THE NATURE OF THE IONIZABLE GROUPS IN PROTEINS.

By HENRY S. SIMMS.

(*From the Department of Animal Pathology of The Rockefeller Institute for Medical Research, Princeton, N. J.*)

(Accepted for publication, February 17, 1928.)

I.

Dissociation Indices of Amino Acids and Peptides.

In Table I will be seen values of the dissociation indices[1] of various divalent amino acids and peptides (*i.e.*, having one free α-amino group and one free α-carboxyl group).

In Table II are the dissociation indices of some trivalent amino acids and peptides which, in addition to the α-amino group and α-carboxyl group, have a third ionizable group.

The various theories of protein structure now under discussion are all in agreement with the assumption that one amino group and one carboxyl group of each amino acid in a protein molecule are bound in some manner which prevents them from ionizing.[2]

Hence we would expect the ionizable groups[3] in the protein molecule

[1] The term "index" is used here, as previously defined, to indicate the negative logarithm of a value; *e.g.*, "dissociation index" = pK = $-\log K$, or "titration index" = pG = $-\log G$.

[2] The "chain" theory, however, assumes that in each protein molecule one α-amino group and one α-carboxyl group are free; but such an assumption is unwarranted by experimental accuracy even if the chain theory were known to be correct and the chains had no branches. In gelatin it would raise the predicted sum of groups from 4.00 to 4.08 which would be in better agreement with the experimental value of 4.05. In egg albumin it would raise the basic groups from 1.6 to 1.75 which would better agree with the experimental value of 2.1. But despite the better agreement it is preferable to simplify our assumptions and attribute the ionizable groups to only the extra groups of trivalent amino acids.

[3] There is still a school of chemists who believe that the charges on a protein molecule are not due to the ionizable groups but to its charge as a colloidal particle. In this article, however, we feel justified in making the unqualified assumption that the charges *in dilute solution* are due to free ionizable groups.

to be due to the "extra" groups in the trivalent amino acids (Table II). It does not follow, however, that the bound groups are always the α groups. For instance in lysine we can conceive of the α-carboxyl and the ϵ-amino group being bound, leaving the α-amino group free to ionize. The titration index will be different than if the α-amino group were bound and the ϵ-amino group were free. However, the data indicate that the latter structure prevails. The same is true of arginine and histidine; the α-amino groups are bound and the extra basic groups are free (in so far as they exist in that form, see below).

TABLE I.

Titration Indices of Divalent Amino Acids and Peptides (and Related Monovalent Electrolytes).

Substance	pG_1' (– COOH)	pG_2' (– NH_2)	Author*
Acetic acid	$4.740-0.9a\sqrt{\mu}$		S.
Glycollic acid	3.82		S.
Amino ethanol		$9.470+0.5a\sqrt{\mu}$	S.
Glycine ethyl ester hydrochloride		7.655	S.
Glycine	$2.365-0.08a\sqrt{\mu}$	$9.715-a\sqrt{\mu}$	S.
Alanine	2.35	9.72	L.S.
Sarconine	2.23	10.01	L.S.
Tryptophane	2.266	9.372	S.
Valine	2.28	9.65	L.B.
Glycyl-glycine	3.12	8.07	L.S.
Sarcosyl-glycine	3.10	8.51	L.S.
Glycyl-sarconine	2.83	8.54	L.S.
Sarcosyl-sarconine	2.86	9.10	L.S.
Alanyl-alanine	3.17	8.42	L.S.
Glycyl-valine	2.28	8.30	L.B.
Glycyl-leucine	3.18	8.29	S.
Glycyl-alanine	3.15	8.25	S.
Glycyl-asparagine	2.9	8.3	S.
Glycyl-glycyl-glycine	3.26	7.91	L.S.
Glycyl-alanyl-alanyl-glycine	3.30	7.9	L.S.
Alanyl-alanine anhydride (enol group = 13.5)			L.B.

* L.S. represents Levene and Simms; L.B., Levene and Bass; and S., Simms.

TABLE II.

Titration Indices of Trivalent Amino Acids.

