HETEROCYCLES, Vol. 4, No. 1, 1976

IMINO ACIDS AND N-HETEROCYCLIC AMINO ACIDS FROM PLANTS

Leslie Fowden
Rothamsted Experimental Station, Harpenden, Herts, United Kingdom.

Plants produce a large number of imino acids and amino acids with a N-heterocyclic group. The chemistry of these compounds is reviewed, and information about their distribution and biogenesis in plants is provided.

This issue of 'Heterocycles' is dedicated to Professor Tsunematsu Takemoto on the occasion of his retirement from the Faculty of Pharmaceutical Sciences of Tohoku University, and it is a pleasure to contribute this review article in appreciation of his contributions to the chemistry of naturally-occurring heterocyclic compounds. Takemoto's researches covered many groups of natural products, including imino acids derived from proline by further substitution on the pyrrolidine ring. Compounds of this type include kainic and allokainic acids from a seaweed, (Digenea simplex) which have pharmaceutical importance as anthelmintics.[1] The present article will survey the known group of imino acids of plant origin, and also give some account of plant amino acids containing a N-heterocyclic nucleus.

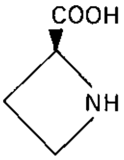
The number of amino and imino acids characterized as plant constituents has increased rapidly in the past 20 years and now some 200 compounds are fully identified.[2] Individual compounds are often restricted (as judged by routine chromatographic survey techniques) to small groups of closely related plants, and information about the distribution of some compounds has been used to strengthen plant taxonomy.

A. Imino Acids

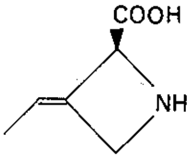
More than 40 naturally-occurring imino acids are known, although a considerable number of these appear confined to micro-organisms, especially to their extra-cellular antibiotic products. Generally, only compounds produced by higher plants will be considered in this article, but an exception is made with polyoximic acid. The imino acids form three main types when classed according to their N-heterocycle: four, five and six atom ring systems are encountered i.e. compounds based on the azetidine, pyrrolidine and piperidine rings. In all cases the C-2 atom shows the L-configuration, but additional substituents introduced onto other ring-C atoms may be in *cis*- or *trans*-configurations relative to the carboxyl group.

1. Compounds based on the azetidine ring

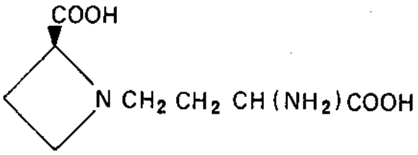
The simplest natural imino acid is L-azetidine-2-carboxylic acid (A2C, I), first isolated and correctly characterized from *Convallaria majalis* (Liliaceae).[3] Subsequently the imino acid has been recognised as a major constituent of many other liliaceous species, some legumes,[4] and of the red alga, *Lophocladia lallemandi*.[5] It also occurs in extremely low concentrations in sugar beet.[6]



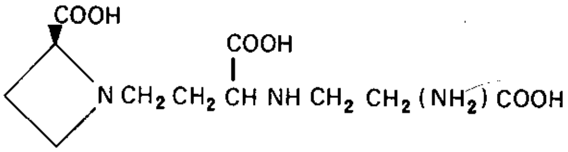
I



II



III



IV

The only simple derivative yet known is polyoximic acid (II), L-3-ethylidene-azetidine-2-carboxylic acid, a component of the polyoxin group of antibiotics produced by Streptomyces cacaoi var asoensis.[7] Structures (III) and (IV), isolated recently from seeds of Fagus silvatica,[8] are N-substituted forms of A2C: they are also N-heterocyclic amino acids. Compound (IV) is confirmed as the correct structure for nicotianamine, which was isolated previously from Nicotiana tabacum.[9]

Labelled precursor feeding experiments have shown that several C_4 amino acids can be converted into A2C in liliaceous plants. Although details of the enzymic pathway are still unknown, it is clear that some form of activation of the Y-C atom must precede cyclization. It is also clear that the C_4 skeleton of methionine is introduced intact into A2C,[10] and S-adenosylmethionine may function as the activated intermediate.

