Develop Externally Validated QNAR Models That Can be Reliabliy Used to Prioritize Nanomaterials for Experimental Testing *(425.045)*

**Development of Materials Fingerprints*
**Predictive Models for Superconductivity*

PIs:

- **Alexander Tropsha, PhD (University of North Carolina at Chapel Hill)**
- **Denis Fourches, PhD (University of North Carolina at Chapel Hill)**

Research Objectives

- **Develop predictive QNAR models that correlate the compositional/physical/chemical/geometrical and biological descriptors of nanomaterials with toxicological endpoints or other valuable properties for semi-conductor industry;**

- **Employ QNAR models for virtual screening of libraries to prioritize materials for experimental testing;**

- **Mine and visualize very large sets of materials using their computed descriptors and other cheminformatics techniques.**

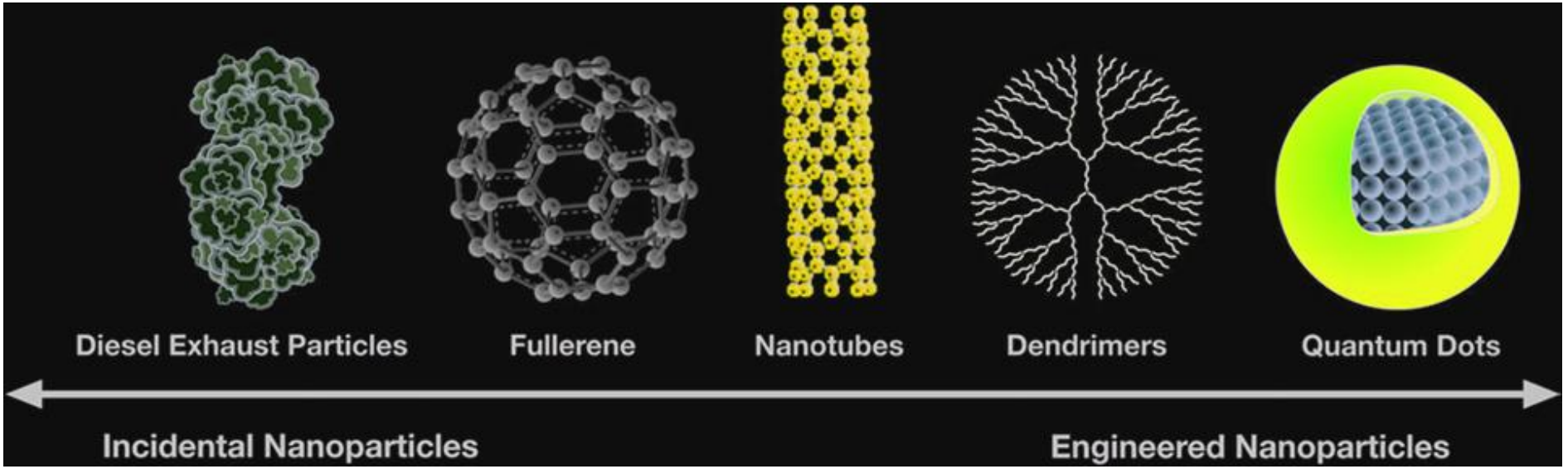
ESH Metrics and Impact

1. **Obtain reliable descriptors of the physical and chemical properties of nanomaterials.**

2. **Develop predictive computational models that correlate physical-chemical descriptors of nanomaterials with their properties and potential toxic effects.**

Impact: Utilize the knowledge gained through above metrics for improved nanomaterials' experimental design and prioritized toxicity testing toward the manufacturing of safe nanomaterials.

Challenges in Modeling of Nanomaterials

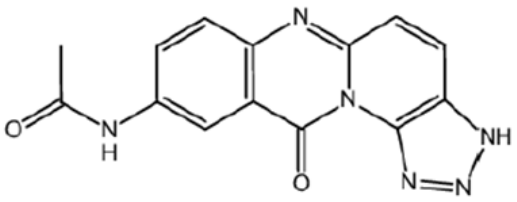


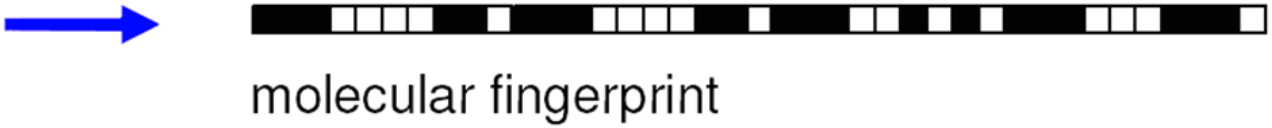
S. Stern and S. McNeil, Toxicological Sciences, 101(1), 4-21, 2008.

- **Structures are very diverse → a real challenge to develop quantitative parameters (descriptors) for (nano)materials.**
- **Systematic physico-chemical, geometrical, structural and biological studies of nanomaterials are nearly absent.**
- **Computational modeling of nanomaterials is only beginning to emerge; best if done in collaboration with experimental scientists.**

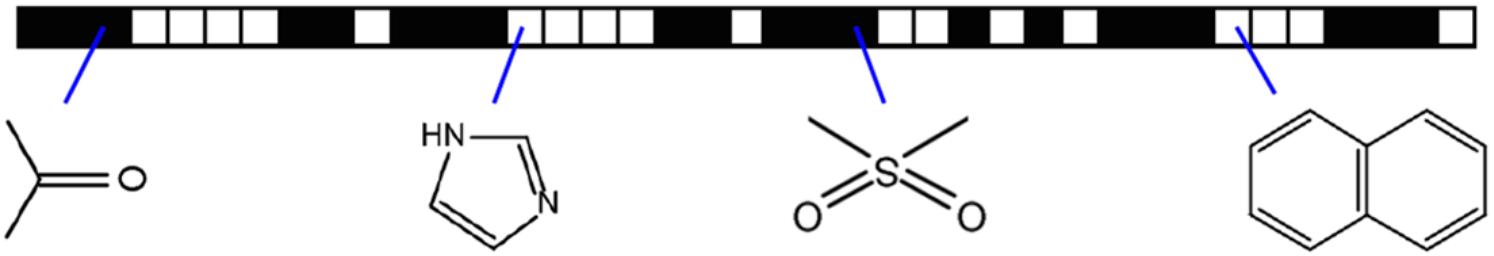
Molecular fingerprints

bit string encodings of
structural features and/or calculated molecular properties.



