

Designing Chemical Reaction Arrays using phactor and ChatGPT

Babak Mahjour¹, Jillian Hoffstadt¹, Tim Cernak^{1,2*}

¹Department of Medicinal Chemistry, University of Michigan

²Department of Chemistry, University of Michigan

ABSTRACT: High throughput experimentation (HTE) is a common practice in the pharmaceutical industry. Medicinal chemists design reaction arrays to optimize the yield of couplings between building blocks and/or pharmacophores. Popular reactions attempted by medicinal chemists include the amide coupling, Suzuki coupling, and Buchwald-Hartwig coupling. We show how the artificial intelligence (AI) language model ChatGPT can automatically formulate reaction arrays for these common reactions based on the literature corpus it was trained on. Furthermore, we showcase how ChatGPT results can be directly translated into inputs for the HTE management software phactor, which enables automated execution and analysis of assays. This workflow is experimentally demonstrated.

Introduction

Synthesis of pharmacophore analogs is a bottleneck in drug development. High throughput experimentation (HTE) is a widely practiced method in drug development for the discovery and optimization of reaction conditions in medicinal chemistry campaigns.¹⁻⁵ Chemists typically design reaction arrays based on conditions found in the literature, found via Google, SciFinder, or Reaxys. The automated generation of reaction arrays to optimize or discover a coupling between two substrates is a contemporary problem.⁶⁻¹² Recently, generative transformers, a form of artificial intelligence (AI), have emerged as interactive language models that can interpret and answer scientific questions via verbal human input.^{13,14} Herein we demonstrate how the general-purpose language model ChatGPT can be utilized to generate initial-guess reaction array designs for specific substrate pairs. Due to the flexibility of the AI model, the output can be directly translated into input files for the HTE management software phactor.¹⁵ We showcase several case studies of using ChatGPT to aid in designing reaction arrays. The AI model's reaction array design for the optimization of a Buchwald-Hartwig carbon-nitrogen coupling between simple substrates is experimentally tested and reported.

Methods and Workflow

A workflow using freely available software to automatically generate and execute reaction arrays that couple substrate pairs for popular reactions is reported. A typical workflow can be summarized in three steps (Figure 1):

- First, have ChatGPT generate reaction array designs for specific substrates.
- Then, translate the output from ChatGPT into an input file for phactor. An interfacing script written in python is provided online.
- Finally, use phactor to create stock solutions and distribute the chemicals into the reaction array, manually or robotically, and then analyze its results.

ChatGPT can be interrogated during the design step to elaborate on experimental details or reasonings.

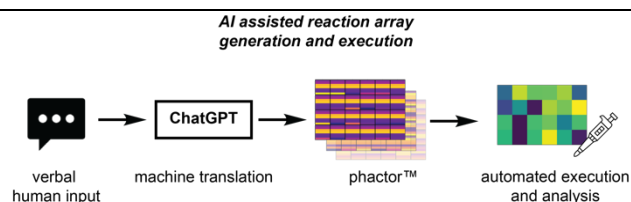


Figure 1. Overview of the ChatGPT to phactor workflow. Verbal input is given by a human to have ChatGPT generate a reaction array design for a particular coupling and substrate pair. The output can be fed directly into phactor to create an assay recipe to be executed robotically or manually.

Case Studies: Literature Informed Reaction Array Generation Via ChatGPT

ChatGPT is a newly released general-purpose AI language model developed by OpenAI. It serves as a conversational model where the user can ask a series of questions and receive answers based on the context of the conversation. While not directly a model for chemistry, ChatGPT has been trained on a large corpus of scientific literature. As such, in its own words, ChatGPT has “knowledge of basic chemistry concepts, such as the periodic table, chemical reactions, acids and bases, and thermodynamics. I can also provide information on more advanced topics, such as organic chemistry, biochemistry, and physical chemistry.” We demonstrate how ChatGPT can be asked to generate reaction arrays for common reaction classes for specific substrates. For each class of reactions (amide, Suzuki, and Buchwald-Hartwig), we ask ChatGPT to develop an experimental design for various pairs of substrates. In the first conversation, we request the generation of a reaction array to optimize an amide coupling between 2-methylbenzoic acid (**1**) and *p*-toluidine (**2**) to form amide **3** (Figure 2).

1. Simple Amide Coupling



Figure 2. The amide coupling ChatGPT is asked to optimize. The reagent classes are specified, but the specific reagents are generated by ChatGPT.

