

MLSCN SC Meeting

Day 1

WELCOME

Welcome – Dr. Insel

- Annual Molecular Libraries Screening Center Network (MLSCN) Steering Committee meeting with the Exploratory Centers in Cheminformatics Research (ECCRs) and the External Scientific Advisory Panel. The Advisors are knowledgeable experts from academia and industry. They're input will help us decide the course of the initiative and suggestions for the Mid-Course evaluation.
- Roadmap was designed as scientific incubator space. The MLSCN is essential as a Roadmap initiative to help transform science and develop new tools for biology. The NIH intends to carry the MLSCN through until 2008.

Hugh Rosen

- This SC Meeting, the centers will provide a subjective presentation of our accomplishments and understanding our current state; the ESP will provide an objective analysis and create measurable metrics to move the network forward
- Ultimate goal of this meeting is to generate real measurable steady state metrics that lead us forward to year three to gage progress for future re-competition
- Joint meeting with the ECCRs will help us communicate the output of the centers to the broader scientific community

SMR UPDATE

MLSMR: Biofocus DPI— Doug Livingston

- Business transition: Reverse merger with infinity pharmaceuticals cash for stocks swap. DPI operations now merged intact to BiofocusDPI
- DPI: public company since 2000, \$80M
- Infinity
 - o Top tier private company
 - o Two phase-1 cancer studies, phase 2 imminent, need timely funding for clinical trials, good pre-clinical pipeline, public access via reverse merger
 - o Infinity also made sacrifices for the merger
- Galapagos will pay \$5.4M in cash
 - o Now DPI is large enough to compete with the rest of the world
 - o 4 sites: Netherlands, UK, Germany, USA
 - o Huge asset for natural products
- SBS Article
 - o Compounds brought in by 4 broad themes
 - Diversity Set (DS): classical collection. Question of representative diversity
 - Targeted Libraries (TL): Targeted on the basis of the claims of the vendors, tremendous opportunity for design

- Natural Products (NP): expensive, what defines a natural product?
List based approach based on dictionary and resources of NP
- Known Bioactive Compounds (Specialty Sets): contains known drugs, toxins, metabolites. problem with inconsistent naming between databases

MLMSR Operations – Tim Lease

- Compound Acquisition
 - Pie chart shows the relative % of each set (DS, TL, NP, SS).
 - As of now, there is nothing from Heidelberg in the collection.
Distribution of the first set of NP will be by this September
 - Commercial sources
 - Academic sources from CMLD centers and solicitation
 - Selection is in micro-clusters and functionality filters
 - First 100K compounds contained fewer DS and Tim Lease. The NIH selected 130K compounds, but only 90K passed. 24K failed DPI's QC procedure. Filling in 66K with clusters and NP/known bioactives (1000 in process of purchasing by sigma). 2nd 100K will come in by the first half of next year. 200K by this time next year.
- MLMSR Sample Stores
 - Three sets, majority in Long-term store in N₂ -20°C, Working store DMSO solution in 96-format, short term storage prior to sending compounds off to the centers
 - DPI sends pre-tiered bar-coded vials to the vendors. Compounds are then filled. DPI dissolves compounds into a volatile solvent system and transfers some of that solution to a plate which is dried down and sent for QC. Sample in polypropylene tubes, dry down.
 - DMSO in inert N₂ atmosphere to prevent water uptake from the air and in -20°C freezer
 - Ready to plate to send to the SC, retrieve compounds en mass or cherry pick. Plate under N₂ and store in freezer till ready to send out.
- QC is done at the repository
 - Weight, solubility, Impurity
 - 5-10mg
 - QC solubility to ensure proper transfer. Sample is solubilized into chloroform methanol, a highly volatile solvent which is easy for high throughput. Compounds need to be in solution or in fine, uniform solution. 5% of the compounds are insoluble in Chloroform methanol but are soluble in DMSO.
 - Question: does DPI measure water solubility for fine uniform solution?
 - Answer: DPI uses software to calculate aqueous solubility
 - LC-MS incoming store samples and when centers ask for compound for hit follow up. Purity 90% AUC by UV
 - MS gives evidence for, not proof, for identity
 - LC is one indication of purity, but not an absolute purity measure.

- Each method gives different purity indication
 - In DPI's lab differences between ELS and UV depends on wavelength used
 - 6 minute runs for HT QC
- Data Processing
 - Use standards and blanks
 - AUC: ELS and UV aligned
- Discussion of SMR QC issues
 - NCGC ran independent QC measures. The first set of results matched DPI's values; the second set of results did not.
 - [Jim Inglese] NCGC does not think the differing QC results is a serious problem
 - Pittsburgh also ran independent QC measures. They are concerned that the purity of the cherry picked set of compounds was inconsistent with DPI. Pitt had a much longer LC run time. The samples will be sent back to DPI and re-analyzed.
 - [Pittsburgh] requested 6 chemically distinct compounds from DPI. The chemistry requires the integrity of the compounds prior to launching a chemistry campaign. Most of the compounds retained biological activity, but chemical analysis did not hold up. Extensive QC plans, 3/6 of compounds had less than 80% purity.
 - Scripps has had the same general concerns
 - [Catherine Peishoff] at CATHERINE PEISHOFF, we re-synthesize or re-purify for full re-confirmation of structure before proceeding
 - NOTE: Centers do not have access to the synthetic screens
 - [Jamie Driscoll] HTS WG decided that DPI will send out new samples from the working store that meet the original 90% purity scheme. In the interest of time, centers will work with samples of lesser purity until higher purity samples can be acquired.
 - [Doug Livingston] centers should try to pursue purchasing compounds from the original supplier
 - When the individual centers are doing their own QC, is there any concern about each center's storage procedure?
 - Answer: Individual storage procedures may effect the QC measurements. DPI plans to follow up on procedure at each of the centers.
 - **Action Item: centers which internally QC incoming compounds, please send results to DPI**
 - Does the chloroform go through aluminum to remove any residual non-volatiles (HCL)? Concerned with long-term storage in the chloroform methanol with impurities.

MLSMR Operations [continued]

- Maintenance QC: 10% of the library will be sampled each year to gage degree of compound degradation
 - o Using lot acceptance testing to pick batches or lots of compounds. Choose statistical sample and see the number of failures/passes. Lot size DPI is choosing is 10%. 14 lots of 5K tubes based on manufacture dates, lot definition may evolve to structure, etc.
 - o 4% AQL (failure) means 96% pass
 - o Can select for whatever compound attributes they want
 - o Must accept or reject entire lot
 - o Analyzed 2 lots so far (200 compounds of 5000 samples in the lots). 14> = acceptance, 15< = fail
 - 1st set 189 passed, 11 failed lot passes
 - 2nd set 191 passed, 9 failed lot passes
 - No concentration studies here, but DPI plans to do other concentration studies
- Compound delivery
 - o DPI's website contains a password protected PDF document
 - o Centers have two options for compound delivery; single shipment of the full set delivered once a year with updates staggered throughout or two separate shipments twice a year with updates in second shipment
 - o 455 nMol total as 10mM in DMSO
 - o 121 total shipments.
 - o All centers have received 10K started sets.
 - o 17 hit follow-up cherry picked sets
 - o Lipinski presentation
 - Sample purity/integrity for HTS
 - Compound quality is contingent on storage conditions at the centers
 - DMSO exposed to the environment soaks up H₂O and reduces solubility of the compounds, causing compounds to precipitate out of solution. As compounds thaw, even more compound comes out of solution.
 - o HTS WG suggestions:
 - Minimize freeze thaw cycles by making daughter plates upon first thaw
 - Store working copy at room temp and discard after 3 months
 - Work in dry environment
 - o DPI has offered to send out daughter plates in order to remove liquid handling responsibilities from the centers. This will result in an increased number of shipments each year. Two options increase the concentration and reduce the volume (45uM x 1 mL) or reduce the concentration and increase volume (4.5uM x 10 mL)
- Discussion
 - o [Larry Sklar] it would be convenient if DPI could do a hybrid of delivery options.

- DPI is open to the most convenient process for delivery.
 - **Action Item: send suggestions to DPI for compound delivery**
- Should DPI over-compensate compound selection to take into account attrition rate?
 - DPI accounted for this in the second 100K compound acquisition attempt
- Apart from QC issues, why was there such a high compound dropout rate for the first cycle of compound acquisition?
 - A difficult vendor and a vendor who promised compound and then reneged as they ran out of compounds
- How is DPI filling in 1st set of 100K?
 - Differences in criteria; relaxed excluded functionality filters but maintained selection criteria. DPI is filling in clusters of compound singletons in the existing library.
- Is DPI using NIH sponsored sources of Natural Products?
 - Yes, DPI is using them. Unfortunately, Natural Products are surrounded by extensive IP issues.
 - NOTE: As of now, there are no NPs in the library. First set of 500 NPs came in 2 weeks ago. Sigma's 1100 compounds will be added soon.

FIRST YEAR OVERVIEW

Strength, Weaknesses, Opportunities, Threats (SWOT) Analysis

- Background:
 - Purpose:
 - SWOT analysis for strategic planning purposes to assess the current status, what has been achieved, and future prospects for growth.
 - Timing:
 - Very important: we want to get our feet wet and be involved in the development of the process. Now that it has been a year, we can do this analysis justice.
 - Secondly, the NIH staff is in the process of issuing a competitive renewal. This is an opportunity for the centers to provide input which will go into the development of the competitive renewal.
 - Additionally, we have a mid-course evaluation of the program.
 - Given today's time restriction:
 - The SC will analyze the network as a whole (not individual centers), and address general weaknesses, strengths, threats from a high altitude perspective. Individual detailed issues can be addressed at the workgroup level.
- Definitions:
 - Strength (internal factor): Factors that provide advantages for sustained success
 - Example: Location, management, experiences, technology, etc.

- Weaknesses (internal factor): Factors that undermine strength and future plans. Factors for a false sense of security
- Opportunities (external factor)
 - Potential prospects that may become strengths
 - Awareness of changes in policy, the field, scientific disciplines, etc.
- Threats (external factor)
 - prediction of future challenges and pitfalls
 - What are our competitors doing?
- NIH and the centers/users have different viewpoints. The NIH has discussed these issues and would like to get feedback from the centers and the SC.
 - KEY:
 - Black = NIH
 - Teal = SC Members

Strengths

- Open access to resources and data
- No constraints on targets/diseases
- Network strength of providing multiple perspectives
- Lack of profit motive provides flexibility
- Centers are part of the larger Roadmap which contains other components that can be leveraged
- In position to bring diverse fields together (biologists, chemistry, informaticists)
- In position to provide molecular probes and other broadly applicable tools to biologists
- Can run the same assay target across multiple different formats with the same compounds library and gather data about those platforms
- Integrate biological data from phenotypic screens to help elucidate and map pathways
- Strength of PubChem database
- Diverse library
- Limited chemistry experience will lead to probe development for intractable compounds that industry would typically avoid
- [Jim Rothman] Simplification: (1) public accessibility (2) network strength of integrating multiple perspectives

Weakness

- Chemistry – Limited medchem experience
- Limited practical experience on hit to lead
- Limited informatics for large data sets
- Perception of the community regarding data release policies
- Threat posed by strong egos to the network operations
- Inability to function as a network—the network does not capitalize off of the individual strengths of each center
- Projects lack long-term benefit and stability in network

