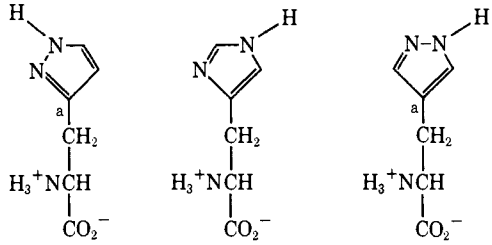
The Crystal and Molecular Structure of β-(Pyrazolyl-3)-L-alanine[1,2]

N. C. Seeman, E. L. McGandy, and R. D. Rosenstein*

Contribution from the Departments of Biochemistry and Crystallography, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.
Received April 28, 1971

Abstract: The crystal structure of β-(pyrazolyl-3)-L-alanine, $\overline{CHCHNHNC}CH_2CH(NH_3^+)COO^-$, has been solved using Patterson superposition techniques. Monoclinic crystals, obtained from an ethanol–water solution, had lattice constants at room temperature of a = 4.62 (2), b = 7.55 (2), c = 10.06 (2) Å, and β = 98.61 (1)°. The space group of the crystals was $P2_1$, with two molecules in the unit cell; D_x = 1.451 g cm^{-3} and D_m = 1.419 g cm^{-3}. Intensity data were collected by a diffractometer; the structure was refined by full-matrix least squares to an R factor of 0.078. The structure has two separate regions: one hydrophilic, which makes use of all four possible hydrogen bonds, and the other hydrophobic, involving a characteristic (perpendicular) stacking of bases. The labile proton of the pyrazole ring has been located, associated with the nitrogen atom furthest from the alanyl moiety.

The replacement of an individual amino acid subunit, within a given polypeptide, is one of the most powerful methods of evaluating its chemical role in biochemical systems. β-(Pyrazolyl-3)-L-alanine (Pyr(3)-Ala) (I) is ideally suited as a substitute for L-histidine (II). The side chains are predictably isosteric, and both



I II III

terminate in five-membered aromatic rings containing two nitrogen atoms. Furthermore, rotation about bond a of I would allow one of the two nitrogen atoms of the pyrazole ring to occupy the same spatial position, with respect to other polypeptide atoms, as a nitrogen of the imidazole ring of histidine.

Finn and Hofmann[3] have shown that, within the bovine pancreatic ribonuclease S-peptide–S-protein[4] system, 12-(Pyr(3)Ala) S-peptide (1–14 fragment) manifests binding to S-protein equal to that exhibited by the complete S-peptide (1–20 fragment of the ribonuclease molecule) which contains histidine in position 12. It was also noted[5] that, although the 1–14 fragment of S-peptide had catalytic activity in the S-peptide–S-protein system,[6] the 12-(Pyr(3)Ala)-S-peptide$_{1-14}$ had none, even at elevated peptide–protein ratios. This lack of activity was ascribed to the absence of a second proton on the pyrazole ring of Pyr(3)Ala at physiological pH. On the basis of this and alkylation evidence,[7] His-12 of the S-peptide was assigned a catalytic role in the S-peptide–S-protein system.

In the same spirit, Pyr(3)Ala has also been substituted for histidine in angiotensin,[8] β-corticotropin,[9] and thyrotropin-releasing hormone.[10] In all instances, the substitution lowered but did not destroy the biological activity; the acid–base properties of histidine were thus not inferred as necessary for the biological function of those peptides.

