CHYMOTRYPSINOGEN: A THREE-DIMENSIONAL FOURIER SYNTHESIS AT 5 Å RESOLUTION*

By J. Kraut, L. C. Sieker, D. F. High, and S. T. Freer†

DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF WASHINGTON, SEATTLE

Communicated by Bruno H. Zimm, June 18, 1962

Chymotrypsinogen A is a pancreatic protein of molecular weight 25,000. Although not itself enzymatically active, it is converted to the enzyme π-chymotrypsin when an arginyl-isoleucine bond within the molecule is hydrolyzed by trypsin. The zymogen consists of a single polypeptide chain cross-linked at five points by S—S bonds between half-cystines, one of which constitutes the N-terminal residue. Present knowledge of the amino acid sequence, mode of activation, and other aspects of the chemistry of chymotrypsinogen has been summarized in a recent review by Desnuelle and Rovery.[1]

An investigation of the three-dimensional structure of bovine chymotrypsinogen A was begun at this laboratory in 1957 as the first stage of a long-range program in which it was hoped eventually to elucidate the structures of both enzyme and precursor. The ultimate goal was to explain the physical basis of the conversion process, and perhaps, of chymotrypsin activity itself.

The only possible means, at present, of obtaining the desired structural information is X-ray diffraction crystallography, utilizing the technique of isomorphous replacement.[2–8] The first requirement, therefore, was to prepare single crystals of chymotrypsinogen containing heavy atoms and isomorphous with parent crystals of the native protein.

The Search for Isomorphous Replacements.—Approximately stoichiometric amounts of heavy-atom reagent were added to small dishes containing about 1 ml of mother liquor and protein crystals, and several days were allowed for diffusion. Precession photographs of the soaked crystals were then examined for changes of relative intensity among the reflections. The space groups and unit-cell parameters of the four previously known crystalline forms of chymotrypsinogen had been reported by Bluhm and Kendrew[9] and it was apparent from their data that only crystal-types B and D were suitable for study. However, since we have been unable to grow type B crystals, almost all our attention until quite recently has been concentrated on type D.

Since 1957, approximately 200 heavy-atom compounds were tested in this way for their ability to complex with the protein, and about 25 gm of highly purified chymotrypsinogen was used up. Only for the chloroplatinate, $PtCl_6^{--}$, however, was there satisfactory evidence of adsorption at specific sites in type D crystals. Unfortunately, with only a single isomorphous replacement available, the general hkl phases could not be determined. Furthermore, even approximate methods such as use of the "centroid" phases[10, 11] in Fourier syntheses, or superposition of difference-Pattersons[12] were of no use since the principal heavy-atom site appeared to be in a special position (0.19, 0.25, 0.00). That is to say, the four symmetry-related principal sites within a unit cell happened to be centrosymmetrically distributed, and therefore any such procedure would inevitably have led to a double image containing both the protein molecule and its superimposed enantiomorph.

Meanwhile, Tan and Wilcox,[13] taking the alternate approach of chemical modification of the protein, had prepared large single crystals of the CS_2 derivative of chymotrypsinogen in which the α-amino group had been converted to the dithiocarbamate group. These crystals however were found to be a new type ($P2_12_12_1$, $a = 47.1$, $b = 55.0$, and $c = 87.8$ Å; 4 molecules per unit cell) and therefore of no potential use in forming isomorphous derivatives of type D. Nor did the CS_2-crystals produce derivatives isomorphous with themselves when diffused with heavy-atom compounds.

Finally, in the course of continued fruitless efforts to grow type B crystals, two new types appeared quite by chance. These have been designated E and F. Type E ($P4_1$, $a = 73.4$ and c approximately 57 Å; 4 molecules per unit cell) grew slowly at pH 5 from a preparation which was subsequently found to contain a trace of trypsin as an impurity. The crystals proved to be composed of *neo*-chymotrypsinogen with an amino-terminal threonine. Type F ($P2_12_12_1$, $a = 52.0$, $b = 63.9$, and $c = 77.1$ Å; 4 molecules per unit cell) grew overnight from 10 per cent ethanol at pH 6 and (unlike type E) appeared to be composed of undegraded chymotrypsinogen, as judged by chromatography on carboxymethylcellulose and amino-terminal end group analysis. The rest of the work described in this communication was done with type F crystals.

