Grid Activity in Korea for the discovery of Potential Inhibitors against Diseases on WISDOM Environment

- Hwa-Ja Ryu, Hee-Kyoung Kang, Hanh Thi Thanh Nguyen, Sehoon Lee, Soonwook Hwang, Ana Lucia Dacosta, Vincent Breton, Doman Kim

1 Chonnam National University, Korea, 2 KISTI, Korea
3 LPC-Clermont-Ferrand, campus des Cezeaux, 63177 Aubiere cedex, France.
Virtual screening: Use of high-performance computing to analyse large databases of chemical compounds in order to identify possible candidates.

Phases of a pharmaceutical development:

- Target discovery
  - Target Identification
  - Target Validation
- Lead discovery
  - Lead Identification
  - Lead Optimization
- Clinical Phases (I-III)

*in vitro* experimentation

*in silico* experimentation

Duration: 12 – 15 years, Costs: 500 - 800 million US $
Virtual Screening

- **Computational methods**
  - Pharmacophore based search
  - Structure based docking

- **Requirements**
  - 3D Structure of target
  - Databases of small molecules
  - A method to dock and score bound small molecule
Millions of chemical compounds available in laboratories

Chemical compounds: Chembridge – 500,000

Targets:
Human maltase glucoamylase (2QMJ)

High throughput virtual docking

High Throughput Screening 1-10$/compound, nearly impossible

Molecular docking (Autodock)

Computational data challenge

Hits screening using assays performed on living cells

 Leads

 Clinical testing

 Drug
Large Computational resources

Grid computing is applying the resources of many computers in a network to a single problem at the same time - usually to a scientific or technical problem that requires a great number of computer processing cycles or access to large amounts of data.

Example:
“EGEE (Enabling Grids for E-sciencE) is providing a production quality grid infrastructure spanning more than 30 countries with over 150 sites”
Database of small molecules

- Drug-like: MDDR (MDL Drug Data Report) >147,000 compounds, CMC (Comprehensive Medicinal Chemistry) >8,600 compounds
- Non-drug-like: ACD (Available Chemicals Directory) ~3 millions compounds
- CSD (Cambridge Structural Database, www.ccdc.cam.ac.uk): 264,000 compounds
- Corporates Databases: few millions in pharmaceuticals companies
- Virtual libraries (Combinatorial chemistry)
- ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 4.6 million compounds in ready-to-dock, 3D formats
- Chembridge database (www.chembridge.com): 454,000 compounds
Finding Inhibitors of Human Intestinal Maltase for the Treatment of Type 2 Diabetes
Human Intestinal Maltase (HMA)

- $\alpha$-glucosidase in the brush border of the small intestines responsible for digestion of maltose oligosaccharides into glucose

- Inhibition of the enzyme activity
  - → retardation of glucose absorption
  - → decrease in postprandial blood glucose level

- Important target in treatment of diabetes type 2 and obesity

- $\alpha$-glucosidase inhibitors – Acarbose (Glucobay), Miglitol (Glyset), Voglibose (Voglib) with side-effects

- Need to discover alternative inhibitors with greater potency and fewer side-effects
Binding information of acarbose with human intestinal maltase

Sim L et. al. 2008 J Mol Biol. 375(3):782-92
Filtration process

454,000 chemical compounds from Chembridge

Scoring based on docking score (308,307)

3016 compounds selected

Interaction with key residues

2616 compounds selected

Key interactions binding models clustering

42 compound selected

In vitro test
Statistics of datachallenge deployment on WISDOM production environment

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total numbers of docking</td>
<td>308,307</td>
</tr>
<tr>
<td>Total size of output results</td>
<td>16.3 GBytes</td>
</tr>
<tr>
<td>Estimated duration by 1CPU</td>
<td>22.4 years</td>
</tr>
<tr>
<td>Duration of experiments</td>
<td>3.2 days</td>
</tr>
<tr>
<td>Maximum numbers of concurrent CPUs</td>
<td>4700 CPUs</td>
</tr>
<tr>
<td>Crunching Factor</td>
<td>2556</td>
</tr>
<tr>
<td>Distribution Efficiency</td>
<td>54.4 %</td>
</tr>
</tbody>
</table>
Cloning and expression of human maltase in yeast

