Generative Models for Structured Based Drug Design

Angelos Nikitaras

Department of Computer Science, TU München
Advisor: Johanna Sommer

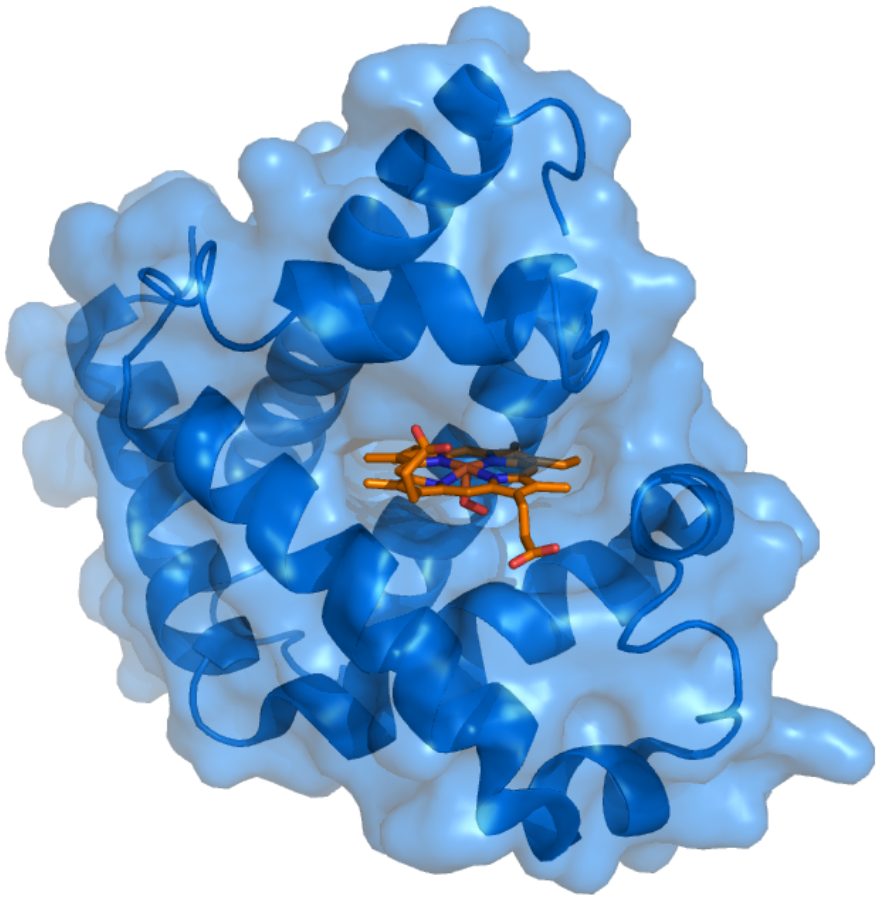
July 23, 2024

Table of Contents

Structured Based Drug Design (SBDD)

Drug design: Find a small molecule(ligand) that binds to a specific protein, modulating its function

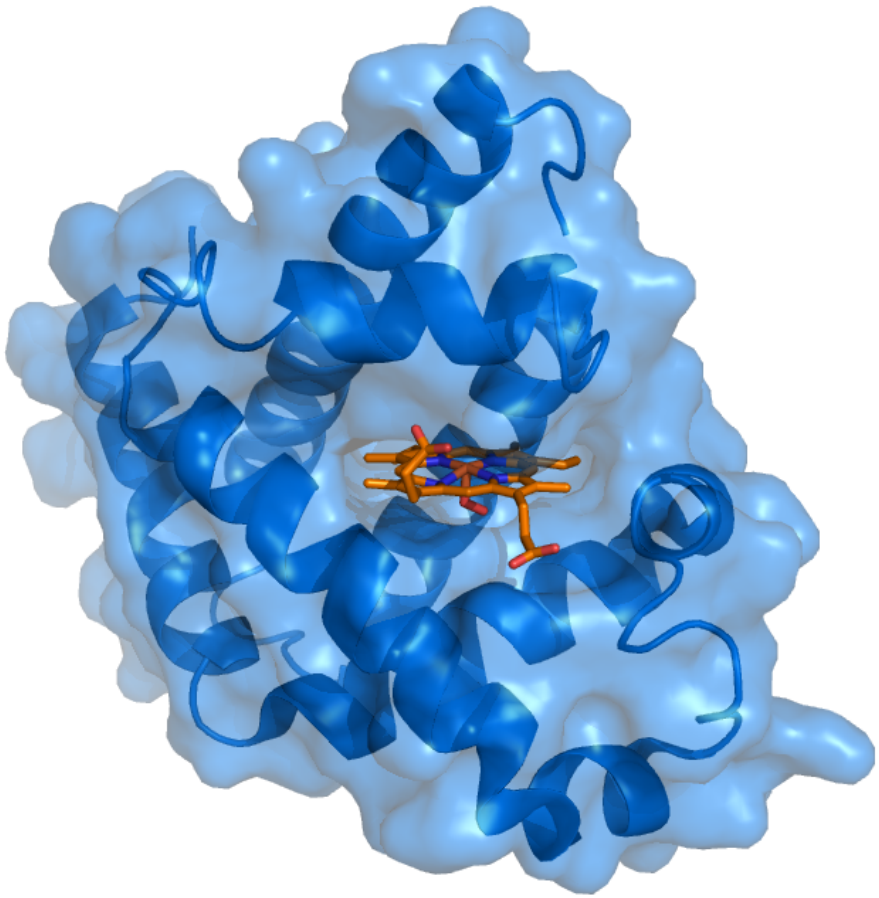
SBDD: Exploit the 3D structure of the protein to design drugs



