BØRGE BAK, JENS JØRGEN LED AND
ERIK JONAS PEDERSEN

REVERSIBLE CHEMICAL CHANGES OF POLYPEPTIDES IN CF_3COOH AS SEEN BY NUCLEAR MAGNETIC RESONANCE SPECTRA

Det Kongelige Danske Videnskabernes Selskab
Matematisk-fysiske Meddelelser **37**, 8

Kommissionær: Munksgaard
København 1969

CONTENTS

Synopsis

Nuclear magnetic resonance spectra of glucagon, $C_{153}H_{225}O_{49}N_{43}S$ (29 amino acid residues) mol. wt. 3482.8, of porcine insulin, $C_{256}H_{381}O_{76}N_{65}S_6$ (51 residues) mol. wt. 5776.6, of bovine insulin, $C_{254}H_{377}O_{75}N_{65}S_6$ (51 residues) mol. wt. 5733.6, of bovine A-chain tetra-S-sulfonate, $C_{97}H_{151}O_{46}N_{25}S_8$ (21 res.) mol. wt. 2659.9 and of bovine B-chain di-S-sulfonate, $C_{157}H_{232}O_{47}N_{40}S_4$ (30 res.) mol. wt. 3560.0, all dissolved in CF_3COOH, show that the primary alcohol groups of all serine residues convert to $-CH_2OCOCF_3$ in 3–4 h at 30° while the secondary alcohol groups of all threonine residues convert to $>CHOCOCF_3$ in 1–2 days. Biologically active glucagon can be recovered in high yield from the heptatrifluoroacetylated derivative. Also, the $COCF_3$ groups of this compound can be replaced by $COCH_3$ groups.

PRINTED IN DENMARK
BIANCO LUNOS BOGTRYKKERI A/S

I. Introduction

Classical chemical studies of polypeptide composition and chemical properties may be supplemented effeciently by the use of accumulated nuclear magnetic resonance (NMR) spectra. This implies, of course, extreme magnetic field stability and homogeneity besides chemical stability of the sample. From a spectroscopic point of view (vide infra) the preferable polypeptide solvent is CF_3COOH. However, we have found that polypeptides containing serine and threonine residues undergo chemical changes in this solvent, their primary and secondary alcohol groups becoming trifluoroacetylated in 1–2 days at 30°. Strictly reproducable spectral work is, therefore, only possible after the elapse of this time. After spectral recording and removal of the solvent biologically active polypeptide can be recovered by hydrolysis as demonstrated for glucagon.

II. Experimental Procedure

^{19}F magnetic resonance (FMR) spectra were recorded at 94.1 MHz, proton magnetic resonance spectra (^1HMR) at 100 and 220 MHz, all on Varian instruments. The polypeptide molarity was about 0.02. Internal standards were $CFCl_2CF_2Cl$ (central peak of *F* triplet) for ^{19}F and $(CH_3)_4Si$ for ^1H.

The FMR spectra were single recordings because the low-field spin-satellite from $^{13}CF_3COOH$ (1% natural abundance) was so close to and, in some cases, inseparable from the ^{19}F resonances of the trifluoroacetylated polypeptides that accumulation of spectra was no advantage. Recording of the equally intense high-field spin-satellite signal before and after each FMR spectrum served as a means of correcting for the intensity contribution of the low-field satellite to the spectrum which was integrated manually (planimeter). Spectral base-line ambiguity etc. introduced an estimated error of 10% in I_{tot} (Table I).

TABLE I. Integrated total intensities of ^{19}F magnetic resonance spectra from trifluoroacetylated polypeptides. Single recordings at 30°.

The integrals are expressed as number of $COCF_3$ groups per mole of substance. Q is the initial quantity of polypeptides in units of 10^{-3} mmole dissolved in 600 µl of CF_3COOH. N_s and N_t are the known numbers of serine, resp. threonine OH groups per molecule of substance. h = hours.