Experimental Section

Blade-shaped crystals from ethanol–water solution were provided by Dr. Klaus Hofmann of the Protein Research Laboratory of the University of Pittsburgh. These crystals were monoclinic, space group $P2_1$, with two molecules of $C_6H_9N_3O_2$ (mol wt, 155 daltons) per unit cell. The density, D_m = 1.42 (2) g cm^{-3}, was measured by flotation in CCl_4 and toluene (D_x = 1.451 g cm^{-3}). A clear triangular prismatic crystal (base = altitude = 0.2 mm, thickness = 0.1 mm) was mounted with {010} parallel to the Φ axis of a Picker automatic four-circle diffractometer. The intensity data and lattice parameters were collected with Cu Kα radiation (λ = 1.5418 Å): a = 4.62 (2), b = 7.55 (2), c = 10.06 (2) Å, and β = 98.61 (1)°. The linear absorption coefficient μ_{Cu} = 6.57 cm^{-1}. Integrated intensities were obtained by means of 10-sec counts using a nine-point unequal interval step scan[11] of 2θ. Balanced (Ross) filters (nickel and cobalt) were used for collecting data below 0.8 reciprocal lattice unit (rlu). The data between 0.8 and 1.8 rlu were collected with Ni filter only, employing a 10-sec count to measure the background.

Solution and Refinement of the Structure. A sharpened, origin-removed Patterson was prepared[12] using normalized structure factors, E, derived on the basis of scale and temperature factors estimated by a Wilson plot.[13] A two-atom symmetry minimum

(1) Supported in part by Research Training Grant No. GM-01728 of the U. S. Public Health Service, National Institutes of Health, and by Grant No. L71 of the Pittsburgh Health Research and Services Foundation.

(2) A preliminary communication of this study was presented at the Summer Meeting of the American Crystallographic Association, Buffalo, N. Y., Aug 1968.

(3) F. Finn and K. Hofmann, *J. Amer. Chem. Soc.*, **89**, 5298 (1967).

(4) F. M. Richards, *Proc. Nat. Acad. Sci. U. S.*, **44**, 162 (1958): RNase S, subtilisin modified beef ribonuclease RNase A; S-peptide, the eicosapeptide obtained from RNase S; S-protein, the protein component obtained from RNase S (see ref 6 in K. Hofmann, *et al.*, *J. Amer. Chem. Soc.*, **88**, 3633 (1966)).

(5) K. Hofmann and H. Bohn, *ibid.*, **88**, 5914 (1966).

(6) K. Hofmann, F. Finn, M. Limetti, J. Montibeller, and G. Zanetti, *ibid.*, **87**, 645 (1965).

(7) A. M. Crestfield, W. H. Stein, and S. Moore, *J. Biol. Chem.*, **238**, 2413 (1963).

(8) K. Hofmann, R. Andreatta, J. P. Buckely, W. E. Hageman, and A. P. Shapire, *J. Amer. Chem. Soc.*, **90**, 1654 (1968).

(9) K. Hofmann, H. Bohn, and R. Andreatta, *ibid.*, **89**, 7126 (1967).

(10) K. Hofmann and C. Y. Bowers, *J. Med. Chem.*, **13**, 1099 (1970).

(11) E. L. McGandy and R. L. Snyder, Abstracts, 26th Pittsburgh Diffraction Conference, Pittsburgh, Pa., 1968.

(12) R. Shiono, Technical Report, Crystallography Laboratory, University of Pittsburgh, 1966.

(13) R. Shiono, Technical Report, Department of Crystallography, University of Pittsburgh, 1969.

