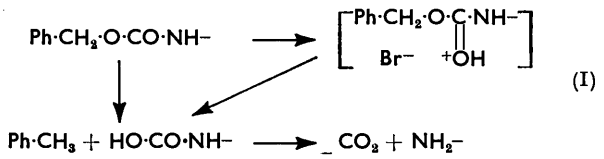
Recent Progress in the Chemistry of Peptides.

THE TILDEN LECTURE, DELIVERED BEFORE THE SOCIETY AT THE IMPERIAL COLLEGE OF SCIENCE AND TECHNOLOGY, LONDON, ON JANUARY 19TH, 1956, AND AT THE UNIVERSITY, MANCHESTER, ON FEBRUARY 9TH, 1956.

By G. W. KENNER.

INVESTIGATION of peptides began, of course, in 1901, when Emil Fischer and Fourneau synthesised glycylglycine and regarded it as the simplest example of the way in which amino-acids are joined together in proteins.[1] During the nineteen-thirties the peptide theory of protein structure was massively supported by the classical investigations of Max Bergmann, who showed that synthetic peptides are attacked by the various proteolytic enzymes in specific fashions.[2]

Synthesis.—The synthetic peptides required for these studies were mostly prepared by the " carbobenzoxy " (benzyloxycarbonyl) method [3]—Bergmann's great contribution to peptide synthesis. The benzyloxycarbonyl protecting group was removed by hydrogenolysis of the benzyl–oxygen link and consequent decarboxylation of the carbamic acid. An alternative



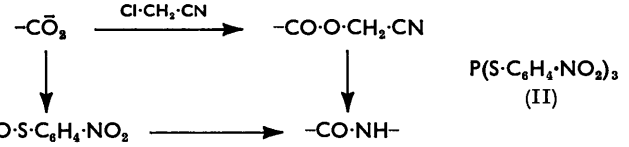
(I)

procedure, which has proved very convenient since its recent introduction,[4] consists in treating the benzyloxycarbonyl derivative with hydrogen bromide in an organic solvent such as acetic acid or nitromethane. The products are benzyl bromide, carbon dioxide, and the hydrobromide of the amino-compound. Presumably the function of the hydrogen bromide is to form an intermediate salt (I), in which the usual sensitivity of the benzyl-carbon atom to nucleophilic attack is greatly enhanced. Another modification of Bergmann's method is to use substituted benzyloxycarbonyl groups in order to obtain more easily crystallised derivatives.[5] Toluene-*p*-sulphonyl was suggested [6] as a protecting group even before discovery of the benzyloxycarbonyl method, but it has been widely used only in recent years since the sodium–liquid ammonia technique [7] has been developed for its removal; ion-exchange resins are valuable aids in working up these reaction mixtures.[8] Toluene-*p*-sulphonyl groups can also be removed with hydrogen bromide in acetic acid containing phenol.[9] A third protecting group, which has been successfully, although less widely, used is phthaloyl.[10,11] This is not an exhaustive catalogue of all the proposed and possible protecting groups, and the availability of this range of groups, which can be removed selectively,[12] will be useful in complex syntheses. The need for amino-protecting groups arises, of course, during extension of the peptide chain by condensation of the carboxyl group at its other end with the amino-group of the addendum.

Hitherto the commonest methods of condensation have involved conversion of the carboxyl group into a mixed anhydride. As this topic was discussed before the Society at a Symposium

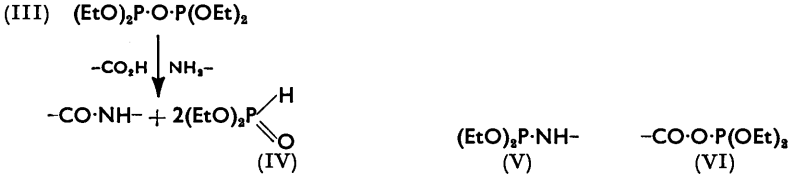
[1] Fischer and Fourneau, *Ber.*, 1901, **34**, 2868.
[2] Cf. Bergmann and Fruton, *Adv. Enzymol.*, 1941, **1**, 63.
[3] Bergmann and Zervas, *Ber.*, 1932, **65**, 1192.
[4] Ben-Ishai and Berger, *J. Org. Chem.*, 1952, **17**, 1564; Anderson, Blodinger, and Welcher, *J. Amer. Chem. Soc.*, 1952, **74**, 5311; Albertson and McKay, *ibid.*, 1953, **75**, 5323.
[5] Channing, Turner, and Young, *Nature*, 1951, **167**, 487; Carpenter and Gish, *J. Amer. Chem. Soc.*, 1952, **74**, 3818.
[6] Schönheimer, *Z. physiol. Chem.*, 1926, **154**, 203.
[7] Du Vigneaud and Behrens, *J. Biol. Chem.*, 1937, **117**, 27.
[8] Swan and du Vigneaud, *J. Amer. Chem. Soc.*, 1954, **76**, 3111; Rudinger, *Coll. Czech. Chem. Comm.*, 1954, **19**, 378.
[9] Weisblat, Magerlein, and Myers, *J. Amer. Chem. Soc.*, 1953, **75**, 3630.
[10] King and Kidd, *J.*, 1949, 3315.
[11] Sheehan and Frank, *J. Amer. Chem. Soc.*, 1949, **71**, 1856; Sheehan, Chapman, and Roth, *ibid.*, 1952, **74**, 3822.
[12] Boissonas and Preitner, *Helv. Chim. Acta*, 1953, **36**, 875.

less than a year ago,[13] I shall not deal with it now, and I shall turn instead to some alternative procedures of recent provenance. The first is very similar in principle; instead of an anhydride, an ester can be used provided that the esterifying group is sufficiently electronegative. This approach has been thoroughly explored by Schwyzer and his colleagues, who have arrived at cyanomethyl as the preferred group.[14] The cyanomethyl esters are prepared from a salt of the acid and chloroacetonitrile, and they react easily with amines in presence of acetic acid.

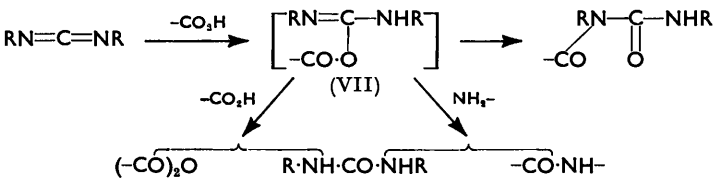


Bodánszky[15] has advocated the use of nitrophenyl esters, but we have preferred to use the *p*-nitrophenyl thiolesters, since these are more reactive and can be prepared directly from the acid by means of tri-*p*-nitrophenyl phosphorotrithioite[16] (II). (Phenyl thiolesters, prepared in a different way, had already been used in similar fashion.[17]) The " activated ester " and the " mixed anhydride " class of synthesis have this in common: activation of the carboxyl group and reaction with the amino-compound in two distinct steps.

