Tetrahedron Vol. 30, pp. 2755 to 2759. Pergamon Press 1974. Printed in Great Britain

THE SYNTHESIS OF PEPTIDES CONTAINING PURINE AND PYRIMIDINE DERIVATIVES OF DL-ALANINE

M. T. DOEL, A. S. JONES* and R. T. WALKER
Chemistry Department, The University of Birmingham, Birmingham. B15 2TT

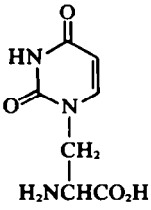
(*Received in the UK 2 November 1973; Accepted for publication 22 November 1973*)

Abstract—The incorporation of the "base-bearing amino acids", DL - β - (uracil - 1 - yl)alanine **1** (Uala), DL - β - (thymin - 1 - yl)alanine **2** (Tala), DL - β - (cytosin - 1 - yl)alanine **3** (Cala) and DL - β - (adenin - 9 - yl)alanine **4** (Aala) into peptides has been studied. The carboxyl group of each of these compounds was protected by the formation of the ethyl ester. The t - butoxycarbonyl (t-BOC) group was suitable for the protection of the α-amino group of **1** but not that of **2**, **3** and **4** because of the formation of ring substituted compounds. Peptides containing the amino acids **1–4** and L-serine were synthesised, however, by the mixed anhydride method; uracil, thymine and adenine residues needed no protection, neither did the cytosine residue provided that it was not present in the intermediate which was treated with ethyl chloroformate to produce the mixed anhydride. In this case reaction with the cytosine residue occurred. By these procedures, four protected dipeptides, namely α - N - t - BOC - L - seryl derivatives of **1–4** and three protected tetrapeptides, namely α - N - t - BOC - L - seryl - DL - β - (thymin - 1 - yl)alanyl - L - seryl - DL - β - (thymin - 1 - yl)alanine, ethyl ester (t - BOC - Ser - Tala - Ser - Tala - OEt), t - BOC - Ser - Aala - Ser - Aala - OEt and t - BOC - Ser - Uala - Ser - Uala - OEt were obtained. For the protection of the α-amino group of **2** or **3** the formyl group was found to be satisfactory.

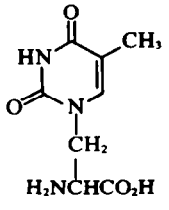
The synthesis of α-amino acids in which the side chains are the purine and pyrimidine bases found in nucleic acids has been described by us[1] and by other workers.[2-12] The object of obtaining these compounds is to incorporate them into peptides to see whether such peptides would have interesting biological properties. The presence of the purine and pyrimidine side chains might make possible interaction with nucleic acids or with enzymes which are concerned with nucleotide or nucleic acid metabolism. This paper reports techniques for the synthesis of peptides containing β - purinyl- or β - pyrimidinyl derivatives of alanine. In order to obtain peptides in which the purine or pyrimidine residues are correctly spaced for interaction with nucleic acids a "spacer" amino acid has been used. For this, serine was chosen because it was considered that it would be more likely than most other simple amino acids to give peptides with high water solubility. Similar work to this has been briefly reported, but only with respect to the uracil derivative **1** and glycine or tyrosine as spacer amino acid.[13]

To date only one of the five known amino acids of this type (**1–5**) (base-bearing amino acids) namely β - (uracil - 1 - yl)alanine (**1**) has been obtained in optically pure form,[2,14] so this work has been carried out on racemic mixtures. It was considered that this was satisfactory for the purpose of developing techniques provided the methods used were not likely to cause racemisation when applied to optically pure compounds.

