Pharmacophores

Next Generation Pharmacophore Modeling: Concepts and Applications

universität wien

Thierry Langer

inte:ligand
www.inteligand.com Vienna, Austria

Acknowledgements

Acknowledgements

universität wien

Christian Doppler Forschungsgesellschaft

imi innovative medicines initiative

Boehringer Ingelheim

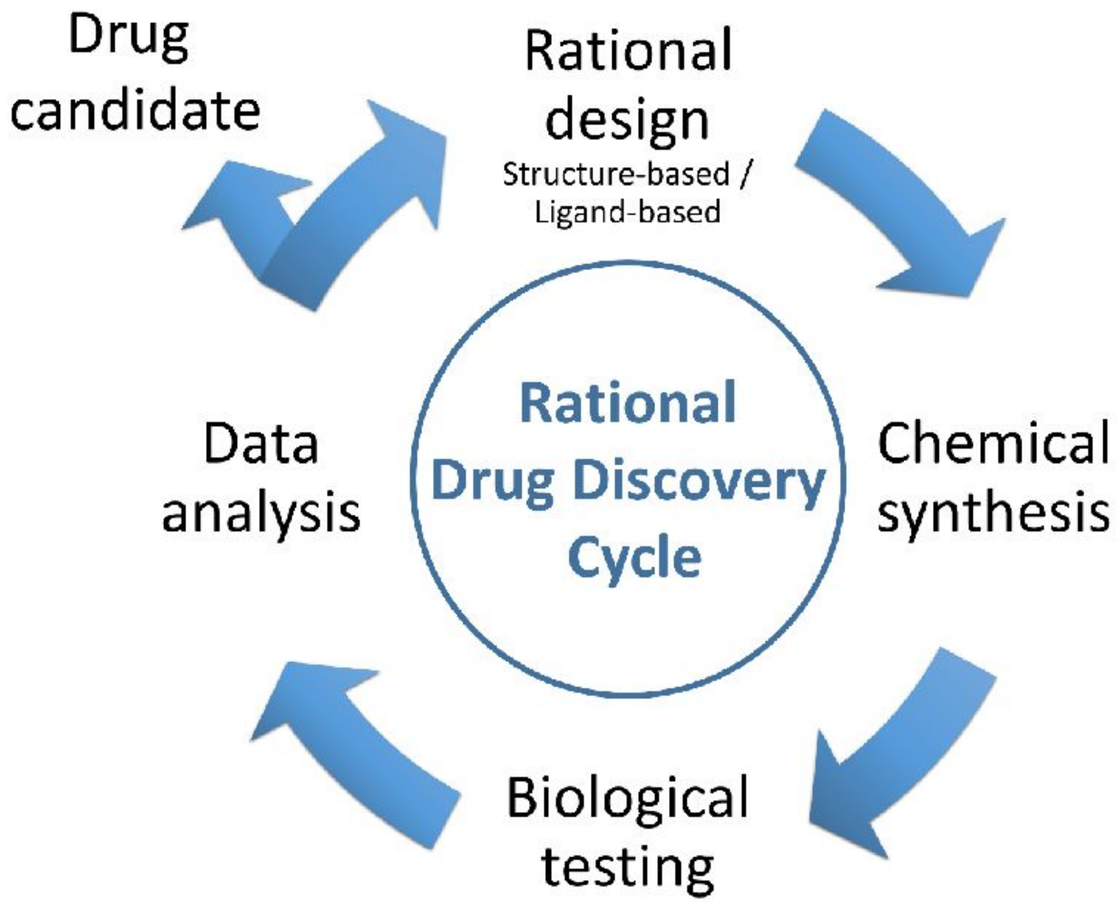
SERVIER moved by you

BASF We create chemistry

janssen PHARMACEUTICAL COMPANIES OF Johnson&Johnson

Never Forget

Only experiments will validate your computational models !



Understanding Ligand Binding

- The most crucial question in early drug discovery
  - Hit identification
  - Hit to lead expansion
  - Lead optimisation

The Important Questions

Data Analysis …

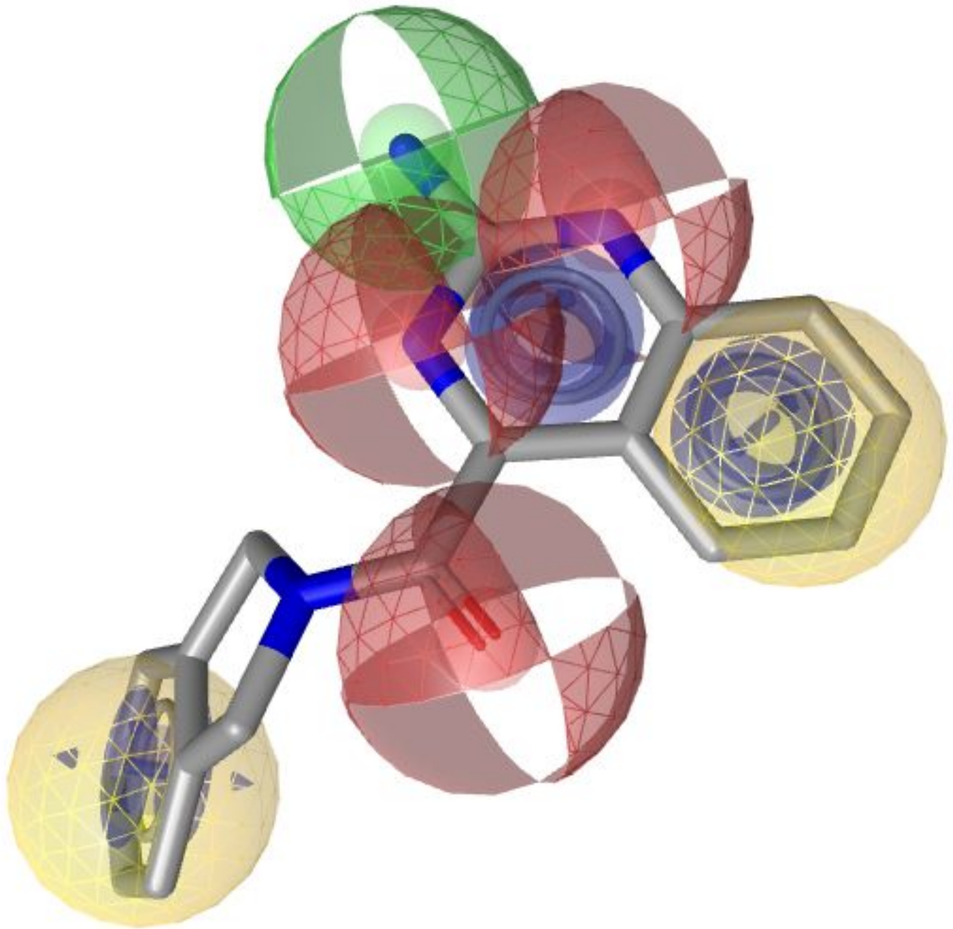
- Which molecule(s) to chose for biological assessment ?
- How to modify molecules for optimising properties ?
- How to build models for reliable medicinal chemistry decision support ?
- How to design molecules void of toxicity risks ?

… and Representation

- How to guide analysis of molecular modelling results ?
- How to make use of MD trajectories for molecular design ?

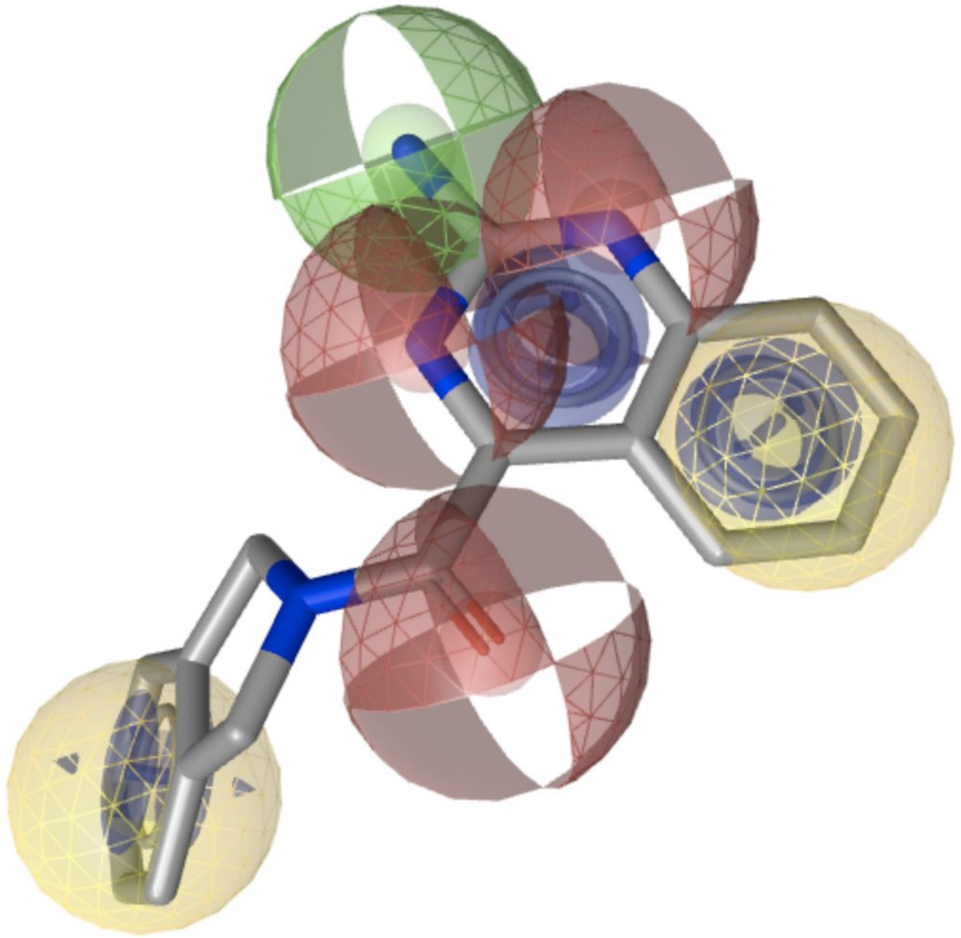
A Possible Way To Do It

- Compare compound structures in view of their preferences for specific molecular interactions
- Annotate molecules with all interaction features possible
- Find out, which of them are the really important ones …



Target Structure Available

- Consider only all those interactions of the ligand within the protein binding site



Schütz D., PhD Thesis,
University of Vienna, 2018

The Pharmacophore Concept

"A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure optimal supra-molecular interactions of a ligand with a specific biological target and to trigger (or block) its biological response."

1933 - 2015

C.-G. Wermuth et al., Pure Appl. Chem. 1998, 70: 1129-1143

Feature-based Pharmacophores

Totality of universal chemical features that represent a defined binding mode of a ligand to a bio-molecular target

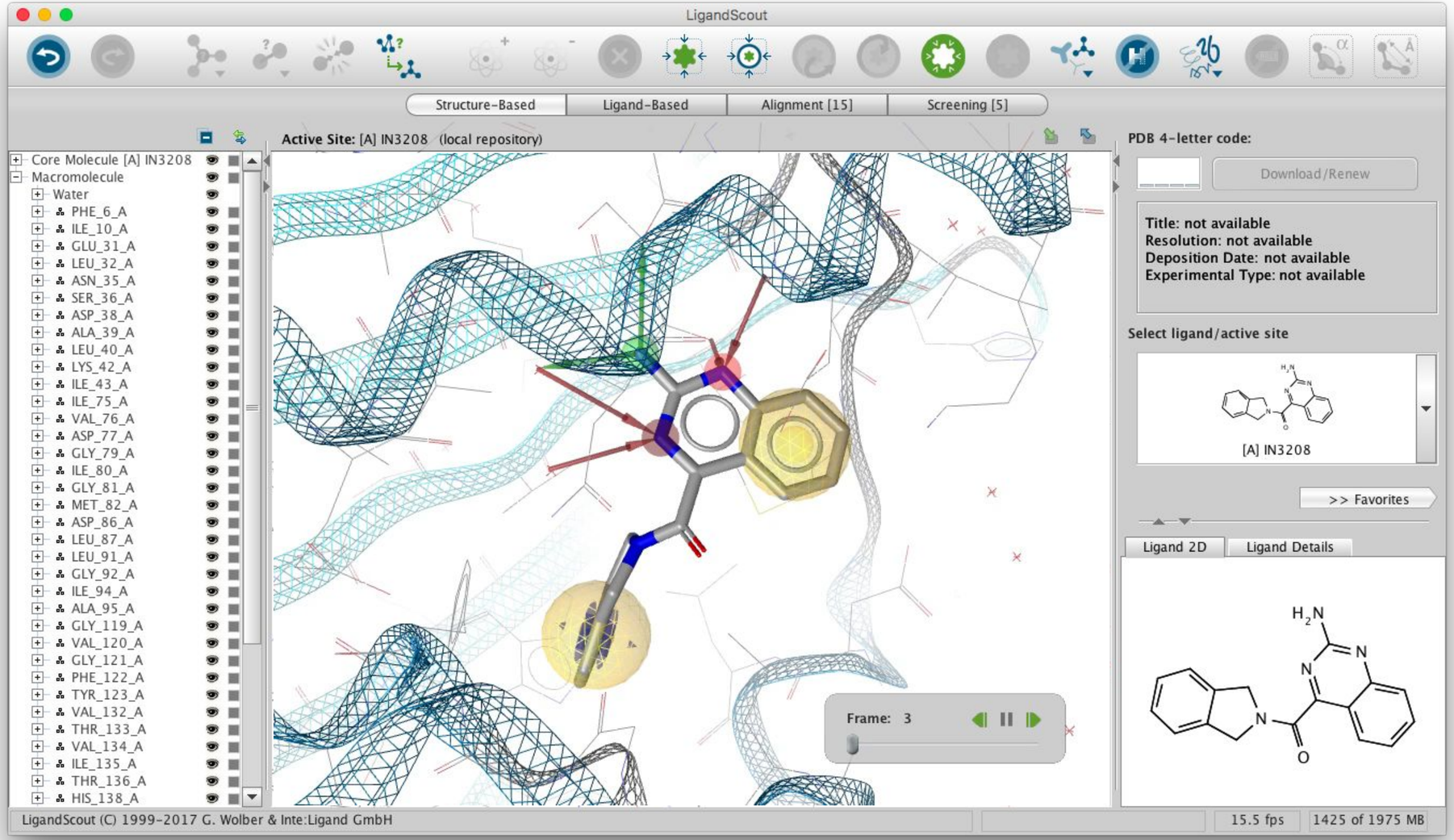
Features: Electrostatic interactions, H-bonding, aromatic interactions, hydrophobic regions, coordination to metal ions, halogen bonds ...

A pharmacophore model is the most suitable data representation method for guiding medicinal chemistry

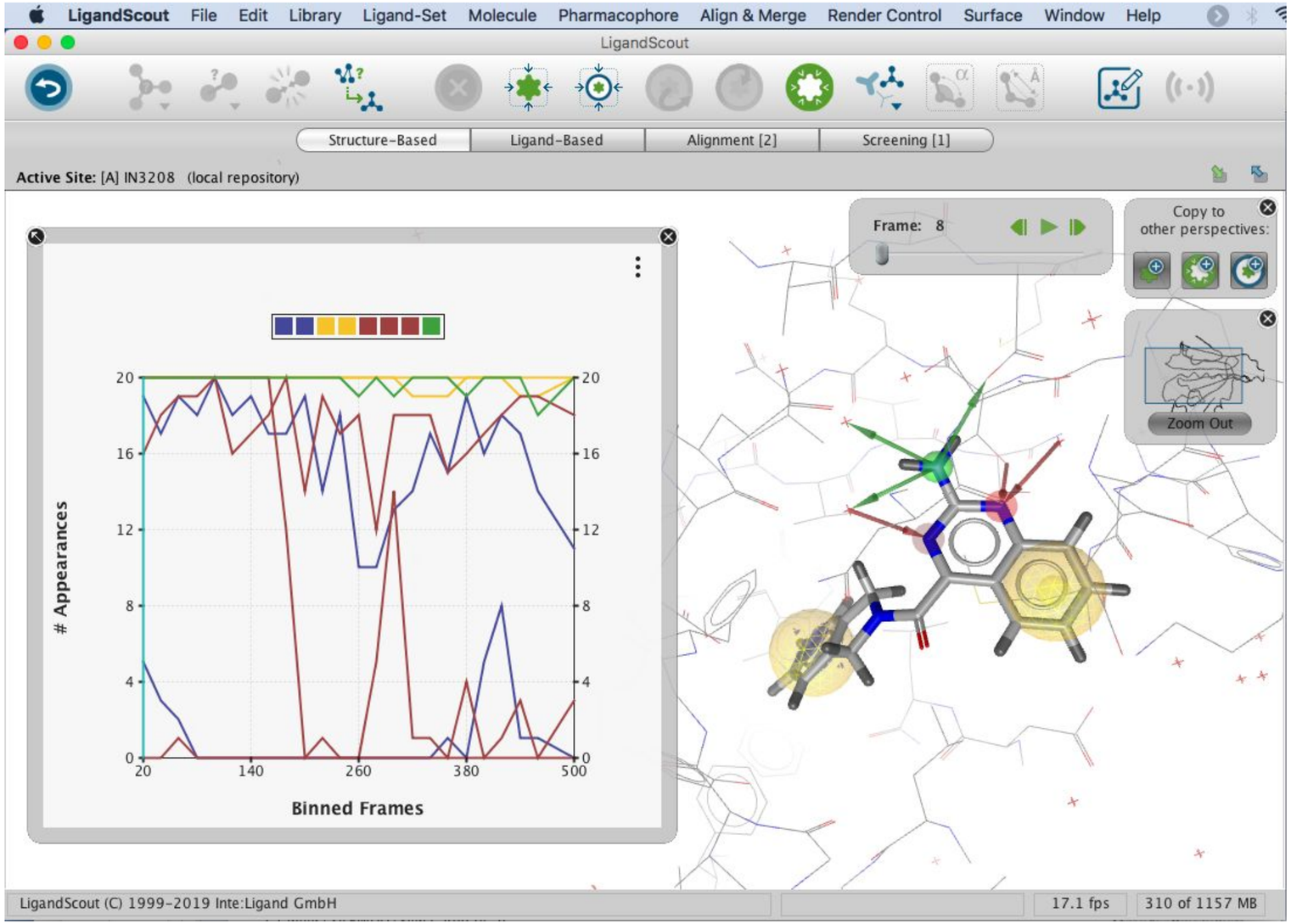
Pharmacophores: An Old Fashioned Concept ?

- Not at all ... there are a lot of new and exciting developments going on ....

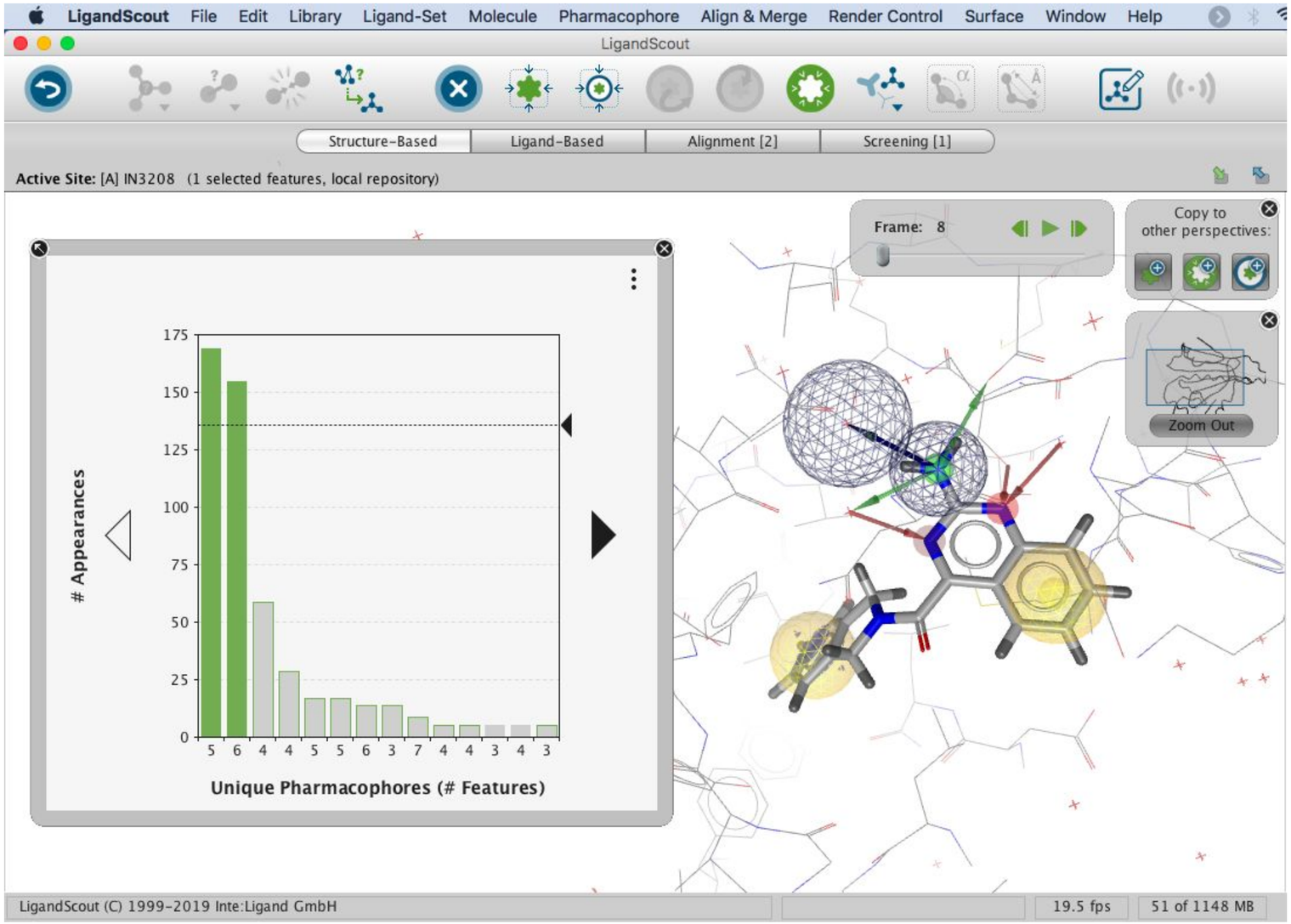
Example: MD Trajectory Analysis



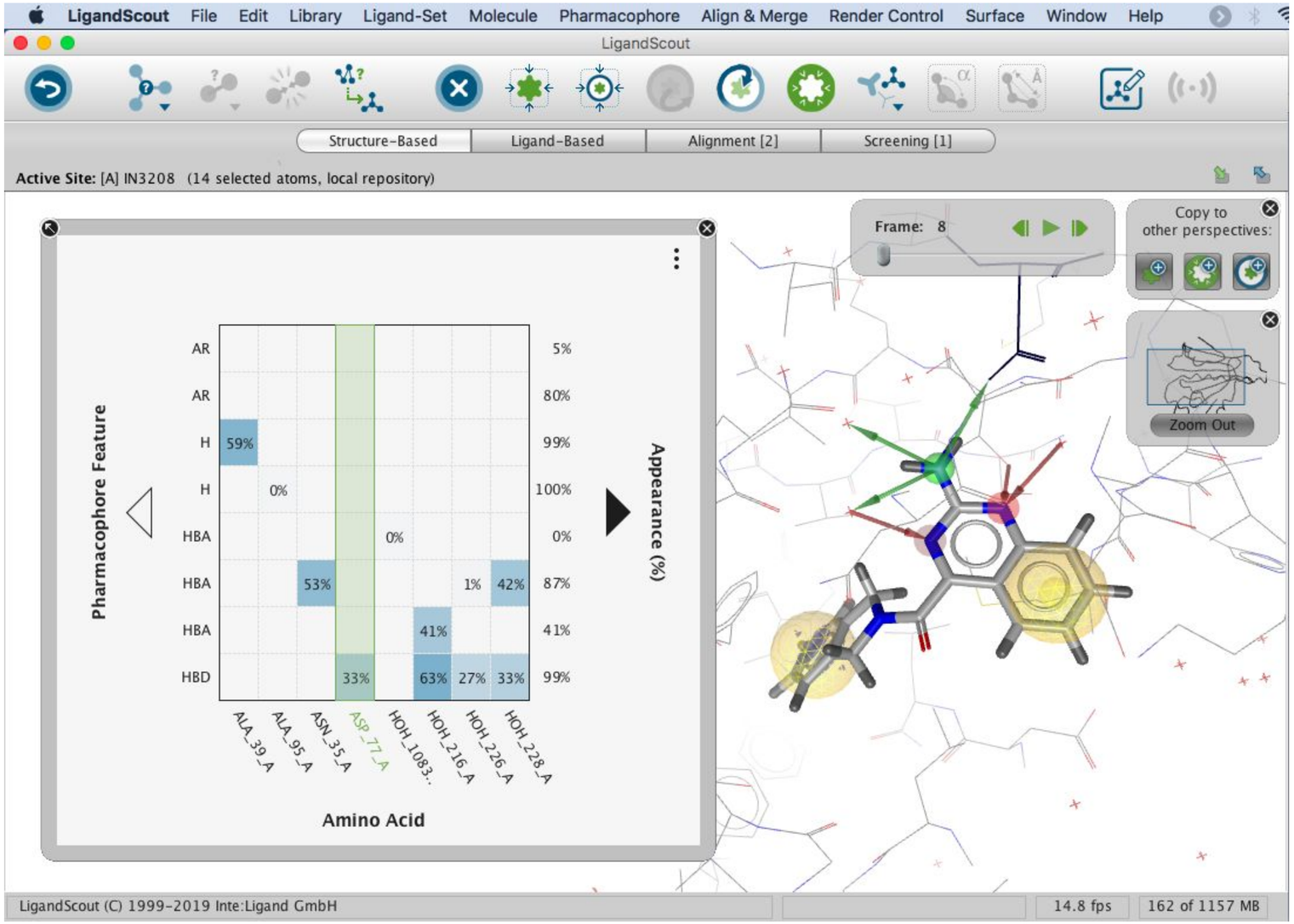
MD Feature Frequency Analysis



Find Models With Specific Features



Determine Interacting Amino Acids

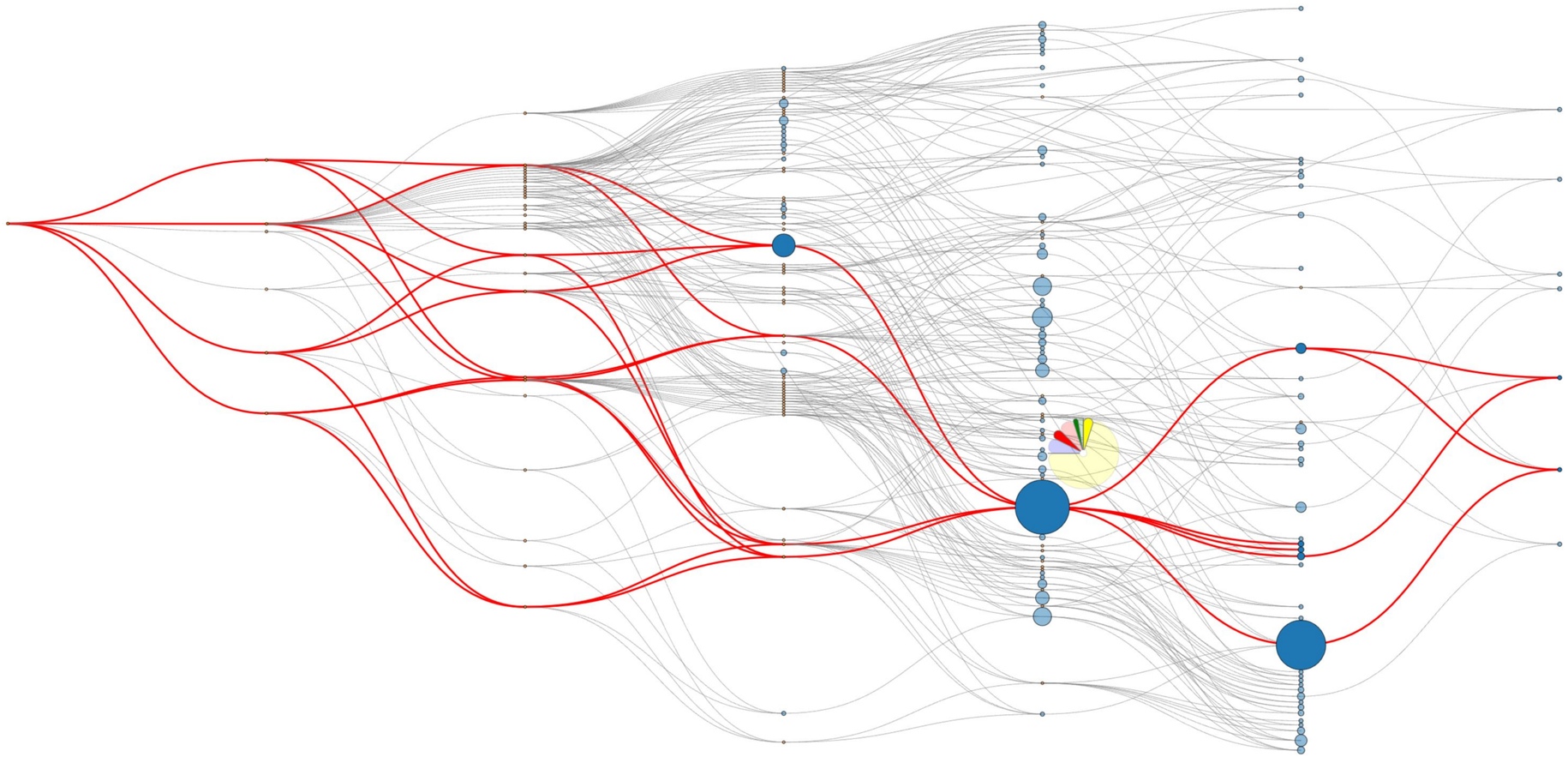


Hierarchical Pharmacophore Network Analysis

Selected Pharmacophores:

Node 0, Frame(s) 13,21,56,64,101,102,105,110,118,125,128,136,147,159,169,173,178,180,184,185,191,192,207,218,221,227,240,255,258,263,275,294,298

Common Fingerprint:

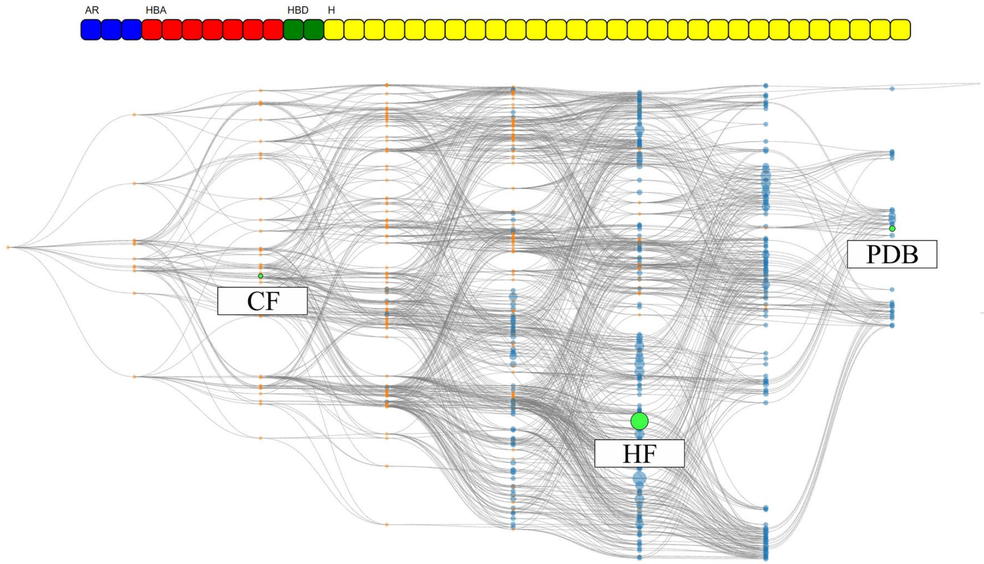


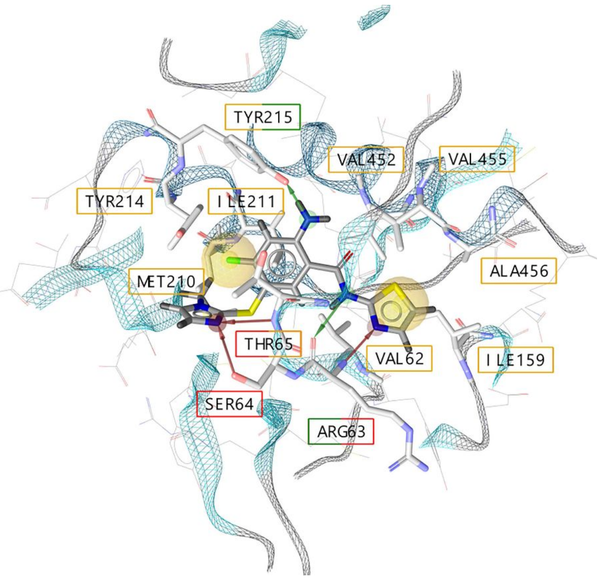
Arthur Garon

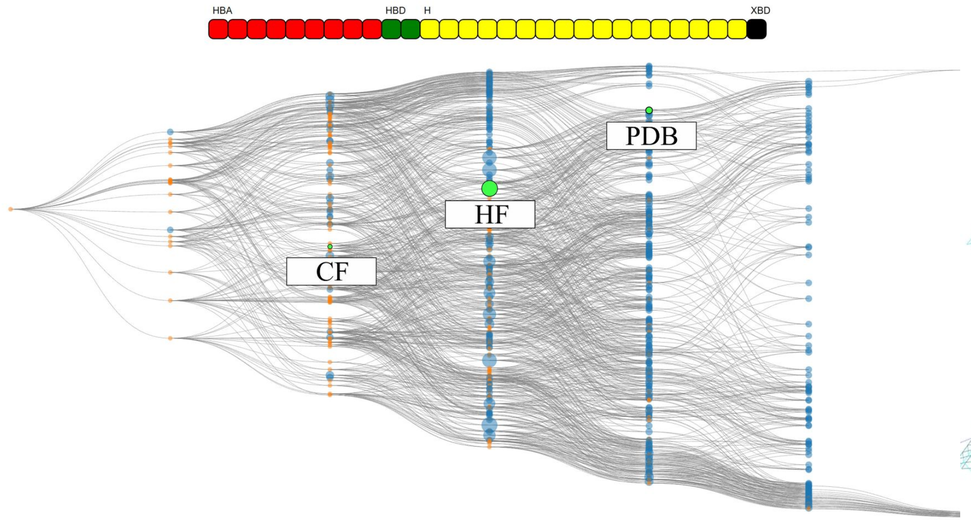
Garon A. et al (2020) [https://doi.org/10.3389/fmolb.2020.599059]