A related study, in which specifically-labelled ^{14}C-isoleucine was introduced into the medium in which Streptomyces cacaoi was grown, has demonstrated that the C_6 branched skeleton of the amino acid is incorporated directly into polyoximic acid; the ethyl residue of isoleucine became the ethylidene group of (II).[11] When an aqueous solution of A2C is heated at 100° for 24 hours, (III) and (IV) are among the products recognised.[8] The production of (III) must take place by a nucleophilic attack of one molecule of A2C with positively charged N upon another with uncharged N. Further polymerization of this type would result in (IV) following an attack on (III) by another molecule of A2C. It is not certain whether similar mechanisms govern the biogenesis of (III) and (IV) in Fagus seeds.

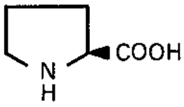
2. Compounds based on the pyrrolidine ring.

The parent compound L-proline (V) is the only imino acid invariably present as a component of protein molecules, where it plays an important contributory role in determining the specific folding of the polypeptide chain and thus the final tertiary structure of the protein molecule. Abnormal replacement of proline residues in protein by analogues impairs their biological activity. In this sense A2C is a particularly effective analogue.[12]

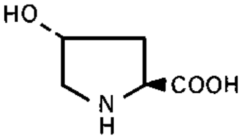
Derivatives of proline consist of a large group having one additional substituent attached to either the C-3 or C-4 atoms, but a few compounds carry substituents on both atoms.

a. *Hydroxy derivatives of Proline*

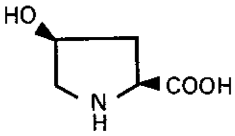
Trans-4-hydroxy-L-proline (VI) is a component of collagen and also of some plant proteins, especially those associated with cell walls.[13] The hydroxyl group is introduced at C-4 by a direct attack by molecular oxygen on a proline residue already present in a polypeptide: the enzymically-catalysed process is dependent upon Fe^{3+}, L-ketoglutarate and ascorbic acid as cofactors. There is no evidence that this isomer of hydroxyproline occurs in an unbound state, but *cis*(*allo*)-4-hydroxy–L-proline (VII) occurs free in *Santalum album*,[14] whilst *trans*-3-hydroxy–L-proline (VIII) represents a major component of the soluble-N fraction of seed and seedlings of *Delonix regia*.[15] In the instances of these unbound hydroxy prolines, no information is available concerning biosynthetic mechanisms: a cyclization mechanism akin to that resulting in proline formation from glutamic acid might operate to form (VII) from *threo*-γ-hydroxy–L-glutamic acid, which is recognised as a plant product.



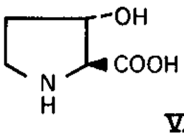
V



VI



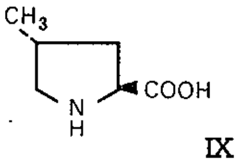
VII



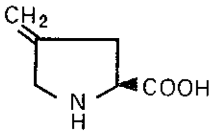
VIII

b. Methyl and similar derivatives of proline

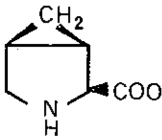
Trans-4-methyl-L-proline (IX) was first isolated from apples[16] and occurs more widely in the fruits of species assigned to the family, Rosaceae. (Cis-4- and cis-3-methyl-L-prolines have been characterized as constituents of antibiotics). The trans-4 isomer could arise by metabolic cyclization of erthyro-γ-methyl-L-glutamic acid, a compound occurring in several plant families but not confirmed in the Rosaceae. A similar biogenetic mechanism might be responsible for the formation of 4-methylene-L-proline (X) from γ-methyleneglutamic acid. (X) was isolated from another rosaceous fruit, loquat (Eriobotrya japonica); by isomerization, it might give rise to cis-3,4-methanoproline (XI), a component of seed of Aesculus parviflora, whilst addition of a molecule of water across the ethylenic link could yield cis-4-hydroxymethyl-L-proline (XII), another imino acid encountered in apple fruits.[19]



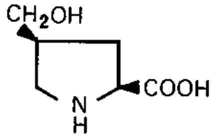
IX



X



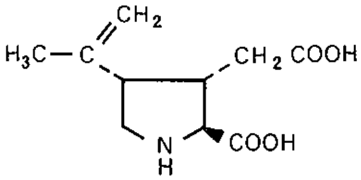
XI



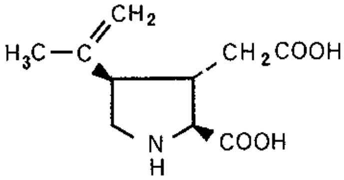
XII

c. Doubly substituted proline derivatives.

