

Screening collection enhancement through open-innovation

Yury Kogej

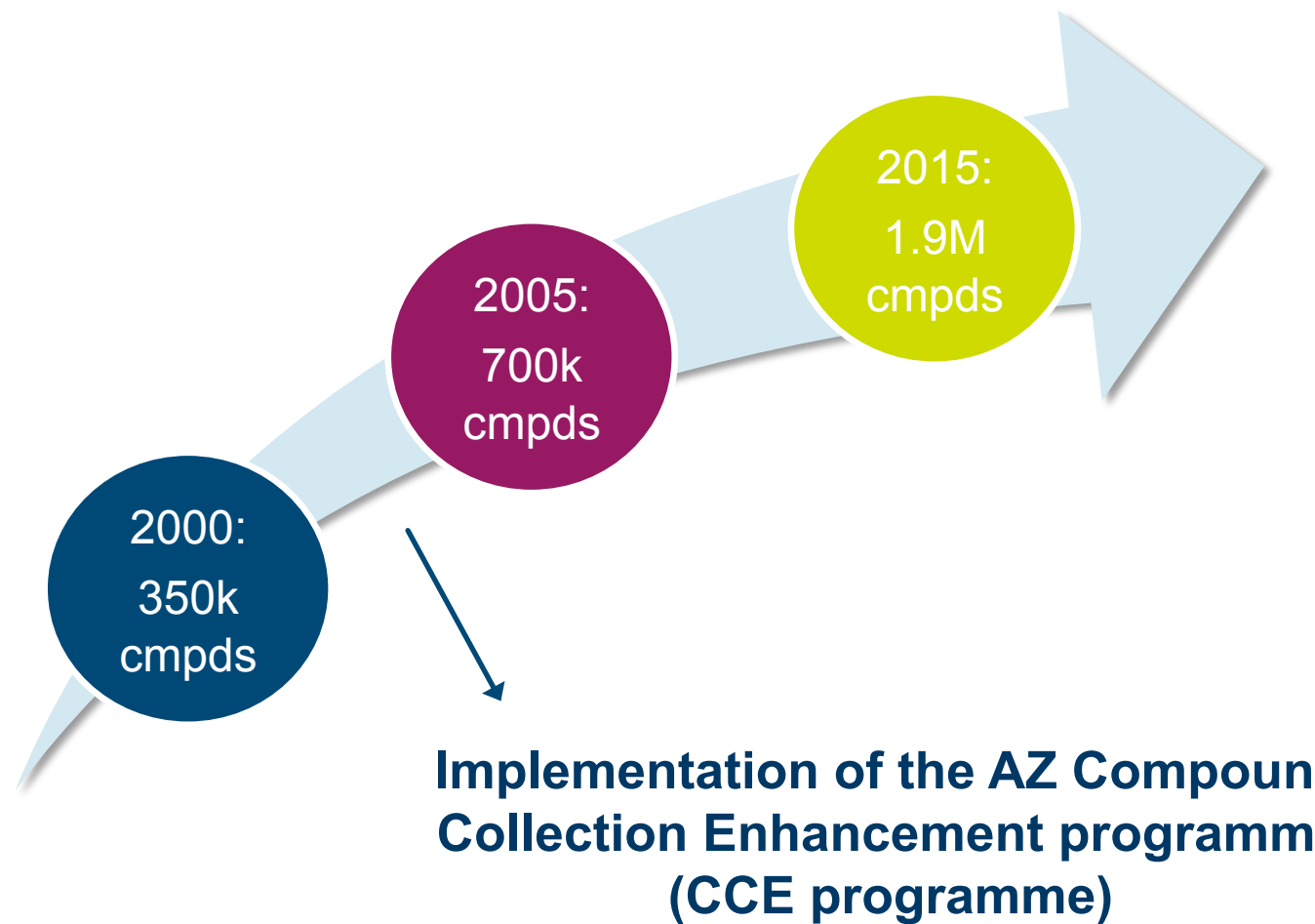
Principal Scientist, Computational Chemistry

8th March 2017

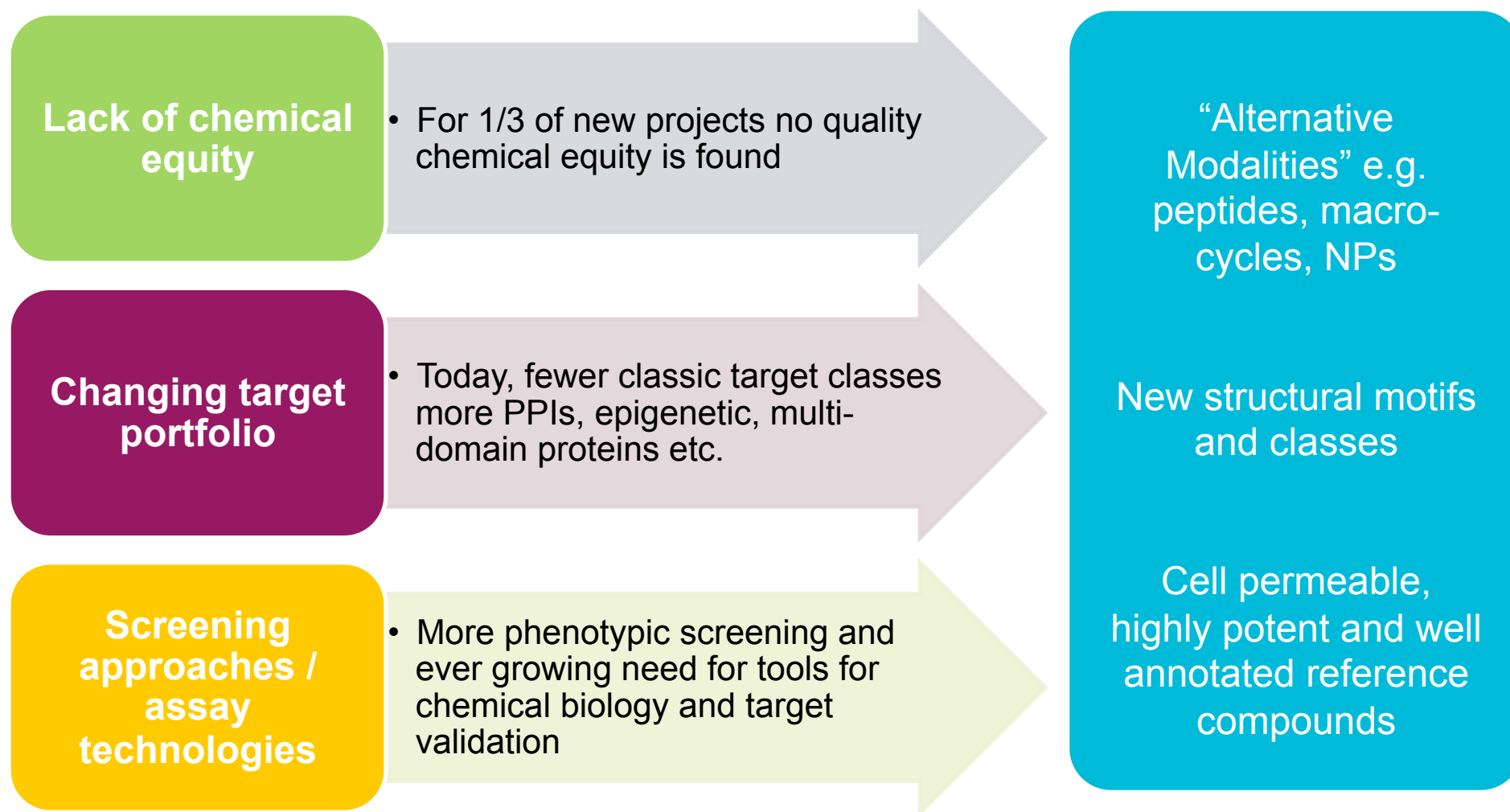


High Throughput Screening

HTS remains one of the few methods to **discover novel chemical equity** and remains the **benchmark for lead generation**
HTS success rates are typically between 50 and 60%



Do we still need to enhance our collection?



