Reports

Genetic Code: Emergence of a Symmetrical Pattern

Abstract: *The triplets of nucleotides which apparently specify the detailed structure of proteins fall into a regular pattern: the 64 combinations of four nucleotides taken three at a time, are resolved into 32 pairs. The second member of each pair is identical with the first, except that in one position a purine is replaced by the other purine or a pyrimidine by the other pyrimidine. Almost all of the reported triplets fit into this pattern, and from it one can predict which amino acids will be found to correspond to the remaining 19 unidentified triplets. This pattern accounts for several of the observations concerning regularities in the data, partially determines the order of the nucleotides in each triplet, and suggests a structural basis for transfer RNA specificity.*

The mathematical-biochemical problem posed by Gamow 9 years ago (*1*) has been considerably clarified. Chromosomal DNA directs the synthesis of messenger RNA, which in turn directs the synthesis of protein, by way of a number of adapter molecules, called transfer RNA's. Some part of each transfer RNA must recognize specifically a region of the messenger RNA (*2*), which Crick calls a "codon." The elucidation of a codon pattern should provide evidence on the mechanism of transfer RNA specificity. How, structurally, does the specific transfer RNA recognize its proper codon combination? An answer to this question is suggested in this paper.

The discovery 2 years ago that a nonliving system could be made to synthesize an artificial protein, polyphenylalanine, with synthetic polyuridylic acid as messenger RNA (*3*), has made possible a rapidly developing experimental attack on this problem. The in vitro synthesis of proteins in a system containing various synthetic copolymers of two or more ribonucleotides as artificial messenger RNA has been reported from two laboratories. Tables have been published giving groups of three nucleotides ("triplets") which have been identified as "coding for" the various amino acids (*4*). There has been much interest and speculation in the number of codons and the possible relationships among them. Various patterns have been propounded which usually accommodate the triplets reported to that date, and predict certain others. The discovery of additional triplets has then required that many of these hypotheses be abandoned or much modified. The rate at which new triplets have been discovered makes it appear that nearly all, if not all $4 \times 4 \times 4$ mathematically possible triplets will ultimately be identified as codons. Will these appear to be a chaotic jumble, or will they fall into some regular pattern?

Some regularity can be seen in the list of published triplets. No more amino acids are usually assigned to a nucleotide combination than there can be alternative rearrangements of it. This supports the expectation that the maximum number will not exceed 64. There are only three exceptions to this, two of which may be the result of laboratory errors. The third exception will be considered later.

It frequently appears that two or more triplets coding for the same amino acid have two of their three nucleotides in common. Roberts has used this property to derive a "doublet" code, which implies that the third nucleotide is irrelevant (*5*). This gives only 16 combinations and requires some supplementary explanation to account for the 20 amino acids. It has also been suggested that the code may be partly doublet and partly triplet.

In the most recent lists from the two laboratories there are a total of 49 different triplets, 26 of them reported by both groups (*4*). This seems like a large enough proportion of the 64 to outline an overall pattern, if one exists. The data can be arranged in various ways, and if one tabulates these triplets as in Table 1, such a pattern emerges clearly. The pattern is this: all 64 triplets occur. Each amino acid is represented by one or more *pairs* of triplets which are identical except for one nucleotide. The nonidentical nucleotides in each pair are the two purines or the two pyrimidines. For example, A*C*C and A*U*C (*6*) code for histidine. Also G*G*U codes for tryptophan, and this pattern predicts that G*A*U will also be found to code for tryptophan.

This pattern seems acceptable stereochemically. Furthermore, it is complete and self-consistent. There are exactly 32 pairs; each uses at least one reported triplet. All 64 possible permutations are used. Of the 49 reported triplets, four were discarded: two because there were four amino acids assigned where there could be only three permutations, one because it was assigned to UUU in conflict with phenylalanine, and only one because it was inconsistent with the pattern presented here. The 19 remaining combinations have been assigned to the various amino acids in such a way as to complete the pattern symmetrically. For example, arginine has a pair, C*C*G and C*U*G, and one unpaired triplet, AAG. The missing triplet could be AA*A* or A*G*G. But AAA is already fully "occupied," and AGG is not. Arginine is therefore assigned the missing AGG. As this process continues, completing all the pairs, the number of remaining alternatives is greatly reduced, but all the missing triplets can be accounted for with only one discrepancy.

