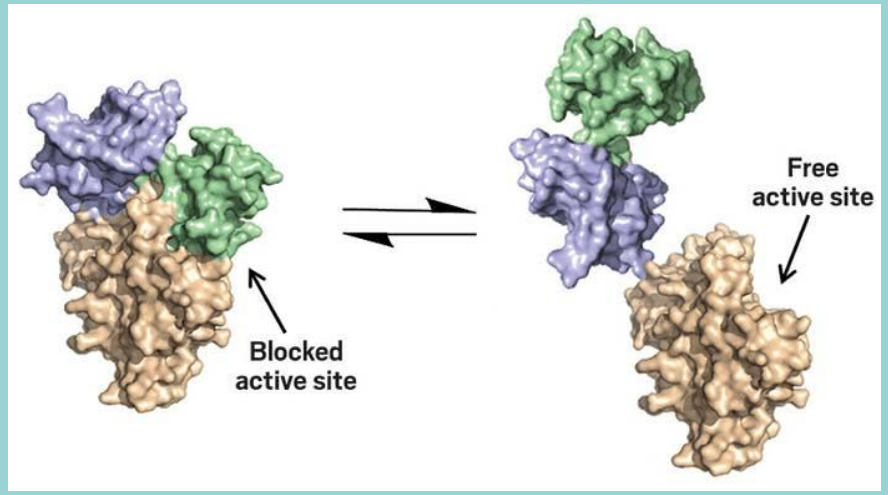
# Protein Binding Site Prediction

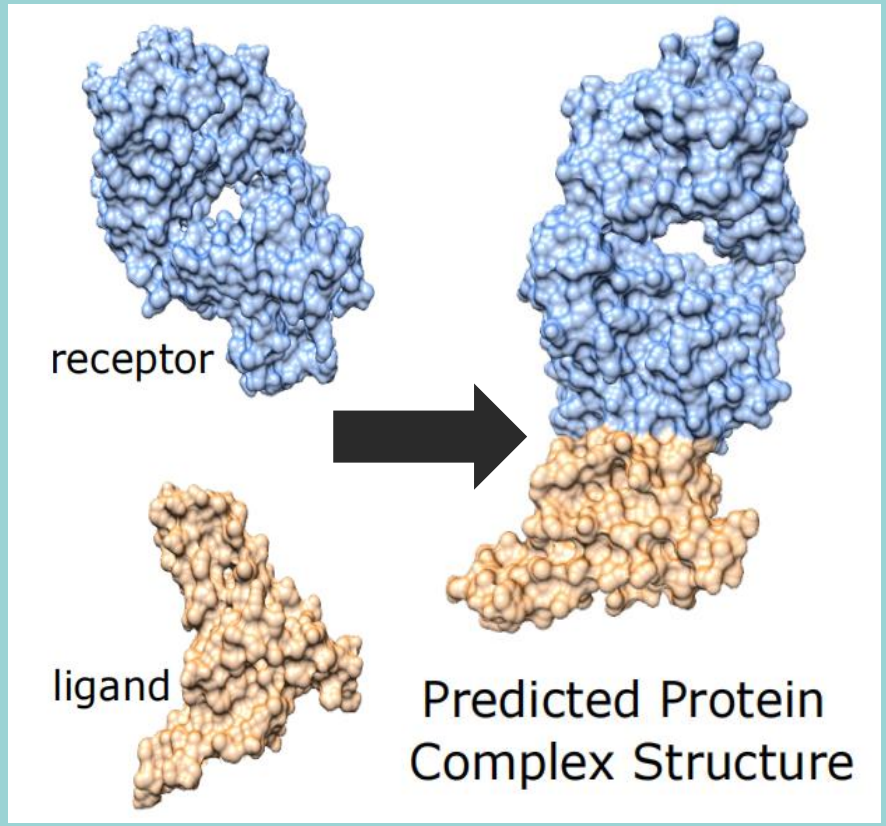
## CS182 final project

周守琛 吴明正 许家睿

# Contents

# Introduction



# Dataset

## PDB(Protein Dataset Bank)

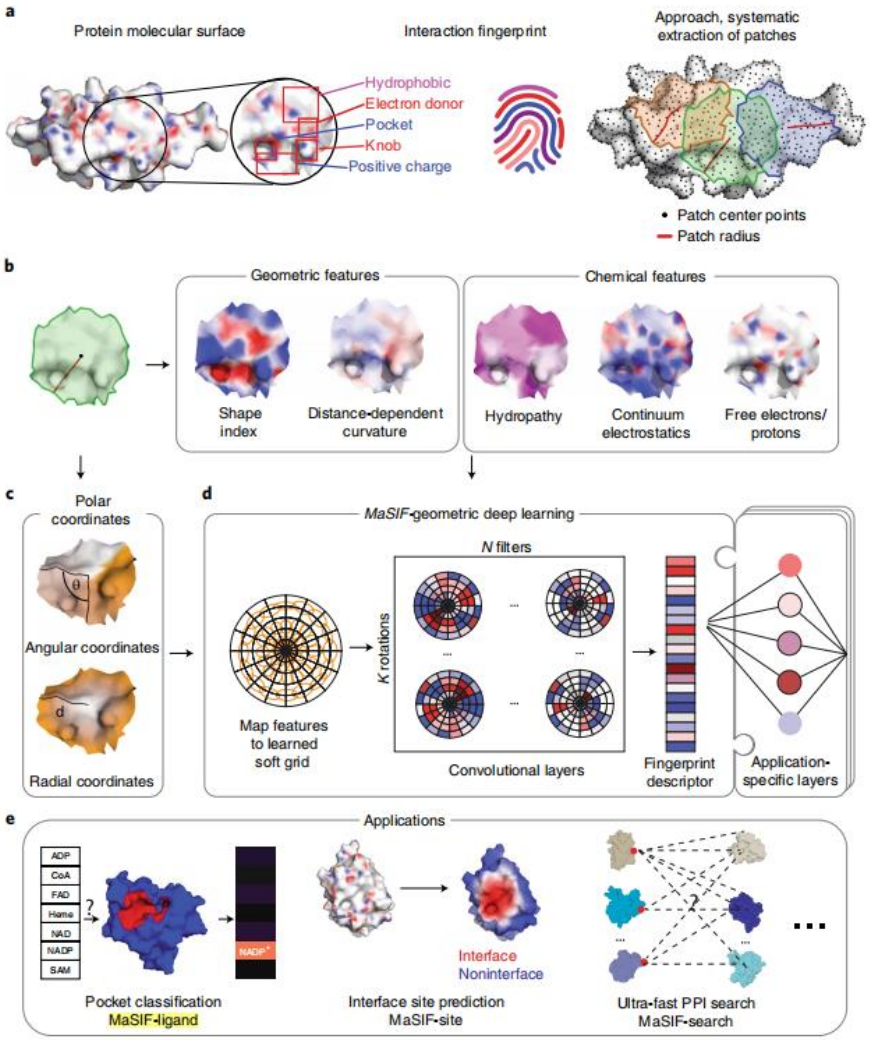
```
HEADER SAMPLE PDB FILE
ATOM 1 N GLY A 1 5.234 8.678 12.456 1.00 20.00 N
ATOM 2 CA GLY A 1 6.345 9.789 13.567 1.00 18.00 C
ATOM 3 C GLY A 1 7.456 10.890 14.678 1.00 16.00 C
ATOM 4 O GLY A 1 8.567 11.901 15.789 1.00 14.00 O
ATOM 5 CB GLY A 1 6.789 9.234 14.901 1.00 16.00 C
ATOM 6 N ALA A 2 7.123 10.567 11.890 1.00 20.00 N
ATOM 7 CA ALA A 2 8.234 11.678 12.567 1.00 18.00 C
ATOM 8 C ALA A 2 9.345 12.789 13.678 1.00 16.00 C
ATOM 9 O ALA A 2 10.567 13.901 14.789 1.00 14.00 O
TER END
```

```
HEADER    PROTEIN                                 20-FEB-22   1XYZ
TITLE     CRYSTAL STRUCTURE OF A SMALL PROTEIN
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: PROTEIN;
COMPND   3 CHAIN: A;
COMPND   4 ENGINEERED: YES;
COMPND   5 MUTATION: YES;
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   3 ORGANISM_COMMON: HUMAN;
SOURCE   4 ORGANISM_TAXID: 9606;
SOURCE   5 GENE: XYZ;
SOURCE   6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE   7 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE   8 EXPRESSION_SYSTEM_STRAIN: BL21(DE3);
