Scalable Enumeration of Pareto-optimal Polymers for Computing Equilibrium Concentrations

Archit Patil[1], Minki Hhan[2], and David Soloveichik[1]

[1]University of Texas at Austin, USA
[2]Korea Advanced Institute of Science and Technology, South Korea

Predicting equilibrium concentrations of molecular complexes is essential for verifying the behavior of engineered DNA systems. However, a finite set of monomer types can in principle generate infinitely many complexes. We study this candidate-enumeration problem in a geometry-free, domain-level abstraction called a domain-monomer system, generalizing Thermodynamic Binding Networks (TBNs) to the unsaturated setting where not every possible bond need be formed. We define *Pareto-suboptimal* polymers as those that can be split into non-interacting parts, and show that restricting attention to *Pareto-optimal* polymers is thermodynamically justified: no Pareto-suboptimal polymer appears in any minimum free-energy configuration, and the total equilibrium concentration of such polymers is small. We prove that there are finitely many Pareto-optimal polymers and exactly characterize them via a Hilbert basis computation, extending prior work from the saturated TBN model. To scale this approach to large systems, we develop a framework that restricts the number of different monomer types that a single polymer contains, and uses combinatorial covering designs to reduce the number of Hilbert basis computations required. We benchmark the method on several families of DNA molecular programming systems, demonstrating order-of-magnitude speedups over direct computation while recovering nearly all equilibrium-relevant polymers.

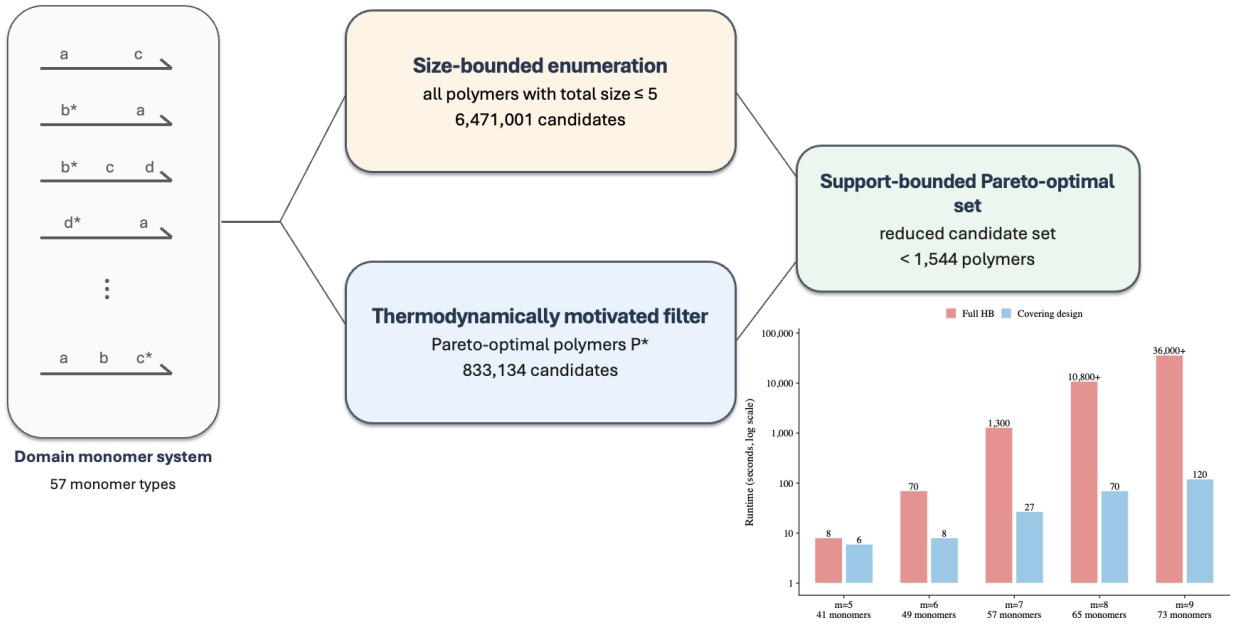

Figure 1: Overview of the support-bounded Pareto-optimal polymer enumeration pipeline. Our method reduces the candidate set for polymers that may occur in significant concentration at equilibrium by combining thermodynamic Pareto optimality with support-bounded covering-design enumeration, allowing much faster equilibrium analysis while being accurate. In the runtime graph, Full HB represents the time required to compute the set of Pareto-optimal polymers, while covering design represents the time required to compute the corresponding support-bounded Pareto-optimal set.