Structured Based Drug Design (SBDD)

Drug design: Find a small molecule(ligand) that binds to a specific protein, modulating its function

SBDD: Exploit the 3D structure of the protein to design drugs



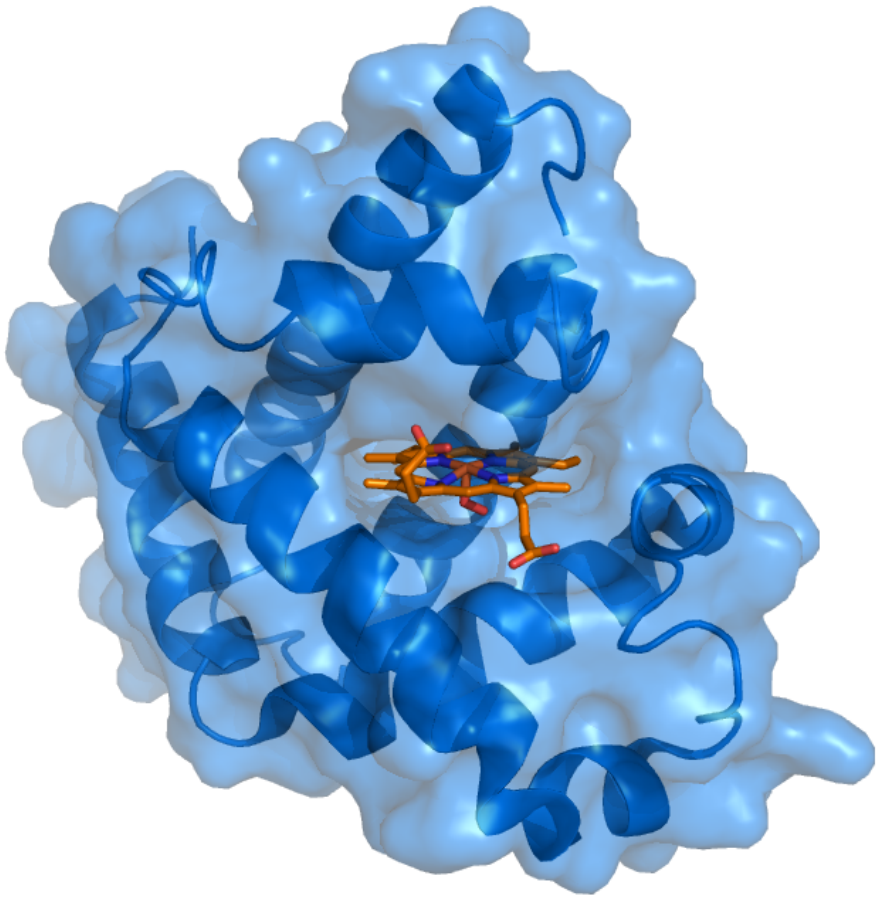
Traditional methods:

- Insufficient exploration of the *enormous* chemical space
- Huge cost

Structured Based Drug Design (SBDD)

Drug design: Find a small molecule(ligand) that binds to a specific protein, modulating its function

SBDD: Exploit the 3D structure of the protein to design drugs



Traditional methods:

- Insufficient exploration of the *enormous* chemical space
- Huge cost

Idea: Use generative models!

Deep Generative Models - Creating new data

panda mad scientist mixing sparkling chemicals, artstation

Basic idea:
Learn the distribution of the data and then sample from it

Deep Generative Models - Creating new data

panda mad scientist mixing sparkling chemicals, artstation

Figure taken from [8]

