# 3DLinker: An E(3) Equivariant Variational Autoencoder for Molecular Linker Design

ICML 2022

Department of Artificial Intelligence Korea University

Younghan Son

Sep 2022

MAI Lab EDICAL

# Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices

## AI/ML-Enabled Medical Devices

Devices are listed in reverse chronological order by Date of Final Decision. To change the sort order, click the arrows in the column headings.

Use the Submission Number link to display the approval, authorization, or clearance information for the device in the appropriate FDA database. The database page will include a link to the FDA's publicly available information.

Search:

Show 50 entries

| Date of Final Decision | Submission Number | Device | Company | Panel (Lead) | Primary Product Code |
|---|---|---|---|---|---|
| 06/17/2021 | K203514 | Precise Position | Philips Healthcare (Suzhou) Co., Ltd. | Radiology | JAK |
| 06/16/2021 | K202718 | Qmenta Care Platform Family | Mint Labs, Inc., D/B/A. QMENTA | Radiology | LLZ |
| 06/11/2021 | K210484 | LINQ II Insertable Cardiac Monitor, Zelda AI ECG Classification System | Medtronic, Inc. | Cardiovascular | MXD |
| 06/10/2021 | K203629 | IDx-DR | Digital Diagnostics Inc. | Ophthalmic | PIB |
| 06/02/2021 | DEN200069 | Cognoa Asd Diagnosis Aid | Cognoa, Inc. | Neurology | QPF |
| 05/19/2021 | K210237 | CINA CHEST | Avicenna.AI | Radiology | QAS |
| 04/30/2021 | K210001 | HYPER AiR | Shanghai United Imaging Healthcare Co. Ltd. | Radiology | KPS |

MAI Lab EDICAL

# [배진건의 신약이야기] 나스닥 히어로 된 '키메라'...비결은 '죽음의 키스'

입력 2020.10.08 17:26 수정 2020.10.12 08:38 가가

배진건 이노큐어테라퓨틱스 수석부사장

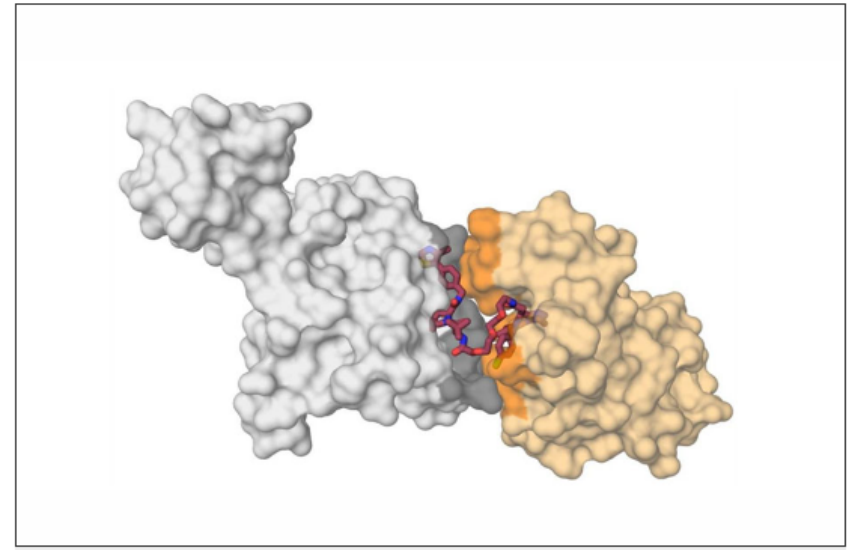

최근 2년간 나스닥에 3개 회사의 상장을 이끈 원천 기술(플랫폼 테크놀로지)이 있다. 바로 프로탁(PROTAC)이다. 프로탁은 프로티알리시스 타깃팅 키메라(PROteolysis-TArgeting Chimera)의 약자로, 특정 단백질의 분해를 유도하는 기술이다.

오늘의 주요뉴


## 페이스오프, 그 죽음의 입맞춤

프로탁의 발전 모습을 최신 논문에서 살펴볼까요? 올해 <네이처 화학생물학(Nature Chemical Biology)>에 실린 논문[7]에는 E3 효소와 프로탁 소분자, 표적 단백질이 결합한 3차원 구조가 실렸습니다. 'MZ1'이라고 명명된 이 특정 프로탁 구조에서는 E3 효소(VHL)와 결합하는 데 VH032라는 소분자가 쓰였으며, 분해 대상이 되는 표적 단백질인 Brd4(이 단백질은 급성 골수형 백혈병 치료에 이용될 표적 단백질로 발굴되었는데 다른 암 치료제에 활용될 가능성도 있다[8])와 결합하도록 JQ1이라는 소분자가 선택됐습니다. 두 소분자들은 연결부(PEG)로 이어져, "VH032-PEG linker-JQ1"가 MZ1라는 프로탁을 구성합니다(**그림6**).

그림에서 보듯이 MZ1으로 명명된 프로탁에 유비퀴틴 단백질인 E3 효소와 분해 대상 단백질 Brd4가 근접합니다. 특히 MZ1 프로탁은, E3와 Brd4 두 단백질 간의 물리적 거리를 충분히 줄여 두 단백질 간에 직접 결합(PPI: Protein-Protein Interaction)이 일어나도록 유도했습니다.



