Chemical and Biological Data – From Compound Selection to Mode of Action Analysis (and Back Again)

Andreas Bender, PhD
Natural Philosopher for Molecular Informatics
Centre for Molecular Science Informatics
University of Cambridge
Fellow of King’s College, Cambridge
Outline

- Chemical and biological data – what is out there (and how can we use it?)

- Understanding modes of action of phenotypic readouts

- Why we need a chemical and biological view of modes of action

- Using chemical and gene expression data for MoA analysis and compound selection
Outline

- Chemical and biological data – what is out there (and how can we use it?)
Core Data Considered: Chemistry, Phenotype, Targets / Mode of Action
So what’s the point of it all?
We would like to answer questions!

- “What is the reason upon treatment with A for phenotypic effect B?”
  -> *Mode of Action*

- “Which compound should I make to achieve effect C in a biological system?”
  -> *Chemistry*

- “Does patient D or patient E respond better to drug F?”
  -> *Phenotype / Phenotype Change*
Outline

- Understanding modes of action of phenotypic readouts
Starting from *in vivo* efficacy we can predict the MoA, based on ligand chemistry.

Exploiting known bioactivity data for new decisions: Target predictions

- The models enable **automated prediction** of the targets or target families of orphan ligands **given** only their chemical structures.

```
Chemogenomics Database
Ligand 1—Target 1
Ligand 1—Target 2
Ligand 2—Target2
...
Ligand N—Target N
```

Orphan compound → Target Class Models → Predicted Targets

Public model with AZ: Mervin et al.. J Cheminf. 2015
How do you describe molecules?
E.g. using ‘Circular fingerprints’

- Each fingerprint feature represents a *central atom and its neighbors*
- Abstract enough due to losing connectivity (but keeping atom types quite concrete); disjoint (but ‘overlapping’ features) … weakness in symmetry and repeat units (terpenes, etc.)

RC Glen, A Bender, CH Arnby, L Carlsson, S Boyer, J Smith
*IDrugs* 2006, 9:199-206
Prediction Examples: Gleevec, Ruboxistaurin

- **Gleevec** (Novartis),
  - Launched
  - Targets Bcr-Abl, c-kit, PDGFRb

- **Ruboxistaurin** (Lilly/Takeda), Phase III
  - PKCb
Case study of *in silico* mode-of-action analysis

- Rat (organism-level) screen, with Eli Lilly
Understanding rat sleep data

- Project with Eli Lilly  Work by Georgios Drakakis
- Male Wistar rats   ACS Chem Biol. 2017

- Treated with ~500 sleep-inducing compounds, dozens of readouts from EEG/EMG, Abdominal Minimitter, Cage that define ‘good sleep’

- Q: What are bioactivity profiles associated with compounds inducing good sleep?

- Going from single to multiple targets (polypharmacology), and from single to multiple simultaneous MoA hypotheses for given phenotype
## Efficacy and side-effect readouts used to define desired phenotype

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Description</th>
<th>Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>NREM6hr</td>
<td>Cumulative non-REM sleep in the first 6 hours post dosing compared to vehicle</td>
<td>Efficacy</td>
</tr>
<tr>
<td>Sleep6hr</td>
<td>Cumulative total sleep in the first 6 hours post dosing compared to vehicle</td>
<td>Efficacy</td>
</tr>
<tr>
<td>LBout</td>
<td>Longest sleep bout in the first 6 hours post dosing compared to vehicle</td>
<td>Efficacy</td>
</tr>
<tr>
<td>AvgAvgBout</td>
<td>Average of the first 6 average hourly sleep bouts post dosing compared to vehicle</td>
<td>Efficacy</td>
</tr>
<tr>
<td>RebIns</td>
<td>Rebound insomnia; the cumulative non-REM sleep between hours 6-9 hours post dosing compared to vehicle</td>
<td>Side-effect</td>
</tr>
<tr>
<td>REMinh</td>
<td>REM sleep inhibition; the cumulative REM-sleep in the first 12 hours post dosing compared to vehicle</td>
<td>Side-effect</td>
</tr>
<tr>
<td>LMinh</td>
<td>Locomotor inhibition; the cumulative locomotor Activity per minute of Wake (MOW) time in the first 6 hours compared to vehicle</td>
<td>Side-effect</td>
</tr>
</tbody>
</table>
Decision trees learn receptor bioactivity profiles associated with ‘good’ and ‘bad’ sleep.
Bioactivity profiles give 6 MoA hypotheses for prospective testing (5 were selected)

