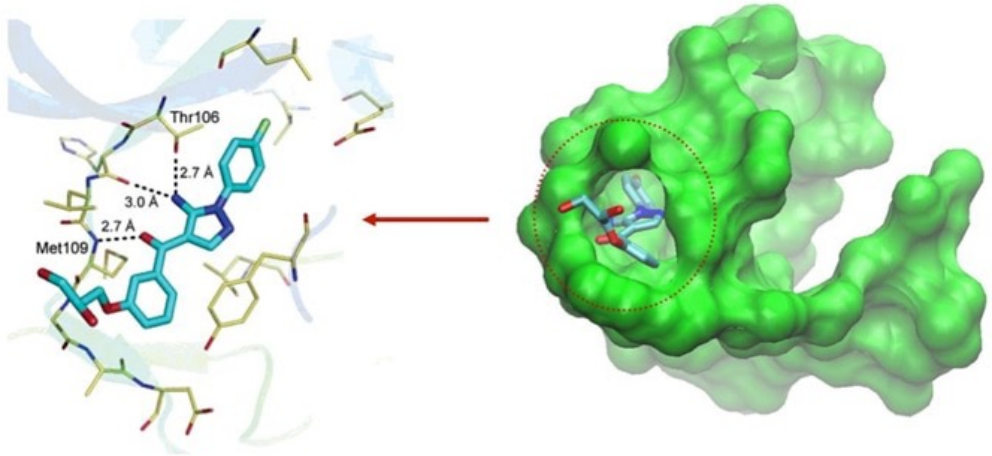
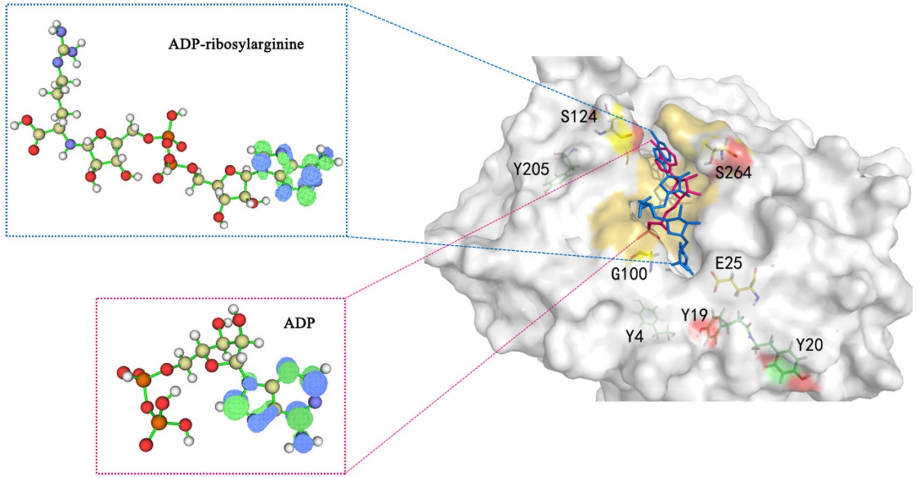
Ligand binding site prediction

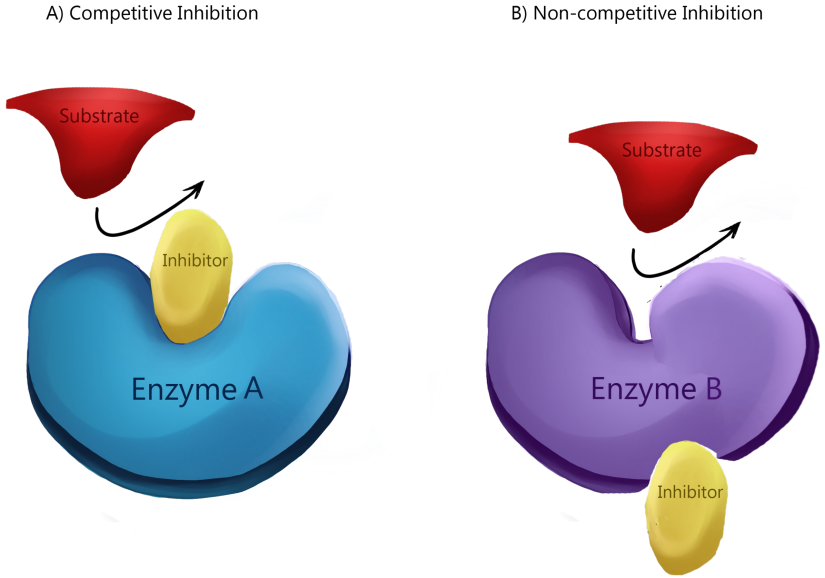
College of Pharmacy, Seoul National University

