
"Large-Scale Information Extraction for Biomedical Modelling and Simulation"



Martin Hofmann-Apitius
Head of the Department of Bioinformatics
Fraunhofer Institute for Algorithms and Scientific
Computing (SCAI)

Fraunhofer Society



- Founded 1949
- Europe´s largest applied research organisation
- 60 Research Institutes
- 17.000 Employees
- Annual Budget about 1,5 Billion Euro
- Financial model: 40 % industry collaborations
40 % public funding
20 % institutional funding

Seite 2

*Joseph von Fraunhofer (1787 – 1826)
Scientist, Inventor and Entrepreneur

Fraunhofer-Campus Schloss Birlinghoven



Institutes

- Algorithms and Scientific Computing SCAI
- Intelligent Analysis and Information Systems IAIS
- Applied Information Technology FIT

600 Scientists, 200 Students

Linked to Universities Bonn, Aachen
and Cologne



Fraunhofer: Applied Research for Industrial Applications

Fraunhofer stands for:

- sustainable (applied) research
- focus on contract research and innovation
- bridging between excellent academic research and industrial application
- clear mission towards improving and fostering innovation
- research done with the idea in mind to generate added value in a commercial sense

SCAI Department of Bioinformatics: R&D in a nutshell

Fraunhofer SCAI Department of Bioinformatics R&D activities:

1. Information extraction in the **life sciences**:
 - I. Text Mining - Recognition of named entities & relationships in text
 - II. Image Mining - Reconstruction of chemical information from chemical structure depictions
2. Disease modelling (focus on neurodegenerative diseases)
3. eScience, Grid-/Cloud- Computing and HPC (Clus^t)

Seite 5

Making Scientific Content
available for Computing

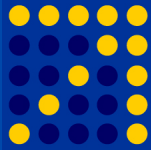
People

Fraunhofer SCAI Department of Bioinformatics currently comprises:

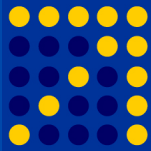
- 10 scientists
- 2 scientific software developers
- 6 PhD students
- > 6 Master students

- predominantly computer scientists & biologists
- additional scientists and PhD students via University of Bonn

The History of High Performance Computing in BioMedicine at SCAI: Once upon a time



- WISDOM stands for Wide In Silico Docking on Malaria
 - Goals
 - Computational goals
 - To show the relevance of computational grids in biomedical applications
 - Biological goals
 - Risk of Malaria, Post genomic era, number of targets, Finding new compounds in a cost effective way
 - Method establishment
 - Establishing virtual screening technology on computational grids



- WISDOM-I: First large scale docking against plasmepsin on computational grids
 - Achievements:
 - 80 CPU years in 45 days (Computational side)
 - Identified three novel scaffolds (biological side)
 - Docking on grids (Method establishment)
- WISDOM-II: Second assault on 4 different proteins (DHFR; Tubulin; two different Plasmepsins) implicated in malaria
 - Achievements:
 - 413 CPU years in 90 days (Computational side)
 - Result analysis is in progress (biological side)
 - Docking on grids (Method improved)

VS Explorer: a Tool for Analyzing “Grid Scale” Ranking Lists

The image displays the VS Explorer software interface, which is used for analyzing grid-scale ranking lists. The main window shows a large grid of data with columns for 'Number', 'SMILES', 'name', and multiple 'scenario' columns. A black circle highlights a specific row in the main grid, and an arrow points to a detailed view of that row in the 'Focus' window.

Main Window Data (Selected Row):

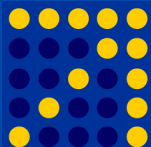
| Number | SMILES | name | scenario1 | scenario2 | scenario3 | scenario4 | scenario5 | scenario6 | scenario7 | scenario8 | scenario9 | scenario10 |
|--------|--|--------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| 25 | <chem>C1=CC=C(C=C1)C2=CC=CC=C2C3=CC=CC=C3</chem> | ZINC00603011 | -28.92 | -29.88 | -28.66 | -28.08 | -27.14 | -28.66 | -28.08 | -28.91 | -28.92 | -29.88 |

Focus Window Data (Selected Row):

| Number | SMILES | name | scenario1 | scenario2 | scenario3 | scenario4 | scenario5 | scenario6 | scenario7 | scenario8 | scenario9 | scenario10 |
|--------|--|--------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| 28 | <chem>C1=CC=C(C=C1)C2=CC=CC=C2C3=CC=CC=C3</chem> | ZINC00607811 | +00.01 | +00.01 | +00.01 | +00.01 | +00.01 | +00.01 | +00.01 | +00.01 | +00.01 | +00.01 |

Other Focus Window Data:

| Number | SMILES | name | scenario1 | scenario2 | scenario3 | scenario4 | scenario5 | scenario6 | scenario7 | scenario8 | scenario9 | scenario10 |
|--------|---------------------------------------|-------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| 62 | <chem>C1=CC=C(C=C1)C2=CC=CC=C2</chem> | ZINC0062... | +00.01 | -14.24 | +00.01 | | | | | | | |
| 63 | <chem>C1=CC=C(C=C1)C2=CC=CC=C2</chem> | ZINC0062... | +00.01 | -15.52 | +00.01 | | | | | | | |
| 64 | <chem>C1=CC=C(C=C1)C2=CC=CC=C2</chem> | ZINC0062... | -35.64 | -37.31 | -37.16 | | | | | | | |
| 398 | <chem>C1=CC=C(C=C1)C2=CC=CC=C2</chem> | 1abe_ara | -13.80 | -13.64 | -13.55 | -14.66 | -13.55 | -13.55 | -14.63 | -13.80 | -13.80 | -13.64 |
| 399 | <chem>C1=CC=C(C=C1)C2=CC=CC=C2</chem> | 2cpp_min | -6.48 | -6.55 | -6.27 | -6.55 | -7.04 | -7.04 | -6.34 | -7.04 | -7.04 | -6.51 |
| 400 | <chem>C1=CC=C(C=C1)C2=CC=CC=C2</chem> | 1tmn | -18.78 | -18.10 | -17.50 | -19.67 | -16.91 | -16.91 | -19.67 | -19.34 | -20.34 | -17.95 |



Achievements

Completed the screening of 4.3 million compounds against four different targets

Identified more than 200 new “interesting chemical scaffolds” that have been tested in “wet lab assays”

Validated biological activity of 13 novel “virtual screening hits”

Submitted patents for 4 new compounds / scaffolds

Lessons Learned from WISDOM and other Workflows

1. Infrastructures such as EGEE (EGI) can be productively used for *in silico experimentation* in biomedicine
2. Post-Processing (searching where the data have been stored; which jobs have been successfully finished) of generated data (docking results) took much more time than simulating the ligand-target interaction
3. Support of complex workflow design and workflow optimisation in compute infrastructures is still a challenge



Scientific Workflow Optimization on the Grid

Sonja Holl

*Jülich Supercomputing Centre (JSC)
Institute for Advanced Simulation (IAS)
Forschungszentrum Jülich, Germany*



Motivation

- ▶ Scientific workflows have many parameters
- ▶ Optimal settings not known a priori and data dependent
- ▶ Often several methods for the same task are available
- ▶ Parameters have constraints and non-linear dependencies
- ▶ Testing many parameter settings requires HPC
- ▶ Currently try and error based optimization, lack of tools

- ▶ Our primary goal: Optimizing the *quality* of a scientific workflow

Workflow Optimization

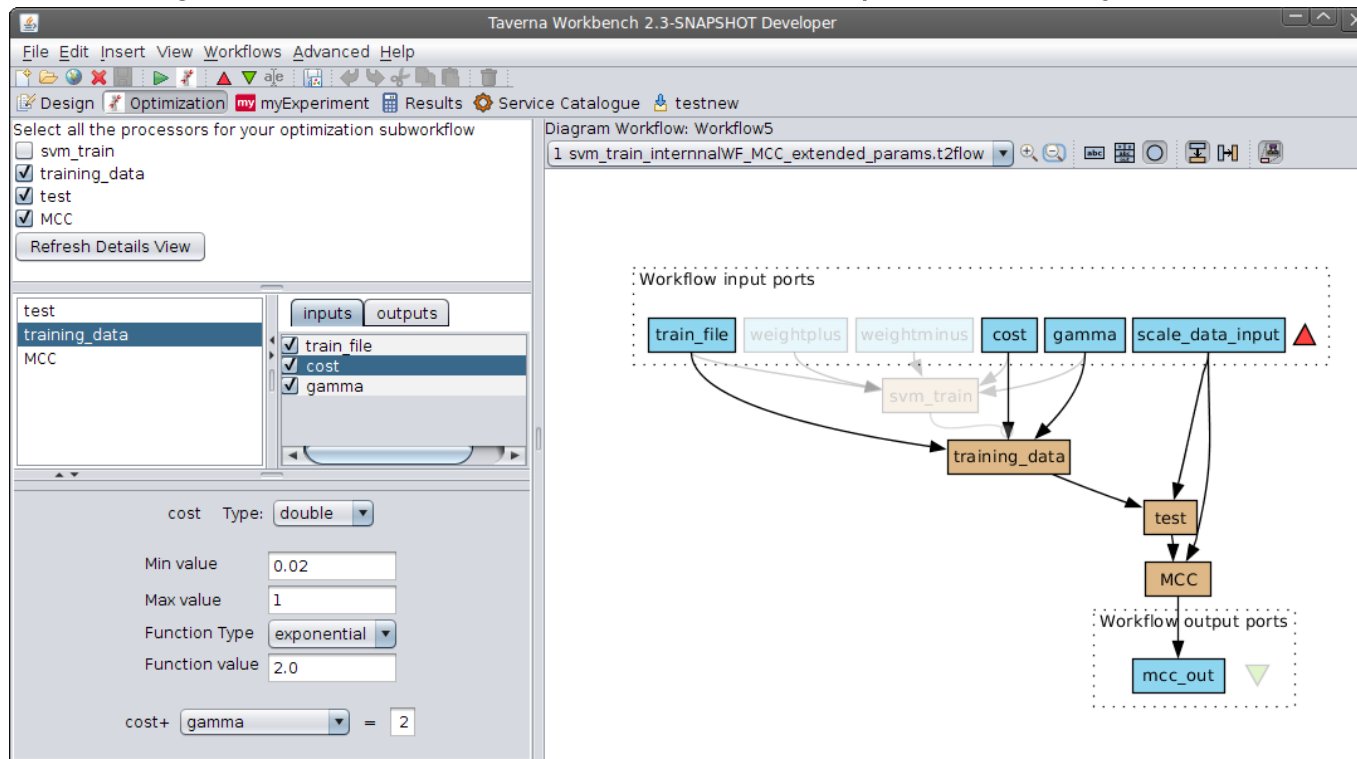
Find workflow settings that optimize the output quality