function,[14,15] SMF, located an arbitrary two-atom rigid body with respect to the symmetry element. The multiple minimum function, MMF, of order four (with imposed symmetry) based on the two-atom SMF implied further atomic positions, and higher order MMF's were therefore calculated. Possible positions for seven atoms were implied by the recurrence of peaks in the MMF's. Fourier and difference Fourier calculations indicated that two of these seven atoms were false and suggested positions for six others. At this point the Pyr(3)Ala molecule was recognized from the geometry of the 11 derived positions. The atomic positional and isotropic thermal parameters were refined by full-matrix least squares,[12] minimizing $\Sigma w(\Delta F)^2$ to a residual, $R = \Sigma||F_o| - |F_c||/\Sigma|F_o| = 0.16$. Anisotropic refinement of the temperature factors reduced R to 0.085. Difference maps revealed all nine hydrogen atoms, the positions of which were arbitrarily idealized at distances of 1.08 and 1.03 Å from the carbon and nitrogen atoms, respectively, to which they were attached. In further refinement, these positions were left fixed, and the hydrogen atoms were arbitrarily assigned the isotropic temperature factors of the atoms to which they were bonded. Two more cycles of least-squares refinement were then carried out, weighing each reflection by $w(F) = 1/\sigma^2(F) = 4F^2/\sigma_c^2$ (where σ_c^2 is the counting-statistical variance (Poisson) of the observations), resulting in a final R factor of 0.078 (weighted $R = 0.077$). The structure factors are listed in the microfilm edition of this journal,[16] the final atomic coordinates and estimated standard deviations are listed in Table I, bond lengths in Table II, bond angles in Table III, and rms amplitudes of vibration in Table IV. The 50% probability thermal ellipsoids and the numbering convention are shown in Figure 1.

Table I. Atomic Coordinates and Estimated Standard Deviations for β-(Pyrazolyl-3)-L-alanine

Atom	x	y	z
O-1	0.1163 (8)	−0.1390[a]	0.0293 (5)
O-2	0.0101 (10)	−0.1030 (10)	0.2322 (5)
N-1	0.3846 (10)	0.1687 (10)	0.0303 (5)
N-2	0.5814 (12)	0.4891 (10)	0.2247 (5)
N-3	0.7892 (13)	0.5747 (10)	0.3042 (6)
C-1	0.1393 (14)	−0.0619 (11)	0.1396 (7)
C-2	0.3526 (13)	0.0931 (11)	0.1625 (6)
C-3	0.2620 (4)	0.2345 (12)	0.2523 (6)
C-4	0.4937 (12)	0.3616 (11)	0.3037 (6)
C-5	0.6482 (13)	0.3747 (12)	0.4305 (6)
C-6	0.8390 (15)	0.5087 (14)	0.4284 (7)
H-1–N-1	0.400	0.045	−0.011
H-2–N-1	0.203	0.236	−0.013
H-3–N-1	0.590	0.234	0.058
H-4–C-2	0.542	0.023	0.211
H-5–C-3	0.240	0.178	0.348
H-6–C-3	0.060	0.299	0.211
H-7–C-5	0.635	0.277	0.508
H-8–C-6	1.039	0.546	0.490
H-9–N-3	0.866	0.701	0.289

[a] Fixed parameter (polar space group). All hydrogen atom positions were arbitrarily fixed. The numbers in parentheses are the estimated standard deviations in the last decimal places.

Table II. Bond Lengths and Estimated Standard Deviations for Pyr(3)Ala

Atom 1	Atom 2	Distance, Å
C-1	O-1	1.243 (8)
C-1	O-2	1.219 (8)
C-1	C-2	1.525 (9)
C-2	N-1	1.477 (8)
C-2	C-3	1.498 (9)
C-3	C-4	1.472 (10)
C-4	N-2	1.348 (9)
C-4	C-5	1.369 (10)
C-5	C-6	1.344 (10)
C-6	N-3	1.332 (10)
N-2	N-3	1.323 (9)

Hydrogen bonds

Atom			Distance	Angle
i	j	k[a]	i–k, Å	jik, deg
N-1	H-1	N-2a	2.925	37.6
N-1	H-2	O-1b	2.729	12.3
N-1	H-3	O-1c	2.857	28.5
N-3	H-9	O-2d	2.776	7.2

[a] Symmetry code: a, $\bar{x}, -1/2 + y, \bar{z}$: b, $\bar{x}, 1/2 + y, \bar{z}$: c, $1 - x, 1/2 + y, \bar{z}$: d, $1 + x, 1 + y, z$.

