

EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction

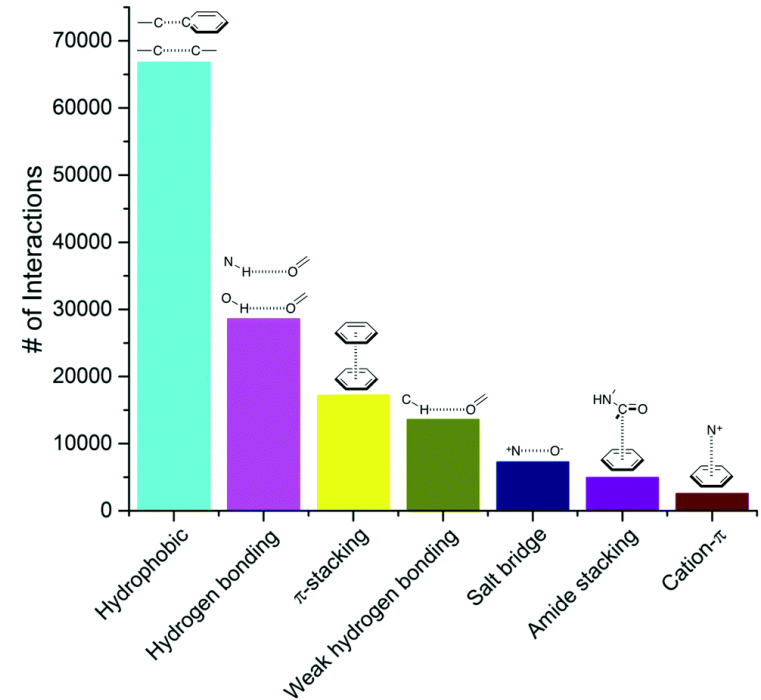
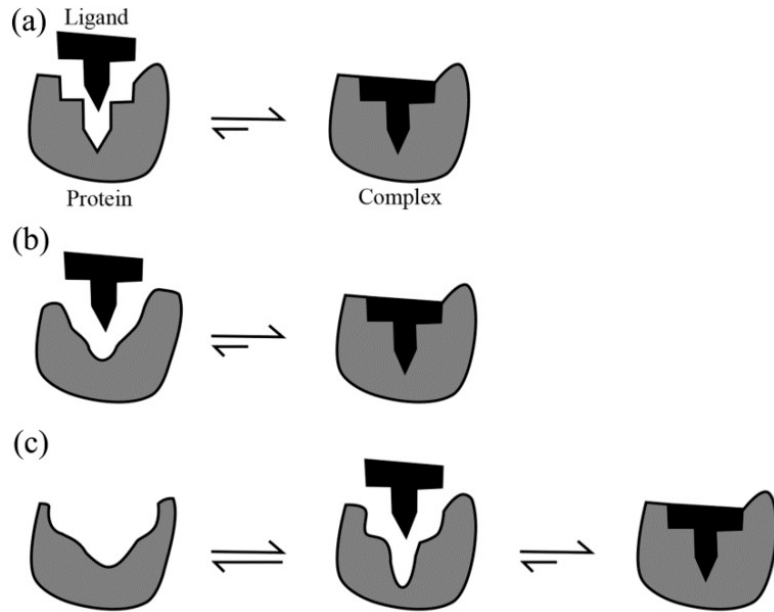
A Deep Learning Approach to Predict Protein-Ligand Binding Structures

Fabian Molls

Warum ist die Vorhersage von Protein-Ligand-Bindungen wichtig?

- Klinische Relevanz
- Ist ein zentrales Element in der Arzneimittelentwicklung
 - Je schneller desto günstiger

Herausforderung: Komplexität der Molekülinteraktionen



Traditionelle Lösungen

- Molekulares Docking
 - Berechnungen auf Basis der räumlichen Struktur, ungenau
- Molekulare Dynamik
 - Berechnung auf Basis von Inter und Intra-molekularen Kräften
 - Sehr rechenaufwändig aufgrund von Sampling und Scoring

