Supplementary Information

Smart Mechanochemistry: Optimizing Amino Acid Acylation with One Factor at a Time, Design of Experiments and Machine Learning Methods

Adrien Gallego,[a,b] Matthieu Lavayssiere,[a] Xavier Bantreil,[a,c] Nicolas Pétry,[a] Julien Pinaud,[b] Olivia Giani*[b] and Frédéric Lamaty*[a]

[a] IBMM, CNRS, ENSCM, Université de Montpellier, France

[b] ICGM, CNRS, ENSCM, Université de Montpellier, France

[c] Institut Universitaire de France (IUF)

* olivia.giani@umontpellier.fr
frederic.lamaty@umontpellier.fr

Table des matières

General information

All reagents were purchased from Sigma Aldrich or BLD Pharmatech and used without further purification.

The milling experiments were carried out in a vibrating Retsch Mixer Mill 400 operated at 30 Hz. Milling load is defined as the ratio between the mass of reactant over the free volume of the jar.

NMR analyses were performed at the UAR PAC Balard. ^{1}H NMR spectra were recorded on a Bruker AVANCE 400 MHz or 500 MHz and are reported in ppm using deuterated solvents ($CDCl_3$ at 7.26 ppm and DMSO-d_6 at 2.50 ppm) purchased from Cambridge Isotopes Laboratories (Eurisotop). Data are reported as s = singlet, br. s = broad singlet, d = doublet, t = triplet, dd = doublet doublet, m = multiplet, coupling constant in Hz, integration. ^{13}C NMR spectra were recorded on a Bruker AVANCE 101 MHz or 126 MHz spectrometer and are reported in ppm using solvent as an internal standard ($CDCl_3$ at 77.16 ppm and DMSO-d_6 at 39.52 ppm).

HPLC conversion was measured on a ThermoFischer Vanquish Core LC using a Chromolith® HighResolution RP-18 endcapped 50-4.6 mm column and a linear gradient of 0 to 100% CH_3CN/0.1% TFA in H_2O/0.1% TFA over 3 min, UV lamp detection at 214 nm. Flow rate: 3 mL/min. To monitor reactions in a ball-mill, a sample was taken from milling jar, dissolved in 1 mL of a mixture CH_3CN/H_2O and submitted to HPLC analysis.

HRMS analyses were performed on UPLC Acquity H-Class from Waters hyphenated to a Q-Tof mass spectrometer (Synapt G2-S from Waters) with a dual ESI source.

Melting points were measured on an Auto Melting Point Apparatus (MP120) from Hanon Instruments.

DoE was performed through the Ellistat software (version 7.8.7 2024/03), available at https://www.ellistat.com

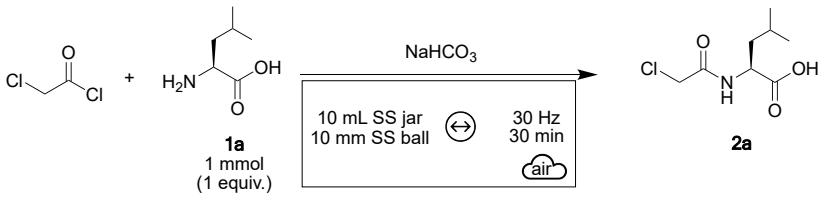
Regarding the reaction time, the values indicated by the DoE and BO were rounded to the nearest 5 minutes for experimental considerations.

The NMR adjusted yield values are reported as: mean value of several experiments ± 1.96*standard error (corresponding to a 95% confidence interval).

Determination of the NMR adjusted yield

In this part, the reaction described on Scheme 1 is considered. After completion of the milling, the reaction media is portioned in between ethyl acetate and a 1M HCl aqueous solution. The

aqueous phase is extracted with ethyl acetate. The organic phases are combined, dried over Na_2SO_4 and concentrated under reduced pressure.



*Scheme 1 - Acylation of L-Leucine **1a** under mechanochemical conditions*

A ^{1}H NMR analysis is performed on the crude whose mass is m_{tot} = 246 mg.

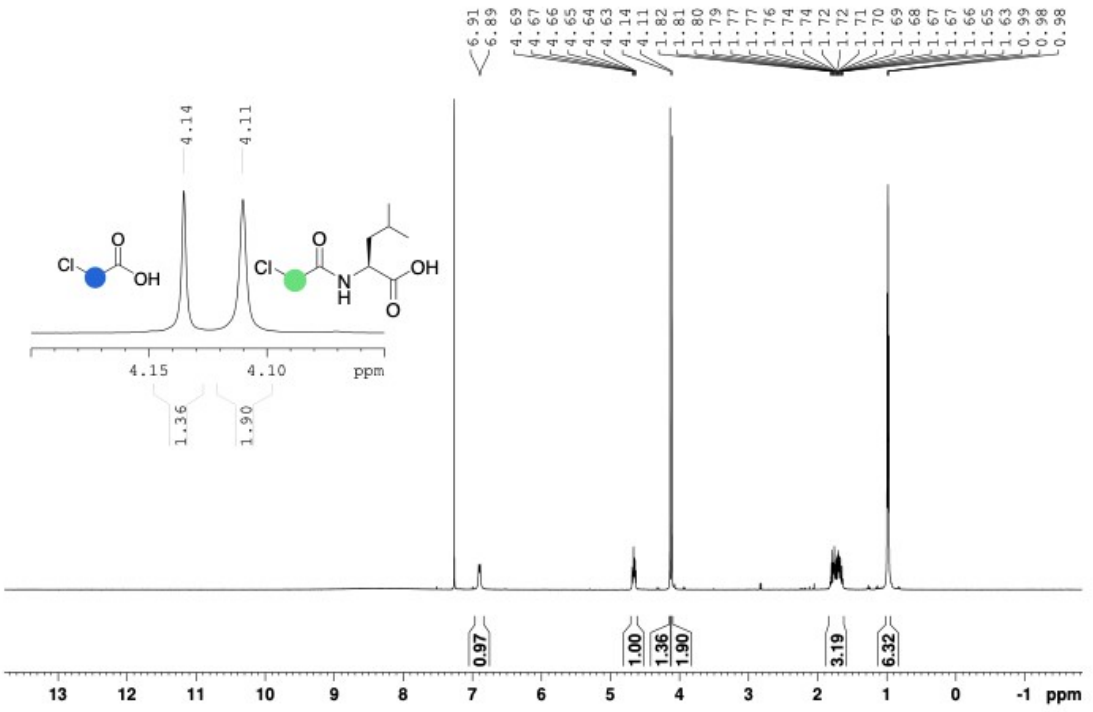


Figure 1 - ^{1}H NMR spectrum of the crude product

The NMR spectrum reveals a peak at 4.14 ppm, corresponding to the -CH_2 of chloroacetic acid. The proportion of desired product in the crude could be estimated thanks to the following equation.

$$m_{product} = \frac{\frac{Integration}{Number\ of\ protons} \times MW_{product}}{\frac{Integration}{Number\ of\ protons} \times MW_{product} + \frac{Integration}{Number\ of\ protons\ of\ impurity} \times MW_{impurity}} \times m_{tot}$$

In this case:

$$m_{product} = \frac{\frac{1.90}{2} \times 207.65}{\frac{1.90}{2} \times 207.65 + \frac{1.36}{2} \times 94.5} \times 246 \approx 186\ mg$$

The NMR adjusted yield can therefore be calculated:

$$NMR\ adjusted\ yield = \frac{m_{product}}{MW_{product}} \times \frac{1}{n_{H-Leu-OH}} = \frac{186}{207.65} \times \frac{1}{1} = 89\%$$

Design of Experiments

Response Surface Methodology (RSM) utilizing a Composite Face-centered (CCF) design was applied to evaluate the influence of three critical variables – chloroacetyl chloride quantity, $NaHCO_3$ quantity and milling time – on the NMR-adjusted yield of the acylation reaction.

The CCF design enables systematic exploration of the experimental space through three distinct types of design points:

(a) **Factorial points**, representing combinations of variables at low (-1) and high (+1) levels, which define the edges of the experimental cube
(b) **Axial points**, positioned at the center of each cube face ($\alpha = \pm 1$)
(c) **Center points**, located at the midpoint of all factors, used to assess experimental error and model stability

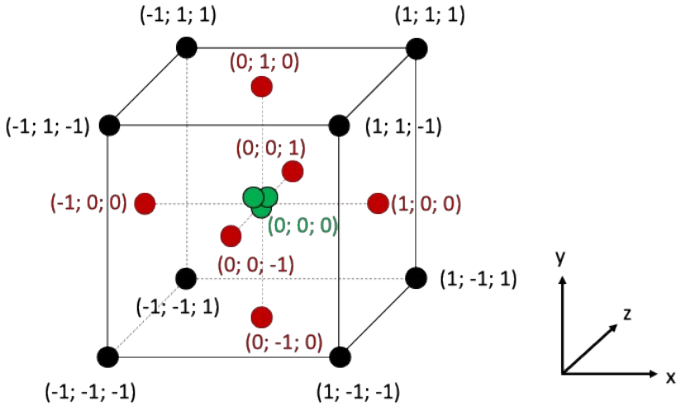


Figure 2 - Composite face-centered design

In the following tables, several statistical parameters are reported:

- Student's **t** parameter, used to assess whether the difference between means is statistically significant;
- Fischer's **F** parameter, used to compare two variances or to test whether a group of variables significantly explains the variation in a model;
- **p-value**, interpreted as the probability of obtaining the observed results, or more extreme ones, if the null hypothesis were true. A small p-value suggests the effect is statistically significant.

First DoE

At the beginning, unsuitable ranges of variations were set leading to surface response showed in Figure 3. The best value is reached at the edges of the reaction space. As the optimum is not clearly defined, an extension of the reaction space was necessary.

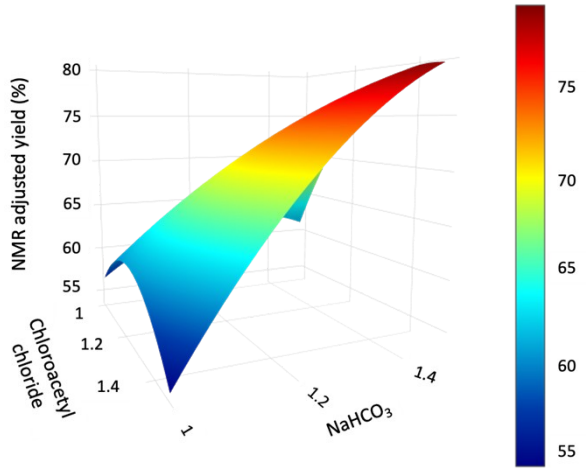
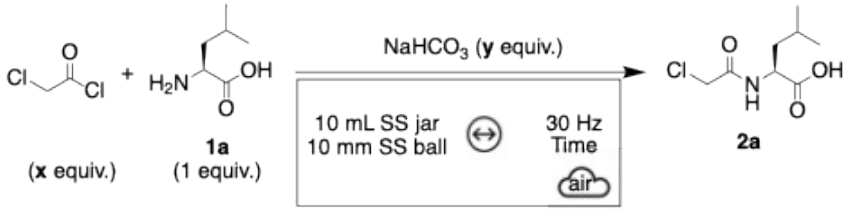


Figure 3 - Correlation between amount of chloroacetyl chloride, amount of NaHCO3 and NMR adjusted yield with reaction time fixed at 30 min

Second DoE

The experimental data related to this DoE is reported in Table 1.

Table 1 - Experimental data, including replicates, used in the DoE



Entry	**x** (equiv.)	**y** (equiv.)	Reaction time (min)	NMR adjusted yield (%)
1	1	1	5	50
2	1	1	5	50
3	1	1	5	56
4	1	1	60	65
5	1	1	60	55
6	1	1	60	56
7	1	4	5	55
8	1	4	5	51
9	1	4	60	60
10	1	4	60	61

11	2	1	5	62
12	2	1	60	66
13	2	4	5	65
14	2	4	5	90
15	2	4	60	82
16	2	4	60	83
17	1.5	2	30	91
18	1.5	2	30	90
19	1.5	2	30	90
20	1.5	2	30	94
21	1.5	2	30	88
22	1	2	30	56
23	1	2	30	56
24	2	2	30	89
25	2	2	30	85
26	1.5	1	30	68
27	1.5	4	30	86
28	1.5	4	30	100
29	1.5	2	5	93
30	1.5	2	5	87
31	1.5	2	60	81
32	1.5	2	60	84

According to the experimental results, the following mathematical model was established:

Equation 1 - Quadratic model

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{11}X_1^2 + b_{22}X_2^2 + b_{33}X_3^2 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3$$

Where:

- Y is the response i.e., NMR adjusted yield;
- X_1, X_2 and X_3 are the coded levels for the variables, respectively amount of chloroacetyl chloride, $NaHCO_3$ quantity and milling time;
- b_0 is a constant coefficient;
- b_1, b_2 and b_3 are the coefficient for the linear effects;
- b_{11}, b_{22} and b_{33} are the coefficient for the quadratic effects;
- b_{12}, b_{13} and b_{23} are the coefficient for the interaction effects.

