CROMER, D. T. & MANN, J. B. (1968). *Acta Cryst.* A**24**, 321–324.
FERNANDEZ, J., SOCIAS, C., ARRIANDIAGA, M. A., TELLO, M. J. & LOPEZ ECHARRI, A. (1982). *J. Phys. C*, **15**, 1151–1159.
JOHNSON, C. K. (1965). *ORTEP*. Report ORNL-3794. Oak Ridge National Laboratory, Tennessee.
KIND, R., PLESKO, S., AREND, H., BLINC, R., ZEKS, B., SELIGER, J., LOZAR, B., SLAK, J., LEVSTIK, A., FILIPC, C., ZAGAR, V., LAHAJNAR, G., MILIA, F. & CHAPUIS, G. (1979). *J. Chem. Phys.* **71**(5), 2118–2130.
KITAIGORODSKII, A. I. (1961). *Organic Chemical Crystallography*, p. 177. New York: Consultants Bureau.
SEGERMAN, E. (1965). *Acta Cryst.* **19**, 789–796.
SOCIAS, C. (1980). Doctoral Thesis, Univ. Pais Vasco, Spain.
SOCIAS, C., ARRIANDIAGA, M. A., TELLO, M. J., FERNANDEZ, J. & GILI, P. (1980). *Phys. Status Solidi A*, **57**, 405–410.
STEWART, J. M., KRUGER, G. J., AMMON, H. L., DICKINSON, C. W. & HALL, S. R. (1972). The XRAY system – version of June 1972. Tech. Rep. TR-192. Computer Science Center, Univ. of Maryland, College Park, Maryland.
STEWART, R. F., DAVIDSON, E. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 3175–3187.
ZUÑIGA, F. J. & CHAPUIS, G. (1981). *Cryst. Struct. Commun.* **10**, 533–540.
ZUÑIGA, F. J. & CHAPUIS, G. (1983). To be published.

Acta Cryst. (1983). B**39**, 625–637

[Leu^5]enkephalin: Four Cocrystallizing Conformers with Extended Backbones that Form an Antiparallel β-Sheet

BY ISABELLA L. KARLE AND JEROME KARLE

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375, *USA*

DONALD MASTROPAOLO AND ARTHUR CAMERMAN

Departments of Medicine and Pharmacology, University of Washington, Seattle, Washington 98195, *USA*

AND NORMAN CAMERMAN

Biochemistry Department, University of Toronto, Toronto, Canada M5S 1A8

(*Received* 25 *January* 1983; *accepted* 14 *April* 1983)

Abstract

[Leu^5]enkephalin (Tyr-Gly-Gly-Phe-Leu) grown from *N*,*N*-dimethylformamide(DMFA)/water crystallizes with four quite different conformers side-by-side in the asymmetric unit. The four conformers with extended backbones form an infinite antiparallel β-sheet. β-sheets related by the twofold screw axis are separated by a 12 Å spacing. Side groups protrude above and below the β-sheets and are entirely immersed in a thick layer of solvent occupying the volume between β-sheets. The crystal, stable only in contact with its mother liquor, appears to be a hybrid consisting of rather rigid sheets of peptide molecules separated by spaces filled with mobile solvent, thus having some resemblance to molecules in solution. Many solvent molecules are completely disordered and may be fluid. The space group is $P2_1$ with $a = 18{\cdot}720\,(4)$, $b = 24{\cdot}732\,(6)$, $c = 20{\cdot}311\,(5)$ Å, $\beta = 115{\cdot}86\,(1)°$, $V = 8462$ Å^3. Composition of the asymmetric unit includes four enkephalin molecules, eight water molecules, eight DMFA molecules, plus an unknown number of disordered solvent molecules: $4C_{28}H_{37}N_5O_7.8H_2O.8C_3H_7NO.X$. The R factor from a restrained least-squares refinement (with six-parameter thermal factors) is 11·9% for 8155 data with $|F_o| > 0$. The procedure used for phase determination and structure analysis is described. Parameters for an extensive antiparallel β-sheet are presented.

Introduction

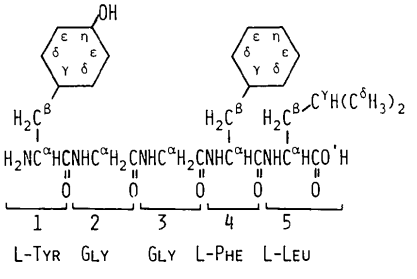
Endogenous enkephalin, a linear pentapeptide functioning as a natural analgesic with opiate-like activity, occurs in the brain as [Leu^5]enkephalin and [Met^5]-enkephalin in varying proportions, depending upon the species (Hughes, Smith, Kosterlitz, Fothergill, Morgan & Morris, 1975). The extreme flexibility of peptides, as compared to the more rigid molecules of morphine and other opiates, necessitates the delineation of probable conformations for the peptide to facilitate structure–activity studies. In the present study of the crystal structure of [Leu^5]enkephalin, the serendipitous occurrence of four molecules of the peptide, each having a different conformation, and a large amount of solvent surrounding the peptide may give good indications for the preferred conformations in solution.

0108-7681/83/050625-13$01.50 © 1983 International Union of Crystallography

The structure contains more than 210 independent C, N and O atoms in space group $P2_1$. It provides an excellent vehicle for the application of formulas and techniques for direct phase determination and demonstrates that the practical limit with respect to size of a structure containing only light atoms has not been reached. This particular structure lies in an area intermediate between small-molecule crystallography and protein analysis. Small structures are normally considered to be those that have about 100 or fewer non-hydrogen atoms in the asymmetric unit. The intermediate range is an area in which few structure analyses have been reported.

In this paper, the procedures for determining the peptide will be described. In addition, the use of restrained least-squares refinement and the location of the solvent molecules will be discussed. The solvent, consisting of H_2O and *N*,*N*-dimethylformamide molecules (DMFA), comprises about 25% of the total content of the cell. Many of the solvent molecules are moderately well-ordered while others, occurring in special regions, are grossly disordered. It is conceivable that some of the disorder may be resolved by further study of the data. The low thermal factors for the atoms in the peptide molecules indicate that the peptide conformations are well established.

The labelling of atoms in an individual enkephalin molecule is as follows:



Experimental procedure

Leucine-enkephalin ($C_{28}H_{37}N_5O_7$) was crystallized from a mixture of DMFA and water maintained at a temperature of 318 K. Upon standing in air the crystals lose solvent of crystallization and their diffraction pattern changes to powder rings. Consequently, for collection of X-ray photographs and diffractometer intensity data, a single crystal was sealed in a capillary tube along with some mother liquor. X-ray data indicated space group $P2_1$ with unit-cell dimensions as in the *Abstract*. The density of the crystals was not measured but calculations assuming a reasonable density for such a molecule suggested that the asymmetric unit consisted of four enkephalin pentapeptide molecules plus 40–50 solvent atoms, a total of over 200 non-hydrogen atoms. X-ray intensities were collected on a Picker four-circle diffractometer with a θ–2θ scan to a resolution of $2\theta = 110°$ with Cu radiation on a crystal measuring approximately 0·6 × 0·2 × 0·1 mm. A total of 10942 independent reflections were measured of which 5965 had intensity greater than $2\sigma_I$. Three standard reflections monitored periodically during data collection indicated no measurable crystal deterioration. Absorption corrections, based on the variation in intensity of a reflection on the φ axis as a function of φ angle and ranging from 1·01 to 1·42, were applied to the data.

Normalized structure factors $|E|$ were obtained from the measured intensities by means of a *K*-curve where the *K*-curve was fit analytically by the exponential function $k\exp(Bs^c)$ (Karle, 1976). For the enkephalin data, $B = 13·6$ and $c = 2·7$. A value of 2·0 for c is equivalent to the Wilson plot.

Peptide structure determination

Phase determination

Since the asymmetric unit contained more than 200 C, O and N atoms with no heavier atom present, solving the crystal structure was not a routine procedure. Attempts utilizing several computerized direct-phasing multiple-solution methods were uniformly unsuccessful. The successful structure determination, although somewhat circuitous, made use of a combination of several approaches.

An immediately obvious departure from a usual distribution of $|E|$ values was the large number of reflections, 83 with unusually high $|E|$ values ranging from 3·0 to 8·3, that indicated a considerable regularity in the structure. In addition to the direct phase determination employing the Symbolic Addition Procedure, a Patterson map using $(|E|^2 - 1)$ values for coefficients was calculated employing a special set of only 55 reflections with high $|E|$ values. The particular reflections chosen were those that occurred in sets with varying k indices such as $\overline{17},k,8$ $(0 \le k \le 7)$ and $\overline{16},k,4$ $(0 \le k \le 6)$, all with $|E| > 2·4$, in order to determine the implications of these sets concerning the structure. The peaks in this vector map, Fig. 1, were consistent with a β-sheet structure. When all the observed data were included in the $(|E|^2 - 1)$ map, the vectors defining the β-sheet were sufficiently masked by vectors from other parts of the structure to make the analysis of the implications of the map much more uncertain. It is possible that in other structures such an abbreviated $(|E|^2 - 1)$ map derived from only special sets of reflections may give indications of characteristic features of the structure.