Table III. Bond Angles and Estimated Standard Deviations for Pyr(3)Ala

i	j	k	$\angle_{ijk}$, deg[a]
O-1	C-1	O-2	125.2 (6)
O-1	C-1	C-2	117.4 (6)
O-2	C-1	C-2	117.3 (6)
C-1	C-2	N-1	108.2 (5)
C-1	C-2	C-3	113.9 (6)
N-1	C-2	C-3	110.2 (5)
C-2	C-3	C-4	114.9 (6)
C-3	C-4	N-2	121.7 (6)
C-3	C-4	C-5	128.7 (6)
N-2	C-4	C-5	109.7 (6)
C-4	N-2	N-3	104.3 (6)
N-2	N-3	C-6	113.1 (6)
N-3	C-6	C-5	105.9 (7)
C-4	C-5	C-6	107.0 (6)

[a] The numbers in parentheses are the estimated standard deviations in the last decimal place.

Table IV. Root-Mean-Square Amplitudes of Vibration for the Heavy Atoms of β-(Pyrazolyl)-3)-L-alanine

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
O-1	0.0542	0.0584	0.0596	−0.0219	0.0107	−0.0154
O-2	0.0841	0.0565	0.0765	−0.0191	0.0408	−0.0052
N-1	0.0438	0.0437	0.0491	−0.0005	0.0070	0.0049
N-2	0.0535	0.0603	0.0524	−0.0099	0.0099	0.0004
N-3	0.0572	0.0526	0.0751	−0.0142	0.0173	−0.0058
C-1	0.0424	0.0458	0.0557	−0.0012	0.0018	0.0000
C-2	0.0447	0.0416	0.0481	0.0016	0.0035	0.0053
C-3	0.0493	0.0457	0.0478	−0.0033	0.0127	−0.0010
C-4	0.0502	0.0337	0.0546	−0.0003	0.0120	0.0030
C-5	0.0544	0.0505	0.0514	−0.0084	0.0066	−0.0078
C-6	0.0565	0.0860	0.0460	−0.0082	0.0041	−0.0206

Discussion

Since the amino acid was crystallized from a neutral solvent, the zwitterionic form, $PyrCH_2CHNH_3^+COO^-$, was predictable. This was confirmed by the presence of a third hydrogen atom attached to N-1 and the absence of any hydrogen atoms attached to either of the oxygen atoms, and by the length of the C–O distances (1.24 and 1.21 Å). All atoms excluding hydrogens of the pyrazole ring plus the β-carbon were within 0.013 Å of the least-squares plane, with the equation $0.7345x - 0.62400y - 0.37362z - 1.16710 = 0$. Similarly, the carboxyl group plus the α-carbon atom were within 0.009 Å of a plane with the equation $0.69429x =$

(14) P. W. R. Corfield, Technical Report, Crystallography Laboratory, University of Pittsburgh, 1965.

(15) P. W. R. Corfield and R. D. Rosenstein, *Trans. Amer. Crystallogr. Ass.*, 2, 17 (1966).

(16) A table of structure factors will appear following these pages in the microfilm edition of this volume of the journal. Single copies may be obtained from the Business Operations Office, Books and Journals Division, American Chemical Society, 1155 Sixteenth St., N.W., Washington, D. C. 20036, by referring to author, title of article, volume, and page number. Remit check or money order for $3.00 for photocopy or $2.00 for microfiche.

Table V. Distances Calculated for the Histidine Free-Base Series[a]

