Mode of Action Analysis
Using Chemical and Biological Data

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

- Using *in silico* methods to understand modes of action, case studies

- The problem with ‘modes of action’

- Using understanding of MoA to go forward – synergistic compound selection
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*
BUT…This is a very simplified view…

- Links between drugs/targets/diseases are quantitative (and incompletely characterized)
- There are subtle differences in eg compound effects (partial agonists vs full agonists, off-targets, residence times, etc.)
- Effects are state-dependent (variation between individuals, … depends on even what you have eaten in the morning/absorption…), not captured in the data
- Data quality is often not sufficient (biology is inherently noisy; noise+species variation)
- …
- All of this makes assigning labels such as ‘active’, ‘toxic’ etc to compounds very difficult!
Starting from *in vivo* efficacy we can predict the MoA, based on ligand chemistry.

The models enable **automated prediction** of the targets or target families of orphan ligands **given** only their chemical structures.
Prediction Examples: Gleevec, Ruboxistaurin

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

- Ruboxistaurin (Lilly/Takeda), Phase III
  - PKCb
Problem of representation of chemical structure

- No ‘natural’ way of encoding molecules
- Graph-based descriptors are information-rich; however binding is mediated ‘via the surface’ of the molecule

- Too close to the connectivity matrix doesn’t generalize; too abstract not specific enough
- ‘Middle ground’ is needed
- In (many) retrospective studies circular fingerprints gave best performance
How do you describe molecules?
E.g. using ‘Circular fingerprints’

- Each fingerprint represents a central atom and its neighbors
- For each molecule, there are as many fingerprints as heavy atoms in the molecule

RC Glen, A Bender, CH Arnby, L Carlsson, S Boyer, J Smith
IDrugs 2006, 9:199-206
Public target prediction model, based on ~200 mio data points

- Work of Lewis Mervin, with AstraZeneca
- 2015, *J. Cheminformatics* (7) 51
- ChEMBL actives (~300k), PubChem inactives (~200m)
- Can be retrained on in-house data
- 1,080 targets

- https://github.com/lhm30/PIDGIN

Also data is available to everyone!
Training MoA models using in-house SAR data

**ChemConnect**
- Orthologs with 85% sequence similarity from Homologene
- Retain targets with 10 or more active data points
- 9,570,000 actives
- 2,882 Targets

**AZ HTS Datamart**
- 420 HTS screens
- 343 Targets
- **189,500,000** inactives

**PubChem**
- 300,000+ screens
- 2,116 Targets
- Annotated inactives from HTS screens
- **420,000,000 +** inactives

AZ Data and PubChem data combined:
- **603,000,000** inactive data points
- 2,161 Targets
Functional target prediction

• Compounds do not only have a ‘class label’ against a protein

• Modulating a protein can have multiple effects (say, in the simplest case, activating and inactivating/inhibiting effects)

• Needed to map activity types to binary activating/inhibiting labels

• Complicates classification even further – now we have 500-5,000 classes, and two subtypes each!
Problem: Biased data

Typical data looks as follows:
- ~ 500-5,000 classes
- ~ 20-10,000 *actives* per class
- ~ 1,000-1,000,000 *inactives* per class
- ~ 1-100 classes per compound (instance)
- Some classes are diverse, some are not

- No reliable way to estimate underlying distributions (‘background chemical space’), or priors for classes (‘how much’ of chemical space belongs to one class)

- Problem: Estimating class-membership across this type of biased data
Understanding rat sleep data

- Project with Eli Lilly  Work by Georgios Drakakis
- Male Wistar rats  

- 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
Compounds classified, followed by pattern discovery in target space

- Efficacy and side-effect readouts used to define ‘good’ and ‘bad’ compound class

- Target prediction, say:
  - ‘Good’ compound targets: ABC, ABD, ABE
  - ‘Bad’ compound targets: ACE, BCD, BCE

- Decision trees for pattern discovery: Here targets ‘AB’ are associated with efficacy, and tolerable side effects
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

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

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

- Overall 78% correct on target level, ~71% on phenotypic level (‘positive predictive value’)
What did we learn?

