Electronic Supplementary Material (ESI) for Reaction Chemistry & Engineering.

Automated kinetic model identification via cloud services using model-based design of experiments

Electronic Supplementary Information (ESI)

S1 LabBot operation controllers

The four timer objects controlling the LabBot operations are: the request, temperature, steady-state, and sample analysis timers. These timer objects are discussed as in the following sections.

S1.1 Request timer

The request timer object searches for request *i* submitted by the external user in the campaign directory in the cloud, where *i* is the current epoch. If the request of the current epoch has been found and – for epochs later than the first – the HPLC has analysed the sample of the preceding epoch (*i*-1, for *i*>1), the LabBot first checks if the requested process conditions respect its operational limits. In the case the external user has requested for the LabBot to run outside its limits, the LabBot communicates this by uploading a message in text format to the campaign folder in the cloud. Otherwise, (a) it uploads a csv file with the process conditions of the waiting period, (b) it updates the epoch and (c) it triggers the temperature timer object. Figure 3 illustrates the flow chart of the *get request* timer object.

S1.2 Temperature timer

The temperature timer object compares the current reactor temperature to the requested. If the difference is lower than one Celsius degree, (a) it starts the reaction by setting the flow rates of the pumps according to the request, (b) it uploads a csv file with the process conditions of the heating up or cooling down period, (c) it triggers the steady state timer and then it stops to maintain the requested temperature. Otherwise, it sets the pumps to stand-by conditions, to

avoid needless consumption of raw materials and to reduce the waste, until the reactor reaches the requested temperature.

S1.3 Steady state timer

The steady state timer object calculates the residence time based on the requested pump flow rates and the reactor volume. Then it compares the time period elapsed from the reaction start to the time period needed for reaching steady state – this is assumed to be four times the residence time. If the time elapsed is longer than the one needed for reaching steady state, (a) it sends a sample to the HPLC, (b) it sets the pumps to stand by conditions, (c) it uploads a csv file with the process conditions of the time period elapsed from reaction start and sampling, then it stops. Otherwise, it waits for the process to reach steady state.

S1.4 Sample analysis timer

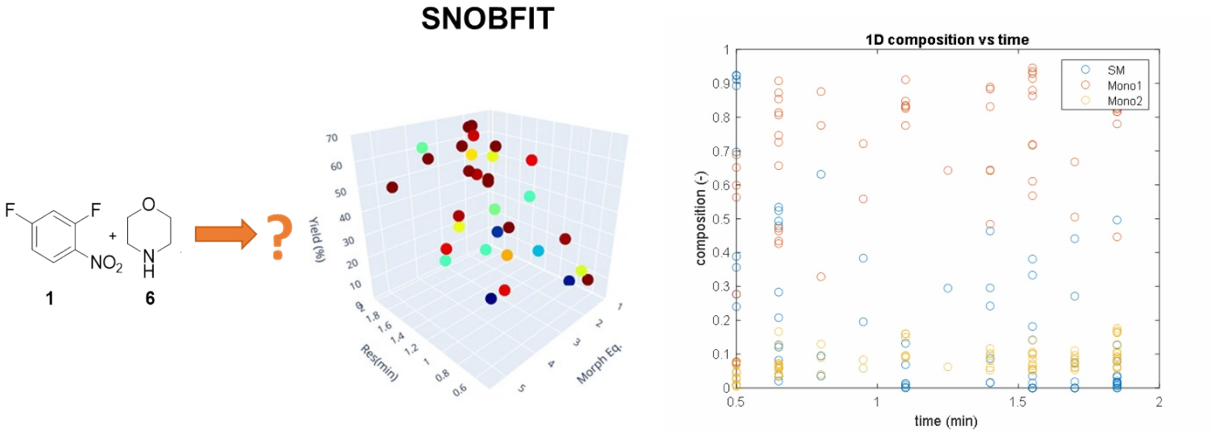
The sample analysis is the last stage of one operation loop. This means that apart from the first epoch, while the request timer object searches for the request of the epoch *i*, the sample analysis timer object needs to search for the HPLC chromatogram of the epoch *i*-1.