Substance	pG′ $\alpha-COOH$	pG′ $\alpha-NH_2$	Third group		Approximate pG′ in protein	Author*
			Group	pG′		
Aspartic acid	$2.05-0.6a\sqrt{\mu}$	$10.00-2.3a\sqrt{\mu}$	$-COOH$	$3.87-1.2a\sqrt{\mu}$		S.
Glycyl aspartic acid	2.81	8.60	$-COOH$	4.45	3.5	S.
Aspartyl glycine	2.10	9.07	$-COOH$	4.53		L.S.
Glutamic acid	2.11	9.45	$-COOH$	4.06		L.S.
Histidine	1.46	9.41	$=NH$	6.06	6.1	S.
Arginine	2.29	9.64	$=NH$	8.15	8.1	S.
Tyrosine	2.24	10.28	$-OH$	9.21	9.4	S.
Lysine	2.04	9.06	$-NH_2$	10.45	10.6	L.S

*L.S. represents Levene and Simms; and S., Simms.

TABLE III.

Content of Amino Acids in Gelatin.

Amino acid		Weight	Molecular weight	Mols per 61,500 gm.	Mols per 2500 gm.
		per cent			
Divalent	Glycine	25.5	75	209	8.5
	Alanine	8.7	89	60	2.4
	Leucine	7.1	131	33	3.4
	Serine	0.4	105	2	0.1
	Phenylalanine	1.4	165	5	0.2
	Proline	9.5	115	51	2.1
	Oxyproline	14.1	131	66	2.7
Total divalent		66.7		426	17.4
Trivalent	Aspartic acid	3.4	133	16	0.64
	Glutamic acid	5.8	147	24	1.0
	Histidine	0.9	155	4	0.15
	Arginine	8.2	174	29	1.2
	Tyrosine	0.01	181	0.03	0
	Lysine	5.9	146	25	1.0
Total trivalent		24.2		98	4.0
Total amino acids		90.9		525	21.4

II.

The Predicted Ionizable Groups of Proteins.

In Table III are the percentages of various amino acids in gelatin as found by Dakin.[4] Since we are interested only in the trivalent amino acids these are listed in Table IV in the column of "predicted" equivalents (per 2500 gm., an arbitrary unit weight). The third column gives the predicted values for gelatin and the eighth column for egg albumin.

TABLE IV.

Equivalents (per 2500 gm.) of the groups found, compared with the equivalents predicted from the content of the respective amino acids. The values for gelatin must be multiplied by 24.6 to obtain the equivalents per molecular weight of 61,500 (Kunitz); and the egg albumin values should be multiplied by 13.5 for molecular weight of 33,800.

1		2	3	4	5	6	7	8	9	10
			Gelatin					Egg albumin		
				1st titration		2nd titration				
Groups	Sources	Approximate indices*	Predicted	Found	Difference	Found	Difference	Predicted	Found	Difference
Acidic	Dicarboxylic acids	$pG_1' = 3.5$*	1.65	1.75	+0.1	1.75	+0.1	?	1.6	?
	Tyrosine	$pG_5' = 9.4$	0	0	0	0	0	0.6	0.4	−0.2
Basic	**Unknown**	$pG_2' = 4.6$	**0**	**1.0**	**+1.0**	**1.05**	**+1.05**	**0**	**0.9**	**+0.9**
	Histidine	$pG_3' = 6.1$	0.15	0.2	0.05	0.15	0	0.3	0.3	0
	Arginine	$pG_4' = 8.1$	**1.2**	**0.2**	**−1.0**	**0.2**	**−1.0**	**0.7**	**0.3**	**−0.4**
	Lysine	$pG_6' = 10.6$*	1.0	0.9	−0.1	0.9	−0.1	0.6	0.6	0
Total acid groups ("base-binding capacity")			1.65	1.75	+0.1	1.75	+0.1	?	2.0	?
Total basic groups ("acid-binding capacity")			2.35	2.30	−0.05	2.30	−0.05	1.6	2.1	+0.5
Sum of **arginine** + **"4.6 group"**			**1.2**	**1.2**	**0**	**1.25**	**0.05**	**0.7**	**1.2**	**0.5**

* The titration indices (pG′ values) found, were those given above except in three cases: pG_1' in egg albumin is 2.9 (instead of 3.5). pG_6' in egg albumin was 10.8, and in the first titration of gelatin was 10.4 (instead of 10.6 as found in the second titration).

[4] Dakin, *J. Biol Chem.*, 1920, xliv, 499.