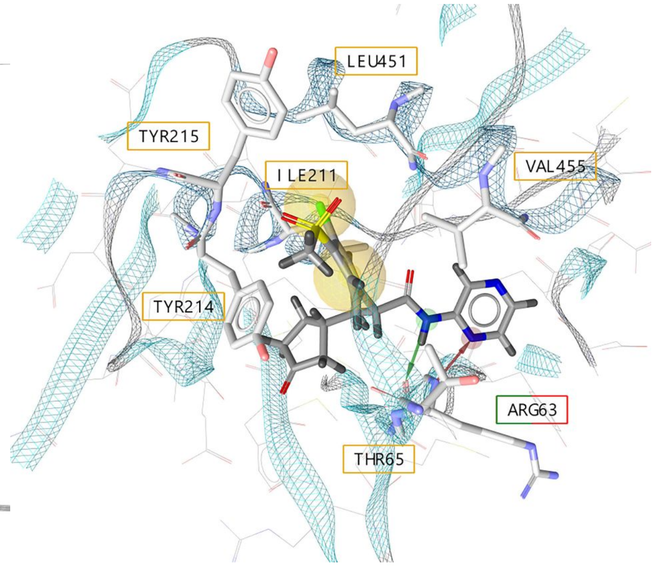
Glucokinase Case Study

- Analysis of binding modes of ligands in human glucokinase: Difference in ligand behaviour, inducing formation of active or inactive form becomes evident from hierarchical pharmacophore network analysis, provides guidance for further design





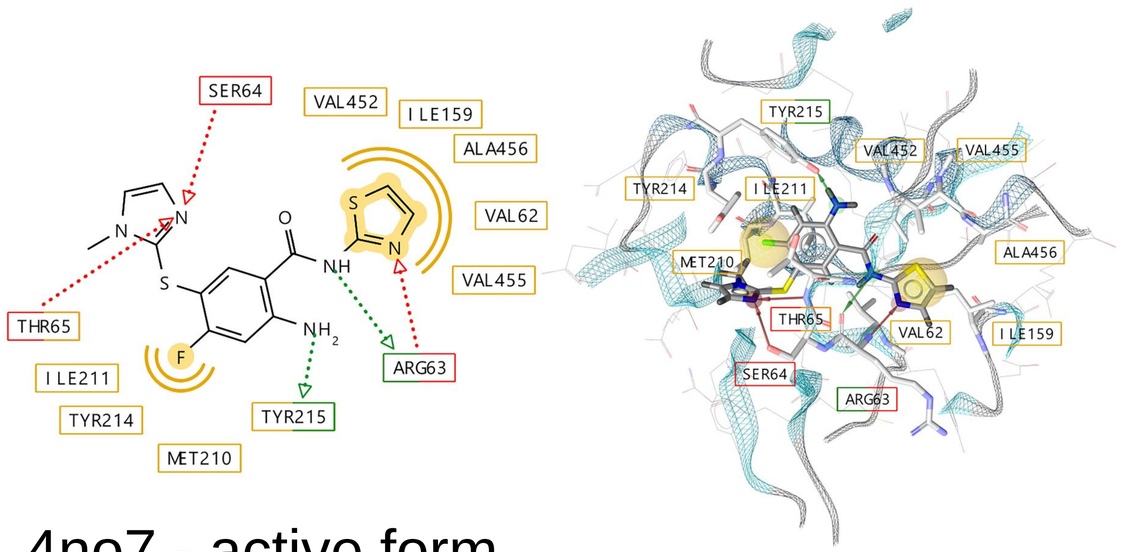


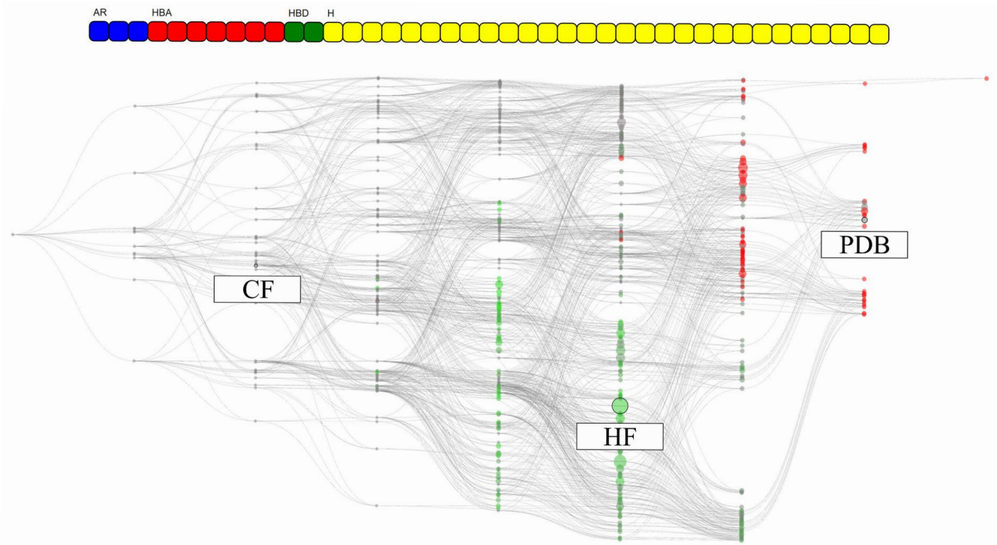


Garon A. et al (2020) [https://doi.org/10.3389/fmolb.2020.599059]

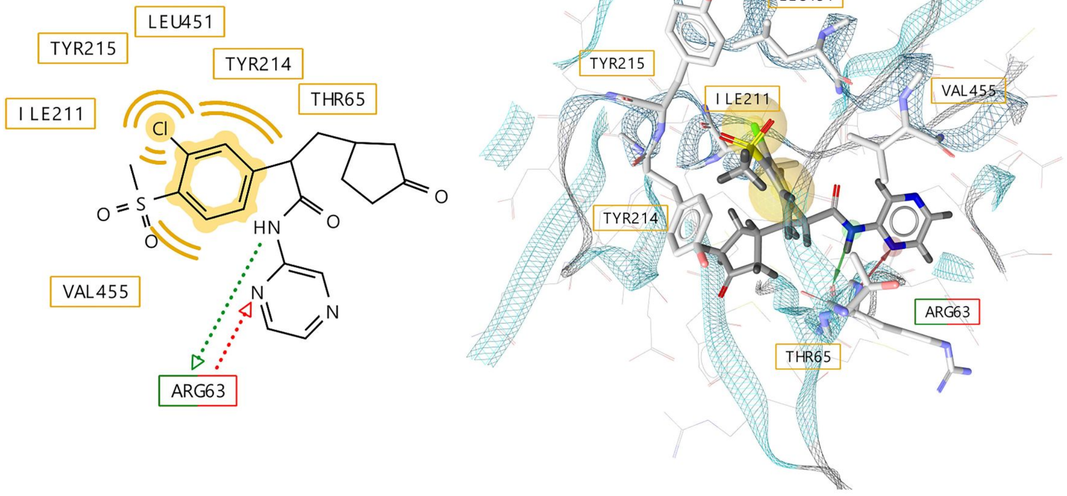
Glucokinase Case Study

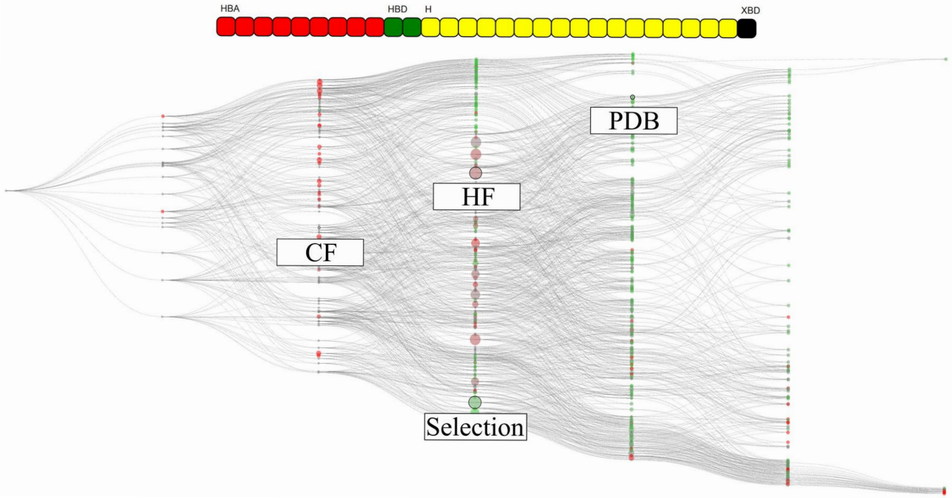
1v4s - inactive form





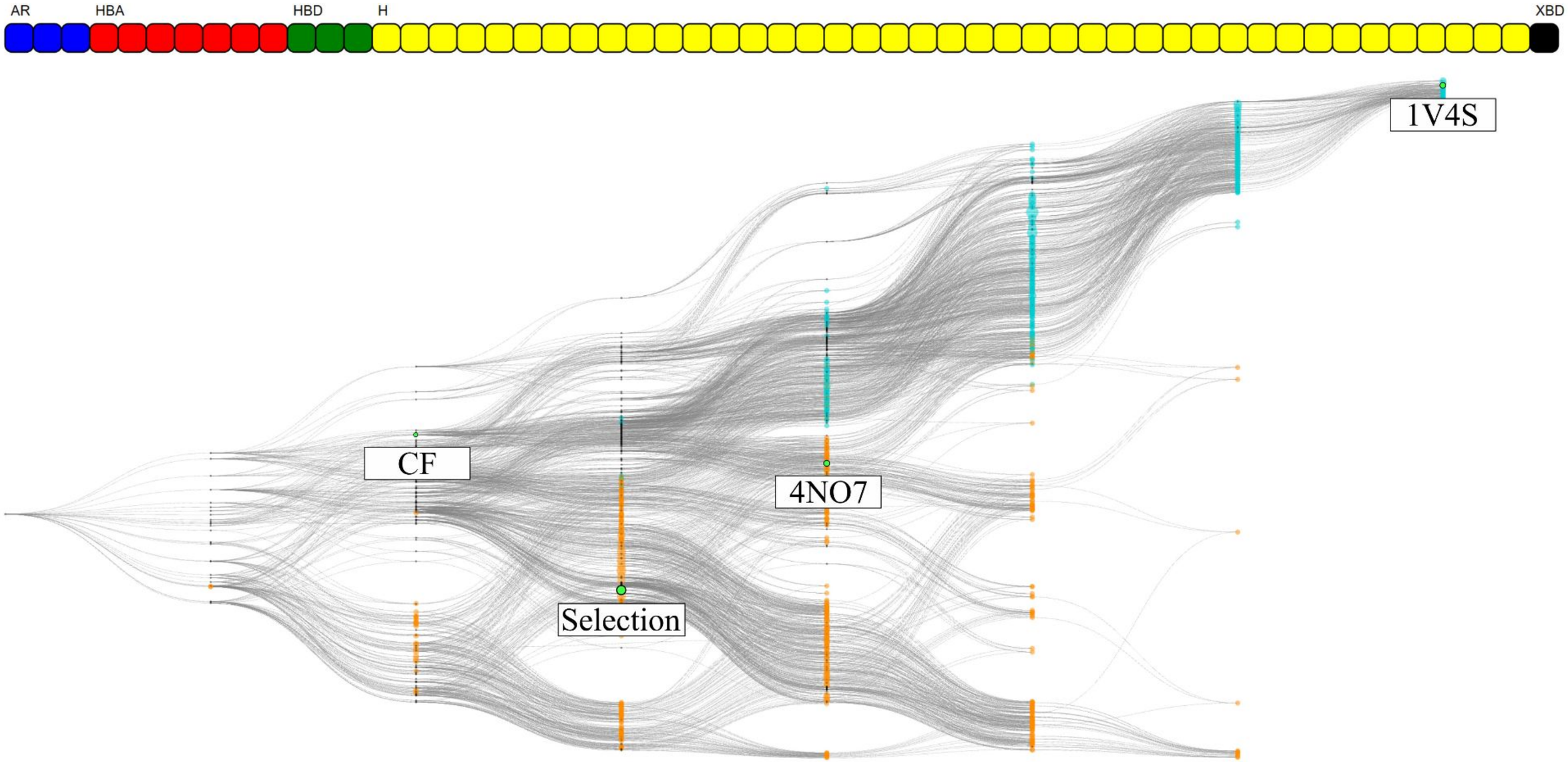
4no7 - active form





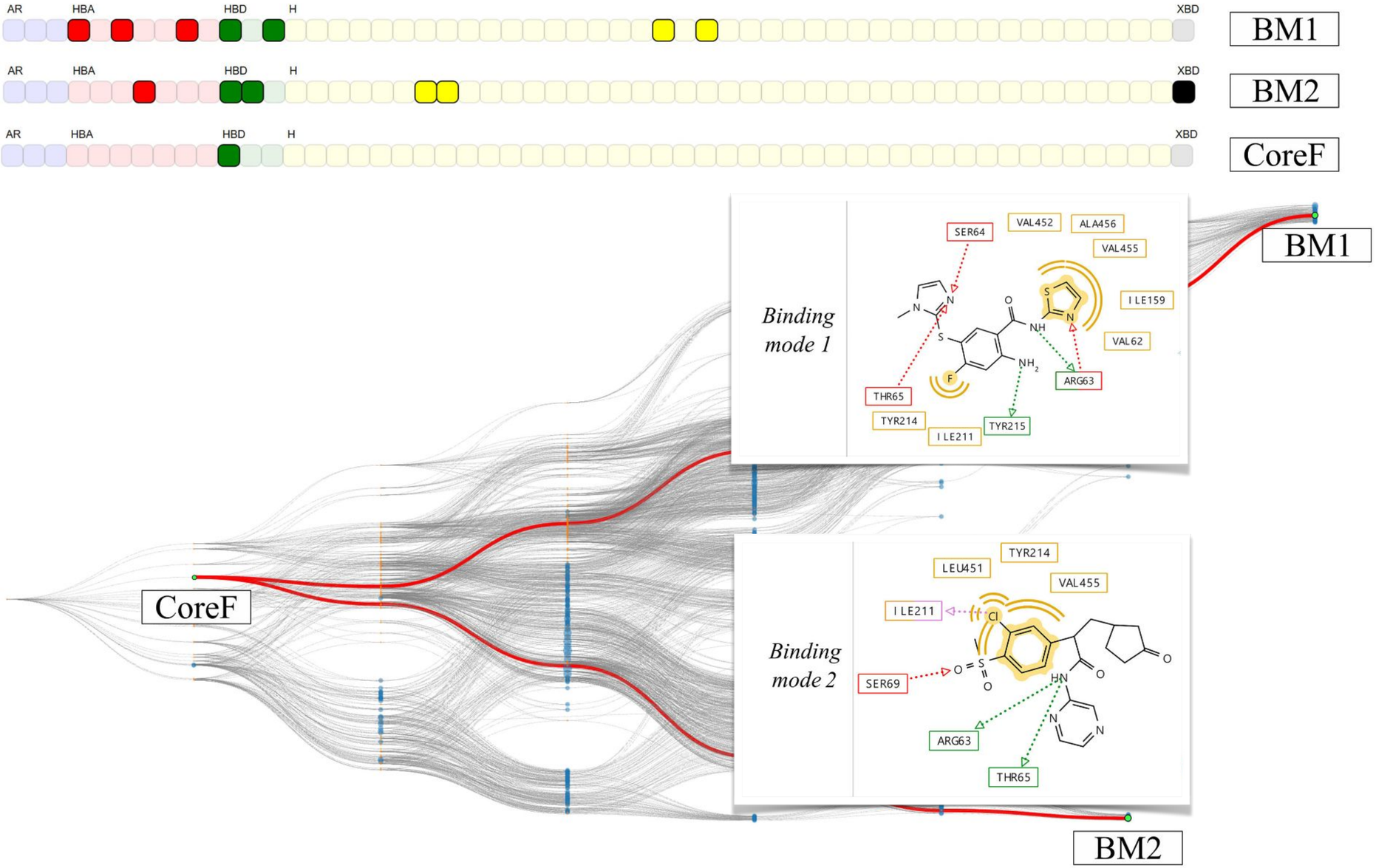
Garon A. et al (2020) [https://doi.org/10.3389/fmolb.2020.599059]

Glucokinase Case Study



Garon A. et al (2020) [https://doi.org/10.3389/fmolb.2020.599059]

Glucokinase Case Study



Garon A. et al (2020) [https://doi.org/10.3389/fmolb.2020.599059]

LigandScout

- Fully integrated molecular design package
- High end GUI & command line tools
- Workflow integration into KNIME
- World wide user basis
- Leading solution for pharmacophore based modelling
- Next generation capabilities in active development:
  - faster high throughput on giga-scale libraries
  - interactive pharmacophore-based clustering ...

LigandScout Scientific Articles

- Around 3600 papers*
  - structure-based modelling
  - ligand-based modelling
  - virtual screening
- Hit identification
- Fragment-based design
- Lead structure optimisation
- Protein-Protein Interactions
- Drug repurposing
- Profiling (side-effects)

Protein Interface Pharmacophore Mapping Tools for Small Molecule Protein: Protein Interaction Inhibitor Discovery

Arnout Voet[1,*], Eleanor F. Banwell[2], Kamlesh K. Sahu[1], Jonathan G. Heddle[2] and Kam Y. J. Zhang[1]

[1]*Zhang Initiative Research Unit, and* [2]*Heddle Initiative Research Unit, Advanced Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan*

Abstract: Protein:protein interactions are becoming increasingly significant as potential drug targets; however, the rational identification of small molecule inhibitors of such interactions remains a challenge. Pharmacophore modelling is a popular tool for virtual screening of compound libraries, and has previously been successfully applied to the discovery of

Pharmacophore-Based Discovery of Small-Molecule Inhibitors of Protein–Protein Interactions between HIV-1 Integrase and Cellular Cofactor LEDGF/p75

Laura De Luca,*[a] Maria Letizia Barreca,*[b] Stefania Ferro,[a] Frauke Christ,[c] Nunzio Iraci,[b] Rosaria Gitto,[a] Anna Maria Monforte,[a] Zeger Debyser,*[c] and Alba Chimirri[a]

The cellular protein lens epitheliun
transcriptional coactivator p75 (LED
in HIV integration. The protein–pro
tween HIV-1 integrase (IN) and its
may therefore serve as targets for
anti-HIV drugs. In this work, a struc
model for potential small-molecu
LEDGF/p75 interaction was develo
software. The 3D model obtained
ing of our in-house chemical datab
identification of compound CHIBA
for further optimization. The rationa

Identification of the first non-peptidic small molecule inhibitor of the c-Abl/14-3-3 protein–protein interactions able to drive sensitive and Imatinib-resistant leukemia cells to apoptosis

Valentina Corradi[a,†], Manuela Mancini[b], Fabrizio Manetti[a], Sara Petta[b], Maria Alessandra Santucci[b], Maurizio Botta[a,*]

[a] *Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, Via Aldo Moro 2, I-53100 Siena, Italy*
[b] *Dipartimento di Ematologia e Scienze Oncologiche "Lorenzo e Ariosto Seràgnoli", Università di Bologna, Via Massarenti 9, I-40138 Bologna, Italy*

ARTICL

Article history:
Received 28 Jun
Revised 3 Augus
Accepted 4 Augu
Available online

Therapeutic Discovery

Molecular Cancer Therapeutics

New Use for an Old Drug: Inhibiting ABCG2 with Sorafenib

Yinxiang Wei[1,3], Yuanfang Ma[3], Qing Zhao[1,4], Zhiguang Ren[1,3], Yan Li[1], Tingjun Hou[2], and Hui Peng[1]

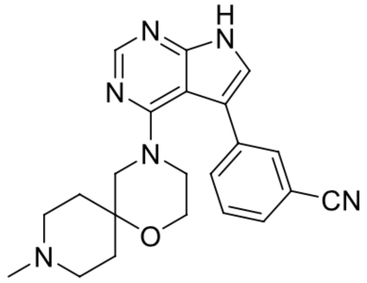
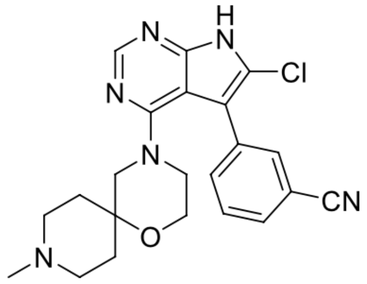
Abstract

Human ABCG2, a member of the ATP-binding cassette transporter superfamily, represents a promising target for sensitizing MDR in cancer chemotherapy. Although lots of ABCG2 inhibitors were identified, none of them has been tested clinically, maybe because of several problems such as toxicity or safety and pharmacokinetic uncertainty of compounds with novel chemical structures. One efficient solution is to rediscover new uses for existing drugs with known pharmacokinetics and safety profiles. Here, we found the new use for

* https://scholar.google.com, 25 January 2025

LigandScout Success Story @Pfizer

Journal of Medicinal Chemistry

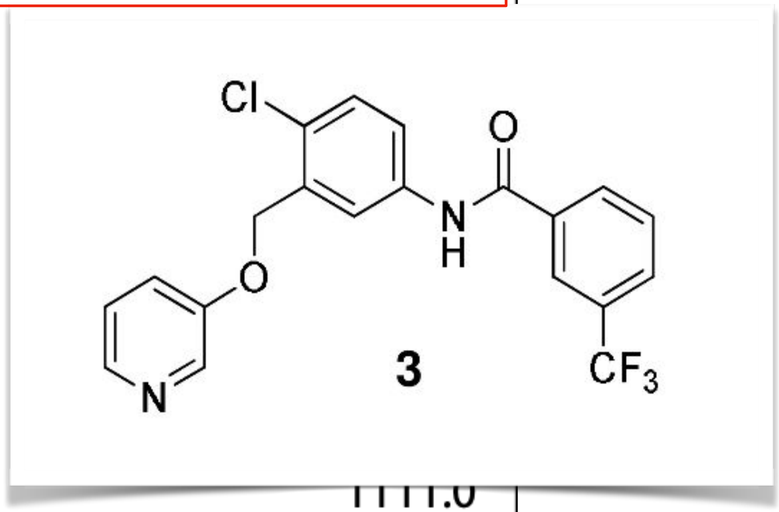
Compound Number	Structure	HPK1 Biochemical Ki (LogD/LipE)	Jurkat pSLP76 cell IC_{50} (LipE)	Jurkat WT IL-2 cell IC_{50}/EC_{50} (max % response)	Jurkat HPK1 KO IL-2 cell IC_{50}/EC_{50} (max % response)	PKCθ Biochemical Ki (window)[a]	PKCη Biochemical Ki
1		17 ± 9 nM (1.4/6.4)	1,960 ± 121 nM (4.4)	IC_{50} = 323 ± 181 nM (28%)	IC_{50} = 259 ± 6 nM (15%)	22 ± 11 nM (1x)	2.4 ± 0.4 nM
2		2.9 ± 1.7 nM (2.0/6.5)	806 ± 105 nM (4.1)	IC_{50} = 1340 ± 148 nM (28%)	IC_{50} = 453 ± 279 nM (24%)	33 ± 14 nM (11x)	7.9 ± 0.6 nM

AI-based Design into Binding Site

ChemMedChem | Communications
doi.org/10.1002/cmdc.202000786
Chemistry Europe
European Chemical Societies Publishing

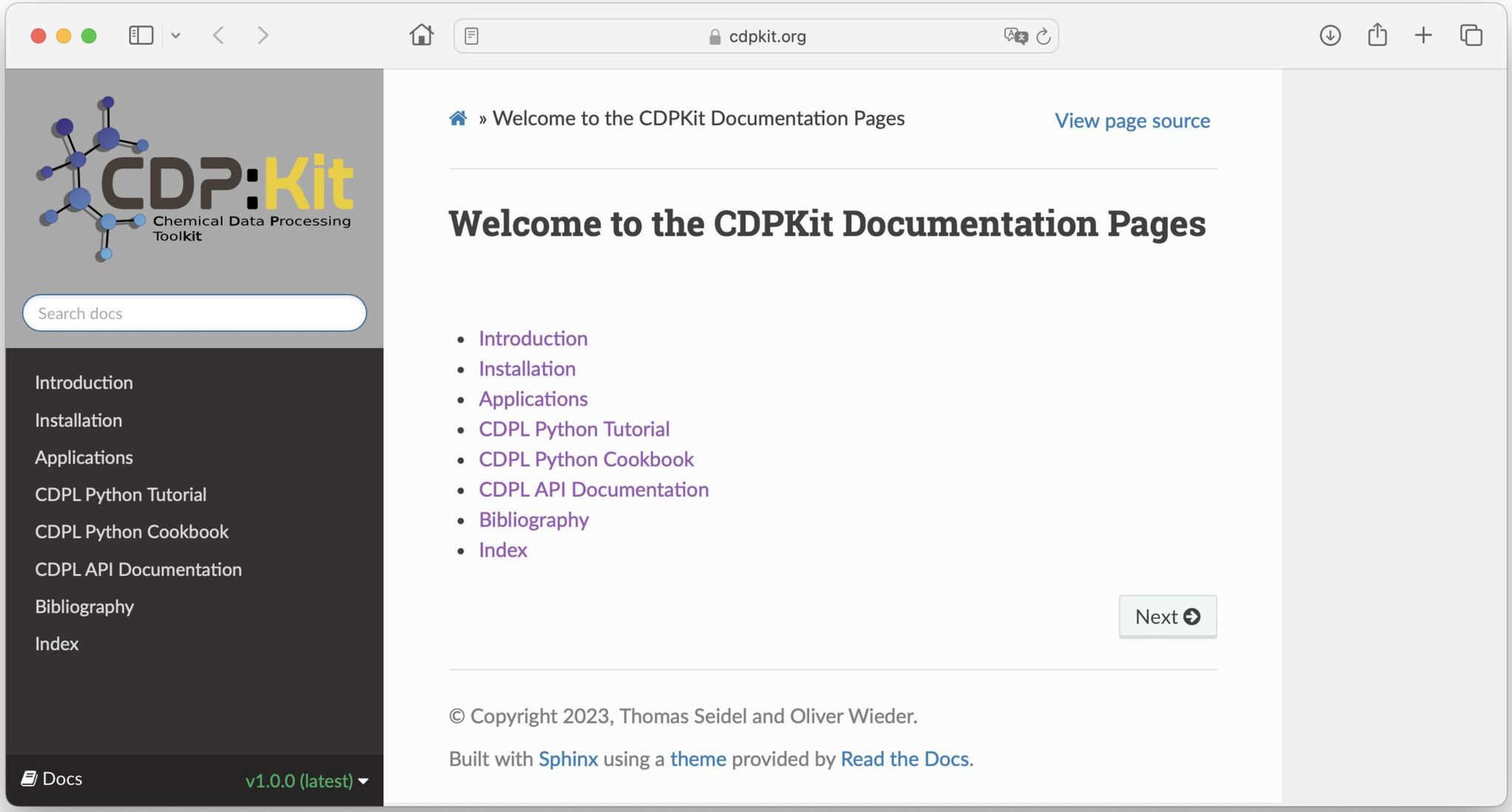
Table 1. DDR1 inhibitory activity of the synthesized compounds.

Compound	Pharmacophore score[a]	Binding affinity score [kJ/mol][b]	IC_{50} [nM][c]
1	0.96	−50.51	1005.9
2	0.95	−52.67	2239.4
3	0.83	−47.13	92.5
4	0.96	−51.97	
5	0.86	−37.29	
6	0.84	−40.83	
7	0.85	−54.38	
8	0.85	−51.78	
9	0.85	−49.46	
7a	0.85	−53.06	
8a	0.85	−51.15	
9a	0.85	−54.83	1111.0



3

Our Open Source Toolkit: CDPKit



https://github.com/molinfo-vienna/cdpkit

Our Open Source Toolkit: CDPKit

universität wien

- Chemical Data Processing Toolkit
  - Implemented in C++
  - Contains all basic functionality for chemoinformatics workflows
- Comprises the Chemical Data Processing Library (CDPL) together with command line software tools
- CDPL C++ API and Python-interfacing layer provided
- Integration of major machine learning libraries
  - scikit-learn, PyTorch, TensorFlow
- Represents our standard environment for prototyping novel algorithms
- Follow-up implementation into LigandScout and KNIME platforms

Thomas Seidel

Oliver Wieder

inte:ligand
www.inteligand.com Vienna, Austria

Bigger Haystack -> Better Hits ?

- Need for faster conformer generator and alignment procedure

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Open Access

pubs.acs.org/jcim

Article

High-Quality Conformer Generation with CONFORGE: Algorithm and Performance Assessment

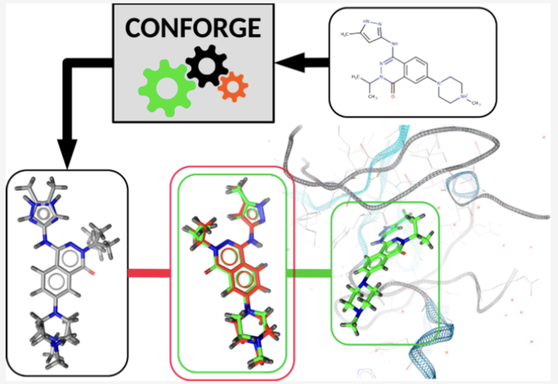
Thomas Seidel,* Christian Permann, Oliver Wieder, Stefan M. Kohlbacher, and Thierry Langer

Cite This: *J. Chem. Inf. Model.* 2023, 63, 5549–5570

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Knowledge of the putative bound-state conformation of a molecule is an essential prerequisite for the successful application of many computer-aided drug design methods that aim to assess or predict its capability to bind to a particular target receptor. An established approach to predict bioactive conformers in the absence of receptor structure information is to sample the low-energy conformational space of the investigated molecules and derive representative conformer ensembles that can be expected to comprise members closely resembling possible bound-state ligand conformations. The high relevance of such conformer generation functionality led to the development of a wide panel of dedicated commercial and open-source software tools throughout the last decades. Several published benchmarking studies have shown that open-source tools usually lag behind their commercial competitors in many key aspects. In this work, we introduce the open-source conformer ensemble generator CONFORGE, which aims at delivering state-of-the-art performance for all types of organic molecules in drug-like chemical space. The ability of CONFORGE and several well-known commercial and open-source conformer ensemble generators to reproduce experimental 3D structures as well as their computational efficiency and robustness has been assessed thoroughly for both typical drug-like molecules and macrocyclic structures. For small molecules, CONFORGE clearly outperformed all other tested open-source conformer generators and performed at least equally well as the evaluated commercial generators in terms of both processing speed and accuracy. In the case of macrocyclic structures, CONFORGE achieved the best average accuracy among all benchmarked generators, with RDKit's generator coming close in second place.



Thomas Seidel

CONFORGE

- Novel conformer generator based on LigandScout iCon
- Available in CDPKit
- Improved small compound performance accuracy
  - Especially for complex compounds such as macrocycles
- Integration into LigandScout XT
  - Faster database generation & possibly better screening hits
    - Approximately two times faster than iCon Fast
  - Additional changes in database format (LDB2)
    - Reduced file size (ca. 60% of LDB; 1 billion compounds currently 16 Tb)
  - Together with G3PS alignment: Basis for P4-based exa-scale VS

molecules

MDPI

Article

Greedy 3-Point Search (G3PS)—A Novel Algorithm for Pharmacophore Alignment

Christian Permann [1,2], Thomas Seidel [1,*] and Thierry Langer [1,2]

[1] Department of Pharmaceutical Sciences, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria; christian.permann@univie.ac.at (C.P.); thierry.langer@univie.ac.at (T.L.)
[2] Inte:Ligand GmbH, Clemens Maria Hofbauer-Gasse 6, 2344 Maria Enzersdorf, Austria
* Correspondence: thomas.seidel@univie.ac.at

Abstract: Chemical features of small molecules can be abstracted to 3D pharmacophore models, which are easy to generate, interpret, and adapt by medicinal chemists. Three-dimensional pharmacophores can be used to efficiently match and align molecules according to their chemical feature pattern, which facilitates the virtual screening of even large compound databases. Existing alignment methods, used in computational drug discovery and bio-activity prediction, are often not suitable for finding matches between pharmacophores accurately as they purely aim to minimize RMSD or maximize volume overlap, when the actual goal is to match as many features as possible within the positional tolerances of the pharmacophore features. As a consequence, the obtained alignment results are often suboptimal in terms of the number of geometrically matched feature pairs, which increases the false-negative rate, thus negatively affecting the outcome of virtual screening experiments. We addressed this issue by introducing a new alignment algorithm, Greedy 3-Point Search (G3PS), which aims at finding optimal alignments by using a matching-feature-pair maximizing search strategy while at the same time being faster than competing methods.

Keywords: pharmacophore alignment; pharmacophore modelling; virtual screening; greedy algorithm; drug design

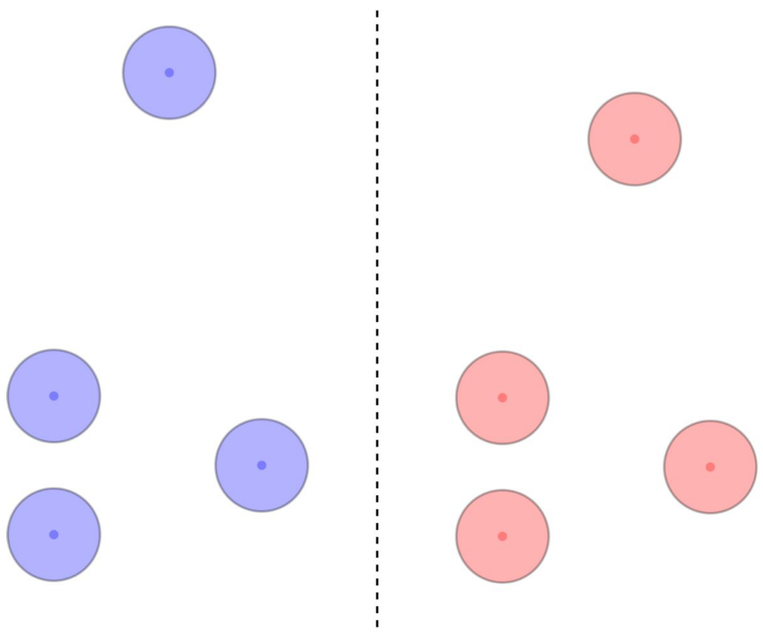
Citation: Permann, C.; Seidel, T.; Langer, T. Greedy 3-Point Search (G3PS)—A Novel Algorithm for Pharmacophore Alignment. *Molecules* **2021**, *26*, 7201. https://doi.org/10.3390/molecules26237201

1. Introduction

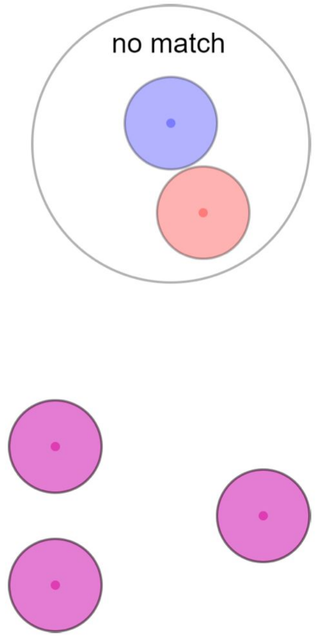
Christian Permann

G3PS Alignment

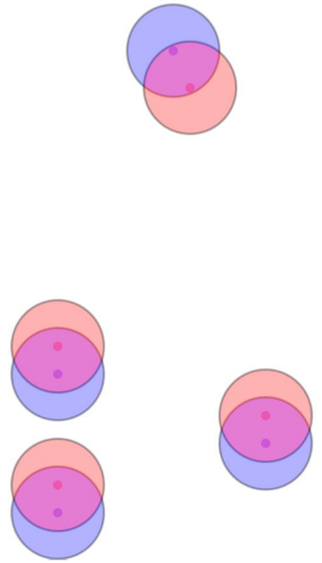
- How can we improve on previous method ?