A method of condensing the carboxyl and the amino-group in a single stage was supplied by Anderson and his colleagues[18] in 1952. Tetraethyl pyrophosphite (III), their reagent, is an unusual anhydride, for its hydration permits the phosphorus atoms to pass from the tervalent to the more stable quinquevalent state. [Diethyl phosphite is less stable than its tautomer diethyl phosphonate (IV).] This probably accounts for the fact that the corresponding



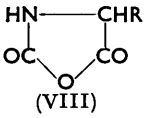
amides (V) yield carboxyamides through reaction with carboxylic acids.[19] The mixed anhydrides (VI), which may be formed from the pyrophosphite and carboxylic acids, react in the normal way with amines,[20] and thus the desired peptide is the product of either reaction path. Different techniques for the exploitation of tervalent phosphorus compounds have been developed by Goldschmidt and his colleagues.[21]



More recently Sheehan and Hess[22] have produced a particularly simple technique of direct condensation. Carbodi-imides[23] generally react with carboxylic acids giving either an *N*-acylurea[24] or the acid anhydride, presumably through intermediate formation of the *O*-acylurea

[13] Kenner, *Chem. Soc. Special Publ.* No. 2, 1955, p. 103.
[14] Schwyzer, Iselin, and Feurer, *Helv. Chim. Acta*, 1955, **38**, 69; Schwyzer, Feurer, Iselin, and Kägi, *ibid.*, p. 80; Iselin, Feurer, and Schwyzer, *ibid.*, p. 1508.
[15] Bodánszky, *Nature*, 1955, **175**, 685; Bodánszky, Szelke, Tömörkény, and Weisz, *Chem. and Ind.*, 1955, 1517.
[16] Farrington, Kenner, and Turner, *Chem. and Ind.*, 1955, 601.
[17] Wieland, Schäfer, and Bokelmann, *Annalen*, 1951, **573**, 99.
[18] Anderson, Blodinger, and Welcher, *J. Amer. Chem. Soc.*, 1952, **74**, 5309.
[19] Anderson, Blodinger, Young, and Welcher, *ibid.*, p. 5304.
[20] Anderson and Young, *ibid.*, p. 5307.
[21] Süs, *Annalen*, 1951, **572**, 96; Goldschmidt and Lautenschlager, *ibid.*, 1953, **580**, 68; Goldschmidt and Obermeier, *ibid.*, 1954, **588**, 24.
[22] Sheehan and Hess, *J. Amer. Chem. Soc.*, 1955, **77**, 1067.
[23] Khorana, *Chem. Rev.*, 1953, **53**, 145.
[24] For some examples in the peptide series, see Khorana, *J.*, 1952, 2081.

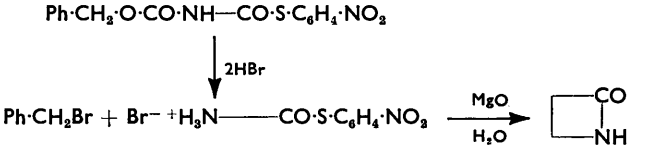
(VII). However, in presence of an amine these reaction sequences may be superseded by that leading to the amide. Di*cyclo*hexyl carbodi-imide is thus a promising reagent for peptide synthesis, although considerable quantities of *N*-acylureas have been isolated as by-products in some cases.[25]


(VIII)

The cyclic *N*-carboxy-α-amino-acid anhydrides ("Leuchs anhydrides") (VIII) have been used in syntheses of quite a different pattern. Although this route can be used for the addition of a single amino-acid residue to a peptide chain,[26] it is better adapted to the production of "poly-α-amino-acids."[27] The mechanism of this process was thoroughly discussed at last year's Symposium.[28]

There is one general point about peptide synthesis which is so important that I make no apology for considering it now despite the previous discussion[13] at the Symposium. Unless the residue concerned is glycine, there is an asymmetric carbon atom next to the carboxyl group undergoing the condensation process. Now, α-acylamino-acids are particularly susceptible to racemisation at this point when the carboxyl group is "activated," for example, by conversion into a mixed anhydride. Indeed racemisation can occur even when apparently it is the amino-group which is activated, as in "phosphite-amide" synthesis.[29] The extent of racemisation depends not only on the type of condensation procedure but also on the nature of the α-acyl substituent. The benzyloxycarbonyl group gives exceptionally resistant derivatives and therefore the safest practice[30] is to confine application of the newer syntheses, which are all, apart from the carbodi-imide method, known to cause racemisation under some conditions, to the benzyloxycarbonylamino-acids and to use the classical azide method for condensation at the carboxyl groups of di- and higher peptides. Such a drastic limitation means almost discarding the new syntheses and accordingly we have made a scrupulous examination of the extent of racemisation caused by our own two methods during the condensation of benzyloxycarbonylglycyl-L-alanine with either L-phenylalanylglycine or its ethyl ester.[31] With the first method,[32] which involves mixed anhydrides of sulphuric acid, racemisation could not be detected at all when dry dimethylformamide was the condensation medium, but in aqueous media there was an appreciable amount, which increased with the pH. However, less than 1% of the product was racemised when the pH was kept below 7 by means of solid magnesium carbonate. Under these conditions the nitrophenyl thiolester method caused much racemisation. This was disappointing, because retention of asymmetry during a different synthesis had been observed earlier.[16] An obvious conclusion is that each condensation should be treated on its own merits, but we believe that the sulphuric anhydride method will be generally free from racemisation during reactions in anhydrous media and we hope that it will also serve well for reactions in aqueous solutions. Theoretical considerations[31] accord with the fact that the thiolesters, which couple more slowly and therefore have a longer exposure to the aqueous media, suffer more racemisation.

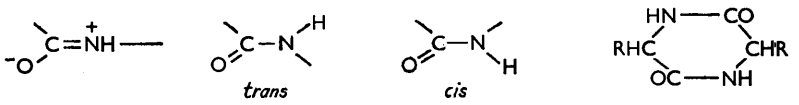
Cyclic Peptides.—Although thiolesters are not likely to be general reagents for peptide synthesis, their stability towards hydrolysis makes them valuable for the synthesis of cyclic peptides,[33] since a terminal amino-group, liberated from its benzyloxycarbonyl derivative by hydrogen bromide, has time to attack the thiolester under conditions of high dilution.* If possible, the ring should be closed at the carboxyl group of a glycine residue in order to avoid racemisation.



* Linear polycondensation also occurs with thiolesters in more concentrated solutions.[34]

[25] Khorana, *Chem. and Ind.*, 1955, 1087.
[26] Bailey, *J.*, 1950, 3461; cf. also Rudinger and Sörm, *Coll. Czech. Chem. Comm.*, 1951, **16**, 214; Zaoral, Rudinger, and Sörm, *ibid.*, 1953, **18**, 530; Zaoral and Rudinger, *ibid.*, 1955, **20**, 1185; Honzl and Rudinger, *ibid.*, p. 1191.
[27] Katchalski, *Adv. Protein Chem.*, 1951, **6**, 123.
[28] Ballard and Bamford, *Chem. Soc. Special Publ.* No. 2, 1955, p. 25.
[29] North and Young, *Chem. and Ind.*, 1955, 1597.
[30] Boissonas, Guttman, Jaquenoud, and Waller, *Helv. Chim. Acta*, 1955, **38**, 1491.
[31] Farrington, Ph.D. Thesis, Cambridge, 1955.
[32] Kenner and Stedman, *J.*, 1952, 2069.
[33] Kenner and Turner, *Chem. and Ind.*, 1955, 602.
[34] Wieland and Schäfer, *Annalen*, 1952, **576**, 104.