- ▶ Identify user requirements (Parameter, Constraints)
- ▶ Optimize relevant parts of a workflow
- ▶ Parameter optimization, Component set optimization
- ▶ Measure output quality (Fitness Function, User Feedback)

- ▶ Use the Grid to speed up optimization
(UNICORE Grid middleware)

The UNICORE Taverna Optimization Plugin

- ▶ Easy, GUI-based setup of WF and optimization tasks
- ▶ Plugin enables optimization within Taverna WFMS
- ▶ Parallel execution of WF optimization runs via UNICORE
- ▶ Link to major biomedical in silico labs (caBIG; myExperiment)



Workflow Optimization – Techniques 1/2

*Find optimal **input parameters** for WF applications*

- ▶ Varying free application parameters
- ▶ Realized via evolutionary algorithm (JGAP-Framework)

*Advanced **Parameter Optimization***

- ▶ Dependent parameters
 - Discrete set of parameter values (files, flags, integer)
 - Dependencies: Fixed combinations, Mathematical dependency, Logical dependency

Workflow Optimization – Techniques 2/2

Component Set Optimization:

Replace tool or algorithm with an equivalent one

- ▶ What is "equivalent"?
- ▶ Ontologies for tasks, inputs, and outputs required

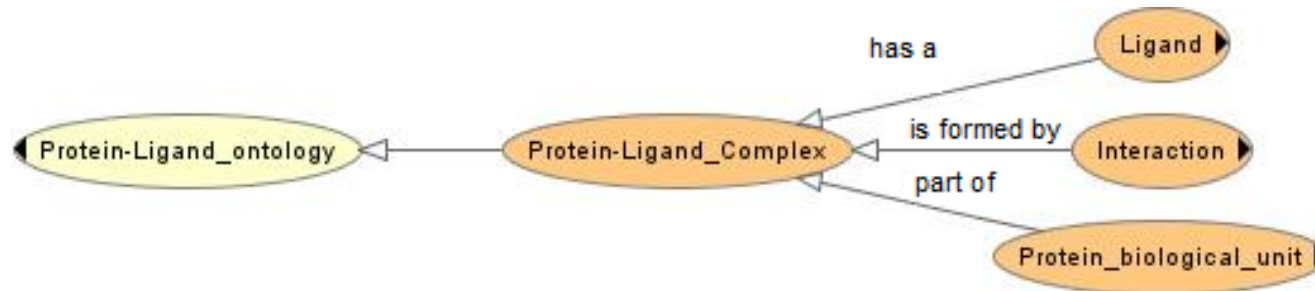
Future work

- ▶ Workflow structure optimization
- ▶ Adaptive optimization – learning from community runs
- ▶ Supervised optimization - User feedback as fitness function

Linking Methods and Algorithms to Scientific Objectives and Knowledge

The PLIO Ontology

PLIO root concepts (total concepts: 371)



- PLIO concept coverage
 - **Biophysical forces**, e.g. Van der Waals, electrostatics, etc.
 - **Interaction descriptors** e.g. pharmacophore, interaction fingerprint, etc.
 - **Experimental techniques**, e.g. NMR, X-ray, etc.
 - **Prediction and simulation methods**, e.g. MD, docking, etc.
 - **Classification of ligand activities**, e.g. biological activity, binding activity, etc.
 - **Classification of ligand modes of action**, e.g. agonist, inhibitor, etc.
 - **Classification of binding sites**, e.g. allosteric site, orthosteric site, etc.
 - **Structure-Activity-Relationships**, e.g. QSAR, COMFA, etc.

Each Concept has:

Reference

"<http://www.shodor.org/chemviz/glossary.html>"

Synonym

"activation energy
Energy of activation
Arrhenius activation energy
activation barrier
activation barriers
E(a)"

Formula

" $k=A\exp(-E_a/RT)$ "

Ea - activation energy (Arrhenius activation energy)
k - rate constant
A - the pre-exponential factor or simply the prefactor
R - the gas constant
T - temperature"

software

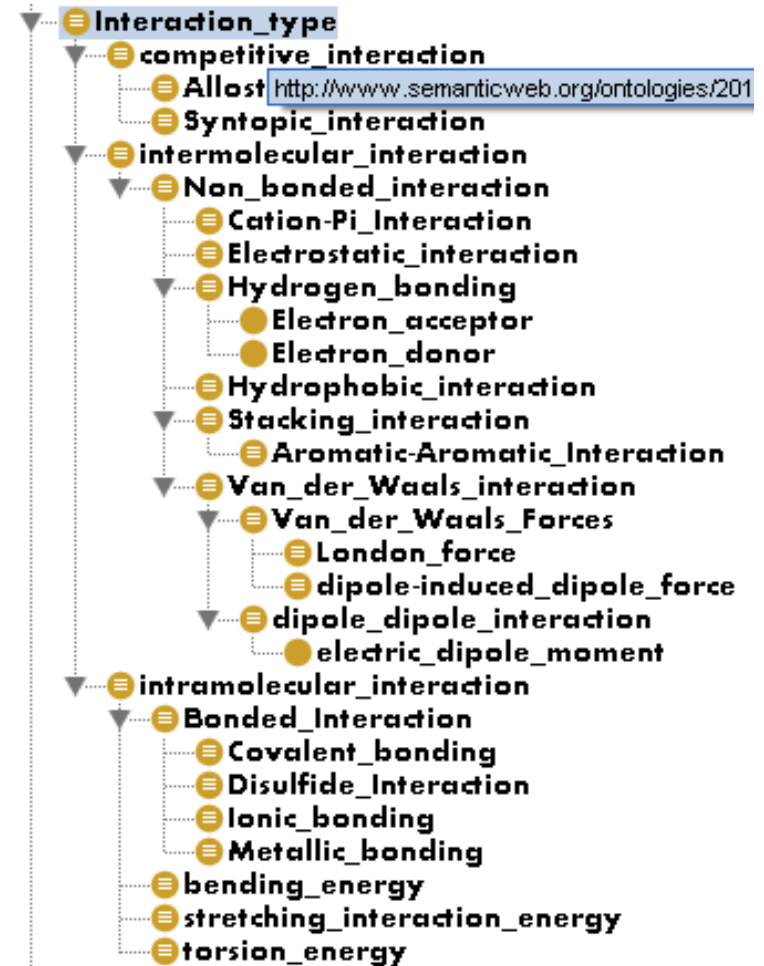
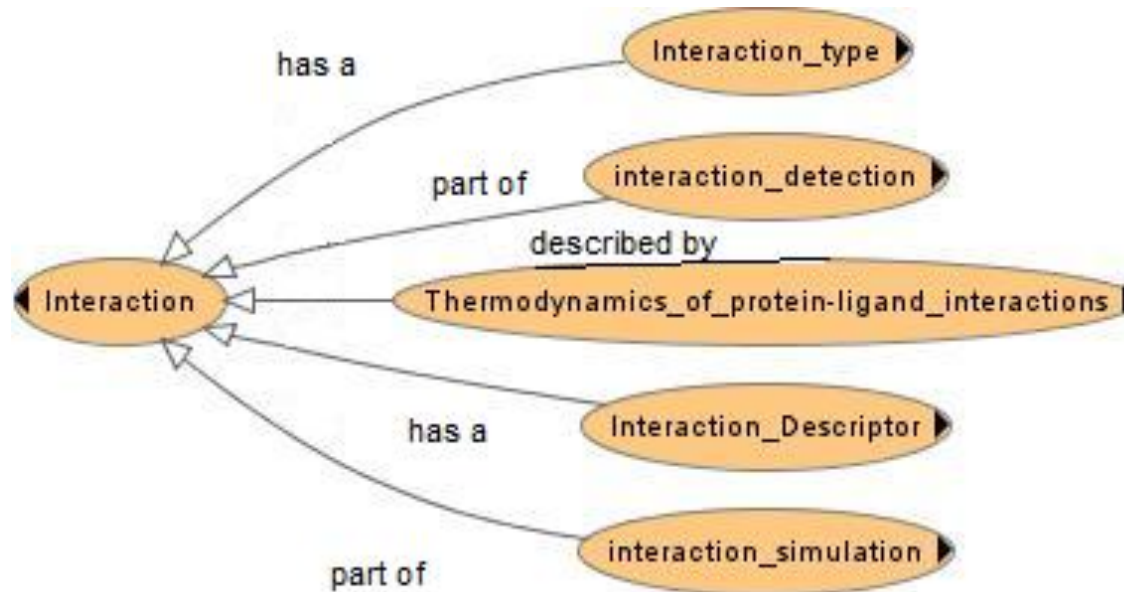
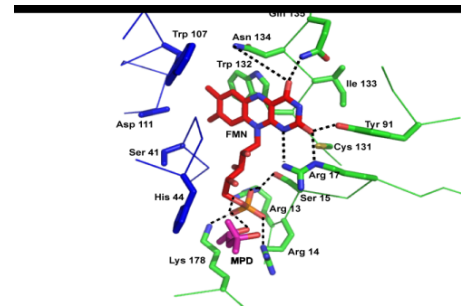
"Activation energy
Arrhenius calculation from two temperatures.
http://www.calctool.org/CALC/chem/kinetics/act_en
"

isDefinedBy

"Activation energy can be thought of as the height of the potential barrier separating two minima of potential energy (of the reactants and products of a reaction). For a chemical reaction to proceed at a reasonable rate, there should exist an appreciable number of molecules with energy equal to or greater than the activation energy.

