4th Annual Forum for SMEs
Information Workshop on European Bioinformatics Resources
Munich, October 18th-19th 2010

Rafael Jimenez
rafael@ebi.ac.uk

ENFIN - EnVISION and DAS
ENFIN Network of Excellence

• Brings together experimentalists and computational biologists to develop the next generation of informatics resources for systems biology

• Funded by the European Commission within its FP6 programme under the thematic area ‘Life sciences, genomics and biotechnology for health’

• 20 partners in 13 countries

• www.enfin.org
EnCore

- ENFIN Platform to enable mining data across various domains, sources, formats and types
- Integrates database resources and analysis tools across different disciplines
**Diverse service world**

<table>
<thead>
<tr>
<th>External data sources</th>
<th>Access interfaces</th>
<th>SOAP, REST, Java API, Perl API, FTP, GUI, …</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Results</td>
<td>XML, CSV, Plain Text, JSON, …</td>
</tr>
</tbody>
</table>

- Multiple manual connections
- Multiple technologies
- Multiple result files which have to be combined manually
- Much work to reproduce
Standardized EnCORE world

External data sources

API, WS access

EnCORE services

Standard EnXML format

EnVISION pages

User

Heterogeneous external world

Standardised EnCORE world

EnXML
EnCORE services
From Inputs to Outputs

Protein
- Database IDs
- Sequences

Web Services

Results / EnXML

Positive

Negative

Input / EnXML

Protein identifications
- pride

Microarray probe mapping to Uniprot
- probe2uniprot

Protein Identifier Cross Reference Service
- picr

Biological Pathways
- reactome
- kegg

Microarray experiments
- arrayExpress

Protein sequence information
- uniprot
- uniprot2proteinAnnotations

Biological models
- biomodels

Gene ontology enrichment
- gGost

Cellular location
- cellmint

Protein domain analysis
- domaination

Protein function prediction
- funcnet

Molecular interactions
- uniprot2molecularInteractions

• Experiment: Identifies the result
• Sets: Contains the structure of the result
• Molecules: Includes the results
• Features: Describe details of the result
EnXML structure

- **Experiment**: Identifies the result
- **Set**: Contains the structure of the result
- **Molecule**: Includes the results
- **Feature**: Describe details of the result
# EnXML Schema

## enXML schema documentation

**schema location:** [http://www.enfin.org/encore/schema/enxml-v1_2_5.xsd](http://www.enfin.org/encore/schema/enxml-v1_2_5.xsd)

## Table of Contents

- Schema Document Properties
- Global Declarations
  - Element: `entries`
- Detailed Definitions
  - Complex Type: `aliasType`
  - Complex Type: `attributeType`
  - Complex Type: `biSourceType`
  - Complex Type: `enType`
  - Complex Type: `enReferenceType`
  - Complex Type: `entriesType`
  - Complex Type: `entryType`
  - Complex Type: `experimentType`
  - Complex Type: `featureType`
  - Complex Type: `moleculeType`
  - Complex Type: `nameType`
  - Complex Type: `parameterType`
  - Complex Type: `partSourceType`
  - Complex Type: `refType`
  - Complex Type: `sourceType`
  - Complex Type: `xrefType`

## Schema Document Properties

<table>
<thead>
<tr>
<th>Target Namespace</th>
<th><a href="http://ebi.ac.uk/enfin/core/model">http://ebi.ac.uk/enfin/core/model</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>1.2.5</td>
</tr>
</tbody>
</table>

### Element and Attribute Namespaces

- Global element and attribute declarations belong to this schema's target namespace.
- By default, local element declarations belong to this schema's target namespace.
- By default, local attribute declarations have no namespace.