Direct phase determination by the Symbolic Addition Procedure (Karle & Karle, 1966, 1968) proceeded in

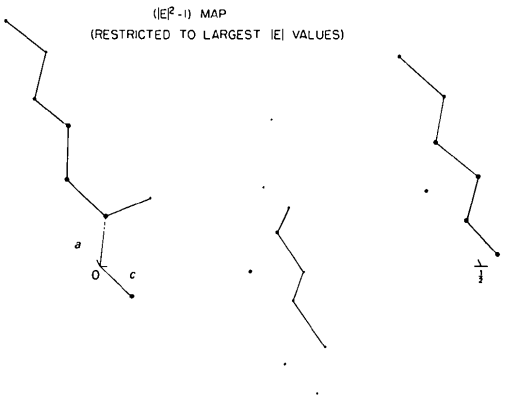


Fig. 1. The highest peaks in a Patterson function computed with $(|E_h^2| - 1)$ values for coefficients. A special set of only 55 reflections were used, each with $|E| > 2{\cdot}4$. The vectors indicated by the abbreviated $(|E^2| - 1)$ map are consistent with a β-sheet structure.

the usual fashion for space group $P2_1$.* An origin was selected by assigning a phase value of zero to reflections $\bar{1}0,0,19$, $\bar{1}7,0,12$ and 014. The phases of reflections 020 and 060 were readily determined to be π in the initial stages. Four other symbolic assignments, a, b, c, and d, were required for the phase determination to proceed. They were assigned to $\bar{1}6,0,4$, $7,1,13$, 170 and $\bar{3}35$, respectively. All assignments were made to reflections with $|E| > 4{\cdot}0$. From the application of the Σ_2 relationship and examination of the values from multiple phase indications, it was not possible to make an unambiguous evaluation of any of the symbols. At this point, the Σ_3 relationship for non-centrosymmetric crystals† (Karle, 1969), proved to be useful in evaluating symbols a, b and c as π, π and 0, while symbol d remained indeterminate. Accordingly, since up to this point all phases would have values of 0 or π, as in a centrosymmetric crystal, symbol d was chosen to be $+\pi/2$, a choice that effectively selected the enantiomorph. In the resulting E map the strongest peaks represented segments of four parallel chains. The atoms, although near $y \sim \frac{1}{4}$, were not coplanar but showed the characteristic pleats of a pleated sheet. With the aid of the indications from the restricted $(|E|^2 - 1)$ map, forty peaks in the E map, calculated with ~1800 reflections, were selected consistent with an antiparallel β-sheet. Phases for reflections with $|E| > 1{\cdot}5$ calculated from this partial structure, and accepted if $|F_{calc}| > 0{\cdot}18|F_{obs}|$, were used for phase extension by the tangent formula for reflections with $|E| > 1{\cdot}2$. The phase values were *not* recycled in the attempt to refine their values, since in space group $P2_1$, particularly if a heavier atom is present, or, as in this example, where many atoms lie near a plane parallel to (010), recycling has the tendency to shift the phases toward centrosymmetric values (Karle, Gilardi, Fratini & Karle, 1969; Karle, 1974). The 84 atoms in the backbones of the four conformers were found in two stages of partial-structure development by use of the tangent formula and calculation of E maps.

The satisfactory appearance of the β-sheet, the many hydrogen bonds that were indicated and the relatively low R factor of 46% for the initial stages of phase extension by the tangent formula suggested that the structure could be correct up to this point. The location of the 76 atoms in the side chains was very difficult, however, since the partial-structure technique did not give any further information. From the known sequence of the [Leu^5]enkephalin, it was possible to place the twelve C^β atoms fairly accurately. From difference maps and models, the positions of most of the atoms in the side chains were eked out. Cycles of restrained least-squares refinement (Konnert, 1976; Konnert & Hendrickson, 1980) were interspersed with difference maps that showed possible water molecules separated by distances appropriate to hydrogen-bond separations. The R factor, however, could not be reduced below 29% for the 190 to 210 atoms that were included in the refinement process and this high value, in fact, indicated that the structure was incorrectly placed with respect to a proper origin.

At this point it was decided to relax the space group to $P1$ (Karle & Karle, 1971). The data base was doubled with $|E_{hkl}| = |E_{\bar{h}k\bar{l}}|$ and $|E_{\bar{h}kl}| = |E_{h\bar{k}l}|$. About 150 atoms comprising the backbones and most of the side chains in the apparently misplaced structure determined above were used as the partial structure for computation of the initial phases in $P1$. Refinement and extensions of phases with the tangent formula and calculation of an E map based on the derived phases, with $|E|_{min} = 1{\cdot}2$, yielded peaks that reproduced the original backbones but not the side chains. Thirty peaks in the other half of the cell, *i.e.* near $y = \frac{3}{4}$, were assigned to the backbones of additional enkephalin molecules potentially related to the original molecules by a symmetry operation. With these additional atoms, the twofold screw operation, for the original space group $P2_1$, could be satisfied either by the original position of the origin, or by a new position shifted by $x - 0{\cdot}068$ and $z - 0{\cdot}135$, or by another new position shifted by $x - 0{\cdot}096$ and $z - 0{\cdot}257$, or even largely satisfied by origins in several other locations. Another round of the partial-structure development in $P1$, with the addition of the 30 new peaks, yielded another 30 or so new atomic positions that could be more definitely related to the original backbone atoms by a twofold screw placed at $x - 0{\cdot}096$ and $z - 0{\cdot}257$ with respect to the original location.

* The Symbolic Addition Procedure is programmed for interactive use on a computer (program *SAPPI*; R. Gilardi, Naval Research Laboratory, Washington, DC).

† See Appendix.

The next step involved the return to space group $P2_1$ with a shift to the new origin and a partial structure development based initially only on the backbone atoms. Since some of the side-chain atoms in the original structure may have been in error, they were omitted at this stage so as not to prejudice the calculation. This time, with the correct placement of the atoms for the backbone with respect to an appropriate origin, use of the tangent formula only for phase extension and not for phase refinement produced atomic positions for the side chains quite rapidly. After three cycles of partial structure development, positions were obtained for 156 of the 160 atoms in the four peptide molecules and seven atoms in solvent molecules.

The initial incorrect placement of the enkephalin chains with respect to an origin may have been caused by a subset of phases derived from reflection $\overline{16},3,4$ ($|E| = 3{\cdot}4$). The phase of $\overline{16},3,4$ was accepted from two strong Σ_2 triplets, each of which indicated an erroneous phase displaced by nearly π from the correct value. Under such circumstances, when a large subset of phases is inconsistent with the remainder of the phases, there is no combination of values for the symbolic phases that will produce a correct E map. The best map may be displaced from a correct origin.

Refinement

Several cycles of restrained least-squares refinement were computed with the newest version of the program *RESLSQ* (Flippen-Anderson, Gilardi & Konnert, 1983) that included a six-parameter thermal refinement. Atomic scattering factors were taken from *International Tables for X-ray Crystallography* (1974). Cycles of refinement were interspersed with difference

Table 1. *Fractional coordinates and B_{eq} values for the four enkephalin conformers*

The e.s.d.'s for the backbone atoms are near 0·0009, 0·0010 and 0·0009 for x, y and z, respectively, while the e.s.d.'s for the atoms in side chains increase up to 0·0015, 0·0015 and 0·0015 for x, y and z, respectively.