This last point is illustrated in Table 2 in which the same 64 assignments are retabulated. Here one can readily confirm that each triplet has exactly as many amino acids assigned to it as there can be permutations.

Any pattern of this sort is open to the suspicion that it may merely resemble the true pattern. It would be pointless to compute a "probability" that it could have occurred "by chance," because the data obviously fall into some sort of pattern—they are not random. But there is a suitable test. One can attempt to construct similar-appearing patterns of 32 pairs in which AC and GU are paired, or AU and CG. This attempt was made;

Table 1. In this pattern there are 32 pairs having the two purines or the two pyrimidines in a certain position (illustrated as if in the center). It includes all 64 possible configurations of three nucleotides. It accommodates 45 of the 49 published triplets. *The 19 triplets necessary to complete these 32 pairs have been added and are printed in italics.* Three are discarded because they are internally inconsistent with the published list (too many amino acids for one triplet). One is discarded because it is inconsistent with this pattern. Nineteen remaining triplets are predicted, as shown. The actual order of the nucleotides is still undetermined, except AUU for tyrosine and GUU for cysteine.

Amino acid	Pairs of triplets		Amino acid	Pairs of triplets		
Alanine	CCG*† CUG†	CAG† *CGG*	Leucine	UAU*† UGU*†	CUU*† *CCU*	(UUU* discard)
Arginine	CCG*† CUG†	AAG† *AGG*	Lysine	AAA*† AGA*	AAU*† *AGU*	(ACA* discard)
Asparagine	ACA*† AUA†	CAU† *CGU*	Methionine	AUG*† *ACG*		
Aspartic acid	ACG† AUG†		Phenylalanine	UUU*† UCU†		
Cysteine	GUU*† *GCU*		Proline	CCC*† CUC*†	CAC*† CGC*	
Glutamic acid	AAG*† *AGG*	GAU*† *GGU*	Serine	CUU*† CCU*†	CAG† *CGG*	CGU* *CAU*
Glutamine	AAC*† *AGC*	(AGG† discard)	Threonine	ACC* AUC†	AAC*† *AGC*	(CCG† discard)
Glycine	GUG*† GCG†	GAG† *GGG*	Tryptophan	GGU*† *GAU*		
Histidine	ACC*† AUC†		Tyrosine	AUU* *ACU*		
Isoleucine	AUU*† *ACU*	AAU† *AGU*	Valine	GUU*† *GCU*		

* Reported by Nirenberg's group (*4*); † Reported by Ochoa's group (*4*).

these patterns cannot be constructed without discarding an unreasonable number of well-established data. For example, the matching triplet for phenylalanine-UUU would have to be UGU or UAU, respectively. Both laboratories have identified each of these with three other amino acids. Only UCU, as in the proposed pattern, is free to represent phenylalanine. About seven such conflicts developed in each attempt. In the purine-pyrimidine pattern the one triplet assignment which had to be discarded in this way was reported from one laboratory only. This appears to be a moderately strong indication that this pattern (which incidentally "makes sense" chemically) is not just a contrived modification of some "partially doublet" code.

This test indicates that if there is a pattern of this general sort in the 64 possible triplets, the existing 46 data (after excluding the three which are inconsistent in any case) are more than sufficient to determine that pattern. It appears that this many data could fit only into a true pattern. Previously, when there were too few data it was possible to devise an almost endless number of patterns in which they could be accommodated.

Another test is possible which incidentally tends to validate many of the reported triplets: in 32 cases, if a particular triplet had been missing from the data the same pattern would have been derived, and the missing triplet would have been predicted. In 13 cases a blank would have been left in the pattern.

The pattern at this stage depends only on the published tables of triplets, not on data from amino acid "mutants," and so forth. If these additional clues are used, a beginning can be made in determining the order of the nucleotides within each triplet.

In the experiments which have yielded the triplet codons, no order can be determined. Thus, in Table 1 "AAG" means, "AAG, AGA, *or* GAA." In Table 2, "ACG" means "ACG, AGC, CAG, CGA, GAC, *and* GCA." In Fig. 1, however, these orders have been assigned, and "CAA-threonine" means that exact order, with the provision that the evidence is not rigorously conclusive, and it *might* be AAC. If the purine-pyrimidine link is always in the same position, and if this position were known, the sequence would be determined in all those cases where the remaining two are the same (ACA = ACA). The experimentally determined sequences, AUU for tyrosine and GUU for cysteine (*7*), require one of the codons for isoleucine to be UUA and for valine to be UUG if the special link is in the middle, or UAU and UGU if it is at the right end. It evidently cannot be at the left end. For illustration, it is assumed to be in the middle.

