RoadRunner
A publicly available bioactivity database

Steve Mathias, Jeremy Yang, Cristian Bologa, Tudor Oprea

UNM Biocomputing
MLI in Numbers

NIH Roadmap Initiative

Molecular Libraries Initiative

4 Chemical Synthesis Centers

CombiChem
Parallel synthesis
DOS
4 centers + DPI
100k–500k compounds

MLSCN (9+1)
9 centers
1 NIH intramural
100 x 10 = 1000 assays

PubChem (NLM)

ECCR (6)
Exploratory Centers

Predictive ADMET (8)

~300,000 compounds

SAR matrix

> 1000 assays

http://nihroadmap.nih.gov

Not renewed
NM MLSC (summary of 3 years)
U54MH074425

• 23 primary targets (55 assays) screened
• 38 targets total pipeline
• ~2 million datapoints loaded into PubChem
• Current throughput: 150,000 samples/week
• first 6-plex (small GTP-ases) in the Roadmap
• 2nd 6-plex (Bcl-2) near completion
• ~2.4 million datapoints awaiting upload
• 33 peer-reviewed papers (published, in preparation) associated with the NM MLSC grant
• 8 new chemical probes reported
Integrated Discovery Teams

Target Development
Eric Prossnitz
Larry Sklar
Bruce Edwards
Dan Cimino*

Screening and Automation
Bruce Edwards
Danuta Wlodek
Susan Young
Irena Ivnitski-Steele
Mark Carter*
TBN
Herbert Tanner
Deepti Kumar

Bead Assemblies
Peter Simons
Eric Prossnitz
Zurab Surviladze
Anna Waller*
Tione Buranda**
Yang Wu**
TBN

Probe Chemistry
Jeff Arterburn (NMSU)
James Herndon (NMSU)
Alex Kornienko (NMT)
Matt Parker (ChemDiv)
Ilya Okun (ChemDiv)
Wei Wang

Cheminformatics
Tudor Oprea
Cristian Bologa*
Steve Mathias*
Jeremy J. Yang*
Dan Fara*
Oleg Ursu
Andrei Leitão
Ramona Rad
Lili Ostopovici
Srinajana Chemburu*
V. Niranjan Kumar*

Administrative
Virginia Salas
Rae Ramirez*
Terry Foutz*
PART TIME**
CONSULT**
HyperCyt

384 wells/10 min
1 μl/sample

**Some Screens From NMMLSC**

- **FPR Ligands**
- **FPRL1 Ligands**
- **VLA-4 Allosteric act/inhib**
- **Bacterial Virulence Inhibitors**
- **Proteasome degradation act/inhib**
- **GPR30 Ligands**
- **ERα Ligands**
- **ERβ Ligands**
- **Prostate Cancer differentiation**
- **GTP-ase 6-plex, GTP-ase 4-plex**
- **ABC Efflux Pump Duplexes (x2)**
- **Prostate cell differentiation**
- **Bcl-2 Family 6-plex**

- **SAR by commerce**
- **Virtual Screen**

<table>
<thead>
<tr>
<th>Selection</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPR Ligands</td>
<td>cell</td>
</tr>
<tr>
<td>FPRL1 Ligands</td>
<td>cell</td>
</tr>
<tr>
<td>VLA-4 Allosteric act/inhib</td>
<td>cell</td>
</tr>
<tr>
<td>Bacterial Virulence Inhibitors</td>
<td>cell</td>
</tr>
<tr>
<td>Proteasome degradation act/inhib</td>
<td>bead</td>
</tr>
<tr>
<td>GPR30 Ligands</td>
<td>cell</td>
</tr>
<tr>
<td>ERα Ligands</td>
<td>cell</td>
</tr>
<tr>
<td>ERβ Ligands</td>
<td>cell</td>
</tr>
<tr>
<td>Prostate Cancer differentiation</td>
<td>cell</td>
</tr>
<tr>
<td>GTP-ase 6-plex, GTP-ase 4-plex</td>
<td>bead</td>
</tr>
<tr>
<td>ABC Efflux Pump Duplexes (x2)</td>
<td>cell</td>
</tr>
<tr>
<td>Prostate cell differentiation</td>
<td>cell</td>
</tr>
<tr>
<td>Bcl-2 Family 6-plex</td>
<td>cell</td>
</tr>
</tbody>
</table>
Discovery of a Potent GPR30 Agonist

