The effect of pyridinecarboxaldehyde functionalisation on reactivity and N-terminal protein modification

Lydia J. Barber[a] and Christopher D. Spicer*[a]

[a]Department of Chemistry and York Biomedical Research Institute, University of York, Heslington, YO10 5DD, UK

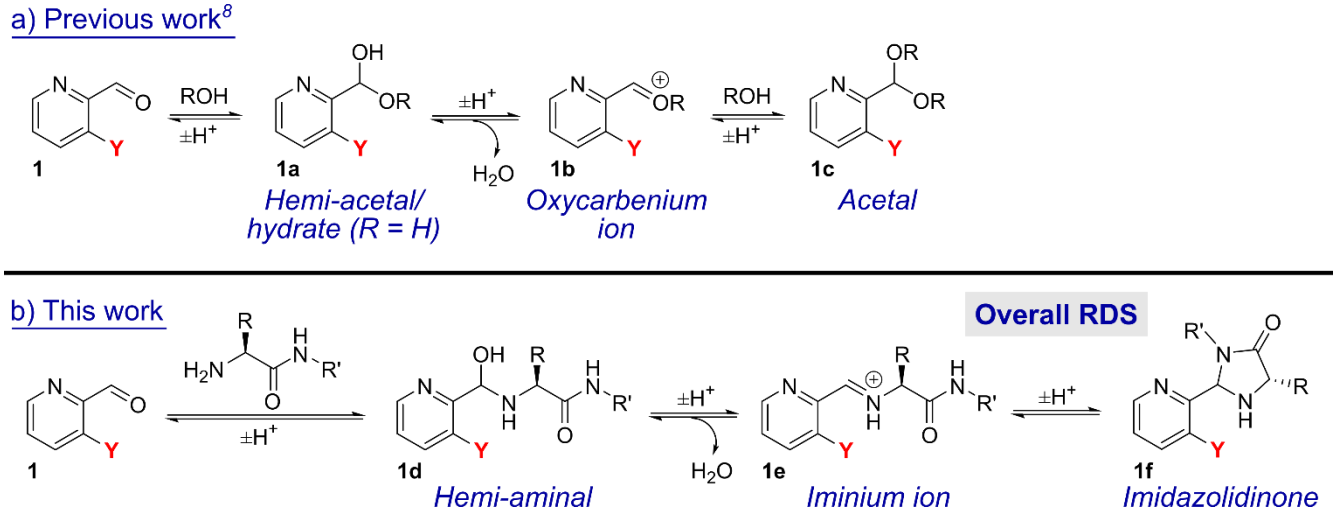
* chris.spicer@york.ac.uk

ABSTRACT: The site-selective modification of protein N-termini represents a powerful strategy for producing homogeneous bioconjugates. 2-Pyridinecarboxaldehydes have emerged as a leading reagent class in this area, but conjugation suffers from relatively slow rates and a degree of reversibility. In this work, we therefore studied the effects of pyridinecarboxaldehyde functionalisation on N-terminal modification, providing insight into the factors governing relative contributions from competing reaction pathways and design criteria for second generation reagents for protein labelling. Importantly, this insight allowed us to identify several candidate reagents which provide both accelerated and more stable protein labelling, enabling further applications of this powerful technology.

Introduction

The N-termini of proteins offer unique handles for site-selective protein modification. In eukaryotes a large proportion of the proteome is post-translationally modified at the N-terminus, but in most bacterial and secreted proteins, including antibodies, the α-amine is chemically and sterically accessible. In recent years this has led to a surge in interest in technologies targeting this α-amine for modification, with applications ranging from N-terminal proteomics,[1,2] to the development of protein-based therapeutics[3,4].

Foremost amongst these technologies has been the development of 2-pyridinecarboxaldehydes (e.g. **1**), as first reported by MacDonald *et al* in 2015.[5] 2-PCAs first form intermediate imines with N-terminal α-amines, which then subsequently cyclise with the adjacent primary amide of the protein backbone to form imidazolidinones (**Scheme 1**). These reactions are complicated by the ability of PCAs to form hydrates in the aqueous media used for protein modification. The nature of PCA functionalisation can significantly alter both the rate and reversibility of each of these steps.[6] However, in recent work we have demonstrated that even improved 2-PCA derivatives undergo significant levels of dissociative cleavage in the presence of competitive peptides, limiting applications where long-term stability is required.[7] Moreover, imidazolidinones form with relatively slow kinetics, necessitating long reaction times and high reagent loadings that may be detrimental for some protein targets, with cyclisation of a protonated iminium ion the rate-determining step.



Scheme 1: Reactions studied in this work, governing modification of protein N-termini with 2-pyridinecarboxaldehydes (2-PCAs) to form imidazolidinone conjugates.

