PSICQUIC
The PSI Common QUery Interface

Interesting facts
pronounced “psykick”, but also known as “pisquick”
spelled in 40 different ways (PSIQUIC, PSICIQK, QPSICUI…)
About me

- **Name:** Bruno Aranda
- **Affiliation:** European Bioinformatics Institute (EBI)
- **Role:** Software Engineer at the IntAct Team

- **IntAct** provides Molecular Interaction data through its open source services and tools.

- One of such tools is **PSICQUIC.**
INTRODUCTION

Take a deep breath
What is PSICQUIC?

Client

PSICQUIC services

Interaction databases

Curation

Publications

Observation error

Sample
What is PSICQUIC?

• Proteomics Standards Initiative Common QUery InterfaCe.

• Community effort to standardise the way to access and retrieve data from Molecular Interaction databases.

• PSICQUIC is a specification of a web service.

• Resources already implementing PSICQUIC are listed in a registry.

• Based on the PSI standard formats (PSI-MI XML and MITAB)

• Documentation: http://psicquic.googlecode.com
Justification

• So basically, if all interaction databases provide data in the same way, we can have…
“The one Client to rule them all”
More than 14 million binary interactions available using PSICQUIC
What can I do?

METHODS
Web Service Methods

- **getByInteraction**
  Retrieves interactions by using an interaction AC.

- **getByInteractionList**
  Retrieves interactions by using a list of interaction AC.

- **getByInteractor**
  Retrieves interactions by using a participant identifier.

- **getByInteractorList**
  Retrieves interactions by using a list of participant identifiers.

- **getByQuery**
  Retrieves interactions by using a Molecular Interaction Query Language (MIQL) query (full text searches)
Web Service Methods

Other metadata methods:

- **getVersion**
  Returns the version of the web service implementation.

- **getSupportedDbAcs**
  Returns the supported database identifiers

- **getSupportedReturnTypes**
  Returns the list of available format types for the results.

- A limited number of interactions can be fetched. It is possible to retrieve large datasets using **pagination**. Most methods have two additional parameters:
  - *First result*: Index for the first result to retrieve.
  - *Max results*: Number of interactions returned per query.
Take a shower before going to sleep?

SOAP AND REST (PROTOCOLS)
How can I access PSICQUIC?

As PSICQUIC is a Web Service, you can access the data:

• **Via SOAP**
  - A WSDL file exists, and it is the same for all the databases.
  - IntAct has developed a Java client, but any other languages can be used.
  - You can use it to get interactions in two standard formats: PSI-MI XML and PSI-MI TAB.

• **Via REST**
  - Retrieving data directly by using a URL
  - Easy to access and data can be obtained just using an internet browser.
  - Effective for scripting.
SoapUI: Executing a query

Use query brca2 and resultType psi-mi/tab25
PSICQUIC: REST query

http://www.ebi.ac.uk/Tools/webservices/psicquic/intact/webservices/current/search/query/species:rat
Standards and more

FORMATS
Default formats

• The default formats are:
  • PSI-MI XML 2.5.4 (psi-mi/xml25)
  • PSI MITAB 2.5 (psi-mi/tab25)
  • Compressed MITAB (tab25-bin) – Only REST
  • Count (count) – Only REST

• New formats will be included in the future (work in progress):
  • BioPAX (biopax)
  • And other RDF formats (rdf-xml / rdf-n3 / rdf-n3-triple / rdf-turtle)