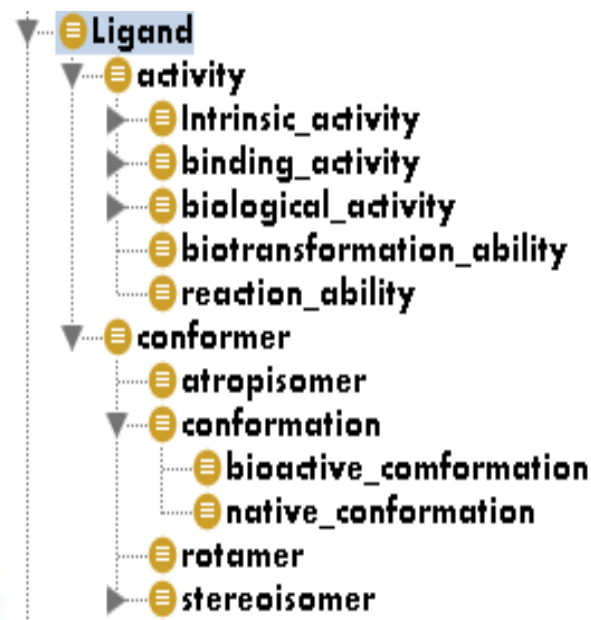
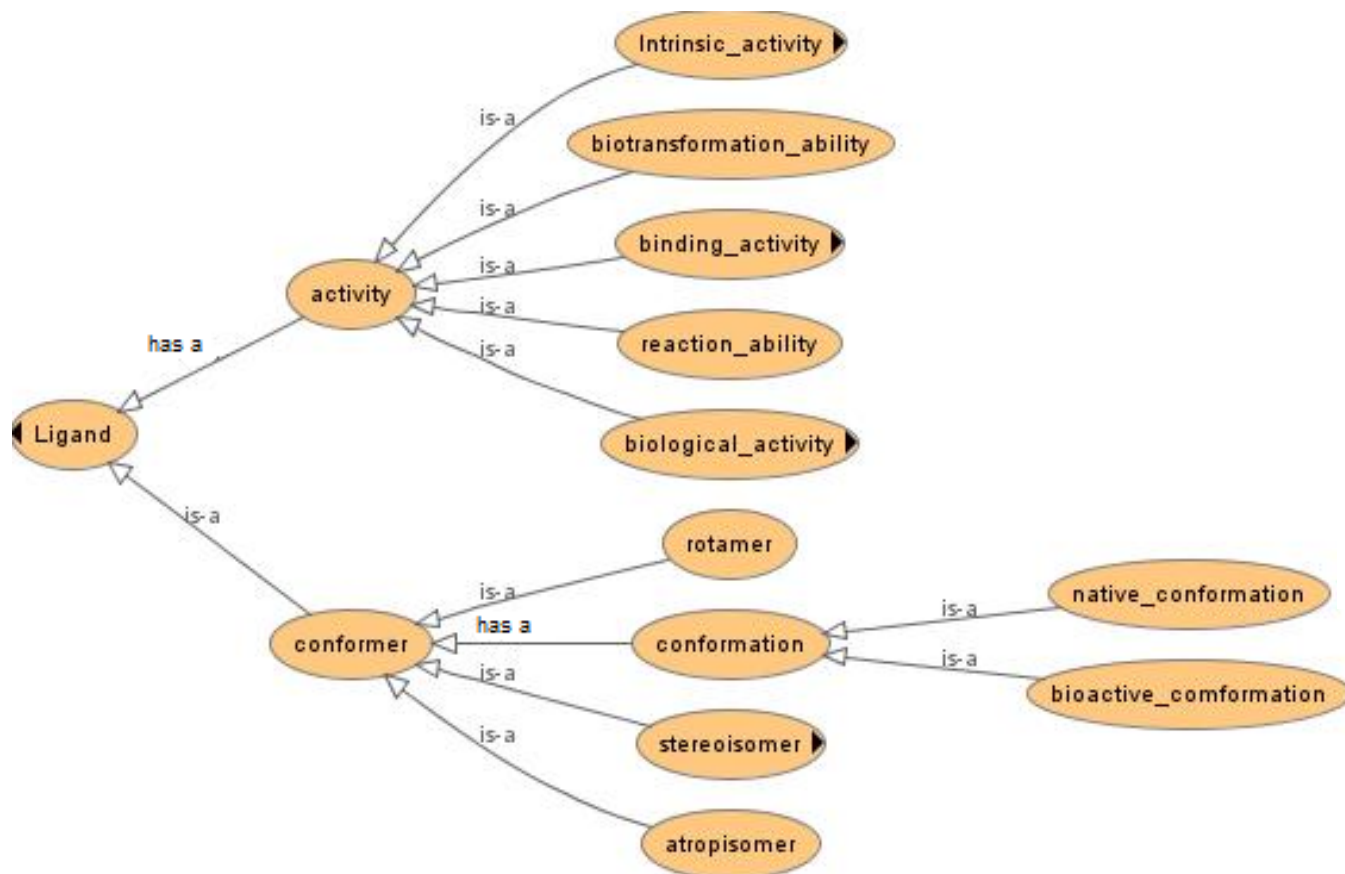
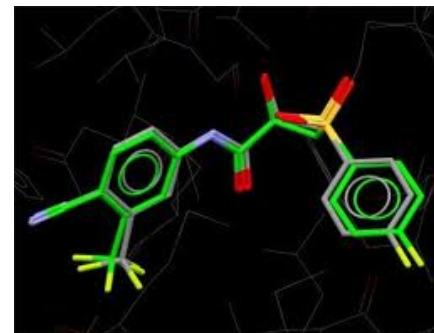
Interaction concept

picture taken from: <http://www.topsan.org/Proteins/JCSG/3ek3>



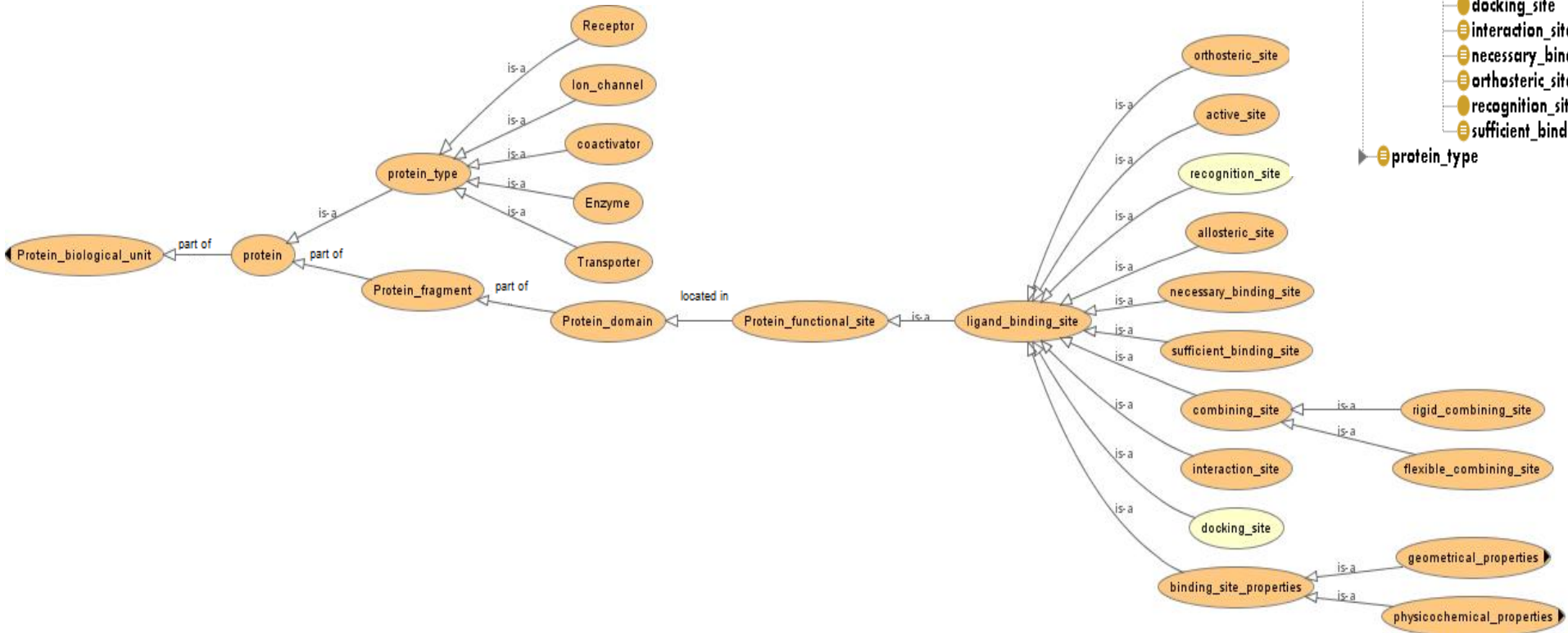
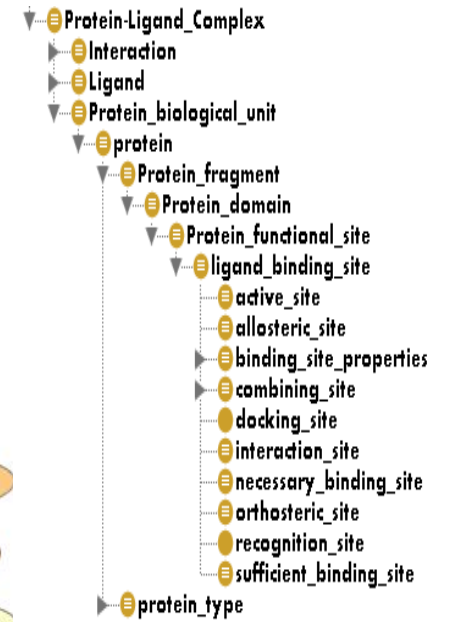
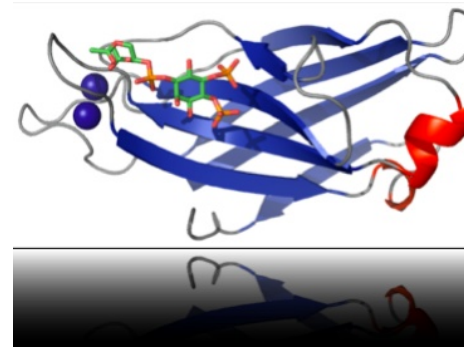
Ligand concept

