Double-Ended Synthesis Planning with Goal-Constrained Bidirectional Search

NeurIPS 2024

Kevin Yu (kyu3@mit.edu), Jihye Roh, Ziang Li, Wenhao Gao, Runzhong Wang, Connor W. Coley

Molecules can be obtained by a variety of means

paclitaxel

10-DAB III

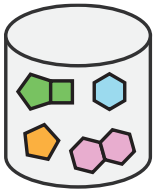
semi-synthesis

isolation

total synthesis

isolation

Pacific yew
(endangered)

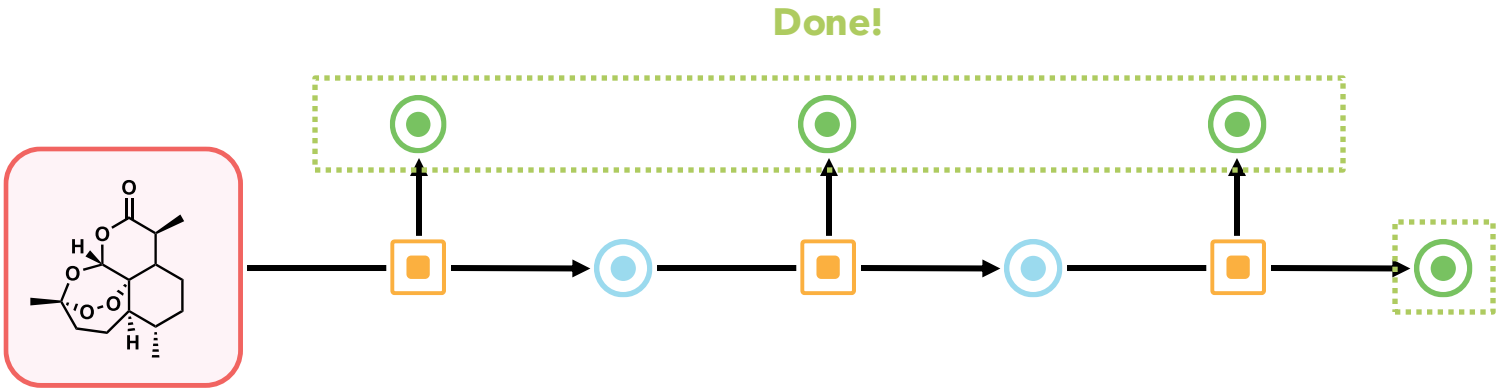


European yew

Synthesis planning problem formulation

Given: target molecule t + set of purchasable building blocks $\mathcal{B}$,

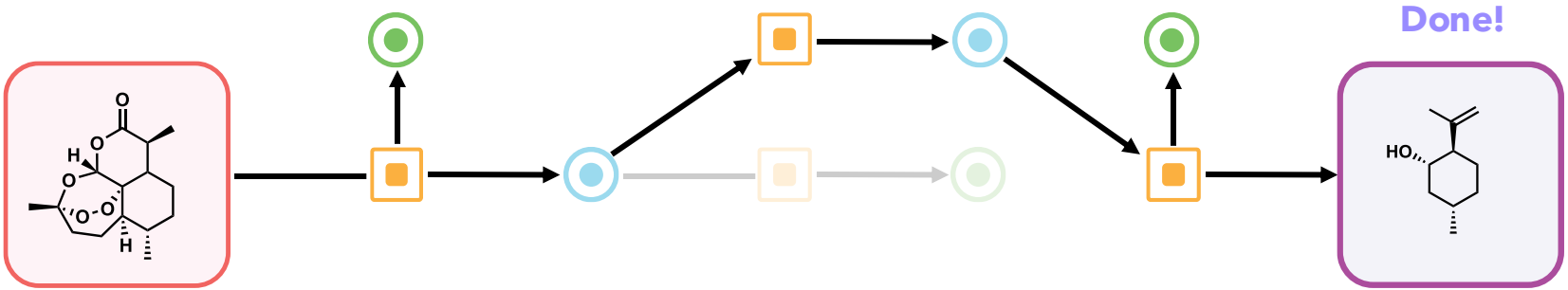
Construct: a sequence of valid chemical reactions forming a synthesis tree with root node t such that all leaf nodes are members of $\mathcal{B}$



The problem can be redefined to include structure goals ("starting material-constrained planning")

