEAT>** 120 C **<WAIT>** 24 h **<COOL>** **<WASH>** DMF 30 mL **<REPEAT>** 2x **<WASH>** DMF 30 mL **<REPEAT>** 2x **<WASH>** CH2Cl2 30 mL **<CENTRIFUGE>** **<DRY>** reduced pressure 2.410-3 bar **<EXTRACT>** nitrogen atmosphere Soxhlet device 5 mL unstabilized CH2Cl2 molecular sieve **<DRY>** reduced pressure 2.410-3 bar **<YIELD>**	**<ADD>** ZrOCl2·8 H2O 0.028 g 0.086 mmol **<DISSOLVE>** DMF 5 mL **<SONICATE>** **<ADD>** acetic acid 980 uL 17 mmol 4,4'-(anthracene-9,10-diyl)dibenzoic acid (PAP diacid) 0.036 g 0.086 mmol **<MIX>** **<APPARATUSACTION>** oven 120 °C **<WAIT>** 24 h **<COOL>** **<WASH>** DMF 30 mL **<REPEAT>** 2x **<WASH>** CH2Cl2 30 mL **<CENTRIFUGE>** **<DRY>** reduced pressure 2.4·10-3 bar **<EXTRACT>** Soxhlet device 5 mL unstabilized CH2Cl2 molecular sieve **<DRY>** reduced pressure 2.4·10-3 bar **<YIELD>**

Example 3: The solvothermal synthesis of NH2-MIL-101(Al) was carried out according to a slightly modified literature synthesis28. In a 50 mL glass reactor, aluminum(III) chloride hexahydrate (0.51 g, 2.11 mmol, 1.0 eq.) and 2-amino terephthalic acid (0.56 g, 3.09 mmol, 1.5 eq.) were dissolved in 30 mL N,N-dimethylformamide (DMF) in an ultrasonic bath. The sealed glass reactor was kept for 72 h in a preheated oven at 403 K. The resulting yellow powder was filtered under vacuum and washed with acetone. To remove organic species trapped within the pores, the samples were extracted in boiling methanol for 8 h and stored at 373 K.			
<SYNTHESIZE> NH2- FCC-101(Al) **<DISSOLVE>** aluminum(III) chloride hexahydrate 0.51 g 2.11 mmol 2-amino terephthalic acid 0.56 g 3.09 mmol N,N-dimethylformamide 30 mL **<WAIT>** 72 h 403 K **<FILTER>** **<WASH>** acetone **<REMOVE>** **<EXTRACT>**	**<SYNTHESIZE>** NH2-MIL-101(Al) **<DISSOLVE>** aluminum(III) chloride hexahydrate 0.51 g 2.11 mmol 2-amino terephthalic acid 0.56 g 3.09 mmol N,N-dimethylformamide 30 mL **<WAIT>** 72 h 403 K **<FILTER>** **<WASH>** Acetone **<REMOVE>** **<EXTRACT>**	**<ADD>** aluminum(III) chloride hexahydrate 0.51 g 2.11 mmol 2-amino terephthalic acid 0.56 g 3.09 mmol N,N-dimethylformamide (DMF) 30 mL **<MIX>** **<HEAT>** 403 K **<WAIT>** 72 h **<FILTER>** **<WASH>** acetone **<EXTRACT>** boiling methanol 8 h **<RECOVER>** 373 K	**<ADD>** aluminum(III) chloride hexahydrate 0.51 g 2.11 mmol 2-amino terephthalic acid 0.56 g 3.09 mmol N,N-dimethylformamide (DMF) 30 mL **<MIX>** **<HEAT>** 403 K **<WAIT>** 72 h **<FILTER>** **<WASH>** acetone **<EXTRACT>** boiling methanol 8 h **<RECOVER>** 373 K

Example 4: The preparation of colloidal gold was similar to that described by Turkevich, Stevenson, and Hillier,33: 100 mL of HAuCl, solution containing 50 mg of gold was added to 850 mL of distilled water boiling in a 2-L flask. When the solution was boiling again, 50 mL of 1% sodium citrate solution was added while stirring vigorously. After continuous boiling for 30 min, it was allowed to cool and made up to 1 L.

