ANNUAL REPORTS ON
NMR SPECTROSCOPY

Edited by

E. F. MOONEY

VOLUME 6B

ANNUAL REPORTS ON
NMR SPECTROSCOPY

Edited by

E. F. MOONEY

Anacon (Instruments) Limited, Bourne End, Buckinghamshire, England

VOLUME 6B

)

LIST OF CONTRIBUTORS

L. CAVALLI, *Societa' Italiana Resine, Centro Ricerche Analisi, SEAN, Via Trento, 106-20099 Sesto S.G., Milan, Italy*

W. A. THOMAS, *Roche Products Limited, Welwyn Garden City, Hertfordshire, England*

ACKNOWLEDGMENTS

CONTENTS

NMR and Conformations of Amino Acids, Peptides and Proteins[*]

W. A. THOMAS

Department of Chemistry, University College of Swansea,
Singleton Park, Swansea, S. Wales†

I. INTRODUCTION

In the last fifteen years the determination of the solution conformations of organic molecules has been revolutionized by the continual improvements in NMR instrumental and theoretical and practical advances in the field. The recent breakthrough in the building of commercial spectrometers able to routinely examine ^{13}C or ^{15}N nuclei promises to inject even more interest in the topic of molecular conformation for a long time to come. The potential of. the technique as applied to biological or biochemical problems has been realised particularly in the last two years (1970–71) and the

[*] An appendix to this chapter can be found on p. 223.

† Present address: Roche Products Limited, Welwyn Garden City, Hertfordshire, England.

domain of the X-ray crystallographer has been infiltrated to a degree previously thought impossible.

In this review, the progress made in the conformational analysis of amino acids and peptides, using NMR as the major tool, is examined, to the end of December, 1971, with a few relevant references after this date also incorporated. The reviewer must apologise for not mentioning papers appearing·in biological and biochemical journals which are not covered by Chemical Abstracts and with which he is not familiar.

The time-lag in the development of the NMR technique in analysing the conformations of polypeptides and proteins was, in the main, because of the apparent lack of information available from complex spectra in aqueous solutions, insolubility in typical NMR solvents, severe line-broadening in the spectra of large peptides (proteins and enzymes) and a general disbelief that these problems could be overcome. Introduction of super-conducting magnets, computers and more recently, ^{13}C and ^{15}N facilities have overcome many of the early problems. In this review, the progress made in determining the conformations of amino acids and their derivatives, small peptides, cyclic peptides, polyamino acids and proteins will be dealt with in that order. For a more complete coverage of biological applications of NMR, several other reviews on various aspects of the subject have appeared, (1 to 13) which complement the material in this chapter.

II. AMINO ACID CONFORMATIONS

Most of the twenty or so common amino acids were examined using NMR as early as 1957–59. (14-16) Comprehensive and up-to-date coupling constants and chemical shifts have been tabulated. (1, 2) However, the conformations of these molecules were not thoroughly investigated by NMR techniques until the spate of papers following the valence-bond calculations of Karplus. (17, 18) Following the apparently successful treatment of simple cyclic organic compounds, Abraham and McLauchlan analysed completely the
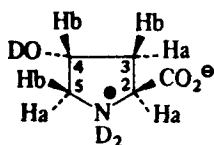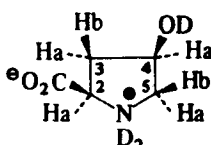
[1]   [2]

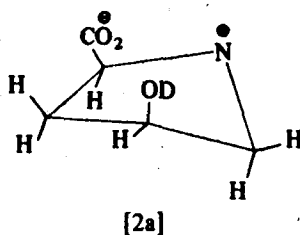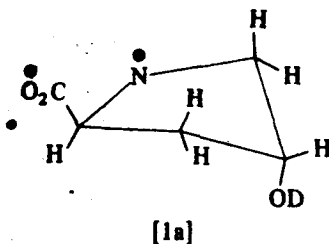|			
J_{2a3a}	7·66 Hz	J_{2a3a}	10·48 Hz
J_{2a3b}	10·44	J_{2a3b}	3·48
J_{3a3b}	−14·06	J_{3a3b}	−14·23
J_{3a4b}	1·41	J_{3a4a}	4·71
J_{3b4b}	4·31	J_{3b4a}	2·09
J_{4b5a}	1·22	J_{4a5a}	4·57
J_{4b5b}	4·09	J_{4a5b}	0·94
J_{5a5b}	−12·69	J_{5a5b}	−12·50
J_{3a5a}	1·6	J_{3b5b}	2·0