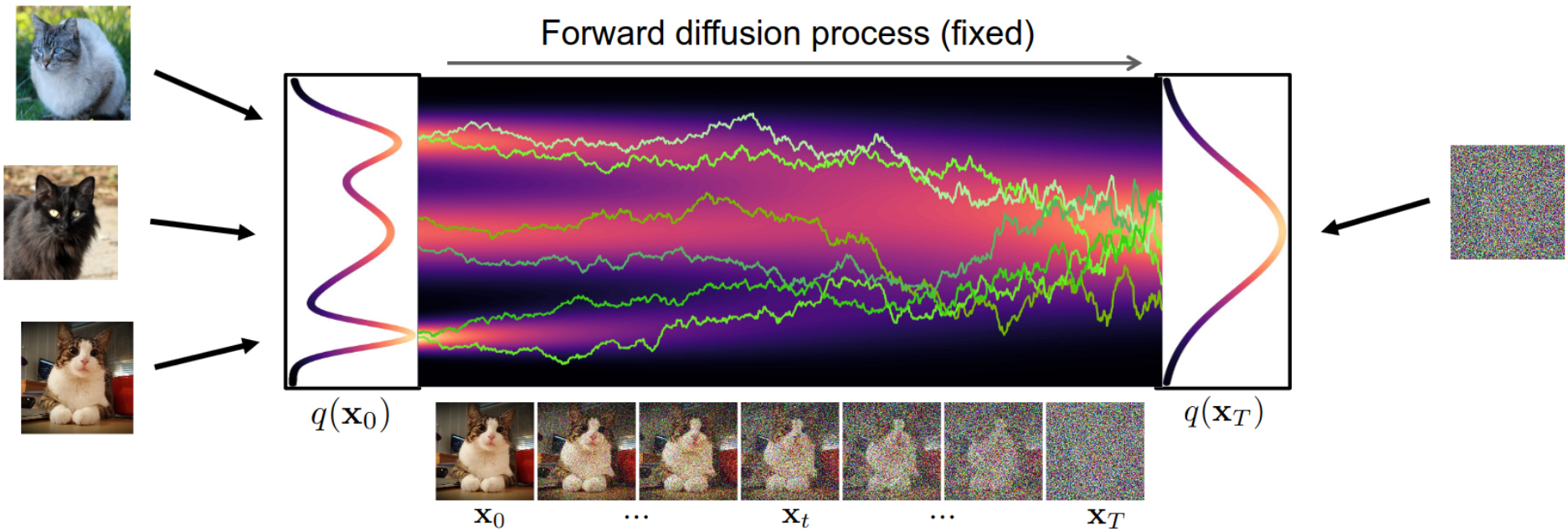
Basic idea:
Learn the distribution of the data and then sample from it

For SBDD:
Capture the conditional distribution of a ligand forming a complex with a given protein

Table of Contents

Diffusion Models - Generating data from noise

- **Forward process:** Add noise gradually $\rightarrow$ pure noise at the end
- **Backward process:** A Neural Network learns to reconstruct the original data
- **Generation:** Sample pure noise and use the backward process



Ho et al., Denoising diffusion probabilistic models. NeurIPS, 2020
Figure taken from [14]

Graph Neural Networks - Operating on molecules

- Molecules are graphs $\rightarrow$ **GNNs**

Graph Neural Networks - Operating on molecules

- Molecules are graphs $\rightarrow$ **GNNs**
- Vanilla GNNs are not enough $\rightarrow$ We need **equivariance**