• (so it will be possible to use PSICQUIC in the semantic web!)
<entrySet minorVersion="4" version="5" level="2">
  <entry>
    <source releaseDate="2009-04-17T01:00">
      <names>
        <shortLabel>European Bioinformatics Institute</shortLabel>
      </names>
      <attributeList>
        <attribute name="postalAddress">
          Welcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom
        </attribute>
      </attributeList>
    </source>
    <interactionList id="201">
      <interaction id="201">
        <names>
          <shortLabel>P51587-EBI-539895-P51587-Q06609-1</shortLabel>
        </names>
        <primaryRef refType="identity" id="EBI-297231" db="intact"/>
        <experimentList id="202">
          <experimentDescription id="202">
            psi-mi/xml25
          </experimentDescription>
        </experimentList>
      </interaction>
    </interactionList>
  </entry>
</entrySet>
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:bp="http://www.biopax.org/release/biopax-level3.owl#"
 xmlns:owlmi="http://purl.org/obo/owl/MI#"
 xmlns:owl="http://www.w3.org/2002/07/owl#"
 xmlns:psimi="http://www.ebi.ac.uk/~intact/psimi.owl#"
 xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
 xmlns="http://www.ebi.ac.uk/intact/
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:base="http://www.ebi.ac.uk/intact/"
 <owl:Ontology rdf:about=""
 </owl:Ontology>
 <bp:MolecularInteraction rdf:about="http://purl.uniprot.org/intact/EBI-1639937">
 <bp:participant>
  <bp:Protein rdf:about="http://purl.uniprot.org/uniprot/Q9EXW9">
   <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Q9EXW9
   </rdfs:label>
  </bp:Protein>
 </bp:participant>
 <bp:participant>
  <bp:Protein rdf:about="http://purl.uniprot.org/uniprot/P51587">
   <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    P51587
   </rdfs:label>
  </bp:Protein>
 </bp:participant>
 <bp:interactorType>
  <bp:InteractionVocabulary rdf:about="http://purl.obo/owl/MI MI_0915">
   <bp:name>physical association</bp:name>
   <rdfs:label>physical association</rdfs:label>
  </bp:InteractionVocabulary>
 </bp:interactorType>
 <bp:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
  F51587-Q9EXW9-1
 </bp:name>
 <bp:MolecularInteraction rdf:about="http://purl.uniprot.org/intact/EBI-260802">
 <bp:participant>
  <bp:Protein rdf:about="http://purl.uniprot.org/uniprot/Q9W157">
   <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Q9W157
   </rdfs:label>
  </bp:Protein>
 </bp:participant>
</bp:MolecularInteraction>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:bp="http://www.biopax.org/release/biopax-level3.owl#"
  xmlns:owlmi="http://purl.org/obo/owl/MI#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:psimi="http://www.ebi.ac.uk/~intact/psimi.owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns="http://www.ebi.ac.uk/intact/
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.ebi.ac.uk/intact/"
>
  <rdf:Description rdf:about="http://purl.uniprot.org/intact/EBI-1639937">
    <bp:participant rdf:resource="http://purl.uniprot.org/uniprot/Q9BXW9"/>
    <bp:participant rdf:resource="http://purl.uniprot.org/uniprot/P51587"/>
    <bp:interactorType rdf:resource="http://purl.org/obo/owl/MI#NI_0915"/>
    <bp:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PS1587-Q9BXW9-1</bp:name>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">P51587-Q9BXW9-1</rdfs:label>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">MolecularInteraction</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.uniprot.org/intact/EBI-260802">
    <bp:participant rdf:resource="http://purl.uniprot.org/uniprot/077135"/>
    <bp:interactorType rdf:resource="http://purl.org/obo/owl/MI#NI_0915"/>
    <bp:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">077135-Q9W157-1</bp:name>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">MolecularInteraction</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.uniprot.org/intact/EBI-2461712">
    <bp:participant rdf:resource="http://purl.uniprot.org/uniprot/P51587"/>
    <bp:participant rdf:resource="http://purl.uniprot.org/uniprot/094776"/>
    <bp:interactorType rdf:resource="http://purl.org/obo/owl/MI#NI_0914"/>
    <bp:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">094776-P51587-1</bp:name>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">094776-P51587-1</rdfs:label>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">MolecularInteraction</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.uniprot.org/intact/EBI-930908">
    <bp:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">P94102</bp:name>
    <bp:interactorType rdf:resource="http://purl.org/obo/owl/MI#NI_0915"/>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">P94102-081303-1</rdfs:label>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">MolecularInteraction</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.uniprot.org/intact/EBI-077135">
    <bp:participant rdf:resource="http://purl.uniprot.org/uniprot/077135"/>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">077135</rdfs:label>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Protein</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.uniprot.org/intact/EBI-930908">
    <bp:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">P94102</bp:name>
    <bp:interactorType rdf:resource="http://purl.org/obo/owl/MI#NI_0915"/>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">P94102-081303-1</rdfs:label>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">MolecularInteraction</rdfs:label>
  </rdf:Description>
</rdf:RDF>
Where are the services?