	L-Histidine (Edinburgh)	L-Histidine (Pittsburgh)	L-Histidine (Pittsburgh)	DL-Histidine (Edinburgh)	β-(Pyrazolyl-3)-L-alanine (Pittsburgh)
Ref	*b, c*	*b, c*	*d*	*e*	*f*
Space group	$P2_1$	$P2_1$	$P2_12_12_1$	$P2_1/c$	$P2_1$
a,b,c	5.17, 7.39, 9.46	5.172, 7.384, 9.474	5.177, 7.322, 18.870	8.983, 8.087, 9.415	4.623, 7.548, 10.063
β	97.93	97.162		97.65	98.61
Data collection	Cu Kα, film	Mo Kα, diffractometer	Mo Kα, diffractometer	Cu Kα, film	Cu Kα, diffractometer
R, %	10.0	11.1	3.4	10.8	7.8
Bond distance, Å					
C-1–O-1	1.23 (2)	1.21 (1)	1.247 (2)	1.248 (6)	1.219 (9)
C-1–O-2	1.26 (2)	1.25 (1)	1.250 (2)	1.252 (6)	1.243 (8)
C-1–C-2	1.55 (2)	1.55 (1)	1.545 (2)	1.529 (6)	1.525 (9)
C-2–N-1	1.51 (2)	1.48 (1)	1.493 (2)	1.482 (6)	1.477 (9)
C-2–C-3	1.53 (2)	1.53 (1)	1.536 (3)	1.537 (6)	1.498 (9)
C-3–C-4	1.49 (2)	1.50 (1)	1.505 (3)	1.503 (6)	1.472 (9)
C-4–N-2	1.36 (2)	1.39 (1)	1.382 (2)	1.385 (6)	1.348 (9)
N-2–N-3	1.33 (2)	1.34 (1)	1.327 (3)	1.314 (6)	[1.323 (9)]
N-3–C-6	1.38 (2)	1.34 (1)	1.339 (3)	1.359 (6)	[1.332 (10)]
C-6–C-5	1.40 (2)	1.38 (1)	1.374 (3)	1.374 (6)	[1.344 (11)]
C-5–C-4	1.40 (2)	1.39 (1)	1.361 (3)	1.374 (6)	1.369 (10)

[a] In β-(pyrazolyl-3)-L-alanine, N-3 and C-5 are reversed from their positions in histidine, but all other atoms are in identical positions, making Pyr–Ala an isostructural analog of histidine. All bonds involving these reversed atoms are marked by square brackets, and the atoms are labeled according to the Pyr–Ala numbering scheme. The numbers in parentheses are the estimated standard deviations in the last decimal places. [b] M. Harding and A. Hoy.[20] [c] J. J. Madden, E. L. McGandy, and N. C. Seeman.[20] [d] J. Madden, E. L. McGandy, and N. C. Seeman, submitted for publication in *Acta Crystallogr.* [e] P. Edington, 1969, Ph.D. Dissertation, University of Edinburgh. [f] This work.

$0.62432y + 0.24998z + 1.10131 = 0$. Thus, all of the heavy atoms of the structure, with the exception of N-1, lie in two planes, which meet at an angle of 142.89°.

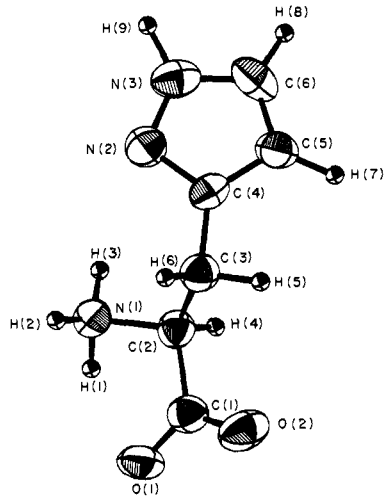


Figure 1. The molecular structure of β-(pyrazolyl-3)-L-alanine, showing the atomic numbering system and 50% probability thermal ellipsoids (drawn by a Calcomp plotter using program ORTEP (C. K. Johnson, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1965, Report No. ORNL-3794)).