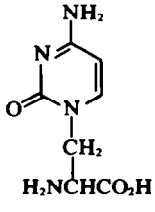
The blocking of the carboxyl groups by formation of the ethyl esters was straight forward, the ethyl esters of DL - β - (uracil - 1 - yl)alanine (**1**) (Uala), DL - β - (thymin - 1 - yl)alanine (**2**) (Tala), DL - β - (cytosin - 1 - yl)alanine (**3**) (Cala) and DL - β - (adenin - 9 - yl)alanine (**4**) (Aala) being readily formed. The use of the t-butoxycarbonyl group (t-BOC) for the protection of the α-amino group of the base-bearing amino acids has been investigated. Compound **1** was successfully converted into the



1

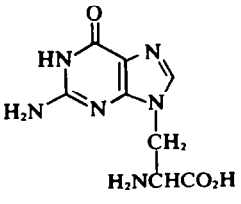


2



3

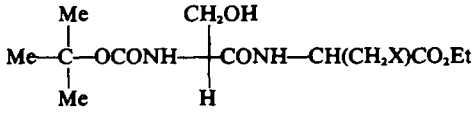
4



5

α-N-t-BOC derivative by treatment with an excess of t - butoxycarbonyl azide at room temperature. At 55°, however, an additional compound, tentatively identified as the 3 - N - t - BOC derivative, was formed. The α - N - compound was readily converted into the free amino acid by the standard procedure for the removal of the t-BOC group whereas the suspected 3 - N - compound was resistant. Attempts to obtain the α - N - t - BOC derivative of **2** by a similar procedure were unsuccessful because the major product was a compound which was probably the 3 - N - t - BOC derivative. Compounds **3** and **4** also gave products which were formed by reaction of the t - butoxycarbonyl azide with the N atoms of the heterocyclic bases; only small amounts of the required α - N - compounds were produced. Thus it appeared that the t-BOC group was only applicable to the uracil derivative. For the protection of the α-amino group of **2** and **3** the formyl group was found to be suitable, α - N - formyl - DL - β - (thymin - 1 - yl)alanine and α - N - formyl - DL - β - (cytosin - 1 - yl)alanine being obtained in good yield.

Despite this reactivity of the thymine, cytosine and adenine residues it was found that their corresponding amino acids could be incorporated into peptides without the need to protect the heterocyclic base moieties. This is shown to be the case by the following syntheses all of which were carried out by the use of the mixed anhydride procedure in which the carboxyl function is activated by reaction with ethyl chloroformate. α - N - t - BOC - L - serine was condensed with DL - β - (thymin - 1 - yl)alanine ethyl ester to give the protected dipeptide **6** (t - BOC - Ser - Tala - OEt). In a similar manner t - BOC - Ser - Cala - OEt (**7**) and t - BOC - Ser - Aala - OEt (**8**) were obtained. The fact that the



6; X = thymin-1-yl-
7; X = cytosin-1-yl-
8; X = adenin-9-yl-
9; X = uracil-1-yl-

mixed anhydride could be formed without substitution on a thymine or adenine side chain was demonstrated by the synthesis of the protected tetrapeptides, **10** and **11** from the partially deprotected dipeptides obtained from **6** and **8** by standard procedures. (All of these compounds are almost certainly mixtures of diastereomers but in the chromatographic systems used they were not separable into their components.) However attempts to obtain a protected tetrapeptide similar to **10** and **11** but containing cytosine residues gave mainly a product, which from it UV absorption spectrum appeared to be substituted at the 4-N position of the cytosine residue. It would appear in view of the successful synthesis of **7** that this must have occurred during the formation of the mixed anhydride. The uracil-containing analogues of **6** and **10** namely **9** and **12** were obtained without problems because of the resistance of the uracil ring to acylation.

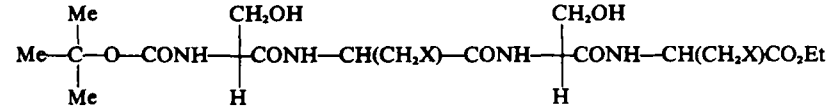
The structures of the protected base-bearing amino acids and the dipeptides and tetrapeptides **6–12**, particularly with regard to the absence of acylation on the base residues were established by UV and NMR spectroscopy. In the case of the uracil and thymine derivatives, absence of acylation at N-3 was shown by a decrease in ϵ_{max} with no change in λ_{max} and a bathochromic shift of about 10 nm in λ_{min} and an increase in ϵ_{min} upon changing the pH from 1-12. 1,3 - Disubstituted uracils show no change in spectrum under these conditions. The NMR spectra of the uracil- and the thymine-containing derivatives showed the presence of the proton at N-3. In the case of the cytosine and the adenine derivatives the absence of absorption peaks at about 300 nm shows that acylation had not occurred on the base residues. In the case of the adenine derivatives the NMR spectra showed the presence of the NH_2 protons at C-6.

