# Potential Research Directions:
## Intelligent Scheduling for Automated Chemical Workflows with Focus on Molecular Solar Thermal Energy Storage

Context Map Analysis
Supervisor: Kasper Moth-Poulsen

January 28, 2026

## Executive Summary

This document identifies potential research directions for a Master's/PhD thesis focused on developing intelligent scheduling systems for automated chemical workflows. The research builds upon two complementary foundation areas:

- **Domain Foundation**: Molecular Solar Thermal (MOST) energy storage systems, specifically photoswitch characterization platforms for norbornadiene-quadricyclane and azobenzene systems
- **Technical Foundation**: Autonomous chemistry research encompassing robotic platforms, AI/ML-driven optimization, and real-time control systems

Each research direction connects specific technologies from the autonomous chemistry landscape to concrete applications in MOST photoswitch development, with explicit references to source publications via context map tags; one will be selected.

# 1 Research Direction 1: Bayesian Optimization for Adaptive Multi-Sample Photoswitch Characterization

## 1.1 Core Problem

The current automated photo-isomerization platform (Moth-Poulsen `184.001`) operates sequentially on individual samples. When screening a library of photoswitch candidates, the system provides no mechanism to prioritize characterization based on early indicators of promising performance. This results in uniform resource allocation across all candidates regardless of their potential.

## 1.2 Proposed Solution

Develop an adaptive scheduling system that integrates Bayesian optimization (Perplexity `001.037`) with multi-objective criteria (Perplexity `001.038`) to dynamically prioritize sample characterization order. The system would:

1. **Early data acquisition**: Measure initial photoconversion quantum yield and absorption onset for each sample in the queue
2. **Gaussian process modeling**: Build surrogate models predicting full characterization outcomes from preliminary measurements
3. **Acquisition function optimization**: Use expected improvement criteria to select which sample receives detailed thermal kinetics characterization next
4. **Multi-objective balancing**: Simultaneously optimize for high energy density (Moth-Poulsen `099.001`), solar spectrum match (Moth-Poulsen `171.001`), and kinetic stability (Moth-Poulsen `151.001`)

## 1.3 Technical Approach

**Workflow Integration:**

- Extend the existing Python GUI (Moth-Poulsen `184.001`) with a Bayesian optimization layer using GPyTorch or BoTorch
- Implement non-blocking experiment queues that yield control during thermal back-conversion waiting periods
- Interface with the 16-position Knauer valve system for automated sample switching

**Performance Metrics:**

- Reduction in total laboratory time to identify top 10% candidates from a 50-sample library
- Number of full characterizations required to achieve 95% confidence in Pareto frontier identification
- Comparison against random, round-robin, and greedy (best-first) scheduling baselines

### 1.4 Connection to Autonomous Chemistry

This direction directly implements high-throughput screening methodologies (Perplexity `001.026`, `001.032`) while leveraging Bayesian optimization principles proven in materials discovery contexts. The adaptive sampling strategy mirrors approaches in autonomous robotic platforms like ARChemist (Perplexity `001.010`) and ORGANA (Perplexity `001.009`).

### 1.5 Expected Outcomes

- 40-60% reduction in characterization time for photoswitch library screening
- Validated Bayesian optimization framework generalizable to other molecular property optimization problems
- Open-source scheduling software compatible with the Digital Discovery platform ecosystem

# 2 Research Direction 2: Control Barrier Functions for Safety-Constrained Automated Workflows

## 2.1 Core Problem

Automated chemical systems face safety-critical constraints: temperature-sensitive samples can degrade (Moth-Poulsen `098.001` shows solvent-dependent thermal stability), flow cells can clog causing pressure build-up, and LED arrays can overheat. Current automation frameworks use ad-hoc timeout mechanisms and boolean error flags, providing no formal guarantees that safety constraints remain satisfied during schedule execution.