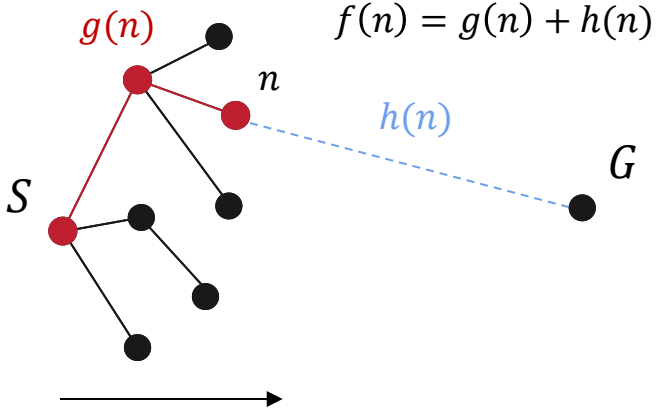
Given: target molecule t *and starting material* s + set of purchasable building blocks $\mathcal{B}$,

Construct: a sequence of chemical reactions forming a synthesis tree with root node t such that *one leaf node is* s and all other leaf nodes are members of $\mathcal{B}$



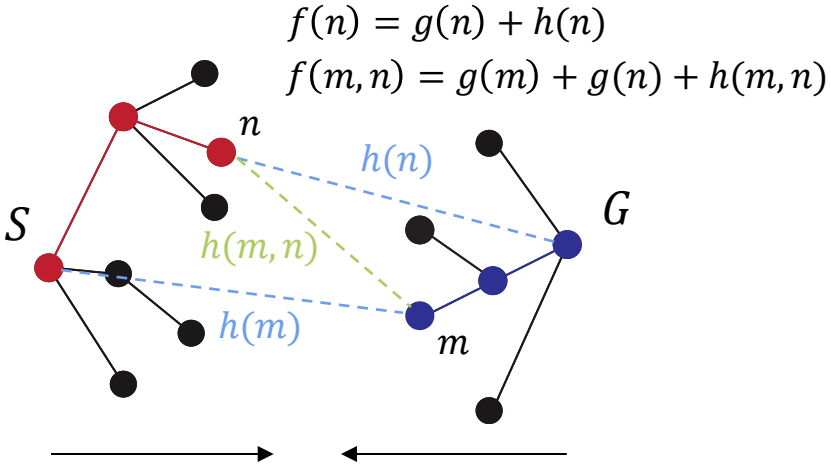
Bidirectional search algorithms are naturally suited for goal-constrained planning

