



HAL
open science

Computer Aided Molecular Design for bio-sourced molecules. The IBSS Tool

Vincent Gerbaud

► **To cite this version:**

Vincent Gerbaud. Computer Aided Molecular Design for bio-sourced molecules. The IBSS Tool. Journée d'échange sur la formulation des agents et substances de biocontrôle 2023, INRAE, Jun 2023, Maisons-Alfort, France. hal-04142541

HAL Id: hal-04142541

<https://ut3-toulouseinp.hal.science/hal-04142541>

Submitted on 27 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

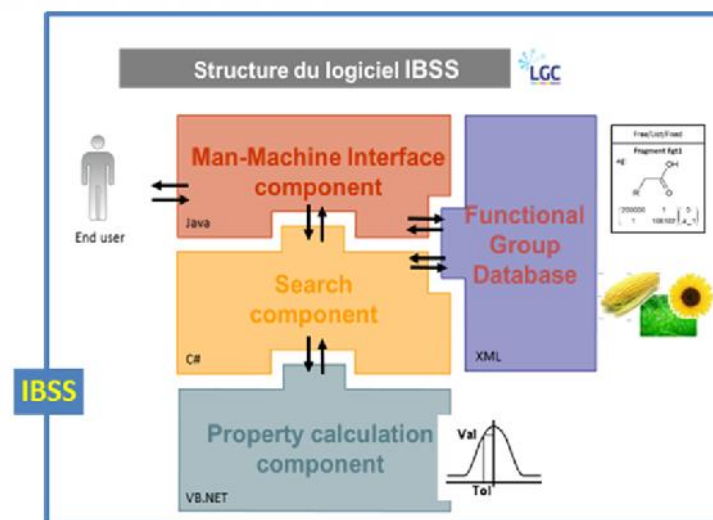
Computer Aided Molecular Design for bio-sourced molecules. The IBSS Tool

Contact person: Vincent.Gerbaud@ensiacet.fr

Co-authors: Juliette Heintz, Nishant Pandya, Moises Teles dos Santos

<http://lgc.cnrs.fr>

INBIOSYNOLV (IBSS) IN CONTEXT



J. HEINTZ, J.P. BELAUD D, V. GERBAUD. Chemical enterprise model and decision-making framework for sustainable chemical product design. *Computers in Industry Special Issue*, 65(3), 505-520, 2014.

J. HEINTZ, J.P. BELAUD, N. PANDYA, M. TELES DOS SANTOS, V. GERBAUD. Computer aided product design tool for sustainable chemical product development. *Comput. Chem. Eng.*, 71, 362-376, 2014.

- **Motivation:**

- Finding new molecules or substituting existing molecules by greener and safer ones

- **Challenges for substitution**

- Multiple molecular requirements

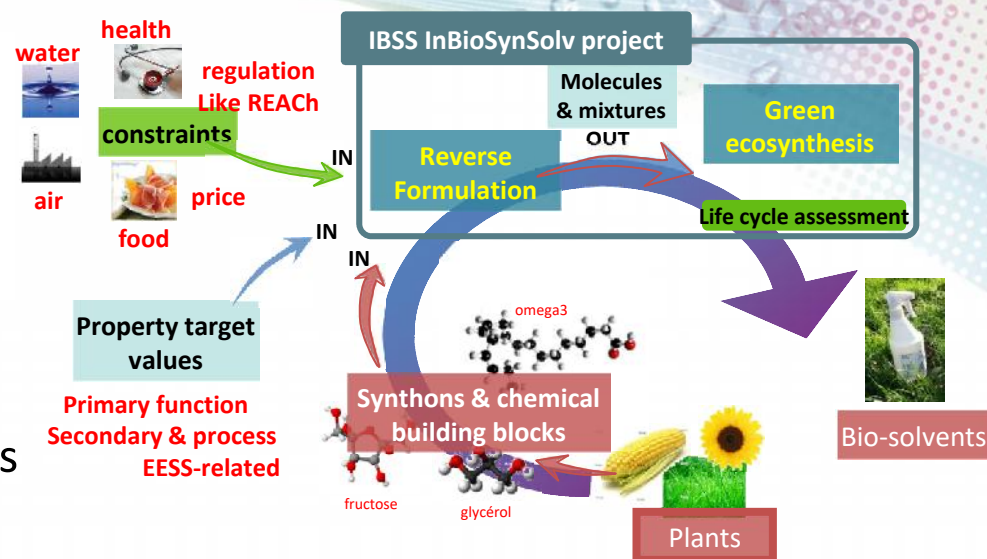
- Low price of chemical synthesis
- Keep at least one key property (eg. solubilizing an active principle)
- Keep process & usage properties (surface tension, viscosity, vapor pressure)
- Improve EHS / Environment Health Safety properties

» *Improve market added value: “green product issued from renewable materials”*

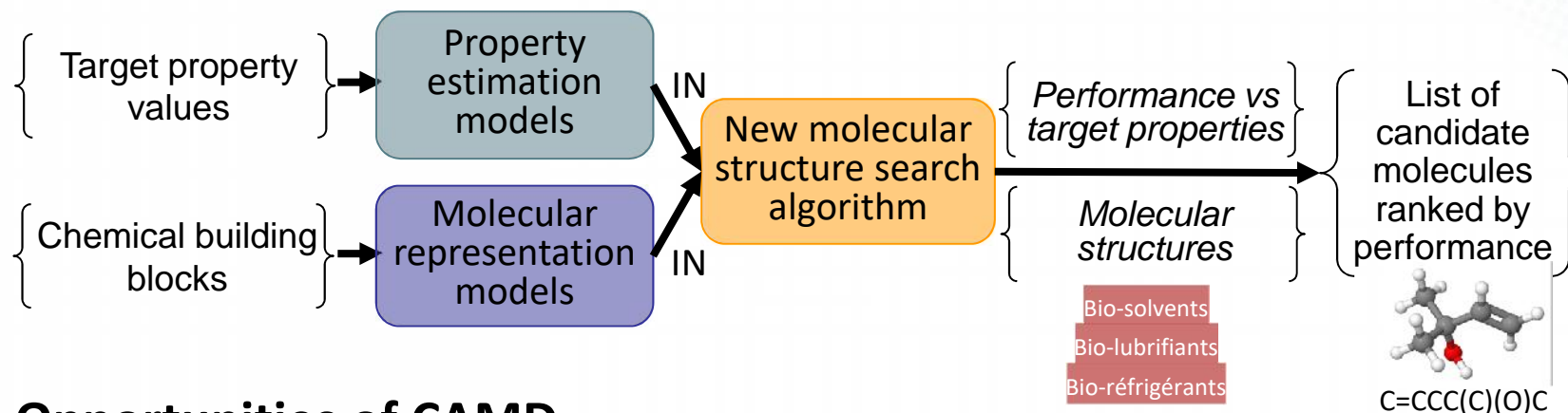
- Need of a screening method that will explore many chemical structures and match multiple properties at the same time.

