Domain specificity

Jinho Kim
Cellular Signal Integration Group (C-SIG)
Center for Biological Sequence Analysis (CBS)
Department of Systems Biology
Technical University of Denmark (DTU)
• Computational modeling of domains specificity
• Evolution of domains specificity
• Using software pipelines
Protein devices: logical gates/switches built from protein modules

Protein modules: domains (eg. SH2) and linear-motifs (eg. phosphorylation site)

Proteome: 80% of proteins in higher eukaryotes are modular

Cellular protein circuitry **compute** dynamic responses to environmental cues

Computational Modeling of Specificity of PDZ Domains
PDZ domain

- PDZ domain is one of the most commonly found protein-protein interaction domains in various organisms from bacteria to humans.

- By recognizing specific C-terminal residues of ligands, PDZ domains play critical roles in key biological processes such as signal transduction, synapse formation, protein trafficking, and cell polarity maintenance.
PDZ proteins play important roles in the postsynaptic density (PSD).

W. Feng, M Zhang, Nat. Rev. Neurosci 2009
Human PDZ domain-ligand interaction network (PDZNet)
1,212 interactions between 97 PDZ proteins and 596 ligands
Construction of PDZNet

A. Experimental datasets of PDZ domain-peptide interactions

- PSD-95
- SAP102
- ERBIN
- CASK
- Frizzled-1
- Stargazin
- ERBB4
- Caspr2
- Syndecan-1

Protein array → PDZBase → Phage display

Build a quantitative model of PDZ domain-peptide interactions

Generate Position Weight Matrices (PWMs) and calculate binding scores of PDZ protein-ligand interactions

B. PPI databases

- HPRD
- MIPS
- DIP

Integration → Select PDZ domain-mediated interactions

C. PDZNet

- CELR3
- NMDZ1
- CLD1
- JAM1
- AKA10
- NHRF3
- PSD-95
- ZO1
- PICK
- ODO2
- DLG5
- HGS
- NOS
- ARG1
- SRY
- NHRF2
- CTBP1

PDZ protein → Ligand → Binding score

Affinity contribution

PWM of PSD-95

Position

Affinity contribution

Binding score
The model identifies known binding PDZ domain-ligand interactions.

### Known ligands of PSD-95_1 PDZ domain

<table>
<thead>
<tr>
<th>Species</th>
<th>Ligand</th>
<th>C-terminal</th>
<th>Binding score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>PMCA4b</td>
<td>ETSV</td>
<td>13.05</td>
</tr>
<tr>
<td></td>
<td>PMCA2b</td>
<td>ETSL</td>
<td>9.49</td>
</tr>
<tr>
<td>Mouse</td>
<td>Frizzled-4</td>
<td>ETVV</td>
<td>13.80</td>
</tr>
<tr>
<td></td>
<td>Frizzled-1</td>
<td>ETTV</td>
<td>13.49</td>
</tr>
<tr>
<td></td>
<td>Frizzled-2</td>
<td>ETTV</td>
<td>13.49</td>
</tr>
<tr>
<td></td>
<td>Frizzled-7</td>
<td>ETAV</td>
<td>13.19</td>
</tr>
<tr>
<td></td>
<td>Sema4c</td>
<td>ESSV</td>
<td>10.61</td>
</tr>
<tr>
<td></td>
<td>BAI1</td>
<td>QTEV</td>
<td>10.52</td>
</tr>
<tr>
<td></td>
<td>Stargazin</td>
<td>TTPV</td>
<td>9.77</td>
</tr>
<tr>
<td>Rat</td>
<td>Kv1.4</td>
<td>ETDV</td>
<td>14.72</td>
</tr>
<tr>
<td></td>
<td>ERBB4</td>
<td>NTVV</td>
<td>12.67</td>
</tr>
<tr>
<td></td>
<td>NMDAR2A</td>
<td>ESDV</td>
<td>12.28</td>
</tr>
<tr>
<td></td>
<td>NMDAR2B</td>
<td>ESDV</td>
<td>12.28</td>
</tr>
<tr>
<td></td>
<td>SynGAP</td>
<td>QTRV</td>
<td>9.79</td>
</tr>
<tr>
<td></td>
<td>PKC-A</td>
<td>QSAV</td>
<td>8.99</td>
</tr>
<tr>
<td></td>
<td>Sec8</td>
<td>ITTV</td>
<td>7.93</td>
</tr>
<tr>
<td></td>
<td>GluR6</td>
<td>ETMA</td>
<td>6.00</td>
</tr>
</tbody>
</table>
In-vivo ligands binding affinities of PSD-95 PDZ1 domain correlates with binding scores