## Declared Namespaces

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Namespace</th>
</tr>
</thead>
<tbody>
<tr>
<td>xml</td>
<td><a href="http://www.w3.org/XML/1998/namespace">http://www.w3.org/XML/1998/namespace</a></td>
</tr>
<tr>
<td>xs</td>
<td><a href="http://www.w3.org/2000/XMLSchema">http://www.w3.org/2000/XMLSchema</a></td>
</tr>
<tr>
<td>trs</td>
<td><a href="http://ebi.ac.uk/enfin/core/model">http://ebi.ac.uk/enfin/core/model</a></td>
</tr>
</tbody>
</table>

## Global Declarations

### Element: `entries`

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>trs:entriesType</td>
</tr>
<tr>
<td>Nullable</td>
<td>no</td>
</tr>
<tr>
<td>Abstract</td>
<td>no</td>
</tr>
<tr>
<td>Documentation</td>
<td>enxml root:element</td>
</tr>
</tbody>
</table>
EnCORE services
Example

Input/Query

EnCORE dataset

• Database ID (Uniprot ID)
  P37173

Program/Service

EnCORE webservice

• Encore webservice
  uniprot2molecularInteractions

Output/Results

EnCORE results

• Experiment: ID4
• Sets: (1) EBI-296235, (2) EBI-1033040, (3) EBI-902913, EBI-902937, (4) EBI-296166, EBI-296246, (5) EBI-902913
• Molecules: (1) O35613, (2) P10600, (3) P07200, (4) Q9UER7, (5) Q99K41
• Features: No features
# EnCore services

Example (Result on a table)

## Input/Query

**P37173**

## Program/Service

**Enfin-IntAct**

## Output/Results

<table>
<thead>
<tr>
<th></th>
<th>Interactor A</th>
<th>Interactor B</th>
<th>Interaction IDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P37173</td>
<td>O35613</td>
<td>EBI-296235</td>
</tr>
<tr>
<td>2</td>
<td>P37173</td>
<td>P10600</td>
<td>EBI-1033040</td>
</tr>
<tr>
<td>3</td>
<td>P37173</td>
<td>P07200</td>
<td>EBI-902913, EBI-902937</td>
</tr>
<tr>
<td>4</td>
<td>P37173</td>
<td>Q9UER7</td>
<td>EBI-296166, EBI-296246</td>
</tr>
<tr>
<td>5</td>
<td>P37173</td>
<td>Q99K41</td>
<td>EBI-902913</td>
</tr>
</tbody>
</table>
EnCore services
Building workflows

Input selection
Webservice
Result
Positive result
Negative result
EnCore Services
Taverna

Workflow Inputs:
- protein_id_list
- split_id_list
- utility_input_xml
- EnfinXmlFromProteinList
- utility_output_xml
- enXml-out

Workflow Outputs:
- picr_options_xml
- utility_input_id_list
- readme_options_xml

Split_string_into_string_list_by_regular_expression
- utility_inputs
- EnfinXmlFromProteinList
- utility_output_xml
- picr_inputs
- MapsToUniprot
- picr_output_xml
- intact_input_xml
- FindInteractionPartners
- intact_output_xml
- reactome_input_xml
- FindReactomePathways
- reactome_output_xml
- output
EnVision interface
Input Form

Query
P07200,Q99K41,P37173,P37023,Q13131,A3QNQ0,Q9Y6C2,P98170,A2AI38,Q8CGZ0,Q13287,Q8WTW2,P61812

Default workflow
EnVision2 interface

- Results for Pride, Uniprot, Intact, Reactome, CellMint, PICR, Biomodels, ...

