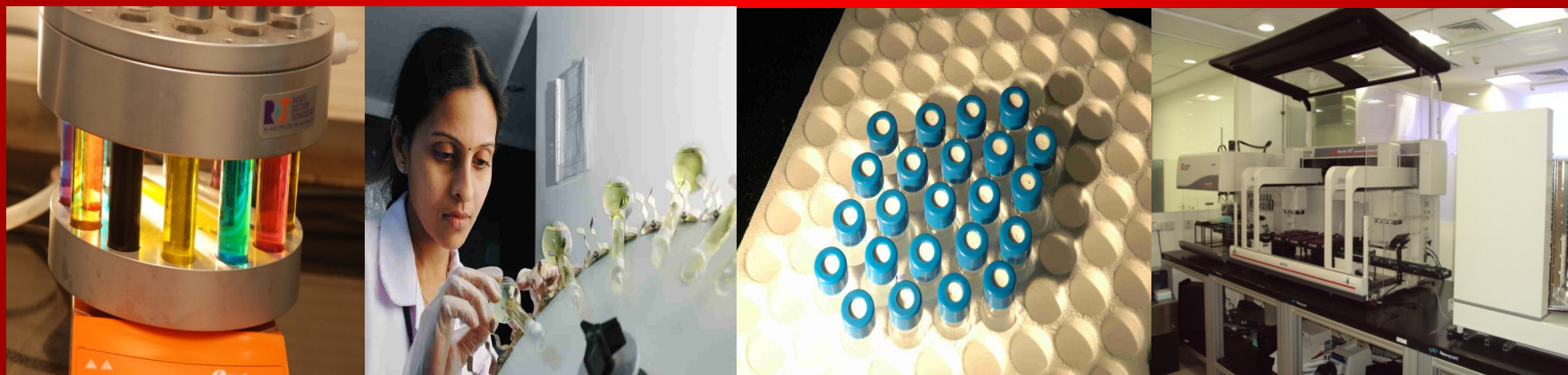
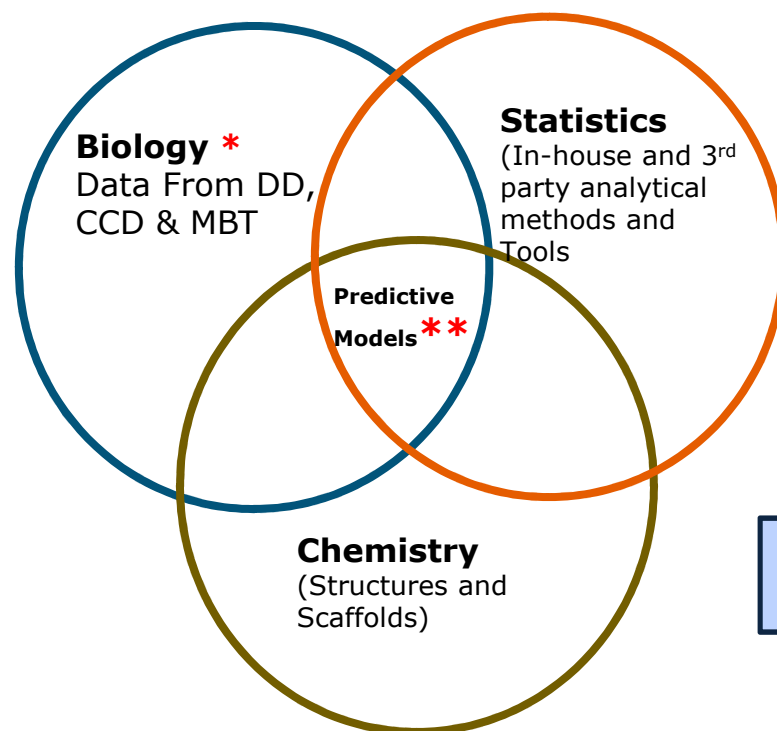


Utility of GVK BIO's
Clinical Candidate Database,
Drug Database, and
Mechanism Based Toxicity Database



Over View - Workflow



* Developed by GVK BIO
* * Few times GVK offered as service

➤ Software:

- **ADMET/ Efficacy Models:** Discovery Studio - C2 ADME & TOPKAT
- **ADMET/ Efficacy Models:** Schrodinger
- **Cheminformatics:** Pipeline-Pilot, Clustering Tools and Spotfire