<SYNTHESIZE> gold **<ADD>** HAuCl 100 mL gold 50 mg water 850 mL 2-L sodium citrate **<STIR>** **<WAIT>** 30 min **<COOL>**	**<SYNTHESIZE>** gold **<ADD>** HAuCl 100 mL gold 50 mg water 850 mL 2-L sodium citrate 50 mL **<STIR>** **<WAIT>** 30 min **<COOL>**	**<ADD>** HAuCl 100 mL gold 50 mg **<MIX>** distilled water 850 mL **<HEAT>** boiling **<ADD>** sodium citrate solution 50 mL **<STIR>** vigorously **<WAIT>** 30 min **<COOL>** **<ADJUSTPH>** 1 L	**<ADD>** HAuCl 100 mL gold 50 mg distilled water 850 mL **<HEAT>** boiling **<ADD>** sodium citrate solution 1% 50 mL **<STIR>** vigorously **<HEAT>** boiling 30 min **<COOL>** **<ADJUSTPH>** 1 L

Example 5: A mixture of methyl 4-bromo-2-fluorobenzoate (116.5 g, 0.5 mol), 1H-pyrrolo[2,3-b]pyridin-5-ol (67 g, 0.5 mol) and K 2 CO 3 (138 g, 1.0 mol) in DMF (500 mL) was heated at 95° C. for about 16 h. The reaction mixture was cooled to ambient temperature, filtered and the filtrate was diluted with DCM (1 L). The resulting solution was washed with H 2 O (500 mL×2) and concentrated. The residue was recrystallized from EA (200 mL) and PE (400 mL) and the cake (68 g) was collected as the first batch. The filtrate was concentrated and dissolved with EA (500 mL). The solution was washed with H2O (200 mL x 2), concentrated, slurried with EA (25 mL) and PE (25 mL) at reflux for 1 h, cooled to ambient temperature, and filtered to give the second batch of product (38 g). The two batches of product were combined to give the product (106 g, 61.3%) as a brown solid. MS (ESI, m/e) [M+1] + 346.9, 348.9.

<ADD> 2 CO 3 (methyl 4-bromo-2-fluorobenzoate 116.5 g 0.5 mol 1H-pyrrolo[2,3-b]pyridin-5-ol 67 g 0.5 mol **<DISSOLVE>** 3 (DMF 500 mL **<HEAT>** **<WAIT>** 16 h **<COOL>** **<FILTER>** **<ADD>** DCM 1 L **<WASH>** 2 O 500 H **<CONCENTRATE>** **<PRECIPITATE>** **<RECOVER>** **<CONCENTRATE>** **<DISSOLVE>** **<WASH>** 2) H2O 200 mL **<CONCENTRATE>** **<HEAT>** 1 h **<COOL>** **<FILTER>** **<YIELD>** 61 (	**<ADD>** 2 CO 3 (methyl 4-bromo-2-fluorobenzoate 116.5 g 0.5 mol 1H-pyrrolo[2,3-b]pyridin-5-ol 67 g 0.5 mol **<DISSOLVE>** 3 (DMF 500 mL **<HEAT>** 95 ° C **<WAIT>** 16 h **<COOL>** **<FILTER>** **<ADD>** DCM 1 L **<WASH>** 2 O 500 mLx2 H **<CONCENTRATE>** **<PRECIPITATE>** **<RECOVER>** **<CONCENTRATE>** **<DISSOLVE>** **<WASH>** 2) H2O 200 mL **<CONCENTRATE>** **<HEAT>** 1 h **<COOL>** **<FILTER>** **<YIELD>**	**<ADD>** methyl 4-bromo-2-fluorobenzoate 116.5 g 0.5 mol 1H-pyrrolo[2,3-b]pyridin-5-ol 67 g 0.5 mol K2CO3 138 g 1.0 mol DMF 500 mL **<HEAT>** 95 C **<WAIT>** 16 h **<COOL>** ambient temperature **<FILTER>** **<MIX>** DCM 1 L **<WASH>** H2O 500 mLx2 **<CONCENTRATE>** **<MIX>** EA 200 mL PE 400 mL **<PURIFY>** EA 500 mL **<WASH>** H2O 200 mL x 2 **<CONCENTRATE>** **<MIX>** EA 25 mL PE 25 mL **<HEAT>** reflux **<COOL>** ambient temperature **<FILTER>** **<MIX>** **<YIELD>** brown solid 106 g 61 %	**<ADD>** methyl 4-bromo-2-fluorobenzoate 116.5 g 0.5 mol 1H-pyrrolo[2,3-b]pyridin-5-ol 67 g 0.5 mol K2CO3 138 g 1.0 mol DMF 500 mL **<HEAT>** 95 °C **<WAIT>** 16 h **<COOL>** ambient temperature **<FILTER>** **<WASH>** H2O 500 mLx2 **<CONCENTRATE>** **<MIX>** EA 200 mL PE 400 mL **<PURIFY>** EA **<WASH>** H2O 200 mL x 2 **<CONCENTRATE>** **<MIX>** EA 25 mL PE 25 mL **<HEAT>** reflux **<WAIT>** 1 h **<COOL>** ambient temperature **<FILTER>** **<MIX>** **<YIELD>** brown solid 106 g 61.3 %