The stereochemistry of cyclisation is interesting. Mesomeric participation of the dipolar structure causes an amide to prefer the planar conformation. Of the two possibilities the *trans* is presumably more stable and occurs in open-chain peptides,[35] but there is evidently no serious bar to the *cis*-conformation, for cyclic dipeptides, dioxopiperazines, are formed from dipeptide



esters with notorious ease. A molecular model of a cyclic tripeptide shows that in this too the amide groups must be in the *cis*-form and that there is considerable congestion around the three hydrogen atoms on the α-carbon atoms; this would prohibit the existence of a cyclic tripeptide composed of both D- and L-amino-acids. Models of cyclic tetrapeptides can be constructed with some difficulty using various combinations of *cis*- and *trans*-conformations. All the amide groups are *trans* in the models of the penta- and hexa-peptides and there is no marked congestion. The thiolester cyclisation has been applied to several peptides composed of L-leucine and glycine and the yields of crystalline cyclic product obtained under comparable conditions from the tri-, tetra-, penta-, and hexa-peptides were **16**, **2**, **44**, and **44**% respectively.[36] These results agree with our deductions from the models and with the composition of the mixture of cyclic glycine peptides, in which the hexapeptide predominates, obtained from *N*-carboxyglycine anhydride;[37] moreover, *cyclo*hexaglycyl is also the major condensation product from diglycylglycine azide in dilute solution.[37,38] * Results of greater value should be provided by our current investigation of the effect of the asymmetry at the α-carbon atoms in the cyclisation process. The relatively rigid cyclic peptides are themselves stereochemically interesting, particularly with regard to their resistance to peptidases; their study may throw light on the conformation adopted by the peptide during enzymic attack.

Glutamic Acid Peptides.—The general problem of peptide synthesis is sometimes complicated by the side-chain groups. The important case of cystine peptides was discussed at last year's Symposium.[39] Two other amino-acids which present difficulties, histidine[40] and arginine,[41] have recently received attention. I shall only deal now with glutamic acid.

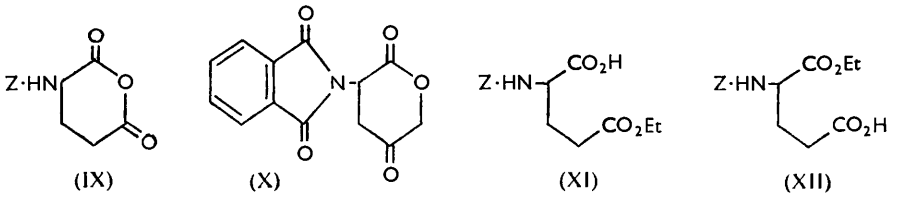
The problem here is to condense selectively either the α- or the γ-carboxyl with the amino-group of a peptide or amino-acid. It is the α-linkage which occurs in proteins, but derivatives in which the γ-carboxyl group is involved in chain formation are well known, namely, glutathione, the fermentation *Lactobacillus casei* factor,[42] and bacterial poly-D-glutamic acid.[43] The classical solution [3] of the α-problem is to use the cyclic anhydride (IX) of benzyloxycarbonyl-L-glutamic acid, which is attacked predominantly at the α-carbonyl group, from which electrons are withdrawn by the neighbouring urethane group. On the other hand, γ-derivatives are formed from the cyclic phthaloyl anhydride (X),[10] owing perhaps to the bulk of the phthalimido-substituent.† However, mixtures are produced, at least from the benzyloxycarbonyl

* [Added in proof.] *cyclo*Tetraglycyl has now been synthesised by Schcoyzer, Iselin, Rittel, and Sieber (*Helv. Chim. Acta*, 1956, **39**, 872).

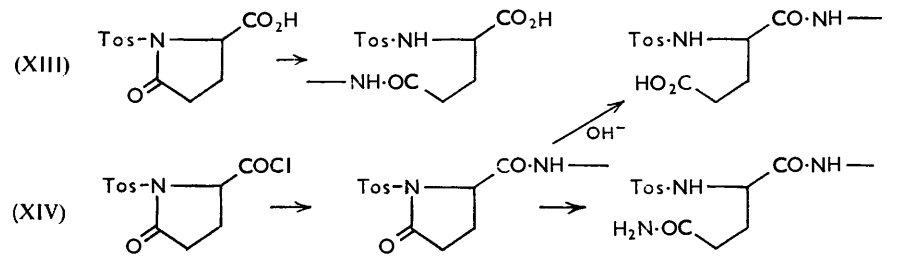
† It is interesting that the anhydrides (IX) and (X), which are unlikely to be in equilibrium with the isomeric oxazolones, can be prepared in the optically active forms, whereas only racemic anhydrides have been obtained from other acyl-L-glutamic acids [44] with the exception of the trifluoroacetyl compound.[45]

[35] Pauling and Corey, *Fortschr. Chem. org. Naturstoffe*, 1954, **11**, 180.
[36] Turner, Ph.D. Thesis, Cambridge, 1955.
[37] Bamford and Weymouth, *J. Amer. Chem. Soc.*, 1955, **77**, 6368.
[38] Sheehan and Richardson, *ibid.*, 1954, **76**, 6329; Sheehan, Goodman, and Richardson, *ibid.*, 1955, **77**, 6391.
[39] Du Vigneaud, *Chem. Soc. Special Publ.* No. 2, 1955, p. 49.
[40] Holley and Sondheimer, *J. Amer. Chem. Soc.*, 1954, **76**, 1326; Fischer and Whetstone, *ibid.*, p. 5076.
[41] Gish and Carpenter, *ibid.*, 1953, **75**, 5872; Anderson, *ibid.*, p. 6081; Hofmann, Rheiner, and Peckham, *ibid.*, p. 6083; Van Orden and Smith, *J. Biol. Chem.*, 1954, **208**, 751.
[42] Semb, Boothe, Angier, Waller, Mowat, Hutchings, and SubbaRow, *J. Amer. Chem. Soc.*, 1949, **71**, 2310, and previous papers.
[43] Bruckner, Kovács, and Nagy, *J.*, 1953, 148; Waley, *J.*, 1955, 517.
[44] An example and references to other cases are given by Clayton, Kenner, and Sheppard, *J.*, 1956, 371.
[45] Weygand and Leising, *Chem. Ber.*, 1954, **87**, 248.

anhydride.[46,47] The crystalline γ- (XI) [48,49] and α-derivative (XII),[46,49] ethyl esters of benzyl oxycarbonyl-L-glutamic acid, can serve as starting materials for other syntheses,[50] but these are attended by the risk of various rearrangements [51] which I cannot discuss now.