The values of the coefficients are reported in Table 2.

Table 2 - Values for regression coefficient for the response surface quadratic model defined by Eq. 1

Coefficient	Value	Standard error	t-value	p-value [a]

b_0	-114.714	25.65	-4.473	0.000
b_1	218.052	38.7	5.635	0.000
b_2	14.2162	9.34	1.522	0.142
b_3	0.372565	0.343	1.086	0.289
b_{11}	-68.788	12.72	-5.408	0.000
b_{22}	-3.54584	1.721	-2.060	0.051
b_{33}	-0.00341554	0.004244	-0.805	0.430
b_{12}	4.2091	2.468	1.706	0.102
b_{13}	-0.0922225	0.1356	-0.680	0.504
b_{23}	0.0168229	0.04359	0.386	0.703

[a] Significant if p-value < 0.05

An analysis of variance was also performed.

Table 3 - Analysis of variance for the response surface quadratic model

Source	Degree of freedom	Mean square	F-value	p-value [a]
X_1	1	1586.6	31.7484	0.0000
X_2	1	115.79	2.3168	0.1422
X_3	1	58.977	1.1801	0.2891
X_1^2	1	1461.5	29.2434	0.0000
X_2^2	1	212.15	4.2450	0.0514
X_3^2	1	32.364	0.6476	0.4296
$X_1 X_2$	1	145.4	2.9095	0.1021
$X_1 X_3$	1	23.111	0.4625	0.5036
$X_2 X_3$	1	7.4448	0.1490	0.7032
Residual	9	49.976		

[a] Significant if p-value < 0.05

DoE including OFAT results

The corresponding Response Surface is shown in Figure 4.

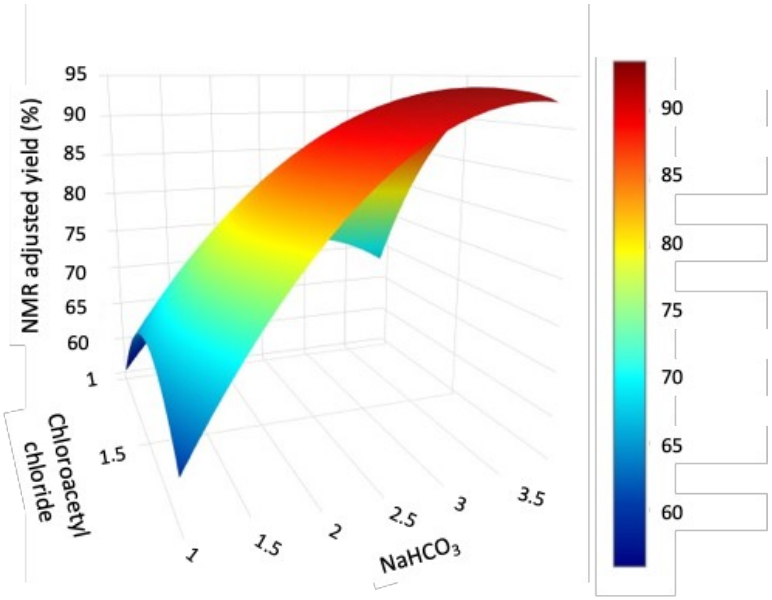


Figure 4 - Correlation between amount of chloroacetyl chloride, amount of $NaHCO_3$ and NMR adjusted yield with reaction time fixed at 30 min for the set OFAT + DoE

By implementing the results of the experiments performed in the OFAT section, a new mathematical model was established:

$$Y = b'_0 + b'_1X'_1 + b'_2X'_2 + b'_3X'_3 + b'_{11}X'^2_1 + b'_{22}X'^2_2 + b'_{33}X'^2_3 + b'_{12}X'_1X'_2 + b'_{13}X'_1X'_3 + b'_{23}X'_2X'_3$$

Table 4 - Values of coefficients of the model for the DoE including the OFAT data

Coefficient	Value	Standard error	t-value	p-value [a]
b'_0	-55.1263	14.8	-3.726	0.000
b'_1	125.7650	18.99	6.623	0.000
b'_2	15.8624	6.049	2.622	0.011
b'_3	0.393205	0.2112	1.862	0.067
b'_{11}	-39.9022	6.521	-6.119	0.000
b'_{22}	-4.33298	1.061	-4.085	0.000
b'_{33}	-0.00407138	0.002621	-1.553	0.125
b'_{12}	7.22697	2.357	3.066	0.003
b'_{13}	-0.0574181	0.1034	-0.555	0.581
b'_{23}	-0.0130537	0.04149	-0.315	0.754

[a] Significant if p-value < 0.05

The corresponding analysis of variance is reported in Table 5.

Table 5 – Analysis of variance (ANOVA)

Source	Degree of freedom	Mean square	F-value	p-value [a]
X_1	1	1469.6	43.8623	0.0000
X_2	1	230.4	6.8767	0.0108
X_3	1	116.18	3.4674	0.0670
X_1^2	1	1254.7	37.4469	0.0000
X_2^2	1	559.23	16.6911	0.0001
X_3^2	1	80.844	2.4129	0.1251
$X_1 X_2$	1	314.87	9.3979	0.0031
$X_1 X_3$	1	10.33	0.3083	0.5806
$X_2 X_3$	1	3.3157	0.0990	0.7541
Residual	66	33.505		

[a] Significant if p-value < 0.05

Implemented Bayesian optimization

General information for the Bayesian optimization

The implementation of the optimization workflow relied on several essential Python libraries. **Pandas** and **NumPy** were used for efficient data manipulation and numerical operations, respectively, forming the backbone of the data preprocessing and feature engineering steps. The **scikit-learn** library provided the core functionality for Gaussian Process Regression (GaussianProcessRegressor) along with kernel definitions such as RBF (Radial Basis Function) and ConstantKernel, as well as normalization tools like MinMaxScaler to ensure the features were properly scaled before model training. To support statistical computations, particularly in the acquisition function (e.g., Expected Improvement), the **SciPy** library offered access to probability distributions (scipy.stats.norm) and distance metrics (scipy.spatial.distance.cdist).

For cheminformatics tasks, the **RDKit** library was employed to compute molecular fingerprints using GetMorganGenerator, enabling similarity calculations and metadata-based modeling through tools such as Chem and DataStructs.

In terms of visualization, **Matplotlib** and **Seaborn** were used, **Matplotlib** served for general plotting, while **Seaborn** provided statistical visualizations such as heatmaps and correlation matrices, facilitating the interpretation of model behavior and experimental trends.

Note that **RDKit** had to be used on **Jupyterlab.**

Data preparation

The data frame of the currently evaluated reaction `df_hplc` is given the alpha value `alpha_hplc = 5e-2` to consider the elevated noise inherent to the reaction.

The data frame of the previously evaluated reaction `df_Leu` is given the alpha value `alpha_Leu = 8e1` calculation of this value is explained in Improved stabilization of the surrogate.

The data frames are concatenated and scaled

```
df_all = pd.concat([df_Leu, df_hplc], ignore_index=True)
features = ["Eq NaHCO3", "Eq ClACl", "ML", "time (min)"]
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(df_all[features])
```

The combined objective is defined as a weighted sum of either HPLC conversion or NMR adjusted yield and the corresponding 1/PMI value, both variables having therefore the same range [0;1] and optimization direction (the variables are named after “yield” and “NMR yield” as an artifact of the **1a** reaction).

```
w_yield = 0.5
w_PMI = 0.5
Y_combined = (w_yield * df_all["NMR yield"].to_numpy() +
              w_PMI * df_all["1/PMI"].to_numpy()).reshape(-1, 1)
```

Surrogate model generation

For information on `Kernel` choices and `GaussianProcessRegressor` parameters read `sklearn.gaussian_process` documentation, a simple radial basis function kernel is used with a vector with the same number of dimensions as the inputs X as length scale parameter (anisotropic variant of the kernel).

```
kernel = ConstantKernel(1.0, (1e-2, 1e5)) *
RBF(length_scale=np.ones(4), length_scale_bounds=(1e-2, 1e7))

gp = GaussianProcessRegressor(kernel=kernel, n_restarts_optimizer=50,
alpha=alpha, normalize_y=True)

gp.fit(X_scaled, Y_combined)
```

Length scale of each parameter can give similar information than p values (see above) on the order of magnitude of the influence of each input variable on the function the GP is modeling. To extract use :

```
rbf_kernel = gp.kernel_.k2
if hasattr(rbf_kernel, 'length_scale'):
    print("Optimized length scales:", rbf_kernel.length_scale)
```

Identify the most informative experimental conditions

The role of the acquisition function is to quickly identify what are the best conditions and guide the optimization to the global maxima. A commonly used acquisition function is Expected Improvement as it is good at balancing exploration and exploitation.

```
Parameters:
        X_candidates_df: pd.DataFrame
            Candidate points at which to evaluate the acquisition
function
```

```
        X_candidates_scaled: pd.DataFrame
            Candidate points at which to evaluate the acquisition
scaled
        model: GaussianProcessRegressor
            Trained GP model
        y_best: float
            Best observed objective value so far
        xi: float
            Exploration/exploitation parameter
Returns:
        ei: np.ndarray
```

Classical expected improvement would be defined as below

```
def expected_improvement(X_candidates_df, X_candidates_scaled, model,
y_best, xi=0.01):

    X_candidates = X_candidates_df.values

    mu, sigma = model.predict(X_candidates, return_std=True)
    sigma = sigma.reshape(-1, 1)
    mu = mu.reshape(-1, 1)

    with np.errstate(divide="warn"):
        Z = (mu - y_best - xi) / sigma
        ei = (mu - y_best - xi) * norm.cdf(Z) + sigma * norm.pdf(Z)
        ei[sigma == 0.0] = 0.0

    return ei.flatten()
```

But this function, as it relies on the best observed objective value so far, is very subject to a "lucky" event in a noisy data set.

An alternative is to replace the best observed objective value `y_best` by the best mean value `mu.max()`.

```
def expected_improvement_mu(X_candidates_df, X_candidates_scaled,
model, y_best, xi=0.01):

    X_candidates = X_candidates_df.values

    mu, sigma = model.predict(X_candidates, return_std=True)
    sigma = sigma.reshape(-1, 1)
    mu = mu.reshape(-1, 1)
    mu_best = mu.max()

    with np.errstate(divide="warn"):
        Z = (mu - mu_best - xi) / sigma
        ei = (mu - mu_best - xi) * norm.cdf(Z) + sigma * norm.pdf(Z)
```

```
        ei[sigma == 0.0] = 0.0

    return ei.flatten()
```

The link between yield and PMI imply that the best best value of the PMI can only be obtained if `NMR_Yield = 100%`, but there is no reciprocity as the best value of yield can be obtained using excess reactants. PMI is dependent on the `NMR_Yield` and the mass of reactants, and the theorical max PMI value of each candidate points can therefore be calculated as:

```
max_PMI_value = (1 * mw_Prod)/(1 * mw_Phe+mw_NaHCO3 *
     X_candidates_df['Eq NaHCO3']+mw_ClACl*X_candidates_df["Eq ClACl"])
```

with `mw_Phe`, `mw_ClACl`, `mw_NaHCO3`, `mw_Prod` the molecular weight of the reactants and product of interest.

