Applications of QSAR in drug discovery

Jonna Stålring

Computational Toxicology

Safety Assessment

AstraZeneca R&D

AstraZeneca

The Pharmaceutical Discovery Process

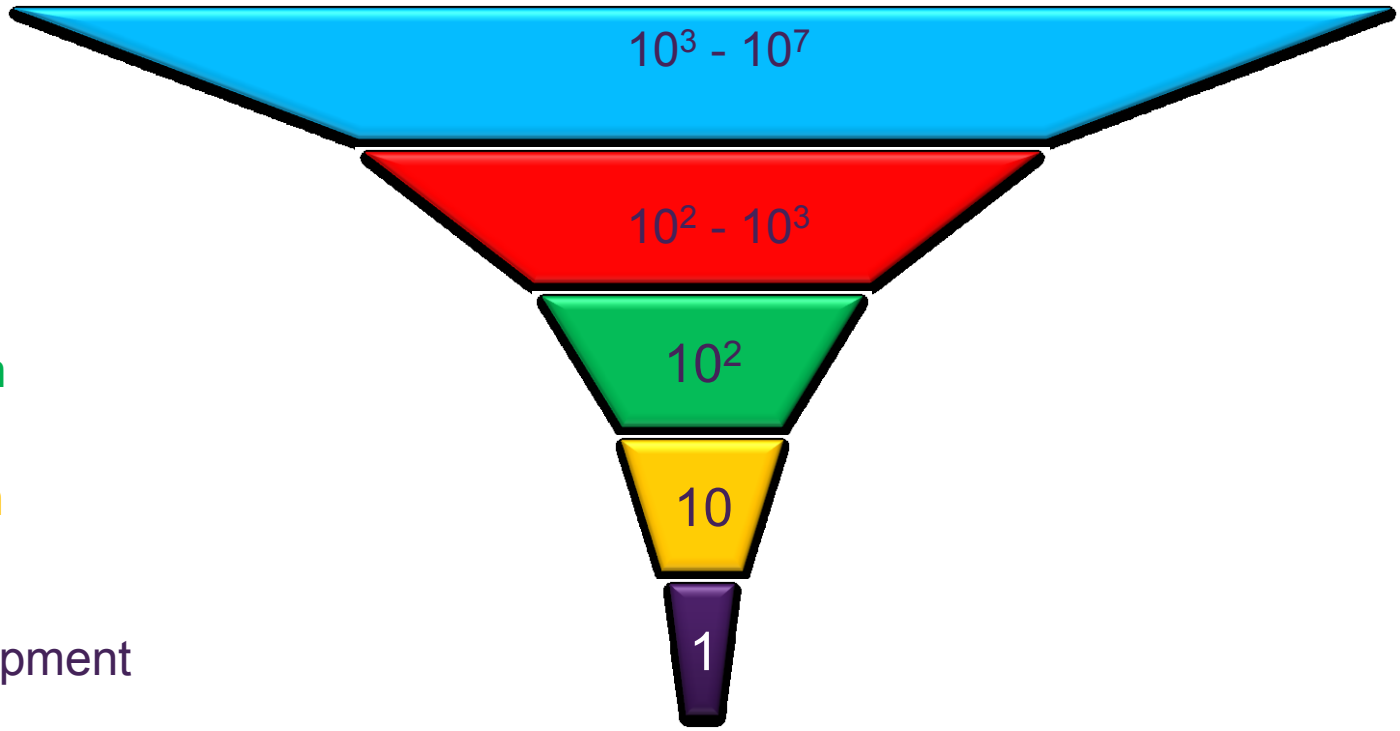
HTS
Virtual Screening
Library Design

Hit Identification

Lead Identification

Lead Optimization

Preclinical Development



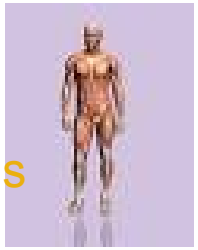
AstraZeneca

Experimental Hierarchy

Method

Experimental Requirements

Clinical Trials

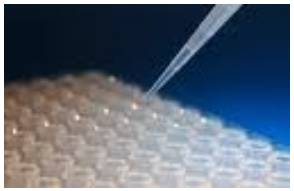


Humans

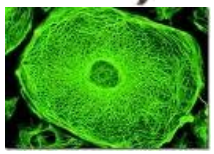
In Vivo

Animals

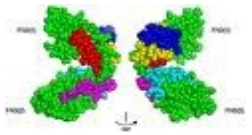
In Vitro



Cells



Biomolecules



Molecular Structure

In Silico

AstraZeneca

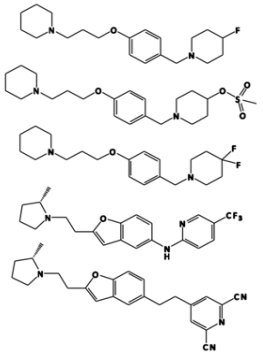
Frontloading

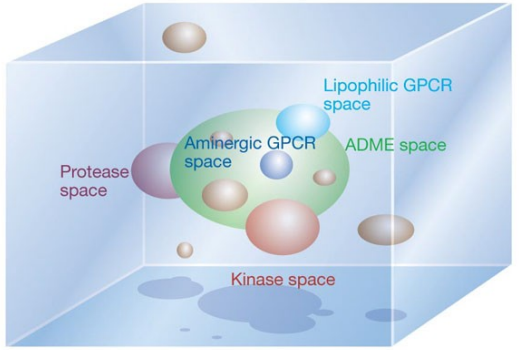
Mathematical Modeling

QSAR

AstraZeneca

Descriptive versus Predictive QSAR



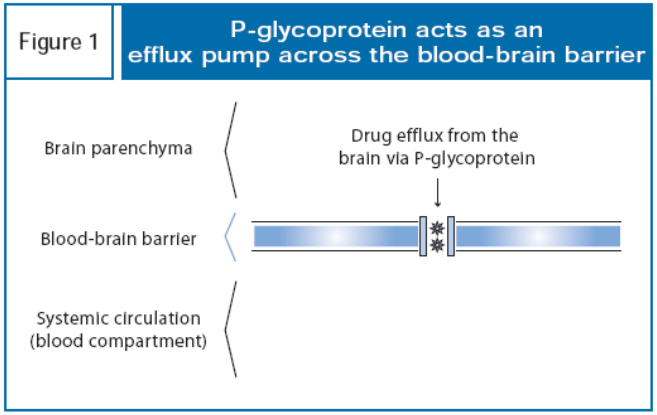


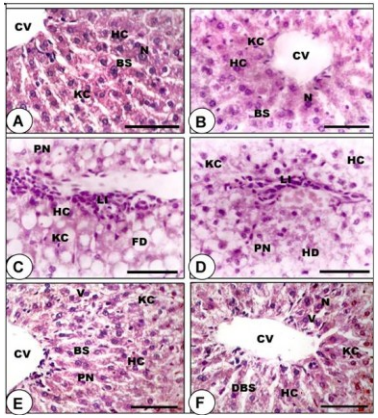
Descriptive QSAR/ Local QSAR/ SAR	Predictive QSAR/ Global QSAR/ QSAR
Predicting congeneric chemical series	Predicting greatest possible part of chemical space
Impact on activity of small structural change	Which compounds are most likely to be active