Prompt: “Generate a reaction array with 3 coupling agents, 2 bases, 2 solvents and 2 additives to optimize an amide coupling between 2-methylbenzoic acid and p-toluidine. Add each reagent into a single table, where each row has a column for the reagent class and a column for the reactant name. As well, include specific recommendations and values for other reaction parameters for this screen. Respond succinctly. Finally, cite some literature references from which this screen was designed with DOIs.”

ChatGPT Response:

Reagent Class	Reactant Name
Coupling Agent	HATU
Coupling Agent	EDC
Coupling Agent	PyBOP
Base	DIPEA
Base	TEA
Solvent	DMF
Solvent	DCM
Additive	HOAt
Additive	DMAP

For reaction parameters, some specific recommendations for this amide coupling reaction could include:

- Temperature: 0-25°C (refrigeration)
- Time: 2-24 hours
- Stoichiometry: 1.2-2.0 equivalents of coupling agent, 1.0-1.5 equivalents of base, and 1.0-1.2 equivalents of amine relative to carboxylic acid
- Concentration: 0.1-0.5 M
- Inert atmosphere: nitrogen or argon

References:

- Li, H.; Banerjee, S.; Liao, L.; Pei, X.; Chen, X.; Miao, Q.; Huang, W. An efficient one-pot approach for the synthesis of unsymmetrical ureas via domino reaction of isocyanides with amines using EDC as the coupling agent.
- Kamber, D. N.; Jeon, J.; Lee, S. H.; Kim, S. K. Efficient N,N'-dimethylformamide catalyzed synthesis of amides from carboxylic acids using HATU as a coupling reagent.
- Hwang, S. H.; Park, S. H.; Yang, J. E.; Yoon, Y. J.; Kim, D. S.; Lee, D. W. High yield synthesis of peptide-PEGylated sunitinib via EDC-mediated conjugation.

We note the DOI and citation references did not match the article titles and were omitted from the response. Furthermore, many of the references are not readily available online via Google search. This is because ChatGPT is a language model rather than a knowledge model. Despite this, the design is still reasonable. Further examples of “references” for each design are included in the Supporting Information. We also note that the model struggled to accurately recreate the SMILES of compounds.

2. Late-Stage Amide Coupling

Here we showcase how the conversation can be continued with a different substrate for the same coupling. In the same dialog, we ask it to refer to the previous question but to instead optimize the amide coupling between Sitagliptin (**4**) and carboxylic acid **2** to form late-stage amide **5** (Figure 3).

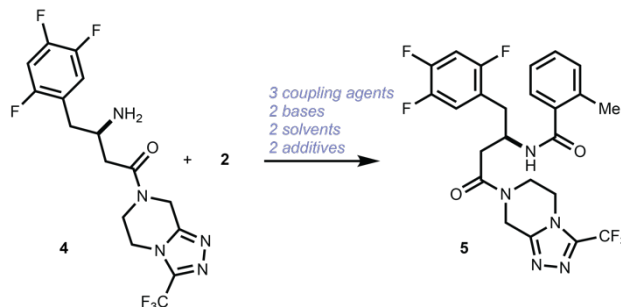


Figure 3. The late-stage amide coupling ChatGPT is asked to optimize. The resultant design is different than with the simple substrates.

Prompt: “Generate a similar reaction assay that instead optimizes the coupling of Sitagliptin with 2-methylbenzoic acid. Respond in the same format.”

ChatGPT Response:

Reagent Class	Reactant Name
Coupling Agent	HBTU
Coupling Agent	TBTU
Coupling Agent	PyAOP
Base	DIEA
Base	TBD
Solvent	DMF
Solvent	DCE
Additive	DMAP
Additive	HOAt

For other reaction parameters, the following values are recommended:

- Temperature: 0-25°C (refrigeration)
- Time: 2-24 hours
- Stoichiometry: 1.2-2.0 equivalents of coupling agent, 1.0-1.5 equivalents of base, and 1.0-1.2 equivalents of amine relative to carboxylic acid
- Concentration: 0.1-0.5 M
- Inert atmosphere: nitrogen or argon

3. Simple Suzuki Coupling

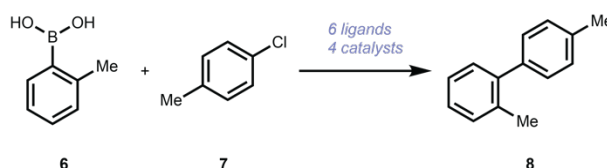


Figure 4. The Suzuki coupling ChatGPT is asked to optimize.

