[CONTRIBUTION FROM THE VIRUS LABORATORY, UNIVERSITY OF CALIFORNIA, BERKELEY]

Determination of C-Terminal Amino Acids and Peptides by Hydrazinolysis[1]

BY CHING-I NIU AND H. FRAENKEL-CONRAT

RECEIVED JUNE 6, 1955

The hydrazinolysis method of Akabori and Ohno, followed by isolation of the C-terminal amino acids as DNP derivatives, has been applied to a number of proteins. In several instances it has been possible to isolate C-terminal peptides. With several previously studied proteins, results have been confirmatory. For tobacco mosaic virus, ovalbumin and ribonuclease, prolylalanylthreonine, valylserylproline and serylvaline, respectively, were found as C-terminal sequences.

Akabori and co-workers proposed hydrazinolysis as a method for the identification of C-terminal amino acids. Upon heating proteins with anhydrous hydrazine, all but the C-terminal amino acids were transformed into hydrazides which were separated as the benzal derivatives from the originally C-terminal and now free amino acids.[2] Ohno later made use of fluorodinitrobenzene (FDNB) and separation of acidic from neutral dinitrophenyl (DNP) products as a means of separating the free amino acids from the hydrazides.[3-5] In the present study a combination of the two procedures has given the best results. The DNP derivatives of the C-terminal amino acids were identified and determined by the two-dimensional chromatographic technique of Levy.[6,7] The qualitative and quantitative significance, and the limitations of this technique have been studied in some detail. The method also has been applied to a number of proteins. The scope of the method was further extended when it was found possible, in some instances, to isolate C-terminal peptides after incomplete hydrazinolysis.

Experimental

Materials.—Tobacco mosaic virus (TMV) was prepared by the standard method of differential centrifugation. Dethreoninated TMV was obtained by high-speed centrifugation of the virus (2%) after exposing it for several hours to 0.01% carboxypeptidase.[8] TMV protein was prepared with sodium dodecyl sulfate as previously described.[9] Dethreoninated TMV protein was prepared by detergent treatment of the dethreoninated virus. For the preparations of the masked and H.R. strain of TMV[8] we are indebted to C. A. Knight of this Laboratory. Samples of ovalbumin were kindly placed at our disposal by J. L. Stokes of the Western Regional Laboratory and by G. Perlman of the Rockefeller Institute for Medical Research. DFP-chymotrypsin was kindly supplied by E. F. Jansen of the Western Regional Research Laboratory. The other proteins were commercial preparations. The arginine peptides were kindly given to us by F. Carpenter, Biochemistry Department, the others are from the Emil Fischer collection. Hydrazine was distilled from calcium oxide and analyzed as suggested by Ohno.[4]

Procedure.—Samples of 0.2–0.4 μmole of protein were dried at 100° or *in vacuo* at 78°. To these was added 0.5 ml. of anhydrous hydrazine in a dry chamber and the tubes were sealed and heated for various time periods, usually 10 hr. at 100°. The excess hydrazine was removed in a vacuum desiccator containing concentrated sulfuric acid and the residue was dissolved in 1.0 ml. of water; about 0.2 ml. of benzaldehyde was added and the mixture shaken at room temperature for two hr. The benzaldehyde layer together with the condensation product of the hydrazides with benzaldehyde was removed by centrifugation. The clear supernatant which comprised all the amino acids and traces of amino acid hydrazides was then quantitatively transferred to a test-tube, diluted with 0.5 ml. of 6% $NaHCO_3$ solution, and 3.0 ml. of 2.5% FDNB in ethanol was added. The dinitrophenylation was carried out at room temperature for 80 minutes with shaking. The reaction mixture was then diluted with 20 ml. of water, acidified with 2 *N* HCl, and extracted with 2 × 20 ml. of ethyl acetate. The ethyl acetate solution was then diluted with two volumes of ether[10] and extracted with 2 × 50 ml. of 2% $NaHCO_3$ and the $NaHCO_3$ solution washed once with 20 ml. of ethyl acetate; it was then acidified and extracted with 2 × 15 ml. of ethyl acetate. The solvent was removed *in vacuo*, and from the residue dinitrophenol was sublimed at 60° below 1 mm. with a cold finger containing ice. The non-volatile sample was then dissolved in ethyl acetate and quantitatively transferred to Whatman No. 1 paper, for two-dimensional paper chromatography, according to Levy.[6]

