

# Green Chemistry

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



[www.rsc.org/greenchem](http://www.rsc.org/greenchem)



# Green Chemistry

## PAPER

### A “top-down” *in silico* approach for designing *ad hoc* bio-based solvents: Application to glycerol-derived solvents of nitrocellulose

Laurianne Moity<sup>a</sup>, Valérie Molinier<sup>\*a</sup>, Adrien Benazzouz<sup>a</sup>, Benjamin Joossen<sup>a</sup>, Vincent Gerbaud<sup>b</sup> and Jean-Marie Aubry<sup>\*a</sup>

Received 00th January 20xx,  
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Potentially effective glycerol-based solvents for nitrocellulose have been designed using a top-down *in silico* procedure that combines Computer Assisted Organic Synthesis (CAOS) and Molecular Design (CAMD). Starting from a bio-based building block – glycerol – a large number of synthetically feasible chemical structures have been designed using the GRASS (GeneratoR of Agro-based Sustainable Solvents) program. GRASS applies well-selected industrial chemical transformations to glycerol together with a limited number of relevant co-reactants. Then, the most promising structures are considered as lead compounds for further modification *in silico* thanks to the IBSS (InBioSynSolv) program, which generates derivatives with alkyl, cycloalkyl, alkene, cycloalkene or phenyl substituents. Finally, IBSS ranks all the candidates according to the value of their overall performance function to best fit the predefined specifications, *i.e.* (i) high solubilisation of nitrocellulose, (ii) slow evaporation and non-flammability (iii) low toxicity and environmental impact. This general strategy enables the highlighting of the most relevant solvent candidate derived from any building block for a given application. To validate the approach, 15 commercially available solvents derived from glycerol were confronted with nitrocellulose and led to highlight diacetin as an effective and safe solvent.

## Introduction

The search for alternative solvents is a topic of major interest from both academic and industrial points of view.<sup>1,2,3,4</sup> In some applications, especially those without appropriate safety precautions for consumers, finding alternatives to harmful traditional solvents has become mandatory. This is the case for instance for paint removing, degreasing or coating formulations that traditionally used plethora of solvents with a poor environmental footprint, such as halogenated, aromatic or glycol derivatives, which are now banned. For these formulations, in which the role of solvent is crucial and involves many different physical and chemical properties, an efficient substitution methodology is required to find the best alternative candidate(s). Among the so-called “green” solvents, which are non-toxic and have a good environmental footprint, the bio-based ones are prepared from molecules readily available from the biomass feedstock.<sup>5,6,7</sup> Despite a dynamic and renewing supply of bio-based chemicals, there is a need to develop new solvent structures to match specific properties, which should be guided by efficient *in silico* tools.<sup>8,9,10</sup>

We have recently developed a Computer Assisted Organic Synthesis (CAOS) program named GRASS, standing for GeneratoR of Agro-based Sustainable Solvents. This program generates structures from a chosen bio-based building block by applying industrially relevant transformations to any starting molecule.<sup>11</sup> The relevance of this tool has been exemplified starting from an emerging building block, itaconic acid, from which several families of potential solvents have been highlighted. The strength of this tool is that all the structures

generated are supposed to be experimentally feasible, since they are obtained through well-selected chemical transformations. However, as GRASS requires the choice of co-reactants, the structures obtained should only be regarded as “lead compounds” that may not be the best candidate within a family to match the required properties. Also, only the first generation compounds can reasonably be examined because of the combinatorial explosion afterwards and the lack of properties calculation tools to sort out the molecules.

Another approach to generate virtual structures of potential solvents is the IBSS (InBioSynSolv) program. IBSS is a Computer Aided Product Design (CAPD) tool that defines a set of target values for selected properties and search for new molecules – individual or mixtures – that best satisfy them. The structures are optimized during the search thanks to a genetic algorithm and only the best candidates are proposed.<sup>12</sup> IBSS can generate structures “from scratch” or starting from selected building blocks, including ones coming from the biomass feedstock.<sup>13</sup> A drawback of this approach is that no safeguard ensures the viability of the generated structures. Therefore, the best structures proposed might be virtually excellent but practically non-feasible at an industrial scale.

In the present work, we have taken advantage of both approaches by combining the ability of GRASS to generate rational structures and the potentialities of IBSS to predict properties and to rank and optimize molecules thanks to a performance function. Here, IBSS was allowed to perform only slight structural modifications to the alkyl substituents introduced by GRASS at the first generation – chain length increase or branching, introduction of unsaturations or phenyl groups – to match the performances defined in the specifications.

To highlight this strategy, we have chosen the design of glycerol-based solvents for nitrocellulose as a relevant case. Nitrocelluloses

<sup>a</sup> Univ. Lille, CNRS, Centrale Lille, ENSCL, Univ. Artois, UMR 8181 - UCCS - Unité de Catalyse et Chimie du Solide, F-59000 Lille, France.

E-mail: jean-marie.aubry@ensc-lille.fr, valerie.molinier@ensc-lille.fr

<sup>b</sup> Université de Toulouse, INP, CNRS, LGC (Laboratoire de Génie Chimique), ENSIACET, 4 Allée Emile Monso, F-31432 Toulouse, France.

with a low substitution degree are used for coatings, especially for nail varnishes and wood lacquers. In these formulations, the polymer is dissolved in a solvent - or a mixture of solvents - that allows low-viscosity formulations, easy spreading and formation of a glossy film after evaporation. In nitrocellulose lacquers, the solvents used are referred to as "lacquer thinners" and are usually mixtures of butyl acetate and aromatic solvents (xylene, toluene), which are all flammable and for the latter two, exhibit risks for human safety. This is therefore a good example of solvent substitution. We have chosen to look for a solvent derived from glycerol, which is an archetypical building block from the biomass feedstock as it received considerable interest in the last decade.<sup>14,15,16,17</sup> Several solvents derived from glycerol are already described in the literature or available on the market, especially glycerol esters, glycerol carbonate and glycerol acetals.<sup>18,19,20,21,22,23</sup>

The specifications here are for the solvent to exhibit high solubilisation capacity towards nitrocellulose, high flash point to be considered as "non-flammable" and middle-range boiling point for slow evaporation ensuring good film formation and levelling. Additional prediction methods regarding EHS (Environmental, Health, Safety) profile are also included in the evaluation. Finally, to validate the approach, 15 commercially available solvents derived from glycerol were confronted with nitrocellulose to search for effective and safe solvents able to replace conventional solvents.