In Fig. 1, Curve *B*, we see the way the titration curve of gelatin would look if these groups occurred in gelatin in the quantities "predicted" by the amino acid content. The isoelectric point would be around pH 8.0 instead of 4.7. The experimental curve (*A*) of gelatin has an isoelectric point at 4.7 and has a different shape between pH 4 and 10.

We have analyzed the experimental curve of gelatin (Curve *A*, Fig. 1) in order to see if it could be resolved into groups with the

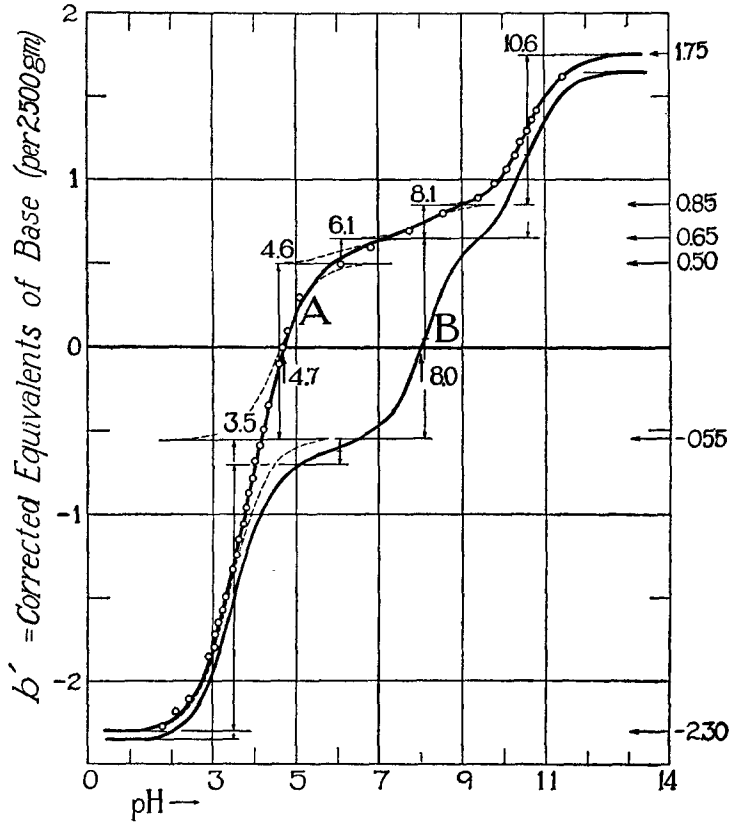

FIG. 1. Gelatin titration curves. Curve *A*. Experimental points and curve corresponding to values marked "found" in Table IV. Curve *B*. Curve "predicted" from trivalent amino acids (see Table IV).

indices given in Table IV. Our method was similar to that of E. J. Cohn[5] except that we restricted ourselves to titration indices which the free ionizable groups should have in the protein molecule, and furthermore we did not consider the isoelectric point as a transition between one ionizable group and another. It has been shown by Northrop[6] that gelatin contains at least two constituents not chemically bound. It is probable that gelatin and other proteins consist of mixtures of numerous constituents. In speaking of the "gelatin molecule" we recognize that we are dealing with a mixture of different molecules. This is kept in mind in the construction of the "predicted" curve and the analysis of the experimental curve. The isoelectric points do not represent transition points between ionic species as is the case with pure chemical substances having groups close together.

III.

The Actual Groups of Proteins.

The analysis of the experimental gelatin curve (Fig. 1, Curve *A*, and Columns 4 and 6 of Table IV) shows that the total number of acid groups agrees within 0.1 equivalent (per 2500 gm.) with the value predicted from the content of acidic amino acids.

The total number of basic groups agrees (within 0.05 equivalent) with the predicted value.[7] However, the quantity of arginine is a whole equivalent (1.0 eq.) too low in quantity. Furthermore, *the deficiency in arginine is compensated by the existence in gelatin of one equivalent of a basic group of unknown source with a titration index* (pG'_3) *at 4.6.*

A glance at the last three columns of Table IV shows that the same is true of egg albumin (see Fig. 2). Arginine group is 0.4 equivalent

[5] Cohn, *Physiol Rev.*, 1925, v, 349.

[6] Northrop, *J. Gen. Physiol.*, 1926–27, x, 161.