60 MHz spectra of aqueous solutions of 4-hydroxy-L-proline [1] and the corresponding *allo*-4-hydroxy-D-proline [2]. (19, 20)

[1a]   [2a]

It was argued that the presence of large vicinal coupling constants in the spectra of both compounds implied that both molecules were relatively rigid in solution and not interconverting rapidly between two or more conformations (leading to time-averaging and hence decrease in the observed coupling constants). Using the accepted "Karplus" type equations:

$$J = K_1 \cos^2 \phi \quad . \quad 0 \leqslant \phi \leqslant 90° \tag{1}$$

$$J = K_2 \cos^2 \phi \quad 90° \leqslant \phi \leqslant 180° \tag{2}$$

where K_1 and K_2 were constants for each particular C–C fragment, and ϕ was the dihedral angle between the C–H bonds concerned, it was proposed that the conformations [1a] and [2a] were the most likely for these molecules in aqueous solution i.e. envelope conformations with four out of five ring atoms in a plane. The assignments for the protons fit these conformations, additional support coming from the presence in each molecule of 4-bond couplings between pseudo-equatorial protons. A similar investigation of these two compounds in acid and alkaline media suggested that in alkaline solution more than one conformation was present in each case, the increase in mobility perhaps arising in part due to the change $ND_2 \rightarrow N-D$. (21) These papers were criticized by Robertson *et al.* in later work on a series of 4-substituted prolines. (22) In the

light of present-day knowledge of stereochemical factors affecting NMR data, it seems that quantitative use of "Karplus equations" was unjustified at the time, though the qualitative features of the proposed conformations were probably correct. Five-membered rings have always proven difficult in conformation-determining exercises. In the paper published by Robertson and his co-workers (22) it was concluded that though the difference between *cis* and *trans* 4-substituted prolines could be rigorously defined by NMR, the individual conformations of these compounds were not easily defined. Similar problems have arisen with 3-methyl, (23, 24) 3-hydroxy, (25) 3,4-epoxy and 3,4-dihydroxy prolines, (26) where J_{2a3a} and J_{2a3b} often have similar magnitudes, making assignment difficult. Abraham and Gatti, however, claimed some success with the determination of the conformation of 3,4-cyclopropyl proline. (27)

In acyclic amino acids of the type $RCH_2CH(\overset{\bullet}{N}H_3)CO_2^{\ominus}$ [3], assuming the classical staggered forms, there are three possible rotational isomers [3a, b and c]. In none of the isomers are H_2 and

[3a]   [3b]   [3c]
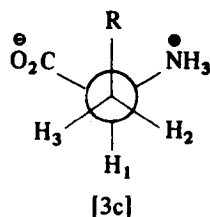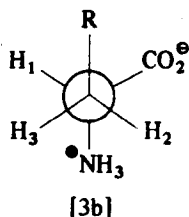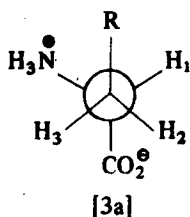
H_3 magnetically equivalent. the β methylene protons thus being diastereotopic. Therefore no matter what the rate of rotation about the $C_\alpha - C_\beta$ bond and no matter what the mole fractions n_a, n_b and n_c may be, protons 1, 2 and 3 should show an ABC type of multiplet in the NMR spectrum. This is usually so except in cases of accidental equivalence, the extent of the non-equivalence depending on the state of ionization of the molecule and the nature of R. Pachler (28, 29) discovered that in amino acids of the type [3] the average vicinal coupling constant $\frac{1}{2}(J_{12} + J_{13})$ was largely independent of the group R and was always approximately 6.3 Hz in magnitude. This figure was arrived at by several independent routes. Making the assumption that $J_{gauche} = J_{12}$ in [a] $= J_{13}$ in [b] $= J_{12}$ or J_{13} in [c] and that $J_{trans} = J_{13}$ in [a] $= J_{12}$ in [b], it was found possible to derive the mole fractions n_a, n_b and n_c using the following equations:

$$n_a + n_b + n_c = 1 \tag{1}$$

$$J_{12} = n_a J_g + n_b J_t + n_c J_g \tag{2}$$

$$J_{13} = n_a J_t + n_b J_g + n_c J_g \qquad (3)$$

$$\therefore n_a = \frac{J_{13} - J_g}{J_t - J_g} \qquad (4)$$

$$n_b = \frac{J_{12} - J_g}{J_t - J_g} \qquad (5)$$

$$n_c = 1 - (n_a + n_b) \qquad (6)$$

If suitable values for J_t and J_g can be inserted in the equations, the mole fractions n_a, n_b and n_c can be estimated. However, the errors implicit in the treatment may be considerable for the following reasons; (i) the initial assumption of one *gauche* and one *trans* coupling constant may be considered doubtful, since it is well-known that electronegativity effects are maximum when the electronegative group is *trans* to one of the coupling protons. It is therefore unlikely that all the *gauche* couplings are identical (for an example note the differences in the vicinal coupling constants in 4-phenyl-1,3-dioxan (30)). (ii) the initial values of J_t (13.6 Hz) and J_g (2.6 Hz), although suggested by perfectly rational argument, seem to be suspect in that J_t in particular has rarely been found to be greater than *ca.* 12 Hz even in systems where no electronegative substituents exist. Conversely the values J_t (6.27) and J_g (2.84) derived from a modified Karplus equation (1) seem to have J_t too small. An intermediate value of the order of 10.5 Hz for J_t and 2.7 Hz for J_g would seem to be more realistic. However, the errors made in the residence times by varying J_t and J_g are not $> 25\%$ over the range suggested, so that the treatment does apparently give a reasonable qualitative picture of the situation. Another assumption made, however, in order that the β-methylene protons be assigned is that the rotamer {a} is always the

Scheme 1

(2S, 3R)β-^{2}H phenylalanine
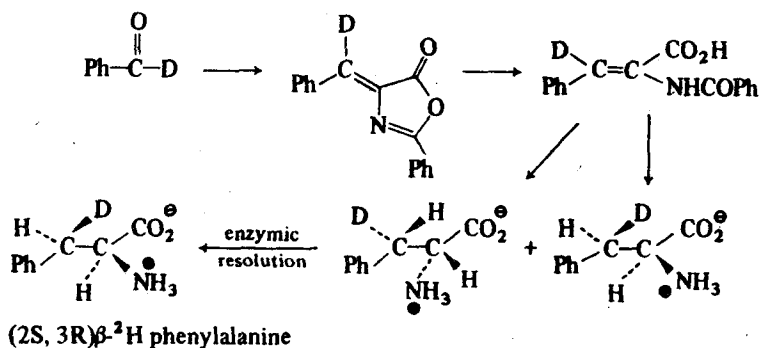
favoured form. If this is the case, then the larger coupling constant will always be J_{13} involving a *trans* coupling in rotamer [a]. The only unambiguous method of assigning H_2 and H_3 is to stereospecifically substitute one of the β-protons for deuterium. This experiment has been performed in the case of phenylalanine ([3], R = Ph) and tyrosine ([3], R = p-C_6H_4OH) by the following synthetic scheme (31, 32, 33) (Scheme 1)
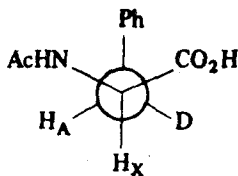
The vicinal coupling constants of the specifically deuterated phenylalanine derivatives [4] and [5] clearly indicate a marked preference for the rotamer ([4a] or [5a]) with the phenyl and carboxyl groups *trans.* These experiments confirm the previous assignments in phenylalanine, and also for tyrosine which appears to adopt a similar preferred rotational form.
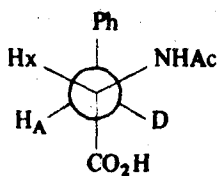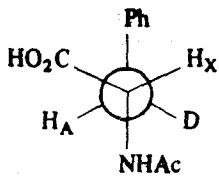
J_{AX}(ave.) = 4·9 Hz
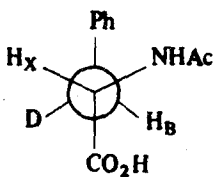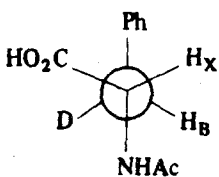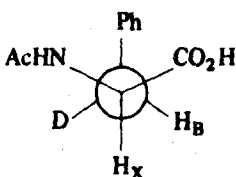
J_{BX}(ave.) = 8·7 Hz

The most detailed analyses of the spectra of acyclic amino acids have been performed on phenylalanine, tyrosine, tryptophan ([3], R = 3-indolyl), histidine ([3], R = 5-imidazoyl) and serine ([3], R = OH). In phenylalanine (34, 35) Cavanaugh found that the temperature variation of the vicinal coupling constants depended critically on the concentration: at low concentration the two vicinal coupling constants diverged with increasing temperature; at high concentration they converged. It was concluded that deviations from classical staggered conformations were small but that the anion rotamer energies varied with temperature and concentration, the two

less favoured rotamers becoming more stable at lower temperatures and lower concentrations.