Aside from these, the orderings in Fig. 1 are not rigorously determined. Any nucleotide pair might be exchanged with the diagonally corresponding one and still be consistent with Table 2. For example, lysine-AAU and isoleucine-UAA might be interchanged. However, some possibilities seem much more plausible than others. It seems reasonable to expect that a mutation will often involve the change of a single link. The "allele" pairs in homologous protein sequences might most frequently be amino acids which could replace one another in this

Table 2. Entries of Table 1 rearranged, to emphasize the number of amino acids reported and predicted (printed in italics) for each triplet. There are as many entries for each triplet as there are possible permutations of the three nucleotides — one, three, or six.

Triplets	Reported			Predicted			Discard
AAA CCC GGG UUU	Lys*† Pro*† Phe*†			 *Gly* 			 Leu*
AAC AAG AAU	Asn*† Arg† Asn†	Gln*† Glu*† Ilu†	Thr*† Lys* Lys*†				Lys*
CCA CCG CCU	His*† Ala*† Pro*†	Pro*† Arg*† Ser*†	Thr* Pro*	 *Leu*			 Thr†
GGA GGC GGU	Gly† Gly† Gly*†	 Try*†		*Arg* *Ala* *Glu*	*Glu* *Ser*		Gln†
UUA UUC UUG	Ilu*† Leu*† Cys*†	Leu*† Phe† Leu*†	Tyr*† Ser*† Val*†				
ACG ACU AGU CGU	Ala† Asn† Asp† Ala†	Asp† His† Glu*† Arg†	Ser† Thr† Met*† Ser*	*Gln* *Ilu* *Ilu* *Cys*	*Met* *Ser* *Lys* *Val*	*Thr* *Tyr* *Try* *Asn*	

* Reported by Nirenberg's group (*4*); † Reported by Ochoa's group (*4*).

way. For example, alanine can change to serine if one of its nucleotides changes from G to U. This can occur in two different pairs. However, if the alanines at C . . . G were exchanged with serine and arginine at G . . . C, there would then be no codons of alanine and serine that have two letters in common. Since the combination alanine-serine is the most frequently occurring "allele" pair (*8*), the arrangement as shown is much preferred. Similarly, if valine and cysteine were exchanged, the numerous allele pairs val-ala, val-ilu, val-leu (*6*), and others would not be producible by single-link "interchanges." (This, with the previously mentioned determination of the sequence valine = UUG, constitutes support for this procedure.) By comparing each possible alternative in this way with a list of about 300 "alleles" from hemoglobin and other proteins, the tentative arrangement shown in Fig. 1 was derived.

AAA*† AGA*	LYS	AAC*† *AGC*	gln	AAG*† *AGG*	glu	AAU*† *AGU*	lys
ACA*† AUA†	ASN	ACC*† AUC†	his	ACG† AUG†	asp	*ACU* AUU*†	TYR
CAA*† *CGA*	thr	CAC*† CGC*	PRO	CAG† *CGG*	ala	*CAU* CGU*	ser
CCA* CUA†	thr	CCC*† CUC*†	PRO	CCG*† CUG†	ala	CCU*† CUU*†	ser
GAA† *GGA*	arg	GAC† *GGC*	ser	GAG† *GGG*	GLY	*GAU* GGU*†	try
GCA GUA*†	met	GCC*† GUC†	arg	GCG† GUG*†	GLY	*GCU* GUU*†	CYS
UAA† *UGA*	ilu	UAC† *UGC*	asn	UAG*† *UGG*	glu	UAU*† UGU*†	LEU
UCA UUA*†	ILU	*UCC* UUC*†	leu	*UCG* UUG*†	VAL	UCU† UUU*†	PHE

Fig. 1. A symmetrical pattern in the genetic code. The purine-pyrimidine pairs in the genetic code, with order tentatively determined. In one position of each pair, the two purines (A and G), and the two pyrimidines (C and U) are equivalent, reducing the 64 triplets to 32 "codons." Amino acids in capitals have their sequences unambiguously determined, assuming that the special position is in the middle. Amino acids in lower case could possibly belong to the pairs diagonally opposite, CAA-gln, AAC-thr, and so forth. The frequencies of amino acid "alleles" suggest the assignments indicated. The determinations made by Nirenberg's group are marked with an asterisk; those done by Ochoa's group are indicated with a dagger. Those amino acids predicted by the pattern, but not yet reported, are printed in italics.