- > 30 millions molecules are registered but 10 000 are well described with properties
- “trial and error” or “high-throughput experiments” methods are inefficient and costly and not adapted to multiple requirements

→ **What about CAMD Computer Aided Molecular Design?**



- **Computer Aided Molecular Design (Brignole & Gani, 1983)**



- **Opportunities of CAMD**

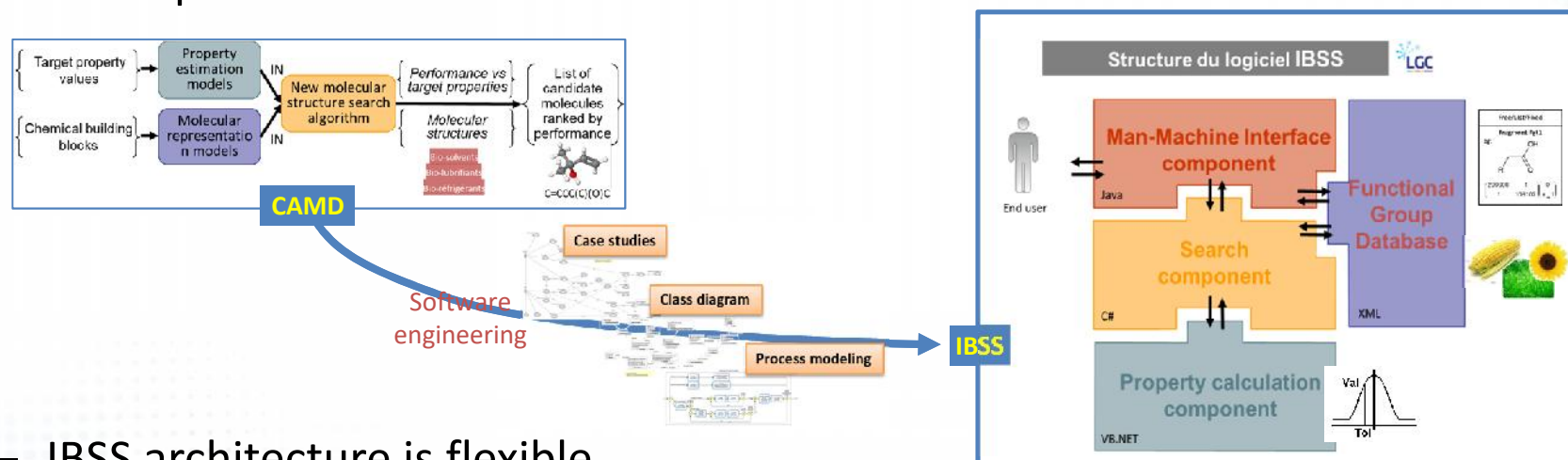
- Can explore millions of chemical structures in a few hours
- Can handle multiple properties at the same time

- **Limitations of CAMD**

- CAMD remains a screening technique → needs further experimental validation
- CAMD needs property prediction models related to molecular structures
 - Property prediction models : group contribution, QSAR, QSPR, ...
 - Accuracy of prediction is unequal: from very good to approximate
 - ! Property models accuracy is never better than experimental measurements

- The IBSS software follows the CAMD concepts but is adapted to deal with mixtures where biobased synthons can be imposed

– A component-based software



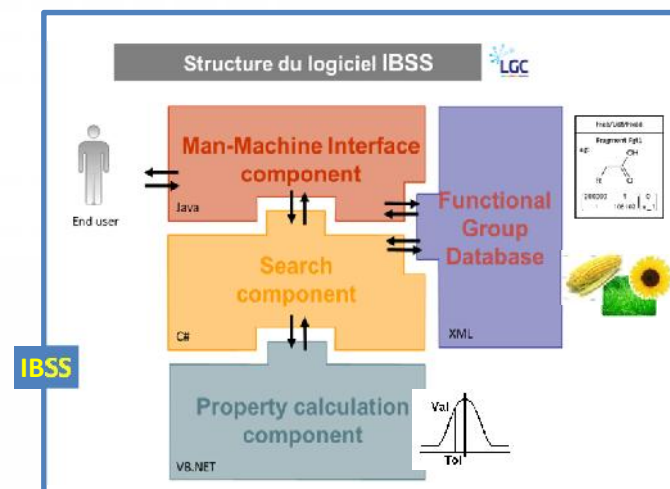
– IBSS architecture is flexible

- New property models can be added into property library
- New candidate synthons and other chemical building blocks can be added into the functional group database

– As of today, IBSS is licenced by INPT University to the Toulouse SATT but it is not distributed so far by the Toulouse SATT.

- Ask any of the authors for conditions of a further test.

INBIOSYN SOLV (IBSS) IN PRACTICE

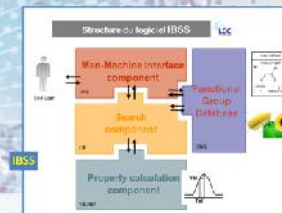


V. GERBAUD, TELES DOS SANTOS M., PANDYA N., AUBRY J.M. Computer aided framework for designing bio-based commodity molecules with enhanced properties. *Chem. Eng. Sci.*, 159, 177-193, **2017**.

L. MOITY, MOLINIER V., BENAZZOUZ A., JOOSSEN B., GERBAUD, V., AUBRY J.M. A "top-down" in silico approach for designing ad hoc bio-based solvents: Application to glycerol-derived solvents of nitrocellulose. *Green Chem.*, 18, 3239-3249, 2016.