A new chat was started for this example. Here we ask for an optimized reaction array to form biaryl **8** from the Suzuki coupling between boronate **6** and chloride **7** (Figure 4). A typical ligand versus catalyst screen is generated, as requested. The model performed well at pulling ligands and catalysts from the literature corpus, as well as generating additional parameters needed for the assay.

Prompt: “Generate a 6 ligand by 4 catalyst reaction array to optimize an Suzuki coupling between o-tolylboronic acid and 1-chloro-4-methylbenzene. Separate each parameter into a single table, where each row has a column for the reagent class and a column for the reactant name. As well, include

specific recommendations and values for other reaction parameters for this screen. Respond succinctly. Finally, cite some literature references from which this screen was designed with DOIs.”

ChatGPT Response:

Reagent Class	Reactant Name
Ligand	XPhos
Ligand	SPhos
Ligand	RuPhos
Ligand	BrettPhos
Ligand	Buchwald ligand
Ligand	DavePhos
Catalyst	Pd ₂ (dba) ₃
Catalyst	Pd(OAc) ₂
Catalyst	PdCl ₂
Catalyst	Pd(dppf)Cl ₂
Substrate	o-tolylboronic acid
Substrate	1-chloro-4-methylbenzene

Conditions: The reaction should be performed under nitrogen atmosphere, and solvents such as DMF, DMSO, or toluene could be used. The reaction should be performed at 100-120°C for 12-24 hours, with a 1:1 ratio of the two substrates, and 2 mol% of Pd catalyst. The ratio of the Pd catalyst to the ligand should be 1:2.

We note that the model does not discuss the addition of a base, so we directly ask in the same conversation: “What base should I use?”

The model responds:

A base is typically not necessary for Suzuki coupling reactions since the reaction proceeds via a transmetalation step rather than a deprotonation step. However, if a base is needed, a common choice is potassium carbonate (K₂CO₃) or cesium carbonate (Cs₂CO₃) at a loading of 1.2-2 equiv.

4. Simple Buchwald-Hartwig Coupling

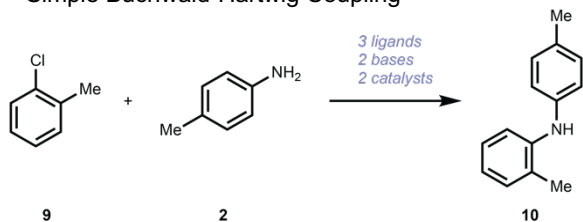


Figure 5. The simple Buchwald-Hartwig coupling ChatGPT is asked to optimize.

Again, a new conversation is started for this example. In this final example, we ask the model to generate a three factorial screen to optimize the carbon-nitrogen coupling between chloride **9** and aniline **2** to form compound **10** (Figure 5).

Prompt: “Generate a 3 ligand by 4 catalyst by 2 base reaction array to optimize an Buchwald-Hartwig coupling between 1-chloro-2-methylbenzene and p-toluidine. Add each reagent into a single table, where each row has a column for the reagent class and a column for the reactant name. As well, include specific recommendations and values for other reaction parameters for this screen. Respond

succinctly. Finally, cite some literature references from which this screen was designed with DOIs.”

ChatGPT Response:

Reagent Class	Reactant Name
Ligand	XPhos
Ligand	SPhos
Ligand	RuPhos
Catalyst	Pd ₂ (dba) ₃
Catalyst	Pd(dppf)Cl ₂
Catalyst	P(t-Bu) ₃
Catalyst	Pd(OAc) ₂
Base	K ₃ PO ₄
Base	NaO ^t Bu

Other recommended reaction parameters:

- Solvent: DMF
- Temperature: 100°C
- Reaction time: 24 hours
- Stoichiometry: 1:1 ratio of 1-chloro-2-methylbenzene and p-toluidine
- Concentration: 0.1 M

We note the incorrect output “P(t-Bu)₃”, which we correct to the palladium complex in our experimental execution. We ask some additional questions to clarify the experimental details.

Prompt: What catalyst, ligand, and base concentrations should I use for this reaction?

ChatGPT Response:

The specific concentrations of catalyst, ligand, and base to use for the reaction depend on various factors such as the desired reaction rate, reaction yield, and the specific reactants being used. However, as a general guideline, a common catalyst concentration range for Buchwald-Hartwig couplings is 0.5-5 mol%. Ligand concentrations can range from 1-10 mol%. Base concentrations can range from 0.5-2 equiv. It's important to note that these concentrations may need to be optimized for specific substrates and conditions.

Prompt: How about the order of addition for the reagents?