Many ways to access novel chemistry

Internal Design

- Internal design, external synthesis by CRO

\$\$\$

Cross-screening Consortia

- “Boomerang project” with Bayer Pharma HC
- IMI European Lead Factory: Access other Pharma and innovative academic design

Collection Exchange

- Equivalent number of compounds exchanged with partners, e.g. Sanofi

Collection Leasing

- Accessing complementary external collections for certain number of targets or time

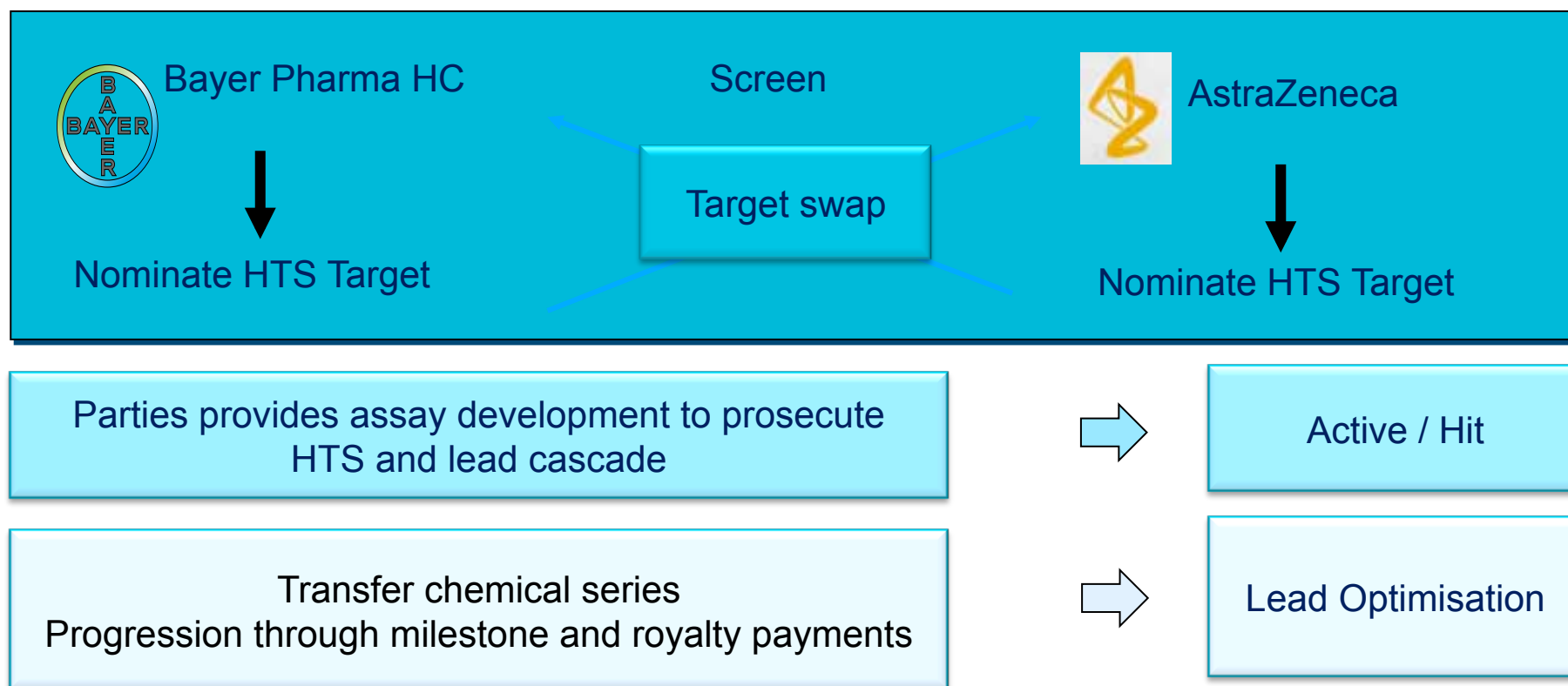
Open-innovation

Bayer Pharma HC - AstraZeneca 'Boomerang' project

a successful example of peer-peer collaboration

Pioneering Joint initiative established between AstraZeneca and Bayer in 2010 (alliance extended until 2016), based on mutual trust and shared values

Enables both parties to seek chemistry starting points not available in their internal collections



Bayer and AstraZeneca - Origin of libraries

Bayer Collection



2.7M structures

- SCHERING part of the collection (875.000 structures) was cleaned and expanded library with mainly purchasable external compounds (2003-2005)
- Huge investments at BAYER between 2000 and 2007 to expand library based on proprietary building blocks
- Compound design based on favorable PhysChem properties and undesirable groups filtering
- Realization through external collaborations and internal combinatorial chemistry
- 1/3 classical medchem structures from optimization projects
- 2/3 combichem compounds

AstraZeneca Collection



1.4M structures

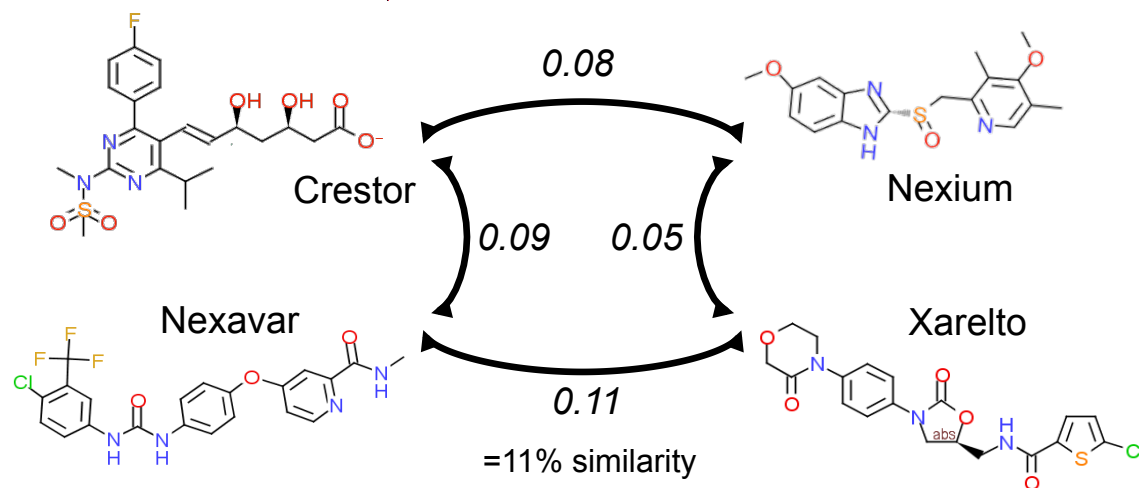
- AstraZeneca screening collection underwent major review (structural and sample quality) in 2001/2002
- Strict classification on Phys-Chem and structural features
- Major acquisition campaign in 2002
- Three consecutive *Compound Collection Enhancement* programmes (2003-2005, 2006-2008, and 2008-2011)
- Internal design from Lead Generation chemistry, outsourced production of small libraries, no combichem
- >80% proprietary compounds

- The Bayer collection consists of MedChem designed proprietary classical and combinatorial compounds
- The Astra collection consists of MedChem designed project and enhancement compounds and acquisitions

Comparison of Bayer and AstraZeneca collection

On-site “workathon” at AstraZeneca Sweden with Bayer + AstraZeneca scientists (in 2011)
Collaborative calculations and analysis effort of pre-prepared fingerprint and data-files
3 days of compute time
1920 cores / Infiniband QDR 40Gbit interconnect
...resulting in 150 GB of data