In the cation, however, the vicinal coupling constants varied very little with temperature and concentration, suggesting that rotamer energies do not vary to the same extent as in the anion. Similar results were found for the tryptophan anion, but for histidine and tyrosine slight differences in behaviour were noted. (36) However, in all cases the two less favourable rotamers became less stable with increase in temperature. These variations may arise from solvent-solute and solute-solute interaction, the precise nature of which were not defined. Although the concentration dependence of the NMR data for serine was not evaluated, changes in the vicinal coupling constants with temperature and pH were observed. (37) Throughout the pH range the rotamer ([3c], R $\rightleftharpoons$ OH) with the $-OH$ and $-CO_2^{\ominus}$ groups *gauche* was always the most stable in contrast to most amino acids, the dominance of this form increasing in the case of the cation. Again the Pachler treatment (28, 29) was used in all this work to derive the mole fractions of the rotational isomers.

The sulphur-containing amino-acids have received much attention, (38-41) the $-SH$ group of cysteine ([3] R = SH), cystine, cysteic acid, lanthionine and djenkolic acid providing novel conformational features. The sulphur substituent is bulky, and in cysteine has an acidic proton which, in alkaline solution, can be removed. Severe repulsion between the $-S^{\ominus}$ and $-COO^{\ominus}$ groups makes the conformation [3b or c] extremely unlikely. This feature is also true in aspartic acid and similar amino acids. Martin and Mathur (39) followed a different procedure from Pachler in examining cysteine, considering limiting cases in the three rotational isomers. However, the assumption of one J_g and one J_t was used once more, and the values adopted by Pachler did not fit the treatment for cysteine. In L-cystine $[-SCH_2CH(\overset{\bullet}{N}H_3)\overset{\bullet}{C}OO]_2$ in acid solution, it was suggested that there is a stabilizing interaction between the two moieties, which is absent in *meso*-cystine.

In valine [6] $[Me_2CH\ CH(\overset{\bullet}{N}H_3)\overset{\ominus}{C}OO]$ where the methyl groups

[6a]   [6b]   [6c]
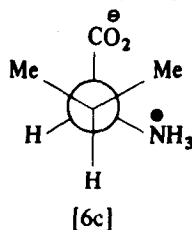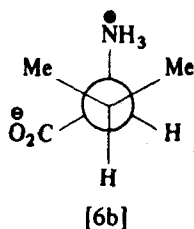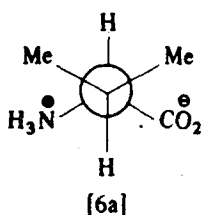
are diasterotopic the vicinal coupling constant varied from 5.1 Hz at
pH 14.0 to 4.3 Hz at pH 1.0. (42, 43) Although the presence of a
bulky group between the two methyl groups as in [6b] and [6c]
would seem to be sterically unfavourable, the magnitude of the
coupling does not indicate a marked preference for the rotamer [6a].

Several papers have appeared (44, 45, 46) in which it has been
pointed out that the presence of an impurity in the racemic
threonine [7] $[CH_3 CH(OH)CH(N\overset{\bullet}{H}_3)C\overset{\bullet}{O}O]$ examined by Aruldhas
(47) led to erroneous conclusions. By analogy with previous work,
(47) it was clearly shown that separate rotameric forms are not seen
in the NMR spectrum of threonine at room temperature and the
impurity in the sample examined by Aruldhas arose from the
presence of some of the *allo* isomer.

