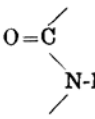
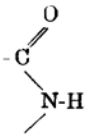
On the Configurations of the Peptide Linkage in Proteins and Some Other Substances, as Revealed by Their Near-Infrared Absorption Spectra.

By Masamichi TSUBOI.

(Received June 29, 1949.)

As continuation of his investigation on the hydrogen bondings in compounds with the peptide linkage,[1] the writer made 3μ region infrared absorption measurements of some proteins, diketopiperazine, some amides, some lactams, etc., in the solid or the liquid state. In this paper the writer reports some results of these measurements and gives interpretations of them in the light of the knowledge, previously obtained,

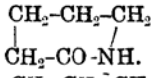
(1) M. Tsuboi, This Bulletin, **22** (1949), 215.

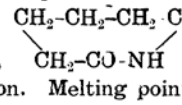
of the 3μ region absorptions of amides in carbon tetrachloride solutions.[1] A conclusion reached is that the peptide linkage in proteins is of the *trans* configuration,  , but not of the *cis* configuration, 

Experimental.

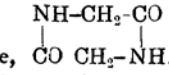
The infrared monochromator and other apparatus used in the present investigation are the same as those described in the previous paper.[1] The samples examined are:

N-Methylacetamide, $CH_3CONHCH_3$. Purified by vacuum distillation. Melting point: 28°C.

δ-Valerolactam,  . Purified by vacuum distillation. Melting point: 40°C.

ε-Caparolactam,  . Supplied by K. Hoshino and H. Yumoto. Purified by vacuum distillation. Melting point: 68°C.

Formamide, $HCONH_2$. Purified by vacuum distillation.

Diketopiperazine,  . Prepared by A. Mukai. Purified by repeated recrystallization from hydrochloric acid using methanol.

Poly-ε-capramide ("Amilan").[2] $\ldots(CH_2)_5CONH(CH_2)_5CONH\ldots$. Supplied by K. Hoshino and H. Yumoto. Dried over calcium chloride.

Egg albumin. Prepared according to the method of Kekwick, and purified by recrystallization by H. Sugano. A suitable film for the infrared absorption measurements was obtained by spreading a drop of aqueous solution of egg albumin on a cover glass (the salt contained being removed by dialysis), and drying it first over calcium chloride and then over phosphorus pentoxide.

Horse serum albumin. Prepared according to the method of Kekwick, and purified by recrystallization by S. Nakamura. A film of this protein was prepared in the same way as that of egg albumin.

Results and Interpretations.

The results obtained are shown in Fig. 1, where $\log I_0/I$ is plotted against wave-length, I_0/I being the ratio of the intensity of light transmitted by blank space to that by the film.

Before proceeding to the interpretation of these results, the assignments of N-H absorption bands of compounds with the peptide linkage, described in detail in the previous paper,[1] may be summarized in Table 1. These relations were obtained by studying the effects of the concentrations and the temperatures of carbon tetrachloride solutions of N-methylacetamide and several other amides and δ-valerolactam on the 3μ region absorption spectra. Compounds with the peptide linkage show the strong tendency of forming intermolecular and intramolecular N-H...O=C bondings; accordingly most of N-H's and C=O's in these compounds

(2) Material of synthetic fibers, made by Tōyō Rayon Co.

