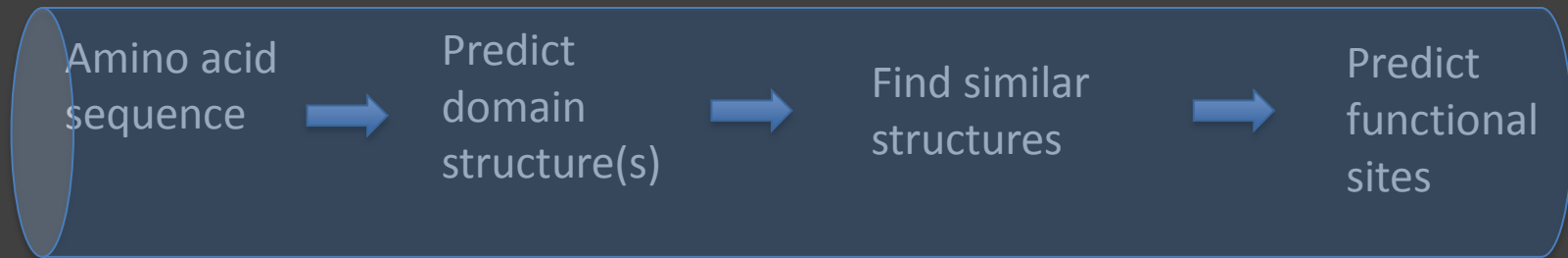


# A Functional Annotation Pipeline for Orphan Sequences

Daron M. Standley

WPI Immunology Frontier Research  
Center (IFReC), Osaka Univ.

# General Approach



Program flow is a pipeline

# Pipeline [MAFFTash|MSThread|Spanner|SeSAW]

**Start** Progress PrepMAFFTash Results MAFFTash Results MSThread Results Spanner Results SeSAW Results

## Pipeline Input

### Query (Amino Acid Sequence)

GVPIALYNRGHLGSGHFGAVIKASLDDGTLAAKVPYSQIVPNADATSAELEAGISSARAELVKTIQELDVRDKLVAKGLTLTETVSOYGLPLCOMTLTLPENKATVVRRGSRLF  
VVSKEVMLLPLIDGSALNSLVQSOPPFQRAVAREAILALAKLHELGFAGHDVKLNNMMIDVHGFHMLDMGSRVPVDSCVSEEDKYLRWAPELAKSQHTSOKTCLKRGALDV  
WALGLAIFEFVCFNRLPYSLSNLPSSFWSRVEHLSRLRLSDFSVKDCNESDPAVMGIVVQFLNPDQERPELPKFVNSYTFQOAPGVTSHLTRIPTTELS