picture taken from <http://coolvod.net/software/927-ccdc-gold-suite-412.html>



Protein concept

picture taken from: <http://homepage.univie.ac.at/nicolas.coudevylle/Home.html>



Productive Use of Large Compute Infrastructures

High Throughput Extraction of Scientific Information from Full Text Sources:

The UIMA-HPC Project

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

UIMA-HPC

Efficient Information Extraction Workflows in many-core environments



Vision

Scientific Challenge:

The knowledge in Chemistry, Biology and Pharmaceutical Sciences grows with impressive speed. As a result, the number of publications in these areas is reaching unparalleled dimensions. However, knowledge is being communicated in non-standardised ways.

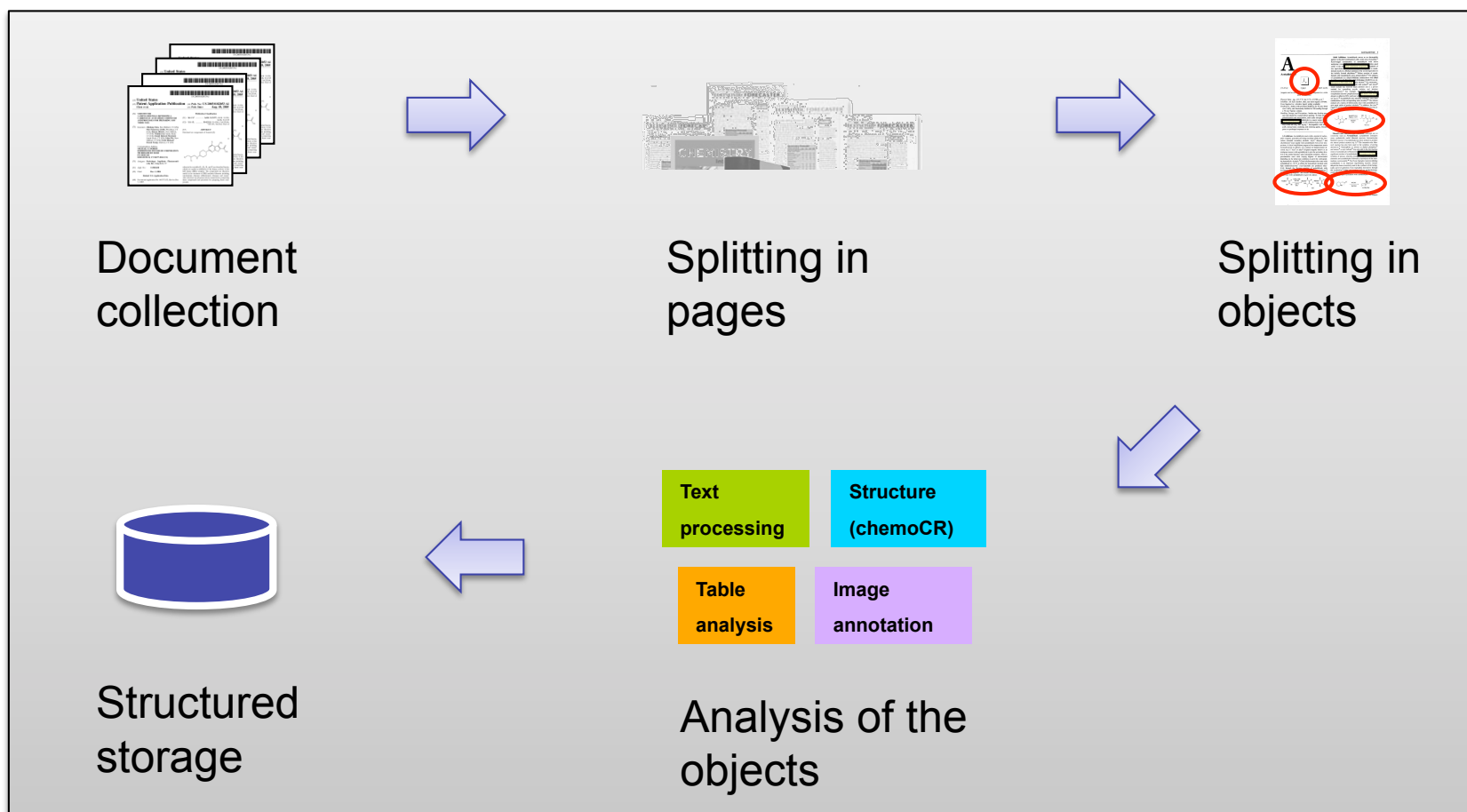
Relevant knowledge sources are not well standardized, let alone is the knowledge structured. This limits the ability to query knowledge sources.

Problem-solving approach:

Development of technology that – based on HPC – allows for high throughput extraction of structured information from unstructured knowledge sources

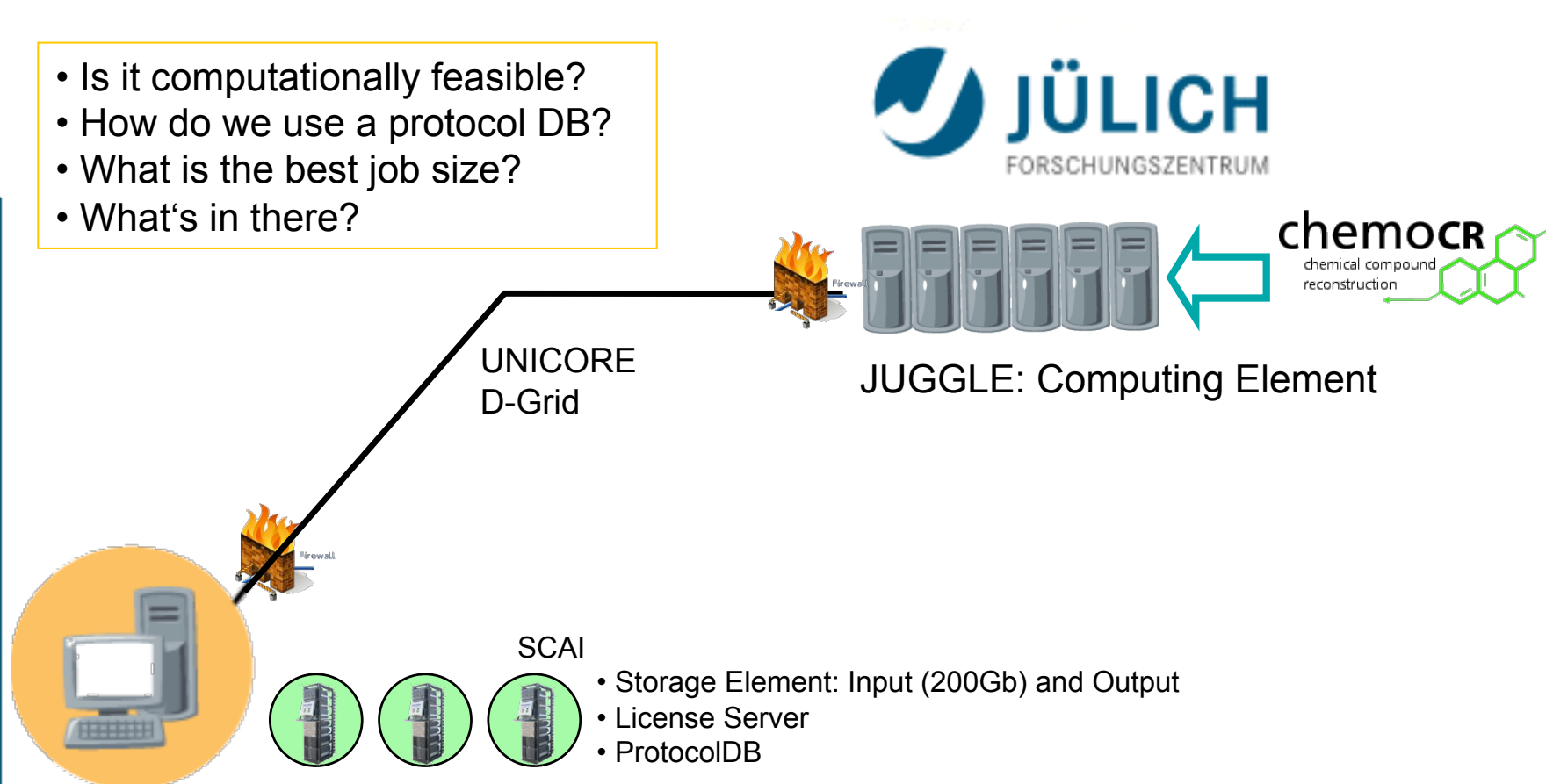
Structured Knowledge

Use case scenario: automatic patent structuring



Previous study: The grand patent challenge 2009

- Is it computationally feasible?
- How do we use a protocol DB?
- What is the best job size?
- What's in there?