$$B_{eq} = \tfrac{4}{3}\sum_i \sum_j \beta_{ij}\mathbf{a}_i.\mathbf{a}_j.$$

Molecule *A*	x	y	z	B_{eq} (Å^2)	Molecule *B*	x	y	z	B_{eq} (Å^2)
N_1	0·4583	0·2735	0·0969	4·7	N_1	−0·2865	0·3155	0·3118	5·4
C_1^α	0·3873	0·3015	0·0997	4·6	C_1^α	−0·2388	0·3407	0·2767	4·8
C_1'	0·3170	0·2651	0·0702	4·4	C_1'	−0·1504	0·3372	0·3299	5·7
O_1	0·2978	0·2428	0·0115	5·5	O_1	−0·1292	0·3316	0·3981	8·0
N_2	0·2743	0·2619	0·1078	4·6	N_2	−0·1006	0·3374	0·3002	6·8
C_2^α	0·1990	0·2331	0·0807	5·4	C_2^α	−0·0133	0·3371	0·3476	9·4
C_2'	0·1474	0·2650	0·1055	7·7	C_2'	0·0253	0·3097	0·3070	6·3
O_2	0·1639	0·2688	0·1696	12·0	O_2	−0·0036	0·3096	0·2415	9·5
N_3	0·0821	0·2900	0·0565	5·4	N_3	0·0960	0·2878	0·3499	6·4
C_3^α	0·0303	0·3237	0·0762	5·4	C_3^α	0·1379	0·2599	0·3142	9·1
C_3'	−0·0534	0·3022	0·0413	4·7	C_3'	0·2261	0·2593	0·3603	5·0
O_3	−0·0868	0·2891	−0·0229	7·3	O_3	0·2576	0·2568	0·4270	5·7
N_4	−0·0870	0·2979	0·0860	5·0	N_4	0·2656	0·2573	0·3178	4·3
C_4^α	−0·1682	0·2772	0·0646	5·5	C_4^α	0·3519	0·2476	0·3461	3·9
C_4'	−0·2168	0·3196	0·0797	5·2	C_4'	0·3874	0·2908	0·3185	4·0
O_4	−0·1902	0·3464	0·1374	5·8	O_4	0·3563	0·3043	0·2534	5·8
N_5	−0·2895	0·3238	0·0285	5·3	N_5	0·4505	0·3158	0·3669	4·3
C_5^α	−0·3500	0·3553	0·0399	4·7	C_5^α	0·4940	0·3550	0·3452	5·1
C_5'	−0·3804	0·3207	0·0849	5·1	C_5'	0·5391	0·3254	0·3099	5·1
O_5	−0·3612	0·2729	0·0975	6·8	O_5	0·5520	0·2769	0·3173	6·5
O_5'	−0·4295	0·3445	0·1028	6·0	O_5'	0·5672	0·3548	0·2762	6·7
C_1^β	0·3697	0·3520	0·0505	6·9	C_1^β	−0·2596	0·3996	0·2564	5·5
C_1^γ	0·2981	0·3848	0·0428	6·9	C_1^γ	−0·2369	0·4400	0·3162	5·8
C_1^δ	0·2237	0·3696	−0·0111	8·4	C_1^δ	−0·2826	0·4511	0·3503	6·6
C_1^ε	0·1582	0·3998	−0·0176	8·7	C_1^ε	−0·2635	0·4927	0·4023	9·8
C_1^ζ	0·1667	0·4422	0·0274	12·7	C_1^ζ	−0·1963	0·5220	0·4229	10·2
C_{11}^η	0·2404	0·4551	0·0810	12·4	C_{11}^η	−0·1472	0·5094	0·3903	11·4
C_{11}^δ	0·3057	0·4274	0·0881	12·0	C_{11}^δ	−0·1681	0·4710	0·3346	10·1
OTyr	0·0986	0·4709	0·0169	17·5	OTyr	−0·1735	0·5621	0·4739	14·9
C_4^β	−0·1643	0·2288	0·1129	7·3	C_4^β	0·3652	0·1940	0·3177	5·0
C_4^γ	−0·1183	0·1827	0·1030	8·8	C_4^γ	0·3238	0·1468	0·3356	6·2
C_4^δ	−0·1571	0·1406	0·0547	11·5	C_4^δ	0·3634	0·1194	0·4018	7·6
C_4^ε	−0·1185	0·0940	0·0429	13·0	C_4^ε	0·3244	0·0796	0·4244	10·3
C_4^ζ	−0·0396	0·0919	0·0832	13·0	C_4^ζ	0·2464	0·0678	0·3764	12·6
C_{41}^η	0·0005	0·1343	0·1327	15·6	C_{41}^η	0·2087	0·0919	0·3095	13·1
C_{41}^δ	−0·0380	0·1787	0·1425	13·3	C_{41}^δ	0·2443	0·1328	0·2893	9·3
C_5^β	−0·4189	0·3677	−0·0352	6·0	C_5^β	0·5512	0·3859	0·4123	6·1
C_5^γ	−0·3951	0·4039	−0·0834	7·4	C_5^γ	0·5162	0·4183	0·4520	9·2
C_5^δ	−0·4647	0·4090	−0·1616	10·2	C_5^δ	0·5827	0·4343	0·5287	16·6
C_{51}^ε	−0·3728	0·4636	−0·0504	11·9	C_{51}^ε	0·4727	0·4640	0·4120	15·3

Table 1 (*cont.*)

Molecule *C*	*x*	*y*	*z*	B_{eq} (Å²)	Molecule *D*	*x*	*y*	*z*	B_{eq} (Å²)
N_1	0·4764	0·2734	0·6469	4·6	N_1	−0·3012	0·2952	0·7717	5·3
C_1^{α}	0·4002	0·2481	0·5896	4·3	C_1^{α}	−0·2246	0·2669	0·8094	4·8
C_1'	0·3436	0·2418	0·6236	4·8	C_1'	−0·1668	0·2888	0·7820	4·3
O_1	0·3689	0·2382	0·6901	5·6	O_1	−0·1910	0·2973	0·7166	6·9
N_2	0·2686	0·2405	0·5778	5·1	N_2	−0·0926	0·2985	0·8303	5·3
C_2^{α}	0·2059	0·2319	0·6013	6·5	C_2^{α}	−0·0318	0·3138	0·8079	5·5
C_2'	0·1357	0·2662	0·5585	5·7	C_2'	0·0397	0·2778	0·8487	4·8
O_2	0·1207	0·2812	0·4972	8·5	O_2	0·0631	0·2673	0·9135	7·1
N_3	0·0878	0·2765	0·5900	6·6	N_3	0·0793	0·2610	0·8108	4·4
C_3^{α}	0·0169	0·3074	0·5521	6·8	C_3^{α}	0·1493	0·2277	0·8437	5·2
C_3'	−0·0356	0·3108	0·5909	5·1	C_3'	0·2057	0·2447	0·8123	4·4
O_3	−0·0115	0·2959	0·6544	6·9	O_3	0·1809	0·2448	0·7445	5·6
N_4	−0·1058	0·3315	0·5514	4·9	N_4	0·2811	0·2544	0·8568	4·5
C_4^{α}	−0·1676	0·3376	0·5786	4·6	C_4^{α}	0·3394	0·2694	0·8308	4·6
C_4'	−0·2392	0·3045	0·5296	5·4	C_4'	0·4184	0·2462	0·8833	4·5
O_4	−0·2598	0·2997	0·4627	7·3	O_4	0·4377	0·2438	0·9497	5·9
N_5	−0·2812	0·2840	0·5604	4·4	N_5	0·4678	0·2351	0·8540	4·7
C_5^{α}	−0·3525	0·2520	0·5228	4·8	C_5^{α}	0·5505	0·2198	0·8971	4·7
C_5'	−0·4124	0·2680	0·5527	5·2	C_5'	0·6019	0·2476	0·8624	5·0
O_5	−0·3877	0·2706	0·6217	6·3	O_5	0·5688	0·2562	0·7944	5·6
O_5'	−0·4824	0·2741	0·5092	6·0	O_5'	0·6725	0·2587	0·9044	6·4
C_1^{β}	0·4186	0·1921	0·5683	5·1	C_1^{β}	−0·2350	0·2058	0·7945	7·7
C_1^{γ}	0·4607	0·1513	0·6246	5·2	C_1^{γ}	−0·1542	0·1759	0·8203	10·1
C_1^{δ}	0·4188	0·1099	0·6376	6·4	C_1^{δ}	−0·1283	0·1607	0·7670	11·9
C_1^{ε}	0·4574	0·0687	0·6869	7·2	C_1^{ε}	−0·0552	0·1351	0·7924	13·3
C_1^{η}	0·5403	0·0668	0·7216	6·9	C_1^{η}	−0·0055	0·1287	0·8637	13·5
C_{11}'	0·5815	0·1083	0·7093	6·8	C_{11}'	−0·0271	0·1463	0·9171	13·3
C_{11}^{ζ}	0·5416	0·1484	0·6620	5·7	C_{11}^{ζ}	−0·1029	0·1691	0·8981	12·9
OTyr	0·5787	0·0246	0·7689	8·7	OTyr	0·0689	0·1044	0·8922	17·5
C_4^{β}	−0·1912	0·3961	0·5756	5·8	C_4^{β}	0·3487	0·3304	0·8311	6·7
C_4^{γ}	−0·1307	0·4309	0·6294	7·5	C_4^{γ}	0·2748	0·3580	0·7743	7·2
C_4^{δ}	−0·1205	0·4338	0·7021	9·3	C_4^{δ}	0·2656	0·3681	0·7041	9·8
C_4'	−0·0594	0·4666	0·7530	13·2	C_4'	0·1982	0·3917	0·6503	11·7
C_4^{η}	−0·0104	0·4992	0·7320	12·1	C_4^{η}	0·1411	0·4095	0·6745	13·0
C_{41}'	−0·0249	0·4958	0·6581	11·9	C_{41}'	0·1489	0·4017	0·7407	11·2
C_{41}^{ζ}	−0·0828	0·4614	0·6066	9·8	C_{41}^{ζ}	0·2173	0·3765	0·7937	9·6
C_5^{β}	−0·3307	0·1922	0·5393	6·7	C_5^{β}	0·5644	0·1591	0·8958	6·9
C_5'	−0·3954	0·1509	0·4956	8·5	C_5'	0·5443	0·1252	0·9453	10·8
C_5^{δ}	−0·3695	0·0939	0·5311	12·4	C_5^{δ}	0·5619	0·0634	0·9330	17·5
C_{51}^{ζ}	−0·4123	0·1508	0·4143	11·2	C_{51}^{ζ}	0·5983	0·1410	1·0283	12·3

maps to locate additional atoms. By these means, the *R* factor based on 202 atoms was reduced to 14·4% for 8184 data [$|F_o| > 0$ and $(\sin\theta)/\lambda \leq 0{\cdot}5$ Å^{-1} (d = 1·0 Å)]. At this stage, the difference map contained a large number of peaks consistent with H-atom positions, particularly on the backbones, although the nature of the solvent structure was only partially resolved. Accordingly, idealized positions for 108 H atoms that could be placed unambiguously on the backbones and side chains were added to the refinement with restraints placed on the C–H distances.

A comparison of the observed and calculated structure factors showed that all the large differences occurred for reflections with small indices. Most of the reflections with the large differences, called 'outliers', could be correlated with the region of the cell containing the unresolved solvent molecules. In the case of protein refinement, the very inner data that are most sensitive to unresolved solvent are usually omitted from the refinement (Sielecki, Hendrickson, Broughton, Delbaere, Brayer & James, 1979). In the *RESLSQ* program the outliers (differences of 20σ) are automatically excluded from the refinement. However, it is important to include them in the calculation of a difference map when looking for solvent molecules. Furthermore, it is advantageous to exclude high-angle data from the difference map. The solvent molecules being sought have very large *B* values associated with them and do not contribute to the structure factors with high θ values. In addition, at high θ values, differences between $|F_o|$ and $|F_c|$ are relatively large and add noise that masks the positions of the solvent atoms. In the present structure analysis, the difference maps were calculated with 3800 reflections with $\theta_{max} = 35°$ (d = 1·35 Å).