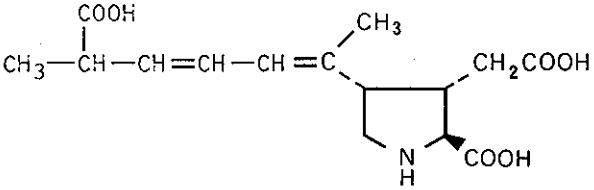
The most important members of this class consist of L-γ-kainic acid (XIII) and its isomer, L-γ-allokainic acid (XIV), isolated from the seaweed, Digenea simplex,[20-22] and the related compound, domoic acid (XV), obtained from another seaweed, Chondria armata.[23] There is no information providing an insight into the metabolic pathways involved in the biosynthesis of these acidic derivatives.



XIII



XIV



XV

3. Compounds based on the piperidine ring.

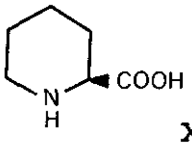
The parent imino acid of this group, L-piperidine-2-carboxylic acid (L-pipecolic acid, XVI) occurs in high concentration in a few legume species, especially in seeds, and as a minor component of plants of several other families. Biosynthesis from lysine proceeds rapidly in some plant tissues.[24]

Unlike A2C, pipecolic acid exhibits little or no metabolic antagonism towards proline, never seems to become incorporated into protein molecules, but may substitute for proline in the actinomycin series of small antibiotic peptides.[25]

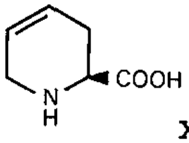
L-4,5-dehydropipecolic acid (baikiain, XVII) occurs together with (XVI) and its hydroxy derivatives (see below) in seeds of *Baikiaea plurijuga*[26] and certain other legume species.

a. Hydroxy derivatives of pipecolic acid.

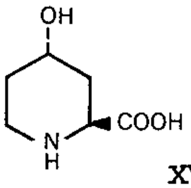
Trans-4- and *trans*-5-hydroxy-L-pipecolic acids (XVIII and XIX, respectively) conceivably arise by addition of a molecule of water across the C=C linkage of (XVII). The initial isolation of the *trans*-5 derivative was from the palm, *Rhapis flabelliformis*,[27] and *Baikiaea plurijuga*,[28] whilst the 4-hydroxy isomer was obtained from *Acacia* species.[29] More recently, two dihydroxypipecolic acids have been characterized: *cis*-4,5-dihydroxy-L pipecolic acid (2(S), 4(R), 5(S)) (XX) is a constituent of leaves of *Calliandra haematocephala* (Mimosaceae)[30] whilst the *trans*-4,5-dihydroxy derivative (2(S), 4(S), 5(S)) (XXI) is present in leaves of the legume, *Derris elliptica*,[31] together with (XVI), (XVIII), (XIX) and (XX).



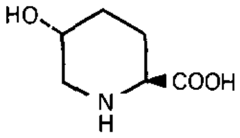
XVI



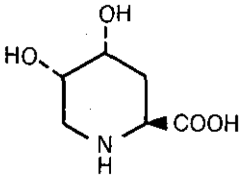
XVII



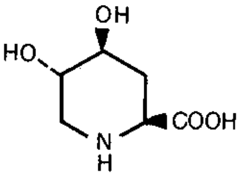
XVIII



XIX



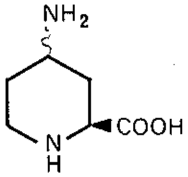
XX



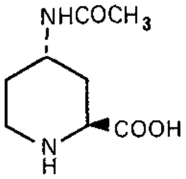
XXI

b. Amino derivatives of pipecolic acid

A 4-aminopipecolic acid (XXII) of unknown stereochemical configuration was isolated from Strophanthus scandens,[32] a species that also contains trans-4-hydroxy-L-pipecolic acid. A recent report has shown that trans-4-acetylamino-L-pipecolic acid (XXIII) occurs in C. haematocephala leaves.[31] The simple 4-amino derivative is then likely to be the trans-L isomer.



