The GRIDs @ Novartis

Prof. Manuel C. Peitsch
Global Head of Information and Knowledge Management at NIBR
Agenda

- Information Integration and the Knowledge Space (a Data GRID?)
- How can in silico science contribute to the Drug Discovery process and in pharma?
- The NIBR/IK@N HPC concept and status update.
- E-Collaboration environment
- Concluding remarks
The Challenges of Drug Discovery

- Mechanism-based Drug Discovery
  - Understanding Disease
  - Pathways elucidation and association
  - Target validation
  - Clinical PoC
- Exploit X-nomics
- Reduce project life cycle
- Increase PoS after D3 (Lead optimisation)
Geographical complexity.

Novartis Pharmaceuticals Research Worldwide Community 2002
3000 Scientists

- New Jersey
  - Oncology
  - Arthritis
  - Functional Genomics

- Cambridge/Boston
  - (build-up phase)
  - Metab.
  - Cardiov.
  - Dis.
  - Infectious Diseases TA

- Horsham London
  - Resp. diseases
  - Chronic pain

- Vienna
  - Dermatology/
  - Immunopathology

- Gaithersburg
  - Gene therapy

- La Jolla
  - GNF
  - Functional Genomics

- Basel
  - FMI
  - Nervous System
  - Transplantation
  - Oncology
  - Arthritis/Bone
  - Functional Genomics
  - Ophthalmics

- Singapore
  - NITD

- Japan
  - Oncology
  - Arthritis
Organisational complexity.

Planned Launches 2002 – 2006

<table>
<thead>
<tr>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>≥ 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elidel®</td>
<td>myforti® (EU)</td>
<td>Prestige®</td>
<td>EPO906</td>
<td>FIT270</td>
</tr>
<tr>
<td>Zelma® (US)</td>
<td>Visudyne® (J)</td>
<td>Xolar®</td>
<td>IOL670</td>
<td>NH3104</td>
</tr>
<tr>
<td>Glivec®/Gleevec® (GIST)</td>
<td>Enlacinone Triple Clothin</td>
<td>Cilastin®</td>
<td>Ocular Therol®</td>
<td>Femara® adj.</td>
</tr>
<tr>
<td>Diene® CHF</td>
<td>Citrazini® (US)</td>
<td>Intercept®</td>
<td>myforti® (US)</td>
<td>Prestige® New formulations</td>
</tr>
<tr>
<td>Zonesta® Bone met. treat</td>
<td>Lepad®</td>
<td>Zental® (J)</td>
<td>Zelma® (EU)</td>
<td>LAF237</td>
</tr>
<tr>
<td>Brilinta® LA (ADHD)</td>
<td>Lexio®</td>
<td>Zentral® (J)</td>
<td>Zentral®/Zentral®</td>
<td>PKC415</td>
</tr>
<tr>
<td>Lofenad® 10/20 mg (US) Hyperension</td>
<td>Viopteg® EU</td>
<td>Zental®/Zental®</td>
<td>Zentral®/Zental®</td>
<td>PKI308</td>
</tr>
<tr>
<td>Co-Doven® Hypertension (150/25; 140/15 mg)</td>
<td>Glicence®/Gleven®</td>
<td>Trilep®</td>
<td>Zentral®/Zentral®/Gerd</td>
<td>Laran® Tianeptapril</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NME</td>
<td>LCM</td>
<td>NME rollout</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NIH Novartis Media Conference / 2003-10-17
Data and Information complexity.

Raw data from instruments.

Analysed data.
The Vision

Enable and transform the Drug Discovery process through:

- Comprehensive and reliable Data and Information
- Seamless information integration for easy navigation
- Turning Data into Knowledge using in silico science
- Simulate biomolecular processes using in silico science
- E-Collaboration and v-communities

Data Information and Knowledge GRID → Knowledge Space

Knowledge Space Portal

Computational life science

People Networks
• The Vision and the challenges in Drug Discovery
  
  • How can in silico science contribute to the Drug Discovery process and in pharma?
  • The NIBR/IK@N HPC concept and status update.
  • E-Collaboration environment
  • Concluding remarks
The Space metaphor

**Map and navigation metaphor**
Connectivity map between discrete objects. The maps are published and updated to reflect changes in the space. All towns must be connected by the system of roadways. If not there will be isolated pockets of people (ideas, knowledge, and wasted resources). Reasonable number of “roads” between any two points. Travel service help travelers scope out new adventures.