In the refinement, restraints were imposed only on bonded distances and next-bonded distances, corresponding to restraints on bond angles; and, in addition, the eight phenyl rings were restrained to be relatively planar. No restraints were imposed upon hydrogen-bond lengths. Restraints were relaxed gradually as more solvent atoms were added and as the *R* factor decreased. At the present stage of refinement with *RESLSQ*, $R = 11{\cdot}9\%$ for 8155 data with $|F_o| > 0$ and weighted $R = 7{\cdot}9\%$ for 9081 data. [The data used for refinement were arbitrarily cut off at $(\sin\theta)/\lambda =$

0·5 Å^{-1}, *i.e.* $d = 1·0$ Å.] There are 34 outliers with low indices not included in the R factors. Included in the structure factor calculations are eight O atoms from water molecules, eight DMFA molecules, and six other atoms that could be either water O atoms or fragments of DMFA, and 108 H atoms. There are still some unfilled voids in the cell and there is a large region with disordered solvent that has not been resolved.

During the course of refinement parallel attempts to improve atom positional and thermal parameters were made using conventional full-matrix and block diagonal least-squares procedures. It is of interest to note that neither of these methods was as successful as the *RESLSQ* technique in reducing the discrepancy index. In addition, the unrestrained least-squares cycles produced large fluctuations in thermal parameters and ended in generally higher B values than was the case for the restrained program.

Coordinates from the *RESLSQ* refinement for the four enkephalin molecules are listed in Table 1* while the more approximate coordinates for the water oxygens and eight DMFA molecules are listed in Table 2. Bond lengths for the four enkephalin molecules are shown in Table 3. The values are entirely consistent with those determined for other peptide molecules.

Table 2. *Fractional coordinates and B_{eq} for the solvent molecules*

Oxygens in water molecules are represented by *W*. Coordinates for the DMFA molecules are prefixed by a number distinguishing the eight molecules. The symbol *X* denotes unknown atoms in the solvent region. E.s.d.'s for *W*(1) to *W*(4) are of the order of 0·0012, 0·0011 and 0·0011, for *x*, *y* and *z*, respectively. For atoms with B_{eq} near 20 Å^2, the e.s.d.'s are near 0·0023, 0·0017 and 0·0021, and they increase with increasing values of B_{eq}.

	x	*y*	*z*	B_{eq} (Å^2)
W(1)	−0·3185	0·2163	0·2223	12·8
W(2)	0·5142	0·2023	0·2150	9·6
W(3)	−0·4633	0·4463	0·0889	11·3
W(4)	−0·4885	0·4525	0·2098	8·2
W(5)	0·0497	0·1252	0·5071	30·2
W(6)	−0·0239	0·4740	0·1795	80·0
W(7)	0·1281	0·0804	0·0805	84·6
W(8)	0·7233	0·2041	0·3684	32·7
1O	0·4122	0·4670	0·9381	30·3
1C(1)	0·3436	0·4771	0·9353	26·5
1N	0·3055	0·5059	0·8682	21·5
1C(2)	0·2262	0·5213	0·8585	26·2
1C(3)	0·3305	0·5177	0·8129	25·7
2O	0·9750	0·4373	0·9338	32·7
2C(1)	0·9032	0·4280	0·9338	29·1
2N	0·8810	0·4534	0·9850	23·9
2C(2)	0·8133	0·4283	0·9755	31·0
2C(3)	0·9555	0·4774	1·0413	30·1
3O	1·0089	0·1515	0·6324	31·6
3C(1)	0·9313	0·1638	0·5943	47·9
3N	0·8840	0·1177	0·5601	56·4
3C(2)	0·8019	0·0961	0·5059	65·6
3C(3)	0·9452	0·0751	0·6034	58·9
4O	1·0025	0·1556	0·3800	35·6
4C(1)	0·9358	0·1661	0·3232	39·2
4N	0·8742	0·1266	0·3234	49·0
4C(2)	0·8487	0·0983	0·3818	53·4
4C(3)	0·7894	0·1374	0·2562	49·7
5O	0·2105	0·1018	0·0198	33·1
5C(1)	0·2763	0·1069	0·0147	25·5
5N	0·3447	0·1133	0·0828	17·8
5C(2)	0·4000	0·1141	0·0456	20·6
5C(3)	0·3577	0·1142	0·1560	19·9
6O	0·0997	0·4389	0·2995	53·8
6C(1)	0·1264	0·4263	0·2499	51·4
6N	0·1985	0·4513	0·2585	39·2
6C(2)	0·2399	0·4047	0·2446	36·2
6C(3)	0·2175	0·5077	0·2504	36·0
7O	0·1730	0·0380	0·8342	78·1
7C(1)	0·1923	0·0942	0·7891	26·0
7N	0·2829	0·0779	0·7953	12·9
7C(2)	0·2948	0·1159	0·7551	21·6
7C(3)	0·3266	0·0294	0·8418	20·8
8O	−0·2685	0·4034	0·7981	17·2
8C(1)	−0·3331	0·4224	0·7452	23·3
8N	−0·3403	0·4817	0·7392	26·7
8C(2)	−0·4156	0·5150	0·6955	23·3
8C(3)	−0·2554	0·5017	0·7880	35·5
X(1)	0·4117	0·3784	0·6251	16·9
X(2)	0·4866	0·4734	0·6728	40·1
X(3)	0·3807	0·4887	0·6473	56·4
X(4)	0·2802	0·3996	0·4949	50·2
X(5)	0·4636	0·3924	0·6546	47·2
X(6)	0·9137	0·4392	0·1873	30·4

Description of the structure

The cell

[Leu^5]enkephalin crystallizes from a mixture of water and DMFA in a monoclinic cell, space group $P2_1$, with four distinctly different conformers per asymmetric unit and a quantity of solvent comprising 25–30% of the contents of the cell. There is no subcell or pseudo-halving of the cell axes as occurred in the $C2$ modification of [Leu^5]enkephalin (Blundell *et al.*, 1979). The volume per asymmetric unit in the $P2_1$ crystal is 4231·0 Å^3 as compared to 3375·3 Å^3 in the $C2$ crystal. Furthermore, the conformations of the peptide molecules in the $P2_1$ cell with extended backbones are entirely different from the folded conformation reported for the $C2$ cell (Smith & Griffin, 1978; Blundell *et al.*, 1979).

In the present structure, the four conformers lie antiparallel to each other as shown in the stereodiagram in Fig. 2 and schematically in Fig. 3. Not only are the backbones somewhat staggered with respect to each other, but molecules *C* and *D* are rotated by about 180° through the long axes of the molecules as compared to molecules *A* and *B*, respectively. Through extensive lateral and head-to-tail hydrogen bonding, utilizing all the NH and CO moieties, an infinite hydrogen-bonded sheet is formed. It is a somewhat more complex β-sheet than the classical antiparallel pleated sheet exhibited by L-Ala-L-Ala-L-Ala (Fawcett, Camerman & Camerman, 1975). The backbones (but not the side chains) of conformers *A* and *B* are approximately related by a pseudo twofold rotation

* Lists of observed and calculated structure factors and anisotropic thermal parameters for the four enkephalin molecules have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 38501 (44 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 3. *Bond lengths* (Å) *in four enkephalin molecules*

The e.s.d.'s range up to 0·03 Å for bonds in the backbone. Restraints in the refinement were relaxed in the later cycles and the low B values for the backbone atoms indicate that they are well placed in the cell. The scatter from the averages of 20 similar bonds, shown below, is near 0·014 Å. A comparison can be made with the structure analysis of [Phe^{4}Val^6]antamanide (Karle, 1977), also containing a large void with unresolved solvent, with $R = 12·1\%$ for anisotropic full-matrix refinement, in which the e.s.d.'s are 0·017 Å for backbone bonds.

i	N_i–C_i^α	C_i^α–C_i'	C_i'–O_i	C_i'–N_{i+1}	C_i^α–C_i^β	C_i^β–C_i^γ	C_i^γ–C_i^δ	C_i^δ–C_i^ε	C_i^ε–C_i^ζ	C_i^ζ–OTyr
Molecule *A*										
1	1·522*	1·486	1·218	1·328	1·544	1·514	1·364 1·397	1·392 1·353	1·353 1·371	1·393
2	1·455	1·495	1·203	1·345						
3	1·461	1·506	1·218	1·315						
4	1·478	1·504	1·247	1·307	1·530	1·495	1·364 1·397	1·436 1·377	1·342 1·418	
5	1·475	1·531	1·228	1·272†	1·540	1·528	1·560 1·601			
Molecule *B*										
1	1·500*	1·534	1·270	1·313	1·517	1·485	1·343 1·404	1·406 1·397	1·349 1·382	1·361
2	1·493	1·476	1·198	1·340						
3	1·454	1·502	1·221	1·358						
4	1·480	1·491	1·236	1·315	1·510	1·529	1·396 1·413	1·415 1·370	1·387 1·364	
5	1·454	1·513	1·220	1·260†	1·521	1·478	1·423 1·561			
Molecule *C*										
1	1·527*	1·504	1·225	1·302	1·534	1·469	1·369 1·385	1·390 1·356	1·397 1·370	1·388
2	1·462	1·484	1·209	1·333						
3	1·432	1·506	1·223	1·309						
4	1·492	1·514	1·247	1·301	1·508	1·461	1·406 1·396	1·416 1·415	1·420 1·408	
5	1·451	1·544	1·270	1·228†	1·532	1·535	1·561 1·539			
Molecule *D*										
1	1·474*	1·516	1·221	1·327	1·536	1·555	1·416 1·457	1·385 1·415	1·346 1·384	1·389
2	1·447	1·519	1·220	1·346						
3	1·442	1·512	1·247	1·323						
4	1·455	1·509	1·237	1·329	1·517	1·523	1·383 1·375	1·387 1·409	1·428 1·303	
5	1·458	1·577	1·260	1·252†	1·526	1·478	1·607 1·591			
Average										
	1·462 ±0·014 *1·505 ±0·019	1·511 ±0·014	1·231 ±0·017	1·324 ±0·014 †1·253 ±0·013	1·526 ±0·010					

* Terminal NH_2–C^α.
† Terminal C_5'–O_5' instead of C_5'–N_{i+1}.

axis parallel to **b**. The ends of these two molecules are bridged by lateral hydrogen bonds to four interjected water molecules, $W(1)$–$W(4)$. The backbones of conformers *C* and *D* are also approximately related by a different pseudo twofold rotation axis. However, there are no pseudo symmetry elements between backbones of molecules *B* and *C* or *D* and *A*.

