Hit and Lead Discovery with Explorative RL and Fragment-based Molecule Generation

PRESENTED BY : BERNARD MOUSSAD

UNIVERSITY LIBRARIES
VIRGINIA TECH

Problem Area: Looking for "Hits" 

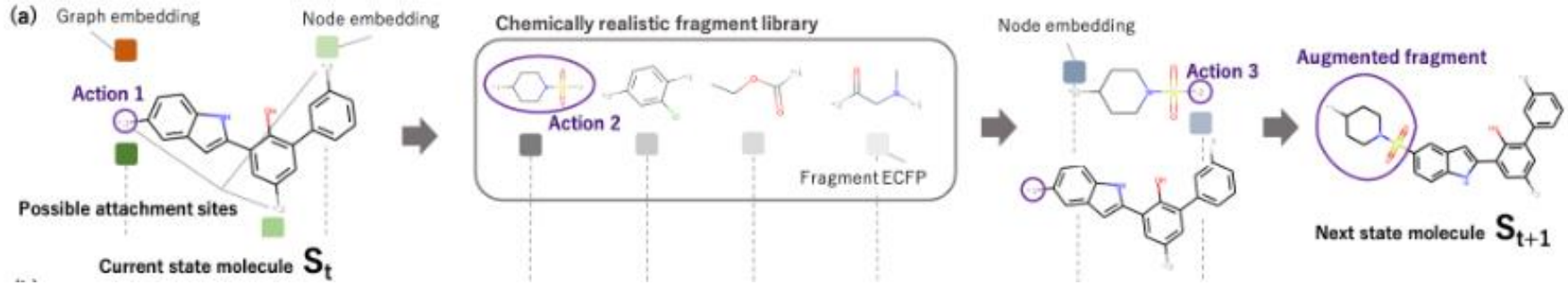
- Drug discovery is about finding "hits", or molecules with desired therapeutic potentials
- Previously performed via brute-force until generative models entered the field
- Reinforcement Learning (RL) methods have produced metrically promising results, but these products have not been guaranteed to be synthesizable
- Docking scores have helped to be a more complex scoring strategy due to the nonlinear nature
  - Can still result in toxic/unstable molecules

Related Works

- SMILES-based methods: Inherently suffer from unrealistic generated molecules
- Motif-based methods: Not compatible with scaffold-based generation because these methods depend on a motif-wise decomposition order
  - This isn’t relevant for docking score
  - Can result in bias in generated molecules

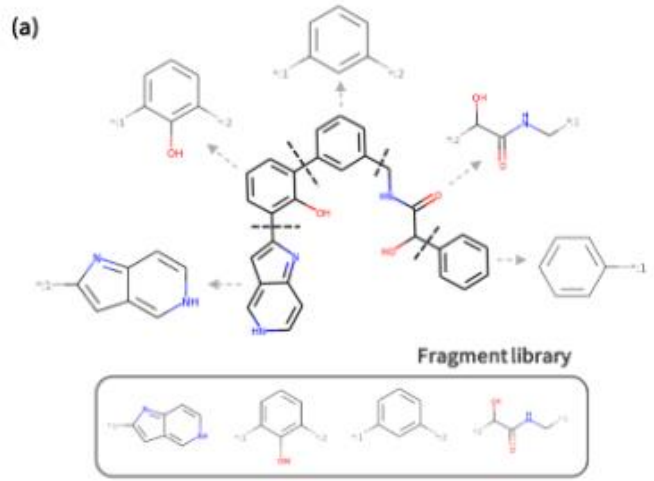
Fragment Based generative RL with Explorative Experience replay for Drug design (FREED): Generation Method