And the theoretical best `Y_combined` value of each candidate points can therefore be calculated as:

```
    max_y_value = 1*w_yield+w_ae *
        ((1 * mw_Prod)/ (1 * mw_Phe+mw_NaHCO3 *
         X_candidates_df['Eq NaHCO3']+mw_ClACl *
         X_candidates_df["Eq ClACl"]))
```

The mu expected improvement function is then penalized if the corresponding `max_y_value` is lower than the prediction from the `model`. It is, we believe, a convenient way to automatically actualize a constrained EI function that would remove all the candidates that could not improve the pareto in any way.

```
def tempered_expected_improvement(X_candidates_df, X, model, y_best,
xi=0.01):
    mu, sigma = model.predict(X, return_std=True)
    sigma = sigma.reshape(-1, 1)
    mu = mu.reshape(-1, 1)
    y_best = mu.max()
    beta = 8
    max_y_value = 1*w_yield+w_ae * \
        ((mw_Prod)/(mw_Phe+mw_NaHCO3 *
         X_candidates_df['Eq NaHCO3']+mw_ClACl*X_candidates_df["Eq
ClACl"]))
    max_y_value = max_y_value.values.reshape(-1, 1)

    with np.errstate(divide="warn"):
        Z = (mu - y_best - xi) / sigma
```

```
        ei = (mu - y_best - xi) * norm.cdf(Z) + sigma * norm.pdf(Z) * \
            (2 / (1 + np.exp(beta - (max_y_value*100/ y_best) * beta)))
        ei[sigma == 0.0] = 0.0
        ei[ei < 0.0] = 0.0
    return ei.flatten()
```

Minimizing overlap between the explored conditions

This function `select_top_diverse_candidates` selects the top `n` candidate points from a set of candidates (`X_candidates`) by balancing high `tempered_expected_improvement` value and the distance between points

```
def select_top_diverse_candidates(X_candidates, ei_values, n=2):
    sorted_idx = np.argsort(-ei_values)
    selected = [sorted_idx[0]]

    for _ in range(1, n):
        remaining = [i for i in sorted_idx if i not in selected]
        if not remaining:
            break
        dists = cdist(X_candidates[remaining], X_candidates[selected])
        min_dists = dists.min(axis=1)
        next_best_idx = remaining[np.argmax(min_dists)]
        selected.append(next_best_idx)

    return selected
```

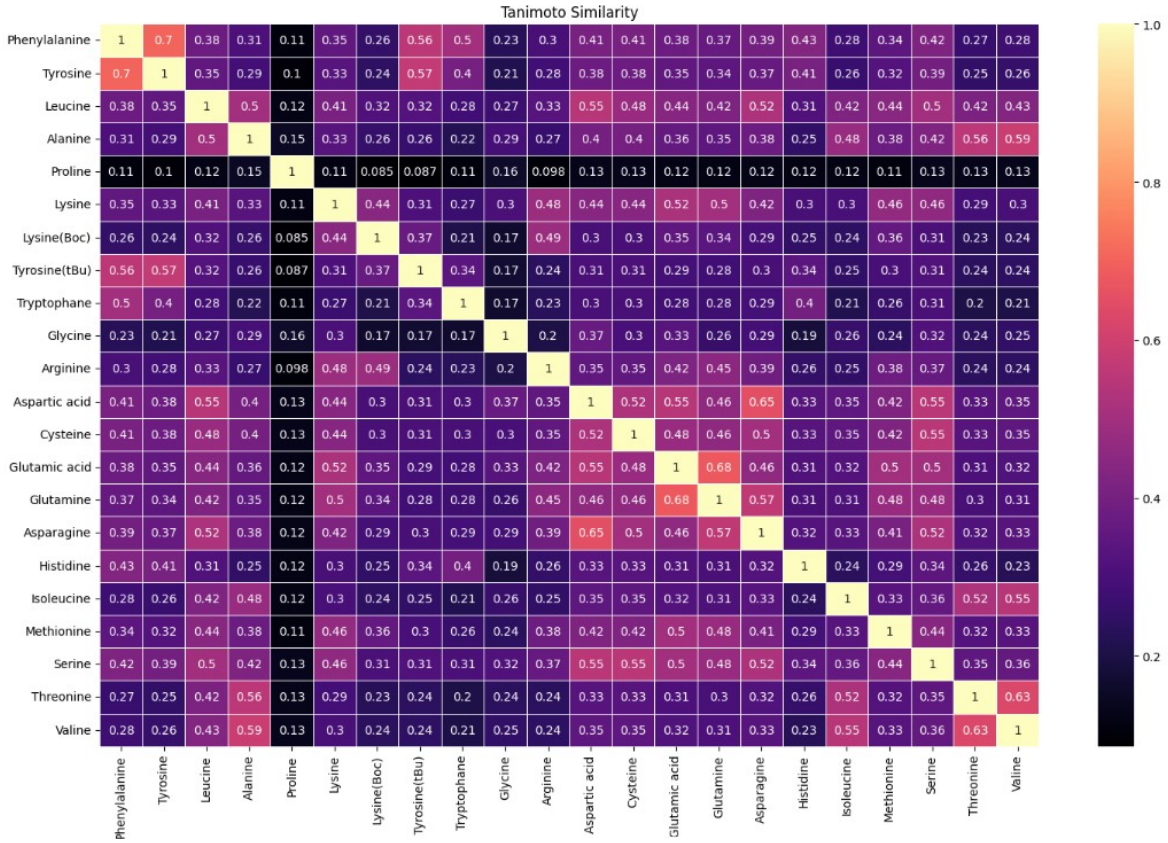
Improved stabilization of the surrogate

Once the acylation of both **1a** and **1b** was achieved we had now two data sets of 'low fidelity' data. To use both to stabilize the optimization the acylation of **1c** we needed to order them by their likelihood. The simplest approach other than giving equivalent value to each dataset would be to suggest a structure-reactivity relationship, to that end a multitude of descriptors allow to quantify the similarities between two molecules and machine learning tools have been developed. For this application we have used the Tanimoto index on the Morgan fingerprint of each amino acid, conveniently calculated using `DataStructs.TanimotoSimilarity`.

```
smiles_dict = {
    "Phenylalanine": "N[C@@H](Cc1ccccc1)C(O)=O",
    "Tyrosine": "N[C@@H](Cc1ccc(O)cc1)C(O)=O",
    "Leucine": "NC(CC(C)C)C(=O)O",
    "Alanine": "C[C@H](N)C(O)=O",
    "Proline": "OC(=O)[C@@H]1CCCN1",
```

```
    "Lysine": "NCCCC[C@H](N)C(O)=O",
    "Lysine(Boc)": "CC(C)(C)OC(=O)NCCCC[C@H](N)C(O)=O",
    "Tyrosine(tBu)":"N[C@@H](Cc1ccc(cc1)OC(C)(C)C)C(=O)O",
    "Tryptophane": "N[C@@H](Cc1c[nH]c2ccccc12)C(O)=O",
    "Glycine": "NCC(O)=O",
    "Arginine": "N[C@@H](CCCNC(N)=N)C(O)=O",
    "Aspartic acid": "N[C@@H](CC(O)=O)C(O)=O",
    "Cysteine": "N[C@@H](CS)C(O)=O",
    "Glutamic acid": "N[C@@H](CCC(O)=O)C(O)=O",
    "Glutamine": "N[C@@H](CCC(N)=O)C(O)=O",
    "Asparagine": "N[C@@H](CC(N)=O)C(O)=O",
    "Histidine": "N[C@@H](Cc1c[nH]cn1)C(O)=O",
    "Isoleucine": "CC[C@H](C)[C@H](N)C(O)=O",
    "Methionine": "CSCC[C@H](N)C(O)=O",
    "Serine": "N[C@@H](CO)C(O)=O",
    "Threonine": "C[C@@H](O)[C@H](N)C(O)=O",
    "Valine": "CC(C)[C@H](N)C(O)=O",
}

mols = {name: Chem.MolFromSmiles(smi) for name, smi in
smiles_dict.items()}
generator = GetMorganGenerator(radius=2, fpSize=2048)
fps = {name: generator.GetFingerprint(mol) for name, mol in
mols.items()}

names = list(fps.keys())
sim_mat = pd.DataFrame(index=names, columns=names)

for i in names:
    for j in names:
        sim_mat.loc[i, j] = DataStructs.TanimotoSimilarity(fps[i],
fps[j])

sim_mat = sim_mat.astype(float)

plt.figure(figsize=(15, 10))
sns.heatmap(sim_mat, annot=True, cmap="magma", linewidths=0.5)
plt.title("Tanimoto Similarity")
plt.tight_layout()
plt.show()
```

output :



For the acylation of **1c-g,** the data frame of the currently evaluated reaction `df_hplc` is given the alpha value `alpha_hplc = 5e-2` to consider the elevated noise inherent to the reaction.

The data frame of the previously evaluated reaction `df_Aaa` is given the alpha value `alpha_Aaacurrent_Aaa = 3e1/ Tanimoto_Aaacurrent_Aaa.`

Where `Tanimoto_Aaacurrent_Aaa` correspond to the Tanimoto value between the currently evaluated amino acid and the one from the previous dataframe.

Experimental assays for Bayesian optimization

Acylation of *L*-Leucine **1a**

Simple BO

The BO algorithm was initiated with the five experiments presented in Table 6 which were designed using a simple centered factorial DoE model.

Table 6 - Experiments used for the initialization of the BO algorithm

Entry	Chloroacetyl chloride (equiv.)	$NaHCO_3$ (equiv.)	Reaction time (min)	NMR adjusted yield (%)
1	1	1	60	65
2	2	1	5	62
3	1	2	5	62
4	1.5	1.5	30	76
5	2	2	60	84

The BO algorithm suggested the 5 following experiments. Only the two showing the highest expected improvement (EI) were run. In our case, it corresponds to Candidate#1 and candidate #2 (in green).

Iteration 1 (Entries 6 and 7)
Optimized kernel: 0.976**2 * RBF(length_scale=[0.505, 3.96e+06, 0.353])
Optimized length scales: [5.05137619e-01 3.96412011e+06 3.52558553e-01]

--- Coordinates of Best Predicted Point ---
→ w_yield = 1
→ w_ae = 0
→ Eq NaHCO3 = 1.99
→ Eq ClACl = 1.87
→ Time = 57 min
Predicted Combined Objective (mu) = 83.4494
real Combined Objective (mu) = 84.0000

--- Top 5 Diverse High-EI Candidates --

Candidate #1
→ Eq NaHCO3 = 2.10
→ Eq ClACl = 2.97
→ Time = 51 min
μ value = 82.2289
max μ value = 1.0000
Expected Improvement = 1.1527

Candidate #2
→ Eq NaHCO3 = 5.93
→ Eq ClACl = 1.01

→ Time = 10 min
μ value = 69.8000
max μ value = 1.0000
Expected Improvement = 0.2017

Candidate #3
→ Eq NaHCO3 = 1.02
→ Eq ClACl = 2.97
→ Time = 30 min
μ value = 67.4958
max μ value = 1.0000
Expected Improvement = 0.0123

Candidate #4
→ Eq NaHCO3 = 5.99
→ Eq ClACl = 1.01
→ Time = 40 min
μ value = 69.8000
max μ value = 1.0000
Expected Improvement = 0.2017

Candidate #5
→ Eq NaHCO3 = 6.00
→ Eq ClACl = 1.11
→ Time = 21 min
μ value = 69.8000
max μ value = 1.0000
Expected Improvement = 0.2017

Iteration 2 (Entries 8 and 9)
Optimized kernel: 0.922**2 * RBF(length_scale=[0.136, 1.04e+05, 0.336])
Optimized length scales: [1.36043932e-01 1.04196350e+05 3.35932315e-01]

--- Coordinates of Best Predicted Point ---
→ w_yield = 1
→ w_ae = 0
→ Eq NaHCO3 = 2.18
→ Eq ClACl = 1.29
→ Time = 49 min
Predicted Combined Objective (mu) = 87.3445
real Combined Objective (mu) = 88.0000

--- Top 5 Diverse High-EI Candidates ---

Candidate #1
→ Eq NaHCO3 = 2.43
→ Eq ClACl = 2.95
→ Time = 50 min
μ value = 86.1016
max μ value = 1.0000

Expected Improvement = 1.2030

Candidate #2
→ Eq NaHCO3 = 5.99
→ Eq ClACl = 1.10
→ Time   = 10 min
μ value   = 71.1044
max μ value = 1.0000
Expected Improvement = 0.0000

Candidate #3
→ Eq NaHCO3 = 1.04
→ Eq ClACl = 1.29
→ Time   = 30 min
μ value   = 71.1803
max μ value = 1.0000
Expected Improvement = 0.0014

Candidate #4
→ Eq NaHCO3 = 5.91
→ Eq ClACl = 1.14
→ Time   = 40 min
μ value   = 72.1654
max μ value = 1.0000
Expected Improvement = 0.1386

Candidate #5
→ Eq NaHCO3 = 1.06
→ Eq ClACl = 2.96
→ Time   = 19 min
μ value   = 67.2866
max μ value = 1.0000
Expected Improvement = 0.0000

Iteration 3 (Entry 10)
Optimized kernel: 0.957**2 * RBF(length_scale=[0.207, 2.73e+05, 0.402])
Optimized length scales: [2.07200884e-01 2.72974326e+05 4.02326860e-01]

--- Coordinates of Best Predicted Point ---
→ w_yield   = 1
→ w_ae   = 0
→ Eq NaHCO3 = 2.67
→ Eq ClACl = 2.04
→ Time   = 50 min
Predicted Combined Objective (mu) = 92.4879
real Combined Objective (mu) = 93.0000