» 그림7. 그림6에서 확인한 단백질들과 프로탁의 결정 구조. 프로탁(MZ1)으로 인해 두 단백질 (회색: E3 효소, 오렌지: 표적 단백질 Brd4)이 서로 결합하고 있다.[9] 출처/ Xavier Lucas, Alessio Ciulli, University of Dundee

3차원 그림을 살펴보면, 프로탁 소분자의 '중매' 또는 '연결' 구실 덕분에 마치 E3 효소와 표적 단백질 Brd4가 입맞춤하는 듯이 서로 결합하는 모습을 볼 수 있습니다(**그림7**).[9] 그러나 둘의 입맞춤은, 우리가 단백질 분해 과정에서 보았듯이, 표적 단백질을 죽음으로 이끕니다.

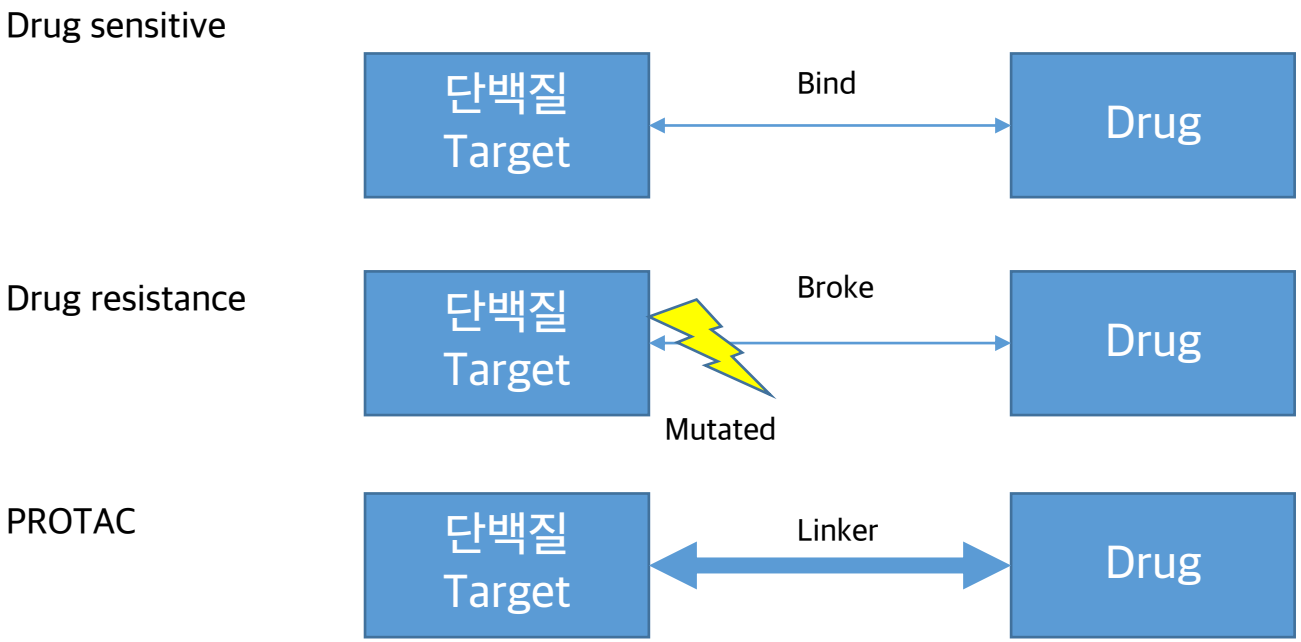
프로탁을 이용하면, 기존 신약 개발 방법으로는 접근할 수 없었던 많은 질병 관련 단백질을 약물 표적으로 이용하여, 현재는 존재하지 않는 질병 치료제의 개발을 앞당길 수 있을 것으로 기대됩니다. 물론, 새로운 신약 개발 방법에 의해 고안된 소분자들이 최종적으로 미식품의약국(FDA)의 허가를 받은 신약으로서 실제로 환자 치료에 쓰일 때까지는 많은 시간이 더 필요하지만, 멀지 않은 미래에 우리 가족과 친구들 중 누군가가 그 혜택을 받는 꿈을 꾸

- 이 과정에서 일어나는 약물과 단백질의 결합은 매우 안정적인 '공유결합'이 아니기에 분리될 수 있고(가역적이고, reversible), 따라서 약효를 유지하려면 일반적으로 몸 전체에 높은 약물 농도를 유지해야 합니다. 하지만 약물 효능을 지키려고 약물 농도를 높이면 약물이 예기치 못한 다른 단백질과 결합할 수 있는데, 이런 상황은 종종 약물 부작용으로 나타납니다.

- 그렇다면, 약 85% 이상의 질병 관련 단백질들을 표적으로 하는 약물이 아직까지 개발되지 못하는 이유는 무엇일까요? 여러 이유가 있겠지만 가장 중요한 이유는 현재의 약물 개발 방법으로는 이 단백질들을 표적으로 삼아 그 기능을 억제하기가 어렵기 때문입니다. 예를 들어, 어떤 단백질들은 약물이 잘 결합할 수 있는 '소수성 포켓(hydrophobic pocket)'이라는 구조를 갖추고 있지 않거나, 또는 그 전통적인 약물 표적 구조가 단백질의 기능과 상관없는 곳에 있기도 합니다. 질병과 관련한 어떤 효소들은 활성부위(active site)에 다른 보조인자(cofactor)가 워낙 강하게 결합하는 특성을 지니기에, 효소 활성부위에 결합해야 하는 약물을 개발하기가 어렵기도 합니다. 또한 알츠하이머를 비롯해 여러 뇌질환의 원인으로 알려진 '타우 단백질의 엉킴(Tau tangles)'과 같은 경우에, 약물의 결합으로 '단백질 엉킴'을 제거할 방법이 아직 존재하지 않습니다.[3]

