Molec. Gen. Genetics 100, 358—363 (1967)

Coat Proteins of Strains of two RNA Viruses: Comparison of their Amino Acid Sequences

B. WITTMANN-LIEBOLD and H. G. WITTMANN
Max-Planck-Institut für Molekulare Genetik, Berlin-Dahlem

Received November 10, 1967

Summary. The amino acid sequences of four strains of tobacco mosaic virus isolated in different parts of the world are compared. The differences between the strains are discussed with respect to special proteinchemical features (such as beginning of the chain, deletion of amino acids, number of different amino acids, sizes and distribution of regions with invariable amino acids) and with respect to the possibility of deducing the most probable nucleotide sequence for the coat protein cistron of tobacco mosaic virus.

The complete amino acid sequences of the two RNA bacteriophage strains fr and f_2 are compared. According to their coat proteins three groups of phages can be formed: 1) MS 2, f_2, M 12 and R 17, 2) fr and 3) $Q\beta$.

During the last years the protein structures of strains of tobacco mosaic virus (TMV) and of an RNA-bacteriophage have been intensively studied. The results up to now allow a comparison of the amino acid sequences of four TMV strains and of two strains of an RNA bacteriophage.

I. Tobacco Mosaic Virus

The amino acid sequences of the following TMV strains are shown in Fig. 1:

V = strain "*vulgare*", isolated by JOHNSON (1926). The first but later corrected amino acid sequence was determined by ANDERER et al. (1960) and TSUGITA et al. (1960). The sequence in Fig. 1 is according to ANDERER, WITTMANN-LIEBOLD and WITTMANN (1965).

D = strain "*dahlemense*", isolated by MELCHERS (1940); the sequence was determined by WITTMANN-LIEBOLD and WITTMANN (1963).

U2 = strain "*U2*", isolated by WILDMAN et al. (1951); the sequence was determined by WITTMANN (1965) and RENTSCHLER (1967).

H = strain "*Homes rib grass*", isolated by HOLMES (1941); the sequence was determined by JAUREGUI-ADELL, HINDENNACH and WITTMANN (to be published), WITTMANN, HINDENNACH and WITTMANN-LIEBOLD (to be published) and FUNATSU (to be published).

Comparing the amino acid sequences of the four TMV strains the following can be pointed out:

1. The protein chain of all so far known phytopathogenic viruses begins with an N-acetyl-group. This is not true for strain U2: The first amino acid is free and not acetylated.

2. The coat proteins of the strain V, D, U2 and of about 300 mutants which have been analyzed (WITTMANN, 1962; WITTMANN-LIEBOLD and WITTMANN, 1965; FUNATSU and FRAENKEL-CONRAT, 1964) consist of 158 amino acids, whereas

strain H has a deletion of two amino acids. In the region between 146—149 *four* positions are free (compared with the three other strains) but only *two* amino acids are present in strain H.

3. The same amino acid at the same position in all four TMV strains (e.g. tyrosine in position 2, isoleucine in 4; etc.) occurs only in 36% of the 158 positions. Furthermore, if one considers in which positions amino acid replacements are present in chemically induced and spontaneous mutant in addition to the four strains, the number of invariable amino acids is reduced to 30%.

4. The positions with the invariable amino acids are not randomely distributed. Two big clusters occur: one consisting of 10 amino acids in positions 113—122 and another with 8 amino acids in position 87—94. These findings probably reflect the restriction of amino acid replacements by the spatial structure of the protein chain.

5. If one compares the codons for two or more amino acids which occur in a given position, it is found that a considerable portion (50% of the amino acid replacements) can only be "explained" by altering more than one nucleotide per codon.

6. A comparison of the amino acid sequences of the four TMV strains allows conclusions about the evolutionary relationship of these strains.