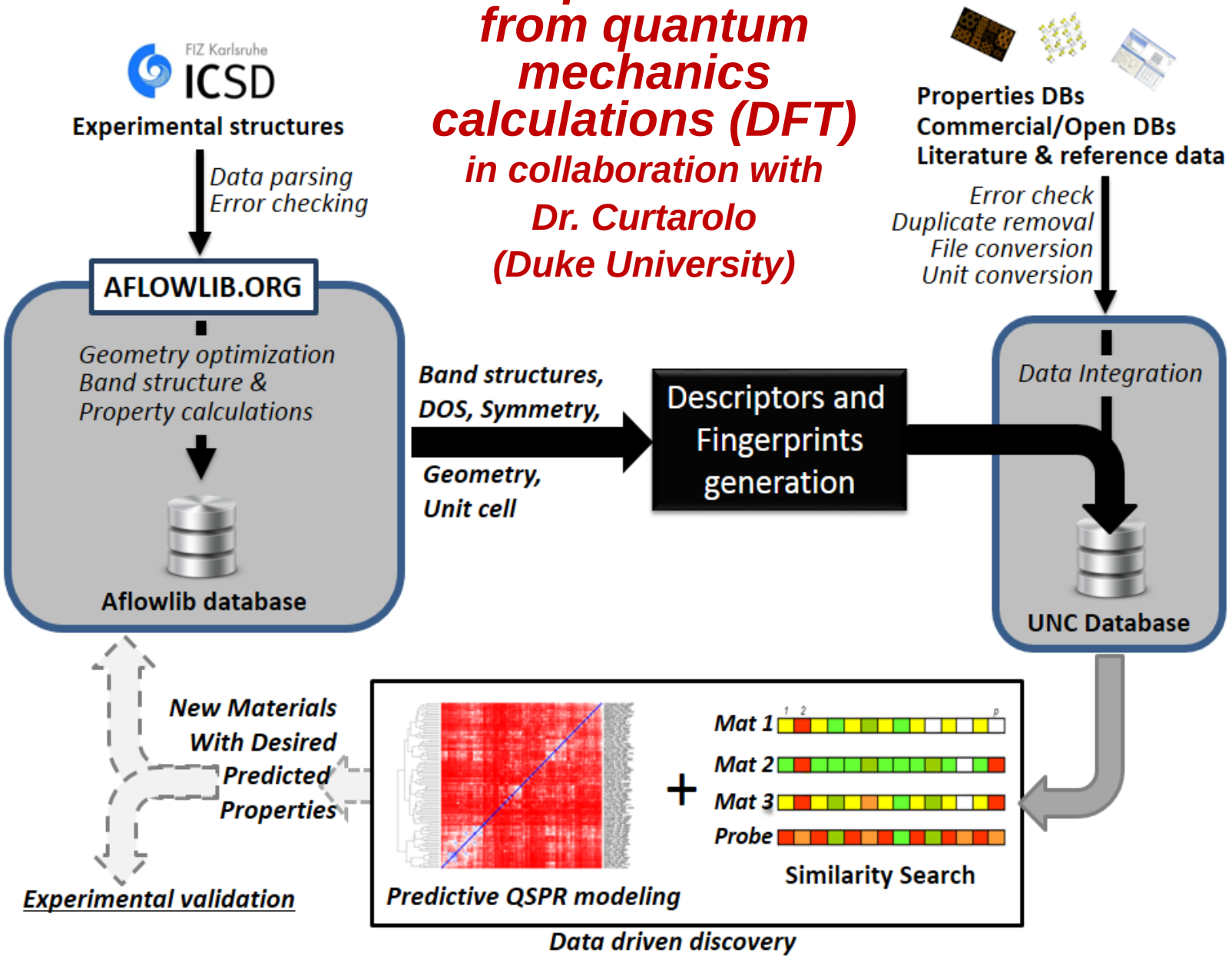


2D Fragment-based, keyed fingerprints: **each bit position monitors**
the presence or absence of structural fragments
MACCS (166 bits), BCI (e.g. 1,052 bits)



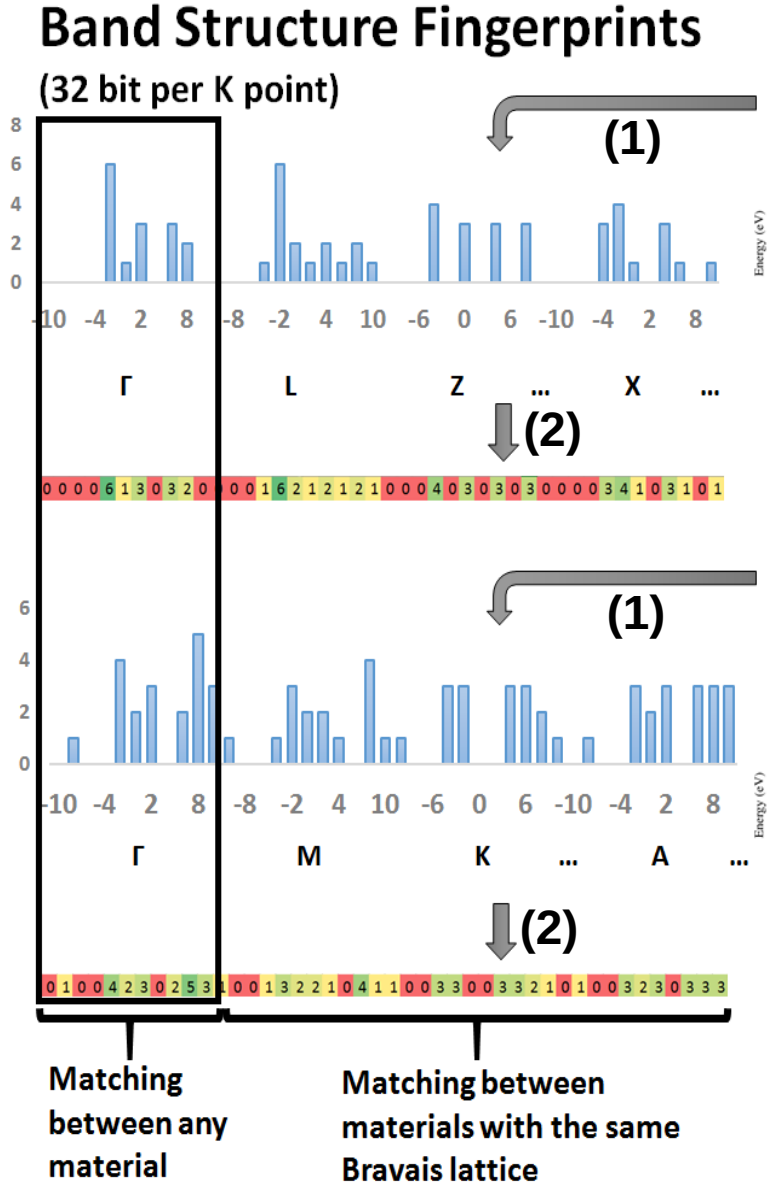
Descriptors derived from quantum mechanics calculations (DFT)