# Why PROTAC?

- Weak binding affinity of drug and target protein leads to drug resistance



# Linker

- Predicted linker coordinate must 3d aligned(plausible)

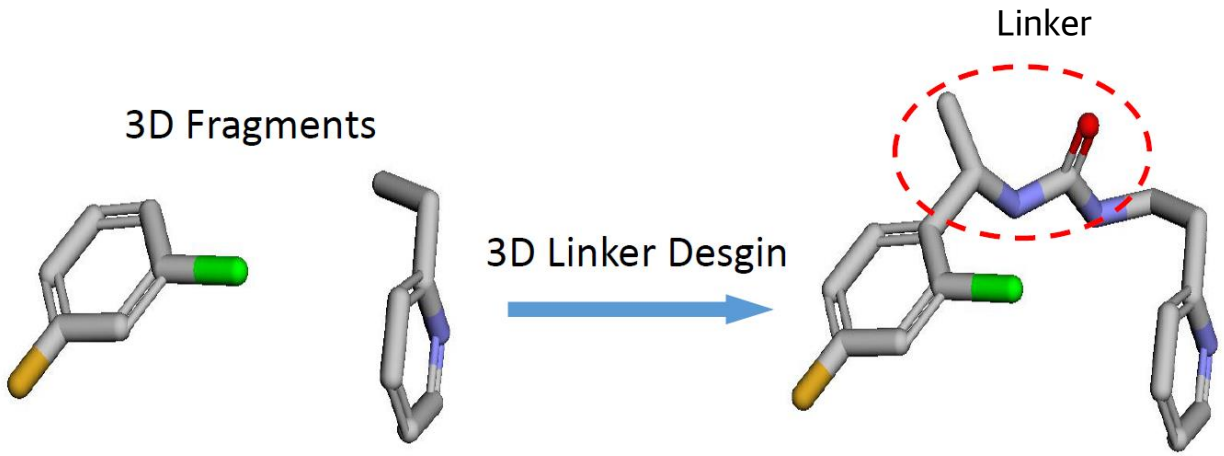



Figure 1: 3D linker design problem: given two fragments' graph with 3D coordinates (left), the goal is to generate a linker graph with 3D coordinates to link these two fragments (right). The 3D coordinates of the generated linker must align with the two fragments, otherwise they cannot link.

# Graph

Node, Edge, Node type

$$G = (\mathcal{V}, \mathcal{E}, X)$$

True Graph/coordinate | Given 2 Fragments

$$p(G, \boldsymbol{R} | G_{\mathrm{F}}, \boldsymbol{R}_{\mathrm{F}})$$

Invariant, Angle, distance

Input Translate = Output Translate

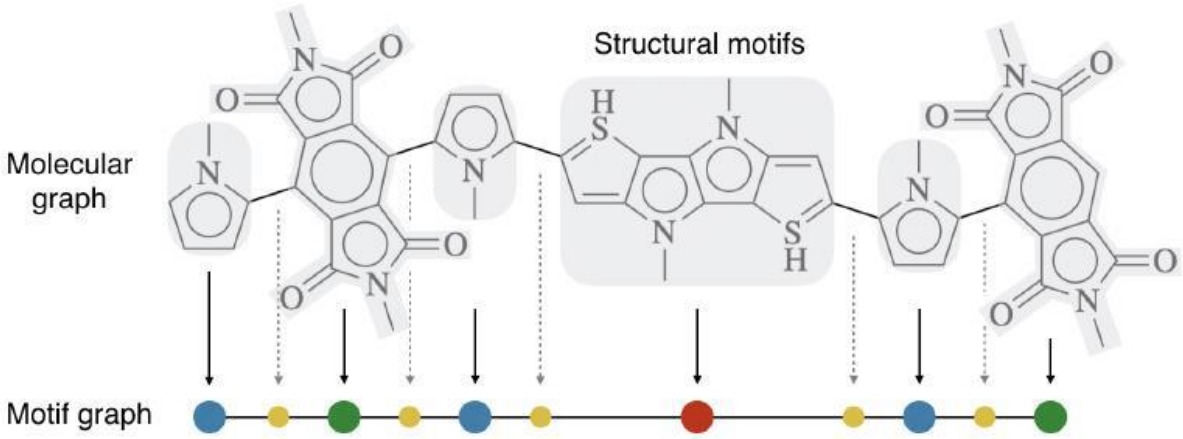
$$\pi^{\mathcal{Y}}(g)\phi(x) = \phi(\pi^{\mathcal{X}}(g)x)$$

Equivariant, Coordinate

$$p(G, \pi(g)\boldsymbol{R} | G_{\mathrm{F}}, \pi(g)\boldsymbol{R}_{\mathrm{F}}) = p(G, \boldsymbol{R} | G_{\mathrm{F}}, \boldsymbol{R}_{\mathrm{F}})$$