Beispiele: GLIDE, GNINA

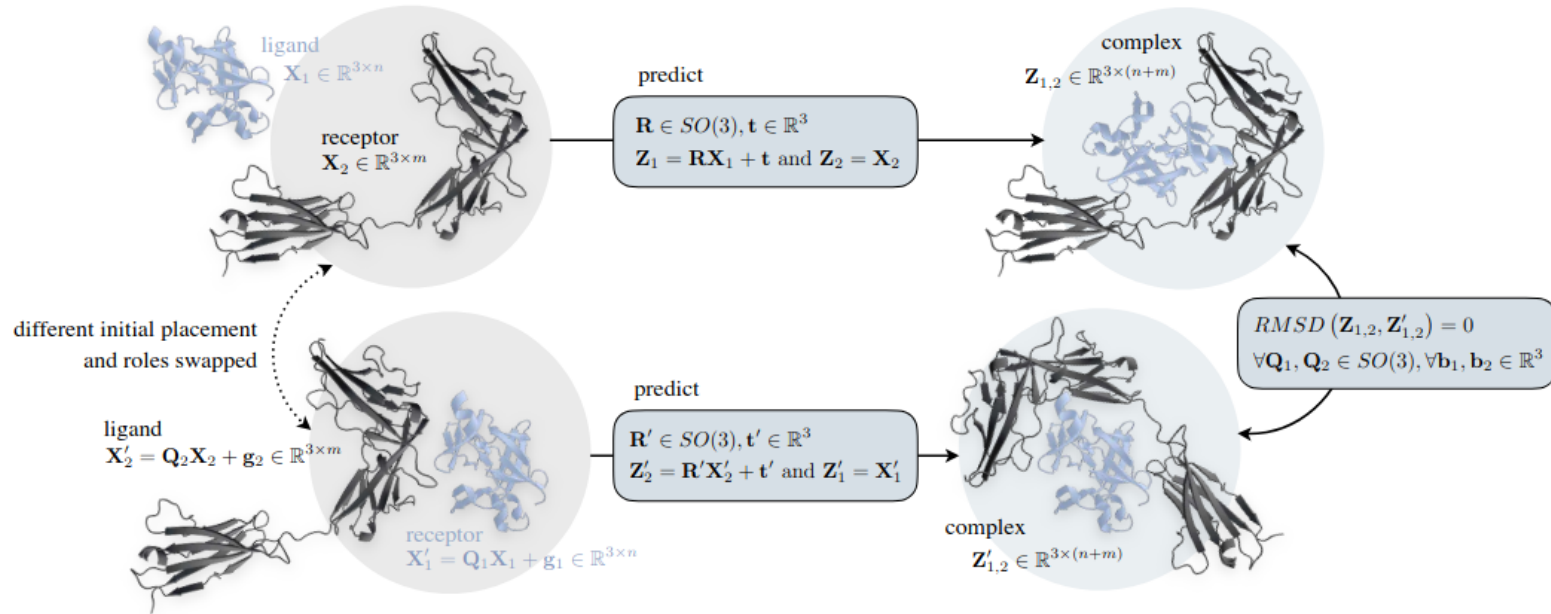
Neue Idee: geometrisches Deep Learning

- Ein NN auf die Geometrie und Symmetrie des Liganden und Proteins trainieren
- Features zur Abbildung von molekularen Kräften

Graph Neural Networks (GNNs)

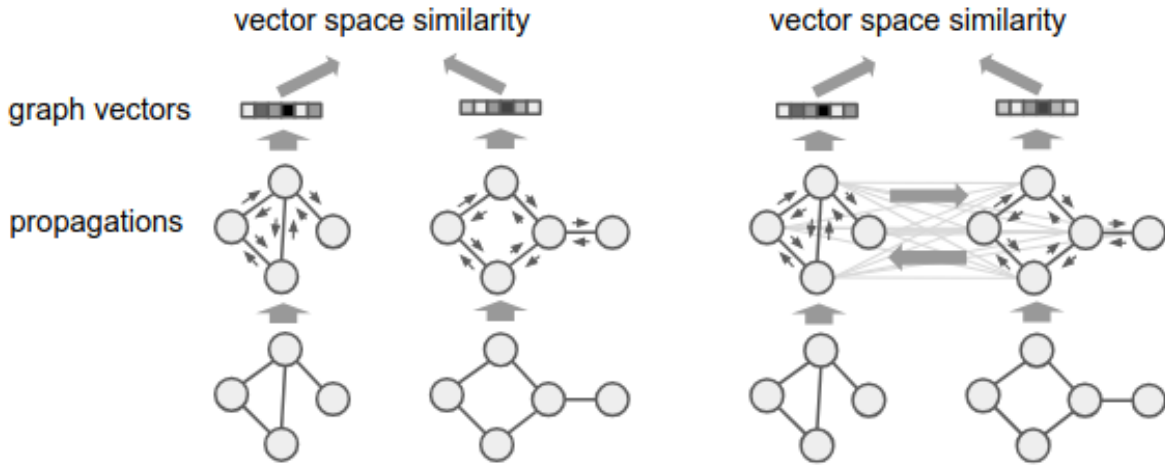
- (K-NN) k-nearest neighbor Graph
- $G = (V, E)$
 - $V = \text{Knoten} \ \& \ E = \text{Edges}$
 - Mit X als 3D Koordinaten und F für Features
- Vorteil: Flexibel & Skalierbar