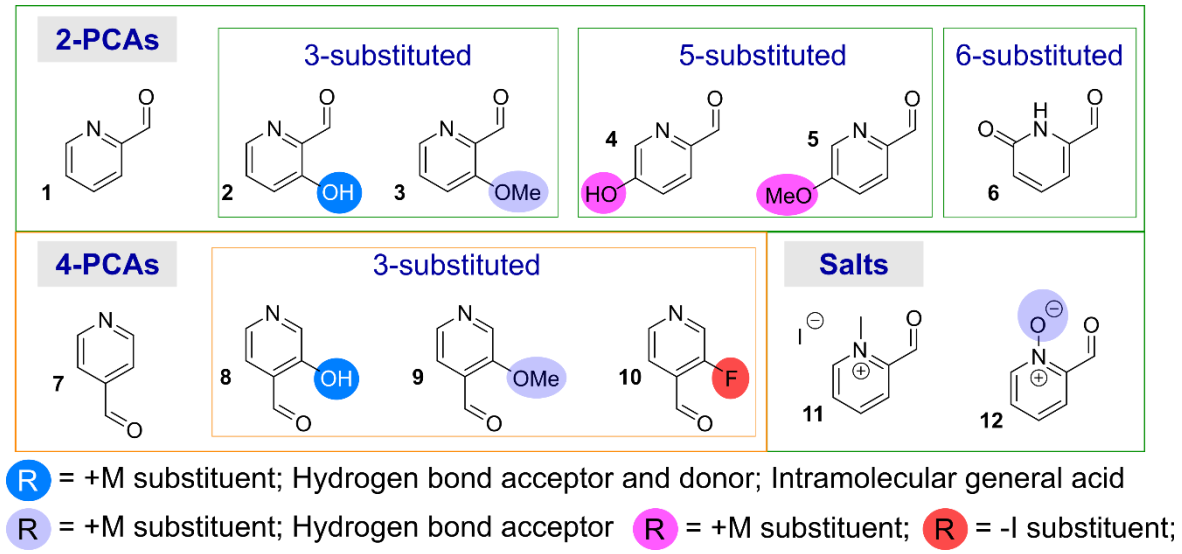


Figure 2: PCA substrates **1**-**12** used in this study.

In this work, we therefore set out to better understand the factors governing the complex, multi-step equilibria that ultimately lead to N-terminal modification, with a view to designing improved PCA reagents for protein labelling.

Results and discussion

Reagent design

We identified 12 PCA-based reagents that would provide diverse substrates for studying N-terminal modification (**Figure 1**), partially inspired by the work of Barman *et al* who previously studied the effects of PCA substitution on hydration and acetal formation (**Scheme 1a**).[8] Their work highlighted the complex interplay between various factors, including the activation of the ring by electronic contributions, intramolecular acid-base catalysis/hydrogen bonding effects, and steric factors, even within the relatively simple confines of a single reversible attack of water on the electrophilic aldehyde. We anticipated that these factors would have an amplified effect when hydrate formation was coupled to imine formation and subsequent cyclisation to form an imidazolidinone (**Scheme 1b**). The substrates could be broadly separated into 2-PCAs (**1**-**6**), 4-PCAs (**7**-**10**), and 2-PCA salts (**11**-**12**) (**Figure 1**). In an aqueous environment, **6** is expected to be found predominantly as its 2-pyridone tautomer, rather than the hydroxypyridine form, as shown.[9]

We chose to study both 2- and 4-PCAs to allow the contributions from electronic effects within the pyridine ring to be distinguished from the potential roles of the nitrogen as a general base or hydrogen-bond acceptor. Notably, in their initial report on N-terminal protein modification MacDonald *et al* found 2-PCA **1** to be more efficient for the modification of angiotensin than 4-PCA **7** (84% vs 28% conversion).[5] However, we have found the relative reactivities of different PCA derivatives to be protein dependent, and so further investigation was warranted.[7]

Hydroxyl functionalisation of PCAs has been previously shown to accelerate imine formation in organic solvents,[10] while Barman *et al* showed that hydrate formation at neutral pH is reduced[8]. Depending on the substitution pattern, we expected hydroxyl groups to have the potential to serve as: (a) resonance contributors to aldehyde electrophilicity (*ortho*- and *para*-hydroxy, **2**, **4**, **8**); (b) hydrogen bond acceptors or donors (*ortho*-hydroxy only, **2** and **8**); and/or (c) general acid catalysts. In contrast, corresponding methoxy derivatives (**3** and **9**) would serve only as hydrogen bond acceptors, while also making a weaker resonance contribution to the pyridine ring and enhancing steric congestion. Similarly, we expected fluoro substitution (**10**) to inductively activate the aldehyde to nucleophilic attack. Methylated pyridinium **11** would be expected to have similar properties, while negating the basicity and H-bonding properties of the nitrogen, while pyridine N-oxide **12** would be similarly electrophilic but would retain hydrogen bond acceptor potential. As a distinct analogue, the pyridone nitrogen of **6** would reverse the hydrogen bonding capabilities of the nitrogen, from acceptor to donor, while also negating any role played by basicity in dictating reaction outcome. However, the possibility of reactions shifting the pyridine-pyridone equilibrium in favour of the pyridine tautomer could not be discounted.