The protecting groups could be removed from these protected peptides by standard procedures. The resulting peptides gave positive reactions with ninhydrin and migrated as expected upon paper electrophoresis. They were not characterised further because they were mixtures of diastereomers. In one case namely L - seryl - DL - β - (uracil - 1 - yl)alanine there was slight separation of the two diastereomers on the ion exchange resin of the amino acid analyser. The results obtained above, however, do show that these base-bearing amino acids can be incorporated into peptides.

EXPERIMENTAL

Chromatography. The following solvents were used (1) Butan - 1 - ol:ethanol:water (4:1:5) organic phase. (2) Butan - 2 - ol saturated with water. (3) Butan - 1 - ol:acetic acid:water (4:1:1). (4) Propan - 2 - ol:ammonia (*d*, 0·88):water (35:15:3).

Amino acid analysis of peptides. The normal conditions of peptide hydrolysis, namely 6N HCl at 110° for 36 h were satisfactory for peptides containing β - (uracil - 1 - yl)alanine, β - (thymin - 1 - yl)alanine and β - (cytosin - 1 - yl)alanine, but not in the case of peptides containing β - (adenin - 9 - yl)alanine. In this case a mixture of compounds, including glycine and unidentified UV absorbing



10; X = thymin-1-yl
11; X = adenin-9-yl
12; X = uridin-1-yl

compounds was obtained; only traces of β - (adenin - 9 - yl)alanine were present in the hydrolysate. A satisfactory hydrolysis of peptides containing this amino acid was obtained by using 60% perchloric acid at 130° for 15 min.

The base-bearing amino acids separated well from each other and from the common naturally-occurring amino acids on the ion exchange column of the automatic amino acid analyser. β - (Uracil - 1 - yl)alanine followed closely by β - (thymin - 1 - yl)alanine is eluted just after cysteic acid and before aspartic acid. β - (Adenin - 9 - yl)alanine and β - (cytosin - 1 - yl)alanine were eluted in that order after the ammonia peak and before ornithine and lysine.

Base-bearing amino acids. DL - β - (Uracil - 1 - yl)alanine, DL - β - (thymin - 1 - yl)alanine, DL - β - (cytosin - 1 - yl)alanine and DL - β - (adenin - 9 - yl)alanine were prepared as previously described.[1]

Ethyl esters. These were obtained from the amino acids by esterification with ethanol and dry HCl. The uracil and adenine derivatives were obtained as pure, crystalline hydrochlorides by crystallisation from aqueous EtOH. The thymine and cytosine derivatives were freed from traces of unchanged amino acids by ion exchange chromatography on Dowex-1 at pH 9 and obtained as free bases. The results are summarised in Table 1.

The compounds had similar UV absorption spectra to those of their amino acids.

α - N - t - Butoxycarbonyl - DL - β - (uracil - 1 - yl)alanine. DL - β - (Uracil - 1 - yl)alanine (5g) was dissolved in water (100 ml), MgO (2g) was added and to the vigorously stirred mixture, a soln of t - butoxycarbonyl azide (14·3 g) in dioxan (100 ml) was added. After stirring for 2 days at 20° the soln was diluted with water (100 ml), adjusted to pH 10–11 and extracted with ether (2 × 25 ml). The aqueous layer was adjusted to pH 4 with citric acid, cooled to 0° and extracted with amyl alcohol (5 × 100 ml). The amyl alcohol layer was washed with NaCl aq (3 × 20 ml), ice-cold water (3 × 20 ml) and dried over $MgSO_4$. The solvent was then evaporated off under reduced pressure to give a residue which was converted into a solid by repeated addition and vaporation of portions (100 ml) of ether–acetone (1 : 1 v/v). This solid was crystallised from EtOAc-light petroleum to give the *required product* (4·5 g) (60%) m.p. 183°(d) (Found: C, 47·9; H, 5·5; N, 14·0. $C_{12}H_{17}N_3O_6$ requires; C, 48·2; H, 5·7; N, 14·1%); λ_{max} 265 nm (ε, 8950); λ_{min} 233 nm at pH 1; λ_{max} 265 nm (ε, 6600) λ_{min} 243 nm at pH 12. The NMR spectrum was consistent with the assigned structure including a singlet at δ 11·34 for the ring proton on nitrogen at position 3. The com-