Safer routes are available through protection of one carboxyl group as its benzyl ester [52] or by use of the crystalline α-azide γ-amide of benzyloxycarbonylglutamic acid.[53] Nevertheless a completely new approach, which has been made quite independently in two laboratories,[54] is very welcome. It has already proved fruitful and may provide a general solution of the problem. Its basis is the circumstance that the primary dehydration product from toluene-*p*-sulphonyl-L-glutamic acid is not the customary cyclic anhydride but the optically active pyrrolidone (" tosyl pyroglutamic acid ") (XIII).[55] Because this is a diacylimide, the ring



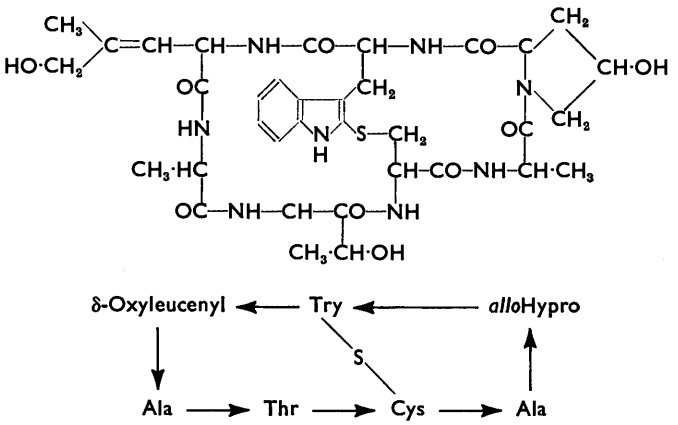
will open when it is attacked by nucleophilic reagents, yet these will prefer to react with the exocyclic carbonyl group in the corresponding acid chloride (XIV), which can also be prepared directly from toluene-*p*-sulphonyl-L-glutamic acid. As sketched in the diagram, ways appear open to all the various glutamyl derivatives, but as yet only some of these have been exploited. Earlier, the pyrrolidone ring had been used for simultaneous protection of the amino- and the γ-carboxyl function in glutamic acid,[56] but in the absence of the toluene-*p*-sulphonyl group opening of the ring is more difficult and not always satisfactory.[57]

Another application of these toluene-*p*-sulphonylpyrrolidones has been to the synthesis of L-proline from L-glutamic acid.[58] The same conversion takes place in micro-organisms, but by a different path.[59]

[46] Le Quesne and Young, *J.*, 1950, 1954.
[47] Wieland and Weidenmüller, *Annalen*, 1955, **597**, 111.
[48] Abderhalden and Nienburg, *Z. physiol. Chem.*, 1933, **219**, 155; Bergmann and Zervas, *ibid.*, 1933, **221**, 53.
[49] Rudinger, *Coll. Czech. Chem. Comm.*, 1951, **16**, 615.
[50] Hegedüs, *Helv. Chim. Acta*, 1948, **31**, 737; Le Quesne and Young, *J.*, 1950, 1959; cf. also Harington and Mead, *Biochem. J.*, 1935. **29**, 1602.
[51] Cason, *J. Org. Chem.*, 1948, **13**, 227; Chase and Hey, *J.*, 1952, 553; Battersby and Robinson, *J.*, 1955, 259; Sachs and Brand, *J. Amer. Chem. Soc.*, 1954, **76**, 1815.
[52] Sachs and Brand, *ibid.*, 1953, **75**, 4608; Clayton, Kenner, and Sheppard, *J.*, 1956, 373.
[53] Sondheimer and Holley, *J. Amer. Chem. Soc.*, 1954, **76**, 2816.
[54] Swan and du Vigneaud, *ibid.*, 1954, **76**, 3110; Rudinger, *Coll. Czech. Chem. Comm.*, 1954, **19**, 375.
[55] Harington and Moggridge, *J.*, 1940, 706.
[56] Angier, Waller, Hutchings, Boothe, Mowat, Semb, and SubbaRow, *J. Amer. Chem. Soc.*, 1950, **72**, 74.
[57] Chambers and Carpenter, *ibid.*, 1955, **77**, 1523.
[58] Pravda and Rudinger, *Coll. Czech. Chem. Comm.*, 1955, **20**, 1.
[59] Ehrensvärd, *Ann. Rev. Biochem.*, 1955, **24**, 282.

Peptides of Natural Origin.—What I have said so far will have shown that considerable ingenuity and effort have been expended during the last half-dozen years on the synthesis of peptides. It may justly be asked what purpose has been behind them. Certainly not the provision of enzyme substrates: the classical methods were adequate, although the new ones may be helpful. Rather it was the belief shared by many working in this field that the number of relatively small peptides with interesting biological properties would shortly grow appreciably. The principal factor in the realisation of this prediction has been the vast improvement in techniques for the purification of peptides. These can be classified under two headings. There are those which use partition between aqueous and organic solvents, either continuously in packed columns [60] or discontinuously in Craig's apparatus for counter-current distribution;[61] others, namely, zone electrophoresis [62] and ion-exchange chromatography,[60,63] take advantage of the electric charges which are borne by most peptides in aqueous solutions.

There are now too many known peptides [64] for me to mention them all individually, and I shall just draw attention to a few examples in each of three groups, classified by biological activity. Amongst the hormones the most completely investigated examples are the insulins and oxytocin. The structures are also known of the two vasopressins [65] and β-corticotropin,[66] one member of the family of peptides composing the adrenocorticotrophic hormone (ACTH). It should not be supposed that all the " protein hormones " will turn out to be peptides with molecular weights less than 6000, like those just mentioned, but nevertheless further progress along these lines can be expected.[67] There are various peptide antibiotics with molecular weights about 1200; the exact structures of gramicidin-S [68] and tyrocidine-A [69] are known and



much progress has recently been made with bacitracin-A.[70] The third group are powerful toxins. Comparatively little is still known about the neurotoxins of snake venoms,[71] but the complete structure of one fungal toxin, phalloidin (molecular weight 770), has been elucidated by the brilliant investigations of T. Wieland.[72]

