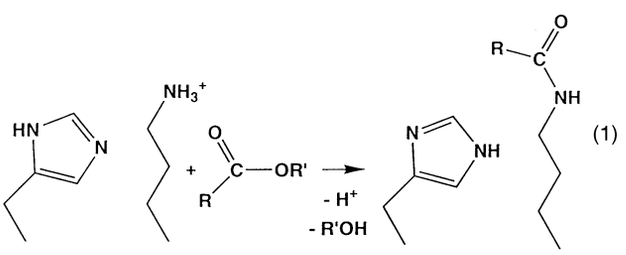
Site selectivity in self-catalysed functionalization of helical polypeptide structures

Klas Broo,[a] Malin Allert,[a] Linda Andersson, Pernilla Erlandsson,[a] Gunnar Stenhagen,[b] Joakim Wigström,[a] Per Ahlberg*,[a] and Lars Baltzer*,[a]

[a] *Department of Organic Chemistry, Göteborg University, 412 96 Göteborg, Sweden*
[b] *Department of Organic Chemistry, Chalmers University of Technology, 412 96 Göteborg, Sweden*

Histidine side chains in helical structures catalyse the acylation of flanking lysine, ornithine and 1,3-diaminobutyric acid residues provided they are in positions *i* − 3 and *i* + 4, but not in positions *i* − 4, *i* − 1, *i* + 2, *i* + 3, relative to a histidine in position *i*, in a novel site-selective functionalization reaction that enhances the potential of polypeptide and protein design.

We recently reported that a histidine side chain catalyses the site-selective acylation of a flanking ornithine residue in a helical structure [eqn. (1)] using mono-*p*-nitrophenyl fumarate.[1]



The mechanism was determined[2] and it was found that an initial rate-limiting attack of the deprotonated form of the histidine side chain to form an acyl intermediate was followed by a fast intramolecular transfer of the acyl group to form an amide at the side chain of the flanking Orn. The reaction was first discovered in a designed polypeptide with 42 amino acid residues, RA-42, that folds into a helix-loop-helix dimer.[1] To explore the generality of the reaction we have further investigated the selectivity and reactivity in other helix-loop-helix motifs with 42 amino acid residues and in model peptides with 20 residues that in solution form helical structures. All model peptides were derived from helix I in RA-42 (Fig. 1).

We now wish to report that the site selectivity in this reaction is high, that only two sites in the vicinity of the histidine are acylated and that the flanking residue may be Lys, Orn or 1,3-diaminobutyric acid (DAB). His-11 catalyses the acylation of Lys-15, Orn-15, DAB-15 and Lys-8, but not Lys-7, Lys-10, Lys-13 and Lys-14 (Fig. 1). Positions 9 and 12 were not investigated since they were part of the hydrophobic core of the amphiphilic helices.

So far, site selectivity has been explored by us in helices, although β sheets can also be expected to function as templates in design and functionalization. The helical content is conveniently assessed by CD spectroscopy[3] and the mean residue ellipticity of the model peptides under reaction conditions was between −9 000 and −20 000 deg cm^2 dmol^{-1}, which corresponds to 25–60% of helix.[4,5] Typical reaction conditions for model peptides with 20 amino acids and for helix-loop-helix motifs[1] were 0.5–1 mM concentration of peptide in 10 vol% 2,2,2-trifluoroethanol at 290 K and pH 5.85. The trifluoroethanol solution was needed to ensure helical conformation for the shorter peptides.[6] The mean residue ellipticity of the helix-loop-helix dimers had larger negative values than −19 000 deg cm^2 dmol^{-1} corresponding to more than 60% helix.

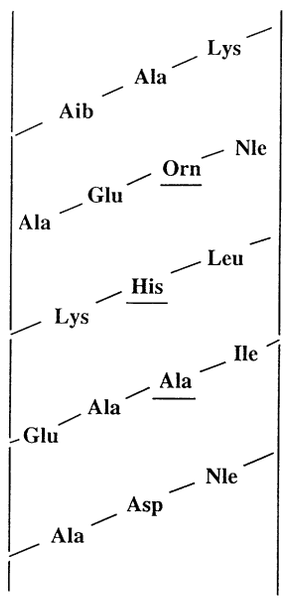


Fig. 1 Schematic representation of helical structure illustrating the geometric relationship between His-11 and positions that are acylated by mono-*p*-nitrophenyl fumarate. The amino acid sequence is that of helix I in RA-42 given in three-letter code. Aib is α-aminoisobutyric acid and Nle is norleucine. All residues that surround His-11 in space have been searched for possible acylation except residues that form part of the hydrophobic core. Only in positions 8 and 11 are DAB, Lys or Orn acylated. The residues shown are those from Ala-3 to Lys-19 and Ala-8 is underscored to indicate that a lysine in that position would have been acylated.