THE REGISTRY
The PSICQUIC Registry

• It contains a list of the PSICQUIC services from different providers.

• It is a web service itself, and it can be accessed remotely using REST.

• Information can be found about the services, such as the URLs to use, number of interactions provided, versioning, etc.

• The Registry can be found at:
http://www.ebi.ac.uk/Tools/webservices/psicquic/registry/registry?action=STATUS
# PSICQUIC Registry

<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>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>YES</td>
<td>416,124</td>
<td>1.1.5</td>
<td><a href="http://cicblade.dep.usal.es">http://cicblade.dep.usal.es</a></td>
<td><a href="http://cicblade.dep.usal.es">http://cicblade.dep.usal.es</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein imported spoke clustered</td>
<td></td>
</tr>
<tr>
<td>ChEMBL</td>
<td>YES</td>
<td>581,858</td>
<td>1.1.0</td>
<td><a href="http://www.ebi.ac.uk/Tools">http://www.ebi.ac.uk/Tools</a></td>
<td><a href="http://www.ebi.ac.uk/Tools">http://www.ebi.ac.uk/Tools</a></td>
<td>Example</td>
<td>NO</td>
<td>smallmolecule-protein internally curated mimix curation spoke evidence</td>
<td></td>
</tr>
<tr>
<td>InnateDB</td>
<td>YES</td>
<td>9,909</td>
<td>1.1.5</td>
<td><a href="http://imex.innatedb.com/">http://imex.innatedb.com/</a></td>
<td><a href="http://imex.innatedb.com/">http://imex.innatedb.com/</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein internally curated spoke rapid curation evidence</td>
<td></td>
</tr>
<tr>
<td>DIP</td>
<td>YES</td>
<td>20,769</td>
<td>1.1.6-SNAPSHOT</td>
<td><a href="http://imex.mbi.ucla.edu/p">http://imex.mbi.ucla.edu/p</a></td>
<td><a href="http://imex.mbi.ucla.edu/p">http://imex.mbi.ucla.edu/p</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein internally curated mimix curation spoke evidence</td>
<td></td>
</tr>
<tr>
<td>BinGrid</td>
<td>NO</td>
<td>337,957</td>
<td>1.1.6-SNAPSHOT</td>
<td><a href="http://tyerslab.bio.ed.ac.uk">http://tyerslab.bio.ed.ac.uk</a></td>
<td><a href="http://tyerslab.bio.ed.ac.uk">http://tyerslab.bio.ed.ac.uk</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein internally curated rapid curation spoke evidence</td>
<td></td>
</tr>
<tr>
<td>MINT</td>
<td>YES</td>
<td>124,473</td>
<td>1.1.5</td>
<td><a href="http://mnt.bio.uniroma2.it">http://mnt.bio.uniroma2.it</a></td>
<td><a href="http://mnt.bio.uniroma2.it">http://mnt.bio.uniroma2.it</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein internally curated mimix curation spoke evidence</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/Tools">http://www.ebi.ac.uk/Tools</a></td>
<td><a href="http://www.ebi.ac.uk/Tools">http://www.ebi.ac.uk/Tools</a></td>
<td>Example</td>
<td>NO</td>
<td>protein-protein smallmolecule-protein nucleicacid-protein</td>
<td></td>
</tr>
</tbody>
</table>

More than 14,000,000 binary interactions available from 13 different sources

**PSICQUIC: Registry**
http://www.ebi.ac.uk/Tools/webservices/psicquic/registry/registry?action=STATUS
Registry Tagging system

• The registry classifies the different services with tags.
Querying the registry

• The registry can be accessed with the browser or programmatically (it is a **web service**).
• Instructions on how to use it can be found here:
• Querying by tags is work in progress at the moment (it will be explained tomorrow).
Examples

REAL APPLICATIONS
PSICQUIC Applications

• It is clear the value of PSICQUIC to application developers, so indirectly the end-user is benefited too.

