acrocentric, some indication of the presence of a terminal or subterminal centromere was obtained in several cases.

The count of 168 for *P. marinus* appears to be the highest diploid number yet recorded for a vertebrate species for which there are at least several counts. This value is close to that obtained for *Lampetra aepyptera*[8] and to the 3 counts recorded for *E. reissneri* by SASAKI and HITOTSUMACHI[6]. The counts obtained for 3 *Lampetra* species from Italy by ZANANDREA and CAPANNA[4] are also extremely high, ranging from 142 for *Lampetra zanandreai* to 156 for *Lampetra fluviatilis*. While the chromosome numbers of lampreys are generally very high, more work is required to establish the reason for the very considerable intraspecific differences recorded for *L. fluviatilis*[2–4] and *E. reissneri*[5,6]. The fact that far higher diploid numbers, than are apparently present in any other vertebrate group, have now been recorded for species from 3 different genera of lampreys lends support to the view that polyploidy played a role in lamprey evolution, as has already been suggested by OHNO, WOLF and ATKIN[9], HOWELL and DENTON[8], and ROBINSON and POTTER[1].

The diploid number of 168, and the presence of at least a few metacentric chromosomes in the karyotype of *P. marinus*, should be considered in the context of polyploidy. OHNO et al.[9] suggested that the ancestral vertebrate genome contained 48 acrocentric chromosomes, the situation found in hagfishes and several teleosts. They thus postulated on the basis of NOGUSA's[5] counts, that *E. reissneri* was a tetraploid. *P. marinus* may therefore represent an octoploid condition in which some reduction in the total number of chromosomes has occurred through centric fusions. While the numbers of chromosomes in lampreys is now beginning to form a pattern, it is difficult to make generalizations about the centromere position. This is epitomized by the conflict between the descriptions for *E. reissneri* with respect to this character, and the difference between *P. marinus*, which possesses some metacentric chromosomes, and *L. aepyptera* which apparently does not. It is quite evident that further work is required on the chromosome morphology of lampreys before the significance of centromere position can be confidently discussed in terms of the overall pattern of karyotype evolution within the group.

Distribution of counts for the diploid chromosome number of *Petromyzon marinus* in gill epithelium

Diploid Nos.:	163	164	165	166	167	168	169	170	171
No. of counts:	1	4	3	3	2	5	1		1

Résumé. L'étude des chromosomes somatiques de la lamproie, *Petromyzon marinus* L. a montré que l'équipment chromosomique diploïde de cette espèce est composé de 168 chromosomes très petits, dont queleues-uns des plus grands sont métacentriques. Le chiffre semble représenter le plus grand nombre de chromosomes trouvé dans une espèce de Vertébré. On a comparé les chromosomes de *P. marinus* avec ceux des autres espèces de lamproies.

I. C. POTTER[10] and B. ROTHWELL

School of Biological Sciences,
Bath University of Technology,
Bath BA2 7AY (Somerset, England), 14 November 1969.

[9] S. OHNO, V. WOLF and N. B. ATKIN, Heriditas *59*, 169 (1968).
[10] Holder of a Science Research Council Fellowship.

Prediction and Explanation of Chemical Mutation Sites

I wish to report a possible cause for the location of sites on nucleic acid molecules particularly susceptible to chemical mutation. As a model of the approach, I propose a consideration of the formation of mutations in tobacco mosaic virus (TMV) vulgare coat protein by the action of nitrous acid on TMV ribonucleic acid[1].

Since one quickly observes that, for example, of the many arginine residues present in this protein, only 4 were replaced in mutation, it appeared therefore that only the RNA chemical environment about the codon for arginine might disclose a possible explanation for the reaction specificity. This same argument may presumably hold true for all other chemical mutations.

Since TMR-RNA codes directly its coat protein, a knowledge of the protein sequence allows the sequence of the RNA to be written – with, however, several triplets per amino-acid since the genetic code is degenerate. Which of the actual triplets comprise the RNA can only be determined in the laboratory and has not yet been done.

Clearly, neighboring codon patterns which fit the RNA sequence described herein containing all possible codons will obviously fit the single true sequence. Therefore, the RNA sequence about each of the codon corresponding to each mutated amino acid was examined with all of the codons listed. Continuing with the above example, in the case of TMV coat protein, arginine underwent the largest number of mutations (4), all to glycine.

Consideration of the codons of these 2 amino acids immediately indicate the location of the nucleotide which has undergone reaction with nitrous acid, as follows:

Arg	*Gly*
AGA	GGU
AGG	GGC
CGU	GGA
CGA	GGG

The change corresponds to position one of the triplets and involves a base-pair substitution of the transition type.