## Results and discussion

### 1. Top-down approach used to design the most promising bio-solvents for a given solute

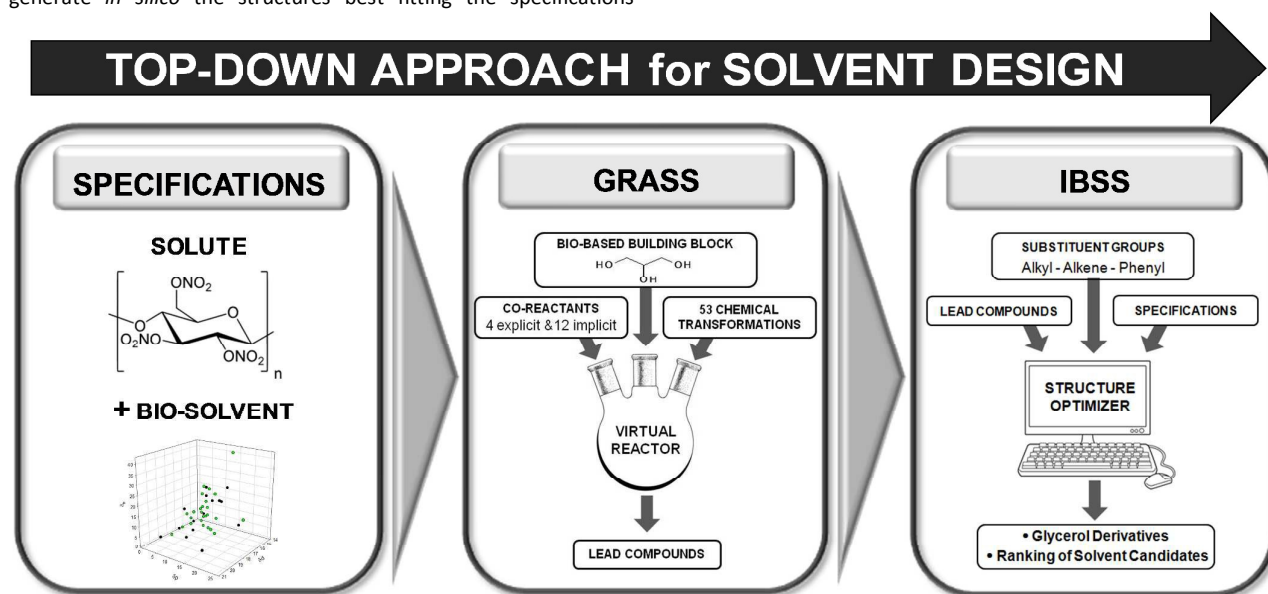
We are using here a "top-down" approach to propose promising solvents for a given solute, which means that we first define the properties that the solvent should comply with and we then generate *in silico* the structures best fitting the specifications

following the three-step process depicted in Figure 1. The final goal is to point out the most promising bio-based solvents satisfying at the same time technical and EHS-related properties.

In the first step, the target solute is chosen. Here, the methodology is illustrated with the example of a highly polar polymer, nitrocellulose, for which the behaviour against classical solvents is well-described by the Hansen solubility parameters approach.<sup>24</sup> Specifications related to the solubilisation of nitrocellulose are therefore based on its so-called Hansen solubility sphere, as described in the next section. Additional technical and EHS-related specifications are also defined at this first step.

In the second step, plausible generic structures derived from a given molecule are generated thanks to a home-made software called GRASS for Generator of Agro-based Sustainable Solvents.<sup>11</sup> A readily available petro- or bio- based starting building block is introduced into a virtual reactor in which it can undergo 53 industrially-relevant chemical transformations in the presence of a limited number of simple co-reactants. Here, an archetypical starting molecule, *i.e.* glycerol, is chosen. The generic molecules obtained at this stage are industrially plausible and represent "lead compounds" that serve as inputs for the third step.

In the last step, the previously developed IBSS (InBioSynSolv) software<sup>12</sup> has a twofold use: to generate a large number of simple derivatives from the "lead compounds" generated by GRASS and to rank all candidates according to their compliance with the specifications. The list of ranked molecules may contain already known solvents or new structures. It should be carefully examined in the eye of an organic chemist in order to select the most accessible candidates. To validate this approach, 15 commercially available solvents derived from glycerol have been confronted with nitrocellulose.



**Figure 1** Schematic representation of the top-down strategy used to find out *in silico* the most promising glycerol-derived solvents for nitrocellulose using GRASS and IBSS softwares

## 2. Specifications and Performance function

To build up an overall performance function *Perf* suitable for evaluating the fitness of each candidate with respect to the  $n_p$  target properties  $p$ , the function defined by Eq. 1 can be used

$$Perf = \sum_{p=1}^{n_p} w_p \cdot F_p \left/ \sum_{p=1}^{n_p} w_p \right. \quad (1)$$

Where  $w_p$  is the weight of each individual property and  $F_p$  is the individual performance function for the property  $p$ . The relevant properties  $p$  that will drive solvent selection have to be defined. In the case study of the present work – solvent for nitrocellulose coating – the main properties for the solvent are i) good solubilising power for nitrocellulose ii) non-flammability and convenient evaporation profile to ensure correct film quality iii) acceptable EHS profile. For each of these characteristics, a calculable property has to be defined and a target value should be chosen, according to functionality (solubilisation, evaporation) or regulatory constraints (flammability, EHS indexes). The properties calculated are the ones presented in Table 1 along with their weight in the total performance *Perf* (Eq. 1)

Their selections are clarified below. Figure 2 shows the evolution of the associated individual performance functions  $F_p$ , described by a Gaussian-type function (Eq. 2).

$$F_p(x) = \exp \left[ \ln(Val) \cdot \left( \frac{Prop - x}{Tol} \right)^2 \right] \quad (2)$$

Where *Prop* is the target property value, whereas *Tol* and *Val* are respectively the tolerance and the property performance value at tolerance. Those values are reported in Table 1 for each of the selected properties. Other functions like desirability functions could also be used.<sup>13</sup>

**Table 1** Properties  $p$  predicted by IBSS, which are used for calculating the individual ( $F_p$ ) and the overall (*Perf*) performance functions. *Prop*, *Tol*, *Val* and  $w_p$  are respectively the symbols for the target value, the tolerance, the performance at tolerance and the weight that appears in equations (1) and (2)

Property, $p$	<i>Prop</i>	Prediction method	<i>Tol</i>	<i>Val</i>	$w_p$
<b>Solubilisation</b>					
RED (Hansen approach)	0	MB2010 <sup>24</sup>	0.2	0.8	4
<b>Volatility / Flammability</b>					
Boiling Point	100 – 300 °C	MG2001 <sup>25</sup>	3	0.8	2
Flash Point	> 60 °C	CPN2006 <sup>26</sup>	3	0.8	2
<b>Environmental and Safety</b>					
Toxicity (LC <sub>50</sub> )	< 3	MY2001 <sup>27</sup>	1	0.8	0.5
Biodegradability (log K <sub>ow</sub> )	< 3	MG2002 <sup>28</sup>	0.3	0.8	0.5
Bioaccumulation (BCF)	< 100	VK1975 <sup>29,30</sup>	1	0.8	0.5

The *ability to solubilise nitrocellulose* is evaluated using the Hansen approach.<sup>24</sup> The three Hansen solubility parameters,  $\delta_p$ ,  $\delta_h$  and  $\delta_d$ , define a three-dimensional space in which all solvents and solutes can be located.<sup>31,10</sup> A solute can be visualized as a point surrounded by its solubility sphere having a R radius. All solvents and mixtures of solvents located inside this volume are likely to solubilise the

solute. The closer the solute and solvents parameters are, the higher the solubility should be. The solute-solvent distance D in the Hansen space is defined according to Equation 3.

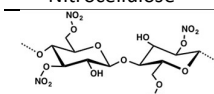
$$D = \sqrt{4(\delta_{d, \text{Solvent}} - \delta_{d, \text{Solute}})^2 + (\delta_{p, \text{Solvent}} - \delta_{p, \text{Solute}})^2 + (\delta_{h, \text{Solvent}} - \delta_{h, \text{Solute}})^2} \quad (3)$$

The ratio between the distance D and the radius R of the solubility sphere is called the “Relative Energy Difference” (RED = R/D) and allows a fast screening of molecules. This is the property calculated by IBSS to evaluate the solubilising capacities of the structures generated. RED < 1 indicates that a molecule is inside the sphere and is likely to have a high affinity with the solute while higher values of RED indicate a poor affinity. The closer to 0 the RED value is, the better the solvent should be, that is why the target value for this property is set to 0. Nitrocellulose has already been experimentally characterized by Hansen by testing its behaviour towards 88 classical solvents.<sup>24</sup> Here, a grade of nitrocellulose, with 13.42 % N and a MW of 69000 g.mol<sup>-1</sup> is investigated. From the experimental solubility data determined with conventional solvents we inferred a centre and a radius of the solubility sphere of the polymer, given in Table 2 and used for the calculation of RED.