Further examples of successful conformational analysis with cyclic
imino acids are found with the six-membered pipecolic acid deriva-
tives [8] and [9]. Shoolery and Virtanen (48) from the 100 MHz
spectrum of the aqueous solution of 4-hydroxy pipecolic acid [8]
and 5-hydroxy pipecolic acid [9] established that the carboxyl and
hydroxyl functions were *trans*. The α-proton splittings for [8] of *ca.*
11 and 5 Hz indicated an axial orientation for the α-proton, and the
narrow septet for the $-CH(OH)$ proton ($\Sigma J = 14$ Hz) clearly sug-
gested an equatorial proton coupled to four neighbours with similar
small couplings (axial-equatorial or equatorial-equatorial). The α-
proton was similarly oriented in 5-hydroxy pipecolic acid [9] but
the $-CH(OH)$ proton was a considerably broader septet ($\Sigma J \simeq$
28 Hz). Thus the chair conformations shown were suggested for
these two molecules. Pipecolic acid [10] itself adopts a similar chair
conformation as expected, though acetylation (49) or nitrosation
(50) of the nitrogen inverts the ring into the other chair form with
the α-carboxyl function axial.

[8]   [9]   [10]
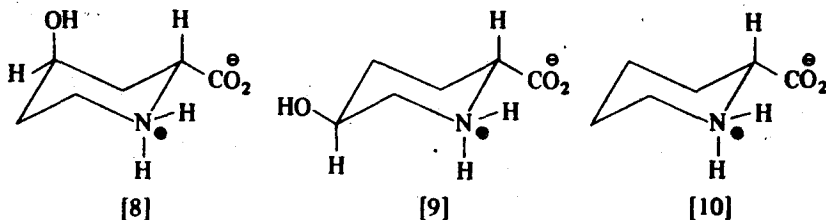
Similar arguments were put forward in assigning the configurations
and conformations of the novel piperazic acids, products of hydroly-
sis of monamycin. (51, 52) The *N*-2,4-dinitro phenyl derivatives of
the chlorine containing piperazic acid [11] and the hydroxyl

containing compound [12] were shown to possess the configurations and conformations indicated using arguments similar to those put forward for the pipecolic acid derivatives. The large $NH \cdot CH$ coupling ($J \simeq 11$ Hz) in both cases (dimethyl sulphoxide-d_6 solvent) suggested that the N–H proton was axially-oriented, presumably to avoid steric interaction with the neighbouring bulky DNP group. Again incorporation of these amino acids into peptide functions reversed the chair conformations, forcing the α-carboxyl functions into the axial position on account of unfavourable steric interactions between the neighbouring amide groups for the equatorial conformation.

A complete analysis of the complex 5-spin spectrum of azetidine-2-carboxylic acid (53) [13] (previously published but not analysed


[11]　　　　　　　　　[12]　　　　　　　　　[13]
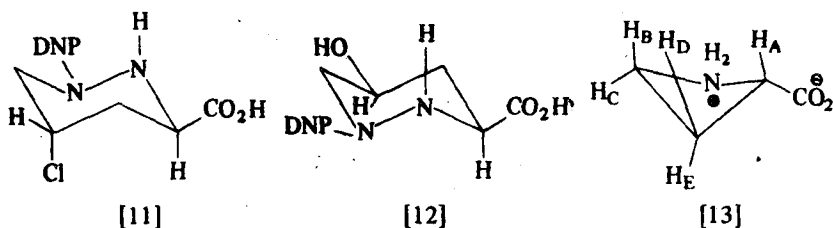
(54)) suggests that as in the crystal, (55) some buckling of the four-membered ring is present.

Although vicinal coupling constants in four-membered rings are often unpredictable, and affected by other factors than dihedral angle and electronegativity relationships, the vicinal couplings are consistent with buckling of the ring as shown [13]. Thus $J_{CD} = 6.02$ Hz, $J_{BE} = 8.23$ Hz, $J_{CE} = 9.74$ Hz and $J_{BD} = 9.34$ Hz in the C_β–C_γ fragment, the difference between J_{CD} and J_{BE} suggesting dihedral angle differences.