- The program is not reaching the scientific community to bring in interesting assays into the centers. Investigators think the activation hurdle to enter the network is too great.
- X01 mechanism does not do the network justice—need some sort of incentive for assay submission. Given the tight funding environment, investigators will not work on something that won't bring money into the lab.
- Data release policy is a deterrent for assay submitters.
- Centers need to focus more on outreach.
- Need to make coherent case for what chemical probes can do for investigators
- Need more publications to spread awareness of the power of small molecules as tools for chemical biology. As of now, don't have that one shining example to show off as incentive to the scientific community
- Need to reach out to the commercial industry as well.
- Not clear whether the investigators or if the centers are responsible for follow-up/secondary assays for hits verification.
- Limited interaction of the assay PIs and the centers. Lacking mechanism to discuss the critical path from a screen into a broadly useful probe/screening data. Reviewers should score applications based on the proposed critical path.
 - o [Lucile White] Assay PI – Center interaction document details process of collaboration
- Definition of a probe and requirement to produce 10 probes a year. Need selective/potent probes. Definition is too vague.
- Summary:
 - o Communication issues.
 - o Resource issues: limited chemistry, experience, informatics
 - o Follow up process issues: unclear what the product is
 - o Incentives for investigators: grant mechanisms/IP issue

Opportunity (things we *could* be doing, not necessarily what we *should* be doing)

- Untapped assay resources of the community
- Create unique, highly productive screening library
- Excellent environment for assay development
- SMR is an appropriate repository for academic compounds
- Produce a great deal of data on compounds outside the rule of 5
- Unrealized promise of large PubChem database
- Great potential across the network for developing collaborations (ex: Burnham/Pitt)
- Use HTS methods to find probes and tools (different than pharmaceutical companies) especially for novel targets and drugs
- Promise of breakthrough in orphan diseases studies
- Opportunity to explore novel chemistry (beyond Carbon based chemistry, beyond carcinogens and mutagens)→ unencumbered by in vivo setting
- Major opportunity of learning from the use of probes
- Other types of screens (non-chemical libraries (ex: siRNA/RNAi))

- [Advisor] have you been approached by any pharmaceutical companies to donate compounds that not want to work on (ex: compounds with large MW, intractable, etc)?
 - o [Carson Loomis] has not been an issue yet. Does speak to our need to communicate with commercial entities.
- [Catherine Peishoff] in response to incentive of pharmaceutical companies to donate compounds; there is a pressure to synthetically modify compounds that will be starting points for therapeutics; compounds that will require too much time/energy could be good probe candidates. Also applies to assays that are ready to HTS, but are not screened due to time/resource pressures.
 - o [Chris Austin] good ideas, but doesn't work well with upper management/lawyers. Scientific and operational issues are relevant to both sides, but it needs a champion.
 - o Is there a possibility of a consortium between pharmaceutical companies to donate compounds that aren't feasible starters?
 - [Catherine Peishoff] most pharma companies have bad experiences with consortia, i.e., loss of control and influence. Caveat: in terms of data release, they may be inclined to participate if they had access to data 90 days earlier than publicly available

Threats

- Profit motive may restrict interesting science
- Competition for assays with institutions lacking a data sharing plan
- Unknown importance of generated probes to the community
- PubChem is not user friendly and lacks data mining tools
- The community may interpret PubChem data as polished secondary data. This may lead to some frustration among the investigator community as well as undermine PubChem's credibility
- Academic screening centers
- No established critical path to identify probes in timely fashion
- To be able to hire and retain personnel
- Cost of the assays/reagents
- Public perception of the Roadmap
- Ability to successfully deliver the purpose of the Roadmap/MLSCN to congress and public
- Threat posed by strong egos to the network operations
- How do we differentiate the network from the academic screening centers vs. in-house screening centers?
 - o [Ralph Garippa] want to prevent "us v. them" scenario. Maybe we could monitor the literature and see if investigators who have already utilized in-house screening centers or academic labs would like to run their assay on an additional 66K compounds within the network
- Failure to define success in achievable terms

CENTER HIGHLIGHT SERIES

NMR Screening – Maurizio Pellecchia

- A Million Little Pieces
 - o Lots of work to find false positives
- why fragment-based drug discovery (FBDD)
 - o higher potency hits and high affinity ligands
 - o FBDD: progressive optimization of ligands

SC WORKING GROUP REPORTS

Chemistry WG – Donna Huryn

- SMR
 - o Strategy for 2nd 100K compound Acquisition
 - Filling in clusters availability for SAR for initial screen
 - Structure/Function Criteria
 - Loosened up Ro5 criteria
 - Include NP-like functionality
 - Appreciate potential to expand definition of probe-like compounds
 - Solicitation of academic compounds
 - o Availability of spectral data/synthetic protocols from vendors
 - o Availability of LC/MS on DPI cherry picked sample
 - o Differences in LC/MS
- compound profiling assays
 - o Strongly endorsed by the chemistry WG
 - o on properties (Fluorescence, DTT, solubility)
 - o allows hit prioritization allows ID of probes better than existing probes
- Chemistry Core Process/Strategies
 - o What (confirmation, LC/MS, selectivity)
 - o What are the parameters of this campaign
 - Time, quantity, number of analogs
 - o Who will screen synthesized analogs
 - o In addition to 20-30 mg
- conclusions:
 - o too early to force standardization on protocol
 - o communication is essential to avoid duplicating resources
 - **Suggestion: restricted website which includes LC/MS data, compounds currently being worked on, analogs purchased, etc.**
- future topics
 - o false positives and artifacts
 - o where to put characterization assay data
 - o input to the PI-center communication document for chemistry resources if/when they will be provided
 - o increase interface between workgroups
 - o stopping criteria

- best practice discussion

HTS/Assay Implementation WG – Lucile White

- Guidelines for HTS Assays – Paul Johnston
 - To address the quality of assays entering the center
- Dictionary of Terms – Dave Weaver
 - In the packet
- Inter-center QC Assay – Network
 - All of the centers have completed the assay and submitted data to Carson
- Minimum data fields – Peter Hodder
 - Issue moved to IWG
- Essential PubChem upload data – Peter Hodder
- Screening chemical synthesis follow-ups – Steve Brown
- Work together to overcome network and individual center problems
 - Hour of each meeting in set aside for the centers to discuss specific questions relating to MLSCN assays and other assays
- Safety in handling compounds if unknown toxicity –Paul Johnston/Tim Lease
- Network website for pricing
- Framework of Assay PI-center collaboration – Doug Auld
 - Added role of SO and also toughened it up
 - **Resolution: Adopted by the SC**
 - **NOTE: definition of probe will evolve over time**
 - **Action item: post document on the website. Let the assay provider know what is expected. Best to have this document presented as the policy of the network. Caveat: document is only suggestions; NIH cannot endorse it without creating expectations among assay providers.**
 - [Advisor] this document suggests a hand-off of the assay to the network.
 - Lucile White: this document tried to build in collaborations. Tried to recommend having assay PI visit the centers to build trust/collaborations.
- Future topics:
 - Develop a common way to determine the cost of a screen – Haiyan Fu
 - What is the best way to store/handle compounds on site – Steve Vasile
 - Discussion about how to do large screens—Lucile White
 - Define Assay Go/no-go stopping rules
- Network Issues
 - A lot of cross cutting topics across working groups. We need advice on how these cross cutting issues should be discussed
 - NIH would like to encourage WGs to communicate between centers in more informal settings

Science Officers WG – Ron Margolis

- Background:
 - SOWG comprised of Scientific Program Officers from 6 different NIH Institutes

- SO is an intermediary between the NIH Project Team and the Assay providers.
- This group has monthly meetings and is privy to center progress reports and other internal information by the network
- Assays are assigned to individual SOs
- Upon receipt of the MTA, in many cases, a teleconferences between the SO, Assay Provider, and Center PI has been held
- Feedback
 - Are the SOs useful?
 - Is the potential role of the NIH SO understood?
 - Is there a need to modify or better define the role of the SOs?
 - Is the requirement to release assay results within 2 weeks of the completion of the assay campaign acting as a disincentive to submit assays?
- SO suggestions:
 - HTS/Assay development WG should consider developing criteria for stopping difficult assays
 - Assays that don't perform as expected
 - Unproductive assay
 - Unreasonable costs
 - Centers could develop common principles and communicate them to the provider up from so that a clear understanding exists at the very start of the assay campaign
 - Relevant SO should follow assay progress via monthly tracking reports
 - To stay informed of assay campaign progress in case questions arise
 - Is more outreach necessary to affect outreach?
 - Resources for follow-up to screening campaigns
 - For immediate follow-up
 - Small grants R03/R21 Pilot and Feasibility studies
 - Very Institute specific
 - NIH-guide and Institute websites
 - Significant resources are available for mostly translational
 - On MLSCN website
 - [Is the data from these NIH supported programs going to PubChem?](#)

MLI PT ACTIVITIES

MLIIG website

- www.Mli.nih.gov
- Educate the public about the organization structure, the governance, center capabilities
- Secure internal website for document repository, assay solicitation/development
- **Action Item: Welcome additional suggestions/comments on content of static pages and also for restricted sections**

- Within the next week, will have calendar of MLSCN events and bulletin boards
- Login
 - o In a few days, every SC member will receive a login
 - o Differential access to different levels (SC/WG/etc)
- Questions:
 - o Will there be a structure searching capability?
 - Answer: not planned for now, but may be a possibility at a future data

Mid-course review

- Mid course review will
 - o Occur before initiatives are approved for the next phase of funding
 - o Inform decisions about the future course of the initiative
 - Transition into the centers
 - Completed
 - o General and initiative-specific criteria
 - o Address information needs of the NIH-decision makers
 - o mid-course review will determine the future scope of the project
- Different components
 - o NCGC, SMR, 9 extramural centers
 - o ECCRs
 - o Tech development initiatives
 - o Imaging probe development center
- Timeline of the MLI
 - o Late October/early-November of this year for feedback to inform the re-competition of the MLSCN
- Questions
 - o Will the initiative be compelling to the stakeholders (public)?
 - o Does this initiative cover all of the ICs?
 - o Can the NIH afford not to do it?
 - o Is the science meritorious?
 - o Does the progress of the initiative meet the primary objectives?
 - o Transforming?
 - o Long range benefits and consequences of the program
- Formal evaluation process
 - o Feasibility study to analyze progress and future of the initiative (September 2005) with group of outside experts
 - o Battelle compiled input from experts and interviews and made a logic model plan for the NIH to use
 - o Outcome measures (short term to long term)
 - o Key questions were proposed by the Roadmap implementation panel
- Progress Outcome Data will be gathered by scientific program managers and PT members
 - o Data compiled and analyzed by third party
 - o Will have third party Advisory panel to review and make recommendations

- Present this plan with recommendations to the RICC by mid-August
- Plan to re-compete the screening centers
- Logic model
 - Focus on short term outcome measures in accordance to where we are right now
- Key questions
 - Innovation
 - What are the unique, innovative, products of the MLI?
 - What are the products of the MLSCN beyond the development of chemical probes i.e.: global effects of the MLSCN
 - Directing of scientific research in the community
 - Education
 - Technology
 - NP
 - Outreach
 - Collaboration
 - Chemists and biologists? Private sector?
 - Dissemination and Use
 - To what extent are the products been used by the research community at large?
 - Value
 - Impact
- Future directions and configuration of the MLSCN
 - What aspects of the network should we reconfigure in the remaining two years of the pilot phase?
 - How should we position the network to maximize transition to the second full funding (production) phase?
- **Action item: SC members provide feedback about these questions and information about their center based on the presentation slides**

RFA-RM-06-003 “Pilot Scale Libraries for HTS”

- premise: novel proteins require novel molecular probes
- NP as well synthesis
- RFA for pilot scale libraries
 - 6 new three year grants (September 2006)
 - 2005 PSL grantees
 - Progress on slide
 - John Porco (BU): 146 compounds
 - K.H. Lee (UNC-Chapel Hill): 250 NP ready for submission
 - Mark Kurth (UC-Davis): 1100 compound structure submitted for review
 - Kevin Burgess (Texas A&M): 135 submitted for review
 - Jeff Wright (UNC-Wilmington): September first submission
 - Stuart Schreiber (MIT): September first submission
- There is a NIH Project Team for chemical diversity