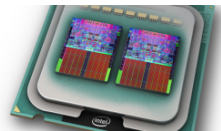
Technical Issues and Pitfalls

- User accession rights (files, scheduler, installed tools and libs, ...)
- Firewall (ports: MySQL, denial of service attack, time outs, ...)
- Missing files (NFS down, package lost, not installed, ...)
- Too many requests on license server
- Too many connections in database
- Ressources (reservation, priorities, ...)

UIMA AS in the context of HPC

Support of many-core architecture

- several instances of a service
- eff. usage of shared memory (JVM)
- asynchronous execution



Support of clusters

- several remote services (eg SOAP)
- communication via JMX and http

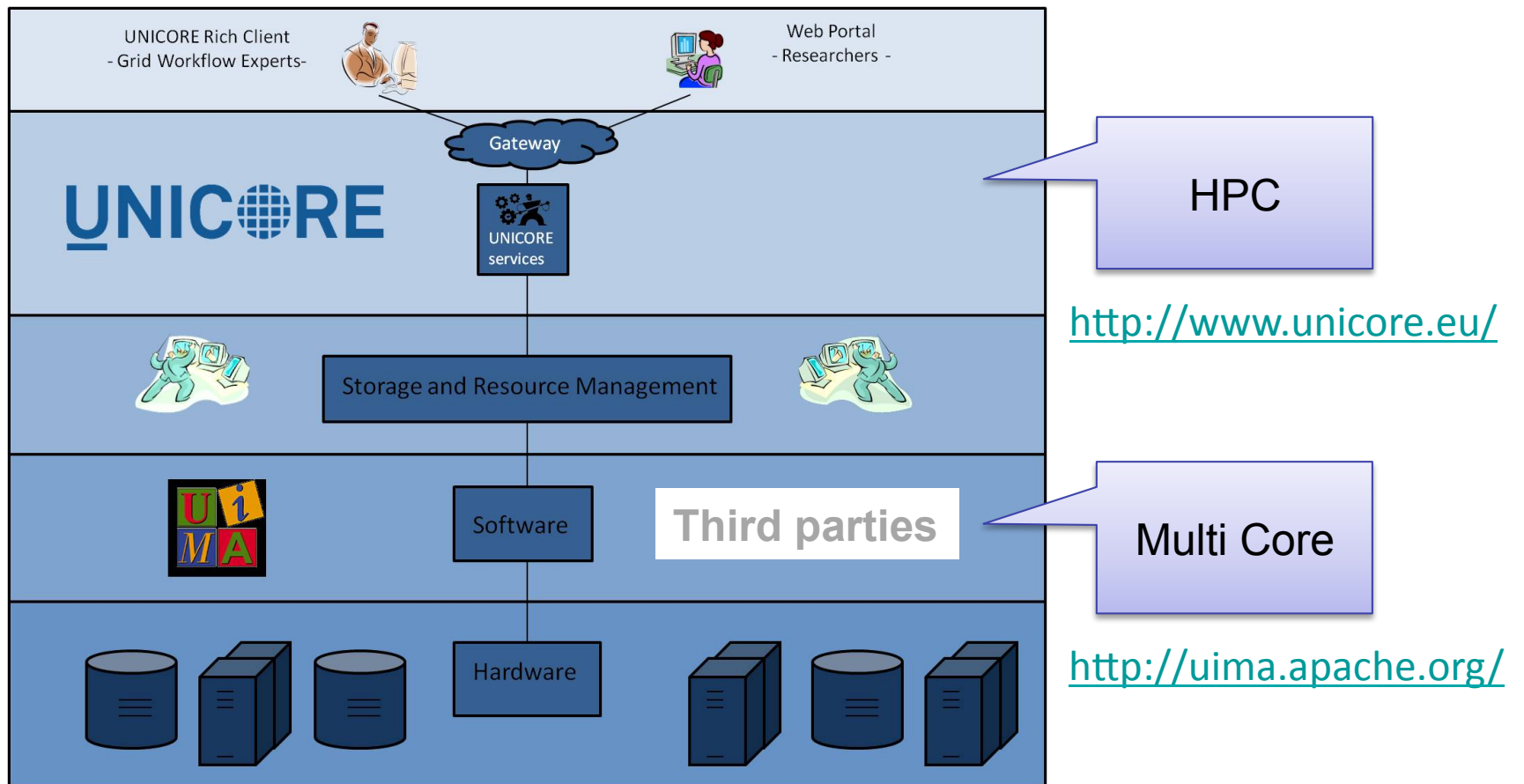


Control via pre-configured parameters

- CAS pool size
- casMultiplier poolSize
- ...