Identification, Estimation and Correction Factors.—After elution with 1% bicarbonate the amount of DNP-amino acid in each spot was estimated spectrophotometrically. Readings of the O.D. were taken at 360 and 390 mμ. Approximate, uncorrected amounts of all DNP-amino acids other than prolines were calculated from the readings at 360 mμ, using 17,000 as the molar extinction coefficient. Proline and hydroxyproline were calculated from the readings at 390 mμ. The identity of each spot was generally established by its position in the two-dimensional pattern.[6] Further confirmation was obtained through reconversion to the free amino acid by hydrolysis in ammonia[11] and chromatographic identification in butanol–acetic acid–water (4:1:5) or buffered phenol–cresol[12] for all the amino acids except leucine and isoleucine, which, in turn, were resolved with *t*-amyl alcohol–*p*H 6 phthalate in an extended run, also according to Levy.[7]

The ratio of O.D. readings at 360 and 390 mμ was an additional guide to the nature of many spots, as indicated on Table I. DNP-Proline was further characterized by the ease of its hydrolysis with acid, and identification of the free amino acid by chromatography and detection with isatin.[13] The presence of peptides, suggested by the chromatographic position and O.D. ratio (Table I) of unknown spots, was proven by chromatography after acid hydrolysis (5.7 *N* HCl, 5–8 hr. reflux). The N-terminal amino acid was identified as the ether-soluble DNP derivative[6] unless it was proline, and the water-soluble components of the peptide, including N-terminal proline, were characterized chromatographically, and with selective spray reagents whenever possible. Their occurrence in stoichiometric amounts could be ascertained by analysis as DNP derivatives.[6] Of general interest appears the case of ovalbumin, where the single DNP peptide spot observed besides DNP-proline

(1) Aided by grants from The National Foundation for Infantile Paralysis, Lederle Laboratories, and The Rockefeller Foundation.

(2) S. Akabori, K. Ohno and K. Narita, *Bull. Chem. Soc., Japan*, **25**, 214 (1952).

(3) K. Ohno, *J. Biochem. (Japan)*, **40**, 621 (1953).

(4) K. Ohno, *ibid.*, **41**, 345 (1954).

(5) S. Akabori, K. Ohno, T. Ikenaka, A. Nagata and I. Haruma, *Proc. Japan Acad.*, **29**, 561 (1953).

(6) A. L. Levy, *Nature*, **174**, 126 (1954).

(7) H. Fraenkel-Conrat, J. I. Harris and A. L. Levy, in "Methods of Biochemical Analysis," Vol. II, Ed. Glick, 1955, p. 375.

(8) J. I. Harris and C. A. Knight, *Nature*, **170**, 613 (1952); *J. Biol. Chem.*, **214**, 215 (1955); C. A. Knight, *ibid.*, **214**, 231 (1955).

(9) H. Fraenkel-Conrat and B. Singer, THIS JOURNAL, **76**, 180 (1954).

(10) The addition of ether facilitates the extraction of di-DNP-tyrosine into bicarbonate. It was not always done, but is recommended when tyrosine is present.

(11) A. G. Lowther, *Nature*, **167**, 767 (1951).

(12) A. L. Levy and D. Chung, *Anal. Chem.*, **25**, 396 (1953).

(13) R. Acher, C. Fromageot and J. Jutisz, *Biochim. Biophys. Acta*, **5**, 81 (1950).

proved to contain two DNP amino acids after hydrolysis; the unhydrolyzed material was thereupon resolved into two peptides by chromatography according to Blackburn,[14] which were composed of DNP-valine, serine and proline and DNP-serine and proline, respectively.