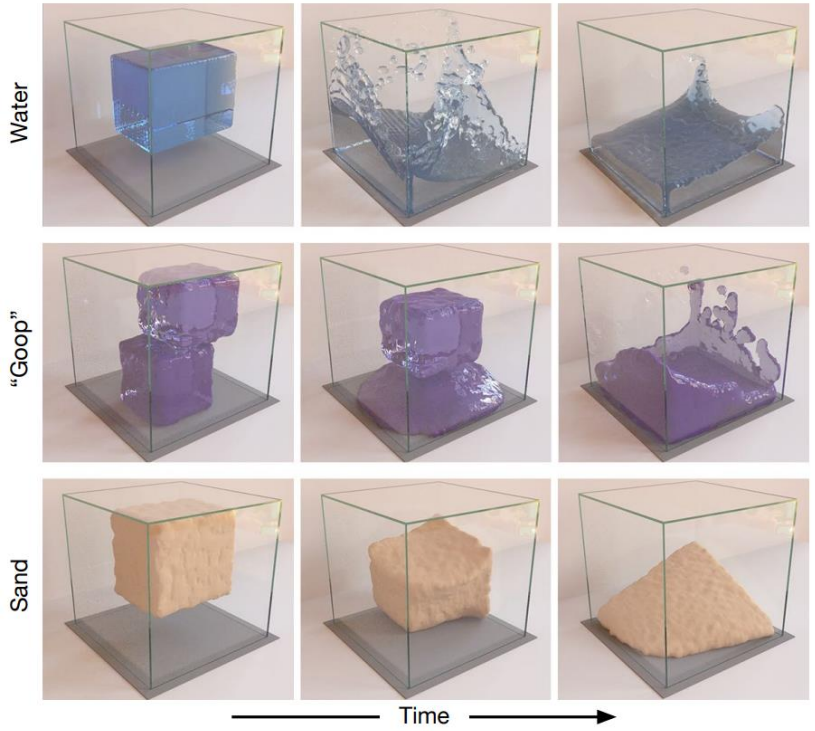
# Generative model

- VAE: JT-VAE(Jin, 2018), Hierarchical VAE(Jin, 2020), generate graph conditionally.
- Auto regression(step-by-step): Hierarchical VAE, Graph-RNN, generate nodes and edges sequentially.