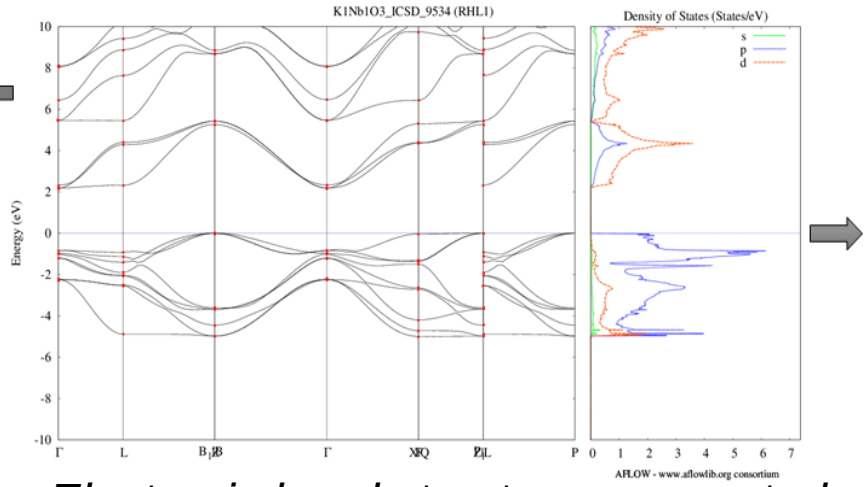
in collaboration with Dr. Curtarolo (Duke University)



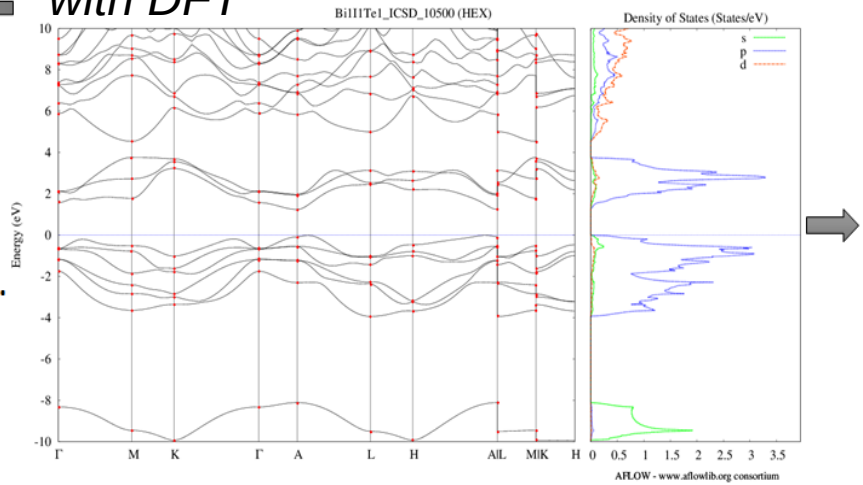
Material Fingerprints

Band structures are transformed into band distribution plots **(1)** and then converted into materials fingerprints **(2)**.



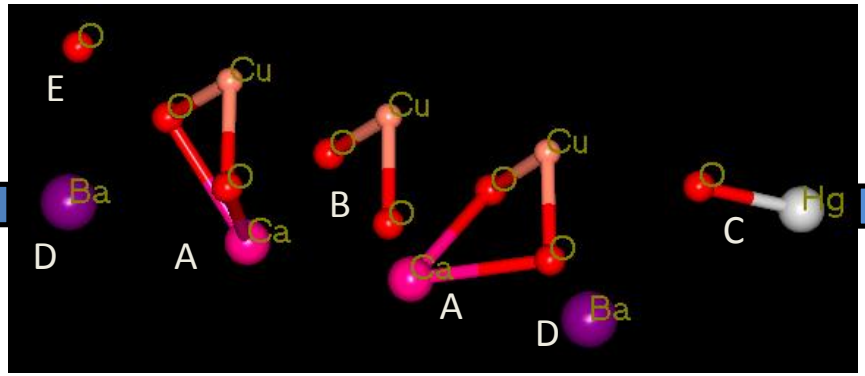


Electronic band structures computed with DFT



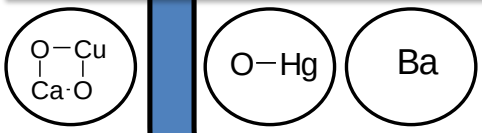
DOS Fingerprints (256 bit per DOS)

SiRMS Fragmental Fingerprints

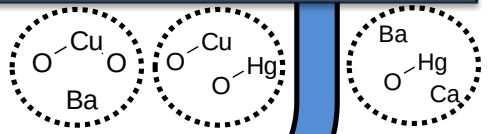


$Ba_2Ca_2Cu_3HgO_8 = A_2BCD_2E$

Traditional simplexes for material constituents taken separately
D(A), D(B), etc.



"**Mixture**" simplexes for material constituents taken separately
D(A+B), D(A+B+C), etc.