[7] This agreement is not found if we use the content of histidine found by Van Slyke (0.5) instead of the value of Dakin (0.15). The latter value agrees with the amount of histidine group (0.2) in gelatin. However, the relation between arginine and 4.6 group is not affected by the value assumed for histidine.

too low, while there is 0.9 equivalent of the "4.6 group" in egg albumin.[8]

The nature of the "4.6 group" is unknown. It is undoubtedly a basic (amino) group. The aliphatic amino groups have indices which range from 12.0 for diethyl amine to 7.65 for glycine ester hydrochloride. Unsaturated bases are weaker: guanidine has a basic group at 6.0; arginine at 8.15, and histidine at 6.06. However, bases as weak as 4.6 are not found except in the very unsaturated systems of aniline (4.6), cytosine (4.6), and isocystosine (4.0). The latter bases have an amino group attached to a conjugated unsaturated cyclic system.

Hence we are safe in concluding that a basic group with a conjugated unsaturated (and perhaps cyclic) structure occurs in gelatin

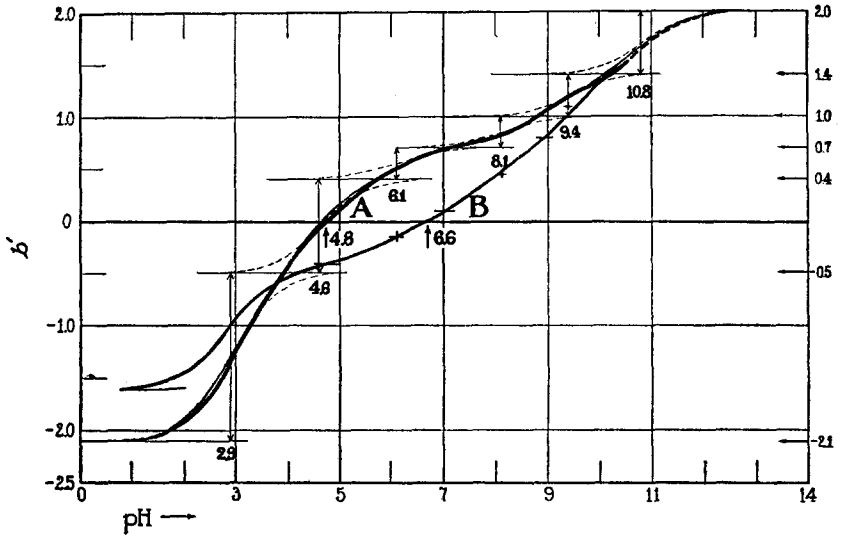

FIG. 2. Egg albumin titration curves. Curve *A*. Experimental curve (heavy line) and "found" curve (light line, see Table IV). Curve *B*. Curve "predicted" from trivalent amino acids.

[8] It would be interesting to know if the discrepancy of 0.5 eq. in the sum of arginine plus 4.6 group in egg albumin is due to an error in the arginine value found on hydrolysis (Column 8). In a subsequent paper it will be shown that the 2.2 eq. of arginine derived from edestin exist as "prearginine."

and egg albumin and that this group is disrupted on hydrolysis. It seems likely that this 4.6 group produces arginine on hydrolysis.

It will be noted that the agreement between the deficiency of arginine and the amount of the 4.6 group is quantitative in the case of gelatin while there is a discrepancy of +0.5 equivalent for egg albumin. Hence we must consider the alternative hypothesis, namely, that the 4.6 group does not produce arginine on hydrolysis and that part of the arginine group in the protein molecule is bound in some unknown manner so that it does not ionize. This seems unlikely, however and does not agree with the data on edestin.[8]

Since the material in proteins which dissociates as a weak base at pH 4.6 appears to give arginine on hydrolysis, we will refer to it as "*prearginine.*"

IV.

Deaminized Gelatin.

Hitchcock[9] showed that the loss of nitrogen on deamination of gelatin agreed with the decrease in acid-combining capacity. It seemed desirable to determine which amino groups in gelatin are removed on deamination. We therefore prepared some deaminized gelatin and titrated it.

If we compare the curve of deaminized gelatin in Fig. 3*A* with that of gelatin it is obvious that the process of deamination has principally removed the free lysine group and that the prearginine, the arginine, and the histidine groups are not materially affected.