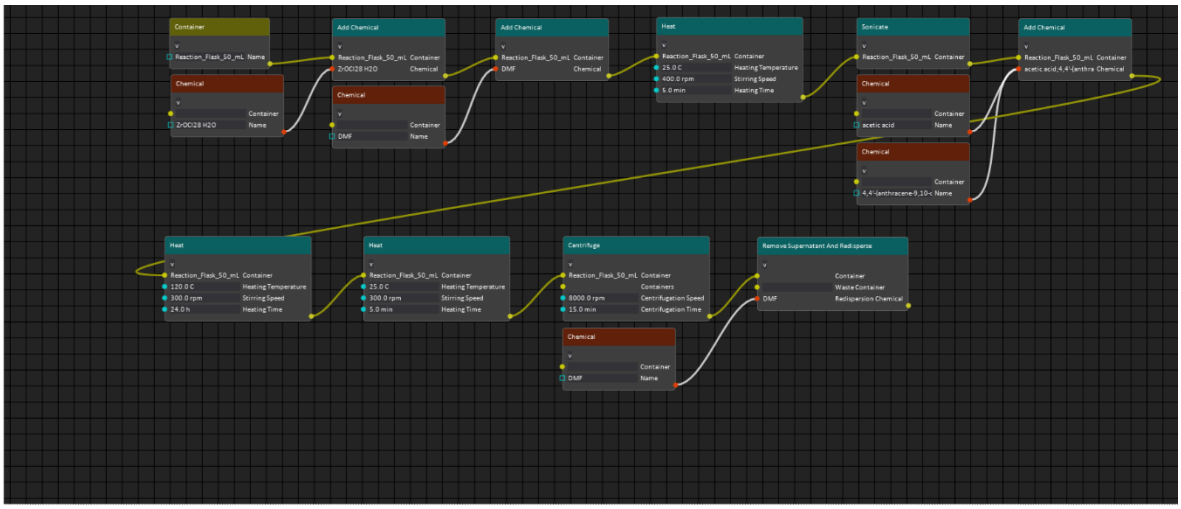
Example 6: 17-ethynylestradiol (888 mg, 3 mmol) and (S)-3-azidopropane-1,2-diol (352 mg, 3 mmol) were suspended in a 1:1 mixture of water and tert-butyl alcohol (12 mL). Sodium ascorbate (0.3 mmol, 300 µL of freshly prepared 1 M solution in water) was added, followed by copper(II) sulfate pentahydrate (7.5 mg, 0.03 mmol, in 100 µL of water). The heterogeneous mixture was stirred vigorously overnight, at which point it cleared and TLC analysis indicated complete consumption of the reactants. The reaction mixture was diluted with water (50 mL), cooled in ice, and the white precipitate was collected by filtration. After washing the precipitate with cold water (2×25 mL), it was dried under vacuum to afford 1.17 g (94 %) of pure product as an off-white powder.

