3/17/63

A SYMMETRICAL PATTERN IN THE GENETIC CODE

The triplets of links in the genetic material which specify the detailed structure of proteins have now been identified in sufficient numbers to fall into a regular pattern.

Richard V. Eck

The author is a biologist in the Laboratory of Biology, National Cancer Institute, Bethesda, Maryland.

The chemical structure of the genetic mechanism must have several very special characteristics. It must contain large, complex molecules capable of representing a large amount of information in their possible alternative configurations. It must be autocatalytic - capable of producing exact replicas of itself. It must do this with great reliability, yet be able to mutate occasionally, and still be autocatalytic in its mutated form. The Watson-Crick model demonstrates how these properties follow from the structure of DNA.

But this accounts only for the self-reproduction and mutation of the genetic material. It must also have quite a different property. It must catalyse specifically some material *other than itself*, which enters into the living chemistry of the organism and results in the phenotypic expression of that "gene". If it is a gene for black rather than brown hair, for example, it must specifically produce some chemical which affects the chain of events which finally results in the production of pigment and its deposition in the hair. The "one gene-one enzyme" formulation has led to the expectation that this non-genic product is protein.

This second problem is more intricate than the first. The reproduction of the genes was pictured as some sort of mold-and-cast system, and so it has proved to be. But for a "template" to determine the structure of some product fundamentally dissimilar from itself requires a more elaborate apparatus. Proteins are long, unbranched chains, made of twenty different kinds of links - amino acids. The exact order of hundreds of these links appears to be genetically determined for each of thousands of different proteins. The DNA of the chromosomes also occurs as long, unbranched chains, but with only four different kinds of links - nucleotides. There are two kinds of nucleotides, two purines and two pyrimidines, which differ markedly in size. The specific matching of one purine with one pyrimidine - adenine with thymine, cytosine with guanine - is the basis for the Watson-Crick model. But how could there possibly be such a matching between the nucleotides and the amino acids, especially in view of their different numbers? This problem was posed in 1954 by Gamow (1) as a mathematical challenge to biochemistry. For each link in a protein chain being constructed, there are twenty possibilities. How can this be specified by a DNA chain containing only four different kinds of links? Evidently more than one, seemingly at least three nucleotides would have to combine in the determination of a single aminc acid link, perhaps in some overlapping fashion.

This mathematical puzzle remains valid, and several facets of it have been clarified. The DNA produces "messenger RNA" apparently by some means similar to the Watson-Crick mechanism. RNA contains nucleotides corresponding to those in DNA, with thymine replaced by uracil. The messenger RNA goes go the protein-producing organelles in the cytoplasm and determines the production of specific protein molecules, according to its detailed sequence of nucleotides. It does not do this directly, but vie a number of adapter molecules, called "transfer RNA's". Presumably each amino acid has one or more transfer RNA which specifically attaches to it. Another part of the transfer RNA attaches to its specific nucleotide "codon" (triplet?) wherever it may occur in the polyribonucleotide chain (2). Thus each amino acid is held in its proper place while the polymerizing mechanism links it to the two adjacent amino acids, determined by the sequence of nucleotides in the information-carrying RNA. 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 two years ago that a non-living system could be made to synthesize an artificial protein, polyphenylalanine, using as messenger RNA synthetic polyuridylic acid (3), has made possible a rapidly developing experimental attack on this problem. The in vitro synthesis of proteins using 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. Judging by the rate at which new ones have been discovered it appears that nearly all, if not all 4 x 4 x 4 = 64 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 due to laboratory errors. The third exception will be considered later.

It frequently appears that the two or more triplets found to code 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, of course, 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.

The Purine-Pyrimidine Pattern

In the most recent lists from the two laboratories there are a total of 49 different triplets, 26 of them reported by both (4). This seems like a large enough proportion of the 64 to outline an over-all 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 non-identical nucleotides in each pair are the two purines or the two pyrimidines. For example, ACC and AUC have been found to code for histidine. GGU codes for tryptophan, and this pattern predicts that GAU 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 using 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, CCG and CUG, and one unpaired triplet, AAG. The missing triplet could be AAA or AGG. 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 as noted.

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; 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 only one triplet assignment had to be discarded in this way, and it 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 accomodated.

Determination of Order

The pattern at this stage depends only on the published tables of triplets, not on data from amino acid "mutants", etc. 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 table 3, 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 sequence AUU for tyrosine and GUU for cysteine (6) requires one of the isoleucine codons to be UUA and 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 table 3 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 amino acids which could replace one another in this way might be expected to be found most frequently as "allele" pairs in homologous protein sequences. 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 having two letters in common. Since alanine-serine is the most frequently-occurring "allele" pair (7), the arrangement as shown is much preferred. Similarly, if valine and cysteine were exchanged, the numerous allele pairs val-ala, val-ilu, val-leu, 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 three hundred "alleles" from hemoglobin and other proteins, the tentative arrangement shown in table 3 was derived.

There are some other uncertain details in table 3. 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.

Even if there were no basis for choice in the alternative positions indicated in table 3, it would contain much information about sequence. For each triplet in table 2 there are one, three, or six possible sequences. Table 3 reduces these to two alternatives at most.

This pattern predicts 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 triplets is erroneous its assigned pair might represent "nonsense". Or, in a few cases the pairs might be sub-divided, one of the two triplets being "nonsense".