Hydrate formation

We first set out to study the degree of hydration for each reagent under conditions relevant to N-terminal protein modification (**1**→**1a**). Though these reactions are tolerant of a range of conditions, most commonly they are performed at near neutral pH at 37 °C in a phosphate or similar buffer. Under these conditions,

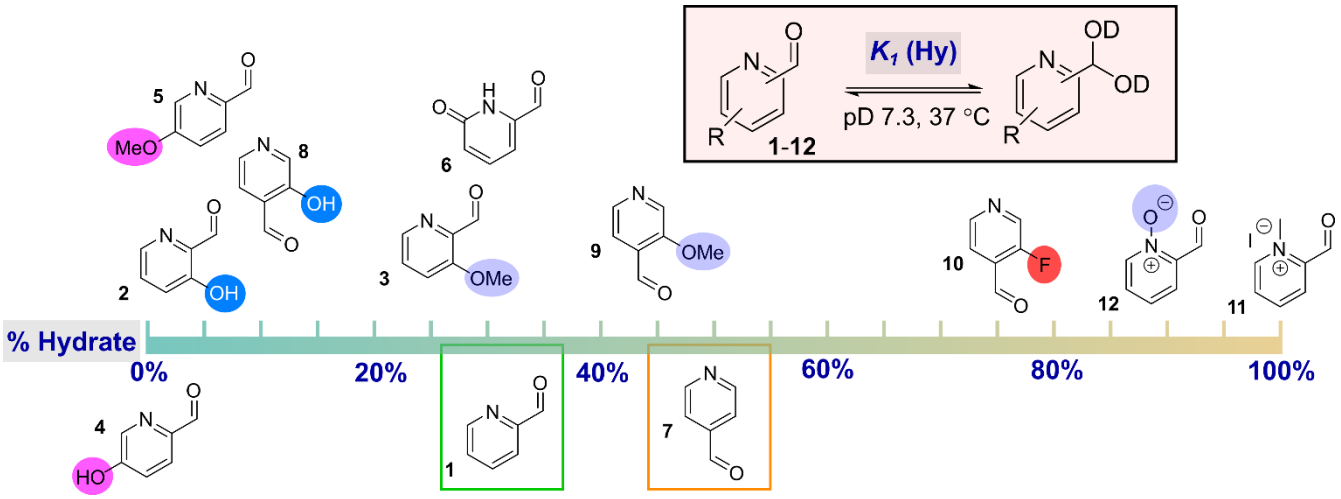


Figure 2: Percentage hydrate formation at 37 °C in pD 7.3 phosphate buffer for PCAs **1**-**12**.

the pyridine will be deprotonated (pyridinium pK_a ~3-5)[11] while the hydroxyls may be partially deprotonated, as discussed further below. It is important to note that this buffering distinguishes these experiments from those previously performed by Barman, whereby hydrate formation was studied in pure water at 25 °C.[8] Each reagent was incubated at a concentration of 50 mM in a deuterated sodium phosphate buffer (pD 7.3) at 37 °C, allowing the equilibration of hydrate formation, which proceeds on a fast time-scale relative to N-terminal imidazolidinone formation.[7,8] Ratios of aldehyde to hydrate were calculated by integration of characteristic peaks in ^{1}H NMR spectra, as detailed in the supporting information.

The degree of hydration ranged from 1% to >99% depending on the substrate (**Figure 2**). The distribution of reagents was largely consistent with the predicted electrophilicity of the aldehyde, with fluorinated (**10**), N-oxide (**12**), and N-methylated (**12**) PCAs exhibiting increased hydrate formation relative to the parent compounds **1** and **4**. In contrast, electron donating substituents decreased hydrate formation in all cases. This effect was far more pronounced for *ortho*-hydroxy PCAs (**2** and **8**, 28-37% decrease) than the *ortho*-methoxy analogues (**3** and **9**, 4-7% decrease). This result is consistent with the observations of Barman *et al*, where differences were found to be more significant than would be predicted based solely on differences between Hammet sigma values of a hydroxy or methoxy substituent. They postulated that contributions from hydroxyl anions resulting from partial deprotonation may have amplified deactivation of the carbonyl. We therefore calculated pK_a values for the hydroxyl groups of **2** (6.5), **4** (6.1), and **8** (6.9), which indicate that these compounds will be significantly deprotonated under the experimental conditions, and the effects of the anion dominate as predicted. Indeed, the degree of hydrate formation between the three analogues was consistent with the small differences in the hydroxyl pK_as. However, the aldehyde of *para*-methoxy analogue **5** was similarly deactivated, potentially indicating steric factors may also be contributing to the surprisingly small decrease in hydrate formation seen for the *ortho*-methoxy derivatives.

When comparing 2- and 4-substituted PCAs, 4-substituted analogues exhibited higher levels of hydrate formation in all cases (7-19% increase), again consistent with the observations of Barman *et al*.[8] Interestingly, pyridone **6** also exhibited comparable behaviour to 2-PCA **1**.