Table 1

Ethyl ester of	Yield %	M.p. (d)°	Found (%) C	H	N	Requires (%) C	H	N	Formula	R_f in solvent 1	4
DL-β-(uracil-1-yl) alanine HCl	94	195	41·4	5·2	15·9	41·1	5·3	15·9	$C_9H_{13}N_3O_4HCl$		0·75
DL-β-(thymin-1-yl) alanine	91	132	49·6	6·6	17·5	49·8	6·2	17·4	$C_{10}H_{15}N_3O_4$	0·58	
DL-β-(adenin-9-yl) alanine HCl	94	193–195	39·4	5·8	27·3	39·4	5·6	27·6	$C_{10}H_{14}N_6O_2.H_2O.HCl$		0·68
DL-β-(cytosin-1-yl) alanine	88	176–177	41·6	6·7	19·7	41·2	6·9	21·3	$C_9H_{14}N_4O_3.2H_2O$	0·40	

Table 2.

Ethyl ester of α-N-t-butoxycarbonyl-L-seryl-DL-	M.p. (d)°	Formula	Found (%) C	H	N	Requires (%) C	H	N
-β-(thymin-1-yl)alanine **6**	83–84	$C_{18}H_{28}N_4O_8$	50·4	6·0	13·1	50·5	6·5	13·1
-β-(cytosin-1-yl)alanine **7**	99–100	$C_{17}H_{27}N_5O_7$	49·3	6·5	17·3	49·4	6·5	17·0
-β-(adenin-9-yl)alanine **8**	112	$C_{18}H_{27}N_7O_6$	49·9	5·4	22·7	49·4	6·2	22·4
-β-(uracil-1-yl)alanine **9**	86–87·5	$C_{17}H_{26}N_4O_8$	50·7	6·6	13·7	49·4	6·3	13·5

Table 3.

Compound	UV absorption pH 1 λ_{max}	ϵ_{max}	λ_{min}	ϵ_{min}	pH 12 λ_{max}	ϵ_{max}	λ_{min}	ϵ_{min}
6	268 nm	9,780	236 nm	1,550	270 nm	7,566	246 nm	3,180
7	278 nm	11,900	240 nm	1,400	273 nm	8,650	250 nm	5,000
8	257 nm	14,500	234 nm	3,050	260 nm	14,750	235 nm	3,950
9	265 nm	8,900	232 nm	1,520	265 nm	6,300	243 nm	3,150

Table 4.

Compound	Yield (%)	Solvent used for isolation	RF Solvent 1	RF Solvent 4	Serine Base-bearing amino acid	UV absorption spectrum pH 1 λ_{max}	pH 1 λ_{min}	pH 1 ϵ_{260}	pH 12 λ_{max}	pH 12 λ_{min}
t-BOC-Ser-Tala-Ser-Tala-OEt **10**	33	2	0·94	0·78	0·87	269 nm	236 nm	16,630	270 nm	246 nm
t-BOC-Ser-Aala-Ser-Aala-OEt **11**	25	3	0·84	0·79	0·91	257 nm	234 nm	28,220	260 nm	235 nm
t-BOC-Ser-Uala-Ser-Uala-OEt **12**	30	2	—	0·63	0·96	265 nm	232 nm	17,290	265 nm	245 nm

Table 5. δ (ppm)