Virtual and biomolecular screening converge on a selective agonist for GPR30

Cristian G Bologa¹,⁷, Chetana M Revankar²,³,⁷, Susan M Young³, Bruce S Edwards³,⁴, Jeffrey B Arterburn⁵, Alexander S Kiselyov⁶, Matthew A Parker⁶, Sergey E Tkachenko⁶, Nikolay P Savchuk⁶, Larry A Sklar³,⁴, Tudor I Oprea¹ & Eric R Prossnitz²,³

Estrogen is a hormone critical in the development, normal physiology and pathophysiology⁴ of numerous human tissues. The effects of estrogen have traditionally been solely ascribed to estrogen receptor α (ERα) and more recently ERβ, members of the soluble, nuclear ligand-activated family of transcription factors³. We have recently shown that the seven-transmembrane G protein-coupled receptor GPR30 binds estrogen with high affinity and resides in the endoplasmic reticulum, where it activates multiple intracellular signaling pathways⁴. To differentiate between the functions of ERα or ERβ and GPR30, we used a combination of virtual and biomolecular screening to isolate compounds that selectively bind to GPR30. Here we describe the identification of the first GPR30-specific agonist, G-1 (1), capable of activating GPR30 in a complex environment of classical and new estrogen receptors. The development of compounds specific to estrogen receptor family members provides the opportunity to increase our understanding of these receptors and their contribution to estrogen biology.

Virtual & Biomolecular Screening Workflow

**ER compounds**

- **Initial set** (144k library)
- ROCS
- Fingerprint
- **Docking**
- 202 compounds
- **Emory assays** 713, 733, 737
- 418 compounds
- Composite set (620 compounds)

**G-like compounds**

- **G-1**
- Substructure search
- 56 compounds

A. Leitão et al., *in preparation*
There are 14 potential probes for 3 estrogen receptors.

Of these, we identified 7 types so far.

We’ve only begun to address poly-pharmacology.
Other Probes Reported by UNM

PI: Bruce Edwards
FPRL1 Antagonist **BB-V-115**
PubChem CID **6622773**
Ki = 174 nM

PI: Bruce Edwards
FPR Antagonist **3570-0208**
PubChem CID **3092570**
Ki = 112 nM

PI: Eric Prossnitz
GPR30 Antagonist **G-15**
PubChem CID **3136844**
K_d = 20 nM

PI: Todd Thompson
Prostate Cell Differentiation Activator Assay
PubChem CID **3240581**
EC_{50} = 770 nM

PI: Hattie Gresham
Small Molecule Inhibition of Staphylococcus Aureus Virulence
PubChem CID **3240990**
EC_{50i} = 125 nM
Roadrunner History

- Given the NM MLSC output, the Cheminformatics team was under pressure to address unmet needs
  - Homegrown project started ca. 2005

- Roadrunner serves informatics requirements from the Biology, Assay Development, Screening, Cheminformatics and Chemistry components of the **NM MLSC** (part of the **NIH Roadmap**)

- Uses:
  - PostgreSQL (http://www.postgresql.org/),
  - CHORD from gNova (http://www.gnova.com/).
  - OEChem from OpenEye (http://www.eyesopen.com/)

- Key: Simplicity of use; web-based
Roadrunner remains in cheminformatics
Its non-commercial codes are open-source & available by request
Roadrunner Capabilities