There are some other uncertain details in Fig. 1. Different amino acids might have been discarded. For example, glutamine at AGG might have been retained, and arginine at AAG discarded (Table 2). Furthermore, glycine at GAG might exchange places with glutamine, if it were retained at AGG. There are, however, only a few such alternatives, and at each choice there seemed some good clue to the selection. Another source of uncertainty is the possibility of experimental error. Of the 32 pairs, five consist of an unreported (predicted) triplet and a triplet reported from only one laboratory. Any of these might prove to be in error.

One would expect the amino acids which occur only once in the pattern to be the ones which occur least frequently in proteins. This is the case, with four exceptions: aspartic acid and valine occur more frequently, isoleucine and arginine less frequently than average. Perhaps more complete data will show that there is another pair for each of the latter, in place of wrongly assigned pairs for the former two.

Even if there were no basis for choice in the alternative positions indicated in Fig. 1, the tabulation would contain much information about sequence. For each triplet in Table 2 there are one, three, or six possible sequences. Fig. 1 reduces these to two alternatives at most.

From this pattern one may predict all the amino acids coded by the remaining 19 ordered triplets, suggesting that there may be no "nonsense" combinations. This is not a strong inference, however. If one of the reported amino acids is erroneous its assigned pair might represent "nonsense." Or, in a few cases the pairs might be subdivided, so that one of the two triplets would be "nonsense."

It would not be inconsistent to find 64 transfer RNA's, one for each triplet. From the pattern of 32 pairs one might infer, however, that there may be only 32 transfer RNA's and that each one responds indiscriminately to both of its specific triplets. The specificity of the attachment site would reside in one of the four nucleotides in one position, a purine or a pyrimidine in another position, and one of the four nucleotides in a third position. These three determinants would occur in three specific (not necessarily adjacent) positions on the messenger RNA chain. The two triplets of each pair would presumably be indistinguishable to the transfer RNA, which might recognize the third (middle?) nucleotide only by its size. In this sense each codon would consist of a pair of triplets. One might facetiously call this a "two-and-a-half-letter" code. On this interpretation there would be only one transfer RNA for aspartic acid, cysteine, glutamine, histidine, methionine, phenylalanine, tryptophan, tyrosine, and valine. There would be two for each of the other amino acids except serine, which would have three (barring errors). The experimental determination of the number of transfer RNA's for each amino acid would be a powerful check on the correctness of this pattern, and therefore on the validity of the individual reported triplets. The two specific transfer RNA's reported for leucine are consistent with this prediction (*2*).

A strong check will also be provided by each subsequent discovery of a triplet which fits, since the symmetry of this pattern leaves little room for rearrangements.

The six cases where the same amino acid belongs to both related pairs accounts for the evidence on which Roberts based his "doublet" code (*5*). One might also interpret this pattern as a code which is partly doublet and partly triplet. This could be tested experimentally. If only one transfer RNA can be found for alanine, glycine, isoleucine, proline, and threonine, and only two for serine, these will represent "doublets" in Roberts' sense. On the other hand, if the purine-pyrimidine pairs are fundamental there should be a specific transfer RNA for every one of the 32 pairs, if not a separate one for every triplet.

There are four crucial tests which could readily be made with isolated transfer RNA's. Is the transfer RNA of proline which responds to CAC the

same as the one which responds to CGC? Similarly for aspartic acid, ACG and AUG; glycine GCG and GUG; and leucine UAU and UGU. If these should prove to be identical, this pattern would be validated.

That leucine as well as phenylalanine (*4*) is coded by UUU is of considerable interest. Is it an accident caused by some abnormal condition in the in vitro situation, or is it of fundamental significance? It raises the possibility that perhaps in the presence of some other source of information not normally present in the artificial situation there might be a second pattern of some sort. Perhaps some of the discarded triplets (Tables 1 and 2) are other ambiguities of this kind rather than errors. Evidence which seems contrary to this conjecture is that the same transfer RNA of leucine responds to UGU and to UUU (*2*).