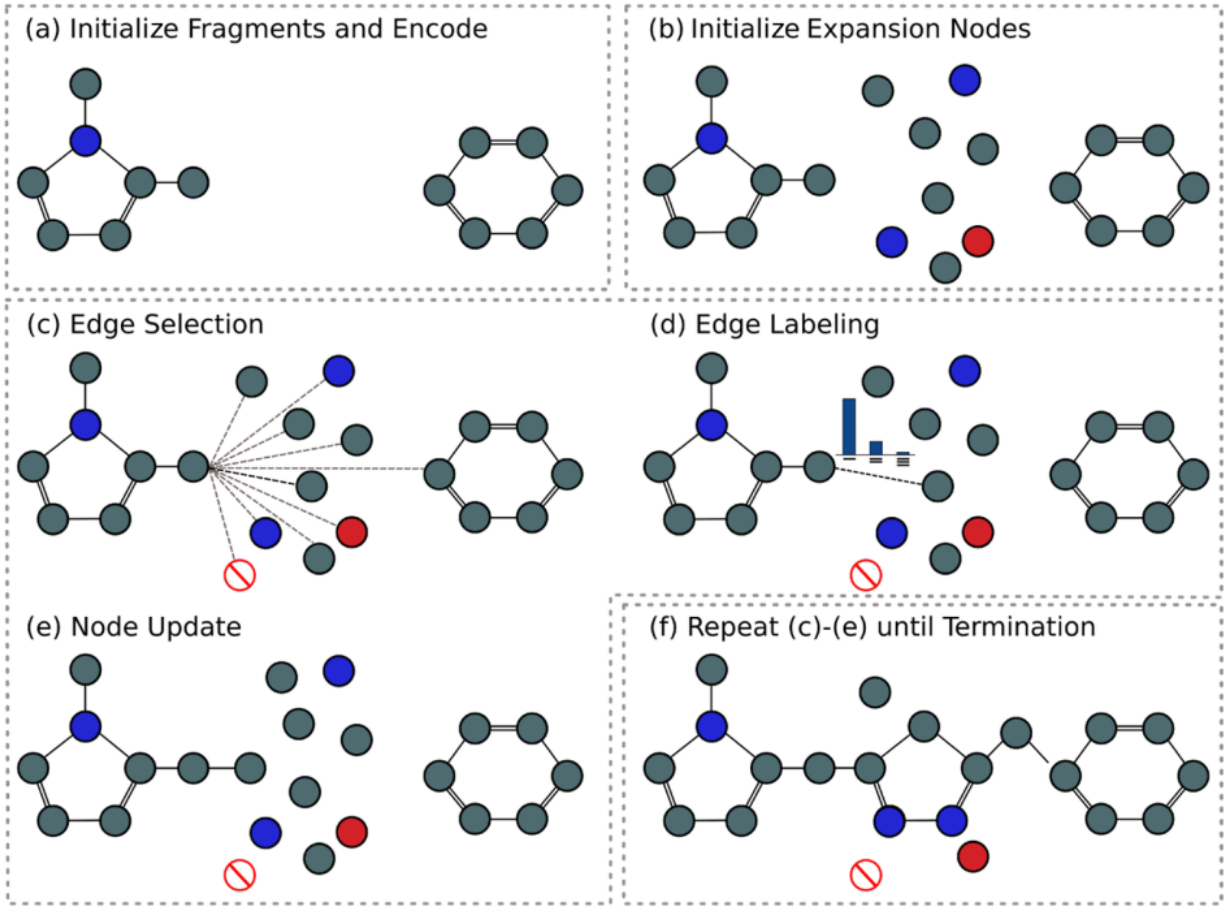
# 3D based model

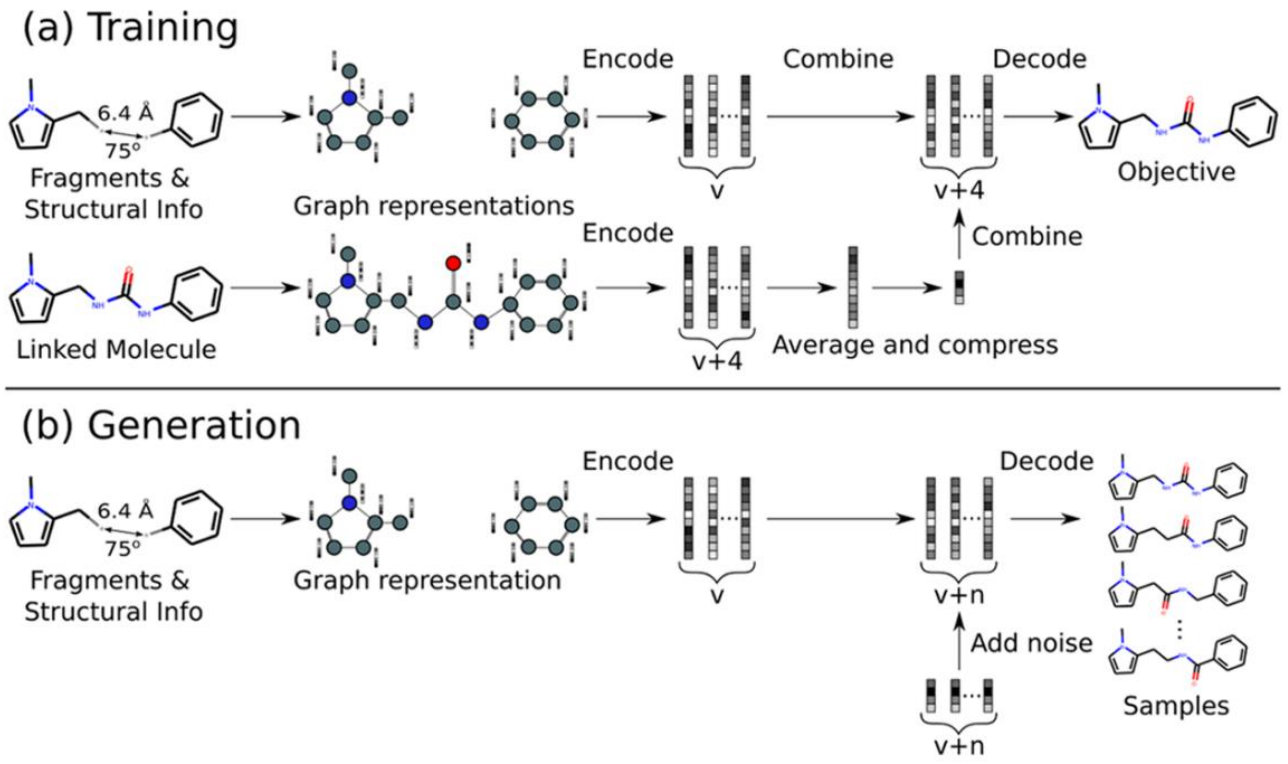
- Point cloud based
- Cannot use conditional term



# DeLinker

- Anchor node must be given, Equivariant not used



# Equivariant Message Passing

- Equivariant: VN-MLP (ICCV 21)

Linear

$$\boldsymbol{q} = W \cdot \boldsymbol{v} \in \mathbb{R}^{n'_v \times 3}, \quad \boldsymbol{k} = U \cdot \boldsymbol{v} \in \mathbb{R}^{n'_v \times 3},$$

Linear

ReLU Like function

$$\boldsymbol{v}' = \boldsymbol{q} - \mathrm{diag}\left\{ \mathbb{1}_{\langle \boldsymbol{q}, \boldsymbol{k} \rangle < 0} \odot \langle \boldsymbol{q}, \frac{\boldsymbol{k}}{\|\boldsymbol{k}\|} \rangle \right\} \cdot \frac{\boldsymbol{k}}{\|\boldsymbol{k}\|},$$

- Invariant: Mixed Feature-MP

# Invariant Message Passing Layer

- 1) Embedding

Equivariant

Invariant

$$h'_j = \phi_1(h_j, \|\text{VN-MLP}_1(\boldsymbol{v_j})\|) \in \mathbb{R}^{n_h},$$
$$h''_j = \phi_2(h_j, \|\text{VN-MLP}_2(\boldsymbol{v_j})\|) \in \mathbb{R}^{n_v},$$
$$\boldsymbol{v}'_j = \text{diag}\{\phi_3(h_j)\} \cdot \text{VN-MLP}_3(\boldsymbol{v}_j) \in \mathbb{R}^{n_v \times 3}.$$