Juyong Lee

Ligand binding sites are critical in regulating protein functions

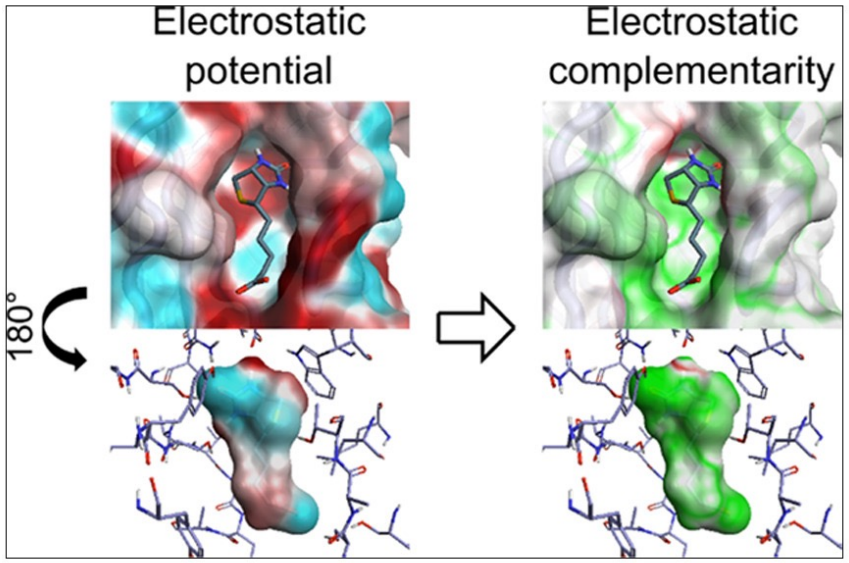


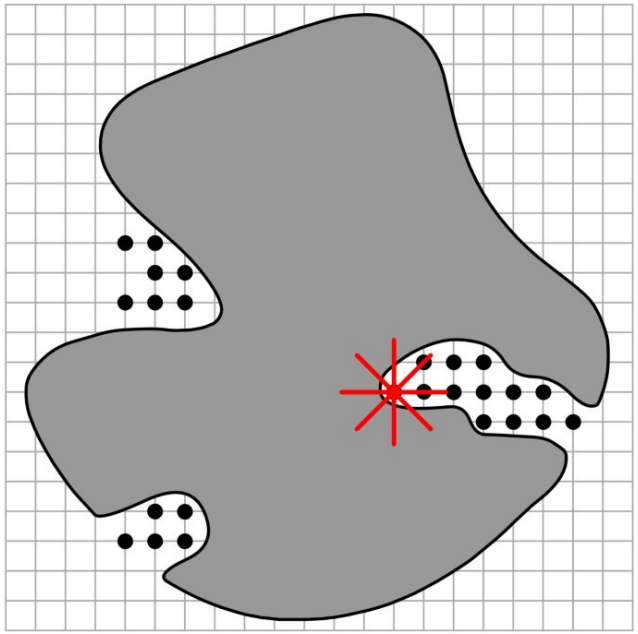


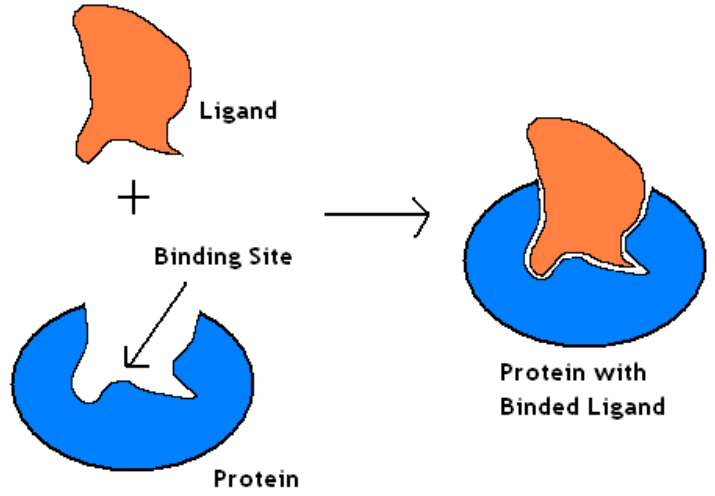


- Identifying exact ligand binding site (pocket) is critical for drug discovery
- Two types of mechanisms
  - Competitive / Non-competitive inhibition