In all peptides His-11 was flanked by DAB, Lys or Orn residues in positions 7, 8, 10, 13, 14 or 15. The second-order rate constants were determined in helix-loop-helix motifs for some flanking residues. A histidine residue in a helix-loop-helix dimer that was not flanked by Arg, Lys, Orn or His was previously shown to catalyse the formation of *p*-nitrophenol with a second-order rate constant of 5.3×10^{-3} M^{-1} s^{-1} in 10 vol% TFE at pH 5.85. The rate constant correlated very well with that of 4-methylimidazole[1] (3.4×10^{-3} M^{-1} s^{-1}) if the difference in pK_a and a Brønsted coefficient β of 0.8[7] for nucleophilic catalysis of ester hydrolysis was taken into account.[2] RA-42, a polypeptide with 42 amino acids that folds into a helix-loop-helix motif in solution, where Orn-15 flanks His-11, reacts with a second-order rate constant of 5.1×10^{-2} M^{-1} s^{-1} in aqueous solution at pH 5.8 and 290 K. LA-42 and MA-42 differed by only one amino acid residue from the sequence of RA-42. In the former, Lys-15 replaced Orn-15 and in the latter Arg-15 replaced Orn-15. The second-order rate constants were

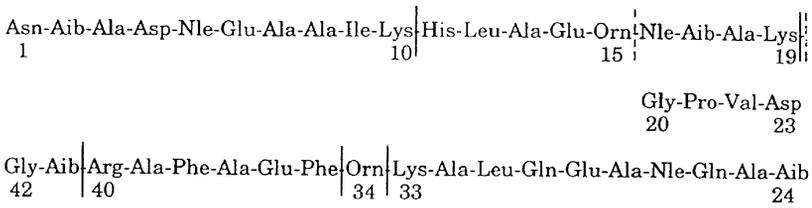


Fig. 2 Schematic representation of the result of trypsin cleavage of RA-42 after reaction with mono-*p*-nitrophenyl fumarate. Residues 1–19 and 24–42 were designed to fold into helical structures and the sequence from 20–23 was designed to form the loop. Trypsin is known to cleave polypeptides on the C-terminal side of basic residues but will not cleave at that site if the basic residue has been modified. The vertical solid lines have been used to indicate cleaved peptide bonds. Vertical dashed lines indicate peptide bonds that would have been cleaved in the absence of acylation, but where no cleavage was observed. Segments Asn-1 to Lys-10, His-11 to Lys-33, His-11 to Lys-19 and F-35 to Arg-40 were found by LC ESMS showing that Lys-10, Lys-33 and Arg-40 were not acylated. The observation of segment His-11 to Lys-33 as well as His-11 to Lys-19 shows that Lys-19 is acylated to some extent as indicated by using both a solid and a dashed vertical line. The fragments His-11 to Orn-15 or Nle-16 to Lys-33 were not found and Orn-15 is therefore acylated prior to Lys-19.

5.6×10^{-2} M^{-1} s^{-1} for LA-42 and 3.1×10^{-2} M^{-1} s^{-1} for MA-42. Arg, Lys or Orn in position $i + 4$, relative to a histidine in position i, thus increases the second-order rate constant by factors of 3–6 in aqueous solution at pH 5.85 and with factors as large as 38 in 30 vol% TFE, most likely by hydrogen bonding or other electrostatic stabilization in the transition state. The similarity in the rate increase between Arg, Lys and Orn substituted peptides was surprising since the side chains of the stabilizing residues differ in hydrogen bond donating ability (different pK_a values[8]) and length. In contrast, protonated flanking histidine side chains have recently been shown to provide rate enhancements of more than 10^3 over 4-MeIm catalysed reactions.[9] The reaction rates of the shorter peptides were found to be concentration dependent and these peptides were only used for studies of reaction products.