Predictions Concerning Transfer RNA's

It is consistent with this pattern that there could be 64 transfer RNA's, one for each triplet. The pattern of 32 pairs suggests, 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, as mentioned above.) 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 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.

Unsettled Questions

The finding 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 fundemental 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 32 codons. Perhaps some of the triplets discarded in making this pattern are other ambiguities of this kind rather than errors. Evidence which seems contrary to this is the finding 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, they would require extra information, which might reside in still other links (as an overlapping quadruplet code?).

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 poly-dinucleotides, etc., 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 using the mechanism reported by Chamberlin and Berg and by Otaka *et al*. (8).

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 lasttwo years there have been a remarkable number of ad hoc proposals to account for the data currently at hand. When there were about twelve 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 but that some circumstance is obscuring it in the in vitro experiments. By permitting some normally hidden potentialities to be expressed, this would produce too many triplets. All of these were attempts to account for the number 20 and to guess the pattern at a stage when an indefinitely large number of patterns were yet mathematically possible. For this reason the probabilities were strongly against success in this approach.

The point of view which led to this pattern was from the opposite direction: Whatever the pattern may be, there are 64 triplets. In the end some of them may be "nonsense", or even non-existent in nature. Some may prove to be equivalent to others, etc. But whatever those details may be they will consist of some sub-pattern 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 vs. U (e.g. aspartic acid-ACG and AUG). 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 non-overlapping 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, using the then currently available data as it continued to increase in amount. Then Brenner (9) concluded that all overlapping triplet codes were inconsistent with the data, and the attention of most protein cryptographers was diverted to non-overlapping 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_ (10) have been taken to disprove overlapping codes, but this is true only for a certain sub-class of such codes. Furthermore it includes the unstated assumption that there is no other source of genetic information. (The results of Crick _et_ _al_ (10) have also been taken as evidence for a triplet code but this inference is valid only for non-overlapping codes.) A large number of overlapping codes are still mathematically possible (11), and several are even structurally plausible which involve a regular folding or coiling of the RNA strand so that the non-adjacent nucleotides of the codons assume their specific positions.

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", etc. 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 protein sequence data now available may be barely sufficient for this (7). 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 two thousand of the potential eight thousand tripeptide sequences have been reported. If, say, three thousand of the eight thousand were "forbidden" according to some pattern, could we see that pattern? Perhaps this seemingly obscure problem will seem simple when it is solved.

Summary

The accumulation of experimental results from the system of Nirenberg and Matthaei has now reached about 75% of the total possible, if the "triplet" concept is correct. Considered as a mathematical puzzle this has proved to be sufficient to determine an apparently unique solution: The sixty-four combinations of four nucleotides taken three at a time, are resolved into thirty-two pairs. The second member of eachpair 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 it predicts which amino acids will be found to correspond to the remaining nineteen unidentified triplets. This pattern accounts for several of the observations concerning regularities inthe data. It partially determines the order of the nucleotides in each triplet and suggests a structural basis for transfer RNA specificity. Whether the three nucleotides of each triplet are adjacent in the nucleic acid chain, and whether they somehow impose constraints on the possible sequences of amino acids which they determine, is yet to be worked out.

Bibliography

1. G. Gamow, *Nature* 173, 318 (1954).

2. B. Weisblum, S. Benzer, and R. W. Holley, *Proc. Nat. Acad. Sci.* 48, 1449 (1962).

3. M. W. Nirenberg and J. H. Matthaei, *Proc. Nat. Acad. Sci.* 47, 1588 (1961).

4. O. W. Jones, Jr. and M. W. Nirenberg, *Proc. Nat. Acad. Sci.* 48, 2115 (1962); J. Wahba, S. Gardner, C. Basilio, R. S. Miller, F. Speyer, and P. Lengyel, *ibid.* 49, 116 (1963); Two additional triplets were reported at a meeting: J. Abelson, *Science* 139, 774 (1963).

5. R. B. Roberts, *Proc. Nat. Acad. Sci.* 48, 897 (1962).

6. A. J. Wahba, C. Basilio, J. F. Speyer, P. Lengyel, R. S. Miller, and S. Ochoa, *Proc. Nat. Acad. Sci.* 48, 1683 (1962).

7. R. V. Eck, *J. Theoret. Biol.* 2, 139 (1962).

8. M. Chamberlin and P. Berg, *Proc. Nat. Acad. Sci.* 48, 81 (1962); E. Otaka, H. Mitsui, and S. Osawa, *ibid.* 48, 425 (1962).

9. S. Brenner, *Proc. Nat. Acad. Sci.* 43, 687 (1957).

10. F. H. C. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin, *Nature*, 192, 1227 (1961).

11. R. Wall, *Nature* 193, 1268 (1962).

R.V.Eck

A symmetrical pattern in the genetic code

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?		

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. 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. 19 remaining triplets are predicted, as shown. The actual order of the nucleotides is still undetermined, except AUU for tyrosine and GUU for cysteine.

* = Reported by Nirenberg's group.

= Reported by Ochoa's group.

? = Predicted by this pattern.

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

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

* = Reported by Nirenberg's group.

= Reported by Ochoa's group.

A symmetrical pattern in the genetic code

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

Table 3. 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, e.g., CAA-gln; AAC-thr, etc. The frequencies of amino acid "alleles" suggest the assignments indicated.

* = Reported by Nirenberg's group.

= Reported by Ochoa's group.