Descriptors of material

$n_A*D(A) + n_B*D(B) + \ldots$

$n_A*D(A+B) + n_B*D(B+C+E) + \ldots$

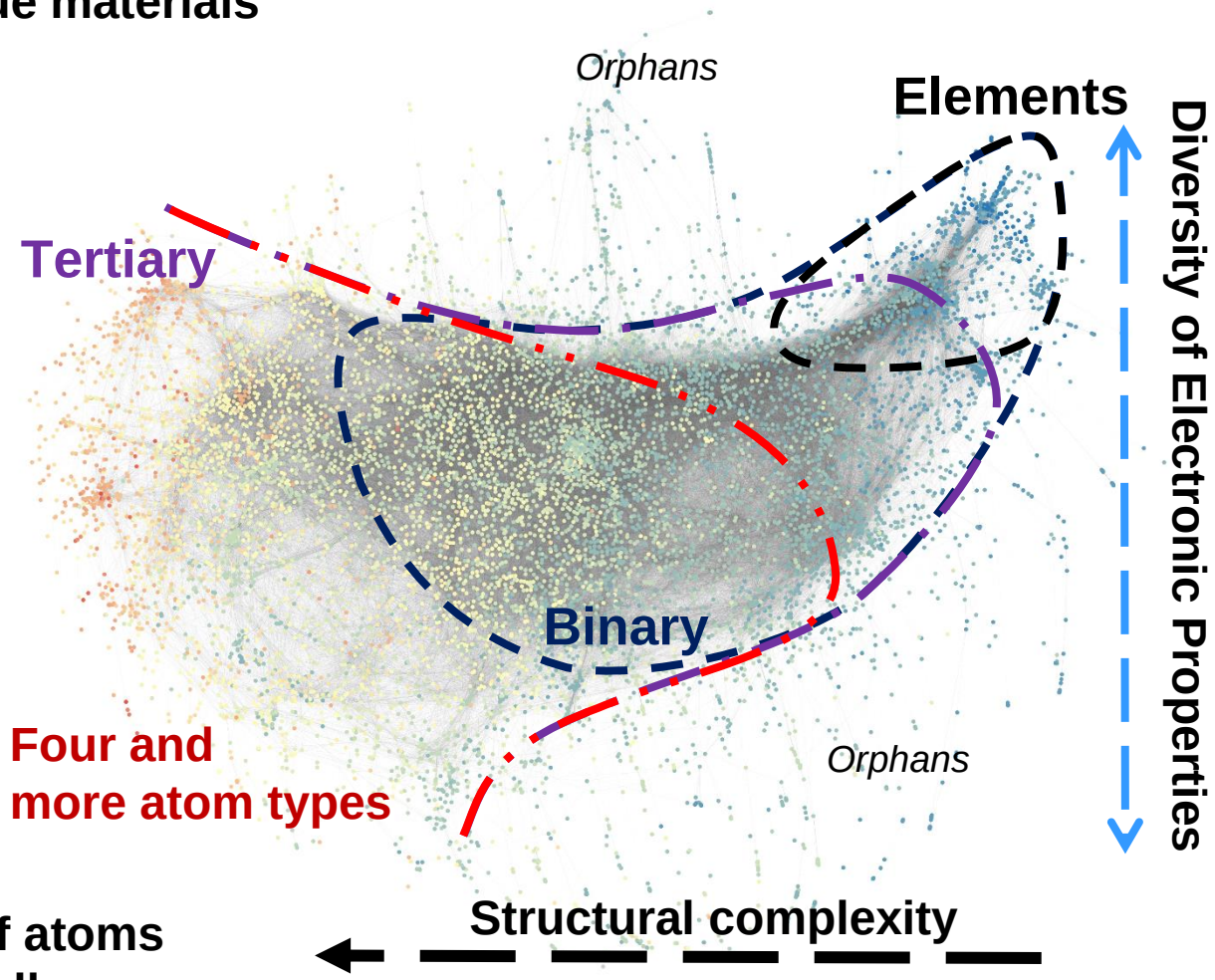
n_A, n_B, etc., are stoichiometric fractions of constituents A, B, …; $n_A < n_B < \ldots$

Dr. Eugene Muratov (UNC-CH)

Materials cartography

ADDAGRA Network* based on DOS fingerprints

17,420 unique materials



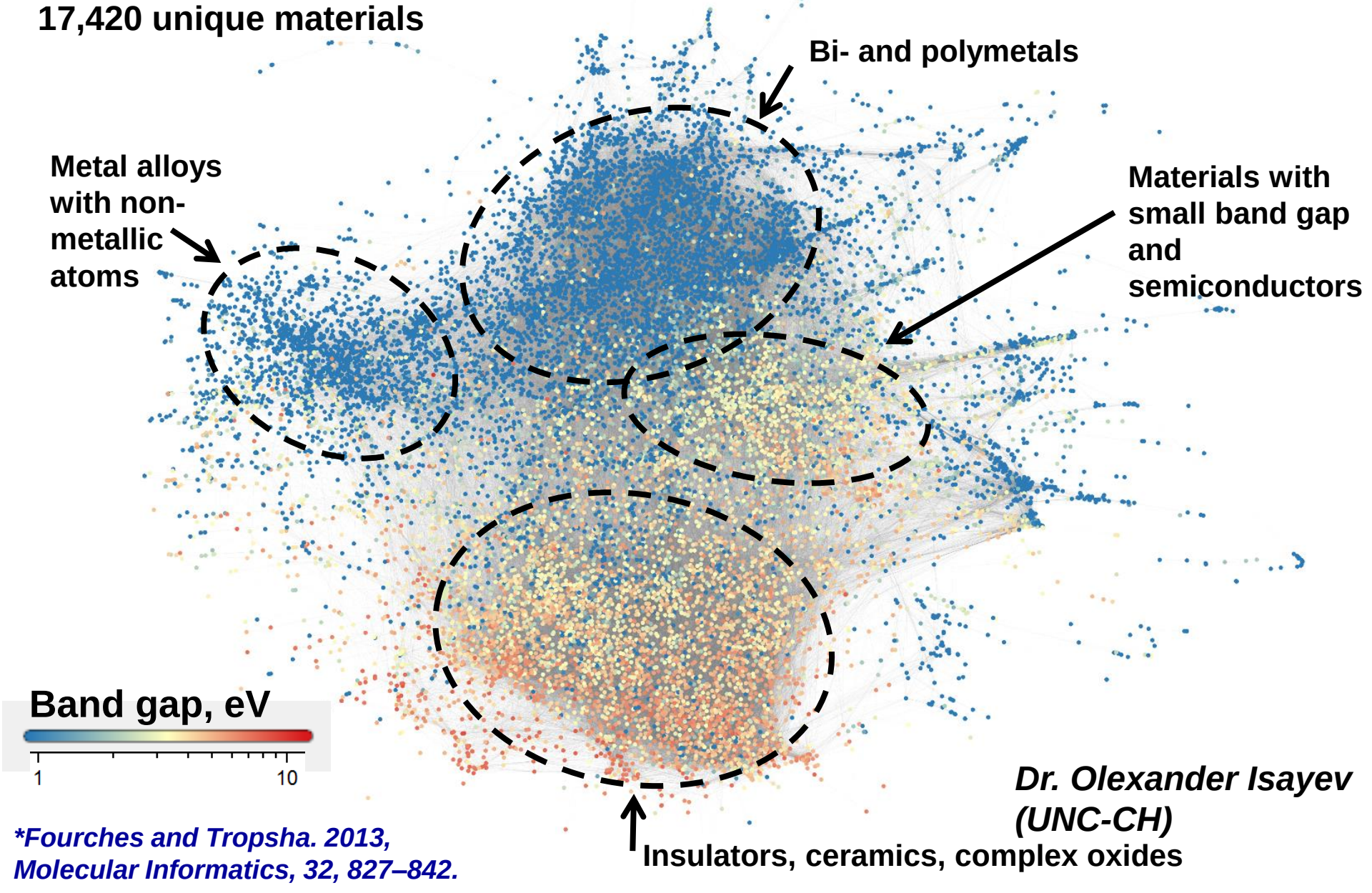
Dr. Olexander Isayev (UNC-CH)