	Q	1h	2h	5h	23h	71h	N_s+N_t	N_s	N_t
Porcine insulin	8.72	1.09	2.39	3.99	5.61		5	3	2
Bovine insulin	8.66	1.29	2.64	3.92	4.75		4	3	1
Bovine insulin A chain S-sulfonate	8.65	~0	1.10	1.82	2.58		2	2	0
Bovine insulin B chain S-sulfonate	8.43	~0	0.58	1.43	2.66		2	1	1
Porcine glucagon	5.73	0.76	2.69	4.07	6.90	7.68	7	4	3
Bovine glucagon	5.37	~0	3.06	4.15	7.22	7.38	7	4	3

As to ^{1}HMR spectra, usually 50 were accumulated when working at 100 MHz. The temperature at the sample site was 30°. The spectra taken at 220 MHz were single recordings, the temperature at the sample site being 10–15°.

The quality of the insulins and the bovine insuline A and B chains was as described earlier[1]. Two times recrystallized porcine and bovine glucagon were given to us by the NOVO RESEARCH INSTITUTE. The glucagons were dried to constant weight *in vacuo* at room temperature. For a sample it was checked that no further loss of weight occurred at 100°.

III. ^{19}FMR Spectra at 94.1 MHz

The glucagon molecule is especially rich in serine (4) and threonine (3) residues. Porcine and bovine glucagon were included in this investigation to follow their trifluoroacetylation, but also to put their alleged chemical identity[2] to spectroscopic tests. Comparison of spectra 1 and 3 (or 2 and 4) shows the expected slight dependance of chemical shifts on concentration (Table II). Within the limits of error spectra 1 and 2 and 3 and 4 are identical, one positive test of porcine and bovine glucagon identity. The FMR spectrum of hepta-trifluoroacetylated glucagon appears as 7 separate peaks (fig. 1), 4 of which are fully developed in 3–4 h (marked 'f'), the remaining 3 appearing much slower('s'). Serine and threonine amino acids react with similar relative rates, but much slower on an absolute scale. The 'f' signals are, therefore, probably FMR signals from $-CH_2OCOCF_3$ groups. This assignment is verified by noting that in spectra 5, 6, 7, and 8 (Table II) 3, 3, 2, and 1 'fast' signals are recorded corresponding to the authentic

TABLE II. Chemical shifts in ^{19}FMR spectra of trifluoroacetylated porcine glucagon (spec. 1 and 3), bovine glucagon (spec. 2 and 4), porcine insulin (spec. 5), bovine insulin (spec. 6), of bovine A-chain S-sulphonate (spec. 7), and of bovine B-chain S-sulphonate measured in cps relative to internal $CFCl_2CF_2Cl$. Q as in Table I. Single recordings at 30°.

Spectrum number	Q								
1	5.73	266.6^f		278.6^f	280.4^f	*287^f	310.6^s	312.2^s	329^s
2	5.37	265.6^f		278.2^f	280.4^f	*287^f	310.7^s	313.1^s	329^s
3	14.3	268.0^f		*281^f	*281^f	288.6^f	312.4^s	314.4^s	330^s
4	14.3	268.2^f		*281^f	*281^f	289.1^f	312.2^s	315.2^s	330^s
5	8.72	268.4^s	272.1^f		275.4^f	287.6^{f+s}			
6	8.66	270.0^f	276.1^f		*284.2^f	287.8^s			
7	8.65	265.0^f	274.2^f						
8	8.43					*287^s	289.3^f		

f: resonance fully developed in 3–4 h.
s: resonance fully developed in 1–2 days.
* interference with ^{19}F resonance in $^{13}CF_3COOH$ (1 % natural abundance).