**Metaphorical mapping**

- **Universe**: All Data / Information / Knowledge relevant to Novartis.
- **Galaxies**: Exploratory / Development / Regulatory / Manufacturing / Commercial.
- **Solar sys**: Disease / Targets / Hits&Leads / Patients
- **Planets**: Bioinformatics / Chemoinformatics / Clin Pharm IS / etc …
- **Continents**: Genetics / Genomics / Proteomics / Libraries / etc …
- **Countries**: SNP / Sequence databases / compound databases / etc …
- **Cities**: Genebank / SWISS-PROT / Beilstein / etc …
The Knowledge Space

• The Knowledge Space consists of:
• The **collection** of all types of data and information within the scope of interest defined by a particular business.
  ➢ Thus, the Knowledge Space is composed of:
    • Databases, information sources, document/knowledge bases, etc… with relevance to us.
    • There is no conceptual difference between internal and external data/information. This is only a matter of tagging it as either.
• The **Meta Data** and the **Knowledge Map** which describe the collection in terms of content and location of content.
The Knowledge Space Portal

- The "Knowledge Space Portal" will, via a single customizable interface
  - Federate heterogeneous data resources and provide precise organization of the content
  - Provide quick and intuitive access to information
  - Provide data extraction, analysis and exploration tools
  - Allow data integration, data exchange and interoperability of applications
  - Provide mechanisms for data capture and annotation
  - Provide knowledge sharing and collaborative tools
Our Knowledge Space

Navigators

Analytics

Expert Systems

Retrieval

Mining

Meta D/I

Ontologies

Thesaurii

Program Book

Research

Documentation

Literature

Comp. Inf.

Bioinformatics

Chemistry

Biology

RDS

Internet

Other

K maps
About the Ultralink/Ultraaction and the Ultralinker

- The Ultralink is an “intelligent” context-sensitive Hyperlink (the classical link provided by standard Web Browsers)
- This “intelligence” is created by the Ultralinker
- In the Portal, the Ultralink looks like a Hyperlink on a shaded background.
- By clicking on an Ultralink you will go to another page, but get a menu.
- This menu will offer a number of actions/options which can be taken, some will in fact be a chain of logical consecutive actions, normally done manually one after the other.
  - The Ultralink allows direct interaction between any type of entity (gene name, compound name, disease name, company name, etc... with the appropriate tools and components of the Business Logic layer.
  - The Ultralink allows to select any portion of text and use it as input to a text mining-assisted search “à la Ulix”, using a single mouse click.
  - And more to come....
- The Ultralink allows easy navigation in a coherent region of the Knowledge Space.
How the Ultralinker works

• The Ultralinker is a Web service.
• The Ultralinker analyses any information (such as a complete web page) it receives for items it recognises using text mining and pattern recognition tools.
• Each recognised item is mapped onto the Knowledge Map.
• The Knowledge Map (meta Knowledge) will tell what can be done with that item e.g.
  ➢ If a gene name is recognised then you can: search for its sequence in databases; do sequence similarity searches; search for it in target databases; search for it in genetic disorder databases; translate its codons into amino acids, map it onto the transcriptome; produce a 3D structure by comparative modelling; look for hits from HTS; etc... any thing you can do with this gene name which is coherent and scientifically valid...It can be chained/automated in a single click (Ultraaction or work-flow) or provided as separate items.
• The Ultralinked will create a menu that will be sent to the User interface.
• Easy to configure
• Recognises “zoned” documents.
### Search Results: “gleevec” in “INTERNET” (49 documents)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Title</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>National Cancer Institute - Gleevec (ST1571)</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>DocCarta - Charting Pathways of Life</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>CIVILSUPPORT.COM Gleevec Resource Center</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>Gleevec Formally OK'd for Rare Cancer</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>UnOfficial Gleevec Site Leukemia - Chronic Myeloid Leukemia (CML)</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>Yahoo! Groups: Gleevec</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>Newly Approved Drug Therapies (599): Gleevec, Novartis</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>Med Ad News</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>ClinicalTrials.gov - Information on Clinical Trials and Human ...</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>Top 10 gleevec links: cheapest prices, side effects, info, buy ...</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>CNN.com - Health - Cancer pill wins marketing approval - May 10 ...</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>Gleevec Press Release</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>Gleevec.com - Prescribing Information</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>Gleevec and Chronic Myeloid Leukemia - Cancer: health and medical ..</td>
<td>Google search</td>
</tr>
<tr>
<td>100</td>
<td>Molecular Cancer</td>
<td>Full text</td>
</tr>
</tbody>
</table>

### Resultset Highlighting
- Keywords
  - Products (2)
  - Diseases (6)
  - Companies (4)
  - Targets (1)

### Resultset Tools
- Filtering
- Clustering
- Data Analysis
- Graph Navigator

### Search Statistics
- **Search Returned**: 49 documents for “gleevec”
- **Documents by source**: Google search - 49
- **Attributes**:
  - Products: 159
  - Diseases: 143
  - Targets: 58
  - Companies: 30
  - Molecules of action: 17
Search sequence: e.g. on the Internet 2

Introduction

Gleevec™ (also called STI571 or imatinib mesylate) is approved by the U.S. Food and Drug Administration for the treatment of some forms of adult and pediatric chronic myelogenous leukemia (CML), and for the treatment of a rare form of stomach cancer called gastrointestinal stromal tumor (GIST).