➤ Customized *Insilico* ADME and Toxicity Models

- ✓ Aqueous Solubility and logP Models
- ✓ Blood Brain Barrier
- ✓ Intestinal Absorption
- ✓ Plasma Protein Binding
- ✓ Half life and Tmax
- ✓ Cmax
- ✓ log K_{hsa} (human serum albumin binding)
- ✓ Caco-2 and MDCK cell permeability
- ✓ Metabolism
- ✓ CYP2D6 Binding
- ✓ Hepatotoxicity
- ✓ hERG Model
- ✓ Carcinogenicity Model
- ✓ Mutagenicity Model
- ✓ LD50 Models
- ✓ Inhalational LC50
- ✓ Skin Sensitization

Services provided to Sponsor

Service 1: Generation of Half life Model

Objective: Developing of Half life Models using set of compounds from CCD and DD databases

Research Plan:

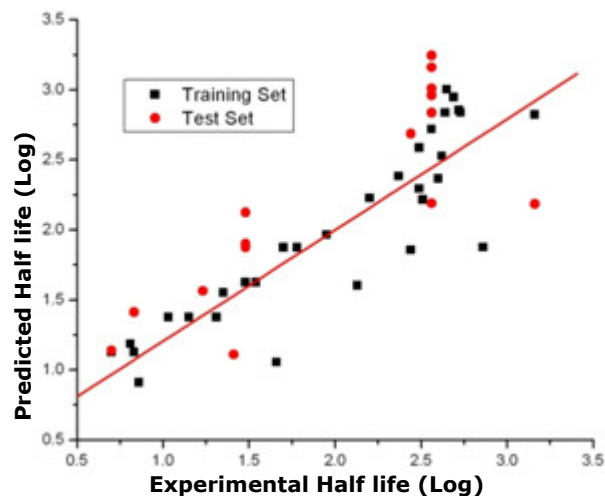
- **GVKBIO Scope**
 - Assemble known information from CCD and DD
 - Preparation of data sets (Training and Test Set)
 - Model generation
 - Model validation
 - Filter compounds using the best model
 - Compound prioritization
- **Sponsor Scope**
 - Assemble known information
 - Supplement with internally generated compounds with half life data
 - Supply new compounds to apply the models and prioritize

Available descriptors

- Set of 398 descriptors
 - 254 molecular descriptors created with DS software (belonging to classes: structural, physico-chemical, topological etc)
 - 144 physico-chemical descriptors created with Schrödinger
- 323 kept after pre-filtering: removed descriptors with constant value in more than 95% of all compounds

Generation and validation of Half Models

Selected best three models with 5 and 6 parameters based internal and external validation



$$\begin{aligned} \text{Log (Half life)} = & -2.0445 \\ & - 0.45877 * D_{88} \\ & + 0.43853 * D_{159} \\ & - 0.85509 * D_{166} \\ & + 0.2779 * D_{283} \\ & + 0.02375 * D_{307} \\ & + 0.01341 * D_{311} \end{aligned}$$

best 5 descriptor model

Model	Descriptors selected	R^2	Q	F	SE	Rank
A1	92, 159, 166, 307, 311	0.8942	0.8679	62.2419	0.2626	3
A2	88, 159, 166, 307, 311	0.8923	0.8742	60.9417	0.2648	5
A3	55, 159, 166, 307, 311	0.8916	0.8685	60.4704	0.2656	6
A4	88, 159, 166, 283, 307, 311	0.9128	0.8922	64.0905	0.2411	1
A5	108, 155, 175, 262, 263, 311	0.9042	0.8650	57.5442	0.2520	3
A6	79, 157, 163, 283, 307, 311	0.9017	0.8773	55.8025	0.2553	2

best 6 descriptor model

Conclusions

- GVK BIO developed two best models with 5 descriptor and 6 descriptor
- 6 descriptor half life model is the best model and used for compound filter and prioritization
- Sponsor supplied 250 compounds to predict their half life data
- 15/250 were suggested by *Insilico* half life model for experimental studies
- **5/15 showed better half life in experimental studies**

Service 2: Generation of Solubility Models

Objective: Improve the solubility of a Lead compound, Poor soluble compound, using *Insilico* Solubility Model

Research Plan:

- GVKBIO Scope
 - Assemble known information from MCD, CCD and DD
 - Preparation of data sets (Training and Test Set)
 - Model generation and validation
 - Designing Analogues of Lead compound
 - Predicting the solubility of new designs of Lead compound
 - Compound prioritization
- Sponsor Scope
 - Provided Lead compound