TABLE I

ABSORPTION CHARACTERISTICS OF DIFFERENT DNP COMPOUNDS

	Ratio of O.D. at 390/360 mμ
DNP-proline	1.40
DNP-prolyl peptides	1.05
DNP-hydroxyproline	1.20
Dinitrophenol	0.80
Other DNP-amino acids	0.60–0.65
Di-DNP-ornithine	0.55–0.60
Di-DNP-lysine	0.55–0.60
Di-DNP-tyrosine	0.55–0.60
Di-DNP-cystine	0.55–0.60
Other DNP-peptides	0.50–0.55

For the purpose of quantitative evaluation of the data obtained with our procedure, it was necessary to determine to what extent known amounts of amino acids were recovered after treatment with hydrazine, benzaldehyde, FDNB, chromatography, etc. To this end various mixtures were analyzed. Most amino acids were recovered in 40 to 60% yield (Table II). Considering the number of steps and manipulations involved, many of these values are not unexpectedly low. In confirmation of Locker's results,[15] cystine and cysteine were lost during hydrazinolysis, and methionine was converted to methionine sulfoxide when benzaldehyde treatment was involved. Arginine, however, was converted to ornithine in moderate yield. Aspartic acid and glutamic acid were recovered in rather poor yield. Asparagine and glutamine were largely destroyed, and what little remained was lost in the hydrazide fraction under the present scheme of isolation.

TABLE II

RECOVERY OF AMINO ACIDS AS DNP-DERIVATIVES AFTER HYDRAZINOLYSIS (10 HR. PERIOD)[a]

	Average	Range		Average	Range
Alanine	0.45	0.38–0.53	Lysine	0.47	0.35–0.56
Arginine[b]	.25	.18– .32	Methionine[c]	.56	.50– .61
Aspartic acid	.19	.13– .25	Phenylalanine	.59	.54– .64
Glutamic acid	.16	.11– .23	Proline	.52	.43– .61
Glycine	.37	.25– .47	Serine	.39	.34– .51
Histidine	.43	.31– .47	Threonine	.52	.45– .59
Hydroxyproline	.50	.43– .51	Tryptophan	.39	.30– .50
Isoleucine	.54	.49– .60	Tyrosine	.44	.35– .49
Leucine	.54	.46– .58	Valine	.76	.72– .82

[a] DNP amino acids were estimated using 17,000 as molar extinction coefficient; no corrections were applied. The recovery factors for threonine and alanine reported in the previous paper[26] differ from the present ones in that a molar extinction coefficient of 15,600 was assumed in the calculation. Recoveries were also slightly higher without the benzaldehyde treatment, but cleaner chromatograms are obtained under the present conditions. Recoveries after 24 hr. of hydrazine treatment were in general about 20% lower, except for the dicarboxylic amino acids which were largely destroyed. [b] Converted to and estimated as di-DNP-ornithine. [c] Recovered as DNP-methionine sulfoxide when benzaldehyde treatment was involved.

The recovery factors for all amino acids as listed on Table II are the mean values from 4 to 10 experiments in the presence or absence of equimolar protein. The values vary to a certain extent with each experiment. The most reliable recovery value for a given C-terminal amino acid is probably obtained by parallel hydrazinolysis of the protein alone and with the addition of an equimolar amount of the amino acid in question. The recovery of amino acids added to proteins prior to hydrazinolysis was not consistently or appreciably different from that of the amino acids alone.

The average recovery factors, as listed on Table II, have been used in calculating the amounts of C-terminal amino acid in proteins from the observed O.D. of the spot eluates (Table III). This should automatically correct for all losses, as well as differences in absorption peaks, molar extinction coefficients, etc. However, no correction is possible for the effect, be it protective or harmful, of the peptide linkage from which the C-terminal amino acid is only gradually released. The fact that close to stoichiometric amounts have been obtained in most instances suggests that this is not generally an important source of error. Hydrazinolysis of alanyldiglycine, alanylglycine, leucylalanylglycine, arginylleucine and arginylglutamic acid yielded 0.39, 0.39, 0.39, 0.64 and 0.25 equivalents of the expected C-terminal amino acid, respectively. Thus it appears that recoveries of C-terminal amino acids may be at times slightly higher than those listed for the free amino acids.

Results and Discussion

Most of our present knowledge of N-terminal sequences of proteins is based on two independent chemical methods,[16,17] and was recently confirmed in a few instances by complete structural elucidation.[18–20] In contrast, our knowledge of C-terminal groups and sequences is almost entirely based on results obtained with carboxypeptidase.[7] This technique appears reliable when one is dealing with a single peptide chain; however, even then a definite sequential allocation is not always possible when several amino acids are released at similar rates.[21] Negative evidence also is of little value.