SE(3)-Equivariance



$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i^1 - \mathbf{r}_i^2)^2}$$

Graph Matching & Message Passing



$$\mathbf{H}^{(l+1)} = \sigma(\hat{\mathbf{A}}\mathbf{H}^{(l)}W^{(l)})$$

Where:

- $\mathbf{H}^{(l)}$ is the matrix of node features at layer l .
- $\hat{\mathbf{A}}$ is the normalized adjacency matrix of the graph, sometimes with added self-loops (i.e., the node can also communicate with itself).
- $W^{(l)}$ is a learnable weight matrix at layer l .
- σ is a non-linear activation function (e.g., ReLU).
- The operation $\hat{\mathbf{A}}\mathbf{H}^{(l)}$ aggregates information from neighboring nodes.

Figure 2. Illustration of the graph embedding (left) and matching models (right).

Attention-Mechanismus

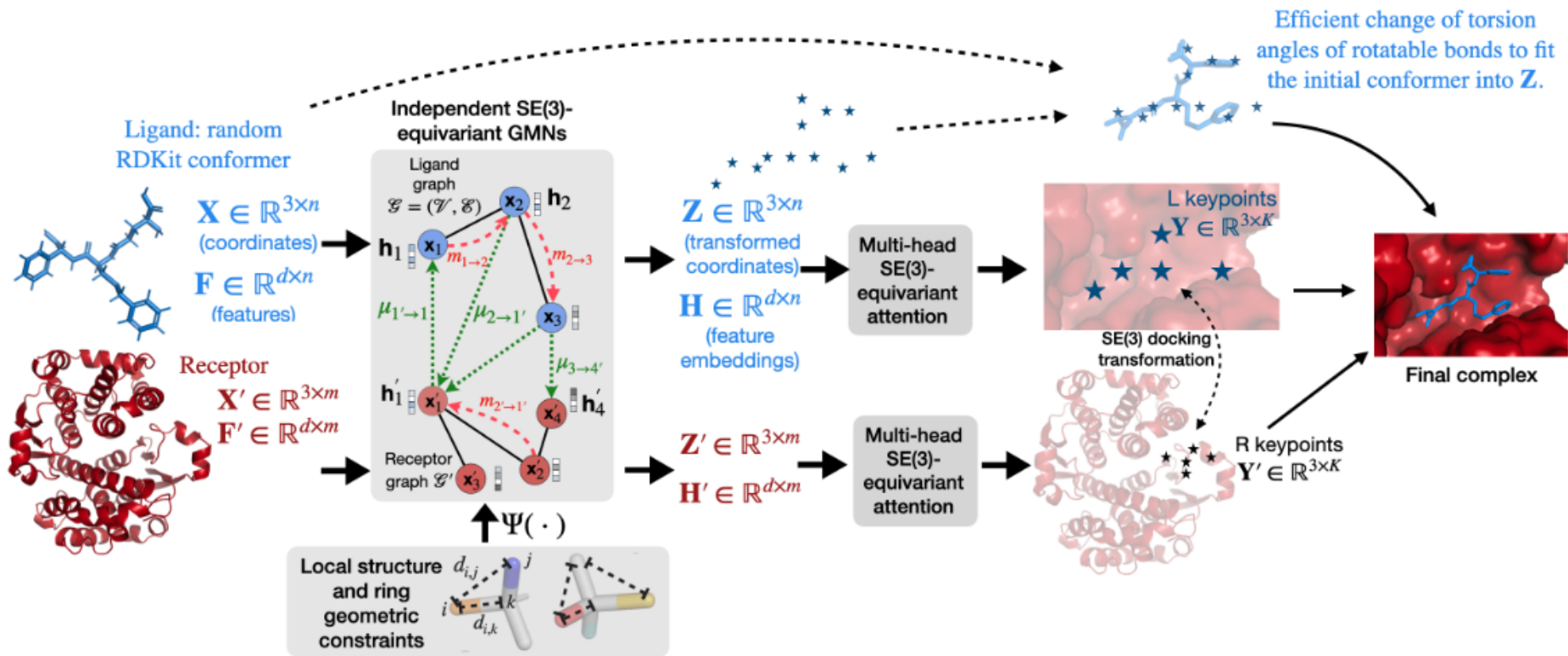
- Attention ermöglicht es dem Modell, sich auf **relevante Teile** der Eingabedaten zu konzentrieren und andere Teile zu ignorieren.

