INSULIN MOLECULES: THE EXTENT OF OUR KNOWLEDGE

DOROTHY CROWFOOT HODGKIN

Chemical Crystallography Laboratory, South Parks Road, Oxford, U.K.

ABSTRACT

The recent calculation of an electron density map for rhombohedral 2 Zn insulin provides us with information on the whereabouts of all the atoms in the molecules as they are arranged in these crystals. Within the precision of atom placing, still rather low, it is possible to see the way in which the insulin molecules aggregate to form dimers and hexamers, and that within the two molecules of the dimer, the atomic arrangement is not exactly the same. Various ideas about the structure and aggregation of insulin and other large molecules can be examined in the light of our findings.

It gives one a curious feeling to listen to many discussions on the structure and behaviour of macromolecules and polyelectrolytes, viewed as objects of various general shapes, with surface charges and solvent atmospheres of estimated density and thickness—and then to think of the evidence we now have about the detailed structure of insulin and its surroundings in crystals, which transforms the way in which we can describe the characteristics of this particular medium sized polyelectrolyte.

The evidence we have is derived from the study by x-ray diffraction of one particular modification of insulin crystals[1]. Crystals of this variety are rhombohedral and form in the presence of zinc; they were first obtained by J. J. Abel in 1925 and were shown to contain zinc by D. A. Scott. Much later experiments by Schlichtkrull, seeking to prepare very uniformly grown insulin crystals for clinical use, showed that the minimum amount of zinc necessary for crystallization corresponded to two zinc atoms per rhombohedral unit cell. Each unit cell was found by preliminary x-ray studies to contain a protein molecular weight of about 36000, which corresponds to six insulin molecules of the weight defined by Sanger's work on the chemical structure of the molecule[2]. In addition, there is some 30 per cent of solvent of crystallization present, which is largely, if not entirely, water. Into this water, it is possible to diffuse a number of heavy atom-containing molecules and ions and so to form a series of isomorphous derivatives, measurements on which provided data for calculating the phase relations corresponding to the diffraction spectra obtained from the 2-zinc insulin crystals themselves. One of the most important of the heavy atom-containing derivatives was made a little differently from the rest. Zinc insulin crystals were left overnight in EDTA solution which removed the zinc from the crystals. The zinc-free

I. The primary structure of pig insulin.

crystals were then left overnight in 0.01 molar lead acetate solution. Lead ions enter at the sites originally occupied by zinc and also at sites between the zinc ions and at the edge of the molecule. From calculations on the intensities of the diffraction spectra from the lead- and zinc-free crystals it is possible to place the lead ions rather precisely and the relations found lead easily to the finding of the other heavy atom sites.

The combined calculations on five different isomorphous insulin derivatives have led so far to the calculation of a three-dimensional electron density distribution from which the arrangement of the atoms in the crystals can be derived. The resolution achieved—2.8Å—does not permit atoms to be seen

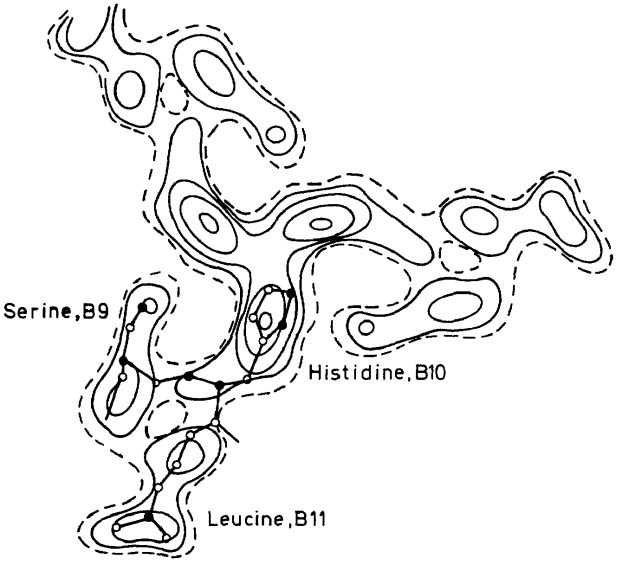

Figure 1. Section of the electron density map in the region of histidine B10; filled circles mark the atomic positions nearest the section.