Imine formation

We next looked to investigate the formation of the key imine intermediate on the pathway to N-terminal labelling. Formation of this imine goes through the hemi-aminal intermediate **1d**, which subsequently undergoes dehydration to form iminium ion **1e**. Hemi-aminal formation is competitive with hydrate formation, and expected to be promoted by electron-withdrawing substituents in the same manner, while electron-donating substituents would be expected to better stabilise the iminium species. The pK_a of this iminium is strongly influenced by the substitution of the pyridine. Crugeiras *et al* previously reported that an iminium formed from 4-PCA **7** had a pK_a ~ 5, while the *ortho*-hydroxy analogue **8** had a pK_a ~ 9, and so at neutral pH the relative protonation degree of the iminium, and thus propensity to undergo either hydrolysis or imidazolidinone formation, is likely to differ greatly.[12] The authors demonstrated that the significant increase in iminium pK_a for **8** was due to a combination of resonance activation of the imine by the *ortho*-hydroxyl substituent, and intramolecular hydrogen bonding, in an analogous fashion to the contribution of hydrogen bonding to the stabilisation of the oxocarbenium intermediate necessary for acetal formation reported by Barman *et al*.[8]

The α-amine of protein N-termini is less basic than the ε-amine of lysine sidechains (ammonium pK_a of

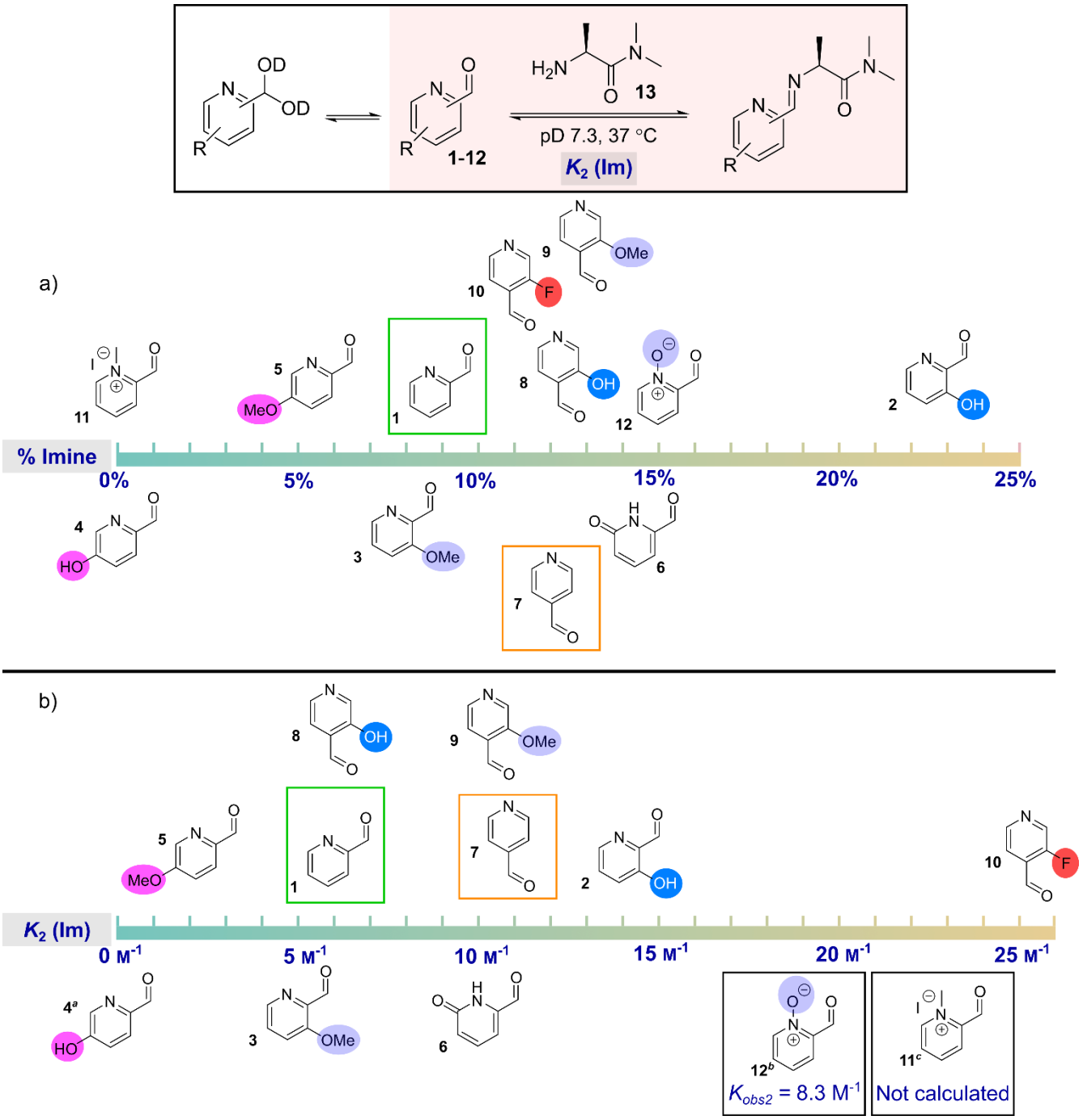


Figure 3: Reaction of PCAs **1**-**12** (25 mM) with alanine dimethylamide **13** (25 mM) at 37 °C in pD 7.3 buffer. a) Percentage imine formation; b) Calculated K_2 (Im) values for imine formation; [a]No imine was observed for **4** so K_2 (Im) could not be calculated but approaches 0 M^{-1}; [b]No aldehyde was observed for **12** and so K_{obs2} for hydrate-imine equilibrium is given; [c]No aldehyde or imine was observed for **11**, and so K_2 could not be calculated.