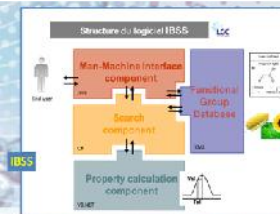
M. BERGEZ-LACOSTE, S. THIEBAUD-ROUX, P. DE CARO, J.F. FABRE, V. GERBAUD, Z. MOULOUNGUI. From chemical platform molecules to new biosolvents: Design engineering as substitution methodology. *Biofuels, Bioproducts & Biorefining.*, 8, 438-441, 2014.

What is IBSS good for?

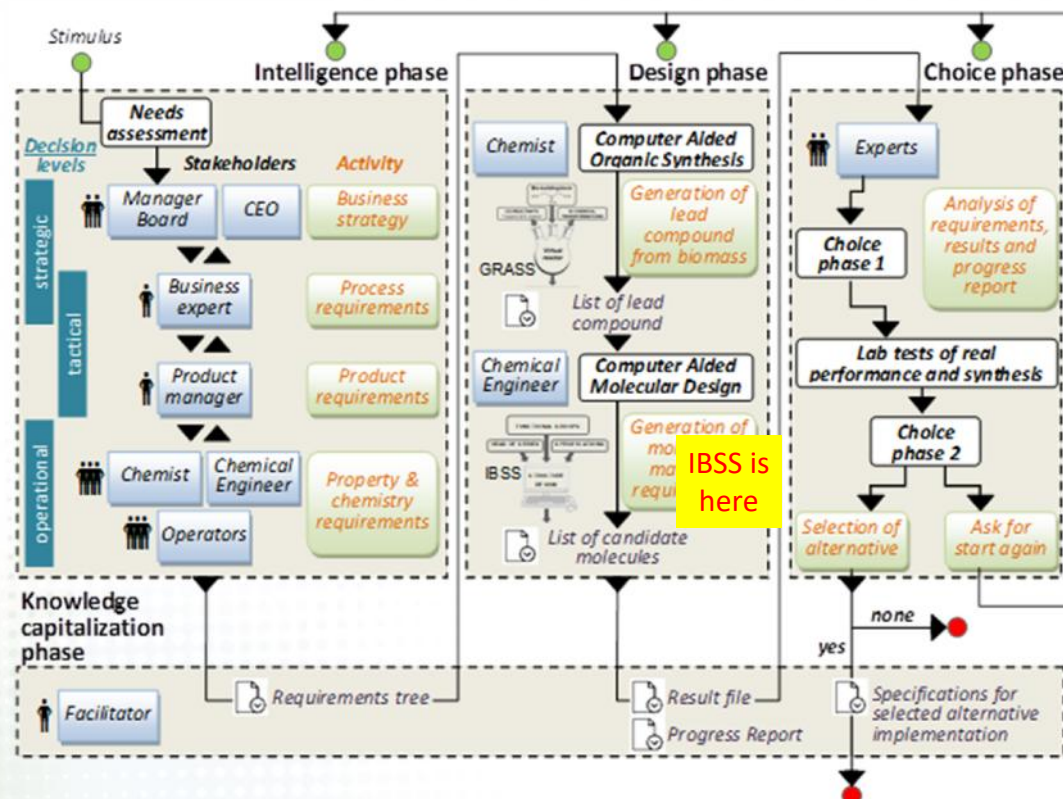


- **For design**
 - Designing molecules or mixtures that match multiple property targets aggregated in a performance function
 - *explore the chemical world efficiently*
 - Designing molecules from any set of building blocks or biasing the design towards molecules that contain bio-sourced synthons
 - *design biosourced molecules*
- **For property estimation**
 - Evaluate the performance of a list of molecules with respect to an objective function consisting of multiple property target values.
 - *estimate never measured properties; compare property model with measured values*
 - Use property prediction models of different accuracy (and computer time) for the same property within a multilevel accuracy scheme
 - *compare property models*
- **For screening**
 - Find how chemical group & functions impact the performance value (and rule out chemical functions that are the least interesting!)
 - *get the ideal fingerprint of your future molecule*
 - Find if some chemicals in your company should undergo chemical improvements to match target properties.
 - *Assess the opportunity to improve the company's portfolio value*

How to properly use IBSS

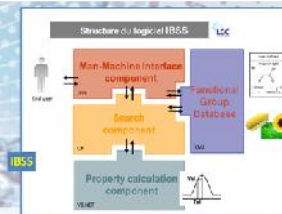


1. Gather a panel of field experts for defining requirements and analyzing results



V. GERBAUD, TELES DOS SANTOS M., PANDYA N., AUBRY J.M. Computer aided framework for designing bio-based commodity molecules with enhanced properties. *Chem. Eng. Sci.*, 159, 177-193, 2017.

→ Define which properties are important and what are the target values



2. Run IBSS for screening chemicals with multiple properties

1. Select property prediction models from a library of 39 models



- Acoustic Factor
- Enthalpy of vaporization at 298K
- Enthalpy / Standard enthalpy of Fusion
- Enthalpy / Standard enthalpy of formation at 298K
- Entropy / Entropy of Melting
- Entropy / Entropy of Vaporization at the normal boiling point
- EHS / Acute Toxicity (96-h LC50)
- EHS / Aqueous Solubility
- EHS / Bioconcentration Factor
- EHS / Environmental Impact Index
- EHS / Environmental Waste index

→ set target values (or range)

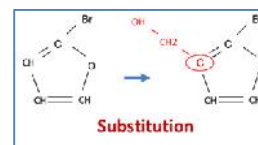
→ define property weights

→ define the performance objective function

2. Select building blocks from the database
(> 100 chemical groups + > 300 biosourced synthons)

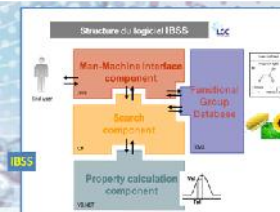


3. Run the search for new structures



4. Analyse results and select the best candidates worth more advanced investigation





2. Run IBSS for screening chemicals with multiple properties



1. Select property prediction models from a library of 39 models

→ set target values (or range)

→ define property weights

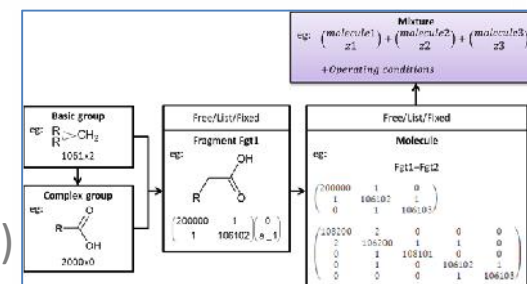
→ define the performance objective function