<table>
<thead>
<tr>
<th>Protein Targets</th>
<th>Polypharmacological Bioactivity Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>D(2) dopamine receptor</td>
<td>1</td>
</tr>
<tr>
<td>Histamine H1 receptor</td>
<td>1</td>
</tr>
<tr>
<td>5-hydroxytryptamine receptor 2A</td>
<td>1</td>
</tr>
<tr>
<td>Transient receptor potential cation channel subfamily A member 1</td>
<td>NA</td>
</tr>
<tr>
<td>D(1B) dopamine receptor</td>
<td>NA</td>
</tr>
<tr>
<td>Muscarinic acetylcholine receptor M4</td>
<td>NA</td>
</tr>
<tr>
<td>α-1A adrenergic receptor</td>
<td>NA</td>
</tr>
<tr>
<td>Muscarinic acetylcholine receptor M1</td>
<td>NA</td>
</tr>
<tr>
<td>D(4) dopamine receptor</td>
<td>NA</td>
</tr>
</tbody>
</table>
Prospective validation on both target and phenotypic level

- 7 marketed drugs/drug combinations were selected which are predicted to modulate sleep, are dissimilar to the training set, but were not annotated with this side effect

- 5 out of 7 marketed drugs (71%) tested increased sleep parameters (a sixth led to hyperactivity!)

- 21 out of the 27 predicted targets (78%) were validated

- Overall 78% correct on target level, ~71% on phenotypic level (across 4 MoA classes)
Outline

- Why we need a chemical *and* biological view of modes of action (… and how little we sometimes understand of how drugs work)
‘Mode of action’ – what is this actually?

- Previous example: *Ligand* binding to a *target*
- Easy to understand by humans, very nice… *but!*

- Many years ago I looked at biological readouts (Western blots, gene expression data, …) of compounds with supposedly ‘similar’ pharmacology, and those were (vastly!) different

- *Hence, a protein-based view is often insufficient to describe the pharmacology of a compound on its own*

- Biased signalling, off-targets, binding kinetics, permeability/pharmacodynamics etc… many reasons
How little we sometimes know about how drugs work...

- Hypothesis: A CNS-active drug works by modulating neurotransmitters in the CNS (specific neurotransmitter, specific region) – basis of much current research

- We* compiled information from 15,777 research articles and neurotransmitter changes from experiments (comprising 110,674 rats)

- Drug class (ATC code - antipsychotic, stimulant, ...), etc., neurotransmitter, region

*Neurochemical Fingerprints of Psychiatric Drugs.
Hamid R. Noori, Lewis Mervin, Vahid Bokharaie, Özlem Durmus, Lisamon Egenrieder, Stefan Fritze, Britta Gruhlke, Hans-Hendrik Schabel, Sabine Staudenmaier, Nikos K. Logothetis, Andreas Bender, Rainer Spanagel (under preparation)
www.syphad.org (publicly, freely accessible)
So what do sedatives, stimulants, antipsychotics, … have in common?

- You would assume that disease, and treatment (mode of action of drugs), are in some way ‘defined’

- According to working hypothesis, drugs are similar on the target level and/or the functional level (which we looked at here)

- So let’s look at the data…
Neurotransmitter (functional) similarity within and between ATC classes

Neurotransmitter changes are vaguely correlated with use (ATC codes) … but only very weakly. So what is their ‘mode of action’???
Eg beta-blockers *high chemical similarity, low response similarity* (‘me too’), antidepressants assumed to hit D2 (but profiles much more diverse)…
So... how should we define the mode of action of a compound?