Results per service

<table>
<thead>
<tr>
<th>Pride AC</th>
<th>Pride Experiment Name</th>
<th>Pubmed ID</th>
<th>Identified Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>3643</td>
<td>Upper extremity SCX fraction 04</td>
<td>Not Available</td>
<td>SHH HUMAN</td>
</tr>
<tr>
<td>342</td>
<td>Plasma Proteome (GPM10100000838)</td>
<td>Not Available</td>
<td>SHH HUMAN</td>
</tr>
<tr>
<td>8427</td>
<td>Human plasma dataset, Tao_PlasmNG-15_26Nov04_Firefly_0204-4</td>
<td>16335952, 16654767</td>
<td>SHH HUMAN</td>
</tr>
<tr>
<td>3671</td>
<td>Lower extremity SCX fraction 05</td>
<td>Not Available</td>
<td>SHH HUMAN</td>
</tr>
<tr>
<td>345</td>
<td>Plasma Proteome (GPM10100000841)</td>
<td>Not Available</td>
<td>TGFR2 HUMAN</td>
</tr>
<tr>
<td>8670</td>
<td>Human Plasma Proteome profile(YPRC)</td>
<td>Not Available</td>
<td>SHH HUMAN</td>
</tr>
<tr>
<td>3670</td>
<td>Lower extremity SCX fraction 04</td>
<td>Not Available</td>
<td>SHH HUMAN</td>
</tr>
<tr>
<td>8689</td>
<td>Human Plasma Proteome profile(YPRC)</td>
<td>Not Available</td>
<td>TGFR2 HUMAN</td>
</tr>
<tr>
<td>14</td>
<td>HUPO Plasma Proteome Project, Lab # 2 Expt # 18</td>
<td>15188391, 15061371, 15502245, 16052519, 16052621, 16052623, 18052624, 18104056, 18104057</td>
<td>SHH HUMAN</td>
</tr>
</tbody>
</table>

Download

To download this data, click the Download Button above. The data will be saved as a tab separated value file (*.tsv)

All your proteins in this dataset could be identified using Pride.
EnVision2 Pathways result

Positive results

Pathways with Proteins in dataset:

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Participating Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Scrolling by NFκ</td>
<td>23</td>
</tr>
<tr>
<td>2 Gene Expression</td>
<td>24</td>
</tr>
<tr>
<td>3 Cell Cycle, Mitotic</td>
<td>21</td>
</tr>
<tr>
<td>4 Scrolling in immune system</td>
<td>21</td>
</tr>
<tr>
<td>5 HIV infection</td>
<td>17</td>
</tr>
<tr>
<td>6 TRKA signalling from the plasma membrane</td>
<td>13</td>
</tr>
<tr>
<td>7 Hemostasis</td>
<td>16</td>
</tr>
<tr>
<td>8 Acon guidance</td>
<td>16</td>
</tr>
<tr>
<td>9 Diabetes pathways</td>
<td>14</td>
</tr>
<tr>
<td>10 Host Interactions of HIV factors</td>
<td>14</td>
</tr>
</tbody>
</table>

To see all proteins that are involved in at least one pathway just click [here](#).

Download

To download this data, click the Download Button above. The file is in a tab separated value file (*.tsv).

Set of proteins that were not found in any pathway:
- MARK4_HUMAN, CYTB_HUMAN, PANPA_HUMAN, AT2A2_HUMAN, FLNB_HUMAN, CDK3_HUMAN, M4K1_HUMAN, MELK_HUMAN, TAR81_HUMAN, CSK2B_HUMAN, PIG_TMY1_HUMAN, FBP_HUMAN, ALBU_PONAB, EF1A1_PONAB, PHL_HUMAN, GNAH2_HUMAN, SOAT1_HUMAN, FL23_HUMAN, ICYP4_HUMAN, NQO2_HUMAN, CYR1A_HUMAN, CLK1_HUMAN, R327L_HUMAN, CSK2B_MOUSE, FA368_HUMAN, AT2A1_HUMAN, FHH_BOVIN, C104_HUMAN, XPO6_HUMAN, GON1L_HUMAN, CYTB_HUMAN, ELPI1_HUMAN, ERD3_HUMAN, G0G1H1_HUMAN, C1104_HUMAN, MOUSE_CNDT_HUMAN, R512_HUMAN, CFK2R_RAT, HCL81_HUMAN, VESE3B_HUMAN, UB10_HUMAN, XENLA_HUMAN, TNPO2_HUMAN, RN213_HUMAN, MGN_HUMAN, RAT,
EnVision2, molecular interactions

**Molecular Interactions - Dataset: set3**

It provides molecular interaction data from different molecular interactions databases. It makes use of PSICQUIC.