The torsion angles (using the IUPAC convention[17]) are: (1) N-1′–C-2–C-1–O-1 = 25.7° (Ψ_T^1); (2) N-1′–C-2–C-1–O-2 = −156.0° (Ψ_T^2); (3) N-1–C-2–C-3–C-4 = −73.2° (χ^1); (4) C-2–C-3–C-4–N-2 = 74.7° ($\chi^{2,1}$); (5) C-2–C-3–C-4–C-5 = −106.1° ($\chi^{2,2}$). When the amino acid is incorporated into a peptide, only angle 1 is involved in the secondary structure of the peptide. This angle is the Ψ angle of the peptide linkage, and the value found here is within the sterically allowed regions for polyalanine.[18]

The unit cell is divided into two distinct regions by a plane parallel to (001) and passing through $z = 0.40$. The portion of the unit cell between 0.4 and 0.6 is distinctly hydrophobic; the remainder of the cell can be described as hydrophilic (Figure 2). The hydrophobic region consists of C-5 and C-6 and the hydrogen atoms bonded to them. The pyrazole planes related by the

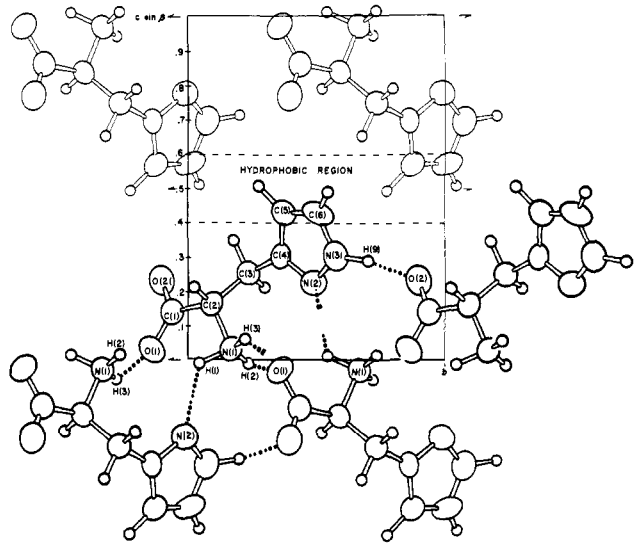


Figure 2. Projection along the *a* axis, showing the molecular arrangement. The hydrophobic region of the unit cell lies between z = 0.4 and 0.6; the hydrophilic region is defined by the hydrogen bonding (drawn by a Calcomp plotter using program ORTEP).

(17) IUPAC–IUB Commission on Biochemical Nomenclature, *Biochemistry*, **9**, 3471 (1970).

(18) G. N. Ramachandran, G. Ramakrishnan, and V. Sasisekheran, *J. Mol. Biol.*, 7, 95 (1963).

screw axis are nearly prependicular with a dihedral angle of 101.6°, and thus the hydrophobic region exhibits typical herringbone packing. There are four hydrogen bonds in the hydrophilic region, providing full employment for the hydrogen atoms attached to N-1 and N-3, and comprising three hydrogen-bonding schemes. Two of these schemes are single head-to-tail links. One, N-1–H-1 → N-2, connects screw-related molecules. The other, N-3–H-9 → O-2, is between a pair related by lattice translation one unit cell diagonally along *a* and *b*. The third scheme is an infinite zig-zag chain ... → O-1 ← H-2–N-1–H-3) → O-1 ← ... along the *a* axis, between molecules screw related tail-to-tail. (Distances and angles are given in Table II.) The structure is thus held together in the *x* direction by the ribbon-like hydrogen bonding scheme, in the *z* direction by the N–N hydrogen bond and the hydrophobic interaction, and in the *y* direction by the N-3–O-2 hydrogen bond.

The standard deviations of the bond lengths are too great for this structure analysis to furnish evidence relevant to the postulate of Mighell and Riemann[19] that the longer C–N bond within the pyrazole ring should be the one to the protonated nitrogen.

Biological Implications

The hydrogen bond between N-3 and O-2 suggests that a similar bond could be formed between N-3 and a phosphate oxygen in the ribonuclease enzyme–substrate system. However, contrary to the stituation with histidine, donation of this hydrogen atom catalytically is not feasible at physiological pH, since N-2 is unlikely ($pK \approx 2$)[5] to be protonated. Hence, although binding of substrate could occur, catalysis and cleavage would be prohibited.