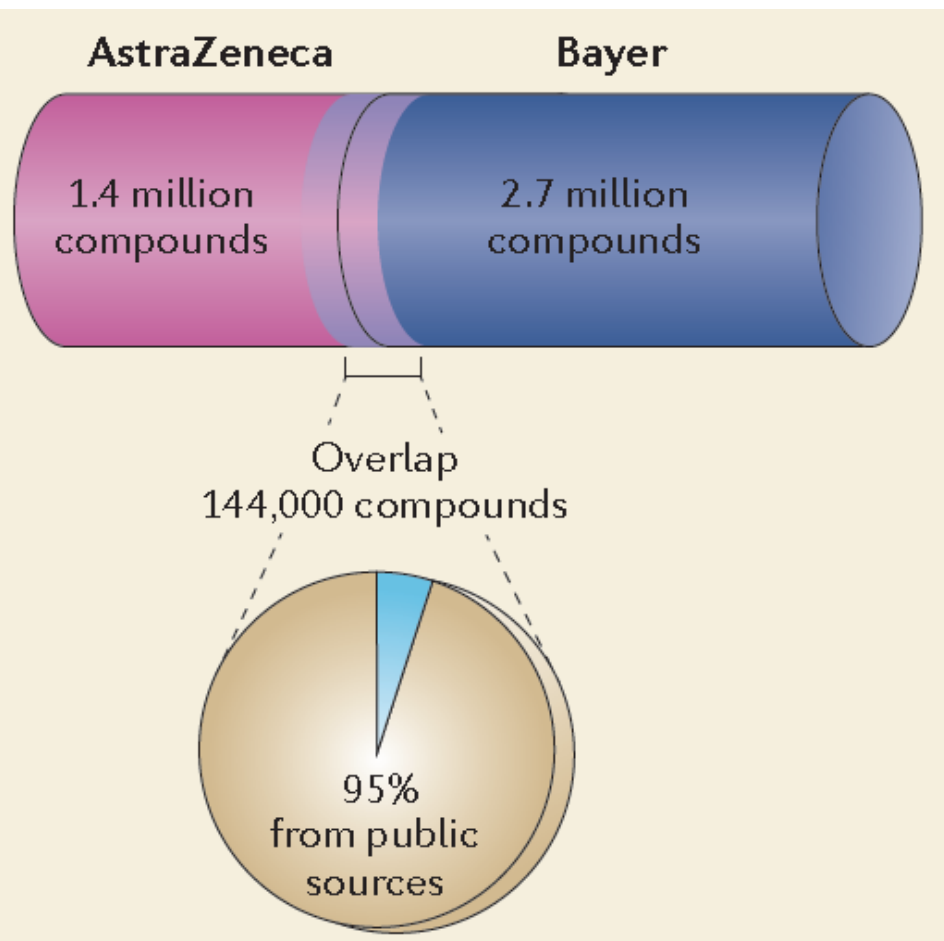
our laptops and >>50 cups of coffee...
very fruitful scientific discussions on methods and results!



Similarity is based on comparison of fingerprints using the *tanimoto distance*
CFP4 Fingerprints are derived with Pipeline Pilot.
Duplicate fingerprints for each collection were eliminated prior to computation

- Diverse research areas and projects led to structural diverse libraries
- Fingerprint and Tanimoto Index (ranging [0-1]) map this diversity

Overlap of Bayer Pharma HC and AstraZeneca collection identical fingerprints



- 3.3% of the total collection (Bayer + AZ is overlapping*)
- 95% of the overlap are public domain compounds

⇒ Screening of > 4.2 Millions unique cmpds

*) As we are not sharing structures for analysis the overlap is based on exact match of molecular fingerprints (ECFP4). This is an overestimate of identity as a small fraction of non-identical compounds will have the same fingerprint

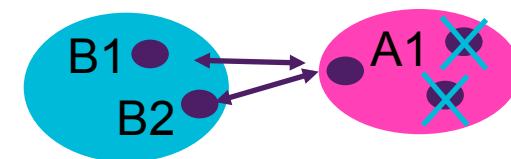
Big pharma screening collections: more of the same or unique libraries? The AstraZeneca-Bayer Pharma AG case

Kogej T, Blomberg N, Greasley PJ, Mundt S, Vainio MJ, Schamberger J, Schmidt G, Huser J. Drug discovery today (2013).

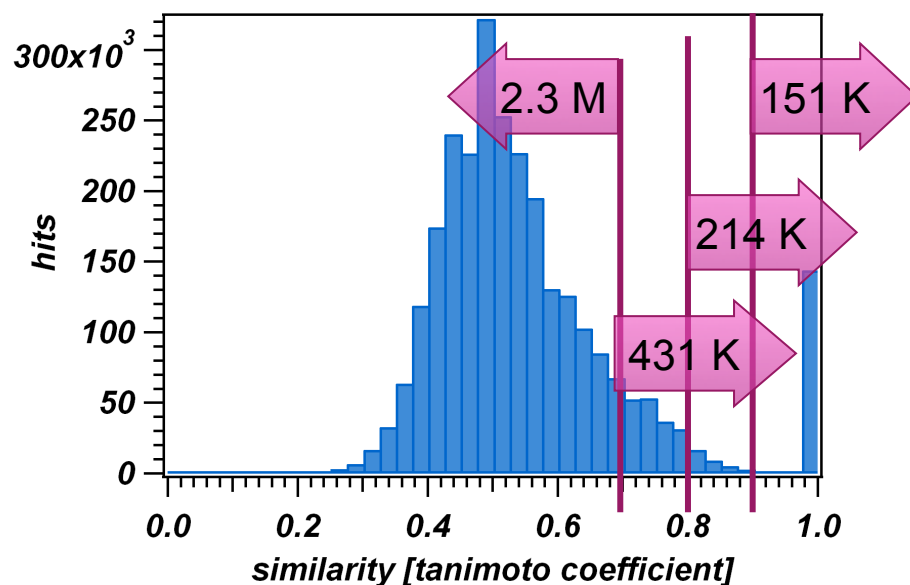
Similarity of Bayer and AstraZeneca collection

Nearest Neighbour Distribution

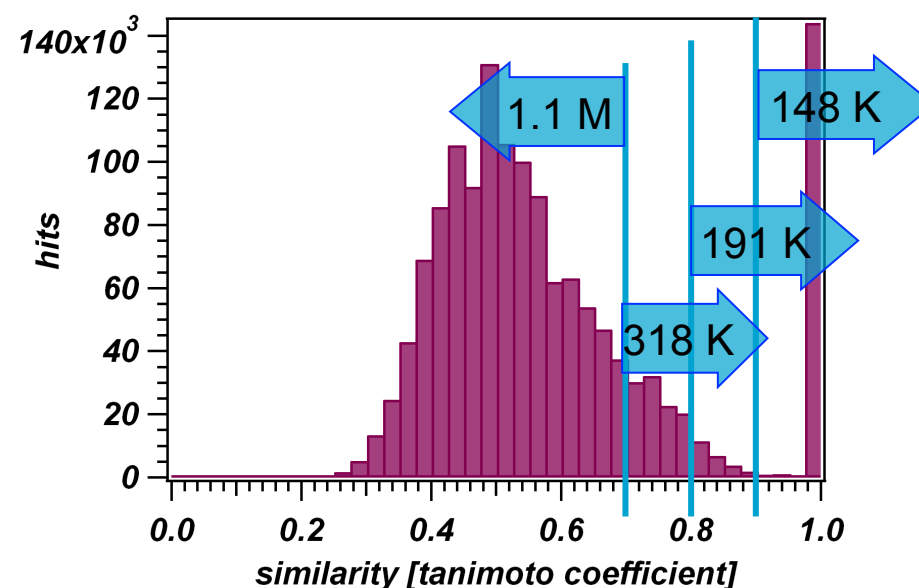
NN



Bayer collection



AstraZeneca collection



"The AZ business case":
2.3 M Compounds are "new" to Astra

"The Bayer business case":
1.1 M Compounds are "new" to Bayer

*A ECFP4 similarity cutoff of 0.70 seems high, but the main objective of this study was to access that the two collections had complementary or alternative chemical series that might not contain necessarily totally novel chemotypes or being in a different part of the chemical space. We should also consider that some subtle changes in the similarity value can dramatically change the target space of the compound. e.g. slightly reduce/increase nitrogen basicity.