(a) Two pharmacophores



(b) RMSD or volume-based alignment

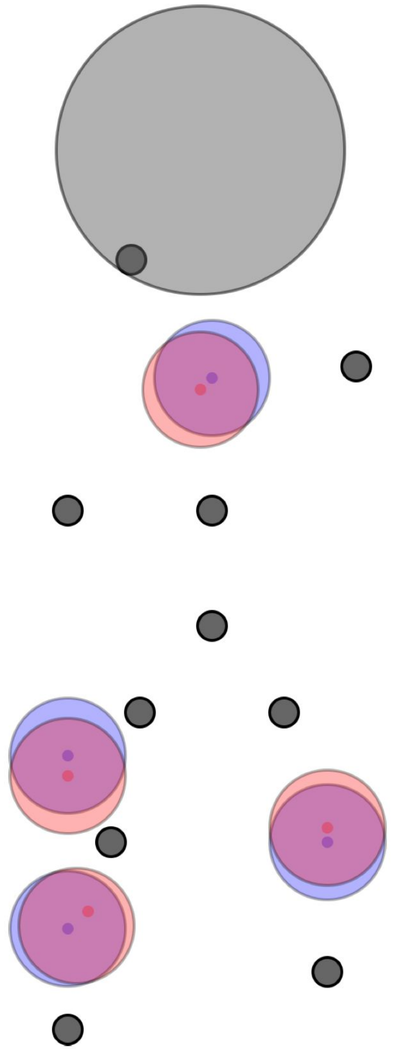


(c) Desired alignment

G3PS Alignment

- Another common pitfalls
  - Failing on exclusion volume spheres
    - Barely touching exclusion volume invalidates alignment
  - Allowing omission of m out of f features
    - for each value r = {0..m}
      - 1. Create all combinations with (f − r) features
      - 2. Perform alignment for each combination
    - #Alignments = Pm f = O(f (m)) → exponential in m

$$\sum_{r=0}^{m} \binom{f}{r} = O(f^{(m)}) \rightarrow$$



G3PS Alignment

- New algorithm is based on iterative improvement of heuristic guesses
- Runtime
  - Much faster than previous methods
  - Independent of number of potentially omitted features
  - User-defined number of computed guesses: Runtime vs Accuracy tuneable
- Found Alignments
  - More accurate retrieval according to user query
  - Optimised for maximum number of matched features
  - Adapts also to exclusion volume spheres

Heuristic Alignment Trees

- Steers alignment between faster runtimes & more accurate retrieval
- Represents the number of (three-point) starting configurations
- Higher number will better sample solution space
  - more correctly identified hits
- Smaller number will result in a faster screening
- Runtime increases less than linearly with number of tries
  - e.g. changing from 20 to 50 - results in approximately 25% longer runtime
- Recommended settings
  - 50 for generic fast run, 300 for generic accurate run

Some More Examples

- New graph neural network architecture for better prediction of standard molecular properties (logP, logS, logD)
- New protein-ligand descriptors for machine learning
- Novel representation of pharmacophore models for machine learning
- Extending pharmacophore based virtual screening for exa-scale libraries

molecules

MDPI

Article

Improved Lipophilicity and Aqueous Solubility Prediction with Composite Graph Neural Networks

Oliver Wieder [1,*], Mélaine Kuenemann [2], Marcus Wieder [1], Thomas Seidel [1], Christophe Meyer [2], Sharon D. Bryant [3] and Thierry Langer [1]

1 Department of Pharmaceutical Chemistry, University of Vienna, Althanstraße 14, A-1090 Vienna, Austria; marcus.wieder@univie.ac.at (M.W.); thomas.seidel@univie.ac.at (T.S.); thierry.langer@univie.ac.at (T.L.)
2 Servier Research Institute-CentEx Biotechnology, 125 Chemin de Ronde, 78290 Croissy-sur-Seine, France; melaine.kuenemann@servier.com (M.K.); christophe.meyer@servier.com (C.M.)
3 Inte:Ligand Software Entwicklungs und Consulting GmbH, 74B/11 Mariahilferstrasse, 1070 Vienna, Austria; bryant@inteligand.com (S.D.B.)
* Correspondence: oliver.wieder@univie.ac.at

Abstract: The accurate prediction of molecular properties, such as lipophilicity and aqueous solubility, are of great importance and pose challenges in several stages of the drug discovery pipeline. Machine learning methods, such as graph-based neural networks (GNNs), have shown exceptionally good performance in predicting these properties. In this work, we introduce a novel GNN architecture, called directed edge graph isomorphism network (D-GIN). It is composed of two distinct sub-architectures (D-MPNN, GIN) and achieves an improvement in accuracy over its sub-architectures employing various learning, and featurization strategies. We argue that combining models with different key aspects help make graph neural networks deeper and simultaneously increase their predictive power. Furthermore, we address current limitations in assessment of deep-learning models, namely, comparison of single training run performance metrics, and offer a more robust solution.

Keywords: AI, deep-learning; neural-networks; graph neural-networks; cheminformatics; molecular property; machine-learning; computational chemistry; lipophilicity; solubility

check for updates

Citation: Wieder, O.; Kuenemann, M.; Wieder, M.; Seidel, T.; Meyer, C.; Bryant, S.D.; Langer, T. Improved Lipophilicity and Aqueous Solubility Prediction with Composite Graph Neural Networks. *Molecules* **2021**, *26*, 6185. https://doi.org/10.3390/molecules26206185

universität wien

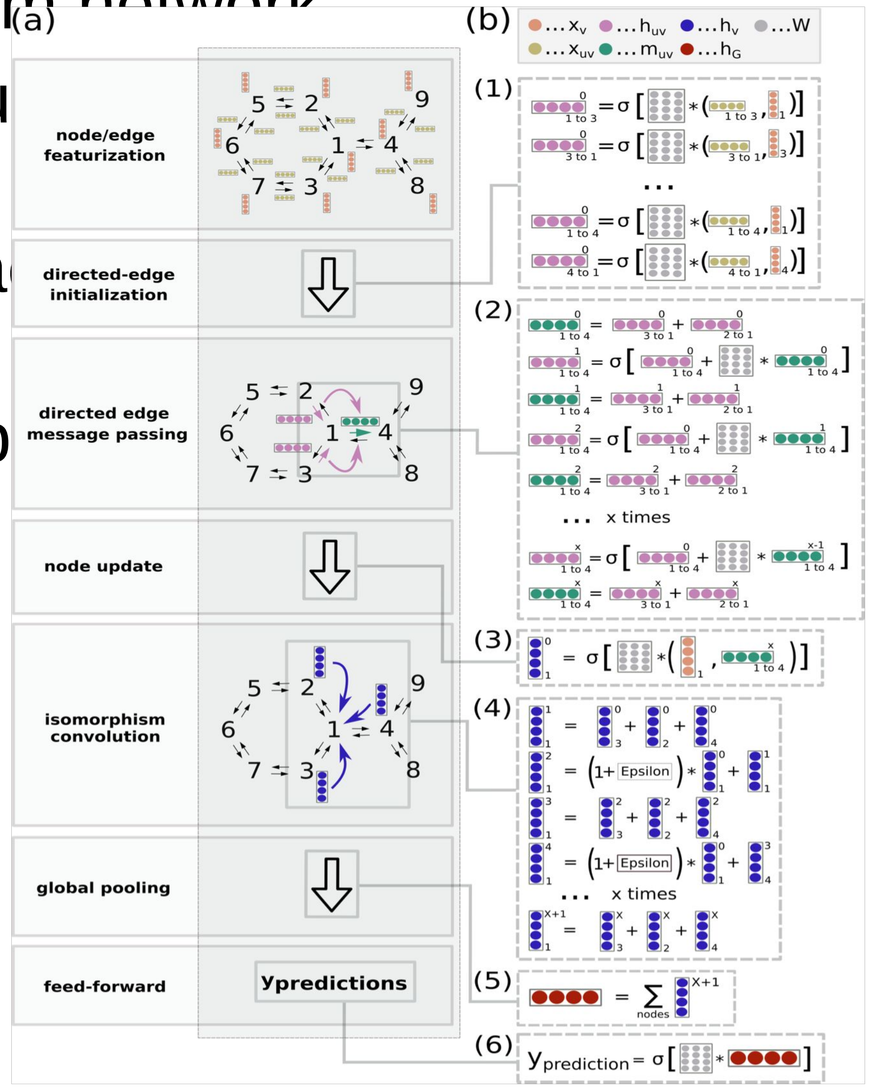
Oliver Wieder

https://doi.org/10.3390/molecules26206185

inte:ligand www.inteligand.com Vienna, Austria

Novel GNN Architecture: D-GIN

- D-GIN: Directed edge graph isomorphism network
- Composed of two distinct sub-architectu
  GIN)
- Achieves drastic improvement in accura
  architectures
- Successfully used for logP, logS, logD p



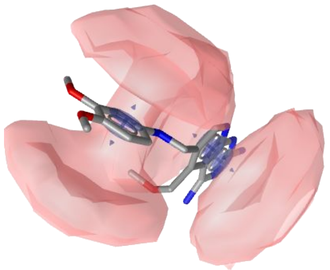
https://doi.org/10.3390/molecules26206185

How To Rank Pharmacophore Features ?

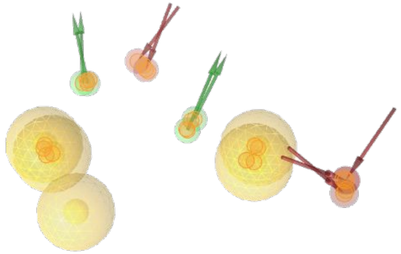
- Ligand-based modelling
  - Explain quantitative differences in binding (K_i, IC_{50})

- Structure-based modelling
  - Geometry of the features: Delta to optimum
  - Environment
  - Analysis of dynamics / kinetics of binding: MD Simulations
    - MD trajectory analysis with pharmacophore models
    - Construction of hierarchical pharmacophore feature trees

Next Gen Pharmacophores: QPhARs

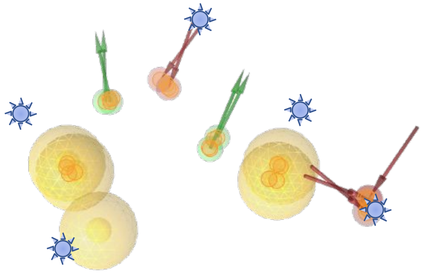


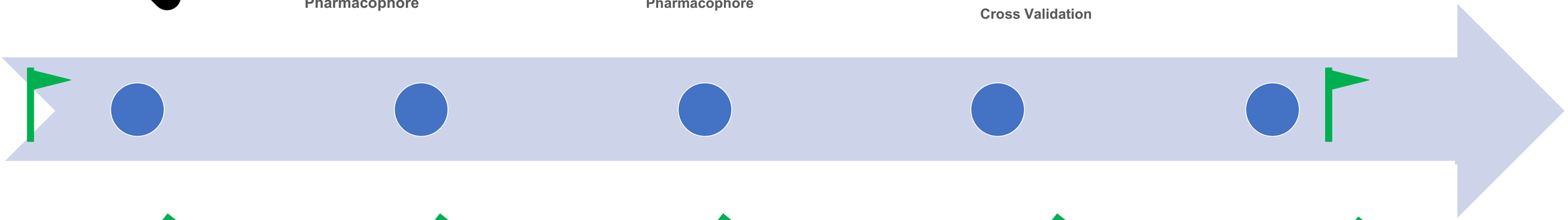
GRAIL* Based Pharmacophore



LigandScout Merged Pharmacophore

Cross Validation





Kohlbacher *et al. J Cheminform* (2021) 13:57
https://doi.org/10.1186/s13321-021-00537-9

Journal of Cheminformatics

METHODOLOGY Open Access

QPHAR: quantitative pharmacophore activity relationship: method and validation

Stefan M. Kohlbacher, Thierry Langer and Thomas Seidel*

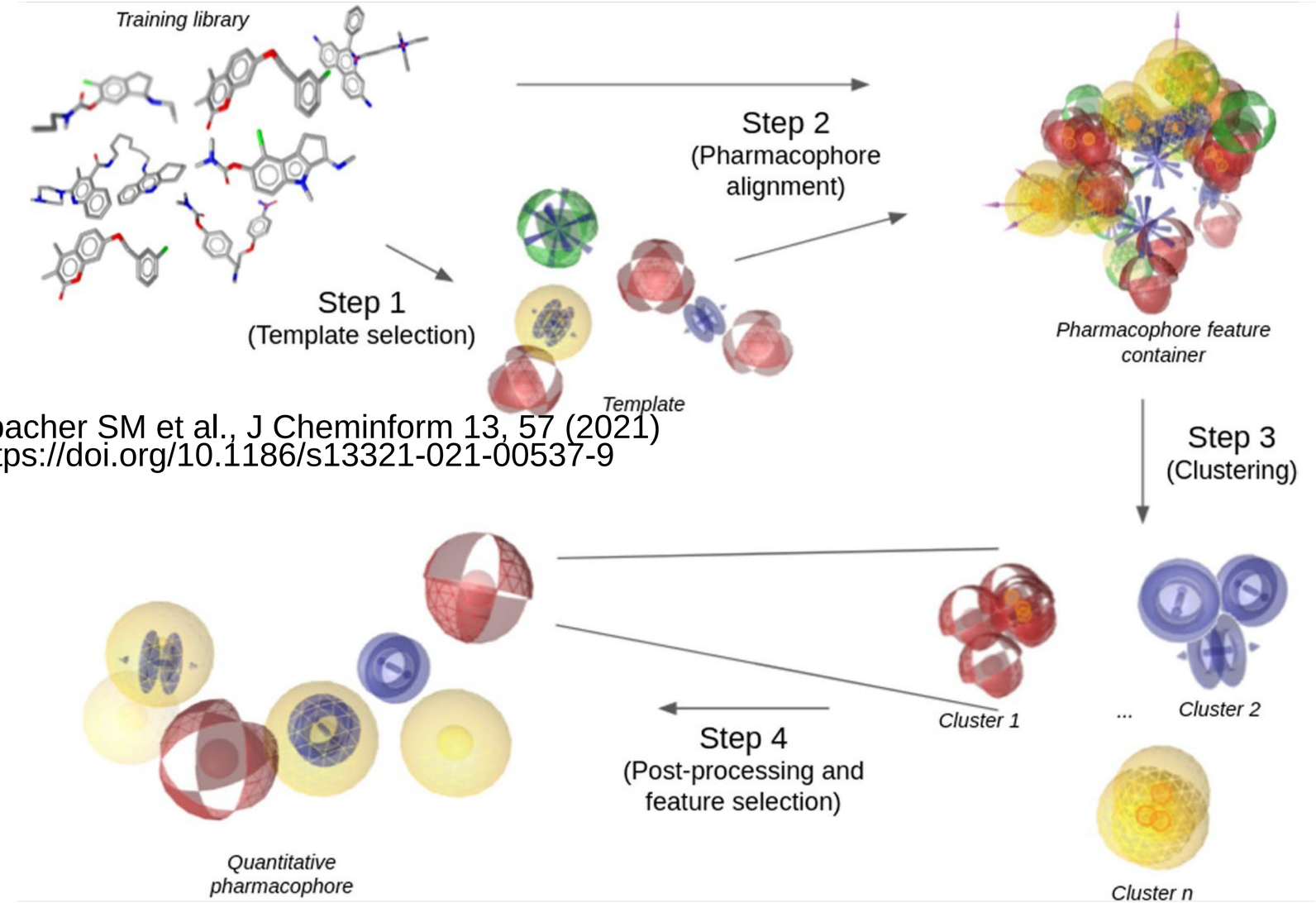
Abstract

QSAR methods are widely applied in the drug discovery process, both in the hit-to-lead and lead optimization phase, as well as in the drug-approval process. Most QSAR algorithms are limited to using molecules as input and disregard pharmacophores or pharmacophoric features entirely. However, due to the high level of abstraction, pharmacophore representations provide some advantageous properties for building quantitative SAR models. The abstract depiction of molecular interactions avoids a bias towards overrepresented functional groups in small datasets. Furthermore, a well-crafted quantitative pharmacophore model can generalise to underrepresented or even missing molecular features in the training set by using pharmacophoric interaction patterns only. This paper presents a novel method to construct quantitative pharmacophore models and demonstrates its applicability and robustness on more than 250 diverse datasets. fivefold cross-validation on these datasets with default settings yielded an average RMSE of 0.62, with an average standard deviation of 0.18. Additional cross-validation studies on datasets with 15–20 training samples showed that robust quantitative pharmacophore models could be obtained. These low requirements for dataset sizes render quantitative pharmacophores a viable go-tomethod for medicinal chemists, especially in the lead-optimisation stage of drug discovery projects.

Keywords: Pharmacophore, QSAR, Regression, Machine learning, Quantitative-pharmacophore-model

Stefan Kohlbacher

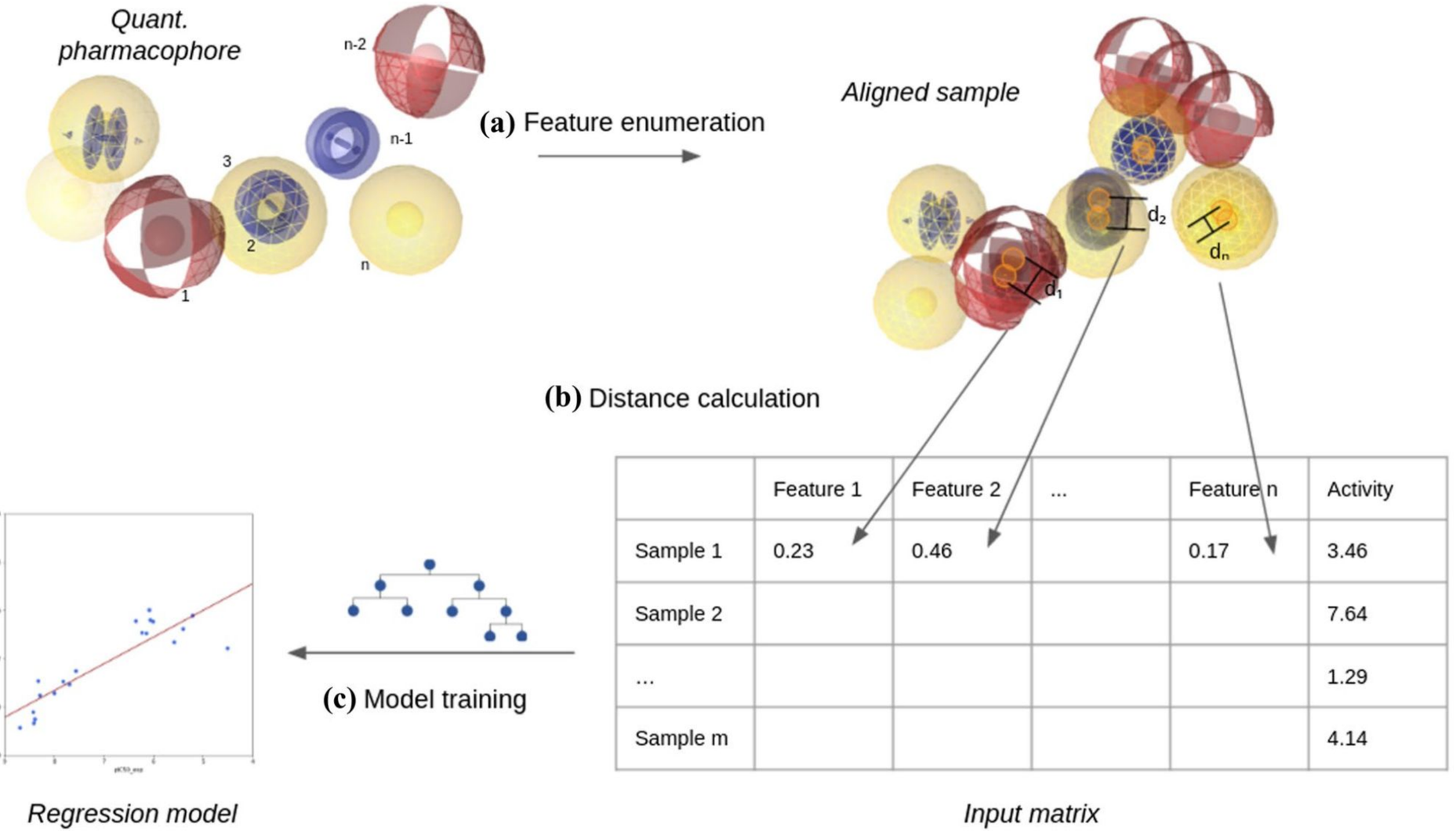
QPhAR Training Steps



Kohlbacher SM et al., J Cheminform 13, 57 (2021)
https://doi.org/10.1186/s13321-021-00537-9

https://github.com/StefanKohlbacher/QuantPharmacophore

QPhAR Model Generation



	Feature 1	Feature 2	...	Feature n	Activity
Sample 1	0.23	0.46		0.17	3.46
Sample 2					7.64
…					1.29
Sample m					4.14

https://github.com/StefanKohlbacher/QuantPharmacophore

pharmaceuticals

MDPI

Article

Applications of the Novel Quantitative Pharmacophore Activity Relationship Method QPhAR in Virtual Screening and Lead-Optimisation

Stefan Michael Kohlbacher, Matthias Schmid, Thomas Seidel * and Thierry Langer

Division of Pharmaceutical Chemistry, Department of Pharmaceutical Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria

* Correspondence: thomas.seidel@univie.ac.at; Tel.: +43-1-4277-55051

Abstract: Pharmacophores are an established concept for the modelling of ligand–receptor interactions based on the abstract representations of stereoelectronic molecular features. They became widely popular as filters for the fast virtual screening of large compound libraries. A lot of effort has been put into the development of sophisticated algorithms and strategies to increase the computational efficiency of the screening process. However, hardly any focus has been put on the development of automated procedures that optimise pharmacophores towards higher discriminatory power, which still has to be done manually by a human expert. In the age of machine learning, the researcher has become the decision-maker at the top level, outsourcing analysis tasks and recurrent work to advanced algorithms and automation workflows. Here, we propose an algorithm for the automated selection of features driving pharmacophore model quality using SAR information extracted from validated QPhAR models. By integrating the developed method into an end-to-end workflow, we present a fully automated method that is able to derive best-quality pharmacophores from a given input dataset. Finally, we show how the QPhAR-generated models can be used to guide the researcher with insights regarding (un-)favourable interactions for compounds of interest.

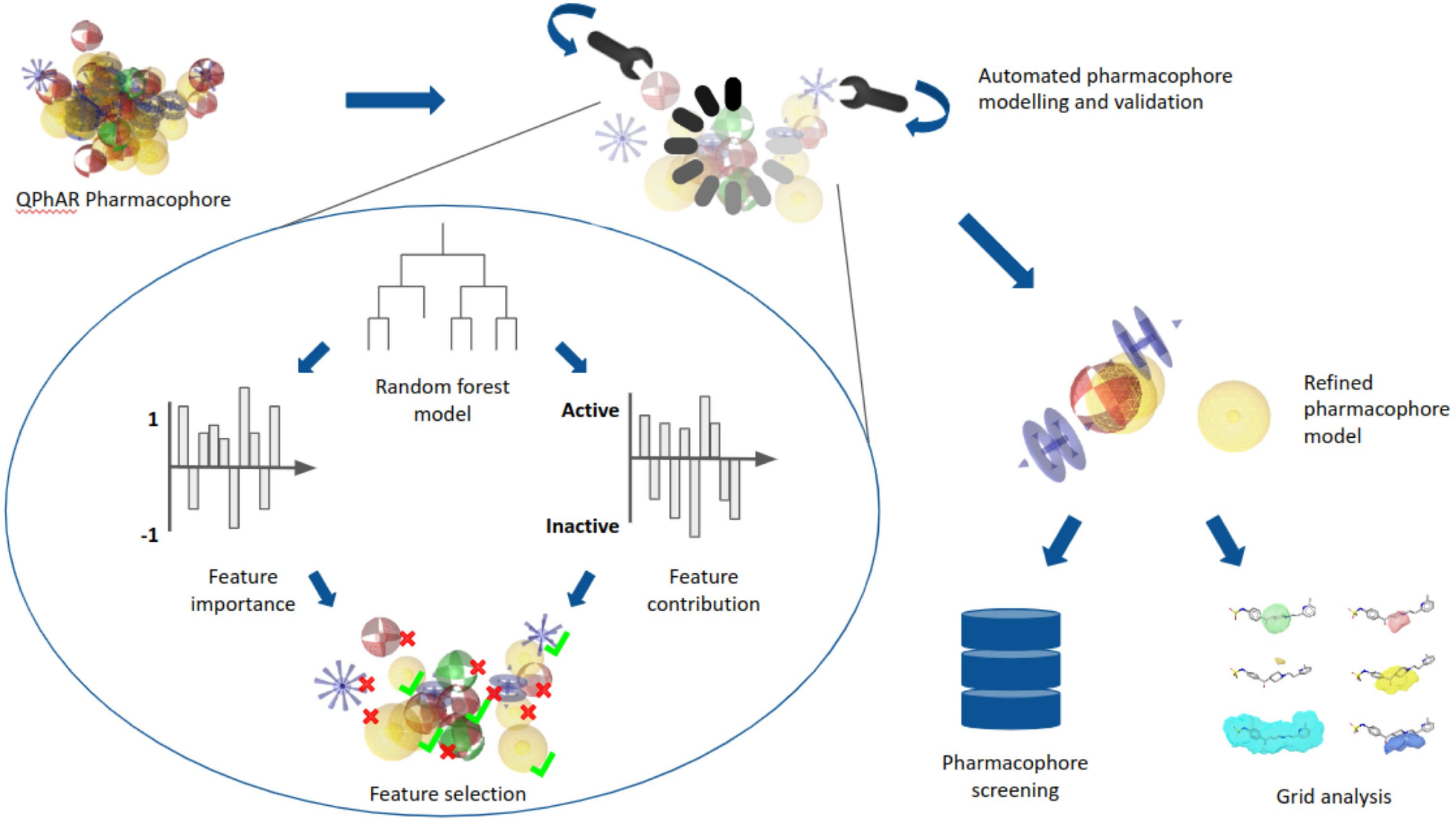
Keywords: pharmacophore; pharmacophore modelling; quantitative pharmacophore; QSAR; machine learning; pharmacophore optimisation; NeuroDeRisk

check for updates

Citation: Kohlbacher, S.M.; Schmid, M.; Seidel, T.; Langer, T. Applications of the Novel Quantitative Pharmacophore Activity Relationship Method QPhAR in Virtual Screening and Lead-Optimisation. *Pharmaceuticals* **2022**, *15*, 1122. https://doi.org/10.3390/ph15091122

https://doi.org/10.3390/ph15091122

Automated QPhAR Model Building



https://github.com/StefanKohlbacher/qphar-applications

Received: 13 October 2022 | Revised: 9 February 2023 | Accepted: 4 March 2023

DOI: 10.1002/minf.202200245

RESEARCH ARTICLE

molecular informatics

A new set of KNIME nodes implementing the QPhAR algorithm

Stefan M. Kohlbacher[1] | **Gökhan Ibis**[2] | **Christian Permann**[1,2] | **Sharon Bryant**[2] | **Thierry Langer**[1,2] | **Thomas Seidel**[1]

[1]Division of Pharmaceutical Chemistry, Department of Pharmaceutical Sciences, University of Vienna, Vienna, Austria

[2]Inte:Ligand GmbH, Vienna, Austria

Correspondence
Thomas Seidel, Division of Pharmaceutical Chemistry, Department of Pharmaceutical Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria.
Email: thomas.seidel@univie.ac.at

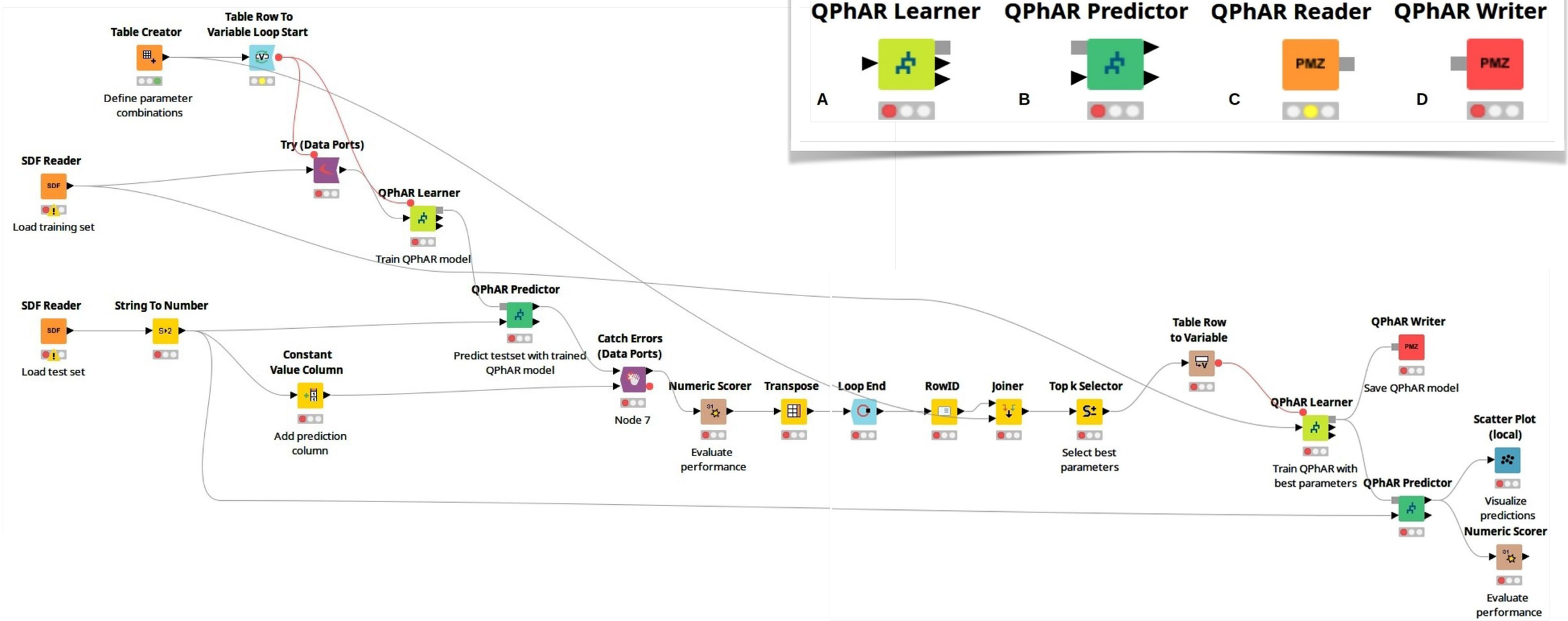
Funding information
NeuroDeRisk; Innovative Medicines Initiative 2 Joint Undertaking, Grant/Award Number: 821528

Abstract

Dissemination of novel research methods, especially in the form of chemo-informatics software, depends heavily on their ease of applicability for non-expert users with only a little or no programming skills and knowledge in computer science. Visual programming has become widely popular over the last few years, also enabling researchers without in-depth programming skills to develop tailored data processing pipelines using elements from a repository of predefined standard procedures. In this work, we present the development of a set of nodes for the KNIME platform implementing the QPhAR algorithm. We show how the developed KNIME nodes can be included in a typical workflow for biological activity prediction. Furthermore, we present best-practice guidelines that should be followed to obtain high-quality QPhAR models. Finally, we show a typical workflow to train and optimise a QPhAR model in KNIME for a set of given input compounds, applying the discussed best practices.