Few compounds	As many compounds as possible
Free Wilson Analysis, PLS	Advanced statistical methods
Late discovery	Early and late discovery

AstraZeneca

Predictive QSAR within Pharma

- Physiochemical properties (QSPR)
  - pKa
  - Solubility
  - logP
- DMPK
  - Permeability
  - Plasma protein binding
- Safety
  - Liver toxicity
  - Carcinogenicity
  - Seizure
  - Off target activities



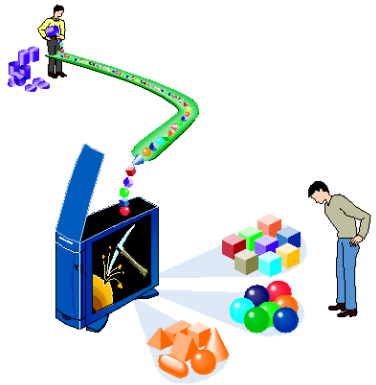


AstraZeneca

QSAR for Safety Risk Assessment

REMOVAL

Library Design

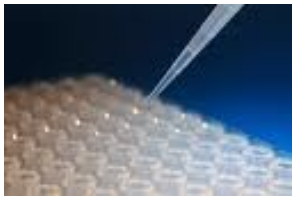


Positive prediction ⟶ Remove compound
Negative prediction ⟶ Keep in library

Few false positives

REMOVAL EXP PRIO

Early Discovery



Positive prediction ⟶ Remove compound/ Early in vitro testing
Negative prediction ⟶ Keep compound

Few false negatives

EXP PRIO

Late Discovery

Positive prediction ⟶ Early in vivo testing
Negative prediction ⟶ Postponed in vivo test

Few false negatives

AstraZeneca

hERG Classifier – Pfizer 2005[a]

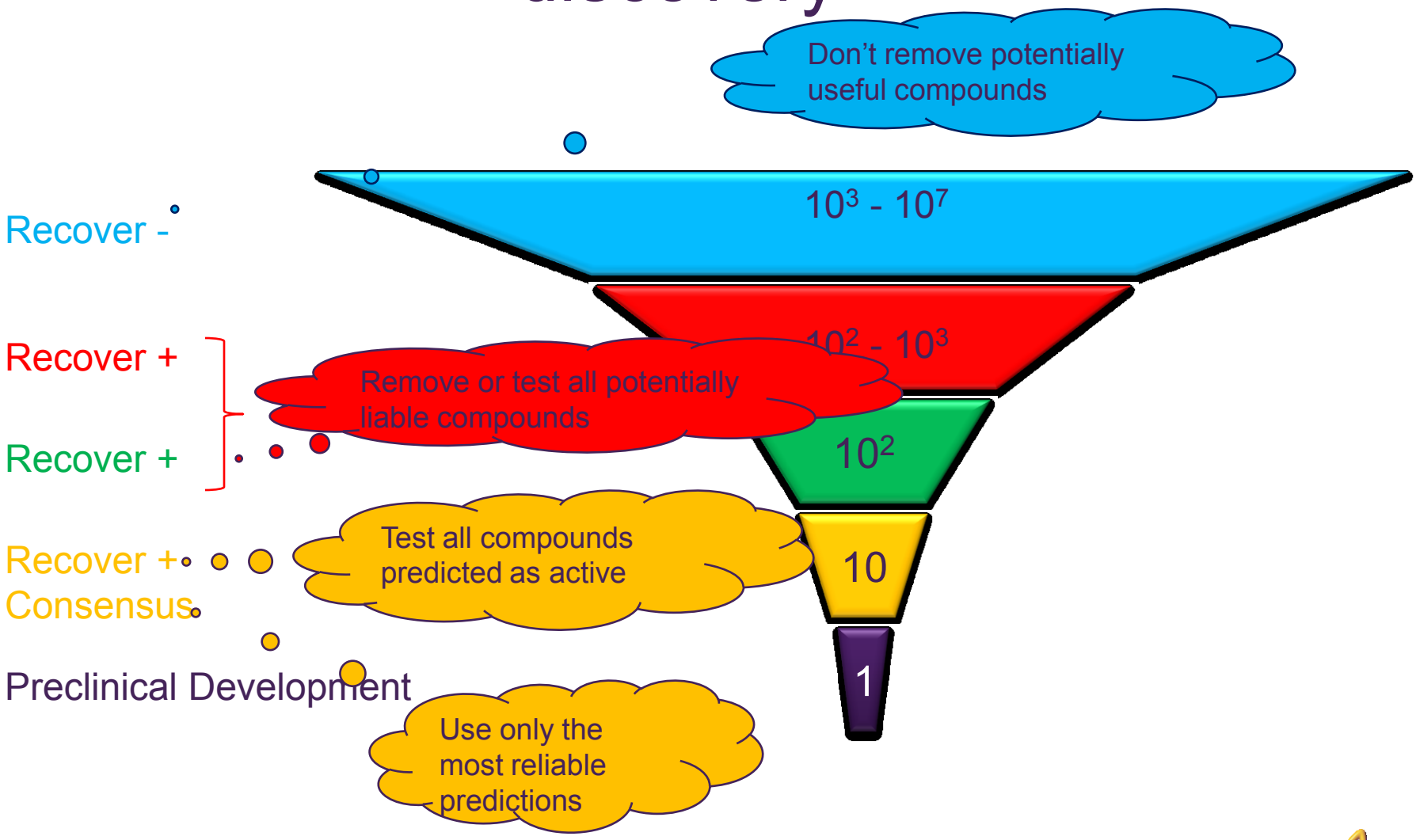
- 60 000 in house compounds
- 12 000 selected as external test
- Balanced binary classifiers