Four heavy-atom reagents were soon found to produce useful isomorphous derivatives of type F: $PtCl_6^{--}$, $UO_2^{++}(P_2O_7^{-x})_n$, HgI_4^{--}, and $IrCl_3$. The nature of the species which actually bind to the protein is not yet known. The two pairs $PtCl_6^{--} + HgI_4^{--}$ and $UO_2^{++}(P_2O_7^{-x})_n + HgI_4^{--}$ were also used, giving a total of six different sets of isomorphous replacement data upon which the final 5 Å phases were ultimately based. It is estimated that the unit-cell parameters had not changed by more than about 0.2 per cent from the native to the derivative crystals.

It is noteworthy that the volume available per molecule in type F is 64,000 Å^3, as compared with 53,000 Å^3 in type D. This represents an increase of about 50 per cent in the space accessible to small molecules diffusing into type F crystals and may in part account for the relative ease with which they produce isomorphous derivatives.

Data Collection.—Integrated intensities were measured with a General Electric XRD-5 diffractometer using a 2θ-scanning technique. The instrument was equipped with an automatic shutter to reduce unnecessary X-ray exposure. Data were collected out to a Bragg spacing of 5 Å, which includes 1,262 independent reflections for the native crystal and for each derivative. Six standard reflections were monitored every few hours and the crystal was replaced when these showed intensity changes of about 10 per cent, presumably resulting from deterioration under the influence of X-rays. In the typical case, three crystals were used to obtain a complete set of 5 Å data for a given derivative. Data from different crystals were placed on a common scale by comparison of the standard reflections. Absorption corrections have not been computed but have been minimized by the use of small (0.2 mm) and approximately isometric crystals.

Scale Factors, Occupancies, and Temperature Factors.—Reflection amplitudes F_H for a given heavy-atom derivative were put on the same scale as the F of the native protein by applying the scale factor[14]

$$k = \frac{1}{2}\frac{\sum FF_H}{\sum F_H^2}\left\{1 - \left[1 - \frac{3}{4}\frac{\sum F^2 \cdot \sum F_H^2}{(\sum FF_H)^2}\right]^{1/2}\right\}.$$

The summations are over all 1,262 reflection indices. Apparently scale factors are satisfactorily approximated by this expression since they differed by less than 4 per cent from the final scale factors resulting from least-squares refinement. Occupancies, or the effective number of electrons added at the heavy-atom sites, and their apparent temperature factors were estimated by plotting log $\langle(kF_H - F)^2\rangle$ versus $(\sin\theta/\lambda)^2$ in the manner of the customary type of Wilson plot. Occupancies were obtained from the intercepts at $(\sin\theta/\lambda)^2 = 0$ and apparent temperature factors (based on the assumption of point atoms at $T = 0$) from the slopes. These initially-estimated occupancies proved to be as much as 40 per cent high in some cases, probably owing to the effects of random errors in the intensity measurements. Apparent temperature factors were in the range $B = 100$ to 150 Å^2. The corresponding apparent temperature factor of the native protein itself was 80 Å^2.

Location of Heavy-Atom Sites.—Three-dimensional difference-Patterson syntheses were computed using $(kF_H - F)^2$ as coefficients for the platinum, uranyl, and mercury derivatives; the iridium derivative was not discovered until later. Although the Pattersons of the others were readily interpretable in terms of single major heavy-atom sites, the site locations for the platinum derivative were somewhat uncertain. The difficulty appears to have been due to the presence of a number of subsidiary sites and was easily overcome by the use of difference-Fourier syntheses. In these and in all subsequent difference-Fouriers, the coefficients were of the type $(kF_H - F)e^{i\phi}$, with ϕ being the current estimate of the parent phase. The procedure actually followed was this: the centric 0kl, h0l, and hk0 parent phases were assigned on the basis of the mercury data and were then used in computing the difference-Fourier *a*-, *b*-, and *c*-axis projections for the platinum and uranyl derivatives. The uranyl major site was plainly visible, establishing its location relative to the mercury site, but the platinum sites, although clearly evident, were sufficiently overlapped so that it was thought desirable to confirm their locations with three-dimensional maps. To this end, all 5 Å phases, including the noncentric, were then assigned on the basis of both the uranyl and mercury derivative data, and full three-dimensional difference-Fouriers were computed for all derivatives for which data had thus far been collected. This now included the mercury-uranyl double derivative and very satisfactorily confirmed the relative uranyl and mercury locations. Furthermore, the individual platinum sites and their approximate relative occupancies were easily distinguishable, as well as a few subsidiary sites of low occupancy for the other derivatives. Four sets of isomorphous replacement data were now usable for phase determination. The iridium and mixed mercury-platinum sites were located by difference-Fouriers at a subsequent intermediate stage of refinement when the required data became available, and these derivatives were included in the final series of refinement cycles.

