**תיאור המשימה:**

* צריך לבחון אם ניתן להשתמש בgeom למשימת pre-training.
* קודם אנחנו רוצים לאמן מודל עם chemprop על הדאטה סט שמכיל המון מולקולות (מולקולות שלא בהכרח קשורות למשימה הסופית שלנו)
* המטרה היא לעשות self-supervised learning , ז"א להכריח את המודל לחזות משהו לגבי המולקולות בדאטה סט שלנו, ובמסגרת האימון, המודל שיצרנו ע"י chemprop בעצם לומד את הייצוד הדו-ממדי ותלת-ממדי של המולקולות.
* אחרי שהרשת נוירונים אומנה בשלב הpre-training , ניתן להשתמש בה ולבצע fine tuning על המשימה שלנו, עם המודולטורים והPPI.
* המטרה היא שכשנתחיל לאמן את הרשת נוירונים, המשקולות ההתחלתיים של המודל לא יהיו משקולות רנדומליים, אלה משקולות אחרי אימון על הדאטה סט של geom.
* לסיכום, לבחון האם ניתן להשתמש ב-geom למשימת pre-training בעזרת מודל מchemprop. ז"א האם עבור כל SMILES יש לנו איזה שהוא וקטור של מספרים (embedding vector) כך שנוכל להשתמש בזה בהמשך.

**רשימת מושגים ועוד:**

**GEOM** - Geometric Ensemble Of Molecules.

**Conformers**

Conformers are different spatial arrangements of atoms within a molecule. They arise due to the rotation of single bonds and other forms of molecular flexibility. Conformers are important because they affect a molecule's properties and behaviour in various environments. Understanding conformers helps in predicting molecular interactions, reactivity, and overall behaviour.

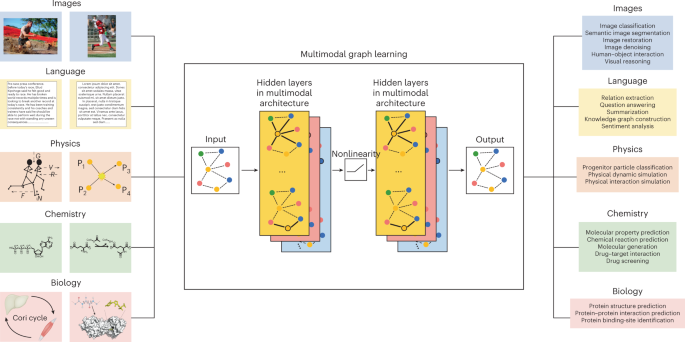
**Conformer Ensembles**

Conformer ensembles refer to collections of molecular structures that represent different conformations or spatial arrangements of the atoms within a molecule. Molecules can exist in various conformations due to rotation around single bonds and other types of flexibility. Conformer ensembles are used in computational chemistry to account for this flexibility when studying molecular properties like energy, geometry, and interactions. They are essential for understanding how molecules behave in different environments or during chemical reactions.

**PPI Modulators**

Modulators are molecules or compounds that have the ability to modify or influence protein-protein interactions (PPIs). These molecules can either enhance or inhibit the interactions between proteins, thereby altering the biological processes in which these interactions are involved. Modulators play a crucial role in drug discovery and therapeutic development, as they can be targeted to intervene in specific disease pathways or cellular processes by modulating protein-protein interactions.

**Multi-Modal Learning**

Multimodal learning, in the context of machine learning, is a type of deep learning using a combination of various modalities of data, such as text, audio, or images, in order to create a more robust model of the real-world phenomena in question.In biology and predicting PPI modulators, multimodal learning involves combining different types of biological data, such as genomics, proteomics, structural, chemical, and pathway information, to improve the accuracy of identifying molecules that can modify protein-protein interactions. It helps create more comprehensive models for drug discovery and understanding biological processes.

**MultiPPIMI Framework (from the paper)**

A screenshot of a computer

Description automatically generated

Overview of the MultiPPIMI framework.

The modulator representation consists of the concatenation of pre-trained structural embeddings and physicochemical properties of compounds. The PPI representation is constructed by concatenating ESM2 embeddings and physicochemical features of two protein chains. The modulator and PPI representations are input into a bilinear attention network, enabling the learning of inter-molecular interactions. The joint representation f is decoded using a fully connected network to predict the probability p of PPI target-modulator interactions.