- Using only *proteins* to explain mode of action (or design compounds) probably only works in narrowly defined cases (eg viral proteins involved in cellular entry; inhibiting blood clotting cascades, etc.)

- Using biological readouts is likely better, *but*...

- Even using ‘biological readouts’ is insufficient – they need to be disease-related (*hypothesis-driven!*)

- So... what is the ‘mode of action’ of (eg) many CNS-active compounds...? I don’t think we understand this fully... (at least I don’t!)
- Using chemical and gene expression data for MoA analysis and repurposing
Using gene expression and on-target activity to understand interactions in mesothelioma treatment

- Abo1 is a complex herbal mixture, based on Cynara scolymus (ie, artichoke), shows clinical efficacy in mesothelioma

- How can we understand interactions (synergy?) based on chemical and biological information?

- Complex mixtures (hundreds of different chemical species), which are characterized by (a) clinical studies, (b) chemically and (c) biologically

- Work of Nitin Sharma, with Jacopo Lucci (Aboca)
Using chemical *and* RNASeq data for mode-of-action analysis

- Project with Natural Bio-Medicine / Aboca (Italian Complex Natural Products company)

- Observation (eg): fraction of Cynara scolymus (ie, artichoke) induces apoptosis in mesothelioma (*Pulito et al.*, *Oncotarget* 2015)

- Q: Can we provide MoA/biological evidence for admission as food supplement, medical device, food for medical purpose, … whatever the relevant category is that applies?

- Complex mixtures (hundreds of different chemical species), which are characterized by (a) clinical studies, (b) chemically and (c) biologically
Using chemical *and* RNASeq data for mode-of-action analysis

- Target predictions are weak on quantitative aspects of predictions, as well as analyzing compound combinations (and interactions), *but* they are easily understandable

- Hence, we use an integrated bioinformatics (RNA-Seq after compound treatment) *and* cheminformatics approach to provide mode-of-action analysis

- Important: choice of model system, parameters (concentration, time, etc.) to ensure physiological relevance
RNA-Seq readouts (left) and on-target readouts to understand MoA

Fraction of plant extract

RNA-Seq

- Time: 6 and 24 hrs
- Dose: 12 and 25 µg/ml
- Control, 6hrs and 24hrs
- Biological Triplicate samples

- FastQC, Bowtie2, Tophat, HTSeq, DESeq2

Differentially Expressed Genes

- IPA

Significant Pathways

- IPA

- IPA

Chemical Composition (47 Compounds)

- PIDGIN

Target Prediction

Reported Activity in ChEMBL

Pathway Analysis
Target prediction and RNA-Seq data gives ‘integrated view on Mode of Action’…
So what did we learn?

- Ligand-target activities are insufficient by themselves to understand MoA and synergy
- Biological readouts are needed, gene expression currently appears to have good signal-to-cost trade-off
- However, trying to understand the MoA of a complex mixture of compounds is still rather … complex
Combined gene expression / on-target activity analysis for compound selection

- Select compounds based *both* on gene expression and target prediction profiles
“BioStateConverter”
(work of Yasaman KalantarMotamedi)

- Compound-Disease mapping via gene expression data
- Drug should *invert* gene expression profile of disease
- This ‘returns the system to the healthy state’
  (better seen as *signal*, not necessarily interpreted mechanistically)
Data Sources

- ConnectivityMap (1,300 compounds to Affymetrix chips)
- LINCS (12,000 compounds to 1,000-gene expression signatures)

- Many issues with the data (dose/timepoint variability; reproducibility of controls, etc.)