Refinement.—Parameters adjusted in the refinement were coordinates and occupancies of heavy-atom sites, and one scale factor, k, for each derivative. Temperature factors for the heavy-atom sites were not adjusted since, when varied, they showed what were thought to be unreasonably large changes and standard deviations, probably owing to the short range of sin θ/λ and interaction with the

occupancies. At present, only the fixed over-all isotropic temperature factors obtained from the Wilson plots have been applied.

The residual minimized in the least-squares refinement was

$$\sum[kF_H(\text{obs}) - F_H(\text{calc})]^2 = \sum[kF_H(\text{obs}) - |Fe^{i\phi} + \mathbf{f}_H|]^2,$$

where $\mathbf{f}_H$ is the calculated contribution of the heavy-atom sites. The summation is understood to include all reflection indices, hkl, for all heavy-atom derivatives. The program cycle consisted of a stage of full-matrix least-squares adjustment of parameters alternating with a stage of parent phase (ϕ) redetermination. In the latter step, the partial sums that would be contributed to the above residual by each hkl were obtained at 5° intervals of ϕ, and the most probable ϕ, i.e., that for which the partial sum is minimum, was assigned as the current estimate of the phase of that parent reflection.

Several subsidiary sites appeared to be present in the difference-Fouriers calculated at intermediate stages of the refinement and were introduced into the calculations with small occupancies. Some of these dropped below about 5 electrons and were eliminated; others remained or even showed increases in their occupancies.

Although intensity data had been collected for two additional heavy-atom derivatives, the p-chloromercuribenzene sulfonate and the mixed platinum-uranyl, these failed to refine adequately and were not used in the final series of calculations.

Refinement was continued until no appreciable changes in any parameter could be obtained. In the final cycle, the centroid phase and figure of merit, m, of Blow and Crick[15] were also calculated for each reflection. In all, the equivalent of about a dozen full-matrix least-squares cycles have been run.

Some over-all statistics from the last refinement cycle are:

$$R = \frac{\sum|kF_H(\text{obs}) - F_H(\text{calc})|}{\sum kF_H(\text{obs})} = 10.4\%,$$

with a range of 9.2–12.7% for the six derivatives taken separately;

$$\text{mean } m = 0.81;$$

$$\text{mean } |\phi \text{ (centroid)} - \phi \text{ (most probable)}| = 15.8°;$$

$$\text{r.m.s. } |kF_H(\text{obs}) - F_H(\text{calc})| = 42 \text{ electrons},$$

with a range of 36–53 electrons;

$$\text{r.m.s. } |\mathbf{f}_H| = 81 \text{ electrons},$$

with a range of 65–102 electrons.

Absolute Intensity Calibration.—The intensities of the parent reflections were placed on an absolute scale by calibration of the diffractometer with twelve selected reflections from a crystal of *myo*-inositol. The calculated structure factors were obtained from the structure refinement of Rabinowitz and Kraut.[16] A further check of the calibration was made with the 004 and 006 reflections of NaCl. Crystal volumes were calculated from carefully measured dimensions and allowance was made for absorption effects. The absorption correction, in fact, introduces the most serious source of uncertainty in the calibration procedure, which probably

amounts to about 10 per cent in terms of electron density. Knowledge of the absolute scale of intensities is important for comparing absolute electron densities of various features in different protein molecules and for estimating the extent of substitution at heavy-atom sites.

Results and Discussion.—Three-dimensional Fourier syntheses of the native protein were calculated, using the most probable phases in one case and the centroid phases with weighting factors m in the second. The two maps differed only in minor details, and the essentially arbitrary decision was made to work with the most probable Fourier at present, partly because its electron density values have a more straightforward significance. The grid used was 48*ths* by 48*ths* by 60*ths*, corresponding to a 1.3 Å grid spacing.