Classifier	Concordance	Sensitivity	Specificity
ANN	85	86	83
Naïve Bayesian	82	84	80
Recover +	82	92	75
Recover -	85	79	89
Consensus (86% predicted)	89	91	87

a) O'Brien *et al*, J. Med. Chem., 2005, 48

AstraZeneca

Applications within Pharmaceutical discovery



AstraZeneca

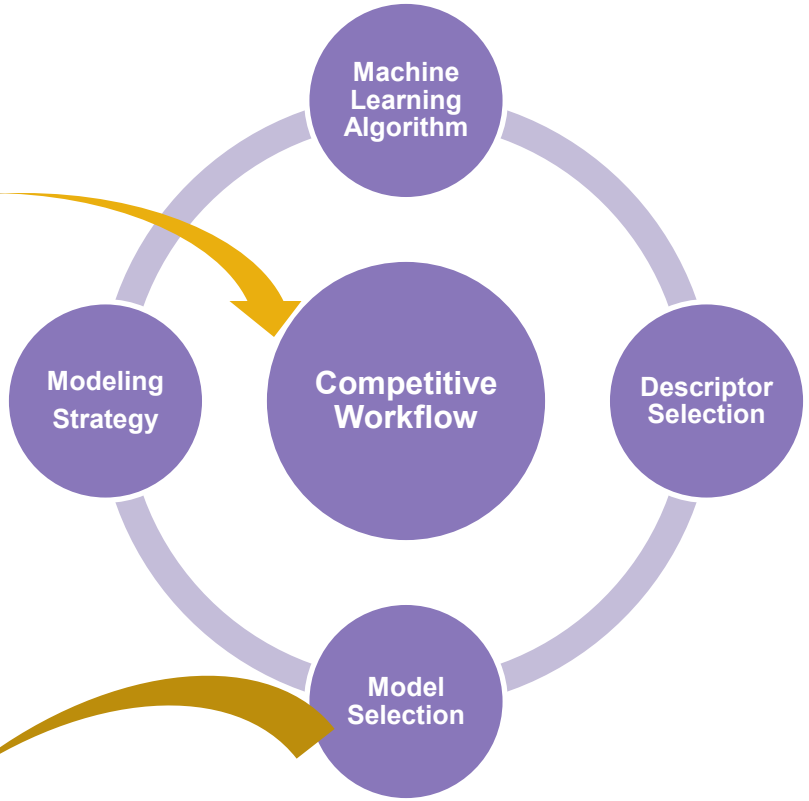
Process for global QSAR model development and maintenance