6-8, vs 10 for ε-amines) due to the electron withdrawing effects of the proximal amide, influencing protonation state and thus reactivity at neutral pH. To best simulate this pK_a, while also preventing imidazolidinone formation, we chose to use alanine dimethyl amide **13** as a reaction partner. While the use of a secondary, rather than primary, amide will influence both imine formation and protonation, we expected the impact to be small when compared to the use of the corresponding carboxylic acid as a model substrate.

PCAs **1**-**12** were mixed at a 1:1 ratio with alanine dimethyl amide **13** at a concentration of 25 mM in deuterated sodium phosphate buffer (pD 7.3) and incubated at 37 °C to allow equilibration. Integrations of characteristic aldehyde, hydrate, and imine peaks in the ^{1}H spectra were again used to calculate the ratios of the three species (**Figure 3a**, see Supporting Information for further details). Notably, we did not observe any signals originating from a hemi-aminal species in any experiment.

From the data obtained, the equilibrium constant, K_2, for imine formation could be calculated (see SI for further details, **Figure 3b**). Considering first the electron deficient derivatives, *ortho*-fluoro PCA **10** displayed the highest K_2 of all derivatives tested, driven by favourable amine addition. For pyridinium oxide **12**, high levels of imine formation were also observed, to the extent that no aldehyde was observed in the experiment. As a result, K_2 could not be calculated, and the observed equilibrium constant, K_{obs2}, between hydrate and imine was instead calculated (but cannot be directly compared to K_2 calculated for the other derivatives). In contrast, no imine was observed for methylated pyridinium **11**, in this case because of the high levels of competitive hydration which depleted aldehyde availability for imine formation.

Within the electron-rich derivatives, *para*-hydroxy PCA **4** was notable in failing to generate detectable imine, dictated by the reduced electrophilicity of the aldehyde and less favourable hemi-aminal formation. In contrast, *ortho*-hydroxy 2-PCA **2** had the highest levels of imine formation (22%) amongst all the derivatives tested. This contrasting behaviour between the *ortho*- and *para*-hydroxy isomers indicates the hydroxy group of **2** is able to act as a hydrogen bond acceptor and stabilise the iminium ion formed. For the *ortho*-methoxy derivatives **3** and **9**, similar K_2 values were obtained to the parent PCAs. These results are analogous to the results obtained by Barman *et al* for acetal formation, and indicate a delicate balance between electronic contributions slowing amine attack, but subsequent dehydration being favoured by the hydrogen bonding capacity of these groups, ultimately dictating the overall equilibrium. However, in contrast to the results found for acetal formation, we surprisingly found that *ortho*-hydroxy 4-PCA **8** exhibited a lower K_2 than the parent 4-PCA **7**.

Imidazolidinone formation

Having studied hydrate and imine formation and determined equilibrium constants for each step, we were now in a position to analyse the factors governing imidazolidinone formation. In an analogous experiment to those described above, each PCA derivative was incubated with 1 equiv. of the model dipeptide Ala-Ala at 37 °C. In our previous study, imidazolidinone formation from 2-PCAs was found to be relatively slow, and ^{1}H NMR spectra were therefore recorded at regular intervals over a 16 h time period.[7] Integrations of peaks originating from the two imidazolidinone diastereomers formed were tracked over time, relative to aldehyde, imine, and hydrate peaks. The data obtained allowed us to analyse imidazolidinone formation at two levels, each providing complementary insights into the reaction.

A first analysis considered solely the cyclisation of the imine to form the imidazolidinone product. Since cyclisation is rate-limiting and several orders of magnitude slower than imine and hydrate formation, K_1 and K_2 could be used to build and fit a kinetic model to the NMR data, to obtain both forward and reverse first-order rate constants for cyclisation, k_3 and k_{-3} respectively, as well as the equilibrium constant for cyclisation, K_3, where relevant (see Supporting Information for details, **Figure 4**). For PCAs **4** and **11**, imidazolidinone formation was observed despite the lack of imine formation seen in the previous experiment. In this scenario, a steady-state approximation was applied to calculate the observed rate constants, k_{obs4} and k_{obs5} (second-order), and k_{obs-4} and k_{obs-5} (first-order), for reversible formation of the imidazolidinone products from the aldehydes or hydrates respectively. However, these constants could not be directly compared to k_3 and k_{-3}, and so interpretation of the reactivity of these analogues could not be undertaken during this first analysis.