*Pichia pastoris*

**Conditions for HMA expression**

→ Culture 500 ml in 2 L flask at 30°C and 200 rpm
→ 0.5% methanol
→ ~4 days
→ enzyme reaction: 90 min at 37 °C (50 mM maltose)
# Inhibitory activity of the identified hits with HMA

<table>
<thead>
<tr>
<th>Compound No.</th>
<th>Lowest energy</th>
<th>M.W (g/mol)</th>
<th>clogP</th>
<th>Ki (µM)</th>
<th>IC50 (µM)</th>
<th>Type of inhibition</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>-16.43</td>
<td>473</td>
<td>3.04</td>
<td>19.8 ±1.2</td>
<td>58±4</td>
<td>competitive</td>
</tr>
<tr>
<td>18</td>
<td>-16.44</td>
<td>429</td>
<td>3.56</td>
<td>19.6±0.9</td>
<td>55±3</td>
<td>competitive</td>
</tr>
<tr>
<td>Acarbose</td>
<td>-12.62</td>
<td>645.605</td>
<td>-6.655</td>
<td>19.4</td>
<td>52±4</td>
<td>competitive</td>
</tr>
</tbody>
</table>
Hydrogen bond interactions with Key residues of two hit compounds in active site of protein

In active site of HMA

Hydrogen bond interaction of 2 compounds
Finding Inhibitors of Plasmepsin IV of Malaria
Malaria, a dreadful disease is caused by the protozoan parasite, *Plasmodium*. *Plasmodium* species include *P. falciparum*, *P. vivax*, *P. malariae*, and *P. ovale*. One of the crucial drug targets in malaria is plasmepsin. The aspartic proteases of *Plasmodium* species are known as plasmepsins. Plasmepsins are involved in the hemoglobin degradation inside the food vacuole during the erythrocytic phase of the life cycle. Ten different isoforms (PMI, II, III, IV, V, VI, VII, IX, X, and HAP) exist. Plasmepsin II and IV is responsible for initial attack on the hemoglobin α-chain between the residues **Phe 33 and Leu 34** in the hinge region.

![Hemoglobin degradation in Plasmodium falciparum](chart)

- **Hemoglobin (Hb)**
  - **Heme**
    - Oxidation
    - **Hematin**
      - Polymerization
      - **Hemzoin** (malarial pigment)
        - **Amino acids**
          - **Falcipain, plasmepsin**
          - **Falcilysin, aminopepdidases**
    - **Large fragments**
      - **Plasmepsins I, II, IV and HAP**
Activity assay of recombinant plasmepsin IV (PM IV)

Fig. Enzyme activity (%) as a function of the reaction time of PM IV with 5 µM FRET substrate.

Fig. Reaction velocity of PM IV as a function of [S]. The line represents fit of data points to the Michaelis-Menten equation.

PM IV ($K_m = 3.7 \pm 0.3$ µM and $k_{cat} = 0.869$ s$^{-1}$)

PM II ($K_m = 4.2 \pm 0.5$ µM and $k_{cat} = 0.894$ s$^{-1}$)
Table. Measured IC$_{50}$ values of the 30 tested compounds and of three reference inhibitors used for comparison