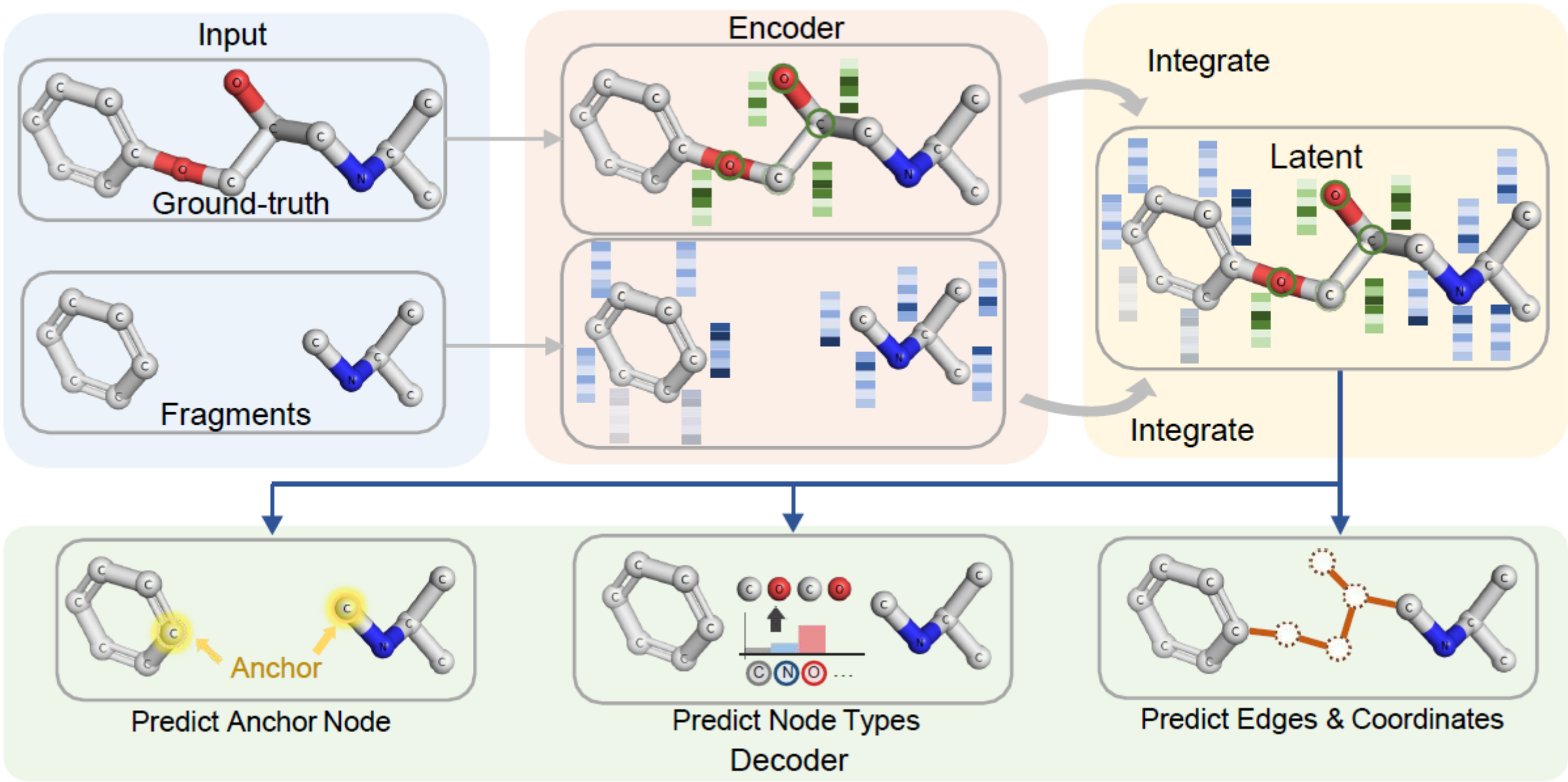
- 2) Message

Trained distance Kernel

$$m^h_{i \leftarrow j} = \text{Ker}_1(\|\boldsymbol{r}_{i,j}\|) \odot h'_j,$$
$$\boldsymbol{m}^{\boldsymbol{v}}_{i \leftarrow j} = \text{diag}\{\text{Ker}_2(\|\boldsymbol{r}_{i,j}\|)\} \cdot \boldsymbol{v}'_j + \left(\text{Ker}_3(\|\boldsymbol{r}_{i,j}\|) \odot h''_j\right) \cdot \boldsymbol{r}^{\top}_{i,j},$$

- 3) Update

$$\tilde{h}_i = \text{GRU}(h_i, \sum_{j \in N(i)} m^h_{i \leftarrow j}),$$
$$\tilde{\boldsymbol{v}}_i = \text{VN-MLP}_4(\boldsymbol{v}_i, \sum_{j \in N(i)} \boldsymbol{m}^{\boldsymbol{v}}_{i \leftarrow j}).$$


Input
Ground-truth
Fragments
Encoder
Integrate
Latent
Integrate
Anchor
Predict Anchor Node
Predict Node Types
Predict Edges & Coordinates
Decoder