Dataset Collection

- 138 compounds [minimum of 3 aromatic Rings, Molecular Weight $\geq$ 300 Grams] with experimental solubility
- 119 compounds are considered as a training set to generate solubility model and 19 Drugs used as a internal set [Test Set-1] and 513 compounds as external set [Test Set-2]

We carefully checked the original literature sources to ensure that the data met the following criteria:

- (1) The compound must be drug or drug-like and in solid state at room temperature
- (2) The solubility value must be reported as the solubility of unionized species (intrinsic solubility) at or around 25°C.

Descriptors Selection:

- More than 200+ molecular descriptors were computed using the discovery studio & Schrödinger software's those includes descriptors of all types such as Topological, Geometrical, Charge, Thermodynamic descriptors and quantum-chemical descriptors related to solubility. Such as the number of H donors (D), the number of H acceptors (A), and the number of hetero-atoms (H) present in the molecular structure). Only 10% of the total descriptors for which the variance was the highest were considered for further analysis.

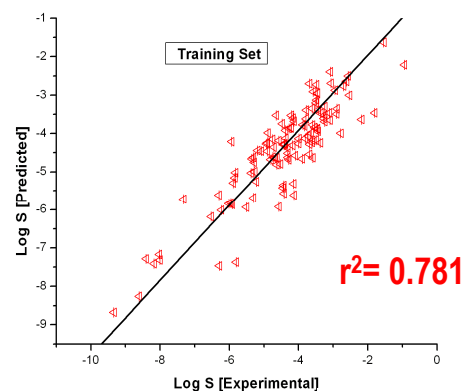
Model Description

Solubility= 53.739
 - 1.1603e-005 * AM1_Eele
 - 5.3506 * GCUT_SLOGP_2
 - 0.96899 * Kier1
 + 10.703 * PEOE_RPC+
 + 0.02893 * PEOE_VSA-5
 - 12.427 * VAdjMa
 + 0.70355 * a_acc
 + 0.19584 * a_aro
 + 0.70025 * a_heavy

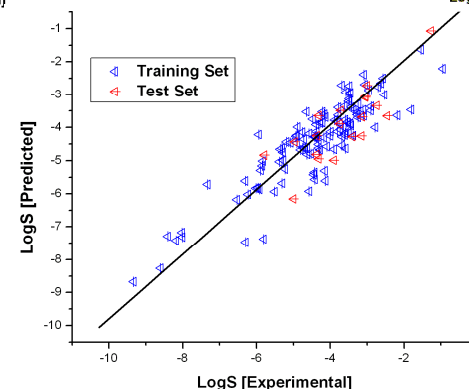
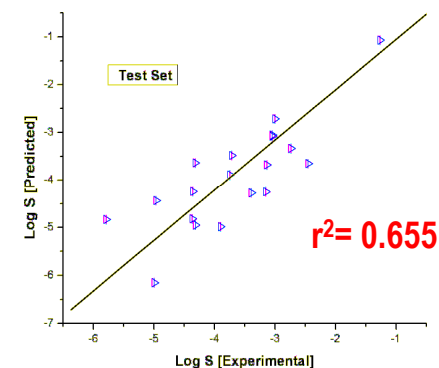
Description:

AM1_Eele :	Electronic energy (kcal/mol)
GCUT_SLOGP_2 :	LogP GCUT (1/3)
Kier1 :	First kappa shape index
PEOE_RPC+ :	Relative positive partial charge
PEOE_VSA-5 :	Total negative 5 vdw surface area
VAdjMa :	Vertex adjacency information (mag)
a_acc :	Number of H-bond acceptor atoms
a_aro :	Number of aromatic atoms
a_heavy :	Number of heavy atoms

Training Set



Test Set-1



Component	
r^2_{cv}	0.781
r^2_{adj}	0.762
$r^2_{(pred)}$	0.852
RMS Residual Error	0.690
No of observations	120
Friedman L.O.F	1.99

Conclusion

- GVK BIO developed and validated the Solubility Model
- Analogues (150) of Lead compound (Sponsor supplied Lead compound) were generated
- Insilico Solubility Model predicted 25/150 compounds, having good solubility than Lead compound
- **12/25 showed better solubility than the lead in the experimental studies**