occurrance of 3, 3, 2, and 1 serine residues in the respective insulins and 'chains'. In all five cases the final total intensity of the FMR spectrum, I_{tot}, was close to $N_s + N_t$, the total number of serine and threonine residues per molecule. Since the discrepancy is systematically > 0 it may be evidence of a so far unidentified chemical reaction, but in view of the experimental uncertainty (10 %) of our intensity measurements we shall refrain from any interpretation of the difference which in no case amounts to 1 $COCF_3$ group

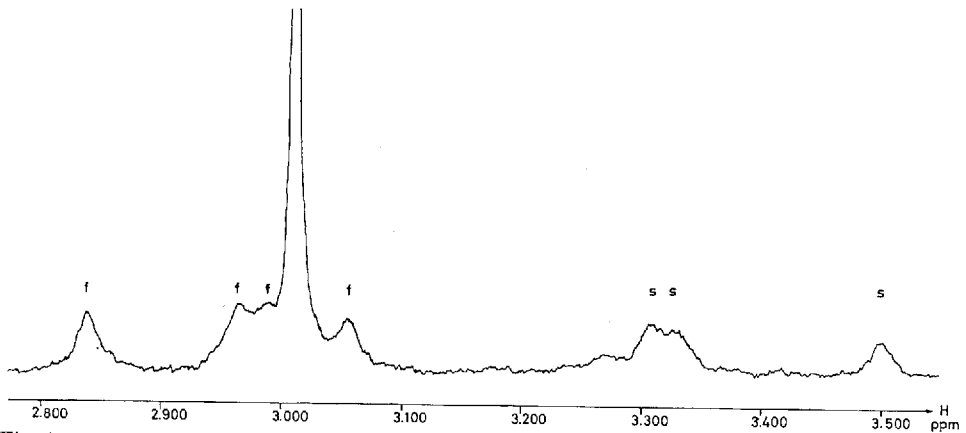

Fig. 1. 10 accumulated FMR spectra of hepta-trifluoroacetylated glucagon in CF_3COOH. f = 'fast' $-CH_2OCOCF_3$ resonance, s = 'slow' $>CHOCOCF_3$ resonance. The strong line at 3.01 ppm/internal $CFCl_2CF_2Cl$ is the $^{13}CF_3COOH$ low-field spin satellite.

per molecule. Of whatever nature the possible change is, it is reversible by hydrolysis just as the trifluoroacetylation (vide infra).

In a physically denatured polypeptide the main determining factor for the chemical shift of a specific 'serine' or 'threonine' $COCF_3$ group is the chemical nature of the adjoining amino acid residues. The fact that identical FMR spectra are obtained from hepta-trifluoroacetylated porcine and bovine glucagon is a first sign of their chemical identity. The occurrance of 7 separate peaks (fig. 1) verifies the chemical sequence determination[2] inasfar as no serine or threonine residue is placed in identical chemical surroundings. The same conclusion applies to the insulins in agreement with Sanger's formulae[3].

IV. ^{1}HMR Spectra at 100 and 220 MHz

A complete analysis of the ^{1}HMR spectrum of glucagon is being published[4]. In the analysis, integrated intensities of a series of separate spectral bands were used as demonstrated earlier[1,5] for the insulins. Thus, a specific spectral band consists of superimposed single protons resonances (p.r.) from a known number of protons in known positions in the molecule.

As a result of progressing trifluoroacetylation the polypeptide ^{1}HMR spectra should be time-dependant. With glucagon as an example this is illustrated in fig. 2. Due to the reaction

$$\begin{array}{l} \mathrm{XCH_\beta OH + CF_3COOH} \\ \quad\quad | \\ \mathrm{-HNCHCO-} \end{array} \rightarrow \begin{array}{l} \mathrm{XCH_\beta OCOCF_3 + H_2O} \\ \quad\quad | \\ \mathrm{-HNCHCO-} \end{array}$$