Many ways to access novel chemistry

Internal Design

- Internal design, external synthesis by CRO

Cross-screening Consortia

- “Boomerang project” with Bayer Pharma HC
- IMI European Lead Factory: Access other Pharma and innovative academic design

Collection Exchange

- Equivalent number of compounds exchanged with partners, e.g. Sanofi

Collection Leasing

- Accessing complementary external collections for certain number of targets or time

Open-innovation



Collection exchanges

Principle: *quid pro quo* exchange of compounds

no cash payments

no royalties

Identify set of
compounds suitable for
sharing



Join meeting



2 ICore7, 8 threads, SSD
2 days (including travel)



Physically exchange the
samples

not AZ patents

not restricted by other alliances

not in active projects

high quality physical sample and
sufficient amount (no stock depletion)

2 days

Standalone computer (not network, clusters)

Isolated room

Sharing structures

AstraZeneca: Sanofi exchange

210k novel screening compounds
200 µL of 10 mM solution (>10y HTS)

Strong relationship foundation for
future collaboration

Positive example of cross Pharma
collaboration

THE WALL STREET JOURNAL.

Drug Giants Seek Edge By Sharing Secrets
AstraZeneca PLC and Sanofi SA have agreed to share thousands of their proprietary chemical compounds with each other, an unusual deal that shows the creative lengths to which pharmaceutical companies will go to pursue new drugs.

FINANCIAL TIMES

Sanofi-AstraZeneca chemicals swap takes open route to R&D

The deal — the biggest of its kind among major pharmaceuticals groups — highlights an increased openness to co-operation in an industry criticised in the past for secrecy and beggar-thy-neighbour practices.

The Daily Telegraph

AstraZeneca and Sanofi agree novel drugs-sharing deal
Britain's AstraZeneca and French pharmaceutical company Sanofi have agreed a landmark deal to share data - for free - in the hope that it could lead to breakthrough treatments for disease.

SCRIP

Intelligence
Is AstraZeneca/Sanofi Library Exchange Open Innovation's Future?



REUTERS

Sanofi, AstraZeneca swap compounds in new twist on open drug R&D

What we were/are interested to ?

structurally **diverse organic scaffolds** from low to high molecular complexity

diverse heterocycles including ones from known privileged structures

a **fragment screening set** of small heterocycles, small undecorated polycycles, plus various common organic linkers and branching groups

linear and cyclic **peptides** with established bioactivity

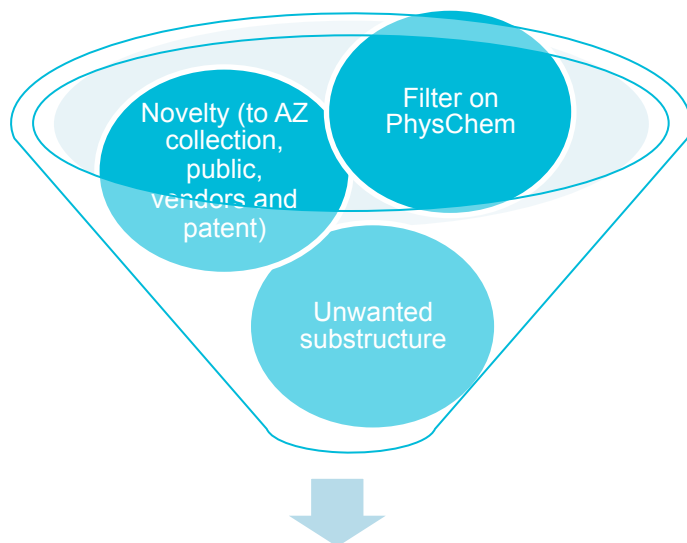
novel oligomers based on **helix, strand, turn and coil structures**

known drugs and clinical candidates

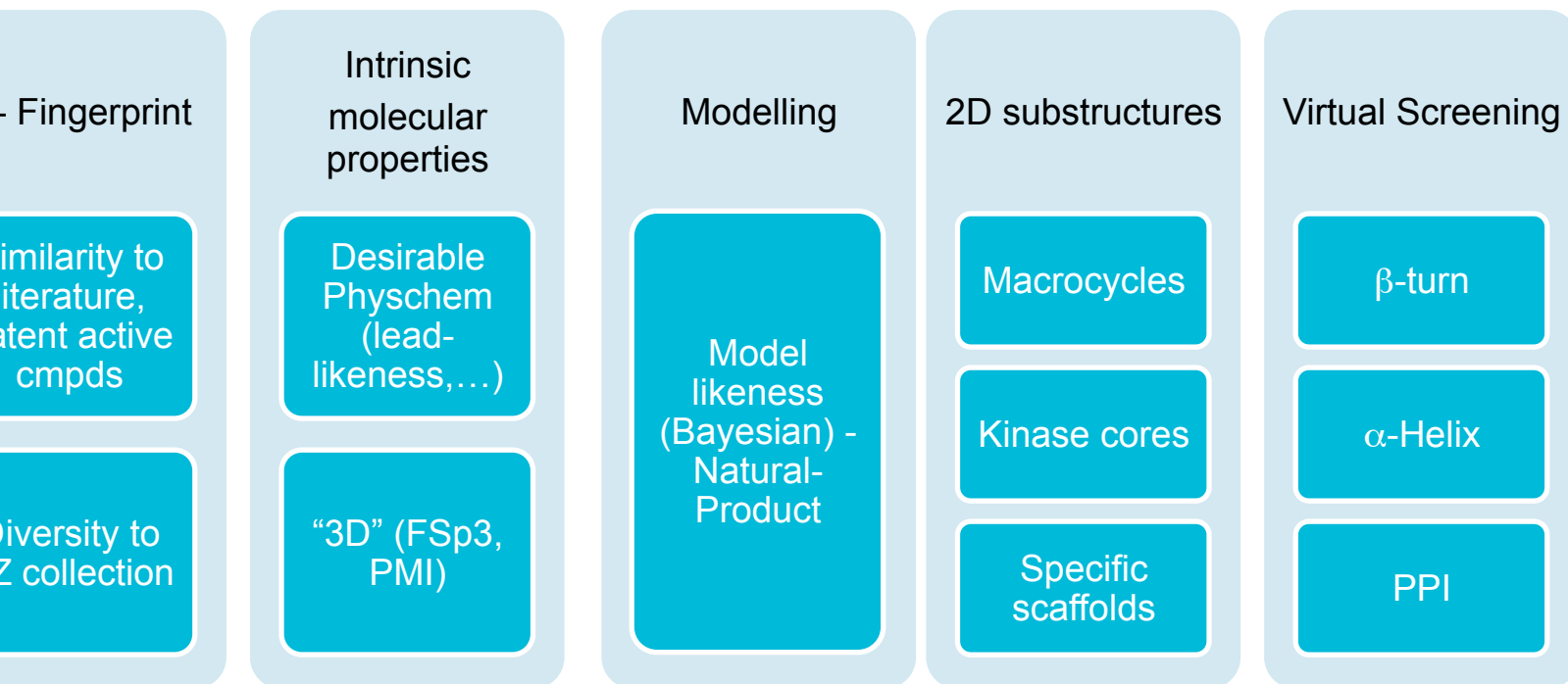
diverse collection of isolated **natural products**