The need for a reliable chemical method for the identification of C-terminal amino acids and sequences is evident. Three main reactions have been intensively studied in recent years. The final verdict of some investigators of the reduction methods have been pessimistic[22,23]; the thiocyanate method has given poor results in the hands of most investigators.[22,24] Hydrazinolysis, in combination with the FDNB method, appeared to represent the most promising approach.[25] This hope appears borne out by the results obtained in the present study. Analysis of a few di- and tripeptides yielded qualitatively and quantitatively good results. From the recovery results previously discussed, it is evident that the presence of C-terminal cystine, cysteine, asparagine or glutamine cannot be ruled out under our present experimental conditions. This limitation on the validity of negative evidence is, however, more circumscribed and predictable, than that inherent in the carboxypeptidase method. The most important question is that concerning the validity of positive findings. If non-terminal amino acids could appear free in the hydrazinolysis mixture,

(14) S. Blackburn and A. G. Lowther, *Biochem. J.*, **48**, 126 (1951).
(15) R. H. Locker, *Biochim. Biophys. Acta*, **14**, 533 (1954).
(16) F. Sanger, *Biochem. J.*, **39**, 507 (1945); **45**, 563 (1949).
(17) P. Edman, *Acta Chem. Scand.*, **4**, 277, 283 (1950).
(18) F. Sanger and H. Tuppy, *Biochem. J.*, **49**, 481 (1951).
(19) F. Sanger and E. O. P. Thompsen, *ibid.*, **53**, 353 (1953).
(20) P. H. Bell, THIS JOURNAL, **76**, 5565 (1954).
(21) C. B. Anfinsen, R. R. Redfield, W. E. Cholate, J. Page and W. R. Carroll, *J. Biol. Chem.*, **207**, 201 (1954).
(22) M. Jutisz, *Bull. soc. chim. biol.*, **36**, 109 (1954).
(23) A. C. Chibnall and M. W. Rees, in "The Chemical Structure of Proteins," Ciba Symposium (Dec. 1952), J. and A. Churchill Ltd., London, 1953, p. 70.
(24) V. H. Baptist and H. B. Bull, THIS JOURNAL, **75**, 1727 (1953).
(25) A similar attack is by ammonolysis as proposed by Chambers and Carpenter (*ibid.*, **77**, 1527 (1955)), but more side reactions appear to occur at the high temperature used for ammonolysis.

TABLE III

C-TERMINAL AMINO ACIDS AND PEPTIDES OF PROTEINS

Protein[a]	Reaction time, hr.	Terminal amino acid Identity	Equiv./mole[b]	Terminal peptide Identity	Equiv./mole[b]
TMV (common strain) (18,000)	10	thr	1.0	....	
TMV protein (18,000)	5	thr	0.8	....	
....	10	thr	1.0	....	
....	22	thr	1.0	....	
TMV protein dethreoninated[c] (18,000)	4	ala	0.1	pro-ala	0.5
	5	ala	.4	pro-ala	.3
	10	ala	.9	pro-ala	.1
TMV (masked strain) dethreoninated[c] (18,000)	4	ala	0.2	pro-ala	.4
	10	ala	1.0	pro-ala	.1
TMV (H.R. strain) dethreoninated[c] (18,000)	4	ala	0.3	....	
	10	ala	1.0	....	
TMV (H.R. strain) protein dethreoninated[c] (18,000)	4	ala	0.4	....	
	10	ala	0.9	....	
Insulin[d] (6000)	10	ala	1.1	....	
Chymotrypsin, DFP-(23,000)	10	tyr	0.8	....	
		leu	1.1	....	
β-Lactoglobulin (40,000)	10	ileu	2.1	....	
Salmine (7000)	10	arg	0.9	....	
Bovine serum albumin (69,000)	10	ala	1.1	....	
Ovalbumin (45,000)	6	pro	0.7	val-ser-pro, ser-pro	.3
	10	pro	0.8	val-ser-pro, ser-pro	.2
	16	pro	1.1	trace of peptide	
	24	pro	0.9	trace of peptide	
Ribonuclease (13,400)	4.5	val	.4	ser-val	.4
	10	val	.8	ser-val	.2
	24	val	1.0	trace of peptide	
Soybean trypsin inhibitor (25,000)	10	ser	0.5	....	
....		leu	.3	....	
....	24	ser	.4	....	
....		leu	.4	....	