$X = H_\beta$ for serine $\quad$ $X = CH_3$ for threonine

the 8 p.r. of the 8 H_β in the four glucagon serine residues and the 3 p.r. of the 3 H_β in the three threonine residues should undergo downfield chemical shifts. From the analysis of the complete spectrum it follows that the spectral intensity between A and D (fig. 2) corresponds to 42 p.r. per molecule. Since the rapidly disappearing 'shoulder' between C and D corresponds to *ca.* 6 p.r. whereas the slowly occurring band between A and B represents *ca.* 2.5 p.r., one interpretation is obvious. This was fully confirmed by studies at 220 MHz (fig. 3). Here, the area between C and D corresponds to 10.5 p.r. for a newly prepared sample. After a week (practical reasons preventing observation after 2 days) this area had shrunk to *ca.* 3.5 p.r. so that *ca.* 7 p.r. have moved. Between A and B there are practically no signals at

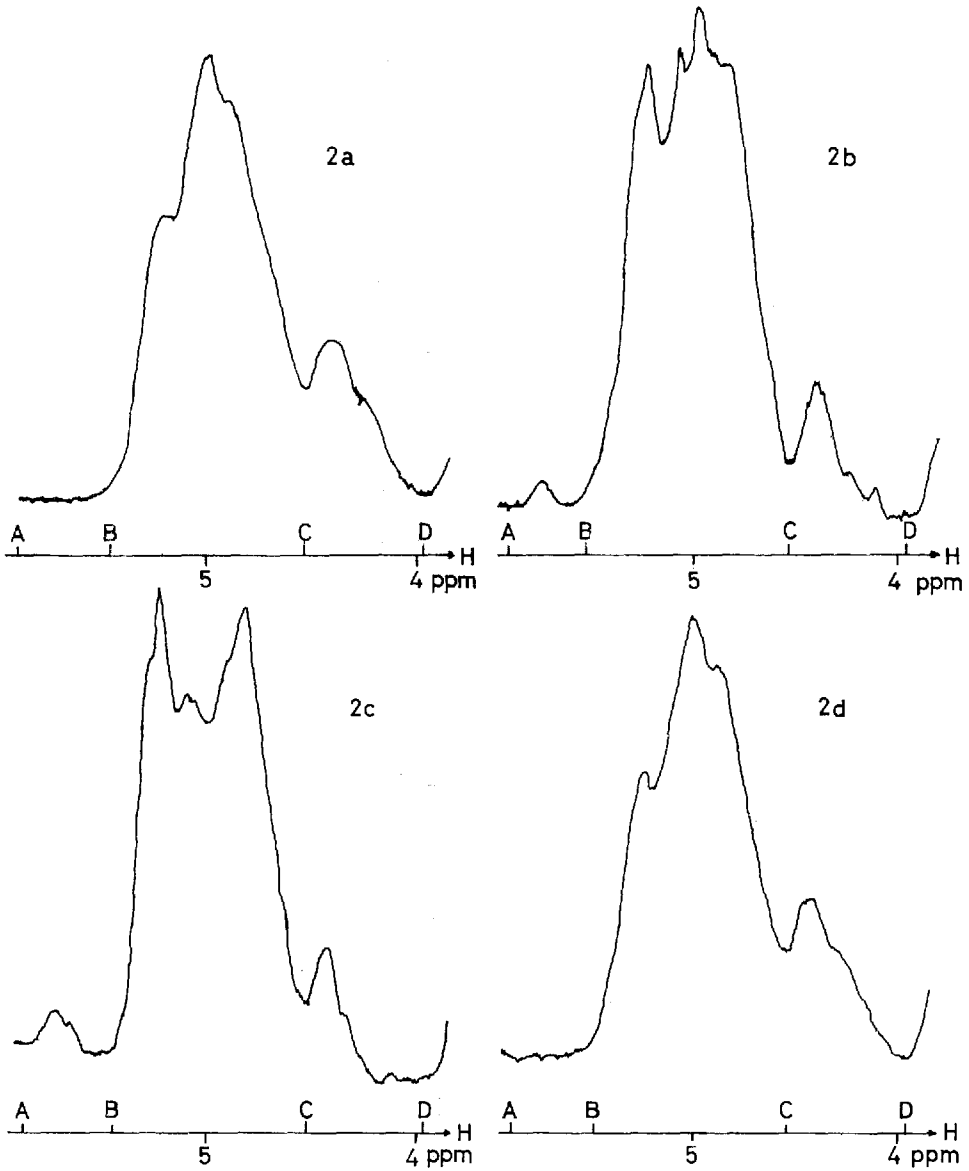

Fig. 2. Section of glucagon ^{1}HMR spectrum at 100 MHz. 2a: fifty accumulated spectra recorded 1–3 h after preparation of the solution in CF_3COOH. 2b: fifty accumulated spectra recorded 3–5 h after preparation. 2c: after 2 days. 2d: after hydrolysis (see text) and recording 1–3 h after dissolving in CF_3COOH again.