This pattern seems to be good evidence for a triplet code, since in it triplets are necessary and sufficient. But if the above-mentioned ambiguities prove to be fundamental, extra information, which might reside in still other links (as an overlapping quadruplet code?) would be required.

If this pattern were the complete code, any simple type of overlapping should be immediately evident by substituting the triplets for the amino acids in a few of the known protein sequences. This does not appear to be the case.

If it were possible to make regularly ordered synthetic polymers such as polydinucleotides, and so forth, the question of whether the three elements of each triplet are adjacent in the chain could be settled. Also, such polymers could be used to study the possibility of overlapping codes. It seems possible that some regular RNA's could be synthesized from synthetic DNA's by way of the mechanism reported by Chamberlin and Berg and by Otaka *et al.* (*9*).

In retrospect, it seems that this simple pattern could have been discovered with fewer clues, and we may wonder why it was not found earlier. In the last 2 years there have been a remarkable number of *ad hoc* proposals to account for the data currently at hand. When there were about 12 triplets identified it was expected by some that the total number would be exactly 20—one for each amino acid. Later a number of special combinations were considered, such as combining the three nucleotides without regard to order, and others of this sort which were mathematically possible but structurally unimaginable. Then there was the "high-U" code, chemically implausible but mathematically capable of accounting for the results to that date. Recently it was proposed that there may be some simple pattern in vivo which is obscured by some circumstance of the in vitro experiments. By permitting some normally hidden potentialities to be expressed, this would produce too many triplets. All these attempted to account for the number 20 and to guess the pattern at a stage when an indefinitely large number of patterns was yet mathematically possible. For this reason the probabilities of success were slight.

In contrast to this approach, I took a more general point of view: whatever the pattern may be, mathematically there are 64 possible triplets. In the end some of them may be "nonsense," or even nonexistent in nature. Some may prove to be equivalent to others, and so forth. But whatever those details may be they will consist of some subpattern of the 64 mathematically possible triplets. Now, with a total of 49 different triplets identified, surely the pattern must be visible! I tabulated these triplets in various ways and shortly this detail appeared: Five amino acids had two triplets, with two letters in common. Another had three, the third being unrelated to the first two. Of these six examples four contained the alternatives C and U (for example aspartic acid-A*C*G and A*U*G). It was not surprising that none contained G in such alternatives, since the G polymers had given the most experimental difficulty and most of the combinations having two G's were as yet unassigned. From this observation Table 1 followed directly and the puzzle practically solved itself.

A similar history has occurred in the related problem of overlapping or nonoverlapping codes. Gamow suggested that the necessary amount of information might be reduced by some systematic constraint on the sequences of amino acids. This could be caused by the same nucleotides serving in more than one triplet simultaneously (*1*). At that time there was a moderate amount of protein sequence data available. Several curious overlapping codes were proposed, each of which was followed enthusiastically and then disproved mathematically, on the basis of the then currently available data as it continued to accumulate. Then Brenner (*10*) concluded that all overlapping triplet codes were inconsistent with the data, and the attention of most protein cryptographers was diverted to nonoverlapping codes, where it has remained ever since. However, the problem was not approached in its most general form. Brenner's computation, the results of the single-step mutations, and the results of Crick *et al.* (*11*) have been taken to disprove overlapping codes, but this is true only for a certain subclass of such codes. Furthermore it includes the unstated assumption that there is no other source of genetic information. (The results of Crick *et al.* (*11*) have also been taken as evidence for a triplet code but this inference is valid only for nonoverlapping codes.) A large number of overlapping codes are still mathematically possible (*12*), and several which involve a regular folding or coiling of the RNA strand so that the non-adjacent nucleotides of the codons assume their specific positions can be visualized structurally.