High throughput screens

Global databases

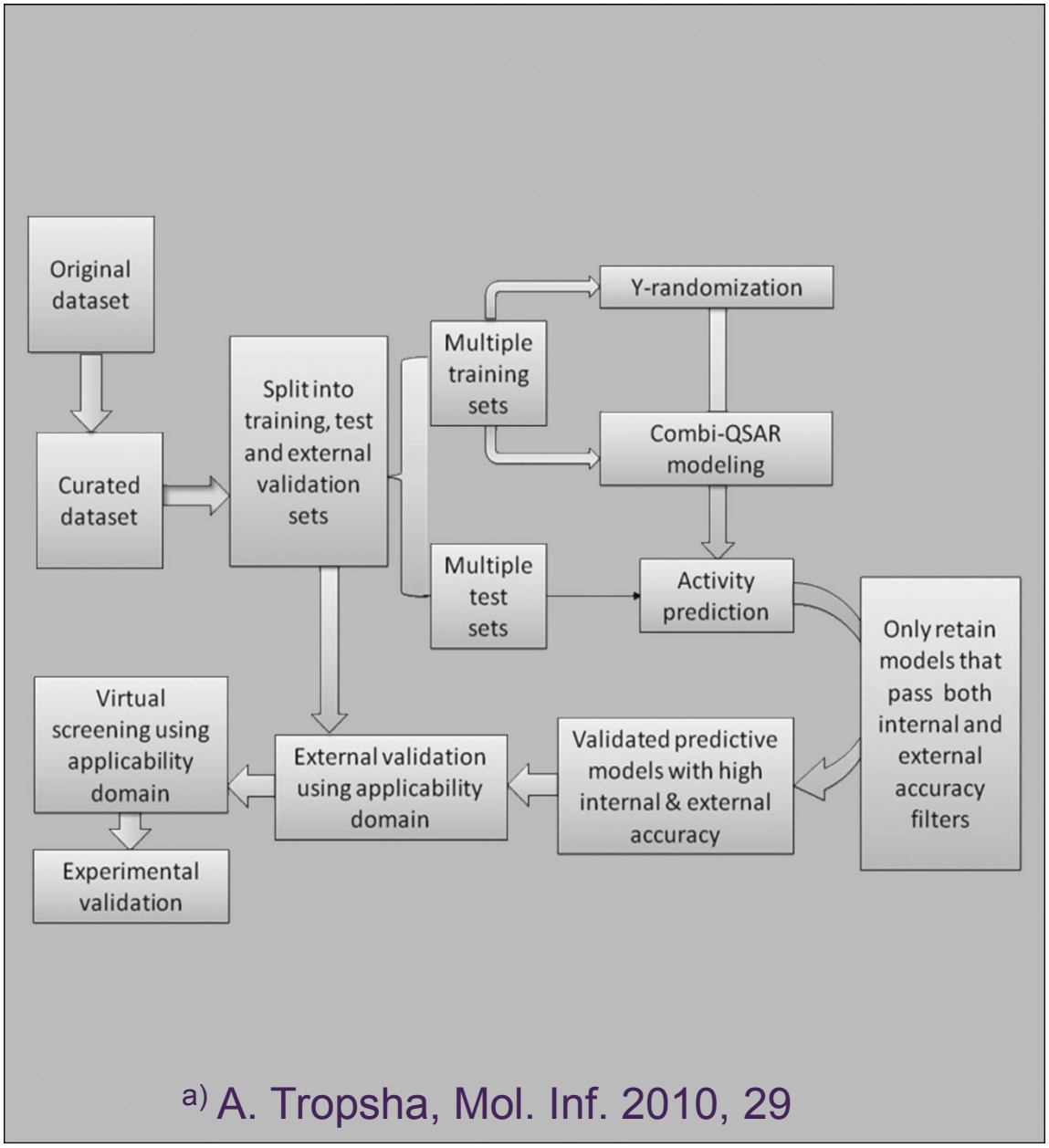
AZorange

Statistical Modeling



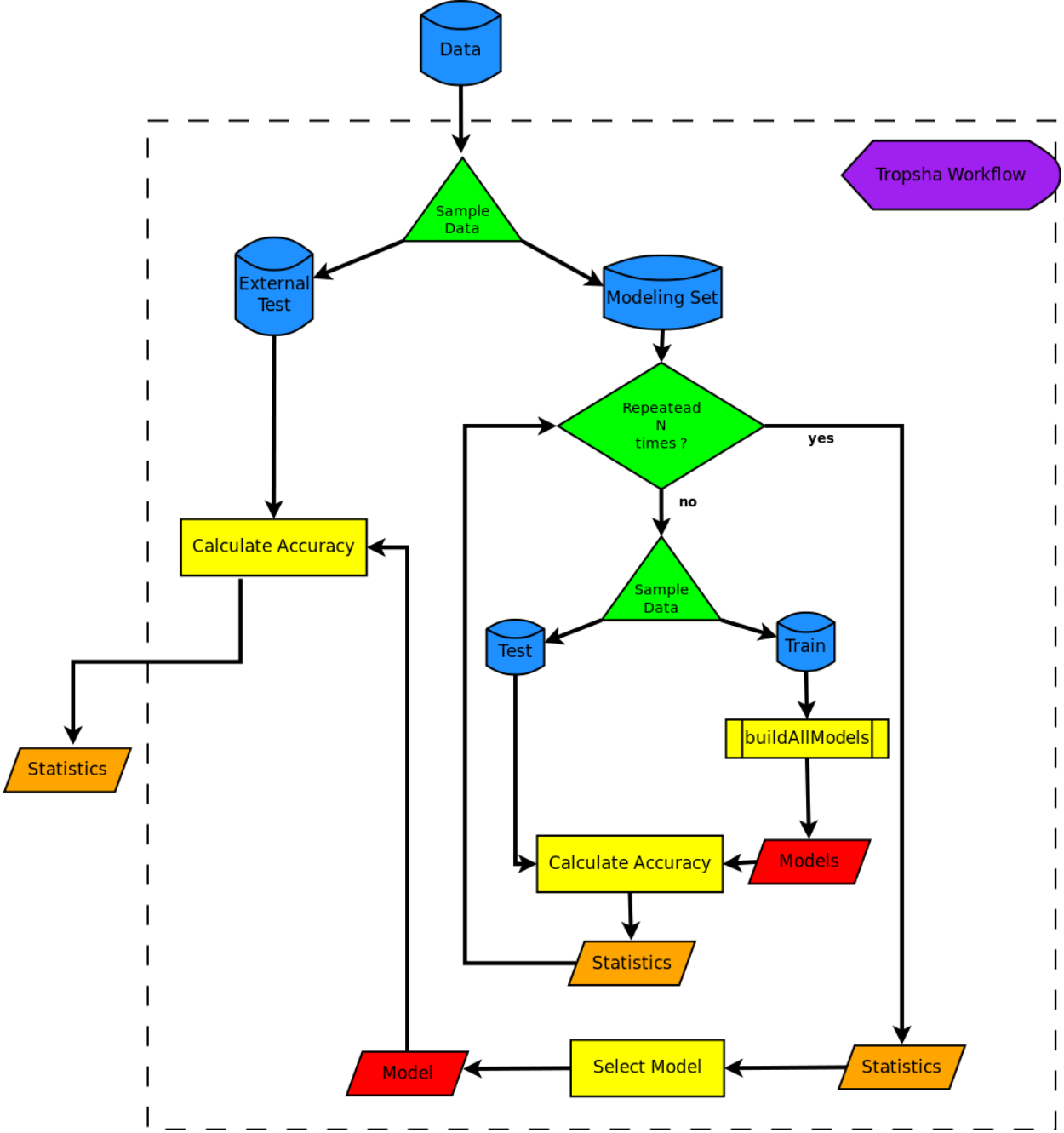
AstraZeneca

Best Practices for QSAR Model Development, Validation and Exploitation
Review by Alexander Tropsha 2010 [a]