<table>
<thead>
<tr>
<th>Mol.</th>
<th>IC$_{50}$ (nM)</th>
<th>Mol.</th>
<th>IC$_{50}$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PM II$^{[b]}$</td>
<td></td>
<td>PM IV</td>
</tr>
<tr>
<td>1</td>
<td>305.1±1.5</td>
<td>18</td>
<td>187.1±3.1</td>
</tr>
<tr>
<td>2</td>
<td>5.5±2.0</td>
<td>19</td>
<td>n.i.</td>
</tr>
<tr>
<td>3</td>
<td>6.4±0.7</td>
<td>20</td>
<td>189.0±1.4</td>
</tr>
<tr>
<td>4</td>
<td>42.6±1.5</td>
<td>21</td>
<td>57.3±0.4</td>
</tr>
<tr>
<td>5</td>
<td>236.4±0.7</td>
<td>22</td>
<td>n.i.</td>
</tr>
<tr>
<td>6</td>
<td>145.2±2.4</td>
<td>23</td>
<td>87.5±0.1</td>
</tr>
<tr>
<td>7</td>
<td>4.3±0.6</td>
<td>24</td>
<td>4.4±0.8</td>
</tr>
<tr>
<td>8</td>
<td>62.1±0.6</td>
<td>25</td>
<td>122.9±1.1</td>
</tr>
<tr>
<td>9</td>
<td>118.1±1.9</td>
<td>26</td>
<td>146.4±1.0</td>
</tr>
<tr>
<td>10</td>
<td>8.8±0.8</td>
<td>27</td>
<td>201.1±1.3</td>
</tr>
<tr>
<td>11</td>
<td>n.i.</td>
<td>28</td>
<td>7.6±1.1</td>
</tr>
<tr>
<td>12</td>
<td>237.4±1.5</td>
<td>29</td>
<td>1831.3±1.9</td>
</tr>
<tr>
<td>13</td>
<td>1087.6±0.7</td>
<td>30</td>
<td>38.9±2.4</td>
</tr>
<tr>
<td>14</td>
<td>9.5±1.1</td>
<td></td>
<td>n.i.</td>
</tr>
<tr>
<td>15</td>
<td>96.1±0.2</td>
<td></td>
<td>RS367</td>
</tr>
<tr>
<td>16</td>
<td>30.0±1.8</td>
<td></td>
<td>RS370</td>
</tr>
<tr>
<td>17</td>
<td>n.i.</td>
<td></td>
<td>Pep.A</td>
</tr>
</tbody>
</table>

[a] IC$_{50}$ values taken from Asojo et al. (2002)
[b] IC$_{50}$ values taken from Degliesposti et al. (2009)
Finding Inhibitors of 3CL-pro of SARS
3CL-pro of SARS coronavirus

- The global outbreak of SARS (Severe Acute Respiratory Syndrome) in 2002-2003 with high mortality (app. 10%)

- Possibility of re-emergence in the future

- No existence of effective therapy for this viral infection

- 3CL-pro (chymotrypsin-like cystein protease) : Attractive target for the development of antiviral drugs directed against SARS-CoV
  - Essential for the viral life cycle
  - Availability of number of 3D structures
  - Preparation of the enzyme in large quantities for *in vitro* test
Redocking of 3CL-protease of SARS

Docking parameters:
Ga_run=50, Ga_pop_size=250, Ga_num_generation=27000, Ga_num_evaluation=5000000

<table>
<thead>
<tr>
<th>Name of ligands</th>
<th>Docking energy</th>
<th>RMSD (Å)</th>
<th>CPU time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CY6_2ALV</td>
<td>-16.47</td>
<td>3.23</td>
<td>12h50m49.70s</td>
</tr>
<tr>
<td>CYV_2QIQ</td>
<td>-16.23</td>
<td>2.52</td>
<td>15h13m34.45s</td>
</tr>
<tr>
<td>KCQ_2Z3E</td>
<td>-9.07</td>
<td>2.41</td>
<td>2h32m03.29s</td>
</tr>
<tr>
<td><strong>ZU3_2ZU3</strong></td>
<td><strong>-17.07</strong></td>
<td><strong>2.08</strong></td>
<td><strong>14h54m45.36s</strong></td>
</tr>
<tr>
<td>ZU3_2ZU4</td>
<td>-16.85</td>
<td>3.78</td>
<td>14h35m32.64s</td>
</tr>
<tr>
<td><strong>ZU5_2ZU5</strong></td>
<td><strong>-20.54</strong></td>
<td><strong>1.09</strong></td>
<td><strong>14h311m04.53s</strong></td>
</tr>
<tr>
<td>ECQ_2Z3D</td>
<td>-9.80</td>
<td>2.5</td>
<td>2h44m25.91s</td>
</tr>
<tr>
<td>OCQ_2Z3D</td>
<td>-10.42</td>
<td>2.29</td>
<td>3h11m07.29s</td>
</tr>
</tbody>
</table>