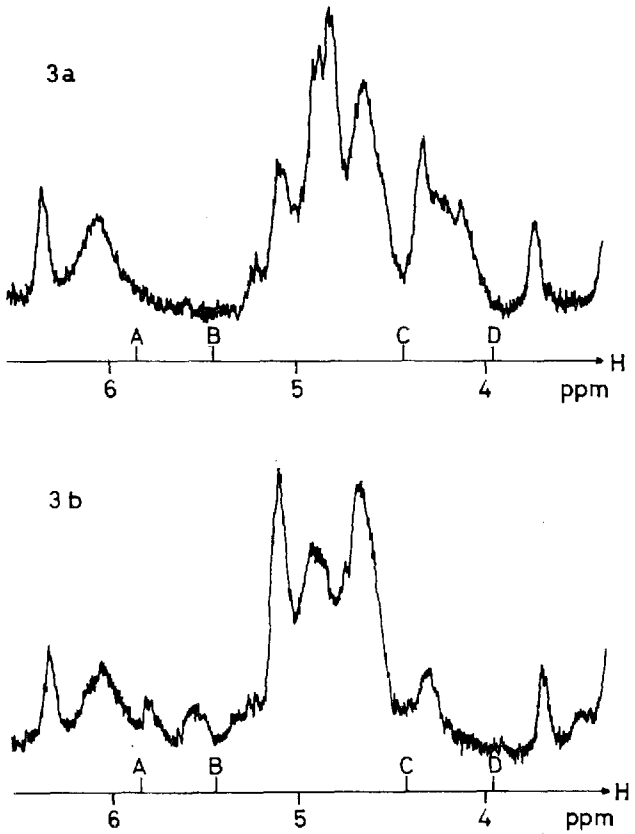

Fig. 3. Section of glucagon ^{1}HMR at 220 MHz. 3a: single recording immediately after preparation of the solution. 3b: single recording 1 week (see text) after preparation.

the beginning, but after a week two distinct signals had appeared between 5.66 and 5.90 ppm/TMS with a total intensity corresponding to *ca.* 3.5 p.r. These signals were positively identified as originating from threonine H_β-protons. By spin-coupling they produce CH_3 resonance doublets at *ca.* 1.50 ppm/TMS. One at a time these doublets became singlets by irradiating the glucagon sample at 5.668, 5.827, and 5.882 ppm downfield from TMS. Serine residues are unable to produce this effect(no CH_3 groups).

Porcine and bovine glucagon identity was checked spectroscopically by demonstrating the identity of their ^{1}HMR spectra at 100 MHz. To test the sensitivity the 0.02 m glucagon solution was afterwards made 0.02 m with respect to leucine, one of the amino acids occurring as residue. The expected spectral intensity changes were observed. The identity and the sensitivity checks were, of course, carried out after chemical 'maturing' for 2 days.