Compound	Ester $-CH_3$	t-BOC $-CH_3$	Thymine $-CH_3$	Serine $-OH$	C_5-H	C_6-H	Adenine $-NH_2$	Ring $-NH$	C_8-H	Adenine C_2-H
t-BOC-Ser-Tala-OEt **6**	1·12(t)3	1·40(s)9	1·74(s)3	3·33(s)1	—	7·33(s)1	—	11·20(s)1	—	—
t-BOC-Ser-Cala-OEt **7**	1·20(t)3	1·40(s)9	—	3·34(s)1	5·74(d)1	7·51(d)1	—	—	—	—
t-BOC-Ser-Aala-OEt **8**	1·10(t)3	1·40(s)9	—	3·39(s)1	—	—	7·26(s)2	—	8·08(s)1	8·23(s)1
t-BOC-Ser-Uala-OEt **9**	1·20(t)3	1·41(s)9	—	3·56(s)1	5·49(d)1	7·49(d)1	—	11·34(s)1	—	—
t-BOC-Ser-Tala-Ser-Tala-OEt **10**	1·15(t)3	1·39(s)9	1·80(s)6	3·40(s)2	—	7·34(s)2	—	11·82(s)2	—	—
t-BOC-Ser-Aala-Ser-Aala-OEt **11**	1·12(t)3	1·40(s)9	—	3·39(s)2	—	—	7·24(s)4	—	8·10(s)2	8·23(s)2
t-BOC-Ser-Uala-Ser-Uala-OEt **12**	1·23(t)3	1·42(s)9	—	3·40(s)2	5·43(d)2	7·50(d)2	—	11·38(s)2	—	—

pound was chromatographically homogeneous in solvent 1 (R_f 0·45) and solvent 4 (R_f 0·72).

α - N - Formyl - DL - β - (thymin - 1 - yl)alanine. DL - β - (Thymin - 1 - yl)alanine (1 g) was suspended in 98% formic acid (50 ml), Ac_2O (2 ml) was added and the mixture heated at 100° for 30 min. The soln was evaporated to dryness under reduced pressure and the residue crystallised from water to give the *required product* (0·74 g) (65%) m.p. 147°(d) (Found: C, 43·6; H, 4·7; N, 17·0. $C_9H_{11}N_3O_5$ 0·5 H_2O requires; C, 43·6; H, 4·8; N, 16·9%); λ_{max} 270 nm (ε, 9900) λ_{min} 236 nm at pH 1; λ_{max} 270 nm (ε, 7750), λ_{min} 246 nm at pH 12. The compound ran as a monobasic acid on paper electrophoresis at pH 6·8 and was chromatographically homogeneous in solvent 4 (R_f 0·58).

α - N - Formyl - DL - β - (cytosin - 1 - yl)alanine. DL - β - (Cytosin - 1 - yl)alanine (1 g) was treated as described above to give the *required product* (0·78 g) (68%) m.p. 142°(d) (Found: C, 42·1; H, 4·1; N, 25·2. $C_8H_{10}N_4O_4$ requires C, 42·5; H, 4·4; N, 24·8%); λ_{max} 281 nm (ε, 11·8 × 10^3), λ_{min} 241 nm at pH 1; λ_{max} 274 nm (ε, 8600), λ_{min} 252 nm at ph 12.

Protected dipeptides of L-serine and base-bearing amino acids. To a soln of α - N - t - butoxycarbonyl - L - serine (0·65 g, 3·1 mmole)[15] in THF (20 ml) and N - methylmorpholine (0·31 g, 3·1 mmole), at − 15° there was added ethyl chloroformate (0·33 g, 3·1 mmole). The soln was stirred at − 15° for 3 min and then a soln of the ethyl ester of the appropriate base-bearing amino acid (6·2 mmole) in dry DMF (30 ml) and Et_3N (0·63 g, 6·2 mmole) added. The mixtures were kept for 30–60 min at − 15°, then allowed to warm to room temp and the solvent removed *in vacuo* at 30°. Examination of the reaction mixtures by chromatography showed that in each case a new UV absorbing compound, which did not react with ninhydrin, had been formed. The *required compounds* were obtained as chromatographically homogeneous white powders in 50–60% yields after column chromatography on silica gel using either solvent 2 (**7** and **8**) or $CHCl_3$–EtOH (9:1) (**6** and **9**) as eluant. Attempts to crystallise the compounds failed, probably because they were mixtures of diastereomers. Amino acid analysis showed that they contained 1 mole of base bearing amino acid per mole of serine. Analytical and spectral data are given in Tables 2 and 3.