### Table view

#### Pubmed ID selection

- **10821830**
- **11483955**
- **11850637**
- **12456677**
- **15267232**
- **15711537**
- **16189514**
- **16530041**
- **16539403**
- **17721511**
- **18480843**
- **18674809**
- **IM-11703**
- **IM-11825**
- **IM-11910**
- **unassigned384**

### Download results

- **Download ALL the interactions**
- **Download SELECTED interactions**

### Molecular interaction(s) with at least one Protein from the query set:

<table>
<thead>
<tr>
<th>Sel.</th>
<th>Interactor A</th>
<th>Interactor B</th>
<th>Interaction (Experiment) IDs</th>
<th>Interaction Database</th>
<th>Pubmed IDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>Q13131</td>
<td>Q9Y478</td>
<td>MINT-6771238</td>
<td>MINT</td>
<td>18480843</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EBI-1181432</td>
<td>INTACT</td>
<td>16306228</td>
</tr>
<tr>
<td>1</td>
<td>Q13131</td>
<td>CHEBI:49960</td>
<td>325079</td>
<td>CHEMBL</td>
<td>15711537</td>
</tr>
<tr>
<td>2</td>
<td>Q13131</td>
<td>CHEBI:38940</td>
<td>325079</td>
<td>CHEMBL</td>
<td>15711537</td>
</tr>
<tr>
<td>3</td>
<td>Q13131</td>
<td>CHEBI:104346</td>
<td>325079</td>
<td>CHEMBL</td>
<td>15711537</td>
</tr>
<tr>
<td>4</td>
<td>Q13131</td>
<td>CHEBI:15738</td>
<td>325079</td>
<td>CHEMBL</td>
<td>15711537</td>
</tr>
<tr>
<td>5</td>
<td>Q13131</td>
<td>P11021</td>
<td>EBI-1181432</td>
<td>INTACT</td>
<td>16306228</td>
</tr>
<tr>
<td>6</td>
<td>Q13131</td>
<td>P05141</td>
<td>EBI-1181432</td>
<td>INTACT</td>
<td>16306228</td>
</tr>
<tr>
<td>7</td>
<td>Q13131</td>
<td>Q16543</td>
<td>EBI-1181432</td>
<td>INTACT</td>
<td>16306228</td>
</tr>
<tr>
<td>8</td>
<td>Q13131</td>
<td>P07900</td>
<td>EBI-1181432</td>
<td>INTACT</td>
<td>16306228</td>
</tr>
<tr>
<td>9</td>
<td>Q13131</td>
<td>P35396</td>
<td>EBI-1809579</td>
<td>IM-11910</td>
<td></td>
</tr>
</tbody>
</table>

EnCore
Adapting EnCORE to Standards and Federation
Molecular Biology Database resources

- Human Genes and Diseases: 14%
- Proteomics Resources: 8%
- Nucleotide Sequence Databases: 9%
- RNA sequence databases: 2%
- Protein sequence databases: 10%
- Organelle databases: 2%
- Other Molecular Biology Databases: 3%
- Plant databases: 8%
- Immunological databases: 2%
- Genomics - Databases (non vertebrate): 9%
- Structure Databases: 9%
- Metabolic and Signaling Pathways: 8%

~1440 resources

Traditional EnCore approach

~1440 resources

Domain 1

Domain 2

Domain 3

Domain 4

Domain 5

Domain ...

Resources: 1440
New EnCore approach
Standards and Federation

Domain 1
External data sources
Federated systems / Standards
EnCORE wrapper
EnVISION pages

WS
WS
Web interface
New EnCore approach
Standards and Federation
New EnCore approach
Standards and Federation

- Less development
- More sources
- Data integration per domain
- Comparable results
- Automatic inclusion of new data sources
- More stable formats
- Validation
- Extra value to the original data
New role for EnCore and EnVision
Extra value to the original data

- Integration of sources
  - Clustering results
  - Data analysis
- Interconnect results
- More visualization
Centralization VS Federation