--- Top 5 Diverse High-EI Candidates ---

Candidate #1

→ Eq NaHCO3 = 2.93
→ Eq ClACl = 2.38
→ Time   = 52 min
μ value   = 91.6635
max μ value = 1.0000
Expected Improvement = 1.54824063

Candidate #2
→ Eq NaHCO3 = 5.96
→ Eq ClACl = 1.09
→ Time   = 10 min
μ value   = 71.0319
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #3
→ Eq NaHCO3 = 1.02
→ Eq ClACl = 2.46
→ Time   = 31 min
μ value   = 71.3372
max μ value = 1.0000
Expected Improvement = 0.00000031

Candidate #4
→ Eq NaHCO3 = 6.00
→ Eq ClACl = 1.05
→ Time   = 41 min
μ value   = 67.3914
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #5
→ Eq NaHCO3 = 5.96
→ Eq ClACl = 2.90
→ Time   = 21 min
μ value   = 69.1805
max μ value = 1.0000
Expected Improvement = 0.00000000

Iteration 4 (Entry 12)
Optimized kernel: 0.895**2 * RBF(length_scale=[0.179, 7.63e+06, 0.378])
Optimized length scales: [1.78543818e-01 7.62782322e+06 3.78202495e-01]

--- Coordinates of Best Predicted Point ---
→ w_yield   = 1
→ w_ae   = 0
→ Eq NaHCO3 = 2.43
→ Eq ClACl = 3.91
→ Time   = 49 min
Predicted Combined Objective (mu) = 90.6571

real Combined Objective (mu) = 93.0000

--- Top 5 Diverse High-EI Candidates ---

Candidate #1
→ Eq NaHCO3 = 2.43
→ Eq ClACl = 1.89
→ Time   = 45 min
μ value   = 90.2853
max μ value = 1.0000
Expected Improvement = 0.75401096

Candidate #2
→ Eq NaHCO3 = 5.90
→ Eq ClACl = 3.96
→ Time   = 10 min
μ value   = 71.0895
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #3
→ Eq NaHCO3 = 1.00
→ Eq ClACl = 1.03
→ Time   = 27 min
μ value   = 70.8053
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #4
→ Eq NaHCO3 = 5.90
→ Eq ClACl = 3.91
→ Time   = 60 min
μ value   = 70.5102
max μ value = 1.0000
Expected Improvement = 0.00648610

Candidate #5
→ Eq NaHCO3 = 5.95
→ Eq ClACl = 3.88
→ Time   = 36 min
μ value   = 67.4420
max μ value = 1.0000
Expected Improvement = 0.00000000

Iteration 5 (Entry 14)
Optimized kernel: 0.923**2 * RBF(length_scale=[0.193, 1.96, 0.386])
Optimized length scales: [0.19294497 1.95936135 0.38638683]

--- Coordinates of Best Predicted Point ---
→ w_yield   = 1

→ w_ae   = 0
→ Eq NaHCO3 = 2.43
→ Eq ClACl = 3.99
→ Time   = 51 min
Predicted Combined Objective (mu) = 90.7935
real Combined Objective (mu) = 93.0000

--- Top 5 Diverse High-EI Candidates ---

Candidate #1
→ Eq NaHCO3 = 2.47
→ Eq ClACl = 4.00
→ Time   = 52 min
μ value   = 90.7668
max μ value = 1.0000
Expected Improvement = 0.96136377

Candidate #2
→ Eq NaHCO3 = 5.96
→ Eq ClACl = 1.05
→ Time   = 10 min
μ value   = 71.0374
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #3
→ Eq NaHCO3 = 1.02
→ Eq ClACl = 3.94
→ Time   = 31 min
μ value   = 72.7072
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #4
→ Eq NaHCO3 = 5.92
→ Eq ClACl = 1.01
→ Time   = 41 min
μ value   = 68.0123
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #5
→ Eq NaHCO3 = 5.98
→ Eq ClACl = 1.07
→ Time   = 21 min
μ value   = 67.9989
max μ value = 1.0000
Expected Improvement = 0.00000000

Iteration 6 (Entry 15)

Optimized kernel: 0.947**2 * RBF(length_scale=[0.194, 1.22e+06, 0.384])
Optimized length scales: [1.94288544e-01 1.21582161e+06 3.83761479e-01]

--- Coordinates of Best Predicted Point ---
→ w_yield = 1
→ w_ae   = 0
→ Eq NaHCO3 = 2.45
→ Eq ClACl = 1.46
→ Time   = 51 min
Predicted Combined Objective (mu) = 89.7417
real Combined Objective (mu) = 93.0000

--- Top 5 Diverse High-EI Candidates ---

Candidate #1
→ Eq NaHCO3 = 2.64
→ Eq ClACl = 3.59
→ Time   = 60 min
μ value   = 88.1711
max μ value = 1.0000
Expected Improvement = 0.67936026

Candidate #2
→ Eq NaHCO3 = 5.83
→ Eq ClACl = 1.04
→ Time   = 10 min
μ value   = 71.0636
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #3
→ Eq NaHCO3 = 1.05
→ Eq ClACl = 3.99
→ Time   = 35 min
μ value   = 72.7744
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #4
→ Eq NaHCO3 = 5.98
→ Eq ClACl = 1.02
→ Time   = 47 min
μ value   = 71.1912
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #5
→ Eq NaHCO3 = 5.99
→ Eq ClACl = 1.00
→ Time   = 23 min

μ value   = 67.1591
max μ value = 1.0000
Expected Improvement = 0.00000000

Iteration 7 (Entry 16)
Optimized kernel: 0.983**2 * RBF(length_scale=[0.194, 7.58e+06, 0.413])
Optimized length scales: [1.94437424e-01 7.58240158e+06 4.13360878e-01]

--- Coordinates of Best Predicted Point ---
→ w_yield   = 1
→ w_ae   = 0
→ Eq NaHCO3 = 2.62
→ Eq ClACl = 3.77
→ Time   = 60 min
Predicted Combined Objective (mu) = 91.4365
real Combined Objective (mu) = 93.0000

--- Top 5 Diverse High-EI Candidates ---

Candidate #1
→ Eq NaHCO3 = 2.67
→ Eq ClACl = 3.97
→ Time   = 60 min
μ value   = 91.3952
max μ value = 1.0000
Expected Improvement = 0.78223392

Candidate #2
→ Eq NaHCO3 = 5.90
→ Eq ClACl = 1.34
→ Time   = 10 min
μ value   = 71.0282
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #3
→ Eq NaHCO3 = 1.06
→ Eq ClACl = 1.20
→ Time   = 35 min
μ value   = 72.7220
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #4
→ Eq NaHCO3 = 5.99
→ Eq ClACl = 1.05
→ Time   = 47 min
μ value   = 71.3404
max μ value = 1.0000
Expected Improvement = 0.00000000

Candidate #5
→ Eq NaHCO3 = 1.05
→ Eq ClACl = 3.99
→ Time = 22 min
μ value = 69.1961
max μ value = 1.0000
Expected Improvement = 0.00000000

The experiments ran for this section are combined in Table 7.

Table 7 - Conducted experiments for the acylation of L-Leucine using BO

Entry	Chloroacetyl chloride (equiv.)	$NaHCO_3$ (equiv.)	Reaction time (min)	NMR adjusted yield (%)
6	2.97	2.1	50	88
7	1.25	5.89	10	71
8	2.95	2.43	50	93
9	1.14	5.91	40	67
10	2.38	2.93	50	86
11	2.46	1.02	30	72
12	1.89	2.43	45	88
13	3.91	5.9	60	79
14	4	2.47	50	89
15	3.59	2.64	60	93
16	3.97	2.67	60	88

Combination of OFAT/DoE and BO

The Bayesian algorithm was initiated by the results from OFAT and DoE.

Table 8 - Experimental data used to initiate the Bayesian algorithm

Entry	Chloroacetyl chloride (equiv.)	$NaHCO_3$ (equiv.)	Reaction time (min)	Milling load (mg/mL)	NMR adjusted yield (%)
1	1.58	1.5	60	43.56	57
2	1.2	1.5	60	39.27	64
3	1.3	1.5	60	40.40	74
4	1.3	1.5	30	40.40	68
5	1.5	1.5	30	42.66	76
6	1.3	1.5	5	40.40	63
7	1.3	1.5	10	40.40	62
8	1.3	1.5	15	40.40	72
9	1	1	5	32.81	50
10	1.3	1.3	5	38.72	70

11	1	1.3	5	35.33	55
12	1	1.5	5	37.01	56
13	1.3	1.4	5	39.56	62
14	1.3	1.2	5	37.88	71
15	1.3	1	5	36.20	60
16	1.3	1	15	36.20	60
17	0.9	1.3	5	34.20	49
18	1.3	1.5	5	40.40	68
19	1.5	1.5	60	42.66	82
20	1.5	1.5	60	42.66	76
21	1.3	1.5	5	40.40	70
22	1.3	1.5	5	40.40	69
23	1	1	5	32.81	50
24	1	1	5	32.81	56
25	1.3	1.3	15	38.72	71
26	1.3	1.3	15	38.72	72
27	1.3	1.3	15	38.72	46
28	1.3	1.3	15	38.72	52
29	1.5	1.5	5	42.66	82
30	1.5	1.5	5	42.66	85
31	1.5	1.2	15	40.14	62
32	1.5	1.2	15	40.14	80
33	1.3	1.3	15	38.72	69
34	1.3	1.3	15	38.72	68
35	1.5	1.5	5	42.66	76
36	1.5	1.5	5	42.66	76
37	1.5	1.2	15	40.14	63
38	1.5	1.2	15	40.14	82
39	1.6	1.5	60	43.79	85
40	1.6	1.5	60	43.79	78
41	1.5	1.4	15	41.82	81
42	1.5	1.4	15	41.82	93
43	1.5	1.3	15	40.98	73
44	1.5	1.3	15	40.98	69
45	1.5	1.2	15	40.14	71
46	1.5	1.2	15	40.14	76
47	2	1.5	60	48.31	75
48	2	1.5	60	48.31	76
49	1.5	1	5	38.46	53
50	1	1	60	32.81	65
51	1.5	1	60	38.46	68
52	1	1.5	60	37.01	55

53	1.3	1.3	60	38.72	78
54	1	1.3	30	35.33	59
55	1.5	1.3	30	40.98	79
56	1.3	1	30	36.20	66
57	1.3	1.3	30	38.72	70
58	2	1	5	44.11	62
59	1	2	5	41.21	62
60	2	2	5	52.51	75
61	2	1	60	44.11	66
62	1	2	60	41.21	64
63	2	2	60	52.51	96
64	1	1.5	30	37.01	58
65	2	1.5	30	48.307	75
66	1.5	1	30	38.46	68
67	1.5	2	30	46.86	91
68	2	2	60	52.51	98
69	2	2	60	52.51	95
70	1.5	2	30	46.86	90
71	1.5	2	30	46.86	90
72	2	2	60	52.51	83
73	2	2	60	52.51	85
74	1.5	2	30	46.86	94
75	1.5	2	30	46.86	88
76	1.7	2	15	49.12	81
77	1.7	2	15	49.12	95
78	1.8	2	30	50.25	93
79	1.8	2	30	50.25	85
80	1	4	5	58.02	55
81	1	4	5	58.02	51
82	1	4	60	58.02	60
83	1	4	60	58.02	61
84	2	4	5	69.31	65
85	2	4	5	69.31	90
86	2	4	60	69.31	82
87	2	4	60	69.31	83
88	1.5	4	30	63.66	86
89	1.5	4	30	63.66	100
90	1	2	30	41.21	56
91	1	2	30	41.21	56
92	2	2	30	52.51	89
93	2	2	30	52.51	85
94	1.5	2	60	46.86	81

95	1.5	2	60	46.86	84
96	1.5	2	5	46.86	93
97	1.5	2	5	46.86	87
98	1.9	3.4	30	63.14	107
99	1.9	3.4	30	63.14	92
100	1.9	3.4	30	63.14	90
101	1.9	3.4	30	63.14	95
102	1.5	2.5	5	51.06	71
103	1.5	2.5	5	51.06	74
104	1.5	2.5	60	51.06	81
105	1.5	2.5	60	51.06	83
106	1	2.5	30	45.41	61
107	1	2.5	30	45.41	60
108	2	2.5	30	56.71	90
109	2	2.5	30	56.71	107
110	1.5	1.5	15	42.66	76
111	1.5	1.5	10	42.66	68
112	1.5	1	15	38.46	79

A series of 10 additional experiments was then conducted whose conditions are summarized in Table 9.