Publications by Clients

Publication 1

- **Reference:**

- Drug Discov Today. 2011, 16(15-16) 646-653

- **Title:**

- Molecular clinical safety intelligence: a system for bridging clinically focused safety knowledge to early-stage drug discovery – the GSK experience

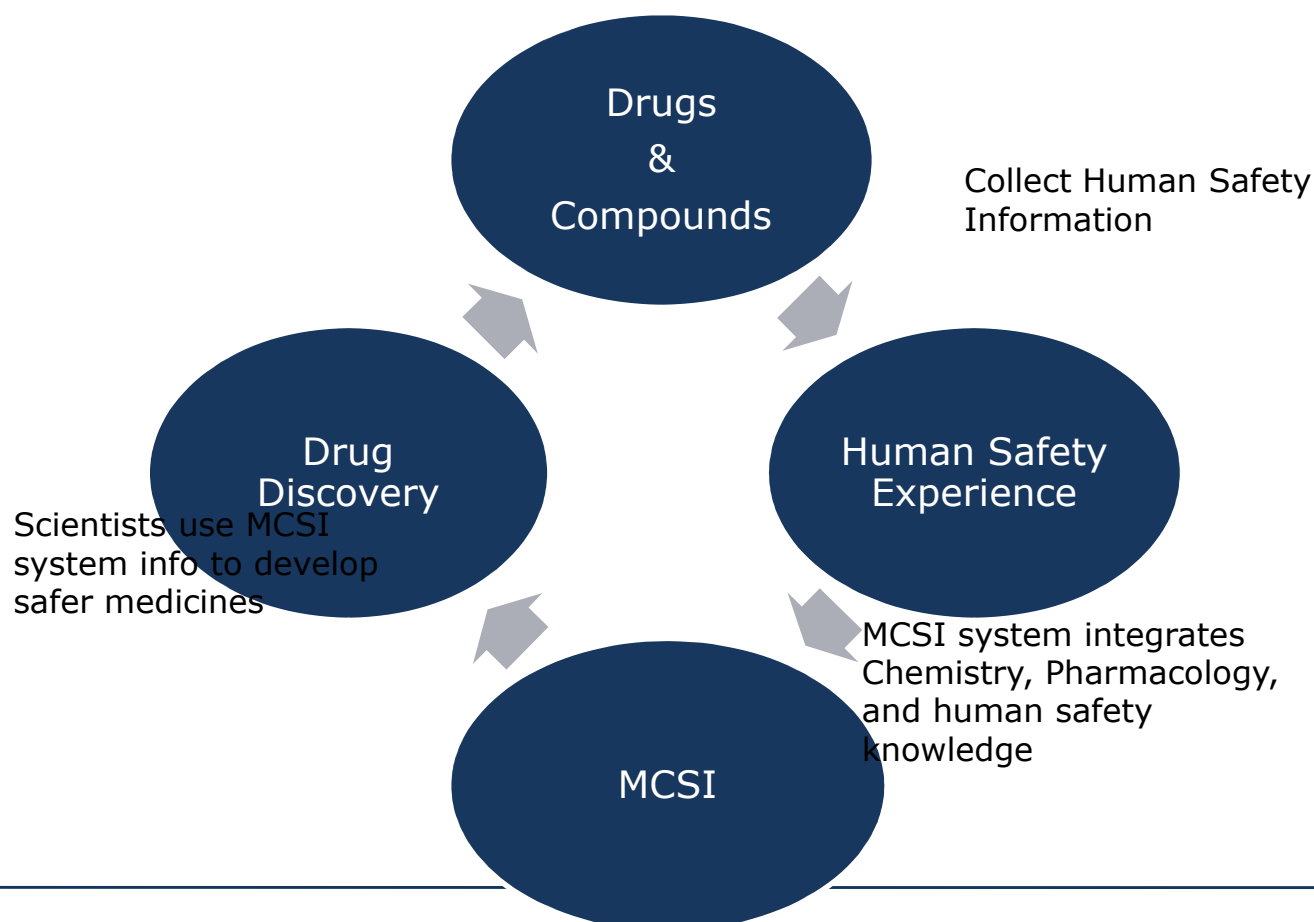
- **Authors:**

- Dana E. Vanderwall, Nancy Yuen, Mohammad Al-Ansari, James Bailey, David Fram, Darren V.S. Green, Stephen Pickett, Giovanni Vitulli, Juan I. Luengo and June S. Almenoff