Unidirectional A* search



$\arg\min_n(f(n))$

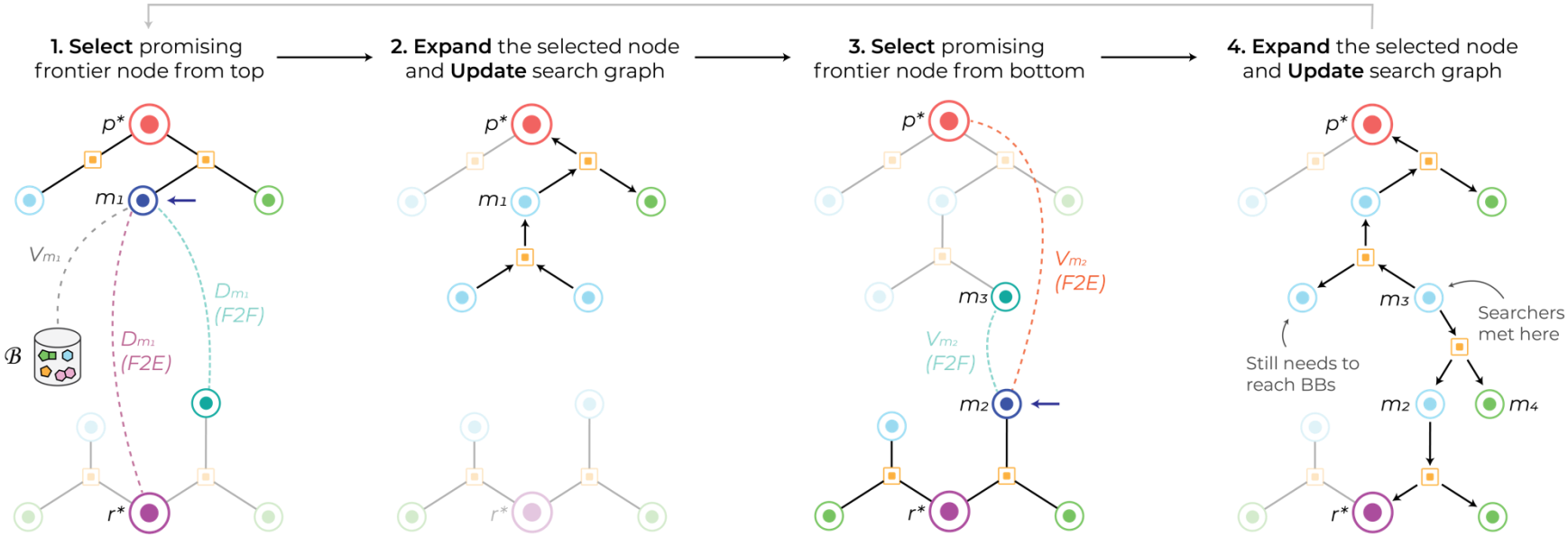
Bidirectional A* search



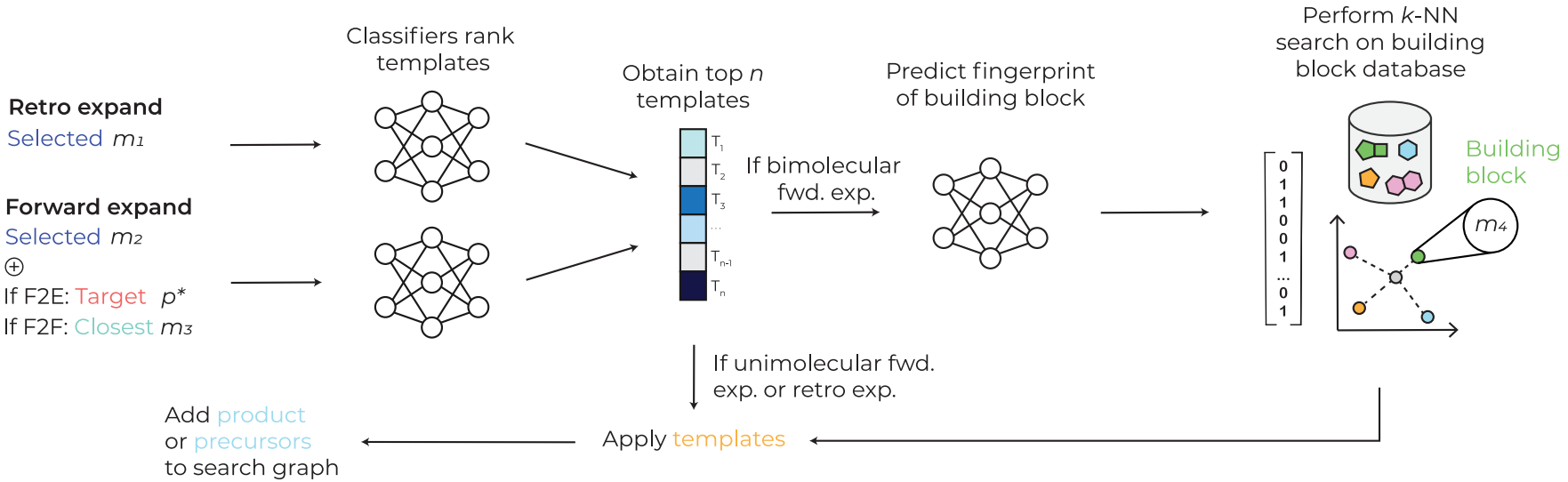
"Front-to-end" $\arg\min_n(f(n))$

"Front-to-front" $\arg\min_{m,n}(f(m,n))$

DESP (Double-ended Synthesis Planning) plans syntheses with bidirectional A*-like search

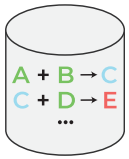


DESP utilizes a conditional forward expansion policy

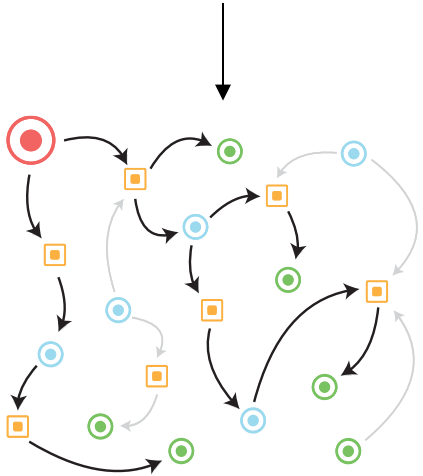


1. Gao et al. "Amortized tree generation for bottom-up synthesis planning and synthesizable molecular design". *ICLR*, 2022