7. The knowledge of the amino acid replacements in chemically induced and spontaneous mutants and of the amino acid sequences of the TMV strains (Fig. 1) allows the deduction of the most probable codon for those amino acids which are coded by several codons, e.g. serine, leucine etc. Thus it is possible to deduce the most probable nucleotide sequence for the TMV coat protein cistron.

The points listed above and summarized briefly will be discussed elsewhere in detail.

II. RNA Bacteriophages

After Loeb and Zinder (1961) had shown the existence of bacteriophages which contain RNA as genetic material, several similar bacteriophages were isolated in various parts of the world.

The amino acid sequences of two such phages, namely f_2 and fr, have been determined and are given in Fig. 2. The homology between their amino acid sequences is so strong that f_2 and fr can be considered as two strains of the same bacteriophage. The strain f_2 was isolated by Loeb and Zinder (1961) and its amino acid sequence was determined by Weber et al. (1966, 1967). The strain fr was isolated by Hoffmann-Berling et al. (1963) and its sequence determined by Wittmann-Liebold et al. (1966 and to be published). Both strains have the same number of amino acids, namely 129. Large regions of both protein chains are identical. The number of positions with different amino acids is 18% and this figure is about the same as the difference in amino acids between the TMV strains V and D. 60% of the amino acid differences between f_2 and fr can be "explained" by alteration of only one nucleotide per codon.

The hitherto analyzed RNA bacteriophages can be divided into three groups according to the homology between their coat proteins:

1. MS2, M12, R17 and f_2 are very strongly related to each other: MS2 and M12 seem to be identical and f_2 and R 17 differ from these by only one amino acid replacement (Lin et al., 1967).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
V	Acetyl	Ser	Tyr	Ser	Ile	Thr	Thr	Pro	Ser	Gln	Phe	Val	Phe	Leu	Ser	Ser	Ala	Try	Ala
D	"	"	"	"	"	"	"	Ser	"	"	"	"	"	"	"	"	"	Val	"	"
U2		Pro	"	Thr	"	Asn	"	"	"	"	"	"	"	Tyr	"	"	"	Ala	Tyr	"
H	"	Ser	"	Asn	"	Thr	Asn	Ser	Asn	"	Tyr	Gln	"	Phe	Ala	Ala	Val	Try	"	"

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
V	Asp	Pro	Ile	Glu	Leu	Ile	Asn	Leu	Cys	Thr	Asn	Ala	Leu	Gly	Asn	Gln	Phe	Gln	Thr
D	"	"	"	"	"	Leu	"	Val	"	"	Ser	Ser	"	"	"	"	"	"	"
U2	"	"	Val	"	"	Ile	"	Leu	"	"	Asn	Ala	"	"	"	"	"	"	"
H	Glu	"	Thr	Pro	Met	Leu	"	Gln	"	Val	Ser	"	"	Ser	Gln	Ser	Tyr	"	"

	38	39	40	41	42	43	44	45	46	47	48	48	50	51	52	53	54	55	56
V	Gln	Gln	Ala	Arg	Thr	Val	Val	Gln	Arg	Gln	Phe	Ser	Gln	Val	Try	Lys	Pro	Ser	Pro
D	"	"	"	"	"	"	Thr	"	"	Gln	"	"	Glu	"	"	"	"	Phe	"
U2	"	"	"	"	"	"	"	"	"	"	"	Ala	Asp	Ala	"	"	"	Ser	"
H	"	Ala	Gly	"	Asp	"	"	Arg	"	"	"	"	Asn	Leu	Leu	Ser	Thr	Ile	Val

	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
V	Gln	Val	Thr	Val	Arg	Phe	Pro	Asp	Ser	Asp	Phe	Lys	Val	Tyr	Arg	Tyr	Asn	Ala	Val
D2	"	Ser	"	"	"	"	"	Gly	Asp	Val	Tyr	"	"	"	"	"	"	"	"
U2	Val	Met	"	"	"	"	"	Ala	Ser	Asp	Phe	Tyr	"	"	"	"	"	Ser	Thr
H	Ala	Pro	Asn	Gln	"	"	"	Asp	Thr	Gly	"	Arg	"	"	Val	Asn	Ser	Ala	Val