as individual peaks, but, in general, the shapes of the peaks and strands of electron density present recognizable views of the different amino acid residues and peptide chains in the sequence found by Sanger, I. The interpretation of the map began with the strong peaks defining the zinc ion positions. These were surrounded by three peaks of the general form expected for histidine residues attached to the metal; the correlation of neighbouring peaks with serine B9 and leucine B11, as shown in *Figure 1*, defined the histidine residue attached to zinc as B10 and led rapidly to the placing of the rest of the B chains. The solution was helped by the fact that these chains proved to consist essentially of three regions of well defined normal geometry, an initial extended stretch followed by ten residues in an α-helix, followed again by an extended region. The A chain, on the other hand, pursues a very involved course which proved more difficult to define. That it forms a small compact loop, resting on the B chain α-helix, was easy to see but our first ideas about the arrangement of the long residues between A12 and A19 have had to be modified with more careful study of the map during this last

year[3]. The problem is illustrated by *Figure 2* which shows the alternative ways in which the atoms have been assigned to fit the electron density pattern. In general the new arrangement (b) fits the electron density better than (a). The one peak which is left unoccupied by any density in (b) is of similar weight to many peaks which seem clearly due to water molecules, hydrogen bonded to negative atoms on the molecular surface. We therefore consider it here to be due to water, hydrogen bonded to the peptide carboxyl group.

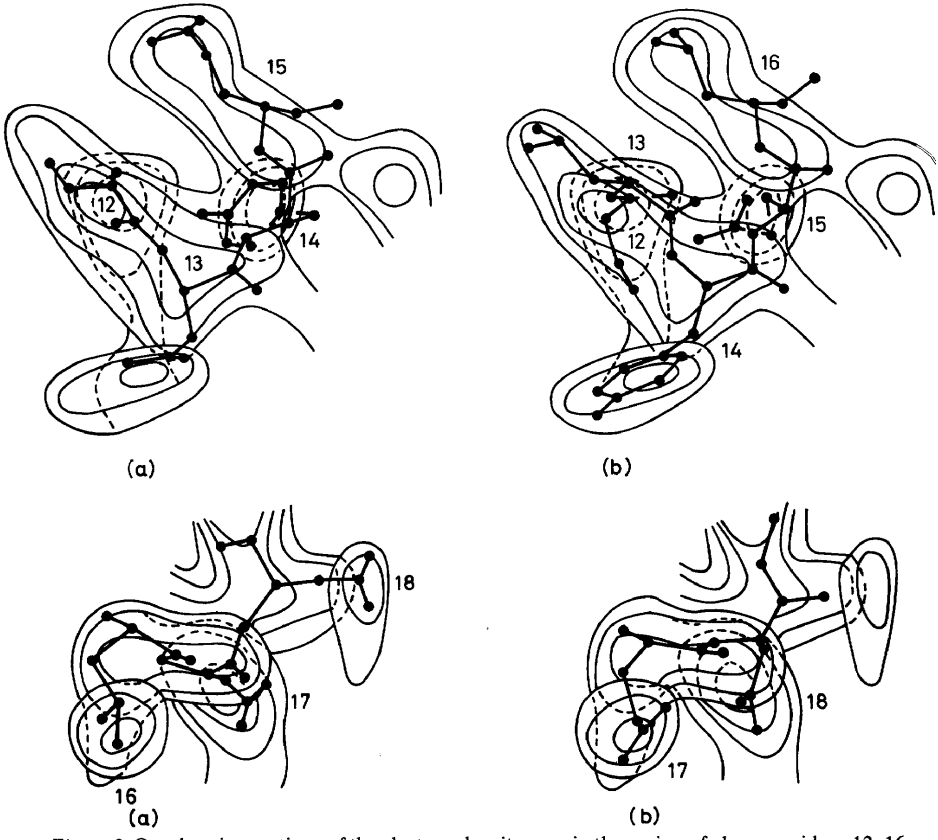

Figure 2. Overlapping sections of the electron density map in the region of above, residues 12–16, below, residues 17–19; (a) first interpretation, (b) second interpretation.