### Template (PDB id and Chain id)

1s9jA

Run

<http://sysimm.ifrec.osaka-u.ac.jp/pipelineDemo/>

# Conserved Domain Database (CDD) uses sequence alignment and predicts Ser/Thr kinase function



HOME SEARCH SITE MAP

NewSearch

Structure Home

3D Macromolecular Structures

Conserved Domains

Pubchem

BioSystems

## Conserved Domains

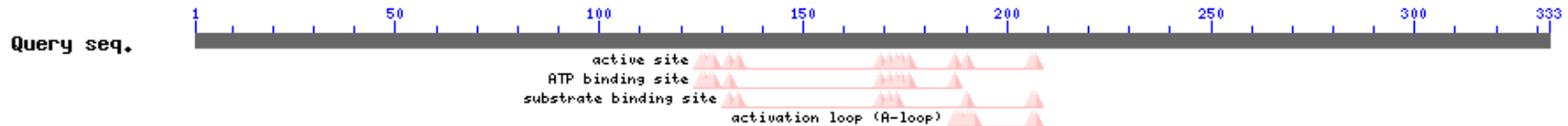


## Conserved domains on [lcl|9261]

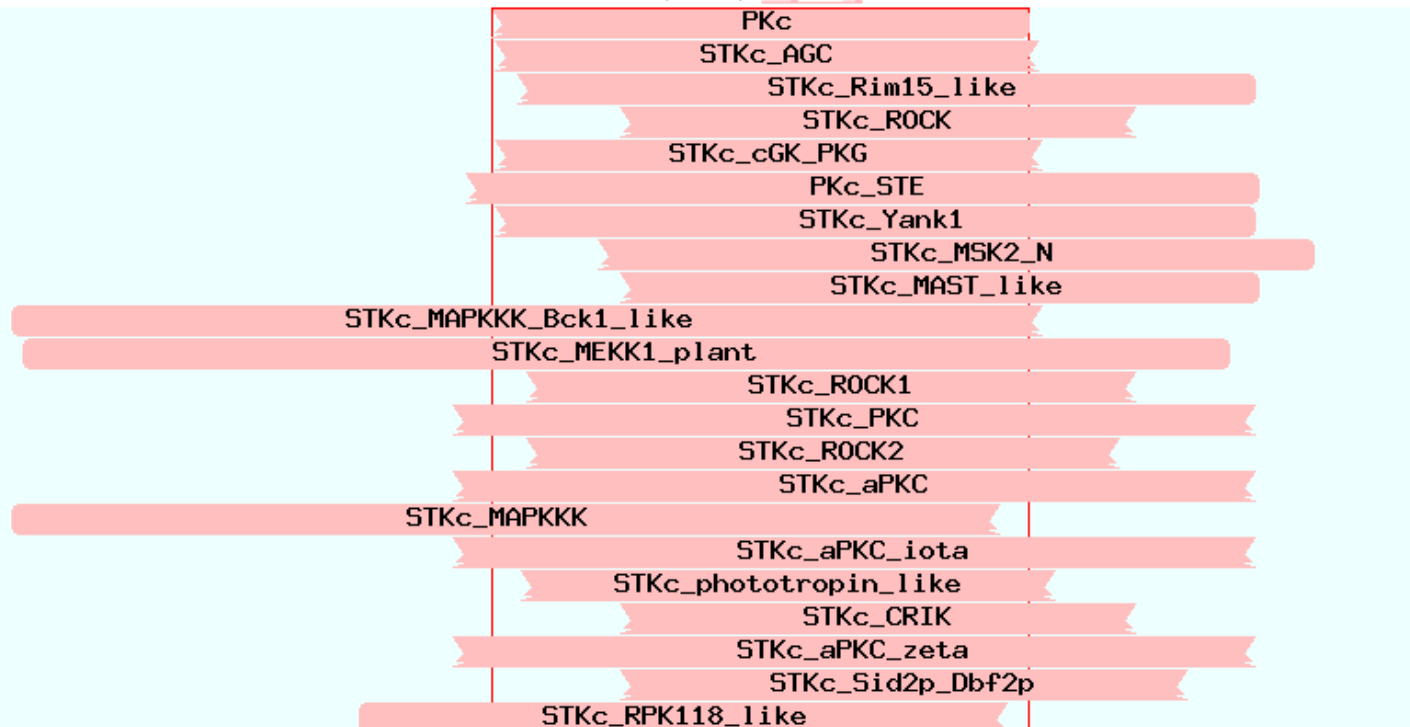
[View concise result](#)

Local query sequence

### Graphical summary [show options >](#)



### Non-specific hits



# Pipeline [MAFFTash|MSThread|Spanner|SeSAW]

[Start](#) **[Progress](#)** [PrepMAFFTash Results](#) [MAFFTash Results](#) [MSThread Results](#) [Spanner Results](#) [SeSAW Results](#)

## Running Progress

Preparing query data  
Preparing template data  
Preparing template msa data  
Preparing query msa data  
Running MSThread



# Pipeline [MAFFTash|MSThread|Spanner|SeSAW]

[Start](#)[Progress](#)[PrepMAFFTash Results](#)[MAFFTash Results](#)[MSThread Results](#)[Spanner Results](#)[SeSAW Results](#)

## PrepMafftAsh Results

### Query

```
>query_pd_1280811594495_query  
GVPIALYNRGHLGSGHFGAVIKASLDDGTLAAKVPYSQIVPNADATSAELEAGISSARAEKVKTIRQELDV RDKL VAKGLTLTETVSQYGLPLCQMTLTL PENKATVVRRGSRLFVVSKEV  
>UniRef90_B6KC39
```