# Encoder

$$q_\psi(z^h, \boldsymbol{z}^{\boldsymbol{v}} | G_{\mathrm{F}}, G, \boldsymbol{R}_{\mathrm{F}}, \boldsymbol{R})$$

$$\forall i \in \mathcal{V}_{\mathrm{L}},$$

$$\mu_i^h = \phi_4(\tilde{h}_i), \quad (\sigma_i^h)^2 = \phi_5(\tilde{h}_i),$$

$$\boldsymbol{\mu}_i^{\boldsymbol{v}} = \text{VN-MLP}_5(\tilde{\boldsymbol{v}}_i), \quad (\sigma_i^{\boldsymbol{v}})^2 = \phi_6(\tilde{h}_i)$$

$$\forall i \in \mathcal{V}_{\mathrm{F}},$$

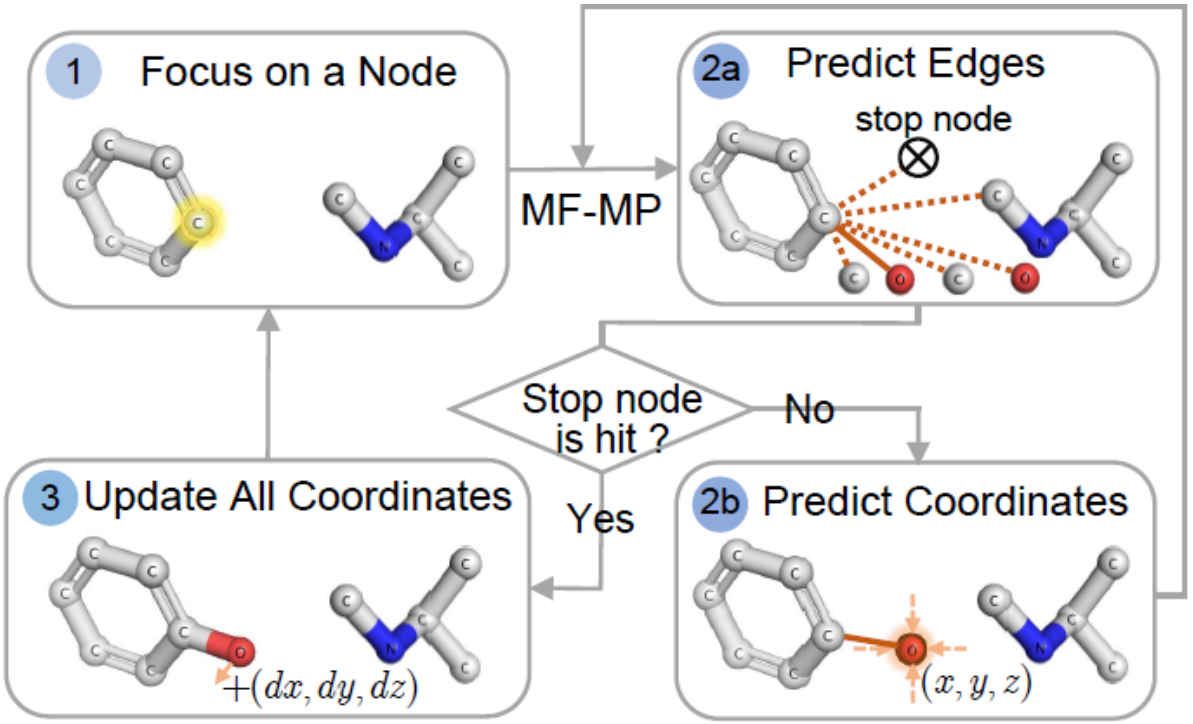
$$z_i^h = \phi_7(\hat{h}_i), \quad \boldsymbol{z}_i^{\boldsymbol{v}} = \text{VN-MLP}_6(\hat{\boldsymbol{v}}_i)$$

# Decoder

$$p_\theta(G, \boldsymbol{R}|G_\mathrm{F}, \boldsymbol{R}_\mathrm{F}, z^h, \boldsymbol{z^v})$$

$$p_\theta(G, \boldsymbol{R}|G_\mathrm{F}, \boldsymbol{R}_\mathrm{F}, z^h, \boldsymbol{z^v}) = p_\theta(\mathcal{E}, X, \boldsymbol{R}|\mathcal{E}_\mathrm{F}, X_\mathrm{F}, \boldsymbol{R}_\mathrm{F}, z^h, \boldsymbol{z^v})$$

$$= \underbrace{p_\theta(a_1, a_2|z^h, \boldsymbol{z^v})}_{\text{Anchor}} \cdot \underbrace{p_\theta(X|z^h)}_{\text{Node Types}} \cdot \underbrace{\prod_{t=0}^{T-1} p_\theta(\mathcal{E}_{t+1}, \boldsymbol{R}_{t+1}|\mathcal{E}_t, \boldsymbol{R}_t, X, a_1, a_2, z^h, \boldsymbol{z^v})}_{\text{Edges and Coordinates}},$$




Masking out impossible edge, anchor node????

MAI Lab EDICAL

# Results

Table 1: Performance metrics for generated molecules.

| Metrics | Valid (%) | Recovered (%) | Pass 2D filters (%) | RMSD | Unique (%) | Novel (%) |
|---|---|---|---|---|---|---|
| 3DLinker (given anchor) | **99.20** | **94.69** | **90.35** | **0.079** | 29.24 | 32.21 |
| 3DLinker | 98.67 | 93.58 | **90.37** | **0.079** | 29.42 | 32.48 |
| DeLinker+ConfVAE | 98.38 | 81.56 | 89.92 | 1.356 | 44.67 | 39.51 |
| GraphAF+ConfVAE | 34.24 | 20.39 | 82.01 | 1.239 | 84.11 | **78.34** |
| GraphVAE+ConfVAE | 15.07 | 0.56 | 85.88 | 1.056 | **85.52** | 61.48 |

# Results

- Compute color similarity scores between two 3D molecules based on the overlap of their pharmacophoric features
- Shape similarity score is a simple volumetric comparison between the two 3D molecules.