Once the chromatogram has been found, (a) the sample analysis timer object calculates the species concentrations by using the chromatogram data and the response factors of each species which have been given to the LabBot, (b) it uploads a csv file reporting the species concentrations and the retention time and peak area of each species, (c) it updates its epoch and wait for the next chromatogram.

S2 Generation of preliminary rate expressions

Each new process to be introduced into the LabBot will require some degree of familiarisation (e.g., reaction phase and chemical analytics) to set up the experiment, validate the analysis and then propose a preliminary rate expression. For previously unstudied systems an initial optimisation for yield of the desired product provides a good system test and can provide an

initial operating point, and reasonable experimental bounds. Stable Noisy Optimization by Branch and FIT (Snobfit), an optimizer developed for optimization problems with noisy and expensive to compute objective functions, is used to acquire initial understanding of a new chemical system [Huyer and Neumaier, 2008]. Snobfit assumes the experimental data obey first-order rate expression with respect to the key starting material, and then, by using a Branch and Fit optimisation algorithm [Morrison et al., 2016], generates a design space where the assumed first-order rate expression is valid. Figure S1 shows a typical result profile of a Snobfit optimisation of yield of X, from which the optimum within the experimental design space of a chemical system can be found. This initial data set can then be processed to generate the master curves – the rate law concentration profiles – as shown in Figure S2 for both the reactant and product, illustrating successful preliminary investigation in the LabBot system, and that the analytical method can detect significantly different compositions used to inform an initial kinetic scheme. Outside the design space, the reaction conditions do not yield useful information.



. (a) (b) (c)

Figure S1: Data from the initial optimisation of the reaction between **1** and **6**. The Snobfit algorithm optimises yield within predefined parameter ranges (b). The resulting data set (c) is

hard to interpret, giving the team no feedback as to the behaviour of the system as function of the parameters, nor on reproducibility of the data.