Property	Factor
Enthalpy of vaporization at 298K
Enthalpy of vaporization at the normal boiling point
Enthalpy / Standard enthalpy of fusion
Enthalpy / Standard enthalpy of formation at 298K
Entropy / Entropy of Melting
Entropy / Entropy of Vaporization at the normal boiling point
EHS / Acute Toxicity (96-h LC50)
EHS / Aqueous Solubility
EHS / Bioaccumulation Factor
EHS / Environmental Impact index
EHS / Environmental Waste index

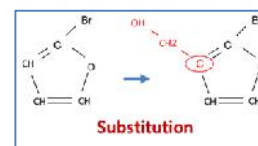


2. Select building blocks from the database and tailor the structure

(> 100 chemical groups + > 300 biosourced synthons)



2. Run the search for new structures



3. Analyse results and select the best candidates worth more advanced investigation

performance: 0.64156963588004

Molecular weights: 175.472g/mol

pe1

pe2

pe3

pe4

pe5

pe6

pe7

pe8

pe9

pe10

pe11

pe12

pe13

pe14

pe15

pe16

pe17

pe18

pe19

pe20

pe21

pe22

pe23

pe24

pe25

pe26

pe27

pe28

pe29

pe30

pe31

pe32

pe33

pe34

pe35

pe36

pe37

pe38

pe39

pe40

pe41

pe42

pe43

pe44

pe45

pe46

pe47

pe48

pe49

pe50

pe51

pe52

pe53

pe54

pe55

pe56

pe57

pe58

pe59

pe60

pe61

pe62

pe63

pe64

pe65

pe66

pe67

pe68

pe69

pe70

pe71

pe72

pe73

pe74

pe75

pe76

pe77

pe78

pe79

pe80

pe81

pe82

pe83

pe84

pe85

pe86

pe87

pe88

pe89

pe90

pe91

pe92

pe93

pe94

pe95

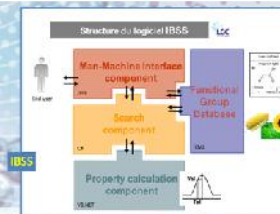
pe96

pe97

pe98

pe99

pe100



3. Post IBSS:

1. Because property prediction (and performance) evaluation can be inaccurate and mixture non linear effects are not accounted by simplest methods,
 - It is recommended to use advanced prediction methods (eg. VLE thermodynamics calculation, COSMO-RS...) on a selected list of molecules that seem worth more investigation

2. Perform experimental validation

• What the expert IBSS user can do:

– INPUT

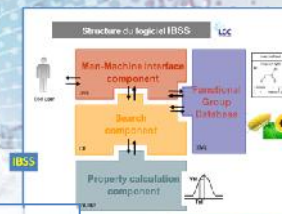
- Tailor the search algorithm parameters
- Define (and save for later) chemical structures
 - Renewable synthons
 - Building block chemical groups
- Define (and save for later) mixture parameters
 - Structure of the molecule or of the mixture

» Are some molecules fixed? Are some fragments imposed in a molecules? What are the list of building blocks to build the fragments?

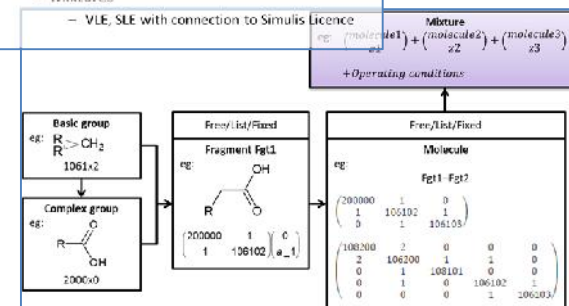
- Define (and save for later) an objective function that is used for computing a performance value

– OUTPUT

- Retrieve a list of candidate molecule or mixture
 - Performance, composition, molecular structure, 3D
 - » e.g.: a molecule improved from the in-house "glycerol29 synthon" with 9 property targets. Performance=84%

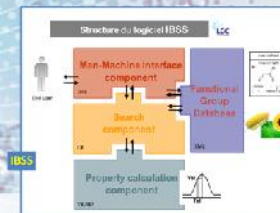


- **Property library model Component**
 - Property estimation models
 - 39 pure compound props, 10 mixture props
 - Physico-chemical props
 - » Solubility parameters, Boiling melting flash temperatures, molar volume, viscosity, heat capacity, enthalpy, ...
 - Environment indicators
 - » BCF, LCS0, Water solubility, KOW
 - 2 equilibrium properties
 - SLE and VLE
 - Model types
 - Pure compounds
 - Group contribution methods (1, 2 3rd order)
 - (extension to QSAR and QSPR planned)
 - Mixtures
 - VLE, SLE with connection to Simulis Licence



performance= 0.844196963588004
 Molecular Weight: 175.272g/mol
 Boiling Point: 570.343693162174K
 Flash Point: 442.995694777013K

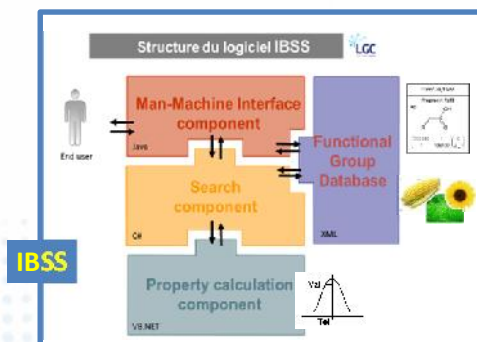
	0	1	-CH--	1	0	1
1	0	0	1	-CH--	1	0
	0	0	0	1	CH3-	0
	0	0	1	0	0	CH3-



