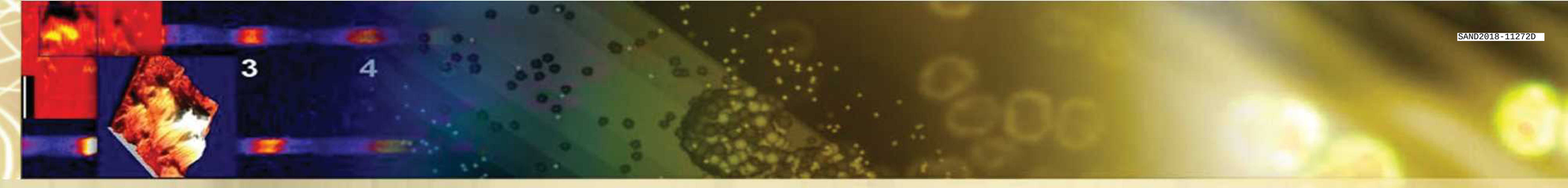
SANDIA BIOSCIENCE



SAND2018-11272D

Computational tools for advancing discovery of biofuels

Leanne Whitmore, Alexander Landera, Corey Hudson & Anthe George

Motivation

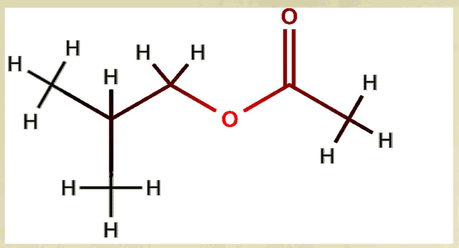
- Increase efficiency of both fuels and engines thereby decreasing the use of petroleum; specifically we aim to discover novel biofuels with the properties that will enhance the performance of the next generation fuels
- Properties that determine how a compound will perform include Research Octane Number (RON) and Reid Vapor Pressure (RVP)
- Also it is necessary to have successful and efficient biological production of the biofuel candidates in chassis microbial organisms
- These tasks are time inefficient and costly to perform on a large set of potential biofuel candidates, a suite of computational tools including, BioCompoundML/FeatureCreature, equation state models (EOS) and RetSynth were built to rapidly screen large sets of compounds for optimum fuel properties and to obtain production pathways for each compound

BioCompoundML/FeatureCreature

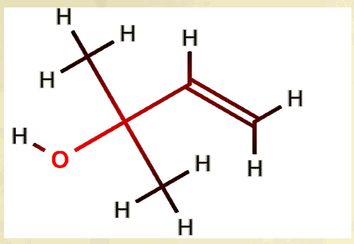
The machine learning classification algorithm Random Forest was used by BioCompoundML to predict whether a compound has a high or low RON value. Additionally, using the interpretability algorithm local interpretable model agnostic explainability (LIME) FeatureCreature can describe why a molecule has a low or high RON prediction.

compound	measured RON	prediction	probability pf high RON
ocimene	72.9	not high RON	0.463
3,7-dimethyl-3-octanol	76.3	not high RON	0.76
myrcene	82.5	not high RON	0.799
α-pinene	83.3	not high RON	0.63
(R)-(+)-limonene	87.6	not high RON	0.695
linalool	96.7	unclear	0.869
eucalyptol	99.2	high RON	0.916
butyl acetate	100.7	high RON	0.99
isoamyl acetate	101	high RON	0.967
methyl benzoate	101.1	high RON	0.998
2-methyl-3-buten-2-ol	103.5	high RON	0.967
isobutyl acetate	108.7	high RON	0.977

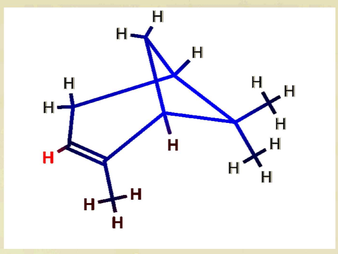
Predicted RON results for compounds which we were then able to validate. High RON is designated to be above 94.4.