One next lists all of the RNA regions corresponding to the mutation in question. Thus, in this case, the RNA

[1] A. TSUGITA and H. FRAENKEL-CONRAT, Proc. Nat. Acad. Sci. *46*, 636 (1960); A. TSUGITA, J. Mol. Biol. *5*, 284 (1962); A. TSUGITA and H. FRAENKEL-CONRAT, J. Mol. Biol. *4*, 73 (1962).

Table I. Arginine → Glycine

Position in protein chain	Protein sequence and corresponding codons								
					4 6				
No. 42–50	*Thr*	*Val*	*Val*	*Gln*	*Arg*	*Gln*	*Phe*	*Ser*	*Gln*
	ACU	GUU	GUU	CAC	CGU		UUU	UCU	
	ACC	GUC	GUC	CAG	CGC		UUC	UCC	
	ACA	GUA	GUA		CGA	CAA		UCA	CAA
	ACG	GUG	GUG		CGG	CAG		UCG	CAG
					AGA			AGU	
					AGG			AGC	
					6 1				
No. 57–65	*Gln*	*Val*	*Thr*	*Val*	*Arg*	*Phe*	*Pro*	*Asp*	*Ser*
	CAA	GUU	ACU	GUU	CGU	UUU	CCU	GAU	UCU
	CAG	GUC	ACC	GUC	CGC	UUC	CCC	GAC	UCC
		GUA	ACA	GUA	CGA		CCA		UCA
		GUG	ACG	GUG	CGG		CCG		UCG
					AGA				AGU
					AGG				AGC
					1 2 2				
No. 118–126	*Thr*	*Val*	*Ala*	*Ileu*	*Arg*	*Ser*	*Ala*	*Ileu*	*Asn*
	ACU	GUU	GCU	AUU	CGU	UCU	GCU	AUU	AAU
	ACC	GUC	GCC	AUC	CGC	UCC	GCC	AUC	AAC
	ACA	GUA	GCA	AUA	CGA	UCA	GCA	AUA	
	ACG	GUG	GCG		CGG	UCG	GCG		
					AGA	AGU			
					AGG	AGC			
					1 3 4				
No. 130–138	*Val*	*Glu*	*Leu*	*Ileu*	*Arg*	*Gly*	*Thr*	*Gly*	*Ser*
	GUU	GAA	CUU	AUU	CGU	GGU	ACU	GGU	UCU
	GUC	GAG	CUC	AUC	CGC	GGC	ACC	GGC	UCC
	GUA		CUA	AUA	CGA	GGA	ACA	GGA	UCA
	GUG		CUG		CGG	GGG	ACG	GGG	UCG
			UUA		AGA				AGU
			UUG		AGG				AGC
		↑	↑		↑		↑		
		position + 9 all G	position + 5 all pyrimidines		Zero position		position − 7 all pyrimidines		

chain regions which code for the regions about the arginine residues in positions (46), (61), (122), and (134) of the coat protein. Table I illustrates the compared systems.

Finally, one searches Table I vertically for *common* nucleotides – either all one type of nucleotide (purine (P) or pyrimidine (p)) or one nucleotide (A, U, G, C). The position of the nucleotide suffering chemical change was assigned the value zero. The common nucleotides to the left of the site of chemical change were labeled (+) and those to the right (−).

Common patterns were detected and, of course, are most substantiated for the most frequent mutation. The mutation neighboring group patterns are shown in Table II. Note that in 3 cases, a particular nucleotide, G, rather than a *class* of nucleotides was found to be a common feature (see arginine in Table II). This may be related to the recent findings by KOCHETKOV[2] that non-helical regions of tRNA is particularly rich in uridine residues – thus, in the present case, TMV-RNA may be G-rich in non-helical regions.

Table II. Common neighboring nucleotide patterns for observed chemical mutation

Arginine ⟶ Glycine	+9G, +5p, −7p
Asparagine ⟶ Serine	+6p, +4P, −5p, −11P
Proline ⟶ Leucine	+4G, +3P, −2P, −5P
Serine ⟶ Phenylalanine	+10G, +9P, −3P

In order to check whether these common neighboring group patterns were significant, the same analysis was applied to the codons in the RNA chain corresponding to the protein chain about arginine residues which did *not* undergo replacement in the mutation process.

The results in Table III show that the non-mutated arginine codon regions have no common nucleotides. Thus, there are patterns of neighboring codons about the arginine codons which undergo mutation which are *not present* about the non-mutated arginine codons.