ChatGPT Response:

The order of addition for the reagents can also be important for optimizing the reaction conditions. In a typical Buchwald-Hartwig coupling, the base is usually added first to the reaction mixture, followed by the catalyst and the ligand. The organic halide and the amine are then added to initiate the reaction. However, depending on the specific reactants and conditions, the order of addition may need to be adjusted to achieve the desired reaction outcome. Therefore, it's always a good idea to consult the literature or perform some preliminary experiments to determine the optimal order of addition for the specific reaction being performed.

Prompt: How would I quench and work up this reaction?

ChatGPT Response:

After the Buchwald-Hartwig coupling reaction is complete, it can be quenched by adding a quenching agent to the reaction mixture. Common quenching agents include ammonium chloride or hydrochloric acid. The choice of quenching agent may depend on the specific reaction conditions and the nature of the products.

To work up the reaction, the mixture is typically cooled and then diluted with an appropriate solvent such as water or ethyl acetate. The reaction mixture is then filtered to remove any solid catalyst or byproducts. The organic layer is separated and washed with water or brine to remove any remaining inorganic salts. The organic layer is then dried over anhydrous sodium sulfate or magnesium sulfate and concentrated under reduced pressure to yield the crude product. Purification of the product can be done using techniques such as column chromatography or recrystallization.

We comment that the quench of the reaction should be performed after cooling for safety.

Prompt: What do you think the single best combination of ligand and catalyst would be for this reaction, given the reaction array you generated earlier? Respond with a single best guess.

ChatGPT Response:

It's difficult to predict the single best combination of ligand and catalyst for the Buchwald-Hartwig coupling reaction between 1-chloro-2-methylbenzene and p-toluidine without performing experiments or conducting computational studies. However, based on the reaction array I provided earlier, a possible combination of a ligand and a catalyst that could work well together for this reaction is SPhos as the ligand and Pd₂(dba)₃ as the catalyst. This is just a best guess, and it's important to note that the optimal combination of ligand and catalyst may depend on various factors, such as the specific reaction conditions, and could require further optimization to achieve the best results.

This reaction array was executed experimentally with the exact specifications given by ChatGPT. The best hit was scaled up and a calibration curve was developed to map the raw data to assay yields. (Figure 6). The array produced many hits with moderate yield with respect to internal standard. Furthermore, low amounts of impurities and byproducts simplified purification. Well B5 was scaled up and resulted in an 84% isolated yield of product **10**. We compare this result to the best yielding hit found in the literature (98%).¹⁶

The model's output seems to select catalysts and ligands that have performed well in the literature but may be less aware of how they are used together. As such, the resultant array includes unusual combinations of palladium sources and ligands, such as PdCl₂(dppf) and RuPhos. Furthermore, DMF and DMSO are less typical than ethereal solvents for this reaction. Despite this, the assay still produced positive results.

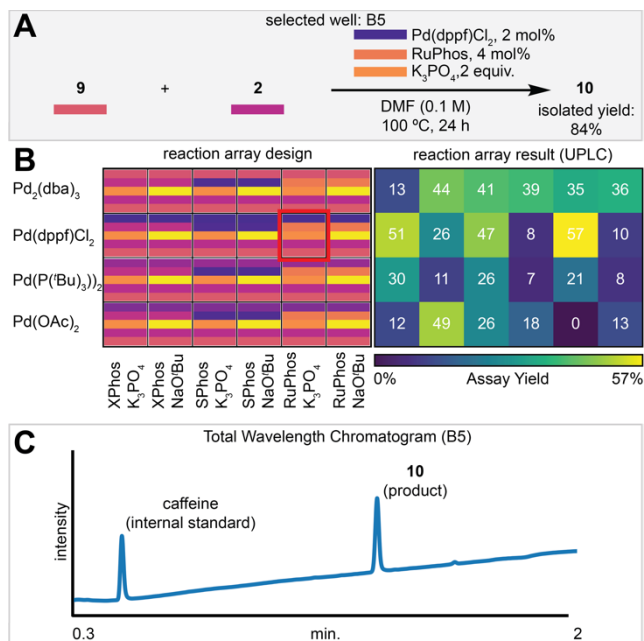


Figure 6. (A) Best hit from the reaction array proposed by ChatGPT. The combination of palladium catalyst complex Pd(dppf)Cl₂ and RuPhos yielded the most product in the array. This trend was consistent with the same catalyst for each ligand used in the experiment. (B) The reaction array design and results. (C) Two minute liquid chromatography UV trace for selected well B5.