Bidirectional A* cost function and forward expansion policy can be learned from extracted multi-step reaction data

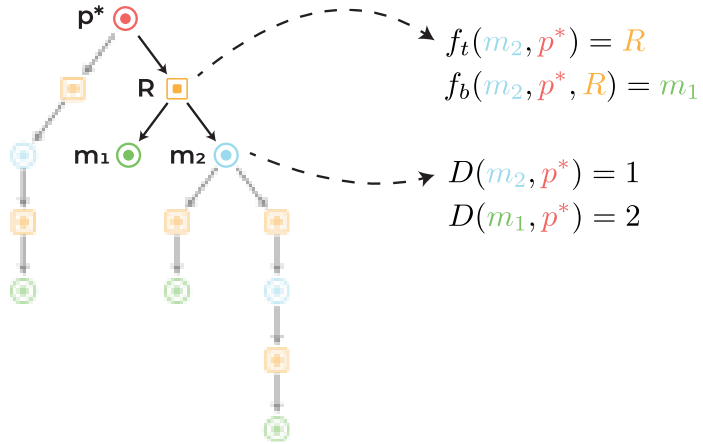


USPTO-Full
~1m reactions
from 1976-2016



Perform enumerative search on non-buyable molecules in synthesis knowledge graph

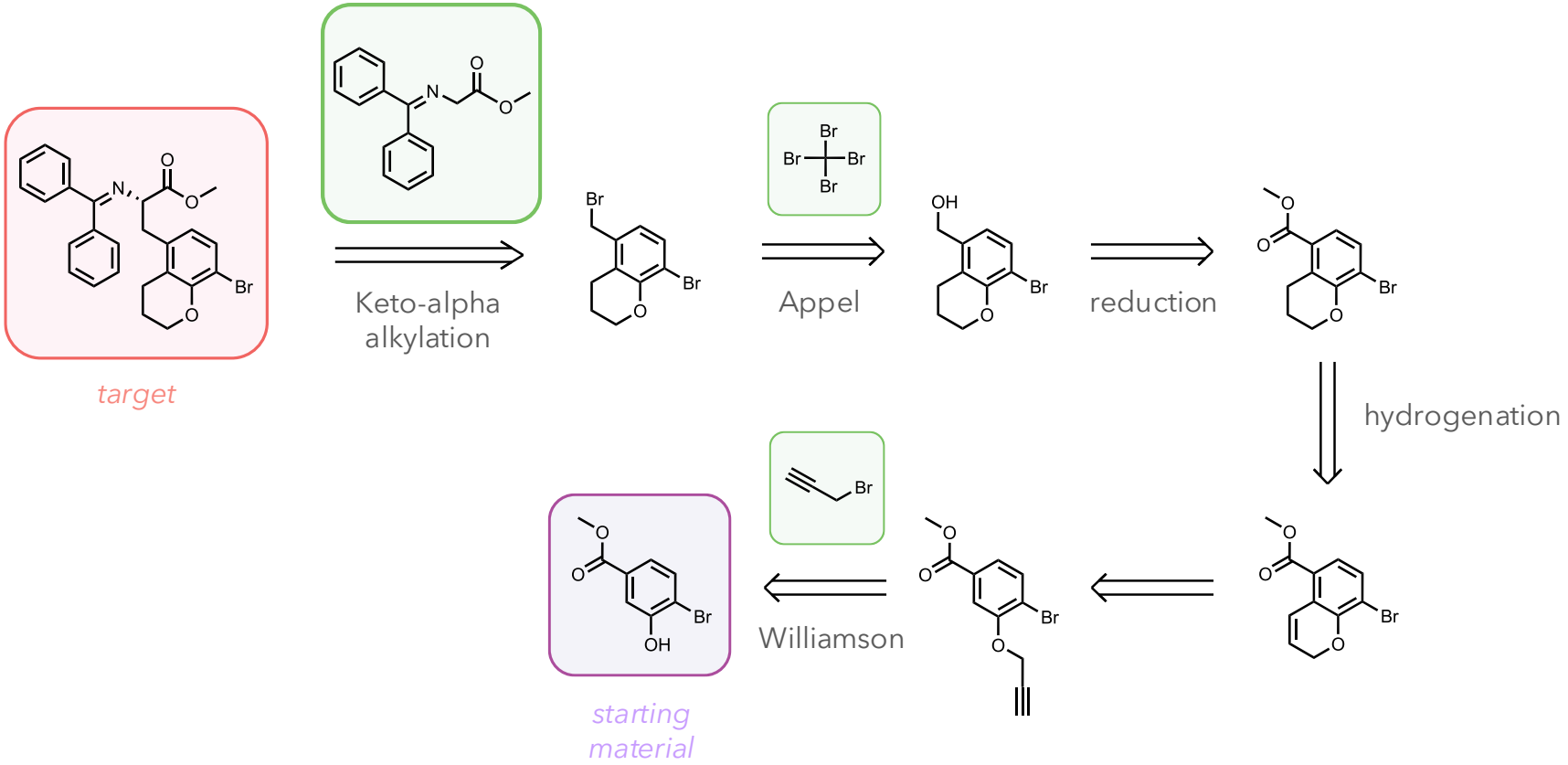
Obtain training examples for "synthetic distance" and conditional forward expansions