Centralized database

Federation

- Database
- GUI
- API
- WS

- DB
- SP

- GUI
- API
- WS

- Standard protocol
# Standards and Federation in EnCORE

<table>
<thead>
<tr>
<th>Protocol (Federated system)</th>
<th>Format</th>
<th>Domain</th>
<th>Subdomain</th>
<th>Some databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAS</td>
<td>DAS</td>
<td>Genome</td>
<td>Functional Genomics</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Structural genomics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Phenotype</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Epigenetics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAS</td>
<td>DAS</td>
<td>Protein sequence</td>
<td>Protein domains</td>
<td>UniProt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Protein motives</td>
<td>Pfam</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Protein function</td>
<td>Interpro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Protein predictions</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Protein identifications</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Protein variations</td>
<td></td>
</tr>
<tr>
<td>DAS</td>
<td>DAS - 3D Structure</td>
<td>Protein structure</td>
<td></td>
<td>Sanger 3DM</td>
</tr>
<tr>
<td>DAS</td>
<td>DAS - Molecular Interaction</td>
<td>Molecular interactions</td>
<td>Protein-Protein interactions</td>
<td>HPRD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Biodiverse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MINT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IntAct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MPIIDB</td>
</tr>
<tr>
<td>PSICQUIC</td>
<td>PSI-MI</td>
<td>Molecular interactions</td>
<td>Protein-Protein interactions</td>
<td>Intact</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>iRefIndex</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MINT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BioGrid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ChEMBL</td>
</tr>
<tr>
<td>BioPax exporter</td>
<td>BioPax</td>
<td>Pathways</td>
<td>Metabolic Signaling</td>
<td>Reactome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
DAS & PSICQUIC Implementation

Service Oriented Architecture

- Service broker
- Registry
  - Find
  - Publish
- Clients
- Service consumer
- Service provider
  - Protocol
  - Annotation sources
- Interact
DAS, how it works

## Available DAS Sources

**Available DAS Services**

<table>
<thead>
<tr>
<th>pos</th>
<th>id</th>
<th>clients</th>
<th>nickname</th>
<th>status</th>
<th>capabilities</th>
<th>coordinateSystem</th>
<th>description</th>
<th>project</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DS_409</td>
<td></td>
<td>uniprot</td>
<td>✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔esum, Protein Structure</td>
<td>go to site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DS_113</td>
<td></td>
<td>alig_pdb_sp</td>
<td>✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔esum, Protein Structure</td>
<td>go to site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DS_114</td>
<td></td>
<td>signalp</td>
<td>✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔esum, Protein Structure</td>
<td>go to site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>DS_115</td>
<td></td>
<td>netphos</td>
<td>✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔esum, Protein Structure</td>
<td>go to site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DS_116</td>
<td></td>
<td>neto glyc</td>
<td>✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔esum, Protein Structure</td>
<td>go to site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DS_117</td>
<td></td>
<td>tmhmm</td>
<td>✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔esum, Protein Structure</td>
<td>go to site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>DS_118</td>
<td></td>
<td>prop</td>
<td>✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔esum, Protein Structure</td>
<td>go to site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DS_119</td>
<td></td>
<td>transmem_pred</td>
<td>✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔esum, Protein Structure</td>
<td>go to site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>DS_120</td>
<td></td>
<td>superfamily</td>
<td>✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔️ ✔esum, Protein Structure</td>
<td>go to site</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EnCore DAS service for protein sequence annotations

- **Service:**
  - **Name:** uniprot2proteinannotations
  - **URL:** [http://www.ebi.ac.uk/enfin-srv/encore/uniprot2proteinannotations/service](http://www.ebi.ac.uk/enfin-srv/encore/uniprot2proteinannotations/service)
  - **Input:** List of Uniprot Acc numbers
  - **Options:** DAS Sources to query
    - Direct input (DAS feature URL) \([0, \ast]\)
    - Registry LABEL \([0, 1]\)
    - Registry source URI (DS_XXX) \([0, \ast]\)
PSICQUIC, how it works