	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
V	Leu	Asp	Pro	Leu	Val	Thr	Ala	Leu	Leu	Gly	Ala	Phe	Asp	Thr	Arg	Asn	Arg	Ile	Ile
D	"	"	"	"	Ile	"	"	"	"	"	Thr	"	"	"	"	"	"	"	"
U2	"	"	"	"	"	"	"	"	"	"	Asn	Ser	"	"	"	"	"	"	"
H	Ile	Lys	"	"	Tyr	Glu	"	"	Met	Lys	"	"	"	"	"	"	"	"	"

	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113
V	Glu	Val	Glu	Asn	Gln	Ala	Asn	Pro	Thr	Thr	Ala	Glu	Thr	Leu	Asp	Ala	Thr	Arg	Arg
D	,,	,,	,,	,,	,,	Gln	Ser	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,
U2	Glx	,,	Asx	Asx	Glx	Ala	Asx	,,	,,	,,	,,	(Asx,	Thr,	Glx,	Glx,	Pro,	Val,	Ile)	,,
H	Gln	Thr	Glu	Glu	Gln	Ser	Arg	,,	Ser	Ala	Ser	Gln	Val	Ala	Asn	Ala	Thr	Gln	,,

	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132
V	Val	Asp	Asp	Ala	Thr	Val	Ala	Ile	Arg	Ser	Ala	Ile	Asn	Asn	Leu	Ile	Val	Glu	Leu
D	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	Val	Asn	,,	,,
U2	,,	,,	,,	,,	,,	,,	,,	,,	,,	Ala	Ser	,,	,,	,,	,,	Ala	,,	,,	,,
H	,,	,,	,,	,,	,,	,,	,,	,,	,,	Ser	Gln	,,	Gln	Leu	,,	Leu	,,	,,	,,

	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151
V	Ile	Arg	Gly	Thr	Gly	Ser	Tyr	Asn	Arg	Ser	Ser	Phe	Glu	Ser	Ser	Ser	Gly	Leu	Val
D	Val	,,	,,	,,	,,	Leu	,,	,,	Gln	Asn	Thr	,,	,,	,,	Met	,,	,,	,,	,,
U2	,,	,,	,,	,,	,,	Met	Phe	,,	,,	Ala	Gly	,,	,,	Thr	Ala	,,	,,	,,	,,
H	Ser	Asx	His	Gly	,,	Tyr	Met	,,	Arg	,,	Glu	,,	,,	← Ala	Ile	→	,,	Pro	

	152	153	154	155	156	157	158
V	Try	Thr	Ser	Gly	Pro	Ala	Thr
D	,,	,,	,,	Ala	,,	,,	Ser
U2	,,	,,	,,	Thr	,,	,,	Thr
H	,,	,,	,,	Ala	,,	,,	,,

Fig. 1. Amino acid sequences of four strains of tobacco mosaic virus

	1				5						10						15					
fr	Ala	—	Ser	—	Asn	—	Phe	—	**Glu**	—	**Glu**	—	Phe	—	Val	—	Leu	—	Val	—	Asn	—
f_2	Ala	—	Ser	—	Asn	—	Phe	—	**Thr**	—	**Gln**	—	Phe	—	Val	—	Leu	—	Val	—	Asn	—

	Asp	—	Gly	—	Gly	—	Thr	—	Gly	—	**Asp**	—	Val	—
f_2	Asp	—	Gly	—	Gly	—	Thr	—	Gly	—	**Asn**	—	Val	—

	20						25						30						35			
fr	**Lys**	—	Val	—	Ala	—	Pro	—	Ser	—	Asn	—	Phe	—	Ala	—	Asn	—	Gly	—	Val	—
f_2	**Thr**	—	Val	—	Ala	—	Pro	—	Ser	—	Asn	—	Phe	—	Ala	—	Asn	—	Gly	—	Val	—