• Handles Assays from small molecule upload to bioactivity data upload (primary; secondary; dose-response plotting – coming soon)
• Chemically cognizant: substr. search, similarity, property calculation, error detection, etc.
• Compound tracking (e.g., volume); plate mapping (heatmap); other HTS Informatics tools
• Post-HTS Analyses: from primary screen to SAR by Commerce and follow-up chemistry
• Tight integration with PubChem
Cheminformatics in Roadrunner

• Implemented: MDL’s 320 and Sunset’s 560 (SMARTS based) fingerprints
• Tanimoto, Tversky, Euclid & Hamming similarity
• Filters (properties, undesired SMARTS)
• Batch similarity via data fusion
• Post-HTS analyses via Scaffold (chemical perception), Clustering & MCS
• These 3 methods may lead to different results
• Goal: enable interactive tools to support hit and probe discovery
Welcome to Roadrunner @ NMMLSC

This Application is for the use of New Mexico Molecular Libraries Screening Center members only. A public Roadrunner demo is available here.

Roadrunner is a chemical database application developed and supported by the Informatics Core of the New Mexico Molecular Libraries Screening Center at the University of New Mexico Health Sciences Center. The Roadrunner Project home page is here.

The Roadrunner NMMLSC database currently contains 222107 substances representing approximately 215175 unique compounds. This represents the NIH Molecular Libraries Small Molecule Repository (MLSMR) from DPI as downloaded from PubChem on 17 September 2007. The screening sets currently (and previously) being screened by the NMMLSC are represented by the various DPI libraries in the following table:

<table>
<thead>
<tr>
<th>Library</th>
<th>Substances</th>
<th>Unique Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLSMR</td>
<td>222107</td>
<td>215175</td>
</tr>
<tr>
<td>DPI-144k</td>
<td>138773</td>
<td>138303</td>
</tr>
<tr>
<td>DPI-100k</td>
<td>97559</td>
<td>97432</td>
</tr>
<tr>
<td>DPI-17k</td>
<td>16322</td>
<td>16324</td>
</tr>
<tr>
<td>DPI-10k</td>
<td>9993</td>
<td>9982</td>
</tr>
</tbody>
</table>

A prototype system for automated post-HTS analysis has been implemented. Try it out and please provide feedback!

Substructure searching with SMARTS!

Please Login to begin using the Roadrunner application.
This page has links to canned queries which allow you to easily browse subsets of data in the Roadrunner database.

**Browse By Library**

- DPI-100k (97559 Substances)
- DPI-10k (9993 Substances)
- DPI-144k (138773 Substances)
- DPI-17k (16322 Substances)
- MLSMR (222181 Substances)
- QS DPlcp1 (399 Substances)
- QS SARI (240 Substances)

**Browse By Activity**

- Substances with Active outcomes in any Assay
- Substances with no Active outcomes in any Assay

**Browse By Compound**

- Lipinski Rule of 5 Compounds This is not a very selective screen on this database.
Structure

Data

ID: 222109
Name: 2-[(3-ethoxy-carbonyl-6-methoxy-quinolin-4-yl)amino]benzoic acid
Synonym: UNM000011032601
Molecular Formula: C_{20}H_{18}N_{2}O_{5}
Molecular Weight: 366.40
SMILES: CCOC(=O)c1ccnc2cccccnc3ccc3c(=O)OOC
Library: QS SAR1
Supplier: ChemDiv(000A-0846)