The substance of the argument against overlapping codes, whether from amino acid sequences or from the single-step mutation data, is that such codes could not contain enough information to account for the observed number of variations in protein sequences. However, all proposed codes have required, explicitly or implicitly, some additional unknown source of information such as "commas," "spacers," "stepping by threes," "forbidden combinations," and so forth. As long as the nature of that additional mechanism is undiscovered the possibility remains that it could contain enough information to supplement an overlapping code. Clear evidence of patterns in the constraints on protein sequences would be a basis for an attack on this problem. The amount of data on protein sequences now available may be barely sufficient for this (*8*). In numerical proportion this problem is at a much earlier stage than that of the nucleotide triplets. Here we required about 45 of the 64 possibilities before the pattern revealed itself. In the protein cryptogram less than 2000 of the potential 8000 tripeptide sequences have been reported. If, say, 3000 of the 8000 were "forbidden" according to some pattern, could we see that pattern? Perhaps this seemingly intricate problem will seem simple when it is solved.

RICHARD V. ECK
Laboratory of Biology, National Cancer Institute, Bethesda, Maryland

References and Notes

1. G. Gamow, *Nature* **173**, 318 (1954).
2. B. Weisblum, S. Benzer, R. W. Holley, *Proc. Natl. Acad. Sci. U.S.* **48**, 1449 (1962).
3. M. W. Nirenberg and J. H. Matthaei, *ibid.* **47**, 1588 (1961).
4. O. W. Jones, Jr. and M. W. Nirenberg, *ibid.* **48**, 2115 (196
5. J. Wahba, S. Gardner, C. Basilio, R. S. Miller, F. Speyer, P. Lengyel, *ibid.* **49**, 116 (1963); two additional triplets were reported at a meetin
6. J. Abelson, *Science* **139**, 774 (1963).
7. R. B. Roberts, *Proc. Natl. Acad. Sci. U.S.* **48**, 897 (1962).
8. Abbreviations: A, adenine; G, guanine; C, cytosine; U, uridine; ala, alanine; arg, arginine; asn, asparagine; asp, aspartic acid; cys, cysteine; gln, glutamine; glu, glutamic acid; gly, glycine; his, histidine; ilu, isoleucine; leu, leucine; lys, lysine; met, methionine; phe, phenylalanine; pro, proline; ser, serine; thr, threonine; try, tryptophan; tyr, tyrosine; val, valine.
9. A. J. Wahba, C. Basilio, J. F. Speyer, P. Lengyel, R. S. Miller, S. Ochoa, *Proc. Natl. Acad. Sci. U.S.* **48**, 1683 (1962).
10. R. V. Eck, *J. Theoret. Biol.* **2**, 139 (1962).
11. M. Chamberlin and P. Berg. *Proc. Natl. Acad. Sci. U.S.* **48**, 81 (196
12. E. Otaka, H. Mitsui, S. Osawa, *ibid.* **48**, 425 (1962).
13. S. Brenner, *ibid.* **43**, 687 (1957).
14. F. H. C. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, *Nature* **192**, 1227 (1961).
15. R. Wall, *ibid.* **193**, 1268 (1962).

20 March 1963

Semiconducting Region of Ytterbium

Abstract. *The resistivity of elemental ytterbium at room temperature rises, by a factor of 11, to a maximum at a pressure of 40 kilobars; a further increase in pressure causes a polymorphic transition; the new phase has a resistivity 80 percent of that of the metal at 1 atmosphere. In the temperature-pressure diagram, the phase boundary has a negative slope. The phase boundary, determined from −190° to 360°C, is a straight line that may be extrapolated nearly to the known α-β transition at 1 atmosphere. Between the transition pressure and 20 kbar, the lowest pressure at which the measurements were made, ytterbium behaved as a semiconductor. The temperature coefficient of resistance is negative; at constant pressure, the resistivity shows the exponential temperature dependence characteristic of a semiconductor. The parameter in the expontial would correspond to an energy gap 0.015 ev at 20 kbar, an increase with pressure to a maximum of 0.080 ev at 37 kbar, and then a decrease to 0.05 ev at 45 kbar.*

In 1954 Bridgman (*1*) published the pressure-volume relationships (to 39 kbar) and the resistivity (to a presumptive 98 kbar) for a number of the rare-earth metals. His pressure scale for the volume work is correct, but on the resistivity work, the scale he used was incorrect. In the following discussion, his figures are corrected where possible to conform to a pressure scale based on the bismuth 1–2 and 6–8 transitions occurring at 25.5 and 88 kbar, respectively.