- Very similar as contract mechanism as a way for us to procure compounds from the scientific community
- **Action Item: talk to your friends about submitting compounds. Contact John Schwab for questions/comments.**
- Discussion:
 - o In process of creating RFA for un-reimbursed compound submission to get data on their compounds
 - o These compounds are a great selling point for the MLSCN. But the average person out there will not have access to these compounds.
 - o One of the advantages the MLSCN has is that they are not dictated by profits. To what extent do the inorganic chemists know about this initiative and the possibilities? How would the study section approach inorganic chemist applications?
 - o **Suggestion: include a notice to the RFA about the acceptance of inorganic compounds**

Solicitation of Assays for HTS in the MLSCN

- MLSCN assay pipeline progress and status
- 17 assays have been accepted in cycle 3
- Cycle 4 applications will be reviewed on Thursday
- Search under depositor category on PubChem,: Molecular Libraries Screening Center Network to see network assay data
- 45 total assays
- 4 centers have tested the full set of compounds
- Very impressive plot of data deposition vs. total number of compounds tested
- Outreach efforts have been successful
- MLSCN Publications
 - o S1P1
 - Phosphate ester into phosphonate
 - ligand mimetic
 - validated as a physiological node
 - o GPR30
 - Extracellular estrogen receptor
- NIH news
 - o Selective S1P1 antagonist
- Research supplements to promote diversity in health-related research
 - o \$75K if underrepresented minority student/young faculty member is hired
- Discussion:
 - o [Catherine Peishoff] is there any provision for funds for the computation chemistry people to do predictive/virtual screens using the library?
Interesting to use these data sets vs. experimental screens

DISCUSSION OF ASSAY RECRUITMENT AND ASSIGNMENT

Discussion Assay Recruitment

- Number of assay submissions have stabilized, it is still too early to gauge the trend yet
- Current application mechanisms have not engaged some top labs around the country
- How many assays should be coming in?
 - o 10 assays the first year, 15 2nd year, 20 3rd year
 - o Depends on how one counts the assays
- Quality of assays/follow through
- What are the barriers?
 - o Investigators are concerned with IP/data release issues and funding
 - How to reconcile this?
 - Provide funding
 - \$3K for travel costs/reagent/scale up costs
 - o data release
 - Value of data is in the context of other data points. Value of the data is in the context of the secondary assays.
 - Primary screening data is not very useful.
 - Early data is not usually worth patenting. Competitiveness is downstream of primary screen. Problem is there is a gap between experienced screeners and academic labs.
 - This is more of a perceived barrier than an actual issue.
 - o Issue of consumer education; scientific community doesn't understand the potential of chemical probes to complement genetic approaches to biology
 - Present data to the community as evidence of the utility of small molecules
 - Outreach!!!
 - o R03/R21 targets are not coming into the Roadmap, even though that was the initial expectation
 - It seems that this mechanism needs to be tied to the Roadmap.
 - It would be nice to see R21 investigators to form relationships with the centers to allow more creative assays
 - R21s will not need to reapply to the X01. Fast track into the screening centers via alternative administrative review to facilitate a rapid transition where appropriate. QC will be a checklist that the center will review. If it is not ready, it will go back for regular X01.
 - Competition with non-MLSCN screening centers
 - o **Action Item: We have talked about a general paper for a widely circulated paper for *Science* or *Nature* and write about experience after a year. We could include in the paper of the perception and the reality supported by screening data.**

Discussion: Assay Assignment

- No major issue with assay assignments
 - o [Paul Johnston] cost issue of screens
 - o [CA] Should there a formal place on the X01 where assay applicants can designate they have worked with a specific center?

PRESENTATION BY MLSCN ADVISOR

The Evolution of High throughput Screening (HTS) Environment at Roche – Ralph Garippa

Overview of Roche's experience with HTS

- Independent evaluation gave us good marks across industry
- Dr. Garippa works at Roche site in Nutley, NJ
 - o Dr. Garippa's background is in *in vivo* pharmacology
 - o rats with ocular conjunctivitis
- Aim is to develop new drugs. Existing drugs include: Boniva (osteoporosis), Tamiflu, Pegasys (Hep C), Xeloda (metastatic breast cancer), Xenical (lipoprotease inhibitor), Rituxan (rheumatoid arthritis)
- Own shares in Genentech and Chugai
- Current pipeline over 100 projects, mostly HTS projects, 59 new molecular entities (NMEs), take advantage of genomic, bio markers
 - o Partnering with new technologies → Affymetrix, etc.

Basics of HTS

- Low throughput
 - o Bench, hand pipettes
- HTS
- uHTS (Ultra-HTS)
 - o kalypsys

Updates

- HTS
 - o Identify a few chemical entities that interact specifically with validated target by screening a large library of chemicals
 - o .5 to 2M compound library
 - o High hit rates. Roche uses virtual screening to limit # of hits
- Random screening rather than targeted screening
 - o To avoid missing novel scaffolds
 - o Large # of compounds = *huge* discount
- Meeting the challenge
 - o New Paradigm
 - HTS as member of Discovery Project Teams
- Plate evolution
 - o 96 → 384 → 1536
 - o Cost savings
- Zeiss uHTS screening platform
 - o Core of the system = reader

- Many liquid handlers
- Small footprint compared to other systems
- Conveyor belt system
 - Plate lifts
 - Liquid handlers
 - Washers
 - Server
 - Optimized readers
- Readers
 - Absorption, Fluorescence Intensity, TRF, HTRF, FP, etc.
 - HCS, translocations?
 - heart of the system:
 - lens array → very powerful tool
- System drastically reduces screening time
 - Can bring in many more assays
 - 2 now, kinetic assays/10 pt dose response
 - Assay development
- HTS formats
 - No longer radiometric
 - Fluorescence
 - TRF
 - More expensive reagents (European)
 - HTRF
 - iMAP (FP)
 - Absorption
 - Good for oncology and metabolic diseases
 - Luminescence
- Cell based
 - GPCRs: Gs, Gi, with cA readout
 - Ion Channels
 - Radiometric binding assays

Manage global inventories

- local
- Goddess plates in Basel, Mother plates at each center
- Cherry picking in Basel
- Avoid freeze-thaw cycles
 - Collaborating with Renk
 - Sealed tubes with 384 well capacities, can access randomly 6K tubes a day. 2 or 8 uM sample of DMSO samples. Can ask for larger amounts.
- plates sealed in foil under inert gas
 - receive compounds after 1-2 weeks
- every three years they reformat the library
 - in constant flux
 - actively curated by the chemistry group

Discussion:

- how many new assays/targets come on board each year per site
 - o capacity to run 18-25 full screens per year per site but steady state is actually 10 -12 per year due to active chemistry
 - o When the library turns over, Roche re-screens.
- How important is z' ?
 - o Z' is important. Roche will never run a screen with a $z' < .5$. Roche doesn't set a hard limit for z'.
 - o We also use Bayesian statistics to run screens. We use the shape of the curve as a cutoff for false +/-
- If Roche is running GPCR, they will run related and unrelated receptor to get EC50s and related selectivity
 - o Look at related classes
- Roche doesn't want to use cell health cytotoxicity on every compound, only for the hits out of the screens

PRESENTATION ON THE GENETIC ALLIANCE

The Importance of Research on Underserved Diseases – the Genetic Alliance

- Genetic Alliances
 - o Coalition of over 600 disease advocacy groups
 - o Most of the diseases are rare genetic diseases
 - o Some of the 600 Advocacy organizations are American Cancer Society, etc. Still some 7K rare diseases with 1200 advocacy groups.
 - o Largest provider of genetic services
- Patient Advocacy
 - o 50 – 60s Alcoholics Anonymous, Cystic Fibrosis Foundation
 - o 70s: what services are missing? What do we need to do?
 - o 80s: Robust changes with increased communication
 - o 90s: coalitions revolving around research
 - o 2003: Genetic Alliances to effect public policy and Orphan Drug Act
 - o 2000: Research advocacy is a huge focus
 - o 2010: Best practice around translational research and delivery
- Consumer/Advocate Perspective
 - o Look at picture as a whole.
 - o Want Bench to bedside to *practice*
- Genetic Alliance BioBank
 - o Infrastructure for advocacy groups to put together blood and tissue samples
 - o Allows collection of epidemiological data, cell lines, medical records, and self reported data
 - o Partner with academic, government, and industry to change the field
 - o Also look into strong patient rights with informed consent and data exchange

- Need control with hierarchical structure
 - o Advocates in the information age with open source and open data archiving
- Focused on prolonged survival, quality of life issues, reduction of cost
 - o we think if we work with scientists, we can work to that place at the end
 - o we really need someone to make an assay for our disease

Pat Terry

- PXE
- GI disease, lost of central vision, cardiovascular disease
- Put together a systems diagram
 - o Launched clinical trials to deal with major morbidity of vision loss
- Hit the wall for a cell based functional assay
- Overcome hurdles such as the informed consent, IRB approval, etc.
- Internet enabled the organization to collect patients, biological materials, etc.
- Eventually resulted in a full genome screen isolating disease to chromosome 16
- Some industry collaborations around the world
- \$15M with academics
- MRP6 Protein Multi-drug resistant protein
 - o Unknown substrate transposed to the extra cellular matrix
 - o Loss of function gene
 - o Patented the IP of the disease gene
 - o Missense mutation analysis
 - Lack of localization, hydrolysis → no genotype/phenotype correlation
 - Effects collagen
 - Disturbance of ATP binding
 - Mimics age related macular degeneration, cardiovascular disease, skin wrinkling, etc.
 - Pathogenic exaggeration
- 48 proteins with ABC binding cassettes
 - o Opportunities to rescue ABCc6
 - o 18 diseases associated with this family of proteins
 - o Known substrates and drugs that are transported by these proteins
- Genetic Alliance wants to infuse again the sense of tempered urgency
 - o this community is stuck trying to find a functional assay
 - o in 11 years they have gone to animal models
- comparative genomics
- control functional elements
- Issue of comparative datasets. Can characterize world wide population off of single operating system
- Want to standardize models, develop best practices, and interact with different communities with different motivations and drives. Part of what they've done is industrialize this process so that others can follow the model and make similar headway.
- International Genetic Alliance:

- Transferring knowledge and technical know-how to the rest of the world
- Genetic Alliance BioBank
- Discussion:
 - Where is this protein normally expressed?
 - Liver, testes, kidneys most expression
 - Mineralization → local deficiency to specific tissue or systemic throughout the body?
 - Debate: some believe it's a metabolic, systemic disease. Elastin.
 - Inhibitors make it worse?
 - Yes. Artificially up-regulate related protein to compensate for the loss of function.
 - Will PXE contribute international funded construction of cell lines to MLSCN?
 - Other challenges
 - [Sklar] Interested in developing multiplexed assays for many members of this family
 - SRI is currently running an MRP1 inhibitor assay

Day 2

INTRODUCTION.MLSCN PROGRESS REPORTS

Discussion of Tracking Reports

- First Year Progress Reports for the MLSCN
- Monthly Tracking report form each center that shows progress of each assigned assay
- Achievements:
 - Facilities completed by June 06
 - One center moving chemistry core
 - Most centers planning for new personnel
 - 7/10 centers requesting yr 2 equipment purchases
 - 2/10 centers looking to outsource chemistry
 - By the end of yr 2, still below 100K
- Tracking Documents
 - Header: different steps chronicling the different phases of compound probe development
 - Phase 1: MTA between NIH and assay PI. Once signed, letters go out to the centers notifying them of assay assignment and corresponding Science Officer
 - Phase 2: assay optimization
 - Most assays need to be tweaked for HTS platform
 - Phase 3: Validation

- Controls, known active compounds, z' (measure of reproducibility of the assay)
 - Once assays meet this criteria, centers notify the NIH assay has been validated
 - Phase 4: screening
 - Requires hit confirmation prior to PubChem Data deposition
 - Run dose response
 - Run secondary assay before committing chemistry. Usually not HT
 - 67 assigned assays, 32 validates, 20 screened, 16 submitted to PubChem
 - Great, considering start point
 - within the last month, additional submissions to PubChem that are not represented in the above overview
- Monthly Tracking Documents
 - Started with R03, now use X01 Resource grant mechanism for assay PIs. Review. X01's now have electronic grant submission
 - Some of the assays have multiple targets. Each target needs to be screened across the entire library. The centers get credit for running two equivalent assays
 - Ex: S1P3 and antagonist screen are both run across the whole library. Commitment of resources, etc justify two assays
- Internal standard cytotoxicity assay
 - SMR provides plate to the centers with different cytotoxic agents. Ask the centers to screen that plate and identify cytotoxic compounds.
 - 7/10 centers have run the plates and used 7 different detection methods to ID cytotoxic compounds.
 - 3 plates: 2 with plate maps, one blinded
 - Compounds chosen so there are molecular targets
 - Hopefully will publish on this data

Discussion:

- Cytotoxic data has not been analyzed yet
- Data should be submitted to PubChem. Although not possible for blinded plate.
- Is there a policy for selecting this assay? Are some signaling pathways vs. receptors?
 - Nope. No restriction on these assays. Looking at the same target through different assays.
- Are there a minimum number of false negatives?
 - Not as concerned with false (-).
- Time from phase 1-7 can vary. Target time?
 - Depends on cost and when the assay will work.
 - Seen cases where well scored assays were difficult to implement and poorly scored assays were easy to run.
 - Resource issues. Reagents difficult/expensive to scale.
- What kind of protocol?