***Fourches and Tropsha. 2013, Molecular Informatics, 32, 827–842.**

Materials cartography

ADDAGRA Network* based on band structure fingerprints

17,420 unique materials

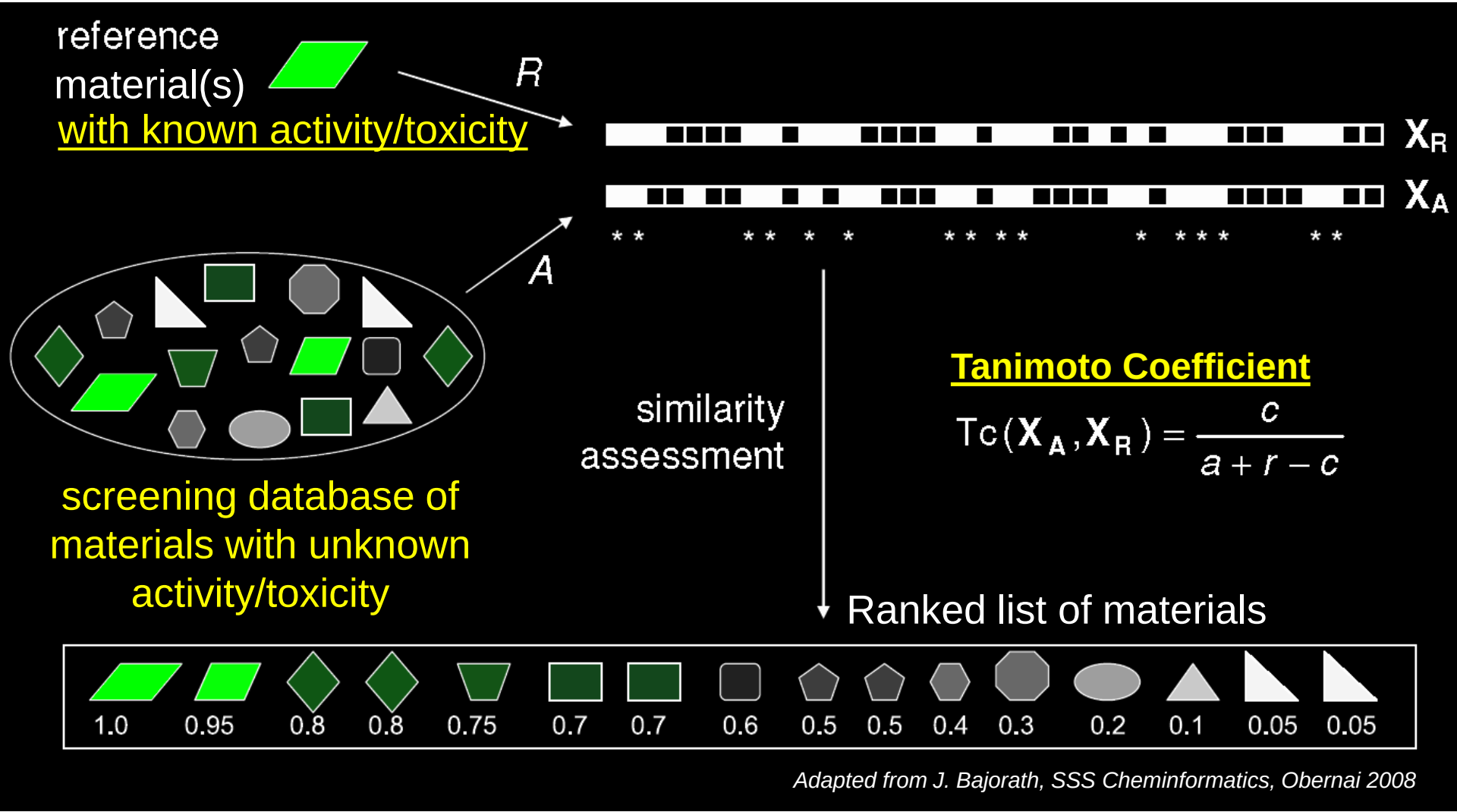


Dr. Olexander Isayev (UNC-CH)

**Fourches and Tropsha. 2013, Molecular Informatics, 32, 827–842.*

Similarity Search

Similarity searching using fingerprint representations of chemicals is one of the most widely used approaches for chemical database mining: it assumes that similar compounds possess similar biological activities.