Ligand binding pockets





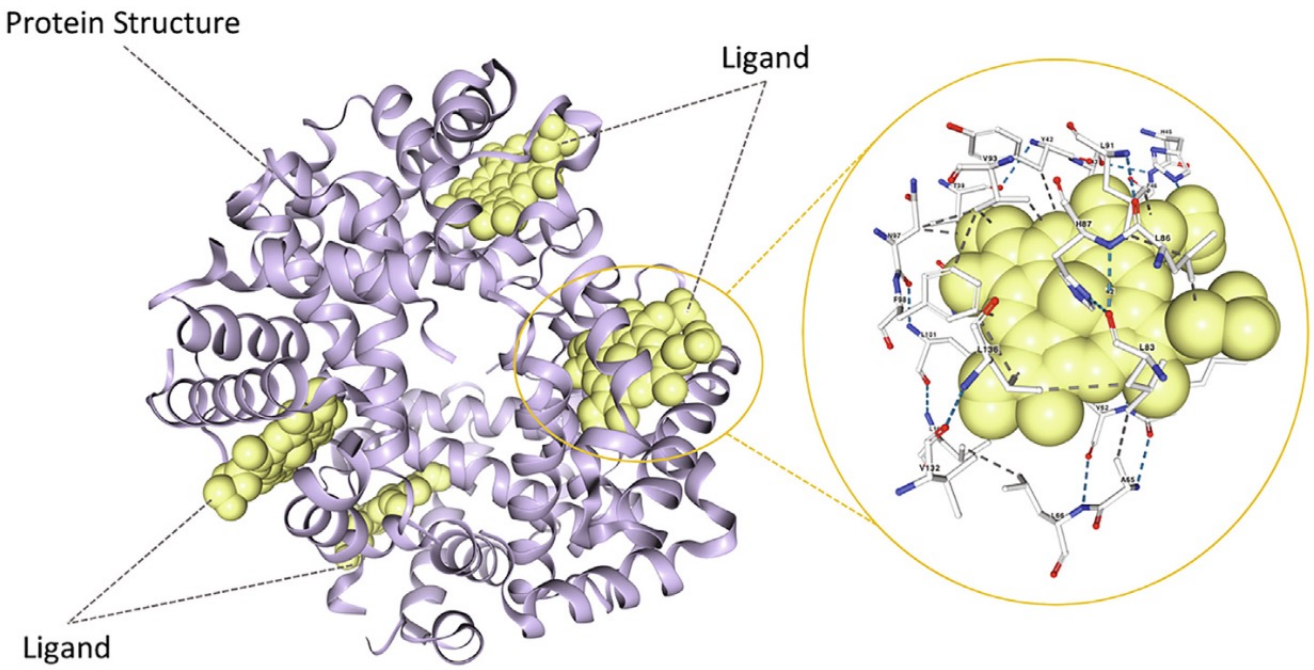


Shape complementarity

- Ligand binding sites (LBS) tend to form **concave** surfaces
- Shape complementarity is important
- Electrostatic & van der Waals interactions are important
  - H-bond, pi-pi stacking, pi-cation interaction