	Ala	—	Glu	—	Try	—	Ile	—	Ser	—	Ser	—	Asn	—
f_2	Ala	—	Glu	—	Try	—	Ile	—	Ser	—	Ser	—	Asn	—

	40						45						50									
fr	Ser	—	Arg	—	Ser	—	Gln	—	Ala	—	Tyr	—	Lys	—	Val	—	Thr	—	Cys	—	Ser	—
f_2	Ser	—	Arg	—	Ser	—	Gln	—	Ala	—	Tyr	—	Lys	—	Val	—	Thr	—	Cys	—	Ser	—

	Val	—	Arg	—	Gln	—	Ser	—	Ser	—	Ala	—	**Asn**	—
f_2	Val	—	Arg	—	Gln	—	Ser	—	Ser	—	Ala	—	**Gln**	—

	55						60						65						70			
fr	Asn	—	Arg	—	Lys	—	Tyr	—	Thr	—	**Val**	—	Lys	—	Val	—	Glu	—	Val	—	Pro	—
f_2	Asn	—	Arg	—	Lys	—	Tyr	—	Thr	—	**Ile**	—	Lys	—	Val	—	Glu	—	Val	—	Pro	—

	Lys	—	Val	—	Ala	—	Thr	—	Gln	—	**Val**	—	**Gln**	—
f_2	Lys	—	Val	—	Ala	—	Thr	—	Gln	—	**Thr**	—	**Val**	—

	75						80						85						90			
fr	Gly	—	Gly	—	Val	—	Glu	—	Leu	—	Pro	—	Val	—	Ala	—	Ala	—	Try	—	Arg	—
f_2	Gly	—	Gly	—	Val	—	Glu	—	Leu	—	Pro	—	Val	—	Ala	—	Ala	—	Try	—	Arg	—

	Ser	—	Tyr	—	**Met**	—	Asn	—	**Met**	—	Glu	—	Leu	—
f_2	Ser	—	Tyr	—	**Leu**	—	Asn	—	**Leu**	—	Gln	—	Leu	—

	95						100						105									
fr	Thr	—	Ile	—	Pro	—	**Val**	—	Phe	—	Ala	—	Thr	—	**Asx**	—	**Asp**	—	Asp	—	Cys	—
f_2	Thr	—	Ile	—	Pro	—	**Ile**	—	Phe	—	Ala	—	Thr	—	**Asn**	—	**Ser**	—	Asp	—	Cys	—

	Ala	—	Leu	—	Ile	—	Val	—	Lys	—	Ala	—	**Leu**	—
f_2	**Glu**	—	Leu	—	Ile	—	Val	—	Lys	—	Ala	—	**Met**	—

	110						115						120						125			
fr	Gln	—	Gly	—	**Thr**	—	**Phe**	—	Lys	—	**Thr**	—	Gly	—	**Ile**	—	**Ala**	—	**Pro**	—	**Asn**	—
f_2	Gln	—	Gly	—	**Leu**	—	**Leu**	—	Lys	—	**Asp**	—	Gly	—	**Asn**	—	**Pro**	—	**Ile**	—	**Pro**	—

	Thr	—	Ala	—	Ile	—	Ala	—	Ala	—	Asn	—	Ser	—
f_2	**Ser**	—	Ala	—	Ile	—	Ala	—	Ala	—	Asn	—	Ser	—

	129			
fr	Gly	—	Ile	—	Tyr	—
f_2	Gly	—	Ile	—	Tyr	—

Fig. 2. Amino acid sequences of two RNA bacteriophage strains

2. The relationship between fr and the strains of the first group is still close, although much less than between the strains within the first group. As mentioned above 18% of the 129 amino acids between fr and f_2 are different.

3. Qβ, which has been isolated in Japan belongs in a third group because it is very different from f_2 in its coat protein (W. KONIGSBERG, personal communication).