With CML, Gleevec works by blocking an abnormal enzyme characteristic of the disease. In GIST, Gleevec blocks a different abnormal enzyme found on the tumor cells.

Gleevec is the first approved drug to directly turn off the signal of a protein known to cause a cell to divide. Other molecular-targeting drugs previously approved by the FDA interfere with proteins associated with other cancers, but not with proteins that directly cause the disease.

Gleevec is being investigated for its effectiveness against other kinds of cancer, as well, including lymphocytic leukemia and hypereosinophilic syndrome (HES).

Novartis, the maker of Gleevec, has a Patient Assistance Program (PAP) to help patients who are ineligible for financial assistance in covering the cost of the drug. The company has set up a website (www.gleevec.com), as well as a toll-free number to call for more information about PAP or the website: 1-877-GLEEVEC (1-877-453-3932).
Search sequence: e.g. on the Internet 3

Introduction

Gleevec™ (also called STI571 or imatinib mesylate) is approved by the U.S. Food and Drug Administration for the treatment of some forms of adult and pediatric chronic myelogenous leukemia (CML), and for the treatment of a rare form of stomach cancer, stromal tumor (GIST).

With CML, Gleevec works by blocking an abnormal enzyme. Gleevec blocks a different abnormal enzyme.

Gleevec is the first approved drug to directly target a molecularly targeted drugs previously used to treat other cancers, but not with proteins that drive the growth of leukemia cells.

Gleevec is being investigated for its effectiveness in lymphoid leukemia and myelodysplastic syndrome.

Related Pages

Leukemia Home Page
NCT's gateway for information about leukemia.

Search sequence: e.g. on the Internet 3
The Agenda includes:

1. The Vision and the challenges in Drug Discovery
2. Information Integration and the Knowledge Space (a Data GRID?)
3. The NIBR/IK@N HPC concept and status update.
4. E-Collaboration environment
5. Concluding remarks
Computational aspects of Drug Discovery

- Bioinformatics Lab
- Macromolecular Structure & Function Lab
- Computational Chemistry Lab

- Target finding
- Target validation
- Lead finding
- Lead optim.
In Silico Drug Discovery: areas of application

Human data → DNA Sample Sequencing → SNP → Translate & Map/Align → SAP → Model & Map → Structures & Modelling templates → Kinases NR Proteases

DB → DB → DB → DB → DB

Disease association
Validated Targets

Virtual Drug Discovery
- In Silico Docking
- In Silico “Chemogenomics”
- Virtual Library Design
- Predictive MedChem
- Tox PK/PK ADME modelling

Functional and Structural insights

Compounds → QSAR → Proteins
In silico drug discovery pipeline: major aims

• Ensure that we have a flexible software architecture which can accommodate any algorithm any time.
• Ensure low cost software infrastructure and freedom to operate.
• Ensure “GRID-ability” and full exploitation of HPC architecture despite hungry software providers.
• Promote internal algorithm development, new science and new ideas should be testable in known environment.
• Allow D0/D1/D2 scientists to run low cost “validations” in silico experiment.
• Break down the boundary between “biologists” and “chemists” and “screeners”.

NIBR / Manuel Peitsch
In Silico Drug Discovery: goals

• **Long Term Goal**
  - Develop a semi-automated, human-supervised In Silico Drug Discovery pipeline.

• **What for?**
  - Enable scientists to quickly and easily find compounds binding to a particular target protein
    - Use the compound as a tool to do pre-clinical target validation.
    - Use the compounds as hits for hits to leads programs.
    - Accelerate target selection and improve probability of success.

• **Specific plans for 2004**
  - Grow GRID computing beyond NIBR.
  - Various key components available at Novartis.
    - Protein modelling
    - Docking
    - Chemoinformatics
  - Create draft pipeline by chaining existing tools and methods; Explore adequate workflow.
In Silico Drug Discovery: the time axis

- 1990: 
  - Productive Automated Protein modelling email server

- 1995: 
  - SETI@Home
  - GeneCrunch

- 2000: 
  - 3D-Crunch
  - First PC-GRID at Novartis
  - Docking in production at Novartis

- 2005: 
  - Productive
  - In Silico Drug Discovery and Chemogenomics pipeline

Key Events:
- 1990: SETI@Home recognised as a leading new concept (ComputerWorld Award)
- 1995: SWISS-MODEL and 3D-Crunch recognised as a leading new concept (ComputerWorld Award)
- 2000: First automated Modelling and Docking pipeline at Novartis
- 2005: Various automated ToxCheck and other CIx services

Other Highlights:
- GeneCrunch recognised as a leading new concept (ComputerWorld Award)
- Full Transcriptome Modelling at Novartis
- PC-GRID using public resources
- Large-scale automated bioinformatics analysis of genome sequence/information
- Large-scale automated prediction of protein structures
The „Kitty Hawk Project“: will it fly?