<ADD> 17-ethynylestradiol 888 mg 3 mmol (S)-3-azidopropane-1,2-diol 352 mg 3 mmol water tert-butyl alcohol 12 mL **<SYNTHESIZE>** Sodium ascorbate **<DISSOLVE>** water **<ADD>** **<STIR>** **<REMOVE>** **<ADD>** water 50 mL **<COOL>** **<RECOVER>** **<WASH>** water 2x25 mL **<DRY>** **<YIELD>**	**<ADD>** 1 :1 of 17-ethynylestradiol 888 mg 3 mmol (S)-3-azidopropane-1,2-diol 352 mg 3 mmol water tert-butyl alcohol 12 mL **<SYNTHESIZE>** Sodium ascorbate **<DISSOLVE>** water **<ADD>** **<STIR>** **<REMOVE>** **<ADD>** water 50 mL **<COOL>** **<RECOVER>** **<WASH>** water 2x25 mL **<DRY>** **<YIELD>**	**<ADD>** 17-ethylesyntradiol 888 mg 3 mmol (S)-3-azidopropane-1,2-diol 352 mg 3 mmol water 6 mL tert-butyl alcohol 6 mL **<MIX>** sodium ascorbate 0.3 mmol 300 uL 1 M solution in water copper(II) sulfate pentahydrate 7.5 mg 0.03 mmol 100 uL water **<STIR>** overnight **<WASH>** cold water 25 mL cold water 25 mL **<DRY>** vacuum **<YIELD>** off-white powder 1.17 g 94 %	**<ADD>** 17-ethynylestradiol 888 mg 3 mmol (S)-3-azidopropane-1,2-diol 352 mg 3 mmol water tert-butyl alcohol 12 mL **<MIX>** sodium ascorbate 0.3 mmol 300 uL 1 M solution in water copper(II) sulfate pentahydrate 7.5 mg 0.03 mmol 100 uL water **<STIR>** overnight **<WASH>** cold water 2x25 mL **<DRY>** vacuum **<YIELD>** off-white powder 1.17 g 94 %

Example 7: A 100 ml 2-necked round bottom flask was charged with a mixture of aqueous HI (6.8 ml, 7.58M) and aqueous H3PO2 (1.7 ml, 9.14M). The liquid was degassed by passing a stream of nitrogen through it for 1 min and keeping it under a nitrogen atmosphere throughout the experiment. SnI2 (372 mg, 1 mmol) was dissolved in the mixture upon heating the flask to 120oC using an oil bath, under constant magnetic stirring, forming a bright yellow solution. To the hot yellow solution was added solid CH3NH3I (159 mg, 1 mmol) which dissolved immediately. The solution was evaporated to approximately half its original volume by heating at 120oC. The stirring was discontinued and the solution was left to cool back to room temperature. Upon cooling, black, elongated, rhombic dodecahedral (12 faces) crystals of the title compound were precipitated. The crystals were left to grow inside the mother liquor for a further 24 h under a nitrogen atmosphere before being filtered and washed copiously with degassed EtOH. Yield 70-90%.