The convenient *evaporation profile* is evaluated through the boiling point value. The solvent should be able to evaporate, but not too quickly to avoid formation of cracks in the nitrocellulose film.<sup>32</sup> The higher limit has been set to 300°C, which may appear a bit high but is frequently encountered for high boiling point solvents used as coalescing agents for paints (dimethyl phthalate: 282°C, di(ethylene glycol) hexyl ether: 260°C, dimethyl isosorbide: 234°C). The lower limit has been set to 100°C, since toluene, which is currently used in lacquer thinners, has a boiling point of 111°C. The *non-flammability* criterion is handled by calculating the flash point and setting a target value higher than 60°C which is the lower limit for a compound to be classified as non-flammable.

Finally, *toxicity* is evaluated by setting a limit value for the 96 h fathead minnow acute toxicity – log<sub>10</sub> LC<sub>50</sub> (in mol/L) lower than 3. This selected target value corresponds to LC<sub>50</sub> > 100 mg/L for a molecule weighting 100 g/mol, labelled as category 4 *i.e.* not harmful for aquatic environment in the United Nation Global Harmonization System of Hazard Classification and Labeling.<sup>29</sup> *Biodegradability* is evaluated from the K<sub>ow</sub> octanol-water partition coefficient. Nonpolar compounds usually have values of log K<sub>ow</sub> > 4 and polar compounds have values of log K<sub>ow</sub> < 1.5. A target value log K<sub>ow</sub> < 3 is set. As regards to the Bioconcentration factor BCF, a value greater than 500 L/kg indicates potential bioaccumulation. It is often correlated to the K<sub>ow</sub> value.<sup>29</sup> Typically, a value of 3 < log K<sub>ow</sub> < 4 corresponds to 100 < BCF < 500. A target value BCF < 100 is set.

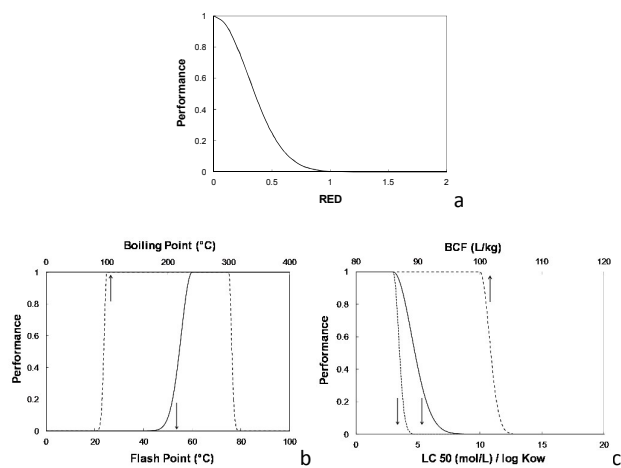
**Table 2** Position of nitrocellulose in the Hansen space (experimental centre and radius of the sphere expressed in MPa<sup>1/2</sup>)

Nitrocellulose <sup>(*)</sup>	$\delta_d$	$\delta_p$	$\delta_h$	R
	18.4	14.4	11.0	8.3

(\*) MW = 69000 g.mol<sup>-1</sup> and 13.42 % N

The potential solvent structures generated are compared and ranked using the performance function (Eq. 1) that takes into account the calculated RED, boiling point, flash point and the EHS

indexes. It should also be pointed that a molecular weight limit has been set to 250 g/mol because “small” molecules are more likely to be liquid. Indeed, no melting point calculations have been performed since this property is poorly predicted by current methods, the average error being greater than 16°C.<sup>33</sup>



**Figure 2** Gaussian-type performance functions for properties listed in Table 1. (a) Solubilising power of a solvent expressed as its normalized distance from the center of the Hansen solubility sphere of nitrocellulose (b) Solvent boiling point and flash point and (c) EHS descriptors for toxicity (LC<sub>50</sub>), biodegradability (log K<sub>ow</sub>) and bioaccumulation (BCF) of solvents

### 3. Generation *in silico* of synthetically feasible “lead compounds” derived from glycerol with GRASS software

In this paper, GRASS was deliberately stopped at the first generation to illustrate the methodology and to avoid combinatorial explosion. Solvents are commodity chemicals that should be obtained in a limited number of synthetic steps. 2<sup>nd</sup> or 3<sup>rd</sup> generation derivatives could also be considered, however, in the present paper, GRASS is used only as a proof of concept to generate synthetically feasible “lead compounds” further optimized by the IBSS program. Glycerol only possesses two primary and one secondary alcohol functional groups, therefore only 16 transformations, summarized in Table 3, among the initial 53 encoded in GRASS,<sup>11</sup> are relevant to transform the glycerol molecule.

**Table 3** Transformations of alcohols encoded in the GRASS program (numbers refer to those listed in Moity et al., 2014)

Nb.	Alcohol	Co-reactant	Final function	Name of transformation
T1	$\text{>C-OH}$	$\text{R}-\text{C}(=\text{O})\text{OH}$	$\text{R}-\text{C}(=\text{O})\text{O}-\text{C}$	Esterification
T8	$\text{>C-OH}$	$\text{H}_2\text{C}=\text{CH}_2$	$\text{H}-\text{C}(\text{OH})-\text{C}(\text{OH})-\text{CH}_2$	Hydro-alkoxy-addition
T10	$\text{>C-OH}$	$\text{R}'-\text{C}(=\text{O})-\text{R}$	$\text{R}'-\text{C}(\text{OR})-\text{C}(\text{OR})-\text{R}$	Acetalization
T14	$\text{>C-OH}$	$\text{H}_2\text{C}=\text{CH}_2$	$\text{HO}-\text{C}(\text{OH})-\text{CH}_2$	Alcoholysis
T21	$\text{H}-\text{C}(\text{OH})-\text{C}$	$-\text{H}_2\text{O}$	$\text{>C}=\text{C}$	Dehydration
T32	$\text{H}-\text{C}-\text{OH}$	$\text{O}_2$	$-\text{C}(=\text{O})\text{OH}$	Oxidation

T33	$\text{H}-\text{C}-\text{OH}$	$\text{O}_2$	$\text{>C}=\text{O}$	Oxidation
T40	$\text{>C-OH} + \text{>C-OH}$	$\text{CO}_2$	$\text{O}=\text{C}(\text{O}-\text{C})_2$	Carbonation
T41	$\text{>C-OH}$	$\text{NH}_3$	$\text{>C}-\text{NH}_2$	Amino-de-hydroxylation
T42	$\text{>C-OH}$	$\text{HN}(\text{R})\text{R}'$	$\text{>C}-\text{N}(\text{R})\text{R}'$	Amino-de-hydroxylation
T43	$\text{>C-OH}$	$\text{R}-\text{OH}$	$\text{>C}-\text{O}-\text{R}$	Alkoxy-de-hydroxylation
T52	$\text{>C-OH}$	$\text{CO}$	$\text{>C}-\text{C}(=\text{O})\text{OH}$	Alcohol carbonylation
T54	$\text{>C-OH}$	$\text{SO}_3$	$\text{>C}-\text{O}-\text{SO}_3\text{H}$	Alcohol sulfation
T56	$\text{>C-OH}$	$\text{R}'-\text{C}(=\text{O})-\text{R}$	$\text{R}'-\text{C}(\text{OR})-\text{C}(\text{OR})-\text{R}$	Hemiacetalization
T58	$\text{HO}-\text{C}(\text{OH})-\text{C}$	Oxidant	$-\text{C}(=\text{O})\text{OH} + -\text{C}(=\text{O})\text{OH}$	Glycol oxidative cleavage

The industrial transformations encoded in GRASS allow the starting building block (glycerol in the present case) to react with implicit or explicit co-reactants, listed in Table 4. Implicit co-reactants are already encoded in the transformations whereas explicit ones need to be provided to the program, as described in the original paper.<sup>11</sup> The explicit co-reactants chosen are the most common monofunctional derivatives of the amine, alcohol, carboxylic acid and ketone chemical families that are reacted with the starting building block. This need to select only one representative member of each chemical family necessarily restrains the derivatives provided by GRASS. Other cheap and readily available co-reactants could have been selected in several families (e. g. ethanol in place of methanol, benzaldehyde in place of acetone, etc...), leading to other 1<sup>st</sup> generation derivatives. That is why the compounds provided by GRASS should be regarded only as generic compounds to be further optimized by changing R and R' groups in the IBSS program.