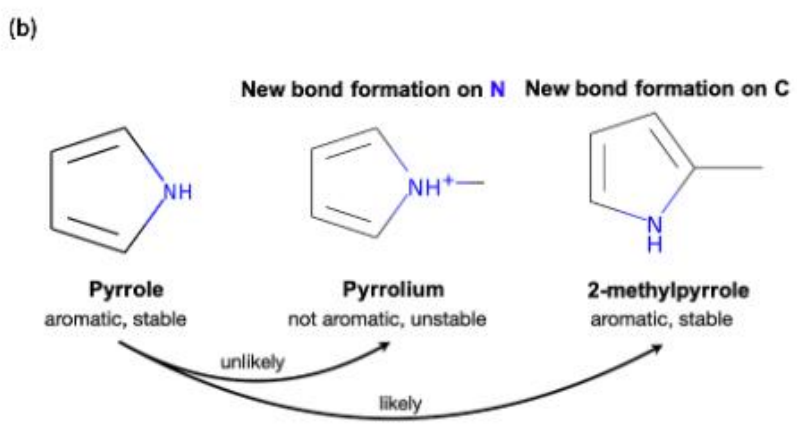
- Implements a “bag of words” approach by using pharmacochemically acceptable molecular fragments to generate molecules with a high docking score
  1. Select where to attach a new fragment in the current state
  2. What fragment to attach
  3. Where on a new fragment to form a new chemical bond



FREED: Generation Method cont.

- Preservation of fragment connectivity information: FREED uses prior knowledge of how the fragments will interact with initial molecules
  - Helps to produce stereochemically valid conformations
  - Helps to identify optimal interaction sites





FREED Policy Network

- Track attachment sites of states to allow policy to know where to attach fragment (1)
- Takes the action from (1) and uses it to predict which fragment should be chosen for (2)
- Track attachment sites of fragments so the policy can choose the attachment site from the fragment side (3)

$$p^{\text{act1}} = \pi_{\text{act1}}(Z^{\text{1st}}), Z^{\text{1st}} = \text{MI}(h_g, H_{\text{att}}) \quad (1)$$

$$p^{\text{act2}} = \pi_{\text{act2}}(Z^{\text{2nd}}), Z^{\text{2nd}} = \text{MI}(z^{\text{1st}}_{\text{act1}}, \text{ECFP}(h_{g_{\text{cand}}})) \quad (2)$$

$$p^{\text{act3}} = \pi_{\text{act3}}(Z^{\text{3rd}}), Z^{\text{3rd}} = \text{MI}(z^{\text{2nd}}_{\text{act2}}, U_{\text{cand}}) \quad (3)$$

Explorative RL: Soft actor-critic (SAC)

- SAC: off-policy actor critic algorithm based on max entropy reinforcement learning
- By requiring the model to be more explorative, it results in a greater number of unique de novo molecules
  - Based on assumption the docking score optimization task requires effective exploration

Explorative RL: Explorative Algorithms

- Prioritized Experience Replay: Novel experience prioritized during sampling batches for RL updates
  - Novel: agent hasn’t visited that state before
- Variants of PER
  - Predictive Error: L2 distance
  - Bayesian Uncertainty
  - Temporal Difference Error

Quantitative Metrics

- **Quality scores**: ratio of accepted, valid molecules to total generated
  - Valid molecules pass these three filters:
    - Glaxo
    - SureChEMBL
    - PAINS
- **Hit Ratio:** Ratio of unique hit molecules to total generated
  - Hits are molecules whose docking scores > median of known active molecule's docking scores
- **Top 5% score:** Avg. score of top 5% scored generated molecules to compare model's ability to produce molecules with better docking scores

Quantitative performance benchmark

- Baselines
  - MORLD and REINEVNT: Models utilized for docking score optimization tasks
  - HierVAE: non-RL VAE for fragment-based molecular generation
    - One-time trained: Trained model on known active molecules
    - Active Learning trained: Trained once on known actives and trained twice on top-scoring generated molecules

Results

Table 1: **Quality scores of the models.** We trained our model and three baseline models with the target fa7 and computed quality scores of the first 3,000 molecules generated during training for each model. The two baseline models REINVENT and MORLD that are jointly trained to maximize filter scores are noted as 'REINVENT w/ filter' and 'MORLD w/ filter'. Standard deviation is given in brackets.