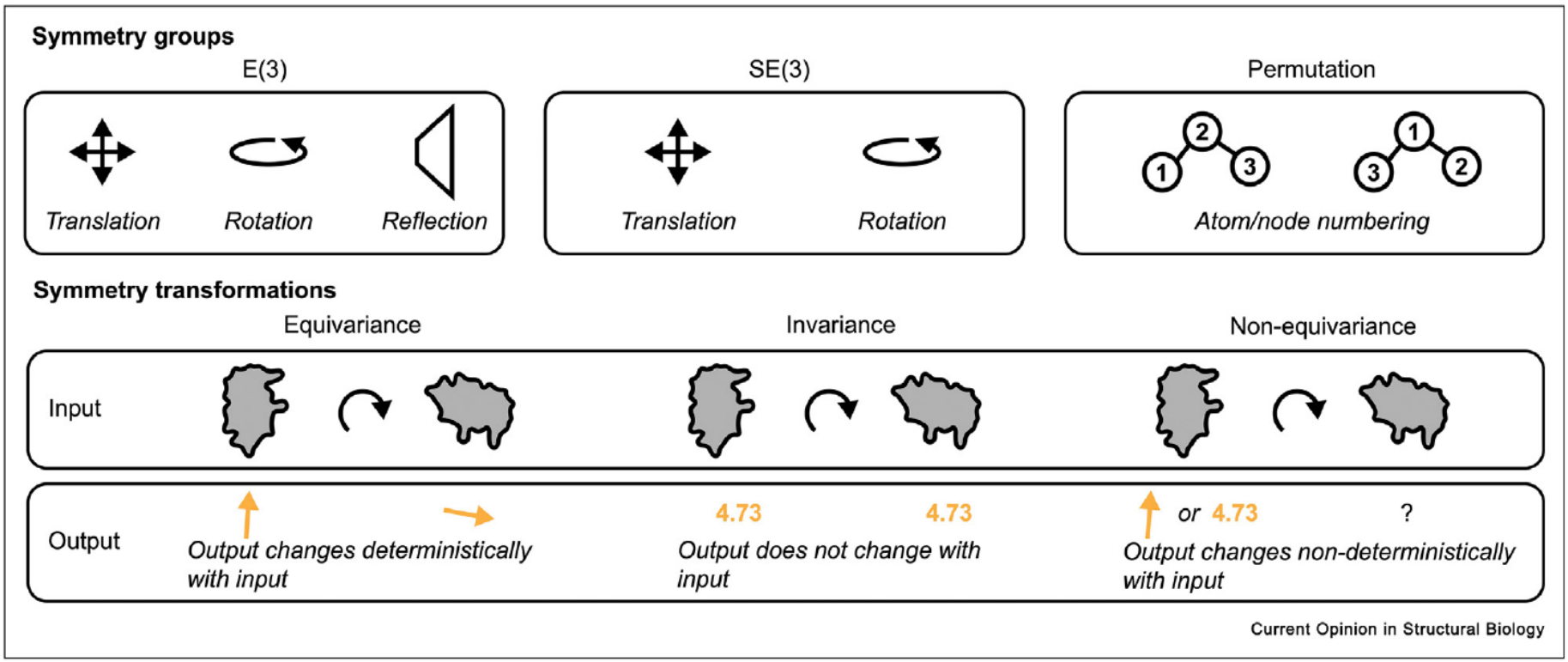
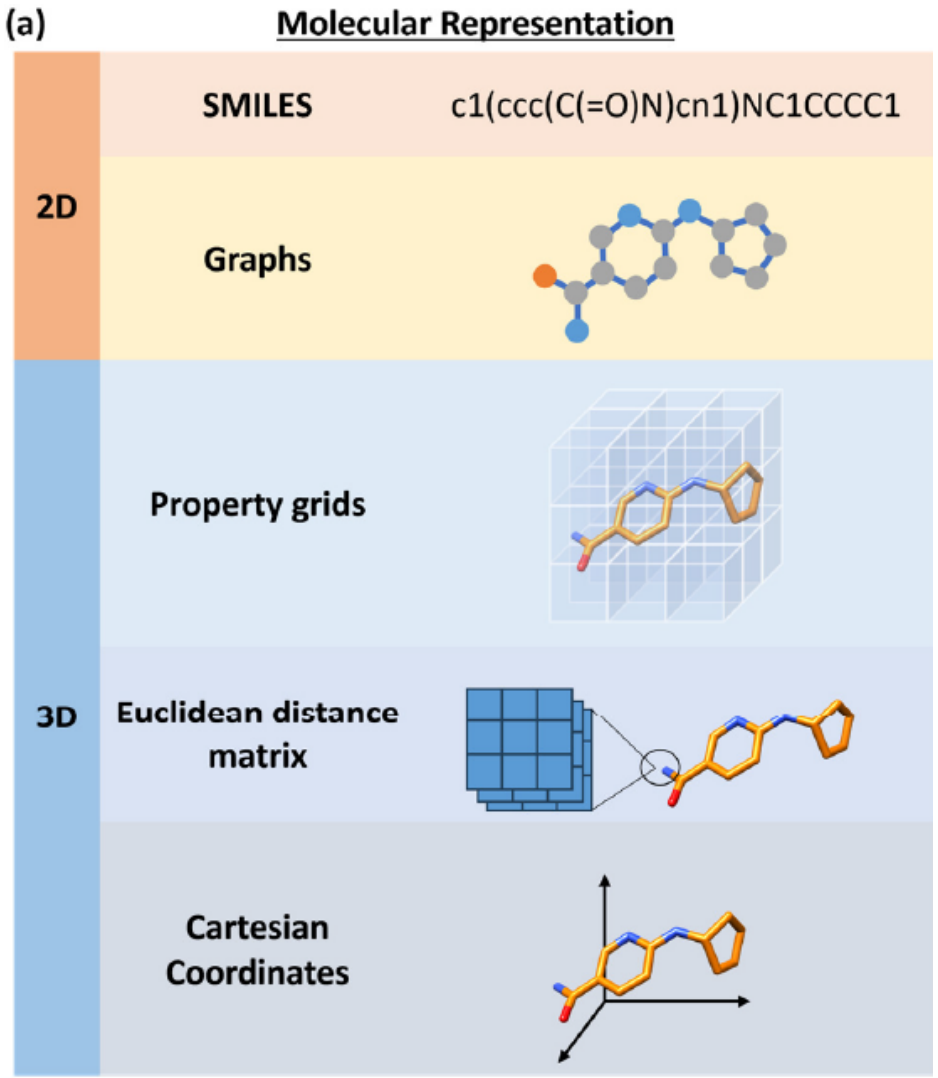


Figure taken from [1]

Table of Contents

Choosing a representation



- **2D/1D representations:** Convert molecules to strings (e.g SMILES), use sequence-based models like RNNs, Transformers [10,16,17]