Identification of reaction products was accomplished by electrospray mass spectrometry (ESMS) or trypsin cleavage followed by LC ESMS. The acylation reaction was carried out by measuring the reaction rate for each peptide with 0.1 mM of mono-*p*-nitrophenyl fumarate[1] followed by the addition of 3–5 times excess of the same substrate. In RA-42 ESMS showed the predominant formation of monoacylated product. Trypsin treatment of the monoacylated reaction product followed by LC ESMS showed the formation of the fragments expected from cleavage of the peptide on the C-terminal side of the residues Lys-10, Lys-19, Lys-33, Orn-34 and Arg-40 (Fig. 2). No cleavage product was observed that corresponded to cleavage at Orn-15. In addition, one fragment (MW 1078) showed the molecular weight corresponding to the fragment His-11 to Lys-19 (980) plus the weight of fumaric acid (116) less the weight of water (18). The formation of an amide at the side chain of Orn-15 was thus established. Some reaction product from RA-42 also showed acylation at the side chain of Lys-19 (Fig. 2) as one fragment showed the molecular weight of the segment from His-11 to Lys-33 plus two fumaryl residues, less the molecular weight of two water molecules (Fig. 2). No other site of acylation was detected. Interestingly, ^{1}H NMR spectroscopic investigations, α-H chemical shift indices[10] and medium-range NOEs,[11] reveal that the structure of RA-42 is not helical at Lys-19 which suggests that some sequence specific conformation is the reason for the observed acylation reaction. The difference in reactivity between Orn-15 and Lys-19 is such that for the addition of one equivalent of substrate only Orn-15 acylation was detected.

The acylation of RA-42 and LA-42 also takes place in 30 vol% TFE where the helix-loop-helices do not form hairpin motifs and are monomeric.[12] The reaction is therefore intramolecular.

Several polypeptides were synthesised and reacted with excess substrate to explore the site selectivity of the reaction and the reaction products were analysed by ESMS or trypsin cleavage followed by LC ESMS. All peptides investigated had a histidine in position 11. LA-42 has the same sequence as RA-42 except that Orn-15 has been replaced by Lys-15. LA-42 is also acylated in position 15. The polypeptide JW-20Lys (Lys-7, Lys-14) showed no acylation, the polypeptides KB-20 (Lys-8, Lys-13), LB-20 (Lys-7, DAB-15) and JW-20Orn (Lys-7, Orn-15) were monoacylated and the polypeptide PB-20 (Lys-8, Orn-15) was monoacylated at Orn-15 showing that only positions 8 and 15 were functionalized by His-11 and that Orn-15 was preferentially acylated over Lys-8. Trypsin cleavage of the reaction product confirms preferential acylation of Orn-15 in the peptide PB-20. In the absence of a flanking residue that could be acylated selectivity was lost. In MA-42, a helix-loop-helix motif, excess substrate led to acylation at several sites. MA-42 has the same amino acid sequence as RA-42 except that Arg-15 replaces Orn-15.

This reaction has thus been carried out in a number of different helical peptides and the reaction appears to be quite general with large potential for the incorporation of functionality in folded polypeptides and proteins. So far we have incorporated amino acid derivatives, fumaryl residues[1] and a nicotinyl residue (dehydrogenase mimic) and the incorporation of carbohydrates is currently being explored. The structural basis for the observed selectivity is probably that unstrained cyclic transitions states are possible for lysine, ornithine and DAB residues in position 15 and for lysine residues in position 8, but not for the other cases reported here. Possible areas of application are the immobilization of proteins and the introduction of non-natural, sensitive or bulky functional groups into peptides and proteins. Surprisingly simple combinations of amino acid residues on the surface of helical structures can accomplish high selectivity and reactivity. The fact that the reaction takes place in aqueous solution as well as in TFE mixtures suggests that peptide catalysed transformations may have played a role in the prebiotic soup at the origin of life.

References

1 L. Baltzer, A.-C. Lundh, K. Broo, S. Olofsson and P. Ahlberg, *J. Chem. Soc., Perkin Trans. 2*, 1996, 1671.
2 K. Broo, L. Brive, A.-C. Lundh, P. Ahlberg and L. Baltzer, *J. Am. Chem. Soc.*, 1996, **118**, 8172.
3 W. C. Johnson, Jr., *Proteins*, 1990, **7**, 205.
4 Y.-H. Chen, J. T. Yang and K. H. Chau, *Biochemistry*, 1974, **13**, 3350.
5 M. Engel, R. W. Williams and B. W. Erickson, *Biochemistry*, 1991, **30**, 3161.
6 J. W. Nelson and N. R. Kallenbach, *Proteins*, 1986, **1**, 211.
7 T. C. Bruice and R. Lapinski, *J. Am. Chem. Soc.*, 1958, **80**, 2265.
8 C. Tanford, *Adv. Protein Chem.*, 1962, **17**, 69.
9 K. Ottosson, L. Brive, P. Ahlberg and L. Baltzer, unpublished results.
10 D. S. Wishart, B. D. Sykes and F. M. Richards, *J. Mol. Biol.*, 1991, **222**, 311.
11 K. Wutrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1985.
12 S. Olofsson and L. Baltzer, *Folding Des.*, 1996, **1**, 347.

Paper 6/05776D
Received 19*th August* 1996
Accepted 23*rd December* 1996