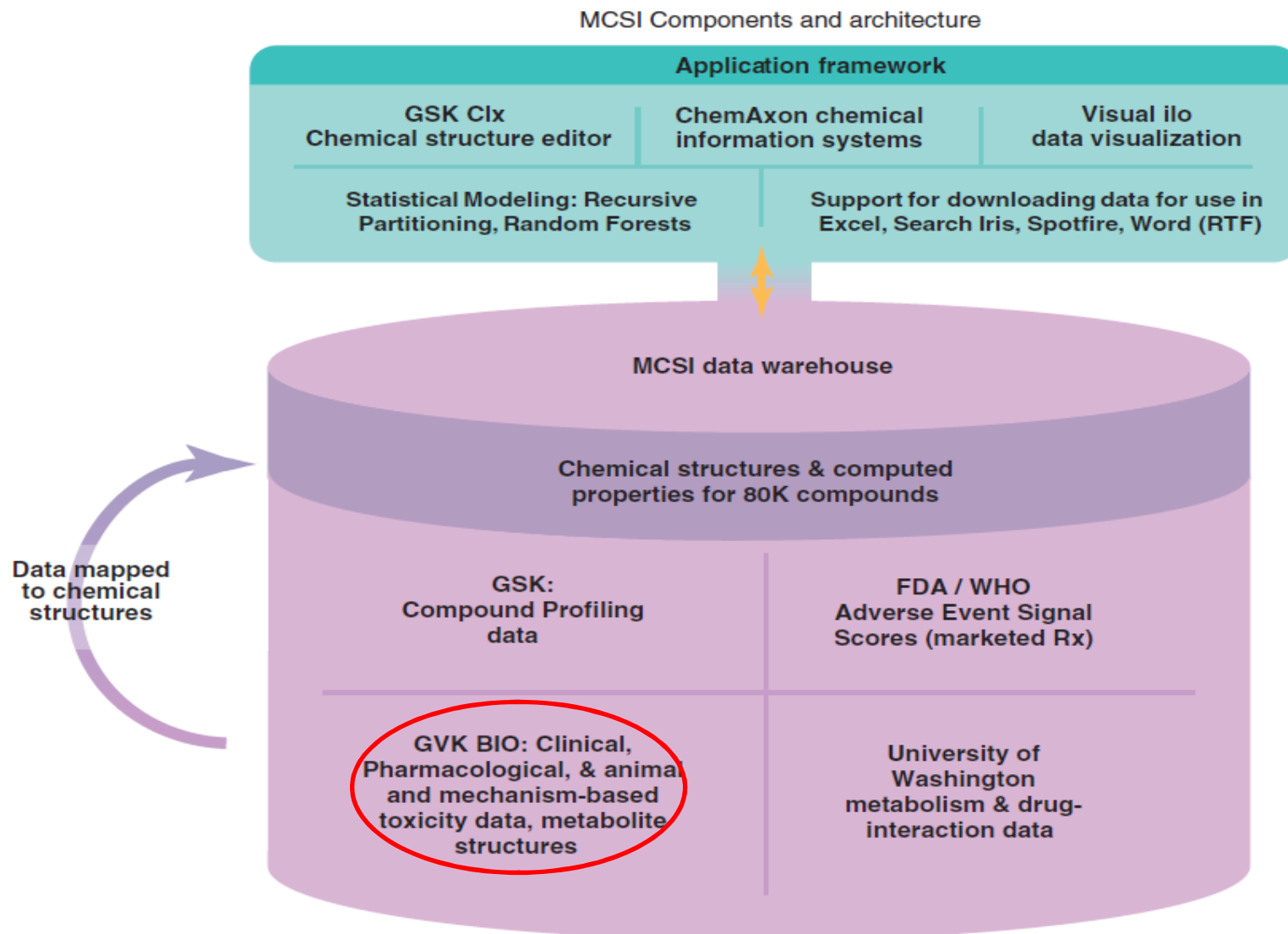
- **Company:**

- GlaxoSmithKline
- Lincoln Safety Group, Phase Forward

- The **Molecular Clinical Safety Intelligence** (MCSI) system first-in-class in-house prototype system combines preclinical, chemistry, toxicology, metabolism, pharmacology and clinical safety information, to improve the detection and management of clinical safety risk in early drug discovery