*Table 9 - Series of 10 experiments suggested by the Bayesian algorithm for the acylation of L-Leucine **1a***

Iteration	Equiv. $NaHCO_3$	Equiv. Chloroacetyl Chloride	Milling Time (min)	Milling Load (mg/mL)	NMR adjusted yield (%)	PMI
1	3.02	1.71	30	57.20	83	3.4
2	2.43	1.65	30	58.43	82	3.1
3	3.12	1.71	30	45.00	86	3.3
4	2.53	1.78	30	45.04	86	3.1
5	3.20	1.85	30	46.01	87	3.4
6	3.80	1.88	30	47.61	90	3.5
7	3.24	1.85	30	47.43	86	3.4
8	1.76	1.70	30	44.77	88	2.6
9	2.80	2.00	60	45.2	79	3.6
10	2.23	2.00	30	49.17	89	2.9

Acylation of *L*-Phenylalanine **1b**

Additionally to the results on *L*-Leucine **1a**, experimental results involving **1b** were used to initiated the algorithm. The corresponding experiments are reported in Table 10.

*Table 10 - Initialization phase of BO with **1b***

Entry	Equiv.	Equiv.	Milling	Milling	HPLC	PMI

	$NaHCO_3$ [a]	Chloroacetyl Chloride [a]	Time (min)	Load (mg/mL)	conversion (%)	
1	1.5	1.4	15	45.22	81	2.3
2	1.5	1.4	15	45.22	94	2.0
3	1.5	1.4	15	45.22	83	2.3
4	1.5	1.4	15	45.22	89	2.1
5	1.5	1.4	15	45.22	80	2.3
6	1.5	1.4	15	45.22	77	2.4
7	1	1	15	36.21	67	2.2
8	1	1	15	36.21	62	2.4
9	1.7	2	15	52.52	87	2.5
10	1.7	2	15	52.52	85	2.6
11	1.9	3.4	30	66.54	92	3.0
12	1.9	3.4	30	66.54	90	3.1

[a] Calculated for 1 equiv. of **1b**

Following the initiation, the experiments reported in Table 11 were part of the iteration study to stabilize the model.

*Table 11 - Iteration phase of BO with **1b***

Iteration	Equiv. $NaHCO_3$ [a]	Equiv. Chloroacetyl Chloride [a]	Milling Time (min)	Milling Load (mg/mL)	HPLC conversion (%)	PMI
1	1.99	2.56	60	55.50	73	3.5
2	1.99	2.45	60	43.10	77	3.3
3	1.95	1.58	15	20.80	66	3.2
4	2.01	1.61	15	35.90	66	3.2
5	1.58	1.46	30	57.26	79	2.4
6	1.58	1.46	35	34.94	82	2.3
7	1.19	1.46	15	20.44	76	2.3
8	5.78	2.41	60	50.00	62	6.2
9	1.15	2.41	45	20.03	73	3.0
10	5.41	1.30	30	49.99	74	4.3
11	5.78	2.41	60	20.00	89	4.3
12	1.26	1.30	45	19.96	77	2.2
13	5.06	2.04	50	20.12	92	3.7
14	3.93	1.59	10	49.94	87	3.2
15	1.04	1.55	20	20.04	70	2.5
16	1.35	1.32	40	36.20	75	2.4
17	3.12	1.86	50	49.92	80	3.3

[a] Calculated for 1 equiv. of **1b**

Finally, the model was tasked with identifying conditions that either maximize HPLC conversion, minimize the PMI value or strike a balance between the two objectives.

*Table 12 - Values of the Pareto front for the acylation of **1b***

Entry	Equiv.	Equiv.	Milling	Milling	HPLC	PMI

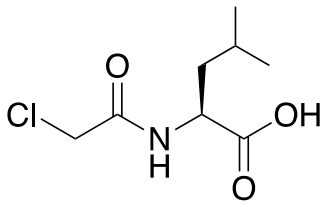
	$NaHCO_3$ [a]	Chloroacetyl Chloride [a]	Time (min)	Load (mg/mL)	conversion (%)	
1	1.5	1.26	30	39.9	76	2.4
2	1.5	1.26	30	39.9	78	2.3
3	1.49	1.39	30	43	84	2.2
4	1.49	1.39	30	43	83	2.2
5	2.04	5.05	30	20.31	96	3.5
6	2.04	5.05	30	20.31	97	3.5
7	2.04	5.26	30	20.53	100	3.5
8	2.04	5.26	30	20.53	100	3.5
9	1.03	1.31	30	22.74	68	2.4
10	1.03	1.31	30	22.74	69	2.3

[a] Calculated for 1 equiv. of **1b**

Experimental procedures and product characterizations

In this section, only the experimental conditions providing the best conversions/NMR adjusted yield are described.

N-(2-chloroacetyl)-*L*-Leucine [688-12-0]



2a

A 10 mL stainless steel (SS) jar was loaded with a 10 mm SS ball, H-*L*-Leu-OH (131.2 mg, 1 mmol, 1 equiv.), $NaHCO_3$ (285.6 mg, 3.40 mmol, 3.4 equiv.) and chloroacetyl chloride (151 µL, 1.90 mmol, 1.9 equiv.). The mixture was subjected to grinding at 30 Hz for 30 min. Then, the mixture was recovered from ethyl acetate and water. The water phase was acidified with a 1M HCl aqueous solution and extracted twice with ethyl acetate. The combined organic phases were washed with brine, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was finally washed with a 1M HCl aqueous solution to yield compound **2a** as a white solid (159.9 mg, 77% yield).

^{1}H NMR (500 MHz, $CDCl_3$) δ 6.93 (d, J = 8.3 Hz, 1H), 4.68 – 4.61 (m, 1H), 4.11 (s, 2H), 1.82 – 1.62 (m, 3H), 0.98 (d, J = 6.3 Hz, 3H), 0.97 (d, J = 6.3 Hz, 3H)

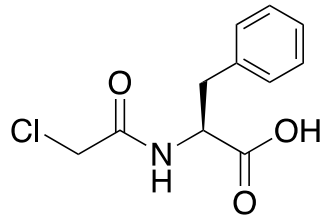
^{13}C NMR (126 MHz, $CDCl_3$) δ 176.8, 166.5, 51.1, 42.5, 41.1, 25.0, 22.9, 21.9

HRMS (ESI): $[M+H]^+$ Calculated for $C_8H_{15}ClNO_3$: 208.0735, found 208.0736

m.p.: 129.0 - 129.7°C

Data in agreement with literature.[1]

N-(2-chloroacetyl)-*L*-phenylalanine [721-65-3]



2b

A 10 mL stainless steel (SS) jar was loaded with a 10 mm SS ball, H-*L*-Phe-OH (38.4 mg, 0.23 mmol, 1 equiv.), $NaHCO_3$ (102.7 mg,

1.22 mmol, 5.26 equiv.) and chloroacetyl chloride (38 µL, 0.47 mmol, 2.04 equiv.). The mixture was subjected to grinding at 30 Hz for 30 min. Then, the mixture was recovered from ethyl acetate and water. The water phase was acidified with a 1M HCl aqueous solution and extracted twice with ethyl acetate. The combined organic phases were washed with brine, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was finally washed with a 1M HCl aqueous solution to yield compound **2b** as a white solid (44 mg, 79% yield).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 – 7.27 (m, 3H), 7.21 – 7.15 (m, 2H), 6.97 (d, J = 7.7 Hz, 1H), 4.94 – 4.87 (m, 1H), 4.06 (d, J = 18.7 Hz, 1H), 4.03 (d, J = 18.7 Hz, 1H), 3.25 (dd, J = 14.0, 5.5 Hz, 1H), 3.17 (dd, J = 14.0, 6.2 Hz, 1H)

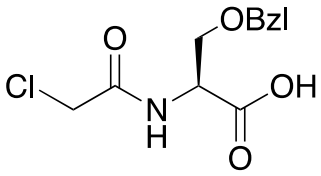
^{13}C NMR (126 MHz, $CDCl_3$) δ 174.7, 166.3, 135.13, 53.3, 42.5, 37.4

HRMS (ESI): $[M+H]^+$ Calculated for $C_{11}H_{13}ClNO_3$: 242.0584, found 242.0579

m.p.: 130.2 – 131.3°C

Data in agreement with literature.[1]

O-benzyl-*N*-(2-chloroacetyl)-*L*-serine [3062-02-0]



2c

A 10 mL stainless steel (SS) jar was loaded with a 10 mm SS ball, H-*L*-Ser(Bzl)-OH (93.6 mg, 0.48 mmol, 1 equiv.), $NaHCO_3$ (228.3 mg, 2.72 mmol, 5.7 equiv.) and chloroacetyl chloride (107 µL, 1.30 mmol, 2.79 equiv.). The mixture was subjected to grinding at 30 Hz for 10 min. Then, the mixture was recovered from ethyl acetate and water. The water phase was acidified with a 1M HCl aqueous solution and extracted twice with ethyl acetate. The combined organic phases were washed with brine, dried over Na_2SO_4 and concentrated under reduced pressure. The oily crude product was finally triturated in Et_2O at 0°C, filtered and dried under vacuum to yield compound **2c** as a white solid (115 mg, 88% yield).

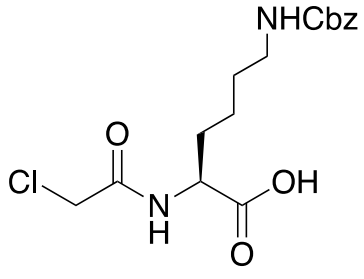
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.43 - 7.27 (m, 6H), 4.80 – 4.72 (m, 1H), 4.57 (d, J = 26.2 Hz, 1H), 4.55 (d, J = 26.2 Hz, 1H), 4.11 (d, J = 25.4 Hz, 1H), 4.08 (d, J = 25.4 Hz, 1H), 3.99 (dd, J = 9.6, 3.0 Hz, 1H), 3.72 (dd, J = 9.6, 3.4 Hz, 1H)

^{13}C NMR (126 MHz, $CDCl_3$) δ 173.9, 166.6, 137.1, 128.7, 128.2, 127.9, 73.6, 68.83, 52.9, 42.5

HRMS (ESI): $[M+Na]^+$ Calculated for $C_{12}H_{14}ClNO_4Na$: 336.0979, found 336.0974

m.p.: 114.8 – 115.4°C

N^6-((benzyloxy)carbonyl)-N^2-(2-chloroacetyl)-*L*-lysine [47376-73-8]



2d

A 10 mL stainless steel (SS) jar was loaded with a 10 mm SS ball, H-*L*-Lys(Cbz)-OH (104.9 mg, 0.40 mmol, 1 equiv.), $NaHCO_3$ (124.5 mg, 4.0 mmol, 3.96 equiv.) and chloroacetyl chloride (86 µL, 1.10 mmol, 2.87 equiv.). The mixture was

subjected to grinding at 30 Hz for 60 min. Then, the mixture was recovered from ethyl acetate and water. The water phase was acidified with a 1M HCl aqueous solution and extracted twice with ethyl acetate. The combined organic phases were washed with brine, dried over Na_2SO_4 and concentrated under reduced pressure. The oily crude product was finally triturated in Et_2O at 0°C, filtered and dried under vacuum to yield compound **2d** as a colorless wax (98 mg, 74% yield).

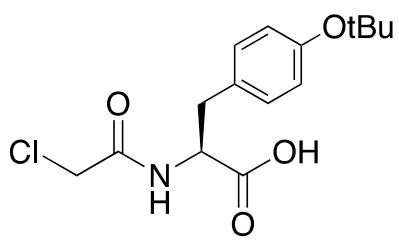
Two rotamers with a ratio 6.7/3.3

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 – 7.27 (m, 6H), 5.21 – 4.99 (m, 3H), 4.67 – 4.56 (m, 1H), 4.07 (s, 2H), 3.31 – 3.07 (m, 2H), 2.04 – 1.72 (m, 2H), 1.62 – 1.30 (m, 4H)

^{13}C NMR (126 MHz, $CDCl_3$) δ 174.6, 174.3, 166.9, 166.5, 158.5, 157.0, 136.4, 136.0, 128.7, 128.4, 128.3, 128.2, 128.1, 67.6, 67.0, 52.5, 42.5, 41.1, 40.6, 31.4, 29.5, 29.1, 22.2, 21.9

HRMS (ESI): $[M+H]^+$ Calculated for $C_{16}H_{22}ClN_2O_5$: 357.1212, found 357.1212

(S)-3-(4-(tert-butoxy)phenyl)-2-(2-chloroacetamido)propanoic acid



2e

A 10 mL stainless steel (SS) jar was loaded with a 10 mm SS ball, H-*L*-Tyr(tBu)-OH (152.7 mg, 0.64 mmol, 1 equiv.), $NaHCO_3$ (133.0 mg, 1.58 mmol, 2.46 equiv.) and chloroacetyl chloride (131 µL, 1.65 mmol, 2.56 equiv.). The mixture was subjected to grinding at 30 Hz for 60 min. Then, the mixture was recovered from ethyl acetate and water. The water phase was acidified with a 1M HCl aqueous solution and extracted twice with ethyl acetate. The combined organic phases were washed with brine, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was finally washed with a 1M HCl aqueous solution to yield compound **2e** as a white solid (178 mg, 87% yield).