<table>
<thead>
<tr>
<th>Species</th>
<th>Ligand</th>
<th>C-terminal</th>
<th>Affinity (μM)</th>
<th>Binding score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rat</td>
<td>Kv1.4</td>
<td>ETDV</td>
<td>0.02</td>
<td>14.72</td>
</tr>
<tr>
<td></td>
<td>GluR6</td>
<td>ETMA</td>
<td>1.5</td>
<td>6.00</td>
</tr>
</tbody>
</table>

(Affinity is from Piserchio et al, 2002)
The binding score correlates with experimental affinity.
Fraction of known PDZ domain-ligand interactions by percentile rank of binding score
Evolution of domain specificity & Interaction rewiring
A strong and intuitive way to understand gene functions

Delete & Look
Comparative genomics is a powerful tool to understand genotype-phenotype relationships.
Systematic analysis of interaction rewiring will provide new insights into phenotypic difference across species.

Interaction rewiring can reconfigure molecular systems without a gain or loss of genes.
“Understanding genotype–phenotype relationships requires that phenotype be viewed as manifestations of network properties, rather than simply as the result of individual genomic variations.” (Nature 2012)
Comparative interactomics
Why only humans are smart?
EVOLUTION BECOMES REVOLUTION

THE RISE OF PLANET OF THE APES

IN CINEMAS AUGUST 4
What is molecular basis of difference in thinking ability across species?
Phylogenetic profile of human PDZ proteins and ligands across 16 fully sequenced species.
Two models describe the difference of PDZ domain-ligand interactions across species.
An example of a PDZ domain-ligand interaction rewiring by sequence mutations – NOS1AP
An example of a PDZ domain-ligand interaction rewiring by sequence mutations – EXOC4
A DNA segment insertion generated a PDZ-binding motif in the C-terminal amino acids of the *Oryzias latipes* EXOC4 protein.
A point mutation generated a PDZ-binding motif in the C-terminal amino acids of the *Macaca mulatta* PBK protein.
Rewiring of PDZ domain-ligand interactions substantially contributed to the functions of nervous systems in vertebrates.
Neurological diseases were most highly associated with mutations of the PDZNet proteins.

Disease classes

- Neurological
- Respiratory
- Gastrointestinal
- Muscular

OMIM

Online Mendelian Inheritance in Man

Bone and cartilage cancer
Breast cancer

P53 TAD
Codons 5–29

P53
Codons 95–289

TP53 tetramer
Codons 318–359

Park, et al. MSB 2009
Neurological diseases were most highly associated with mutations of the PDZNet proteins.
Can we identify mutations affecting PDZ-ligand binding, which causes genetic diseases.

Disruption of the mGluR7a-PICK1 complex is sufficient to induce absence epilepsy–like seizures in rats and mice (Bertaso et al., Nat. Neuroscience 2008).
Can we identify mutations affecting PDZ-ligand binding, which causes genetic diseases.
User friendly web service – Input
http://sbi.postech.ac.kr/pdz/

PDZ domain Specificity Prediction

SUBMISSION

This web service predicts the binding specificity of a query PDZ domain. The program automatically extracts the pocket residues of a PDZ domain from a query sequence and generates a position weight matrix (PWM), based on the binding specificity of a reference set of 133 PDZ domains. By using the PWM, a rank list of potential binding partners is produced from Swiss-prot protein database.

Enter the PDZ domain sequence into the text area and choose a species. Then press `Submit`.

Note: The program takes single PDZ domain currently. If your sequence contains multiple PDZ domains, please divide it such that each sequence fragment have one PDZ domain.

Enter sequence only in FASTA format (example):

PDZ domain sequence

A rank list of predicted ligands are made from **Human**

Species
User friendly web service – Flow diagram & Output

Web server diagram and sample output

Predicted interaction candidate list
PDZ domain-containing 1 (PDZK1) regulates phospholipase C-β3 (PLC-β3) specific activation of somatostatin by forming a ternary complex with PLC-β3 and somatostatin receptors

Jung Kuk Kim, Ohman Kwon, Jinho Kim, Eung-Kyun Kim, Hye Kyung Park, Ji Eun Lee, Kyung Lock Kim, Jung Woong Choi, Seyoung Lim, Heon Seok, Whaseon Lee-Kwon, Jang Hyun Choi, Byoung Heon Kang, Sanguk Kim, Sung Ho Ryu and Pann-Ghill Suh