Particularly intriguing were the results on ytterbium. Bridgman found that at room temperature the resistance of ytterbium increased 11-fold at a pressure of 40 kbar and then rapidly decreased to a value which was 79 percent of that at 1 atmosphere. This value remained nearly constant to about 80 kbar. Vereshchagin *et al.* (*2*) obtained similar results and found that the resistance beyond the peak remained level at least to 200 kbar. Our own work indicates a shallow minimum at about 80 kbar. Both Bridgman and Vereshchagin suggested that the resistance peak represented a polymorphic transition. Recently Hall and co-workers (*3*) verified that a phase change occurs and that the crystal structure changes from face-centered cubic to body-centered cubic. Such a change is known to occur at 1 atmosphere at 798°C.

The large increase in the resistance of ytterbium was such that Bridgman investigated the temperature coefficient of resistance from 0° to 200°C at pressures from 1 atmosphere to 7 kbar. From 1 atmosphere through 6 kbar, the temperature coefficient of resistance was normal for a metal in that the resistance increased with the temperature. At 7 kbar there was a decrease in resistance as the temperature increased from 0° to 100°C, and then the resistance started to increase in the manner characteristic of a metal. Bridgman suggested that ytterbium was being squeezed into a semiconducting state. It is this behavior that is the prime subject of our investigation. The greatest care was taken at temperatures below room temperature. If there was an electrical behavior that was characteristic of semiconduction, it would be most pronounced at the lower temperatures. Unfortunately, our work cannot be made to join that of Bridgman since we were restricted to the pressure region above 20 kbar. In the region between 20 kbar and the α-β phase boundary, ytterbium exhibited electrical characteristics that were characteristic of semiconductors with respect to the magnitude of the resistivity and the temperature coefficient of resistance.

The ytterbium (*4*) was supposedly 99.9 percent pure. Spectral analysis (*5*) showed the following percentages of impurities: Al, 0.015; Ca, 0.03; Fe, 0.025; and Mg, 0.01; if other elements

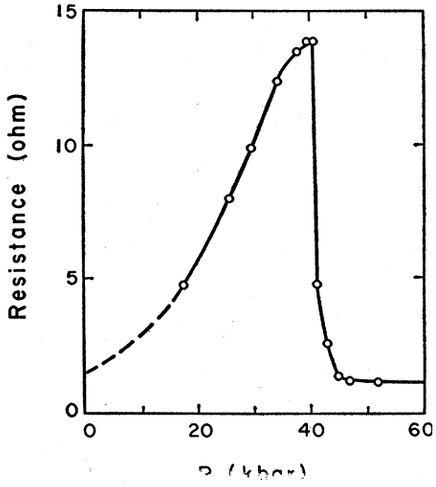

Fig. 1. Resistance-pressure curve for ytterbium at 20°C.

were present the quantities were not detectable by ordinary spectrographic analytical techniques.

The metal was extruded into wire 0.003 inch in diameter; it was then annealed in an argon atmosphere for 15 minutes at 400°C. The Bridgman anvils, silver chloride disks, and pressure measurements have been described (*6*). The resistance of the sample was determined by measuring the voltage drop from the shoulders of the anvils. The current was obtained from a regulated constant-current source. The resistance of the anvils in direct contact was on the order of 30 μohm. There was also a contact resistance, but both these contributions to the resistance were negligible since the resistance of the sample was several ohms. Since the readings were recorded, these factors were neg-

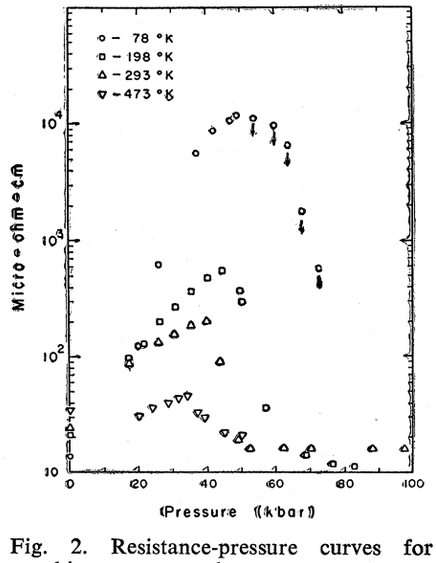

Fig. 2. Resistance-pressure curves for ytterbium at several temperatures.