<ADD> 7.58M) HI 6.8 ml H3PO2 1.7 ml **<DEGASS>** **<DISSOLVE>** SnI2 372 mg 1 mmol **<HEAT>** 120oC **<YIELD>** **<ADD>** CH3NH3I 159 mg 1 mmol **<DISSOLVE>** **<REMOVE>** half its **<HEAT>** 120oC **<WAIT>** **<COOL>** **<PRECIPITATE>** 12 faces dodeca **<WAIT>** 24 h nitrogen **<FILTER>** **<WASH>** EtOH **<YIELD>**	**<ADD>** HI 6.8 ml H3PO2 1.7 ml **<DEGASS>** **<DISSOLVE>** SnI2 372 mg 1 mmol **<HEAT>** 120oC **<YIELD>** **<ADD>** CH3NH3I 159 mg 1 mmol **<DISSOLVE>** **<REMOVE>** half its **<HEAT>** 120oC **<WAIT>** **<COOL>** **<PRECIPITATE>** 12 cells **<WAIT>** 24 h nitrogen **<FILTER>** **<WASH>** EtOH **<YIELD>** 70-90 %	**<ADD>** aqueous HI 6.8 mL 7.58M aqueous H3PO2 1.7 mL 9.14M **<DEGASS>** nitrogen 1 min **<MIX>** SnI2 372 mg 1 mmol **<HEAT>** 120 C **<STIR>** **<ADD>** CH3NH3I 159 mg 1 mmol **<HEAT>** 120 C **<CONCENTRATE>** **<COOL>** room temperature **<PRECIPITATE>** black elongated rhombic dodecahedral 12 **<WAIT>** 24 h **<FILTER>** **<WASH>** degassed EtOH **<YIELD>** black elongated rhombic dodecahedral 12	**<ADD>** aqueous HI 6.8 mL 7.58 M aqueous H3PO2 1.7 mL 9.14 M **<DEGASS>** **<HEAT>** 120 °C **<MIX>** SnI2 372 mg 1 mmol **<ADD>** CH3NH3I 159 mg 1 mmol **<HEAT>** 120 °C **<COOL>** room temperature **<PRECIPITATE>** black octombic dodecahedral (12at) crystals **<WAIT>** 24 h **<FILTER>** **<WASH>** degassed EtOH **<YIELD>** 70-90 %

a)



b)

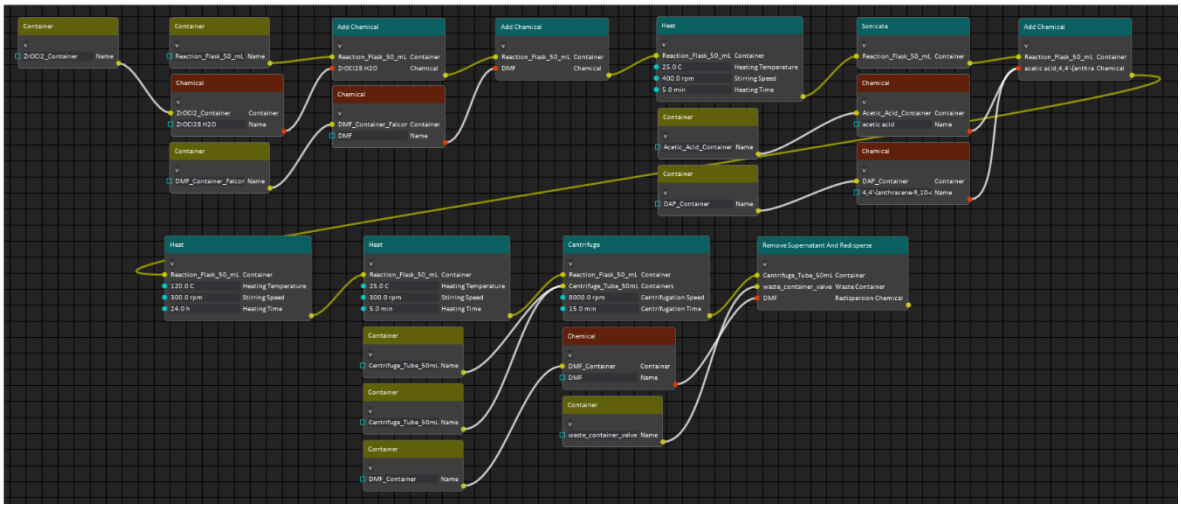


Figure S1: Automatically generated node graphs from the input shown in Table S3, Example 2. a) raw output. For clarity, the reaction flask was renamed and only one washing step is shown (the original experimental procedure, action graph and generated node graph contained 3 washing steps). b) user-modified node graph with the containers for the reactants added manually.