The same analysis was applied to *non*-mutated asparagines, prolines and serines. As with arginine, *in no case* was even a single common nucleotide found present in the codon environments of the corresponding RNA sequences.

These particular constellations of neighboring nucleotides may enhance reactivity of the nucleotide susceptible

[2] N. K. KOCHETKOV, Pure and Appl. Chem. *18*, 257 (1969).

Table III. *Non*-mutated arginine codons and their RNA codon environs

				41				
Thr	*Gln*	*Gln*	*Ala*	*Arg*	*Thr*	*Val*	*Val*	*Gln*
ACU	CAA	CAA	GCU	AGA	ACU	GUU	GUU	CAA
ACC	CAG	CAG	GCC	AGG	ACC	GUC	GUC	CAG
ACA			ACA	CGU	ACA	GUA	GUA	
ACG			GCA	CGC	ACG	GUG	GUG	
			GCC	CGA				
				CGG				
				71				
Phe	*Lys*	*Val*	*Tyr*	*Arg*	*Tyr*	*Asn*	*Ala*	*Val*
UUU	AAA	GUU	UAU		UAU	AAU	GCU	GUU
UUC	AAG	GUC	UAC		UAC	AAC	GCC	GUC
		GUA					GCA	GUA
		GUG					GCG	GUG
				92				
Asp	*Thr*	*Arg*	*Asn*	*Arg*	*Ile*	*Ile*	*Glu*	*Val*
GAU	ACU	AGA	AAU		AUU	AUU	GAA	GUU
GAC	ACC	AGG	AAC		AUC	AUC	GAG	GUC
	ACA	CGU			AUA	AUA		GUA
	ACG	CGC						GUG
		CGA						
		CGG						
				90				
Ala	*Phe*	*Asp*	*Thr*	*Arg*	*Asn*	*Arg*	*Ileu*	*Ileu*
GCU	UUU	GAU	ACU		AAU	AGA	AUU	AUU
GCC	UUC	GAC	ACC		AAC	AGG	AUC	AUC
GCA			ACA			CGU	AUA	AUA
GCG			ACG			CGC		
						CGA		
						CGG		
				112				
Leu	*Asp*	*Ala*	*Thr*	*Arg*	*Arg*	*Val*	*Asp*	*Asp*
UUA	GAU	GCU	ACU		AGA	GUU	GAU	GAU
UUG	GAC	GCC	ACC		AGG	GUC	GAC	GAC
CUU		GCA	ACA		CGU	GUA		
CUC		GCG	ACG		CGC	GUG		
CUA					CGA			
CUG					CGG			
				113				
Asp	*Ala*	*Thr*	*Arg*	*Arg*	*Val*	*Asp*	*Asp*	*Ala*
GAU	GCU	ACU	*AGA*		GUU	GAU	GAU	GCU
GAC	GCC	ACC	AGG		GUC	GAC	GAC	GCC
	GCA	ACA	CGU		GUA			GCA
	GCG	ACG	CGC		GUG			GCG
			CGA					
			CGG					
				141				
Gly	*Ser*	*Tyr*	*Asn*	*Arg*	*Ser*	*Ser*	*Phe*	*Glu*
GGU	AGU	UAU	AAU		AGU	AGU	UUU	GAA
GGC	AGC	UAC	AAC		AGC	AGC	UUC	GAG
GGA	UCU					UCU	UCU	
GGG	UCC					UCC	UCC	
	UCA					UCA	UCA	
	UCG					UCG	UCG	

to mutation by short- and long-range electrostatic interactions, i.e. hydrogen bonding, etc., brought closer by local chain folding.

I wish, therefore, to suggest that such patterns may be found for various chemical mutagens and that such patterns (located most expeditiously by computer) can allow a priori prediction of genetic stability of proteins.

Résumé. En ce qui concerne l'ARN du VMT, les codons voisins y présentent des arrangements analogues. Il s'agit seulement des codons qui subissent une mutation sous l'action de l'acide nitreux, mutation qui conduit à une modification de la protéine de l'enveloppe. Cette observation suggère, que l'examen minutieux de la séquence d'un acide nucléique pourrait permettre de prédire les sites probables où aura lieu la mutation chimique et que les «hot spots» pour une mutation spontanée pourraient aussi avoir des codons voisins caractéristiques.

I. NICHOLSON

Long Island University, Department of Chemistry, C. W. Post College, Brookville, P. O. Greenvale, (New York 11548, USA), 20 October 1969.