V. Hydrolysis and Acetolysis of Trifluoroacetylated Glucagon

Upon hydrolysis one of the spectral bands of glucagon adopts its original shape (fig. 2). This is true for the entire spectrum, but the changes are far less in most other parts of the spectrum. Hydrolysis of 0.01 mmole of heptatrifluoroacetylated glucagon was carried out by removing the solvent CF_3COOH *in vacuo*, adding $^1/_2$ ml of water and 10 μl of CF_3COOH, shaking for 24 h, and finally distilling off water *in vacuo* at room temperature. The remaining glucagon was dissolved in CF_3COOH and its FMR spectrum was immediately recorded. It showed a trace of signals at the location of the 'threonine' $COCF_3$ resonances and a somewhat larger, diffuse band at the location of the 'serine' $COCF_3$ resonances, all amounting to *ca.* 1 $COCF_3$ group per molecule. Part of this is due to trifluoroacetylation starting afresh when glucagon is dissolved in CF_3COOH. Another part is due to incomplete hydrolysis as seen by the occurrance of 'threonine' $COCF_3$ signals.

Once the glucagon derivative

$$(\text{glucagon} - 7\ \text{H})\ (COCF_3)_7$$

is available other derivatives become accessible. We have prepared what appeared to be

$$(\text{glucagon} - 7\ \text{H})\ (COCH_3)_3^{\text{thr}}(COCH_3)_2^{\text{ser}}(COCF_3)_2^{\text{ser}}$$

(ser = serine, thr = threonine). 40 mg hepta-trifluoroacetylated glucagon was dissolved in 250 μl of CF_3COOH, 250 μl of CH_3COOH, and 20 μl of water which catalyzes the exchange of $COCF_3$ with $COCH_3$. After 48 h, 100 μl of $(CH_3CO)_2O$ was added, the quantity necessary to remove water chemically. After another 24 h the sample was evaporated to dryness *in vacuo*. 2 ml portions of di-butylether were repeatedly added and distilled-off in order to remove traces of CH_3COOH which would falsify the next spectrum. The remanence was dissolved in CF_3COOH and its ^{1}HMR spectrum was recorded. It showed a large, extra signal at the usual location of $COCH_3$ resonances with an intensity corresponding to 14–16 p.r. per molecule. At 4 h after the preparation of the solution the FMR spectrum was recorded. Practically no 'threonine' signals were seen, but a diffuse 'serine' signal corresponding to ca. 2 $COCF_3$ groups was recorded, all as expressed in the formula above. There can be little doubt that complete acetylation and, in general, acylation is possible.

Unlike the $COCF_3$ groups the $COCH_3$ groups cannot be removed hydrolytically in slightly acid medium. After removal of the solvent of the sample

above, it was shaken for 24 h with dilute NaOH at pH = 11. Water was distilled-off and replaced by CF_3COOH which was immediately distilled-off in order to remove CH_3COOH formed by the hydrolysis. The residue was slightly trifluoroacetylated glucagon as seen by the ^{1}HMR spectrum of its solution in CF_3COOH.

VI. Biological Activity of Recovered Polypeptides

By immunological determination carried out at the NOVO THERAPEUTIC LABORATORY the amount of glucagon-like substance was found to be 45–55 %. A more specific biological activity determination[6] showed 55 % activity (fidelity limits 48–62 %).

As mentioned above the average composition of our glucagon sample recovered by hydrolysis of the hepta-trifluoroacetylated derivative, was (glucagon – 1 H)$COCF_3$. If one $COCF_3$ group per molecule impairs biological activity the maximum activity of the sample above would be $100(6/7)^7 = 35$ %. If, on the other hand, the introduction of $COCF_3$ groups in a polypeptide is of no or little consequence for its biological activity an established loss of activity must be due to a (so far not identified) chemical change not reversed by hydrolysis.

VII. Conclusions

The slowly occurring FMR signals and the corresponding changes in the ^{1}HMR spectra show that polypeptides dissolved in CF_3COOH undergo chemical changes. The demonstrated trifluoroacetylation of all hydroxy groups in serine and threonine residues is thought to be typical.