Selection process

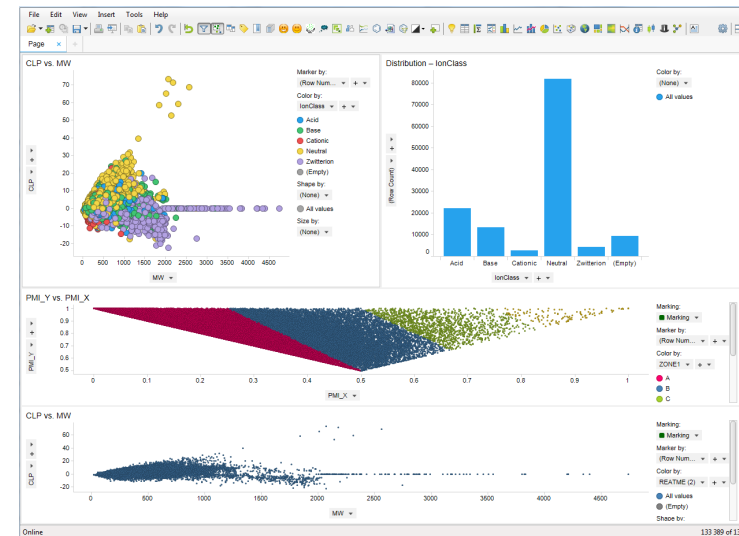
"Filtering OUT"



"Value annotation"



Result visualization



Clustering, grouping

visual inspection

Final selection

cheminformatics assessment of external collections

Novelty to AstraZeneca



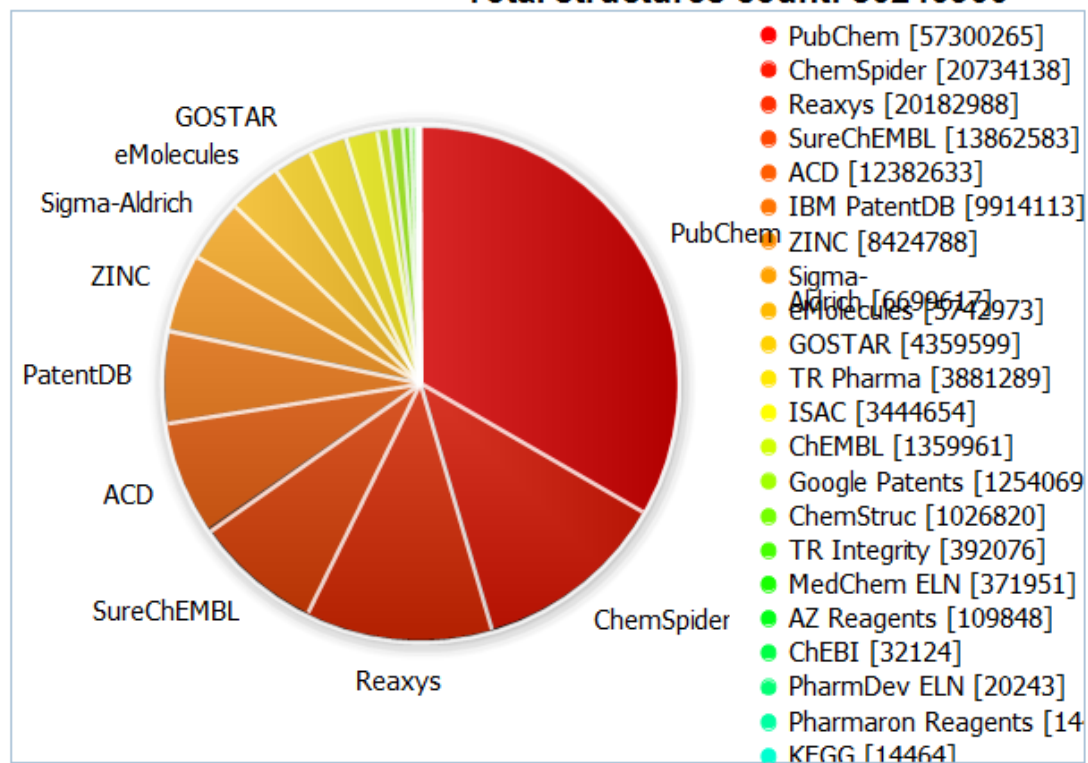
high interest in Sanofi
proprietary compounds

low interest in
publicly available compounds

high interest in "tool compounds"

Chemistry Connect Content

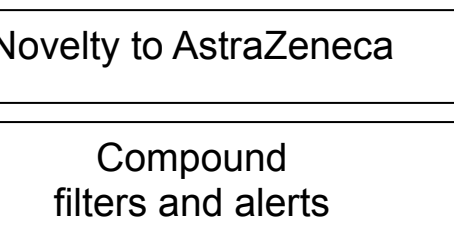
Total structures count: 86246966



Making every SAR point count: the development of Chemistry Connect for the large-scale integration of structure and bioactivity data.

Muresan S, Petrov P, Southan C, Kjellberg MJ, Kogej T, Tyrchan C, Várkonyi P, Xie PH. Drug Discov Today. 2011, 16, 1019-30

cheminformatics assessment of external collections



AZFilters®

Property filters

- $100 \leq \text{MW} \leq 550$
- $-2 \leq \text{ClogP} \leq 6$
- $1 \leq \text{PSA} \leq 160$

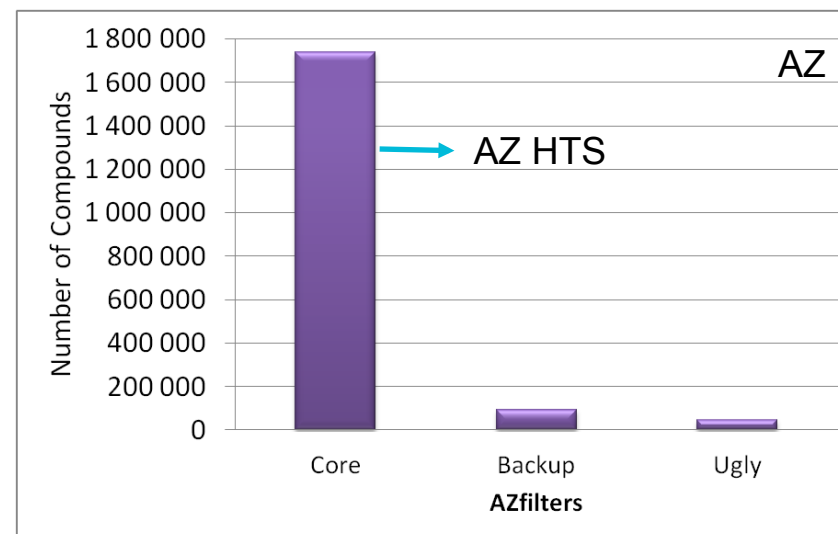
Chemical filters

- ~150 chemistry alerts (ca 500 SMARTS)
- 11 classes (reactives, unwanted structures,...)

	Prop. Filters	Chem. Filters
Core	No violation	No match
Back-Up	1 violation	No match
Ugly	no matter	≥ 1 match
Ugly	≥ 2 violations	no matter

Chemical predictive modelling to improve compound quality
 Cumming JG, Davis AM, Muresan S, Haeblerlein M, Chen H. Nature Review Drug Discovery, 2013, 12, 948-62

- Alerts:
- Genotoxicity via Bursi alert
 - Reactive Metabolite
 - PAINS structures



cheminformatics assessment of external collections

Novelty to AstraZeneca

Compound
filters and alerts

Physchem

Molecular Complexity

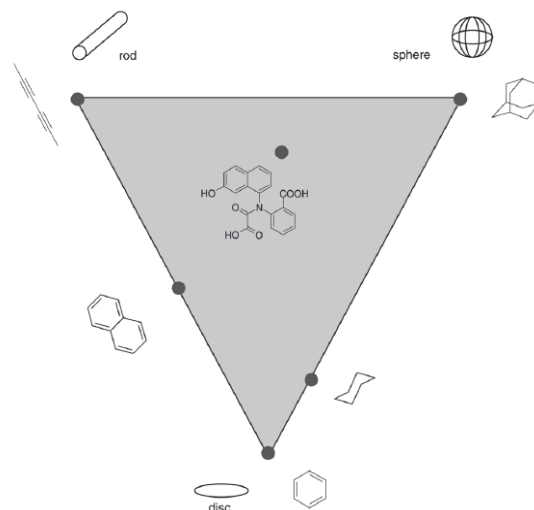


Bioactive Molecules: Perfectly Shaped for Their Target?