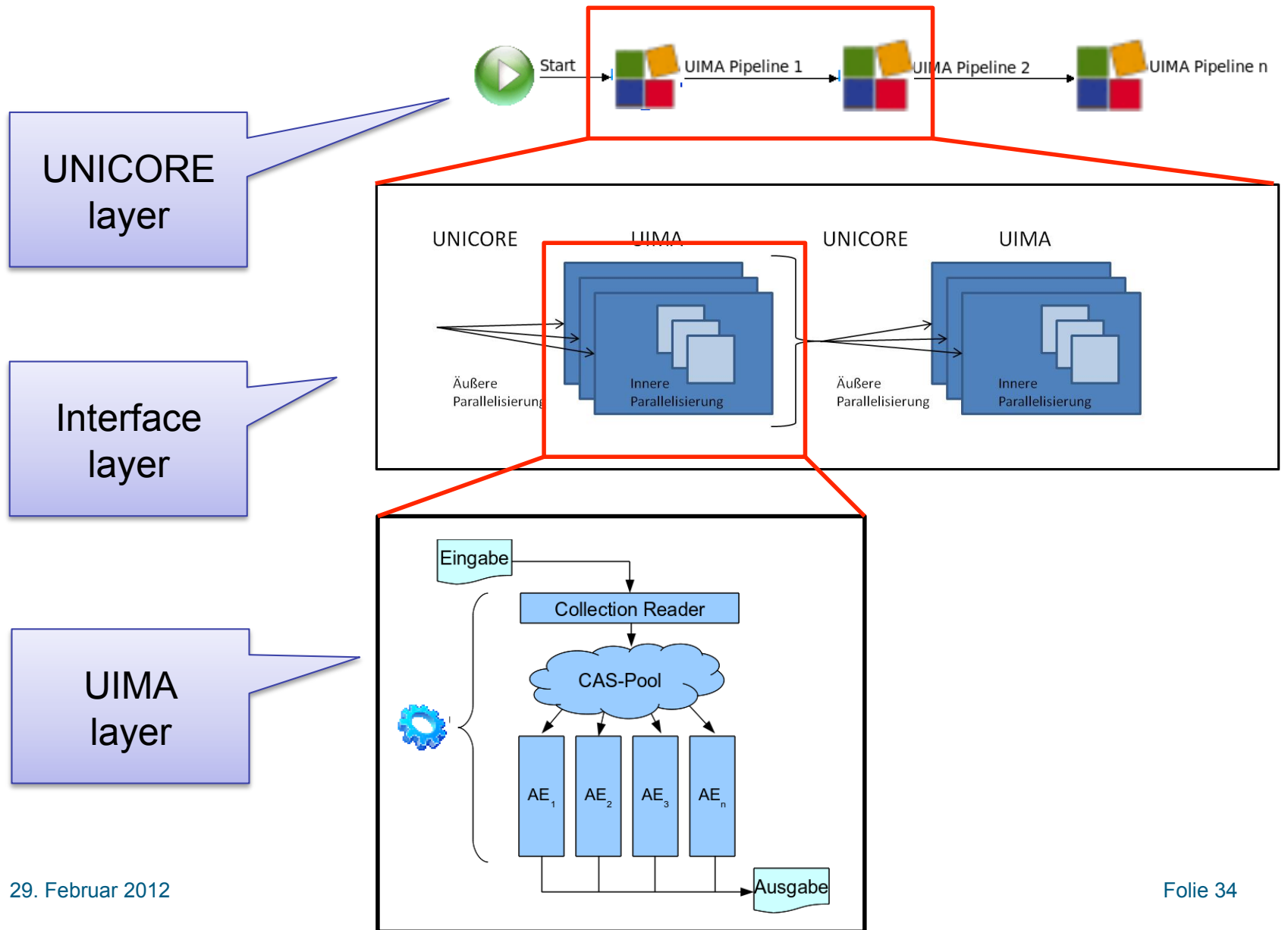
Manual
tuning

Problem-Solving Approach

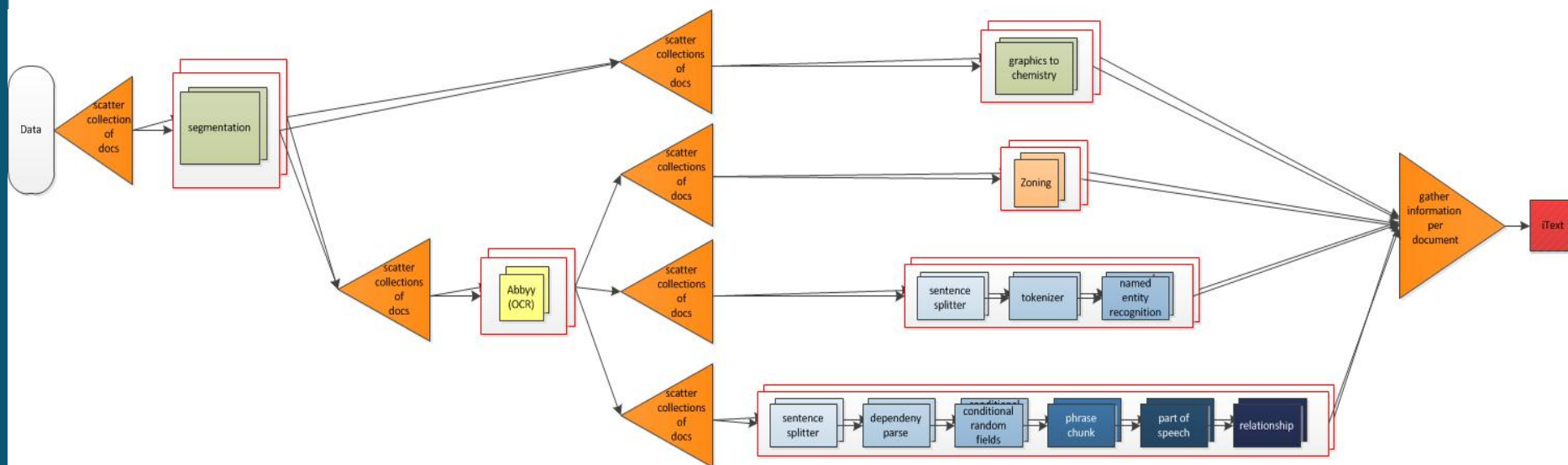


System Architecture

UIMA-HPC



First Prototype: PDF Annotation of Patents



hiekebmc.pdf - Adobe Acrobat Pro

Datei Bearbeiten Anzeige Dokument Kommentare Formulare Werkzeuge Erweitert Fenster Hilfe

Erstellen Zusammenführen Zusammenarbeiten Schützen Unterschriften Formulare Multimedia Kommentar

5372 (1 von 14) 115% Suchen

Füllen Sie bitte das folgende Formular aus. Als Formularverfasser können Sie ein Formular mit der Option "Formular verteilen" im Menü "Formulare" an Empfänger senden. Felder markieren

Bioorganic & Medicinal Chemistry 19 (2011) 5372–5382

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

SAR studies of acidic dual γ -secretase/PPAR γ modulators

Martina Hieke^{a,1}, Julia Ness^b, Ramona Steri^a, Christine Greiner^c, Oliver Werz^d, Manfred Schubert-Zsilavecz^a, Sascha Weggen^{b,*}, Heiko Zettl^{a,*}

^aInstitute of Pharmaceutical Chemistry, Goethe-University Frankfurt, Max-von-Laue-Str. 9, D-60438 Frankfurt am Main, Germany
^bDepartment of Neuropathology, Heinrich-Heine-University Duesseldorf, Moorenstrasse 5, D-40225 Duesseldorf, Germany
^cDepartment of Pharmaceutical Analytics, Eberhard-Karls-University Tuebingen, Auf der Morgenstelle 8, D-72076 Tuebingen, Germany
^dInstitute of Pharmacy, Friedrich-Schiller-University Jena, Philosophenweg 14, D-07743 Jena, Germany

ARTICLE INFO

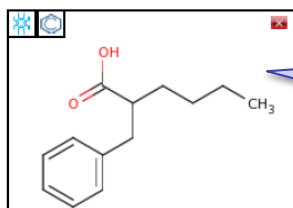
Abstract history:
 Received 25 May 2011
 Revised 25 July 2011
 Accepted 2 August 2011
 Available online 6 August 2011

Keywords:
 SAR
 Alzheimer's disease
 PPAR γ agonists
 γ -Secretase modulators
 Carboxylic acids

1. Introduction

proliferation and B-cell differentiation in pre-
 mice.^{6,7} As an appealing alternative to GSIs, so-

2-benzyl hexanoic acid



Chemical
Popups

Statistics of highlighted entities of class IUPAC (page 1)

| Entity | Frequency |
|---|-----------|
| thioether | 5 |
| benzene | 5 |
| phloroglucinol | 3 |
| pyrimidine | 3 |
| Dihydroxybenzaldehyde | 3 |
| Phloroglucinol | 2 |
| phenylethanol | 2 |
| carbon | 2 |
| sodium hydride | 2 |
| 2-((3,5-diphenethoxyphenoxy)hexanoic acid | 2 |
| phosphonate | 2 |
| 2-benzylidene hexanoic acid | 1 |
| phenol | 1 |
| dimethylthio-carbamoylchloride | 1 |
| 3-hydroxy-4-(2-cyclohexylethoxy)benzaldehyde | 1 |
| glutathione | 1 |
| 2-[(3,5-diphenethoxyphenyl)thio]hexanoic acid | 1 |
| 2-benzyl hexanoic acid | 1 |
| 3-(2-cyclohexylethoxy)-4-phenethoxybenzaldehyde | 1 |
| tetramethylsilane | 1 |
| dimethylthiocarbamoyl chloride | 1 |
| 2-benzylidene hexanoic acid | 1 |

Chemical
Index

Annotations +
Linkouts

Overview

Direct Usage of Unstructured Information Sources for Disease Modelling

From Medline Mining

to

Modelling Neurodegenerative Diseases

Why Modelling of Neurodegeneration?

In 2009 the Federal Government of Germany decided to start a new research centre that focuses on translational research on neurodegenerative diseases. In fact, neurodegenerative diseases (Alzheimer, Parkinson, Multiple Sclerosis; Epilepsy; „rare“ NDDs)

The total costs of Alzheimer is estimated to exceed 20 trillion US\$ in the US in the years between 2020 - 2050. (source: Alzheimer.org). Current costs / year in the US (according to Alzheimer.org): **183 billion US\$**

The incidence rate of Alzheimer and other dementias is almost 50% in the population older than 85 years. Next generation will regularly have a life span of >100 years.

The Starting Conditions

What we have:

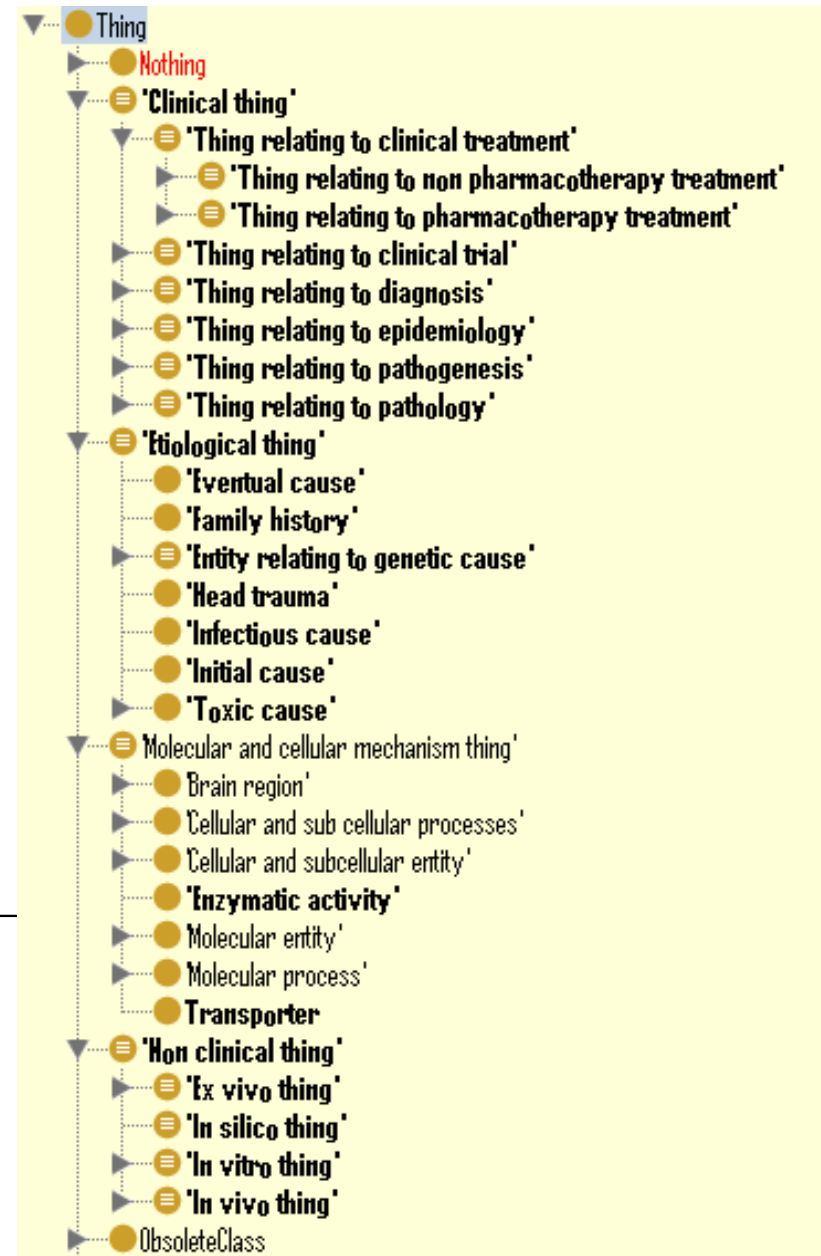
- An ontology capturing relevant knowledge on Alzheimer's Disease (ADO)
- An ontology representing and integrating brain regions and cell types (BRCO)
- A method for the automated identification of hypotheses in text based on regular expressions
- An excellent machinery for biomedical text mining (ProMiner) with top performing gene and protein name recognition

Seite 39

Alzheimer's Disease Ontology (ADO)

Alzheimer's ontology:

- ❑ Captures more than 700 classes/ concepts
- ❑ BFO already implemented



Brain Region and Cell-type Ontology (BRCO)

The screenshot displays the Brain Region and Cell-type Ontology (BRCO) interface. On the left, a hierarchical tree of concepts is shown, with 'Substantia_nigra_pars_compacta_dopaminergic_cell' selected. The right pane shows the details for this concept, including a synonym, a reference, a description, and a list of equivalent classes and superclasses.