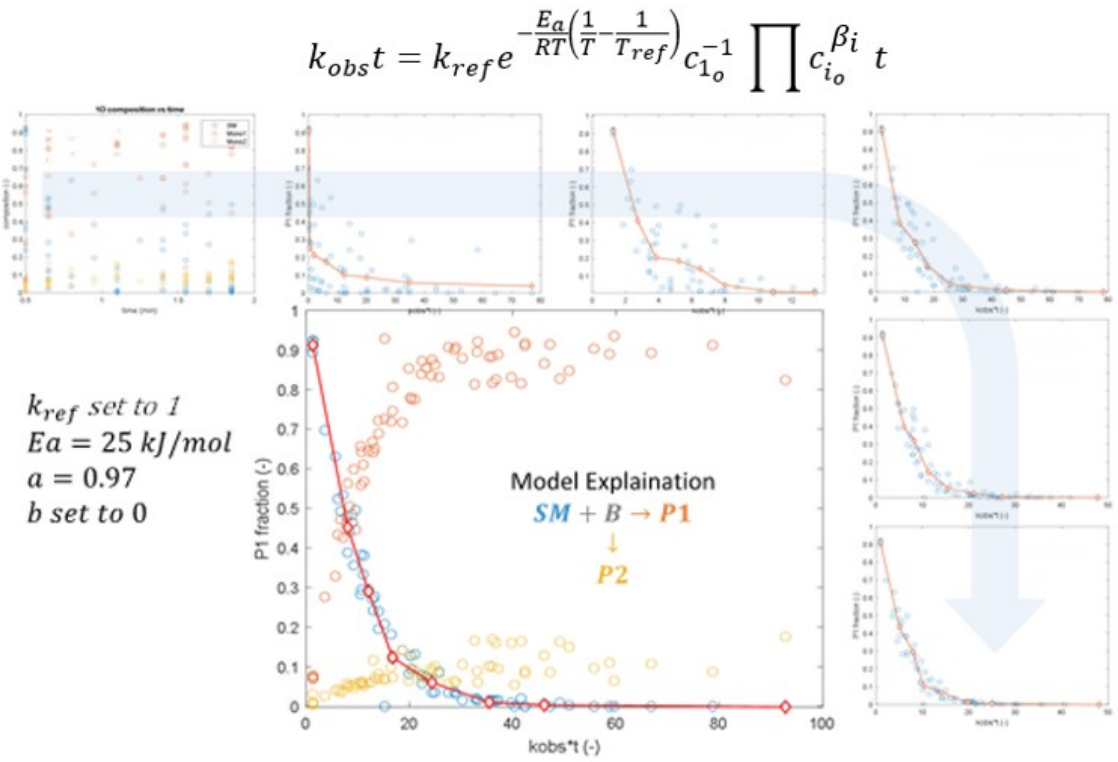


Figure S2: Identification of an initial rate equation from the Snobfit optimisation run. The concentration of starting material is collapsed onto a master curve by iteratively changing the rate parameters ($^{E_a, a\ and\ b}$) to minimise the sum of squares between the data and a regressed piecewise linear function.

Rate expressions defined by Eq. (2.2) in the main article represent a preliminary fit of the key starting material consumption rather than a true kinetic model based on the detailed understanding of the network of reactions controlling the conversion of starting material. The rate equation at the initial conditions of an experiment may be approximated as a first order process:

$$(-r_1) = k_{obs}c_1.\ with\ k_{obs} = k_{ref}e^{-\frac{E_a}{RT}\left(\frac{1}{T}-\frac{1}{T_{ref}}\right)}c_{1_o}^{-1}\prod c_{i_o}^{\beta_i} \qquad \text{(S2.1)}$$

For a plug flow reactor at steady state

$$v_0 dc = -(-r)dV \approx -k_{obs} c dV \quad \text{(S2.2)}$$

Defining the residence time $\tau = \frac{V}{v_o}$ as the ratio of the reactor volume and the volumetric flowrate, an approximate evolution of concentration vs time in the reactor is then:

$$g(k_{obs}\tau) = \frac{c_1}{c_{1_o}} \approx e^{-k_{obs}\tau} \quad \text{(S2.3)}$$

This is a coarse assumption, but suggests that the effect of the processing conditions collapses on a master curve $g(k_{obs}\tau)$ when plotting c_1 versus $k_{obs}\tau$. The task is now to find a set of parameters $\{E_a,\ p_i | i = 1 \ldots n\}$ that minimise the sum of squares (SOS) between the master curve g and the N_{exp} data points, each associated with a set of experimental conditions as captured by $k_{obs_k}\tau_k$:

$$SOS = \sum_{k}^{N_{exp}} \left(c_{1k} - g\left(k_{obs_k}\tau_k\right)\right)^2 \quad \text{(S2.4)}$$

The unknown $g(k_{obs_k}\tau_k)$ can be approximated with a continuous piecewise linear function. Snobfit optimisation provides a preliminary rate expression and bounds on experimental conditions where the preliminary rate expression is valid.

S3 Experimental methods

General methods: Reagents and solvents were obtained from commercial sources and used without purification. The removal of solvent under reduced pressure was carried out on a standard rotary evaporator.

Chromatography: Analytical thin-layer chromatography (TLC) was carried out on pre-coated aluminium plates (silica gel 60 F_{254}) from Merck. Compound spots were visualised under

ultraviolet (UV) light (254 nm), and using ninhydrin or $KMnO_4$ stain solutions. Purification of the products were performed on SiliCycle SiliaSep™ 40–63mm 60 Å flash cartridges using an automated Biotage™ flash chromatography coupled with UV detector at 254 nm.