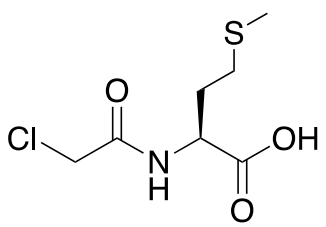
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.10 – 7.03 (m, 2H), 7.00 – 6.91 (m, 3H), 4.89 – 4.80 (m, 1H), 4.05 (d, J = 21.2 Hz, 1H), 4.01 (d, J = 21.2 Hz, 1H), 3.20 (dd, J = 14.2, 5.5 Hz, 1H), 3.12 (dd, J = 14.2, 6.2 Hz, 1H), 1.33 (s, 1H)

^{13}C NMR (101 MHz, $CDCl_3$) δ 174.4, 166.3, 154.6, 129.8, 124.5, 78.8, 53.3, 42.3, 36.6, 28.8

HRMS (ESI): $[M+Na]^+$ Calculated for $C_{15}H_{20}ClNO_4Na$: 336.0979, found 336.0974

m.p.: 119.7 – 121.2°C

(2-chloroacetyl)-*L*-methionine [57230-01-0]



2f

A 10 mL stainless steel (SS) jar was loaded with a 10 mm SS ball, H-*L*-Met-OH (122.0 mg, 0.82 mmol, 1 equiv.), $NaHCO_3$ (143.2 mg, 1.70 mmol, 2.08 equiv.) and chloroacetyl chloride (148 µL, 1.86 mmol, 2.27 equiv.). The mixture was subjected to grinding at 30 Hz for 15 min. Then, the mixture was recovered from ethyl acetate and water. The water phase was acidified with a 1M HCl

aqueous solution and extracted twice with ethyl acetate. The combined organic phases were washed with brine, dried over Na_2SO_4 and concentrated under reduced pressure. The oily crude product was finally triturated in Et_2O at 0°C, filtered and dried under vacuum to yield compound **2f** as a white solid (128.4 mg, 74% yield).

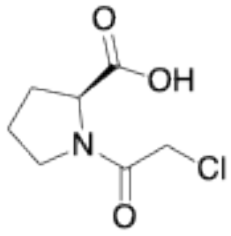
^{1}H NMR (500 MHz, DMSO-*d*$_6$) δ 12.83 (br. s, 1H), 8.55 (d, J = 7.8 Hz, 1H), 4.37 – 4.30 (m, 1H), 4.13 (d, J = 16.1 Hz, 1H), 4.11 (d, J = 16.1 Hz, 1H), 2.54 – 2.42 (m, 2H), 2.04 (s, 3H), 2.03 – 1.95 (m, 1H), 1.92 – 1.84 (m, 1H)

^{13}C NMR (126 MHz, DMSO-*d*$_6$) δ 172.9, 166.2, 51.2, 42.4, 30.5, 39.6, 14.6

HRMS (ESI): $[M+H]^+$ Calculated for $C_7H_{13}ClNO_3S$: 226.0299, found 226.0299

m.p.: 101 – 102.7°C

(2-chloroacetyl)-*L*-proline [23500-10-9]



2g

A 10 mL stainless steel (SS) jar was loaded with a 10 mm SS ball, H-*L*-Pro-OH (25.2 mg, 0.20 mmol, 1 equiv.), $NaHCO_3$ (104.2 mg, 1.24 mmol, 5.67 equiv.) and chloroacetyl chloride (43 µL, 0.54 mmol, 2.48 equiv.). The mixture was subjected to grinding at 30 Hz for 10 min. Then, the mixture was recovered from ethyl acetate and water. The water phase was acidified with a 1M HCl aqueous solution and extracted twice with ethyl acetate. The combined organic phases were washed with brine, dried over Na_2SO_4 and concentrated under reduced pressure. The oily crude product was finally triturated in Et_2O at 0°C, filtered and dried under vacuum to yield compound **2g** as a white solid (20 mg, 52% yield).

Two rotamers with a ratio 8.5/1.5

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.53 (br. s, 1H), 4.61 – 4.57 and 4.57 – 4.51 (m, 1H), 4.12 and 4.04 (d, J = 17.2 Hz, 1H), 4.09 and 4.01 (d, J = 17.2 Hz, 1H), 3.75 – 3.54 (m, 2H), 2.39 – 1.86 (m, 4H)

^{13}C NMR (126 MHz, $CDCl_3$) δ 175.2, 174.8, 166.5, 166.2, 59.7, 59.4, 47.5, 47.3, 41.9, 41.8, 31.4, 28.8, 24.9, 22.4

HRMS (ESI): $[M+H]^+$ Calculated for $C_7H_{11}ClNO_3$: 192.0422, found 192.0422

m.p.: 122.3 – 123.4°C

Data in agreement with literature.[2, 3]

^{1}H and ^{13}C NMR spectra

N-(2-chloroacetyl)-*L*-Leucine [688-12-0]

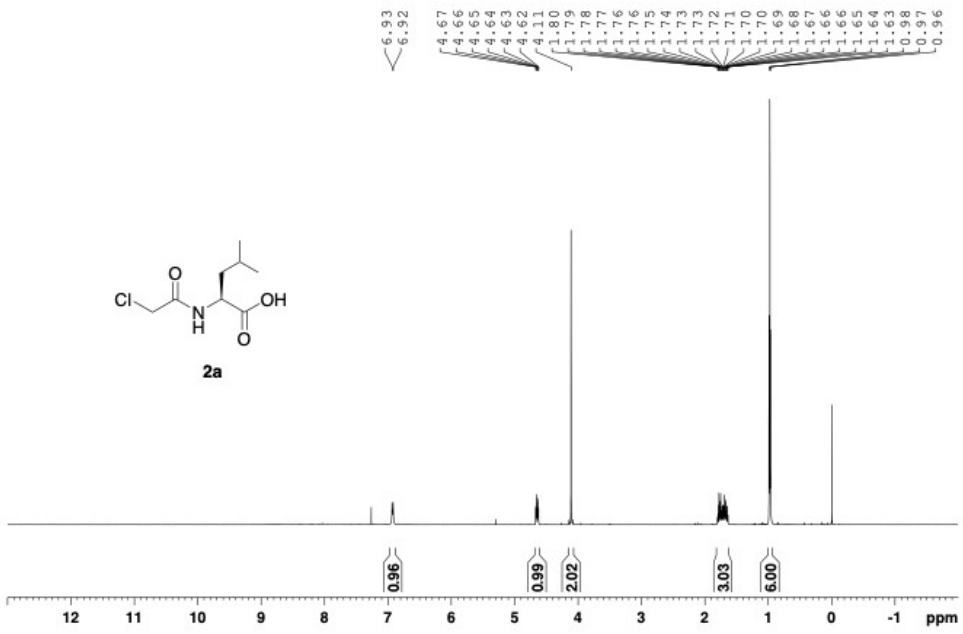


*Figure S5 - ^{1}H NMR spectrum (500MHz, $CDCl_3$) of **2a***

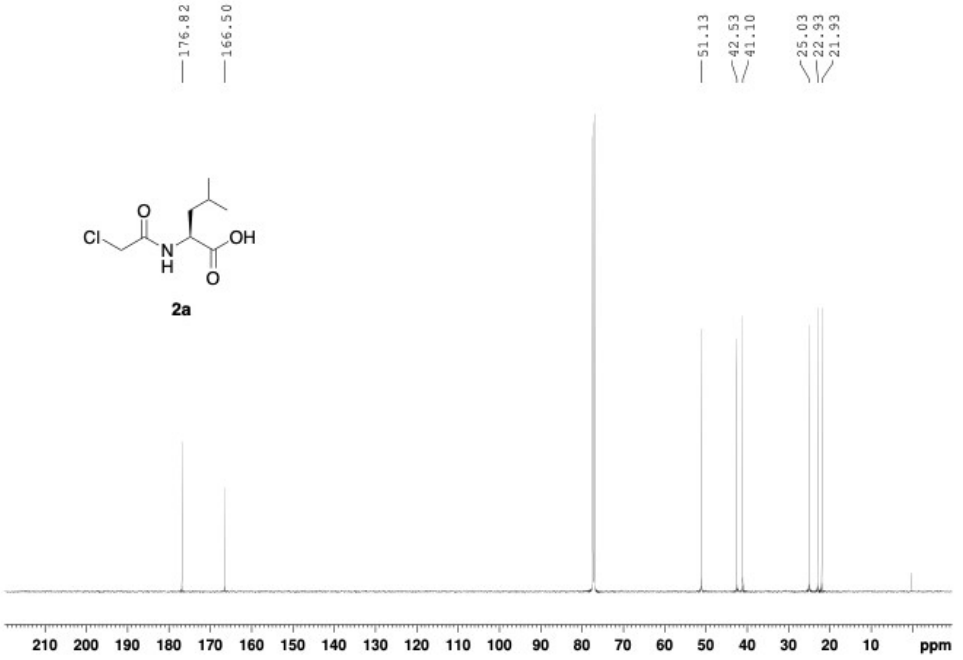


*Figure S6 - ^{13}C NMR spectrum (126MHz, $CDCl_3$) of **2a***

N-(2-chloroacetyl)-*L*-phenylalanine [721-65-3]

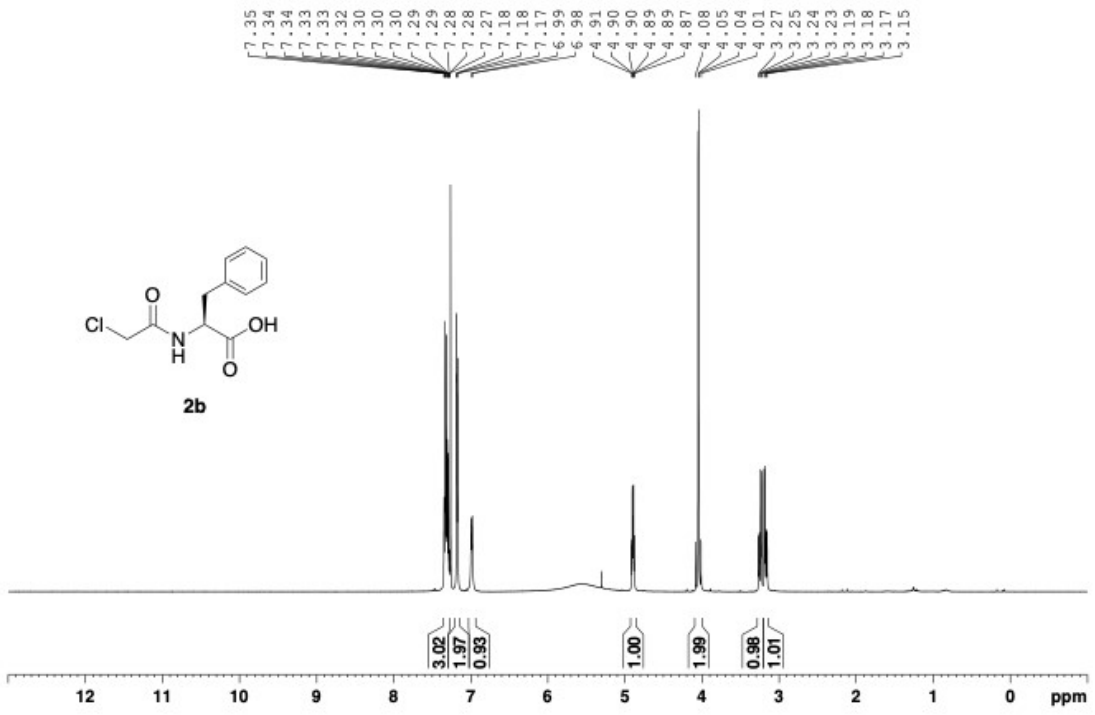


*Figure S7 - ^{1}H NMR spectrum (500MHz, $CDCl_3$) of **2b***

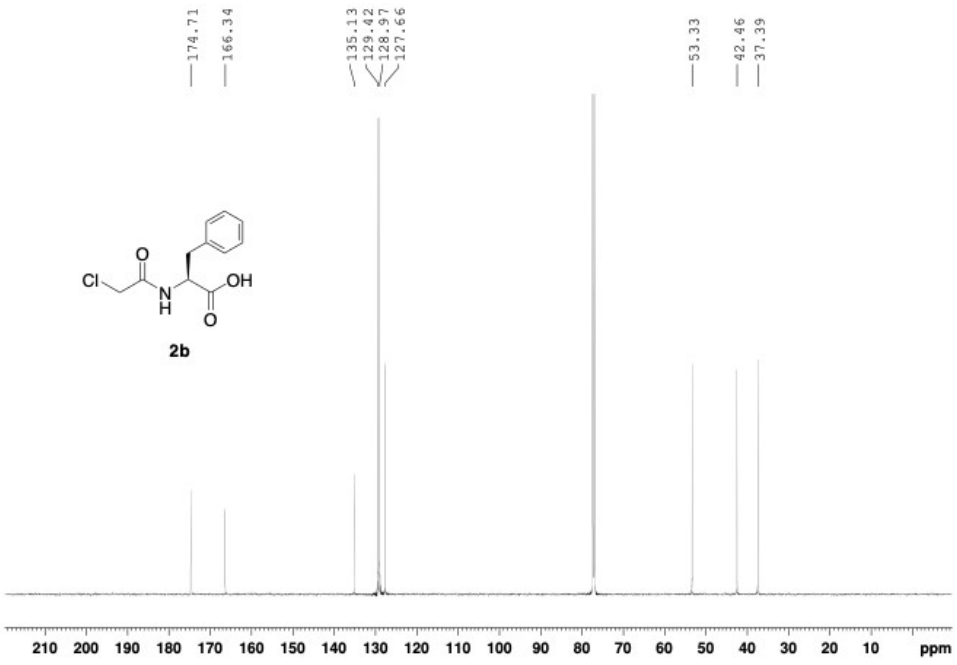


*Figure S8 - ^{13}C NMR spectrum (126MHz, $CDCl_3$) of **2b***

O-benzyl-*N*-(2-chloroacetyl)-*L*-serine [3062-02-0]