Synonym
"Substantia nigra dopaminergic cell, Nigral dopaminergic cell"

reference
"http://neurolex.org/wiki/Category:Substantia_nigra_pars_compacta_dopaminergic_cell"

is DefinedBy
"Nigral dopaminergic cell is a neuron found in the midbrain of vertebrates. These neurons comprise most of the substantia nigra and mainly regulate motor and sensorimotor functions within the brain."

Description:

Soma Location: Substantia nigra pars compacta
Spine density on dendrites: Aspiny Dendrite Quality

Axon Specific Properties
Axon projection laterality: ipsilateral
Location of axon arborization: Neostriatum
Cellular synaptic target: Neostriatum medium spiny neuron
Neurotransmitter: Dopamine

Description: Substantia_nigra_pars_compacta_dopaminergic_cell

Equivalent classes

Superclasses

- Substantia_nigra_pars_compacta


Inferred anonymous superclasses

- has_part some Substantia_nigra_pars_compacta
- has_part some Substantia_nigra_pars_reticulata
- has_part some CA3_alveus
- has_part some Piriform_cortex_layer_1
- has_part some Neocortex_layer_4
- has_part some Chemoarchitectural_part
- has_part some CA1_alveus
- has_part some CA3_stratum_lucidum
- has_part some Hindbrain
- has_part some Piriform_cortex_layer_2
- has_part some Aggregate_regional_part_of_brain
- has_part some Regional_part_of_forebrain
- has_part some Molecular_layer_of_dorsal_cochlear_nucleus
- has_part some Trigeminal_nucleus
- has_part some Regional_part_of_midbrain
- has_part some Composite_part_spanning_multiple_base_regional_parts_of_brain
- has_part some Regional_part_of_hindbrain

Current state: more than 3000 concepts; more than 5000 synonyms

Expression of Speculative Statements in Scientific Text


Kallikrein-related peptidase 6 in Alzheimer's disease and vascular dementia.

 20846516 **Authors:** Ashby, Emma L; Kehoe, Patrick G; Love, Seth **Date:** 2010-12- **Journal:** Brain research **Affiliation:** Dementia Research Group, Institute of Clinical Neurosciences, Clinical Science at North Bristol, University of Bristol, UK.
] *Statistics* *Select ID with comment:*

Human **kallikrein-related peptidase 6 (KLK6)** is highly expressed in the central nervous system. Although the physiological roles of this serine protease are unknown, in vitro substrates include **amyloid precursor protein** and components of the extracellular matrix, which are altered in neurological disease, particularly Alzheimer's disease (AD). We have compared **KLK6** expression in post-mortem brain tissue in AD, vascular dementia (VaD) and controls. We studied the distribution of **KLK6** in the temporal cortex and white matter by immunohistochemistry, and measured **KLK6** mRNA and protein levels in the frontal and temporal cortex from 15 AD, 15 VaD and 15 control brains. Immunohistochemistry showed **KLK6** to be restricted to endothelial cells. After adjustment for variations in vessel density by measurement of factor VIII-related antigen, we found **KLK6** protein and mRNA levels to be significantly decreased in the frontal but not the temporal cortex in AD. In VaD, **KLK6** protein level was significantly increased in the frontal cortex. Our findings **suggest that** an altered **KLK6** expression **may contribute** to vascular abnormalities in AD and VaD.

Hypothesis =  **KLK6** +  **may contribute** +  **AD**

Hypotheses finder \cap AD ontology \cap Human genes and proteins




- Epilepsy Ontology
- Alzheimer Ontology
 - Molecular and cellular mechanis
 - Etiological thing
 - Non clinical thing
 - Clinical thing
 - Thing relating to clinical trial
 - Thing relating to diagnosis
 - Thing relating to epidemiology
 - Thing relating to pathology
 - Thing relating to clinical treat
 - Thing relating to pathogen
 - Mild cognitive Impairmen
 - Stage
 - Moderate cognitive Declin
 - Alzheimer disease
- Parkinson Ontology
- Hypothesis Finder
- Drug Names
- Human Genes / Proteins

Your Search:


- **Fulltext query:**
alzheimer
- **Filtering from Entity Tree:**
 (Boolean OR) (Boolean AND) Boolean NOT
*Alzheimer Ontology:(((Mild cognitive Impairment)))AND
 Hypothesis Finder AND
 Human Genes / Proteins*
- **Display entities in Entity View of type:**
Human Genes / Proteins

Evaluation of plasma Abeta(40) and Abeta(42) as predictors of conversion to Alzheimer's disease in patients with mild cognitive impairment.


 **PubMed** 18486992 **Authors:** Hansson, Oskar; Zetterberg, Henrik; Vanmechelen, Eugeen; Vanderstichele, Hugo; Andreasson, Ulf; Londos, Elisabet; Wallin, Anders; Minthon, Lennart; Blennow, Kaj **Date:** 2010-03 **Journal:** Neurobiology of aging **Affiliation:** Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Sweden. oskar.hansson@med.lu.se

Statistics Select ID with comment:


Numerous studies have shown a marked decrease of beta-amyloid(42) (Abeta(42)) in the cerebrospinal fluid (CSF) of patients with incipient Alzheimer's disease (AD). However, studies on Abeta in plasma are contradictory, and show very marginal differences between patients and controls. Here, we analyzed plasma samples using a new multiplex immunoassay for simultaneous analysis of Abeta(1-40), Abeta(n-40), Abeta(1-42), and Abeta(n-42). The plasma samples were obtained at baseline from two independent cohorts of patients with mild cognitive impairment (MCI) and age-matched controls. In the first cohort, 41% of the 117 MCI cases converted to AD during a clinical follow-up period of 4-7 years. In the second cohort, 14% of the 110 MCI subjects developed AD during a clinical follow-up period of 2-4 years. None of the plasma Abeta isoforms differed between MCI patients that subsequently developed AD and healthy controls or stable MCI patients. The Cox proportional hazards model did not reveal any differences in the probability of progression from MCI to AD related to plasma Abeta levels. In contrast, low levels of Abeta(1-42) in CSF were strongly associated with increased risk of future AD. The absence of a change in plasma Abeta in incipient AD despite the marked change in CSF may be explained by the lack of a correlation between the levels of Abeta(1-42) in CSF and plasma. In conclusion, the results show that CSF biomarkers are better predictors of progression to AD than plasma Abeta isoforms.



Hypothetical Pattern



Human gene

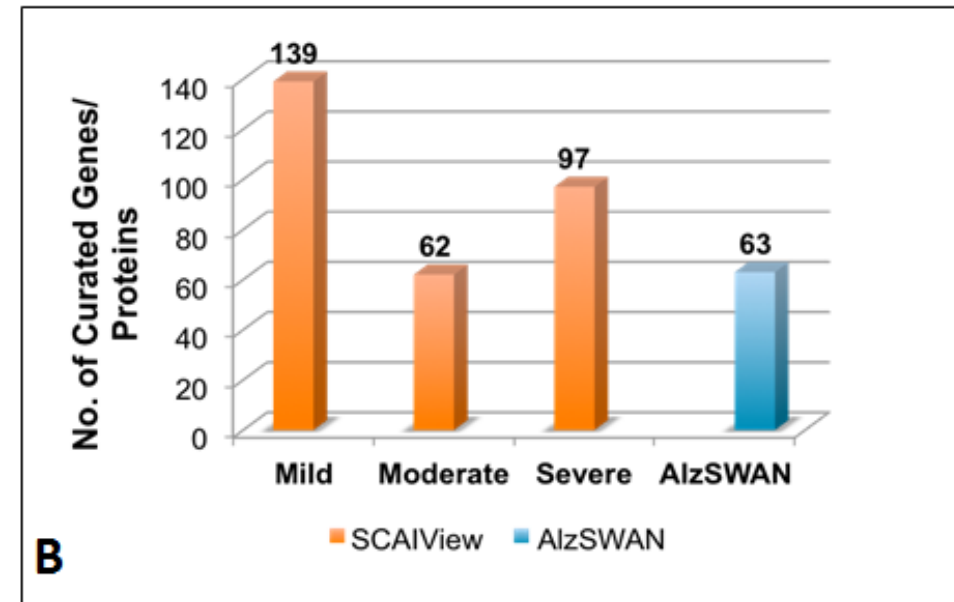
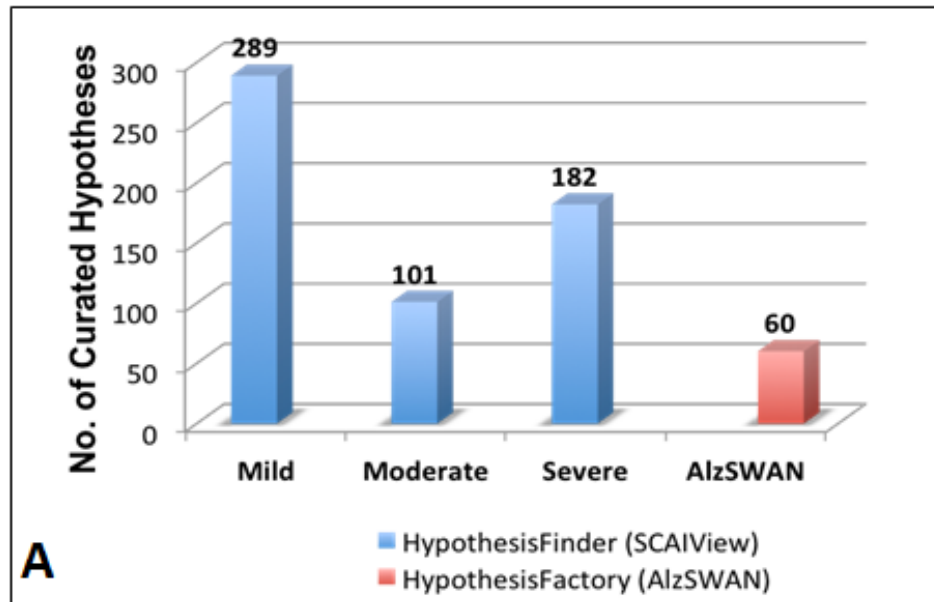