[60] Campbell and Work, *Brit. Med. Bull.*, 1954, **10**, 196; Porter, *ibid.*, p. 237.
[61] Weisiger, " Organic Analysis," Interscience Publ. Inc., New York, 1954, Vol. II, p. 278; Hecker, " Verteilungsverfahren im Laboratorium," Verlag Chemie, Weinheim, 1955.
[62] Tiselius and Flodin, *Adv. Protein Chem.*, 1953, **8**, 461.
[63] Samuelson, " Ion Exchangers in Analytical Chemistry," Wiley, New York, 1953.
[64] Synge, *Quart Rev.*, 1949, **3**, 245; Bricas and Fromageot, *Adv. Protein Chem.*, 1953, **8**, 1.
[65] Du Vigneaud, Lawler, and Popenoe, *J. Amer. Chem. Soc.*, 1953, **75**, 4880; Acher and Chauvet, *Biochim. Biophys. Acta*, 1953, **12**, 487.
[66] Bell, *J. Amer. Chem. Soc.*, 1954, **76**, 5565; Howard, Shepherd, Eigner, Davies, and Bell, *ibid.*, 1955, **77**, 3419. For the structure of corticotropin-A see White and Landmann, *ibid.*, p. 1711.
[67] Gaddum (ed.), " Polypeptides which stimulate Plain Muscle," Livingstone, Edinburgh, 1955; Peart, *Nature*, 1956, **177**, 132.
[68] Synge, *Biochem. J.*, 1945, **39**, 363; Consden, Gordon, Martin, and Synge, *ibid.*, 1947, **41**, 596.
[69] Paladini and Craig, *J. Amer. Chem. Soc.*, 1954, **76**, 688.
[70] Lockhart, Abraham, and Newton, *Biochem. J.*, 1955, **61**, 534; Weisiger, Hausmann, and Craig, *J. Amer. Chem. Soc.*, 1955, **77**, 731, 3123.
[71] Slotta, *Fortschr. Chem. org. Naturstoffe*, 1955, **12**, 433.
[72] Wieland and Schön, *Annalen*, 1955, **593**, 157.

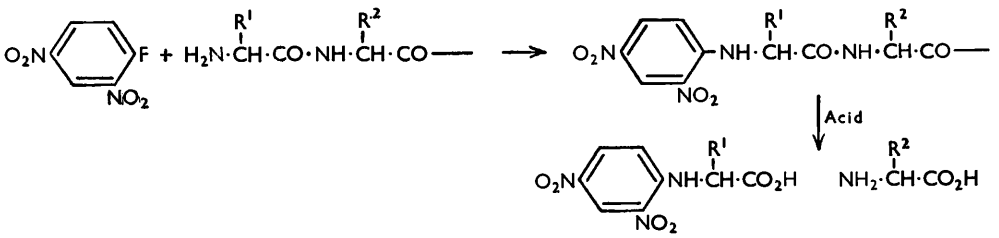
The structure * of phalloidin illustrates the way in which the peptides from lower organisms frequently incorporate amino-acids not found in proteins.[74] Thus, apart from the sulphide cross-link, there is the unique "δ-oxyleucenyl" (2-amino-5-hydroxy-4-methylpent-3-enoyl) residue and the configuration at the secondary hydroxy-group in the pyrrolidine ring is unusual. In other natural products there are residues with usual structures but the D-configuration.

It may be that many other peptides still await detection and isolation because they lack bacteriological or pharmacological activity,[75] but there is little evidence in favour of the obvious possibility that peptides might be definite intermediates in protein synthesis.[76]

The Structures of Insulin and Oxytocin.—The finest example of analytical work in the peptide field is the solution of the insulin problem by Sanger and his colleagues. An adequate account needs a complete lecture, but I shall try to sketch the chief steps.

Insulin has this much to recommend it as a subject of study: it is commercially available in a highly purified form, adequate for virtually all degradations.† Complete purification has been achieved by counter-current distribution [78] and this material has been used for exact analysis of the amino-acid content [79] by the excellent ion-exchange method of Moore and Stein [80] after drastic acid hydrolysis. Analyses had to be made at intervals during the hydrolysis since some amino-acids (serine, threonine, tyrosine) were progressively destroyed, whereas others (*iso*leucine, valine) were liberated only slowly. These phenomena are general and can be compensated by suitable corrections.[81] It is remarkable that a great deal of progress with the structural problem was made before the exact composition of insulin had been determined; indeed the correct amino-acid content had by then been deduced from the degradative studies!

In order to be able to tackle the structural problem, Sanger devised the "DNP" method.[82] This consists in treating the peptide with 1-fluoro-2 : 4-dinitrobenzene, which reacts with amino-groups much more easily than does the chloro-analogue tried much earlier.[83] The newly



formed bond is very stable to acid, although it can be broken by alkali,[84] and consequently the amino-acid residues with free amino-groups can be identified as their yellow dinitrophenyl

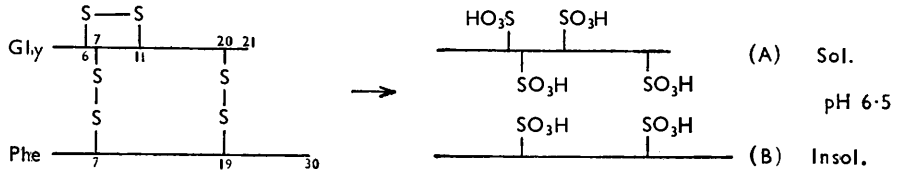
* Comparison of the two formulæ shows how much more expressive and economical is the short hand [73] now widely used. The literal symbols stand for amino-acid residues, *e.g.*, Ala for alany -NH·CH(CH$_3$)·CO–. An arrow symbolises a bond *from* a carbonyl of one residue *to* the imino-group o another. Where dots are used instead of arrows in other formulæ, the residues are always arranged with their imino-groups on the left. Chains ought strictly to be terminated by H· and ·OH, or $^+H_2$· and ·O$^-$, if the end-groups are free, but these terminations are usually omitted in reporting analytical work. Z means benzyloxycarbonyl.

† Cattle, pigs, and sheep produce insulins of slightly different composition. The differences affect only a single region of three amino-acid residues [77] and the following remarks apply to all three substances. That from cattle was used throughout Sanger's work, except for the comparison of species.

[73] Brand and Edsall, *Ann. Rev. Biochem.*, 1947, **16**, 223; Erlanger and Brand, *J. Amer. Chem. Soc.*, 1951, **73**, 3509.
[74] Bricas and Fromageot, *Adv. Protein Chem.*, 1953, **8**, 6.
[75] Synge, "The Chemical Structure of Proteins" (ed. Wolstenholme and Cameron for the Ciba Foundation), Churchill, London, 1953, p. 43.
[76] Ehrensvärd, *Ann. Rev. Biochem.*, 1955, **24**, 300.
[77] Brown, Sanger, and Kitai, *Biochem. J.*, 1955, **60**, 556.
[78] Harfenist and Craig, *J. Amer. Chem. Soc.*, 1952, **74**, 3083.
[79] Harfenist, *ibid.*, 1953, **75**, 5528.
[80] Moore and Stein, *J. Biol. Chem.*, 1954, **211**, 893, and earlier papers.
[81] Smith, Stockell, and Kimmel, *J. Biol. Chem.*, 1954, **207**, 551; Hirs, Stein, and Moore, *ibid.*, **211**, 941.
[82] Sanger, *Biochem. J.*, 1945, **39**, 507; Levy, "Methods of Biochemical Analysis" (ed. Glick), Interscience Publ. Inc., New York, 1955, Vol. II, p. 360.
[83] Abderhalden and Stix, *Z. physiol. Chem.*, 1923, **129**, 143.
[84] Lowther, *Nature*, 1951, **167**, 767.

derivatives, which can be chromatographed on either silica gel [83] or paper.[85] The free amino-groups in insulin turned out to be those of glycine and phenylalanine and the ε-group of lysine, a diamino-monocarboxylic acid, and equivalent amounts of them were detected. Later, using counter-current distribution, Harfenist and Craig [86] isolated the mono(dinitrophenyl) derivative of insulin from partial reaction with fluorodinitrobenzene and showed that the intensity of its absorption at 3500 Å corresponded to a molecular weight of about 6500. In contrast, almost all the usual physical methods had dealt with aggregates and had given values of at least 12,000, which had been regarded as the true molecular weight. The correct figure for insulin from cattle is 5733.