SOURCE   9 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE  10 EXPRESSION_SYSTEM_PLASMID: PET28A;
EXPDTA    X-RAY DIFFRACTION
AUTHOR    A.B., C.D.
REVDAT   2   15-MAR-22 1XYZ    1       SEQADV
REVDAT   1   01-MAR-22 1XYZ    0
JRNL        AUTH   A.B., C.D.
JRNL        TITL   TITLE OF THE PAPER
JRNL        REF    J.MOL.BIOL.         V. 123  456 2022
JRNL        PMID   12345678
REMARK   2
REMARK   2 RESOLUTION.    2.0 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : PHENIX
REMARK   3   AUTHORS     : A.A., B.B., C.C., D.D.
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.0
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 30.0
REMARK   3
ATOM      1  N   ALA A   1      10.932   7.365  -8.260  1.00 20.00           N
ATOM      2  CA  ALA A   1       9.661   7.042  -8.967  1.00 20.00           C
ATOM      3  C   ALA A   1       8.716   8.103  -8.460  1.00 20.00           C
ATOM      4  O   ALA A   1       7.520   7.805  -8.329  1.00 20.00           O
ATOM      5  CB  ALA A   1       9.117   5.727  -8.414  1.00 20.00           C
...
HETATM 1001  O   HOH B 101      16.814   2.363   2.967  1.00 20.00           O
HETATM 1002  O   HOH B 102      12.123   7.078  -0.942  1.00 20.00           O
HETATM 1003  O   HOH B 103       9.232   9.875   3.976  1.00 20.00           O
```