Alzheimer Stage

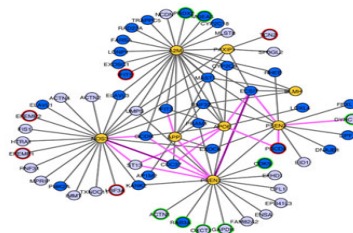
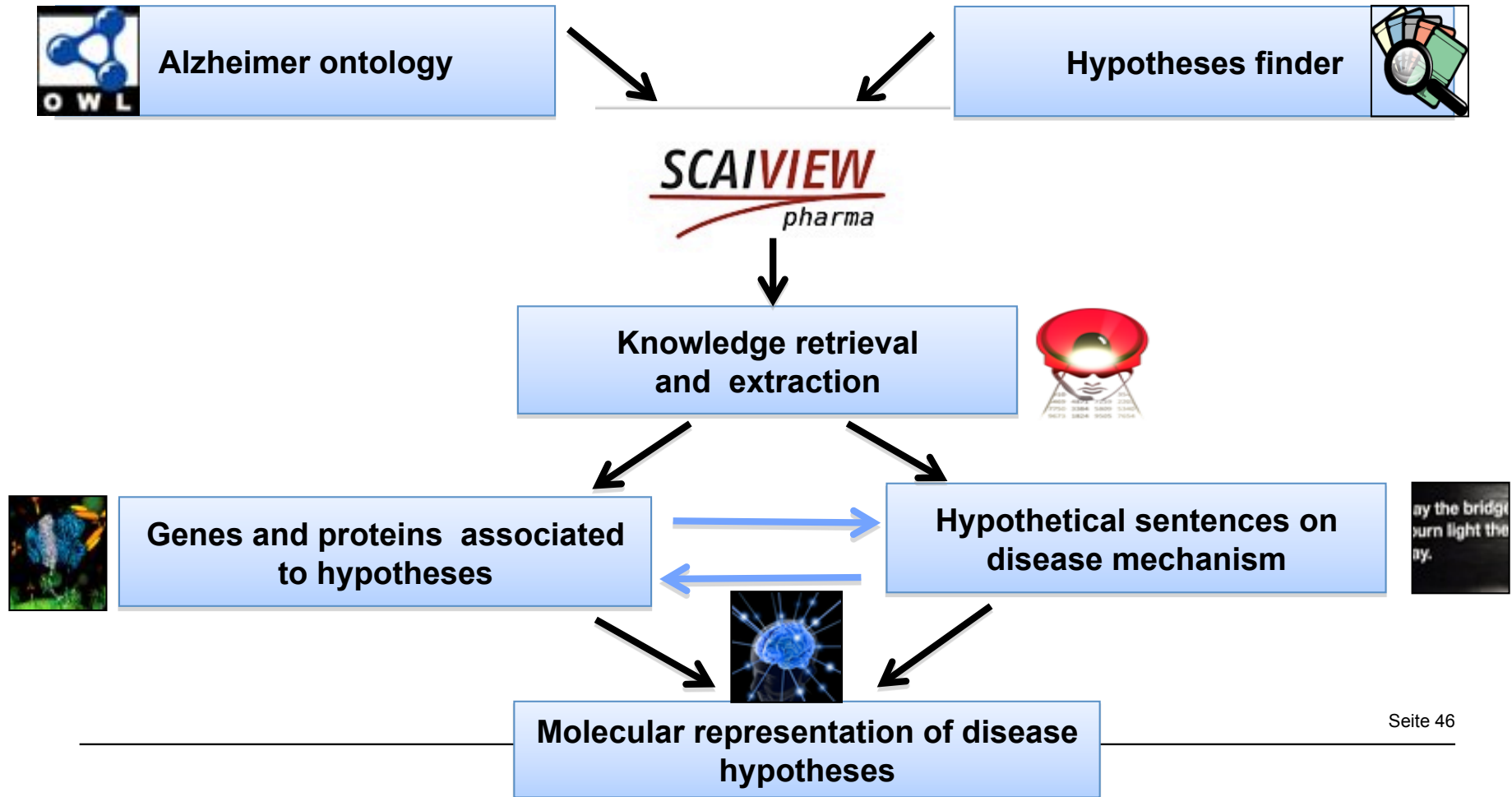
Performance of Hypotheses finder

| S.No | Data type | Source | Precision | Recall | F score |
|------|---|-------------------------------|-----------|--------|---------|
| 1 | 200 abstracts related to Alzheimer's | PubMed | 0.84 | 0.86 | 0.85 |
| 2 | 2 full text articles related to Alzheimer's | Journal of Medical Hypotheses | 0.85 | 0.88 | 0.86 |
| 3 | 143 abstracts related to Alzheimer's | Alzswan/PubMed | 0.90 | 0.97 | 0.93 |
| 4 | 100 abstracts related to Epilepsy | PubMed | 0.96 | 0.91 | 0.94 |
| 5 | 100 abstracts related to Parkinson's | PubMed | 0.90 | 0.93 | 0.92 |

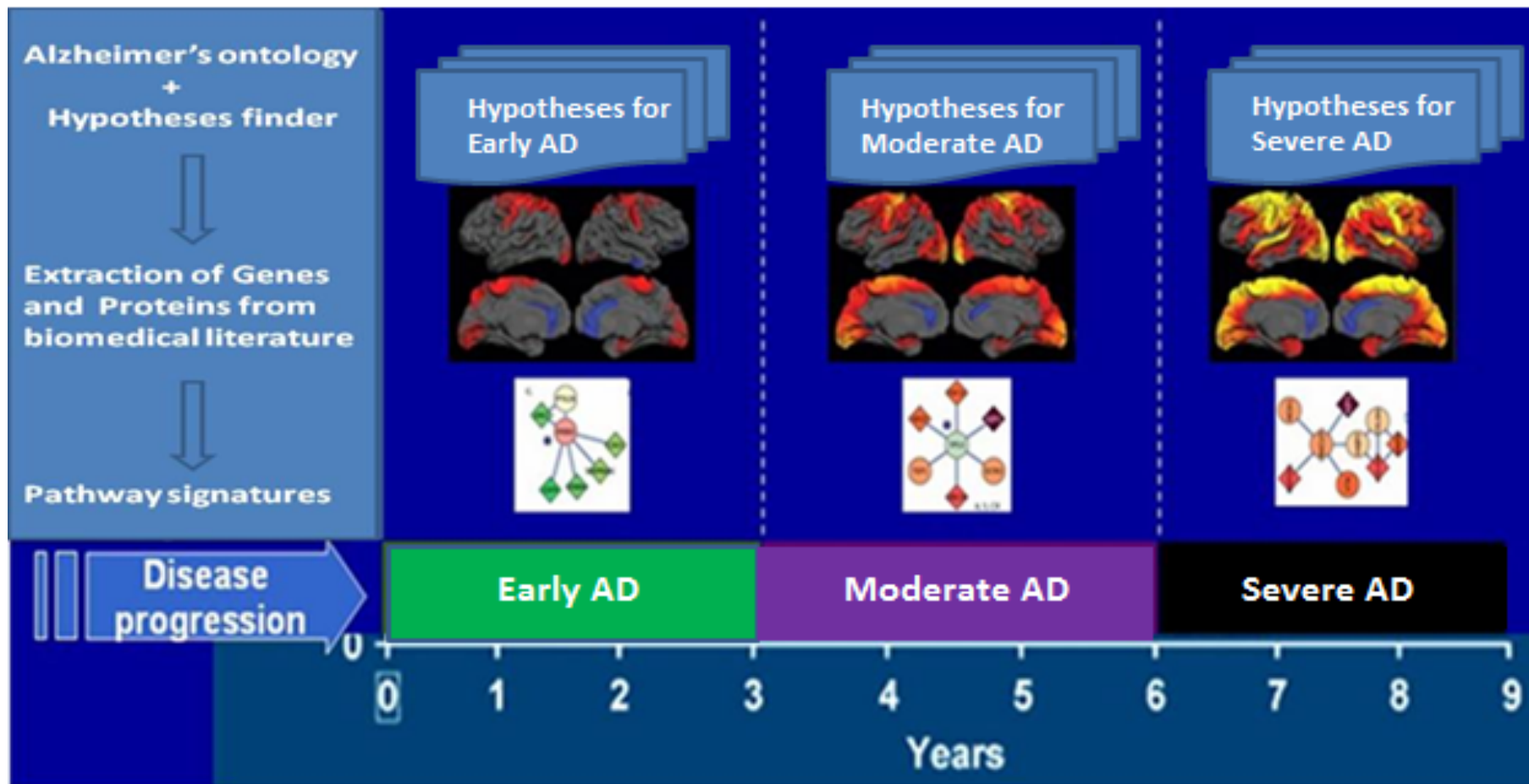
Performance of Hypotheses finder



The Knowledge – Discovery Strategy



Analysis of hypotheses patterns across disease stages



Summary

Fraunhofer SCAI Department of Bioinformatics stands for:

- Advanced technologies in text- and data mining, disease modelling in the area of neurodegeneration and high performance computing
- Trying hard to make the „eScience“ paradigm a “living experience”
- Clear dedication towards contract research and sustainable innovation
- Internal usage of technologies (information extraction; distributed and high performance computing) for biomedical application