NMR spectroscopy: ^{1}H-NMR spectra were recorded on a 400 MHz Avance III HD spectrometer with the residual solvent peak as the internal reference ($CDCl_3$ = 7.26 ppm, d_6-DMSO = 2.50 ppm, CD_3OD = 3.31 ppm). ^{1}H resonances are reported to the nearest 0.01 ppm. ^{13}C-NMR spectra were recorded on 400 MHz Avance III HD spectrometer with the central resonance of the solvent peak as the internal reference ($CDCl_3$ = 77.16 ppm, d_6- DMSO = 39.52 ppm, CD_3OD = 49.00 ppm). All ^{13}C resonances are reported to the nearest 0.1 ppm. The multiplicity of ^{1}H signals are indicated as: s = singlet, d = doublet, dd = doublet of doublet, ddd = doublet of doublet of doublet, t = triplet, q = quadruplet, quint = quintet, sext = sextet, m = multiplet, br = broad, or combinations of thereof. Coupling constants (*J*) are quoted in Hz and reported to the nearest 0.1 Hz. Where appropriate, averages of the signals from peaks displaying multiplicity were used to calculate the value of the coupling constant.

S3.1 Nucleophilic aromatic substitution

Preparation of feed solutions:

2,4-Difluoronitrobenzene **1** (99%, Aldrich), morpholine **2** (≥99.0%, Merck), triethylamine (99% Acros), ethanol (99.8%, VWR) and biphenyl (99.5% GC, Aldrich) were used without further purification.

Reservoir solutions were prepared in volumetric flasks by dissolving the desired reagents in solvent at ambient conditions. Feed A (starting material): 2,4-difluoronitrobenzene (50 mL, 0.46 mol, 2.03 mol L^{-1}) and biphenyl (3.51 g, 22.7 mmol, 0.101 mol L^{-1}) in ethanol (175 mL); Feed B (reactant): morpholine (70 mL, 0.81 mol, 4.17 mol L^{-1}) in triethylamine (124 mL, 0.89 mol, 4.59 mol L^{-1}).

From the reservoir solutions $c_1^0$and c_2^0, the inlet concentrations and residence time can be obtained from the flowrates as:

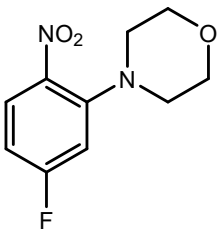
$$c_1(0) = c_1^0 \frac{F_1}{\sum_{i=1}^{3} F_i}; \quad c_2(0) = c_2^0 \frac{F_2}{\sum_{i=1}^{3} F_i}; \quad \tau = \frac{V}{\sum_{i=1}^{3} F_i} \qquad \text{(S3.1)}$$

V is the volume of the reactor.

Analysis: For the analysis, standards of 4-(5-fluoro-2-nitrophenyl)morpholine **3**, *N*-(3-fluoro-4-nitrophenyl)morpholine **4** and 4,4'-(4-nitro-1,3-phenylene) dimorpholine **5** were synthesised and characterised.

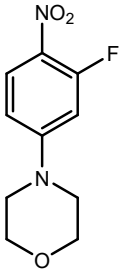
Chemical analytics

To a round-bottomed flask, 2,4-difluoronitrobenzene **1** (5.00 g, 31.4 mmol) in ethanol (150 mL) was added. To this, morpholine **2** (6.02 g, 69.1 mmol) was added and the reaction mixture left to stir at room temperature for 5 hours. The resultant mixture was concentrated *in vacuo*, dissolved in ethyl acetate (100 mL) and washed successively with saturated NH_4Cl solution (100 mL) and brine (100 mL). The organic layer was separated, dried over Na_2SO_4 and concentrated *in vacuo*. The resultant residue was purified by flash column chromatography (10-80% EtOAc/*n*-hexane) to afford *ortho*-**3** (5.53 g, 78%) as an orange oil, *para*-**4** (0.8415 g, 12%) as a bright yellow solid and bis-adduct **5** as an orange solid (0.35 g, 4%).