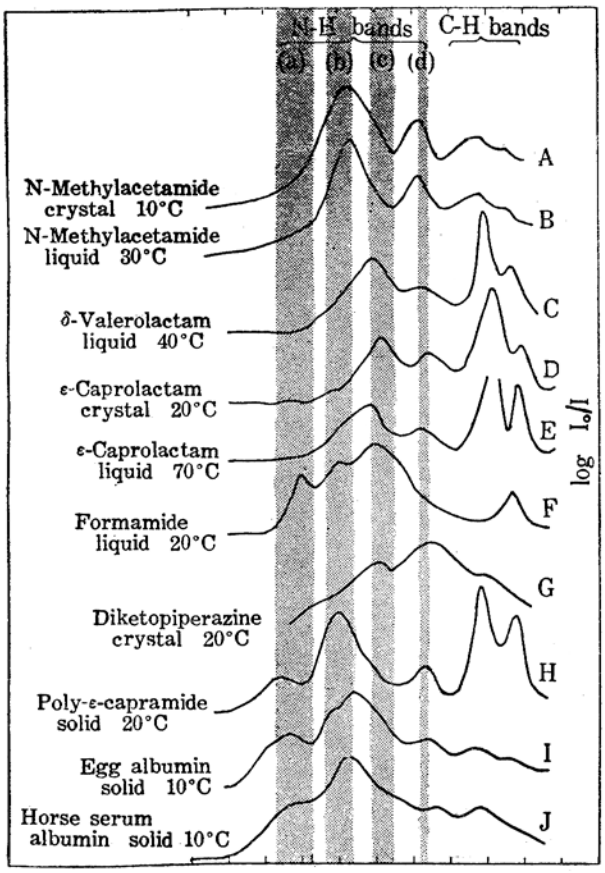
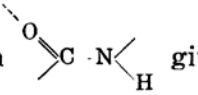
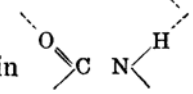
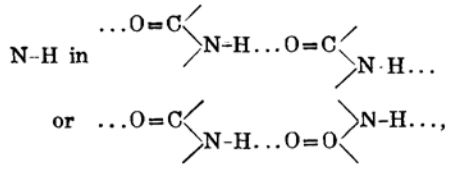
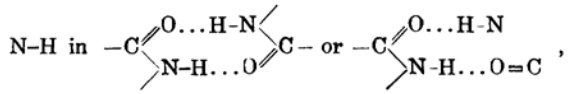


Fig. 1. 3μ region infrared absorption curves of some compounds with peptide linkage.

must form hydrogen bondings in the liquid and the solid states. On the other hand, as is shown in Table 1, N–H in  gives an association band in 2.97~3.04μ region, and N–H in  an association band in 3.09~3.15μ region.* Hence, *we can tell whether the peptide linkage is of the trans configuration or of the cis configuration, according as the N–H association band appears in the 2.97~3.04μ region or the 3.09~3.15μ region.* Using this relation, the configurations of CONH

* This difference in wave-lengths of N–H association bands is attributable to the difference in the positions of H relative to other atoms, as detailed in the author's previous paper.[1]

Table 1.
The assignments of N–H bands of compounds with the peptide linkage.

Wave-lengths (in μ) of absorption maxima	Assignments
(a) 2.83~2.93	Free N-H
(b) 2.97~3.04	 where H on N is *trans* to O.
(c) 3.09~3.15	 where H on N is *cis* to O.
(d) 3.22~3.24	Not uniquely assigned. The band may be assigned to something that accompanies the formation of intermolecular hydrogen bonding N-H...O. In the molecules here concerned H on N may be either *trans* or *cis* to O.

may be determined as shown in Table 2, and the absorption curves in Fig. 1 may be interpreted as follows:

N-Methylacetamide. (Cf. Fig. 1, A and B.) N-Methylacetamide, in the crystalline state, shows the absorption maxima at 3.03μ and 3.22μ, but not at 3.09~3.15μ. This shows that the crystal consists only of *trans* molecules (molecules in each of which H on N is *trans* to O), and

Table 2.
Configurations of CONH, determined from the wave-lengths of absorption maxima of N–H...O bands.

Samples		Wave-lengths (in μ) of absorption maxima of N–H...O bands			Configurations of CONH
N-Methylacetamide	(crystal)	3.03		3.22	trans
,,	(liquid)	3.03		3.22	trans
δ-Valerolactum	(liquid)		3.09	3.23	cis
ε-Caprolactam	(crystal)		3.11	3.24	cis
,,	(liquid)		3.09	3.23	cis
Formamide	(liquid)	3.00	3.10		trans and cis
Diketopiperazine	(crystal)		3.11	3.26	cis
Poly-ε-capramide	(solid)	3.00		3.23	trans
Egg albumin	(solid)	3.04		3.24	trans
Horse serum albumin	(solid)	3.03		2.25	trans

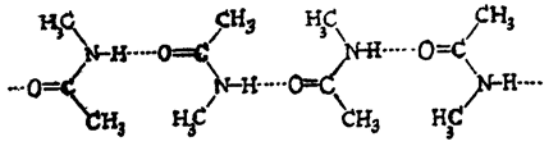


Fig. 2.

has a structure such as shown in Fig. 2. It is interesting to note that, according to C. J. Brown and D. E. Corbridge,[3] acetanilide has the crystal structure (hence the molecular structure) as shown in Fig. 3, which is of the same type as that of N-methylacetamide shown in Fig. 2.