**Table 4** Co-reactants chosen for the generation of “lead compounds” of glycerol-derivatives with the GRASS program

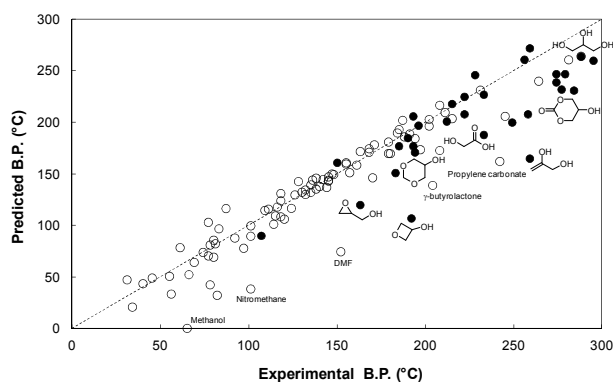
Explicit co-reactants	Methyl amine (R = H, R = Me), methanol (R = Me), acetic acid (R = Me), acetone (R = R = Me)
Implicit co-reactants	H <sub>2</sub> , H <sub>2</sub> O, H <sub>2</sub> O <sub>2</sub> , CO, CO <sub>2</sub> , NH <sub>3</sub> , O <sub>2</sub> , O <sub>3</sub> , SO <sub>3</sub> , ethylene, ethylene oxide, formaldehyde

At the first generation, 66 molecules with a molecular weight lower than 250 g/mol are obtained (Table 5). On one hand, implicit co-reactants react with glycerol according to transformations of Table 3 giving: the oxidation products (T32, T33, T58), the dehydration products (T21), the products of amino-de-hydroxylation (T41), carbonylation (T52) and sulfation (T54). Addition of one, two or three of the hydroxyl group of glycerol on the double bond (T11) provides the five compounds listed for the reaction with ethylene. Carbonation with carbon dioxide (T40) gives the two cyclic carbonates. Ethylene oxide opening (T14) provides the five products of the reaction of one, two or three of the hydroxyl group of glycerol.

On the other hand, the explicit co-reactant methyl amine reacts through amino-dehydroxylation (T42) on one, two or three of the hydroxyl group providing five products. The derivatives obtained by reaction with methanol are the ethers produced by alkoxy-

dehydroxylation (T43) occurring on one, two or three of the hydroxyl group of glycerol. With the same co-reactant, acetalization (T10) involving formaldehyde as implicit co-reactant provides the acetals. Esterification with acetic acid on one, two or three of the hydroxyl groups (T1) provide the five esters. Reaction with acetone gives access to the cyclic ketals (T10) and hemiketals (T56). Among the 66 structures generated by GRASS, 58 have a CAS registry number, which means that GRASS provides indeed synthetically feasible chemical structures. The compounds may or may not be experimentally obtained starting from glycerol. Relevant references describing the preparation of the compounds starting from glycerol and the selected co-reactant (or straightforward derivatives) in reasonable yields are given in Table 5 together with the global performance calculated by IBSS for the target application.

The properties predicted by IBSS may be compared to the experimental values when they are available. For calculating the boiling points, IBSS uses the 3-level group contribution method described by Marrero and Gani<sup>25</sup> that has a reported accuracy of 8 K over 1211 data points. Figure 3 shows the correlation between experimental and predicted boiling points for the 88 solvents of Hansen's original list (white dots)<sup>24</sup> and the 33 glycerol derivatives for which reliable boiling points are reported in the literature at normal or reduced pressure (black dots). In this latter case, the boiling point has been extrapolated to 760 mm Hg using P-T nomograph. There are some noticeable discrepancies in the "traditional solvents" family, particularly for very polar solvents that are suspected to be highly structured, like dimethylformamide (DMF) or propylene carbonate. It should be stressed that Marrero and Gani's prediction method was mainly regressed over molecules issued from the oil & gas industry that did not include much molecules from green feedstocks. Similar discrepancies are observed for the glycerol derivatives obtained by GRASS. Glycerol itself is fairly well-predicted (264°C vs. 288°C) but the boiling points of glycerol carbonate, glycerol formal or glycidol are clearly underestimated, probably because of the strong inter- and intra-hydrogen bonding. For other compounds, particularly the oxidation products, the discrepancy could be due to the wrong experimental determination of high boiling points for temperature-sensitive products. Nevertheless, the overall correlation can be considered as good, which means that IBSS provides an accurate prediction of the boiling points. Regarding the other physical data calculated, not enough experimental data are available for the glycerol derivatives to draw such correlations.



**Figure 3** Comparison between the boiling points predicted by IBSS and the experimental values reported for 88 classical solvents of Hansen's list (white dots) and glycerol-based solvents from the first GRASS generation (black dots)

At this first generation, only 7 out of the 66 structures designed by GRASS have a global performance higher than 60 %. The best structure shows a performance of 69.4 %. It is the diamine resulting from the reaction of glycerol with methyl amine on the two primary positions. This compound is described in the literature (CAS number 83804-96-0, boiling point 304°C), but has not been obtained from glycerol so far and a straightforward and selective synthesis from glycerol is unlikely. The position of substitution and the number of amine substituents impact the performance since the other compounds obtained with methyl amine as co-reactant have performances lower than 60 %.

Other high performers are ethers, esters and acetals in which one alcohol function remains free. None of these compounds is reported to be easily prepared from glycerol in the literature. It is interesting to note that the 1,3-glycerol diethyl ether obtained by addition of ethylene has a higher ranking than the 1,3-glycerol dimethyl ether resulting from reaction with methanol, which tends to show that the length of the lateral chain impacts the performance. At this point, it is also important to stress that all the molecules obtained by virtually reacting glycerol with ethylene can be found in the literature, but they are not experimentally prepared starting from these two reactants. Instead, these ethers are obtained by reaction of glycerol with ethanol, another representative member of the monoalcohol chemical family that could have been chosen in place of methanol.

A second generation with the same co-reactants provides *ca.* 15000 GRASS derivatives, 2675 with a molecular weight lower than 250 g/mol. This generation and the further ones have not been considered. Instead, the most promising families of GRASS compounds obtained at the first generation have been implemented to the IBSS database as complex functional groups to perform slight structural modifications of the substituents allowing a performance increase.

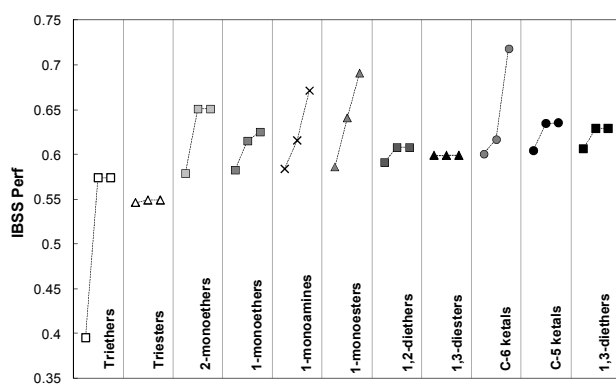
**Table 5** Performances of the 66 structures with MW < 250 g/mol obtained as first generation derivatives of glycerol using the GRASS software. The co-reagents from which they were obtained are indicated: H<sub>2</sub>CO = Formaldehyde, Ethy = Ethylene, EtOx = Ethylene oxide, MA = Methyl amine, MeOH = Methanol, AcOH = Acetic acid, Acet = Acetone and Ox = Oxidant. Relevant references describing an efficient synthetic pathway from glycerol and the given co-reagent are indicated. Compounds without CAS registry number are indicated in italics