The first problem in the interpretation of such maps is the isolation of a single protein molecule. At this point, the absolute electron density scale had yet to be established, so it was necessary to select by trial and error a contour level which was sufficiently low to give a more or less continuous structure to the molecule, yet high enough to allow neighboring molecules to be separated; two narrow bridges of electron density between molecules were easily eliminated. To check on the result, another map was prepared, contoured at the same level, but consisting of sections of constant y instead of z. The same molecule was independently isolated in this by one of us who had not participated in the work on the first map. A further check was the fact that the portion of the Fourier map assigned to a single molecule was afterward found to constitute a complete asymmetric unit, with no important pieces of high electron density omitted and no piece being included in more than one molecule simultaneously.

Other observations which support the correctness of our result are that it agrees with the over-all molecular shape arrived at earlier by Kraut,[17] that all heavy-atom sites which were not wiped out by least-squares refinement are external to the molecule, and finally that the difference-Fouriers of the heavy atoms are well defined and mutually consistent.

Figures 1 and 2 are photographs of balsa-wood models of the chymotrypsinogen molecule at its present stage of refinement. They were constructed by cutting out and joining pieces corresponding to contours within a single molecule on each z-section. The sections are 1.3 Å thick. The contour level in Figure 1 corresponds to 0.12 electron $Å^{-3}$ and is the level which best depicts a single continuous structure. It is probably no coincidence that this electron density agrees closely with the contour level of 0.14 electron $Å^{-3}$ used to outline the hemoglobin molecule.[8] In Figure 2, the contour level corresponds to 0.19 electron $Å^{-3}$; this model was built in order to display the chain segments more clearly.

It should be kept in mind that the structure shown here may be the mirror image of the true protein structure, a matter which will probably be settled in the course of further work.

The molecule is approximately an ellipsoid with axes 50, 40, and 40 Å. A noteworthy feature of the over-all molecular shape is the presence of a distinct hollow (visible at the bottom of Figs. 1 and 2) in the otherwise ellipsoidal outline. This is in agreement with results obtained earlier from calculations based on the very-low-order reflections from type D crystals,[17] although in that case the hollow was somewhat exaggerated, probably owing to the choice of contour level.

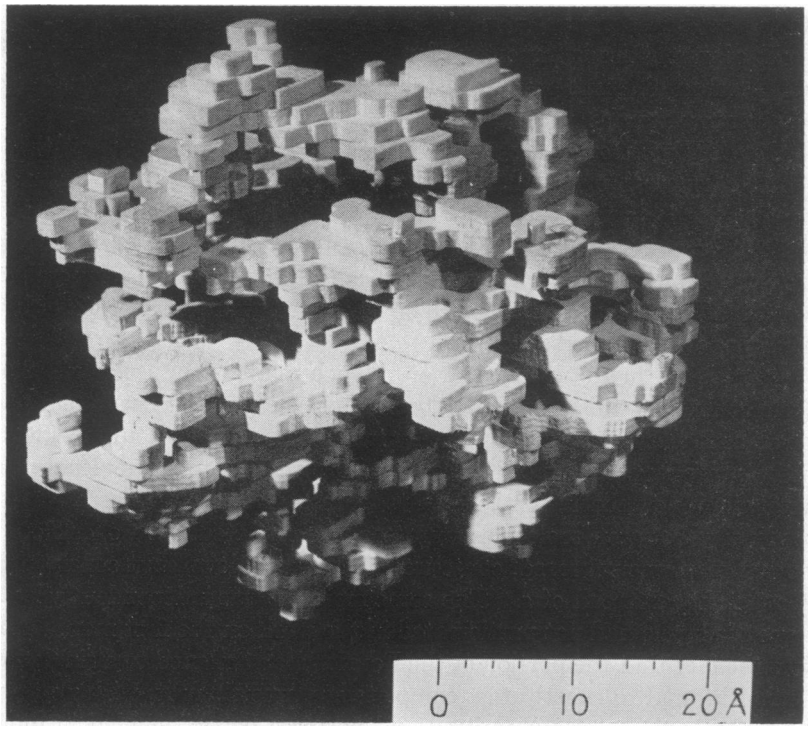

FIG. 1.