Sanger deduced that there are in insulin two kinds of peptide chain, with terminal glycine and phenylalanine residues respectively, and that these are linked by the disulphide bridges of cystine residues. The truth of this belief was demonstrated by oxidising insulin with performic acid and separating the two chains.[87] The actual arrangement of the amino-acid residues in



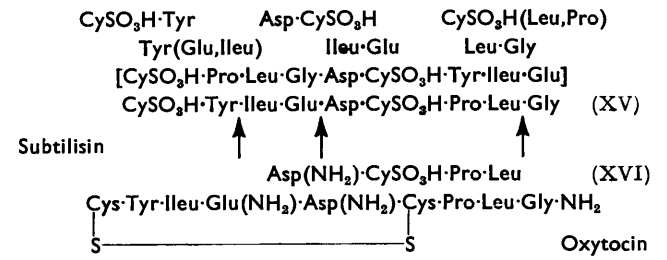
insulin is outlined in the diagram; the numbers refer to the location of different residues along each chain and for this purpose each half of a cystine residue is counted separately. The diagram shows how two of the disulphide bridges are used for interchain links and the third for an intrachain link. The original separation of the two oxidised fragments was based on the fact that the " B " (phenylalanyl) chain has, after oxidation, equal numbers of basic and acidic groups and therefore is insoluble at pH 6·5, whereas the " A " (glycyl) chain is strongly acidic and remains in solution. The techniques of partition chromatography [88] and counter-current distribution [89] have since been applied to this separation.

The sequences of the 30 amino-acid residues in the B chain [90] and of the 21 in the A chain [91] were determined by partial hydrolysis of each chain, separation of the resultant extremely complex mixtures of peptides by electrophoresis and paper chromatography, and analysis of these peptides by total hydrolysis and the DNP method. This work was far too intricate for me to describe here. Instead I shall illustrate the principles by referring to a simpler problem, the structure of oxytocin, a posterior-pituitary hormone. Oxytocin was isolated by du Vigneaud and his colleagues,[92] who showed that it contains one cystine residue together with equivalent amounts of seven other amino-acids and three molecules of ammonia.[93] In addition, they oxidised oxytocin with performic acid and showed that the product contained, instead of the cystine, two residues of cysteic acid and still seven other residues and three molecules of ammonia.[94] Performic-oxidised oxytocin thus presented the same problem as did the B chain of insulin, except that there were 9 residues to be placed in sequence instead of 30.

Tuppy and Michl [95] solved this problem in the fashion in which Sanger and Tuppy had already solved the larger one. Partial acidic hydrolysis gave a mixture of the two tri- and four di-peptides shown below; the parentheses in the tripeptide structures enclose residues of undetermined sequence, because the DNP method identified only the *N*-terminal residue (*i.e.*, that with a free amino-group). With one more piece of information, namely, that the *N*-terminal residue is cysteic acid, these fragments could be reassembled in only the two sequences shown. Enzymic degradation with subtilisin, which attacked the chain at the three points marked by arrows, provided additional degradation products including the tetrapeptide

[85] Biserte and Osteux, *Bull. Soc. Chim. biol.*, 1951, **33**, 50; Levy, *Nature*, 1954, **174**, 126.
[86] Harfenist and Craig, *J. Amer. Chem. Soc.*, 1952, **74**, 3087.
[87] Sanger, *Biochem. J.*, 1949, **44**, 126.
[88] Andersen, *Compt. rend. Trav. Lab. Carlsberg*, 1954, **29**, 49.
[89] Pierce, *J. Amer. Chem. Soc.*, 1955, **77**, 184.
[90] Sanger and Tuppy, *Biochem. J.*, 1951, **49**, 463, 481.
[91] Sanger and Thompson, *ibid.*, 1953, **53**, 353, 366.
[92] Livermore and du Vigneaud, *J. Biol. Chem.*, 1949, **180**, 365; Pierce, Gordon, and du Vigneaud, *ibid.*, 1952, **199**, 929.
[93] Pierce and du Vigneaud, *ibid.*, 1950, **182**, 359; **186**, 77.
[94] Mueller, Pierce, Davoll, and du Vigneaud, *ibid.*, 1951, **191**, 309.
[95] Tuppy and Michl, *Monatsh.*, 1953, **84**, 1011.

(XVI).* This was inconsistent with the nonapeptide sequence in square brackets and confirmed the alternative (XV). By disposing of the three molecules of ammonia on the " spare " carboxyl groups and reconstituting the disulphide bridge, the complete structure of oxytocin is reached with the ambiguity that no evidence has been presented concerning the α- or γ- and α- or β-nature of the glutamyl and the aspartyl link respectively. The same stage was reached independently after a more elaborate series of degradations by du Vigneaud and his colleagues,[96]



but they carried the matter further. By a brilliant total synthesis they prepared material with αα-linkages in the chain and showed that it was identical with the isolated hormone.[97] Those who know Dr. du Vigneaud's own account [39] of this work last year will not wish for mine. I shall do no more than mention that the synthesis depended on protection of the sulphur atoms as benzyl thiolesters [98] and on the " pyrophosphite " [18] and the " tosylpyroglutamyl " [54] method which I have already discussed. Active material has since been synthesised in two other laboratories,[80, 99] but the purification does not seem to have been completed in these cases.

Returning to insulin—the structures of the A and the B chain were completely defined by the products of acidic and enzymic hydrolyses,[90, 91, 100] but there were several possible arrangements of the disulphide bonds. Distinction between these was made difficult by interchange amongst the mixed disulphides, which can occur by two mechanisms,[101] but nevertheless suitable conditions were found for partial hydrolysis without interchange and at last the picture was complete.[102] It is hardly possible to overemphasise the magnitude of this achievement: apart from structures with many repeating units, insulin, containing 254 carbon atoms, is by far the largest molecule of known structure. On the other hand, now that this barrier has been surmounted progress with yet larger molecules can be expected.