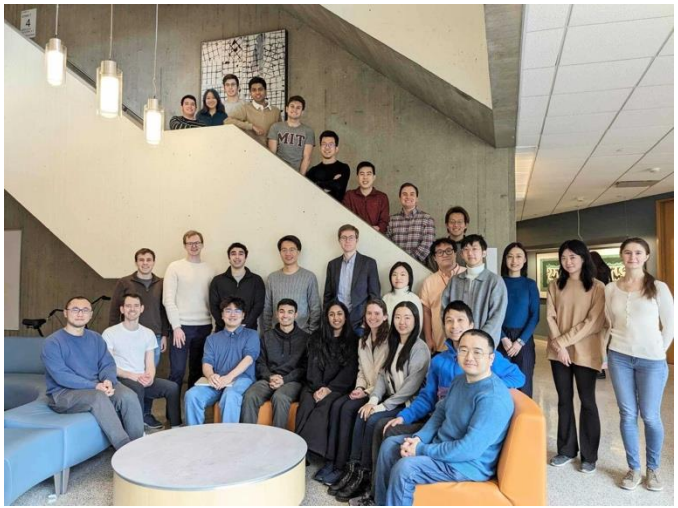
DESP outperforms baseline methods on the starting material-constrained synthesis planning task

Algorithm	**USPTO-190**				**Pistachio Reachable**				**Pistachio Hard**			
	Solve Rate (%) ↑			$\overline{N}$ ↓	Solve Rate (%) ↑			$\overline{N}$ ↓	Solve Rate (%) ↑			$\overline{N}$ ↓
	N=100	300	500		50	100	300		100	300	500	
Random	4.2	4.7	4.7	479	16.0	26.7	40.7	325	6.0	12.0	13.0	452
BFS	12.1	20.0	24.2	413	48.7	57.3	74.0	169	16.0	26.0	29.0	390
MCTS	20.5	32.1	35.3	364	52.0	72.7	85.3	111	27.0	31.0	32.0	361
Retro*	25.8	33.2	35.8	351	70.7	78.0	92.7	73	32.0	35.0	37.0	342
GRASP	15.3	21.1	23.7	410	46.7	51.3	66.7	198	14.0	22.0	29.0	402
Retro*+D	27.4	32.6	37.4	348	77.3	87.3	*96.0*	49	31.0	40.0	42.0	323
DESP-F2E	**30.0**	**35.3**	**39.5**	*340*	*84.0*	**90.0**	*96.0*	*41*	*35.0*	*44.0*	**50.0**	*300*
DESP-F2F	*29.5*	*34.2*	**39.5**	**336**	**84.5**	*88.9*	**97.3**	**38**	**39.0**	**45.0**	*48.0*	**293**

DESP can find routes for targets where single-ended search cannot



Acknowledgements



Center for Computer Assisted Synthesis

MLPDS

Thank you to my co-authors

Jihye Roh
Ziang Li
Wenhao Gao
Runzhong Wang
Prof. Connor Coley

and the rest of the Coley Group!

Dr. Babak Mahjour
Dr. Itai Levin
Dr. Joonyoung Joung
Dr. Jordan Liles
Alex Stoneman
Anji Been
Herry (Tianyi) Jin
Jenna Fromer
Joules Provenzano
Keir Adams
Kento Abeywardane
Magdalena Lederbauer
Mingrou Xie
Mrunali Manjrekar
Nicholas Casetti
Priyanka Raghavan
Shitong Luo
Xiaoqi Sun
Zhengkai Tu
Sourabh Choure
Mun Hong Fong
Huiqian Lin