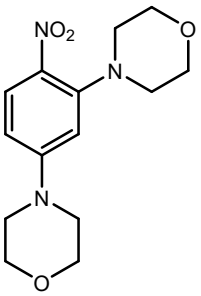


^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.90 (dd, *J* = 9.1, 6.0 Hz, 1H), 6.78 (dd, *J* = 11, 2.6 Hz, 1H), 6.72 (ddd, *J* = 9.5, 7.0, 2.6 Hz, 1H), 3.93 – 3.75 (m, 4H), 3.13 – 2.99 (m, 4H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 165.5 (d, *J* = 260 Hz), 148.6, 148.5, 129.0 (d, *J* = 11 Hz), 108.6 (d, *J* = 24 Hz), 107.2 (d, *J* = 25 Hz), 66.6, 51.7; *m*/*z*

(ESI$^+$) $C_{10}H_{11}FN_2O_3$ [M+H]$^+$, calculated 227.08, found 227.33; in agreement with published data [O'Brien et al., 2012].



^{1}H NMR ($CDCl_3$, 500 MHz) δ 8.09 – 8.00 (t, *J* = 9.1 Hz, 1H), 6.61 (dd, *J* = 9.4, 2.7 Hz, 1H), 6.54 (dd, *J* = 15, 2.7 Hz, 1H), 3.89 – 3.82 (m, 4H), 3.40 – 3.33 (m, 4H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 158.0 (d, *J* = 260 Hz), 155.9, 155.8, 128.2 (d, *J* = 1.3 Hz), 108.2, 101.0 (d, *J* = 26 Hz), 66.2, 46.9; *m/z* (ESI$^+$) $C_{10}H_{11}FN_2O_3$ [M+H]$^+$, calculated 227.08, found 227.32; in agreement with published data [O'Brien et al., 2012].



^{1}H NMR ($CDCl_3$, 500 MHz) δ 8.03 (d, *J* = 9.3 Hz, 1H), 6.47 (dd, *J* = 9.4, 2.6 Hz, 1H), 6.33 (d, *J* = 2.6 Hz, 1H), 3.91 – 3.87 (m, 4H), 3.87 – 3.82 (m, 4H), 3.37 – 3.30 (m, 4H), 3.10 – 3.03 (m, 4H); ^{13}C NMR ($CDCl_3$, 126 MHz) δ 155.2, 149.4, 133.1, 129.7, 107.0, 103.4, 66.9, 66.4, 52.3, 47.3; *m/z* (ESI$^+$) $C_{14}H_{19}N_3O_4$ [M+H]$^+$, calculated 294.15, found 294.39; in agreement with published data [O'Brien et al., 2012].

S3.2 Homogeneous amide formation

Feed preparation

To a 250 ml volumetric flask, di-2-ethylhexylamine (DiEHA) 1 (7.03 g, 0.13 M), Isobutyric Acid 5 (3.54 g, 0.16 M) was mixed in acetonitrile 100 mL). The solution was then sonicated for 30 mins until well mixed. EDC. HCl 6 (7.03 g, 0.15 M) was added and acetonitrile was added to a total volume 250 mL. Another 45 mins of sonication was conducted until the solid

of EDC.HCl was well dispersed. The mix solution was then stirred for 6 hours and 1000 rpm to ensure the solution was well mixed. The mix solution was then transferred into a 500 mL separation funnel for separation process as Fig. S3 below.

10% sulphuric acid (50 mL) was added and mixed for 5 mins (step 1). Then hexane (100 mL) was added and mixed for another 5 mins (step 2). A bilayer was formed after adding hexane (step 3), then the sulphuric acid layer was taken out until only hexane with the product layer was left in the separation funnel (step 4). The remaining solution was washed with saturated sodium bicarbonate solution (20 mL) for twice then the solution was separated between saturated sodium bicarbonate and hexane+product (step 5). The hexane+product was washed with distilled water and then separated between these hexane+product and water (step 6). Magnesium sulphate was added in order to solidify the hexane (step 7). The solid was removed by filtration to leave dried hexane with the product (step 8). Then the product was concentrated *in vacuo* resulting DEHiBA 7 (11.06 g) as product.