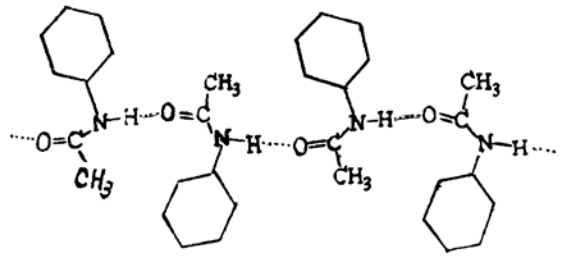


Fig. 3. After C. J. Brown and D. E. Corbridge.

N-Methylacetamide, in the liquid state, also shows the absorption maxima at 3.03μ and 3.22μ with nearly the same intensities (relative to those of the C–H bands) and widths of the bands as in the crystalline state. This shows that the molecule of N-methylacetamide is linked in nearly the same manner in the liquid state as in the crystalline state, at least at the temperatures a little above its melting point.

δ-Valerolactam. (Cf. Fig. 1, C.) δ-Valerolactam, in the liquid state, shows absorption maxima at 3.09μ and 3.23μ, i. e. at wave-lengths, respectively, shorter than 3.11μ and 3.24μ, which are the wave-lengths of absorption maxima of δ-valerolactam in carbon tetrachloride solution.[1] This fact may be explained as due to some structural deformation of the ring dimers (to which the 3.11μ and 3.24μ bands were assigned[1]), caused by closer mutual approach of the molecules of δ-valerolactam.

ε-Caprolactam. (Cf. Fig. 1, D and E.) ε-Caprolactam, in the crystalline state, shows absorption maxima at 3.11μ and 3.24μ, but not at 2.97~3.04μ. This shows that in ε-caprolactam molecules in the crystalline state, H on N is *cis* to O.

ε-Caprolactam, in the liquid state, shows two N–H association bands whose maxima lie at 3.09μ and at 3.23μ; hence, in this molecule H on N is *cis* to O, also in the liquid state, although these two bands are shorter in wave-lengths, weaker in absorption intensities (relative to those of the C–H bands), and broader than the 3.11μ and 3.24μ bands of ε-caprolactam in the crystalline state.

(3) C. J. Brown and D. E. Corbridge, *Nature*, **162** (1948), 72.

Formamide. (Cf. Fig. 1, F.) Formamide, in the liquid state, gives N–H association bands at 3.00μ and 3.10μ. They may be assigned to the stretching vibrations of N–H's, the 3.00μ band to that in which H on N is *trans* to O, and the 3.10μ band to that in which H on N is *cis* to O. Besides, it also gives a band at the free N–H region. This reveals that N–H's of formamide are partially in free state even in its pure liquid state—an unusual phenomenon for an associated liquid.

Diketopiperazine. (Cf. Fig. 1, G.) Diketpiperazine, in crystalline state, shows N–H bands at 3.11μ[4] and 3.26μ, but not at 2.97~3.04μ, revealing that the CONH of this substance is of the *cis* configuration. This is in agreement with the experimental results of X-ray analysis by Go and Ohashi[5] and by Corey[6] (see Fig. 4).

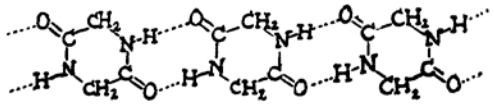
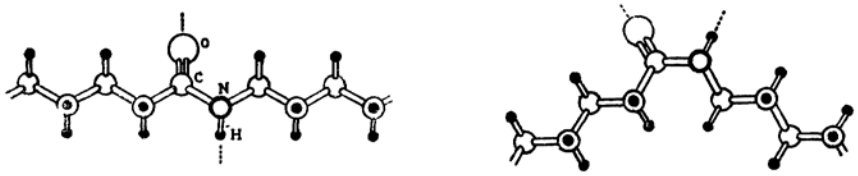


Fig. 4. After R. B. Corey

Poly-ε-capramide. (Cf. Fig. 1, H.) There are absorption maxima for poly-ε-capramide at 3.00μ and at 3.23μ, but none in the 3.09~3.15μ region.[7] This shows that H on N in poly-ε-capramide is *trans* to O. It follows then that the chain ...$-CH_2CH_2CONHCH_2CH_2-$... in this substance must be of the form as shown in Fig. 5 (A), but not of the form as shown in Fig. 5 (B).



(A) H on N is trans to O. (B) H on N is cis to O.

Fig. 5.