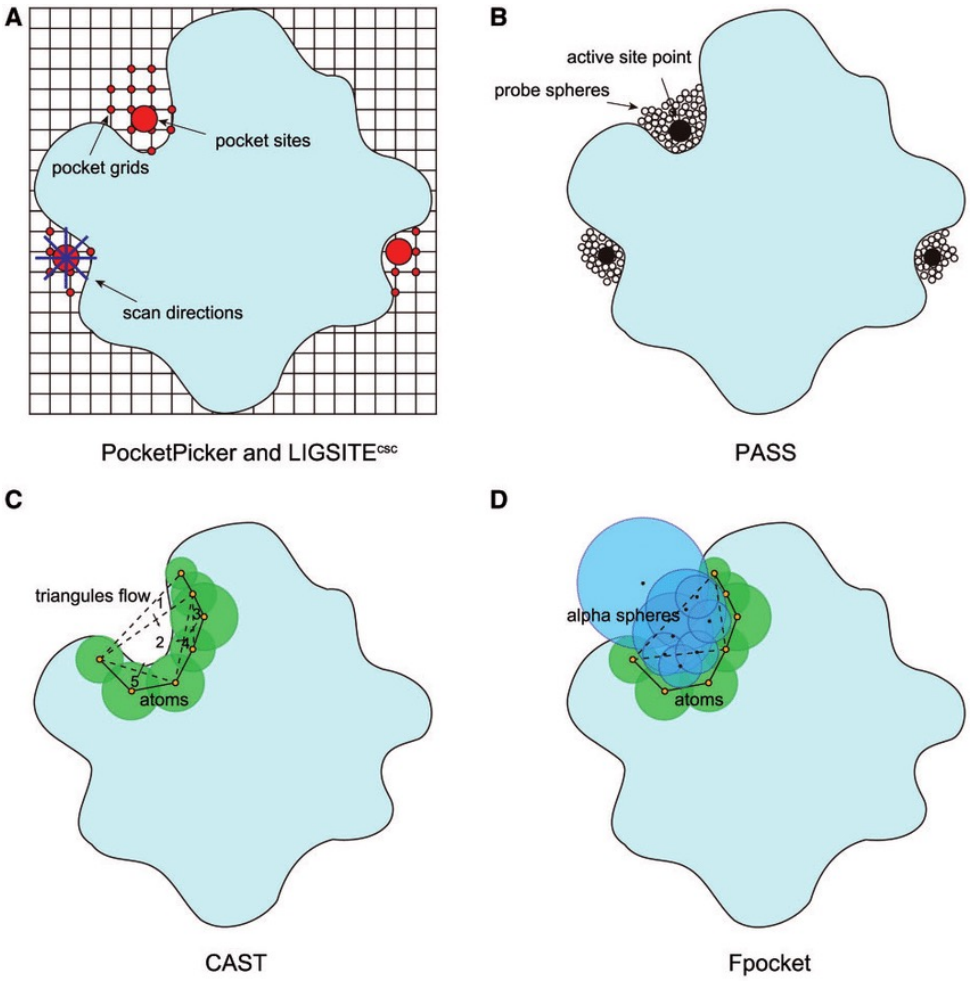
Ligand binding site prediction algorithms

- Structure-based algorithms
  - Geometry/Energy-based
  - Machine-learning-based
- Similarity-based algorithms



Conventional binding site detection

- Geometry-based methods find **concave** sites by
  - Rolling a probe
  - Using 3D grid
  - Using triangles



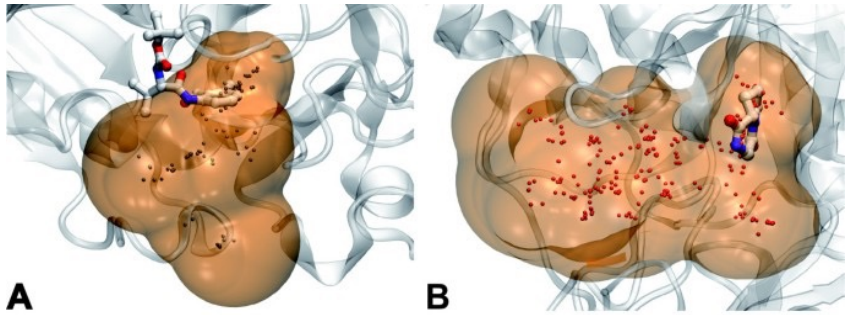
Alpha shape and Delaunay triangulation in studies of protein-related interactions, *Briefings in Bioinformatics,* 2012

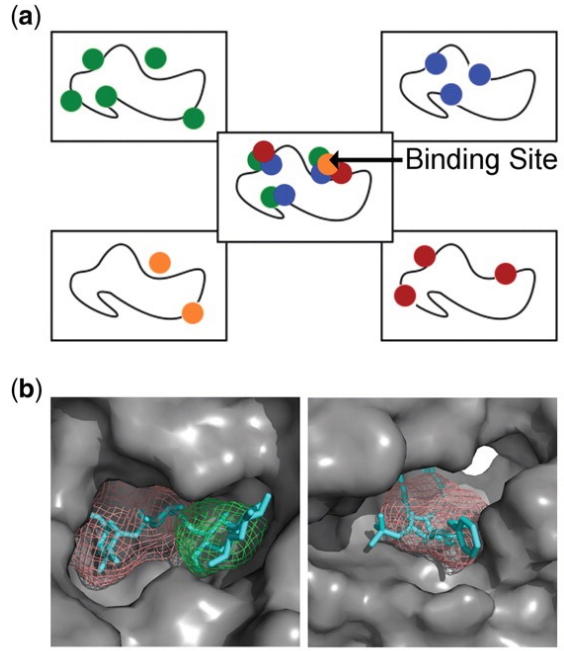
Traditional structure-based method

Table 1
Published 3D structure-based LBS prediction methods.