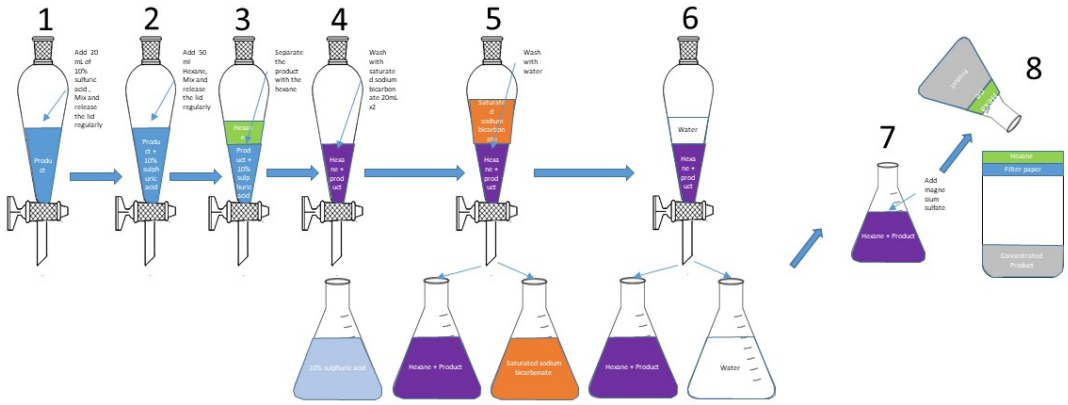


Fig S3. The separation process of DEHiBA product

Chemical analytics

The concentrated DEHiBA was analysed using Gas Chromatography – Mass Spectrometry (GC-MS) Agilent Technologies 7890B and Bruker AV3HD 9.4 T 400 MHz 1H Nuclear Magnetic Resonance (NMR) spectrometer AV3HD-400 series and then presented in Figs. S4 and S5, respectively. From Fig. S4, it shows that DEHiBA product was pure and based on Fig. S5, the pure DEHiBA from experiment was similar to the pure