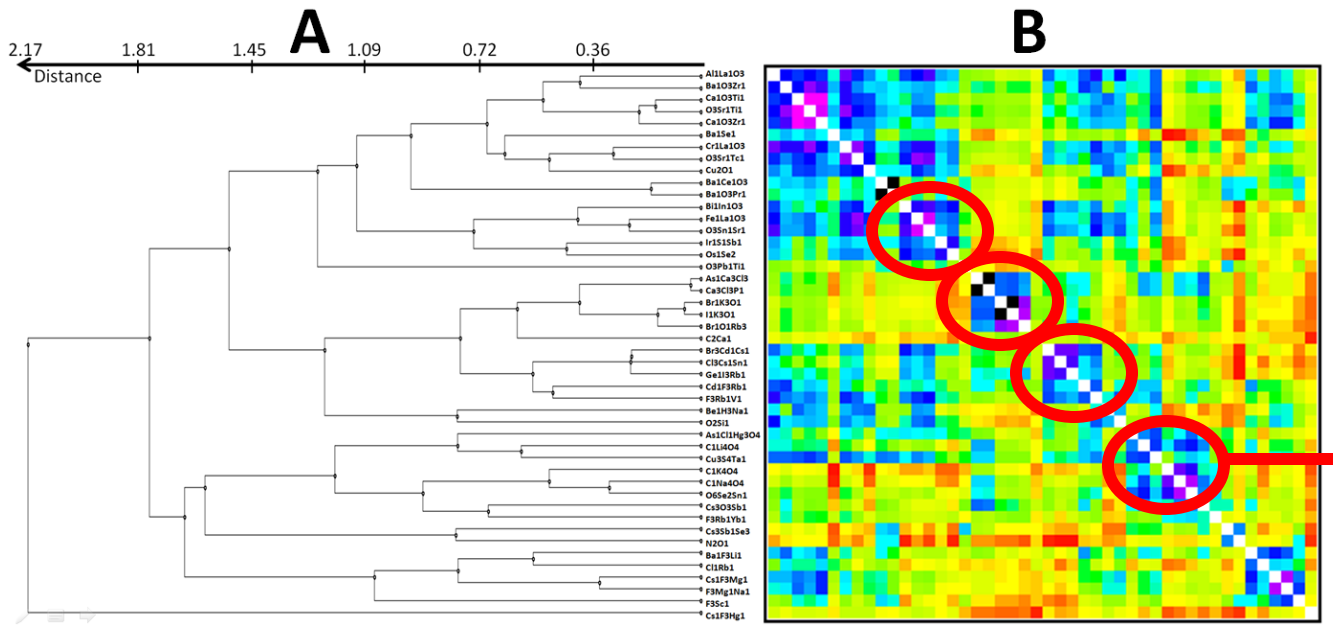
Adapted from J. Bajorath, SSS Cheminformatics, Obernai 2008

Materials Similarity based on their Fingerprints

Tanimoto similarity coefficient S between materials A and B is calculated as follows:

$$S_{A,B} = [\sum_{j=1}^{j=n} x_{jA}x_{jB}]/[\sum_{j=1}^{j=n}(x_{jA})^2 + \sum_{j=1}^{j=n}(x_{jB})^2 - \sum_{j=1}^{j=n} x_{jA}x_{jB}]$$

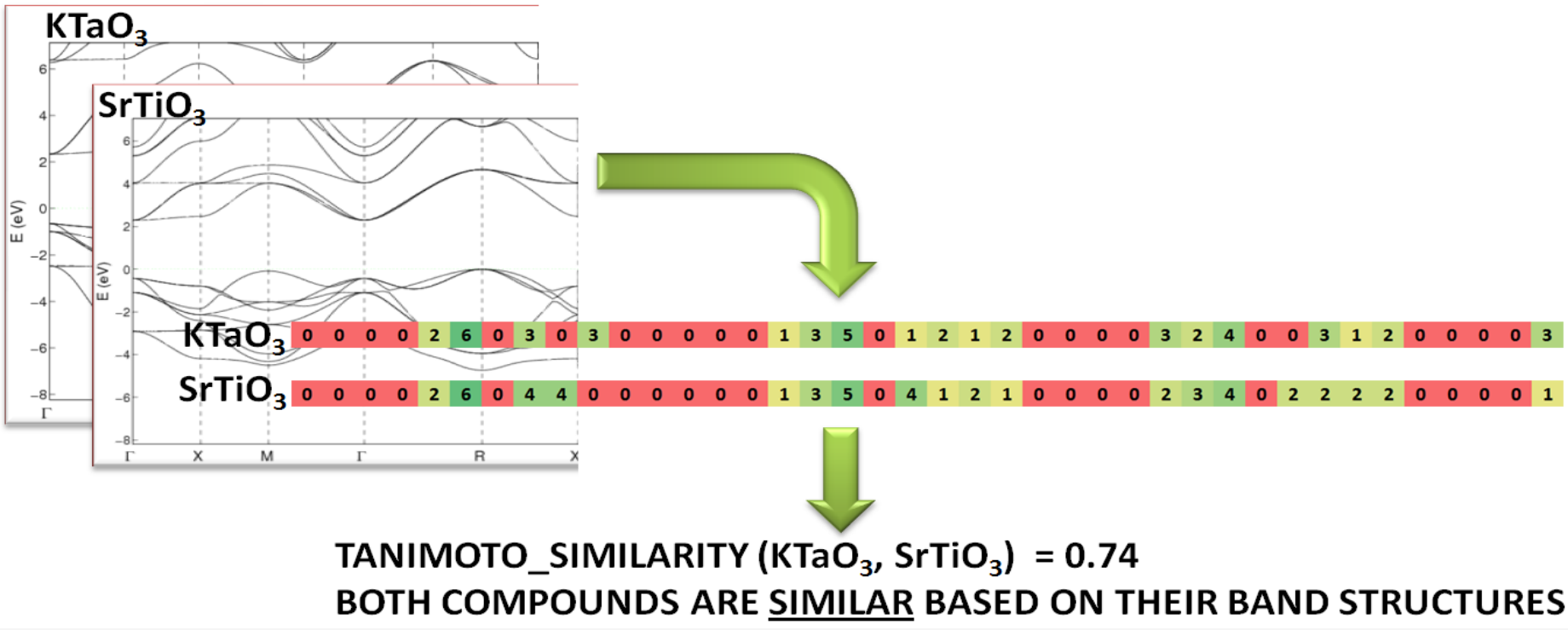
with x_j is the value of the j^{th} descriptor and *n* the total number of descriptors in the fingerprints. Tanimoto similarities are ranging from 0 (no similarity between materials A and B) to 1 (A and B are identical).



Hierarchical clustering of 46 materials according to their fingerprints.