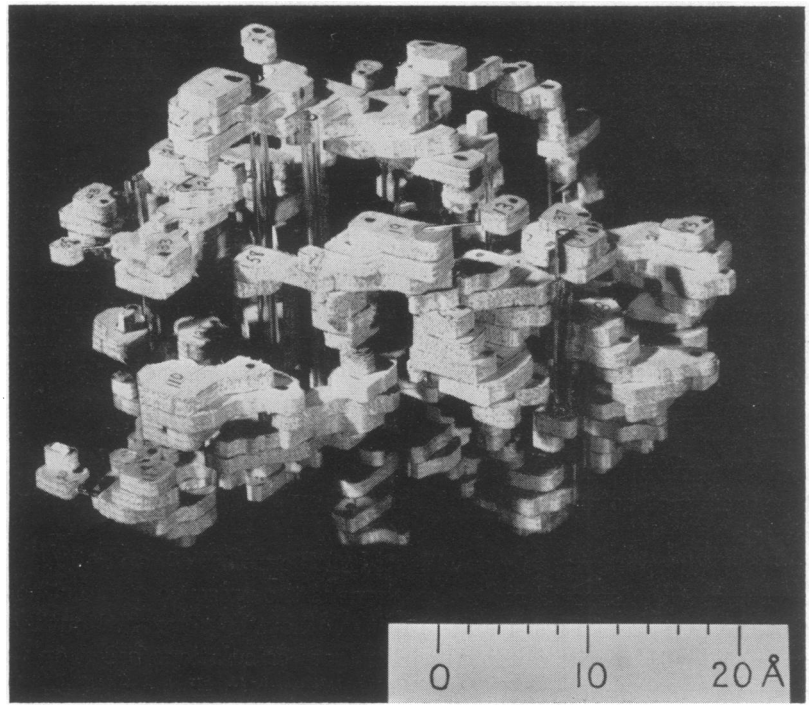

FIG. 2.

The most striking aspect of the molecular structure is its great complexity. There is only one segment, on the outside of the molecule (visible at the top of Figs. 1 and 2), which seems to be α-helix, or at least which appears to be reasonably straight and to have the thickness and high electron density characteristic of the α-helical portions of hemoglobin and myoglobin. This segment extends only 15 Å, or enough for about three turns. The rest of the molecule seems to be almost all bends, and probably for this reason it is not yet possible to trace a single backbone chain continuously with any degree of assurance. There is the further complication, not present in hemoglobin or myoglobin, of the existence of five disulfide cross-links which would probably be difficult to distinguish from main chain at this resolution. It is possible, however, to measure very approximately a total apparent chain-length within the molecule of 650 Å. Given some 243 residues in the molecule and making the extreme assumption that it is composed entirely of either α-helix or completely extended chain at any point, one may calculate that the maximum possible fraction of residues in the α-helical regions must be about 30–40 per cent, corresponding to about 110–150 Å in length. Of course, the amount of α-helix actually present is probably much less than this.

Our general conclusion, then, is that detailed knowledge of the polypeptide-chain folding must await further work at higher resolution but that the over-all appearance of the molecule as it now stands is probably correct.

The nature of the binding sites for the heavy atoms is not yet known. All are found outside the protein molecule. There appear to be two neighboring locations within the asymmetric unit, separated by about 6.5 Å, with a strong tendency to bind heavy-metal groups. One is at the uranyl major site, which also serves as platinum subsidiary site 4 and iridium subsidiary site 2, and the other as at the mercury major site, which in turn also corresponds to the major iridium site and to platinum subsidary site 2. There are in addition four more widely distributed locations which constitute other sites in various derivatives.

The number of heavy-atom sites included in the final series of refinement cycles were 5 for platinum, 2 for uranyl, 2 for mercury, 3 for iridium, 3 for the mixed uranyl-mercury, and 2 for the mixed platinum-mercury. Occupancies ranged from 92 electrons for the mercury major site down to 5 electrons for the least important platinum site. If allowance is made for the electrons of displaced water molecules, it must be concluded that the major heavy-atom sites were nearly saturated.

Further work on the crystal structure of the type F modification of chymotrypsinogen is in progress, in which it is planned, as the next step, to increase the resolution to 4 Å. This involves approximately doubling the number of terms in the Fourier synthesis and ought to clarify some of the details of the chain conformation.