- Open to multiple targets.

MLSCN CENTER REPORTS

Columbia:

- Infrastructure development. Remodeled 6K sq feet of space. Columbia loaned them money for equipment with the expectation forthcoming NIH money will repay the institution
- End of July, max capacity 550K points per week
- 3 units with designed biosafety compatibility. Can use virally infected cells for readouts. Good selling point for assay solicitation.
- HCS and cell imaging (InCell 3000 analyzer)
- Can now run more conventional screens beyond phenotypic assays. New plate reader technology.
- Investment wise, HT cell culture lab important feature for secondary screening and producing a large number of cells.
- Organization
 - o Lars Branden, Thomas Meyer (Assay development)
 - o Taken advantage to build their organization. Met with Nick Tsinoremas and Stephan Schurer (Scripps, FL) to contract out cheminformatics.
- Screening Status
 - o Assigned 3 assays in cycle 1
 - NF-kB completed and data deposition to PubChem. Working on 2ary assays now
 - Huntington's assay: run 1ary screen. Confirmed hit rate of .5%
 - E-selectin: complete on the 1ary level
 - o Second cycle
 - o Third Cycle
 - o All of the first cycles screened with 10K, latter two cycles will screen with full set
- Assays
 - o NF-kB translocation induced by TNF-a. Monitor movement of transcription factor via fluorescence. Hits: Brefeldin A. Other hits that will go into 2ary pathway screening.
 - o Huntington's Disease assay: poly-glutamate protein aggregates. Controversy if disease is by extensive aggregates or microscopic aggregates. Using HCS to visualize the reduction of GFP tagged aggregates. Two hits of reasonable potency; one hit has no other known activity, other hit has broad range activity. Reassuringly, selected Tetracycline for confidence in the screen.
 - o Cytotoxicity assay: Alamar Blue. Toxic compounds will flow to 4 2ary toxicity assays (DNA condensation, DNA Damage, division, apoptosis via Caspase 3 activation → GFP clusters till caspase activated, clipped, GFP spreads)
- Ahead of schedule. Capacity to run non-phenotypic screens → welcome more assay assignments

- Recognizing the vast majority of compounds will not turn into biological probes, cannot gauge success by # of probes. Probes are not the only value of this project. Each hit is an opportunity to populate PubChem and provide resources for chemists/biologists. We need to establish the chemical *and* biological selectivity of these compounds.
 - o Columbia's use of reporter assays (biolum for transcription) and immunostaining and immuno localization. For simplicity, they have focused on HuVEC cells → rich in inter-something signaling also flat, so good for HTS.
 - o In the past 9 months, they have developed 123 biological pathway assays with 63 different readouts (ex: NF-kB translocation). Broad coverage to provide critical info from lary screening efforts. Multiple upstream pathways with multiple ligand sets using a single read-out; so inhibitors that effect downstream will inhibit all three readouts; different placement of inhibitor in the pathway can inhibit one set of ligands but not another set.
 - o BFA hit in NF-kB assay. Whole series of assays can lead to the conclusion that effects structure of Golgi, NF-kB receptor turnover, etc. Find that BFA is effective probe for Golgi. Also that it is pleiotropic.
- Multiple uses for biological pathway assays
 - o Prioritize hits for secondary screening due to limited chemistry resources.
 - Select narrower/more interesting biological effects
 - o Secondary screening for other centers. Establish activity/mechanism of action from protein/enzyme
 - o Intensive biological annotation of selected chemical libraries.
 - o Provide guidance for selecting analogs during probe development
- Proof-of-Principle Pilot program for biological pathway profiling using year 2 direct costs.
 - o Take up to 1K hits from the network → willing to work with other interested centers → b/w now and December for 50 readouts. Columbia will shoulder costs and will make data available on PubChem.

Questions:

- What kind of cheminformatics support will you need?
 - o Columbia has a lower level of chemistry in the budget than they would like to have because costs of cellular screening are more expensive than that of fluorescent screening. Columbia also planned to screen on 10K, now management has asked them to screen 67K (eating away at budget). Hoping combination of outsourcing/supplements will cover.
 - o [Nick Tsinoremas] before you commit chemistry, you have to know what you are going to do. Understanding is to put data from profiling and hits together to decide which hits to actually follow. Prioritize which hits to apply synthetic chemistry to. We need to see what kinds of probes we would like to ultimately see. This is why the chemistry effort will be delayed to the end of the third year.
 - o Cheminformatics at Scripps will also be used for Columbia

- [Brian Roth] Selectivity screening is one of the most essential things and its great that Columbia is moving forward to map these biological pathways. From his experience, most novel findings have come from selectivity screens. Other thing, from his experience, the major things have come from going outside/beyond the original goals of the initiative.
- Tangent
 - o Small number of compounds in the SMR compared to Pharma. Plan?
 - High purity, interesting compounds. Aiming for 300K in the library. Many of the initial compounds failed the QC standards. Some of the urgency in building the library is limited by what the centers are willing to screen. It makes sense to take time and select the best/most interesting possible compounds.
 - Compounds in DMSO, sealed in inert gas -20C and then sent out. Different centers have different storage conditions. Each center has special requirements for their plates.

Burnham – Nick Cosford

- Dose response curves in run parallel
- For biochemical assays, Bio
- HCS: Hamilton Star liquid handlers
- Imaging: InCell 1000
- NMR screening
- Clean room which houses HCS. BSL-2, protein disposal.
- Extensive hardware to support massive amounts of data
 - o CBIS:
 - Web-based
 - Compound registration
 - SAR
 - Plate mapping
 - Provides platform for data to PubChem
 - o SpotFire
 - QC
 - o Pipeline Pilot
 - Data mining and analysis
- Diversity of assigned assays
 - o Assay optimization: biochemical assays have different parameters. Typically to reduce amount of protein needed, increase sensitivity, reduce costs. Ex: MKP-3 assay, reduced protein needs by 40%. Cost/well in line with Roche.
 - o Bfl-1 FP assay: inc assay window and stability
- screening
 - o Yr to date completed 6 screens (including NMR). Two in progress (HaPTP and TNAP screened in parallel)
 - o K-Ras assay didn't result in any hits in the first round of screening
- Hit confirmation

- 1st round, screened 10K compounds, 8 identified hits (some time dependant appears to be a redox process)
- Bfl-1 assay 1ary screen FP assay, 22 confirmed hits. Developed TR-FRET assays
- Hit Validation and SAR optimization
 - Search database of analogs and clusters
 - CBIS: database of commercially available analogs
 - Beginning to ramp up chemistry; synthesized some analogs.
- PubChem Data mining
 - Example from MKP-3 for hits in other assays. Turns out that all hits were active in the MKP-1 assay run at Pitt. Fostered a good collaboration. Hits also active in other assays
 - Five 1ary screening datasets in PubChem (145K data points). Three more assays primed for submission by next week.
- Completed objectives for the year
 - So far, optimized 7 assays, screened 5, NMR-ed 1, 11 X01s submitted, 5 PubChem data depositions
 - Website launched in March of this year

Questions:

- Are the centers only listing confirmed hits as active? Or are they including all hits from the primary screen (may include false-positives)?
 - Burnham: put in all actives from the primary screen. Re-order all hits from 1ary screen up to the number allowed from DPI.
 - 100 point scoring systems. 1ary assay hits 0-20, not hit 20-40. no confirmation 40-50, Confirmation 50-60 points. Then score moves up for IC/EC 50 proportionate to rank order and data.
 - Some cases (ex: imaging based screens) where numerical data good and scoring system low.
 - **Action Item: discuss a uniform scoring system for the informatics users**
 - NCGC's dose response curves with actives and IC50/EC50 on the same line is very useful.

Burnham Continued...

High Content Screening (cell death/cell differentiation = biological focus)

- VCAM Expression on cell surfaces in response to TNF
- Advantageous image based assay
- Algorithm developed by Jeff Price
- Flow diagram to filter out crud (fluorescent compounds)
 - See upregulating/down-regulating compounds
- Reader comparison: InCell 1000 Imager vs. DTX Plate
 - Imager discerned extra number of hits
- Large part of yr 1 activity for algorithm development
- Major interest = HCS to pick up rare events
 - Engineered ESC lines with fluorescence/selection markers to look for compounds that stimulate differentiation
 - See profile of all of the compounds assayed
 - Remove all compounds that don't hit

- Cardiomyocyte Proliferation assay
 - o 1/60K replicate
 - o Green Immunostaining in nucleus indicate compounds that cause stimulation of Cardiomyocytes
- Another interest is b-cell differentiation in pancreatic endocrine progenitor cells
 - o Drive cells down b-cell lineage
- HCS used for sub-cellular localization of assays
 - o Particularly useful for proteins at/near plasma membrane of cells. See protein localization at the surface (b-catenin in adherence junctions of cells and transcription factors)
 - Effects of GSK3b inhibition → can dial up/down presence in nucleus
 - o Quantify data
- PKC-a (signaling pathway localized under plasma membrane in response to phorbol esters)
- HCS in cell physiology
 - o Dynamic Ca imaging → tech development grant
 - o Jeff developing a new instrument that allows people to scan tissue samples while maintaining focus via parallel autofocus
 - o Magnification corrected light to track specimen as it scans
 - o Area sensor and speed of scan (1.84 mm/sec) works at any dry magnification
 - o Goal to get to 20K images/day; 3 color prototype running with lower magnification

Questions:

- images that are not going to the assay endpoint (intracellular localization, precipitate, etc)
 - o Internal knowledge of compounds will be linked to PubChem submission. Eventually can click on hit and see thumbnail of the image.
 - o Trying to work on high resolution data. Huge informatics issue.
 - o Deconvoluting signaling pathways for hits will be difficult.