The improved interpretation of the electron density map followed from the building of a skeletal model at a scale of 2 cm to 1Å and projecting this into the three-dimensional electron density map, drawn at the same scale on transparent sheets, by the mirror device set up in Oxford by Professor F. R. Richards. It was clear that some of the geometry prescribed by the model, e.g. tetrahedral angles at the α-carbon atoms of the peptide chain, should be modified to give a better fit with the observed electron density. In other regions, including the disulphide bonds, there is a good deal of freedom

in fitting the model to the map, and it is very likely that the atomic positions we have now recorded should be modified in various details. Calculations designed to improve the molecular geometry or the potential energy between different groups suggest small changes such as those shown in *Figure 3* which are certainly within the limits of our present accuracy. It ought to be possible to improve our definition of atomic positions by extending the spacing limit of x-ray data collected but that may take some years to achieve.

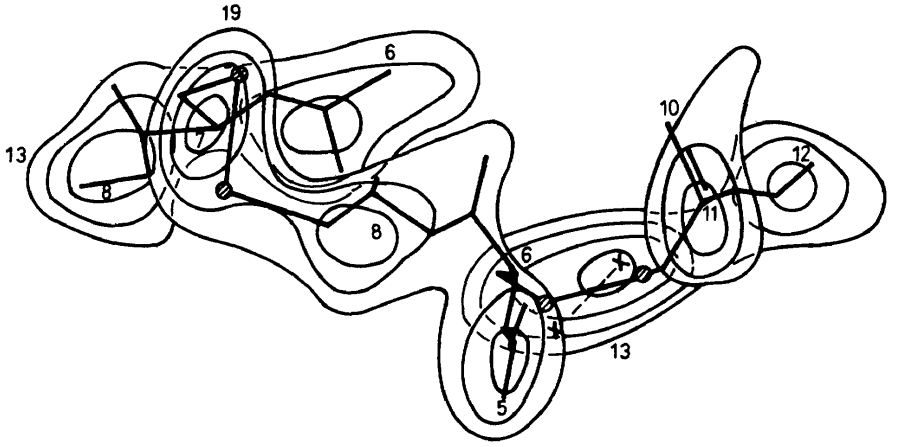

Figure 3. Overlapping sections of the electron density map in the region of the disulphide bonds B7—A7 and A6—A11. The positions marked by circles are those of our present model; those by crosses are given by Dr Michael Levitt's refinement using potential energy functions.

In the meantime we have a three-dimensional view of the arrangement of the atoms in two independent molecules of insulin—since the unit cell contains six molecules and the formal crystal symmetry only requires three. These two insulin molecules are very much alike but not identical; there are small differences between them, particularly in details of the sidechain conformations, which are apparently controlled by packing relations. Each individiaual molecule presents a somewhat irregular shape; some 20 Å across and deep, as seen in *Figure 4* and 28 Å from top to bottom; the initial residues of the B chain hang loose in our crystals but may well adopt another arrangement in another situation. The core, however, is very compact, controlled by contacts between non-polar residues in the centre of the molecule. As drawn in *Figure 4*, the front and back surfaces, top and bottom of the molecule, are studded with polar groups, roughly 5 to 6 Å apart. One can define the exact position of the charged groups relative to one another: there are occasionally suggestively close contacts between positive and negative groups as between arginine, B22 and the A chain carboxyl terminal, asparagine. The two sides of the insulin molecule, on the other hand, include many non-polar residues, phenyl alanine, valine and leucine, for example, and these are evidently involved in the aggregation of the molecules, both in solution and in the crystal to form dimers and hexamers.