Choosing a representation

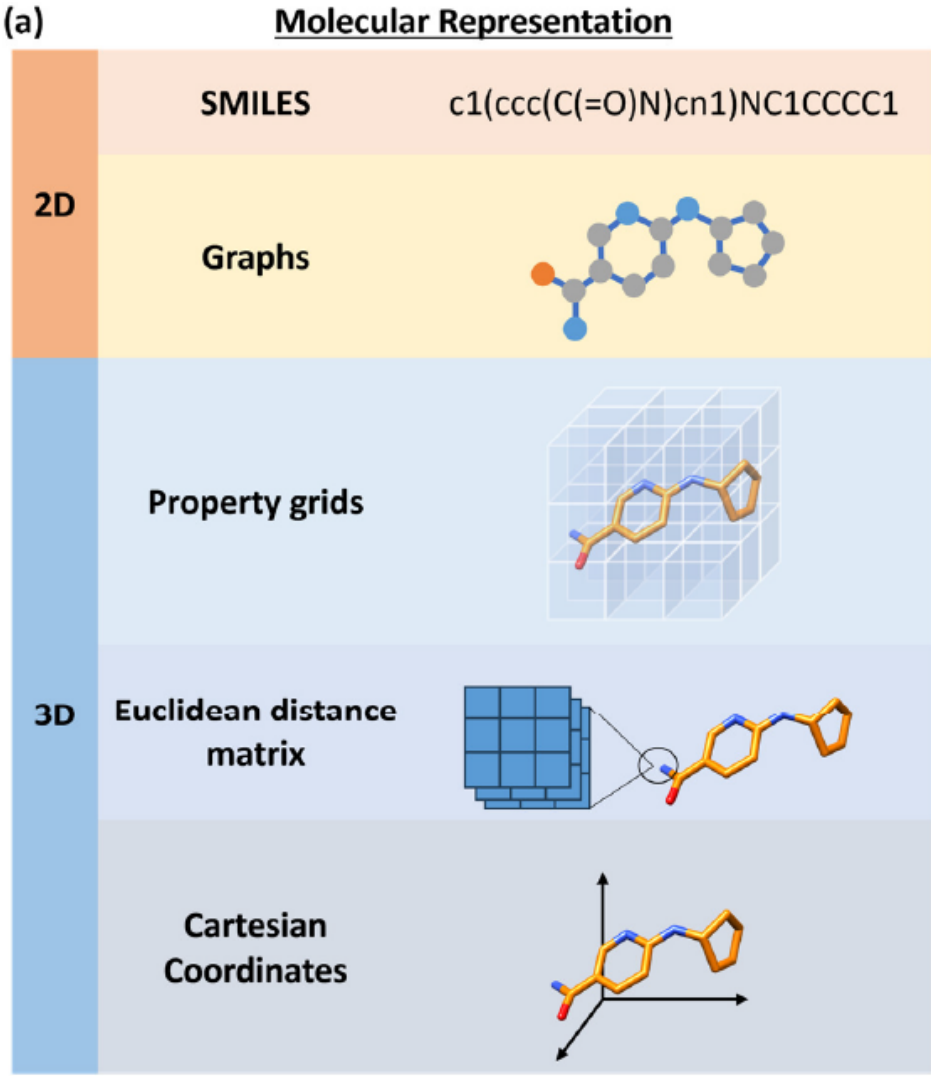
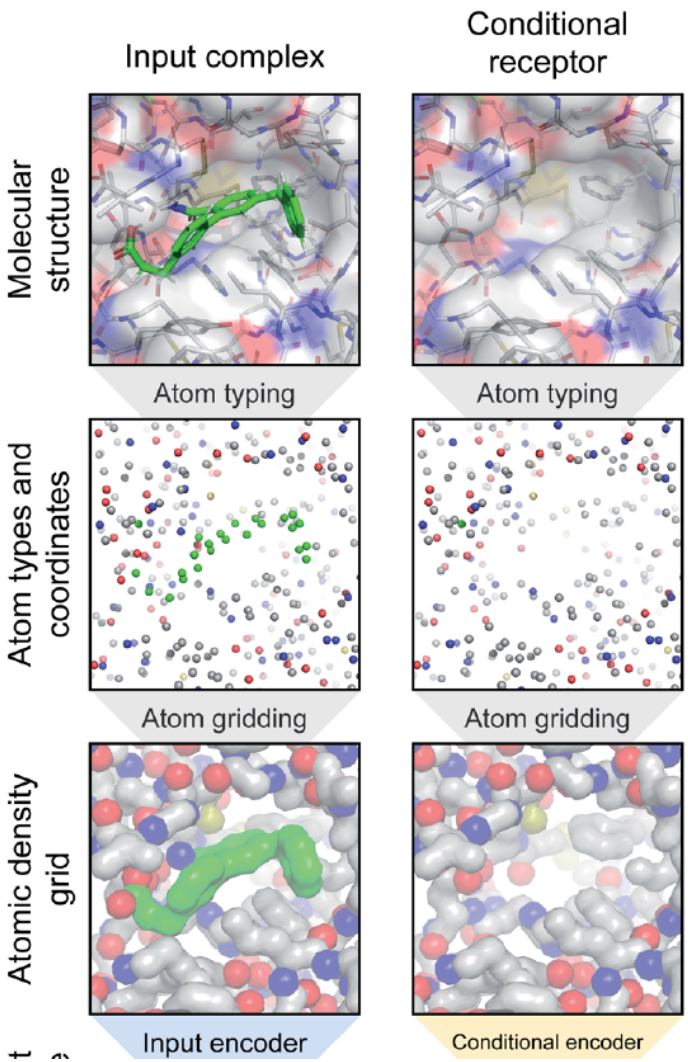