Table 2: $SC_{RDKit}$ score distribution (%) and averaged score.

| Metrics | $SC_{RDKit}$ Fragments | | | |
|---|---|---|---|---|
| | $> 0.7$ | $> 0.8$ | $> 0.9$ | Average |
| 3DLinker (given anchor) | **43.10** | **16.09** | **2.60** | **0.684** |
| 3DLinker | 42.55 | 15.85 | 2.49 | 0.683 |
| DeLinker+ConfVAE | 39.96 | 13.39 | 1.93 | 0.675 |
| GraphAF+ConfVAE | 19.33 | 3.36 | 0.32 | 0.624 |
| GraphVAE+ConfVAE | 13.17 | 2.15 | 0.00 | 0.601 |

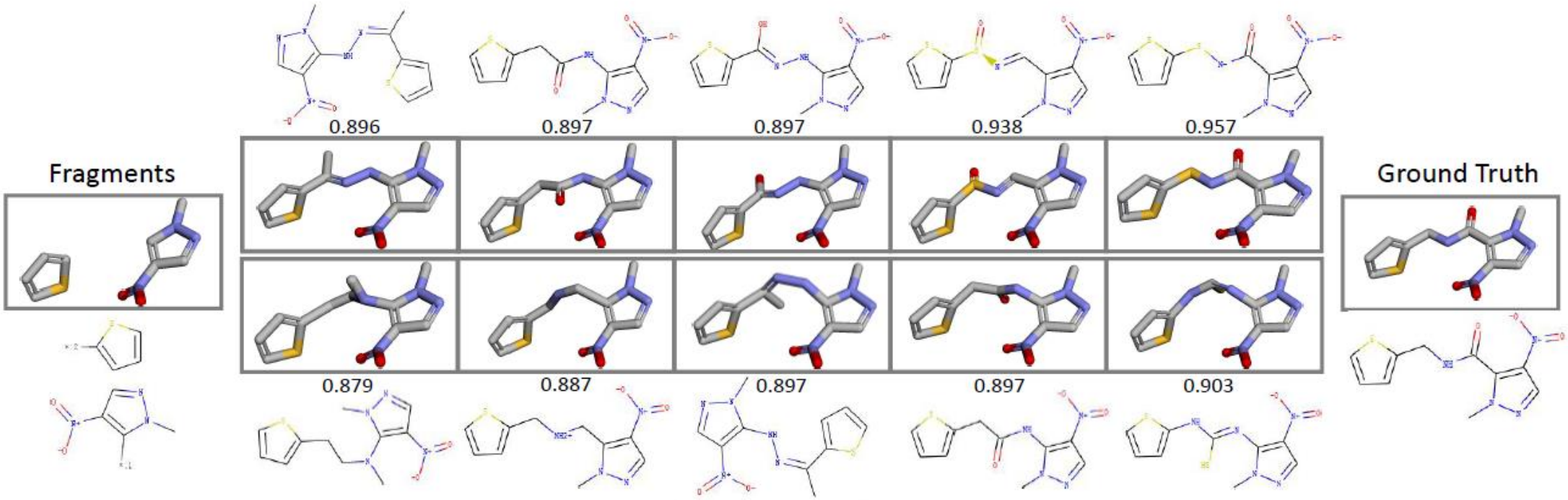



Figure 4: An example of fragment linking. The top-5 similar by $SC_{RDKit}$ Fragments proposed by 3DLinker (first row) and DeLinker+ConfVAE (second row) are shown. Generations from 3DLinker are more realistic and similar to ground truth in terms of $SC_{RDKit}$ and 3D geometry.

# Dataset

- ZINK dataset
- 20 times of MMFF force field optimization using RDKit
- The (fragments, linker) pairs are produced by enumerating all double cuts of acyclic single bonds that are not within any functional groups.
- 365,749 (fragments, linker, coordinates) triplets, training (365,039), validation (351) and test (358).
- 250 samples are made from 358 test set

# Ablation study

- No Equivariant feature & Update all coordinate not fixed one by one

Table 4: Ablation Study. (eqv-) stands for removing equivariant features while (update-) means removing coordinates update strategy.

| Metrics | Valid (%) | Recovered (%) | Pass 2D filters (%) | RMSD | Unique (%) | novel (%) |
|---|---|---|---|---|---|---|
| 3DLinker | 98.67 | 93.58 | 90.37 | 0.079 | 29.42 | 32.48 |
| 3DLinker (eqv-) | 99.42 | 86.59 | 92.68 | 1.352 | 34.58 | 27.02 |
| 3DLinker (update-) | 98.85 | 39.94 | 62.81 | 0.399 | 55.93 | 72.25 |

Table 5: Ablation study. (eqv-) stands for removing equivariant features while (update-) means removing coordinates update strategy.

| Metrics | $SC_{RDKit}$ Fragments | | | |
|---|---|---|---|---|
| | $> 0.7$ (%) | $> 0.8$ (%) | $> 0.9$ (%) | Average |
| 3DLinker | 42.55 | 15.85 | 2.49 | 0.683 |
| 3DLinker (eqv-) | 38.51 | 13.15 | 1.76 | 0.672 |
| 3DLinker (update-) | 37.34 | 10.87 | 1.13 | 0.670 |

# Limitation

- Conditional VAE라고 주장하지만 우리가 생각하는 Condition은 Valency 같은 것 아닐까? 정말 VAE가 맞는가?

# Thank you