Table S4: Automatically generated python code from the node graph in Figure S4.

```
# Automatically created with Minerva Node Editor
import threading
from Minerva import *

if __name__ == '__main__':
    Configuration.load_configuration("Reaction_UiO_Example_Simpilfied.json")

    # Containers
    reaction_flask_50_ml = Container(current_hardware='flaskstation', deck_position=0,
slot_number=1, name='Reaction Flask 50 mL', current_volume=None, max_volume=None,
container_type=ContainerTypeCollection.FLASK_50_ML, has_stirbar=True, is_capped=False)
    zrocl2_container = Container(current_hardware='ot2_holder_15ml', deck_position=0,
slot_number=1, name='ZrOCl2_Container', current_volume='15 mL', max_volume=None,
container_type=ContainerTypeCollection.FALCON_TUBE_15_ML, has_stirbar=False,
is_capped=False)
    dmf_container_falcon_tube = Container(current_hardware='ot2_holder_15ml',
deck_position=0, slot_number=2, name='DMF_Container_Falcon_Tube', current_volume='15 mL',
max_volume=None, container_type=ContainerTypeCollection.FALCON_TUBE_15_ML,
has_stirbar=False, is_capped=False)
    acetic_acid_container = Container(current_hardware='ot2_holder_15ml', deck_position=0,
slot_number=3, name='Acetic_Acid_Container', current_volume='15 mL', max_volume=None,
container_type=ContainerTypeCollection.FALCON_TUBE_15_ML, has_stirbar=False,
is_capped=False)
    dap_container = Container(current_hardware='ot2_holder_15ml', deck_position=0,
slot_number=4, name='DAP_Container', current_volume='15 mL', max_volume=None,
container_type=ContainerTypeCollection.FALCON_TUBE_15_ML, has_stirbar=False,
is_capped=False)
    dmf_container = Container(current_hardware='valve1', deck_position=0, slot_number=8,
name='DMF_Container', current_volume='500 mL', max_volume='500 mL', container_type=None,
has_stirbar=False, is_capped=False)
    centrifuge_tube_50ml = Container(current_hardware='falcon_tube_holder_50mL',
deck_position=0, slot_number=1, name='Centrifuge_Tube_50mL', current_volume=None,
max_volume=None, container_type=ContainerTypeCollection.FALCON_TUBE_50_ML,
has_stirbar=False, is_capped=True)
    centrifuge_tube_50ml_counterweight =
Container(current_hardware='falcon_tube_holder_50mL', deck_position=0, slot_number=1,
name='Centrifuge_Tube_50mL_Counterweight', current_volume=None, max_volume=None,
container_type=ContainerTypeCollection.FALCON_TUBE_50_ML, has_stirbar=False, is_capped=True)
    waste_container_valve1 = Container(current_hardware='valve1', deck_position=0,
slot_number=0, name='waste_container_valve1', current_volume=None, max_volume=None,
container_type=None, has_stirbar=False, is_capped=False)

    # Chemicals
    chemical_zrocl28_h2o = Chemical(container=zrocl2_container, lookup_missing_values=True,
name='ZrOCl28 H2O', lot_number='', supplier='', cas='', smiles='', density=None,
molar_mass=None, mass='0.028 g', molar_amount='0.086 mmol', concentration=None,
mass_concentration='10 mg/mL', volume=None, is_stock_solution=False)
    chemical_dmf = Chemical(container=dmf_container_falcon_tube, lookup_missing_values=True,
name='DMF', lot_number='', supplier='', cas='', smiles='', density=None, molar_mass=None,
mass=None, molar_amount=None, concentration=None, mass_concentration=None, volume='5.0 mL',
is_stock_solution=False)
    chemical_acetic_acid = Chemical(container=acetic_acid_container,
lookup_missing_values=True, name='acetic acid', lot_number='', supplier='', cas='',
smiles='', density=None, molar_mass=None, mass=None, molar_amount='17.0 mmol',
concentration=None, mass_concentration=None, volume='980.0 uL', is_stock_solution=False)
    chemical_4_4_anthracene_9_10_diyldibenzoic_acid = Chemical(container=dap_container,
lookup_missing_values=True, name='4,4-(anthracene-9,10-diyl)dibenzoic acid', lot_number='',