[Download file](#)

### Template

```
>PDBID  
1s9jA  
>PDBID
```

[Download file](#)

## Pipeline [MAFFTash|MSThread|Spanner|SeSAW]

Start	Progress	PrepMAFFTash Results	MAFFTash Results	MSThread Results	Spanner Results	SeSAW Results
-------	----------	----------------------	------------------	------------------	-----------------	---------------

## MafftAsh Results

### Query

File Edit Select View Format Colour Calculate Help

1=query\_pd\_1280811594495\_query  
2=UniRef90\_B6KC39

Alignment has no annotations

Sequence 2 ID: 2=UniRef90\_B6KC39 Residue: GLY (80)



## Template

File Edit Select View Format Colour Calculate Help

10 20 30 40 50 60 70 80

1\_1s9jA.pdb-A  
2\_3enmD.pdb-D  
3\_2dyIA.pdb-A  
4\_3ggfA.pdb-A  
5\_3a7hB.pdb-B  
6\_2l7zA.pdb-A

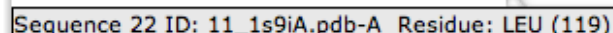
MELKDDDFEKI  
MENFEVKADDLEPI  
TGYLTGGQRYQAEINDLENL  
IADPEELFTKL  
SGLPGMQNLKADPEELFTKL  
HVRRLDLPNEVWEIV