Compound	Co-reagent	Perf %	Compound	Co-reagent	Perf %	Compound	Co-reagent	Perf %
	H <sub>2</sub> CO <sup>39</sup>	58.0		H <sub>2</sub> CO <sup>39</sup>	57.9		NH <sub>3</sub>	57.9
	NH <sub>3</sub> <sup>40</sup>	57.9		-	57.9		NH <sub>3</sub>	57.9
	NH <sub>3</sub>	57.9		NH <sub>3</sub> <sup>41</sup>	57.9		O <sub>2</sub>	57.9
	O <sub>2</sub> <sup>42</sup>	57.9		O <sub>2</sub>	57.9		O <sub>2</sub> <sup>43</sup>	57.9
	O <sub>2</sub>	57.9		O <sub>2</sub>	57.9		-	57.9
	Ox <sup>44</sup>	57.9		O <sub>2</sub>	53.4		O <sub>2</sub>	52.8
	-	50.0		-	36.9		CO	34.8
	CO	34.8		O <sub>2</sub>	32.9		SO <sub>3</sub>	33.3
	SO <sub>3</sub>	33.3		CO	31.6		-	25.9
	Ox	17.3		Ethy	62.9		Ethy	60.8
	Ethy	59.5		Ethy	58.0		Ethy	57.4
	CO <sub>2</sub> <sup>45</sup>	57.9		CO <sub>2</sub> <sup>46</sup>	57.9		EtOx	58.0
	EtOx	57.9		EtOx	45.5		EtOx	40.5
	EtOx	38.5		MA	69.4		MA	59.7
	MA	58.6		MA	58.6		MA	58.4
	MeOH	65.7		MeOH	62.2		MeOH <sup>47</sup>	60.6
	MeOH <sup>48</sup>	59.7		MeOH	59.7		MeOH	59.1
	MeOH <sup>49</sup>	58.3		MeOH	57.9		MeOH <sup>50</sup>	57.9
	MeOH <sup>51</sup>	39.5		AcOH	60.9		AcOH <sup>52</sup>	59.9
	AcOH <sup>53</sup>	58.6		AcOH <sup>54</sup>	57.9		AcOH <sup>55</sup>	54.6
	Acet <sup>56</sup>	58.4		Acet <sup>57</sup>	58.5		Acet	57.9
	Acet	57.9		Acet	57.9		Acet	57.9

#### 4. Evolutionary structure optimisation by IBSS

IBSS has been used to find the R and R' substituents that most increase performance of the GRASS generic compounds shown in the first column of Table 6. These "lead compounds" have been obtained by reaction of glycerol with derivatives of the explicit GRASS co-reactant, i.e. 1-monoamines; mono-, di- and tri-ethers; 1-mono, 1,3-di and tri-esters and the C5- and C6- ketals (see fig 4 and Table 6). IBSS has been asked to search for R substituents among linear or branched alkane, alkene, cycloalkane, cycloalkene and phenyl groups. The maximum number of carbon atoms has been set to 6 in order to limit the molecular weights of the compounds generated.

IBSS provides 92 new compounds with performances higher than the starting GRASS derivatives. A bibliographic search highlighted the compounds that have already been prepared from glycerol or a simple derivative in reasonable or high yields. These are not necessarily the best candidates found by IBSS. Figure 4 illustrates the performance increase obtained thanks to IBSS structure optimization. For each generic family, 3 compounds are shown: the starting GRASS compound, the best IBSS candidate already described in the literature and the very best IBSS candidate. Table 6 shows the structures of these compounds.



**Figure 4** Performance increase after structure optimization by IBSS. For each family, the first point corresponds to the GRASS derivative, the second one to the best structure proposed by IBSS for which a feasible synthetic pathway is described in the literature, and the third one to the very best structure proposed by IBSS. The structures are detailed in Table 6.

**Table 6** Results of the structure optimization by IBSS. For each family, the GRASS head of series and co-reagent are indicated as generic compounds. The performance are given for various substituents: the first line corresponds to the GRASS compound, the second one to the best IBSS compound found in the literature and the last line to the very best structure proposed by IBSS even when it is not known yet. Relevant references describing an efficient synthetic pathway are also indicated.

Head of series	Co-reagent	Substituents		Perf %	Ref
		R	R'		
	R-OH	Me	-	39.5	51
		Et	-	57.4	-
		-	-	-	-
	R-COOH	Me	-	54.6	55
		Et	-	54.9	58
		-	-	-	-
	R-OH	Me	-	57.9	50
			-	65.1	59
		-	-	-	-
	R-OH	Me	-	58.3	49
			-	61.5	60
			-	62.5	-
	R-NH-R'	H	Me	58.4	-
		H		61.6	61
		H		67.1	-
	R-COOH	Me	-	58.6	53
			-	64.1	62
			-	69.1	-
	R-OH	Me	Me	59.1	-
		Et	Et	60.8	63
		-	-	-	-
	R-COOH	Me	Me	59.9	52
		-	-	-	-
		-	-	-	-
	R-COO-R'	Me	Me	58.5	57
		H		61.7	64
				71.8	-
	R-COO-R'	Me	Me	58.4	56
				63.5	65
				63.5	-
	R-OH	Me	Me	60.6	47
		Et	Et	62.9	20
		-	-	-	-

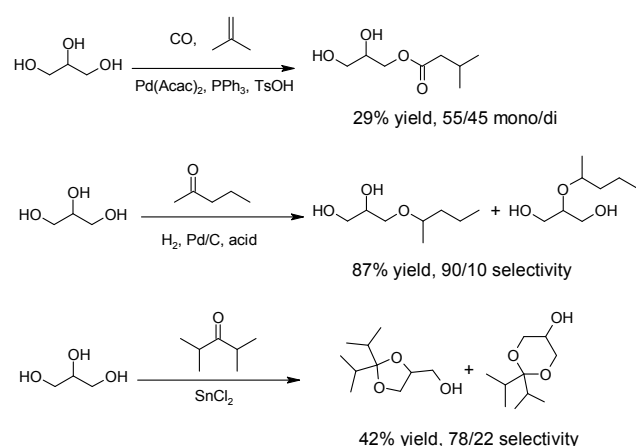
It is worth to notice that a RED value below unity is already calculated for 10 out of the 11 GRASS first generation derivatives in Figure 4, which means that they are themselves supposed to be good solvents of nitrocellulose. The global performance of these first generation compounds ranges from 39.5 % to 60.6 %. The worst score of 39.5 % is for glycerol trimethyl ether whereas glycerol mono- and di-methyl ethers score between 57.9 % and 60.6 %, the best candidate being glycerol-1,3-dimethyl ether. Notice that the glycerol acetals also show high performances, with 58.5 % for the 6-membered ring isomer (1,3-dioxane-type) and 58.4 % for the 5-membered ring one (dioxolane-type). These latter compound is also known as solketal and is already used as coalescing agent in paints or lacquers.<sup>34</sup>

After IBSS optimization, three families rise out in terms of performance greater than 63 %: dioxane-type ketals (nine compounds with performance from 63.0 % to 71.8 %) followed by glycerol esters (nine compounds with performances from 63.0 % to 69.1 %) and then glycerol amines (six compounds with



performances from 63.3 % to 67.1 %). Then come glycerol-2-methyl ether (65.1 % performance) and dioxolane-type ketals (three compounds with performances from 63.4 % to 63.5 %). Several substituents improve significantly the performance for all compounds, especially isoalkyl and isoalkenyl isomers: 3-methylbut-1-ene, isopentyl, isobutyl and isohexyl isomers for monosubstituted compounds and the couple R = isopropyl, R' = ethyl for disubstituted compounds. However, these non-symmetrically disubstituted compounds were not retained since their straightforward preparation from glycerol is unlikely. On the other hand, cyclic substituents apart from 1-cyclopentenyl and apart from 2-cyclopentenyl are not favoured by IBSS to improve the performance.