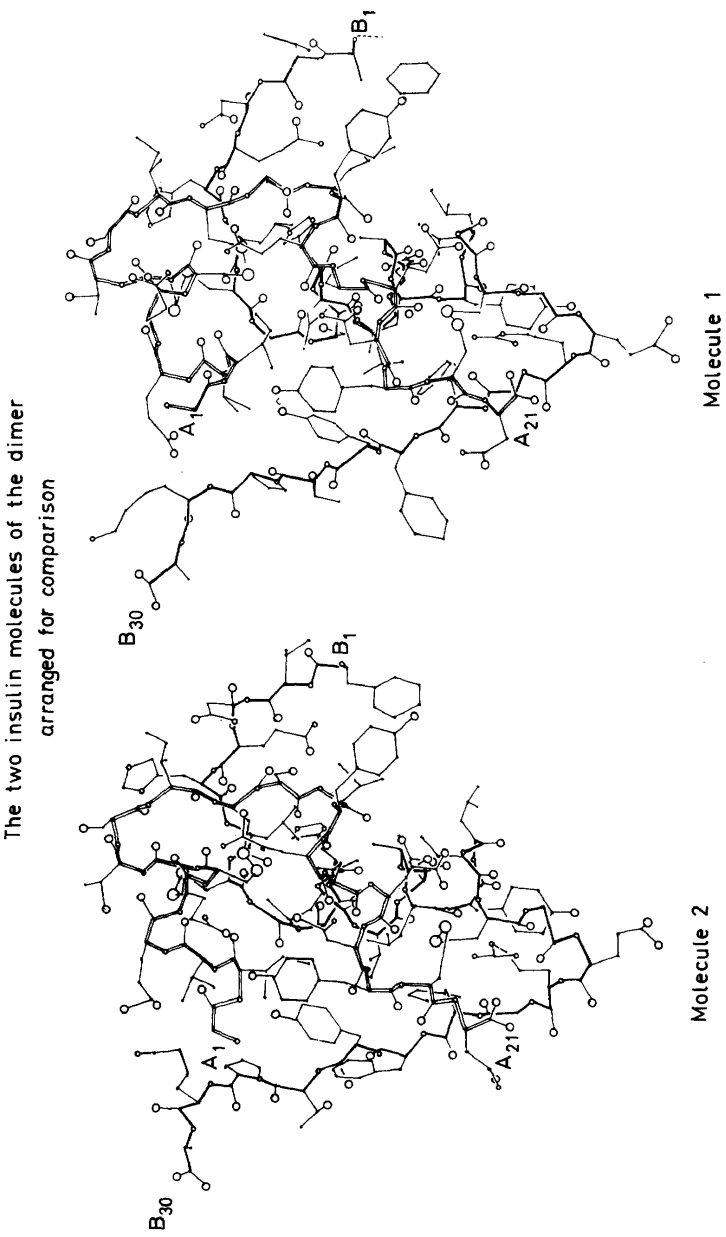

Figure 4. Two insulin molecules seen projected perpendicular to the *c* axis.