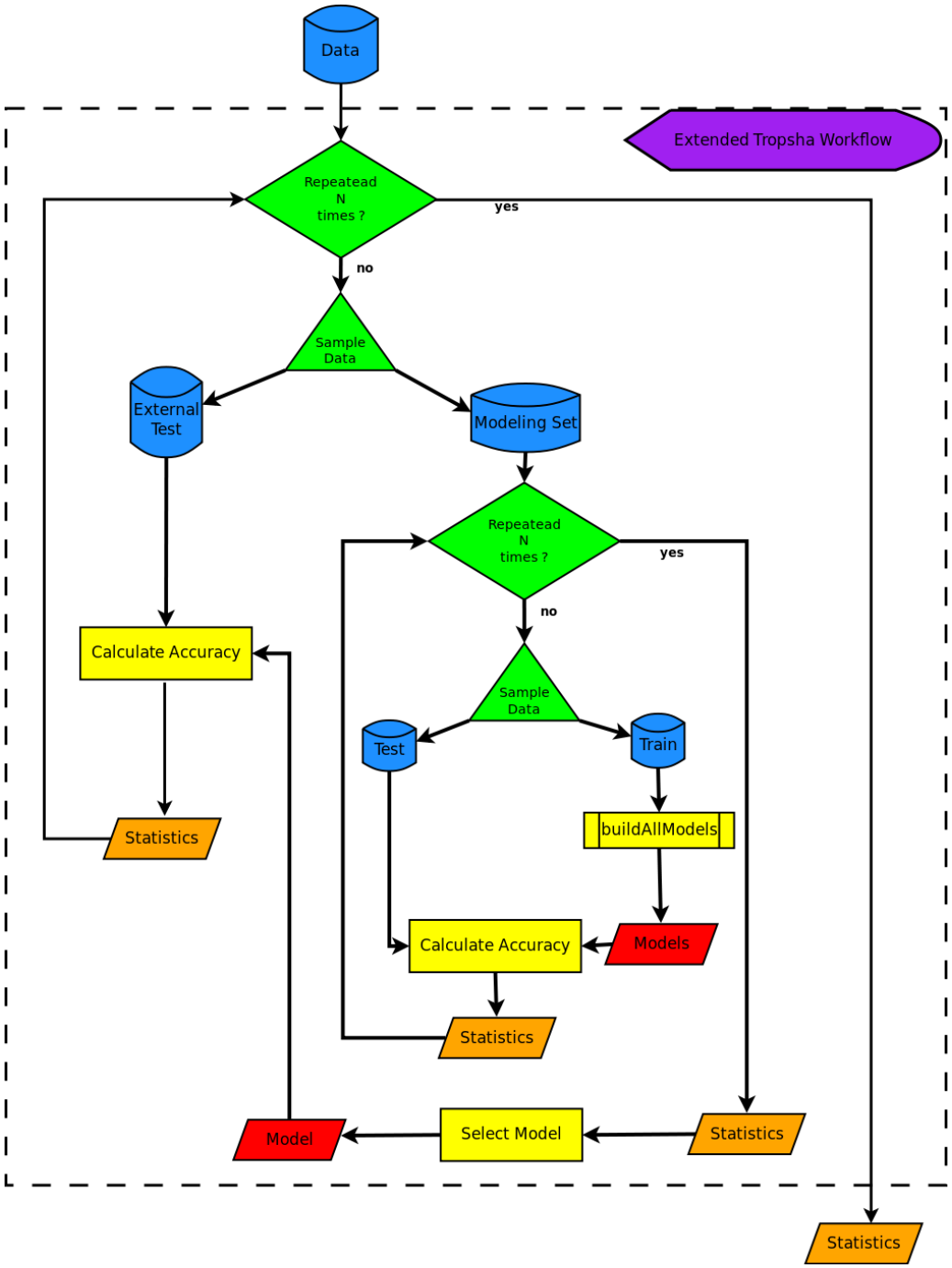


[a)] A. Tropsha, Mol. Inf. 2010, 29

AstraZeneca



AstraZeneca



AstraZeneca

Automated Updating and Validation of global QSARs

Global databases

Internal and Temporal Validation

Internal and Temporal Validation

Internal and Temporal Validation

Original Model

Updated Model

Updated Model

Time

Modeler

Acceptance

AZorange

Prediction

Compound

AstraZeneca

Public Safety Models for Drug Discovery

Public

QSAR TOOLBOX
The OECD QSAR Toolbox for Grouping Chemicals into Categories

OpenQSAR

SEVENTH FRAMEWORK PROGRAMME

OpenTox

eTOX

CADASTER

- Method Development
- Small Public Data Sets

Pharmaceutical

- Standardized experiments
- Large Proprietary Data Sets
- Established Methods

Collaboration

AstraZeneca

Applications of QSAR in drug discovery - Summary

- Drive towards frontloading in pharma
- In silico and in vitro replace in vivo
- Global QSAR mainly in ADMET modeling
- Competitive workflow
- Automatic updating and temporal validation
- Public safety models and methods