$$a_{ij} = \text{Attention}(f(\mathbf{r}_i, \mathbf{r}_j, \mathbf{h}_i, \mathbf{h}_j))$$

Where:

- \mathbf{r}_i and \mathbf{r}_j are the 3D positions of points i and j .
- \mathbf{h}_i and \mathbf{h}_j are the feature vectors of nodes i and j .
- $f(\mathbf{r}_i, \mathbf{r}_j, \mathbf{h}_i, \mathbf{h}_j)$ is a function that computes the relationship between these nodes in a way that is invariant to rotation and translation (e.g., by using only the relative distance $\mathbf{r}_i - \mathbf{r}_j$).

Die Programmarchitektur



Evaluation

METHODS	AVG. SEC.	AVG. SEC.	LIGAND RMSD ↓						CENTROID DISTANCE ↓						KABSCH RMSD ↓	
			PERCENTILES ↓				% BELOW THRESHOLD ↑		PERCENTILES ↓				% BELOW THRESH. ↑		MEAN	MED.
			MEAN	25TH	50TH	75TH	5 Å	2 Å	MEAN	25TH	50TH	75TH	5 Å	2 Å		
QVINA-W	49	-	13.6	2.5	7.7	23.7	40.2	20.9	11.9	0.9	3.7	22.9	54.6	41.0	2.1	1.9
GNINA	247	146	13.3	2.8	8.7	22.1	37.1	21.2	11.5	1.0	4.5	21.2	52.0	36.0	2.2	1.8
SMINA	146	-	12.1	3.8	8.1	17.9	33.9	13.5	9.8	1.3	3.7	16.2	55.9	38.0	2.2	1.9
GLIDE (c.)	1405*	-	16.2	2.6	9.3	28.1	33.6	21.8	14.4	0.8	5.6	26.9	48.7	36.1	2.2	1.9
EQUIBIND	0.16	0.04	8.2	3.8	6.2	10.3	39.1	5.5	5.6	1.3	2.6	7.4	67.5	40.0	2.6	2.3
EQUIBIND+Q	8	8	8.4	2.6	6.6	11.1	38.0	18.7	5.9	1.0	2.5	6.4	68.7	44.6	2.3	1.9
EQUIBIND+Q2	15	15	8.7	2.6	6.8	11.1	40.7	21.6	6.0	1.0	2.4	6.6	70.1	42.7	2.2	1.6
EQUIBIND+S	146	146	8.3	2.1	5.6	10.5	46.4	24.6	6.0	0.9	2.0	6.2	71.0	50.6	2.1	1.8
EQUIBIND-U	0.14	0.02	7.8	3.3	5.7	9.7	42.4	7.2	5.6	1.3	2.6	7.4	67.5	40.0	2.1	1.8

Dissskusion

- Stärken des Ansatzes:

Effizientere und genauere Vorhersagen im Vergleich zu traditionellen Methoden.

Die Fähigkeit des Modells, die Geometrie der Moleküle zu berücksichtigen, führt zu besseren Ergebnissen bei komplexen Bindungsszenarien.

- Limitationen und Herausforderungen:

Wie bei allen Deep-Learning-Ansätzen könnte die Qualität der Vorhersagen von der Größe und Qualität der Trainingsdaten abhängen.

Die Notwendigkeit leistungsfähiger Rechenressourcen, insbesondere für große Datensätze.

Literatur

- EQUIBIND: Geometric Deep Learning for Drug Binding Structure Prediction <https://arxiv.org/pdf/2202.05146>
- INDEPENDENT SE(3)-EQUIVARIANT MODELS FOR END-TO-END RIGID PROTEIN DOCKING <https://arxiv.org/pdf/2111.07786>
- E(n) Equivariant Graph Neural Networks <https://arxiv.org/pdf/2102.09844>
- GEOMETRIC AND PHYSICAL QUANTITIES IMPROVE
- E(3) EQUIVARIANT MESSAGE PASSING <https://arxiv.org/pdf/2110.02905>