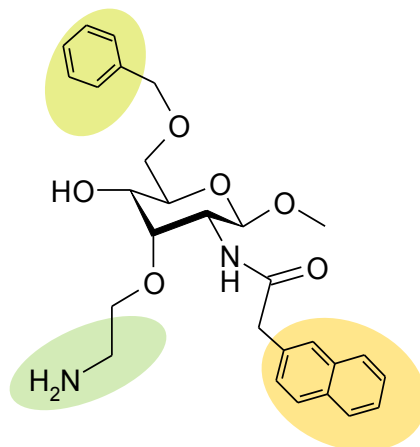
Matthias Wirth^{*[1]} and Wolfgang H. B. Sauer^[2]

Mol. Inf. 2011, 30, 677 – 688

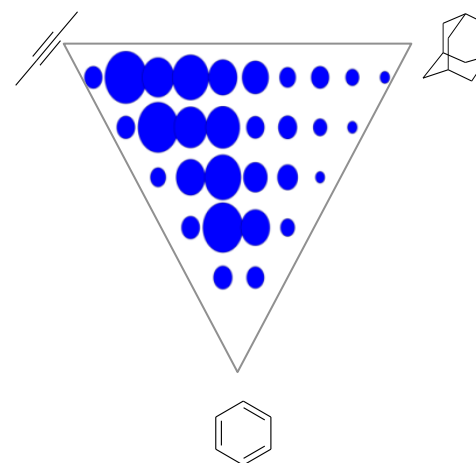
Principal moment of inertia



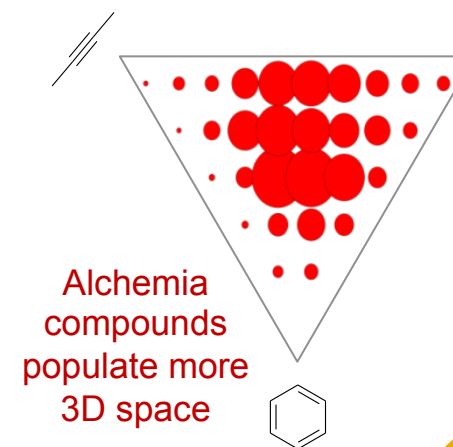
Alchemia



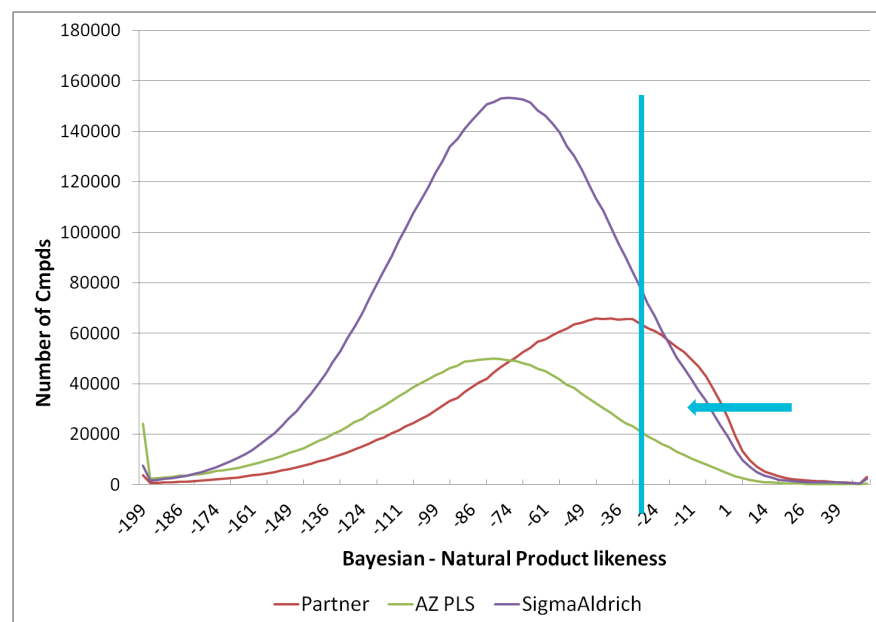
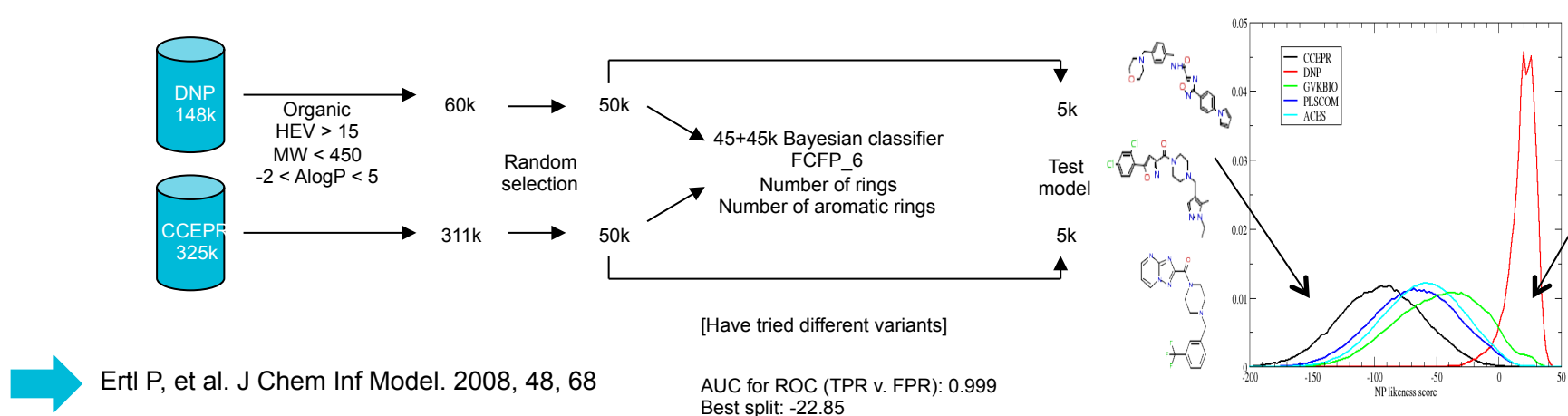
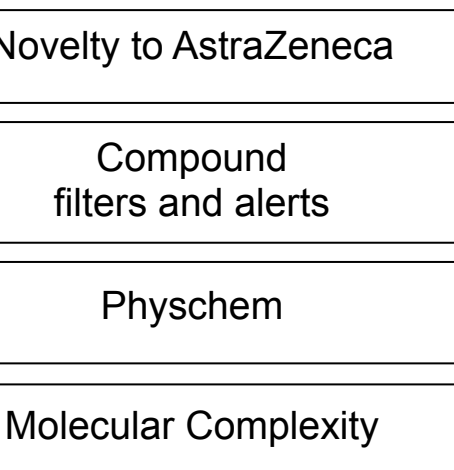
AZ collection



Alchemia 'sub-library'



cheminformatics assessment of external collections



Score > -25 is associated with "natural product like" cmpd

cheminformatics assessment of external collections

Novelty to AstraZeneca

Compound
filters and alerts

Physchem

Molecular Complexity

3D shape/2D search

2D structures:

Macrocycles, spiro scaffold, “ring of the futures”, privileged structures

3D Know mimetics:

-helix i.e. nutlins
β-strand → i.e. peptidomimetics
β-hairpin → macrocycles
PPI Co-crystallize X-ray structure (ref:)

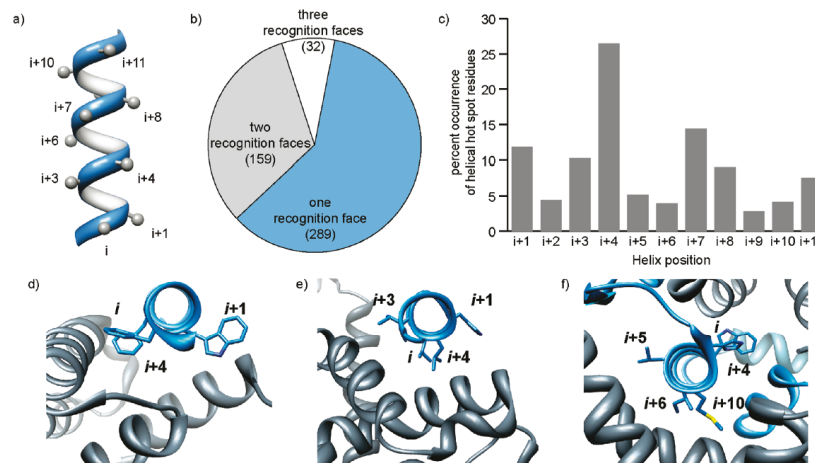
OpenEye
tools

- Filter database to include only compounds with $25 < \text{HAC} < 40$
- Generate omega db (max 100 conf)
- Run ROCS with different tanimoto c depending on the query

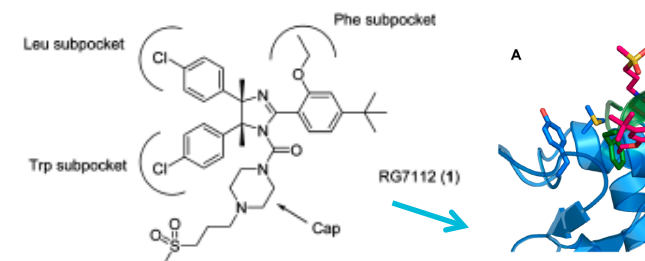


Journal of the American Chemical Society

COMMUNICATION



JACS 2011,133,14220



ACS Med. Chem. Lett. 2013, 4, 660

cheminformatics assessment of external collections

Novelty to AstraZeneca
Compound filters and alerts
Physchem
Molecular Complexity
2D shape/2D search
Fingerprint-based similarity/diversity
Clustering
Visual inspection

Assessment based on 2D fingerprint based (Tanimoto – ECFP4):

- Similarity to reference active compounds from ChEMBL, GVKBio to enrich AZ biological active space
- Diversity to AZ collection

- Ensure maximum coverage of external collection
- Reduce singleton risk
- Use of 2D fingerprint approach and scaffold analysis (Murcko)

Last but not least, review from medicinal chemists



Many ways to access novel chemistry

Internal Design

- Internal design, external synthesis by CRO

Cross-screening Consortia

- “Boomerang project” with Bayer Pharma HC
- IMI European Lead Factory: Access other Pharma and innovative academic design

Collection Exchange

- Equivalent number of compounds exchanged with partners, e.g. Sanofi

Collection Leasing

- Accessing complementary external collections for certain number of targets or time

Open-innovation



<https://openinnovation.astrazeneca.com>

AZ Open Innovation: New Molecule Profiling

Your molecules are welcome!

The screenshot shows the 'New Molecule Profiling' page on the AstraZeneca Open Innovation website. The page has a dark blue header with the 'openinnovation' logo, navigation links (About Us, AZ R&D Focus Areas, What We Offer, Partner With Us, Resources), a 'Newsletter Sign Up' button, and a search bar. A sidebar on the left lists various modules: Overview, Clinical Compound Bank, Pharmacology Toolbox, Target Innovation, New Molecule Profiling (highlighted), How Does it Work, Compound Structure Security Provisions, and R&D Challenges. The main content area features a molecular structure graphic and text explaining the module's objective: to identify novel compounds active in relevant disease biology assays for collaborative work with external investigators. It also includes a section titled 'Why Use AstraZeneca's New Molecule Open Innovation Module?' and a list of resources: Instructions to Authors, Legal Document - Fillable pdf, and Sample Cheminformatics Report.

openinnovation
About Us AZ R&D Focus Areas What We Offer Partner With Us Resources Newsletter Sign Up 192 Proposals Submitted

New Molecule Profiling

The objective of this module is to identify novel compounds active in relevant disease biology assays that serve as the foundation for further collaborative work with external investigators. By providing access to sophisticated cheminformatics and screening technologies, we aim to invite partnerships with top global research talent and ultimately advance the discovery of novel therapeutics to improve patient's lives.

Why Use AstraZeneca's New Molecule Open Innovation Module?

AstraZeneca offers external investigators an exceptional opportunity to access our modern drug discovery screening programme, representing a broad range of assays spanning [therapeutic areas](#) of current

- Instructions to Authors
- Legal Document - Fillable pdf
- Sample Cheminformatics Report

The screenshot shows the AstraZeneca Open Innovation homepage. It features a dark blue header with the 'openinnovation' logo, navigation links, a 'Newsletter Sign Up' button, and a search bar. The main content area has a large banner with the text 'Advancing Research Together' and a list of modules on the right: Clinical Compound Bank, Pharmacology Toolbox, Target Innovation, New Molecule Profiling, R&D Challenges, and Suggestion Box. Each module has a brief description and a 'Learn More' link.

openinnovation
About Us AZ R&D Focus Areas What We Offer Partner With Us Resources Newsletter Sign Up 192 Proposals Submitted

Advancing Research Together

Sharing of ideas and collaboration to push the boundaries of drug discovery and deliver life-changing medicines to patients with previously intractable diseases

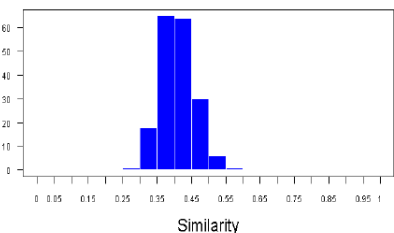
- Clinical Compound Bank** Compounds with evidence of human target coverage and manageable tolerability are available. [LEARN MORE](#)
- Pharmacology Toolbox** Compounds with optimised properties are available for preclinical research to explore disease biology. [LEARN MORE](#)
- Target Innovation** Have an innovative idea for a drug discovery project? Our compound library may be able to help you validate your idea. [LEARN MORE](#)
- New Molecule Profiling** Explore properties and therapeutic potential of compounds from cheminformatic and screening technologies. [LEARN MORE](#)
- R&D Challenges** To expand our problem solving ecosystem, we collaborate on key R&D hurdles and reward innovative solutions. [LEARN MORE](#)
- Suggestion Box** Do you have an idea, technology, or suggestion that fits outside of the above mentioned collaboration offerings? [CONTACT US](#)

More under

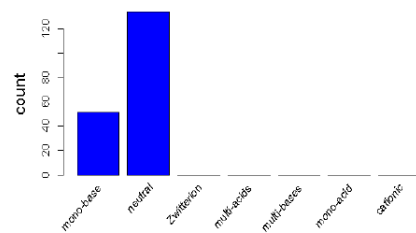
<http://openinnovation.astrazeneca.com>

Similarity

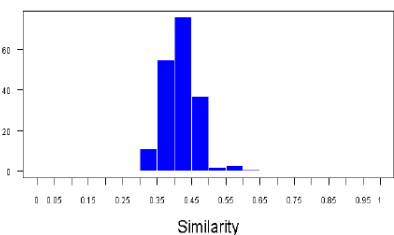
ECFP4 Tanimoto Similarity to AstraZeneca