The question of whether the synthesis of insulin is around the corner has been asked [103] and answered.[104] A major difficulty is that no-one has yet been able to reduce the disulphide bonds of insulin *reversibly*; the tendency of two chains to arrange themselves in the antiparallel fashion may be the cause of the trouble.[105] Consequently the type of synthesis involving oxidation of a reduced form of the hormone, which succeeded with oxytocin, cannot be applied, and the synthesis of insulin must await the development of new approaches to synthetic peptides of cystine. Probably the next stage in the synthesis of polypeptide hormones will be the synthesis of the biologically active fragment of corticotropin which contains 28 amino-acid residues. In this instance the cystine problem is absent, but in compensation there are others.

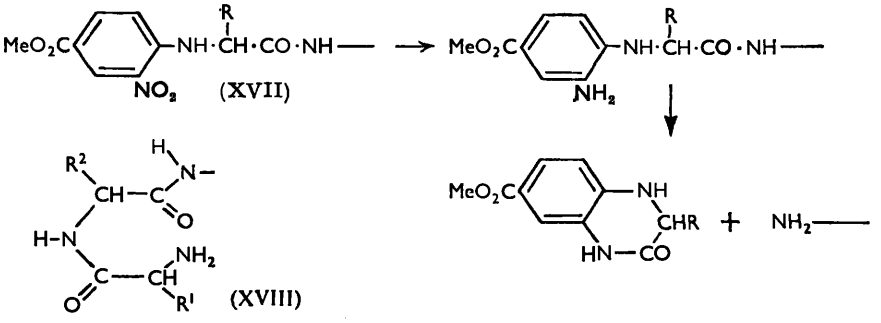
Reactions.—My account of structural analysis of two hormones has, I hope, demonstrated both that the partial hydrolysis method [106] is extremely powerful in the hands of a master of the art and that it can be very complicated. For this reason considerable attention has been paid to possible methods of specific degradation. The proteolytic enzymes have already been used to supplement the stock of acidic-degradation products,[90, 91, 95] and they are likely to be used still more in the future, particularly since highly purified enzyme preparations can now be

* The amide groups, which were lost in the acidic hydrolysis, survived enzymic degradation.

[96] Du Vigneaud, Ressler, and Trippett, *J. Biol. Chem.*, 1953, **205**, 949.
[97] Du Vigneaud, Ressler, Swan, Roberts, and Katsoyannis, *J. Amer. Chem. Soc.*, 1954, **76**, 3115.
[98] Loring and du Vigneaud, *J. Biol. Chem.*, 1935, **111**, 385.
[99] Honzl and Rudinger, *Coll. Czech. Chem. Comm.*, 1955, **20**, 1190.
[100] Sanger, Thompson, and Kitai, *Biochem. J.*, 1955, **59**, 509.
[101] Ryle and Sanger, *ibid.*, 1955, **60**, 535.
[102] Ryle, Sanger, Smith, and Kitai, *ibid.*, 541.
[103] Harington, *Chem. Soc. Special Publ.* No. 2, 1955, p. 64.
[104] Du Vigneaud, *ibid.*, p. 65.
[105] Rydon, *ibid.*, p. 66.
[106] Consden, Gordon, Martin, and Synge, *Biochem. J.*, 1947, **41**, 596; Sanger, *Adv. Protein Chem.* 1952, **7**, 12.

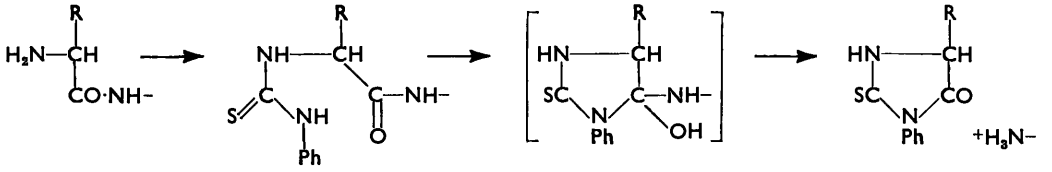
bought. Also many schemes for specific degradation without enzymes have been devised and tested with simple peptides. However, only one of these, that due to Edman,[107] has actually been used to any extent in structural anaylsis.[66, 72, 96, 108] Consequently the main result of these studies has been, in my view, the demonstration that, although the amide group in peptides is normally inert, it can be brought into reaction under very mild conditions provided that the environment is suitable. These reactions of the amide group can be classified under two headings, attack by nucleophilic reagents at the carbonyl group and attack by electrophilic reagents at the nitrogen atom.

An excellent example of the former class is the method for selectively removing the *N*-terminal residue discovered by Holley and Holley.[109] The first step is to attach a 4-methoxy-

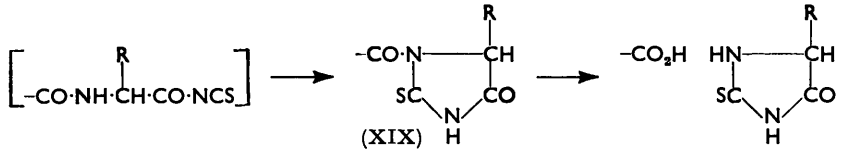


(XVII) (XVIII)

carbonylnitrophenyl group by means of the appropriate fluoro-reagent. Reduction of the intermediate (XVII) then generates an amino-group in a favourable position for attack on the neighbouring amide group and thus liberation of the degraded peptide. In the simple examples described this cyclisation was complete in five hours at room temperature in neutral solution. In contrast, typical conditions for the *partial* hydrolysis of polypeptides are treatment with concentrated hydrochloric acid at 37° for three days or with 0·2N-sodium hydroxide at 100° for five hours. It may well be asked how peptides themselves can be stable, since the terminal amino-group might likewise attack the neighbouring peptide bond forming a dioxopiperazine.



This reaction does occur, but only during very drastic heating in phenolic solvents.[110] At least part of the comparative slowness of this reaction can be ascribed to the necessity, indicated in formula (XVIII), of bringing the terminal amide group into the unfavourable *cis*-conformation.



(XIX)

The intermediate in Edman's degradation[107] is the phenylthioureido-derivative of the peptide. When hydrogen chloride is bubbled into solutions of these intermediates in nitromethane,* the hydrochlorides of the degraded peptides are soon precipitated and the *N*-terminal residue is split off as its phenylthiohydantoin. Rather similar schemes have been devised

* With insoluble compounds aqueous solutions may be used with somewhat less satisfactory results.

[107] Edman, *Acta Chem. Scand.*, 1950, **4**, 283; Landmann, Drake, and Dillaha, *J. Amer. Chem. Soc.*, 1953, **75**, 3638; Fraenkel-Conrat, "Methods of Biochemical Analysis" (ed. Glick), Interscience Publ. Inc., New York, 1955, Vol. II, p. 383.