Future prospects for studying conformational analysis via the NMR spectra of [15]N (56) and [13]C (57, 58) enriched amino acids have been evaluated to some extent. Lichter and Roberts (56) examined the proton spectra of some ^{15}N enriched alanine, phenylalanine and aspartic acid in order to investigate the usefulness of [15]N–[1]H coupling constants as conformational probes.
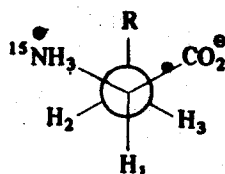
For ^{15}N-alanine (R = H), $J_{NH_1} = J_{NH_2} = 3.1$ Hz; in ^{15}N-phenylalanine (R = Ph), $J_{NH_1} = 3.4$, $J_{NH_2} = 2.9$, and in ^{15}N-aspartic acid (R = COOH), $J_{NH_1} = 3.4$ and $J_{NH_2} = 3.0$ Hz. These couplings are time-averaged for the three rotamers [14a, b and c]. The results, taken with those from other ^{15}N enriched compounds show a trend
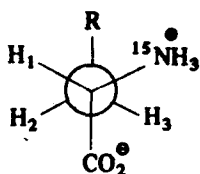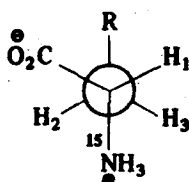
[14a]   [14b]   [14c]

for $^3J_{NH}$ to follow a fairly shallow and somewhat skewed Karplus-type correlation with dihedral angle. However, before this can be taken as a useful technique, in conformation assignment in amino acids and peptides, a lot of "spade-work" is necessary. Fifteen of the naturally-occurring amino acids have been examined and their ^{13}C spectra assigned, using ^{13}C-enriched samples obtained from algae grown in an atmosphere of $^{13}CO_2$. (57) Although proton-decoupled ^{13}C spectra are desirable from the point of view of enhancing the S/N ratio and simplifying the spectra considerably, much conformational information such as one-, two-, three- and four-bond $^{13}C-H$ coupling constants is lost. Potentially, ^{13}C NMR is capable of providing much information about amino acids and peptides which is lacking in ^1H spectra. Tabulated data on amino acids and peptides as provided so far (57, 58) is invaluable in progressing towards this goal.

III. CONFORMATIONS OF *N*-ACYL AND RELATED AMINO ACID DERIVATIVES

Before discussing dipeptide and polypeptide conformations, it is worth examining the progress made in evaluating the conformational features of model compounds, i.e. *N*-acylated amino acids. The partial double-bond character of the amide bond presents a high barrier to rotation, giving rise to geometric isomers as shown [15a, b] normally detectable as an equilibrium mixture in NMR spectra.


[15a]   [15b]
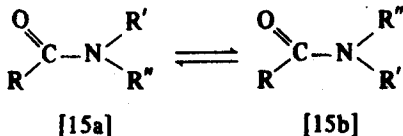
(59) Although the two forms are readily observed as separate species in the NMR spectrum, the assignment of the chemical shifts to one form or the other is by no means obvious. Theory in general proving inadequate, the experimental approach of Paulsen and Todt (60-64)

has enabled a model to be drawn demonstrating the anisotropic effect of the amide group [16]. In this model a proton a experiences greater shielding than in position a'. In the out of plane region, for

[16]
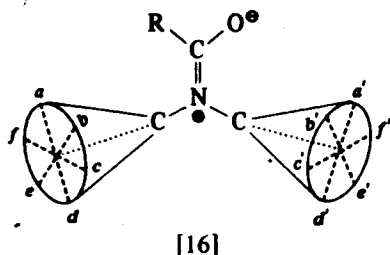
instance, a proton at o' will be more shielded than one at c. In all positions, the experimental data obtained from rigid compounds of known stereochemistry supports this model. Thus in N,N'-dimethyl-formamide the methyl group *cis* to oxygen resonates at lower frequency than the *trans* methyl group, the assignments being proven by Nuclear Overhauser experiments. (65)

In N-acylated amino acids which are examples of monosubstituted amides, the tendency is for the amide bond to strongly prefer the *trans* form ([15a], $R'' = H$) with the N–H bond *trans* to the C=O. (59) *cis* Peptide bonds have rarely been reported in monosubstituted amides of this type in crystalline studies of peptides and proteins, though in small cyclic peptides, as we shall see later, *cis* peptide bonds are often a necessary requirement in the preferred conformation. One of the few examples quoted in the literature demonstrates *cis/trans* isomerism in N-carbobenzoxy derivatives of alanine [17]. (66) It is argued that the *cis* form is stabilized by dimerization through intermolecular hydrogen bonding as shown [18]. This isomerism was not observed at ambient temperatures, when the

[17]   [18]
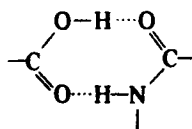
spectrum was that of the averaged species, but was clearly demonstrated at ~0°C.

In the cyclic imino acids, *cis/trans* isomerism about the amide bond is commonly found, since the energy difference between the two forms is governed primarily by the α-carboxy functions, and is