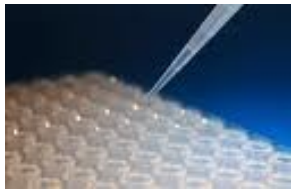
AstraZeneca

Extra Slides

AstraZeneca

Experimental Hierarchy

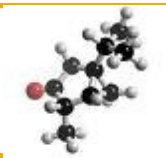
Method



Experimental Requirements

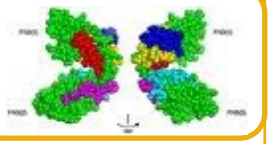
In silico

- Structure



Molecular In vitro

- Biomolecules



Functional In vitro

- Cells

In vivo

- Animals

Clinical

- Humans

AstraZeneca

Multi-mechanistic responses

- Examples – Genotoxicity
- Ames assay, strains
- Local/global modeling
- MTC

AstraZeneca

General Safety Strategy

Computational

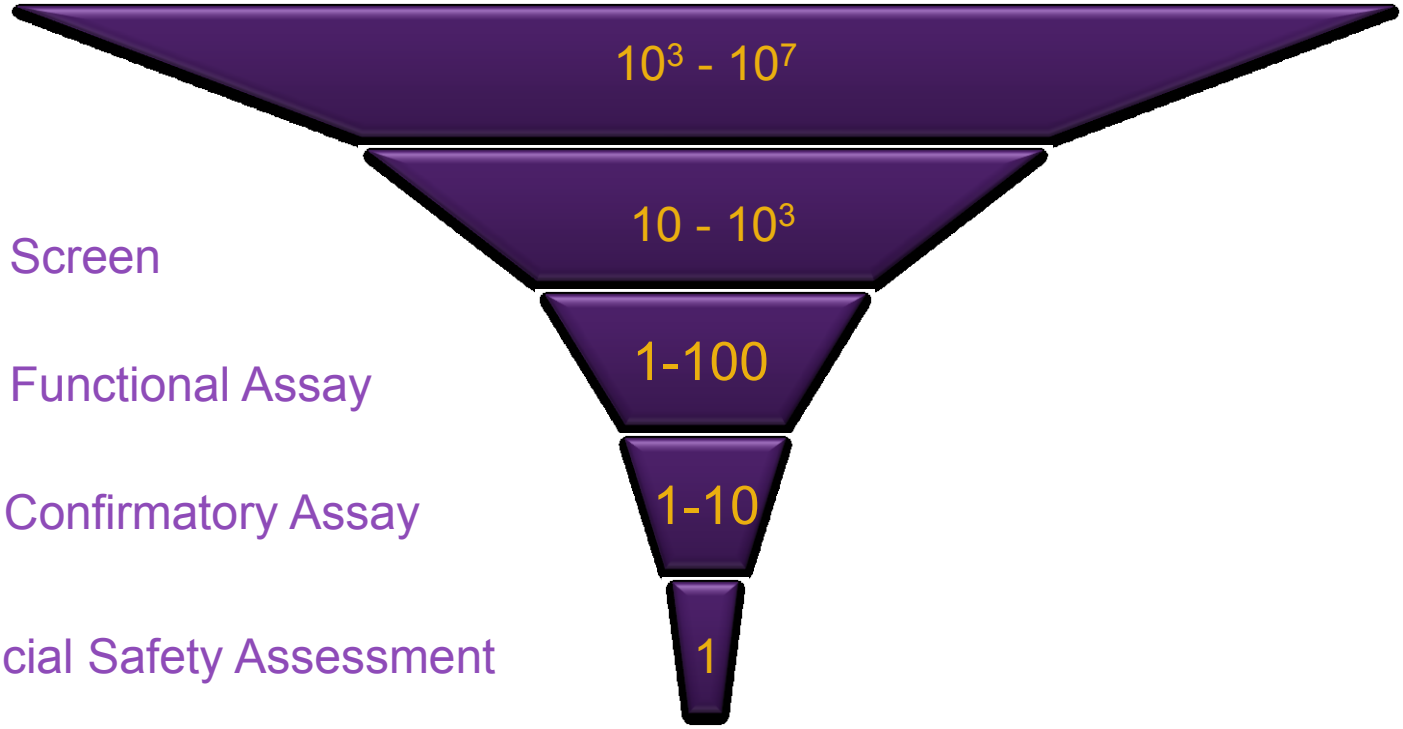
COMPUTATIONAL

In Vitro Screen

In Vitro Functional Assay

In Vivo Confirmatory Assay

Preclinicial Safety Assessment



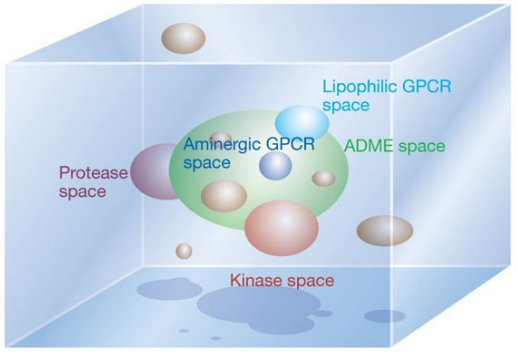
AstraZeneca

Challenges in global toxicological QSAR modeling in pharma

- Diversity of chemical space
- Multi-mechanistic biological processes
- Lack of structural information

AstraZeneca

Diversity of Chemical Space



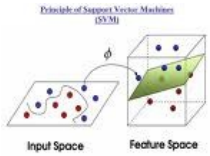
- Merging data from various sources

NIH Molecular Libraries

SAR

Global Classification Data

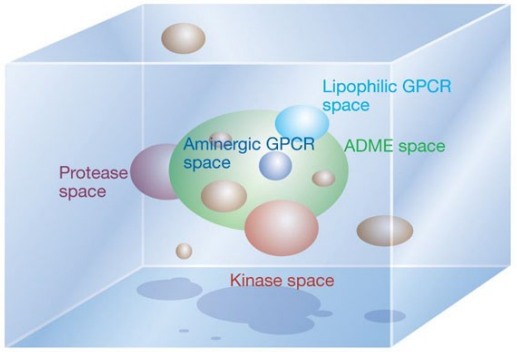
- Automated updating and validation



Predictions & Validation

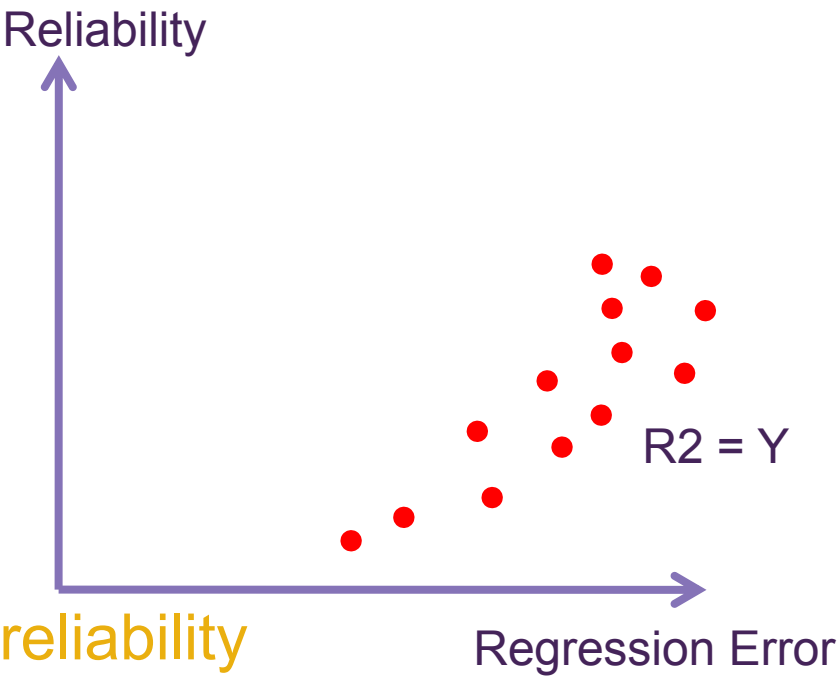
AstraZeneca

Diversity of Chemical Space



- Developing the concept of the applicability domain

"The applicability domain of a QSAR model is the response and chemical structure space in which the model makes predictions with a given reliability"



- Half of the data sets show a positive correlation between reliability metrics and prediction error [a]