User

PSICQUIC services

Interaction databases

Annotation error

Publications

Observation error

Sample
### PSICQUIC Registry

- **13 sources**
- **14,665,530 interactions**

<table>
<thead>
<tr>
<th>Name</th>
<th>Active</th>
<th>Interactions</th>
<th>Version</th>
<th>SOAP URL</th>
<th>REST URL</th>
<th>REST Example</th>
<th>Restricted</th>
<th>Tags</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>App</td>
<td>YES</td>
<td>416,124</td>
<td>1.1.0</td>
<td><a href="http://cicibio.de/dep.usc">http://cicibio.de/dep.usc</a></td>
<td><a href="http://cicibio.de/dep.usc">http://cicibio.de/dep.usc</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein imported interoperated</td>
<td></td>
</tr>
<tr>
<td>CHEMBL</td>
<td>YES</td>
<td>501,858</td>
<td>1.1.0</td>
<td><a href="http://www.ebi.ac.uk/Tosc">http://www.ebi.ac.uk/Tosc</a></td>
<td><a href="http://www.ebi.ac.uk/Tosc">http://www.ebi.ac.uk/Tosc</a></td>
<td>Example</td>
<td>NO</td>
<td>small molecule-protein internally curated</td>
<td></td>
</tr>
<tr>
<td>InnateDB</td>
<td>YES</td>
<td>9,369</td>
<td>1.1.5</td>
<td><a href="http://limex.innatedb.com">http://limex.innatedb.com</a></td>
<td><a href="http://limex.innatedb.com">http://limex.innatedb.com</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein internally curated</td>
<td></td>
</tr>
<tr>
<td>HPR</td>
<td>YES</td>
<td>20,769</td>
<td>1.1.6-snapshot</td>
<td><a href="http://limex.mbiula.edu">http://limex.mbiula.edu</a></td>
<td><a href="http://limex.mbiula.edu">http://limex.mbiula.edu</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein internally curated</td>
<td></td>
</tr>
<tr>
<td>IntAct</td>
<td>YES</td>
<td>228,262</td>
<td>1.1.6-snapshot</td>
<td><a href="http://www.ebi.ac.uk/Tosc">http://www.ebi.ac.uk/Tosc</a></td>
<td><a href="http://www.ebi.ac.uk/Tosc">http://www.ebi.ac.uk/Tosc</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein internally curated</td>
<td></td>
</tr>
<tr>
<td>RefIndex</td>
<td>YES</td>
<td>404,453</td>
<td>1.1.5</td>
<td><a href="http://biocn.sinoc16000lj">http://biocn.sinoc16000lj</a></td>
<td><a href="http://biocn.sinoc16000lj">http://biocn.sinoc16000lj</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein imported bipartite</td>
<td></td>
</tr>
<tr>
<td>MEDPDB</td>
<td>YES</td>
<td>24,286</td>
<td>1.1.5</td>
<td><a href="http://www.jcvi.org/medpdb">http://www.jcvi.org/medpdb</a></td>
<td><a href="http://www.jcvi.org/medpdb">http://www.jcvi.org/medpdb</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein internally curated</td>
<td></td>
</tr>
</tbody>
</table>

Total: **14,665,530 Interactions**
EnVision2, molecular interactions

PSICQUIC client

http://www.enfin.org/
## Clustering Scores

Molecular interaction(s) with at least one protein from the query set:

<table>
<thead>
<tr>
<th>Sel.</th>
<th>Interactor A</th>
<th>Interactor B</th>
<th>Interaction (Experiment) IDs</th>
<th>Interaction Database</th>
<th>Pubmed IDs</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>G9PER7</td>
<td>PC0537</td>
<td>EBI-1562310</td>
<td>INTACT</td>
<td>15364937</td>
<td>0.66561054</td>
</tr>
<tr>
<td>27</td>
<td>G9PER7</td>
<td>G005867</td>
<td>MNT-6801273</td>
<td>INTACT</td>
<td>15364937</td>
<td>0.63422064</td>
</tr>
<tr>
<td>60</td>
<td>P27173</td>
<td>F100300</td>
<td>IDB-120140</td>
<td>INNATEDB</td>
<td>15017145</td>
<td>0.6224111</td>
</tr>
<tr>
<td>61</td>
<td>P27173</td>
<td>P37173</td>
<td>IDB-120176</td>
<td>INNATEDB</td>
<td>15017145</td>
<td>0.6224111</td>
</tr>
<tr>
<td>66</td>
<td>P27173</td>
<td>TOFB3</td>
<td>IDB-120140</td>
<td>REACTOME</td>
<td>15017145</td>
<td>0.0224111</td>
</tr>
<tr>
<td>32</td>
<td>G9PER7</td>
<td>G93D09</td>
<td>EBI-992543</td>
<td>INTACT</td>
<td>18555590</td>
<td>0.61503214</td>
</tr>
<tr>
<td>23</td>
<td>G9PER7</td>
<td>F290500</td>
<td>EBI-1562610</td>
<td>INTACT</td>
<td>10033754</td>
<td>0.59273103</td>
</tr>
<tr>
<td>20</td>
<td>G9PER7</td>
<td>G995683</td>
<td>EBI-519032</td>
<td>INTACT</td>
<td>5743501</td>
<td>0.5566619</td>
</tr>
</tbody>
</table>
4 interactors displayed for Enquart: P07200, P2407736, P37173, Q9UER7, Q934539