• Case studies

- Finding biosourced solvents with low EHS impacts for agroindustry active principles
 - 9 properties (solubilization, surface tension, ... 5 EHS factors)
 - » *Patent: (WO2014132178A2) Furfural derivatives as a vehicle*
- Finding cooling mixture for airplanes
 - 6 properties working from -55°C to +55°C + nonflammable
 - » *Patents: Coolant mixtures for use in heat-exchange devices in the aerospace field (WO2017216493A1) /// Non-flammable or weakly flammable cooling mixtures, characterised by low relative volatility for two-phase heat-exchange systems (WO2017216492A1)*
- Finding solvents for perfume industry
 - » *I. Rodriguez-Donis, V. Gerbaud, S. Lavoine, S. Thiebaut-Roux. Computer aided product design of alternative solvents based on phase equilibrium synergism in mixtures. Comptes rendus de Chimie, 21, 606-621, 2018.*
- Finding greener solvents for photovoltaic panel production
 - » *Wang, J., Rodriguez-Donis, I., Thiebaut-Roux, S., Ibraikulov, O., Leclerc, N., Lévêque, P., Gerbaud, V., Kohlstädt, M., Heiser, T. Selection of green solvents for organic photovoltaics by reverse engineering. Mol. Syst. Des. Eng., 7, 182-195, 2022.*
- Substituting hexane in metabolite extraction from biomass feedstock
 - » *Nehmeh, M., Rodriguez-Donis, I., Cavaco-Soares A., Evon, P., Gerbaud, V., Thiebaut-Roux, Bio-refinery of oilseeds: oil extraction, secondary metabolites separation towards protein meal valorisation – A review. Processes, 10(5), 841, 2022.*

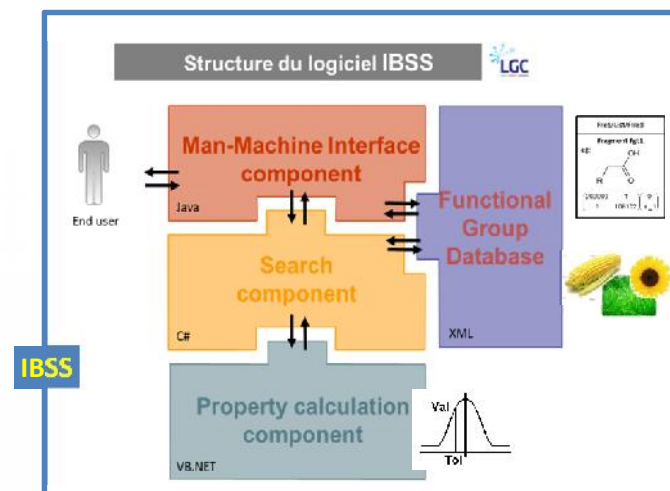
thank you



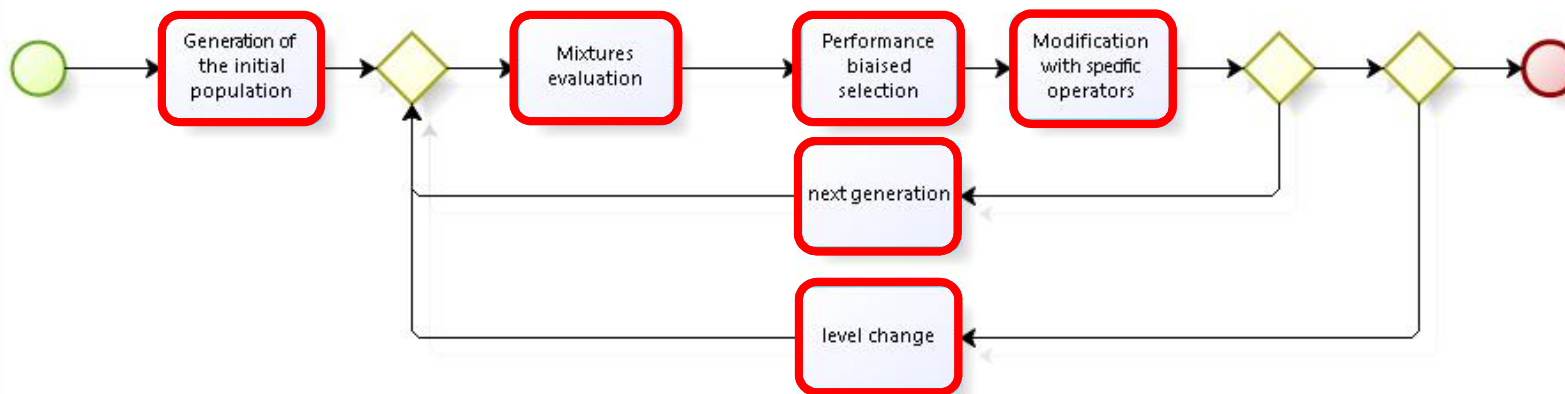
Contact person: Vincent.Gerbaud@ensiacet.fr



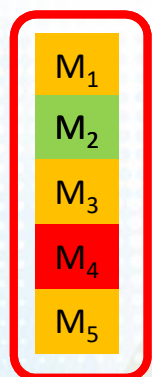
IBSS IN DETAILS



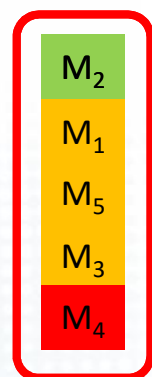
- SEARCH Component  uses a Genetic algorithm



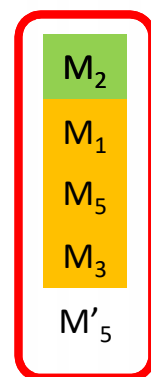
Current generation population



Performance ranking

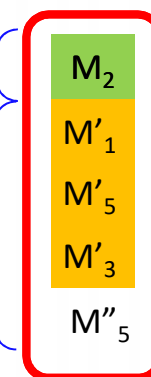


Removal worst + generation new candidate



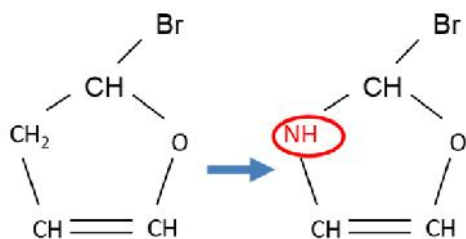
Keep the best (elitism)

Structure modification

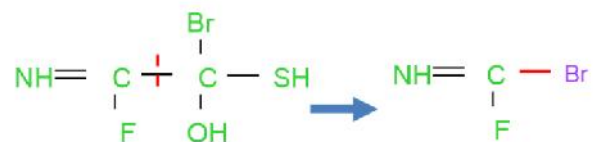


Etc...