Cluster Analysis: within a given cluster, analysis of physical/chemical/ biological properties of materials

Reprofiling materials with the desired properties



TANIMOTO_SIMILARITY ($KTaO_3$, $SrTiO_3$) = 0.74

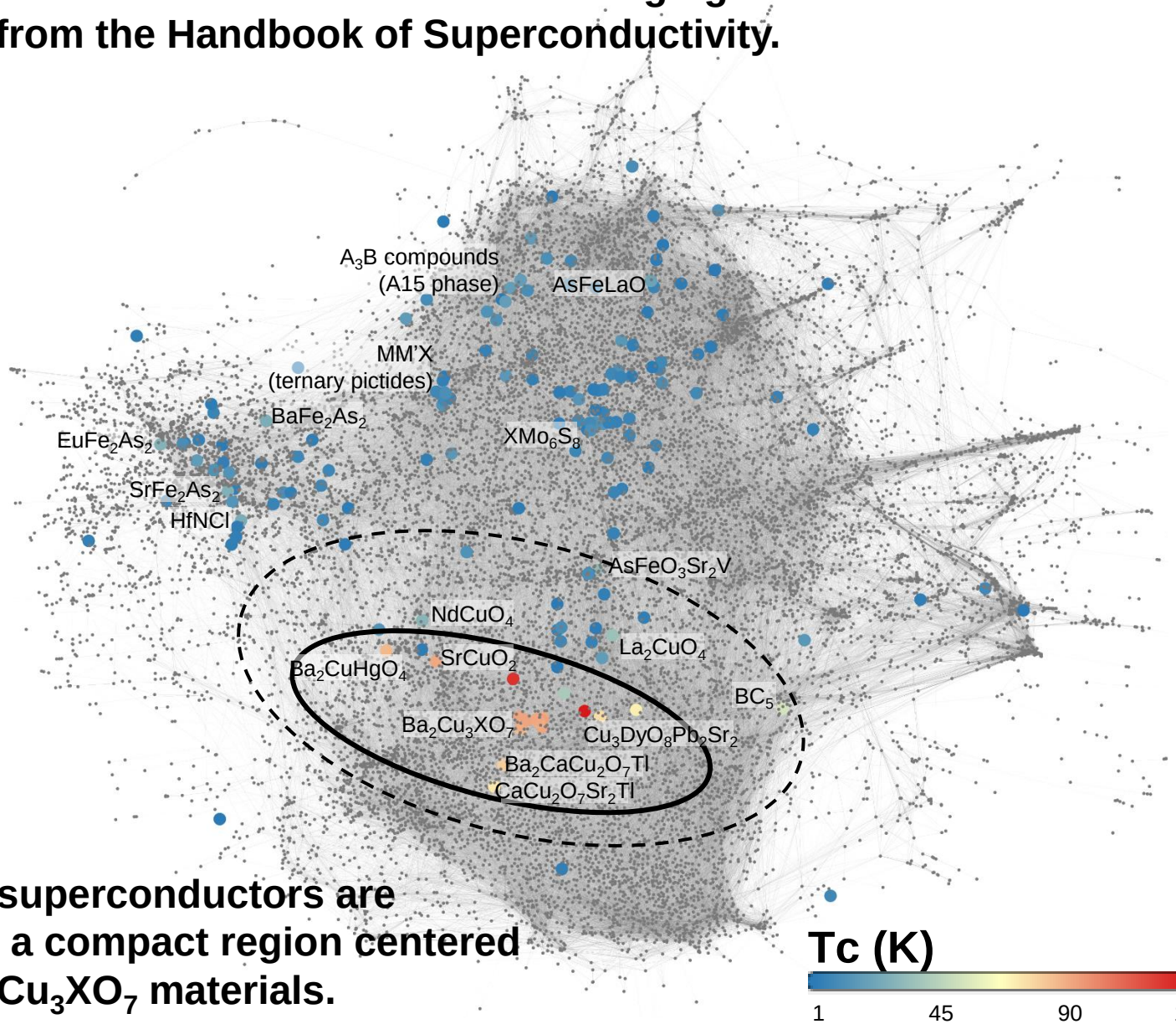
BOTH COMPOUNDS ARE SIMILAR BASED ON THEIR BAND STRUCTURES

"$KTaO_3$ is a promising candidate for superconductivity induced by electrostatic doping because it is similar to the superconductor $SrTiO_3$: [...] have similar band structures, and both exhibit quantum para-electricity"

Ueno et al, Nature Nanotech, 2011, 6, 408-412

Modeling Materials Superconductivity

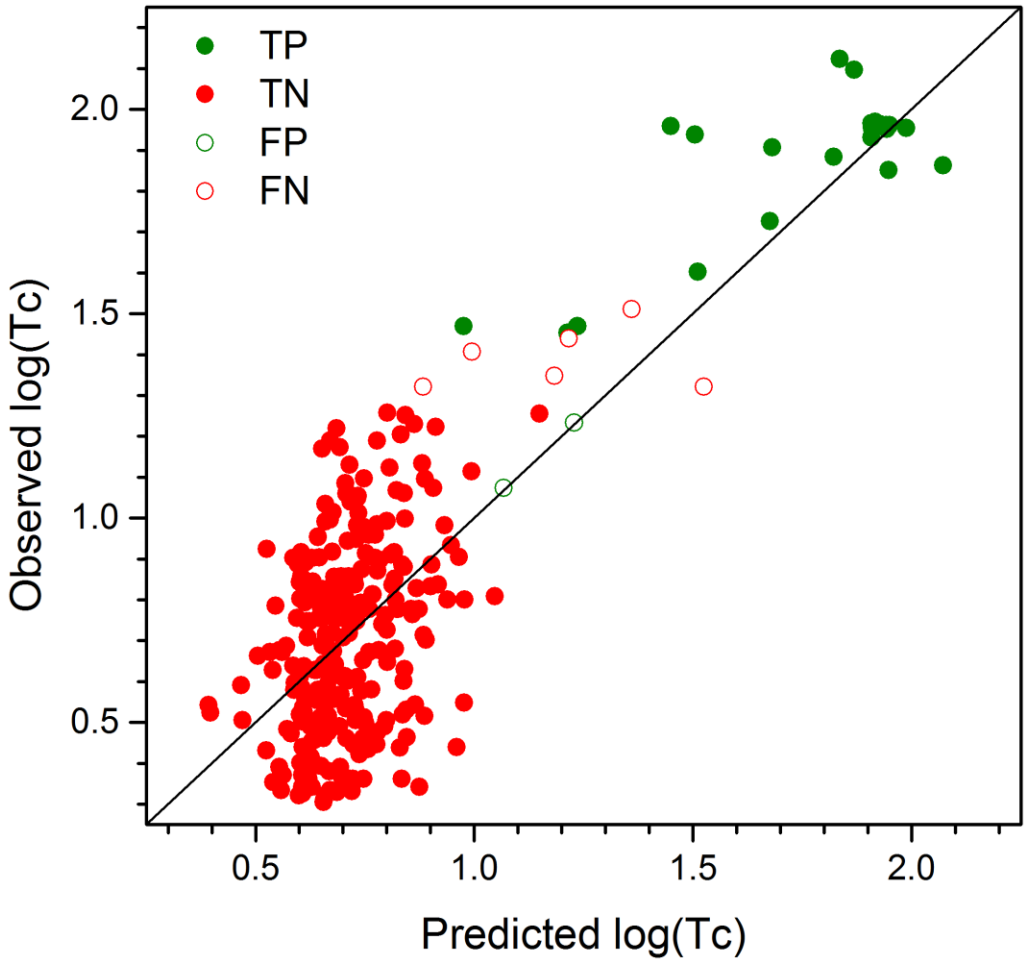
165 materials with continuous Tc values ranging from 2 K to 133 K extracted from the Handbook of Superconductivity.