References

ANDERER, F. A., H. UHLIG, E. WEBER, and G. SCHRAMM: Primary structure of the protein of TMV. Nature (Lond.) **186**, 922—925 (1960).

— B. WITTMANN-LIEBOLD u. H. G. WITTMANN: Weitere Untersuchungen zur Aminosäuresequenz des Proteins im Tabakmosaikvirus. Z. Naturforsch. **20**b, 1203—1213 (1965).

FUNATSU, G., and H. FRAENKEL-CONRAT: Location of amino acid exchanges in chemically evoked mutants of tobacco mosaic virus. Biochemistry **3**, 1356—1362 (1964).

HOFFMANN-BERLING, H., D. A. MARVIN u. H. DÜRWALD: Ein fädiger DNS-Phage (fd) und ein sphärischer RNS-Phage (fr) wirtsspezifisch für männliche Stämme von *E. coli*. Teil I. Z. Naturforsch. **18**b, 876—883 (1963).

HOLMES, F. O.: A distinctive strain of tobacco mosaic virus from plantago. Phytopathology **31**, 1089—1098 (1941).

JOHNSON, J.: Mosaic disease on differential hosts. Phytopathology **16**, 141—149 (1962).

LIN, J. Y., C. M. TSUNG, and H. FRAENKEL-CONRAT: The coat protein of the RNA bacteriophage MS 2. J. molec. Biol. **24**, 1—14 (1967).

LOEB, T., and N. D. ZINDER: A bacteriophage containing RNA. Proc. nat. Acad. Sci. (Wash.) **47**, 282—289 (1961).

MELCHERS, G.: Biologische Untersuchung des „Tomatenvirus Dahlem 1940". Zbl. Biol. **60**, 527—537 (1940).

RENTSCHLER, L.: Aminosäuresequenzen und physikochemisches Verhalten des Hüllproteins des Tabakmosaikvirus. Molec. Gen. Genetics **100**, 96—108 (1967).

TSUGITA, A., D. T. GISH, J. YOUNG, H. FRAENKEL-CONRAT, C. A. KNIGTH, and W. M. STANLEY: The complete amino acid sequence of the protein of tobacco mosaic virus. Proc. nat. Acad. Sci. (Wash.) **46**, 1463—1469 (1960).

WEBER, K., and W. KONIGSBERG: Amino acid sequence of the f_2 coat protein. J. Biol. Chem. **242**, 3563—3578 (1967).

— G. NOTANI, M. WIKLER, and W. KONIGSBERG: Amino acid sequence of the f_2 coat protein. J. molec. Biol. **20**, 423—425 (1966).

WILDMAN, S. G., and R. D. OWEN: The detection and isolation of naturally occuring strains of tobacco mosaic virus by electrophoresis. Science **114**, 463—465 (1951).

WITTMANN, H. G.: Proteinuntersuchungen an Mutanten des Tabakmosaikvirus als Beitrag zum Problem des genetischen Codes. Z. Vererbungsl. **93**, 491—530 (1962).

— Die primäre Proteinstruktur von Stämmen des Tabakmosaikvirus. Teil IV. Z. Naturforsch. **20**b, 1213—1223 (1965).

WITTMANN-LIEBOLD, B.: Aminosäuresequenzen im Hüllenprotein des RNS-Bakteriophagen fr. Z. Naturforsch. **21**b, 1249 (1966).

—, u. H. G. WITTMANN: Die primäre Proteinstruktur von Stämmen des Tabakmosaikvirus. Teil III. Z. Vererbungsl. **94**, 427—435 (1963).

— — Lokalisierung von Aminosäureaustauschen bei Nitritmutanten des Tabakmosaikvirus. Z. Vererbungsl. **97**, 305—326 (1965).

Dr. B. WITTMANN-LIEBOLD
Dr. H. G. WITTMANN
1 Berlin 33, Harnackstr. 23