## 2.2 Proposed Solution

Apply Control Barrier Functions (Perplexity `003.011`) to chemical workflow scheduling, providing mathematically rigorous safety certificates. CBFs would encode:

1. **Temperature envelope constraints**: Ensure sample temperature remains within stability limits during thermal kinetics measurements
2. **Pressure safety margins**: Monitor flow cell pressure and preemptively halt pumping before membrane rupture thresholds
3. **Equipment thermal budgets**: Track cumulative LED irradiation time and enforce mandatory cooling periods
4. **Chemical compatibility**: Prevent solvent cross-contamination through enforced purge sequences

## 2.3 Technical Approach

**Modeling Framework:**

- Formulate discrete-time CBF constraints for chemical workflow state machines
- Implement filtered CBFs (Perplexity `003.002`) to handle non-smooth control inputs (e.g., valve switching, pump on/off)
- Integrate with feasibility assurance mechanisms (Perplexity `003.001`) to guarantee CBF-QP solvability at each scheduling decision

**Safety Constraint Examples:**

- $h_1(x) = T_{\max} - T_{\text{sample}}(t) \geq 0$ (temperature upper bound)
- $h_2(x) = t_{\text{cool}} - t_{\text{LED\_cumulative}} \geq 0$ (thermal management)
- $h_3(x) = P_{\text{rupture}} - P_{\text{cell}}(t) \geq \alpha$ (pressure safety margin)

## 2.4 Connection to Control Theory

Leverages recent advances in CBF handling of uncertainty (Perplexity `003.012`), real-time constraints (Perplexity `003.019`), and robust control (Perplexity `003.010`). Addresses the identified challenge in CBF literature regarding feasibility under complex multi-constraint scenarios.

### 2.5 Expected Outcomes

- Formal proof of safety constraint satisfaction throughout workflow execution
- Reduction in equipment damage incidents and sample loss during unattended operation
- Certified controller suitable for regulatory submission in pharmaceutical/industrial contexts
- Publication-quality contribution to CBF application in chemical automation

# 3 Research Direction 3: Digital Twin-Enabled Predictive Scheduling for Flow Chemistry

## 3.1 Core Problem

Flow-based photoswitch characterization (Moth-Poulsen `177.001`, `184.001`) involves complex fluid dynamics: Taylor dispersion during sample injection, residence time distribution effects, and thermal equilibration transients. The current system treats the flow cell as a black box, leading to conservative wait times and suboptimal throughput.

## 3.2 Proposed Solution

Develop a digital twin (Perplexity `001.002`) of the flow characterization platform that simulates:

1. **Microfluidic transport**: CFD modeling of sample plug propagation through tubing and flow cell (Perplexity `001.042`)
2. **Photochemical conversion**: Spatially-resolved photoconversion accounting for light absorption gradients
3. **Thermal dynamics**: Temperature controller response and sample heating/cooling profiles
4. **Spectroscopic measurement**: Ray tracing to predict measurement artifacts from flow cell geometry

The twin enables **model predictive control** (MPC) for scheduling: simulate candidate action sequences, select the trajectory that minimizes total characterization time while maintaining measurement accuracy.

## 3.3 Technical Approach

**Simulation Components:**

- Physics-based models (Perplexity `001.006`): 1D advection-diffusion PDE for concentration profiles
- Data-driven surrogate models: Neural networks trained on experimental data to correct physics model errors
- Sim-to-real transfer (Perplexity `001.002`): Bayesian calibration using periodic experimental validation runs

**Integration with MATTERIX Framework:** Leverage GPU-accelerated simulation infrastructure (Perplexity `001.002`) and the multi-scale physics approach. Extend to include photochemical reactions specific to norbornadiene-quadricyclane systems (Moth-Poulsen `097.001`, `180.001`).

## 3.4 Connection to Flow Chemistry

Builds on continuous processing methodologies (Perplexity `001.029`), inline PAT (Perplexity `001.046`), and scalable synthesis approaches (Moth-Poulsen `134.001`, `162.001`). Addresses the gap between batch reactor HTE and continuous flow optimization.

### 3.5 Expected Outcomes

- 20-30% throughput improvement through optimized pump sequencing and reduced dead time
- Digital twin validated to within 5% error on key metrics (residence time, conversion efficiency)
- MPC controller that adapts to equipment drift and degradation over multi-day campaigns
- Framework generalizable to other flow chemistry automation platforms

# 4 Research Direction 4: LLM-Driven Protocol Translation for Heterogeneous Workflow Orchestration

## 4.1 Core Problem

Chemical synthesis procedures in literature (e.g., Moth-Poulsen `180.001` "Practical Synthesis of Multi-Site Functionalized Norbornadiene") are written in natural language with implicit assumptions. Translating these into executable scheduler instructions requires expert knowledge and manual protocol decomposition. This creates a bottleneck when integrating new synthetic routes into automated platforms.

