

Nucleotide sequence of capsid, E2 and E1 protein genes of Rubella virus vaccine strain RA27/3

Hira L.Nakhasi*, Desiree Thomas, Dexian Zheng and Teh-Yung Liu

DBB, CBER, FDA, 8800 Rockville Pike, Bethesda, MD 20892, USA
 Submitted April 6, 1989

EMBL accession no. X14871

Complete nucleotide sequence of the 24S subgenomic RNA of the vaccine strain rubella virus RA27/3 was determined. The total number of amino acids which code for the capsid, E2 and E1 proteins is the same as in the wild type virus (1-4). Comparison between the structural proteins of the vaccine strain and the wild strain rubella virus revealed that after attenuation, thirty-one amino acids have changed in the vaccine strain of which twelve each are in the capsid and E1 and seven are in the E2 protein. In contrast, only five amino acids have changed in another vaccine strain of rubella virus HPV77 (5). The higher degree of change in RA27/3 strain might have resulted in this strain being a safer and more effective vaccine than HPV77 strain.

```

5'-CGGCCCTGACCACCCGGCCACCATCGCGCGCTCGAGGAGTTGACAGCCCTACGCCGCGCCAATCTC 70
CACGACGCTGACTAACGCCCTGTACTGCTGGGECCTTAATCTTACTACTCTAACCCGGTTCATCACCACCGTTGTTTCCGCCGATCTGGTGGGTACCACTTTTGCACATTCGGGAGAG 190
Capsid M A S T T P I T M E D L Q K A L E T Q S R V L R A G L T A G A S Q S R
CCCCAGGGTGCCTAATGCTTCTACTACCCCATCACATGGAGACTCCAGAAGGCCCTCGAGACACAAATCCCGGCTCTCGCCGCGGGTCTCACTGCGCGCGCCCTCGCAGTCCGCGC 310
R P R P P R Q R D S S T T G D D S G R D S G G P R R P R G N R G R G Q R R D W S
CGCCCGCGCGCGCGCAGACACCGCGACTCAGCAGCAGGAGATGACTCCGGCCGTGACTCCGAGGGGCCCGCGCCGCAACCGGGCCGTGGCCAGCGCAGGGACTGGTCC 430
R A P P P P E E R Q E S R S Q T P A P K P S R A P P Q Q P Q P P R M Q T G R G G
AGGCCCGCCCGCCCCCGAAGAGCGGCAAGAAAGTCCGTCAGACTCCGGCCCGCAAGCCATCGGGCGCCGCGCACACAGCCCTCAACCCCGCGTATGCAACCCGGCGTGGGGCC 550
S A P R P E L G P T T N P F Q A A V A R G L R P P L H D P T E A P T E A C V T
TCTGCCCGCGCGCCGAGCTGGGGCCAGCCACCCGTTCAAGCAGCCGTGGCCGTGGCTGGCCGCGCTCTCCAGACCCCTGACACCGAGGCCACCCAGGCGCTGGCTGACC 670
S W L W S E G E G A V F Y R V D L H F T N L G T P P L D E D G R W D P A L M Y N
TCATGGCTTTGGAGCGGGCGAAGGGCGGCTCTTTTACCGCGTGCACCTGATTTTACCAACTGGGACCCCGCCACTCGACGAGGACCGCCGCTGGGACCTGGCTCATGTACAAC 790
P C G P E P P A H V V V A Y N Q P A G D V R G V W G K G E R T Y A E Q D F R V G
CCTTGGCGCGCGCCGCGCTCAGCTCGCTCCGGCGTACAACCTGCCCAGCGTCAAGGCGCTTTGGGTAAAGGTGAGCCACTACGCCGAGCAGGATTTCCCGCTGGCC 910
G T R W L R H L L R M P V R L D G D T A P L S P H T T E R S A R N P S R
