orientation in the apolactoferrin cell is related to that in the diferric lactoferrin cell by a rotation of 90° about the *a* axis, 20° about *b*, followed by 20° about *c*. The similarity in unit cells seems to be largely coincidental, although in the *c* direction at least it may be attributed to the ellipsoidal shape of the molecules; a rotation of $\sim 90^\circ$ about *a* does not change the *c* dimensions of the molecule.

Finally, although molecular replacement has been successful in this structure analysis, a recent elaboration of the method, in which molecular dynamics calculations are used to give flexibility to a search model (Brünger, 1990) may make problems of this type much more amenable to solution.*

We wish to thank Dr P. M. Colman for helpful discussions on both the molecular replacement procedures and the rigid-body refinement, and Mrs E. J. Dodson for helpful advice on the molecular replacement methods and the *CCP*4 program package. We are grateful to Dr P. F. Lindley and members of the Birkbeck College transferrin group for help, encouragement and provision of unpublished results, and we gratefully acknowledge financial support from the US National Institutes of Health (grant HD-20859), the Medical Research Council of New Zealand and the New Zealand Dairy Research Institute.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory, and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37051 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

References

- AISEN, P. & LISTOWSKY, I. (1980). Annu. Rev. Biochem. 49, 357-393.
- ANDERSON, B. F., BAKER, H. M., DODSON, E. J., NORRIS, G. E., RUMBALL, S. V., WATERS, J. M. & BAKER, E. N. (1987). Proc. Natl Acad. Sci. USA, 84, 1769–1773.
- ANDERSON, B. F., BAKER, H. M., NORRIS, G. E., RICE, D. W. & BAKER, E. N. (1989). J. Mol. Biol. 209, 711-734.
- ANDERSON, B. F., BAKER, H. M., NORRIS, G. E., RUMBALL, S. V. & BAKER, E. N. (1990). Nature (London), 344, 784–787.
- BAILEY, S., EVANS, R. W., GARRATT, R. C., GORINSKY, B., HASNAIN, S., HORSBURGH, C., JHOTI, H., LINDLEY, P. F., MYDIN, A., SARRA, R. & WATSON, J. L. (1988). *Biochemistry*, 27, 5804–5812.
- BAKER, E. N. & RUMBALL, S. V. (1977). J. Mol. Biol. 111, 207-210.
- BAKER, E. N., RUMBALL, S. V. & ANDERSON, B. F. (1987). Trends Biochem. Sci. 12, 350-353.
- BROCK, J. H. (1985). *Metalloproteins*, Part II, edited by P. HARRISON, pp. 183-262. London: MacMillan.
- BRÜNGER, A. T. (1990). Acta Cryst. A46, 46-57.
- CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178, New York: Gordon and Breach.
- EKLUND, H., SAMAMA, J.-P., WALLEN, L., BRANDEN, C.-I., AKESON, A. & JONES, T. A. (1981). J. Mol. Biol. 146, 561 587.
- HANSON, J. C., WATENPAUGH, K. D., SIEKER, L. & JENSEN, L. H. (1979). Acta Cryst. A35, 616–621.
- JONES, T. A. (1978). J. Appl. Cryst. 11, 268-272.
- KILAR, F. & SIMON. I. (1985). Biophys. J. 48, 799-802.
- Norris, G. E., Baker, H. M. & Baker, E. N. (1989). J. Mol. Biol. 209, 329-331.
- North, A. C. T., Phillips, D. C. & Mathews, F. S. (1968). Acta Cryst. A24, 351-359.
- ROSSENAU-MOTREFF, M. Y. F., SOETEWEY, R., LAMOTE, R. & PEETERS, H. (1971). *Biopolymers*, **10**, 1039–1048.
- SUSSMAN, J. L. (1985). Methods Enzymol. 115, 271-303.
- TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). Acta Cryst. A43, 489-501.
- VIGH, R., CSER, L., KILAR, F. & SIMON, I. (1989). Arch. Biochem. Biophys. 275, 181-184.
- WANG, B. C. (1985). Methods Enzymol. 115, 90-112.