XXII



XXIII

Whilst 4-ketopipecolic acid, which could function as an intermediate between the 4-hydroxy and 4-amino derivatives, has not been identified as a constituent of higher plants, it is a product of microbial metabolism.

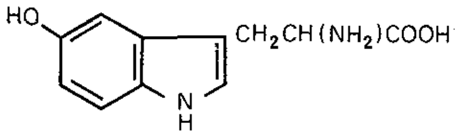
B. N-Heterocyclic Amino Acids

Two amino acids, histidine and tryptophan, containing N-heterocycles are constituents of protein molecules, and therefore are universally synthesized by plants. Normally, they are encountered free in only small concentrations relative to other protein amino acids. The α-N-malonyl derivative of D-tryptophan also occurs free in plants. Both histidine and tryptophan may be regarded as examples of β-substituted alanines, and many of the other known N-heterocyclic amino acids isolated from plants are also of this type.

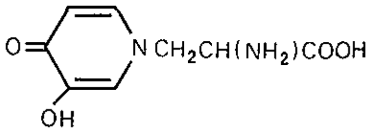
1. Compounds with one ring-N atom

5-Hydroxy-L-tryptophan (XXIV) was isolated from leaves of the legume species, Griffonia simplicifolia.[33] There is evidence suggesting that it is formed by direct hydroxylation of tryptophan. Decarboxylation produces the physiologically active amine, 5-hydroxytryptamine.

The pyridine nucleus is encountered in mimosine (XXV), first isolated from Mimosa pudica,[34] and later shown to be present in high concentration in leaves and seed of many Leucaena species. The amino acid accumulates in seed of L. leucocephala to represent 6-8% of the dry weight. It is toxic to animals causing shedding of the fleece in sheep, loss of hair in cattle, and loss of the tail hair in horses. Liver lesions also occur in sheep after prolonged grazing of Leucaena foliage. The β-D-glucoside derivative may coexist with mimosine.



XXIV

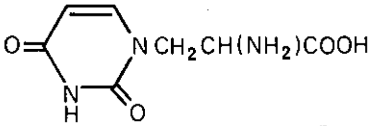


XXV

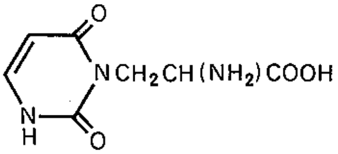
2. Compounds with two ring-N atoms

Three amino acids of this class contain a pyrimidine nucleus. Willardiine (XXVI) was isolated first from seed of Acacia willardiana.[35] Subsequently, it was identified in a few additonal Acacia spp. and it is also a constituent of peas.[36] The isomeric compound, isowillardiine (XXVII), also occurs in peas. Lathyrine (XXVIII) is a characteristic component of a sub-generic group of species of Lathyrus (Leguminaceae), being isolated and characterized initially from L. tingitanus.[37] The ring-NH_2 group confers basic properties on lathyrine, and the compound gives an unusual red chromophore when reacted with ninhydrin.

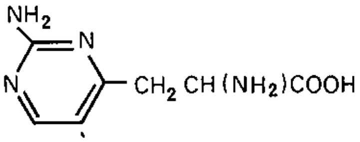
L-β-pyrazol-1-ylalanine (XXIX) is an isomer of histidine. The compound was isolated first from seed of watermelon, Citrullus vulgaris;[38] subsequently, it has been recognised as a component of some, but not all members of the Cucurbitaceae, where its distribution accords with the usual taxonomic classification of the family. The amino acid was the first natural product to be characterized possessing the pyrazole ring system (subsequently alkaloids containing a pyrazole nucleus have been identified). Cucurbit seeds normally contain significant amounts of γ-L-glutamyl-β-pyrazolylalanine together with the free amino acid.