Download ALL the interactions
**MI Score: Molecular interactions confidence score**

We extended the molecular interaction EnCore service to provide confidence scores. We calculated confidence scores using quantifiable parameters that can be consistently found across all the PSI-MI PSICQUIC sources. Confidence scores rely on all the information available for one interaction. Since PSICQUIC just provide binary scores, calculation of confidence scores require clustering binary molecular interactions (look at Mi cluster).

**Score formula**

The equation used for this confidence score has been discussed with molecular interaction domain experts and it can be found in: [http://docs.google.com/Doc?docid=0AG_p-HKWUjHoZGQ5c9NtcrmhfMjJ2ZDdwcDhmag&hl=en](http://docs.google.com/Doc?docid=0AG_p-HKWUjHoZGQ5c9NtcrmhfMjJ2ZDdwcDhmag&hl=en)

**Confidence Score Java API**

- miscore

**Related APIs**

- miclusterScore
- miscoreDistribution

EnCore APIs
Example: Score distribution across several databases
The enfin-core (EnCore) is the integration platform for the ENFIN European Network of Excellence. It aims to integrate an extensive list of database resources and analysis tools in a computationally accessible and extensible manner across different disciplines, facilitating automated data retrieval and processing with a special focus on systems biology. The EnCore platform is available as a collection of webservices with a common standard format (EnXML), easy to integrate in Workflow management software such as Taverna. Additionally EnCore services are also accessible thought EnVision, a web graphical user interface providing elaborated information such as molecular interaction, biological pathways and system biology models. Information about the ENFIN project and the EnCore platform is available in http://www.enfin.org.

For feedback, questions and requests please email: encore-help@ebi.ac.uk
The European Network of Excellence ENFIN is committed to provide a Europe-wide integration of computational approaches in systems biology.

The ENFIN Network runs four major platforms:

- A Joint Research Program covering the fields of Discrete Function Prediction, Network Reconstruction, Systems-Level Modeling
- A Provision of Analysis Tools - EnSUITE
- A Platform for Data Integration - EnCORE
- Training Courses and Workshops on Systems Biology

EnVision
Portal to the integration platform EnCORE

EnSUITE
Analysis tools provision

Documentation
From a wet laboratory perspective

“Best Practice” Guidelines
Computers in systems biology

Login
Username: ____________________________
Password: ____________________________
Login (Send me my password)

Next Event
ECCB10
September 26-29, 2010
Ghent, Belgium

ENFIN Newsletter
Subscribe via email
Newsletter feed
Thank you!

Questions?

ENFIN partners:
- Pascal Kahlem (project coordinator)
- Bernd Brandt (IBIVU)
- Christine Orengo (UCL)
- Andrew Clegg (UCL)
- Ioannis Xenarios (SIB)
- Heinz Stockinger (SIB)
- Jaak Vilo (QURETEC)
- Jüri Reimand (QURETEC)
- Gianni Cesareni (UNITOR)
- Arnaud Ceol (UNITOR)
- James Procter (UNIVDUN)
- Ana Rojas Mendoza (CNIO)