Links

Compound: 216620
Plate Location: 1 Plate
PubChem Substance: 26746690

Bioactivity

Assays: 2 Assays 2 Composite
Bioactivity: 1 Result

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aliphatic Atoms</td>
<td>4</td>
</tr>
<tr>
<td>ALogP</td>
<td>3.79</td>
</tr>
<tr>
<td>ALogS</td>
<td>-4.38</td>
</tr>
<tr>
<td>Andrews Binding Energy</td>
<td>21.4</td>
</tr>
<tr>
<td>Barone and Channon Complexity</td>
<td>543</td>
</tr>
<tr>
<td>Carbons</td>
<td>20</td>
</tr>
<tr>
<td>Charge</td>
<td>0</td>
</tr>
<tr>
<td>Charge Count</td>
<td>0</td>
</tr>
<tr>
<td>Chiral Centers</td>
<td>0</td>
</tr>
<tr>
<td>CLogP</td>
<td>6.385</td>
</tr>
<tr>
<td>Functional Groups</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halide Fraction</td>
<td>0.00</td>
</tr>
<tr>
<td>H-Bond Acceptors</td>
<td>7</td>
</tr>
<tr>
<td>H-Bond Donors</td>
<td>1</td>
</tr>
<tr>
<td>Heavy Atoms</td>
<td>27</td>
</tr>
<tr>
<td>HetC Ratio</td>
<td>0.35</td>
</tr>
<tr>
<td>Heteroatoms</td>
<td>7</td>
</tr>
<tr>
<td>Lipinski Acceptors</td>
<td>7</td>
</tr>
<tr>
<td>Lipinski Donors</td>
<td>2</td>
</tr>
<tr>
<td>Max. Ring Size</td>
<td>6</td>
</tr>
<tr>
<td>Min. Ring Size</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Rings</td>
<td>3</td>
</tr>
<tr>
<td>Polar Surface Area</td>
<td>97.22</td>
</tr>
<tr>
<td>PSA w/ S and P</td>
<td>97.22</td>
</tr>
<tr>
<td>Rigid bonds</td>
<td>20</td>
</tr>
<tr>
<td>Ring Degrees of Freedom</td>
<td>3</td>
</tr>
<tr>
<td>Rotatable Bonds</td>
<td>7</td>
</tr>
<tr>
<td>Simple Molecular Complexity Metric</td>
<td>44.283</td>
</tr>
<tr>
<td>Solubility</td>
<td>poor</td>
</tr>
<tr>
<td>Unbranched Atoms</td>
<td>2</td>
</tr>
</tbody>
</table>
Jeremy Yang will flash-demo Roadrunner

Post-HTS Analysis Search Results

Found 10 Post-HTS Analyses Run By smathias@unm.edu

<table>
<thead>
<tr>
<th>#</th>
<th>Assay</th>
<th>Outcome</th>
<th>Method</th>
<th>Timestamp</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dose Response Assay for Inhibitors of Bacterial Quorum Sensing</td>
<td>active</td>
<td>cluster_mcss</td>
<td>2007-12-13 16:13:22.812483</td>
<td>Report</td>
</tr>
<tr>
<td>2</td>
<td>Dose Response Assay for Formylpeptide Receptor (FPR) Ligands and Dose Response Counter-Screen Assay for Formylpeptide-Like-1 (FPR1L) Ligands</td>
<td>active</td>
<td>cluster_fp</td>
<td>2007-12-13 16:30:51.189134</td>
<td>Report</td>
</tr>
<tr>
<td>3</td>
<td>HTS to identify specific small molecule inhibitors of Ras and Ras-related GTPases specifically Cdc42 wildtype</td>
<td>active</td>
<td>cluster_fp</td>
<td>2007-12-13 16:50:25.92729</td>
<td>Report</td>
</tr>
<tr>
<td>4</td>
<td>HTS to identify specific small molecule inhibitors of Ras and Ras-related GTPases specifically Rab7 wildtype</td>
<td>active</td>
<td>cluster_mcss</td>
<td>2007-12-13 16:55:18.408324</td>
<td>Report</td>
</tr>
<tr>
<td>5</td>
<td>Primary-HTS Assay for Inhibitors of Bacterial Quorum Sensing</td>
<td>active</td>
<td>cluster_fp</td>
<td>2008-01-08 14:33:08.569532</td>
<td>Report</td>
</tr>
<tr>
<td>6</td>
<td>Primary-HTS Assay for Inhibitors of Bacterial Quorum Sensing</td>
<td>active</td>
<td>cluster_fp</td>
<td>2008-01-10 12:35:41.51849</td>
<td>Report</td>
</tr>
<tr>
<td>7</td>
<td>Primary-HTS Assay for Inhibitors of Bacterial Quorum Sensing</td>
<td>active</td>
<td>cluster_fp</td>
<td>2008-01-10 12:48:44.402701</td>
<td>Report</td>
</tr>
<tr>
<td>8</td>
<td>Primary-HTS Assay for Inhibitors of Bacterial Quorum Sensing</td>
<td>active</td>
<td>cluster_fp</td>
<td>2008-01-10 13:05:01.06292</td>
<td>Report</td>
</tr>
<tr>
<td>9</td>
<td>Primary-HTS Assay for Inhibitors of Bacterial Quorum Sensing</td>
<td>active</td>
<td>cluster_fp</td>
<td>2008-01-10 14:30:32.069364</td>
<td>Report</td>
</tr>
<tr>
<td>10</td>
<td>Primary-HTS Assay for Inhibitors of Bacterial Quorum Sensing</td>
<td>active</td>
<td>cluster_fp</td>
<td>2008-01-10 14:38:52.593655</td>
<td>Report</td>
</tr>
</tbody>
</table>
NIH Funding Cheminformatics?