Ionclass Distribution



ECFP4 Tanimoto Similarity to Pubchem



Main considerations for recommendation:

- Novelty
- Molecular complexity
- “Drug discovery friendly”

Structure filters

GenoTox	Count	Percentage
No	185	100.00%
Yes	0	0.00%

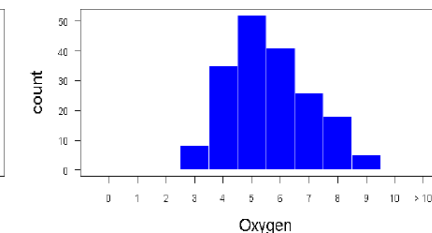
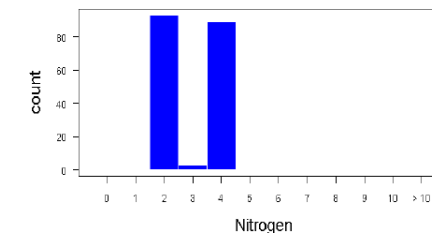
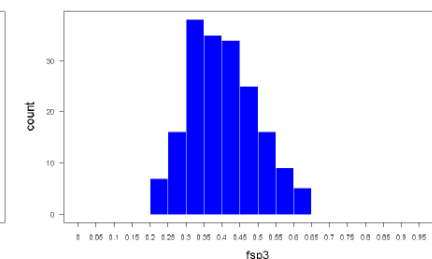
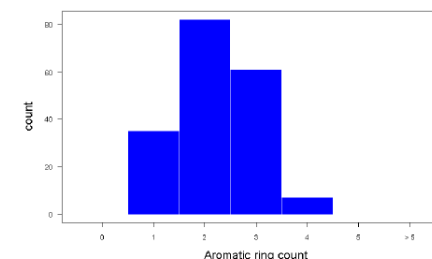
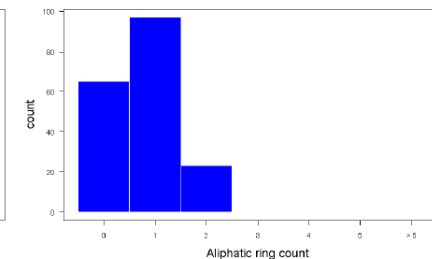
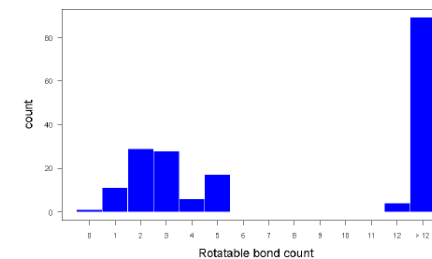
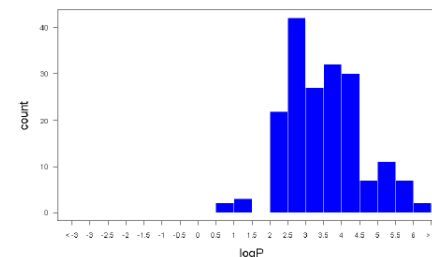
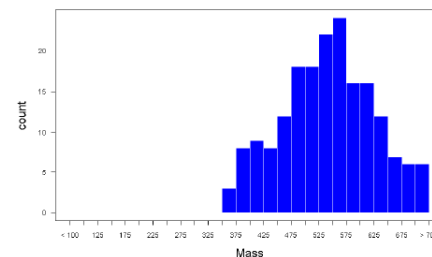
Reactive metabolite	Count	Percentage
No	185	100.00%
Yes	0	0.00%

Risk level	Count	Percentage
No Risk	74	40.00%
Risk	111	60.00%
All Known explosives	0	0.00%

Controlled	Count	Percentage
No	185	100.00%
Yes	0	0.00%

When	Metals and metalloids
After standardization	[]
Before standardization	[]

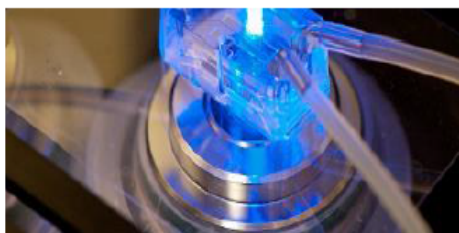
Physchem properties



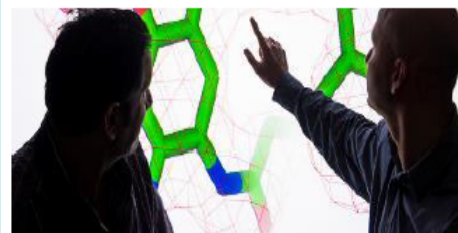
**Being open for
collaboration...**



**...creating an
environment where
science thrives**



**...and challenge
conventional
thinking**



...will enable us to change the way we treat disease and transform lives

What science can do

With special thanks to...

HTS:

Mark Wigglesworth

Martina Fitzek

Marian Preston

Carolyn Blackett

Dave Murray

Kirsty Rich

Matt Collier

OI:

Craig Wegner

Pam Hill

Hitesh

Discovery Sciences:

Mike Snowden

Steve Rees

Dave Smith

Michael Kossenjans

Selmi Nidhal

David Andrews

Clive Green

Kevin Cross

Ian Sinclair

John Cuff

Phil Spencer

Abdul Ingar

Rick Davies

Computational Chemistry:

Ola Engkvist

Hongming Chen

Isabella Feieberg

Thierry Kogej

BD:

Iain Comely

Terry Reed

Duncan Young

Vicki Foster

IMED

Fred Goldberg,

Lena Ripa

Backup

Z hit finding strategy: multiple approaches

Diversity Screening

High-throughput screening (HTS)

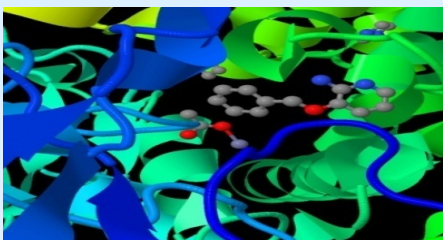
Broad set of assay technologies for all types of targets

Strong capacity for cherry picking



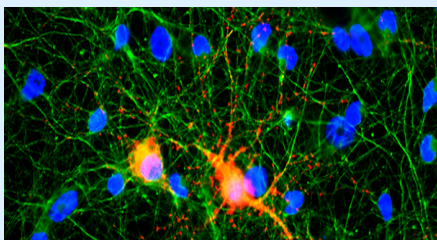
Structural Biology & Fragments

- Large experience in FBLG by NMR & SPR, and assays
- Dedicated team to perform chemistry
- Broad range of biophysical technologies
- Specific fragment library for FBLG



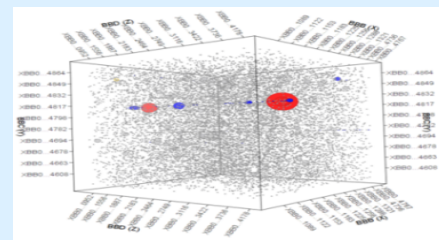
Phenotypic Drug Discovery

- Hit discovery using primary / stem cell assays for finding of pathway specific hits
- iPSC alliance with external partner
- Strategy for target deconvolution



DNA Encoded Library Technologies

- Screening of very large libraries of compounds ($>10^{10}$ molecules)
- Collaboration with external partner
- Build-in selectivity profiling in hit finding strategy
- Proprietary libraries



Public-Private & Private-Private Partnerships

- Hit discovery using academic and industrial partners
- Novel chemical space
- Novel biological space
- Synergies between partners



Key to all approaches is access to high quality screening compounds!