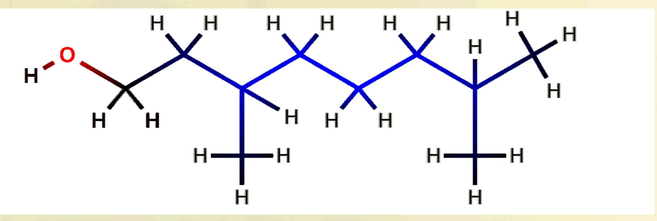
isobutyl acetate



2-methyl-3-buten-2-ol



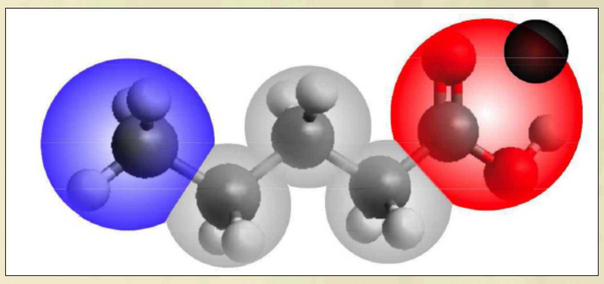
pinene



3,7 dimethyl-3-octanol

Red indicates that substructure of the compound lead to prediction of high RON while blue indicates that substructure lead to prediction of low RON.

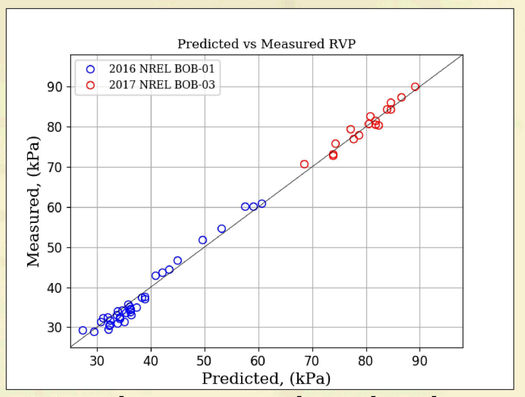
Equation of state (EOS) model



Group contribution method

$\varepsilon, \sigma, s, \lambda^r$

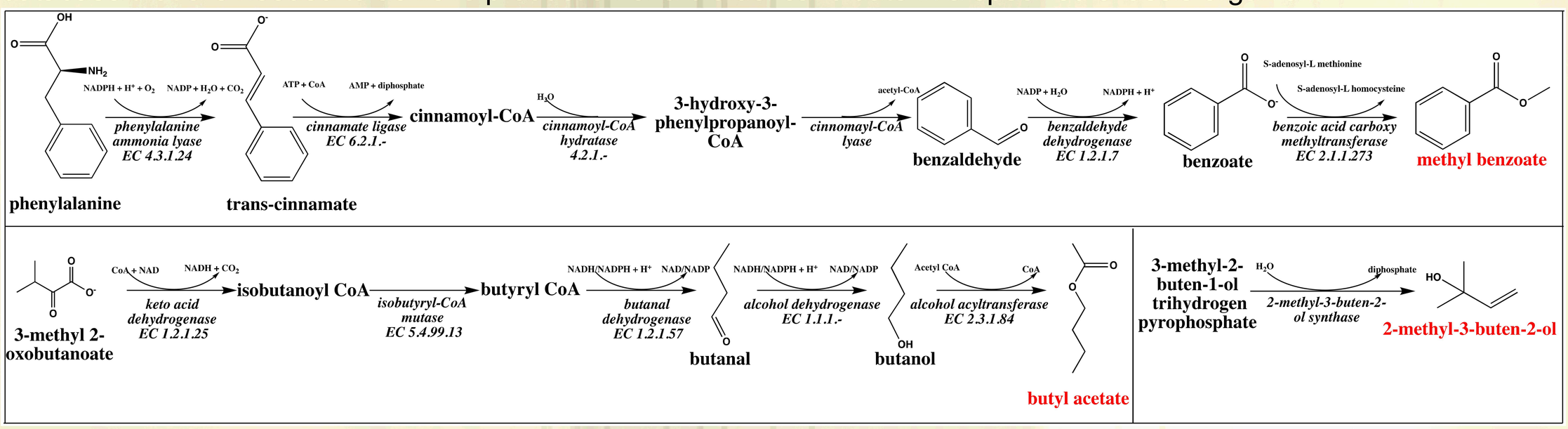
Group parameters



Accurate thermo-physical properties

RetSynth

RetSynth uses a linear programming algorithm to identify the minimal number of reaction/enzyme steps from a constructed metabolic database that are needed to achieve production of a biofuel candidate in a specified chassis organism.



Optimal pathways to target production in *Escherichia coli* K-12. Red indicates target compound.

Impact

- BiocompoundML/FeatureCreature gives insight into molecular behavior and guides decision making
- EOS modeling enables determination of the value proposition of different molecules
- RetSynth establishes viability of target molecule production

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. **SAND Number:**

U.S. DEPARTMENT OF ENERGY

Sandia National Laboratories