MCSI Architecture

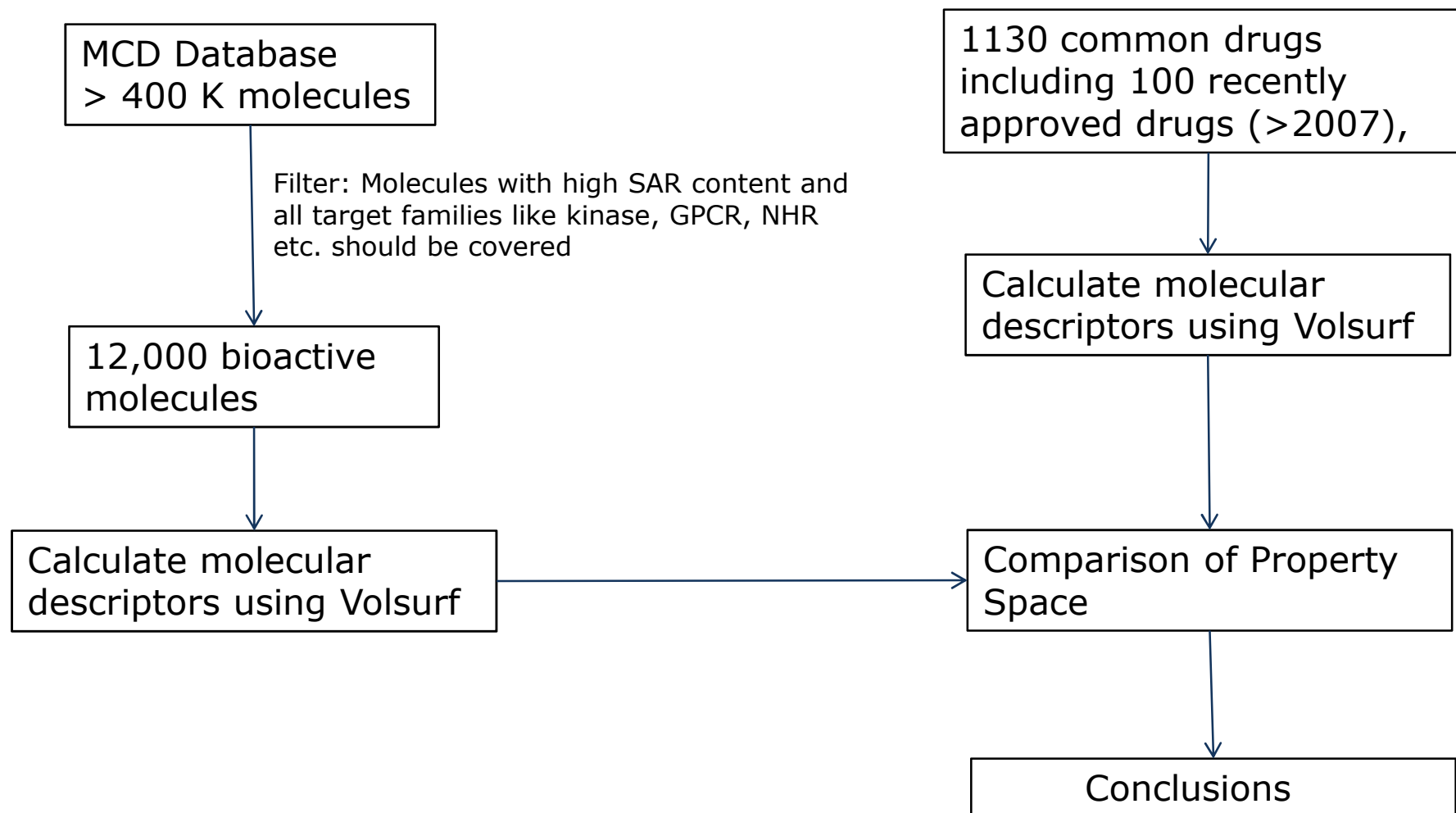


- Query and analysis results are presented as interactive, scrollable tables and graphs that offer point-and-click drilldown to full compound details. After a list of compounds related to the novel compound has been generated the following questions can be evaluated:
 - Are there compounds with similar structures that have associated safety concerns?
 - Which particular structural features appear most associated with a particular adverse event and/or toxicity?
 - Is the biological target for a novel compound associated with adverse events?
 - Is there evidence that the novel compound could be metabolized to toxic intermediates?
 - Are there options to potentially block this metabolism?
 - Is the novel compound likely to have unanticipated biological effects (i.e. off-target activity)?
- Answers for all the points highlighted in blue are predicted using the GVKBIO DD, CCD and MBT Databases
- For Structural similarity searching, GSK is using ChemAxon software