	Glaxo	SureChEMBL	PAINS	validity	uniqueness
MORLD	0.561 (.009)	0.131 (.013)	0.805 (.013)	**1.000** (.000)	**1.000** (.000)
MORLD w/ filter	0.578 (.010)	0.145 (.018)	0.816 (.008)	**1.000** (.000)	**1.000** (.001)
REINVENT	0.773 (.023)	0.667 (.030)	0.769 (.022)	0.813 (.024)	0.988 (.008)
REINVENT w/ filter	0.832 (.034)	0.747 (.040)	0.842 (.034)	0.872 (.028)	0.990 (.007)
HierVAE	0.899 (.027)	0.748 (.024)	0.975 (.006)	1.000 (.000)	0.138 (.006)
HierVAE(AL)	0.975 (.004)	0.795 (.007)	0.893 (.011)	1.000 (.000)	0.131 (.003)
Ours: FREED(PE)	**0.996** (.001)	**0.808** (.049)	**0.991** (.002)	**1.000** (.000)	0.723 (.135)

Results (cont.)

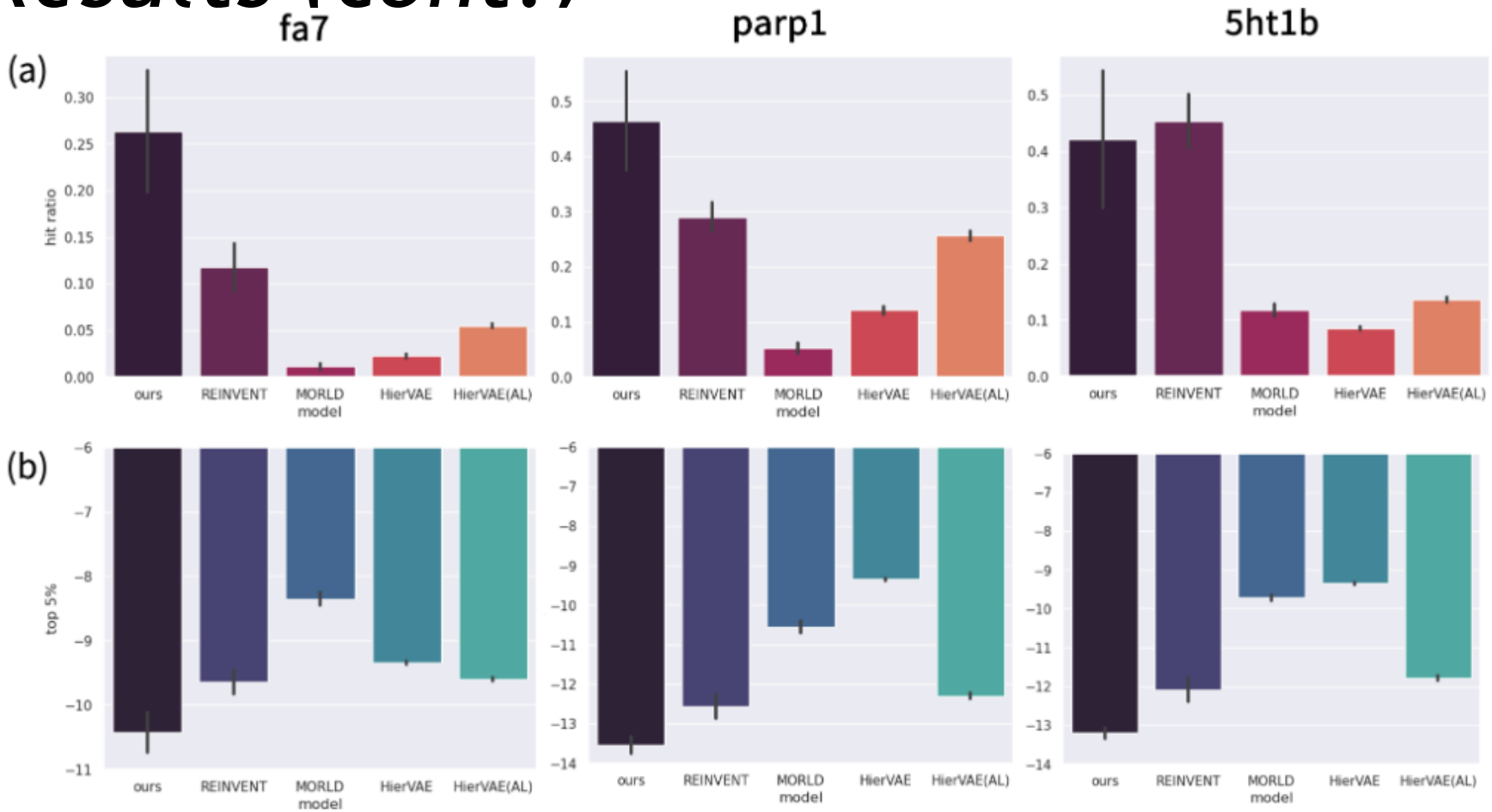


Figure 4: **Hit ratio and top 5% score of our model FREED(PE), REINVENT, MORLD, HierVAE, and HierVAE(AL).** Standard deviation is given as error bars. Higher hit ratios and greater negative value of the top 5% scores indicate better performance.