Acta Cryst. (1991). B47, 1004-1010

Polymorphism of L-Glutamic Acid: Decoding the $\alpha-\beta$ Phase Relationship via Graph-Set Analysis

BY JOEL BERNSTEIN

Department of Chemistry, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

(Received 6 February 1991; accepted 30 July 1991)

Abstract

Graph-set analysis is employed to study the similarities and differences in the hydrogen-bonding patterns of the two polymorphs of L-glutamic acid. There are considerable similarities between the two structures, and only the higher-order graph sets reveal the differences. These are used to gain insight into the crystallization chemistry of the two forms as well as the structural relationship between them and the significantly different molecular conformations found in them.

0108-7681/91/061004-07\$03.00

0 © 1991 International Union of Crystallography

Introduction

L-Glutamic acid (LGA) is known to be at least dimorphic (Hirokawa, 1955), the two polymorphs being designated α and β . The structures of both



forms have been determined (Lehmans, Koetzle & Hamilton, 1972; Hirayama, Shirahata, Ohashi & Sasada, 1980), the β form being apparently the more stable. It is obtained when the material is crystallized slowly *i.e.* under equilibrium conditions. The α form may be obtained by the addition of trace amounts of impurities, especially amino acids, to the crystallizing solution (Hasegawa, Fukuda, Higuchi & Matsubara, 1977; Hiramatsu, 1977*a*,*b*); it is converted to the β form upon standing in solution for a long time. Hirayama, Shirahata, Ohashi & Sasada (1980) attempted to compare the two structures and proposed a structural pathway for the $\alpha \rightarrow \beta$ transformation.

The crystal structures of LGA will clearly be dominated by hydrogen bonds, and determination of the differences in hydrogen-bonding patterns between the two forms provides the way to understanding the structural differences. In this case the hydrogen-bonding patterns are generated from combination of four hydrogen-bond donors and four hydrogen-bond acceptors.

Recently, it has been shown that graph-set analysis can be a very powerful technique for decoding the hydrogen-bonding patterns in organic crystals (Etter, 1990; Etter, MacDonald & Bernstein, 1990; Bernstein, Etter & MacDonald, 1990). In polymorphic systems it can be used to readily identify the differences in structural patterns which distinguish multiple structures of a chemical entity. In this paper we apply this technique to unravel the hydrogenbonding patterns and the polymorphic crystal chemistry of LGA.

Molecular structures

The molecular structures of the two forms are shown in Fig. 1. As for all amino acids the molecule is in the zwitterionic form in both structures. The conformations are clearly very different in the two polymorphs, so that this pair of structures constitutes an example of *conformational polymorphism* (Bernstein, 1987). In quantitative terms the

significant differences in molecular conformation are essentially localized in three torsion angles, N1-C2-C3-C4, C2-C3-C4-C5, C3-C4-C5-O3, with corresponding values for the α and β forms of 178.4, -51.8; 68.8, -73.1; -150.0, -160.7° , respectively. In general, the torsion angles can take on a variety of values, which fall into six quite well defined molecular conformations for rotations about C2-C3 and C3-C4. These have been observed in various crystal structures of salts and complexes of LGA: (\overline{g}, t) , (g, t), (\overline{g}, g) , (g, g), (t, t) and (t, g), where g, \overline{g} and t correspond to approximate torsion angles of 60, -60 and 180° . In the α form the value for the torsion angle about C2–C3 (-105.0°) falls between the g and t conformations, so it may be considered as constituting an outlier from the conformations found in a variety of LGA structures and complexes. The conformation of the molecule in the β form is (\overline{g}) .

Crystal structures

Both forms crystallize in space groups $P2_12_12_1$. Examination of the experimental cell constants (given in parentheses) suggests the following relationship between the two structures: a_{α} (7.068 Å) = c_{β} (6.948 Å); b_{α} (10.277 Å) = $2a_{\beta}$ (2 × 5.519 Å); $2c_{\alpha}$ (2 × 8.755 Å) = h_{β} (17.30 Å). The packing diagrams in Fig. 2 are oriented to reflect the possible structural relationship suggested by the cell constants. Inspection of the two packing diagrams indicates quite clearly that a simple relationship between the two structures does not in fact exist.