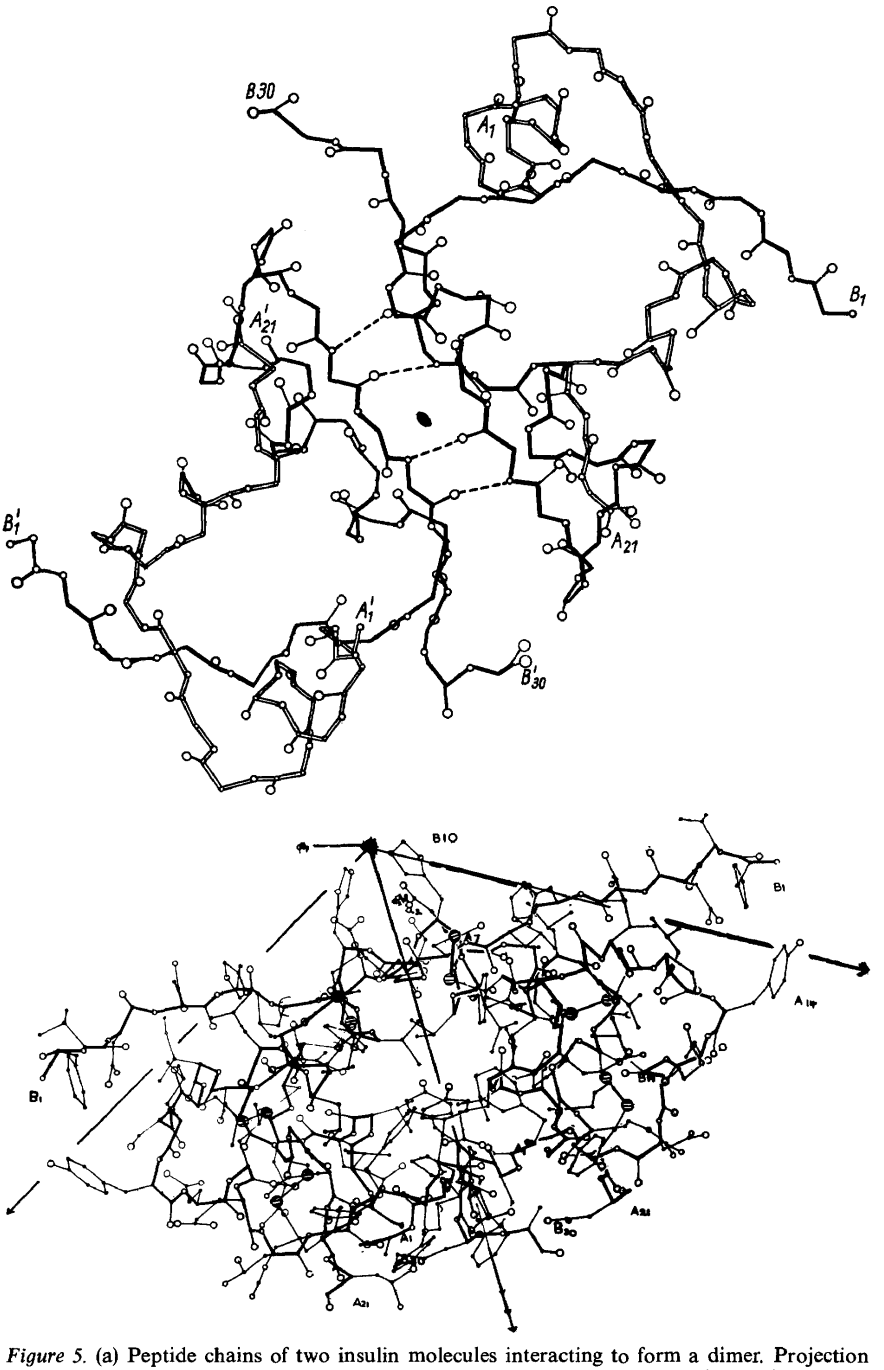

Figure 5. (a) Peptide chains of two insulin molecules interacting to form a dimer. Projection along the pseudo twofold symmetry axis. Hydrogen bonds dotted. (b) Atomic positions in the insulin dimer projected along the threefold axis.

The most extended pattern of contacts between two insulin molecules involves the last eight residues of the B chains. In this region the two molecules are approximately related to one another by a twofold axis which runs perpendicular to the threefold axis of the crystals. Within the contact region of the two terminal B chains, the peptide groups run antiparallel and are hydrogen bonded together in an antiparallel pleated sheet [*Figure* 5(a)]. Both polar and non-polar residues also make contact, like to like. Occasionally this has the effect of destroying the twofold symmetry as in the contact between the B25 phenyl alanine residues [*Figure* 5(b)].

The insulin dimer that results from these interactions is a long ellipsoidal object, not unlike lysozyme in overall dimensions. In the crystal, three such dimers closely pack around two zinc ions to form a hexamer, capable of persisting in solution. Again the contacts between them are very close and include both polar and non-polar interactions, as between the two tyrosine residues, A14, or the two phenyl alanine groups, B1. The hexamer (*Figure 6*) is smoothly spheroidal around the circumference but presents a top and bottom surface from which the three residues, A8, 9 and 10, project, leaving deep grooves between them. These grooves, in the crystal, accommodate the projecting A chain residues of the succeeding molecules, packed along the *c* axis. And again both polar and non-polar residues are involved in the contact.

Figure 6. Model of insulin hexane based on an electron density map at 6 Å resolution, seen along the *c* axis.

The crystal structure, as a whole, is represented by *Figure 7* which shows the atomic positions found in four molecules projected along the *c* axis. The hexamers present an appearance of close-packed spheres; the contacts they make side by side are again close and involve both non-polar and polar groups, often bridged by water molecules. Around the threefold screw axes, which run between the hexamers, are channels, nearly 10 Å across, filled with water, which extend from one side of the crystal to the other. Narrower channels thread their way between the molecules within the hexamers,