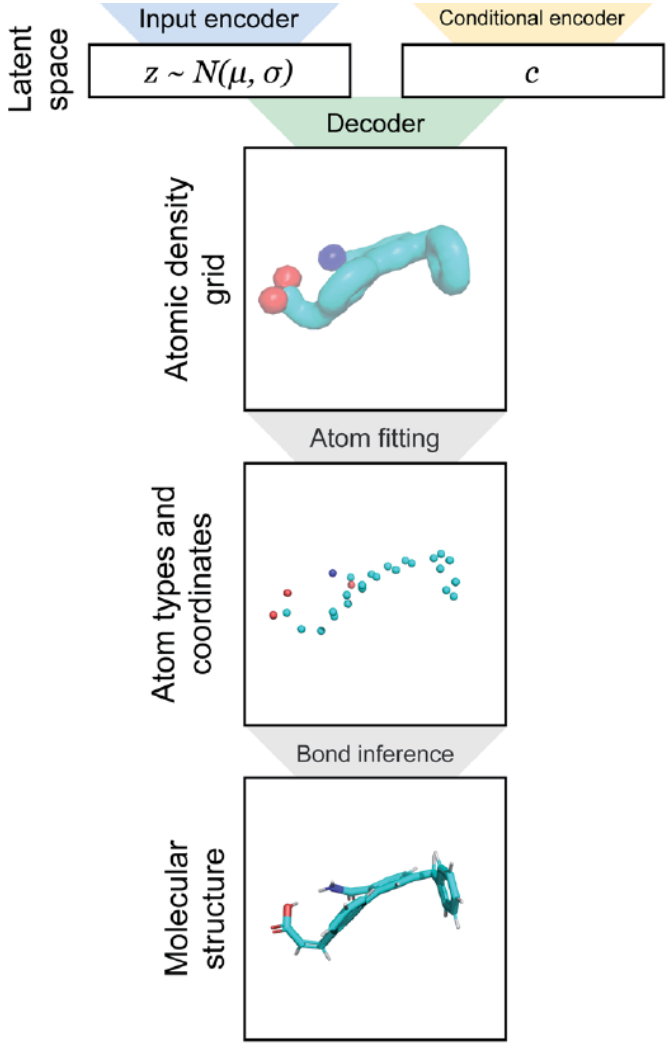
Figure taken from [7]

- **2D/1D representations:** Convert molecules to strings (e.g SMILES), use sequence-based models like RNNs, Transformers [10,16,17]
- **3D representations:** Explicitly model the 3D structure of the molecule

Grid-based models - liGAN



(a)



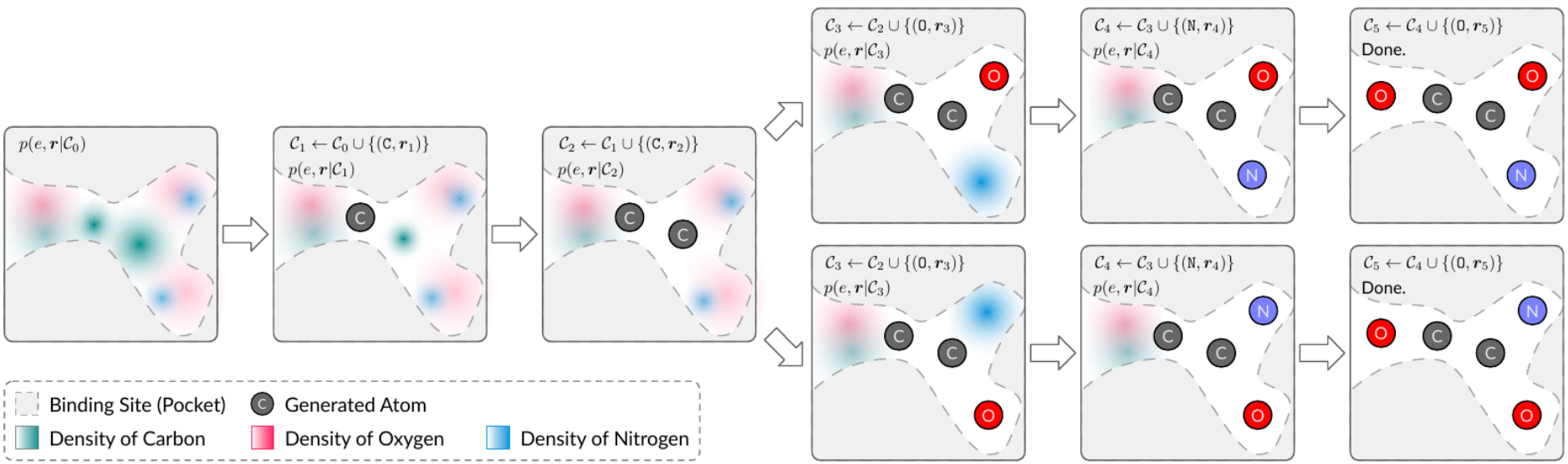
(b)

Ragoza et. al, Generating 3D molecules conditional on receptor binding sites with deep generative models, Chemical Science 2022

Autoregressive models - Luo et. al

Main ideas:

- Model the density $p(e, r|\mathcal{C}_0)$
- Sample atoms one by one
- Rotation/Translation invariant GNNs for **context encoding**
- A **spatial classifier** estimates the probability of finding an atom of type e at position r