The dominant intermolecular interactions in these structures clearly must be the hydrogen bonds, and the structural differences and similarities should be apparent from an examination of the hydrogenbonding patterns. Traditionally, one of the most



Fig. 1. Stereoviews of L-glutamic acid in (a) the α form and (b) the β form. In both cases the view is on the plane of C2--C3--C4, which highlights the conformational differences. Hydrogens have been omitted for clarity.

common methods for comparison of the hydrogenbonding patterns has been the tabulation of the geometric characteristics of the hydrogen bonds, as in Table 1. It is quite apparent from examination of this table that such an approach does not provide very much additional insight into the similarities and differences between the two structures. The application of graph-set analysis is a straightforward method for making those distinctions.

Hydrogen bonding – application of graph theory

The procedures for identifying the different kinds of hydrogen bonds and assigning the graph sets have been outlined previously (Etter, 1990) and an exemplary case has been presented (Bernstein, Etter & MacDonald, 1990). LGA has four hydrogen-bond donors and four potential acceptors. Of the latter, two are carboxylate oxygens, one is a carbonyl oxygen and the fourth is a hydroxyl oxygen. The last is a relatively poor acceptor and in fact does not participate in a hydrogen bond in either structure. The fact that all four donors do participate in hydrogen bonds requires that in both cases one of the remaining three oxygens participates in two hydrogen bonds.

To proceed with the analysis each hydrogen-bond type in turn is isolated from the others and a map is prepared showing all the repetitive occurrences of







Fig. 2. Stereo packing diagrams for the two forms of LGA. The unit cells are oriented to demonstrate the possible geometric relationship between the cell constants in Table 1 (see text). (a) α form, (b) β form.

 Table 1. Hydrogen-bond geometries in the two forms
 of LGA

Interaction	Symmetry*	Distance (Å)	Interaction	Distance (Å)
α form				
H1…O1	3 - 101	1.81	N…O1	2.77
H2…O4	41 11	2.18	N…O4	2.89
H3…O2	2010	1.92	N…O2	2.81
H3O…O2	4100	1.56	O3…O2	2.57
β form				
H1…O2	2000	1.84	N…O2	2.87
H2…O2	1100	1.87	N…O2	2.89
H3…O4	3002	1.90	N…O4	2.92
HO3…O1	4101	1.48	0301	2.52

* Symmetry operations include the following space-group symmetry operator followed by the suitable translations along the *a*, *b* and *c* directions. (1) *x*, *y*, *z*; (2) $\frac{1}{2} - x$, -y, $\frac{1}{2} + z$; (3) $\frac{1}{2} + x$, $\frac{1}{2} - y$, -z; (4) *x*, $\frac{1}{2} + y$, $\frac{1}{2} - z$.

that single type of hydrogen bond. The set of molecules that is connected by this kind of hydrogen bond is the characteristic motif, which is assigned to one of four graph types, determined by whether the motif is intramolecular (S), infinite (C), cyclic (R) or dimeric (D) (Etter, 1990; Etter, MacDonald & Bernstein, 1990). The numbers of proton donors and acceptors involved in the motif are assigned as suband superscripts (the default value is one) and the degree or size of the motif is included in the notation in parentheses. Motifs are always composed of only one hydrogen-bond type, as opposed to networks, which contain multiple types of hydrogen bonds.

These principles are demonstrated in the scheme below, for a hypothetical molecule that contains one acceptor (A) and two hydrogens which can participate in the formation of hydrogen bonds. The hydrogen-bond motif for the first of these hydrogens, H1, is an eight-membered ring, with two donors and two acceptors, hence $R_2^2(8)$. The second hydrogen, H2, participates in a four-atom chain motif with one donor and one acceptor (the designator default values), hence C(4). These two motifs comprise the first-order graph set, $N_1 = C(4)R_2^2(8)$.