GGCAGCGCTGGCAGCAGTCTGCTGGCATCCGCTCGCGCGCTCGACCGCGCACCCCGGCTTCCCCCACACACCCGAGCGCATTGAGACCCGCTCGGCCGCTCATCTGGAGC 1130
I R F G A P Q A F L A G L L L A A V A V G T A R A G L Q P R V D M A A P P H P P
ATCCGCTTCGCTGCCCGAGGCTTCTTTCCGCGGCTCTCTCGCCCGGCTGGCTTGGCAGCGCGCGCGCGGCTCCAGCCCGGTTGATATGGCGCCCGCCCTATGCCCGCG 1250
Q P P R L G Q H Y G H H H H Q L P F L G H D G H H G G T L R V G Q H H R H A S
CAGCCCGCGCTGGCAGCGGAGCATTACGGCACACCCACCATCACTGCGGTTCTCGGGGACGAGGGCCATCATGGCGGCACTTGGCGCTCGCGGAGCATCAGCCGAAACCGCAGC 1370
D V L P G H W L Q G G W G C Y N L S D W H Q G T H V C H T K H M D F W C V E H D
GAGCTGCGCCCGCCAGCTGCTCAAGGCGGCTGGGTGCTCAACCTGAGCAGTGGCAGCAGGGGACTCATGTCTGACACCAAGACACATGACTTTTGGTGTGGAGCAGCAG 1490
R P P P A T P T S L T T A A S T T A A T P A T A P A P C H A G L N D S C G G F
CGACCGCGCGCGCCGACCCGAGCTCTCAACACCGGCGGAATCAAGCAGCGCCGCAACCCCGCCACTGGCGCGCCCTGCCACCGCGCTCAATGACAGCTGGCGGCGCTC 1610
L S G C G P M R L R H G A D T R C G R L I C G L S T T A Q Y P P T R F G Y A M R
TTGTCTGGGTGGGGCGATGCGCCTGGCCGCGCGCTGACACCCGGTGGCTGGTGTATCTGGCGGCTGTCTACCACCGCCAGTACCCGCTACCCCGGTTTGGCTGTCTATGGG 1730
W G L P P W E L V V L T A R P E D G M T C R G V P A H P G T R C P E L V S P M G
TGGGGCTTCCCCCTGGGAATGGTGTCTTACCCCGCGCCCGAAGCGGCTGGACTGTGCTGGCTGCCCGCCACCCAGCCGCTGCCCGGAAGTGGTGGGCGCATGGGA 1850
R A T C S P A S A L W L A T A N A L S L D H A F A A F V L L W P V V L I F M V C
CGCGGACTTGTCTCCAGCTCGGCCCTCTGGCTGCCACAGCAGCGGCTGTCTTGGACACCGCTTGGCGGCTTGTCTGCTGGTCCCGTGGTGGTGGTGGTGGTGGTGG 1970
R R A C T R R G A A A A L T A V L L Q G Y N P P A Y G E E A F T Y L C T A P G C
CGCCGCGCTGTGCCCGCGCGCCCGCCCTCAACCGGCTCTCGCAGGGGTCACACCCCGCCGCTATGGCAGGAGGCTTTCACCTACTCTGCAGTCAACCGGTTGG 2090
A T Q A P P V P V R L A G V R F E S K I V D G G C F A P W D L E A T G A C I C E I
GCCACTCAAGCAGCTGTCCCGTGGCGCTGGCTGGCTGGCTTGGTCAAGATTGTGGACGGCGCTGCTTGGCCCATGGGACCTGGAGCCGCTGATTCAGGATC 2110
P T D V S C E G L G A M V P T A P C A R I W N G T Q R A C T F W A V N A Y S S G
CCCAGTGTCTGTCGAGGGCTTGGGGCTGGTACCACAGCCCTTGGCGCGCATCTGGAATGGCACAGCGCGCTGCACTTCTGGGCTGCAACCGCTACTCTCTGGC 2230
G Y A Q L A S Y F M G G S Y K Q Y H P T A C E V P A F G H S D A A C W G F
GGGTACCGCAGCTGCCCTTACTCAACCTGGCGGAGCTACTACAAGAGTACCACCTACCGGCTGGAGGTTCAACTGCTTGGACACAGCGAGCGGCTGCTGGGGCTC 2350
P T D T V M S V F A L A S Y V Q H P H K T V R V K F H T E T R T V W Q L S V A G
CCCACGACACCGTGTAGGCTGTGGCTCGCTAGCTACGTCAGCAGCCCTCACAAGACCGTCCGGGCTCAAGTTTCATACAGAGACTAGGACTGTCTGGCACTCTCGTGGTGGC 2470