Protected tetrapeptides of L-serine and base - bearing amino acids. The protected dipeptides obtained above were partly deprotected by standard procedures to give compounds with either the carboxyl or amino group free. They were then combined to form protected tetrapeptides by the following procedure.

Partly-protected dipeptide (CO_2H free) (0·5 mmole) was dissolved in dry THF (40 ml), N - methylmorpholine (0·054 ml, 0·5 mmole) was added and the soln cooled to − 15°. Ethyl chloroformate (0·072 ml, 0·75 mmole) was then added and the soln kept at − 15° for 10 min. A soln of partly - protected dipeptide (NH_2 free) (1·0 mmole) in dry DMF (25 ml) at − 15° was then added and the soln kept at − 15° for 1 h. It was then allowed to warm to room temp and the solvents removed at 30° *in vacuo*. The products were then isolated by means of column chromatography on silica gel. The results are summarised in Table 4.

NMR spectra. These were recorded on a 'Varian A100' 100 MHz instrument with deuterochloroform as solvent and tetramethylsilane as internal standard. The results shown in Table 5 give chemical shifts for groups relevant for the assignment of the structures given.

Acknowledgements—The authors thank the Medical Research Council and the Cancer Research Campaign for financial assistance and the Science Research Council for a Research Studentship (to M.T.D.).

REFERENCES

[1]M. T. Doel, A. S. Jones and N. Taylor, *Tetrahedron Letters* 2285 (1969)
[2]J. H. Dewar and G. Shaw, *J. Chem. Soc.* 583 (1962)
[3]A. Kjaer, A. Knudsen and P. Olesen-Larsen, *Acta Chem. Scand.* **15**, 1193 (1961)
[4]N. T. Azorere and Y. P. Shvachkin, *J. Gen. Chem. USSR* **34**, 411 (1964)
[5]S. A. Giller, M. Yu Lidak, Y. Y. Shluke and Y. P. Shvachkin, *Chem. Heterocyclic Compd.* (Acad. Sci. Latvian Rep.) **2**, 124 (1968)
[6]M. Yu Lidak, Y. Y. Shluke and Y. P. Shvachkin, *Ibid.* **2**, 955 (1968)
[7]M. Yu Lidak, R. A. Paegle, K. Y. Petz, M. G. Plate and Y. P. Shvachkin, *Ibid*, **2**, 193 (1968)
[8]S. A. Giller, M. Yu Lidak, R. A. Paegle, M. G. Plate and Y. P. Shvachkin, *First All Union Conference on the Chemotherapy of Malignant Tumors* (*Chem. Sect.*) 103 (1968)
[9]H. de Koning and U. K. Pandit, Rec. Trav. Chim. **90**, 874 (1971)
[10]A. J. H. Nollet, C. M. Huting and U. K. Pandit, *Tetrahedron* **25**, 5971 (1969)
[11]A. J. H. Nollet and U. K. Pandit, *Ibid*, **25**, 5983, 5989 (1969)
[12]H. de Koning and U. K. Pandit, *Rec. Trav. Chim.* **91**, 1069 (1971)
[13]M. Yu Lidak, R. A. Paegle, V. E. Straume, D. E. Shnore and Y. P. Shvachkin, *Khim. Geterotski. Soedin.* **3**, 404 (1971)
[14]R. Gmelin, *Z. Physiol. Chem.* **316**, 164 (1959)
[15]E. Schnabel, *Liebigs Ann* **702**, 188 (1967)