In the α structure, the hydroxyl hydrogen is in a chain motif C(8). Two of the amino hydrogens participate in C(5) chain motifs while the third amino hydrogen is in a C(7) chain motif. The motifs are presented schematically in Fig. 3. These four hydrogen-bond motifs comprise the first-order graph set for the α form. The corresponding analysis for the β form yields precisely the same graph set. Hence the first-order graph sets for the two forms are identical. Figs. 4 and 5 show the first-order network for the α and β forms of LGA, respectively. The identity of the first-order graph sets for polymorphic structures was also observed in two of the three polymorphs of iminodiacetic acid (Bernstein, Etter & MacDonald, 1990).

In such a case it is necessary to turn to higherorder graph sets to determine the hydrogen-bonding patterns which may be used to distinguish between the forms. Higher order graph sets are determined in a manner similar to the first-order ones. However, in this case each motif comprises a graph set of nth order. Determination of the graph set involves starting at one hydrogen-bonded H atom and tracing out a path through the first hydrogen bond, along an atomic path (which can be intra- or intermolecular) until reaching a second hydrogen-bonded H atom, proceeding through its hydrogen bond and continuing along an atomic path until the first hydrogenbonded hydrogen is needed again. The pathway is not necessarily a cyclical one; 'returning to the first atom' simply means that the pathway leads to a chemically identical (and identically labelled with respect to the starting molecule) H atom. In the scheme above, the second-order graph set clearly contains both H1 and H2 in an eight-membered ring with four donors and two acceptors, denoted by $N_2 = R_4^2(8)$.

The first difference between the two polymorphs shows up in the second-order graph sets. For the α form the set is $C_2^2(6)$ while that for the β form is $C_2^1(4)$. The short chain length for the β form is due to two facts: first, one oxygen serves as a single acceptor for both hydrogen bonds, and second, both hydrogen donors are bonded to the same (nitrogen) atom. Incidentally, this interaction forms a screwgenerated chain along the crystallographic c axis. The graph sets for both forms are summarized in Table 2.

The next five orders N_3 - N_7 are given in short form in Fig. 3. All of them for both forms are chains. The



Fig. 3. Hydrogen-bond patterns and the graph sets for LGA.

rules for determining the graph sets (Etter, 1990; Etter, MacDonald & Bernstein, 1990) assign the lowest degree (*i.e.*, shortest) chains to the lowest order and the increasingly longer chains to higher orders. It is important to point out here that the prescription for determining the hierarchy of the orders of graph sets does not in any way determine the chemical or structural priority of the hydrogenbond patterns; it does provide a useful method of analysis and comparison. Moving to the next higher order N_3 , the assignment is identical for the two forms. In N_4 the chain length in the α form is shorter by one atom than in the β form; again in the former



Fig. 4. Disgram of the α form of LGA showing the first-order network.



Fig. 5. Diagram of the β form of LGA showing the first-order network.

 Table 2. Summary of graph-set assignments for LGA polymorphs

Order	α form	$\boldsymbol{\beta}$ form
N_1	C(5)C(5)C(7)C(8)	C(5)C(5)C(7)C(8)
N_2	$C_{2}^{2}(6)$	$C_{2}^{1}(4)$
N_3	$C_{2}^{2}(9)$	$C_{2}^{2}(9)$
N_4	$C_{2}^{1}(9)$	$C_{2}^{2}(10)$
N_5	$C_{2}^{2}(10)$	$C_{2}^{2}(10)$
N_6	$C_{2}^{2}(10)$	$C_{2}^{2}(11)$
N_7	$C_{2}^{2}(11)$	$C_{2}^{2}(11)$
N_8	$R_{3}^{3}(14)$	$C_{3}^{3}(10)$

 $[C_2^1(9)]$ one oxygen acts as an acceptor in two hydrogen bonds. The identity of the graph sets returns in N_5 and N_7 , N_6 again being distinguished by a difference in only one carbon, and again in favor of the α form.

To this point, the graph-set distinction between the two forms at higher levels does not provide a great deal of chemical information which might serve as a basis for understanding the crystallization phenomena described in the Introduction. Through N_7 all of the graph sets are chain structures of identical length or differing in length by only one atom. For this insight we turn to the next higher order graph set N_8 (Table 2). Since all the pairwise combinations have been exhausted through N_7 , we must now turn to triple combinations of hydrogen bonds. The method for determining the graph set is similar to that described for the pairwise combinations, but in this case three hydrogen bonds must be traversed before returning to the original chemically identical H atom.