- We went *in silico* from single targets to multiple targets, and multiple hypotheses, in mode of action analysis

- Able to *understand* (hypothesize) modes of action, *and select* new compounds

- Missing: Functional effects, quantitative activities (to some extent in new versions of *in silico* models), any *in vivo* (PK/PD) properties, etc.
Application: Understanding and predicting cytotoxicity in screening HTS collections

Work of Lewis Mervin, with AstraZeneca (Molndal/Cambridge)

Understanding Cytotoxicity and Cytostaticity in a High-Throughput Screening Collection

Lewis H. Mervin,† Qing Cao,‡ Ian P. Barrett,§ Mike A. Firth,§ David Murray,‖ Lisa McWilliams,‖ Malcolm Hadrick,‖ Mark Wigginsworth,‖ Ola Engkvist,⊥ and Andreas Bender‡,*†

†Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Cambridge, United Kingdom
‡Discovery Sciences, AstraZeneca R&D, Waltham, United States
§Discovery Sciences, AstraZeneca R&D, Cambridge Science Park, Cambridge, United Kingdom
‖Discovery Sciences, AstraZeneca R&D, Alderley Park, Macclesfield, United Kingdom
⊥Discovery Sciences, AstraZeneca R&D, Mölndal, Sweden
Cytotoxicity in compound sets

- Even low level cytotoxicity is linked to adverse events in man, and is hence often undesired (…where not explicitly desired)

- Aims of project three-fold:
  - Predict cytotoxicity of new compounds
  - Gain *chemical* insight into cytotoxic substructures
  - Gain *biological* insight into cytotoxicity-related mechanisms activated by small molecules
Predicting and understanding cytotoxicity of compound libraries

300k compounds profiled with AstraZeneca for cytotoxicity (in dose response)

- Single-Concentration AlamarBlue® (Screen A)
  - 388,000 compounds profiled
    - 296,970 inactives

- Dose-Response AlamarBlue® (Screen B)
  - 25,000 compounds profiled
    - 6,844 actives

- CellTox™ Green Membrane Integrity Assay (Screen C)
  - 6,836 compounds profiled
    - 5,784 actives

CTCS Dataset = Active in Screen B & Screen C
Cytostatic Dataset = Active in Screen B & inactive in Screen C
Non-Toxic Dataset = Inactive in Screen A
Chemical fragments, targets can be used for predictions, interpretation of cytotoxicity.
- The problem with ‘modes of action’
“Mode of action”… words easily said, not so easily verified

- Need to show achievement of effect, via proposed ‘mechanism’
- Involves e.g. target engagement in vivo; ruling out other ‘routes’ of activity

- MoA has different levels – target, gene level, protein level, protein activity level, …
- Operating on eg target level ‘simplifies’ problem, but possibly also oversimplifies it

- Q: What is the desired activity of a small molecule that inverts the disease state (to ‘healthy’)?
Investigating links between indications and neurotransmitter level changes

- Frequent working hypotheses of CNS active drugs: We aim for particular activity on the *target level* and/or the *biomarker level* (e.g. neurotransmitter/brain area level).
- Hoped to be linked to efficacy *in vivo*.

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

- So let’s look at the data…
So what do sedatives, stimulants, antipsychotics, … have in common?

- Hypothesis: “A CNS-active drug of a certain type works by modulating neurotransmitters (specific neurotransmitter(s), specific region(s))”

- We* compiled information from 15,777 research articles (comprising 110,674 rats) from literature:
  - Drug class (ATC code - antipsychotic, stimulant, …), etc., neurotransmitter, region

*Neurochemical Fingerprints of Psychiatric Drugs.
Hamid R. Noori (MPI Tuebingen), 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 revision) www.syphad.org (publicly, freely accessible)
So what do antidepressants, antipsychotics,… have in common?