Ablation Study

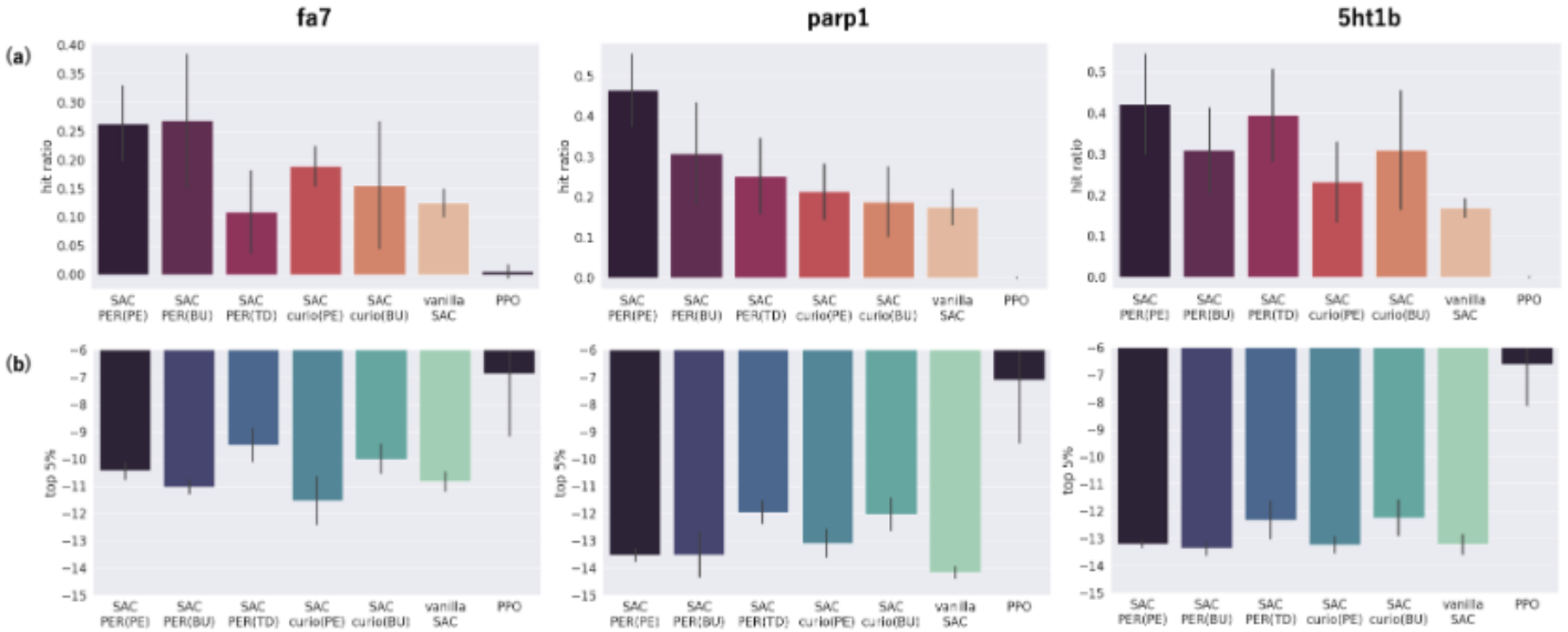


Figure 5: **Hit ratio and top 5% score of ablation studies.** Models can be categorized by whether they use {PER, curiosity-driven exploration(curio)}, and whether they use {predictive error from predictor(PE), Bayesian uncertainty(BU), and TD error from agent(TD)} as priority or intrinsic reward. Standard deviation is given as error bars.