Pave the way towards achieving new functionality by combining available components

17th December, 1903
12 seconds, 36 meters
Let’s make it fly!

December 17, 1903

Fly in 2005?

Grow
applicability

Optimize
components

Grow
reliability

New concepts

100 years of optimization and improvements
Agenda

- The Vision and the challenges in Drug Discovery
- Information Integration and the Knowledge Space (a Data GRID?)
- How can in silico science contribute to the Drug Discovery process and in pharma?
- E-Collaboration environment
- Concluding remarks
The NIBR/IK@N HPC concept and status update

- Key enabler of in silico science
- Please see Pascal’s presentation from yesterday.
Agenda

• The Vision and the challenges in Drug Discovery
• Information Integration and the Knowledge Space (a Data GRID?)
• How can in silico science contribute to the Drug Discovery process and in pharma?
• The NIBR/IK@N HPC concept and status update.

• Concluding remarks
E-Collaboration environment

- **Objectives**
  - Bring people together across sites, time zones and functional areas.
  - Enable collaborations with external partners, both in academia and business.
  - Foster collaboration and virtual projects → the right person on the right project → Availability and location are not skills!

**Status**
- Toolbox state. Each situation has a recommended tool.

- **Future**
  - Fully integrated into our work environment across all of Novartis.
Agenda

• The Vision and the challenges in Drug Discovery
• Information Integration and the Knowledge Space (a Data GRID?)
• How can in silico science contribute to the Drug Discovery process and in pharma?
• The NIBR/IK@N HPC concept and status update.
• E-Collaboration environment
General comments on GRIDs

- To create credibility and support from Senior Mgmt, it is crucial to avoid confusion between Crunch-GRIDs, Data-GRIDs, Information/Knowledge-GRIDs and People-Networks/GRIDs.
  - Clarity in concepts will help focus projects and concrete deliverables.
- They can all be part of a single multi-tiered and coherent strategic framework.
- Whichever you talk about, it will not be perfect for quite some time. Manage expectations!
  - Think “Kitty-Hawk”. The conceptual layers are there, prototype are emerging, improvements and optimization is essential, maturity will follow
- Leadership, transcendence, entrepreneurship and tenacity are the essence of transformation!
  - Concepts are easy to draw on a napkin over beer!
  - But new and great things are hard to achieve!
Comments on Crunch-GRIDs

- A component of an integrated HPC strategy
- Address unmet computational needs
- Create a fertile ground for new in silico science approaches
- Most current software licensing models must change and will change.
  - CREATE your Freedom to Compute!
  - A new opportunity for “freeware”
  - New algorithms and sw implementations are needed
    - Build more, buy less: Commercial software as an impediment to in silico science and its applications.
  - Stop believing that there is a successful business model for specialty scientific software.
Comments on Data-GRIDs

• A component of an integrated storage strategy
• Dissociate from Information/Knowledge GRIDs
• It is about:
  ➢ Replication, distribution vs centralization
  ➢ Optimal access for Crunch and Knowledge GRIDs
  ➢ “Federation” schemes for input and retrieval
• They require:
  ➢ Standards and shared Data as well as Meta-Data models
  ➢ QA&QC: curation requires data officers and discipline
  ➢ Undivided adherence to all shared standards, schemas, principles and data quality.
  ➢ Security
Comments on Information/Knowledge-GRIDs

- Layer above Data-GRIDs
- Expert systems, Semantic-Web/GRID

It is about:
- Ontologies, Semantic Networks, Thesauri and Vocabularies
- Categorization, Text Mining, Correspondence Tables, MetaStores, analytics.
- Context-based UltraLinking and UltraAction (workflows)

They require:
- Standards
- Maintenance of Ontologies, Semantic Networks, Thesauri and Vocabularies
- Undivided adherence to all shared standards, schemas, principles and data quality.
- Security
Comments on People-Networks/GRIDs

• Knowledge Management: explicit vs tacit
• eCollaboration and distributed teams (a team cannot be virtual!)

• It is about:
  ➢ Easy to use and well integrated eCollaboration services.
  ➢ Capability and not technology.

• They require:
  ➢ Standards
  ➢ Integrated into people’s workspace
  ➢ Mobility and the Martini Principle
  ➢ Security