KEYWORDS
KNIME, NeuroDeRisk, pharmacophore modeling, pharmacophores, QPhAR

QPhAR in KNIME Platform



https://github.com/StefanKohlbacher/qphar-knime-nodes

QPhAR Application: Quantitative GABA-A Models

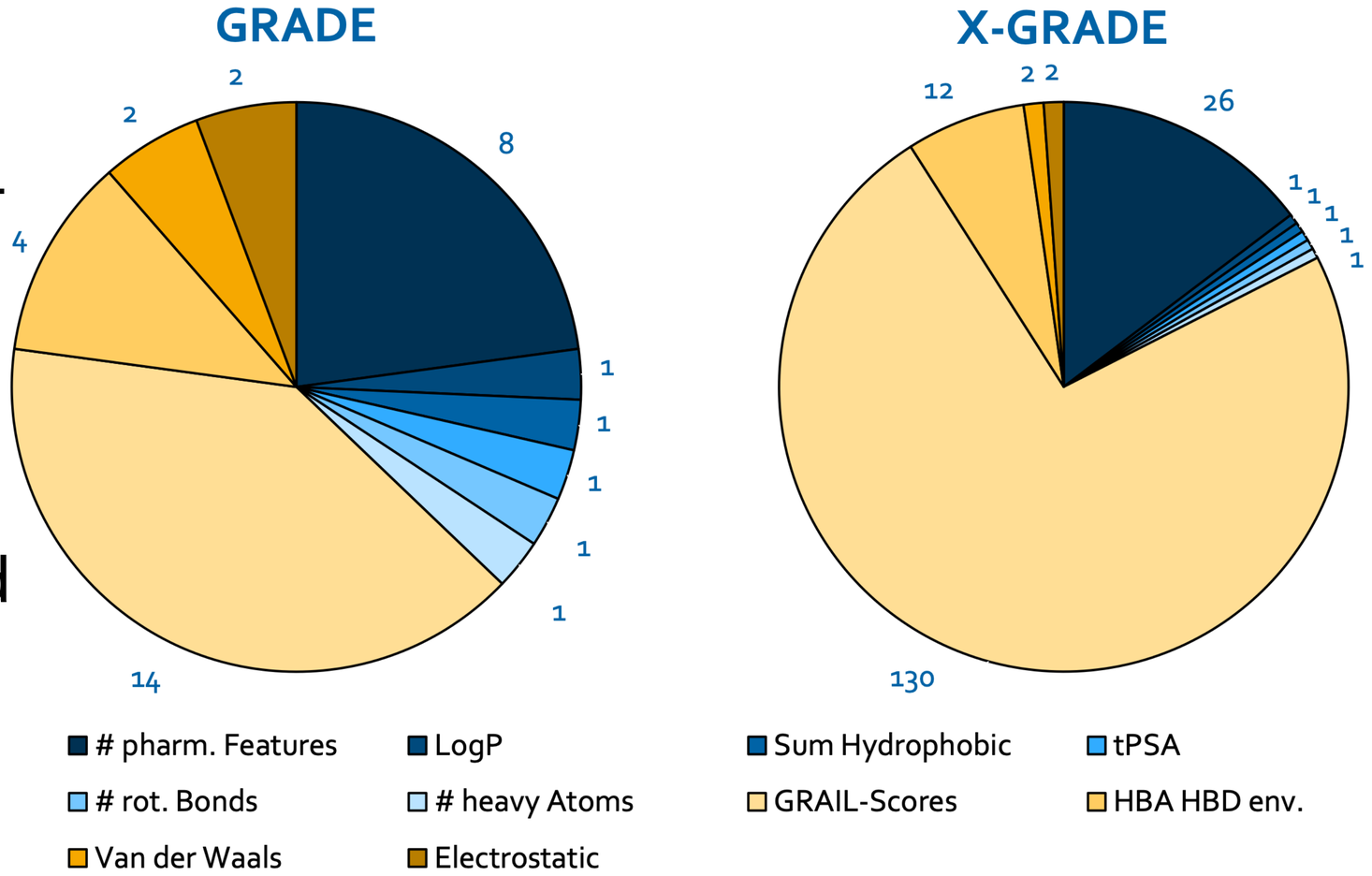
- Usage: Quantitative prediction of drug-induced seizure risks

Assay ID	Nr. Molecules	RMSE
CHEMBL1273617	44	0.51
CHEMBL1787625	45	0.97
CHEMBL3370250	34	0.99
CHEMBL3430052	32	0.26
CHEMBL676826	52	0.69
CHEMBL823916	48	0.62
CHEMBL824296	41	0.40

Kohlbacher SM, et al. (2023) https://doi.org/10.1002/minf.202200245

New Protein-Ligand Interaction Fingerprints

- GRADE & X-GRADE (eXtended GRAil-based DEscriptors)
- Contains defined set of floating point values rather than bit-string
- Useful as input for machine-learning models for QSAR and binding affinity estimation



Christian Fellinger

Fellinger C, et al., submitted

GRAIL Descriptors

JCTC Journal of Chemical Theory and Computation

Article

Cite This: *J. Chem. Theory Comput.* 2018, 14, 4958−4970

pubs.acs.org/JCTC

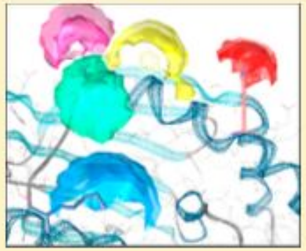
GRAIL: GRids of phArmacophore Interaction fieLds

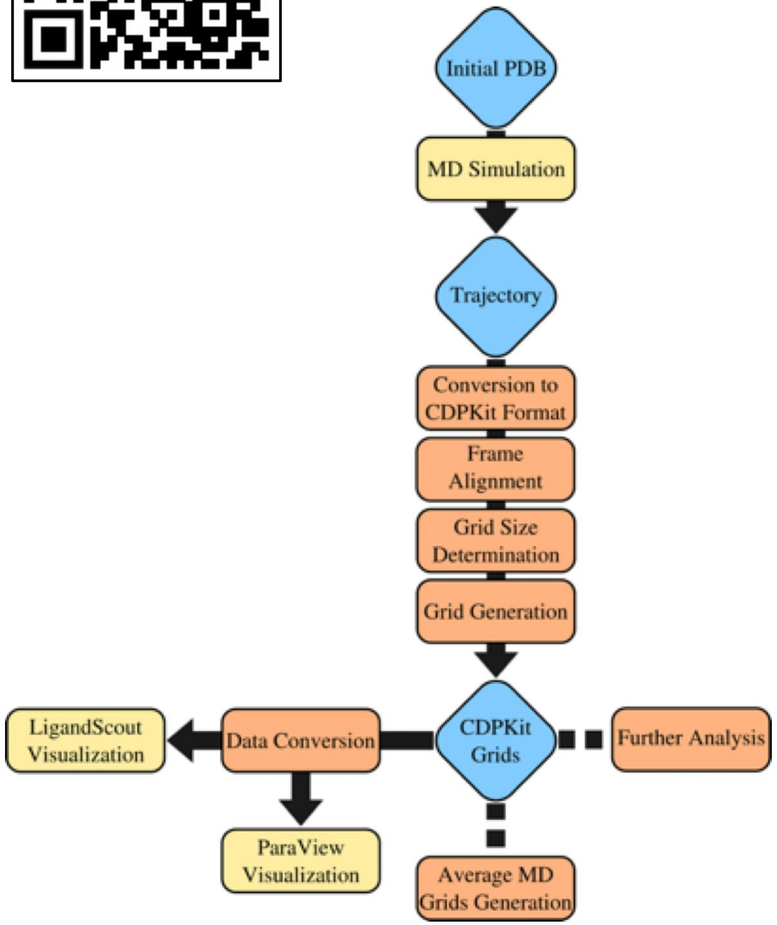
Doris A. Schuetz,[†] Thomas Seidel,[*,‡] Arthur Garon,[‡] Riccardo Martini,[†,‡] Markus Körbel,[‡,∥] Gerhard F. Ecker,[‡] and Thierry Langer[†,‡]

[†]Inte:Ligand GmbH, Mariahilferstrasse 74B/11, A-1070 Vienna, Austria

[‡]Department of Pharmaceutical Chemistry, University of Vienna, UZA 2, Althanstrasse 14, 1090 Vienna, Austria

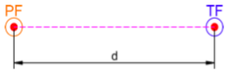
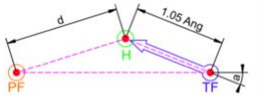
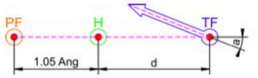
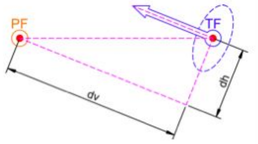
ABSTRACT: In the absence of experimentally derived, three-dimensional structures of receptors in complex with active ligands, it is of high value to be able to gain knowledge about energetically favorable interaction sites solely from the structure of the receptor binding site. For de novo ligand design as well as for lead optimization, this information retrieved from the protein is inevitable. The herein presented method called GRAIL combines the advantages of traditional grid-based approaches for the identification of interaction sites and the power of the pharmacophore concept. A reduced pharmacophoric abstraction of the target system enables the computation of all relevant interaction grid maps in short amounts of time. This allows one to extend the utility of a grid-based method for the analysis of large amounts of coordinate sets obtained by long-time MD simulations. In this way it is possible to assess conformation dependent characteristics of key interactions over time. Furthermore, conformational changes of the protein can be taken into account easily and information thus obtained well-guides a rational ligand design process. A study employing MD trajectories of the oncology target heat shock protein 90 showcases how well our novel approach GRAIL performs for a set of different inhibitors bound to their target protein and how molecular features of the inhibitors are subject to optimization.



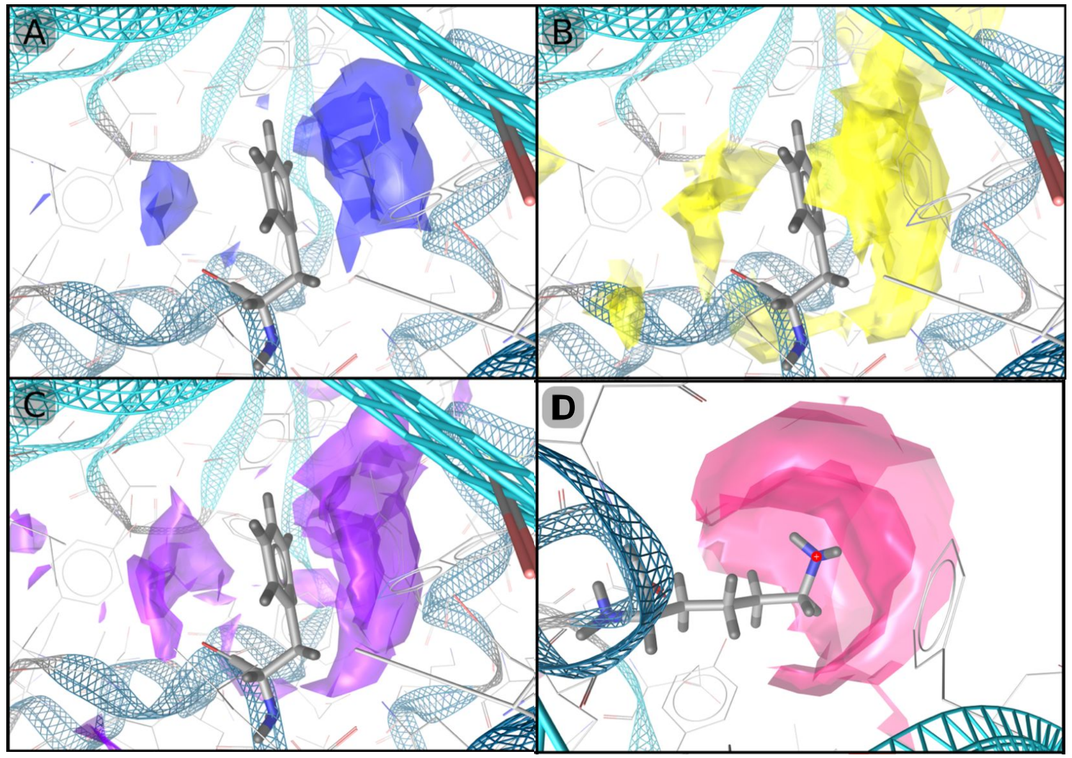


GRAIL Scoring Function

- Grid is defined within the protein binding site, no ligand needed
- At each grid point: Pharmacophore feature probability calculated
- For every interaction type: Optimum distances & angles defined
- Extremely fast, amendable for the analysis of entire MD trajectories
- Script available for CDPKit

Complementary Feature Pair			Interaction Geometry	Distance Scoring Function/ Distance Range	Angle Scoring Function/ Angle Range
Probe Feature Type	Target Feature				
	Type	Geometry			
H	H	Point		*GBF(d)* d = 2.0 - 6.0	-
NI	PI	Point		*GBF(d)* d = 1.5 - 5.5	-
PI	NI	Point		*GBF(d)* d = 1.5 - 5.5	-
AR	PI	Point		*GBF(d)* d = 3.5 - 5.5	-
PI	AR	Plane		*GBF(d)* d = 3.5 - 5.5	*GBF(a)* a = -60° - 60°
HBA	HBD	Vector		*GBF(d)* d = 1.2 - 2.8	*GBF(a)* a = -50° - 50°
HBD	HBA	Vector		*GBF(d)* d = 1.2 - 2.8	*GBF(a)* a = -85° - 85°
AR	AR	Plane		*GBF(d_v)*GBF(d_h)* d_v = 3.5 - 6.0 d_h = 0.0 - 2.8	-

Visualisations: HSP90 Case Study



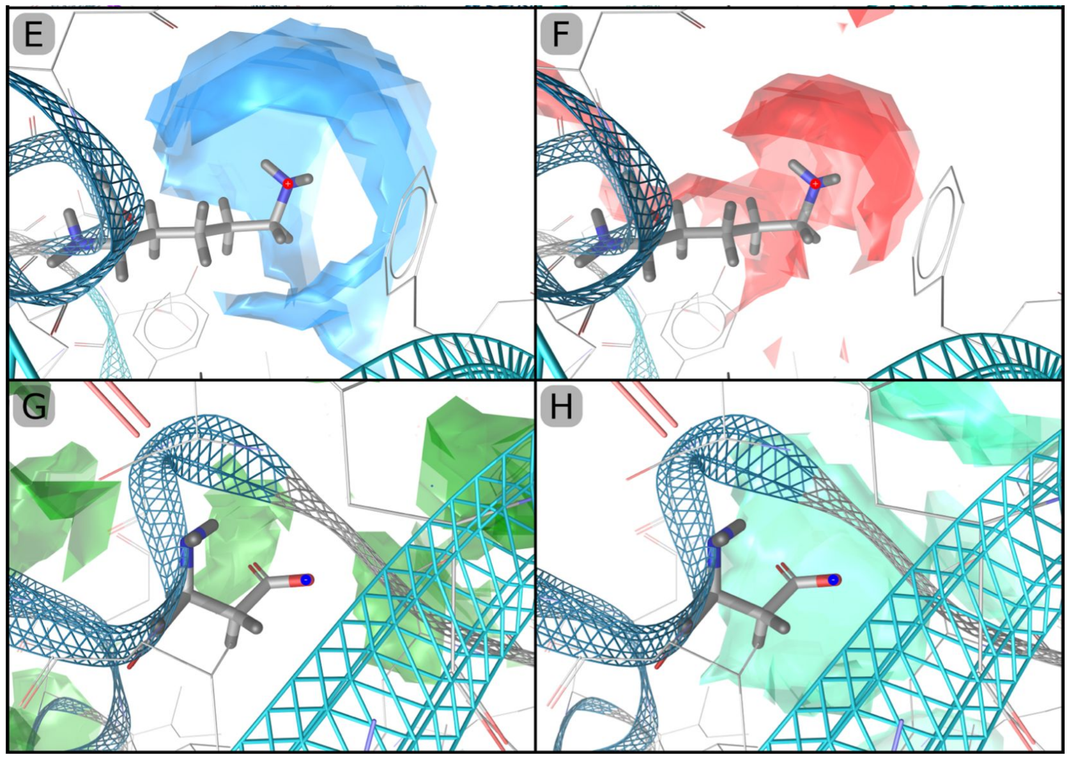
Residue Phe138 interactions:

(A) … aromatic - aromatic

(B) … hydrophobic - hydrophobic

(C) … positiv charge - aromatic

(D) … negative - positive charge



Residue Lys100 and Asp 85 interactions:

(E) … aromatic - positive charge

(F) … HBA - HBD

(G) … HBD -> HBA

(H) … hydrophobic - hydrophobic

GRAIL Use in Lead Optimisation

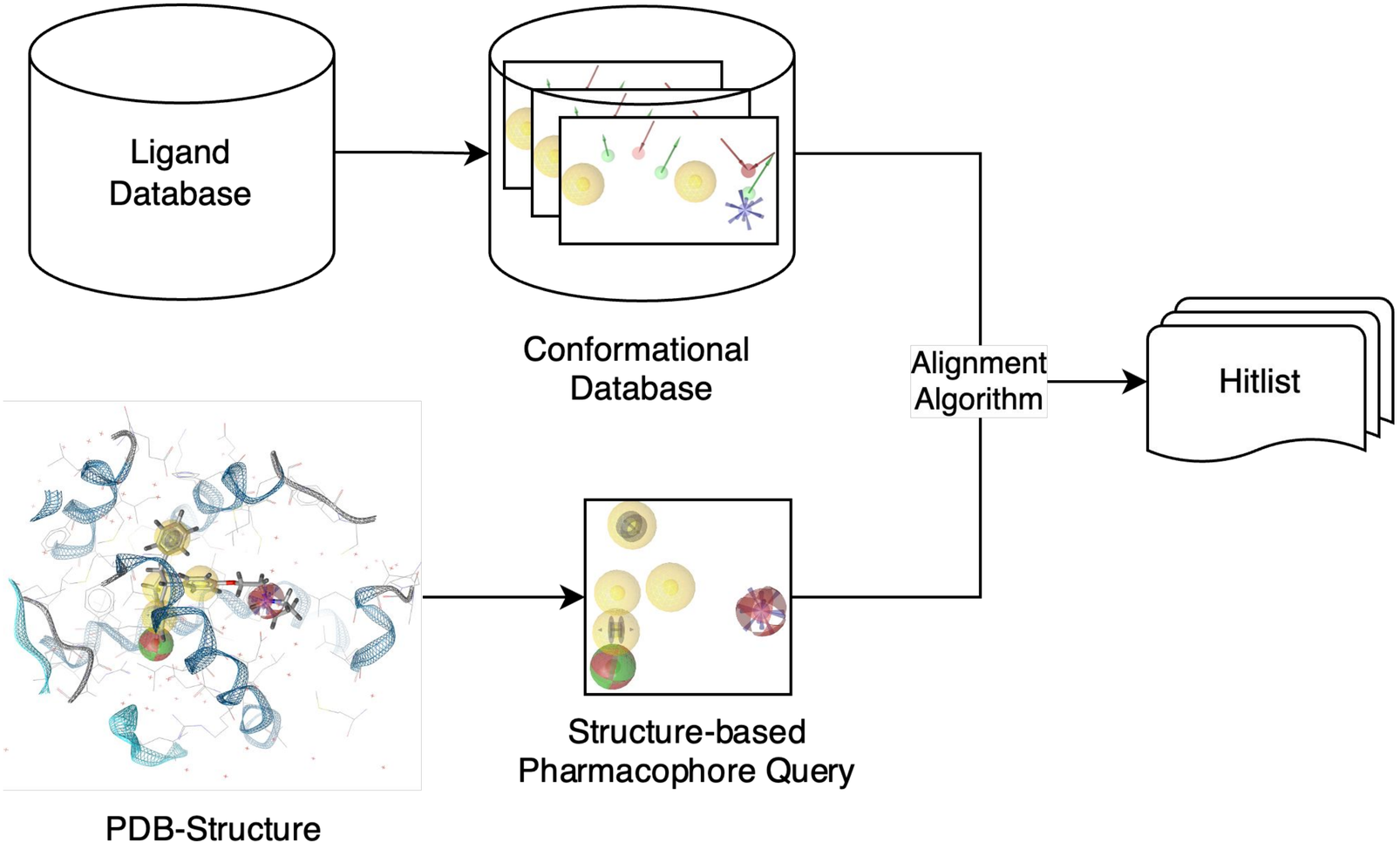
- Easy understandable medicinal chemistry design guidance provided
- Focus on specific regions
  - e.g. replacing entropically disfavoured water molecules with small hydrophobic substituent (“magic methyl positioning”)
- Pharmacophore hotspot feature frequency analysis
  - for prioritising replacement/modifications of molecular substructures
  - providing interaction preference guidance
  - easily adaptable for generative de novo design

Back to the haystack ...

- Is there a way to use high quality pharmacophore information but avoid conformer generation and 3D alignment ?

Next Generation Descriptors for P4 Modeling

- Virtual screening using traditional P4 matching

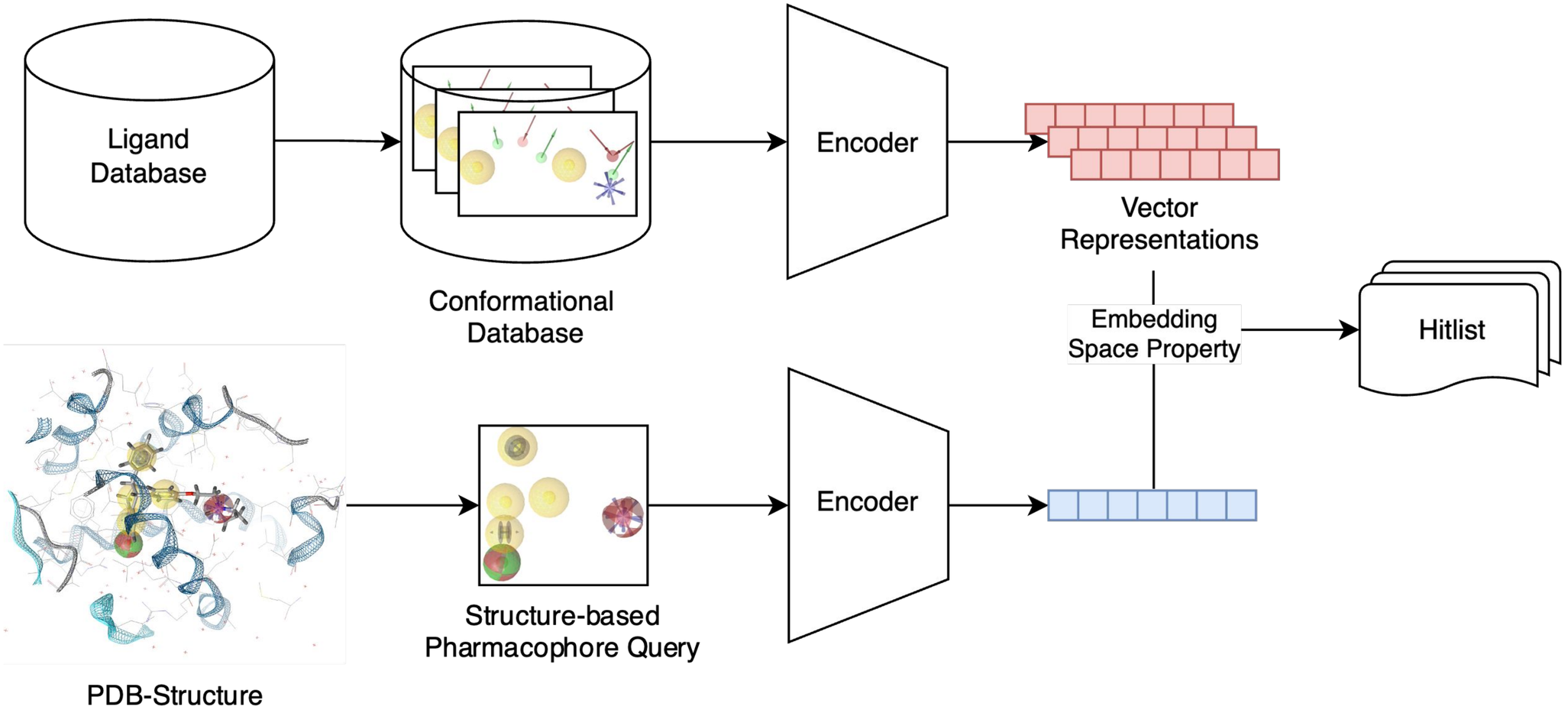


Daniel Rose

New Concept: PharmacoMatch

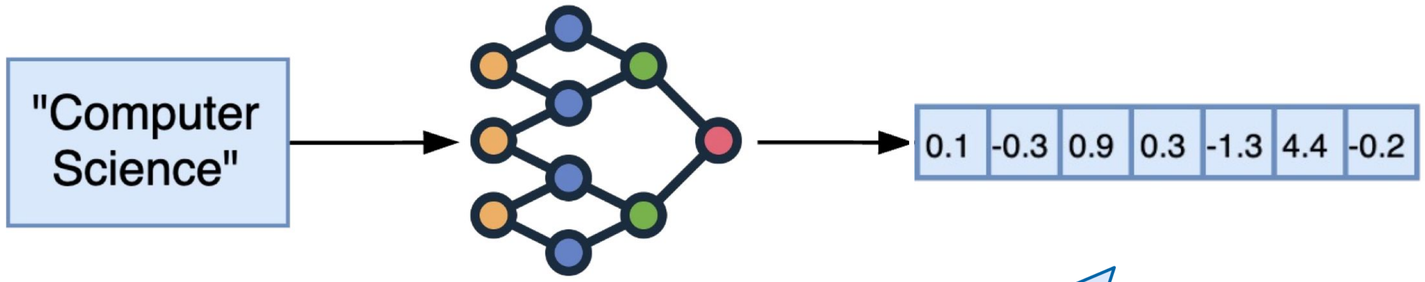
- Contrastive learning framework maps pharmacophore model vectors into an order embedding space -> virtual screening possible, avoiding 3D alignment

Daniel Rose



Rose D. et al., PharmacoMatch, accepted [https://doi.org/10.48550/arXiv.2409.06316]