SELGAGNGGVV  
MELGRGAYGVV  
GEMTGVQV  
ERIGKGSFGEV  
EKIGKGSFGEV  
GELGDGAFGKV

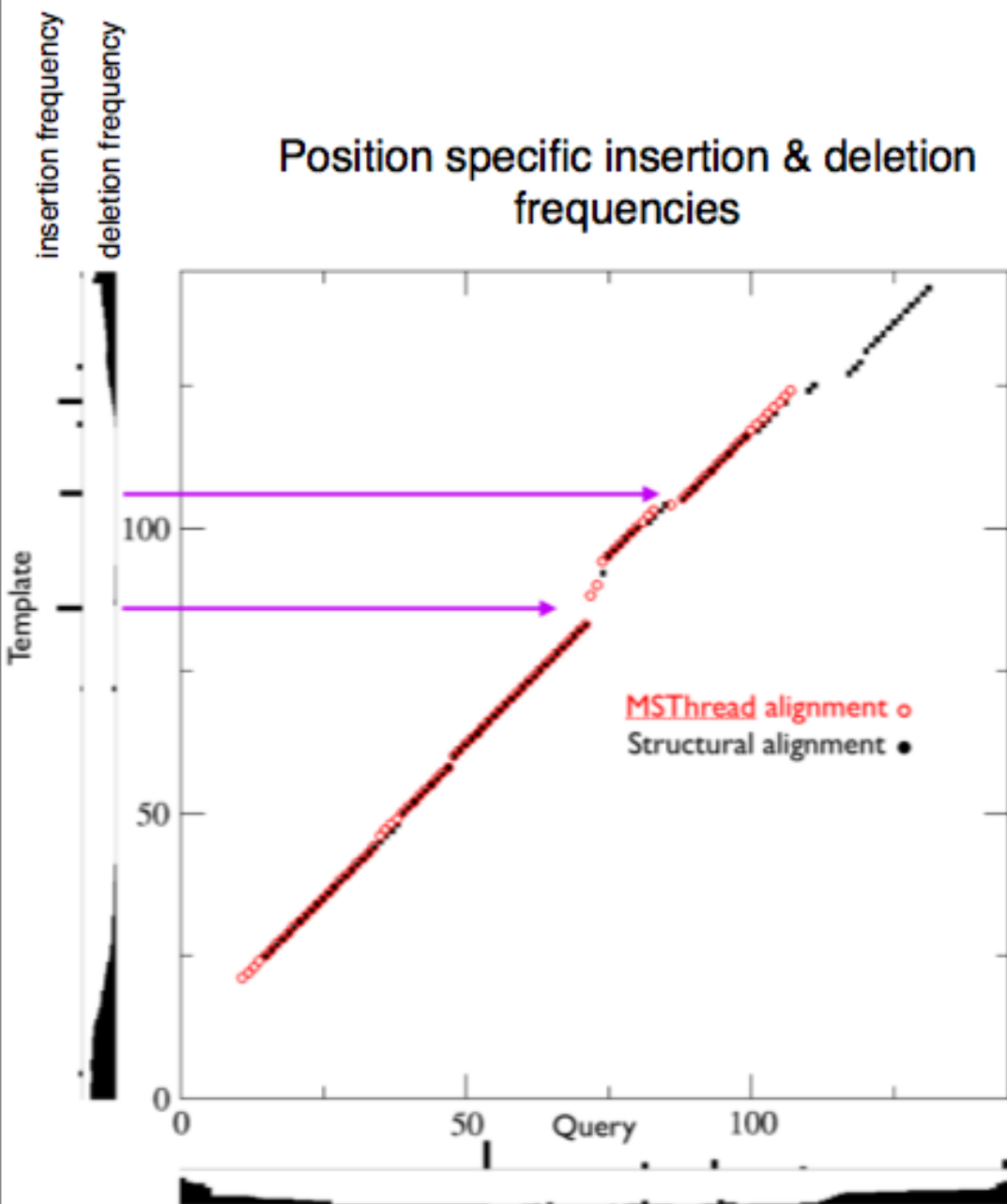
F  
E  
W  
F  
F  
Y

KVSH  
KMRH  
KMRF  
KGI  
KGI  
KAKN

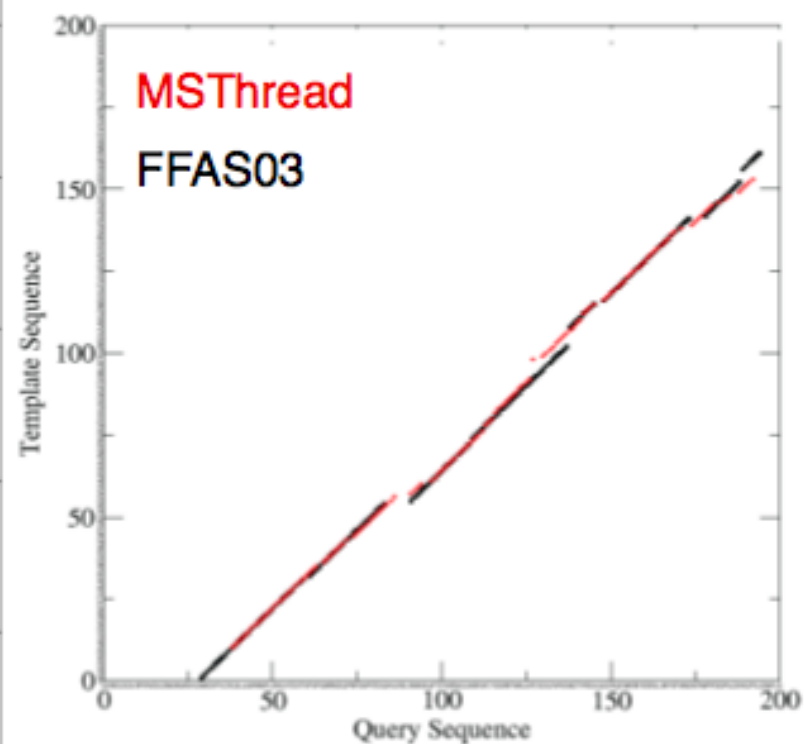
Sequences: 439  
Minimum Sequence Length: 153  
Maximum Sequence Length: 372  
Average Length: 271







zc3h12a-2qip alignment (12% ID)