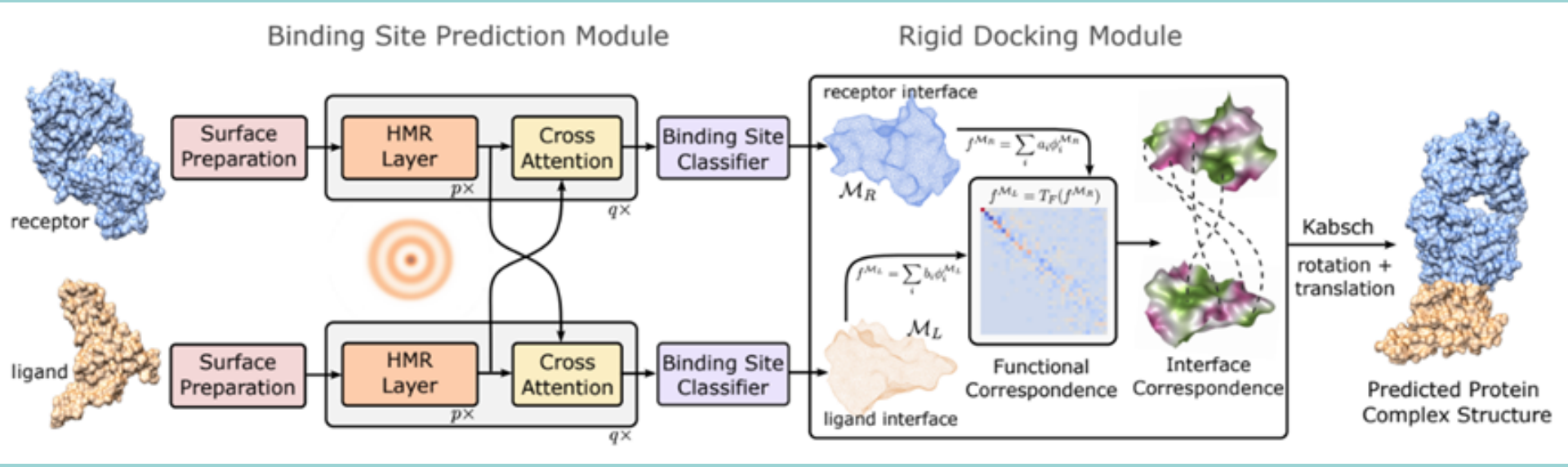
# MaSIF

MoleculAr Surface Interaction Fingerprinting

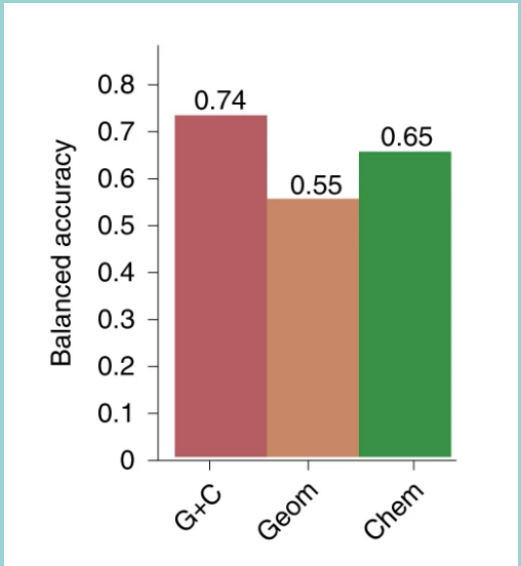
- Geometry feature
- Surface Physical and Chemical Properties

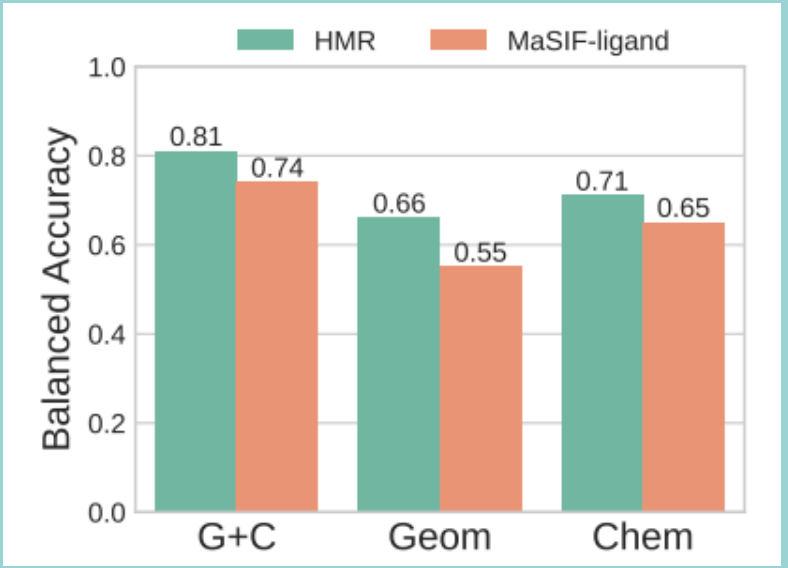


# Ours Pipeline



# Comparison




*https://www.nature.com/articles/s41592-019-0666-6*


# Thank You