All high Tc superconductors are localized in a compact region centered around $Ba_2Cu_3XO_7$ materials.

Modeling Materials Superconductivity

165 materials with continuous Tc values ranging from 2 K to 133 K



-SiRMS Fragments
-Random Forest
-5 fold external cross-validation for a rigorous assessment of models' predictivity.

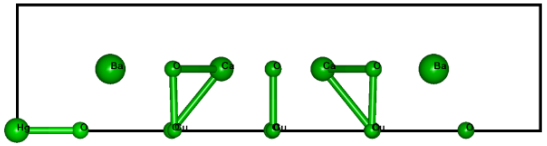
Prediction performances:

Q^2 = 0.69, MAE = 0.17

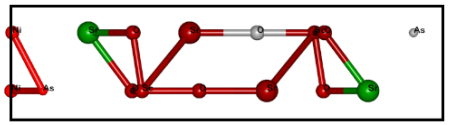
Model Interpretation to Design Novel Materials

Atoms' contributions to logTc as given by the RF model

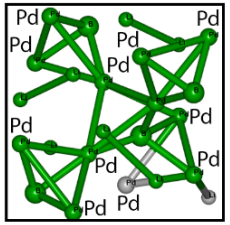
$Ba_2Ca_2Cu_3HgO_8$ (lgTc=2.12)



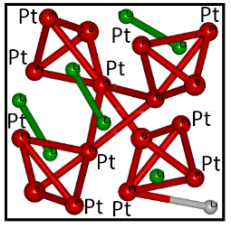
$As_2Ni_2O_6Sc_2Sr_4$ (lgTc=0.44)



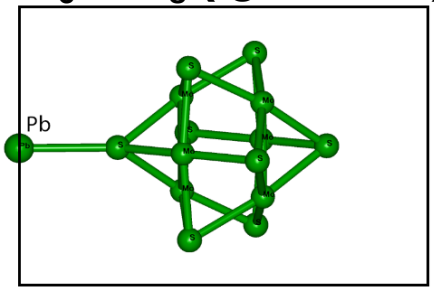
BLi_2Pd_3 (lgTc=0.89)



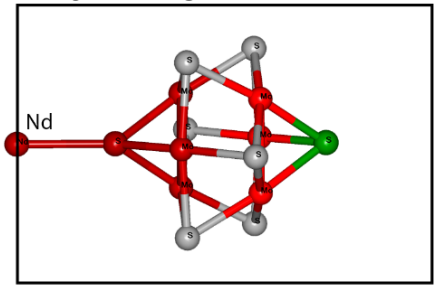
BLi_2Pt_3 (lgTc=0.49)



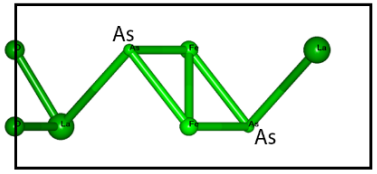
Mo_6PbS_8 (lgTc=1.13)



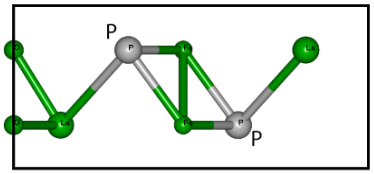
Mo_6NdS_8 (lgTc=0.54)



FeLaAsO (lgTc=1.32)



FeLaPO (lgTc=0.82)



Deliverables

- **Different types of structural and electronic materials fingerprints based on BS, DOS, and SiRMS descriptors;**
- **Visualization tools to navigate the materials space;**
- **Proof-of-concept similarity searches for materials;**
- **QMSPR models of materials superconductivity.**

Future Plans

In Progress

- **Screen the whole Aflowlib library using our superconductivity models;**
- **Prioritize a few interesting materials to be experimentally tested for superconductivity based on our model's predictions;**
- **Assemble a standard set of MNPs fingerprints to build predictive QNAR models.**

Publications, Presentations, and Recognitions/Awards

- **Grant awarded by Office of Naval Research ($125k /year for 3 years)**
- **D.Fourches, A. Tropsha. Quantitative Nanostructure-Activity Relationships: from Unstructured Data to Predictive Models for Designing Nanomaterials with Controlled Properties. In book: *Nanotoxicology: Progress toward Nanomedicine*, 2nd Edition, CRC Press, edited by Nancy A. Monteiro-Riviere, C. Lang Tran. ISBN: 9781482203875, available 03/2014.**
- **Fourches and Tropsha. Using Graph Indices for the Analysis and Comparison of Chemical Datasets. 2013, Molecular Informatics, 32, 827–842.**
- **Cherkasov et al. QSAR Modeling: Where Have You Been? Where Are You Going To? 2014, J Med Chem, In Press.**
- **Golbraikh, Muratov, Fourches, and Tropsha. Dataset modelability by QSAR. 2013, JCIM, In Press.**
- **Fourches et al. Computer-Aided Design of Carbon Nanotubes with the Desired Bioactivity and Safety Profiles. Nature Nanotechnology. *Revised Manuscript Under Review*.**
- **Isayev et al. Materials Cartography. Nature Communications, *Submitted*.**

Industrial Interactions and Technology Transfer

- **Duke University – Dr. Curtarolo**
  DFT calculations for superconductive materials

- **Office of Naval Research – Development of Predictive Materials Fingerprints for designing materials with desired properties**

- ***In Preparation* – NSF grant proposal for materials modeling & design**

- ***In Preparation* – NSF grant proposal for nanomaterials ESH in collaboration with RTI**