Proteins. (Cf. Fig. 1, I and J.) The molecules of proteins have as their constituents various atomic groups that give absorption bands in the 3μ region. They are shown in Table 3, with the wave-lengths of

(4) The 3μ region absorption curve of solid film of diketopiperazine has already been obtained by A. M. Buswell and others in 1940 (*J. Am. Chem. Soc.*, **62** (1940), 2759). They observed an absorption at 3.14μ, but they suspected this to be due to glycine, which might be produced as a dissociation product in the preparation of the film. The present writer observed an absorption at 3.11μ (probably corresponding to that at 3.14μ observed by Buswell) even in a film confirmed to be free from glycine by the K. Kuratani's infrared absorption measurements in the rock salt region.

(5) Go and Ohashi, Paper communicated to Chem. Soc. Japan, at its 60th Annual Meeting, April 1938.

(6) R. B. Corey, *J. Am. Chem. Soc.*, **60** (1938), 1598.

(7) See foot-note (9).

Table 3.

Atomic groups	Wave-lengths (in μ) of absorption maxima	Number of the groups in one molecule	
		Egg albumin	Bovine serum albumin
$\rangle$C-H	3.35~3.60	2050	3390
$\rangle$C-H[a]	3.21~3.30	150	270
CONH	2.97~3.26 (2.83~2.93 in free state)	430	690
N^+-H	3.14~3.28	80	180
N-H[b]	2.83~2.95		
COOH[c]	3.27~3.89 (2.79~2.84 in free state)	$\leqslant$ 50[c]	$\leqslant$ 90[c]
OH	2.83~3.22 (2.75~2.77 in free state)	50	80

(*a*) C-H with unsaturated C.
(*b*) Amino group may be mostly of N^+-H form, that of N-H form being negligible.
(*c*) Carboxyl group may be mostly of COO^- form, that of COOH form being negligible.

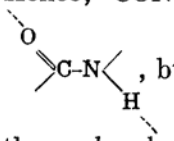
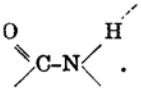
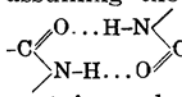
the absorption maxima and the number of each of the atomic groups in one molecule of egg albumin and of bovine serum albumin, estimated from the results of amino acid analyses of these proteins.[8] As shown in the table, COHN is the most abundant of the atomic groups that give absorption maxima in the 2.97~3.26μ region. It is this group that determines the essential forms of the absorption curves of these proteins in the said region. Other groups, such as N^+-H and O-H, being far less in number, their effects upon the forms of the curves must be very little. The group $\rangle$C-H, on the other hand, is comparatively large in number, attaining as many as one third of the number of CONH; but since $\rangle$C-H bands are generally much weaker than N-H association bands in absorption intensities (per one group), its effect is considered to be also negligible. The absorption maxima in the same region of egg albumin and horse serum albumin are at 3.04μ or 3.03μ and at 3.24μ or 3.25μ,[9] but not in the vicinity of 3.11μ, unlike δ-valerolactam, ε-caprolactam, and diketopiperazine, all of which show the absorption maximum near 3.11μ. Other proteins, according to Buswell and others,[10] also show the absorption maxima at nearly the same positions, except salmin.[11]

(8) Block, *Advances in Protein Chemistry*, **2** (1945), 119; Stein and Moore, *J. Biol. Chem.*, **178** (1949), 79.

(9) In the last three curves of Fig. 1, besides the absorption maxima mentioned above, there are weak ones at 2.85~2.88μ, which may probably be attributed to the bound water.

(10) A. M. Buswell, K. F. Krebs and W. H. Rodebush, *J. Phys. Chem.*, **44** (1940), 1126.

(11) The 3.09μ band of salmin may be due to the N-H's in guanidinium groups, as Buswell and others indicated in their paper cited in (10).

Hence, CONH in these proteins must be of the *trans* configuration  , but not of the *cis* configuration  . This is due to the molecular structure of proteins, which has been much debated. On this basis any structural view involving the assumption of the *cis* configuration in the peptide linkage is excluded, such as the view assuming the diketopiperazine structure, or the ring dimer structure,  , etc. On the other hand, the structural model of protein molecule recently proposed by Simanouti and Mizushima[12] is incidentally in accord with the writer's conclusion given above, for in their model the peptide linkage is always and exclusively of *trans* configuration.

The writer wishes to express his sincere thanks to Professor Sanichiro Mizushima and Assistant Professor Takehiko Simanouti for their kind guidance throughout this work.

Chemical Institute, Faculty of Science,
Tokyo University.

---

(12) T. Simanouti and S. Mizushima, *Kwagaku*, **17** (1947), 24, 52; This Bulletin, **21** (1948), 1.