• Reduces the time to implement an application that uses data from the different provides, as all of them are accessed the same way.

• Some of the applications:
  o Cytoscape 2.7.x
  o PSICQUIC View
  o Envision2
  o PSICQUIC Client for Android
  o GMOD client?
<table>
<thead>
<tr>
<th>Name molecule A</th>
<th>Links molecule A</th>
<th>Name molecule B</th>
<th>Links molecule B</th>
<th>Alt. identifiers molecule A</th>
<th>Alt. identifiers molecule B</th>
<th>Aliases molecule A</th>
<th>Aliases molecule B</th>
<th>Species molecule A</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>P51587:EBI-79792</td>
<td>UniProt</td>
<td>Q9BXW9-2:EBI-598678</td>
<td>UniProt</td>
<td>FANCD1; FACC; Fanconi anemia group D1 protein; brc2a_human</td>
<td>Q9BXW9-2</td>
<td>BRCA2</td>
<td></td>
<td>Human</td>
<td>9606</td>
</tr>
<tr>
<td>P51587:EBI-79792</td>
<td>UniProt</td>
<td>Q9BXW9-2:EBI-598678</td>
<td>UniProt</td>
<td>FANCD1; FACC; Fanconi anemia group D1 protein; brc2a_human</td>
<td>Q9BXW9-2</td>
<td>BRCA2</td>
<td></td>
<td>Human</td>
<td>9606</td>
</tr>
<tr>
<td>P51587:EBI-79792</td>
<td>UniProt</td>
<td>Q9BXW9-2:EBI-598678</td>
<td>UniProt</td>
<td>FANCD1; FACC; Fanconi anemia group D1 protein; brc2a_human</td>
<td>Q9BXW9-2</td>
<td>BRCA2</td>
<td></td>
<td>Human</td>
<td>9606</td>
</tr>
<tr>
<td>P51587:EBI-79792</td>
<td>UniProt</td>
<td>Q9BXW9-2:EBI-598678</td>
<td>UniProt</td>
<td>FANCD1; FACC; Fanconi anemia group D1 protein; brc2a_human</td>
<td>Q9BXW9-2</td>
<td>BRCA2</td>
<td></td>
<td>Human</td>
<td>9606</td>
</tr>
<tr>
<td>CHEBI:15422:EBI-1108845</td>
<td>UniProt</td>
<td>CHEBI:15422:EBI-1108845</td>
<td>UniProt</td>
<td>FANCD1; FACC; Fanconi anemia group D1 protein; brc2a_human</td>
<td>CHEBI:15422:EBI-1108845</td>
<td>BRCA2</td>
<td></td>
<td>Human</td>
<td>9606</td>
</tr>
<tr>
<td>CHEBI:15422:EBI-1108845</td>
<td>UniProt</td>
<td>CHEBI:15422:EBI-1108845</td>
<td>UniProt</td>
<td>FANCD1; FACC; Fanconi anemia group D1 protein; brc2a_human</td>
<td>CHEBI:15422:EBI-1108845</td>
<td>BRCA2</td>
<td></td>
<td>Human</td>
<td>9606</td>
</tr>
<tr>
<td>CHEBI:15422:EBI-1108845</td>
<td>UniProt</td>
<td>CHEBI:15422:EBI-1108845</td>
<td>UniProt</td>
<td>FANCD1; FACC; Fanconi anemia group D1 protein; brc2a_human</td>
<td>CHEBI:15422:EBI-1108845</td>
<td>BRCA2</td>
<td></td>
<td>Human</td>
<td>9606</td>
</tr>
<tr>
<td>CHEBI:15422:EBI-1108845</td>
<td>UniProt</td>
<td>CHEBI:15422:EBI-1108845</td>
<td>UniProt</td>
<td>FANCD1; FACC; Fanconi anemia group D1 protein; brc2a_human</td>
<td>CHEBI:15422:EBI-1108845</td>
<td>BRCA2</td>
<td></td>
<td>Human</td>
<td>9606</td>
</tr>
<tr>
<td>Q9FL96:EBI-931034</td>
<td>UniProt</td>
<td>Q7Y1C4:EBI-930753</td>
<td>UniProt</td>
<td>At5g45010; K21C13.20; sem12_arath</td>
<td>At5g01630; q7v1c4_arath</td>
<td>brc2a</td>
<td></td>
<td>Mouse-ear cress</td>
<td>3702</td>
</tr>
<tr>
<td>Q9XR8:EBI-931045</td>
<td>UniProt</td>
<td>Q7Y1C4:EBI-930753</td>
<td>UniProt</td>
<td>At5g4750; F13011.6; sem11_arath</td>
<td>At5g01630; q7v1c4_arath</td>
<td>brc2a</td>
<td></td>
<td>Mouse-ear cress</td>
<td>3702</td>
</tr>
<tr>
<td>Q9XR8:EBI-931045</td>
<td>UniProt</td>
<td>Q7Y1C4:EBI-930753</td>
<td>UniProt</td>
<td>At5g4750; F13011.6; sem11_arath</td>
<td>At5g01630; q7v1c4_arath</td>
<td>brc2a</td>
<td></td>
<td>Mouse-ear cress</td>
<td>3702</td>
</tr>
<tr>
<td>Q9FL96:EBI-931034</td>
<td>UniProt</td>
<td>Q7Y1C4:EBI-930753</td>
<td>UniProt</td>
<td>At5g45010; K21C13.20; sem12_arath</td>
<td>At5g01630; q7v1c4_arath</td>
<td>brc2a</td>
<td></td>
<td>Mouse-ear cress</td>
<td>3702</td>
</tr>
</tbody>
</table>