In order to analyze the data more accurately we drew the comparison curve,[10] C_1 in Fig. 3*B*. The drop at 10.6 indicates that there is less of the lysine group in the deaminized gelatin. Adding 0.75 equivalent of 10.6 group (and lowering the curve 0.75 equivalent) gives the second comparison curve, C_2.

This second curve (C_2) represents the relation between deaminized gelatin and gelatin deprived of 0.75 equivalent of its lysine group.

[9] Hitchcock, *J. Gen. Physiol.*, 1923–24, vi, 95.

[10] Simms and Levene, *J. Biol. Chem.*, 1926, lxx, 319; Levene and Simms, *J. Biol. Chem.*, 1926, lxx, 327. This method could have been used in analyzing the data in Fig. 1, but was not adopted since the exact pG′ values were uncertain.

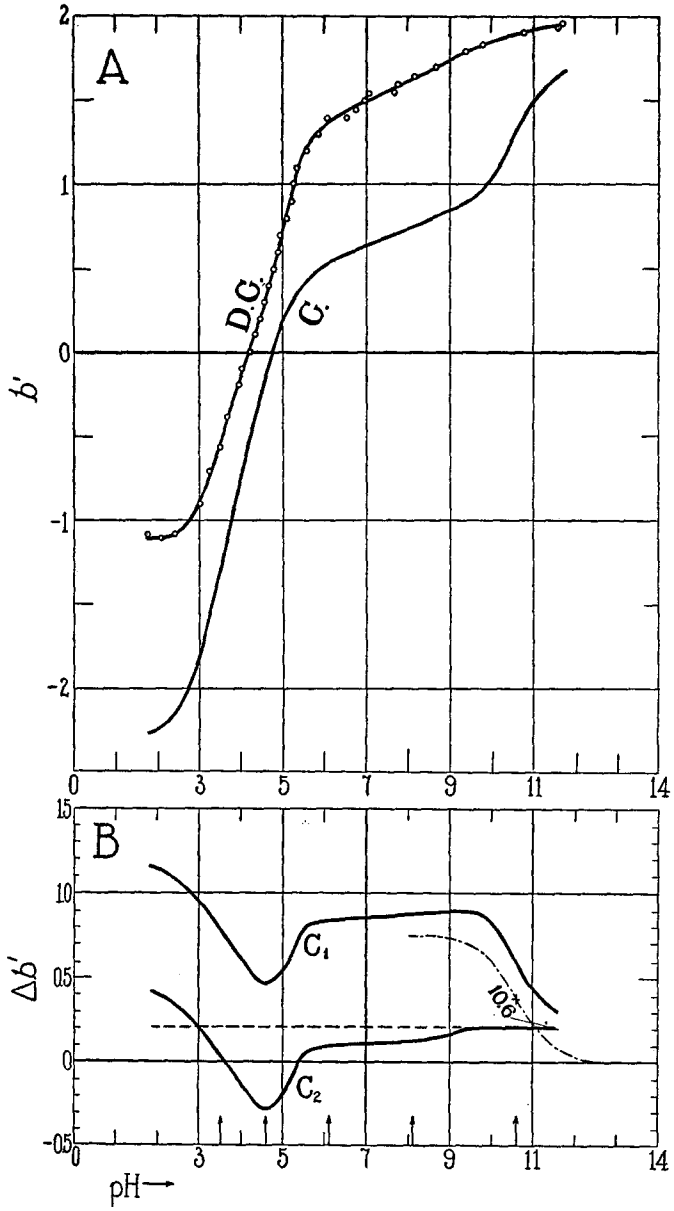

FIG. 3. *A*. Titration curves of gelatin (*G.*) and deaminized gelatin (*D.G.*). *B*. First comparison curve, C_1 (equals *D.G.* minus *G.*); and second comparison curve, C_2 (equals C_1 minus 0.75, plus 0.75 equivalent of 10.6 group).

The drop at 4.6 shows that the index of this group is materially raised (to, say, 5.0 or 5.1) on deamination. There is also a rise of the 6.1 and 8.1 indices and a drop in the acid carboxyl group index (3.5). The height of the curve at pH 11 shows that there is 0.2 equivalent more carboxyl group than before deamination (perhaps due to slight hydrolysis followed by deamination of the amino group produced).