[a] Other protein samples tested but found not to yield C-terminal amino acids: ovomucoid, chymotrypsinogen, trypsin and cytochrome c. For pepsin see text. Values in parentheses represent the molecular or subunit weights used in the calculations. [b] Averages of several runs, calculated with the correction factors of Table II, using those of the N-terminal amino acid in the case of peptides. These values are not strictly quantitative, owing to the high correction factors that had to be applied as well as to the unknown differences in the rate of release of various amino acids under the influence of hydrazine. [c] By means of carboxypeptidase. [d] Once a small amount of aspartic acid was detected. Asparagine, the other known C-terminal amino acid would not be recovered by our technique (see text). [e] Estimated as diDNP-ornithine.

the method would lose most of its usefulness. Empirically this was found not to be the case. Out of more than 15 proteins investigated, only two showed unexpected C-terminal amino acid, and in most cases no other amino acids were detectable in even small amounts (0.1 equiv./mole). When the unexpected proline end group was detected in ovalbumin, it appeared important to exclude its formation through hydrolysis of a particularly hydrophilic peptide linkage; however, proline was found present in the same amount even when the protein and the hydrazine were dried most vigorously. Furthermore, addition of water up to 10% to the hydrazine did not cause the appearance of spurious end groups in other proteins. It thus appears very probable that the observed free amino acids are C-terminal, but absolute proof will have to await more extensive structural elucidation by a variety of methods.

It seems that this is the first instance where hydrazinolysis has supplied definite sequential information for a number of proteins. From certain proteins secondary yellow spots were obtained upon chromatography which were found in amounts inversely related to both hydrazinolysis time and amount of C-terminal amino acid. This suggested that they were peptides, a supposition which was supported by their showing absorption maxima at lower wave lengths than the corresponding C-terminal DNP-amino acid. Final proof of their nature came from chromatographic analysis of the hydrolysates of these spots. The yellow ether extract of such hydrolysates contained the DNP-N-terminal amino acid of the peptide, while the aqueous phase contained the C-terminal amino acid of the protein and in the case of a tripeptide also the penultimate amino acid. In the case of a DNP-prolyl peptide, the proline, as expected, appeared in the aqueous phase. Hydrazinolysis thus has proven more instructive than expected. However, it must be stressed that the survival of C-terminal peptides will depend on their relative labilities toward hydrazine, and that no systematic sequence study can be based on this reaction.

The C-terminal amino acids found for different proteins are listed on Table III. Preliminary results with TMV protein have been previously published.[26] It appears noteworthy that only

(26) C.-I. Niu and H. Fraenkel-Conrat, *Biochim. Biophys. Acta*, **16** 597 (1955).

threonine could be obtained from the original virus or virus protein; and that only after enzymatic release of the threonine was it possible to obtain the new terminal amino acid and peptide. The apparent homogeneity of the sub-units comprising the virus, at least as far as the C-terminal sequence is concerned, is noteworthy, since it is a necessary requirement for further structural elucidation. It is also of interest that the masked strain which is closely related to TMV has the same C-terminal sequence. However, the dethreoninated H.R. strain failed to yield the prolylalanine peptide. Since all of the other dethreoninated TMV proteins gave rise to the peptide on partial hydrazinolysis, the failure of the H.R. strain to do so represents indirect but definite evidence for a structural difference. This difference may be at the fourth amino acid since the resistance of this as all other dethreoninated TMV proteins to carboxypeptidase suggests the presence of the same pro-ala- sequence.

The results with ovalbumin already have been mentioned. The same results were obtained with four different samples, prepared by two procedures.[27] Since no N-terminal amino acid has ever been detected in ovalbumin, a six-shaped structure appears probable. The attack of subtilisin[28] (and contaminated carboxypetidase) appears to be on a particularly labile site in the ring, while the actual C-terminal proline is disregarded by the enzyme for specificity reasons.