Summary.—Six heavy-atom-containing isomorphous derivatives of a new crystalline form of bovine chymotrypsinogen A have been prepared. On the basis of intensity changes observed in the X-ray diffraction patterns from these derivatives, phases of the parent reflections have been obtained out to a Bragg spacing of 5 Å. Three-dimensional Fourier syntheses have been computed which reveal a molecule of approximately ellipsoidal shape, with axes 50, 40, and 40 Å, but with an obvious hollow or depression. The chain conformation is complicated and appears to have little α-helix content in comparison with myoglobin and hemoglobin. It is not yet possible at this resolution to trace a continuous chain throughout the molecule.

We are deeply indebted to Philip E. Wilcox for his guidance in the methods of protein chemistry. The chymotrypsinogen used in this work was very kindly supplied gratis by Herman Cohen of the Princeton Laboratories, Inc., and carefully purified by Professor Wilcox and coworkers before the final crystallization.

We have also benefited from discussions with Lyle H. Jensen and especially from his consistently enthusiastic and optimistic encouragement. William Awad performed the N-terminal end-group analyses mentioned herein, and John Neal and Mary Esterberg operated the diffractometer.

* This project has been supported by continuing grants from the National Institutes of Health, by the Howard Hughes Medical Institute, and by grants of free computer time from the Research Computer Laboratory of the University of Washington.

† After October 1, 1962, the address of the authors will be: Department of Chemistry, University of California (San Diego), La Jolla.

[1] Desnuelle, P., and M. Rovery, *Advances in Protein Chemistry*, **16**, 139 (1961).

[2] Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, *Nature*, **181**, 662 (1958).

[3] Bodo, G., H. M. Dintzis, J. C. Kendrew, and H. Wyckoff, *Proc. Roy. Soc.* (*London*) *A*, **253**, 70 (1959).

[4] Perutz, M. F., M. G. Rossman, A. F. Cullis, H. Muirhead, G. Will, and A. C. T. North, *Nature*, **185**, 416 (1960).

[5] Kendrew, J. C., R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore, *Nature*, **185**, 422 (1960).

[6] Dickerson, R. E., J. C. Kendrew, and B. E. Strandberg, *Acta Cryst.*, **14**, 1188 (1961).

[7] Cullis, A. F., H. Muirhead, M. F. Perutz, M. G. Rossmann, and A. C. T. North, *Proc. Roy. Soc.* (*London*) *A*, **265**, 15 (1961).

[8] Cullis, A. F., H. Muirhead, M. F. Perutz, M. G. Rossmann, and A. C. T. North, *Proc. Roy. Soc.* (*London*) *A*, **265**, 161 (1961).

[9] Bluhm, M. M., and J. C. Kendrew, *Biochim. Biophys. Acta*, **20**, 562 (1956).

[10] Blow, D. M., and M. G. Rossmann, *Acta Cryst.*, **14**, 1195 (1961).

[11] Kartha, G., *Acta Cryst.*, **14**, 680 (1961).

[12] Kartha, G., and G. N. Ramachandran, *Acta Cryst.*, **8**, 195 (1955).

[13] Tan, W., and P. E. Wilcox, to be published.

[14] Kraut, J., unpublished.

[15] Blow, D. M., and F. H. C. Crick, *Acta Cryst.*, **12**, 794 (1959).

[16] Rabinowitz, I. N., and J. Kraut, *Abstracts: A. C. A. Annual Meeting*, paper K-8 (1962).

[17] Kraut, J., *Biochim. Biophys. Acta*, **30**, 265 (1958).

---

*BIOSYNTHESIS OF THE COAT PROTEIN OF COLIPHAGE f2 BY E. COLI EXTRACTS**

BY D. NATHANS,† G. NOTANI, J. H. SCHWARTZ, AND N. D. ZINDER

THE ROCKEFELLER INSTITUTE

Communicated by Fritz Lipmann, June 15, 1962

Nirenberg and Matthaei[1] have discovered an assay system in which RNA serves as an activator of protein synthesis in *E. coli* extracts. RNA fractions from cells,[1] synthetic polyribonucleotides,[2, 3] and viral RNA[1, 4] can all stimulate amino acid incorporation into acid-insoluble products in *E. coli* extracts. Although in each case the product formed is presumed to be a protein or polypeptide whose structure is uniquely determined by the RNA, the products have not as yet been completely