<table>
<thead>
<tr>
<th></th>
<th>2nd PDZ domain</th>
<th>3rd PDZ domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Probable GPR135 (DTSL)</td>
<td>SSTR1 (ITTL)</td>
</tr>
<tr>
<td>2</td>
<td>β-2 adrenergicR (DSLL)</td>
<td>mGluR3 (TSSL)</td>
</tr>
<tr>
<td>3</td>
<td>Probable GPR37 (GTHC)</td>
<td>mGluR2 (TSSL)</td>
</tr>
<tr>
<td>4</td>
<td>Probable GPR19 (NTFV)</td>
<td>P2Y4R (ADRL)</td>
</tr>
<tr>
<td>5</td>
<td>LPAR5 (DSAL)</td>
<td>TSHR (QTVL)</td>
</tr>
<tr>
<td>6</td>
<td>SSTR1 (ITTL)</td>
<td>Probable GPR135 (DTSL)</td>
</tr>
<tr>
<td>7</td>
<td>TSHR (QTVL)</td>
<td>SSTR5 (TSKL)</td>
</tr>
<tr>
<td>8</td>
<td>S1PR2 (NTVV)</td>
<td>SSTR3 (ISYL)</td>
</tr>
<tr>
<td>9</td>
<td>GIPR (ESYC)</td>
<td>SSTR2 (QTSI)</td>
</tr>
<tr>
<td>10</td>
<td>LT4R2 (EWDL)</td>
<td>5HT2C (ISSV)</td>
</tr>
<tr>
<td>11</td>
<td>PTH1R (ETVM)</td>
<td>NMBR (EMAL)</td>
</tr>
<tr>
<td>12</td>
<td>PTH2R (EDVL)</td>
<td>SSTR4 (TTTF)</td>
</tr>
<tr>
<td>13</td>
<td>GPR44 (STSS)</td>
<td>P2Y2R (DIRL)</td>
</tr>
<tr>
<td>14</td>
<td>NPFF2R (SSEI)</td>
<td>LPAR5 (DSAL)</td>
</tr>
<tr>
<td>15</td>
<td>SSTR3 (ISYL)</td>
<td>CNR1 (AEAL)</td>
</tr>
</tbody>
</table>

Kim, et al, JBC 2012
PDZ domains of PDZK1 mediate ternary complex formation.

<table>
<thead>
<tr>
<th>Flag-SSTR1~5</th>
<th>GST</th>
<th>GST-PDZ1</th>
<th>GST-PDZ2</th>
<th>GST-PDZ3</th>
<th>GST-PDZ4</th>
</tr>
</thead>
<tbody>
<tr>
<td>GST Pull-down</td>
<td>SSTR1</td>
<td>SSTR2</td>
<td>SSTR3</td>
<td>SSTR4</td>
<td>SSTR5</td>
</tr>
<tr>
<td>GST Pull-down</td>
<td>GST</td>
<td>GST-PDZ1</td>
<td>GST-PDZ2</td>
<td>GST-PDZ3</td>
<td>GST-PDZ4</td>
</tr>
</tbody>
</table>

Blot: α-Flag Ab

![Diagram showing SSTRs, PLC-β3, PDZK1, and Gαγ structures](image)

**Kim, et al, JBC 2012**

Experiment conducted by Jung Kuk Kim
Linear Motif Atlas for Phosphorylation-Dependent Signaling

Martin Lee Miller,1,2* Lars Juhl Jensen,2,3* Francesca Diella,3 Claus Jørgensen,4 Michele Tinti,5 Lei Li,6 Marilyn Hsiung,4 Sirlester A. Parker,7 Jennifer Bordeaux,7 Thomas Sicheritz-Ponten,1 Marina Olhovsky,4 Adrian Pasculescu,4 Jes Alexander,8 Stefan Knapp,9 Nikolaj Blom,1 Peer Bork,2,10 Shawn Li,6 Gianni Cesareni,5 Tony Pawson,4 Benjamin E. Turk,7 Michael B. Yaffe,8† Søren Brunak,1,2† Rune Linding4,8,11†

(Published 2 September 2008)
C

Redundancy reduction
- Phosphosite
- Peptide
- Full length

Elimination of examples that are too similar at the sequence level

Keep

Discard

Similar peptide

Similar full-length sequence

D

Partitioning
- Training
- Test
- Validation

Round-robin distribution of examples into training, test, and validation sets
Overview of the performance of the NetPhoreost classifiers

![Bar chart showing the area under the receiver operating characteristic curve for the NetPhoreost classifiers. The x-axis represents the area under the curve ranging from 0.5 to 1.0, while the y-axis represents the number of classifiers. The chart demonstrates varying performance levels across different areas.](image-url)
Systematic Discovery of In Vivo Phosphorylation Networks


2007
NetworKIN Algorithm

Effects of Including Substrate Context
Development of Computational Framework

**NetworKIN**

MS identification of phosphorylation sites

Manual annotation of phosphorylation sites

Matching of sequence motifs for kinase families

Construction of a context network from STRING

Related publications:

Linding et al. Cell, 129,

Miller et al. Science Signaling

**NetPhorest**

Protein domain sequences

In vitro specificity assays

Classified phosphorylation sites

Motif atlas

Matching of sequence motifs for kinase families

Design of consensus antibodies

Detection of purification biases

Modeling of signaling networks

Related publications:

http://NetworKIN.info

http://NetPhorest.info