- In our experience data contains sufficient signal for signal detection (but, possibly, less so for ‘modelling’)

- Gene expression data is still ‘difficult’ (regarding conditions, interpretability – less so its generation)
Selected compound induces differentiation of stem cells into cardiac myocytes (by RT-PCR; work with Dr Nasr, Royan Institute, Isfahan)

Selecting combinations active in malaria based on gene expression data

- Combining information from (a) ligand-target predictions and (b) combination screens, mapping both (c) on pathways, and (d) selecting novel synergistic combinations

- Work by Yasaman KalantarMotamedi (Cambridge), Raj Guha, Steve Eastman (NIH)

- KalantarMotamedi et al., Malaria Journal 2018 (in press)
1. Select single active compounds based on GE data
2. Enumerate target predictions, pathways of combinations
3. Combination screen – predict targets, pathways of synergistic combinations
4. Select compound combinations where synergistic pathway combinations are hit
Overall we seem to enrich for synergistic combinations, based on selecting compounds that modulate targets/pathways *predicted* to be involved in synergistic effects.

<table>
<thead>
<tr>
<th></th>
<th>Mild-to-Strong Synergy (gamma&lt;=0.995)</th>
<th>Moderate-to-Strong Synergy (gamma&lt;=0.975)</th>
<th>Strong Synergy (gamma&lt;=0.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain</td>
<td>3D7</td>
<td>DD2</td>
<td>HB3</td>
</tr>
<tr>
<td>True positives</td>
<td>18</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>False Negatives</td>
<td>11</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>False Positives</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Precision</td>
<td>78.3%</td>
<td>81.8%</td>
<td>90.5%</td>
</tr>
<tr>
<td>Recall</td>
<td>62.1%</td>
<td>72.0%</td>
<td>61.3%</td>
</tr>
<tr>
<td>F measure</td>
<td>69.27%</td>
<td>76.59%</td>
<td>73.09%</td>
</tr>
<tr>
<td>Training Set Synergies</td>
<td>42.8%</td>
<td>41.2%</td>
<td>40.0%</td>
</tr>
</tbody>
</table>
So what did we learn?

- We can apparently use information from many different sources (gene expression, ligand-target prediction, pathway annotations, …) to aim to understand and model synergy
- Pathway annotations are able to integrate ligand-target interactions with gene expression/other biological readouts
- How to put those parts together needs to be explored in more detail
Startup ‘Healx’ founded, for ‘data-driven drug repurposing in rare diseases’

- Emphasis on patient groups
- CEO Tim Guilliams, funded by Amadeus and others
- CUE ‘Life Science Startup of the Year’ 2015

www.healx.io; ~3yrs old; 15+ people
Positions available

PhD position with Bayer
“Using Deep Learning in Drug Discovery”
From October 2018 (3 years)

CAMS (Cambridge Alliance on Medicines Safety)
Junior Research Fellowship, Computational Toxicology, for PhD with postdoc experience
3 year position with research budget and co-supervision of several PhD students, working with GSK and AZ
Application deadline in June, start October 2018 -
www.jobs.cam.ac.uk/job/17074
Acknowledgments

Avid Afzal
Azedine Zoufir
Ben Alexander-Dann
Bobby Glen
Chad Allen
Christoph Schlaffner
Dan Mason
Dezso Modos
Erin Oerton
Fatima Baldo
Fredrik Svensson
Fynn Krause
Georgios Drakakis
Kathryn Giblin
Keerthy Kumar
Krishna Bulusu
Leen Kalash
Lewis Mervin
Nitin Sharma
Ranjoo Choi
Richard Lewis
Samar Mahmoud
Salundi Basappa
Sharadul Paricharak
Sharif Siam
Siti Zuraidah Sobir
Stephanie Ashenden
Xianjun Fu
Yasaman Motamedi
Yue Kong

Sebastian Rohrer
Stefan Tresch
Antje Wolf
Klaus-Juergen Schleifer

AstraZeneca

Ola Engkvist
Thierry Kogej
Jay Mettetal
Samantha Hughes

MERCK MILLIPORE

Martin Augustin
Tom Klenka

Institut Pasteur

Therese Malliavin
Isidro Cortes

erc

Ian Stott

Hinrich Goehlmann
Herman van Vlijmen
Joerg K. Wegener

Mike Bodkin
David Evans
Keith Wafford
Suzanne Brewerton

Massimo Mercati
Anna Maidecchi
Jacopo Lucci