After hydrolysis of hepta-trifluoroacetylated glucagon 'genuine' glucagon of ca. 50 % biological activity is obtained. As mentioned in an earlier publication[1] the same seems true for insulin. Therefore, although CF_3COOH is a powerful chemical agent it may only cause chemical changes of minor biological importance.

The use of CF_3COOH in preparative and analytical polypeptide studies taking advantage of NMR techniques can be justified like this: 1° Its ability to dissolve at least a majority of polypeptides and many proteins. 2° The CF_3COOH p.r. is well separated from the polypeptide ^{1}HMR spectra in contrast to H_2O, HDO, H_3O^+ etc. with p.r. occurring in the middle of the spectrum if aqueous medium is applied. 3° CF_3COOH is a very strong acid provoking maximum protonation of the solute. Without this (for instance by studies in aqueous media) solubility differences would necessitate work

at widely different acidities with accompanying changes of chemical shifts of the numerous protons in slow or fast exchange with the solvent proton(s).

The finding of a solvent supplementing CF_3COOH would be of great value. $(CD_3)_2SO$ is under observation, but the isotopic purity of the commercially available compound is less satisfactory (although $>98\,^0/_0$) for work applying accumulated spectra.

The easy access to trifluoroacetylated polypeptides here demonstrated opens routes to numerous derivatives differing in molecular weight, isoelectric point, polarity, total electric charge etc. Since biologically activite polypeptides can be retrieved promising perspectives for, for example, the separation of naturally occurring mixtures of polypeptides by way of their derivatives are obvious.

Acknowledgements

We want to thank Dr. Jørgen Schlichtkrull, head of the NOVO RESEARCH INSTITUTE, Copenhagen, and staff members of the INSTITUTE and of NOVO THERAPEUTICAL LABORATORY for gifts of highly purified insulin and glucagon, and for carrying out the biological tests.

We are grateful to VARIAN ASSOCIATES, Palo Alto, for access to their 220 MHz spectrograph during a short visit (by B.B.). The skillful help of mr. Lewis Cary in recording the spectra is gratefully acknowledged.

Thanks are due to mrs. Annelise Hallund of this laboratory for her able participation in the experimental work.

Note Added In Proof

Recent, not yet completed studies at 220 MHz and $25°$ ($\neq 10-15°$ as employed above) show that whereas the free amino acid tryptophane dissolved in CF_3COOH is stable for at least 48 h, ^{1}HMR signals from the aromatic protons of the tryptophane residue of glucagon start fading discernably after 3 h. After 24 h the easily recognizable signal from H(2) of the tryptophane residue has disappeared. At present the nature of the corresponding chemical change is unknown. Also, it may or may not be reversible during the hydrolysis of 'hepta-trifluoroacetylated glucagon'. This type of change does, of course, not occur for polypeptides such as the insulins with no tryptophane residues.

References

1. B. Bak, E. J. Pedersen, and F. Sundby, J. Biol. Chem. **242**, 2637 (1967).
2. O. K. Behrens and W. W. Bromer, Vitamins and Hormons **16**, 263 (1958) and Private communication.
3. F. Sanger, Biochem. J. **44**, 126 (1947); ibid. **45**, 563 (1949).
4. B. Bak, J. J. Led, and E. J. Pedersen, J. Mol. Spectry. (to be published).
5. B. Bak, C. Dambmann, F. Nicolaisen, E. J. Pedersen, and N. S. Bhacca. J. Mol. Spectry. **26**, 78 (1968).
6. F. Tarding, P. Nielsen, B. Keiser-Nielsen, and Aa. V. Nielsen, Diabetologia **5** (1969) (in the press).

Indleveret til Selskabet den 17. marts 1969.
Færdig fra trykkeriet den 29. september 1969.