On the other hand, we could consider the overall picture of total imidazolidinone formation as a function of time, encompassing the overall complex balance between hydrate, imine, and imidazolidinone formation (see Supporting Information Section 2). This analysis allowed a more qualitative comparison of the data to be performed, including integrating the behaviour of **4** and **11**.

In all cases, electron donating substituents accelerated imidazolidinone formation. Since electron-donation is expected to decrease imine electrophilicity, this effect is most likely due to the increased pK_a of the iminium, and thus higher degree of protonation relative to the parent PCA, which would in turn favour nucleophilic attack, as described above.[12] This effect would be expected to be highest for *ortho*-substituted analogues where hydrogen bonding would serve to increase the pK_a further, and the lower k_3 of *para*-methoxy PCA **5** would appear to support this. Though a comparable k_3 value for **4** could not be obtained, overall conversion to the imidazolidinone product was found to be slow (~20% after 16 h), as it was for **5**, lending further support to this hypothesis. An exception to this general behaviour was *ortho*-methoxy 4-PCA **9** which exhibited a surprisingly low k_3.

Considering the reverse reaction, the favourable acceleration of ring closing induced by electron donating substituents was partially tempered by an analogous increase in ring opening for hydroxy-substituted PCAs. Though K_3 was higher overall than for the parent PCAs, and total imidazolidinone formation was therefore both accelerated and increased, the net result of these factors was that incomplete cyclisation was achieved at equilibrium. In contrast, for the *ortho*-methoxy analogues **3** and **9**, the data fit an irreversible kinetic model indicating that k_{-3} was below the threshold of detection within