http://www.ebi.ac.uk/Tools/webservices/psicquic/view/
Envision2
As an example of PSICQUIC integration
Powerful queries with the Molecular Interaction Query Language

MIQL
Common Query Language

• The Molecular Interactions Query Language (MIQL) allows more powerful and flexible queries.
• It is the default query syntax for PSIQCUCI.
• Designed for fast and effective searches on PSI-MI TAB files.
• All fields (columns) can be searched with specific queries.
• MIQL is a consensus between the different databases, so you should be able to use the same query across different repositories.
miql syntax reference

The MIQL syntax is based on the Lucene syntax[1]. A query is broken into terms and operators:

- **Terms**: single words or phrases (group of words surrounded by quotes). E.g. `brca2 AND "pull down"
- **Fields**: search in specific columns. E.g. `brca2 AND species:human`
- **Term modifiers**: wildcard searches, fuzzy searches, proximity and range searches. E.g. `brca*`
- **Operands**: OR (or space), AND, NOT, +, -. E.g. `brca2 AND rpa1 / brca2 NOT mouse / +brca2 –mouse –expansion:spoke`
- **Grouping and field grouping**: `brca2 AND (mouse "in vitro")`

How to Create Your Service

Bruno Aranda (baranda@ebi.ac.uk)

GMOD Meeting
13th – 16th September 2010
Cambridge, UK
Simplest recipe to implement PSICQUIC

- **Ingredients:**
  - PSI-MITAB compliant file.
  - Subversion: to get the source code.
  - Maven: to run the scripts and start the service.

- **Steps:**
  - Generate the MITAB compliant file.
  - Get the Reference Implementation (RI):
  - Run the script to index the file.
  - Start the service with the script provided.
Lots of possibilities

CURRENT AND FUTURE WORK
Future developments

• Smart PSICQUICs: Identification and removal of redundancy
  • Merger and Cluster PSICQUIC services
• PSICQUIC 2.0
  • Overcome the current limitations and many fancy features:
    • Queries using CV terms not possible in the reference implementation (it is possible in IntAct).
    • PSI-MI XML is created from the MITAB, so no n-nary interactions.
  • New features:
    • Redundancy detection mechanism. ROG/RIG ids by default.
    • Built from PSI-MI XML, so complex data available.
• A GMOD component?

Bruno Aranda (baranda@ebi.ac.uk)
GMOD Meeting
13th – 16th September 2010
Cambridge, UK
IntAct is funded by the European Commission under SLING, grant agreement number 226073 (Integrating Activity) within Research Infrastructures of the FP7, under PSIMEx, contract number FP7-HEALTH-2007-223411 and under APO-SYS, contract number FP7-HEALTH-2007-200767.