Method	Type	Feature	Year
A computational procedure (with no specific name) [39]	Probe Energy-based	Contour surfaces at appropriate energy levels are calculated for each probe and displayed with the protein structure	1985
POCKET [27]	Spatial Geometry Measurement	Place spheres between atoms and surfaces of pockets are modeled using marching cubes algorithm	1992
SURFNET [40]	Spatial Geometry Measurement	Place spheres at the gap between any two protein atoms	1995
LIGSITE [26]	Spatial Geometry Measurement	Set up some regular 3D meshes to cover the target protein	1997
CAST [41]	Spatial Geometry Measurement	Calculate by using alpha shape and discrete flow theory	1998
CASTp [42,43]	Spatial Geometry Measurement	Use alpha shape and the pocket algorithm [44] developed in computational geometry	2003
QSiteFinder [45]	Probe Energy-based	Use the interaction energy between the protein and a simple van der Waals probe	2005
LIGSITE^{CSC} [46]	Spatial Geometry Measurement	An extension and implementation of the LIGSITE algorithm by using the Connolly surface	2006
VISCANA [47]	Probe Energy-based	A total energy of the molecule is evaluated by summation of fragment energies and interfragment interaction energies	2006
Fpocket [48]	Spatial Geometry Measurement	Voronoi tessellation and alpha spheres are used to detect pockets	2009
SITEHOUND [28,49]	Probe Energy-based	The carbon probe and phosphate probe used to detect interaction force between the probe and the protein	2009
MSPocket [50]	Spatial Geometry Measurement	Identify surface pocket regions according to the normal vector directions at the vertices on the surface	2010
FTSite [51]	Probe Energy-based	Use 16 different probes on these grids to detect free energy	2011
SiteComp [52]	Probe Energy-based	Discovery of subsites with different interaction properties and for fast calculations of residue contribution to binding sites	2012
LISE [53]	Spatial Geometry Measurement	Compute a score by counting geometric motifs extracted from substructures of interaction networks connecting protein and ligand atoms	2013
Patch-Surfer2. 0 [54]	Spatial Geometry Measurement	Represent and compare pockets at the level of small local surface patches that characterize physicochemical properties of the local regions	2014
CurPocket [55]	Spatial Geometry Measurement	Compute the curvature distribution of protein surface and identify the clusters of concave regions	2019

Computational and Structural Biotechnology Journal 18 (2020) 417–426



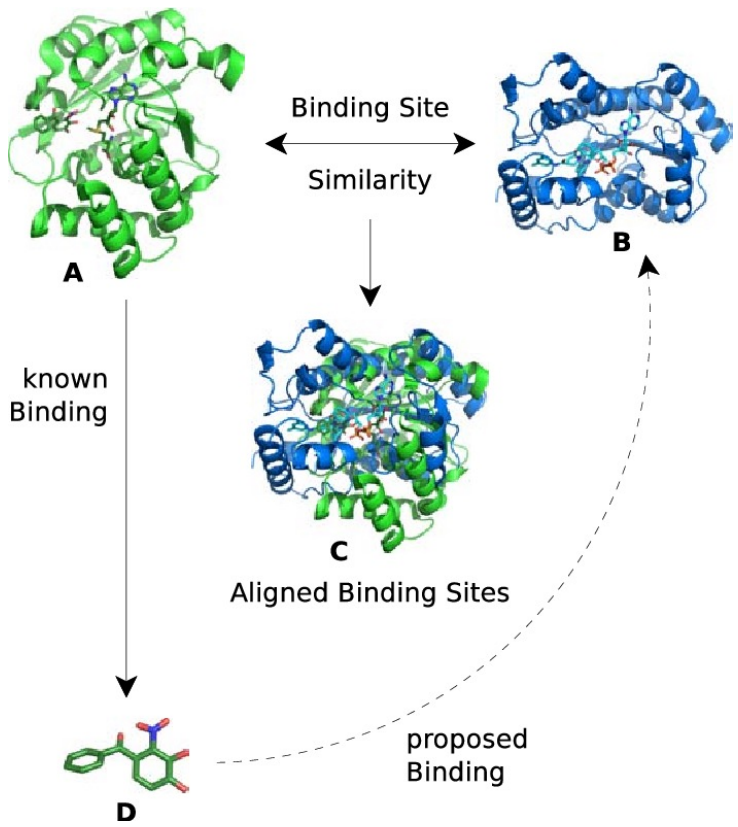


- Search possible binding pockets by calculating energies with probes

Similarity-based method

Table 2
Published template similarity-based LBS prediction methods.