Scripps

- Assay development labs in FL and CA
- Went from empty room to fully operational 1536 robotics system within the last year
- Compound management lab can store .75M samples with automation and cherry picking ability → automated access to all compound in SMR
- Compound QA/QC platform fully operational and integrated with the HTS platform
 - o Can hit pick, reconfirm mass and purity, perform dilution series
 - o FI readouts
 - o Can support broad assay formats
- cytotoxicity screen for the MLSCN: run 1ary screen, hit pick, cluster different families of cytotoxic compounds and IC50s within literature rangers

- Reporter assay for S1P3 GPCR looking for agonists, finding 13 agonists
- Ran NF- κ B assay from Burnham ready in 384 plates, ready for miniaturization
- S1P1 suppression allosteric potentiator assay EC20 with known selective agonist for S1P1. Full deck screened, 515 compounds emerged (agonists and allosteric pot). Can automatically hit pick 1280 compounds. Generated 10 point response curves. Data uploaded to PubChem
 - o Activity of Steve Brown and Peter Hodder
- Also interested in chemical probes with in vitro biological activity with optimal in vivo activity
 - o Screening pharmacokinetics. Allows definition of scaffolds from hits that are worthy of chemical optimization. Prior to chemical optimization, look at physical properties (solubility, tissue penetration). Completed 59 PK studies—IV dosing in SD rat
- Informatics
 - o Challenge to integrate HTS with PubChem
 - o Needed integration and operation systems to account facilities I La Jolla, Jupiter, and palm Beach to publish data in PubChem and integrate analytic tools to pursue subset of leads in a practical way
 - o Use Spotfire, Leadscope (leader for the network)
 - o First center to identify structural discrepancies between PubChem and SD files from DPI
 - o Support automation and downstream analysis of the data
 - BCUT space
 - Lower filter to sub μ M (less than 100 nM) between receptor subtypes
 - In addition, could predict hits coming out of the NIH collection for Rosen's collection
- Chemistry
 - o Library and linear approaches
 - o Only commit chemistry after reconfirmation of hits and detailed counter-screens. Search Zink database and look for SAR. Lots of useful info without committing to synthesis.
 - o Chemical libraries and computation chemistry
- S1P1 antagonism
 - o Screening is the slowest path to solving chemistry problems (selectivity, stability, etc)
 - o Phosphonate analogs
 - o Moved from agonist to antagonist platform
 - o Resolved racemic mixture, R enantiomer more potent. Selective and worked in vivo.
- combined chemical probe development and chemical probe imaging
 - o agonists: lymphocytes stationary
 - o antagonist: mobilizes lymphocytes and pass the endothelial barrier into the sinus
- Summary: based on literature leads and chemical approaches

- Convert screening leads to molecular probes in cell based/animal based systems
- Using informatics to mine leads and cluster in a meaningful way prior to optimization
- SAR via purchase
- Addressed potency, selectivity, etc. issues of compounds
- Chemical approach is acute and reversible with agonist/antagonist

Questions:

- Initial goal was to get a sense of the diversity space covered by the SMR.
- No universal representation of any biological activities, so you need to base biological activities based on chemistry
- Phosphate/phosphonate/alternative head groups on the agonist scaffold put dipole interaction into a salt bridge, so it remains an agonist with altered potency. Can move in and out of the pocket b/c the receptor is volume triggered. Chain length of 15
- Define set of privileged scaffold across this family of receptors. Developing a sub-library to shed light on the family.
- Question of BCUT graph: true did not optimize parameters, blue dots disappeared when [] lowered, red dots (Zink database) remained. Comment on NIH vs. Library
 - Lab development library enriched in actives. NIH picks up higher [] hits that serve as useful leads to the enriched narrower library.

Assay Development for HT—Mark Scheideler

- Aims:
 - Extract assays from ongoing research programs. Provide the resourcing needed to configure them for use in Probe Development projects.
 - Encourage the development of novel assays and assay technologies.
- Emphasis on the Application of Pharmacological Tools in Hypothesis-Driven Research
 - State a biological question that can be addressed through the use small molecule probe.
 - Define the essential attributes of the probe that would make it useful.
 - Propose a plan of assays that can be used to extract compounds with probe attributes from a small molecule compound library.
- May 2006 Council Results
 - Competitive Review, Strong Response Rate
 - 35 awards for \$6 million.
 - 30 R21's and 5 R03's Funded
 - 20 Research Institutes, 3 Biotech Companies, 15 New PI's
- New Announcement: RM07-001
 - Move to Grants.gov E-application process.
 - September 22 Receipt Date, November Review, February Council.
 - Single "Scalable" R21 Mechanism.
 - \$8 million Budgeted for 2 Announcements in FY07.
 - Fast Track MLSCN Entry.

- Opportunity Extended to Prior Awardees.
- Option for IC's to Supplement PI's for Work on
- Completed Projects which enter the MLSCN.
- <http://grants.nih.gov/grants/rfa-files/RFA-RM-07-001.html>

Questions:

- How do facilitate the communication between the centers and the assay PIs?
 - Encourage the WG to sit down and think of a way to approach the assay PIs
- What mechanism is in place to do this?
 - Not one in place yet, trying to institute it next round. 10% of funding, should have access to 10% of what is out there
 - IM can apply to the X01 (separate on ramp)
- Discussion of how this is playing out. How often do investigators follow-up after letter of recommendation from the center and actually enter the network. What is the actual track record for awards back into the center?
 - 2004 tough to gage
 - 2005, monitored closely. Mark Scheideler in contact with assay providers and they are intending to submit in fall. A lot of questions on how to foster collaboration. A few centers looking to provide funds for this.
- Do we have an idea of the success rate of these apps? Attrition rate?
 - Only data that is finished is from the 04 round before the MLSCN was set up. There are some good publications out there.
 - 2005 funded, looking to monitor that detail.
 - Very competitive funding mechanism. R21 apps also require preliminary data. Idea is to take assays that already exists and mold them into a HT format
- There is no contractual obligation for the R21 applicants to submit to the MLSCN?
 - No. Trying to establish an expectation, but these are grants, *not* contracts.
 - Because they are grants, they need to file final reports. Trying for the power of persuasion type thing.
- Any consideration of making this a 2 year program, second year being contingent on submission to the MLSCN
 - [MS] have issues no cost extension. Making it a cost extension changes the scope of the program.

MLII Cheminformatics Initiative

Goals, Opportunities, Challenges

- Secondary goal of the meeting is to provide a forum for network Q&A for technical issues
- Every WG, tried to create cost sharing plans
- Recent activities
 - Data analysis and visualization software
 - MLSMR updates

- WG Cross-participation
- Minimum data to PubChem
- Summary of Break-out session for PubChem
- Presentation by Pitt
- Data Analysis and visualization software
 - Genedata Screener (Ajit, NCGC)
 - Spotfire (Tudor better pricing)
 - ActivityBase XE (beta-module)
 - Pitt center (custom apps and data analysis)
- MLSMR Updates
 - DPI is correcting structures uploaded to PubChem within a one to two week timeline
 - Modifications to SD file formats
- Cross WG participation
 - Chemists oversee functionality filters, other compound selection parameters
 - IWG members should provide input
 - HTS joint as well
- Minimum Data to PubChem
 - Background
 - Assay/HTS conducted survey with no consensus
 - Normalized data sufficient for data mining
 - Upload every piece of data to allow for assay reconstruction
 - SC debated issue in '06
 - Requested IWG input
 - Import everything informatics issues
 - Practicality, data storage and management, useful
 - Not a problem for PubChem to handle, but makes it less user friendly (pain to upload/overwhelm users)
 - Is it useful? We don't know who/how this data will be used.
 - Group decided that raw data is not necessary. If researchers are really interested, they would most likely re-create the assay
 - Recommendation
 - Centers that are running the assay will make the call on what needs to be uploaded.
 - Encourage centers to upload all biologically relevant results
 - Keep more relevant data on the left
 - Evaluate network data mining efforts
 - Gaga/solicit PubChem user needs
 - Yin: more data is better; network's legacy lives in the PubChem archives. Let's not try to predict the user
 - Yang: is it necessary to keep millions of data summaries online?
- Summary:
 - PubChem built for single summary result per sample
 - Use dictate needs

- Currently posted data is sufficient so far
- Some centers will upload raw data regardless of existential philosophy

Questions:

- ECCR meeting: compound structure determination? PubChem assays show up and disappear and definition of actives changes, how should we deal with this?
 - Assays as submitted to PubChem are available for editing later. So if they have additional scientific information, they should also update PubChem
- Need more data and experience to go over annotation of PubChem. Wrong decision now is more dangerous.
 - Some meeting within the next year
- Who is our audience?
 - We need to find out.
 - What do the ECCRs have to say about this?
- Why can't we have both raw and refined data?
 - There is no storage limitation on PubChem.
 - There is a minimum data requirement to include experimental data behind activity calls—so data present to re-run screen.
- Option to create FTP server for database dumps including all data so PubChem can still maintain the clean simplified results
- Breakout session
 - Workflow process outline
 - Goal: ID and rectify problems
 - Functional and computational areas
 - Volunteer to champion this cause to the SC
 - False positive ID in PubChem = “artifact”
 - Add artifact field
 - Activity field would include new state
 - May affect searching in unintended ways
 - future break-out sessions: allow for more time
- Questions:
 - Because the SMR is growing, it might make sense to assign one of the centers to do all of the auto-fluorescence assays

ECCR PRESENTATIONS

- Role outside the network to develop tools and methods for the users of the data that the MLSCN generates and publishes via PubChem
- Debate on what the NIH should emphasize for the P50 and their role in the MLSCN
- Discussion of the role of cheminformatics
- NOTE: folders contain brief description of each ECCR and tools available to the MLSCN

RECCR (RPI)

- Would like to interface with the MLSCN
- Informatics
 - o Data = bottom of the pyramid
 - o Depending on how much data you have 1) data mining 2) models that are effective
- SAR
 - o Need a good way to represent molecules to show what kind of biologic activity they may exhibit
 - o Easier to *a priori* figure out the relationship rather than relying on natural examples
- Molecular Similarities
 - o Use fingerprints to isolate properties of chemicals
 - o Does similar structure relate to function?
 - o Similar in what way?
 - o Use statistics to create predictive/validated models
- emphasis on representing models effectively, use the right model for chemical space, build models rationally
- Tools for mining and modeling, modeling validation, etc.
 - o Tools to help MLSCN people do their work and mold screening efforts is a specific way
- put effort together in different ways
 - o input from different communities
- Goals:
 - o Work with people who mine data and develop tools that are easy to use and easy to apply
 - o Also work in a training capacity
 - o Look at protein-DNA, protein-protein, SM-ligand interactions
 - o New descriptor based on simulations
 - o Validate traditional and novel approaches
 - o Integrating software tools into Pipeline Pilot environment and other workflow environments, so it is easy to integrate into existing systems
 - o Models and methods will be disseminated over the web
 - o Nanotech cluster

RECCR Emphasis I: Descriptors

- look at fields, encoded surfaces
- Encoded surface → look at molecule electronically
 - o Potentials
 - o Not restricted to sub-structures
 - o Look at surfaces: Add descriptors and look at property distribution
 - o Can also encode distributions differently (wavelets, histograms, etc)
- What if molecule shape is the most important thing (not the case for transport properties)?
 - o Shape signatures using electronic properties → PEST (Molecular shape/property Hybrid Encoding) Take 1D distribution and do a 2D

- explosion based upon ray length bouncing in the molecule. Creates a distinctive shape/prediction for the molecule.
 - Electronic PDP very important for docking
- PROLICSS: protein ligand interface surfaces
 - Look for complimentary views of ligand from proteins perspective and vice versa (H-bonding, etc) to make descriptors and predictive models
- Can ID pharmacophores, guide selection of probes (similar and relevant probes)
- Methodology of electronics of SM interactions can also be applied to DNA
 - Technology is used to encode what DNA BP look like in the major and minor groove
 - “Drixel” DNA Descriptors elucidate how proteins bind to major groove and how SM bind to minor groove of DNA
 - Electronic properties of adjacent BP of DNA are highly effected by their environment
- Model Validation
 - Start with data set of things you know and try to generate numerology
- Model Applicability
 - Help understand where we are in chemical space and if a model is applicable there
- Outreach
 - Online modeling tools
 - Downloadable tools
 - Pre-compete descriptor sets for a large number of PubChem datasets
 - Try to get the public to get involved in modeling competitions
 - New journal “Cheminformatics”
- Current RECCR software
 - Drop in to use RECCR descriptors in existing MOE modeling system
- Capabilities summery
 - New molecular descriptors to define molecular
 - Diversity
 - Bioinformatics
 - Within our group local data generation and screening capabilities to test ideals
 - Collaboration with the Cleveland Clinic and Seeker
- **RECCR.chem.RPI.edu**