Vector Embedding Space

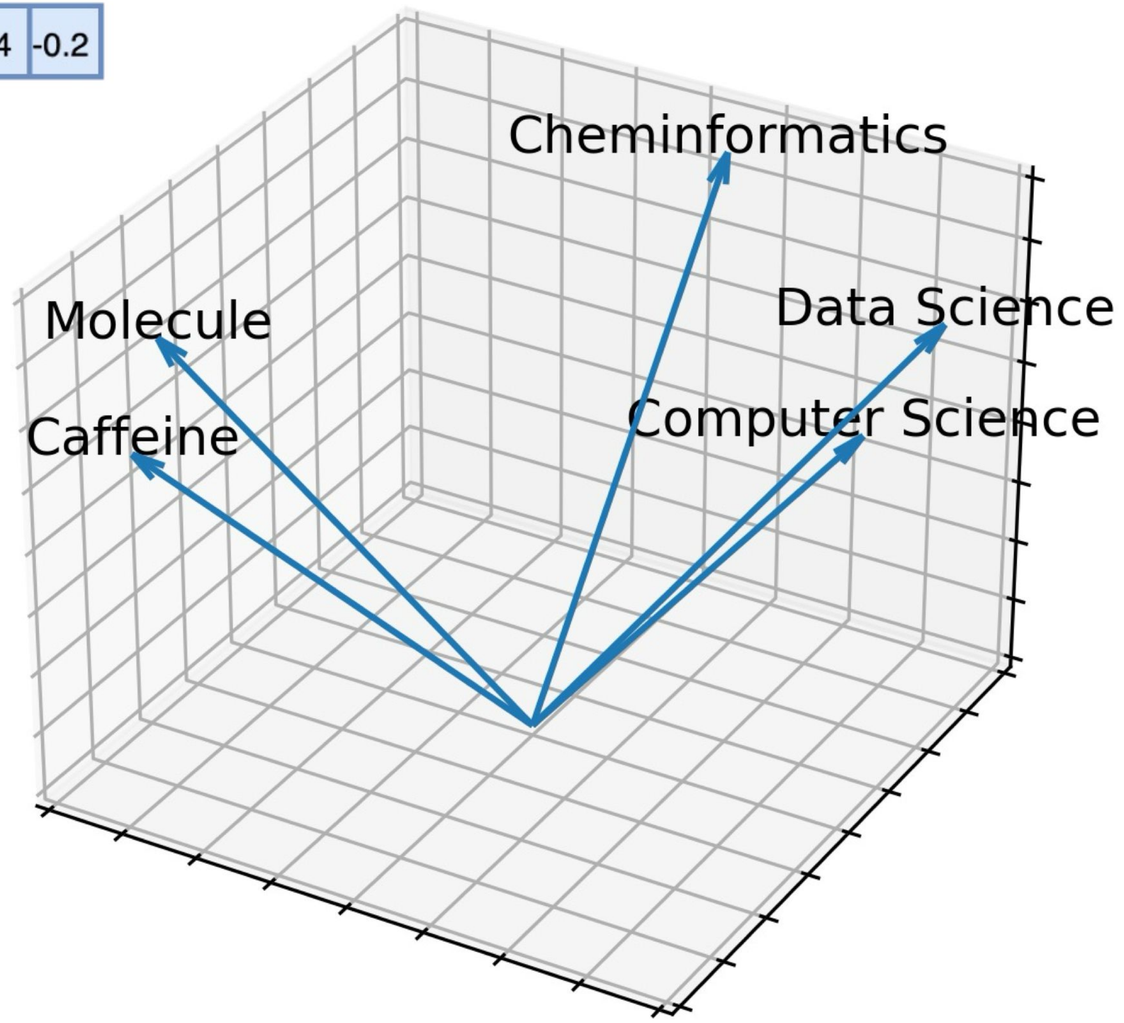


Can be used in…

… Search Engines

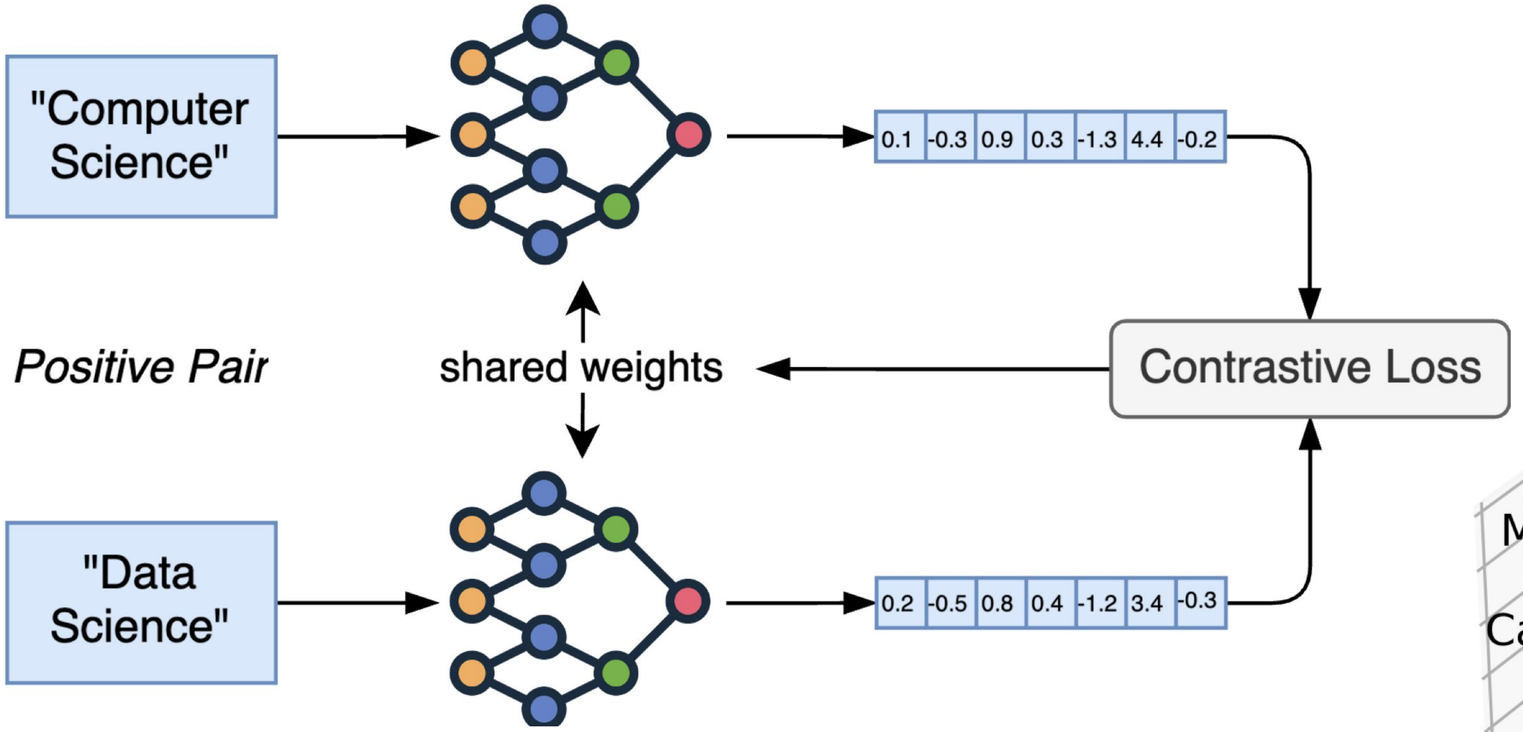
… Recommender Systems

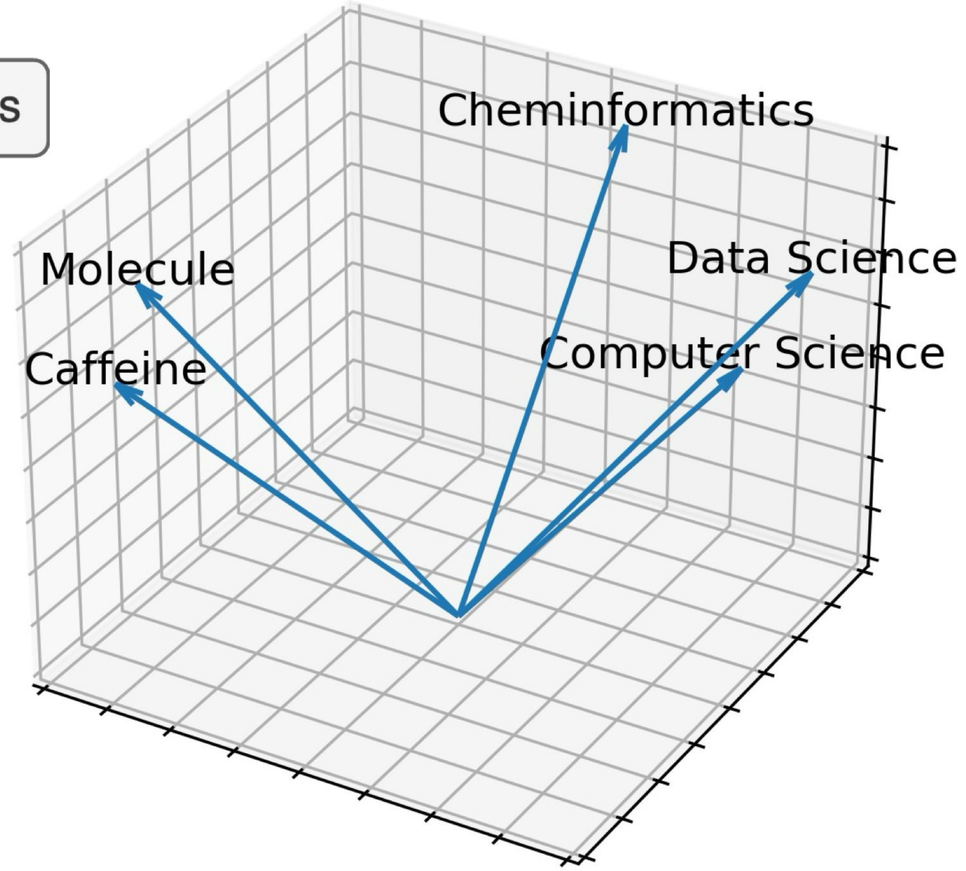
… Chatbots



- RECS: Retrieval of Concepts in Databases

Embedding Model

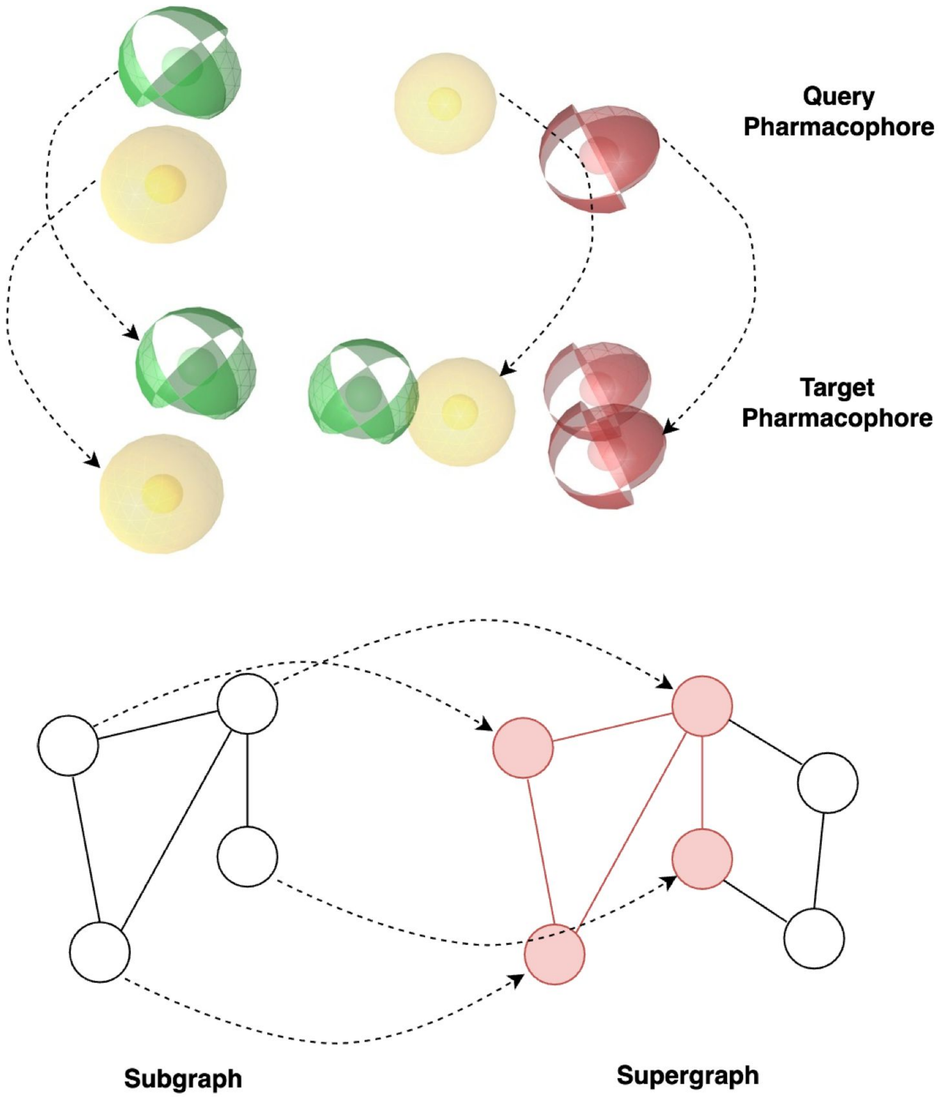


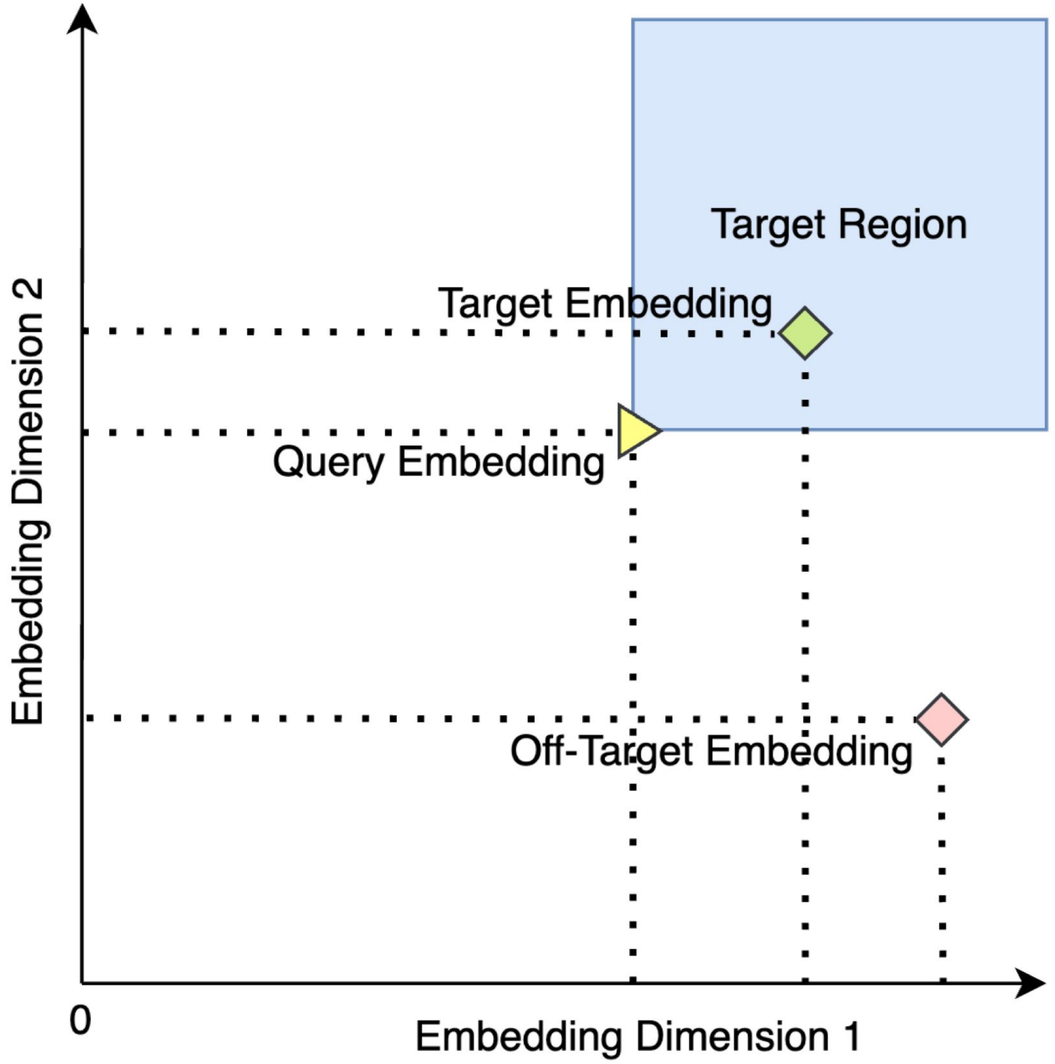


What shall we compare to train a

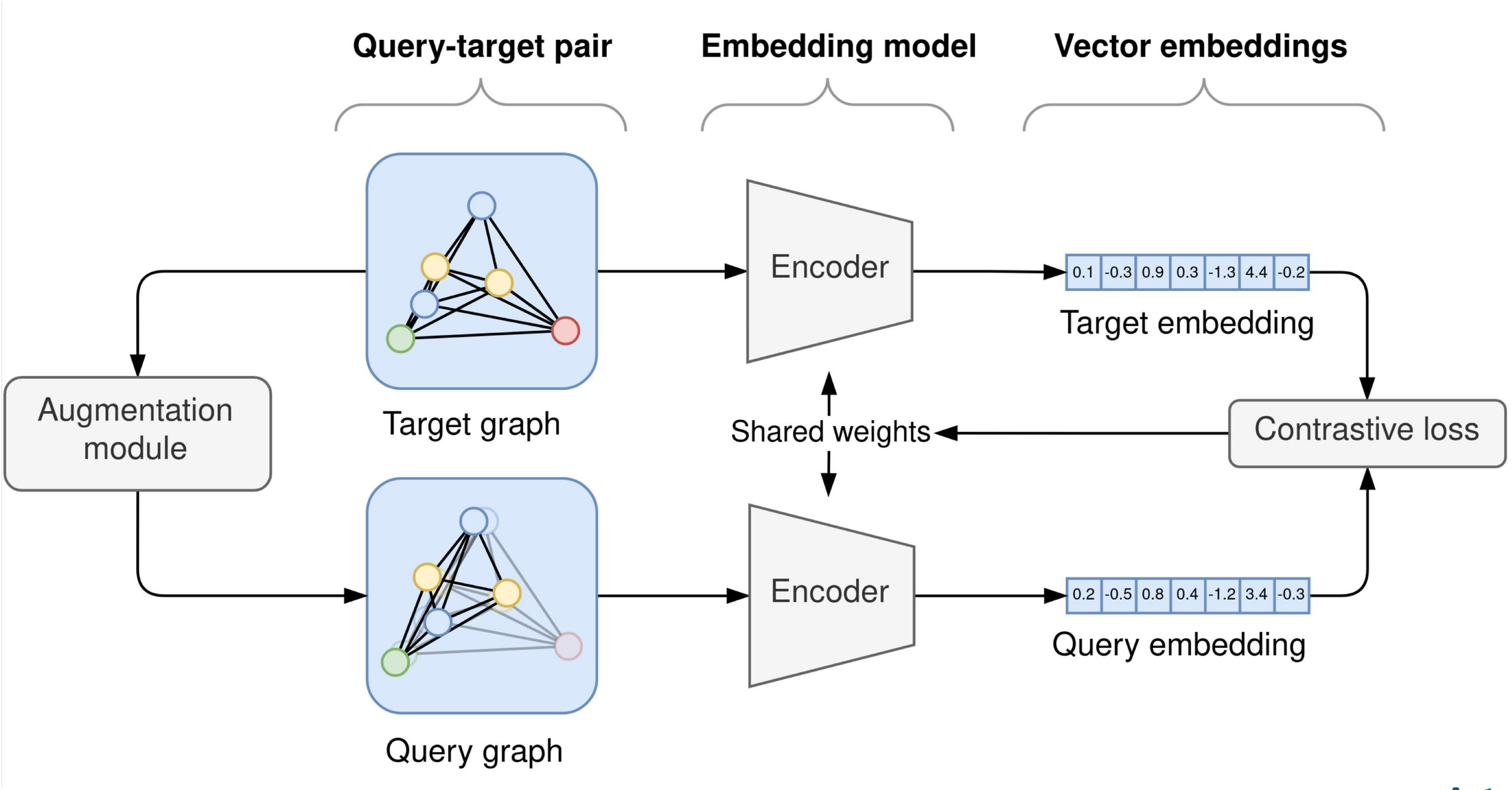
Pharmacophore Embedding Model?

Subgraph Isomorphism & Order Embedding





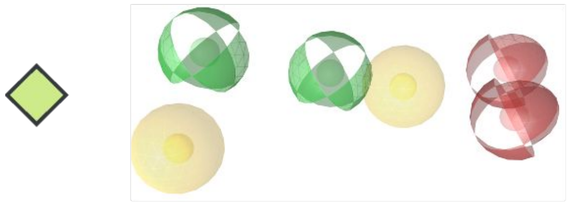
P4 Vector Embedding Model



Virtual Screening

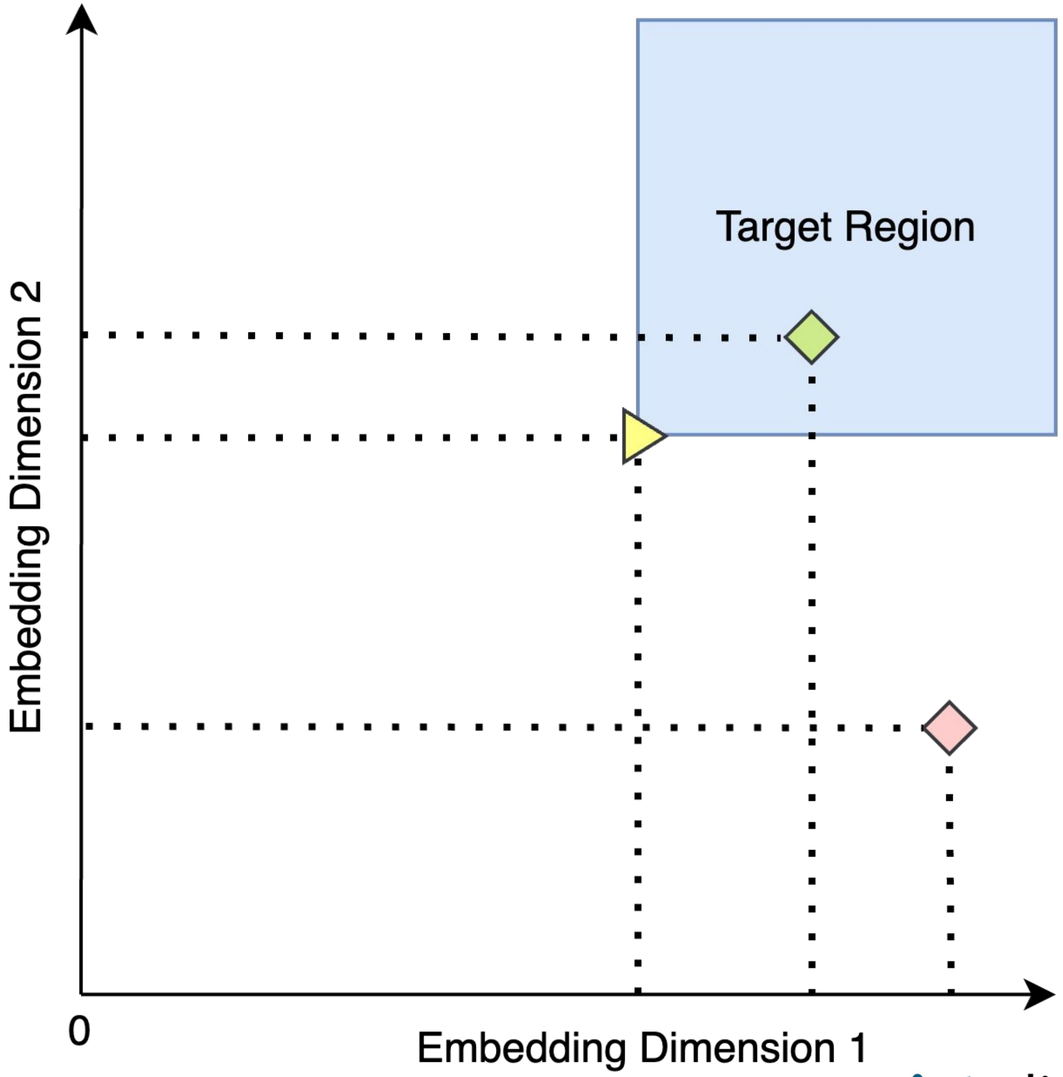
- Neural subgraph matching

Subgraph



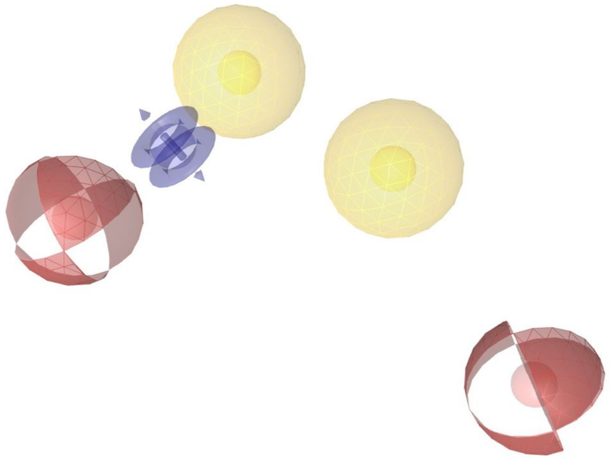
Supergraph

Other Graph

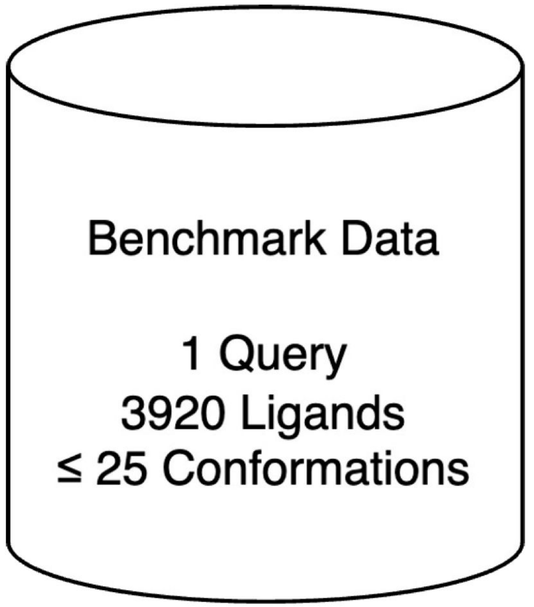


Case Study: Preliminary Results

Arbitrary Query Pharmacophore



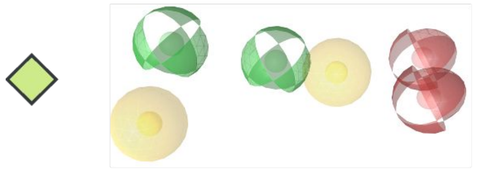
Drug-like Ligand Database



Tran-Nguyen VK, Jacquemard C, Rognan D. (2020). JCIM, 60(9), 4263. doi: 10.1021/acs.jcim.0c00155

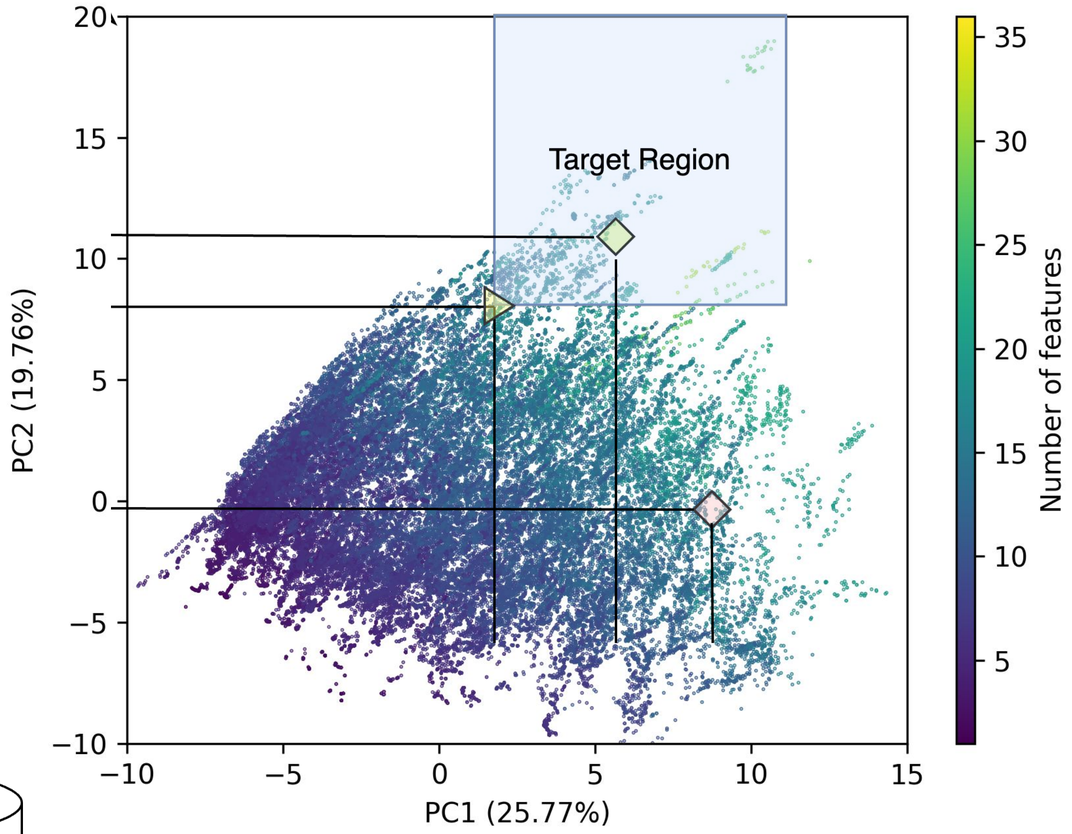
Case Study: Preliminary Results

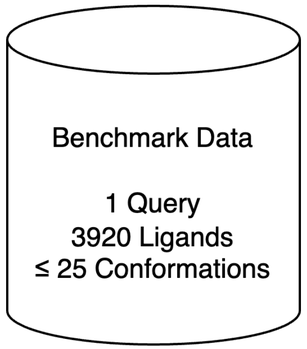
Subgraph



Supergraph

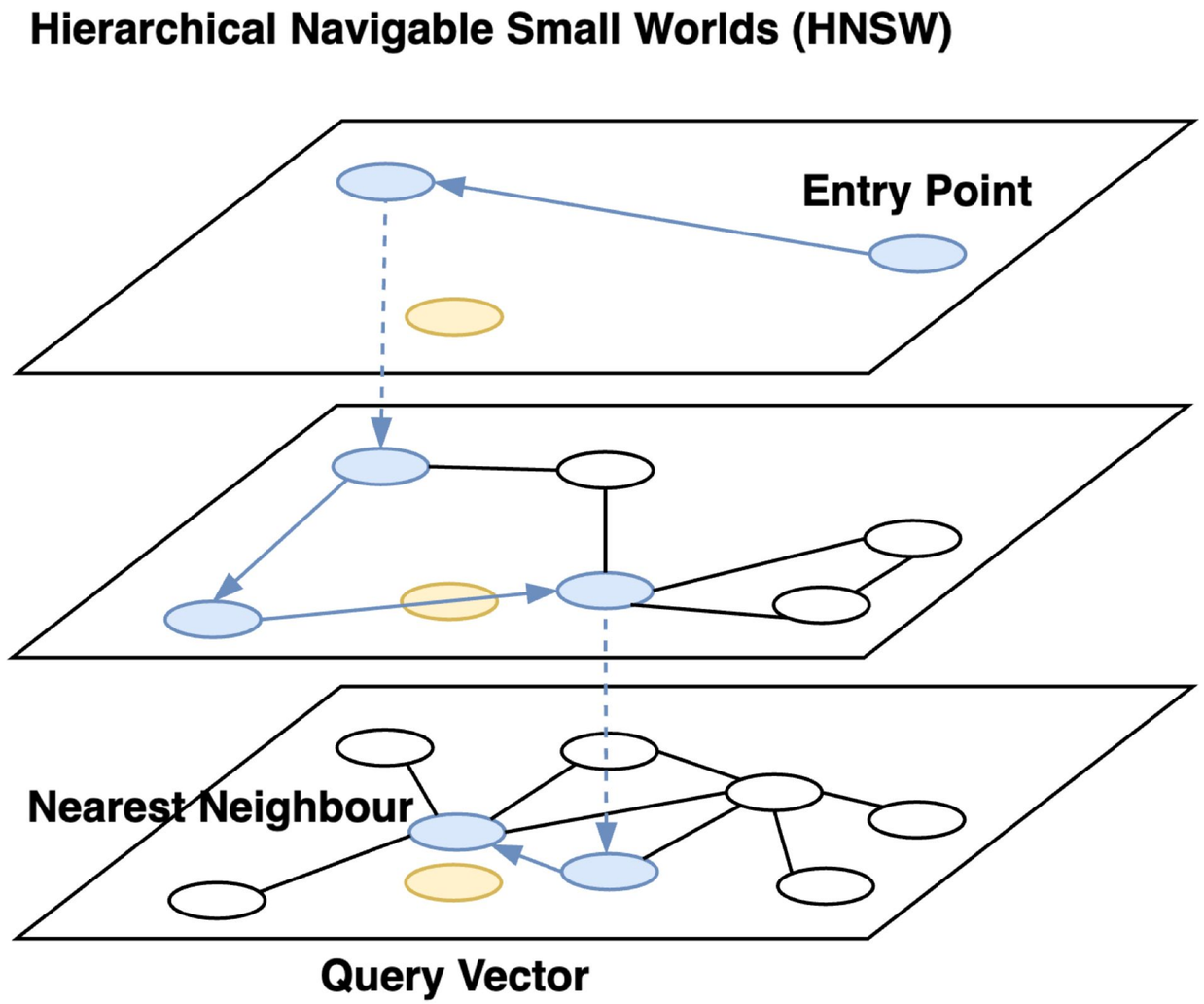
Other Graph





	CDPKit	PhectorMatch
On Disk Space	50 MB	50 MB + 20 MB
Screening Time	2.0 s	0.02 s

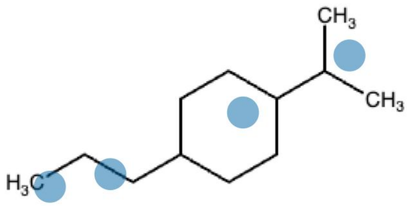
Potential for Faster DB Search

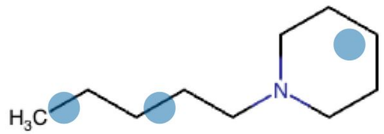


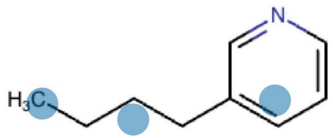
- Vector database for faster retrieval
- Inherent tree structure could be exploited

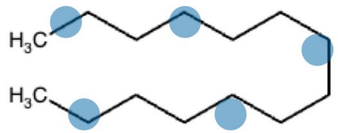
Another Application for P4 Vectors

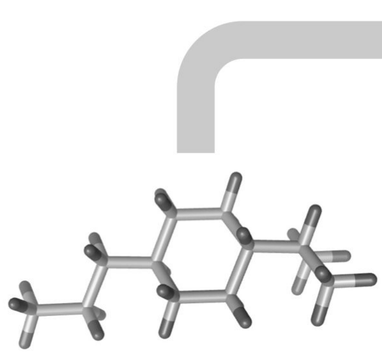
Enhanced Hydrophobic Perception

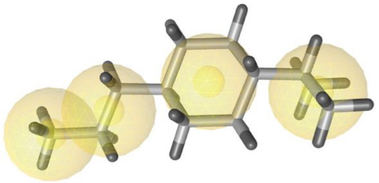




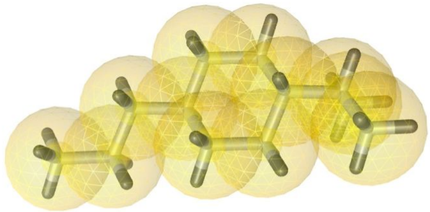








Classical Hydrophobic Features



Atom-typed Hydrophobic Features

Point-Cloud Encoder

Vector Embedding

Cheminformatics Approach:
- Gaussian Shape Description
- Gaussian Shape Alignment

ML Approach:
- Point Cloud Encoder
- Order Embedding Alignment

Example molecules adopted from: Wolber G et al., ,Drug Discov. Today, vol. 13, no. 1-2, pp. 23–29, 2008.

NEURODERISK

DERISKING NEUROTOXICITY

universität wien

The NeuroDeRisk project has received funding from the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under grant agreement No 821528.

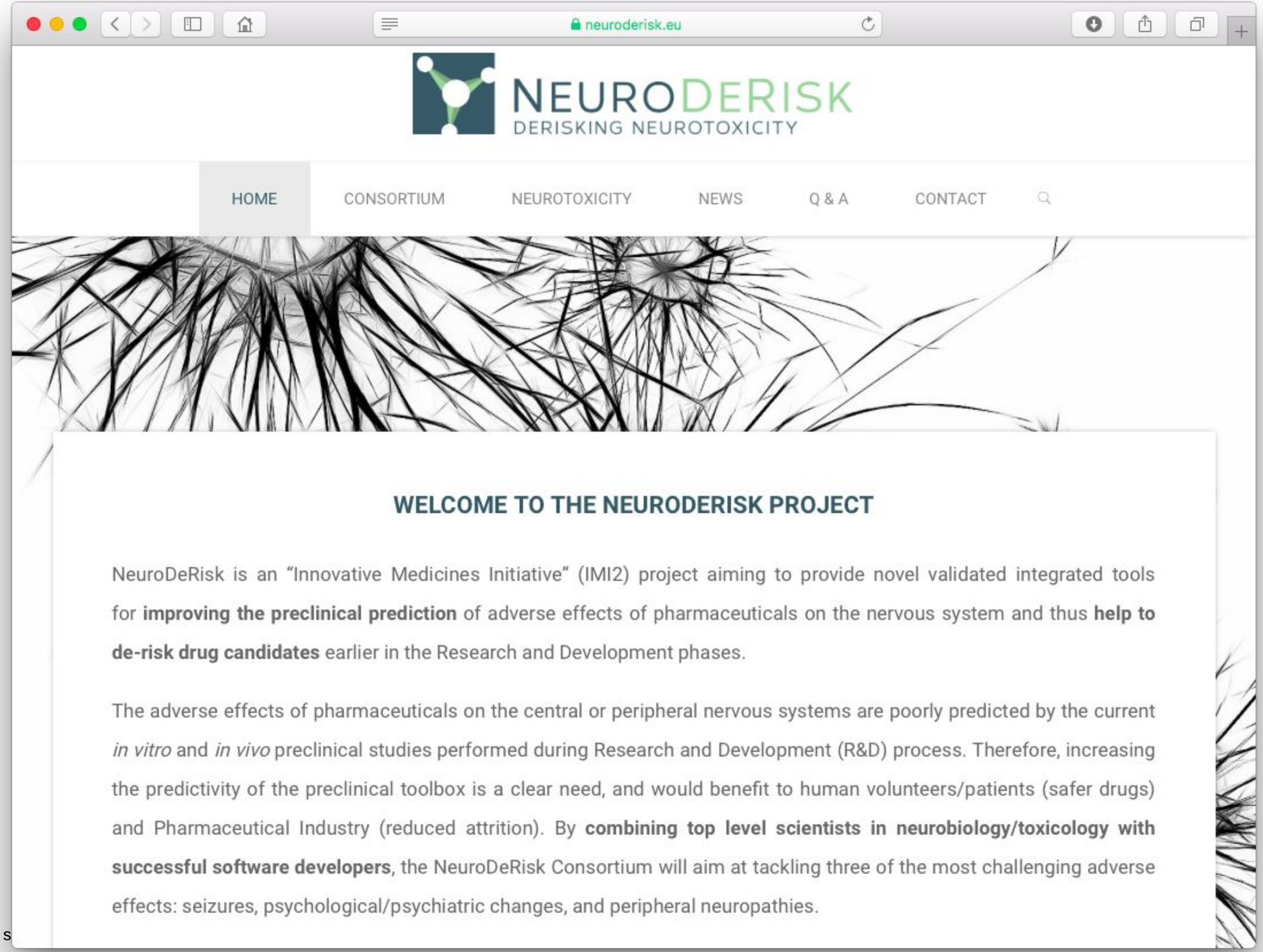
This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

https://neuroderisk.eu

imi innovative medicines initiative

efpia

inte:ligand www.inteligand.com Vienna, Austria


neuroderisk.eu
NEURODERISK
DERISKING NEUROTOXICITY
HOME
CONSORTIUM
NEUROTOXICITY
NEWS
Q & A
CONTACT
WELCOME TO THE NEURODERISK PROJECT
NeuroDeRisk is an “Innovative Medicines Initiative” (IMI2) project aiming to provide novel validated integrated tools for **improving the preclinical prediction** of adverse effects of pharmaceuticals on the nervous system and thus **help to de-risk drug candidates** earlier in the Research and Development phases.
The adverse effects of pharmaceuticals on the central or peripheral nervous systems are poorly predicted by the current *in vitro* and *in vivo* preclinical studies performed during Research and Development (R&D) process. Therefore, increasing the predictivity of the preclinical toolbox is a clear need, and would benefit to human volunteers/patients (safer drugs) and Pharmaceutical Industry (reduced attrition). By **combining top level scientists in neurobiology/toxicology with successful software developers**, the NeuroDeRisk Consortium will aim at tackling three of the most challenging adverse effects: seizures, psychological/psychiatric changes, and peripheral neuropathies.

De-risking Neurotoxicity

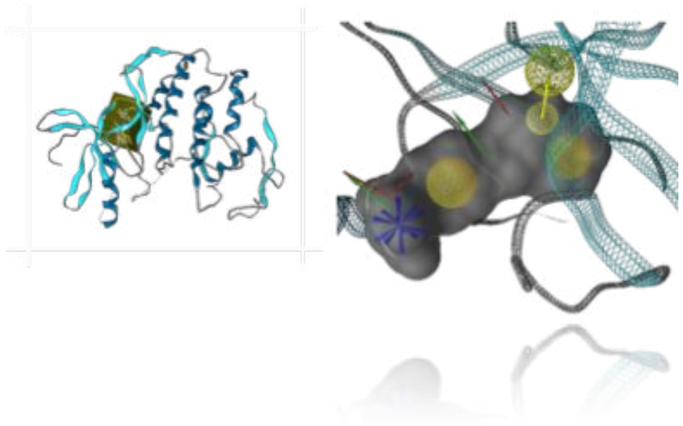
Target Based Approach

- Seizure Risk / Pharmacology
- GABA-A Antagonist
- GABA-A NAM
- GABA-A Channel Blocker
- GABA-A PAM
- GABA-A Agonist
- GABA-A Neurosteroid

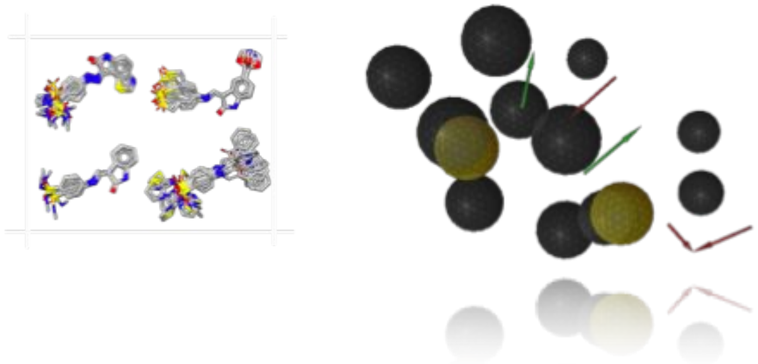
Outcome Based Approach

- Suicidal Ideation
- Drugs associated with reported outcomes (pharmacovigilance)
- RNA Editing (Alcediag)

Structure-Based Modeling

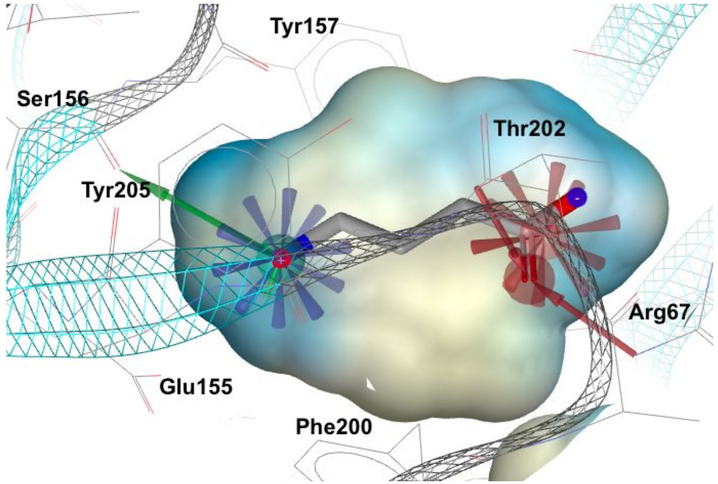


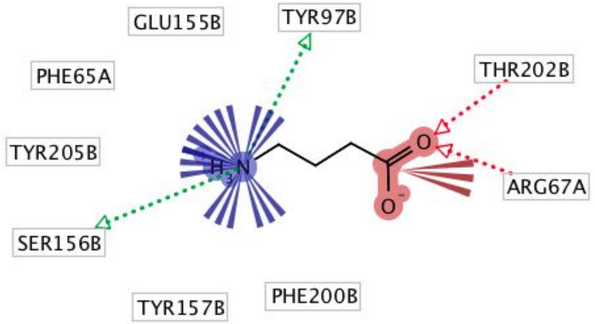
Ligand-Based Modeling



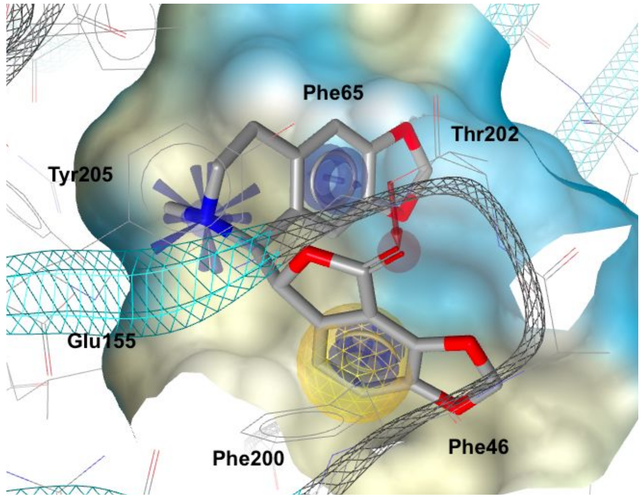
De-risking Neurotoxicity

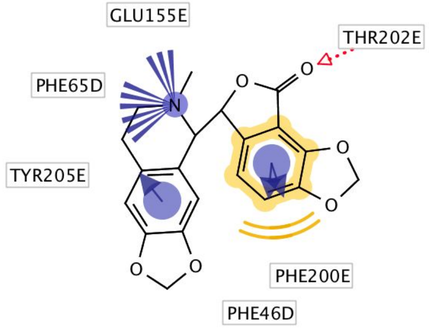
GABA-A orthosteric site, channel SB-modeling examples



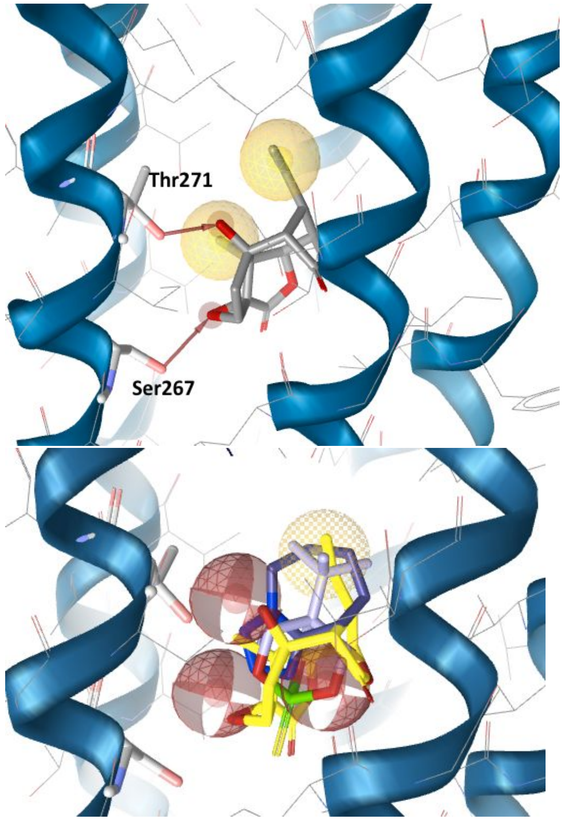


GABA-GS-Agonist
(PDB ID:6huj)





Bicuculline-GS-Antagonist
(PDB ID:6huk)

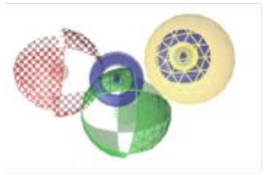


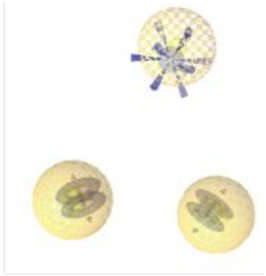
Picrotoxinin-Channel
(PDB ID:6huj)

De-risking Neurotoxicity

Can we identify chemical features in 3D-space associated with suicidal ADEs?