Molecules *A*–*D* are repeated in the **c** direction by translation and continue the β-sheet by lateral hydrogen bonding where the average N···O length for the 20 lateral NH···O hydrogen bonds is 2·900 Å. In the **a** direction, molecules related by translation are joined by head-to-tail NH···O hydrogen bonds. There are two different modes in the head-to-tail contacts. The vicinity of the amine and carboxyl termini of molecules *C* and *D* is completely devoid of water; while near the termini of molecules *A* and *B*, there are four water molecules that are intimately involved in hydrogen bonding with the NH_3 and COO groups of these molecules. Water $W(1)$ bridges AO_5 and BN_1 with hydrogen bonds, while $W(2)$ bridges AN_1 and BO_5 and may form a third hydrogen bond with $W(1)$. Atoms AO_5' and BO_5' are bridged by two intervening water molecules, *i.e.* $AO_5'\cdots W(3)\cdots W(4)\cdots BO_5'$. The hydrogen-bond lengths are listed in Table 4.

Arnott, Dover & Elliott (1967) estimated the values $\varphi = -139°$ and $\psi = +135°$ for the torsional angles about N_i–C_i^α and C_i^α–C_i', respectively, for the regular antiparallel pleated sheet occurring in β-poly-L-alanine. In the [Leu^5]enkephalin structure, most values for the torsional angles (Table 5) cluster near $\varphi = -130°$ and near $\psi = +140°$, in good agreement with Arnott's values. Different values occur for the two Gly residues in conformers *B* and *C* where $\varphi \sim +155°$ and $\psi \sim$

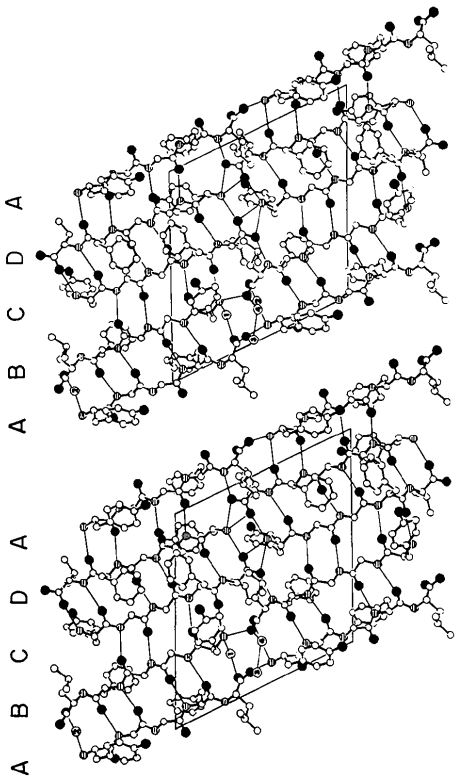


Fig. 2. Four independent molecules of enkephalin, labelled *A*, *B*, *C* and *D*, each with a different conformation, form an infinite antiparallel β-sheet in the crystal. The layer of hydrogen-bonded molecules near $y \sim \frac{1}{4}$ is shown in the stereodiagram. The axial directions are $a \nwarrow$ and $c \rightarrow$. Four water molecules are labelled 1–4, O atoms are indicated by filled circles and N atoms by hatched circles. See Fig. 3 for the schematic representation. There may be an additional hydrogen bond between $W(1)\cdots W(2)$.

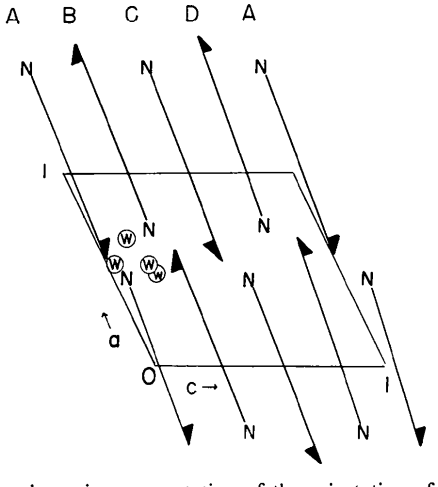


Fig. 3. A schematic representation of the orientation of the four independent enkephalin conformers *A*–*D* that form the antiparallel β-sheet in Fig. 2. The semi-arrowheads represent the carboxyl termini. The backbone of molecule *C* is flipped over as compared to *A* and similarly the backbone of molecule *D* is flipped over as compared to *B*. Four water molecules are represented by *W*.

Table 4. *Hydrogen bonds in the β-sheet formed by enkephalin molecules A, B, C and D*

The e.s.d.'s for hydrogen bonds range from 0·02 to 0·03 Å.

A···*B*		*B*···*C*		*C*···*D*	
$N_2\cdots O_4$	2·871	$N_1\cdots O_4$	2·908	$N_1\cdots O_5$	2·762
$O_1\cdots N_4$	2·788	$O_1\cdots N_4$	2·949	$O_1\cdots N_5$	3·023
$N_4\cdots O_2$	2·863	$N_3\cdots O_2$	2·827	$N_3\cdots O_3$	2·954
$O_4\cdots N_2$	2·993	$O_3\cdots N_2$	3·007	$O_3\cdots N_3$	3·001
		$*N_5\cdots O_5'$	2·798	$N_5\cdots O_1$	2·890
				$O_5\cdots N_1$	2·824
D···*A* (1 + *z*)		*A*···*W*···*B*		*A*···*W*···*W*···*B*	
$O_4\cdots N_1$	2·936	$AN_1\cdots W(2)$	2·785	$AO_5'\cdots W(3)$	2·582
$N_4\cdots O_1$	3·031	$AO_5\cdots W(1)$	2·694	$BO_5'\cdots W(4)$	2·741
$O_2\cdots N_3$	2·825	$BN_1\cdots W(1)$	2·954	$W(3)\cdots W(4)$	2·699
$N_2\cdots O_1$	2·945	$BO_5\cdots W(2)$	2·635		
$*O_5'\cdots N_5$	2·808	$W(1)\cdots W(2)$	3·090		
Head-to-tail					
$AN_1\cdots AO_5'$	2·700	$CN_1\cdots CO_5$	2·807	$DN_1\cdots DO_5$	2·837
$BN_1\cdots BO_5'$	2·691	$CN_1\cdots CO_5'$	3·209	$DN_1\cdots DO_5'$	3·078

* Between molecules where one is shifted by 1 + *x*.

Table 5. *Torsional angles* (°)

E.s.d.'s are near 1·8–2·5°. See note in Table 3.

	Residue	Molecules *A*	*B*	*C*	*D*
Backbone					
ψ_1	Tyr	135	154	155	137
ω_1		172	177	177	173
φ_2		−144	151	141	−131
ψ_2	Gly	114	−155	−157	142
ω_2		−177	180	−178	179
φ_3		−122	154	174	−144
ψ_3	Gly	132	−151	−170	131
ω_3		−179	−170	179	178
φ_4		−122	−128	−119	−147
ψ_4	Phe	139	130	149	152
ω_4		168	174	179	171
φ_5		−79	−72	−141	−141
$*\psi_5'$	Leu	176	167	137	151
Side chains					
χ_{11}	Tyr	177	70	53	169
χ_{12}		−86 / 93	−86 / 99	101 / −85	−102 / 71
χ_{41}	Phe	−63	−55	−71	−68
χ_{42}		−95 / 87	−88 / 87	−82 / 98	−91 / 93
χ_{51}	Leu	−64	−62	−171	−80
χ_{52}		173 / −67	165 / −69	−168 / 68	179 / −63

* Angle ψ_5' represents the terminal angle of the backbone $N_5C_5^{\alpha}C_5'O_5'$ at the carboxyl end of the molecule.

−160°. These values fall in the D region of the φ,ψ map but they can also be considered to be close to the upper left-hand corner of the L region. In that respect the two clusters are not very far from each other.

The infinite pleated sheet near $y \sim \frac{1}{4}$ is repeated near $y \sim \frac{3}{4}$ by the twofold screw operation in space group $P2_1$. Side chains of enkephalin molecules project above and below the sheets into the space between the sheets.

Since the separation between two sheets is ~12 Å there cannot be any direct contact between side groups from adjacent sheets in the form of hydrogen bonds or van der Waals attractions. The large volume between the molecular layers is filled by a layer of solvent molecules, that is, water and DMFA. Figs. 4 and 5 show a cross-section of the layered structure in two slices, one with z between ~0 and ~$\frac{1}{4}$ and the other with z between ~$\frac{1}{4}$ and ~$\frac{1}{2}$, respectively. The two figures together comprise an asymmetric unit.