Broad

- operates at the Broad institute at Harvard and MIT at Cambridge
- adjacent to the MLI CMLD center
- NCI funded database of SM structures and assays: CHEMBANK
 - Similar to PubChem, but not a deposition database per se; very well curated
 - Descriptors
 - Assays performed in Broad’s SC

- Spent time of encoded experimental design based on control meta-data and extending public ontologies that can be used to reference available literature
- Study few years ago using screening data from 40 assays (phenotypic assays and 18 modulators for pathway dissection analysis)
- Extendable scoring system for rendering results of HTS assays comparable
 - Binary outcomes
 - Ranked outcome
 - PubChem has this standard outcome
 - Percent inhibition
 - Characteristic of enzyme assay with positive controls
 - If positive control is not 100% inhibited, assays no longer comparable
 - Mapping to [] domain (ex: EC50)
 - NCGC uses this
 - Does not need data normalization
- Criteria for development of standardized HTS Analysis
 - Amendable to initial measurement at a single dose
 - Does not require foreknowledge of dynamic range
 - Does not depend on biological interpretation per se
 - Does not require a known modulator
 - Parallel screens for Metric space
- developed standardized scoring system based in relationships of SM to mock-treatment distribution

Questions:

- Is there an overlap between ChemBank and PubChem?
 - Through links. ChemBank compounds were contributed to PubChem. So substance records link back to ChemBank. Plan to link assay results in a similar way.
- Particular scoring scheme. Should rescoring methodology added to the PubChem results?
 - Need agreement from centers, IWG, and ECCRs
- legacy data collected before screening center policy was adjusted
 - Analyzed that data to see whether it holds up. Encourage duplicate screening, but reagent constraints result in twice the amount of compound

Indiana University

- Intro:
 - Chemical informatics, a cyber-infrastructure collaboration
 - Collaborations between school of informatics, departments of biology and chemistry at IU Bloomington and IUPUI
 - NSF has an office of cyber-infrastructure running
 - Cheminformatics education at IU (www.informatics.indiana.edu)
- Grid concepts I

- Services are “just” (distributed) programs sending and receiving message with well defined syntax
 - Interfaces (input-output) must be open
 - Service overhead will be just a few milliseconds
- Grid Concepts II
 - Systems are build from contributions from many different groups
 - Grids are distributed in services and data allowing anybody to store their data and produce “their” view
 - “2 level programming model”
 - Grid of grids: (system of systems) use wrapping to integrate Pipeline Pilot, CBIS, Chembank, and OGSA
- Grid capabilities for science
 - Open technologies for any large scale distributed system
 - Service and message specifications
 - User interfaces via portals and portals virtualizing to desktops, email, etc
 - Uniform approach to access distributed supercomputers
- Taverna: typical Grid workflow developed in UK for bioinformatics
 - Being robust and extended by eScience program
 - Perhaps not as good as well known Pipeline pilot
- Work with Scripps for Grid workflow data mining in Earth science
- Workflows using chemical literature: All of PubMed takes a day to run through OSCAR3 on 2048 node Big Red
- Large scale calculations on “all of PubChem/Med”
 - The CDK can currently calc approx 107 descriptors
- MLSCN post-HTS biology decision support: help screening centers by developing services and data repositories (process→informatics→grids)

Questions

- Would the university libraries be interested in this?

NC State

- Our Approach and relevance to MLI Goals
 - Assessment of modeling tools ChemModLab
 - Improvement upon those tools Matrix Decomposition
 - Enable users (you!) PowerMV
- our ultimate goal is to introduce new modeling methods dedicated to informatics and new computational methods (algorithmic development on existing methods)

ChemModLab

- a research enabling platform and planning tool
- take any descriptors and incorporate them into the ChemModLab and guide us where we should go for the future
- ex: certain modeling methods work best with certain descriptors, use one set of methods to reduce something and use another set of descriptors for evaluating QSAR models
- Upload assay data

- Generate Descriptors if necessary
- Build and honestly assess models
 - o Batch process
 - o 10 running
 - o 5 descriptor sets
- Web-Accessible Modeling
 - o K-fold cross validation (double cross-validation) to create predictive activity for each model
 - o Accumulation of curves
 - o Testing order according to predictive activity (x-axis) number of compounds found (y-axis)
 - Color curves represent different methods
 - Ideally we would like to see a single curve. Best curve is far away from the ideal
 - o diversity maps address this to some extent
 - each row represents a compound, compounds are clustered by structure, each column represents a different modeling method, each graph = different descriptor
 - not a lot of grey means no cluster was overlooked
 - grey may mean we are missing complete mechanisms
- websites
 - o prediction and activities
 - o in the future, they will provide predictions for untested compounds (ex: sets from PubChem)
 - o Future: customize output to generate more direct info
 - o Also plan to provide in the same area another CSD file to tell you which Descriptors were used

Action Item:

- **provide relevant assessment criteria feedback**
- See slide for extensive list of methods
 - o Neural networks (old)
 - o Partial regressions (new)
 - o Centroid decomposition
- as of now, use all 2D Descriptors
- looked at 3 SC assays
 - o look at website
 - o submitting data = minor registration process

PowerMV

- Menu-driven web software
 - o Molecular viewer
 - o Computer descriptor site
 - o Similarity searches
 - o Some analysis methods
 - o Links to R, an open source software system

- Planning links to SAS JPP
- screen shot of power MV
 - can include files of various sorts (SD files in 2D format)
 - logistical control on the left
 - can similarity searches

Matrix Decomposition

- Data comes in the form of a matrix (structural descriptors used) and vector (bioassay results)
- QSAR Models aim to decompose X and Y by separating compounds according to mechanisms. Mechanisms are defined by structural methods
- Advantage of thinking of matrix decomposition allows for easier comparisons and all for exploring in different ways
- Algorithmic development
- How to decompose X (what is x?)
 - First option: completely ignore responses
 - Classical, mathematical point of view for decomposition
 - Methods not typically applied to Cheminformatics
 - Second Option: don't ignore the response
 - Here we have options that aren't used at all in cheminformatics
 - Orthogonal
 - Matrices must be non-negative (positive)
 - Scalability = good
 - Allows for the separation of groups of compounds without specifying the groups of clusters you would like
- Inverse regression
 - Invented in statistics literature in the early 90s, ideal for separations
 - Picture a data set with two separate mechanisms
- algorithmic development
 - sparse matrix storage are actually slower than without it
- pooling
 - collaborations with Broad ECCR
 - conducting pooled screens
 - provides efficiency and provides synergy
- **Action Item: please provide feedback, assessment criteria, ease of use, assays, descriptors, favorite methods**
- Questions:
 - Dependence of performance on descriptors based on 2-factor experiment. Analysis of variant studies. Keep in mind of interactions.
 - Are you looking at various classes of descriptors or different distribution of descriptors?
 - Pharmacophores, combined different sets of binary descriptors

MACE

- varied group (chemists, biologists, statisticians)
- If I had \$10K to find compounds, which ones would I get?

- Maximize the biological significance of chemical diversity of PubChem
 - o Increase relevance to biomedical scientists spanning different disease and therapeutic areas

UNC

- Please see slides

JOINT DISCUSSION

- Questions:
 - o Is there a general mechanism for researchers at the MLSCN who might have questions that the ECCRs could help with?
 - Answer is mediated by PubChem. But we need a mechanism so that we can work together and be happy. Need lines of open communication.
 - o Propose ECCRs meet with the IWG over conference calls every now and then
 - So that the ECCRs can address the specific needs of the community
- **Action Item: may be worthwhile for the NIH to coordinate the ECCRs, IWG, and Chemistry WG**
- How do GPCR ligands appear in the context of chemical descriptors?
 - o Push them together
 - o Have open source software so everyone can create their own descriptors to correlate with PubChem
- Unbalanced binary data set?
- PubChem is overly complicated. There is a half day course that you can sign up for.
- Where does PubChem plan to extend to?
 - o Exploratory tools. Working on electrostatic surface shaping and such.
- Problem you raised is that there is also information about what assays have been performed. If you only cluster on fingerprints, you miss out a lot of information and outliers.
 - o Plan to cluster around activity.
- Commercial web services have blazed the path
- Problem with hit assessment b/c of time delay b/w initial data submission after 1ary screen and data after 2ary assays
 - o Can create artifacts due to time dependant inhibition
- Take advantage of this opportunity to meet with each other!!!

MLSCN SC Meeting

DAY 3

ADVISOR PRESENTATIONS

Biomedical Research with a chemical toolbox – Robert Glen

Molecules to probe biological function (chemical toolbox)

- screening database
- pharmacology standards
 - o running standard assays for comparable data
- molecular probes (ex: molecules developed for pathway analysis/modulation)

Tools to Create and Analyze data (complex toolbox)

- software, data, models
 - o informational/knowledge driven

Typical problems in compound discovery

- effectively we have all of the molecules synthesizable (10^{60})
- molecules with relevance to the problem (10^{10})
- molecules we can search efficiently (10^7)
- molecules we can make and test (10^3)
- effectively, library only represents a small part of the total chemical space

Biology can be probed by chemigenomic and related methods

- modulation of gene expression products by chemical activation/inhibition

Concept:

- have available a chemical toolkit that can interact in different ways with a biological system to elicit a response
- compounds will need synthetic chemistry to make more interesting probes
- probes are different than pharma compounds
 - o may want high cell penetration that don't fit Ro5
 - more soluble = less penetration
 - high % of assays are cell based
 - o may want toxic compounds

Complexity

- novel biological targets may be simple (GPCR) or more complex (anti-cancer)
- different properties for different biological effects; need to think in different ways that require different Cheminformatic models

Comments on the current situation

- Currently, the library very small. Expansion will allow more cost effective use of screening investment
- Library composition is key, without right compounds, won't find right probes. Need to choose compounds carefully
- need to tailor compounds in the lib to cell based assays
- Compounds in library are more complicated (>500 MW). Difficult to reconstruct and resynthesize analogs around these compounds

Sociology

- centers want to do everything locally

- needs to be incorporation of ECCRs into the experimental design of the experiments and analysis of the data
- use of commercial software is difficult when mixed with open-source data. Results only reproducible with that software, may be impossible to reconstruct

Software Toolbox Concept

- Several Advantages
 - o Can start small and be expanded
 - o Choice of methods at each stage
 - o Multiple use components
 - o Variable performance of each component
 - o Requires standards
 - o Can be developed by the community as open source
- Components can be any objects
 - o Algorithms, data structures, etc.

Toolbox

- collection of self-contained modules
- can be used for compound selection and design

Already started

- ID different components
- Select models for screening
- Generate similarity models
- Need to make sure these resources are used

Comments

- technology is available to allow distributed access to any software
- some key processes can be pipelined

PubChem knowledge discovery

- already links provide navigation around various concepts in chemistry, biology, medicine
- navigation tools can be a key issue as data sources proliferate
- XML-ising may allow navigation in numerous directions

Can Cheminformatics move to a new paradigm?

What's changing?