[a] Bosnic, The Knowledge Engineering Review (2010), 25: 27-47

AstraZeneca

Multimechanistic Endpoints

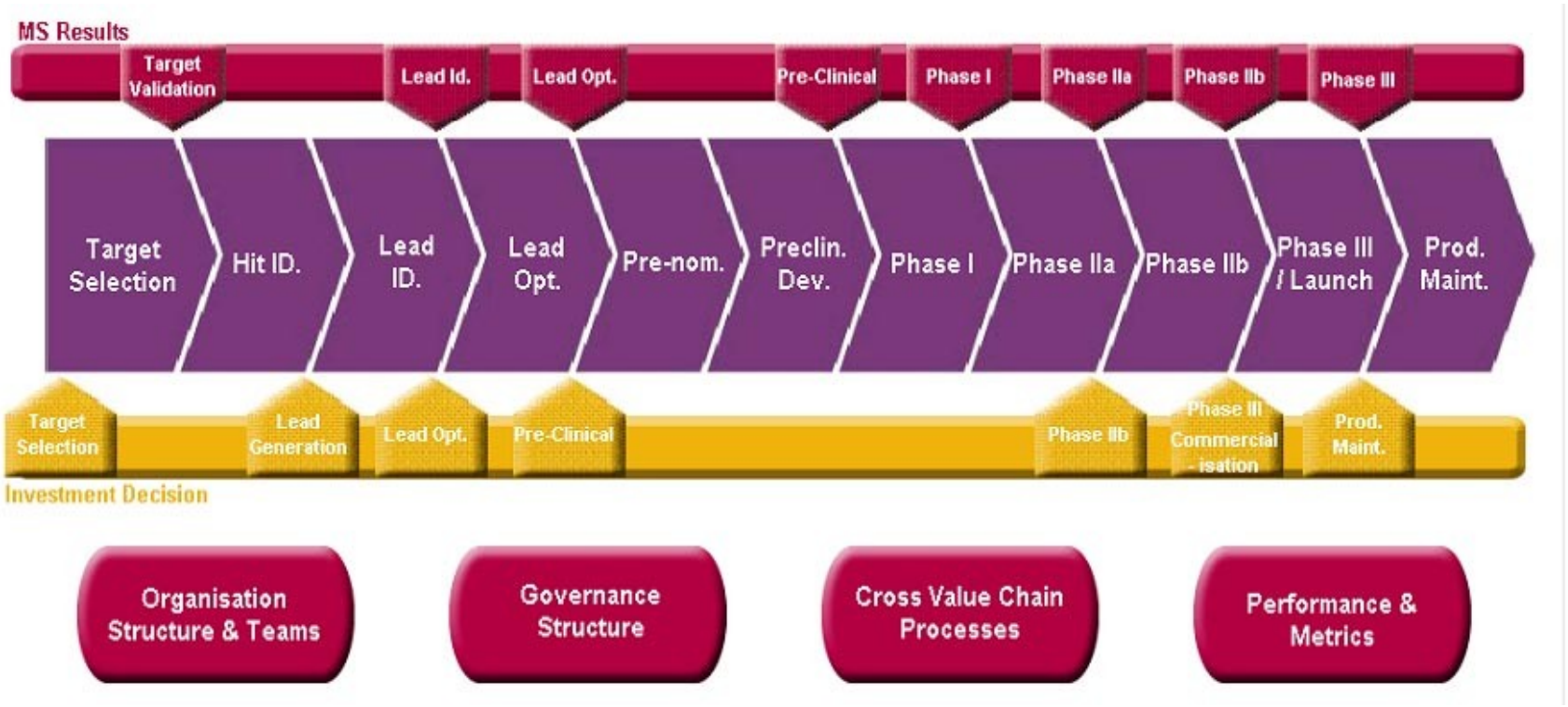
- Local/Global Modeling
- Structural class similarity as descriptor
- MTC