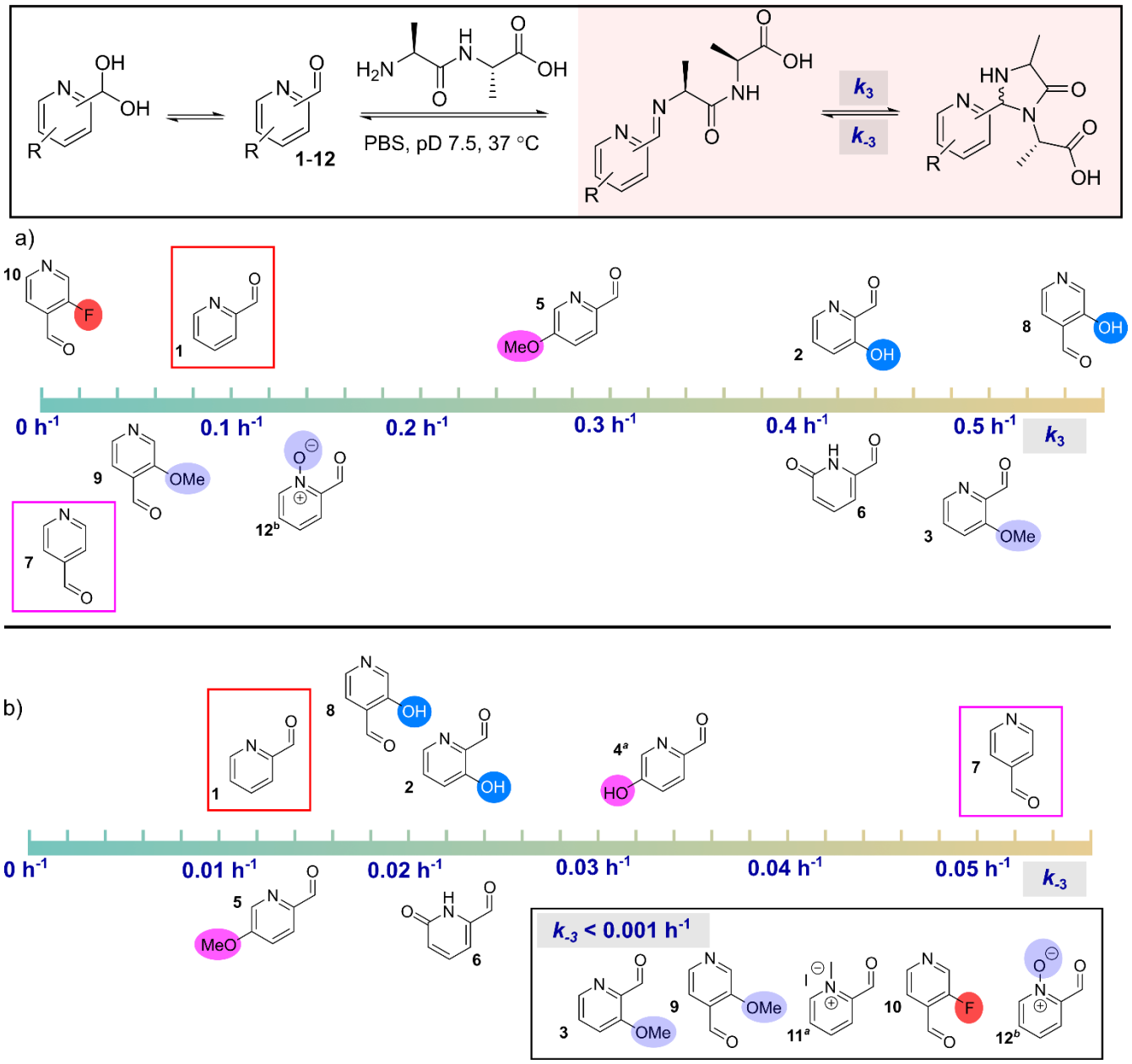


Figure 4: Reaction of PCAs **1**-**12** (50 mM) with di-alanine (50 mM) at 37 °C in pD 7.3 buffer. a) Calculated forward rate constants, k_3. Nb. For PCAs **4** and **11** no imine was shown in previous experiments and so no k_3 could be calculated. Steady-state approximations could be applied to give the second-order observed rate constants k_{obs4} = 0.44 M^{-1} h^{-1} and k_{obs5} = 0.18 M^{-1} h^{-1} respectively, as described in the SI; b) Calculated reverse rate constants, k_{-3}; [a]For **4** and **11** reverse reaction rates, k_{obs-4} and k_{obs-5} respectively, were based on calculations encompassing the steady state approximation detailed above and in the SI; [b]For **12** values of both k_3 and k_{-3} were derived based on K_{obs2} as described in Figure 2 and the SI.

these experiments (< 1 × 10^{-3} h^{-1}). The net effect of this on imidazolidinone formation was most significant for **3**, with equilibrium having not been reached by the end-point of the experiment.

A similar decrease in k_{-3} was observed for the electron deficient derivatives **10**-**12** for which k_{-3} could not be measured (k_{obs5} in the case of **11**). Fluoro PCA **10** exhibited a very slow rate of cyclisation, and overall accumulation of imidazolidinone was slow for both **10** and methyl pyridinium **11**, indicative of the lower degree of imine protonation reducing susceptibility to ring-closing. However, most interestingly pyridine N-oxide **12** exhibited a small increase in k_3 over the parent PCA **1**, while also fitting to an irreversible kinetic model. Collectively, these data highlighted the potential for a number of PCA derivatives, particularly **3** and **12**, to improve N-terminal protein modification.

Protein modification

Each of the PCA derivatives **1**-**12** was screened in the modification of RNase A and myoglobin, as model proteins that we previously showed to give

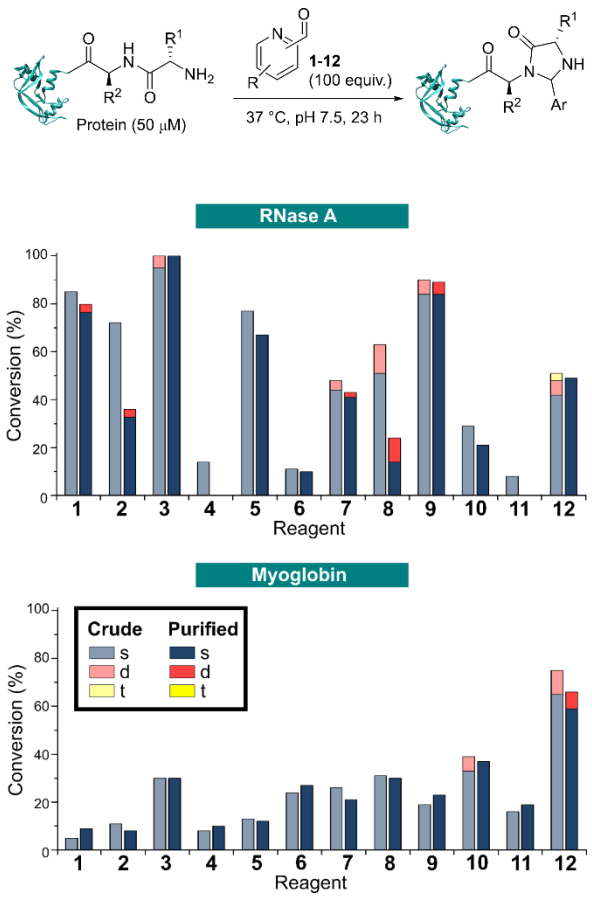


Figure 5: Conversions for the modification of RNase A and myoglobin with PCAs **1**-**12** before (Crude) and after (Purified) overnight dialysis at 4 °C; s = single, d = double, t = triple modification.

variable levels of labelling depending on the reagent used.[7] Conversions were determined on both crude reaction mixtures and after dialysis overnight to remove excess reagent. If imidazolidinone formation was unstable, this dialysis would be expected to lead to a drop in conversion, as well as the hydrolysis of any transient imines formed at lysine residues (**Figure 5**).

For RNase A, we were pleased to observe that methoxy-PCA **3** gave the highest levels of modification among the panel of reagents screened, and furthermore that modification was stable after dialysis. *Ortho*-methoxy 4-PCA **9** was also found to give relatively high conversions and conjugate stability, despite the slow rate of imidazolidinone formation observed during our small molecule studies. In contrast, the *ortho*-hydroxy analogues **2** and **8**, while giving high levels of initial modification, were found to be unstable to extended dialysis, as would be predicted from their increased k_{-3}. For fluoro-PCA **10** and methylated-PCA **11** conversions were low, likely due to their overall slow rate of imidazolidinone formation.