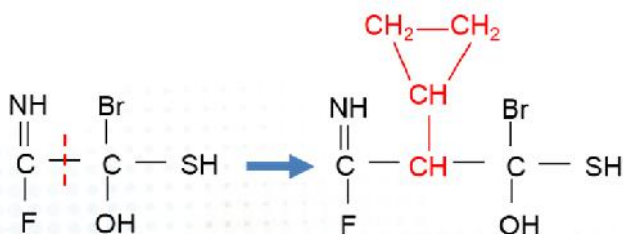
- Operators for modifying molecular structures



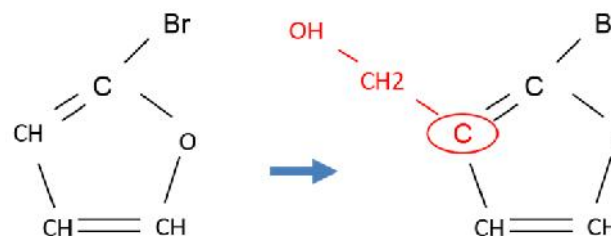
Mutation



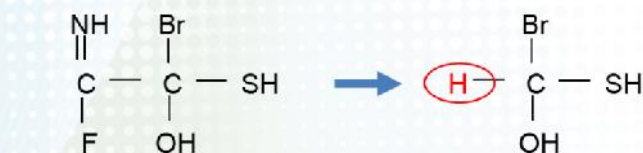
Crossover



Insertion

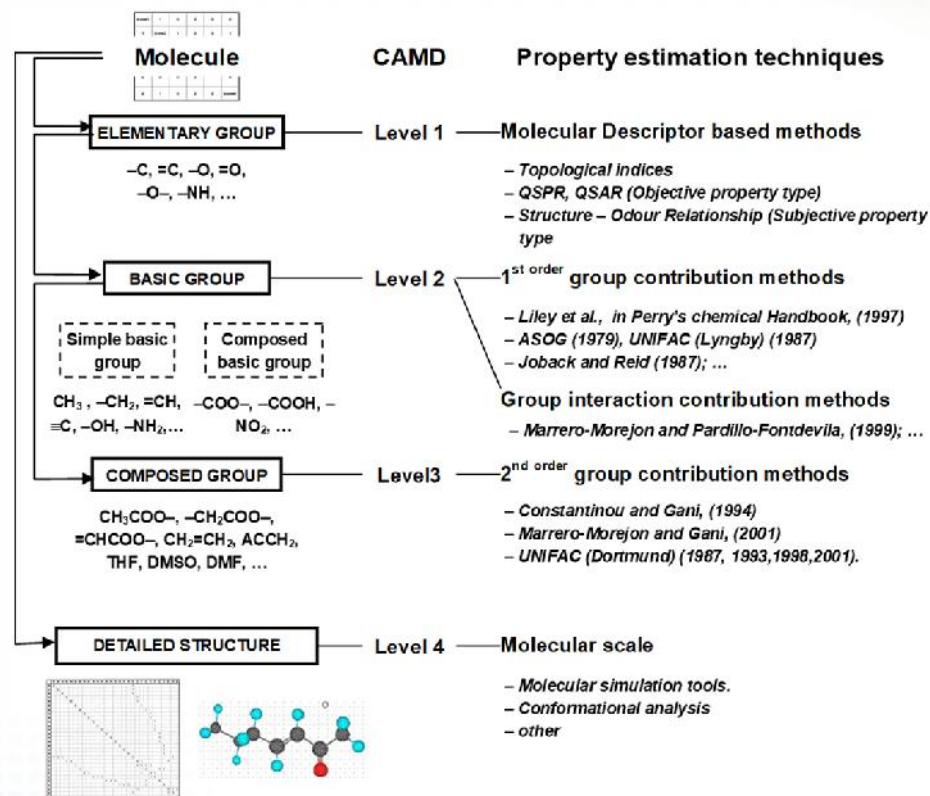


Substitution



Deletion

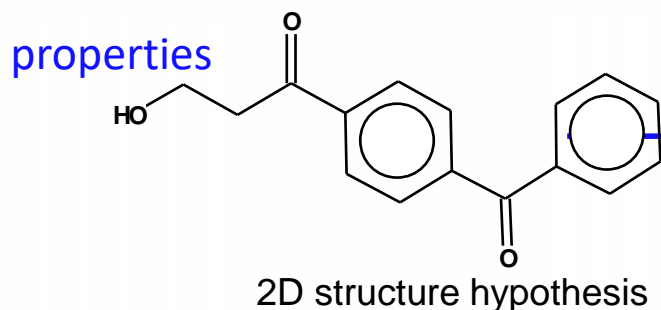
- Pure compound



- Mixture models

- Linear mixing rule is rarely good but non linear models are rare!
- Accurate mixture models require thermodynamic based models
 - Much more computer time demanding and often with dedicated solving strategy -> test in IBSS have concerned small scale problems only.

Molecular representation (ex. chemical groups)



OCCC(=O)c1ccc(cc1)C(=O)c2ccccc2

SMILES (*Simplified Molecular-Input Line-Entry System*)

- Prediction of properties $f(X)$ from group contributions

$$f(X) = \sum_i N_i C_i + w \sum_j M_j D_j + z \sum_k E_k O_k$$

- 2 aC-CO
- 9 aCH
- 1 aC
- 1 OH
- 2 CH2

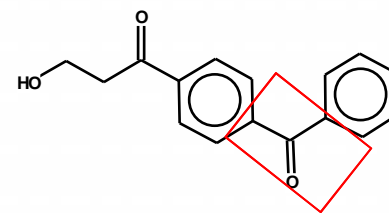
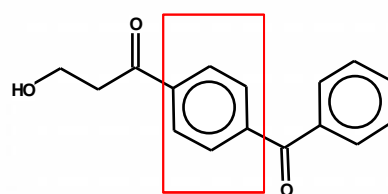
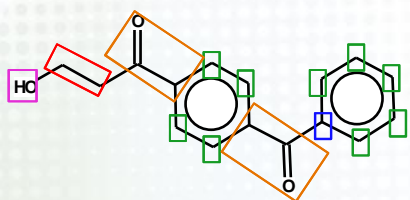
level 1

level 2

level 3

1 AROMRINGS1s4

1 aC-CO-aC (different rings)



- **Property library model Component**

- Property estimation models

- 39 pure compound props, 10 mixture props

- Physico-chemical props

- » *Solubility parameters, Boiling melting flash temperatures, molar volume, viscosity, ehat capacity, enthalpy, ...*