- Outcome Based Approach – No Target
- Suicidal Ideation (5 terms)
- 1492 drugs with suicidal annotations from FAERS, Meta ADEDB and NIH databases (pharmacovigilance)
- Clustering; > 45 LB models created and tested
- Models also generated using confidential experimental data from Alcediag
- Editox – unambiguous IFNa like RNA editing profiles (Alcediag)





Name	#	Matching Features
Bifeprunox	2	
Hydralazine	1	
Reserpine	3	
Rimonabant	1	
Ketoconazole	5	
Sertindole	3	
Taranabant	4	
Aripiprazole	2	
Imipramine	4	
Clomipramine	5	
Diphenhydramine	1	
Fluoxetine	2	
Nortriptyline	3	

Promising results with the IFNα like RNA editing (Editox) datasets

Van der Laan, S, et. al., (2017) Emerging RNA editing biomarkers will foster drug development. Drug Discovery Today, 22(7), 1056. doi:10.1016/j.drudis.2017.01.017

De-risking Neurotoxicity

- 60 3D-pharmacophore models have so far been incorporated and deployed in the NeuroDeRisk in silico toolbox and more are currently developed

5 AMPA

- NDR-IL-AMPA-Agonist-LB
- NDR-IL-AMPA-Antag-Fanapel-6ruq
- NDR-IL-AMPA-Kainate-4u2q-a
- NDR-IL-AMPA-PAM-HI-LB-2
- NDR-IL-AMPA-PAM-Thiazides-LB

6 NMDA N2A

- NDR-IL-NMDA-Agonist-Glu-7eu7
- NDR-IL-NMDA-Agonist-Gly-7eu7
- NDR-IL-NMDA-Antag-Glu-LB
- NDR-IL-NMDA-Antag-Gly-1pbq
- NDR-IL-NMDA-Antag-Gly-LB
- NDR-IL-NMDA-Channel-LB-2

7 Suicidality

- NDR-IL-Suicidality-2
- NDR-IL-Suicidality-3
- NDR-IL-Suicidality-3v1
- NDR-IL-Suicidality-4
- NDR-IL-Suicidality-5
- NDR-IL-Suicidality-SE-ed-1
- NDR-IL-Suicidality-SE-sd-6

23 GABA-AR

- NDR-IL-GABA-A-Barbiturate-LB
- NDR-IL-GABA-A-Channel-LB-6
- NDR-IL-GABA-A-gs-Agonist-6huj-4
- NDR-IL-GABA-A-gs-Agonist-LB
- NDR-IL-GABA-A-gs-Antag-6huk-3
- NDR-IL-GABA-A-gs-Antag-LB
- NDR-IL-GABA-A-NAM-Flumazenil-6d6t
- NDR-IL-GABA-A-NAM-Flumazenil-LB
- NDR-IL-GABA-A-NSteroid-Pregnanolone-5O8f
- NDR-IL-GABA-A-PAM-BDZ-LB-3
- NDR-IL-GABA-A-PAM-Diazepam-6hup-2
- NDR-IL-GABA-A-Z-drug-LB
- NDR-IL-GABA-A-a1-BDZ-site-EFPIA-C24
- NDR-IL-GABA-A-a1-BDZ-site-EFPIA-C38
- NDR-IL-GABA-A-a1-BDZ-site-EFPIA-C42
- NDR-IL-GABA-A-a1-BDZ-site-EFPIA-Cmix-2
- NDR-IL-GABA-A-a2-Flumazenil-site-EFPIA-C8-2
- NDR-IL-GABA-A-a3-Flumazenil-site-EFPIA-C61
- NDR-IL-GABA-A-a3-Flumazenil-site-EFPIA-C74
- NDR-IL-GABA-A-a4-Ro154513-site-EFPIA-C5
- NDR-IL-GABA-A-a5-Flumazenil-site-EFPIA-C47
- NDR-IL-GABA-A-a5-Flumazenil-site-EFPIA-C51
- NDR-IL-GABA-A-a6-Ro154513-site-EFPIA

4 GlyRa3

- NDR-IL-GlyRa3-Antag-LB
- NDR-IL-GlyRa3-Antag-Strychnine-5cfb-d3
- NDR-IL-GlyRa3-Channel-LB
- NDR-IL-GlyRa3-os-Agonist-5vdh-c

3 BBB Transporter

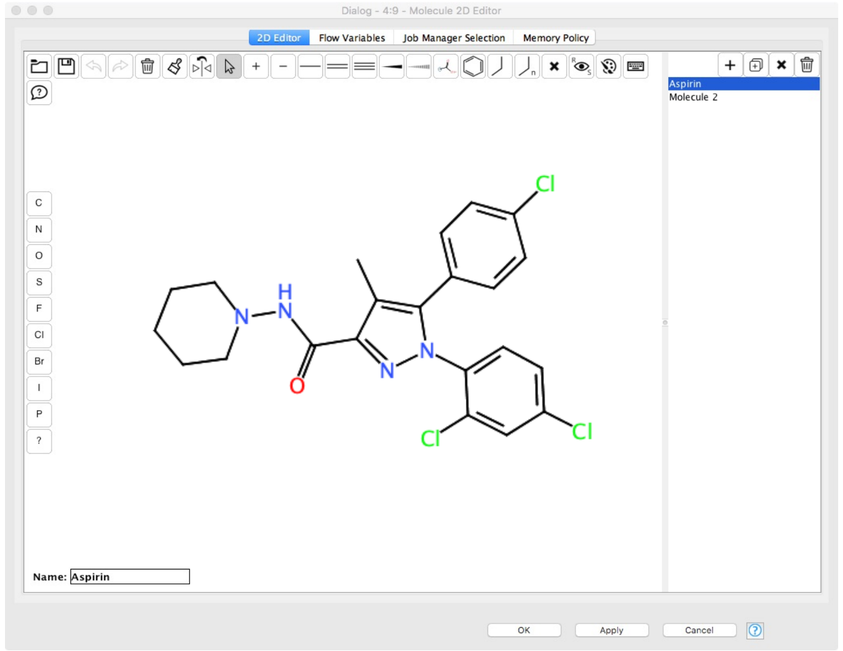
- NDR-IL-BBB-H+CO-Antiporter-C13C11-LB-2
- NDR-IL-BBB-OATP1A2-C18-LB
- NDR-IL-BBB-OATP1A2-C19-LB

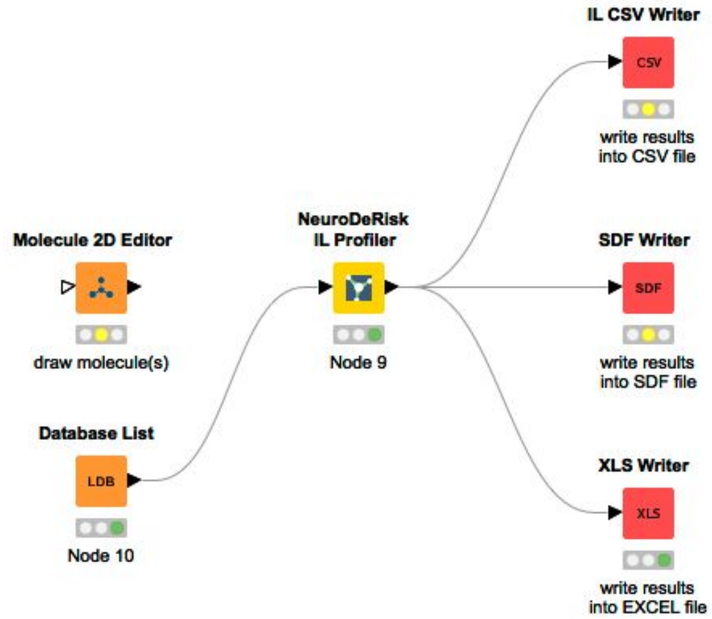
15 PNS

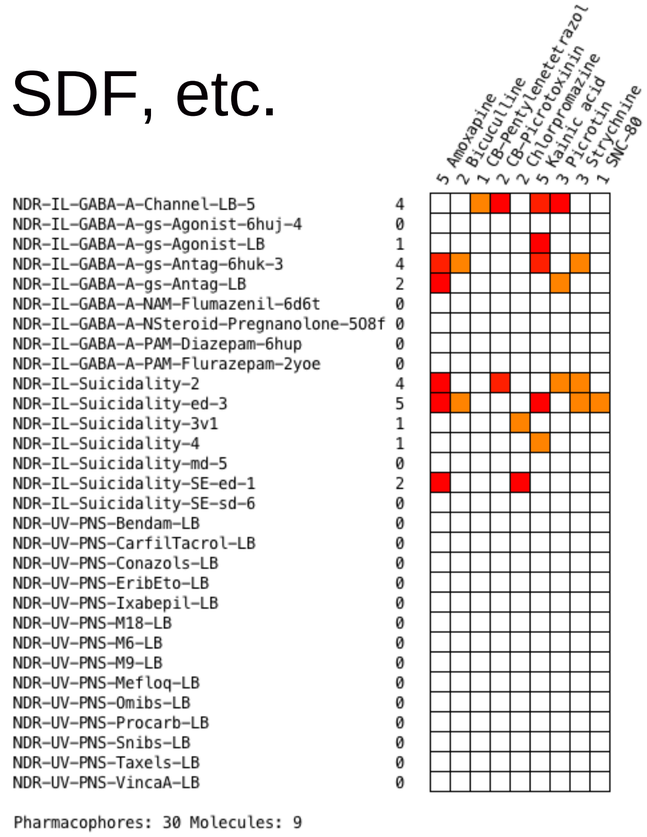
- NDR-IL-PNS-Anthra-C3
- NDR-UV-PNS-Bendam-LB
- NDR-UV-PNS-CarfilTacrol-LB
- NDR-UV-PNS-Conazols-LB
- NDR-UV-PNS-EribEto-LB
- NDR-UV-PNS-Ixabepil-LB
- NDR-UV-PNS-M18-LB
- NDR-UV-PNS-M6-LB
- NDR-UV-PNS-M9-LB
- NDR-UV-PNS-Mefloq-LB
- NDR-UV-PNS-Omibs-LB
- NDR-UV-PNS-Procarb-LB
- NDR-UV-PNS-Snibs-LB
- NDR-UV-PNS-Taxels-LB
- NDR-UV-PNS-VincaA-LB

Neurotoxicity Off Target Prediction

- Profile chemical structures (queries) using 3D-pharmacophore models
- Both models and LigandScout algorithms for profiling are encoded into the NeuroDeRisk IL Profiler node
- Multiple inputs supported including a 2D-editor
- Visualisation and export in different formats: PNG, CSV, XLS, SDF, etc.







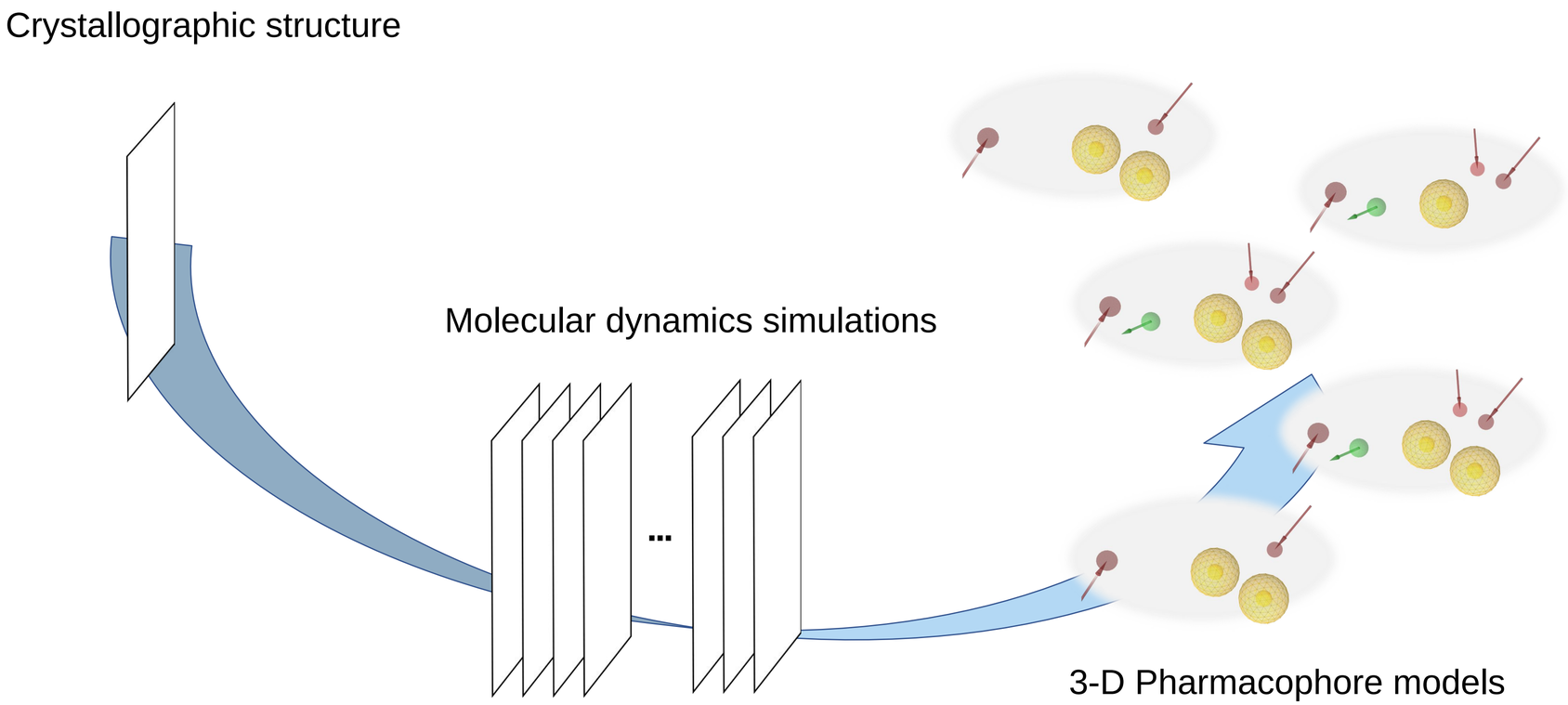
https://neuroderisk.eu/in-silico-toolbox/

Conclusions

- Next generation pharmacophore models are ready for the most difficult challenges in computer aided molecular design
- Highly useful for hit identification & prioritisation
- Amendable for high precision virtual screening in exa scale libraries
- Most comprehensive for medicinal chemistry guidance
- Take into account dynamic effects in binding kinetics and translate directly to design
- Better lead structures void of toxicity risks

Understanding MD Trajectories

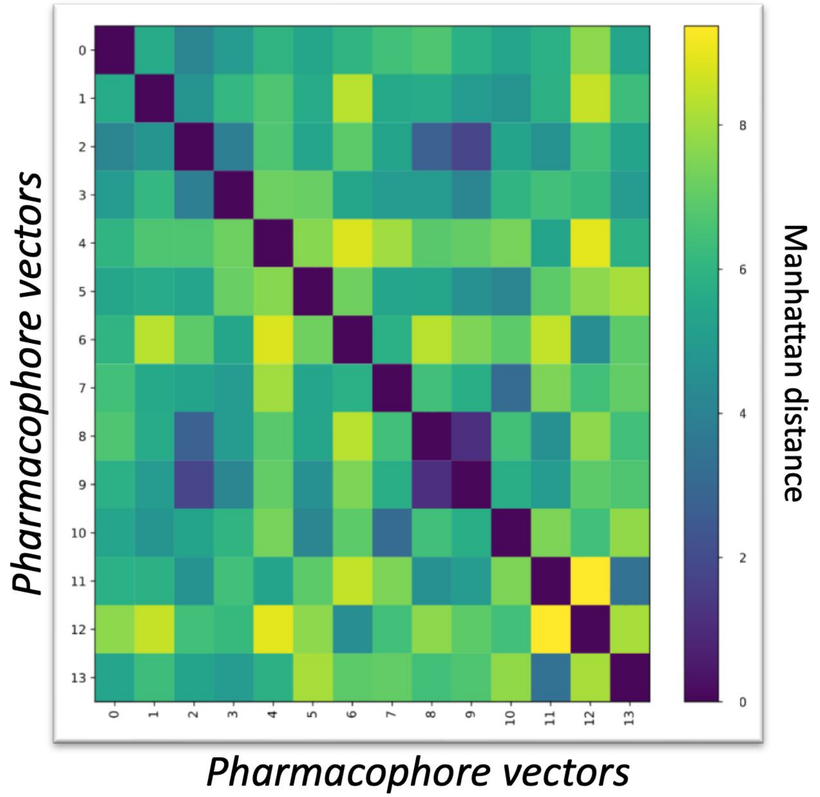
- Define a new way to analyse MD trajectory big data:
  Hierarchical pharmacophore graph representation



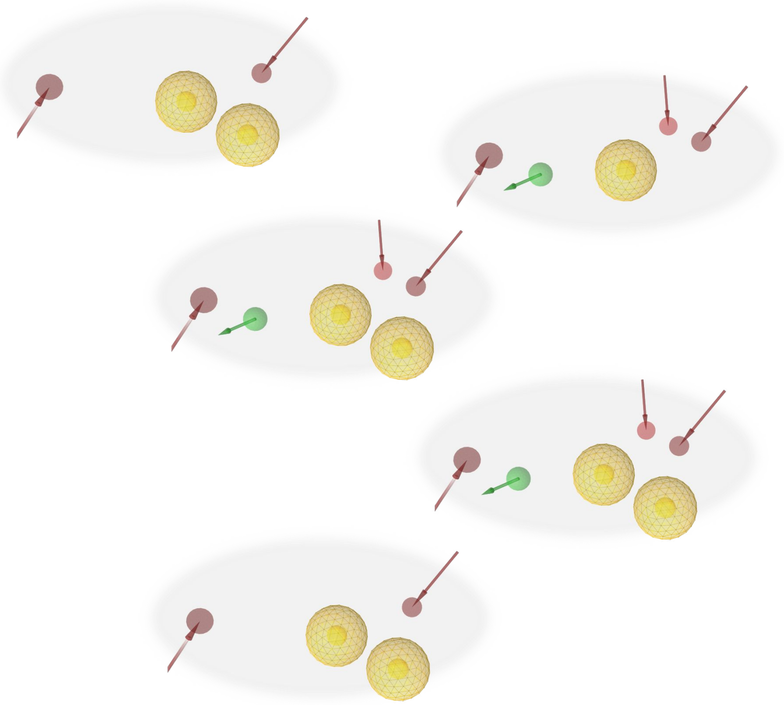
Garon A et al., Models. Front. Mol. Biosci. 7:599059. doi: 10.3389/fmolb.2020.599059

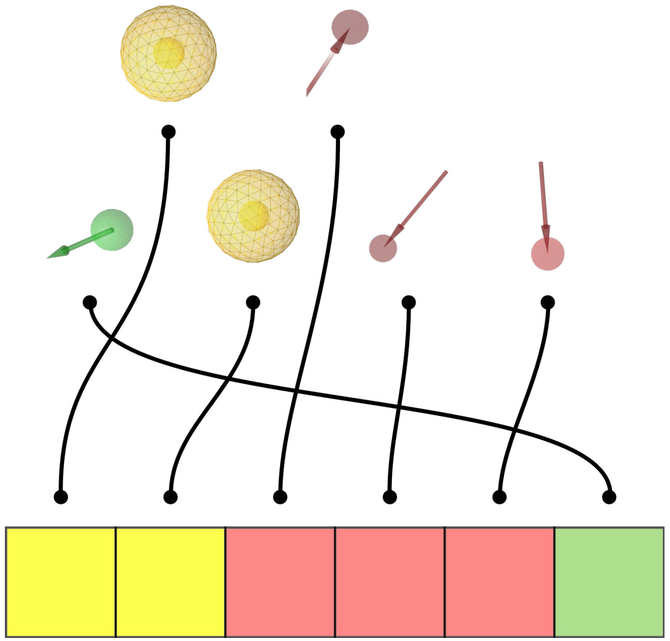
MD P4 Relationship Analysis

- Hierarchical network of pharmacophores
- Layers are organised by multidimensional scaling
  - each node represents one unique pharmacophore model
  - nodes are sorted to best represent the distance between them
  - fingerprint space reduced into one dimension
  - distance between nodes defined by Manhattan metrics

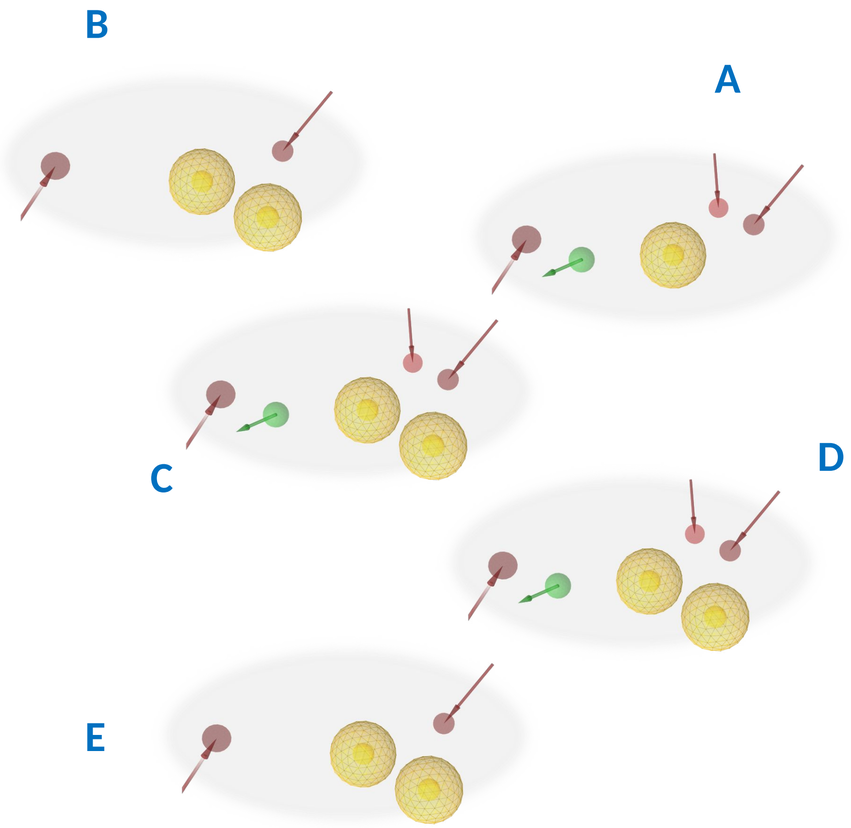


P4 Model Vector Representation



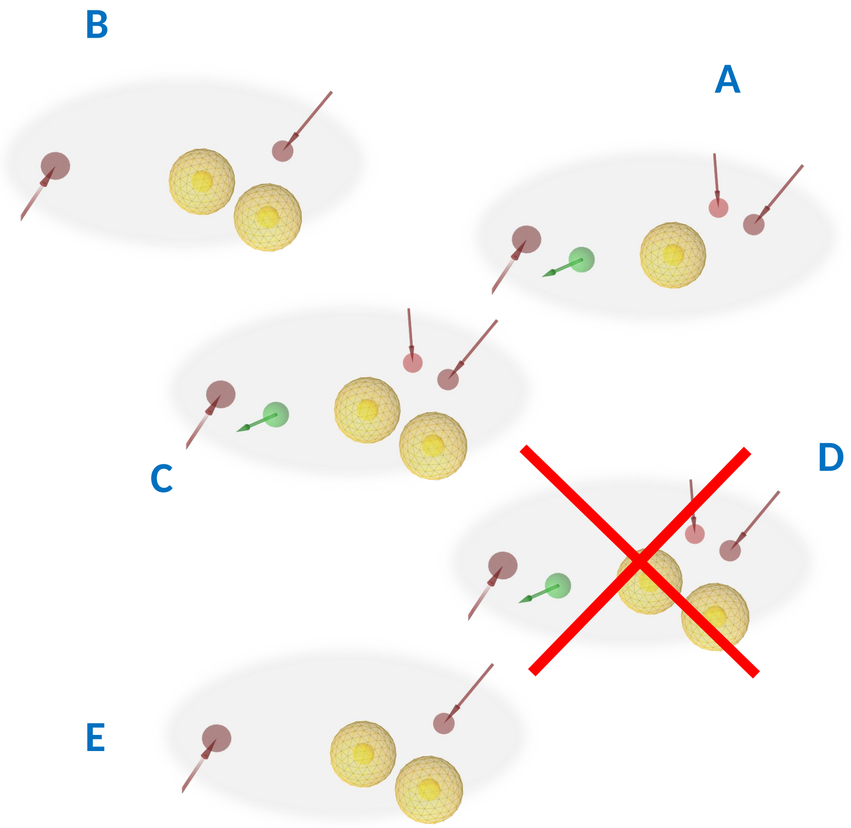


P4 Model Vector Representation



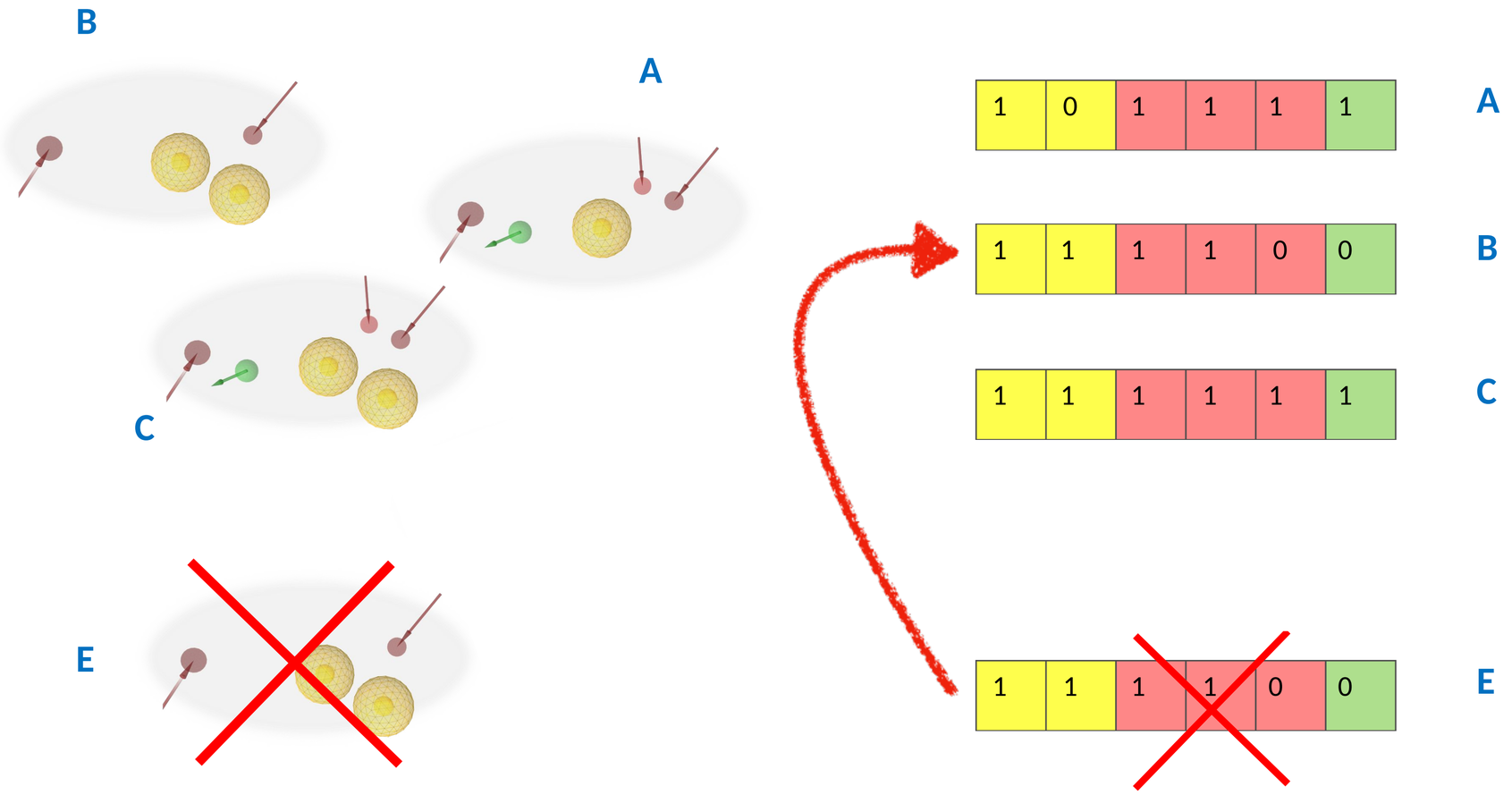
1	0	1	1	1	1	A
1	1	1	1	0	0	B
1	1	1	1	1	1	C
1	1	1	1	1	1	D
1	1	1	1	0	0	E

P4 Model Vector Representation

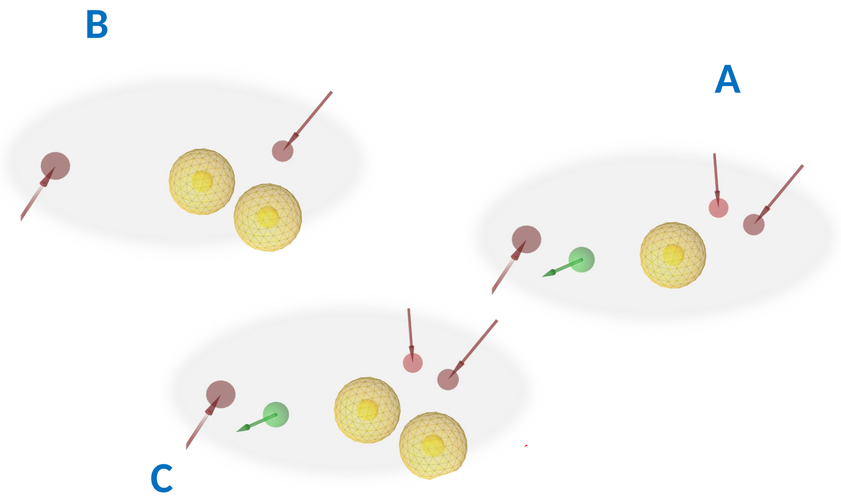


1	0	1	1	1	1	A
1	1	1	1	0	0	B
1	1	1	1	1	1	C
1	1	1	1	1	1	D
1	1	1	1	0	0	E

P4 Model Vector Representation

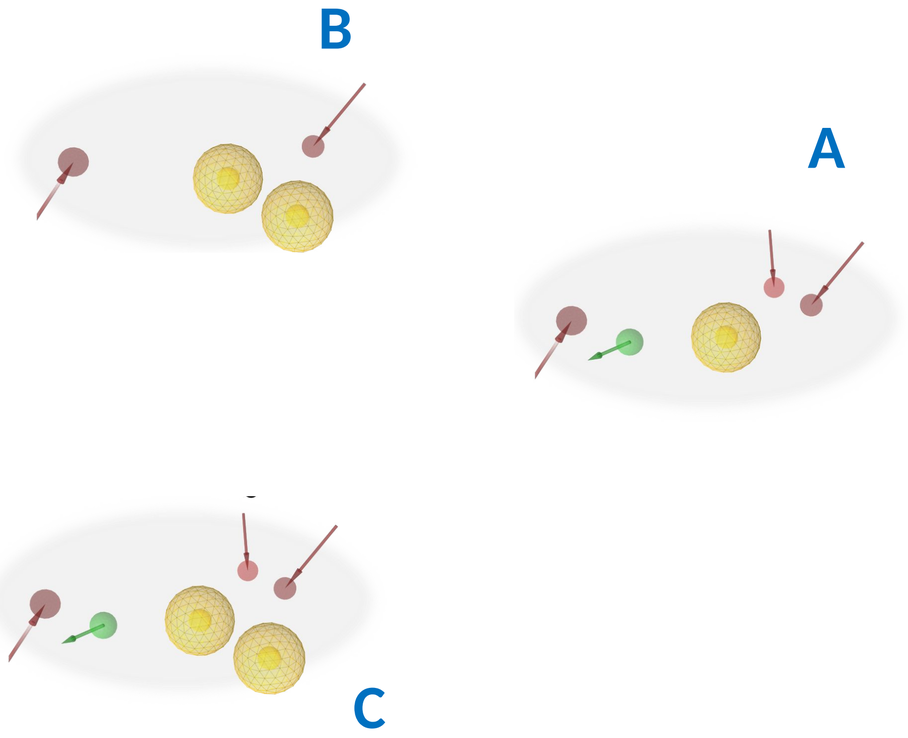


P4 Model Vector Representation

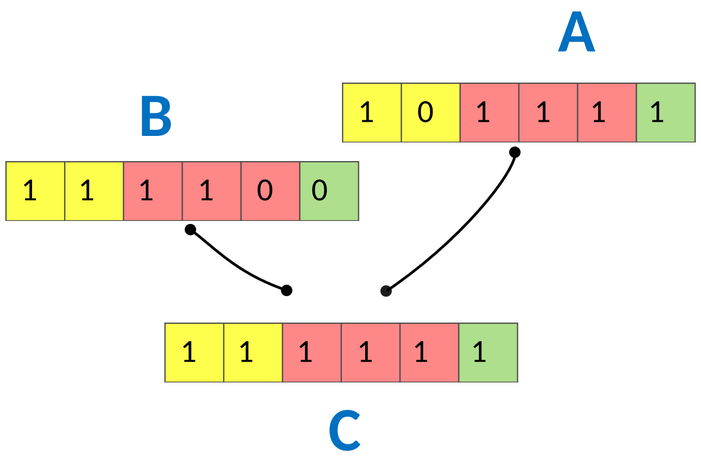


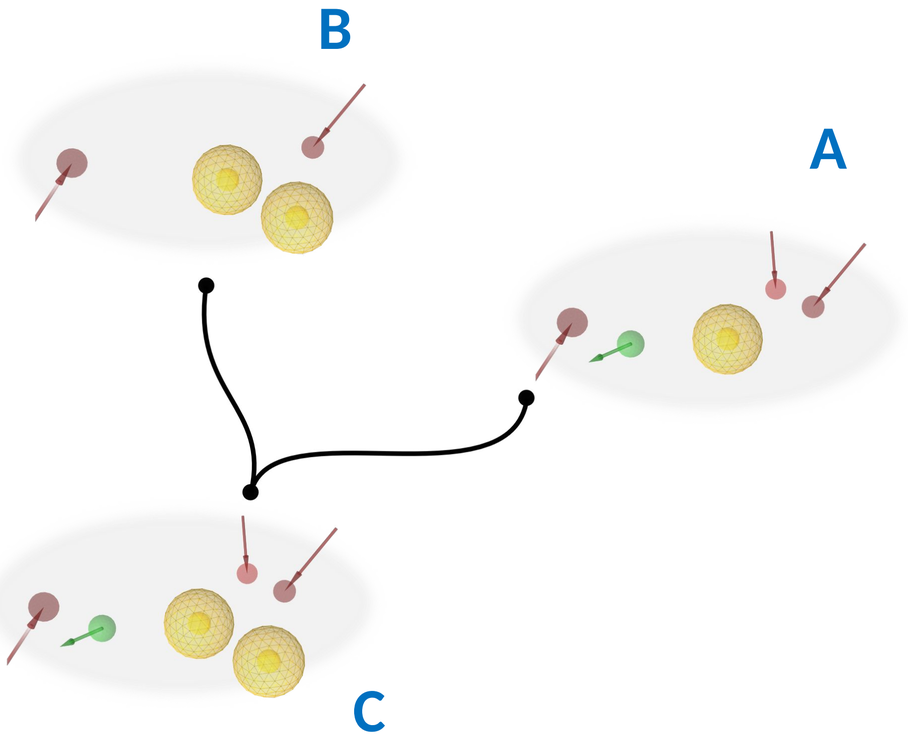
1	0	1	1	1	1	A
1	1	1	1	0	0	B
1	1	1	1	1	1	C