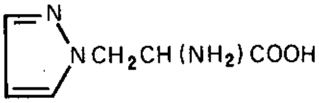
XXVI



XXVII



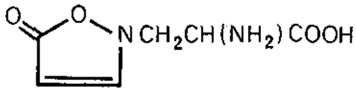
XXVIII



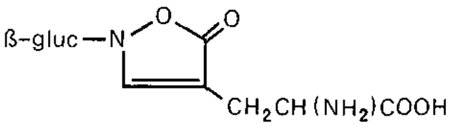
XXIX

3. Compounds with rings containing N and O atoms

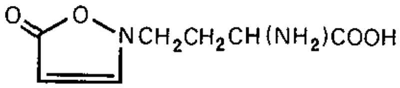
Amino acids containing the isoxazolinone nucleus were first isolated from seedlings of Pisum sativum (pea). Initially, β-isoxazolin-5-one-2-ylalanine (XXX) and β-(2-β-D-glucopyranosyl)-isoxazolin-5-one-4-ylalanine (XXXI) were obtained. Like other members of this group, they are characteristically unstable in alkaline solution and under UV irradiation. Both compounds are found also in seedlings of *Lens culinaris*, *Pisum arvense* and *Lathyrus odoratus*.[39] A member of other isoxazolinone derivatives have been identified as constituents of *Lathyrus odoratus* seedlings: the new compounds include α-amino-γ-isoxazolin-5-one-2-ylbutyric acid (XXXII, the higher homologue of XXX), 2-aminoethylisoxazolin-5-one (the decarboxylation product of XXX) and its γ-glutamyl derivative, 2-cyanoethylisoxazolin-5-one and 2-carboxymethylisoxazolin-5-one. β-Aminopropionitrile, a lathyrogenic constituent of *Lathyrus odoratus* is a photochemical decomposition product of 2-cyanoethylisoxazolin-5-one.



XXX



XXXI



XXXII

Three other amino acids possess a glycine residue substituted on the α-C atom by a heterocyclic ring system: they are tricholomic acid, ibotenic acid and muscazone, products of macro-fungi that are toxic to houseflies.

4. Biosynthesis of β-substituted alanines

Work in several laboratories has established that the C_3 side-chain of many β-substituted alanines has its biogenetic origin in a molecule of serine. The experimental evidence is of two types. Precursor feeding experiments using ^{14}C-labelled serine has confirmed that the C_3 skeleton is introduced into compounds (XXV), (XXVI), (XXVII) and (XXIX). For synthesis to proceed efficiently, activation of the serine molecule is probably necessary. In the case of compounds (XXX) and (XXXI), enzymic synthesis in pea seedling extracts requires O-acetylserine as the C_3 donor,[39,40] and possibly this mechanism is more general.

Biosynthesis of lathyrine follows a quite distinct route. Homoarginine and γ-hydroxyhomoarginine act as precursors, cyclization presumably occurring from the latter by elimination of a molecule of water from the γ-hydroxy and terminal guanidino group.[41]

C. General Considerations

The foregoing account has described more than thirty heterocyclic-N compounds elaborated by plants, and indicated the very diverse biosynthetic pathways utilized by plants. Many of the compounds accumulate particularly in seeds, where they may act as temporary N-storage products, being degraded on germination and growth of the seedling. However, more generally the compounds seem to be metabolized (synthesized or degraded) rather slowly in comparison with most protein amino acids.

A number of compounds show physiological activity, being inhibitory to the growth of bacteria or fungi or toxic to higher forms of life. Therefore, although their precise role in the physiology of the plants producing them frequently remains uncertain, there is an increasing amount of indirect evidence of suggesting they have evolutionary importance in providing a degree of protection against predating insects or other animal life, or in enhancing the competitive ability of the producer plant when establishing a particular niche in the vegetation of a given area.

The compounds described have been recognised largely in the last 20 years, and several new compounds are described each year. Given the increasing sensitivity of chromatographic techniques, it would seem certain that the numbers of imino acids and N-heterocyclic amino acids known as plant products will increase considerably in the next decade.