insertion frequency  
deletion frequency

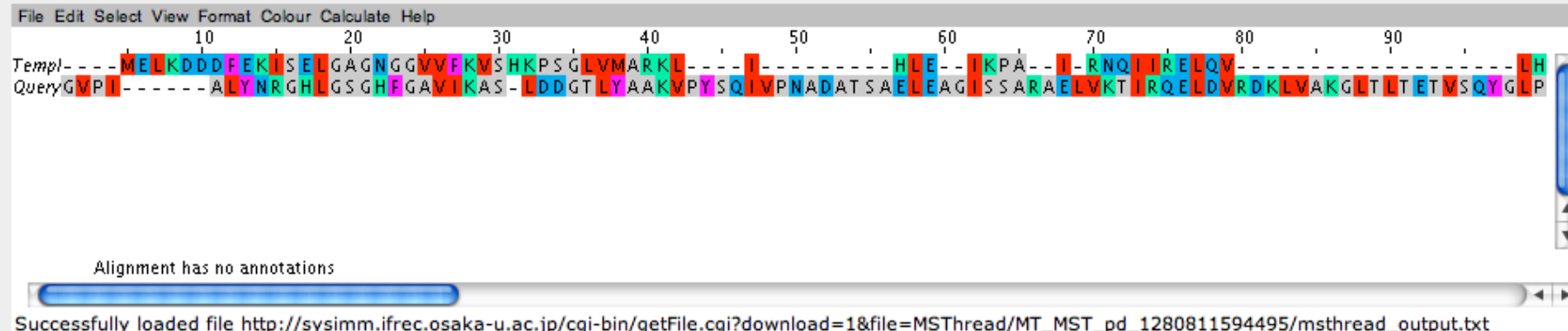
# Pipeline [MAFFTash|MSThread|Spanner|SeSAW]

Start Progress PrepMAFFTash Results MAFFTash Results **MSThread Results** Spanner Results SeSAW Results

## MSThread Result

Please review MSThread result and select which one you wish to run through Spanner

Score : 1223.486



[Download file](#)

Run this result in Spanner >>

# Pipeline [MAFFTash|MSThread|Spanner|SeSAW]

[Start](#) **[Progress](#)** [PrepMAFFTash Results](#) [MAFFTash Results](#) [MSThread Results](#) [Spanner Results](#) [SeSAW Results](#)

## Running Progress

Preparing query data

Preparing template data

Preparing template msa data

Preparing query msa data

Running MSThread

Running Spanner



# Pipeline [MAFFTash|MSThread|Spanner|SeSAW]

[Start](#) [Progress](#) [PrepMAFFTash Results](#) [MAFFTash Results](#) [MSThread Results](#) [Spanner Results](#) [SeSAW Results](#)

Spanner pdbfile will be displayed here, after the result are available



[Download file](#) [Run this result with SeSaw](#)

# Pipeline [MAFFTash|MSThread|Spanner|SeSAW]

**Start** **Progress** PrepMAFFTash Results MAFFTash Results MSThread Results Spanner Results SeSAW Results

## Running Progress

Preparing query data  
Preparing template data  
Preparing template msa data  
Preparing query msa data  
Running MSThread  
Running Spanner  
Running SeSAW



# Pipeline [MAFFTash|MSThread|Spanner|SeSAW]

[Start](#) [Progress](#) [PrepMAFFTash Results](#) [MAFFTash Results](#) [MSThread Results](#) [Spanner Results](#) [SeSAW Results](#)

SeSAW result will be displayed here, after the result are available


[result #1](#) [result #2](#) [result #3](#)