Our finding of C-terminal alanine in bovine serum albumin has since been anticipated in the literature.[29] A peptide was obtained in too low yield for identification. With ribonuclease valine was found C-terminal, in accord with results obtained with carboxypeptidase[21]; however, the next amino acid appears to be serine, contradictory to that study. For chymotrypsin we have confirmed Neurath's conclusion[30,31] that this represents a two-chain protein. β-Lactoglobulin also appears to be a two-chain protein, but both chains seem to terminate in isoleucine and the histidine obtained after carboxypeptidase treatment[31] must represent a penultimate amino acid.[32]

No terminal amino acids were found in chymotrypsinogen and ovomucoid. In the latter, phenylalanine appeared C-terminal by the reduction[33] and thiocyanate[34] methods, and nothing is released by carboxypeptidase. From crystalline soybean trypsin inhibitor serine and leucine were obtained in non-stoichiometric amounts, as well as traces of other amino acids. Carboxypeptidase released only leucine stoichiometrically.[35] Several preparations of pepsin, as well as of denatured and dialyzed pepsin, contained free amino acids when tested by dinitrophenylation. Therefore the complex results of hydrazinolysis are of no significance; however, the absence of stoichiometric amounts of alanine is in contrast to the conclusions of others.[36]

(27) Akabori[5] had found C-terminal alanine, and this question is being further investigated through an exchange of samples. We have recently found the same C-terminal sequence in a preparation kindly sent to us by Prof. Akabori and have been informed by Dr. Ohno that they have now also indentified the C-terminal amino acid as proline.

(28) D. Steinberg, THIS JOURNAL, **74**, 4217 (1952).

(29) W. F. White, J. Shields and K. C. Robbins, *ibid.*, **77**, 1267 (1955).

(30) J. A. Gladner and H. Neurath, *J. Biol. Chem.*, **206**, 911 (1954).

(31) H. Neurath, J. Gladner and E. Davie, in "The Mechanism of Enzyme Action," Ed. McElroy and Glass, Baltimore, Md., 1954, p. 51.

(32) The absence of histidine was also demonstrated, after direct chromatography of the hydrazinolysis residue, by spraying with the very sensitive diazobenzenesulfonic acid reagent.

(33) L. M. Penasse, M. Jutisz, C. Fromageot and H. Fraenkel-Conrat, *Biochim. Biophys. Acta*, **9**, 551 (1952).

(34) R. A. Turner and G. Schmerzler, *ibid.*, 11, 586 (1953).

(35) E. W. Davie and H. Neurath, *J. Biol. Chem.*, **212**, 507 (1955).

(36) M. B. Williamson and J. M. Passmann, *Biochim. Biophys. Acta*, **15**, 246 (1954).

BERKELEY, CAL.

[CONTRIBUTION FROM THE LABORATORY OF CHEMISTRY OF NATURAL PRODUCTS, NATIONAL HEART INSTITUTE, NATIONAL INSTITUTES OF HEALTH, PUBLIC HEALTH SERVICE, U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE]

Alkaloids of the Amaryllidaceae. VI. The Action of Oxidizing Agents on Lycorine and Caranine[1]

BY H. M. FALES, E. W. WARNHOFF AND W. C. WILDMAN

RECEIVED MAY 14, 1955

Oxidation of lycorine with selenium dioxide, *t*-butyl hypochlorite, pyridine perbromide hydrobromide or N-bromosuccinimide has been shown to yield a compound $C_{16}H_{11}NO_3$ for which structure IIa is suggested. A modified Oppenauer oxidation of caranine affords an isomeric compound for which structure IIIa is proposed. Oxidation of lycorine with mercuric acetate has been shown to yield a mixture of IIa and IIIa. The structures of IIa and IIIa are supported by the ultraviolet spectra of synthetic model compounds. A partial formula for caranine is proposed.

Although the principal alkaloid of the Amaryllidaceae, lycorine, has been known for sixty years, the most valuable evidence for its structure was published last year through the joint efforts of research groups in Japan and New Brunswick.[2] In the currently accepted formula for lycorine (I), one of the two alcoholic groups is placed in an allylic position. With the advent of manganese dioxide as a selective oxidizing agent for allylic alcohols, it seemed desirable to study the action of this reagent upon lycorine and other alkaloids of this family. Unfortunately, the insolubility of lycorine in the common solvents under the normal reaction conditions made impossible a direct comparison with other allylic alcohols. However, when lycorine was extracted

(1) Previous paper, W. C. Wildman and C. J. Kaufman, THIS JOURNAL, **77**, 4807 (1955).

(2) L. G. Humber, *et al.*, *J. Chem. Soc.*, 4622 (1954).