Finally, it is interesting to stress that some of the top performers highlighted by IBSS can be easily accessed experimentally through catalytic routes (figure 5). Noticeably, the experimental preparation of the best esters and ethers do not take the GRASS pathway: the access to esters *via* the hydroalkoxycarbonylation of alkenes is encoded within GRASS (Transformation 92 listed in reference<sup>11</sup>) but was not retained in the short list of industrially-relevant reactions; the preparation of ethers *via* the reductive alkylation of ketones has not been included in the list of GRASS transformations so far because it is not industrially applied yet. As discussed in an earlier publication, GRASS is a scalable software that allows adding further co-reagents or chemical transformations newly published in the literature or in patents insofar as they seem applicable to the preparation of compounds on large scale. Interestingly, the best C5 and C6 ketals have recently been obtained by reaction of glycerol with diisopropylketone at room temperature and solvent-free conditions using a recyclable tin catalyst.<sup>35</sup> The authors could prepare a series of ketals with varying ratios of 1,3-dioxane and 1,3-dioxolane isomers depending on the bulkiness of the starting ketones. As we showed in a recent publication devoted to ketals and glycerol acetal, one often gets the two isomers (dioxane and dioxolane) in varying proportions. It is difficult to obtain a single isomer but this is not an inconvenience because their solubilizing properties are very close. However their stability towards acid cleavage and oxidative degradation by oxygen are significantly different.<sup>23</sup>



**Figure 5** Experimental catalytic accesses to top performers in the esters<sup>36</sup> ethers<sup>37</sup> and ketals<sup>35</sup> families

The example described in this paper therefore proves the relevancy of combining GRASS and IBSS approaches to generate *in silico* synthetically feasible structures with predefined properties for a target application.

### 5. Experimental validation of the approach with commercially available glycerol-derived solvents

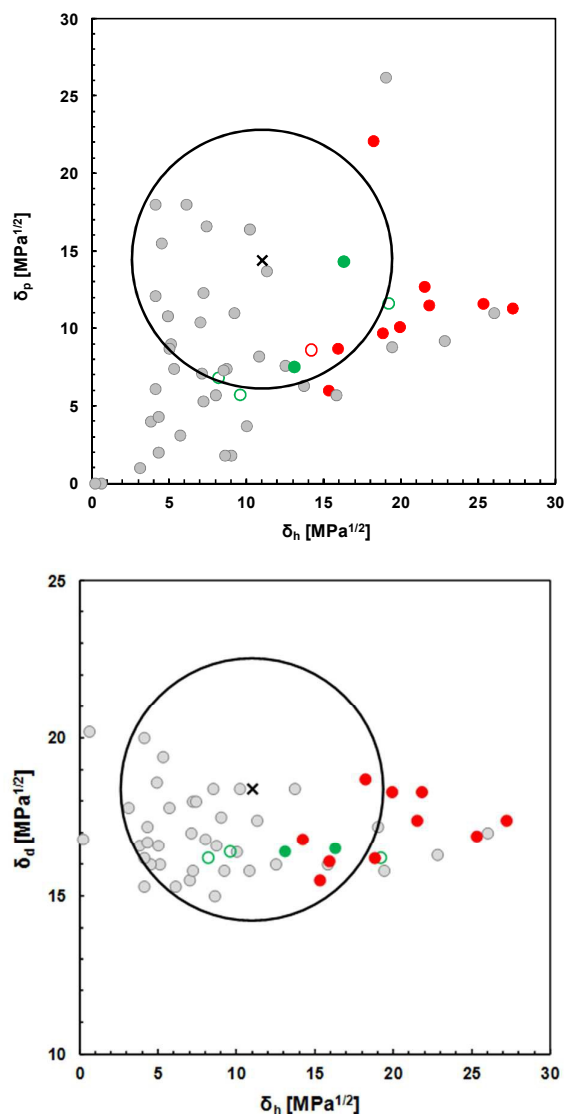
Among the list of the most promising solvents shown in Table 6, several compounds are commercially available. In particular the performance values of solketal, glycerol diacetates (two isomers) and 3-methylamino-1,2-propanediol are close to 60 %. Therefore, to validate the top-down approach described above, we have assessed the ability of these solvents to solubilize nitrocellulose and compared them to 11 other glycerol-based solvents. Table 7 summarizes the results obtained. Only 5 solvents effectively dissolve nitrocellulose (score 1 and 2) whereas most of the others, including solketal, do not interact at all with this polymer (score 6). In principle, all solvents that have a RED value less than 1 should be good solvents for nitrocellulose (score 1 and 2). This is indeed the case for the diacetin isomers and for not for solketal (RED = 0.89) which is an outlier. In the same way all solvents having a RED value higher than 1 are expected to be poor solvents. Actually most of these solvents have a score of 6 but there are three outliers (triacetin, tripropionin and 3-methylamino-1,3-propanediol) which efficiently dissolve nitrocellulose (score 2).

**Table 7** Experimentally determined solubility of nitrocellulose in 15 glycerol-derived solvents characterized by their Hansen solubility parameters ( $\delta_d$ ,  $\delta_p$ ,  $\delta_h$ ) in MPa<sup>1/2</sup>. Solubility parameters were calculated thanks to the Yamamoto Molecular Break (Y-MB) group-contribution method (HSPiP software). The first column expresses our results of the 10 % w/v solubility test using visual inspection as described in the experimental part: from 1 for completely soluble to 6 for no visible effect.

Glycerol derivative	Score	$\delta_d$	$\delta_p$	$\delta_h$	RED
1,2-Diacetin	2	16.5	14.3	16.3	0.79
1,3-Diacetin	2	16.4	7.5	13.1	0.99
Tripropionin	2	16.2	6.8	8.2	1.11
Triacetin	2	16.4	5.7	9.6	1.17
3-Methylamino-1,2-propanediol	1	16.2	11.6	19.2	1.17
3-Ethoxy-1,2-propanediol	6	16.1	8.7	15.9	1.06
Glycerol formal	6	18.3	10.1	19.9	1.19
3-Methoxy-1,2-propanediol	6	16.2	9.7	18.8	1.22
Glycerol carbonate	3	18.7	22.1	18.2	1.27
Glycidol	6	17.4	12.7	21.5	1.30
3-tert-Butoxy-1,2-propanediol	6	15.5	6	15.3	1.33

3-Oxetanol	3	18.3	11.5	21.8	1.35
Solketal	6	16.8	8.6	14.2	0.89
$\alpha,\alpha'$ -Diglycerol	6	16.9	11.6	25.3	1.79
Glycerol	6	17.4	11.3	27.2	2.00

good solvents outside the sphere and the red circle to the sole bad solvent inside the sphere. It is noteworthy that these four outliers are close to the boundary between bad and good solvents. These results are in fair agreement with the predictions of IBSS and allow highlighting diacetin as an effective, inexpensive and non-hazardous solvent of nitrocellulose. Actually, commercial diacetin (B.P. = 240°C, FP = 110°C, safe and unlabelled substance in the EU) is a mixture of the 1,2 and 1,3 isomers, which is inconsequential to the solubility efficiency since both of them are located within the Hansen's sphere of nitrocellulose.



