Sequential Decision-Making in DNA: From POMDPs to Molecular Controllers

Antonio Llano, Shobhit Agarwal, Ethan Goodhart

AA228/CS238 Project Status Update

Motivation: Beyond Static Logic

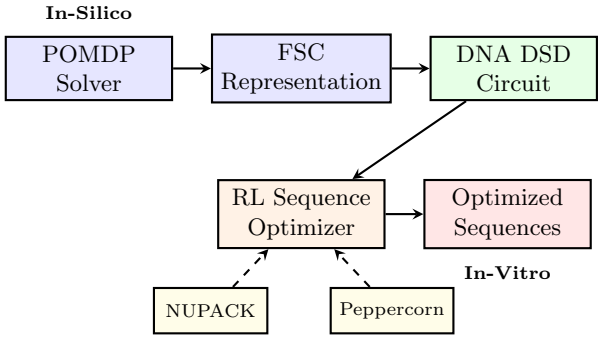
DNA strand displacement circuits have demonstrated Boolean computation (AND, OR gates) and even one-shot Bayesian inference. However, real-world biosensing demands **sequential decision-making**: a therapeutic implant monitoring disease biomarkers over days must *accumulate evidence* before committing to drug release, balancing false-positive risks against treatment delays.

DNA computers operate *in situ* within biological environments, possibly enabling autonomous diagnostics and therapeutics without external instrumentation. *Can we implement principled decision policies that sense, reason, and actuate over multiple sensing cycles at the molecular level?*

From POMDPs to Finite-State Controllers

We initially sought to implement full POMDP belief updates, maintaining probability distributions $b_t(s)$ and computing Bayesian updates at each time step. Prior work demonstrated one-shot Bayesian inference in DNA (Sainz de Murieta, 2012), but sequential belief tracking over multiple episodes faces fundamental biophysical barriers: **1) DNA degradation**, which prevents reliable multi-day state storage. **2) Drift & noise**: Molecular stochasticity and spurious leak compounds over sequential updates, degrading computational fidelity.

Rather than computing beliefs *in DNA*, we now compile optimal POMDP policies into **Finite-State Controllers (FSCs)**: compact automata with discrete memory states. FSCs map directly to DNA strand displacement cascades: each controller node is a molecular state, and observations trigger strand displacement reactions that transition between states (Chen et al., 2013), embodying the optimal policy in molecular hardware without requiring in-vitro belief computation.



Building Topology-Aware RL Gym

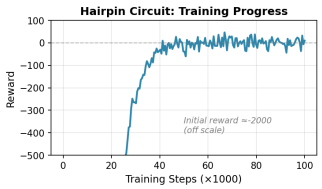
Given an FSC topology, assigning nucleotide sequences is non-trivial. Designs must satisfy (1) **thermodynamic stability**: NUPACK computes minimum free energy (MFE) via nearest-neighbor thermodynamics, (2) **reaction orthogonality**: non-reactant strands don't spuriously bind, and (3) **kinetic feasibility**: Peppercorn enumerates strand displacement pathways and computes rate constants ($k_f > 10^{-3}$ s^{-1} for intended reactions, $k_{\text{leak}} < 10^{-6}$ s^{-1} for leaks).

Why prior methods fail: NUPACK's native design and genetic algorithms optimize thermodynamics only, ignoring kinetics. None close the NUPACK-Peppercorn loop.

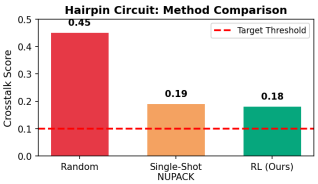
Our approach: OpenAI Gym MDP with *state* = one-hot encoded sequences + NUPACK features (MFE, crosstalk, GC) + Peppercorn features (intended/leak rates); *action* = mutate domain i position j to base b; *reward* $= -w_1$crosstalk $+ w_2 \log k_{\text{int}} - w_3 \log k_{\text{leak}} + w_4$structure. A 21.5K-parameter actor-critic network (69-dim input $\rightarrow$ 64 hidden $\rightarrow$ 64 hidden $\rightarrow$ 64 actions) trains via PPO over 100K steps (2048 rollouts, batch 64), learning heuristics like avoiding homopolymers and balancing GC content through trial-and-error with NUPACK/Peppercorn feedback.

RL Gym Validation: Hairpin Circuit

To validate our optimizer, we tested on a 2-domain hairpin (d1: 8nt, d2: 6nt) forming `((((....))))`. The network trains over 100K steps, converging from random initialization (reward ∼-2000) to local optimum (reward ∼+6) within 20K steps. Each episode involves ∼10 mutations, discovering heuristics: avoid homopolymers (AAAA/GGGG), balance GC near 50%, achieve MFE < -10 kcal/mol.



(a) Training over 100K steps. Converges to reward ∼+6 at 20K steps via ∼10 mutations/episode.



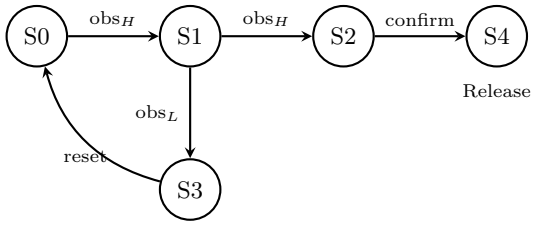
(b) RL (0.18) vs. Single-Shot NUPACK (0.19): comparable on simple benchmark.

Key finding: The 21.5K-parameter network discovers heuristics through trial-and-error, achieving comparable performance to single-shot NUPACK on this well-studied benchmark. While hairpin circuits have estab-

lished heuristic-based tools, our FSC-POMDP topologies are unique; our hope is that RL strategy generalizes where no prior design rules exist.

FSC-POMDP Circuit Design

We designed a 5-node FSC for adaptive drug delivery (inflammatory disease domain). The topology includes: **7 domains** (transition toeholds + branch migration regions) **6 strands** (5 state strands + 1 output signal strand) **4 intended reactions** (state transitions triggered by observations)



Sequence design status: We are currently designing the complexes and domains in dot-parens for our FSC before generating the optimal sequences.

Next Steps: **(1)** Complete FSC-5 sequence optimization (in progress). **(2)** Validate sequences via Peppercorn kinetic simulation. **(3)** Compare FSC policy vs. Boolean baseline in stochastic disease simulator.

Long-term vision: Demonstrate that topology-aware RL sequence design enables compilation of *arbitrary* POMDP policies into DNA circuits, bridging formal decision theory and molecular implementation. Potential applications span adaptive diagnostics, personalized therapeutics, and biosensors that reason under uncertainty.

This work represents the first systematic pipeline from high-level decision problems (POMDPs) to molecular hardware (DNA circuits), with RL as the critical enabler for navigating the vast sequence design space.