- data was disconnected and incomplete
- single use of meta data
- if you wanted to do something to retrieve data, you had to develop your own software
- Modeling
- Reproducibility is key

Future

- transparent

PubChem

- can select compound
- link to additional databases for more information (toxicity database)
- can move from there to specific papers
- amazing how this is linked together
- knowledge discovery is happening

Data quality

- is vital
- standards are key
- need some sort of community review
 - o example: Wikipedia
- reliability of data
 - o ex: solubility of caffeine ☺
 - o 10 different measurements of solubility
 - Variable criteria/conditions

Discussion

- library composition
 - o Constraints too narrow; especially because many of the targets have few existing probes. We want to be able to access chemical space that will open the world up.
- CMLD grant to create complex, interesting compounds. Significant efforts to get the compounds into the library making the MLSMR unique and provide incentive for PIs to submit assays to the Network. Caveat, need 5 mg.
- Lipinski criteria. Screening enzyme in cell based assay vs. enzyme assay will lead to very different results. Choosing the correct library for the assays is very important
 - o Don't want a high barrier when doing cell-based assay
 - o Lo5 doesn't apply
 - More soluble, less likely to go through a lipid membrane
 - If all molecules non-toxic, won't upset the cell and won't elucidate biochemical pathways
 - Therefore, need insoluble, toxic, non Ro5 compliant compounds
- Work in CS community going on in these areas, but need investment to bring this into PubChem. Annotation by robots is possible.

From Hits to Imaging Probes – Problems and Prospects

Lee Josephson

Definitions: Hit is a small molecule that has a biological activity in cell based or free cell assay

Imaging Probe is a hit with a fluorochrome on it. These provide valuable, versatile, and unique information early in the discovery process.

Direct fluorescence cell based assays include kinetics of uptake/extrusion, intracellular distribution, and uptake of the probe in a cell.

Fluorescent Hapten immunochemical methodologies: Tissues or microtissue arrays.

This allows to-fold use of the imaging probe.

Fluorescent hapten visualization – probe is guaranteed to bind, look for distribution of the receptor; receptor with known distribution is used to analyze the binding of the probe.

FHA gives you qualitative data on the binding of probes to tissues and cells.

With tissues with known receptors, you can use this technique to evaluate where the probe goes.

Current fluorescent imaging probes are not based on low molecular weight hits.

Enzyme activities – enzyme activated substrates, covalent modification of active sites.

Agents targeted to membrane proteins.

Ways to get imaging probes: Screen Then Label; Label Then Screen, The Fluorescent Library; Screen To Obtain Scaffold, Design Focused Fluorescent Library

Considerations with designing probes – consider chemistry of the hit, chemistry of the fluorochrome, issues for detection system.

Selecting hits for fluorochrome attachment: need chemically reactive site for attachment, site whose modification doesn't impair binding to molecular target, and modification that doesn't impair membrane permeability.

Searching a database for hits amenable to fluorochrome attachment – given X (biological activity) and R (irrelevant arm, replaced by fluorochrome) variations and interactions, making a R a fluorochrome will be tolerated.

Selecting fluorochromes: you want small, neutral, no bioactivity, and not to be internalized by cells. You always want it to be reasonably priced and chemically reactive for “hit.”

There are few fluorochromes that are neutral at physiological pH. Most of the options are large, highly charged, and multiply charged.

Cell based label and screen strategy with surface modified magneto/fluorescent nanoparticles has been published. Most important part is cell associated fluorescence, which is cheap compared to MRI.

Wortmannin: a low molecular weight, cell permeable imaging probe. It has a long history in drug development – anti-inflammatory and anti-proliferative, but unstable and toxicity. Steroid-like. The 11 position was a likely place for probe attachment, and Biotin or NBD could be attached. There is a quick entry and a peak during imaging. There is an entry and extrusion phenomenon, and this has lead to new insights about designing drugs using Wortmannin. You want a cellular area under the curve. Goal

improved potency and lowered toxicity by changing cellular pharmacokinetics. You can get these results if you have an imaging probe early in the drug development process.

Nanoparticle based MRI contrast agents enhance contrast by selective uptake in normal tissue like lymph nodes. Want to visualize with haptens to see receptors expressed on normal tissue. GRP receptor in normal pancreas was targeted, without the use of cell lines in the development of the agent.

Drug dosing, safety, and efficacy are attained through knowledge of drug disposition (ADME).

Imaging probes provide valuable, versatile, and unique types of information very early in the discovery process.

For visualizing a target with a specific probe, you should only start with something at the nanomolar level. If your goal is to visualize where a compound goes in the design process, you would have to use a higher criterion. At a micromolar you might see too much fluorescence, but you need more than a nanomolar.

In the Hapten visualization method, probes are applied to a fixed cell layer or microtissue array – they don't non-specifically stick to the tissue at hand. When you add charge to a probe, you lower the permeability substantially. Charge is tied to intra versus extracellular, and the method of detection.

Many fluorophores already have a very predetermined pattern of cellular distribution. Their approach has been to select fluorophores for their lack of specificity. They want the fluorophores to act like radio labels and track things without pushing them anywhere.

MLSCN REPORTS

Vanderbilt

Status of the Projects

Library

- Library composition determines the success of the assay
- look at GPCR modulator
 - o interested at SM that are modulating receptor in novel, different ways
 - o should broaden the library to accomplish this
- investigator experience with HTS
 - o close collaboration, focused work with inexperienced investigator resulted in successful screen
 - o another assay is on hold unless the assay is completely redesigned or reconfigured → still under debate

GIRK- based assays

- search for novel agonist of G-coupled GPCRs
- screened about 14K samples for each assay

- determined slopes and looked for outliers
- looking for agonists
- also many other mechanisms where thallium can enter the cell which generates varying results of the screen
- Vanderbilt deposited data into PubChem very early on. As 2ary assays are implementation, PubChem data will be modified
- Hits = anything that changes the Thallium permeability on the cell
- Investigator can look at the name of the assay and derive incorrect conclusions about the nature of a hit
- Use Labcyte Echo 550
 - o Dose-responses Curves
 - In cell-based targets, there will be a natural convolution of compound. Should look at the potency of the compound within the target.
 - o No washing
 - o No carryover (no error propagation)
 - o Cumbersome process
 - Manual
 - Working with manufacturer to build automation and compatibility with Pipeline Pilot
 - o Cherry picking
- interesting compound hit
 - o compounds that look chemically tractable and not like any other known modulators of these pathways
 - o many ways to describe a chemical
 - o analogs in 5 different structures
 - o early stage of chemistry

Ion-channel Assay

- beta subunit of potassium channel
- presence of the beta-unit causes a change in the channels activity
- absorbance-based assay to detect NADPH fluorescence
- Reconfigured the assay to find substrates, inhibitors, potentiators, etc.
- only 10% of the compounds were self-fluoresce
- 1 compound out of the 14K that is active in this assay. See [] dependence
- Investigator found aldehydes that functioned as substrate
- This target is an aldo-keto reductase
 - o Only 4 aldehydes in the library
 - o The hit is not a aldehyde, instead ebselen
- Ebselen
 - o Has interesting known activity on KV1 channels in the cardiac system
 - o Opens K⁺ channel and quiets the system
 - o Interesting that this compound is a substrate of the beta KV subunit and it may be responsible for that known activity
- Future
 - o considering screening higher []
 - o consider cherry picking similar structural classes

New Technology Evaluations in Object-based screening

- multi channel laser scanning plate fluorometer
- high speed object based detection
- collaborating with BlueShift Biotech company creator
- started using it for immunoassays
- might be interesting tool to do a cell health and proliferation assay
- interested in collaborating with other experienced centers to use multi-parameter imaging system

Upcoming infrastructure changes: present lab space

- Quite small. Manual compound management and plate production
- expanding lab space, adding additional screening system
- multiple changes within the next 6-8 months
- 2 automated screening systems
 - o Redundant, but have some differing capabilities
- Room for additional large instrumentation
- Designing separate compound management system
- Done major hiring
 - o Chris Farmer: senior programmer on R&D for Pipeline Pilot
 - o Heading up technology core

Questions:

- **compound with a Selenium in it**
 - o **reinforces the fact we should have diverse compounds in the lib and that we shouldn't go into a screen with preconceived ideas of what we'll find**
- **Curious where Vanderbilt is thinking of going with this. Intractable compound**
 - o **Jeff Conn interested in potentiators or allosteric modulators**
 - o **Hope to find compounds that are not agonists for traditional orthosteric sites**

UNM

Highlights

- HT Flow cytometry potentially multiplex technology that can be done quantitatively and homogeneously without wash steps
- Personnel
- Focused on collaborative pipeline through outreach to bring in HT Flow Cytometry assays into the centers
 - o 4 assays came in ready to go
 - o Rebuilt 5 of them
 - o 3 that are so challenging, going to take on virtual screening approaches
- Pipeline examples/opportunities
- Plans

- Tudor's presentation

HyperCyt

- attach autosampler to flow cytometer, inverted, suspended plates
- 384 plates in 10 minutes
- Just made the system available to the public
- Autosampler moving wall to wall while peristaltic pump is running
- Last 6 month period, developed HyperPlex
 - o Adding this platform to the luminex concept
 - o Using air separated samples and beads to result in 50 plex in 1536 well format → 20M per day per detector
 - o Luminex wants to charge a lot

Personnel

- added protein chemistry expertise for bead based assays
- hired 2 more screeners (one more in the works)
- built team for management technology
- pipeline built for Cheminformatics
- senior medicinal chemists TBD

Status of the Pipeline

- look at it as family of assays that come in cycle by cycle
 - o duplex
 - o allosteric inhibition
 - o some cell based assays
 - o all come in through outreach and collaboration
- completed screening 2 assays
- Looked at efflux pumps and bacterial resistance targets
- Building large multiplex (10-plex) to be evaluated on Thursday
- HT-solubility assay
- Library fluorescence

FPR Family (FPR, FPR1)

- had to optimize cell line, receptor expression
- reformatted assay to run the assays together
- represent opportunities for novel ligands for peptide receptors in inflammation and Alzheimer's

Assay Details

- binding of a Fluor ligand competed by an unlabeled molecule re-competed by Fluorescent compound
- 2 cell populations colored to run through the flow cytometer
- Competitive binding assays
- Ran dataset as a duplex
- Picked compounds that were interesting and positive in dose response curve

- Picked separate hits. Haven't confirmed structures yet. Moving fwd with chemical characterization

Virtual Screening *in silico*

- FPR model
 - o Series of data to compare for screening different libraries
 - o One active with hit % of 1/1000
 - o 10K set, 17 hits.
 - o Class focused group from ChemDiv selected on GPCR
 - o Looking at comparison of libraries/approaches

VLA-4 Allosteric Regulator

- homogeneous binding assays
- have to get VLA-4 expression up to a certain level
- had to pay attention to fluorescence of compounds
- integrins role in metastasis
- allosteric activation/regulation
- Biogen molecule converted to fluor ligand
- Flow cytometry vs. time
- Look for molecules that allosterically limit the ability of divalent cations to bind activated ligand
- Differential binding when integrin in resting state or if activated
- Spikes occur when compounds added

Compound profiling

- flow cytometer has 9 different colors at different wavelengths
- source of color = cell uptake
- assays can be reconfigured to run at different wavelengths to avoid background fluorescence of the library

Bacterial Virulence: agr operon quorum sensing

- peptide pheromones from bacteria when they sense one and other they turn on metabolic pathways, toxins, proteases, capsules
- can be linked to fluorescence

Next in Pipeline

- Estrogen receptors
 - o Classical nuclear receptors
 - o Intracellular GPR30
 - Selective compounds
 - Radio-ligand assays
 - o looked at bioactives in the estrogen disruptive assays
 - o cell permeable assay. Plan for ligand-based virtual screen
- Target collection: prostate cell differentiation
 - o Differentiated cells are not proliferative

Cycle 5: functional assay for drug efflux pumps

- implications in cancer and drug resistance
- Identified 19 actives out of the drug library from pilot study of X01
- Currently evaluating clinical

Bead Based assays

- bead based = 1/3 of the total UNM portfolio
- Start with microspheres and build beads through HT affinity step
- Color code beads for multiplex
- Very general and broadly applicable (DNA, pro, etc)
- Homogenous resolution of free and bound to 500 nM
- pM targets
- Can build assays without every purifying protein
 - o Label proteins from cell lysates
- Real-time analysis

Bead-based Assays for inhibitors of the proteasome degradation cycle

- Implications in cancer
- Ubiquitinated proteins marked for degradation (lids fall off)
- Yeast proteasome, lids labeled with GFP, lids fall off when ATP is added
- Looking for things that inhibit and keep lids on (prevent cell cycle degradation) or activate and make lids fall off

Bcl-2 collaboration with Burnham

- purified all proteins in the Bcl family
- 2 peptide probes
- SM can regulate pro-apoptotic signals
- Task to find adequate quantities of reagents. Made in-house reagents
- Taken each member of the Bcl family on different color coded beads. Run on set of dose response curve.
 - o Bcl-2 terrible data = dead protein on bead serves as a control as the upper limit of the worse possible result

Future:

- multiplex ideas: kinases, GTPases, DNA/RNA-protein, Domain-Domain
 - o in-house collaborations, have proof of principle already
- target families, color coded beads,

Plans:

- from SOPs to current protocols Flow Cytometry
- fill pipeline
- fill personnel gaps
- address bottlenecks in infrastructure
 - o data acquisition
 - o from work station to systems automation
 - o compound handling
 - o optimize integration of chemistry and informatics

Questions:

- UNM is focusing on end-point assays rather than real-time cell based assays. Assays are predicated on the fact that cells will be developed in the wells.
- ECCRs discuss chemical space. Makes sense to screen peptides
 - o Scientifically, makes sense to screen peptides. But UNM cannot screen officially through the network. Question of how to add value to the data.
- Bcl-2 family is the first proof of principle for multiplexing for protein-protein interactions. Are there any plans for G-proteins?
 - o Yes. We have companies signed on through outreach.