V S C N V T T E H P F C N T P H G O L E V Q V L P D P G D L V E Y I M N H T G N
GTGCTGCAACGCTCACCCTGAACCCGTTCTGCAACCGCGCCGACGACAACTCGAGTCCAGTCTGCTGACCTGGGACCTGTTGATACATATGAACACACCGGCAAT 2590
Q O S R W L G S P N C H G P V C A G S V C Q R H S P D C S R T G S R L V G V T P E R P
    
```

```

CAGCAGTCCCAGTGGGGCCCTCGGGAGCCCGAATTGCCATGGCCCGATTGGGCTCCCCGGTTTGCCAACGCCATTCCCCTGACTGCTCGGGGCTTGTGGGGTCAAGCCAGAGCGTCCC 2710
R L R L V D A D D P L L R T A P G P G E V W V T P V I G S Q A R K C G L H I R A 875
CGGTGCGCCTGGTGCACCGCCAGCACCCCTGCTGGCAGTCCCTGGCCCGCGGAGGTGTGGTCACGGCTGTATAGGCTCTCAGGGCCGCAAGTGGCGCTCCACATACGGCT 2830
G P Y G H A T V E M P E W I H A H S T S D P W H P P G P L R L K F K T V R P V A 915
GGACCGTACGGCCATGCTACCGTCGAAATGCCCGAGTGGATCCACGCCCACTTACCAGCGACCCCTGGCACCCACCGGGCCCTTGGGTTGAAGTCAAGACAGTTCGCCCGGTGGCC 2950
L P R A L A P P R N V R V T G C Y O C G T P A L V E G L A P G G G N C H L T V N 955
CTGCCACCGCGTTAGCGCCACCCCGCAATGCGGTGACCCGGTGTACCACTGCGGTACCCCGCGCTGGTGAAGGCTTGCCTCCCGGGGAGGGAAATGCCATCTCACCGTCAAT 3070
G E D V G A F P P G K F V T A A L L N T P P P Y Q V S C G G E S D R A S A R V I 995
GGGAGGATGTCGGGCGCTTCCCGCTGGGAAGTTCGTACCCCGCCCTCTCAACACTCCCCGCCCTACCAAGTCAGCTGGGGGGGAGAGGCATCGCCGAGCGCGCGGGTCAAT 3190
D P A A Q S F T G V V Y G T H T T A V S E T R Q T W A E W A A A H W W O L T L G 1035
GACCCCGCCCGCAATCGTTTACCGGGTGGTGTATGGCACACACCACTGCTGTGTGGAGACCCCGGACACTGGGCGGAGTGGGCTGCTGCCATTTGGTGGCAGCTCACTTGGGC 3310
A I C A L L L A G L L A C C A K C L Y L R G A I A P R 1063
GCCATTTGCCCGCTCTACTTGTGGTCTACTCGCTTGTGTGCCAAATGCTTGTACTACTTGGCGGGCGTATAGCCCGCGTAGTGGGCCCGCGGAAACCCGCACTAGCCACT 3430
AGATCCCGCACCTGTGCTGCATAG - Poly(A) 3456

```

Fig 1. Signal sequences are underlined. The palindromic sequences at the 5' -end and 3' -end of 24SmRNA are shown by a thick underline. Glycosylation sites are shown by (▲). Amino acids different from the wild type strain (Therien) are shown by (●) and different from the vaccine strain (HPV77) are shown by (■).

Acknowledgments: We thank Drs. Arthur Elliott and Suresh Prasad of Merck Sharp and Dohme for providing RA27/3 strain of rubella virus.

*To whom correspondence should be addressed

References:

1. Frey, K.T., et al. (1986) *Virology* 154, 228-232.
2. Frey, K.T., et al (1988) *Gene* 62, 85-99.
3. Vidgren, G., et al. (1987) *J. Gen. Virol.* 68, 2347-2357.
4. Takkinen, K., et al. (1988) *J. Gen Virol.* 69, 603-612.
5. Zheng, D., et al. (1989) *Gene* (in press).