The thermal parameters B_{eq} for the relatively rigid backbone atoms in the enkephalin molecules have values between 4 and 7 Å^2 while the ends of the side chains have values of about 13 and up to 17 Å^2 for some of the tyrosyl hydroxyl groups, Table 1. The solvent molecules, $W(1)$ to $W(4)$, have B_{eq} values of 8 to 13 Å^2. The remainder of the solvent molecules have B_{eq} values that generally exceed 20 Å^2 and many that exceed 35 Å^2, values that indicate very considerable positional disorder.

The integrity of the unit cell is maintained under tenuous conditions. The only firm link between the widely separated β-sheets formed by the peptide molecules occurs *via* the water molecule $W(4)$ that forms a hydrogen bond to the tyrosyl oxygen C'OTyr in molecule C' in the upper sheet and to the carbonyl oxygen BO'_5 in molecule B in the lower sheet. In addition, $W(4)$ forms a third hydrogen bond with $W(3)$ that in turn forms a hydrogen bond to AO'_5 in the lower sheet and to the solvent molecule DMFA 1, Table 6. There are no other direct links *via* hydrogen bonding between the β-sheets.

The remainder of the space between the sheets of peptide molecules is filled mainly with DMFA molecules. The locations of eight DMFA molecules that are moderately ordered have been established thus far. In these more ordered DMFA molecules, the O atoms act as acceptors for hydrogen bonds either with water molecules or OH groups from the three remaining Tyr side groups. That is, DMFA 2 accepts a proton from AOTyr, Table 6. BOTyr donates its proton to $W(5)$ that in turn hydrogen bonds with DMFA 3 and DMFA 4. DOTyr donates its proton to DMFA 5 while $W(6)$ forms hydrogen bonds with DMFA 6 and DMFA 7.

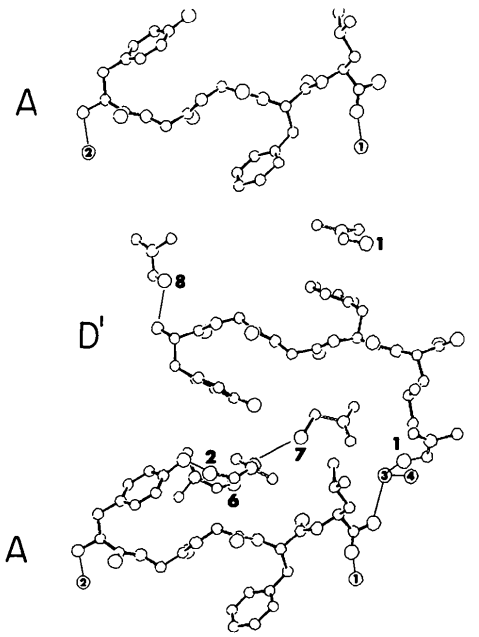


Fig. 4. Enkephalin structure along the b axis with z between ~0 and ~$\frac{1}{4}$. The axial directions are b ↑ and a ←. Represented are: two molecules of conformer A at x,y,z and x, 1 + y, z; conformer D' at $\bar{x}$, $\frac{1}{2}$ + y, 1 − z; DMFA molecules 1, 2, 6, 7 and 8 (large numbers); and water molecules 1, 2, 3, 4. Water molecule 6 is hidden behind DMFA 2.

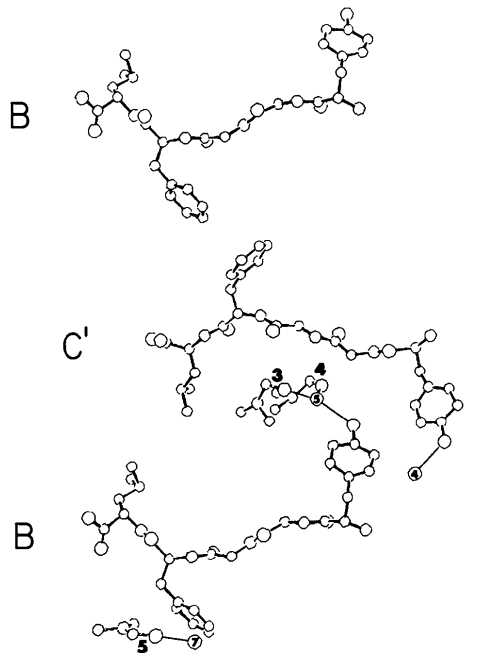


Fig. 5. Enkephalin structure along the b axis with z between ~$\frac{1}{4}$ and ~$\frac{1}{2}$. The axial directions are b ↑ and a ←. Represented are: two molecules of conformer B at x,y,z and x, 1 + y, z; conformer C' at $\bar{x}$, $\frac{1}{2}$ + y, 1 − z; DMFA molecules 3, 4 and 5 (large numbers); and water molecules 4, 5 and 7. Figs. 4 and 5 together represent an asymmetric unit.

Table 6. *Possible hydrogen bonds to solvent molecules (approximate lengths in* Å)

E.s.d.'s are near 0·04–0·06 Å.

AOTyr···O(DMFA 2)	2·34	$W(3)$···O(DMFA 1)	2·98
BOTyr···$W(5)$	2·68	$W(5)$···O(DMFA 3)	3·03
COTyr···$W(4)$	2·62	$W(5)$···O(DMFA 4)	2·45
DN_1···O(DMFA 8)	2·75	$W(6)$···O(DMFA 6)	2·67
DOTyr···O(DMFA 5)	2·79	$W(6)$···O(DMFA 7)	3·11
$W(1)$···$W(8)$	2·74		

(Note that DMFA 5 is shown in Fig. 5 for clarity, since, in Fig. 4, in occupying the symmetry-equivalent position $\bar{x}, \frac{1}{2}+y, 1-z$, it would be hidden behind other solvent molecules.) Solvent molecule DMFA 8 may be connected by a hydrogen bond to DN_1.

All the protons available for hydrogen bonding interact with the above eight DMFA molecules. The remainder of the solvent molecules have no anchor to the peptide sheets and apparently no anchor to discrete positions in the unit cell. Difference maps show that several possible DMFA molecules may be grossly disordered in a diagonal band between enkephalin molecules *C* and *D* near the Gly[2] and Gly[3] residues. Furthermore, as is evident in Fig. 5, along the *b* axis between molecules *B* and *C'* there are large voids surrounded by the nonpolar side chains of the enkephalin molecules. These voids may be filled with rather fluid solvent molecules.

The rather fluid character of the solvent layers lying between the β-sheets of the peptides seems to have a directional effect on the anisotropic thermal factors, especially for the backbone atoms where the B_{22} components are considerably larger than the B_{11} or B_{33} components. The sizes of the B_{ii} values are correlated with the relative strength of the bonding in the *xz* plane as compared to that in the *y* direction. The elongation along the *y* direction of all the thermal ellipsoids for the peptide backbone atoms, the general absence of hydrogen bonds perpendicular to the β-sheet and the disorder or fluidity of the solvent are all consistent with undulations along the β-sheet.

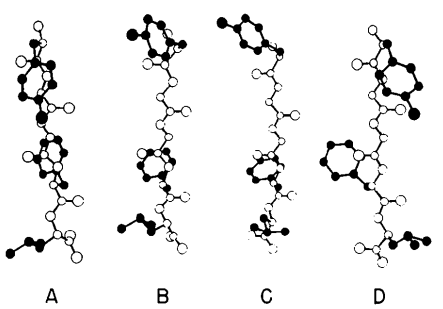


Fig. 6. Four conformations of [Leu⁵]enkephalin occurring in a crystal grown from DMFA. The individual molecules in the drawing have been aligned to illustrate the similarities in the backbone conformations and the differences in the side-chain conformations. The side chains have been shaded.

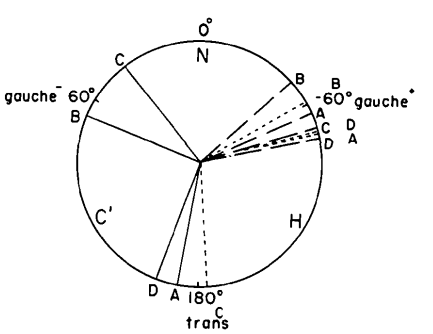


Fig. 7. The χ_{i1} torsional angles about $N_iC_i^{\alpha}C_i^{\beta}C_i^{\gamma}$ for the side chains in the four conformers of [Leu⁵]enkephalin, *A*, *B*, *C* and *D*, shown in Fig. 6. The center of the diagram represents the $C^{\alpha}-C^{\beta}$ axis.

The molecules

In the four individual molecules of enkephalin, the greatest differences in conformation lie not in the backbones, although the pleats in the backbones of molecules *B* and *C* are flatter than those of molecules *A* and *D*, Figs. 4 and 5, but in the conformations of the side chains. In Fig. 6, the four molecules have been aligned with their backbones in similar orientations, and the side chains have been darkened to emphasize the differences in their conformations. The χ_{i1} torsional angles about $N_iC_i^{\alpha}C_i^{\beta}C_i^{\gamma}$ for the twelve side chains are plotted in Fig. 7. In the trimodal distribution, the *gauche*$^-$ position ($\chi_{i1} = +60°$) is the least populated while the *gauche*$^+$ position is most populated. This distribution is similar to that found in a survey of side-chain conformations in proteins (Janin, Wodak, Levitt & Maigret, 1978).