- You would *assume* that diseases, and hence treatments (via their ‘mode of action’), are in some way ‘defined’

- How consistent are changes to neurotransmitter levels, *within* and *between* drug classes?

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

Same use, but no functional readout similarity

Different use, but same response profile (repurposing options?)

Neurotransmitter changes are vaguely correlated with use (ATC codes) … but only very weakly
So… how should we define the mode of action of a CNS-active compound?

- Not really defined on neurotransmitter (so likely also not protein) level

- Using *protein targets* to explain mode of action/ design compounds probably only ‘really’ works in narrowly defined cases (eg infectious diseases, activation of kinases/enzymes, …)

- Using biological readouts is likely better, *but*… they need to be mechanistically related to disease

- Poses problems when developing a design MoA hypothesis – what do we need to target, and how?

- Time and spatially resolved data *might* help
Novel 2-Amino-Chromene-Nitrile that Targets Bcl-2 in Acute Myeloid Leukemia (AML)

Work with Dr Basappa’s and Prof Rangappa’s Groups and Philip Koeffler; first authors are Keerthy, Manoj Garg

Screening of active compounds affecting the proliferation of HL-60 cells from a library of chromene derivatives
**In Silico** Target Predictions Suggest Bcl-2 as a Protein Targeted by this Compound

Note: In some cases – such as here – the predicted target is not necessarily the direct target, often they turn out to be indirectly targeted!
Compound 4g Decreases Expression of Bcl-2 And Increases Levels of Activated, Cleaved Caspase-9 in Human AML Cell Lines

MOLM13, MOLM14, MV4-11 and HL-60 all expressed anti-apoptotic Bcl-2 as determined by Western Blotting

Treatment with compound 4g decreased bcl-2 expression and increased levels of activated, cleaved Caspase-9
Integrated chemical and biological view on compound action..??
- Using gene expression data to understand modes of action, and explain/select synergistic compound combinations
Note on chemical and biological data

- **In chemistry**
  - We can (generally) characterize the system (compound) reasonably well
  - Chemical space is large (say, $10^{63}$ molecules?)
  - Compounds exist in different forms (conformations, etc.).

- **Biology**
  - Operates on ‘different levels’ (spatial, time, context such as cell type and state, etc.)
  - Space is smaller (say 200k proteins?) but highly connected, conditional (different cells, states of a cell/protein, etc.)
  - We (generally) don’t know what the readouts (genes, imaging readouts, ..) mean, where the signal is
  - Technology development & relevance of data don’t always go hand in hand (‘technology push’ not always helpful…)

Combined gene expression / on-target activity analysis for compound selection

- Select compounds based *both* on gene expression and target prediction *profiles*

KalantarMotamedi *et al.* *Cell Death Discovery* 2016
“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)

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; 4yrs old; ~35 people

July 2018 Series A funding
$10m, led by Balderton Capital
Identifying synergistic combinations with Gemcitabine in Pancreatic cancer

- Pancreatic cancer difficult to treat (chemotherapy; targeted treatments erlotinib, larotrectinib, not many other options)

- Gemcitabine frequently used, but efficacy relatively low

- Looking for synergistic combinations

- How? Correlation, anticorrelation, particular pathways, ...

- “Desired combination on pathways level – keeping desired anticorrelation part of activity, finding second drug that increases overall anticorrelation with disease signature”
Criteria for selecting combinations

- Score for (a) reversing *undesired* anticorrelation with disease signature, and (b) taking (resistant) Panc-1-specific differentially expressed genes into account (Panc1 vs BXPC3, Mia Paca-2, HPAFII and HS766T)
LINCS dataset for selection of compounds selective for Panc1 vs epithelial cells

- Gene expression from Panc-1, BXPC3, Mia Paca-2, HPAFII and HS766T cells as signal, selective over human pancreatic ductal epithelial cells

- 20,413 compounds applied to 77 different cell lines including 59 cancer and 10 primary cell lines with eight other cell lines compared to gene expression