## 4.2 Proposed Solution

Develop an LLM-based system (Perplexity `001.021`) that:

1. **Parses literature protocols**: Extract reaction steps, timing constraints, equipment requirements, and safety considerations from unstructured text
2. **Generates workflow graphs**: Convert procedures into directed acyclic graphs (DAGs) with explicit dependencies
3. **Maps to hardware capabilities**: Match abstract operations (e.g., "heat to reflux") to available equipment (temperature-controlled cuvette holder, heating rates)
4. **Synthesizes scheduler instructions**: Output Python/domain-specific language code compatible with the existing automation framework

## 4.3 Technical Approach

**LLM Architecture:**

- Fine-tune Llama-3.1-70B (inspired by ChemAgents architecture, Perplexity `002.004`) on corpus of synthetic organic chemistry procedures
- Implement retrieval-augmented generation (RAG) with chemistry literature database mining (Perplexity `001.024`)
- Multi-agent design: Literature Reader agent, Equipment Mapping agent, Safety Verification agent, Code Generation agent

**Validation Strategy:**

- Test on 20 published photoswitch synthesis protocols from Moth-Poulsen group papers
- Metrics: Syntactic correctness (code compiles), semantic correctness (produces intended product), safety compliance (no hazardous operations)
- Human-in-the-loop validation for critical decision points

## 4.4 Connection to Autonomous Chemistry

Directly implements LLM-driven agent concepts (Perplexity `001.021`, `002.004`) and collaborative intelligence frameworks (Perplexity `001.024`). Addresses the challenge of task coordination (Perplexity `001.055`) across heterogeneous equipment.

### 4.5 Expected Outcomes

- 10x reduction in time required to port literature protocols to automated platform
- Success rate >80% on protocol translation benchmark
- Open-source tool applicable to any chemistry automation framework with API access
- Contribution to standardization efforts (Perplexity `001.052`) in chemical automation languages

# 5 Research Direction 5: Reinforcement Learning for Multi-Instrument Resource Allocation

## 5.1 Core Problem

When characterizing photoswitches, multiple experiments compete for shared resources: spectrometer time, heating stages, LED light sources, and pump access. Static scheduling heuristics (round-robin, priority queues) fail to adapt to:

- Variable experiment durations (Moth-Poulsen `151.001`: some photoswitches require hours for thermal kinetics)
- Equipment failures requiring dynamic rescheduling
- Measurement quality feedback suggesting early experiment termination or extension

## 5.2 Proposed Solution

Frame workflow scheduling as a Markov Decision Process (MDP) and train a reinforcement learning agent (Perplexity `001.073`) to learn optimal resource allocation policies from experience.

**State Space:**

- Current experiment progress (conversion percentages, elapsed time)
- Resource availability (spectrometer: busy/idle, LED: thermal budget remaining)
- Queue status (samples waiting, priority levels)
- Historical performance (average success rate per sample type)

**Action Space:**

- Select next sample to characterize
- Allocate spectrometer measurement frequency (high-frequency for fast reactions, sparse for slow)
- Preempt running experiments to prioritize high-value candidates
- Trigger calibration/cleaning routines

**Reward Function:**

- Maximize number of fully characterized samples per unit time
- Penalize incomplete characterizations and equipment idle time
- Bonus for early identification of high-performance candidates (energy density >0.4 MJ/kg)

### 5.3 Technical Approach

**Algorithm Selection:**

- Proximal Policy Optimization (PPO) or Deep Q-Networks (DQN) depending on action space dimensionality
- Model-based RL: Learn transition dynamics model to enable planning
- Reward shaping (Perplexity `001.073`) to accelerate convergence

**Training Strategy:**

- Simulation-based pre-training using digital twin (Research Direction 3)
- Sim-to-real transfer via domain randomization and system identification
- Online learning from real experimental data with safety constraints (Research Direction 2)

### 5.4 Connection to Autonomous Research

Implements reinforcement learning frameworks proven in materials discovery (Perplexity `001.073`, `001.012`). Builds on autonomous decision-making paradigms (Perplexity `001.017`) and real-time adaptation (Perplexity `003.003`). Aligns with automated research platform concepts (Moth-Poulsen `179.001`).

### 5.5 Expected Outcomes

- 30-40% improvement in multi-sample throughput versus rule-based scheduling
- Learned policy that generalizes across different photoswitch families (NBD, azobenzene, BOD)
- Interpretable policy analysis revealing optimal scheduling principles for human operators
- Integration with broader self-driving laboratory ecosystems