Tudor

- database system for Roadrunner to allow for more flexibility than ActivityBase
- Crude interface, but essentially allow people to visualize structure and data mining of substructures
- Again, goal is to augment PubChem and allow people storing their data to manipulate their results
- Molecular property calculator
- Aggregator vs. Non-Aggregator model development (example)
 - o Downloaded MLSMR
 - o Sunset keys based on fingerprints
 - o Took data from Brian Shoika's lab
 - o Developed 100 randomized models and predicted the rest
 - o Accuracy in the 70-80% range
 - o Accuracy gets better with increased number of compounds
- Predicted MLSMR activities
 - o Took all available actives (~2K). 8% are predicted to be aggregates
 - o NCGC:
 - 3 assays published and possible 1 aggregator
 - Problems are due to the assay, not a fault of the collection
 - o 11.3% of MLSMR are predicted to be aggregators, 90% not
 - o Helpful in trying to annotate PubChem

MLSMR profiling for estrogen receptors

- goal to determine the smallest screening deck which is likely to produce a probe for estrogen receptors
- results: potential novel ligands for each sub-type of Estrogen receptor

Cheminformatics support of NMMLSC

- Roadrunner proposed functionality
- query HTS hits
- Filter HTS hits
- Process HTS hits
- Run predictive properties (solubility, pKA)
- Categorize hit list by user defined, computer-defined sub-structures, scaffolds

- Submit hit query in PubChem and ToxNet automatically
- Modify post-HTS hitlist and score HTS list by family and individual compounds map results to unique scaffolds. Automated QSAR analysis. Extend methods to capture additional information.

Question:

- How to make sure known positive controls are in the repository?
 - o **Action Item: interested if beta-lac assay with these additional compounds.**
- Does anyone have the capability to measure aggregators?
 - o NCGC can. They are running assays with and without detergent to see which ones are falling out. NCGC is not specifically looking for aggregators.
 - o UNM has the ability to develop an assay for aggregation through light scatter. Unless you know the solubility, you don't know if the light scatter is detecting aggregators or not.
- Could you elaborate a bit more on how you came up with the models for detection aggregation
 - o Non-linear model based on different inputs. Pool = 700 aggregators and 300 non-aggregators. Each time it selects an aggregator correctly, develops a model. Run 100 times. Prediction based on % of correct models.
- What about modeling surface tension?
 - o Curt working on it, no solution yet.

ECCR-led discussion on MLSCN and PubChem

- Selection of the compounds. Process of selection and account for high attrition rate?
 - o 1st 100K in the repository meant to be 150K, NIH approved 130K to be acquired. There was significant attrition before the compounds entered the door. One vendor dropped out on availability of compounds. Only 90K entered the door. 66K passed QC processes. 1% of compounds dropped out because molecular weights were incorrect. 15% dropped out because of solubility issues. 10% dropped out due to identity issues (LC/MS)
 - o Compounds selected via 4 classes (TL, DS, SS, NP). TL and DS subject to excluded functionality filters.
- [RPI] productive one-on-one discussions at the bar (MLSCN-ECCR). Suggestions from the centers to initiate interactions between MLSCN and ECCR to set up next phase of interaction.
 - o Collaboration between subgroups. Use model for more centers rather than less to optimize man power. Create standardization and avoid overlap in direction and efforts.
 - o Interaction mechanism between the ECCRs and MLSCN should be formalized via NIH.

- NECCR.org communal website forum for discussion. Top down approach matching needs and capabilities.
 - ECCRs will be a place where the field of cheminformatics will move forward. Tools developed by the ECCRs should be broadly applicable for both probe development and drug discovery.
- American Pharmaceutical Industry—advisors have gone through these exercises and there must be some sort of PharmChem where people can work together on direct projects. PubChem is for global communication. Networks are great, but we will need direct communications.
 - Tudor worked for Astra-Zeneca. Clear that chemist's need customized software to maximize interaction with ECCRs → reason for developing Roadrunner vs. ActivityBase.
- Need more collaborative interactions between biologists, chemists, and cheminformaticists. Screening centers should think about how problems will be solved by one group and what problems will need a more team-based approach. Flag those problems and see if those issues are worthy of ECCR attention.
- Problems of aggregation, promiscuous compounds, etc. are of interest to everyone. No set mechanism to further predictive modeling.
 - GSK started departments of Cheminformatics 8 years ago. Department responsible for infrastructure development via IT and scientists. Department was disbanded. 70+ computational chemists have always been integrated into development. Informatics people were not integrated. Use global computational team and site based models. They can tap into each others expertise, but the people who do the day to day validation support/experiment are not there.
 - Want to look at feasibility of tools, but we also want chemical info coming out opposed to numbers and etc. ECCRs helpful to extract useful information and make useful predictions.
 - Effective to bring in compounds after modeling, synthesized compounds, and lastly hits from the collection.
- What is exciting biologically about each assay? Informatics people need to know how to prioritize biological targets. Once it gets to the informatics people, become statistics.
- ECCRs need to know the importance priority of addressing problems
- Can we use Tudor's model to filter out compounds prior to screening SMR library sets?
- What do the Screening centers hope to get from the screens?
 - Whole plan is to synthetic novel chemistry probes that are selective for their specific biological targets
 - Sounds like selectivity is more important than potency.
- Should the ECCRs work individually with centers or are the ECCRs work more broadly? Example, descriptor types are contingent on assay types.
 - Expectation that the ECCRs will address both.

ESP Suggestions for the MLSCN

August 8, 2006

- A. Strategic Changes:
 - a) At this critical point in the pilot program, centers need guidance/direction from the NIH.
 - b) Need to restate the goals of the MLSCN to centers.
 - i) Few Centers are focused on probe development and more on interesting science, building a HTS facility, latest technology.
 - c) Centers need to develop a critical path mindset to probe development.
 - i) Focus on deliverables
 - (1) Completion of assigned assays
 - (2) Number of successful probes identified
 - ii) Focus on timelines to delivery of results
 - iii) Focus on efficiency and lowering screening costs
 - d) Each center should assign a project manager responsible for the rapid processing to completion of the assigned assays.
 - i) Project manager instills sense of urgency in completing assay.
 - ii) One manager per assay and the manager remains with the project until it is completed (internal assay champion) and then is reassigned to the new assigned assay.
 - e) Now at mid-course in pilot program, time to raise the program to the next level.
- B. Need to immediately address low application rate for X01s
 - a) Any aspect of the network which hinders the assay transfer to the centers should be carefully scrutinized and reconfigured. Ensuring smooth assay transfer is a key deliverable for the pilot study.
 - b) Switch to an R03 with \$25K award to cover reagent costs and travel to centers.
 - c) Potential assay applicants have concern that immediate release of data to PubChem would inhibit their ability to publish screening results and probe identification.
 - d) Provide strong encouragement to assay PI to move from R21 Assay Development grants to MLSCN screening.
 - e) Improve quality and reputation of compound library in MLSMR so outside investigators want their assays screened with MLSCN library.
- C. PubChem:
 - a) Delayed deposition to PubChem is good idea because:
 - i) Provides assay PI with head-start on competition
 - ii) Provides time for assay PI to prepare a publication on assay/MLSCN study.
 - iii) Improves quality of data deposited into PubChem by providing centers more time to confirm results, generate dose dependence data and rule out false positives.
 - b) Centers should be encouraged to deposit their raw data. The raw data and control data may be required to obtain the full benefit of mining PubChem.

- D. Improve MLSMR:
- a) Add novel compounds not found in other libraries (like boron & silicon containing compounds).
 - b) Throw out rule of 5 and loosen restrictive filters.
 - c) Profiling assays will add important value to MLSMR library over other available screening libraries.
- E. Chemistry may be choke point for the MLSCN
- a) How can we achieve the maximum output of quality chemical probes with the limited chemistry resources within the MLSCN?
 - b) Need to carefully select the compound structures worthy of committing our limited chemistry resources.
 - c) Should introduce a pause step after SAR on hits from MLSMR and purchased compounds to determine if synthetic chemistry is required.
 - d) Chemistry evaluation should be standardized across network.
 - e) Pool chemistry resources of MLSCN:
 - i) Suggested that following SAR and before synthetic chemistry, data from hits be submitted to chemistry committee made up of chemists from all centers. Committee would prioritize assays for biological importance and hit potential.
 - ii) The goal would be to direct the limited chemistry resources to the most promising hits.
- F. MLSCN needs to improve communication with pharma
- a) Need to find champions inside pharma to provide assays and compounds to the MLSCN. This can be done but need someone to take the time to walk it through the company.
- G. Major competition is other academic screening centers not pharma.
- a) Combination of cooperative advantage of the network and size of program should provide MLSCN with clear advantages over the competition.
 - (1) One advantage should be quality and importance of MLSMR
 - (2) Targets focused on orphan diseases
 - (3) Demonstrate success in finding probes useful in scientific investigation. Nanomolar compounds needed for target identification.
 - (4) MLSCN chemistry is hampered by limited experience to predict “drug-like” properties. If profiling assays could be generated for every compound in the screening collection, there would be a great deal of interest from industry and academics.
 - b) Need to gather intelligence on competition; number and size of programs.
 - c) Many laboratories without access to chemistry or screening resources; we should target these laboratories.
 - d) Do we know what the assay PIs want from the MLSCN; do they understand what are the deliverables and the prospects for success from screening in the MLSCN?
 - e) Capitalize on and advertise the uniqueness of technology in the centers (qHTS, multiplexing, HCS, ion channels,).

- H. Lack of resources becoming a challenge to running assays
 - a) Resources should not affect critical path
 - b) Need to develop a tactical plan to meet the challenge.

- I. Informatics programs need to be internalized into centers.
 - a) Counterproductive to have separate ECCRs and MLSCN.
 - b) Cheminformatics program must identify what is needed, work out methods and then test with real data. This can be done best within the screening centers.
 - c) Need an infinite feedback loop to test methods and integrate data.
 - d) Center informatics/cheminformatics resources do not need to be in every center but would be most efficiently applied if shared across the network.

P50 cheminformatics q: screening ctrs will generate data & deposit to pubchem. Cheminfo is: what to do w/ pubch data. How to mine the data. Several PIs were confused, work w/ screening ctr – how to work w/ screening ctrs; now NIH has to think about what the major goals are for the cheminfo ctrs – develop software for datamining. [Note added by TIO]