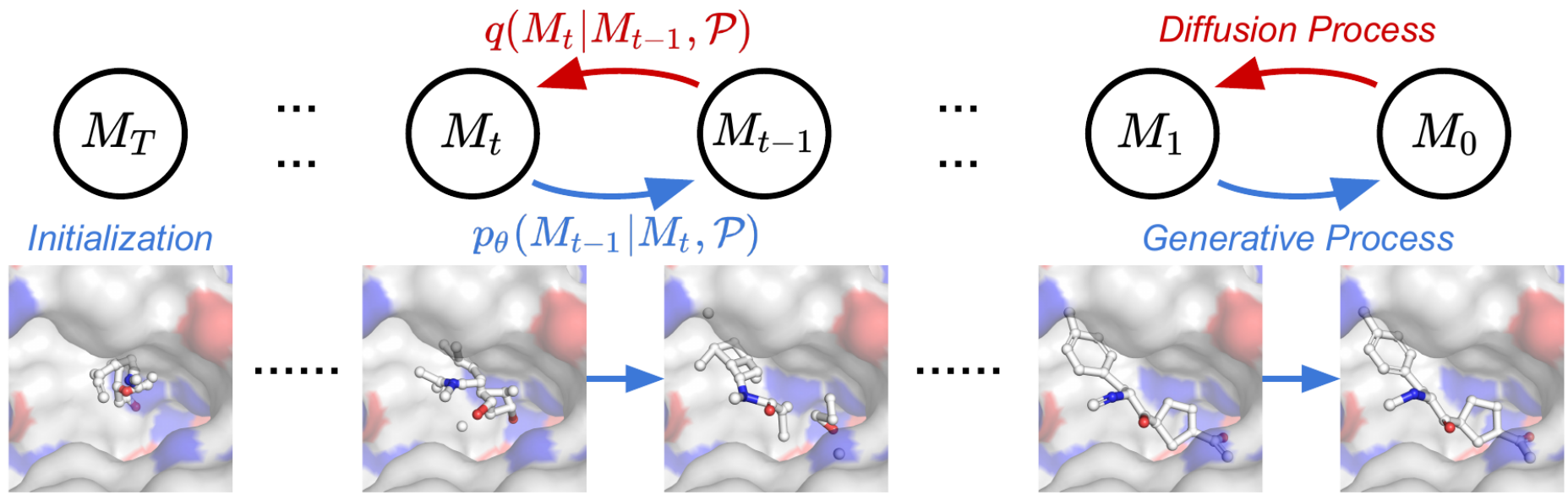


Luo et. al, A 3D Generative Model for Structure-Based Drug Design. NeurIPS 2021

Diffusion models - TargetDiff

Main ideas:

- Diffusion model, invariant likelihood for SE(3) transformations
- Equivariant GNNs for the denoising process
- Diffusion on both positions and atom types, protein remains fixed

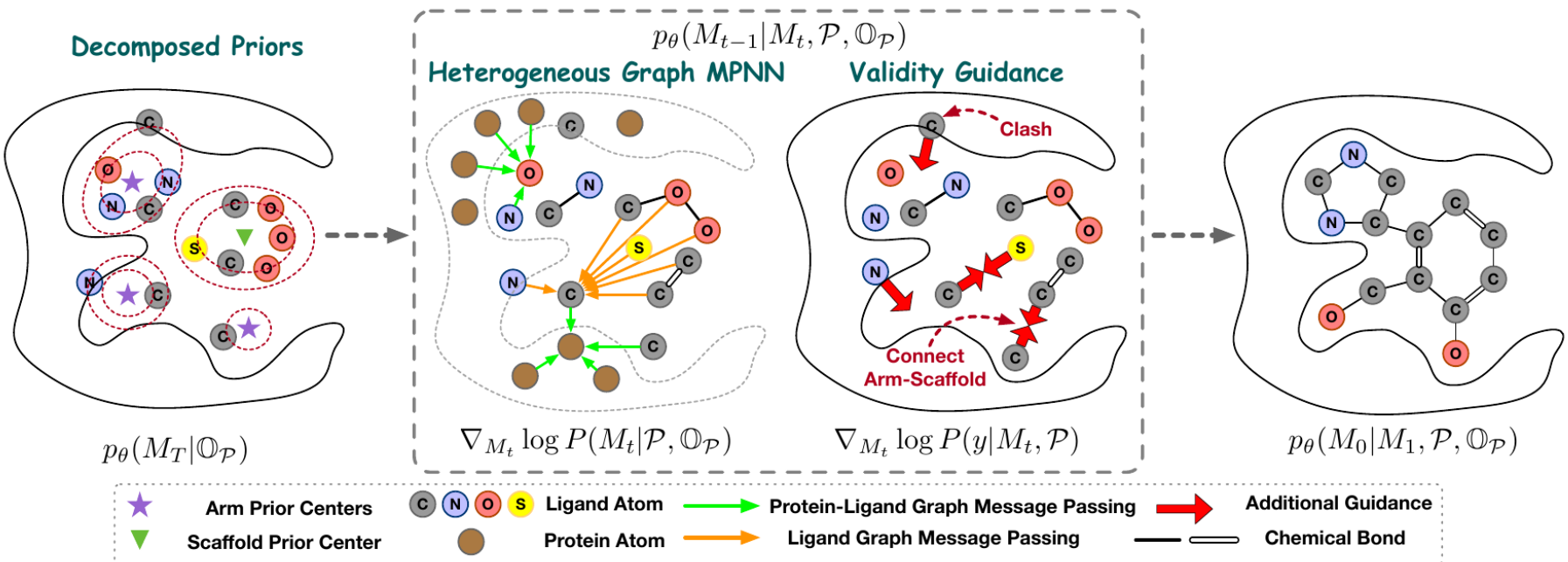


Guan et al, 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction, ICLR 2023