The individual conformations of the side chains do not appear to be correlated with local hydrogen bonding. There may be some correlation with the formation of a β-sheet, however, since the side chains in adjacent molecules in the sheet tend to form clusters above and below the sheet, Fig. 2. It appears possible that the conformation of an enkephalin molecule is readily affected by its microenvironment whether in a crystal as demonstrated in this paper or in a biological milieu during the binding process to a receptor.

Biological relevance of the structure

Predictions of enkephalin solution conformations based on proton magnetic resonance spectra have generally resulted in folded structures, differing primarily in the location of the bend. However, these postulations have been challenged by spectroscopic and theoretical studies (Khaled, Long, Thompson, Bradley, Brown & Urry, 1977; DeCoen, Humblet & Koch, 1977) which have shown that the NMR results can be equally well explained by an extended conformation, and that the extended conformation is the most energetically favored one. Recently (DiMaio & Schiller, 1980) a semi-rigid cyclic analogue of Leu-enkephalin has been described which is biologically active and which is constrained to have a conformation with many characteristics similar to those observed for Leu-enkephalin in this crystal structure determination (*i.e.* extended backbone with Leu and Phe side chains on

opposite sides). The above data suggest: (1) the extended crystal conformation is a favored one in solution (this indication is, of course, reinforced by the presence of much solvent in the crystal), and (2) since the cyclic analogue demonstrates a strong preference for the opiate receptor subclass which binds morphine-related opiates, stereochemical relationships between enkephalins and narcotic analgesics should be sought utilizing the conformations observed in this study. Fuller details will be given elsewhere (Camerman, Mastropaolo, Karle, Karle & Camerman, 1983).

This research was supported in part by NIH grant GM30902, by USPHS grants DA02042 and NS09839, by the University of Washington Graduate School Research Fund and Alcohol and Drug Abuse Institute, and by the Medical Research Council of Canada. We thank Drs Sam Wilkinson and D. H. Coy for supplying leucine-enkephalin and Dr L. H. Jensen for use of his diffractometer.

APPENDIX
Use of Σ_3 formula

The Σ_3 formula gives the sign relationship between two centrosymmetric reflections, given the observed values for a set of intensities. If the sign of one of the reflections is known, that of the second can thus be determined. A virtue of the Σ_3 formula is that if the signs of several centrosymmetric reflections are known, it is possible to obtain more than one indication for an unknown sign. The occurrence of a large number of large normalized structure factor magnitudes permitted the useful application of the Σ_3 formula to the structure determination of enkephalin.

The Σ_3 formula was first written for centrosymmetric crystals (Hauptman & Karle, 1953) and a general formula for both centrosymmetric and noncentrosymmetric crystals was presented later (Karle, 1969). In the latter case, the formula is applicable only to pairs of phases that are real or pure imaginary. It will be evident from the formula that a pair of phases whose sign relationship is being determined must be of the same parity.

Formulas

The Σ_3 formula, where $sE_\mathbf{h}$ means 'sign of $E_\mathbf{h}$', can be written

$$sE_\mathbf{h} = s\sum_\mu \sum_\nu \frac{m_{112}}{m_{200}^{1/2}\, m_{020}^{1/2}\, m_{002}} E_{\mathbf{h}_\nu}(|E_{\mathbf{h}_\mu}|^2 - 1). \quad (A1)$$

This formula gives meaningful information when

$$\mathbf{h} - \mathbf{h}_\nu - 2\mathbf{h}_\mu = 0, \quad (A2)$$

where $\mathbf{h}$ and $\mathbf{h}_\nu$ are the indices of real or pure imaginary reflections. Formula (A1) can also give meaningful information when some particular index that is equal to zero for $\mathbf{h} - \mathbf{h}_\nu$ is nonzero for $\mathbf{h}_\mu$, *e.g.* in space group $P2_1$, $k = 0$ in $\mathbf{h} - \mathbf{h}_\nu$ but takes on all values in $\mathbf{h}_\mu$. In fact, in space group $P2_1$, the sum over μ in formula (A1) means the sum over this latter index. Equation (A2) shows that $\mathbf{h}$ and $\mathbf{h}_\nu$ must be of the same parity. The probability that the sign of $E_\mathbf{h}$ is plus, as determined by (A1), is given by

$$P_+(E_\mathbf{h}) \simeq \tfrac{1}{2} + \tfrac{1}{2}\tanh \frac{m_{112}\, n\sigma_4}{q m_{200}^{1/2}\, m_{020}^{1/2}\, m_{002}\, \sigma_2^2} \times |E_\mathbf{h}| \sum_\mu \sum_\nu E_{\mathbf{h}_\nu}(|E_{\mathbf{h}_\mu}|^2 - 1), \quad (A3)$$

where the E are normalized structure factors, n is the symmetry number of the space group,

$$\sigma_p = \sum_{j=1}^{N} Z_j^p, \quad (A4)$$

N is the number of atoms in the unit cell, Z_j is the atomic number of the jth atom, $q = 2$ when $\mathbf{h}_\mu$ represents indices of real or pure imaginary reflections, $q = 1$ when $\mathbf{h}_\mu$ represents indices of general noncentrosymmetric reflections and

$$m_{a_1\ldots a_r}^{b_1\ldots b_r} = \int_0^1\int_0^1\int_0^1 \xi^{a_1}(x,y,z;\mathbf{h}_1)\ldots\xi^{a_r}(x,y,z;\mathbf{h}_r) \times \eta^{b_1}(x,y,z;\mathbf{h}_1)\ldots\eta^{b_r}(x,y,z;\mathbf{h}_r)\,dx\,dy\,dz. \quad (A5)$$

In deriving formulas (A1) and (A3) when $E_{\mathbf{h}_\mu}$ is a noncentrosymmetric reflection, it was assumed that $m_{12}^{02} = m_{10}^{02}$ and $m_{112}^{000} = m_{110}^{002}$. Before making an application in a particular space group, these relationships among the moments should be verified. The functions ξ and η are found in *International Tables for X-ray Crystallography* (1965) wherein they define the structure factors for centrosymmetric crystals,

$$F_\mathbf{h} = \sum_{j=1}^{N/n} f_{j\mathbf{h}}\, \xi(x_j, y_j, z_j;\mathbf{h}), \quad (A6)$$

and noncentrosymmetric ones,

$$F_\mathbf{h} = X_\mathbf{h} + iY_\mathbf{h}, \quad (A7)$$

where

$$X_\mathbf{h} = \sum_{j=1}^{N/n} f_{j\mathbf{h}}\, \xi(x_j, y_j, z_j;\mathbf{h}) \quad (A8)$$

and

$$Y_\mathbf{h} = \sum_{j=1}^{N/n} f_{j\mathbf{h}}\, \eta(x_j, y_j, z_j;\mathbf{h}). \quad (A9)$$

The calculation of the moments in formula (*A*1) can result in different magnitudes depending upon the characteristics of h, h_ν and h_μ such as dimensionality and other details concerning the interrelationship of these indices. In applying the probability formula (*A*3) no distinction has been made among the coefficients obtained from the evaluation of the moments. Instead, the most conservative value has been used for all calculations resulting in the formula

$$P_+(\mathbf{h}) \simeq \tfrac{1}{2} + \tfrac{1}{2}\tanh \frac{\sigma_4}{q\sigma_2^2} |E_\mathbf{h}| \sum_\mu \sum_\nu E_{\mathbf{h}_\nu}(|E_{\mathbf{h}_\mu}|^2 - 1) \times (-1)^{f(\mathbf{h}_\mu)}, \qquad (A10)$$

where $f(\mathbf{h}_\mu)$ is derived from the calculation of moments, formula (*A*5).

Application of Σ_3 in $P2_1$

The following considerations are germane to the application of the Σ_3 formula in space group $P2_1$.

Sums over μ

The sum over μ in formulas (*A*1) and (*A*10) can be labeled by means of a one- or two-dimensional index according to (*A*2). There is only one type of sum for space group $P2_1$, *i.e.* over k_μ. Consider

$$0{\cdot}5(h - h_\nu) = h_\mu \neq 0$$
$$0{\cdot}5(k - k_\nu) = k_\mu \neq 0$$
$$0{\cdot}5(l - l_\nu) = l_\mu \neq 0, \qquad (A11)$$

where $\mathbf{h} \equiv (h,0,l)$, $\mathbf{h}_\nu \equiv (h_\nu,0,l_\nu)$ and $\mathbf{h}_\mu \equiv (h_\mu,0,l_\mu)$ for one of the definitions of $\mathbf{h}_\mu$ (that used for labeling), but may be defined in general by $\mathbf{h}_\mu \equiv (h_\mu,k_\mu,l_\mu)$. The index k_μ can take on all values provided by the experimental data, giving for the sum over μ in (*A*10).

$$q^{-1} \sum_{k_\mu=0}^{k_\mu(\max.)} (-1)^{k_\mu}(|E_{\mathbf{h}_\mu}|^2 - 1). \qquad (A12)$$

Note that for $P2_1$ there are really two different labels that can be assigned to the sum given by (*A*12), namely, $h_\mu,0,l_\mu$ and $\bar{h}_\mu,0,\bar{l}_\mu$. Note also that in this space group $\bar{h}_\mu,0,l_\mu$ is distinct in phase and magnitude from $h_\mu,0,l_\mu$ and will give rise to a separate sum over k_μ which can also be labeled as $h_\mu,0,\bar{l}_\mu$.

It may turn out that

$$0{\cdot}5(h - h_\nu) = (0,0,l_\mu) \text{ or } (h_\mu,0,0). \qquad (A13)$$

In these cases, the sum in (*A*12) still applies. This sum is the same as that for the Σ_1 formula for space group $P2_1$ and will be so referred to.

