SAND2019-7075C

SANDIA BIOSCIENCE

Computational tools for advancing discovery of biofuels

Leanne Whitmore, Alexander Landera, Corey Hudson & Anthe George

Motivation

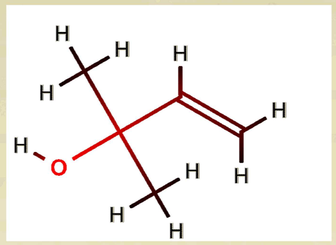
- The goal of this project is to increase efficiency of both fuels and engines thereby decreasing the use of petroleum. Specifically we aim to discover novel biofuels with properties that will enhance the performance of the next generation fuels.
- There are several measurements that determine how a compound will perform as a fuel including Research Octane Number (RON) and Reid Vapor Pressure (RVP).
  - RON is a measure of a compounds resistance to detonation under pressure. For spark ignition engines performance it is better if resistance is high.
  - RVP is a measure of volatility (how easily the gasoline evaporates). For spark ignition engines performance it is better if the volatility is low.
- Also it is necessary to have successful and efficient biological production of the biofuel candidates in chassis microbial organisms.
- Because all of 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

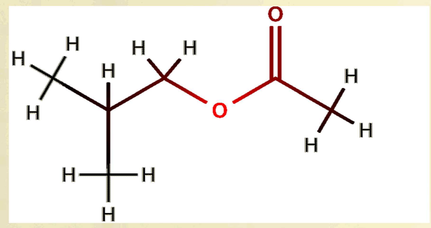
The machine learning classification algorithm Random Forest was used by BioCompoundML to predict whether a compound had a high or low RON value. Additionally, using interpretability algorithms such as local interpretable model agnostic explainability (LIME) FeatureCreature can describe why a molecule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

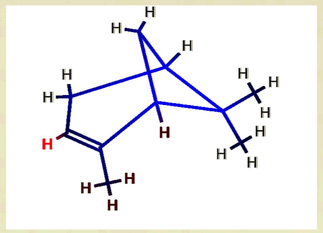
Predicted RON results for compounds which we were then able to validate. High RON is designated to be above 94.4.



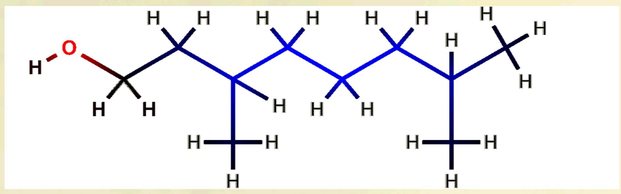
2-methyl-3-buten-2-ol



isobutyl acetate



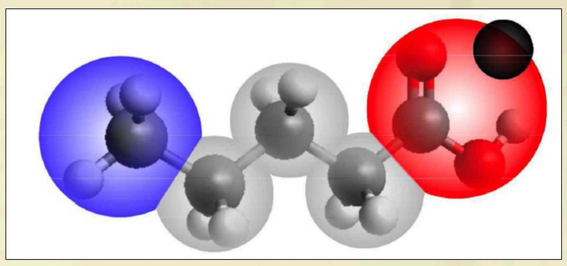
pinene



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

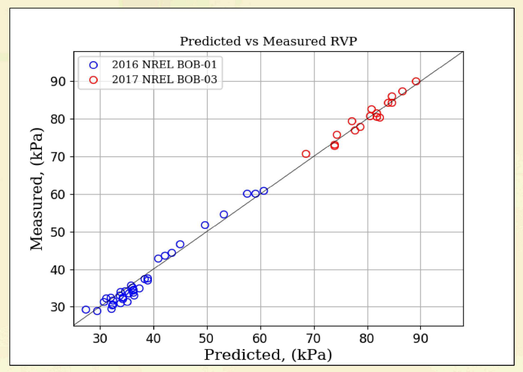


Group contribution method

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

Group parameters

ε – a measure of the strength of interaction between molecule groups
σ – a distance between interacting groups
s – factor that determines importance of group
Λ^r – measure of repulsive interactions between groups

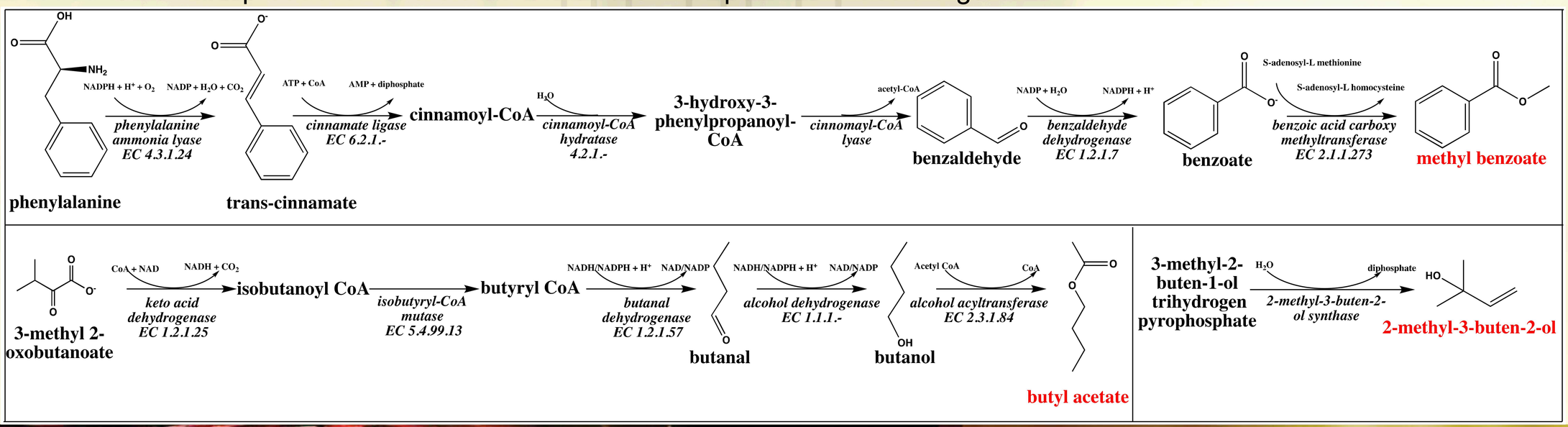


Accurate thermo-physical properties

Group contribution methods break a molecule into smaller, more manageable components. Each component is associated with group parameters, which capture the chemical behavior of that group. When combined with other groups, they can provide accurate thermo-physical properties of fluids.

RetSynth

RetSynth uses a linear programming algorithm to identify the minimal number of reaction/enzyme steps from a constructed metabolic database 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.

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