**Figure 6** Projections onto the  $(\delta_p, \delta_h)$  and the  $(\delta_d, \delta_h)$  planes of the Hansen's solubility sphere of nitrocellulose

Figure 6 shows even more clearly the results of the solubilization experiments. Figure 6a and 6b are the projections onto the  $(\delta_p, \delta_h)$  and the  $(\delta_d, \delta_h)$  planes respectively of the Hansen solubility sphere of nitrocellulose. The cross is the centre of the sphere, grey dots represent conventional solvents used to build the sphere whereas red and green dots correspond to the poor and the effective glycerol-derived solvents respectively. Green circles, correspond to

## Conclusions

Today, the art of organic synthesis has reached such a level of sophistication that chemists are able to synthesize almost any molecule with a given structure. The challenges of modern chemistry have therefore moved to other ambitious objectives including the design of molecules having predefined functional properties. The archetypical example is therapeutic chemistry which has to identify, among thousands of possible derivatives of a lead drug compound, the most effective one with regard to biological activity. These so-called fine chemicals are high value substances, produced in low volumes using multistep processes. The design of a bio-based solvent suitable for solubilising a precise target is derived from the same logic. However, due to the higher tonnages required and the auxiliary role of solvents, their synthesis must imperatively be green and simple. In contrast to the total synthesis of complex organic compounds for which numbers of advanced strategies have been developed, to the best of our knowledge, a systematic methodology for designing bio-based solvents suitable for a given application does not exist.

The coupling of two programs, namely GRASS and IBSS, provides an effective answer to this issue. In the first step, GRASS applies relevant chemical transformations to any starting building block to provide synthetically feasible molecular candidates. In the second step, IBSS generates hundreds of derivatives from these candidates, considered as lead compounds, and predicts their functional properties in order to highlight and rank the top performers which best fit the required specifications. To illustrate our approach we chose glycerol as the starting building block and nitrocellulose as the target solute. However the method can be generalized to any starting molecule and any molecular or macromolecular solute. Furthermore, for the sake of clarity, we limited GRASS to the first generation, while syntheses in two or three steps are quite acceptable even for commodity products such as solvents. Thanks to this limitation, we were able to discuss in detail each step of the process and search in the literature, candidates which are actually obtained from glycerol or one of its precursors. In a real situation, these two programs would generate and sort, in a reasonable time, thousands of virtual compounds. Thus, GRASS would provide, in the third generation, many thousands of structures whose groups R and R' would be exchanged by IBSS. During the last step, the values of the performance function  $F_p$  would be calculated automatically to allow the ranking of candidates according to their ability to meet specifications. Finally, a synthetic chemist should carefully examine the list of the most promising compounds in order to select those that seem the most easily accessible even if the synthesis has never been described before.

## Experimental

### GRASS program

GRASS is a Computer-Assisted Organic Synthesis program that derives from GRAAL, a software developed by Barone *et al.* to generate all possible degradation products of food flavours.<sup>62</sup> The detailed description of the program architecture is given in the original papers.<sup>62,63</sup> We have recently updated this software to generate systematically all plausible molecules from a given bio-based building block, according to well-chosen transformations and co-reactants.<sup>11</sup> Particularly, the careful choice of 53 transformations on the basis of industrial relevancy ascertains that the structures generated are feasible. The input data are the starting building-block – glycerol in the present case – the list of transformations and the list of co-reactants. The outputs are the generated structures which are encoded by their SMILE code.

### IBSS program

The IBSS program is a Computer Aided Product Design (CAPD) tool that was developed for finding sustainable products.<sup>13,64</sup> The principle of the Computer Aided Molecular Design (CAMD) is to define initially a set of target values for several properties and search for new molecules built from chemical blocks, that satisfy the set of properties.<sup>65</sup> The IBSS tool was expanded to search for mixtures and is able to orientate the search towards molecules issued from renewable material feedstocks.<sup>13</sup> In the present work, single molecules are searched, each containing one of the core chemical fragments proposed by the GRASS program.

The problem to solve is then a multi-objective simultaneous search over the optimization variables like the molecules described by a molecular graph  $MG_i$ , the mixture composition  $z_i$  and the mixture operating conditions  $cond_i$ . The performance of each candidate molecule or mixture is calculated from a performance function described above, which takes into account the calculated RED (distance from the centre of the Hansen solubility sphere of nitrocellulose over the radius of this sphere), boiling point, flash point and the EHS indexes mentioned before.

### Solubility of nitrocellulose in glycerol-based solvents

For solubility tests, 100 mg of nitrocellulose are placed in a 1.5 mL glass vial, and 1.0 mL of solvent is added. The mixture is stirred at 60 rpm using a Intelli-Mixer RM-2L rotator (Elmi) during 24 h at  $(25 \pm 1)$  °C. The classification of solubility behaviours is based on visual observations: score 1 for soluble, 2 for almost soluble, 3 for strongly swollen but slightly soluble, 4 for swollen, 5 for little swelling and 6 for no visible effect. A solvent is considered as a good solvent when its score equals 1 or 2.

For building the Hansen solubility sphere of nitrocellulose, 40 well-selected conventional solvents are confronted to nitrocellulose.<sup>31,10</sup> Results were computed with the HSPiP software, version 4.0.05 developed by Abbott and Yamamoto, to calculate the solubility spheres coordinates – centre and radius – of nitrocellulose. Then 15 glycerol-based solvents were tested with the same procedure.

## Acknowledgements

This work was performed within the framework of ANR “InBioSynSolv” project. The authors gratefully acknowledge the gift of nitrocellulose from Pr. Pierre Gareil (Ecole Nationale Supérieure de Chimie de Paris).

## References

- 1 C. Capello, U. Fischer and K. Hungerbühler, *Green Chem.*, 2007, **9**, 927–934.
- 2 P. G. Jessop, *Green Chem.*, 2011, **13**, 1391–1398.
- 3 F. M. Kerton and R. Marriott, *Alternative Solvents for Green Chemistry*, Royal Society of Chemistry, 2013.
- 4 L. Moity, M. Durand, A. Benazzouz, V. Molinier and J.-M. Aubry, in *Alternative Solvents for Natural Products Extraction*, eds. F. Chemat and M. A. Vian, Springer Berlin Heidelberg, 2014, pp. 1–24.
- 5 A. Corma, S. Iborra and A. Velty, *Chem. Rev.*, 2007, **107**, 2411–2502.
- 6 J. J. Bozell and G. R. Petersen, *Green Chem.*, 2010, **12**, 539–554.
- 7 R. A. Sheldon, *Green Chem.*, 2014, **16**, 950–963.
- 8 M. Durand, V. Molinier, W. Kunz and J.-M. Aubry, *Chem. – Eur. J.*, 2011, **17**, 5155–5164.
- 9 L. Moity, M. Durand, A. Benazzouz, C. Pierlot, V. Molinier and J.-M. Aubry, *Green Chem.*, 2012, **14**, 1132–1145.
- 10 A. Benazzouz, L. Moity, C. Pierlot, V. Molinier and J.-M. Aubry, *Colloids Surf. Physicochem. Eng. Asp.*, 2014, **458**, 101–109.
- 11 L. Moity, V. Molinier, A. Benazzouz, R. Barone, P. Marion and J.-M. Aubry, *Green Chem.*, 2014, **16**, 146–160.
- 12 J. Heintz, I. Touche, M. Teles Dos Santos and V. Gerbaud, *Comput Aided Chem Eng.*, 2012, **30**, 702–6.
- 13 J. Heintz, J.-P. Belaud, N. Pandya, M. Teles Dos Santos and V. Gerbaud, *Comput. Chem. Eng.*, 2014, **71**, 362–376.
- 14 M. Pagliaro, R. Ciriminna, H. Kimura, M. Rossi and C. Della Pina, *Angew. Chem. Int. Ed.*, 2007, **46**, 4434–4440.
- 15 A. Behr, J. Eilting, K. Irawadi, J. Leschinski and F. Lindner, *Green Chem.*, 2008, **10**, 13–30.
- 16 C.-H. (Clayton) Zhou, J. N. Beltrami, Y.-X. Fan and G. Q. (Max) Lu, *Chem Soc Rev.*, 2008, **37**, 527–549.
- 17 H. Zhang and M. W. Grinstaff, *Macromol. Rapid Commun.*, 2014, **35**, 1906–1924.
- 18 A. E. Díaz-Álvarez, J. Francos, B. Lastra-Barreira, P. Crochet and V. Cadierno, *Chem. Commun.*, 2011, **47**, 6208.
- 19 Y. Gu and F. Jérôme, *Green Chem.*, 2010, **12**, 1127.
- 20 J. I. García, H. García-Marín, J. A. Mayoral and P. Pérez, *Green Chem.*, 2010, **12**, 426.
- 21 J. I. García, H. García-Marín and E. Pires, *Green Chem.*, 2014, **16**, 1007.
- 22 Y. Gu and F. Jérôme, *Chem. Soc. Rev.*, 2013, **42**, 9550–9570.
- 23 L. Moity, A. Benazzouz, V. Molinier, V. Nardello-Rataj, M. K. Elmekaddem, P. de Caro, S. Thiébaud-Roux, V. Gerbaud, P. Marion and J.-M. Aubry, *Green Chem.*, 2015, **17**, 1779–1792.
- 24 C. M. Hansen, *Hansen Solubility Parameters: A User's Handbook, Second Edition*, CRC Press, 2007.
- 25 J. Marrero and R. Gani, *Fluid Phase Equilibria*, 2001, **183–184**, 183–208.
- 26 L. Catoire, S. Paulmier and V. Naudet, *Process Saf. Prog.*, 2006, **25**, 33–39.
- 27 T. M. Martin and D. M. Young, *Chem. Res. Toxicol.*, 2001, **14**, 1378–1385.