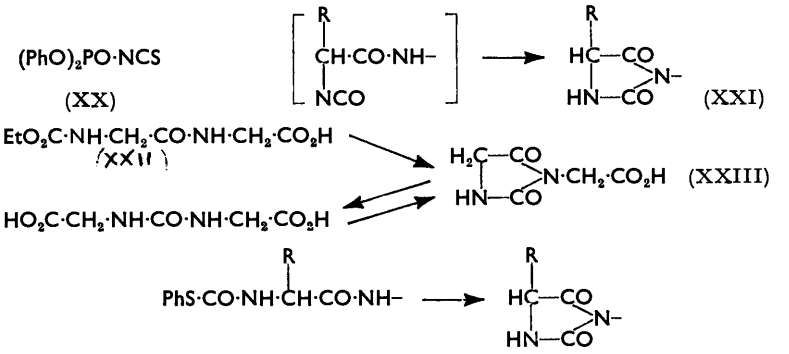
[108] Popenoe and du Vigneaud, *J. Biol. Chem.*, 1953, **205**, 133; Thompson, *ibid.*, 1954, **208**, 565.

[109] Holley and Holley, *J. Amer. Chem. Soc.*, 1952, **74**, 5445.

[110] Lichtenstein, *ibid.*, 1938, **60**, 560; Rydon and Smith, *J.*, 1955, 2542.

by Khorana and myself [111] and by Elmore and Toseland.[112] Another, which I regret that I have not time to describe, was discovered at the same time as Edman's by the late A. L. Levy [113] and was developed by him and Léonis in the most elegant way.[114] *

The second class of reactions is well illustrated by a method for selective removal of the amino-acid residue with a free carboxyl group, the *C*-terminal residue.[115] When the carboxyl group is converted into an acyl *iso*thiocyanate, this cyclises spontaneously to an acylthiohydantoin (XIX), which can be easily hydrolysed. We have shown that the phosphate compound (XX) is an effective reagent for achieving the cyclisation under mild conditions.[116] Similarly, when the amino-group at the other end of the peptide chain is converted into an *iso*cyanate, the isolated product is the isomeric hydantoin (XXI), formed by electrophilic attack of the carbonyl group on the amide-nitrogen atom.[117] A closely related reaction was discovered by Emil Fischer at the beginning of his peptide studies.[118] He hydrolysed ethoxycarbonylglycylglycine (XXII) with alkali and obtained a dicarboxylic acid, which to his surprise did not lose carbon dioxide and give glycylglycine. The simple carbamic acid formula was soon disproved



by esterification to an isomer of ethoxycarbonylglycylglycine ethyl ester, but it was not until 1928 that the mystery was solved.[119] The alkali catalyses condensation between the urethane and the amide group, and the resultant hydantoin (XXIII) is then hydrolysed. Wessely, Schlögl, and their colleagues [120] have used this reaction for structural work. It occurs under rather drastic conditions but the similar elimination of thiophenol from phenylthiocarbonyl peptides takes place very easily and the hydantoin can be isolated.[121]

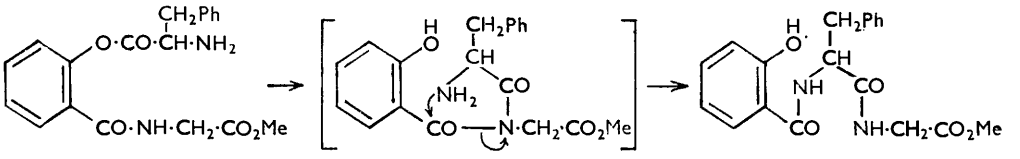
A remarkable rearrangement, which can be regarded as involving both types of reaction, has recently been discovered by Brenner *et al.*[122] The salicylic acid derivative (XXIV) easily isomerises to (XXVI). This becomes less mysterious of we remember that *O*-acetylsalicylamide easily rearranges to *N*-acetylsalicylamide.[123] The analogous rearrangement of (XXIV) would produce the diacylimide (XXV), in which the amino-group is very favourably placed for nucleophilic attack on the salicyl carbonyl group. This interpretation † may be too naïve, and

* [Added in proof.] The similarity of these four schemes is even greater than the above formulæ suggest. The *sulphur* atom first attacks the neighbouring amide group in the phenylthioureido-peptides (Edman, *Nature*, 1956, **177**, 667).

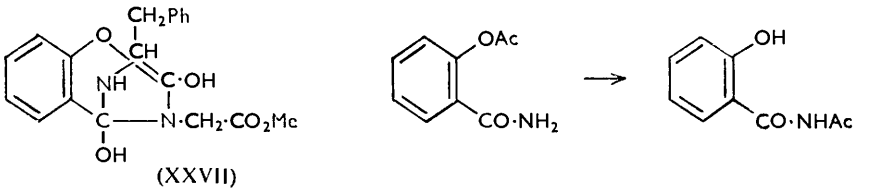
† The same interpretation has now been published by Wieland.[124]

[111] Kenner and Khorana, *J.*, 1952, 2076.
[112] Elmore and Toseland, *J.*, 1954, 4533.
[113] Levy, *J.*, 1950, 404; 1954, 4711.
[114] Léonis and Levy, *Compt. rend. Trav. Lab. Carlsberg*, 1954, **29**, 57, 87
[115] Schlack and Kumpf, *Z. physiol. Chem.*, 1926, **154**, 125.
[116] Kenner, Khorana, and Stedman, *J.*, 1953, 673.
[117] Goldschmidt and Wick, *Annalen*, 1952, **575**, 217.
[118] Fischer, *Ber.*, 1902, **35**, 1097; 1903, **36**, 2097.
[119] Wessely and Kemm, *Z. physiol. Chem.*, 1928, **174**, 306; Wessely, Kemm, and Mayer, *ibid.*, **180**, 64.
[120] Wessely, Schlögl, and Korger, *Monatsh.*, 1952, **83**, 1156; Wessely, Schlögl, and Wawersich, *ibid.*, p. 1426; Schlögl, Wessely, and Wawersich, *ibid.*, 1954, **85**, 957.
[121] Lindenmann, Khan, and Hofmann, *J. Amer. Chem. Soc.*, 1952, **74**, 476; Wessely, Schlögl, and Wawersich, *Monatsh.*, 1952, **83**, 1439.
[122] Brenner, Zimmerman, Wehrmüller, Quitt, and Photaki, *Experientia*, 1955, **10**, 397.
[123] Auwers, *Ber.*, 1905, **38**, 3256; McConnan and Titherley, *J.*, 1906, **89**, 1318; Anschütz, Aschenberg, Kuckertz, Krone, Riepenkröger, and Zerbe, *Annalen*, 1925, **442**, 18.
[124] Wieland, Lang, and Liebsch, *Annalen*, 1955, **597**, 227.

indeed Brenner prefers to consider (XXVII) as the intermediate, but these alternatives are actually quite closely related.



(XXIV) (XXV) (XXVI)



(XXVII)

Although I have far from exhausted the stock of peptide reactions, I have probably given sufficient examples to show that the amide group in simple peptide systems is reactive in suitable environments. We may therefore reasonably hope that the way in which the amide group reacts on certain enzyme surfaces may be explained in chemical terms during the next decade. In the last few years the main aims of peptide work have been different from those of the era dominated by Bergmann, but nevertheless the problems which interested him are not likely to be neglected.

I am very grateful to Sir Alexander Todd, F.R.S., who has most generously given me every opportunity and encouragement to develop my interest in peptides.