- Obs: Not about docking, molecular dynamics and other computational chemistry tools, but about “cheminformatics”!
- INChI (IUPAC/NIST → Wikipedia, ChemSpider) ← Not NIH
- Pubchem, ChemBank (public repository systems) & others
- RoadRunner (by-product of a Screening Center; “open-source”)
- Some cheminformatics funded under “Bioinformatics” as well
- Six Cheminformatics Exploratory Centers (http://neccr.org/)
- Funded to develop full centers, not to provide public resources
  - Indiana: David Wild
  - Massachusetts: Paul Clemons
  - Michigan: Kerby Shedden
  - New York: Curt Breneman
  - North Carolina: Jackie Hughes-Oliver
  - North Carolina: Alex Tropsha
- These P20s are on their final quarter of a 3-year period…
- Is there a future for cheminformatics at NIH?
Future of MLI

NIH Roadmap Initiative

Molecular Libraries Initiative

- 4 Chemical Probe Synthesis Centers
- MLPCN (3+1) HTS Major Centers
- 2-4 Specialty Centers
- 100 x 10 = 1000 assays
- PubChem (NLM)
- CCR (1?) Full Center(s)
- Predictive ADMET (?)

- MLSMR accepts contributions from ~20 synthetic chemistry
  ~10 nat. prod chemistry
  Outreach program

- ~500,000 (?) compounds

- Evolving SAR matrix

- Who will analyze the data?

> 1000 assays
Systems Chemical Biology

One Possible Future for Cheminformatics
The Systems Chemical Biology Interface

Oxaloacetate

EC 1.1.1.37 malate dehydrogenase

EC 2.3.3.9 malate synthase

S-Malate

EC 2.3.3.1 citrate synthase

Citrate

EC 4.2.1.3 cis-aconitase

cis-Aconitate

EC 4.2.1.3 cis-aconitase

Isocitrate

EC 4.1.3.1 isocitrate lyase

Glyoxylate

Enter Ligand for Simulation:

Sketch Structure Here

SMILES

Name

Search Clear

Query Target Structure:

Primary Sequence

SwissProt Nr.

Target Name

PDB Only

Search Clear

Prepare BioXyce Run:

Choose Pathway

Update Kinetics

Use Inhibitors

Run DAKOTA Start Clear

Systems Chemical Biology integrates small molecule modeling with biological networks simulation: BioXyce confirms that blocking malate synthase depletes isocitrate lyase – which makes malate synthase a target for drugs against latent tuberculosis.