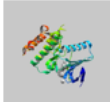
Template	Score	P-value	Identity	Coverage	Domain	Annot	Alignment		Superposition <a href="#">(help)</a>
<a href="#">3h9fA</a>	108.3	4.3e-02	22	64			<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">3c4xB</a>	105.2	4.4e-02	21	63		<a href="#">GO</a>	<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">2dq7X</a>	102.5	4.5e-02	26	65		<a href="#">GO</a>	<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">3b8qB</a>	98.6	4.7e-02	22	61		<a href="#">GO</a>	<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">3lj0B</a>	96.6	4.9e-02	17	62			<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">3a8wA</a>	94.0	5.0e-02	28	62			<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">3gu6A</a>	93.3	5.0e-02	18	65			<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">3hmpA</a>	92.6	5.1e-02	21	65			<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">3g5dB</a>	92.6	5.1e-02	23	65			<a href="#">View</a> <a href="#">Download</a>	<i>f</i>	<a href="#">View</a> <a href="#">Download</a>
<a href="#">3iw4A</a>	92.3	5.1e-02	23	65			<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">2dylA</a>	92.0	5.1e-02	21	57	CATH:1.10.510.10	<a href="#">GO</a>	<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">2wzjD</a>	91.7	5.1e-02	25	64			<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>
<a href="#">1zmvB</a>	91.6	5.1e-02	25	64		<a href="#">GO</a>	<a href="#">View</a> <a href="#">Download</a>	<i>f</i>	<a href="#">View</a> <a href="#">Download</a>
<a href="#">1zmvA</a>	91.3	5.2e-02	25	64		<a href="#">GO</a>	<a href="#">View</a> <a href="#">Download</a>	<i>f</i>	<a href="#">View</a> <a href="#">Download</a>
<a href="#">3h0yA</a>	91.0	5.2e-02	20	60			<a href="#">View</a> <a href="#">Download</a>	<i>f</i>	<a href="#">View</a> <a href="#">Download</a>
<a href="#">3hmoA</a>	90.8	5.2e-02	21	64			<a href="#">View</a> <a href="#">Download</a>		<a href="#">View</a> <a href="#">Download</a>



<b>PDB ID</b>	3h9f <a href="#">sequence information (FASTA format)</a> <a href="#">download PDB format file</a>
<b>Descriptor</b>	Dual specificity protein kinase TTK (E.C.2.7.12.1)
<b>Title</b>	Crystal Structure of Human Dual Specificity Protein Kinase (TTK) in complex with a pyrimido-diazepin ligand
<b>Functional Keywords</b>	TTK, hMPS1, PYT, ESK, kinase, Dual Specificity, phosphotyrosine picked threonine kinase, SGC, Structural Genomics Consortium, ATP-binding, Nucleotide-binding, Phosphoprotein, Polymorphism, Serine/threonine-protein kinase, Transferase, Tyrosine-protein kinase TRANSFERASE
<b>Biological source</b>	Homo sapiens (human)
<b>Organ source</b>	[UNP - TTK_HUMAN] Testis [UNP - TTK_HUMAN] Eye, and Testis [UNP - TTK_HUMAN] Epithelium

Service and Software >>  
jV: Graphic Viewer  
Protein Globe  
ASH  
MAFFTash  
Structure Prediction >>  
CRNPRED  
Spanner  
SFAS  
Derived database >>  
eF-site/eF-seek/eF-surf

  
**Structure Viewers**  
[jV3](#) / [Jmol](#)  
jV3 and Jmol require  
Java(TM)Plug-in 1.5 or  
later.

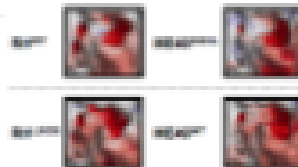
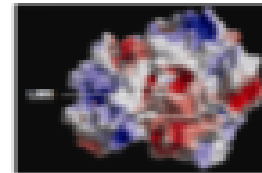
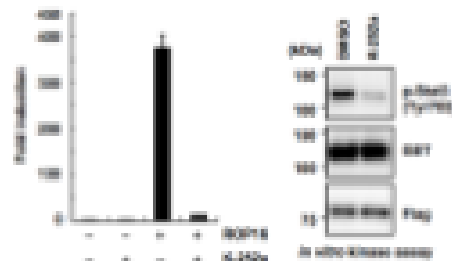
  