Method	Type	Feature	Year
ConSurf [56]	Sequence Template-based	Phylogenetic relationships among the sequences and the similarity between the amino acids are taken into account	2003
A Sequence template-based approach with no specific name [57]	Sequence Template-based	An information-theoretic approach for estimating sequence conservation based on Jensen–Shannon divergence	2007
FINDSITE [58]	Structure Template-based	PROSPECTOR 3 threading algorithm and TMalign tool are used	2008
A two-stage template-based LBS prediction method [59]	Structure Template-based	Construct protein's 3D model and use structural clustering of ligand-containing templates on the predicted 3D model	2009
3DLigandSite [29]	Structure Template-based	MAMMOTH is used	2010
FunFOLD [60]	Structure Template-based	Use an automatic approach for cluster identification and residue selection	2011
COFACTOR [61]	Structure and Sequence Template-based	Use global-to-local sequence and structural comparison algorithm	2012
webPDBinder [62]	Structure Template-based	Search a protein structure against a library of known binding sites and a collection of control nonbinding pockets.	2013
S-SITE [31]	Sequence Template-based	Needleman–Wunsch algorithms are used	2013
TM-SITE [31]	Structure and Sequence Template-based	Mix Structure Template-based and Sequence Template-based method	2013



- Finding structurally similar ligand binding pockets from the PDB database

Machine-learning based methods

Table 3
Traditional machine learning-based LBS prediction and binding affinity research methods.

Method	Machine Learning Algorithm	Year
Knowledge-based QSAR approach [69]	Kernel-Partial Least Squares (K-PLS) [70]	2004
Multi-RELIEF [71]	RELIEF algorithm [72]	2007
SFCscore [73]	multiple linear regression partial least squares analysis	2008
ATPint [74]	Support Vector Machine	2009
ConCavity [75]	K-Means algorithm	2009
MetaPocket [76]	hierarchical clustering algorithm [77]	2009
RF-Score [4]	The Random Forest algorithm	2010
MetaDBSite [78]	Support Vector Machine	2011
NsitePred [79]	Support Vector Machine	2011
NNSCORE [80,81]	Artificial Neural Network (shallow neural network [82])	2011
L1pred [30]	L1-Logreg Regression classifier	2012
TargetS [83]	Support Vector Machine	2013
eFindSite [84]	Support Vector Machine	2013
VitaPred [85]	Support Vector Machine	2013
COACH [31]	Support Vector Machine	2013
LigandRFs [86]	The Random Forest algorithm	2014
OSML [87]	Support Vector Machine	2015
LigandDSES [88]	The Random Forest algorithm	2015
PRANK [89]	The Random Forest algorithm	2015
A method for protein-ligand binding affinity prediction [90]	Gradient Boosting Regressor [91]	2018
SAnDReS [92]	Regression Analysis	2016
P2Rank [93]	The Random Forest algorithm	2018
COACH-D [94]	Support Vector Machine	2018
Taba [95]	Regression Analysis	2019

Table 4
Deep learning-based LBS prediction and binding affinity research methods.

Method	Main Goal	Network Type	Year
A deep learning framework for modeling structural features of RNA-binding protein targets [118]	Binding references modeling of RNA-binding proteins	DBN	2015
DeepBind [119]	Sequence specificities prediction of DNA- and RNA-binding proteins	CNN	2015
DeepDTA [3]	Drug-target interaction identification	CNN	2018
K_{DEEP} [120]	Protein-ligand binding affinity prediction	CNN	2018
DEEPSite [36]	LBS Prediction	CNN	2017
DeepCSeqSite [121]	LBS Prediction	CNN	2019
DeepConv-DTI [122]	Drug-target interaction identification	CNN	2019
DeepDrug3D [35]	Binding pockets characterization and classification	CNN	2019
Onionnet [123]	Protein-ligand binding affinity prediction	CNN	2019

- Predicting the propensity of being ligand binding sites from l earning the structural patterns of the existing LBS

Practice

- P2Rank
  - Machine-learning-based algorithm
  - https://prankweb.cz/
- GalaxySite
  - Using an energy-based algorithm
  - https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=SITE
- Probis-Fold
  - Similarity-based search algorithm using AlphaFold DB
  - http://probis-fold.insilab.org/