- 28 J. Marrero and R. Gani, *Ind. Eng. Chem. Res.*, 2002, **41**, 6623–6633.
- 29 Great Lakes Research Advisory Board, G. Veith and D. Konasewich, *Int. Jt. Comm. IJC Digit. Arch.*, 1975.
- 30 M. Sinha, L. E. K. Achenie and G. M. Ostrovsky, *Comput. Chem. Eng.*, 1999, **23**, 1381–1394.
- 31 A. Benazzouz, L. Moity, C. Pierlot, M. Sergent, V. Molinier and J.-M. Aubry, *Ind. Eng. Chem. Res.*, 2013, **52**, 16585–16597.
- 32 M. Durand, V. Molinier, T. Féron and J.-M. Aubry, *Prog. Org. Coat.*, 2010, **69**, 344–351.
- 33 A. S. Hukkerikar, B. Sarup, A. Ten Kate, J. Abildskov, G. Sin and R. Gani, *Fluid Phase Equilibria*, 2012, **321**, 25–43.
- 34 WO2011042652 A1, 2011.
- 35 M. J. da Silva, M. de O. Guimaraes and A. A. Julio, *Catal. Lett.*, 2015, **145**, 769–776.
- 36 K. A. Suerbaev, E. G. Chepaikin, N. O. Appazov and B. Z. Dzhiembaev, *Pet. Chem.*, 2012, **52**, 189–193.
- 37 Y. Shi, W. Dayoub, G.-R. Chen and M. Lemaire, *Sci. China Chem.*, 2010, **53**, 1953–1956.
- 38 V. R. Ruiz, A. Velty, L. L. Santos, A. Leyva-Pérez, M. J. Sabater, S. Iborra and A. Corma, *J. Catal.*, 2010, **271**, 351–357.
- 39 Y.-H. Joo and J. M. Shreeve, *Chem Commun*, 2010, **46**, 142–144.
- 40 B. Andreeßen and A. Steinbüchel, *Appl. Microbiol. Biotechnol.*, 2011, **93**, 357–365.
- 41 R. M. Painter, D. M. Pearson and R. M. Waymouth, *Angew. Chem. Int. Ed.*, 2010, **49**, 9456–9459.
- 42 M. Trincado, K. Kühlein and H. Grützmacher, *Chem. – Eur. J.*, 2011, **17**, 11905–11913.
- 43 M. Ziolk, I. Sobczak, P. Decyk and L. Wolski, *Catal. Commun.*, 2013, **37**, 85–91.
- 44 M. S. Khayoon and B. H. Hameed, *Appl. Catal. Gen.*, 2013, **466**, 272–281.
- 45 T. W. Turney, A. Patti, W. Gates, U. Shaheen and S. Kulasegaram, *Green Chem.*, 2013, **15**, 1925.
- 46 J.-J. Hu, *Chin. Chem. Lett.*, 2013, **24**, 1106–1108.
- 47 Y. Matsumoto, K. Mita, K. Hashimoto, H. Iio and T. Tokoroyama, *Tetrahedron*, 1996, **52**, 9387–9398.
- 48 L. Birkofer and O. Stuhl, *J. Organomet. Chem.*, 1980, **187**, 21–29.
- 49 E. Quesada, J. Delgado, C. Gajate, F. Mollinedo, A. U. Acuña and F. Amat-Guerri, *J. Med. Chem.*, 2004, **47**, 5333–5335.
- 50 W. Dayoub, M. Sutter, E. Metay, Y. Raoul and M. Lemaire, *ChemCatChem*, 2013, **26**, 5902–5916.
- 51 M. Zieringer, M. Wyszogrodzka, K. Biskup and R. Haag, *New J Chem*, 2012, **36**, 402–406.
- 52 R. Srivastava and N. Venkatathri, *INDIAN J. Chem. Sect. B*, 2004, **43**, 888–891.
- 53 I. Dosuna-Rodríguez and E. M. Gaigneaux, *Catal. Today*, 2012, **195**, 14–21.
- 54 J. Ye, S. Liu, J. Xiang, J. Lei and C. Zhou, *J. Appl. Polym. Sci.*, 2013, **129**, 1915–1921.
- 55 A. W. Pierpont, E. R. Batista, R. L. Martin, W. Chen, J. K. Kim, C. B. Hoyt, J. C. Gordon, R. Michalczyk, L. A. “Pete” Silks and R. Wu, *ACS Catal.*, 2015, **5**, 1013–1019.
- 56 B. Wang, Y. Shen, J. Sun, F. Xu and R. Sun, *RSC Adv.*, 2014, **4**, 18917.
- 57 R. Prager and Z. Yurui, *Aust. J. Chem.*, 1989, **42**, 1003–1005.
- 58 D. Güclü, M. Rale and W.-D. Fessner, *Eur. J. Org. Chem.*, 2015, **2015**, 2960–2964.
- 59 C. Bonini, L. Chiummiento, N. Di Blasio, M. Funicello, P. Lupattelli, F. Tramutola, F. Berti, A. Ostric, S. Miertus, V. Frecer and D.-X. Kong, *Bioorg. Med. Chem.*, 2014, **22**, 4792–4802.
- 60 S. Pariente, N. Tanchoux and F. Fajula, *Green Chem.*, 2009, **11**, 1256.
- 61 I. Notar Francesco, B. Cacciuttolo, M. Pucheault and S. Antonioti, *Green Chem*, 2015, **17**, 837–841.
- 62 F. Shahidi and A. M. Spanier, *Food Flavor and Chemistry: Explorations Into the 21st Century; [the Proceedings of the 11th International Flavor Conference: Recent Advances in Food Flavor Chemistry, 3rd Charalambous Memorial Symposium Held on 29 June to 2 July 2004 in Samos, Greece]*, Royal Society of Chemistry, 2005.
- 63 C.-T. Ho and E. T. Contis, *Recent Advances in Food and Flavor Chemistry: Food Flavors and Encapsulation, Health Benefits, Analytical Methods, and Molecular Biology of Functional Foods*, Royal Society of Chemistry, 2010.
- 64 J. Heintz, J.-P. Belaud and V. Gerbaud, *Comput. Ind.*, 2014, **65**, 505–520.
- 65 L. Achenie, V. Venkatasubramanian and R. Gani, *Computer Aided Molecular Design: Theory and Practice*, Elsevier, 2002.

A “top-down” in silico approach for designing ad hoc bio-based solvents: Application to glycerol-derived solvents of nitrocellulose

By L. Moity, V. Molinier, A. Benazzouz, V. Gerbaud and J.-M. Aubry

Graphical Abstract (8 cm x 4 cm)