![CY6_2ALV](image1.png)  ![CYV_2QIQ](image2.png)  ![ZU3_2ZU3](image3.png)  ![ZU3_2ZU4](image4.png)  ![ZU5_2ZU5](image5.png)
Cloning and expression of 3CL-pro of SARS

Gene synthesis of 3CL-pro:

Amino acid sequence:

EcoRI SGFRKMAFSPGKVEGCMVQVTCTGTTTLNGLWLDDTVYCPRHVICTA
EDMLNPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPK
TPKYKFRVIRIQPQHTFSVLCYNGSPSGVYQCAMRPNHTIKGSLNGSGC
TPKYKFRVIRIQPQHTFSVLCYNGSPSGVYQCAMRPNHTIKGSLNGSGC
FNIDYDCVSFCYMMELPTGVHAGTDLEGKFGFVPQDRQTAQAAGTDTD
TLNVLAWLYAVINGDRFLNFRRTTLNDFNLVAMKYNLEPLTDHVDLGP
LSAQTVLCALMDCAALKEQLQNGMNGRTILGSTILEDEFTPDVVRQCSV

pPICZαA 3.6kb
3CL-932bp

Expression system: *Pichia pastoris*

- Cloning, expression and purification and characterization of 3CL-pro enzyme
- In vitro assay
Finding Inhibitors of Neuraminidase N1 of the *influenza virus* H5N1
Neuraminidase N1 of the influenza virus H5N1

- H5N1 influenza virus: highly pathogenic
  the potential of a constant threat
  (← its high fatality and resistance to the commercially available drugs (Tamiflu® and Relenza®))

- Needs of both effective vaccination and antiviral drug treatment

- Neuraminidase (NA): A glycoprotein on the virion surface
  → Release of the progeny virions from the infected host cells

- Inhibition of NA → Suppressing the viral growth

- In vitro assay of compounds selected from in silico screening by Dr. Wu in Academia Sinica, Taiwan
Cloning and expression of head domain of N1

Gene synthesis

- Amino acids sequence

VKLAGNSSLCPIINGWAVYSKDNSIRIGSKGDVFVIREPFISCSHLECRTFFLTQGALLNDKHSNGTVKD
RSPHRTLMSCPVGEPASPYPESFESVAWSASACH
DGTSWLTIGISGPDNGAVAVLKYNGITDTIKSWR
NNILRTQESECACVNGSCFTVMTDGPSNGQASH
KIFKMEKGVKSVKVELDAPNYHYEECSCYPNA
GEITCVCRDNWHGSNRPVSVSNQNLLEYQIGYIC
SGVFDNPRPNDGTGSCGPVSSNGAYGVKGFSF
KYMGNGVWIGRTKSTNSRSGFEMIWDPWTET
DSSFSVQDIQDIVTDSGYGSFQVQHPETTGLDC
IRPCFWVELIRGRPKESTIWTSGSSISFCVGVSDT
VGWSVPDGAELPFTIDK
Cloning and expression of neuraminidase and its mutant

DNA-plasmid

1. H274F
2. E119A
3. R292K
4. H274Y
5. E119D
6. E119V

Digestion with XhoI & NcoI
**Enzyme *in vitro* tests:**

Young-Min KIM (KRIIBB, Neuraminidases), Hee-Kyoung KANG (CNU, Plasmpsin), Thanh-Hanh NGUYEN (CNU, Data Challenge & *in vitro* test)

**In silico data challenge and analyses (WISDOM):**

- Academia Sinica, Taiwan  
  Simon LIN, Hsin-Yen CHEN, Ying-Ta WU et al.
- CNRS-IN2P3-LPC, Clermont-Fd, France  
  Vincent BRETON, Jean SALZEMANN et al.
- HealthGrid  
  Ana Da COSTA et al.
- SCAI-Fraunhofer Institute, Germany  
  Martin HOFMANN et al.
- Modena University, Italy  
  Giulio RASTELLI et al.
- KISTI, Korea  
  Soon-Wook HWANG, Se-Hoon LEE et al.