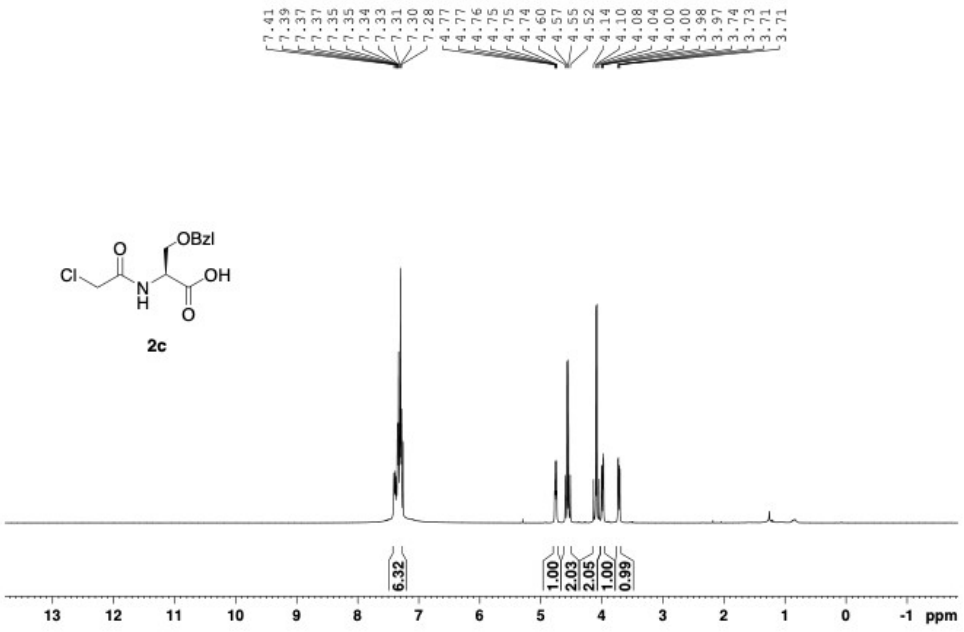


*Figure S9 - ^{1}H NMR spectrum (500MHz, $CDCl_3$) of **2c***

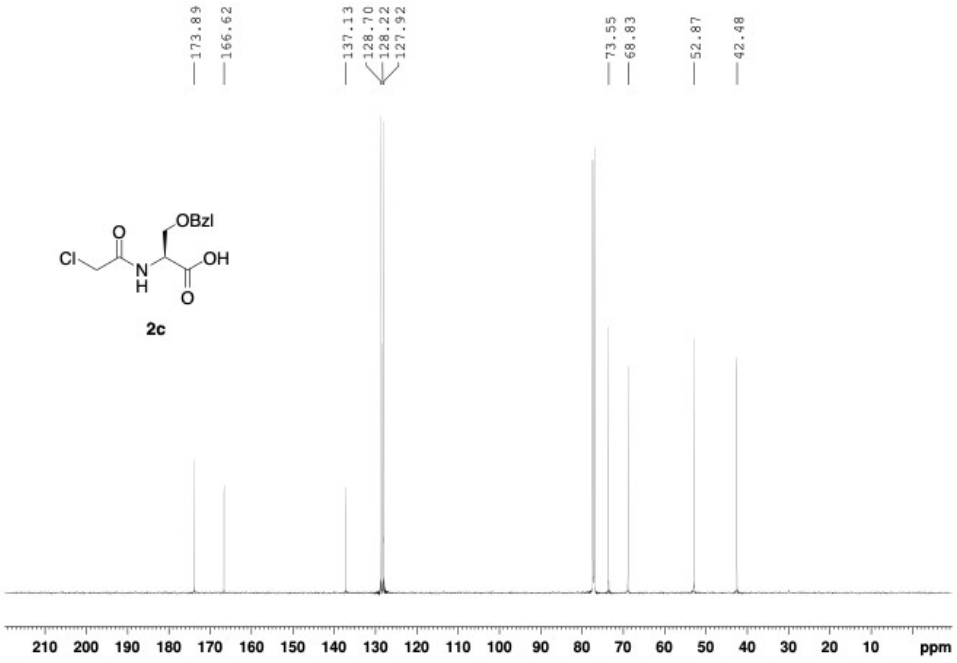


*Figure S10 - ^{13}C NMR spectrum (126MHz, $CDCl_3$) of **2c***

N^6-((benzyloxy)carbonyl)-N^2-(2-chloroacetyl)-*L*-lysine [47376-73-8]