Diffusion models - DecompDiff

Main ideas:

- Equivariant diffusion model
- Decomposed priors over *arms* and *scaffold*
- Both atom and bond diffusion
- Validity guidance



Guan et al., DecompDiff: Diffusion models with Decomposed Priors for Structure-Based Drug Design, ICML 2023

Table of Contents

Evaluation metrics - Results

Evaluation metrics

- Vina Score $\rightarrow$ Binding affinity
- QED $\rightarrow$ Drug-likeness
- SA $\rightarrow$ Synthetic accessibility
- Diversity

Evaluation metrics - Results

Evaluation metrics

- Vina Score $\rightarrow$ Binding affinity
- QED $\rightarrow$ Drug-likeness
- SA $\rightarrow$ Synthetic accessibility
- Diversity

Data sets

CrossDocked [11], synthetically generated dataset

Evaluation metrics - Results

Evaluation metrics

- Vina Score $\rightarrow$ Binding affinity
- QED $\rightarrow$ Drug-likeness
- SA $\rightarrow$ Synthetic accessibility
- Diversity

Data sets

CrossDocked [11], synthetically generated dataset

Method	Vina Score	QED	SA	Diversity
liGAN	-6.33	0.39	0.59	0.66
Luo et al	-6.75	**0.51**	**0.63**	0.70
TargetDiff	-7.80	0.48	0.58	**0.72**
DecompDiff	**-8.39**	0.45	0.61	0.68

Discussion

Current issues:

- Not easy to compare 1D and 3D methods
- Need for a comprehensive evaluation framework for SBDD
- Experimental validation
- Both datasets and evaluation metrics are limited by score functions

Outlook: Close collaboration between domains experts and ML researchers, to combine state-of-the-art methods with domain-specific knowledge

References

[1] Isert, C., Atz, K., and Schneider, G. Structure-based drug design with geometric deep learning. Current Opinion in Structural Biology, 79:102548, 2023.
[2] Luo, S., Guan, J., Ma, J., and Peng, J. A 3d generative model for structure-based drug design. Advances in Neural Information Processing Systems, 34:6229–6239, 2021.
[3] Liu, M., Luo, Y., Uchino, K., Maruhashi, K., and Ji, S. Generating 3d molecules for target protein binding. arXiv preprint arXiv:2204.09410, 2022.
[4] Guan, J., Qian, W. W., Peng, X., Su, Y., Peng, J., and Ma, J. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. arXiv preprint arXiv:2303.03543, 2023.
[5] Schneuing, A., Du, Y., Harris, C., Jamasb, A., Igashov, I., Du, W., Blundell, T., Lió, P., Gomes, C., Welling, M., et al. Structure-based drug design with equivariant diffusion models. arXiv preprint arXiv:2210.13695, 2022.
[6] Guan, J., Zhou, X., Yang, Y., Bao, Y., Peng, J., Ma, J., Liu, Q., Wang, L., and Gu, Q. Decompdiff: diffusion models with decomposed priors for structure-based drug design. arXiv preprint arXiv:2403.07902, 2024.
[7]Tang, Y., Moretti, R., and Meiler, J. Recent advances in automated structure-based de novo drug design. Journal of Chemical Information and Modeling, 64(6):1794–1805,2024. [8] wikipedia.org
[9]Ramesh, Aditya, et al. ”Hierarchical text-conditional image generation with clip latents.” arXiv preprint arXiv:2204.06125 1.2 (2022): 3.
[10] Xu, Mingyuan, Ting Ran, and Hongming Chen. ”De novo molecule design through the molecular generative model conditioned by 3D information of protein binding sites.” Journal of Chemical Information and Modeling 61.7 (2021): 3240-3254.
[11] Francoeur, P. G., Masuda, T., Sunseri, J., Jia, A., Iovanisci, R. B., Snyder, I., and Koes, D. R. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. Journal of chemical information and modeling, 60(9):4200–4215, 2020.
[12]Hu, L., Benson, M. L., Smith, R. D., Lerner, M. G., and Carlson, H. A. Binding moad (mother of all databases). Proteins: Structure, Function, and Bioinformatics, 60(3): 333–340, 2005.
[13]Ragoza, Matthew, Tomohide Masuda, and David Ryan Koes. ”Generating 3D molecules conditional on receptor binding sites with deep generative models.” Chemical science 13.9 (2022): 2701-2713.
[14] CVPR 2022, Tutorial ”Denoising Diffusion-based Generative Modeling: Foundations and Applications, https://cvpr2022-tutorial-diffusion-models.github.io/
[15] Satorras, Garcia, Hoogeboom, Welling. “E(n) equivariant graph neural networks”. ICML 2021.
[16] Grechishnikova, Daria. ”Transformer neural network for protein-specific de novo drug generation as a machine translation problem.” Scientific reports 11.1 (2021): 321.
[17] Wang, Ye, et al. ”cMolGPT: A conditional generative pre-trained transformer for target-specific de novo molecular generation.” Molecules 28.11 (2023): 4430.

Thank you for your attention!