DRUG DESIGN, DISCOVERY AND DEVELOPMENT PLATFORM AND REPOSITORY (D4PR)

Ekene Ezeasor, Tochukwu Eze and Ikemefuna Uzochukwu
Nnamdi Azikiwe University, Awka, Nigeria – (https://d4pr.unizik.edu.ng)
(ezeasor.ekene@unizik.edu.ng, td.eze@unizik.edu.ng, ic.uzochukwu@unizik.edu.ng)
WACREN e-Research Hackfest – Lagos (Nigeria)
Outline

- Abstract
- Scientific problem area
- Identified Problems
- Benefits of using D4PR on the web
- End User Work Flows
- Data Origins, Ingestion and Management
- Process Automation
- D4PR Features
- Computing Model
- Implementation strategy
- Risks and Unknown
- References
ABSTRACT

Nnamdi Azikiwe University-Drug Design and Informatics Group (NAU-DDIG) has established a platform for finding new drug candidates from natural products, existing drugs or chemical databases. We have implemented our Computer Aided Drug Design (CADD) protocols and obtained published and some unpublished results for Neglected Tropical Diseases (NTDs) such as ascariasis, Ebola virus disease, malaria, sickle cell disease and Shiga Toxin producing *Escherichia coli* (STEC) etc. We intend to design and develop a Drug Design, Discovery, Development Platform and Repository (D4PR). D4PR (https://d4pr.unizik.edu.ng) will be a web-based platform containing an integrated set of tools, applications, data repositories etc., which can be accessed via D4PR portal. We intend implementing the front-end of the proposed platform with HTML, CSS and JavaScript while incorporating Bootstrap technology. The back-end will be implemented using PHP scripting language, while integrating standard technologies such as Future Gateways, Open Access Repository, SAML and LDAP for authentication and authorization. The developed application will be hosted on an online web server ported to African Grid Science Gateway (AGSG) cloud computing e-infrastructure. Our platform will help to address the disease burdens, especially from NTDs, in Nigeria and Africa in general.
SCIENTIFIC DOMAIN ADDRESSED

✓ Health
  ✓ Neglected Tropical Diseases (NTDs)
  ✓ Non-communicable diseases (Diabetes mellitus)

✓ Drug Design, Discovery and Development
  ✓ Traditional approaches: Bioprospecting, Serendipity etc.
  ✓ Cost, ≈ $1B and time, ≈ 10 years per drug (Trader, 2014)
  ✓ **Computer Aided Drug Discovery (CADD)**
    • Unpublished Results for malaria, diabetes, sickle cell disease, STEC etc.
  ✓ **Drug Repurposing (Naylor, 2014)**
    ✓ Short circuits the drug R & D process in comparison to de novo drug discovery
    ✓ Cuts down cost & time needed for drug development
    ✓ Near 100% probability of technical success at the discovery stage
    ✓ Drug repurposing for Ascariasis (Uzochukwu et al. 2014)
    ✓ Drug repurposing for Ebola
      (http://www.unesco-mars.com/mars_award_winners.php)

“In 11 hours, the team completed the equivalent of 39 drug design years for a cost of $4,232. During that run, Novartis screened 10 million compounds and identified three promising candidates, … life-saving cancer drug” (Trader, July 2014).
IDENTIFIED PROBLEMS THAT D4PR TRIES TO SOLVE

✓ Process automation
✓ Decision support
✓ Reduction in computational cost
✓ Reduction in computational time
✓ Non-visibility (globally)
✓ Capacity of local investigators/scientists
✓ Collaborative research
BENEFITS OF USING D4PR ON THE WEB

✓ Time and cost saving (Ebola example)
✓ Capacity building in drug discovery, design and development (NAU-DDIG example)
✓ Improved collaborative research and sharing (UI example)
✓ Affords global visibility and awareness
✓ Improved e-learning.
✓ New medicines and improved health outcomes
✓ Boosting production capacities, generating economic growth with attendant multiplier effects (Novartis example)
✓ Strengthening local technical capacity in e-infrastructure
END USER WORK FLOWS

- User login through an identity provider
  - Perform data manipulation operations
  - Perform graphical pattern discovery analyses
- Export data for decision support
- Submit contents (bioscores, binding energy values, project proposals, project documents, multi-media)
<table>
<thead>
<tr>
<th>S/N</th>
<th>DATA</th>
<th>DATA ORIGINS</th>
<th>INGESTION</th>
<th>MANAGEMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phytoconstituents</td>
<td>Local Researchers and review papers</td>
<td>Obtain, process bioscores, store as xlsx</td>
<td>Manual curation, local storage and back up</td>
</tr>
<tr>
<td>2</td>
<td>Existing drugs</td>
<td>DrugBank</td>
<td>Obtain, process bioscores, store as xlsx</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MMV Pathogen Box</td>
<td>Multilateral Malaria Ventures</td>
<td>Obtain, process bioscores etc., store as xlsx</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Binding affinities</td>
<td>Local simulations</td>
<td>Run molecular docking, copy energy values, store as xlsx</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Bioscores</td>
<td>Molinspiration</td>
<td>Process smiles, copy values, store as xlsx</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ADMET parameters</td>
<td>ZINC, DrugBank</td>
<td>Obtain, copy values, store as xlsx</td>
<td></td>
</tr>
</tbody>
</table>
PROCESS AUTOMATION

✓ Bioscore analyses
✓ Binding affinity analyses
✓ Compound Data Mining
D4PR FEATURES

- User Authentication Feature
- Database Query window/data manipulation
- Drug discovery using reference drug bioscores
- Graphical pattern simulation
- Data export and import
- Uploading/retrieval of molecular dynamics trajectories
- Content submission (bioscores, binding energy values, project proposals, project documents, multi-media)
COMPUTING MODEL

- Cloud Computing infrastructure e.g. African Grid Science Gateway (AGSG)
- Online Apache Web Server
- Datastore - MySQL database

Current Working Environment
- A dual core desktop with 4gb RAM, 500GB hard-disk, 32Bit System with 3.0Ghertz of processing speed
  - Molecular docking run time: from 30 sec to 1hr per compound depending on torsions;
  - Molecular dynamics run time: from 5 days to 1 week or more
- Group members' laptops
IMPLEMENTATION STRATEGY

• Authentication and Authorization: SAML, LDAP
• Repository: OAR
• Data Processing and Manipulation: FutureGateway, PHP
• Other Tools: Bootstrap, Jquery, LAMP, JSON
IMPLEMENTATION STRATEGY .....cont'd

Researcher

Federated Identity

D4PR Portal

Future Gateway

Data Storage

OAR

Africa Grid
Science Gateway

Sci-GaIA
www.sci-gaia.eu
RISKS AND UNKNOWNS

✓ Epileptic power supply – may lead to damage to infrastructure and downtime
✓ Attitudes
✓ Poor Internet bandwidth – increases process time and downtime
EAGER TO START CODING WITH GUIDANCE...!
REFERENCES


✓ UNESCO MARS Award: http://www.unesco-mars.com/mars_award_winners.php

REFERENCES ..... (past events)
REFERENCES ..... (past events)
Thank you!

sci-gaia.eu
info@sci-gaia.eu