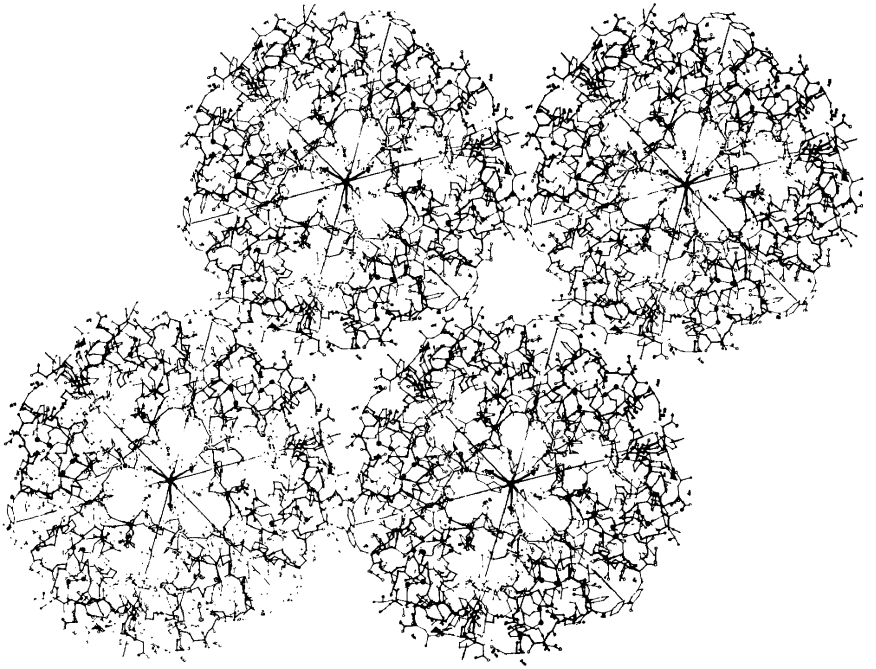

Figure 7. Atomic positions in four insulin hexamers in contact as in the crystal projected along threefold axis.

permitting ions to penetrate through to the centre of the hexamer between the two zinc ions along the threefold axis. In the wider channels, there is a background of low electron density peaks which may present the more probable positions of moving water molecules. At the edges of the molecules, both within and outside the hexamer, are many much more definite peaks which seem almost certainly to indicate the precise positions of hydrogen-bonded water molecules. It is particularly noticeable that there is a trail of electron density running from the hydroxyl group position of every tyrosine in the molecule. Most of these seem to represent two to three water molecules bridging specific gaps, e.g. between the B16 hydroxyl group of one molecule and B9 serine of the second within the insulin dimer or the A19 tyrosine and A1 glycine within one molecule as in *Figure 8*. At 2.8 Å resolution these electron density trails are not resolved to give individual water molecule

positions. It will be very interesting to see how they develop when we have data extending to higher resolution.

There has not yet been time to evaluate properly the many different observations which have been made on insulin in the light of this structure—varied estimates of helix content, for example, and of the differential chemical reactivity of many of the residues. There is certainly a great deal more to be discovered from such an evaluation about the behaviour of the molecule as a

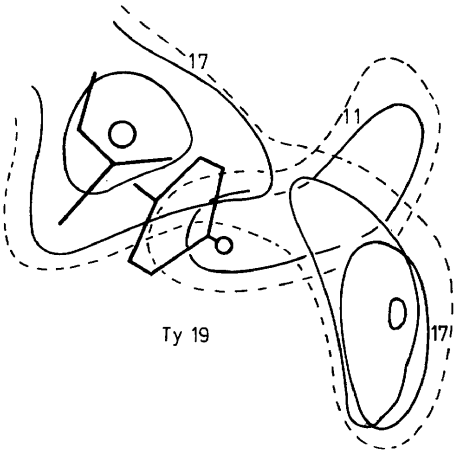

Figure 8. Overlapping sections of electron density map showing electron density streaks in the neighbourhood of A19.

whole, and particularly about its biological activity. It seems extraordinary to know as much about insulin as we now know—the whereabouts of all the atoms in the molecule in space—and still to know so little about what it actually does that makes it so important in our lives.

NOTE: In the above figures, the atomic positions shown are the first full set obtained by the process of matching the model against the map described in the tect. They are in the process of being improved and regularized.

REFERENCES

[1] M. J. Adams, T. L. Blundell, E. J. Dodson, G. G. Dodson, M. Vijayan, E. N. Baker, M. M. Harding, D. C. Hodgkin, B. Rimmer and S. Sheat, *Nature, London*, **224**, 491 (1969).
[2] A. P. Ryle, F. Sanger, L. F. Smith and R. Kitai, *Biochem. J.* **60**, 541 (1955).
[3] T. L. Blundell, E. Dodson, G. G. Dodson, D. C. Hodgkin and M. Vijayan, *Recent Progress in Hormone Research*, **27**, No. 1, in press (1971).