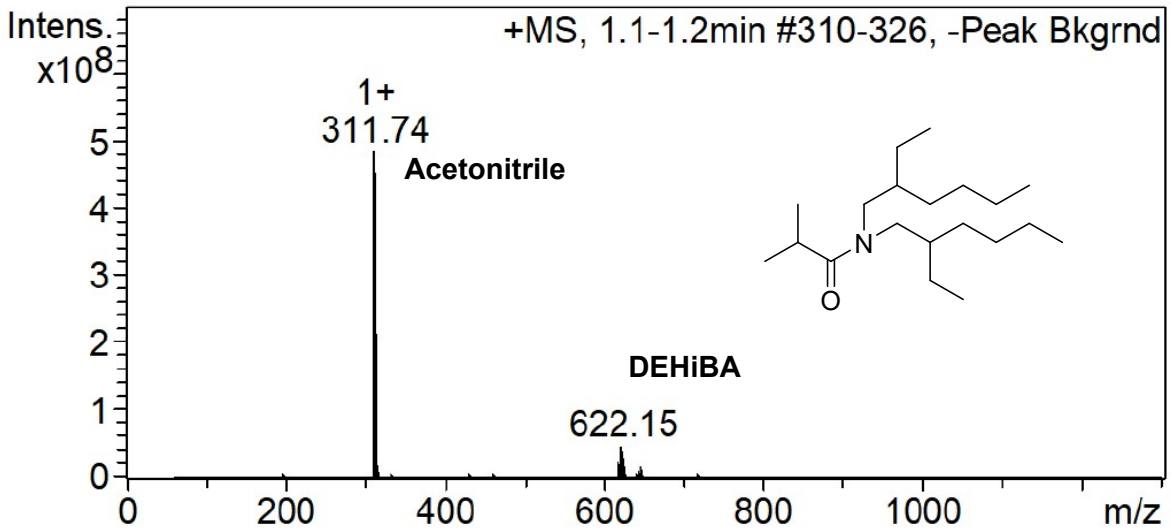


Fig S4. The GC-MS result of DEHiBA product

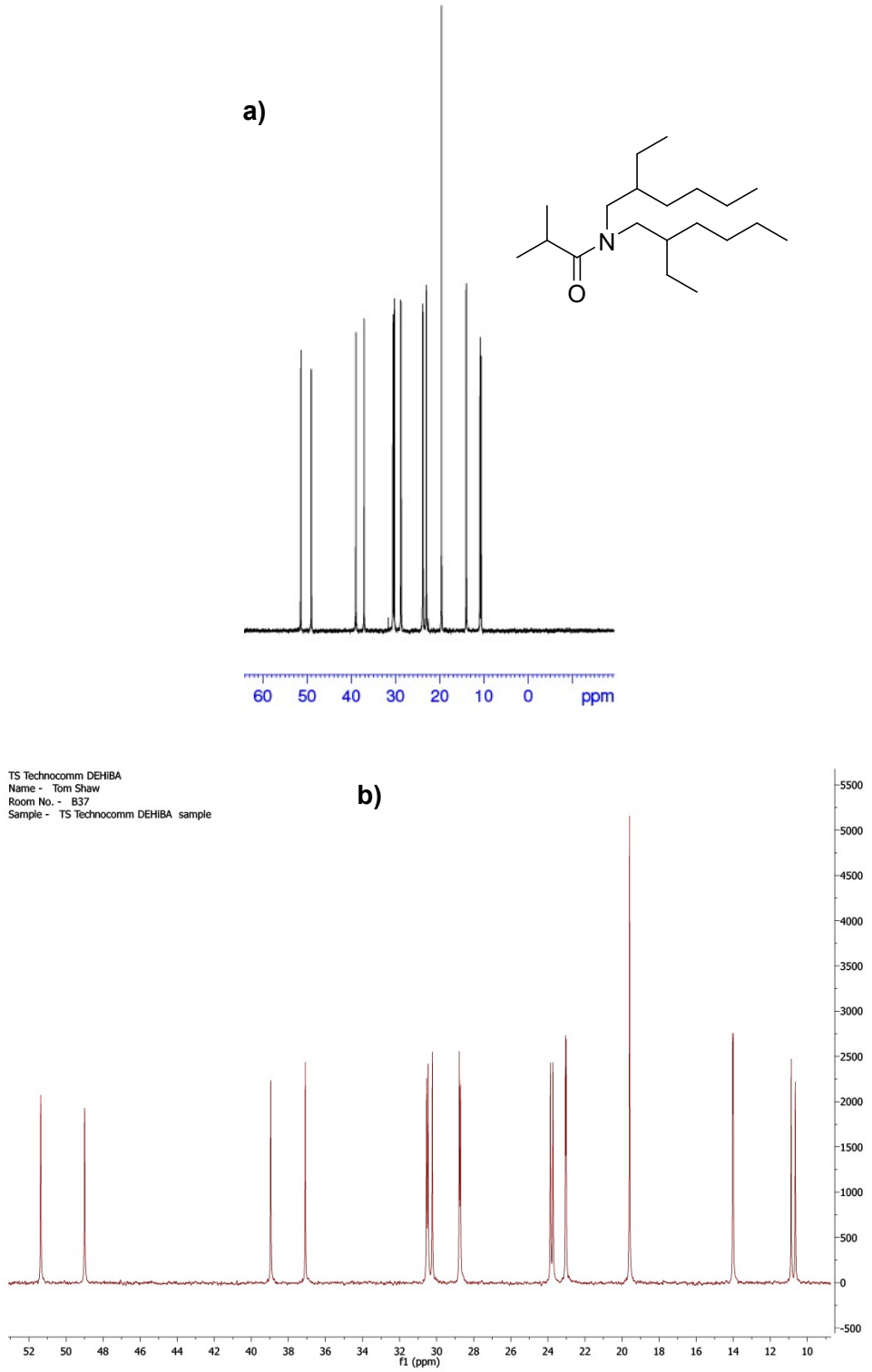


Fig S5. The NMR results of DEHiBA product a) pure DEHiBA from experiment and b) pure DEHiBA from market (Technocorm company)