Case Studies on Drug Design

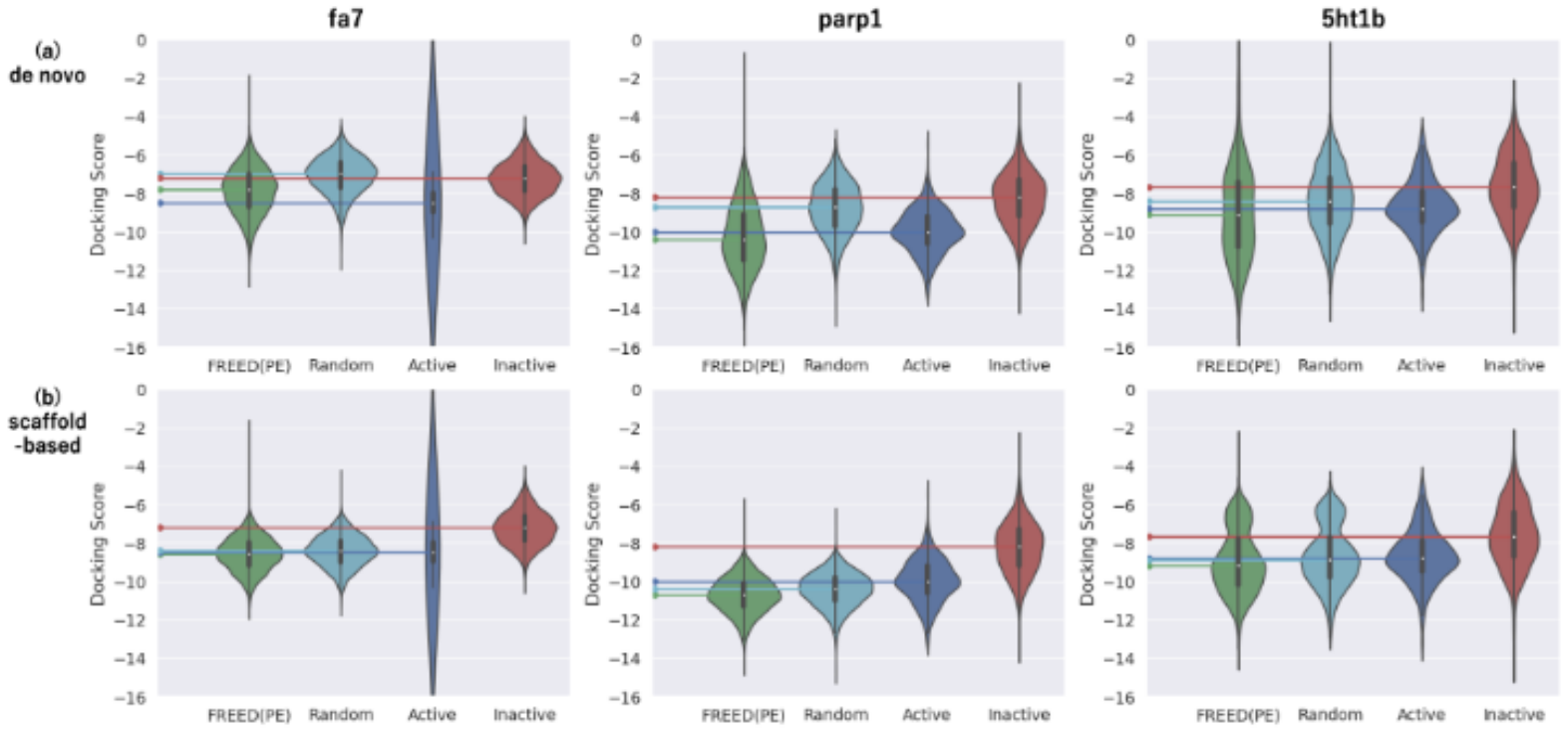


Figure 6: **Docking score distribution of the generated molecules.** Duplicate molecules were removed after gathering 3,000 molecules each from five random seed experiments. "Random" molecules are generated by our fragment-based generation method without training the policy network. "FREED(PE)" molecules are generated by the fragment-based generation method while training the policy network. We also plot known "Active" and "Inactive" molecules from DUD-E (fa7, parp1) or ChEMBL (5ht1b) datasets for comparison. Colored horizontal lines indicate the median of the corresponding distribution. **(a)** *de novo* scenario **(b)** scaffold-based scenario