For the β form N_8 is another chain structure, $C_3^3(10)$ (Fig. 7), perhaps surprisingly short for a motif containing three hydrogen bonds. However N_8 for the α form is $R_3^3(14)$ (Fig. 6), the first hydrogenbonded ring structure encountered in this polymorphic system. This is a very clear and significant structural difference between the two forms. Clearly, in such a complex structure many ring graph sets must exist, especially at higher orders, and two of them, $R_3^3(17)$ and $R_3^3(18)$ are shown in Fig. 8. There are no ring structures in the β form which contain less than 17 atoms. The reason for the absence of smaller ring motifs in the β form is due to the conformation of the molecule. The conformational difference in the two forms is manifested primarily in the torsion angles N-C2-C3-C4 and C2-C3-C4—C5. In the α form these two angles have values which lead to a folded molecular conformation in which the intramolecular distance O2...O3 (on opposite ends of the molecule) is 3.34 Å. This folding is the conformation which allows the molecule to participate in the rather small cyclic $R_3^3(14)$ graph arrangement. The key role of this folding in forming the ring may be seen in Fig. 6. The distances in the β form between oxygen atoms on different carboxyl groups are greater than 6.5 Å, reflecting the clearly more extended conformation found in this structure, so that more than 14 atoms are required for any cyclic arrangement at this level of hydrogen bonds.

Let us now return to the crystallization chemistry in order to examine its relationship to the graph sets which have just been defined. The β form is obtained upon slow crystallization, *i.e.* it is the form obtained under more nearly equilibrium conditions and therefore may be considered to be the more stable one. The α form is obtained under kinetic conditions but converts to β when placed in solution for a long time. Consideration of these facts in the light of the graph sets suggests that the $C_3^3(10)$ chain structure is the thermodynamically more stable of the two arrangements. Under the kinetic conditions of rapid crystallization the closing of a ring arrangement will take place more readily than the attachment of an additional molecule necessary for an infinite chain



Fig. 6. Diagram of the α form of LGA showing the eighth-order network.



Fig. 7. Diagram of the β form of LGA showing the eighth-order network.

arrangement. In energetic terms one would expect the most stable molecular conformation to appear under thermodynamic crystallization conditions. Higher energy conformations would be more likely to be 'trapped' under kinetic crystallization conditions (Bernstein & Schmidt, 1972). Conformational energy calculations on LGA, including non-bonded, electrostatic and torsional energies (Ponnuswamy & Sasisekharan, 1971) indicate that the conformation in the α form is higher in energy by *ca* 0.4 kcal mol⁻¹ (1.7 kJ mol⁻¹). This value is consistent with the crystallization behavior and the molecular conformations found in the two polymorphs.

Another important observation is that the addition of trace amounts of some amino acids or other special compounds causes the formation of the α form. Some very elegant experiments have been carried out over the past few years investigating the role of 'tailor made' impurities on the morphology of a number of materials including amino acids (Addadi et al., 1985). In these studies it was possible to determine, at the molecular level, the influence of the impurities on the rate of growth of specific crystal faces by examining the match between the molecular shape of the impurity and the various crystal faces in question. In the LGA system, a considerable amount of information is lacking but the graph-theory analysis provides some useful sights for some educated speculation.

First, the addition of trace amino-acid impurities inhibits the formation of the β form. In graph-set terms the impurity should inhibit the formation of the infinite $C_3^3(10)$ chain. The α form, in which the molecules can close on themselves should be preferred under these perturbed circumstances of crystallization. When the same crystals are left in solution, again under thermodynamic conditions, the slow dissolution and recrystallization allows for the chain structure to develop at the expense of the ring structure.



Fig. 8. Diagram of the β form of LGA showing two higher-order ring networks.

