Bioinformatics Projects in Kiharalab 2019

Daisuke Kihara

Professor
Department of Biological Sciences
Department of Computer Science
Purdue University, USA
dkihara@purdue.edu
http://kiharalab.org

Active Projects

- Protein-protein docking prediction
- Protein structure prediction
- Protein structure modeling for cryo-EM
- Protein structure/EM map database search
- Virtual drug screening
- Protein function prediction
- Moonlighting proteins
PROTEIN DOCKING

LZerD

normal vector

3DZernike descriptor

6Å

Available at: http://kiharalab.org/proteindocking/

(Venkatraman, Yang, Sael, & Kihara, BMC Bioinformatics, 10: 407, 2009)
3D Zernike Descriptors (3DZD)

- An extension of spherical harmonics based descriptors
- A 3D object can be represented by a series of orthogonal functions, thus practically represented by a series of coefficients as a feature vector
- Compact
- Rotation invariant

A surface representation of 1ew0A (A) is reconstructed from its 3D Zernike invariants of the order 5, 10, 15, 20, and 25 (B-F). (Sael & Kihara, 2009)

$$Z_m^n(r, \theta, \phi) = R_m(r)Y^n_m(\theta, \phi)$$

$$Y^n_m(\theta, \phi)$$: Spherical harmonics, $$R_m(r)$$: radial functions

$$Z_m^n(r, \theta, \phi)$$: polynomials in Cartesian coordinates

Zernike moments: $$\Omega_m^n = \frac{1}{2\pi} \int f(x) Z_m^n(x) dx$$

Zernike Descriptor: $$F_m = \left\{ \sum_{n=0}^{\infty} (\Omega_m^n)^2 \right\}^{\frac{1}{2}}$$

Multi-LZerD

Shape-based score

Physics-based score

Esquivel-Rodríguez, J et al. Proteins, 2012)
Multiple Docking Examples

1AOK
RMSD: 0.53Å

2AZE
RMSD: 0.73Å

1I3O
RMSD: 1.09Å

1VCB
RMSD: 1.57Å

Assembly Order Prediction for Protein Complexes

Urease accessory complex:
Dimer of UreF/UreH dimer is formed first, which binds to G dimer

Modeling the assembly order of multimeric heteroprotein complexes. PLOS Computational 8
Biology 14(11): e1005837
The number of votes for assembly pathways of 4hi0 across generations of the genetic algorithm

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005937

IDR-LZerD Overview

1. Predict multiple (30) conformations for each 9 residue-window based on sequence propensity
2. Dock fragments
(Peterson, Roy, Christoffer, Terashi, Kihara, PLOS Comp. Biol. 2017)
3. Combine docked fragments to create full model
4. Refine
Combining Docked Fragments

Eliminate pairs based on distance and angle cutoffs:

Example of Unbound Predictions

PDB ID 4ah2
20 AA
Rank 1
L-RMSD 5.8 Å
I-RMSD 2.4 Å

PDB ID 1ijj
25 AA
Rank 9
L-RMSD 5.1 Å
I-RMSD 2.5 Å

PDB ID 1l3e
44 AA
Rank 3
L-RMSD 7.4 Å
I-RMSD 6.3 Å

PDB ID 1jya
69 AA
Rank 2
L-RMSD 9.9 Å
I-RMSD 5.4 Å
DOVE: Protein docking model selection using 3D deep learning

(Wang, Terashi, Christoffer, Zhu, Kihara, Bioinformatics, 2019)

PROTEIN STRUCTURE MODELING FOR CRYO-EM
MAINMAST: De novo Structure Modeling for medium (~4 Å) Resolution Maps

EM Map

Cα model

EMD-6555 Porcine circovirus 2
Resolution: 2.9 Å

De-novo modeling: does not require known protein structures.
Fully automated: No need for visual inspection and human intervention.

(G. Terashi & D. Kihara, Nature Communications, 2018)

MAINMAST: (MAINchain Model trAcing from Spanning Tree)

EM map

Find local dense points (LDPs) by Mean Shift

Connect All points by Minimum Spanning Tree

Refine Tree Structure

Thread sequence on the longest path

Cα models ranked by threading

The longest path (red) does not always cover the whole protein structure.
Emap2sec: Detecting Secondary Structures in 6-10 Å EM maps

Examples of Structure Detection in real maps

Resolution
a: 6.0 Å
b: 6.2 Å
c: 6.8 Å
d: 8.3 Å
e: 9.1 Å
f: 7.9 Å
MOLECULAR SURFACE SHAPE ANALYSIS, DRUG SCREENING

Rapid & Seamless Screening of Similar/Complementary 3D Molecules
3D-SURFER 2.0

A web-based tool for real-time protein surface comparison and analysis. The server integrates a repertoire of methods to assist in high throughput screening and visualization of protein surface comparisons. 3D-Zernike Descriptors (3ZD) are utilized for the efficient comparison of protein surfaces.

Statistics of latest release (Last updated on Sep 5, 2018):

- All Entries: 702,197
- Chain Entries: 368,549
- Complex Entries: 84,740
- Domain Entries: 248,908
- All (IACN) Entries: 701,836
- Chain (IACN) Entries: 368,397
- Complex (IACN) Entries: 84,512
- Domain (IACN) Entries: 248,907

http://kiharalab.org/3d-surfer/

(La et al., Bioinformatics 2009)
(Xiong et al., Methods in Mol Biol, 2013)

EM-SURFER

A web-based tool for real-time comparison and analysis of Electron Microscopy (EM) density maps. It compares the shape of EM map isosurfaces, generated using author-recommended contour values. Users can either upload an EM map or choose an existing map in the EMDB and compare it against maps stored in the EMDB. 3D-Zernike Descriptors (3ZD) are utilized for the efficient comparison between EM maps. The web interface is designed in an intuitive manner to aid users.


- All Entries with resolution data: 6,727
- Entries with resolution <5Å: 1,963
- Entries with resolution 5Å and <10Å: 1,070
- Entries with resolution 10Å and <15Å: 804
- Entries with resolution 15Å and <20Å: 787
- Entries with resolution ≥20Å: 1,701

http://kiharalab.org/em-surfer/

(Equivel-Rodriguez et al., BMC Bioinformatics 2015)
Patch-Surfer 2.0: Local Patch-Based Pocket Comparison Method

Data Process

Generate surface
Extract pocket
Segment to patches
Compute descriptors

Database Search

Patch Scores

Ligand Type Prediction

(Sael & Kihara, Proteins, 2012)
(Zhu, Xiong, & Kihara, Bioinformatics, 2015)

Patch-Surfer Retrieval Results for Flexible Ligands: FAD and NAD

flavin adenine dinucleotide (FAD)  FAD  Nicotinamide adenine dinucleotide (NAD)

A  B  C

1cqX  1jr8  1e8g  1k87  1mi3  1s7g

Patch-based: 3rd
Global pocket: 31st
Patch-based: 1st
Global pocket: 18th
Patch-based: 2nd
Global pocket: 16th
Combined Computational & Experimental Approach

PL-PatchSurfer2: Local Surface-Based Virtual Screening

Sequence-Based Function Prediction

(Jain & Kihara, Bioinformatics, 2018)
DextMP: Finding Moonlighting Proteins from Literature Text Information

(Khan, Bhuiyan, & Kihara, Bioinformatics 2017)
iGFP: Group Function Prediction from Functional Relevance Networks with Conditional Random Field

(Khan, Jain, Rawi, Bensmail, & Kihara, Bioinformatics 2018)