A comparison[20] of the structure of Pyr(3)Ala with a series of histidine crystal structures (Table V) demonstrates that, for most chemical purposes, Pyr(3)Ala is isosteric with histidine. However, Pyr(4)Ala (III) could possibly be a better substitute for histidine as a biochemical probe. Use of this isomer would avoid the insertion of a hydrophobic region in a place where it is not present in nature. This is due to the fact that if the ring of Pyr(4)Ala were divided along the C-3, C-4 vector (bond a, III), each side of the ring would have both a nitrogen and a carbon, and thus there would be no extensive hydrophobic region. Therefore, any potential hindrance of binding caused by the introduction of the hydrophobic region of Pyr(3)Ala would probably be avoided by substituting Pyr(4)Ala.

Acknowledgments. The authors wish to thank Mrs. Joan Klinger, Mr. Anthony Boczon, Mrs. Hiroko Takagi, and Mrs. Linda Wasserman for technical assistance. We are grateful to Dr. Robert Stewart of Carnegie Mellon University and Dr. Klaus Hofmann for useful discussions about the structure. We also wish to thank Dr. John Madden for the use of Table V. Use of the computing facilities of the University of Pittsburgh Computer Center is gratefully acknowledged.

(19) A. D. Mighell and C. W. Reimann, *J. Phys. Chem.*, 71, 2375 (1967).

(20) J. Madden, E. L. McGandy, N. C. Seeman, M. Harding, and A. Hoy, *Acta Crystallogr.*, submitted for publication.

Synthetic Peptides Related to Horse Heart Cytochrome *c*. VII. Synthesis and Inhibitory Properties of the 70–80 Undecapeptide[1]

Y. Wolman, A. Schejter, and M. Sokolovsky*

Contribution from the Department of Organic Chemistry, The Hebrew University of Jerusalem, and the Department of Biochemistry, Tel-Aviv University, Israel. *Received July 3, 1971.*

Abstract: The undecapeptide corresponding to the amino acid sequence 70–80 of cytochrome *c* was synthesized. The synthetic peptide inhibited the cytochrome oxidase mediated oxidation of ferrocytochrome *c*. The inhibition was not due to the overall positive electrostatic charge of the peptide, since the $N^{\alpha,\epsilon}$-acylated peptide was an even better inhibitor. No inhibition was observed when smaller peptides of the 70–80 sequence were tested.

Examination of the similarities and differences among the amino acid sequences of eukaryotic cytochrome *c* has led to tentative conclusions concerning structure–function relationships; the most remarkable constant segment of the polypeptide chain is that extending from residue 70 to residue 80, Asn-Pro-Lys-Lys-Tyr-Ile-Pro-Gly-Thr-Lys-Met, as contrasted to the next longest invariant segment which is no longer than two residues.[2] Experiments involving chemical modification and X-ray crystallographic analysis of horse ferricytochrome *c*[3] indicate that the area of residue 70–80 may represent a surface at which cytochrome oxidase binds to cytochrome *c*.

In order to explore this possibility, we synthesized the undecapeptide corresponding to the sequence 70–80 for cytochrome *c* as summarized in Figure 1. The two fragments VIII and XVI were synthesized in the stepwise elongation procedure from the respective C-terminal amino acid, and the *tert*-butyloxycarbonyl group was removed after each step of the elongation by using anhydrous trifluoroacetic acid.[4] The dipeptide X was

* Address correspondence to this author at Tel-Aviv University.

(1) Part VI: Y. Wolman and Y. S. Klausner, *Israel J. Chem.*, **9**, 229 (1971).

(2) C. Nolan and E. Margoliash, *Annu. Rev. Biochem.* 37, 727 (1968).

(3) R. E. Dickerson, T. Takano, D. Eisenberg, O. B. Kallai, L. Samson, A. Cooper, and E. Margoliash, *J. Biol. Chem.*, **246**, 1511 (1971).