AstraZeneca

PLD example

AstraZeneca

Pharmaceutical discovery and development process



AstraZeneca

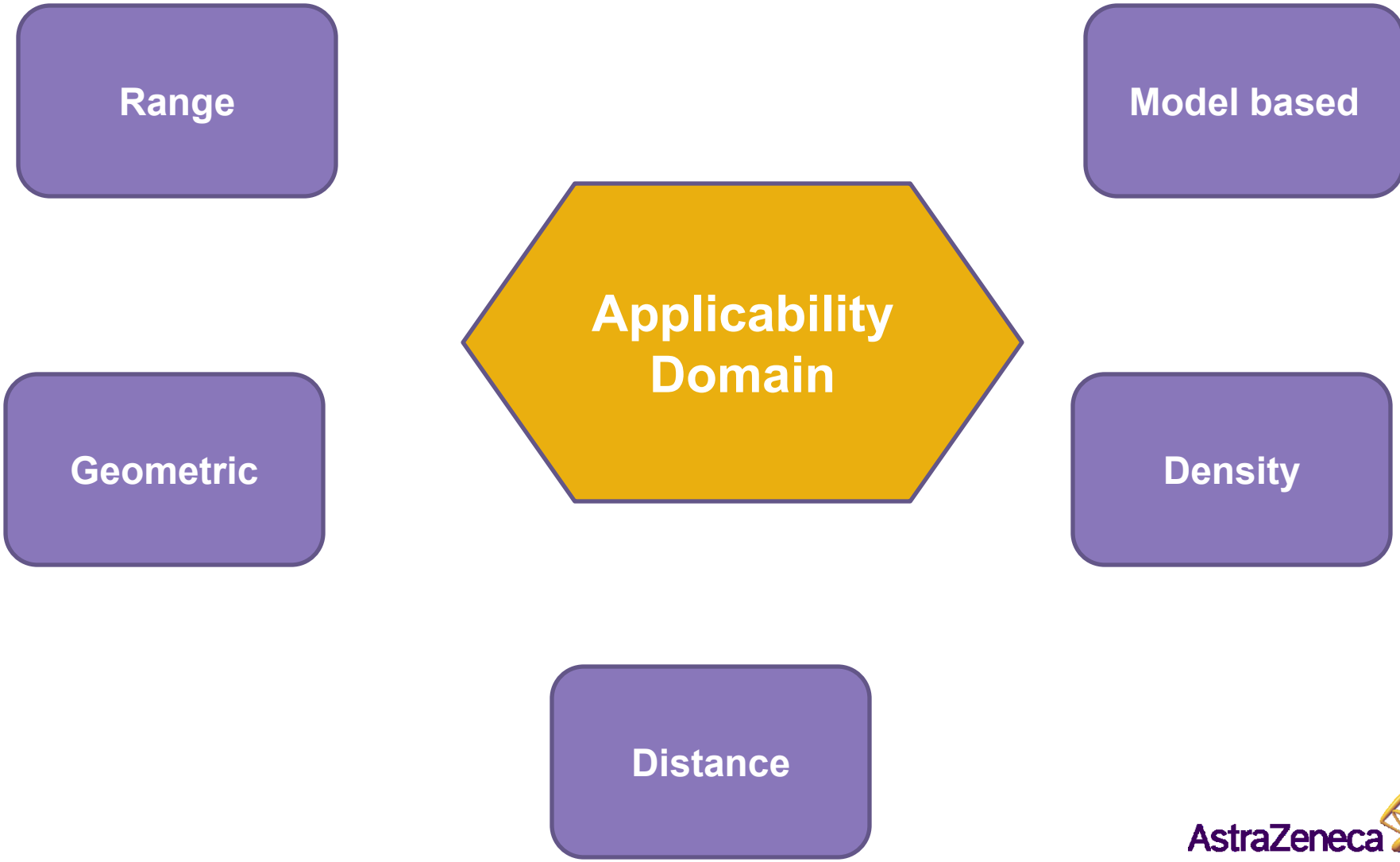
The Applicability Domain

- The third principal of “The OECD principals for QSAR validation”[1] (2004) requires “a defined domain of applicability”
- The ECVAM Workshop [2] in 2005 agreed on the definition:

“The applicability domain of a QSAR model is the response and chemical structure space in which the model makes predictions with a given reliability”

AstraZeneca

Applicability Domain Methods in QSAR literature



AstraZeneca

Distance – “Distance to Model” [3, 5]

- Distance between the test compound and the train set
- Distances;
  - Distance to mean
  - Average distance to all
  - Average distance to near neighbors
  - Count of neighbors
- Molecular representation;
  - Chemical similarity
  - Descriptor space
  - Property space (includes response)
- Metrics; Euclidean, Mahalanobis, City block

AstraZeneca

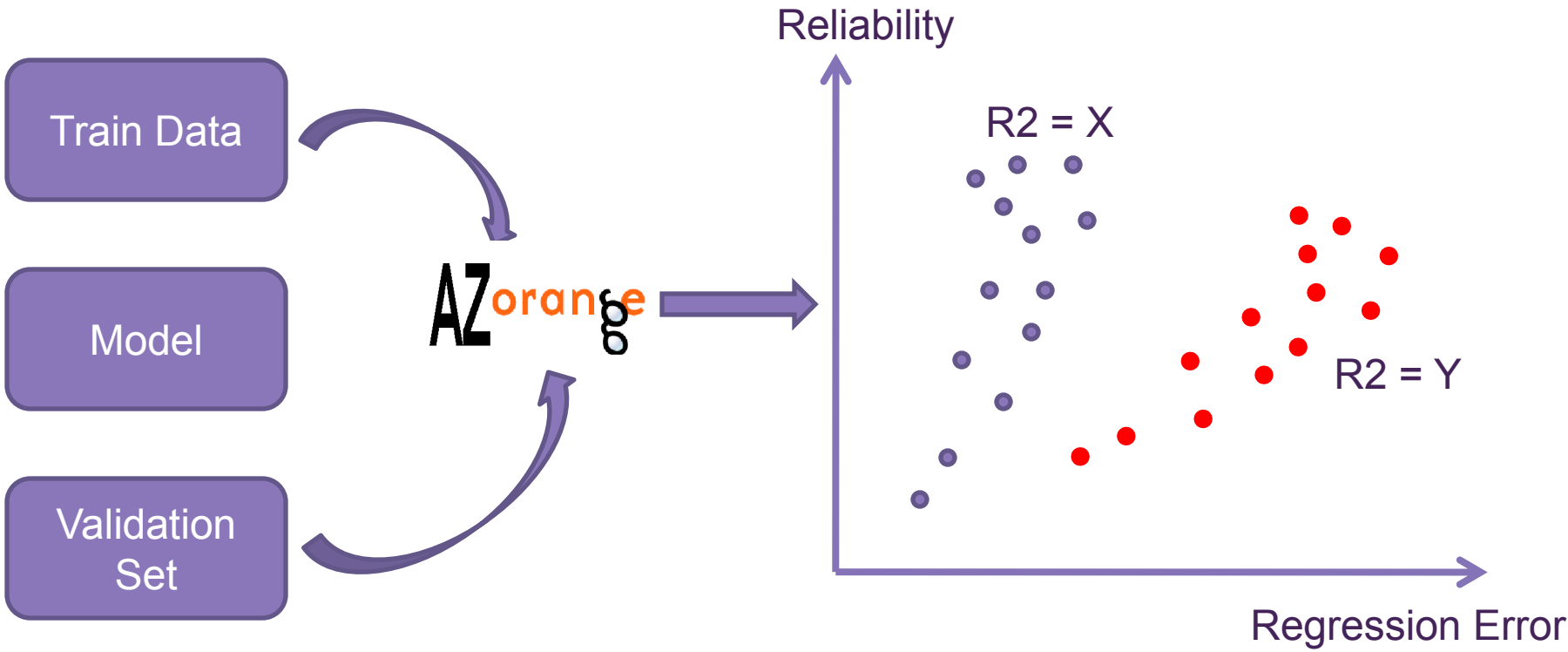
Model Independent Reliability – Proposed by Orange group

- Sensitivity Analysis - 45% positive correlation with prediction error
- Bagging Variance – 53%
- Local Cross-Validation – 44%
- Density Estimation – 34%
- Local Error Modeling – 48%

- Meta classifier selecting reliability method – 55%
- Cross validation selection – 78%

AstraZeneca

AZOrange implementation



AstraZeneca