supplier='', cas='', smiles='', density=None, molar_mass=None, mass='0.036 g',
molar_amount='0.086 mmol', concentration=None, mass_concentration='10 mg/mL', volume=None,
is stock solution=False)
    chemical_dmf_1 = Chemical(container=dmf_container, lookup_missing_values=True,
name='DMF', lot_number='', supplier='', cas='', smiles='', density=None, molar_mass=None,
mass=None, molar_amount=None, concentration=None, mass_concentration=None, volume='30.0 mL',
is_stock_solution=False)

    # Reactions
    def reaction_0():
        reaction_flask_50_ml.add_chemical(chemical=chemical_zrocl28_h2o, withdraw_rate=None,
addition_rate=None, robot_arm=None, capper_decapper=None,
return_container_after_addition=False, return_chemicals_after_addition=True,
bottom_clearance=None, purging_volume='30mL', purging_addition_rate=None, purging_port=None,
priming_volume=None, priming_waste_container=None)
        reaction_flask_50_ml.add_chemical(chemical=chemical_dmf, withdraw_rate=None,
addition_rate=None, robot_arm=None, capper_decapper=None,
return_container_after_addition=False, return_chemicals_after_addition=True,
bottom_clearance=None, purging_volume='30mL', purging_addition_rate=None, purging_port=None,
```

```
priming_volume=None, priming_waste_container=None)
        reaction_flask_50_ml.heat(heating_temperature='25.0 C', stirring_speed='400.0 rpm',
heating_time='5.0 min', temperature_stabilization_time='5min',
maximum_temperature_deviation='2C', cooldown_temperature='40C', active_cooling=True,
temperature_sensor='EXTERNAL', robot_arm=None)
        reaction_flask_50_ml.sonicate(sonicator=None, sonication_time='10.0 min',
sonication_power=50.0, sonication_amplitude=50.0, sonication_temperature='20C',
bottom_clearance=None, container_for_cleaning=None, robot_arm=None, capper_decapper=None)
        reaction_flask_50_ml.add_chemical(chemical=[chemical_acetic_acid,
chemical_4_4_anthracene_9_10_diyldibenzoic_acid], withdraw_rate=None, addition_rate=None,
robot_arm=None, capper_decapper=None, return_container_after_addition=False,
return_chemicals_after_addition=True, bottom_clearance=None, purging_volume='30mL',
purging_addition_rate=None, purging_port=None, priming_volume=None,
priming_waste_container=None)
        reaction_flask_50_ml.heat(heating_temperature='120.0 C', stirring_speed='300.0 rpm',
heating_time='24.0 h', temperature_stabilization_time='5min',
maximum_temperature_deviation='2C', cooldown_temperature='40C', active_cooling=True,
temperature_sensor='EXTERNAL', robot_arm=None)
        reaction_flask_50_ml.heat(heating_temperature='25.0 C', stirring_speed='300.0 rpm',
heating_time='5.0 min', temperature_stabilization_time='5min',
maximum_temperature_deviation='2C', cooldown_temperature='40C', active_cooling=True,
temperature_sensor='EXTERNAL', robot_arm=None)
        reaction_flask_50_ml.centrifuge(containers=[centrifuge_tube_50ml,
centrifuge_tube_50ml_counterweight], centrifugation_speed='8000.0 rpm',
centrifugation_time='15.0 min', centrifugation_temperature='25.0 C',
bottom_clearance_withdrawing=3.0, dropoff_location=None, robot_arm=None, centrifuge=None,
transfer_hardware=None, capper_decapper=None)

centrifuge_tube_50ml.remove_supernatant_and_redisperse(waste_container=waste_container_valve
1, redispersion_chemical=chemical_dmf, sonicator=None, sonication_time='10min',
sonication_power=50.0, sonication_amplitude=50.0, sonication_temperature='20C',
container_for_cleaning=None, robot_arm=None, capper_decapper=None, purging_volume='30mL',
bottom_clearance_withdrawing=None, bottom_clearance_dispensing=None,
bottom_clearance_sonication=None)


    # Start reactions in individual threads
    reaction_thread_0 = threading.Thread(target=reaction_0)
    reaction_thread_0.start()

    # Wait for all threads to finish
    reaction_thread_0.join()
```