Concluding remarks

By providing a straightforward method for classifying the hydrogen-bond patterns is crystal structures, graph-set analysis permits a very simple means for comparing the structures of polymorphic systems. Differences in structural patterns may be readily identified. The application of graph-set analysis to the LGA system yielded significant differences between the hydrogen-bonding patterns only when rather high-order networks were considered. It also yielded a structural rationale for the significant molecular conformation differences between the two polymorphs and the relationship of those conformational differences to the crystal packing. The hydrogen-bond patterns, clearly defined and hierarchically ordered with the aid of graph theory, can also aid in the interpretation of the crystallization chemistry, including relative rates of growth and the influence of impurities. We consider this to be a very general approach to studying the packing patterns of hydrogen-bonded crystals, and look forward to its application to many additional systems.

I am most grateful to Professor M. C. Etter of the University of Minnesota for introducing me to graph sets, and for inspiring and encouraging further work in this area. I also wish to thank Mr Michael

Dorfman for very able programming assistance. This project was funded by a grant from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel.

References

- ADDADI, L., BERKOVITCH-YELLIN, Z., WEISSBUCH, L., VAN MIL, J., SHIMON, L., LAHAV, M. & LEISEROWITZ, L. (1985). Angew. Chem. Int. Ed. Engl. 24, 466-485.
- BERNSTEIN, J. (1987). Organic Solid State Chemistry, Studies in Organic Chemistry, Vol. 32, edited by G. DESIRAJU, pp. 471-518. Amsterdam: Elsevier.
- BERNSTEIN, J., ETTER, M. C. & MACDONALD, J. C. (1990), J. Chem. Soc. Perkin Trans. 2, pp. 695-698.
- BERNSTEIN, J. & SCHMIDT, G. M. J. (1972). J. Chem. Soc. Perkin Trans. 2, pp. 951-955.
- ETTER, M. C. (1990). Acc. Chem. Res. 23, 120-126.
- ETTER, M. C., MACDONALD, J. C. & BERNSTEIN, J. (1990). Acta Cryst. B46, 256-262.
- HASEGAWA, M., FUKUDA, N., HIGUCHI, H. & MATSUBARA, I. (1977). Agric. Biol. Chem. 41, 49-56.
- HIRAMATSU, S. (1977a). Nippon Nogei Kagaku Kaishi, 51, 27-37.
- HIRAMATSU, S. (1977b). Nippon Nogei Kagaku Kaishi, 51, 39-46.
- HIRAYAMA, N., SHIRAHATA, K., OHASHI, Y. & SASADA, Y. (1980). Bull. Chem. Soc. Jpn, 53, 30-35.
- HIROKAWA, S. (1955). Acta Cryst. 8, 637 641.
- LEHMANS, M. S., KOETZLE, T. F. & HAMILTON, W. C. (1972). J. Crvst. Mol. Struct. 2, 225-233.
- PONNUSWAMY, P. K. & SASISEKHARAN, V. (1971). Int. J. Protein Res. 3, 9 18.

Acta Cryst. (1991). B47, 1010-1019

Structural Studies on 5-(n-Alkyl)-Substituted Derivatives of the Plant Hormone Indole-3-acetic Acid

BY BISERKA KOJIĆ-PRODIĆ,* BILJANA NIGOVIĆ, SANJA TOMIĆ, NEBOJŠA ILIĆ AND VOLKER MAGNUS Ruder Bosković Institute, POB 1016, 41001 Zagreb, Yugoslavia

ZLATKO GIBA AND RADOMIR KONJEVIĆ

Botanical Institute, Faculty of Natural Sciences and Mathematics and 'Siniša Stanković' Institute for Biological Research, University of Belgrade, 29 novembra 142, 11060 Belgrade, Yugoslavia

AND WILLIAM L. DUAX

Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, USA

(Received 12 March 1991; accepted 18 July 1991)

Abstract

The biological properties of the plant hormone (auxin) indole-3-acetic acid (IAA) and its ringsubstituted derivatives have so far been rationalized

* To whom correspondence should be addressed.

0108-7681/91/061010-10\$03.00

by a number of contradictory hypotheses based on incomplete structural data deduced mainly by inspection of molecular models. To permit a more detailed insight into structure-activity correlations, we here compare the molecular structures of IAA and 5-(nalkyl)indole-3-acetic acids (alkyl = methyl, ethyl,

© 1991 International Union of Crystallography