rotated about x by 90°

<b>Biological source</b>	TRANSFERASE Homo sapiens (human)
<b>Organ source</b>	[UNP - TTK_HUMAN] Testis [UNP - TTK_HUMAN] Eye, and Testis [UNP - TTK_HUMAN] Epithelium
<b>Total number of polymer chains</b>	1
<b>Total molecular weight</b>	36870.7 (the details in <a href="#">Structural Details Page</a> )
<b>Authors</b>	Filippakopoulos, P. , Soundararajan, M. , Keates, T. , Elkins, J.M. , King, O. , Fedorov, O. , Picaud, S.S. , Pike, A.C.W. , Yue, W. , Chaikuad, A. , von Delft, F. , Arrowsmith, C.H. , Edwards, A. , Weigelt, J. , Bountra, C. , Kwiatkowski, N. , Grey, N. , Knapp, S. , Structural Genomics Consortium (SGC) ( <i>deposition date</i> : 2009-04-30, <i>release date</i> : 2009-05-19) Filippakopoulos, P. , Soundararajan, M. , Keates, T. , Elkins, J.M. , King, O. , Fedorov, O. , Picaud, S.S. , Pike, A.C.W. , Yue, W. , Chaikuad, A. , von Delft, F. , Arrowsmith, C.H. , Edwards, A. , Weigelt, J. , Bountra, C. , Kwiatkowski, N. , Grey, N. , Knapp, S. , Structural Genomics Consortium (SGC)

SeSAW correctly identifies dual specificity (Tyr + Ser/Thr) kinase function  
whereas conventional sequence alignment predicts Ser/Thr kinase function

A single polymorphic amino acid on *Toxoplasma gondii* kinase ROP16 determines the direct and strain-specific activation of Stat3, Masahiro Yamamoto, et al (2009) *JEM*

Accession	Strain Name	Score	Length
AF175	Strain-specific, integer activated, under-activated strain 1	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 2	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 3	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 4	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 5	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 6	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 7	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 8	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 9	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 10	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 11	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 12	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 13	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 14	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 15	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 16	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 17	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 18	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 19	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 20	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 21	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 22	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 23	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 24	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 25	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 26	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 27	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 28	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 29	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 30	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 31	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 32	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 33	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 34	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 35	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 36	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 37	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 38	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 39	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 40	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 41	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 42	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 43	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 44	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 45	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 46	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 47	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 48	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 49	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 50	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 51	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 52	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 53	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 54	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 55	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 56	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 57	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 58	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 59	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 60	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 61	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 62	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 63	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 64	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 65	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 66	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 67	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 68	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 69	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 70	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 71	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 72	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 73	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 74	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 75	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 76	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 77	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 78	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 79	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 80	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 81	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 82	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 83	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 84	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 85	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 86	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 87	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 88	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 89	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 90	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 91	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 92	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 93	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 94	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 95	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 96	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 97	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 98	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 99	109.02	25
AF175	Strain-specific, integer activated, under-activated strain 100	109.02	25



SeSAW correctly distinguished between dual specific (Ser/Thr or Tyr) and Ser/Thr-specific kinases based on our 20% ID homology model. In contrast, sequence alignment alone indicated only Ser/Thr kinase hits. We also modeled the polymorphic position 503 which has been shown to affect the virulence of *T. gondii* and showed that the it sits beneath the predicted substrate binding cavity, leading to characteristic changes in shape and electrostatic potential.

Macrophages from *Zc3h12a*<sup>-/-</sup> mice showed highly increased production of inflammatory cytokines (IL-6 and IL-12p40)



# Conclusions

Our pipeline uses structural modeling to identify remote homologs  
This approach can be more sensitive than sequence alignment

Successful predictions include zc3h12a (Matsushita, et al Nature) and  
ROP-16 (Yamamoto et al (JEM))

Still to do:

Enable bookmarking results

More interaction with other web servers

# Acknowledgements



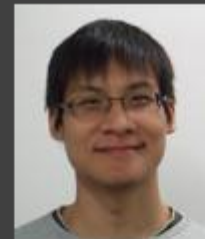
Prof. Haruki Nakamura  
Osaka Univ.



Prof. Kengo Kinoshita  
Tohoku Univ.



Dr. Robert Milewski



Huy Dinh



Ms. Joy Sarmiento



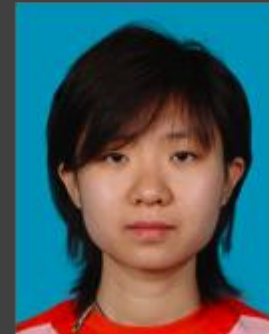
Dr. Osamu Takeuchi



Dr. Masahiro Yamamoto



Dr. Kazutaka Katoh  
Kyushu Univ.



Ms. Ai Han



Mieszko Lis  
MIT