Hierarchical Links Between Models

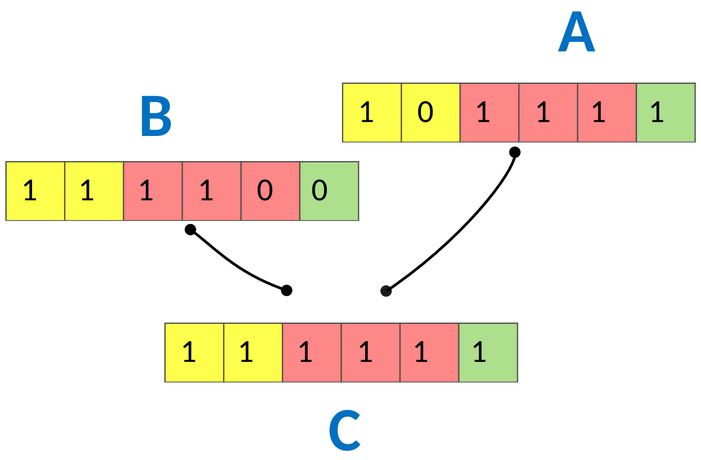


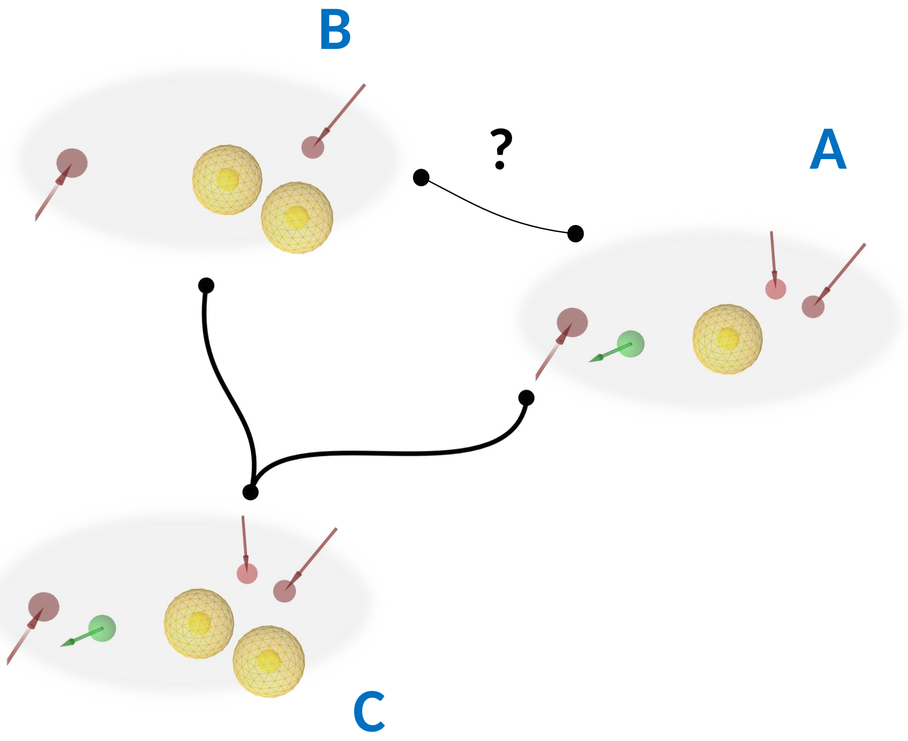
Hierarchical Links Between Models



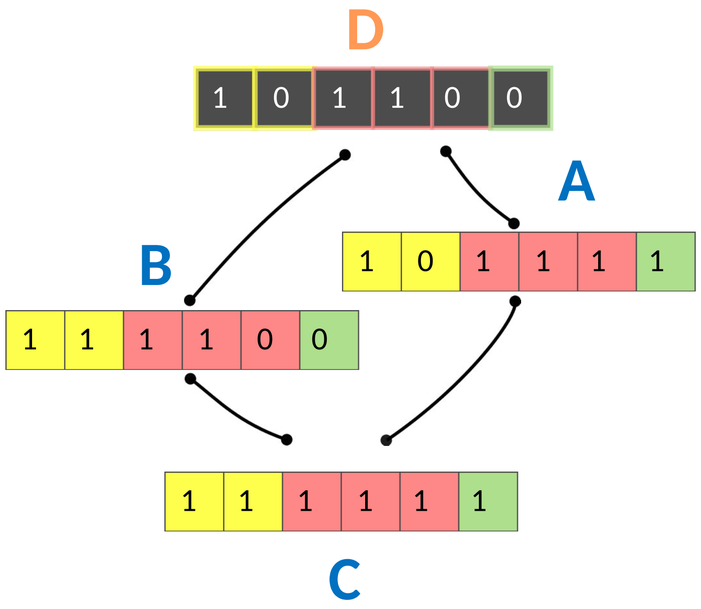


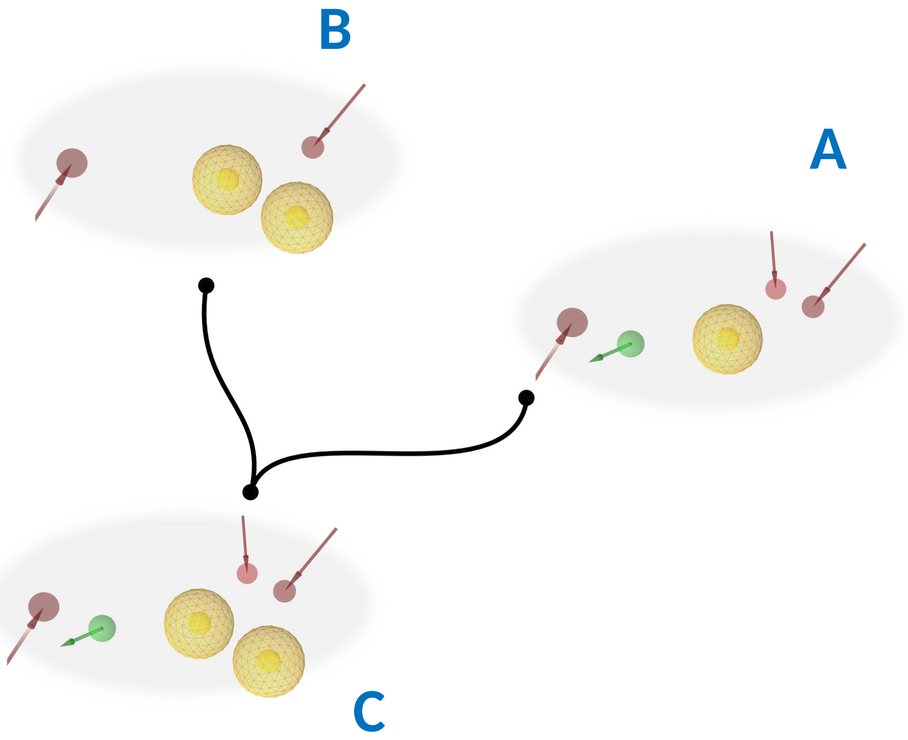
Hierarchical Links Between Models



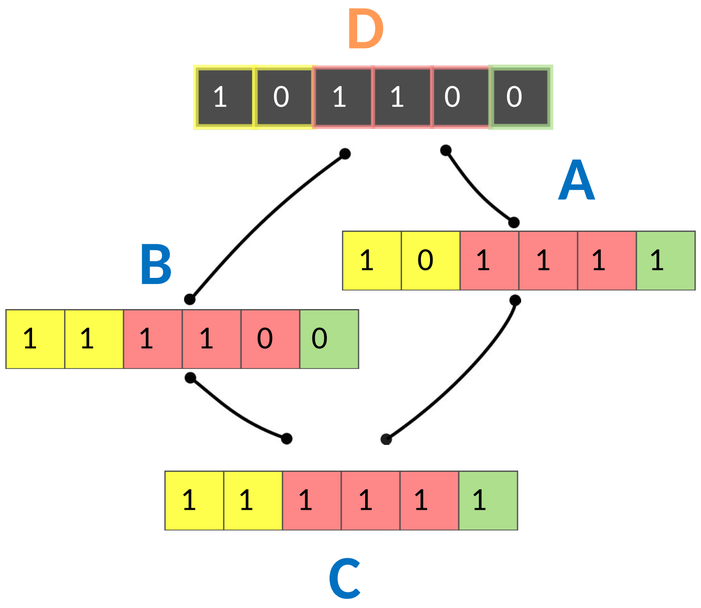


Hierarchical Links Between Models





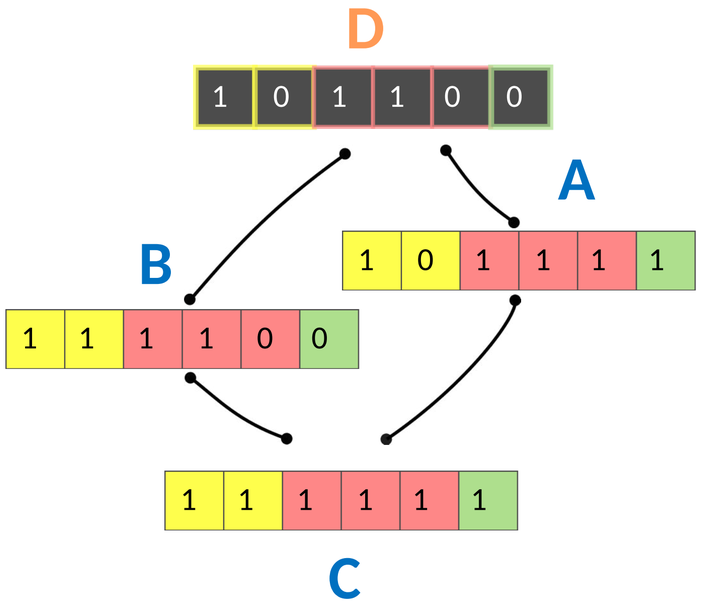
Multi-dimensional Scaling

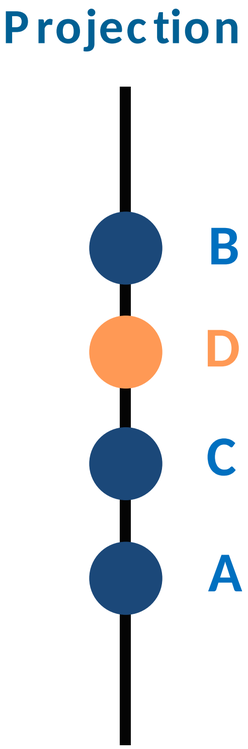


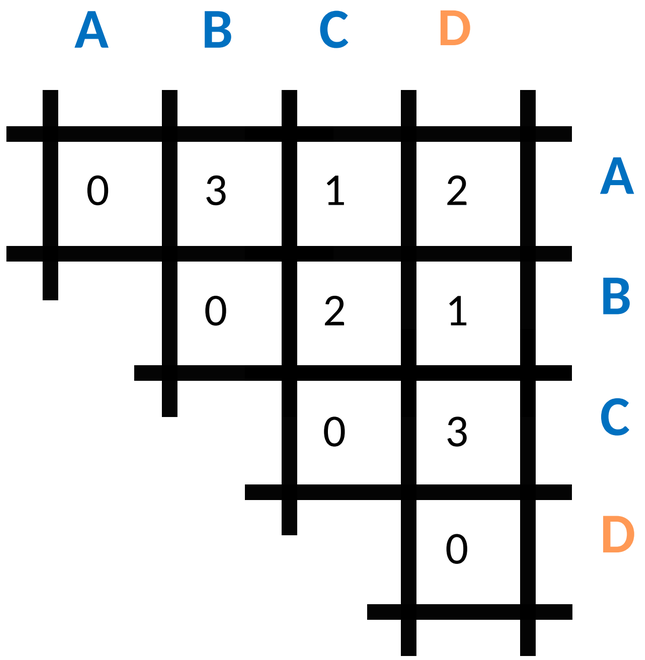
	A	B	C	D
A	0	3	1	2
B		0	2	1
C			0	3
D				0

Distance Matrix

Multi-dimensional Scaling

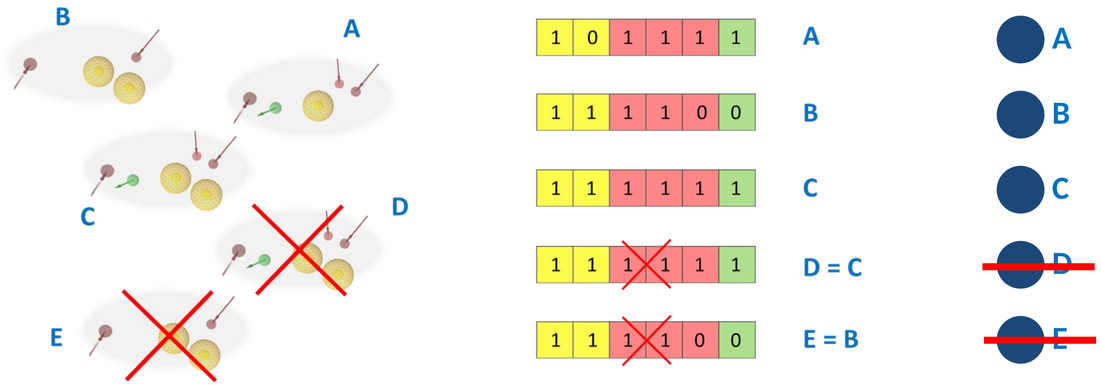


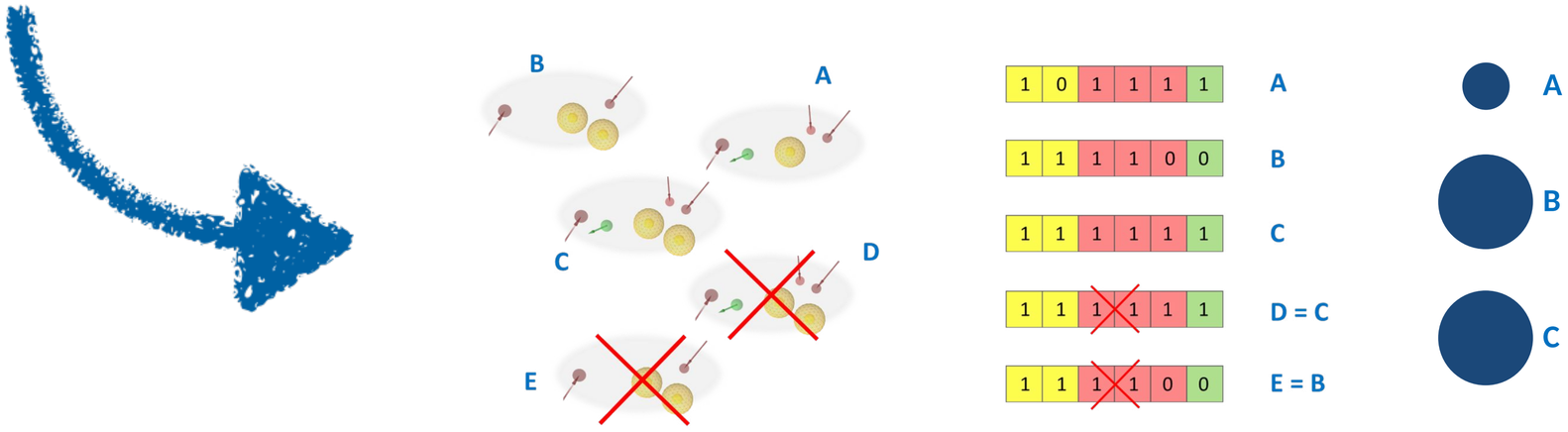




Distance Matrix

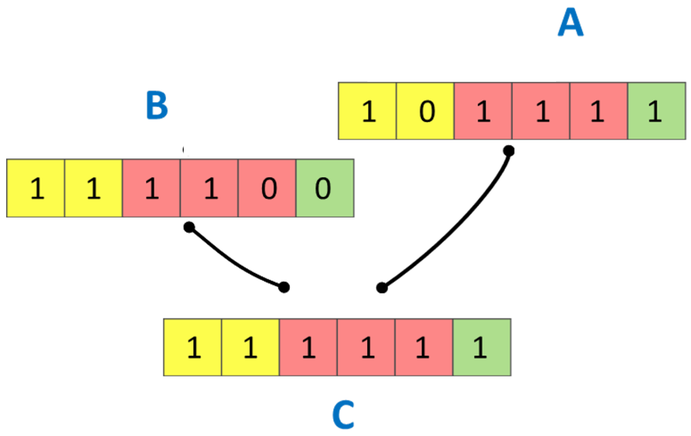
From P4 Model to Vector to Node

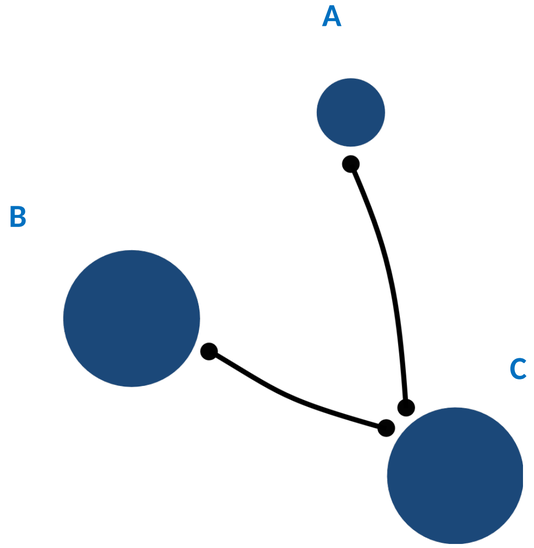




Hierarchical Graph of P4 Models

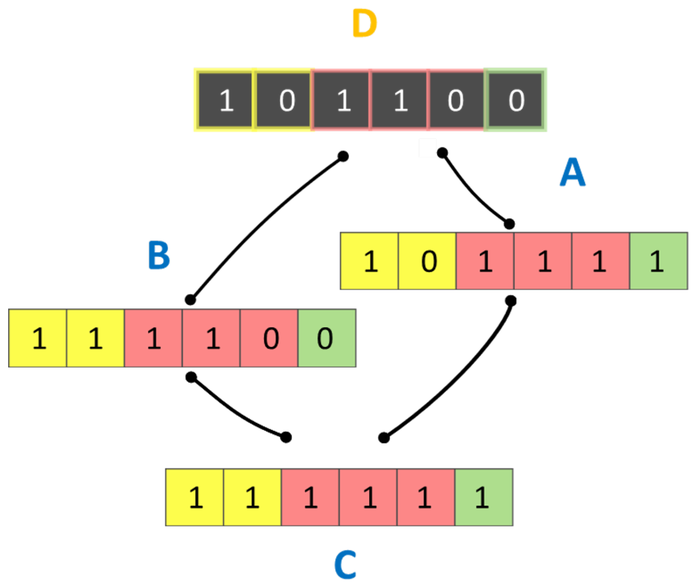
- 3-D Pharmacophore model ⇔ Vector ⇔ Node
- Frequency ⇔ Node size
- Hierarchical link ⇔ Subset/superset of P4 features

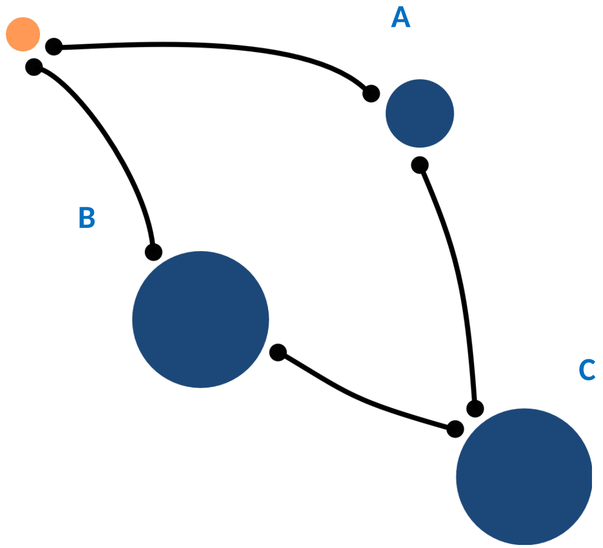




Hierarchical Graph of P4 Models

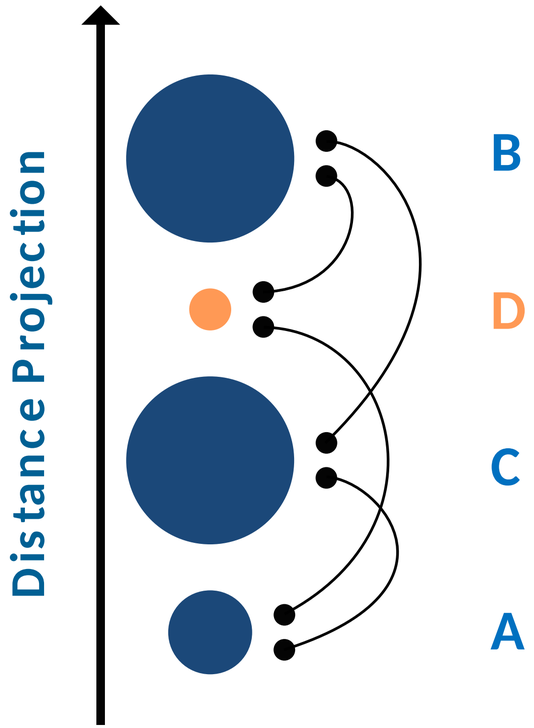
- 3-D Pharmacophore model ⇔ Vector ⇔ Node
- Frequency ⇔ Node size
- Hierarchical link ⇔ Subset/superset of P4 features
- Node color (observed/calculated)





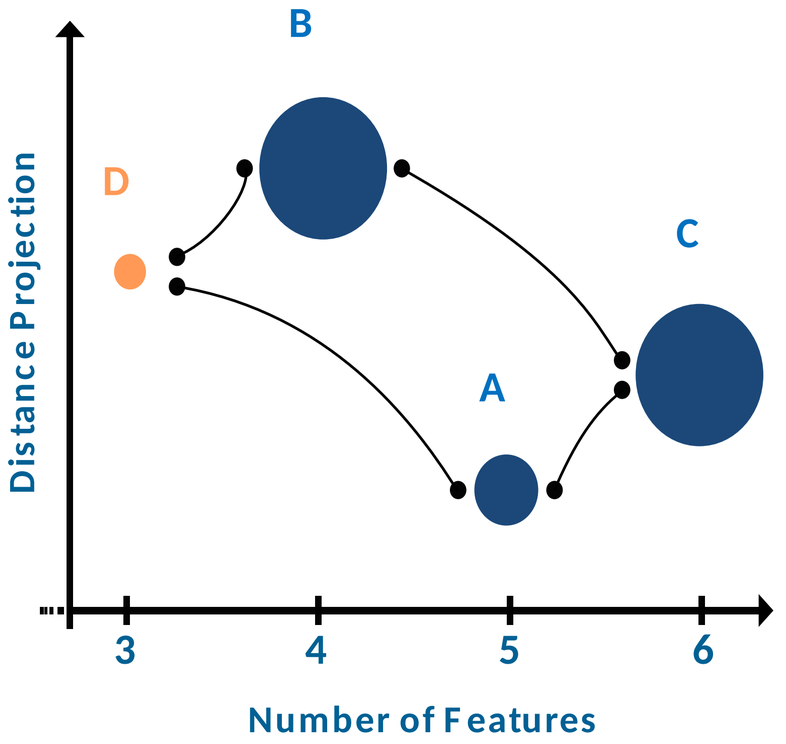
Hierarchical Graph of P4 Models

- 3-D Pharmacophore model ⇔ Vector ⇔ Node
- Frequency ⇔ Node size
- Hierarchical link ⇔ Subset/superset of P4 features
- Node color (observed/calculated)
- Distance projection: Y-Axis

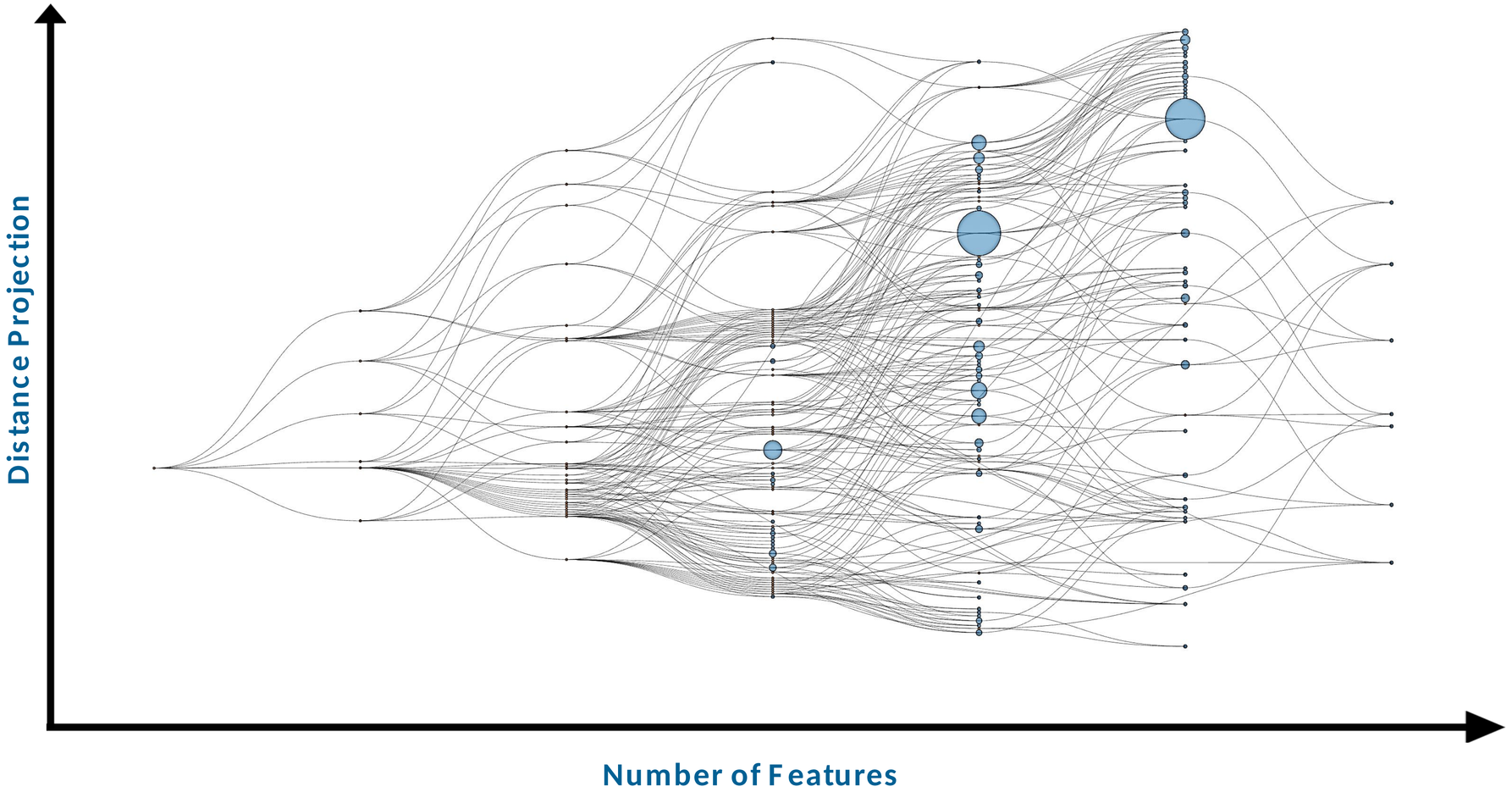


Hierarchical Graph of P4 Models

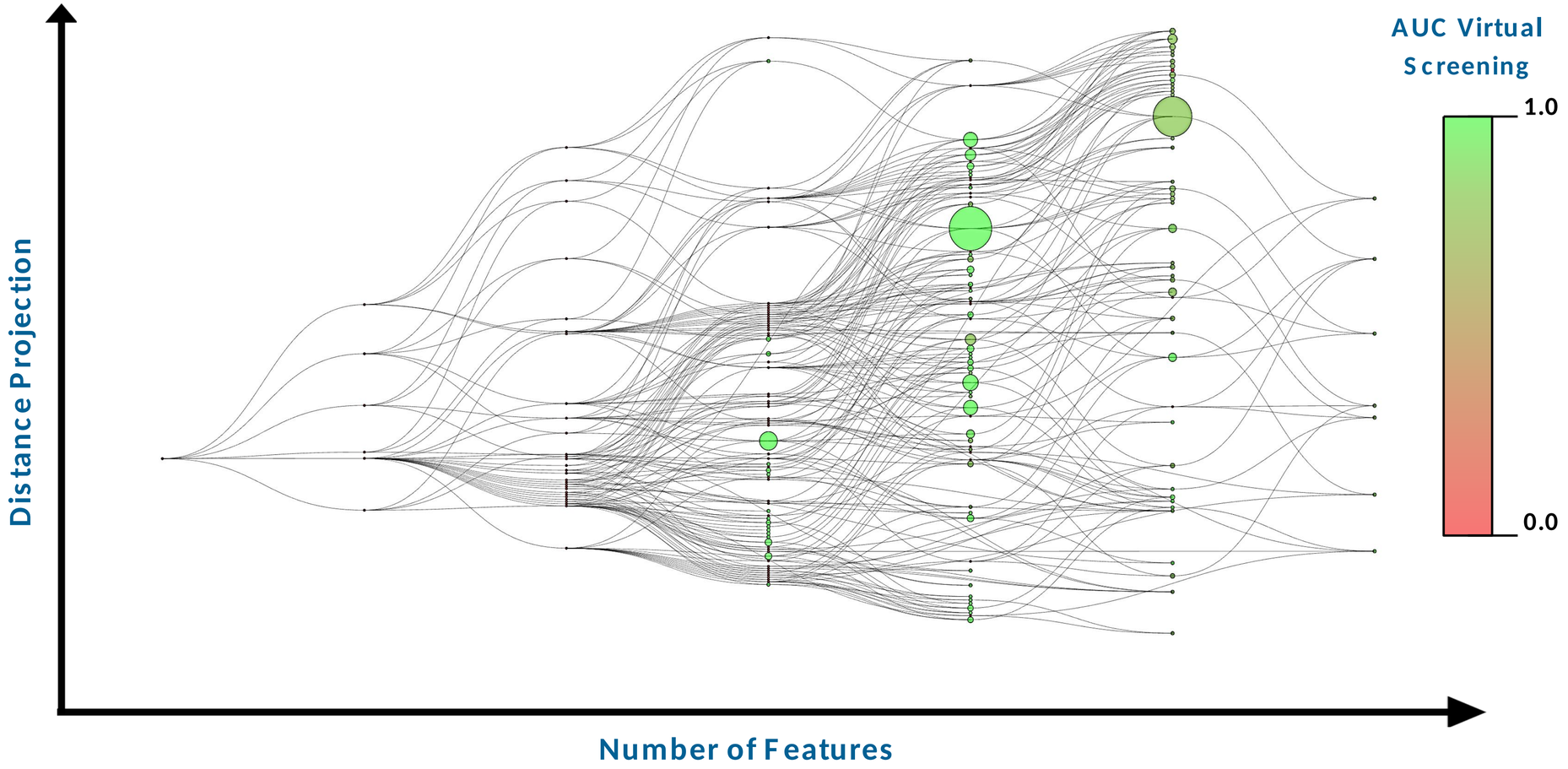
- 3-D Pharmacophore model ⇔ Vector ⇔ Node
- Frequency ⇔ Node size
- Hierarchical link ⇔ Subset/superset of P4 features
- Node color (observed/calculated)
- Distance projection: Y-Axis
- Number of node features: X-Axis