The first step in the application of the Σ_3 formula is to compute a complete set of Σ_1 sums for all possible labels, $h_\mu,0,l_\mu$ and $h_\mu,0,\bar{l}_\mu$ which are associated in pairs with the same sum and similarly for $\bar{h}_\mu,0,l_\mu$ and $h_\mu,0,\bar{l}_\mu$.

Procedure

(1) Select a number of $\mathbf{h}$ associated with the largest $|E_\mathbf{h}|$ and the largest Σ_1 sums labeled with $\mathbf{h}_\mu$.

(2) For each individual pair of $\mathbf{h}$ and $\mathbf{h}_\mu$ (label), there is a corresponding $\mathbf{h}_\nu$. The indices of all the $\mathbf{h}_\nu$ associated with the pairs $\mathbf{h}$ and $\mathbf{h}_\mu$ are found and the probability formula (*A*10) with $E_{\mathbf{h}_\nu}$ replaced by $|E_{\mathbf{h}_\nu}|$ for a specific $\mathbf{h}_\nu$ is computed:

$$P = \tfrac{1}{2} + \tfrac{1}{2}\tanh \sigma_4 \sigma_2^{-2} |E_\mathbf{h}|\, |E_{\mathbf{h}_\nu}|\, \Sigma_1(\mathbf{h}_\mu). \qquad (A14)$$

If P is close to 1, the phases of $E_\mathbf{h}$ and $E_{\mathbf{h}_\nu}$ are probably the same; if close to zero, the phases of $E_\mathbf{h}$ and $E_{\mathbf{h}_\nu}$ are probably opposite, differing by π.

(3) The largest values of P are sorted out and those very close to unity or zero would be expected to be quite definitive. In a structure as complex as enkephalin, $\sigma_4 \sigma_2^{-2}$ is so small that despite the unusually large $|E|$ values present, the values for P computed from eq. (*A*14) are not definitive but, nevertheless, suggestive and can be used in combination with the Symbolic Addition Procedure to imply the more likely values for the symbols. Table 7 shows such an application.

(4) As the application of the Σ_3 formula proceeds, interrelationships develop among phases associated with pairs of indices $\mathbf{h}$ and $\mathbf{h}_\nu$ to the extent that

Table 7. *Illustrations of the application of the* Σ_3 *formula to the enkephalin structure*

P was calculated by use of formula (*A*14). Conclusions were drawn from a comparison of the Σ_3 calculation with numerical and symbolic definitions of phases obtained from application of the Symbolic Addition Procedure.

h	$\lvert E_\mathbf{h}\rvert$	$\mathbf{h}_\nu$	$\lvert E_{\mathbf{h}_\nu}\rvert$	$\Sigma_1(\mathbf{h}_\nu)$	P	Conclusion
$\bar{1}\,0\,19$	8·29	$8\,0\,\bar{9}$	2·80	7·74	0·71	$a = \pi$
$\bar{1}\,0\,19$	8·29	$\bar{1}\,0\,\bar{17}$	2·19	−8·97	0·31	$c = a + b$
$\bar{1}\,0\,19$	8·29	$\bar{1}\,0\,17$	2·19	−12·97	0·24	$c = a + b$
$\bar{1}\,0\,19$	8·29	$4\,0\,9$	2·45	−9·05	0·29	Corrob.*
$\bar{1}6\,0\,4$	5·87	$14\,0\,4$	2·10	12·43	0·68	Corrob.
$\bar{1}6\,0\,4$	5·87	$\bar{1}6\,0\,6$	2·77	−12·97	0·26	$c = a + b$
$\bar{1}6\,0\,4$	5·87	$\bar{1}6\,0\,\bar{4}$	2·08	−38·35	0·09	Corrob.
$\bar{1}7\,0\,4$	3·89	$17\,0\,\bar{12}$	4·26	−16·01	0·21	Corrob.
$13\,0\,7$	3·58	$13\,0\,\bar{1}$	2·28	−38·35	0·17	Corrob.
$8\,0\,9$	3·00	$\bar{8}\,0\,1$	3·57	−38·35	0·12	Corrob.
$\bar{1}6\,0\,6$	2·77	$\bar{1}4\,0\,6$	2·70	−23·39	0·29	Corrob.
$\bar{4}\,0\,2$	3·71	$\bar{6}\,0\,2$	2·57	−23·39	0·25	Not used
$\bar{8}\,0\,9$	2·80	$\bar{8}\,0\,1$	2·63	−38·35	0·20	Not used

* Corrob. implies that a relationship between signs already found in the application of the Σ_2 formula in the Symbolic Addition Procedure is corroborated by the Σ_3 calculation.

opportunities arise to make tests of internal consistency. It is evidently valuable to make use of such tests since they can improve the probabilities considerably.

Remarks

The implications of the column labeled P in Table 7 are that when $P > 0{\cdot}5$ the signs of $E_{\mathbf{h}}$ and $E_{\mathbf{h}_p}$ are likely to be the same and when $P < 0{\cdot}5$ the signs of $E_{\mathbf{h}}$ and $E_{\mathbf{h}_p}$ are likely to be opposite with the probabilities as given. The last two indications in Table 7 were incorrect. In this application, they were not used because the reflections involved had not played a role in the phase determination at the stage that the comparison was made. It should be particularly noted that three essentially independent computations resulted in the relation $c = a + b$. Although the individual probabilities in each of the three computations are rather undefinitive, the combined probabilities imply that the relation has a probability of about 0·98 that it is correct.

The application given in Table 7 involves the comparison of the signs of one $E_{\mathbf{h}}$ and one $E_{\mathbf{h}_p}$ by use of formula (A14) and the probabilities obtained are not strongly definitive except when several of them can be combined. In another type of application, it is seen that if the signs of several $E_{\mathbf{h}_p}$ associated with a given $E_{\mathbf{h}}$ and large Σ_1 sums are known, formula (A10) has the potential to produce values of $P_+(\mathbf{h})$ that are very close to 1 or to zero. In this way, the Σ_1 formula can participate in the later stages of a phase determination as well as the earliest ones, as outlined above. Current procedures for phase determination generally result in a variety of phase sets to which various criteria are applied in order to eliminate incorrect sets. It would seem that the Σ_3 formula could also play such a role since the sets contain phase values for the $E_{\mathbf{h}}$ that could be tested by the Σ_3 formula to determine which of them are most consistent with the implications of the Σ_3 formula.

References

ARNOTT, S., DOVER, S. D. & ELLIOTT, A. (1967). *J. Mol. Biol.* **30**, 201–208.

BLUNDELL, T. L., HEARN, L., TICKLE, I. J., PALMER, R. A., MORGAN, B. A., SMITH, G. D. & GRIFFIN, J. F. (1979). *Science*, **205**, 220.

CAMERMAN, A., MASTROPAOLO, D., KARLE, I. L., KARLE, J. & CAMERMAN, N. (1983). *Nature (London)*. In the press.

DECOEN, J. L., HUMBLET, C. & KOCH, M. H. J. (1977). *FEBS Lett.* **73**, 38–42.

DIMAIO, J. & SCHILLER, P. W. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 7162–7166.

FAWCETT, J. K., CAMERMAN, N. & CAMERMAN, A. (1975). *Acta Cryst.* B**31**, 658–665.

FLIPPEN-ANDERSON, J., GILARDI, R. & KONNERT, J. H. (1983). Program *RESLSQ*. NRL Memorandum Report 5042, Naval Research Laboratory, Washington, DC 20375, USA.

HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* Am. Crystallogr. Assoc. Monogr. 3. Western Springs: Polycrystal Book Service.

HUGHES, J., SMITH, T. W., KOSTERLITZ, H. W., FOTHERGILL, L. A., MORGAN, B. A. & MORRIS, H. R. (1975). *Nature (London)*, **258**, 577–579.

International Tables for X-ray Crystallography (1965). Vol. I. Birmingham: Kynoch Press.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press.

JANIN, J., WODAK, S., LEVITT, M. & MAIGRET, B. (1978). *J. Mol. Biol.* **125**, 357–386

KARLE, I. L. (1974). *J. Am. Chem. Soc.* **96**, 4000–4006.

KARLE, I. L. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 27–70. Copenhagen: Munksgaard.

KARLE, I. L. (1977). *J. Am. Chem. Soc.* **99**, 5152–5157.

KARLE, I. L., GILARDI, R. D., FRATINI, A. V. & KARLE, J. (1969). *Acta Cryst.* B**25**, 1469–1479.

KARLE, I. L. & KARLE, J. (1968). *Acta Cryst.* B**24**, 81–91.

KARLE, I. L. & KARLE, J. (1971). *Acta Cryst.* B**27**, 1891–1898.

KARLE, J. (1969). *Advances in Chemical Physics*, Vol. XVI, edited by I. PRIGOGINE & S. A. RICE, pp. 131–222. New York: Interscience.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–859.

KHALED, M. A., LONG, M. M., THOMPSON, W. D., BRADLEY, R. J., BROWN, G. B. & URRY, P. W. (1977). *Biochem. Biophys. Res. Commun.* **76**, 224–231.

KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

SIELECKI, A. R., HENDRICKSON, W. A., BROUGHTON, C. G., DELBAERE, L. T. G., BRAYER, G. D. & JAMES, M. N. G. (1979). *J. Mol. Biol.* **134**, 781–804.

SMITH, G. D. & GRIFFIN, J. F. (1978). *Science*, **199**, 1214–1216.