- No Panc-1 in LINCS, assumed/hoped MCF-7 differential gene expression extrapolates to Panc-1

- Pathway-based signature matching of disease and compound space
Prospective validation – 9/30 combinations synergistic

- 30 compound combination prospectively tested
- 9 out of 30 compounds showed synergy (according to SUM_SYN_WEIGHTED metric in the Combenefit software using Bliss and Loewe synergy definitions)
Conclusions from pancreatic cancer part

- Gemcitabine+entinostat dose reduction index/DRI\textsubscript{50} = 43, compared to gemcitabine+trichostatin-A DRI\textsubscript{50}=3

- Despite Trichostatine HDAC1 IC\textsubscript{50} of 20nM, entinostat IC\textsubscript{50} of 510nM, so other factors in addition to HDAC inhibition possible relevant

- LINCS-derived Hypothesis (untested!): “Entinostat transcriptional profile in LINCS reverses undesired effect of gemcitabine on chromosome maintenance pathway by down-regulating BRCA1, RFC5, LIG1, POLE2 and PCNA. Only PCNA and POLE2 are down regulated in gene signature profile of Trichostatine-A as well”

- Combination changes mechanism over gemcitabine treatment alone
Understanding synergy in Shexiang Baoxin Pill (SBP)

- SBP is treatment for cardiovascular diseases from Traditional Chinese Medicine; 7 Materia Medica, 22 compounds detected in blood plasma – how do they interact pairwise?
- Modelled based on predicted targets, network information
- Work of Siti Zobir, Ranjoo Choi, Tai-Ping Fan, Dezso Modos (Cambridge)
SheXiang BaoXin Pill (SBP) is a widely-used Chinese prescription medicine for the treatment of cardiovascular diseases in China. It comprises seven materia medica, with “aromatic herbs activating yang, benefiting vital energy and strengthening the heart for treating angina and myocardial infarction caused by ischemia.” MOAs of SBP involve neovascularization through promoting angiogenesis in the heart.

**AIM:** To elucidate mechanism of action of the synergistic pairwise combination in promoting angiogenesis by using in silico and RNA-seq analysis.
a) Compounds mapped to targets

Standardized 22 SBP structures

458 predicted target by PIDGIN + 13 additional targets from literature

Mapped 281 predicted targets onto the network

b) Generation of disease network

Selected proteins which are involved in CHD and angiogenesis in at least two databases

Reactome

Signalink2.0

Network generation 2,371 nodes and 16,336 edges

Synergy score calculations of 231 pairwise SBP combinations

Validated the highest 20 and the lowest 20 combinations \textit{in vitro} using HUVEC

c) Network-based prediction of compound synergies

d) Quantitative validation of synergy predictions

e) Mechanistic analysis

Gene expression and proteomic analysis of highest observed synergy
SBP targets the central nodes of the angiogenesis and coronary heart disease network.
A ginsenoside and an adjuvant compound (cholic acid) or progesterone often show synergy.
Rg3/Rb2 combination synergistic in cell proliferation, tube formation assay

**Biological readouts**

1. Endothelial cell proliferation
2. Rescue of Homocysteine-induced tube damage
3. RNA-seq analysis
4. Validation of key genes and pathways

**Elucidate MOAs**

< Total tube area >

<table>
<thead>
<tr>
<th>Condition</th>
<th>Total Tube Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control, no HCY</td>
<td>20000</td>
</tr>
<tr>
<td>2mM HCY</td>
<td>20000</td>
</tr>
<tr>
<td>Rb2 50µM + 2mM HCY</td>
<td><strong>80000</strong></td>
</tr>
<tr>
<td>Rg3 5µM + 2mM HCY</td>
<td><strong>80000</strong></td>
</tr>
<tr>
<td>(Rb2+Rg3) + 2mM HCY</td>
<td><strong>120000</strong></td>
</tr>
<tr>
<td>bFGF 50ng/ml + 2mM HCY</td>
<td>120000</td>
</tr>
</tbody>
</table>