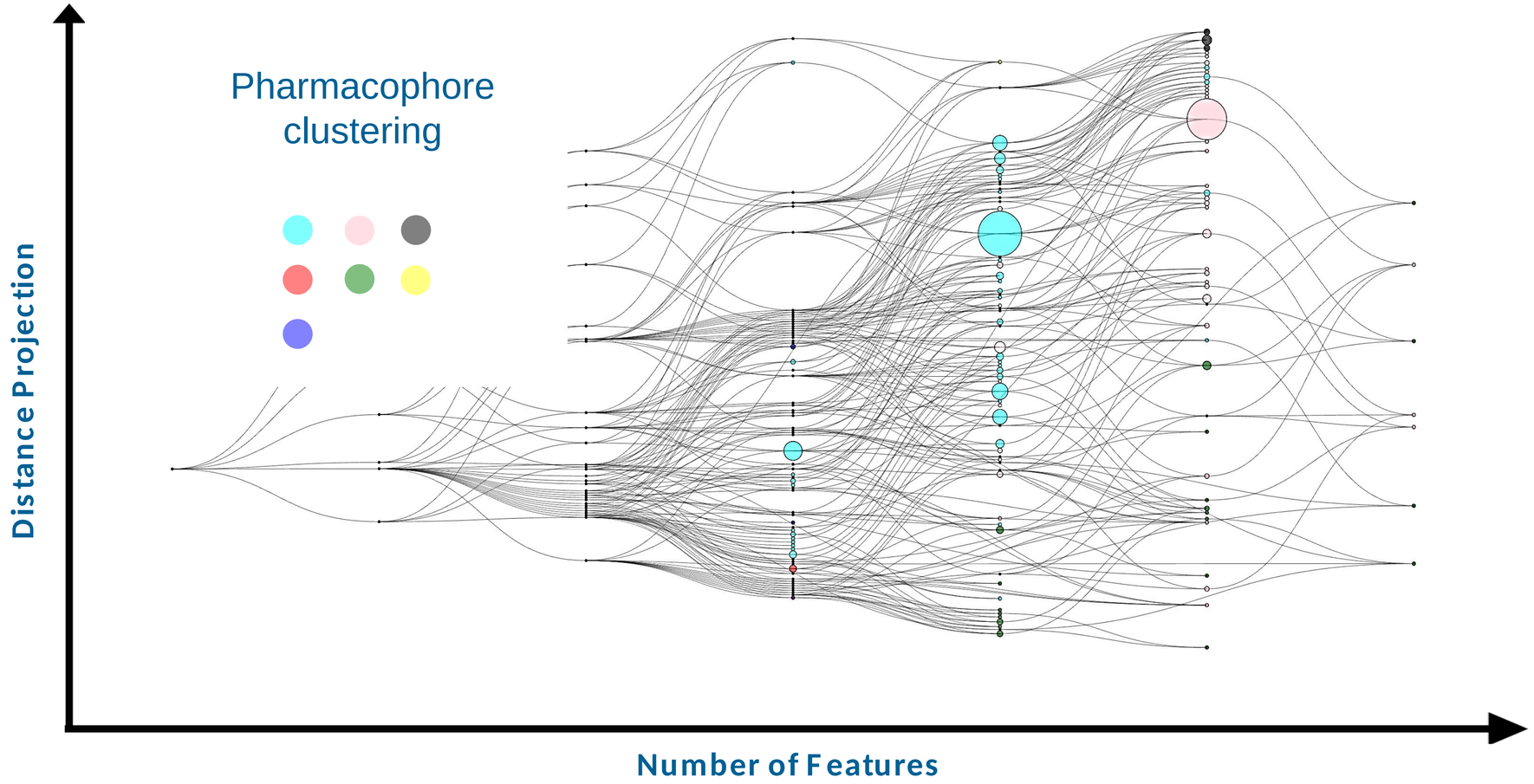
Example



Example - Virtual Screening AUC

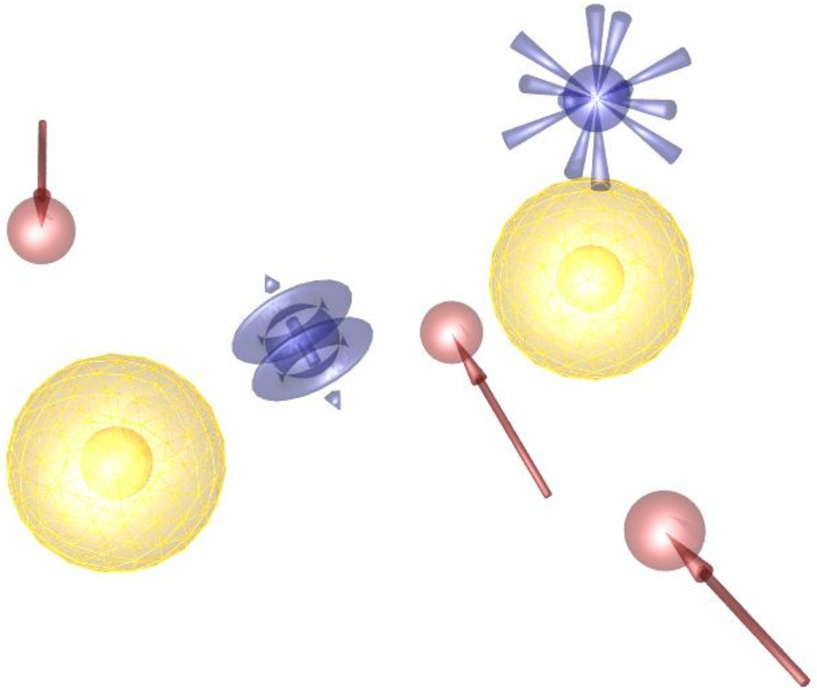


Example - P4 Clustering

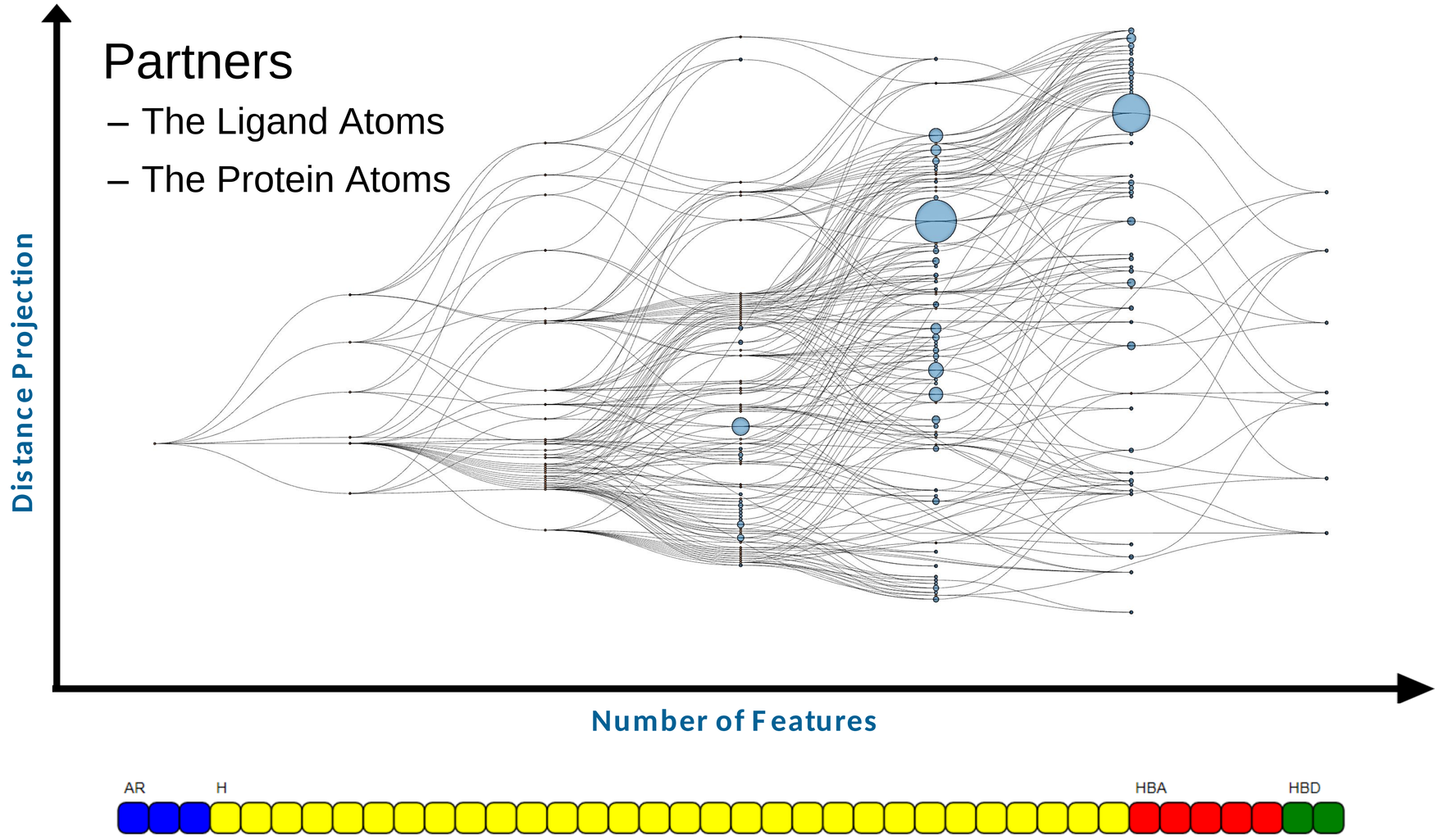


P4 Feature Representation

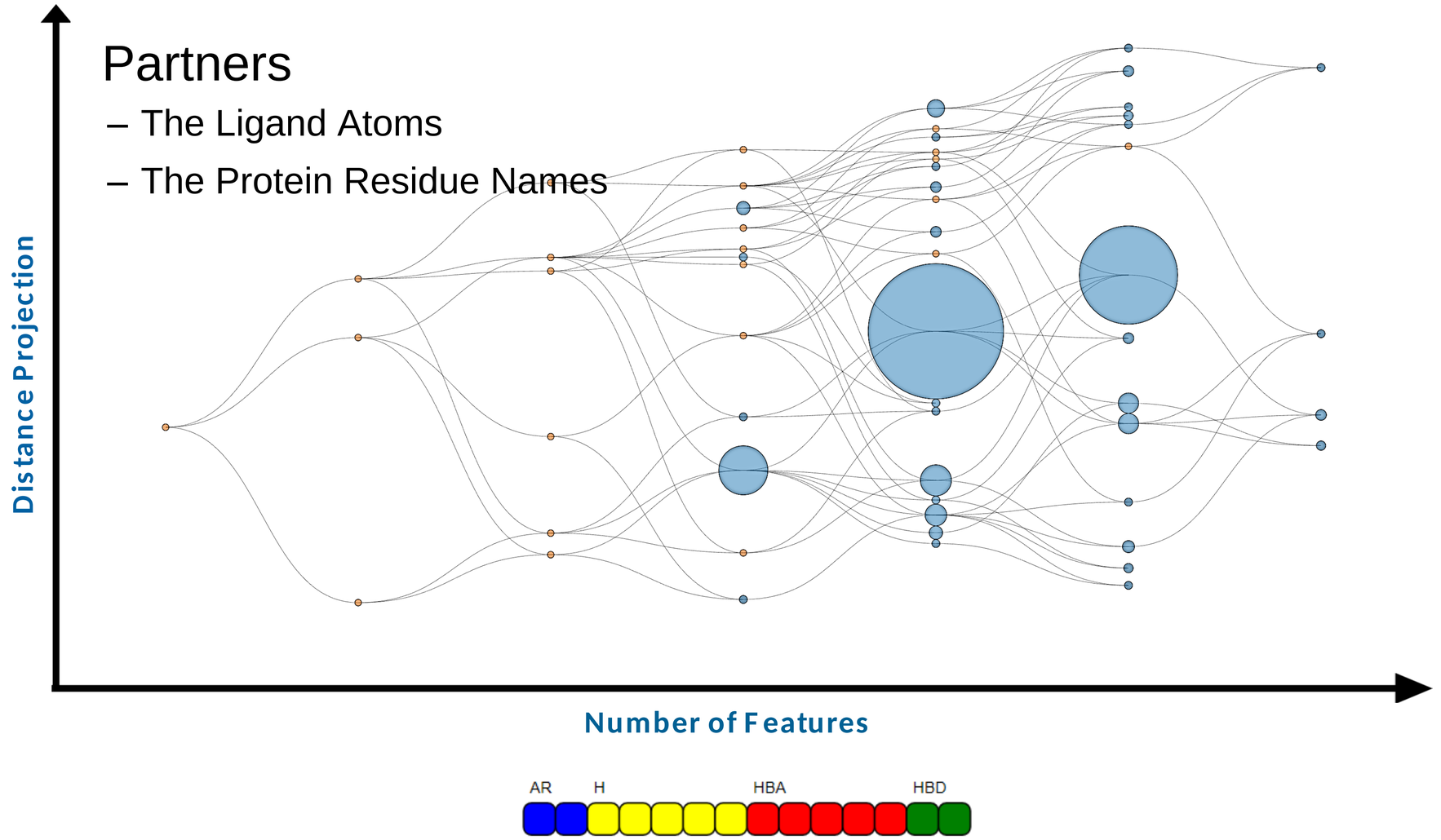
- Type of Interaction
  - Distance, Angle, Direction
- Partners
  - X,Y,Z Coordinates
- Type of Interaction
  - Distance, Angle, Direction
- Partners
  - The Ligand Atoms / None
  - The Protein Atoms / Residue Na



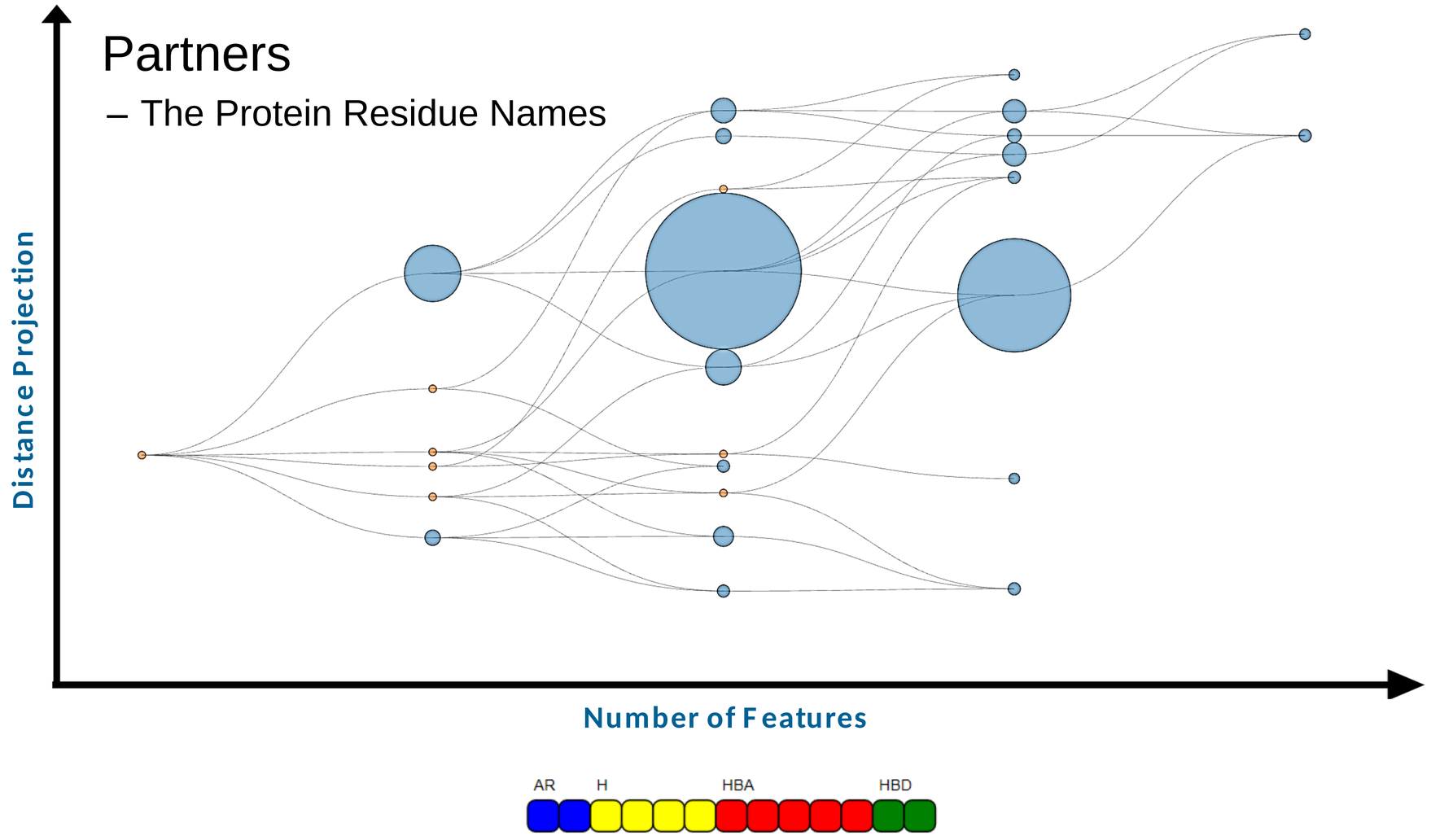
Example - P4 Feature Partners



Tree Simplification

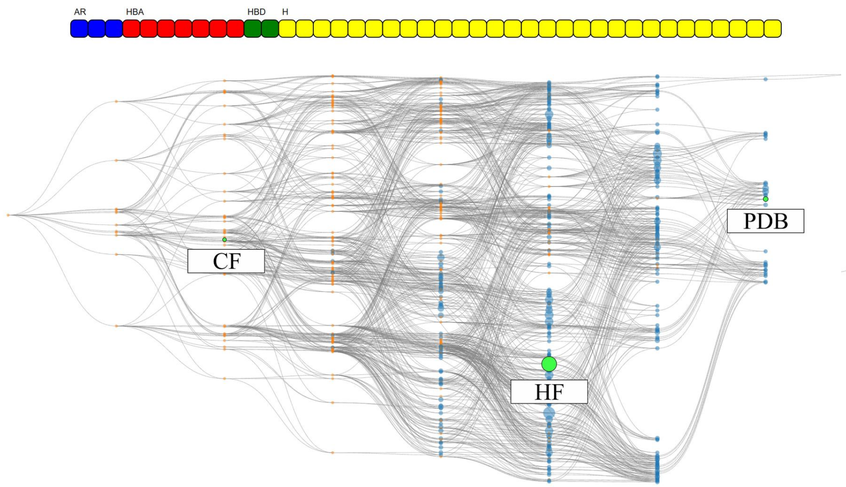


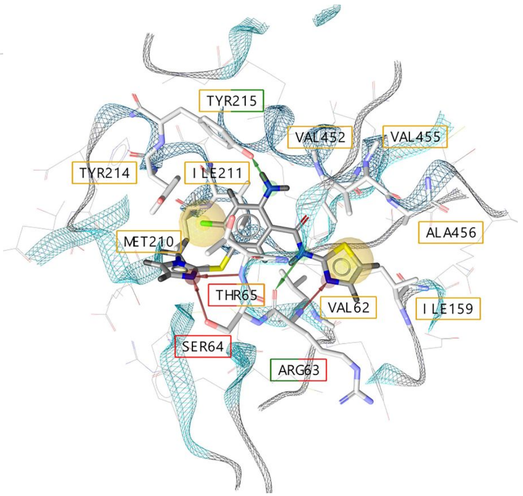
Tree Simplification (2)

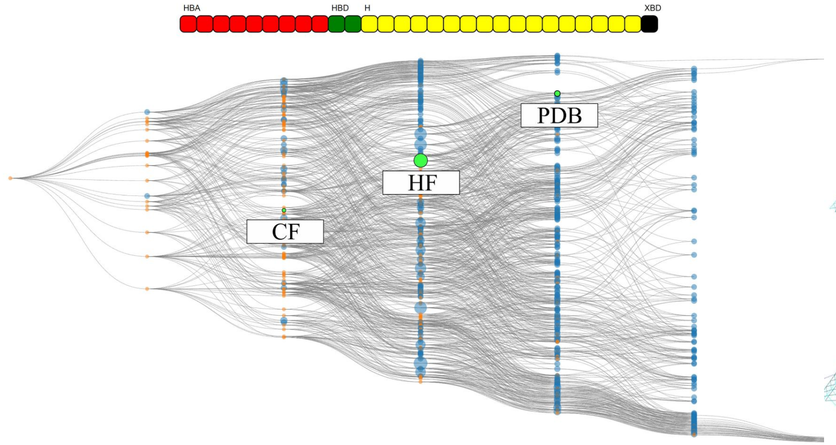


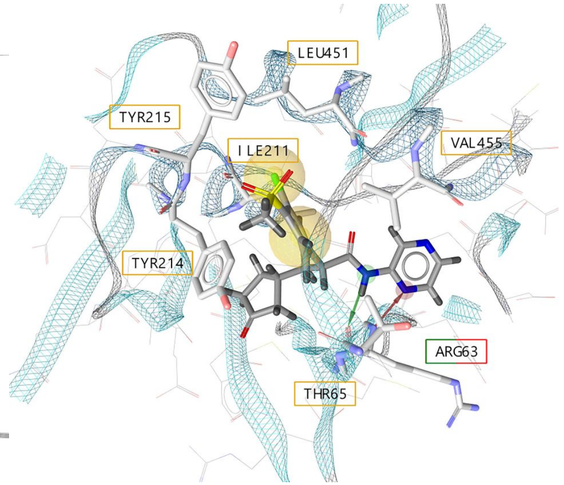
Glucokinase Case Study

- Analysis of binding modes of ligands in human glucokinase: Difference between active & inactive form





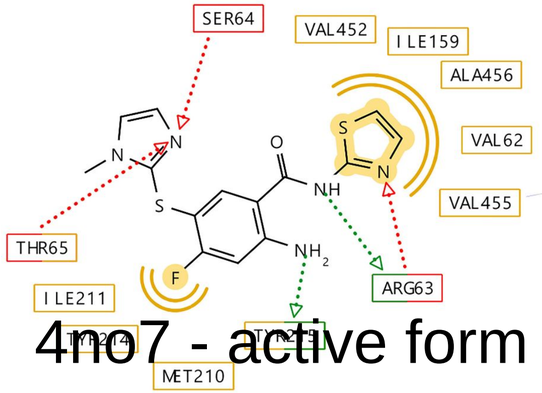


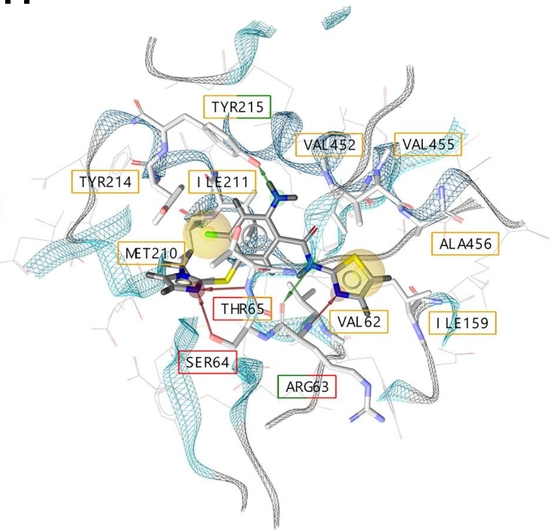


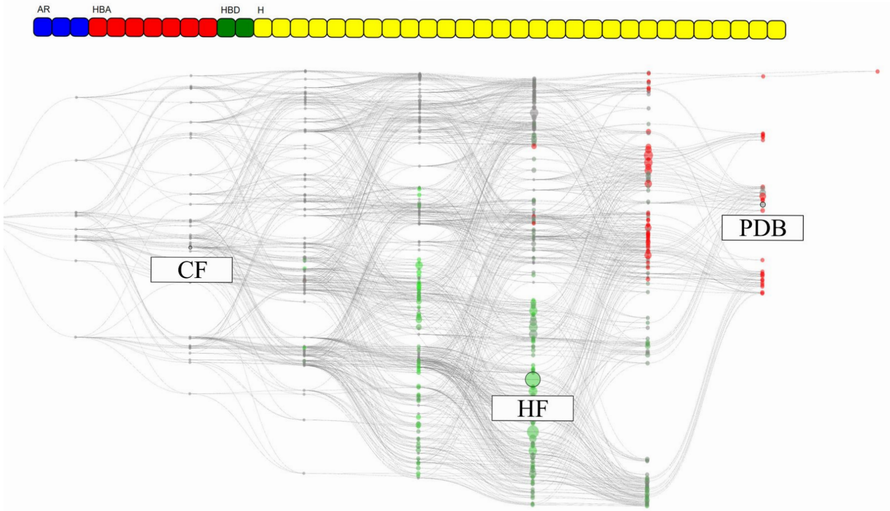
Garon A et al., Models. Front. Mol. Biosci. 7:599059. doi: 10.3389/fmolb.2020.599059

Glucokinase Case Study

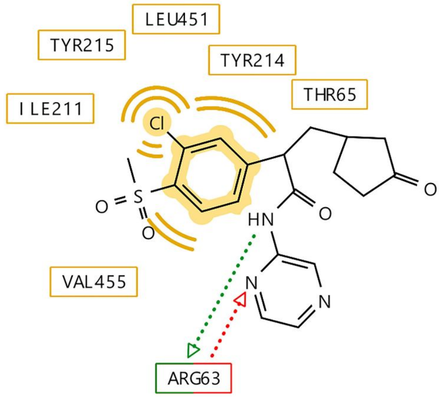
1v4s - inactive form

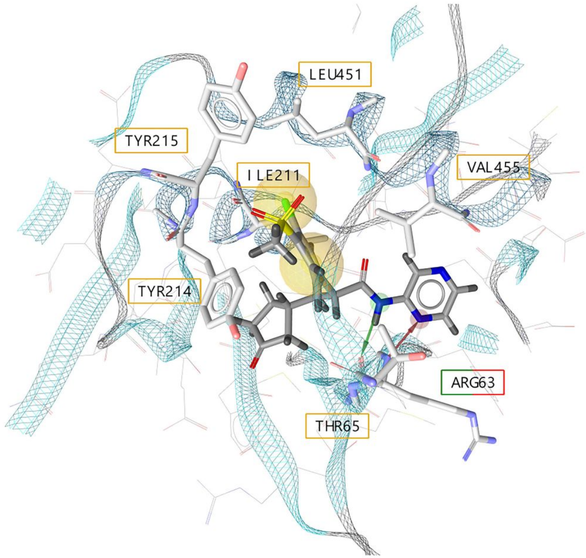


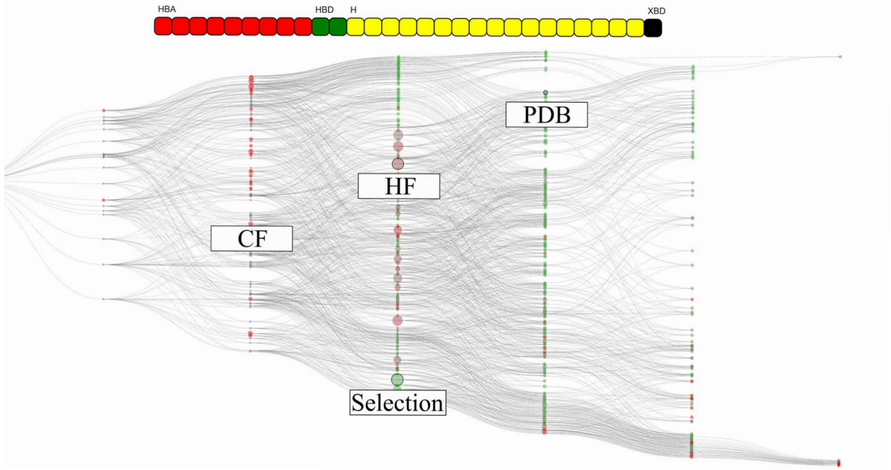




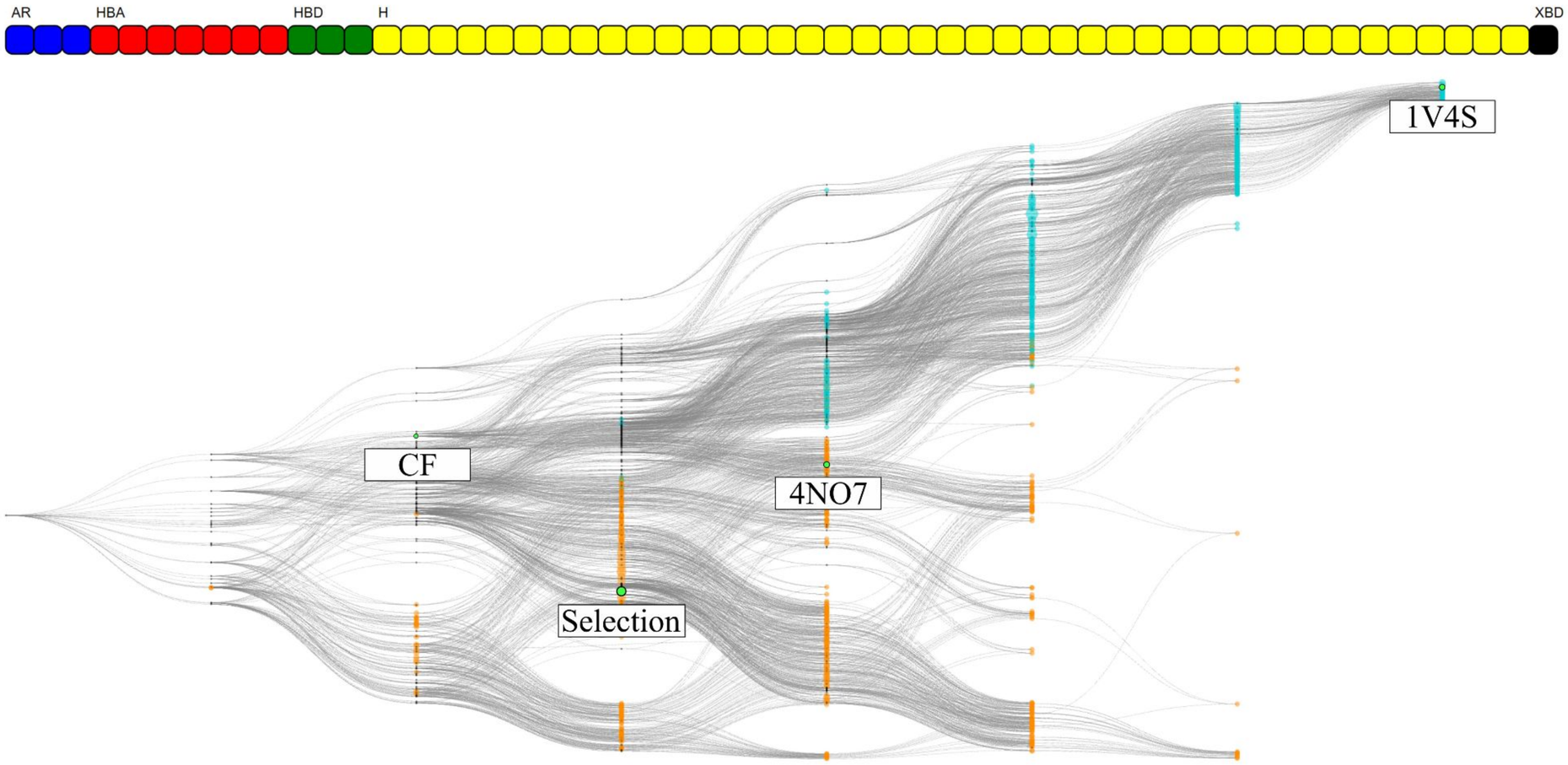
4ne7 - active form



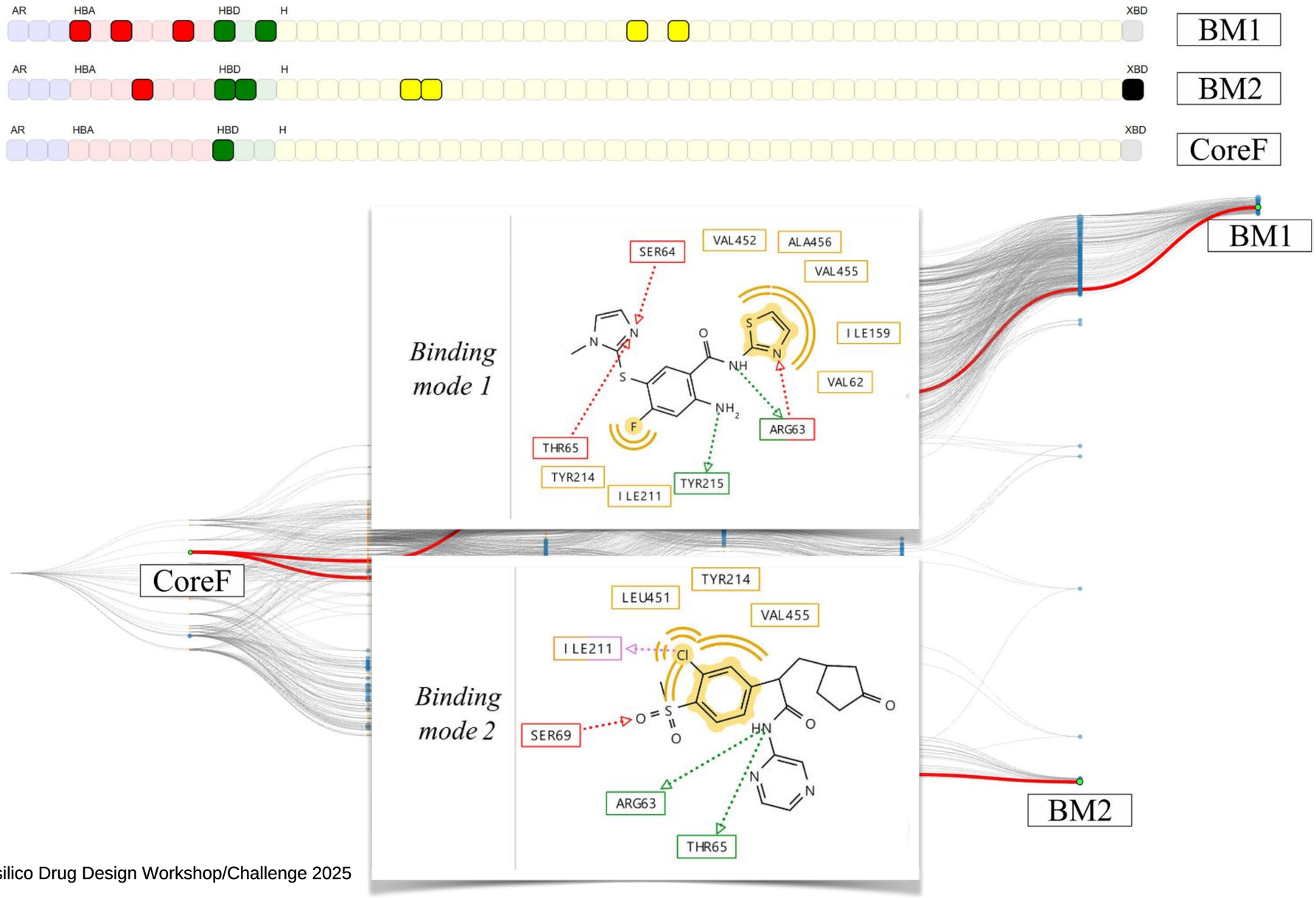




Glucokinase Case Study



Glucokinase Case Study



Conclusions

- Our pattern recognition-base pharmacophore technique is superior to all previous P4 methods with respect to speed and accuracy

➡ Highly useful for hit identification

- The pharmacophore interaction analysis concept is no more limited to static observation but is available in a convenient dynamic approach

➡ Highly useful for lead structure optimisation

Thank you for your attention

https://cheminfo.univie.ac.at

https://CDPKit.org

https://www.inteligand.com

LigandScout XT