The important points to observe are that *deamination removes a large part of the lysine group* and *does not remove the prearginine, the arginine, or the histidine groups.*

This agrees with the observation of Van Slyke and Birchard[11] that the amino nitrogen of proteins equals half the lysine nitrogen.

TABLE V.

Titration of Deaminized Gelatin, 0.0100 M per 2500 Gm. (2.5 per cent).

A. More acid solutions.

pH	$\frac{b-a}{c}$	b'	pH	$\frac{b-a}{c}$	b'
1.773	−3.000	−1.083	4.347	.100	.105
2.103	−2.000	−1.105	4.471	.200	.204
2.431	−1.500	−1.081	4.565	.300	.303
3.005	−1.000	−0.889	4.660	.400	.402
3.223	−0.800	− .733	4.802	.500	.502
3.521	− .600	− .567	4.888	.600	.601
3.670	− .400	− .377	4.936	.700	.701
3.962	− .200	− .188	5.101	.800	.801
4.033	− .100	− .090	5.228	.900	.901
4.213	0	+0.006	5.280	1.000	1.001
4.217	0	.006	5.331	1.100	1.101
4.341	0.100	.105			

B. More alkali solutions.

5.585	1.200	1.200	7.769	1.600	1.600
5.856	1.300	1.300	8.145	1.650	1.650
(6.52)	1.400	1.400	8.648	1.700	1.694
(6.07)	1.400	1.400	9.373	1.800	1.796
6.746	1.450	1.450	9.785	1.850	1.841
6.964	1.500	1.500	10.771	2.000	1.913
(7.07)	1.550	1.550	11.570	2.500	1.944
(7.67)	1.550	1.550	11.693	2.700	1.962

[11] Van Slyke and Birchard, *J. Biol. Chem.*, 1913, xvi, 539.

V.

Hydrolyzed Deaminized Gelatin.

An attempt was made to partially hydrolyze some deaminized gelatin with HCl and determine if the 4.6 group remained intact. The readings were very unsatisfactory and it was not possible to draw any conclusion.

VI.

EXPERIMENTAL.

The data for gelatin are the data without salt in the preceding article of this series.[12] They do not differ materially from the corrected data obtained by others on direct titration of gelatin.

The deaminized gelatin was obtained by treating 100 gm. of isoelectric gelatin in 1 liter of water with 10 gm. of solid $NaNO_3$ and about 15 cc. of glacial acetic acid, heating on steam bath (with occasional stirring) for 4 hours, and then dialyzing through collodion membranes by a method described in another article.[13] The material after 24 hours had a conductivity of 6.4×10^{-5} reciprocal ohms (for 6.7 per cent protein).

This material was titrated by methods previously described.[14] The data are given in Table V.

The experimental curve for egg albumin is taken from the compiled data of Cohn.[5]

VII.

SUMMARY.

Analysis of the experimental titration curves shows that gelatin contains acid groups with dissociation indices at pH 2.9 to 3.5 corresponding quantitatively with the content in dicarboxylic amino acids; and that the acidic group at pH 9.4 in egg albumin agrees with the amount of tyrosine.

[12] Simms, *J. Gen. Physiol.*, 1927–28, xi, 613.
[13] Kunitz and Simms, *J. Gen. Physiol.*, 1927–28, xi, 641.
[14] Simms, *J. Am. Chem. Soc.*, 1926, xlviii, 1239.

The amounts of histidine and lysine present in both these proteins agree quantitatively with basic groups at pH 6.1 and pH 10.4 to 10.6, respectively.

However, the quantity of the arginine group (pH 8.1) in these proteins is considerably less than the amount of arginine found on hydrolysis. This deficiency is compensated (quantitatively with gelatin and approximately with egg albumin) by a basic group at pH 4.6.

The structure of this "4.6 group" should be similar to aniline and cytosine in consisting of an amino group on a conjugated unsaturated (perhaps cyclic) system. It would appear that the 4.6 group is disrupted on hydrolysis, producing arginine, and may be referred to as "prearginine."

The presence of prearginine in proteins, instead of the full amount of arginine, has an important effect on the properties. Otherwise the isoelectric point of gelatin would be 8.0 (instead of 4.7) and of egg albumin 6.6 (instead of 4.8), and the titration curves would be quite different in shape between pH 4 and 10.

Deamination of gelatin produces no decrease in prearginine, arginine, or histidine groups, but removes nearly all of the lysine group.