REFERENCES

1 T. Takemoto, 'Kainic acid', a pamphlet reprinted by the Shiraishi Pharmaceutical Companies, Osaka, Japan, (undated).

2 L. Fowden, In 'Recent Advances in Phytochemistry', edit. by V. C. Runeckles and E. E. Conn, Academic Press, New York, 1974, p. 95

3 L. Fowden, Nature, London, 1955 176, 347

4 M-L. Sung and L. Fowden, Phytochemistry, 1969, 8, 2095

5 G. Impellizzeri, S. Mangiafico, G. Orienta, M. Piattelli, S. Sciuto, E. Fattorusso, S. Magno, C. Santacroce and D. Sica, Phytochemistry, 1975, 14, 1549

6 L Fowden, Phytochemistry, 1972, 11, 2271

7 K. Isono, K. Asahi and S. Susuki, J. Am. Chem. Soc, 1969, 91, 7490

8 I. Kristensen and P. Olesen-Larsen, Phytochemistry, 1974, 13, 2791

9 M. Noma, M. Noguchi and E. Tamaki, Tetrahedron Letters, 1971, p.2017

10 E. Leete, G. E. Davis, C. R. Hutchinson, K. W. Woo and M. R. Chedekel, Phytochemistry, 1974, 13, 427

11 K. Isono, S. Funayama and R. J. Suhadolnik, Biochemistry, 1975, 14, 2992

12 L. Fowden, D. Lewis and H. Tristram, Advanc. Enzymol. 1967, 29, 89.

13 D. T. A. Lamport, Advance. Bot. Res. 1965, 2, 151

14 A. N. Radhakrishnan and K. V. Giri, Biochem. J. 1954, 58, 57.

15 M-L. Sung and L. Fowden, Phytochemistry, 1968, 7, 2061

16 A. C. Hulme and W. Arthington, Nature, London, 1952, 170, 659

17 D. O. Gray and L. Fowden, Nature, London, 1962, 193, 1285

18 L. Fowden, A. Smith, D. S. Millington and R. C. Sheppard, Phytochemistry 1969, 8, 437

19 A. B. Mauger and B. Witkop, Chem. Rev. 1966, 66, 47

20 S. Murakami, T. Takemoto and Z. Shimuzu, J. Pharm. Soc. Japan, 1953, 73, 1026

21 K. Tanaka, M. Miyamoto, M. Honjo, H. Morimoto, T. Sugawa and M. Uchibayashi, Proc. Japan. Acad. 1957, 33, 47

22 Y. Ueno, K. Tanaka, J. Ueyanagi, H. Nawa, Y. Sanno, M. Honjo, R. Nakamori T. Sugawa, M. Uchibayashi, S. Osugi and S. Tatsuoka, Proc. Japan. Acad. 1957, 33, 53.

23 T. Takemoto, K. Daigo, Y. Kondo and K. Kondo, J. Pharm. Soc. Japan, 1966, 86, 874.

24 L. Fowden, J. Exp. Bot. 1960, 11, 302

25 E. Katz and W. A. Goss , Biochem. J. 1959, 73, 458

26 F. E. King, T. J. King and A. J. Warwick, J. Chem. Soc. 1950, p. 3590

27 A. I. Virtanen and S. Kari, Acta Chem. Scand. 1954, 8, 1290

28 N Grobbelaar, J. K. Pollard and F. C. Steward, Nature, London, 1955, 175, 703

29 J. W. Clark-Lewis and P. I. Mortimer, J. Chem. Soc. 1961, p.189

30 M. Marlier, G. A. Dardenne and J. Casimir, Phytochemistry, 1972, 11, 2597

31 G. A. Dardenne, Phytochemistry, 1975, 14, 860

32 W. Schenk and H. R. Schütte, Naturwiss, 1961, 14, 14

33 E. A. Bell and L. E. Fellows, Nature, London, 1966, 210, 529

34 J. Renz, Hoppe-Seyler's Z. physiol. Chem. 1936, 244, 153

35 R. Gmelin, Hoppe-Seyler's Z. physiol. Chem. 1959, 316, 164

36 F. Lambein and R. Van Parijs, Biochem. Biophys. Res. Commun.

37 E. A. Bell and R. G. Foster, Nature, London, 1962, 194, 91

38 F. F. Noe and L. Fowden, Biochem. J. 1960, 77, 543

39 F. Lambein, Phytochemistry, 1975, 14, 856

40 I. Murakoshi, F. Kato, J. Haginawa and L. Fowden, Chem. Pharm. Bull. (Tokyo) 1973, 21, 918

41 E. A. Bell and J. Przybylska, Biochem. J. 1965, 95, 35P.

Received, 8th November, 1975