Moving on to myoglobin, our results highlighted the variability in labelling efficiency between different proteins. Most of the PCAs studied gave generally low conversions (~5-30%), including both the methoxy-PCAs (**3** and **9**) that had performed so well on RNase A. Most significantly though, labelling was considerably higher with pyridine N-oxide **12**, with ~60% conversion after purification, albeit with low levels (~7%) of off-target modification also being observed.

Conclusions

In this paper we have studied the effects of PCA functionalisation on the key reaction steps governing the modification of protein N-termini, specifically competing aldehyde hydration, iminium formation, and finally cyclisation of this intermediate iminium to form the key imidazolidinone species which delivers N-terminal selectivity. Our results indicate a complex and delicate balance exists between factors governing each individual step, and that a consideration of each is important to understand the relative performance of different reagent classes in N-terminal functionalisation.

In particular, we identify two promising reagent classes for protein modification, with contrasting features: on the one hand, electron rich ortho-methoxy substituted 2-PCAs, particularly 2-PCA **3**, can serve to enhance and accelerate imidazolidinone formation, which we hypothesise is due to an increase in imine protonation and thus electrophilicity relative to unsubstituted PCA analogues; on the other, electron deficient pyridinium oxide **12** delivered stable imidazolidinone conjugates, albeit at slower rates. The structural differences between these two reagent classes, and the differences in behaviour observed in subsequent protein labelling experiments, serves to further highlight the complexity of N-terminal modification, and the merit of mechanistic insight provided from small molecule models – in the absence of this insight, it is possible that promising reagents could be overlooked due to poor compatibility with a given target protein during screening. Further research at the interface of fundamental physical organic chemistry and protein biochemistry is therefore vital to continued progress in the bioconjugation field, and in enabling the potential of N-terminal selective labelling to be realised.

ASSOCIATED CONTENT

Supporting Information

The supporting information contains details of all experiments, including small molecule synthesis, NMR studies, and protein labelling.

AUTHOR INFORMATION

Author contributions

LJB performed all experiments. LJB and CDS developed the study, analysed the data, and wrote

the manuscript. CDS supervised and managed the study.

Competing interests

The authors declare no competing financial interests.

ACKNOWLEDGMENT

Dr Ksenia Stankevich is thanked for constructive discussion and feedback on the work. LJB and CDS acknowledge PhD studentship funding from the BBSRC White Rose Mechanistic Biology DTP. This PhD was supported with generous industrial CASE support from Qkine and our thanks go to Dr Marko Hyvönen and Dr Catherine Elton for their support. CDS acknowledges support through a Wellcome Trust Career Development Award (225257/Z/22/Z). We thank Dr Ed Bergstrom and The York Centre of Excellence in Mass Spectrometry for support with peptide mass spectrometry. This centre was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent support from the EPSRC (EP/K039660/1; EP/M028127/1).

Bibliography

1 A. R. Griswold, P. Cifani, S. D. Rao, A. J. Axelrod, M. M. Miele, R. C. Hendrickson, A. Kentsis and D. A. Bachovchin, *Cell Chem. Biol.*, 2019, **26**, 901-907.e6.
2 H. N. Bridge, W. Leiter, C. L. Frazier and A. M. Weeks, *Cell Chem. Biol.*, 2024, **31**, 534-549.e8.
3 R. Sangsuwan, P. Tachachartvanich and M. B. Francis, *J. Am. Chem. Soc.*, 2019, **141**, 2376–2383.
4 J. Sun, X. Liu, J. Guo, W. Zhao and W. Gao, *ACS Appl. Mater. Interfaces*, 2021, **13**, 88–96.
5 J. I. MacDonald, H. K. Munch, T. Moore and M. B. Francis, *Nat. Chem. Biol.*, 2015, **11**, 326–331.
6 B. Koo, N. S. Dolan, K. Wucherer, H. K. Munch and M. B. Francis, *Biomacromolecules*, 2019, **20**, 3933–3939.
7 L. J. Barber, N. D. J. Yates, M. A. Fascione, A. Parkin, G. R. Hemsworth, P. G. Genever and C. D. Spicer, *RSC Chem. Biol.*, 2023, **4**, 56–64.
8 S. Barman, K. L. Diehl and E. V. Anslyn, *RSC Adv.*, 2014, **4**, 28893–28900.
9 J. Frank and A. R. Katritzky, *J. Chem. Soc., Perkin Trans. 2*, 1976, 1428–1431.
10 J. M. Dragna, G. Pescitelli, L. Tran, V. M. Lynch, E. V. Anslyn and L. Di Bari, *J. Am. Chem. Soc.*, 2012, **134**, 4398–4407.
11 C. S. Handloser, M. R. Chakrabarty and M. W. Mosher, *J. Chem. Educ.*, 1973, **50**, 510–511.
12 J. Crugeiras, A. Rios, E. Riveiros and J. P. Richard, *J. Am. Chem. Soc.*, 2009, **131**, 15815–15824.