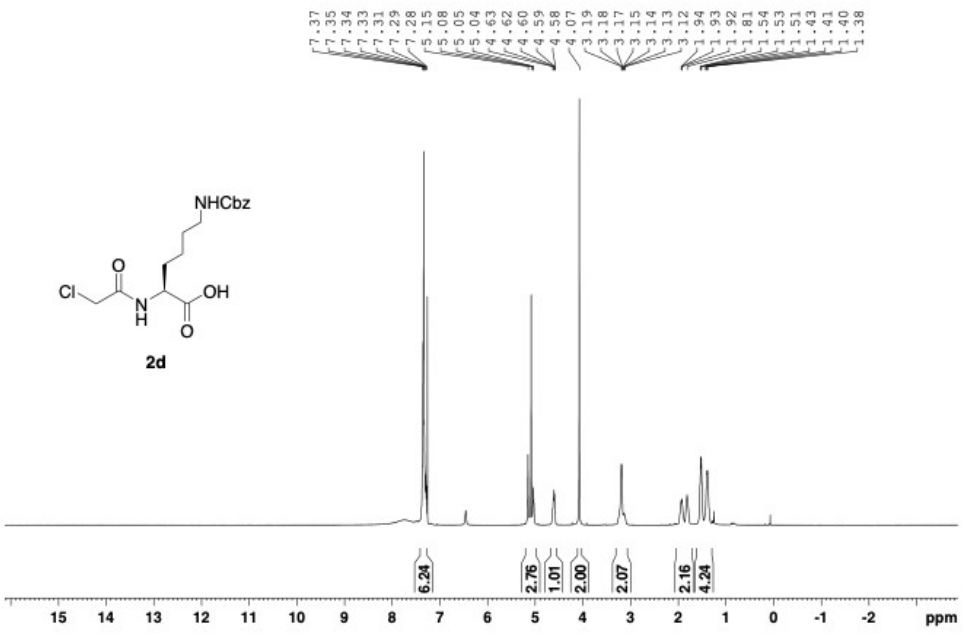


*Figure S11 - ^{1}H NMR spectrum (500MHz, $CDCl_3$) of **2d***

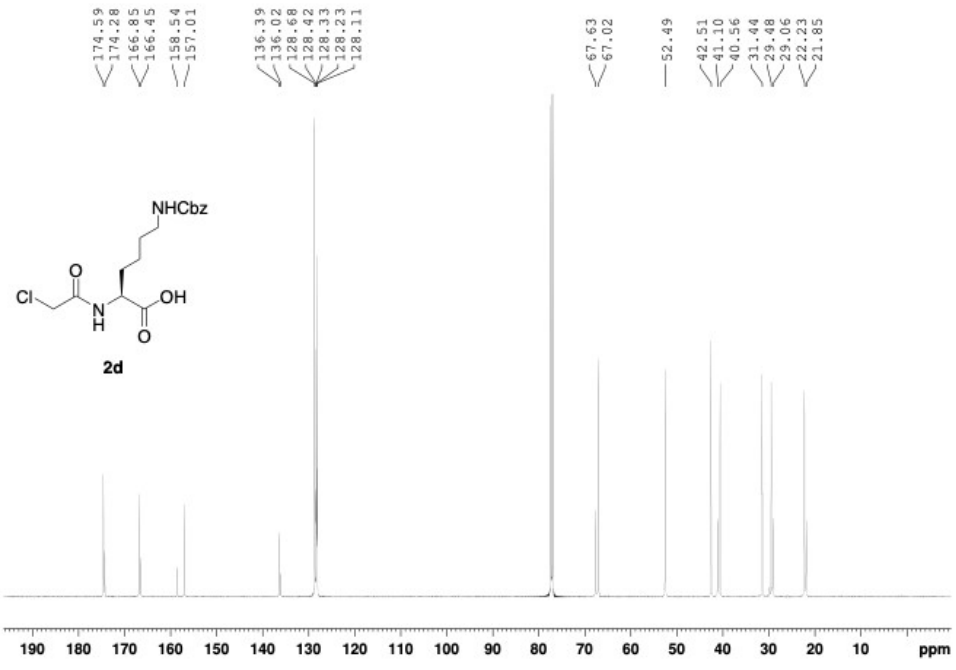


*Figure S12 - ^{13}C NMR spectrum (126MHz, $CDCl_3$) of **2d***

(S)-3-(4-(tert-butoxy)phenyl)-2-(2-chloroacetamido)propanoic acid

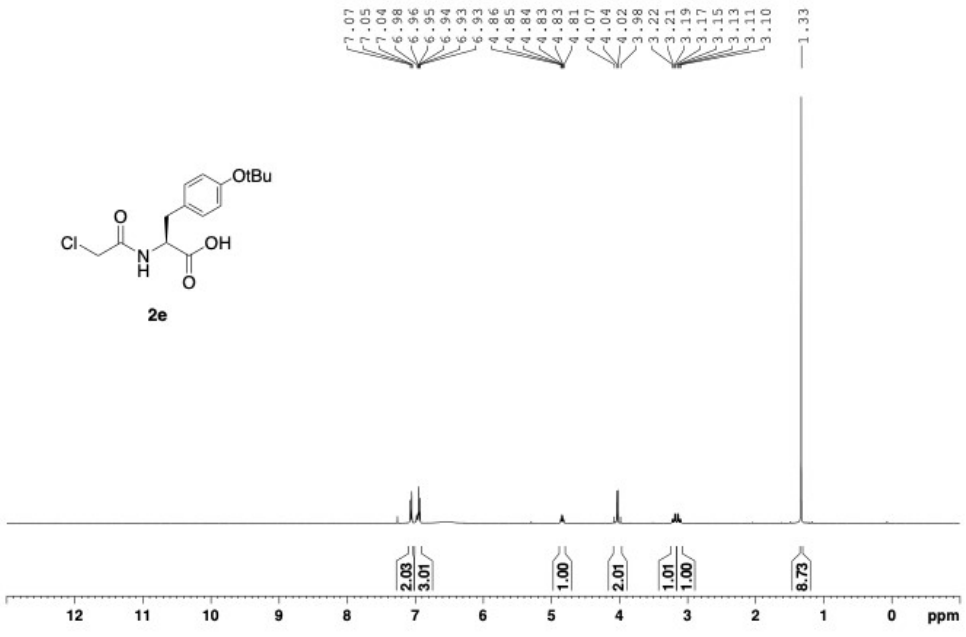


*Figure S13 - ^{1}H NMR spectrum (400MHz, $CDCl_3$) of **2e***

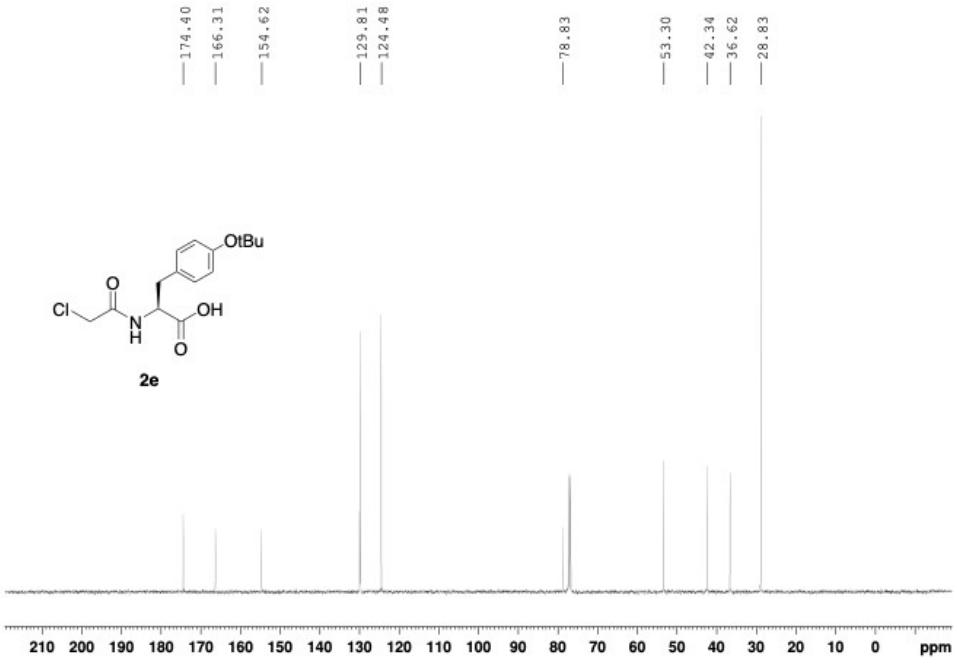


*Figure S14 - ^{13}C NMR spectrum (101MHz, $CDCl_3$) of **2e***

(2-chloroacetyl)-*L*-methionine [57230-01-0]

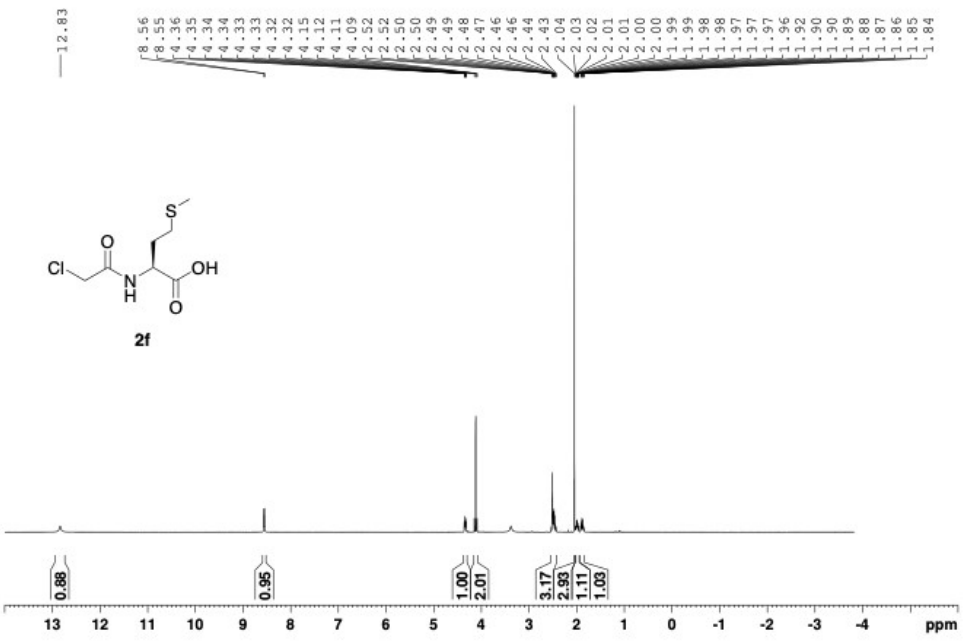


*Figure S15 - [1]H NMR spectrum (500MHz, DMSO-d_6) of **2f***

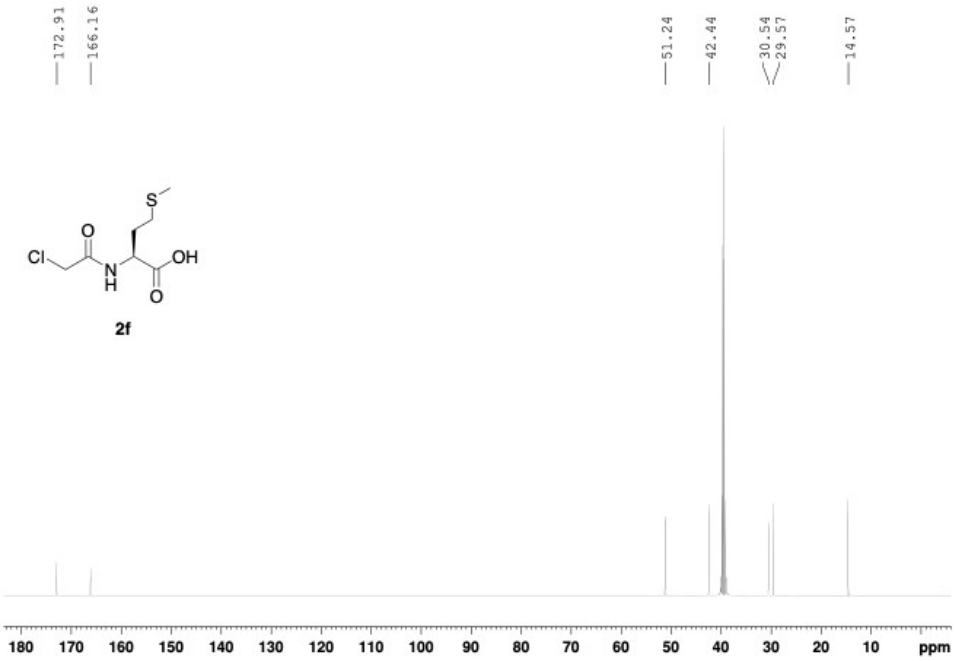


*Figure S16 - [13]C NMR spectrum (126MHz, DMSO-d_6) of **2f***

(2-chloroacetyl)-*L*-proline [23500-10-9]

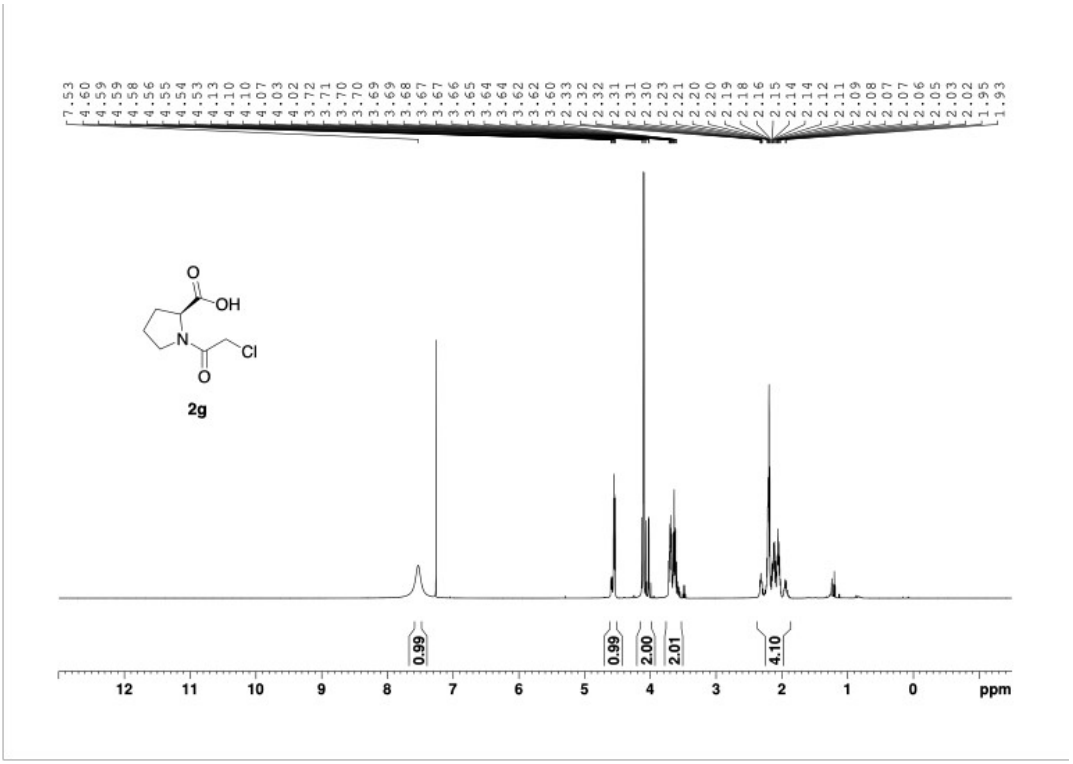


*Figure S17 - ^{1}H NMR spectrum (500MHz, $CDCl_3$) of **2g***

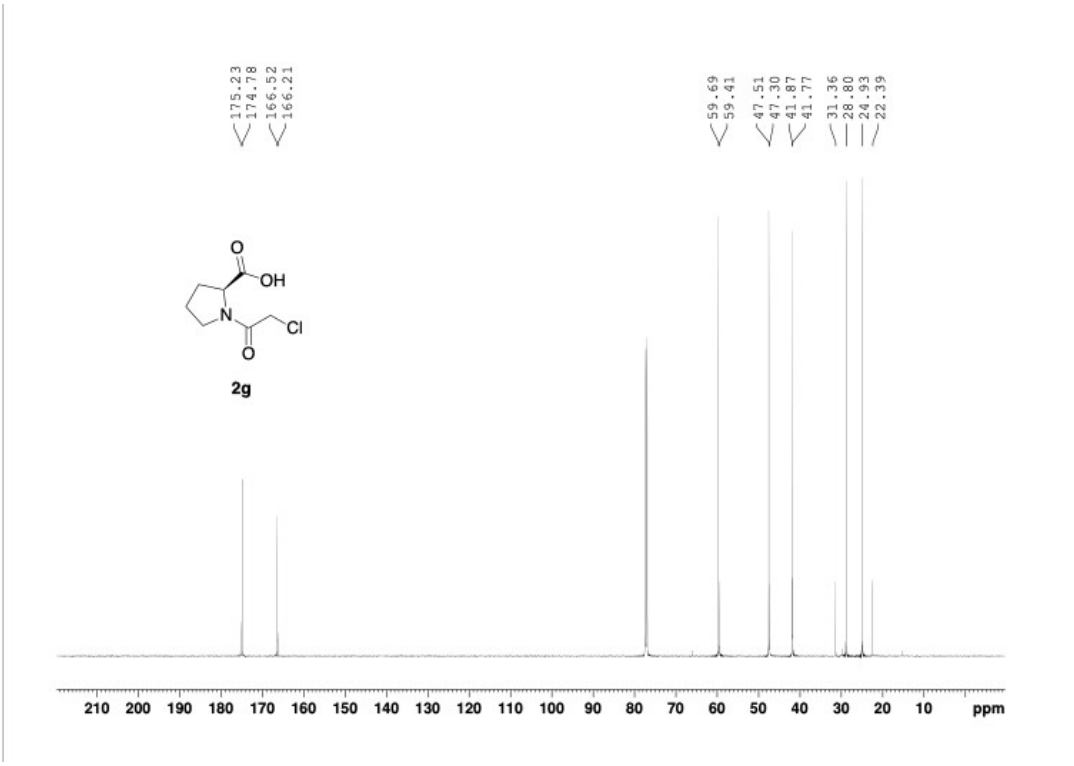


*Figure S18 - ^{13}C NMR spectrum (126MHz, $CDCl_3$) of **2g***

References

1. L. K. Beagle, F. K. Hansen, J.-C. M. Monbaliu, M. P. DesRosiers, A. M. Phillips, C. V. Stevens and A. R. Katritzky, *Synlett*, 2012, **23**, 2337-2340.

2. S. K. Singh, N. Manne and M. Pal, *Beilstein J. Org. Chem.*, 2008, **4**, 20.

3. F. Le Vaillant, M. D. Wodrich and J. Waser, *Chem. Sci.*, 2017, **8**, 1790-1800.