The combination of Pd(dppf)Cl₂ and RuPhos is a surprising result, since there is likely an equilibrating mixture of Pd(dppf)L_n and Pd(RuPhos)L_n complexes. We suspect the proposal to mix these compounds arises as an artifact of making a combinatorial array (phactor) out of popular Pd complex and ligand choices from the literature (ChatGPT). Nonetheless, the observation that this "cocktail" of ligands was the most productive result could be supported by related reports of "cocktail" catalysis in the Buchwald-Hartwig coupling.¹⁷⁻²⁰

Conclusion

The software ChatGPT was utilized to generate reaction array designs for several popular reactions. In one instance, the suggestion from ChatGPT was translated to interface with the high throughput experimentation manager phactor, which facilitated its subsequent execution. Analysis of the array revealed a high yielding reaction.

Corresponding Author

* Tim Cernak – Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, MI, USA 48104; orcid.org/0000-0001-5407-0643. Email: tcernak@med.umich.edu

Notes

T.C. holds equity in Scorpion Therapeutics, and is a co-Founder and equity holder of Entos, Inc.

References

- 1 Mahjour, B., Shen, Y. & Cernak, T. Ultrahigh-throughput experimentation for information-rich

- chemical synthesis. *Accounts of Chemical Research* **54**, 2337-2346 (2021).
- 2 Gesmundo, N. J. *et al.* Nanoscale synthesis and affinity ranking. *Nature* **557**, 228-232 (2018).
- 3 Uehling, M. R., King, R. P., Krska, S. W., Cernak, T. & Buchwald, S. L. Pharmaceutical diversification via palladium oxidative addition complexes. *Science* **363**, 405 (2019).
- 4 Lin, S. *et al.* Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS. *Science* **361**, eaar6236 (2018).
- 5 Buitrago Santanilla, A. *et al.* Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **347**, 49 (2015).
- 6 Shen, Y. *et al.* Automation and computer-assisted planning for chemical synthesis. *Nature Reviews Methods Primers* **1**, 1-23 (2021).
- 7 Torres, J. A. G. *et al.* A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. *Journal of the American Chemical Society* **144**, 19999-20007 (2022).
- 8 Shields, B. J. *et al.* Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89-96 (2021).
- 9 Williams, W. L. *et al.* The evolution of data-driven modeling in organic chemistry. *ACS central science* **7**, 1622-1637 (2021).
- 10 Gao, W., Raghavan, P. & Coley, C. W. Autonomous platforms for data-driven organic synthesis. *Nature Communications* **13**, 1-4 (2022).
- 11 Tu, Z., Stuyver, T. & Coley, C. W. Predictive chemistry: Machine learning for reaction deployment, reaction development, and reaction discovery. *Chemical Science* (2022).
- 12 Shim, E. *et al.* Predicting reaction conditions from limited data through active transfer learning. *Chemical science* **13**, 6655-6668 (2022).
- 13 Luo, R. *et al.* BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics* **23** (2022).
- 14 Ferruz, N., Schmidt, S. & Höcker, B. ProtGPT2 is a deep unsupervised language model for protein design. *Nature communications* **13**, 4348 (2022).
- 15 Mahjour, B. *et al.* Rapid Planning and Analysis of High-Throughput Experiment Arrays for Reaction Discovery. (2022).
- 16 Zhao, X.-Y., Zhou, Q. & Lu, J.-M. Synthesis and characterization of N-heterocyclic carbene-palladium (ii) chlorides-1-methylindazole and-1-methylpyrazole complexes and their catalytic activity toward C–N coupling of aryl chlorides. *RSC advances* **6**, 24484-24490 (2016).
- 17 King, A. K., Brar, A. & Findlater, M. A tertiary phosphine oxide ligand-based recyclable system for the Suzuki–Miyaura and Negishi reactions: evidence for pseudo-homogeneous catalysis. *Catalysis Science & Technology* **13**, 301-304 (2023).
- 18 Prima, D. O. *et al.* Evidence for “cocktail”-type catalysis in Buchwald–Hartwig reaction. A mechanistic study. *Catalysis Science & Technology* **11**, 7171-7188 (2021).
- 19 Ondar, E. E., Burykina, J. V. & Ananikov, V. P. Evidence for the “cocktail” nature of platinum-catalyzed alkyne and alkene hydrosilylation reactions. *Catalysis Science & Technology* **12**, 1173-1186 (2022).
- 20 Fors, B. P. & Buchwald, S. L. A multiligand based Pd catalyst for C– N cross-coupling reactions.