Publication 2

- **Reference:**
 - Drug Discovery Today, 2011, 16 (21-22), 976-984
- **Title:**
 - Evolution of the physicochemical properties of marketed drugs: can history foretell the future?
- **Authors:**
 - Bernard Faller, Giorgio Ottaviani, Peter Ertl, Giuliano Berellini and Alan Collis
- **Company:**
 - Novartis Institutes for BioMedical Research, Basel, Switzerland
 - Novartis Institutes for BioMedical Research, Cambridge, MA, USA

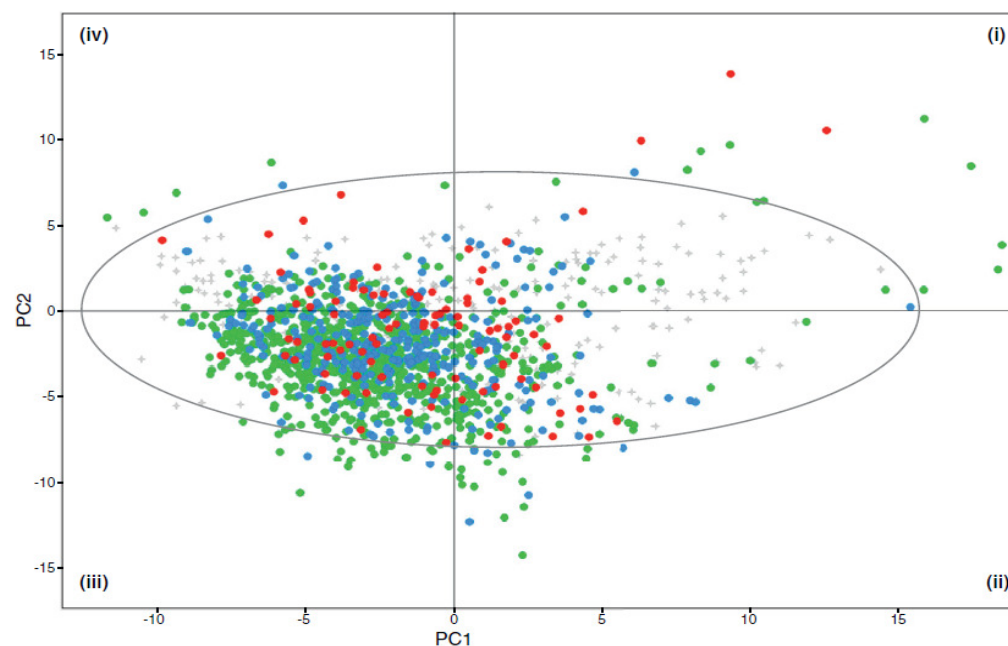
Work Flow



Comparison of Property Space

- Many Greens in left lower quadrant indicates small molecules with localised hydrophilicity and relatively few H-bond donors and/or acceptors (Traditional Drug Space)
- Molecules in the upper left quadrant are larger and more lipophilic
- The upper right quadrant contains the lowest number of drug molecules and is occupied by large hydrophilic compounds
- Lower right quadrant are hydrophilic with localised hydrophobicity and a higher number of Hbond donors and/or acceptors.

Open Stars – Bioactive molecules (MCD Set)
Green – drugs approved before 1983
Blue – drugs approved between 1983 and 2002
Red – drugs approved after 2002



Drug Discovery Today

- Drugs occupy only a fraction of the property space of the bioactives and are unevenly distributed in this physicochemical space.
- When comparison made **between NMEs launched since 2002 and traditional drugs (between 1983 and 2002)** shows a clear shift away from the traditional drug space.
- On an average, **newly launched molecules are larger molecules** but not necessarily more lipophilic than older drugs.
- A clear change in the property space of recent chemical programs compared to traditional drug space like
 - PSA and c Log P are no longer able to segregate orally available molecules and c Log P no longer correlates with brain free fraction.
 - MW increases from below 400 to above 500
 - The fraction of permeable compounds associated with a measurable free fraction in plasma drops from 80% to <10%
- **Can history foretell future? the answer seems negative, if property space of new chemistry programs moving away from the traditional drug space.**



Accelerating Research

▶ Thank You