**Control, no HCY**

2mM HCY

Rb2 50µM + 2mM HCY

Rg3 5µM + 2mM HCY

(Rb2+Rg3) + 2mM HCY

bFGF 50ng/ml + 2mM HCY
Using gene expression data for mechanistic insight (2)

Biological readouts
1. Endothelial cell proliferation
2. Rescue of Homocysteine-induced tube damage
3. RNA-seq analysis
4. Validation of key genes and pathways

Elucidate MOAs

GO analysis of the white module: angiogenesis

GO Shortlisted 7 genes

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>PValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>extrinsic apoptotic signaling pathway</td>
<td>0.014722</td>
</tr>
<tr>
<td>GO:0097191</td>
<td>extrinsic apoptotic signaling pathway</td>
<td>0.014722</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0001525 angiogenesis</td>
<td>0.016897</td>
</tr>
<tr>
<td>GO:0001525</td>
<td>angiogenesis</td>
<td>0.016897</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0090050 positive regulation of cell migration involved in sprouting angiogenesis</td>
<td>0.030449</td>
</tr>
<tr>
<td>GO:0090050</td>
<td>positive regulation of cell migration involved in sprouting angiogenesis</td>
<td>0.030449</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>metabolic process</td>
<td>0.038043</td>
</tr>
<tr>
<td>GO:0008152</td>
<td>metabolic process</td>
<td>0.038043</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>presynaptic membrane assembly</td>
<td>0.038979</td>
</tr>
<tr>
<td>GO:0097105</td>
<td>presynaptic membrane assembly</td>
<td>0.038979</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0070886 positive regulation of calcineurin-NFAT signaling cascade</td>
<td>0.038979</td>
</tr>
<tr>
<td>GO:0070886</td>
<td>positive regulation of calcineurin-NFAT signaling cascade</td>
<td>0.038979</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0043085 positive regulation of catalytic activity</td>
<td>0.049651</td>
</tr>
<tr>
<td>GO:0043085</td>
<td>positive regulation of catalytic activity</td>
<td>0.049651</td>
</tr>
</tbody>
</table>
Validation by RT-PCR – eg CXCL8 is synergistically upregulated (etc)
So what did we learn?

• Predicting targets, using disease networks, connects formulation, chemistry, protein targets and disease biology

• We can use network topology to generally understand and predict synergy, as demonstrated for SBP

• Experimental analysis provides hypothesis for mechanism of synergy
Summary

- Chemical and biological data tell us something different about the ‘mode of action’ of a molecule

- We *can* use target prediction, gene expression data to understand parts of the mode of action of a compound

- … but MoA is not uniquely defined, different data sources provide difference parts of the puzzle

- Gene expression data helps understand MoA, repurposing, help select synergistic compound combinations
Anika Liu
Avid Afzal
Azedine Zoufir
Ben Alexander-Dann
Bobby Glen
Chad Allen
Dezso Modos
Erin Oerton
Fatima Baldo
Fredrik Svensson
Fynn Krause
Georgios Drakakis
Ines Smit
Kathryn Giblin
Keerthy Kumar
Krishna Bulusu
Leen Kalash
Lewis Mervin
Mengwu Xiao
Nitin Sharma
Peter Wright
Ranjoo Choi
Samar Mahmoud
Salundi Basappa
Sharul Paricharak
Sharif Siam
Siti Zuraidah Sobir
Stephanie Ashenden
Xianjun Fu
Yasaman Motamedi

Sebastian Rohrer
Stefan Tresch
Antje Wolf
Klaus-Juergen Schleifer

AstraZeneca

Ola Engkvist
Thierry Kogej
Jay Mettetal
Samantha Hughes

Chris de Graaf
Miles Congreave

Acknowledgments

Ian Stott
Hinrich Goehlmann
Herman van Vlijmen
Joerg K. Wegener

David Evans
Keith Wafford
Suzanne Brewerton

Jacopo Lucci
Jim Harvey