- Environment indicators

- » *BCF, LC50, Water solubility, KOW*

- 2 equilibrium properties

- SLE and VLE

- Model types

- Pure compounds

- Group contribution methods (1, 2 3rd order)

- (extension to QSAR and QSPR planned)

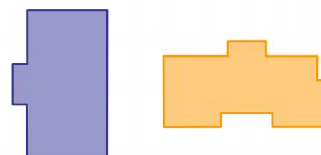
- Mixtures

- VLE, SLE with connection to Simulis Licence

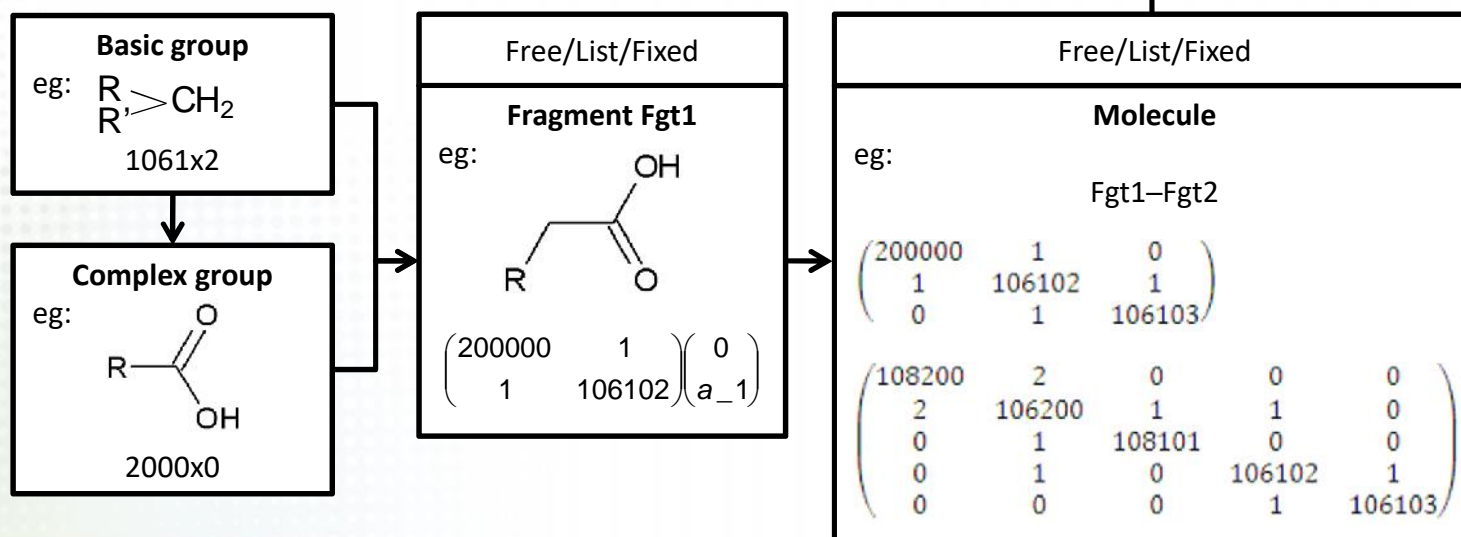
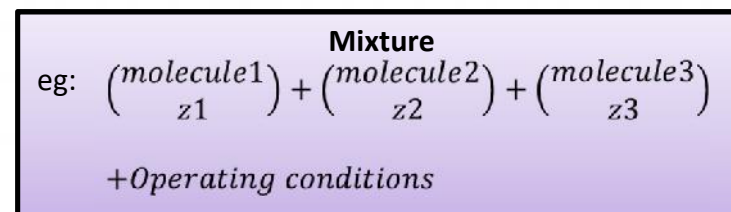
Acentric Factor	
Enthalpy of vaporization at 298K	
Enthalpy of vaporization at the normal boiling point	
Enthalpy / Standard enthalpy of fusion	
Enthalpy / Standard enthalpy of formation at 298K	
Entropy / Entropy of Melting	
Entropy / Entropy of Vaporization at the normal boiling point	
EHS / Acute Toxicity (96-h LC50)	
EHS / Aqueous Solubility	
EHS / Bioconcentration Factor	
EHS / Environmental Impact index	
EHS / Environmental Waste index	
EHS / Health index	
EHS / Safety index	
EHS / LCA index	
EHS / Kow / Octanol - water partition coefficient	
Gibbs energy / Standard Gibbs energy at 298K	
Solubility parameter / Hansen solubility parameter - dispersion (D)	
Solubility parameter / Hansen solubility parameter - Polar(P)	
Solubility parameter / Hansen solubility parameter - H ₂ -bond (H)	
Solubility parameter / Hansen solubility parameter / HSP Distance	
Solubility parameter / Hildebrand Solubility parameter	
Solubility parameter / RED – Relative Energy Difference	
Heat capacity / Cp - Ideal gas	
Pressure / Critical Pressure	
Pressure / Vapor Pressure	
Surface Tension at 298 K	
Surface Tension (as a function of temperature)	
Temperature / Auto ignition temperature	
Temperature / Critical Temperature	
Temperature / Flash Point (in air at atmospheric pressure)	
Temperature / Normal Boiling Point	
Temperature / Normal Melting Point	
Viscosity / Viscosity at 300 K	
Viscosity / Liquid Viscosity as a function of temperature	
Volume / Liquid molar volume as a function of temperature	
Volume / Liquid molar volume at 298 K	
Volume / Critical Volume	

- **Molecular representation scheme**

- Developed to allow biased search towards bio sourced molecules
- Software components concerned are:

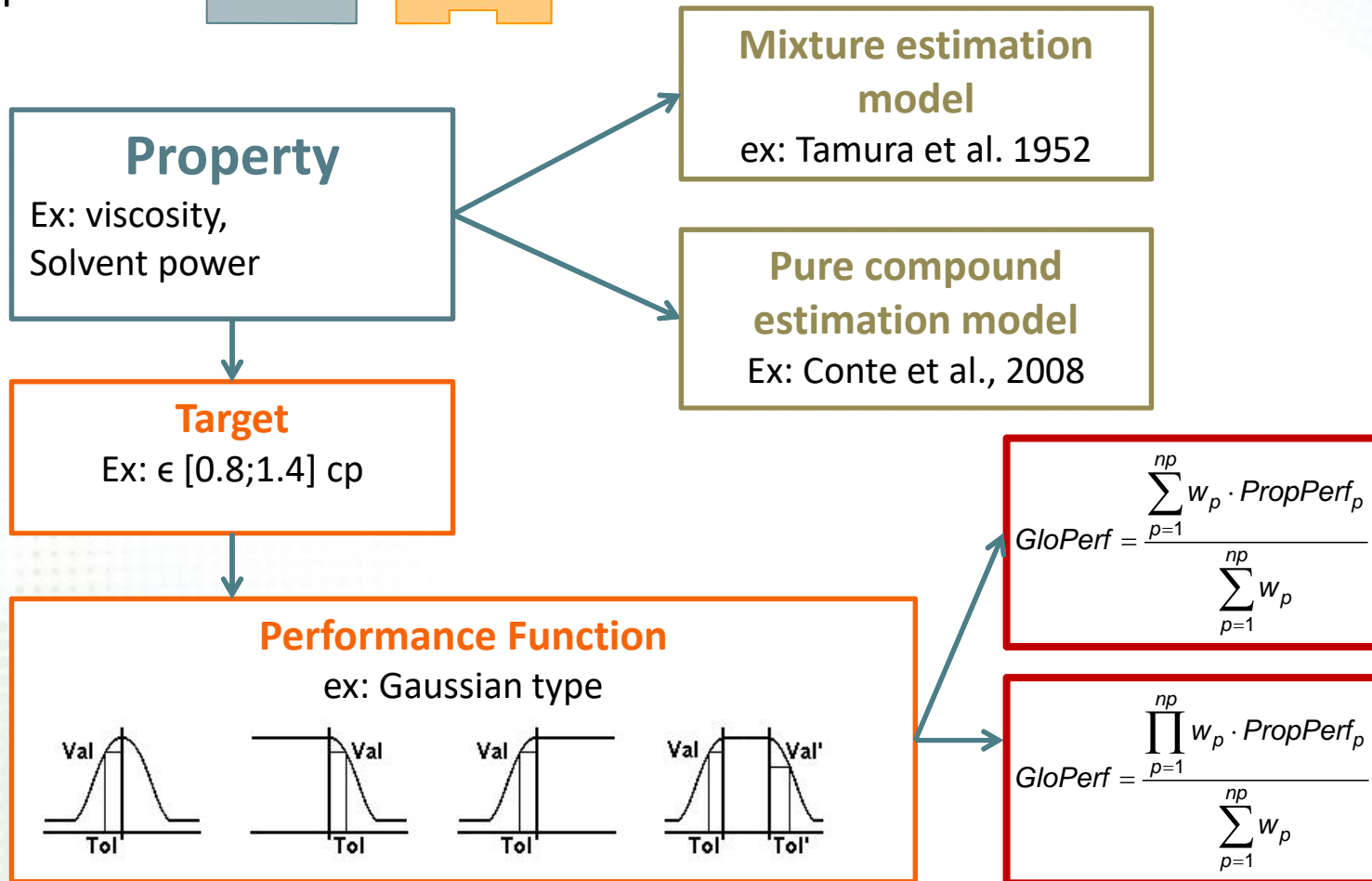


- Almost anything can be set fixed or free
 - molecule, fragment, building block

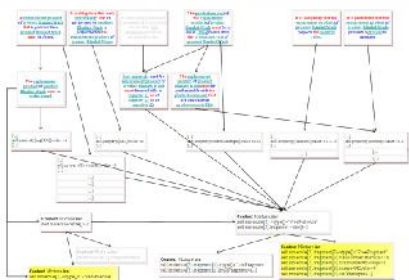


- Performance evaluation

- Components



- IBSS flowchart and file types



Final choice step

```
<?xml version="1.0" encoding="utf 8"?>
<!DOCTYPE [xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www...
<AlgorithmData>
  <Elitism>20</Elitism>
</AlgorithmData>
<MixtureData>
  <ElementData>
    <Id>1</Id>
    <MoleculeSelectionProc><MoleculeSelectionPrb>
      <Fragments>
        <PbFixedFragment>
          <FragmentSelectionPrb>5</FragmentSelectionPrb>
          <Fragments>
            <PbFixedFragment>
              <StrFragmentGraph>2003E0</StrFragmentGraph>
              <StrConnectionVector>a_1</StrConnectionVector>
            </PbFixedFragment>
            <PbFixedFragment>
              <StrFragmentGraph>2001E0</StrFragmentGraph>
              <StrConnectionVector>a_1</StrConnectionVector>
            </PbFixedFragment>
            <PbFixedFragment>
              <StrFragmentGraph>2003E0</StrFragmentGraph>
              <StrConnectionVector>a_1</StrConnectionVector>
            </PbFixedFragment>
            <PbFixedFragment>
              <StrFragmentGraph>2005E0</StrFragmentGraph>
              <StrConnectionVector>a_1</StrConnectionVector>
            </PbFixedFragment>
            <PbFixedFragment>
              <StrFragmentGraph>2005E0</StrFragmentGraph>
              <StrConnectionVector>a_1</StrConnectionVector>
            </PbFixedFragment>
            <PbFixedFragment>
              <StrFragmentGraph>2005E0</StrFragmentGraph>
              <StrConnectionVector>a_1</StrConnectionVector>
            </PbFixedFragment>
          </Fragments>
        </FragmentSelectionPrb>
      </Fragments>
    </MoleculeSelectionPrb>
  </ElementData>
</MixtureData>
```



IBSS software

N°1
 performance= 0.959492806122624
 Molecular Weight: 180.203g/mol
 Flash Point: 356.454776773145K
 Vapor Pressure: 7.47383874841577E-05bar

Rank	Performance	Molecule	SMILES	Computation
1	0.46214809	1	C=CC=CC	1
2	0.46197309	1	OC	1
3	0.46159227	1	O=C=CC	1
4	0.46105202	1	O=C=CC=C	1
5	0.46091179	1	OO	1
6	0.46044936	1	OC=O	1
7	0.449422515	1	C=CC=CC=CC	1
8	0.429743622	1	O=C=C	1
9	0.404817611	1	OC=C	1
10	0.40074117	1	COO	1

Molecular Weight: 180.203g/mol
 Flash Point: 356.454776773145K
 Vapor Pressure: 7.47383874841577E-05bar
 RED: 0.531870158002891