NUCLEOTIDE SEQUENCING OF MOUSE HEPATITIS VIRUS STRAIN JHM MESSENGER RNA 7

Michael Skinner and Stuart Siddell

Institute of Virology
University of Würzburg, F.R.G.

INTRODUCTION

Determination of the nucleic acid sequence of MHV genomic and subgenomic RNAs is clearly a way of predicting the primary structure of encoded polypeptides and at the same time identifying non-coding regions, which may be relevant to the regulation of the viral genes and to the replication of the viral RNA. A comparative analysis of the sequences of different MHV strains, or mutants with altered phenotypes, and of other coronaviruses will also be a method of identifying functionally important regions by their conserved sequences. Such analysis will also be required to relate changes in nucleotide sequence to changes, for example, in host range and pathogenicity.

As the first stage of a project to clone and sequence DNA copies of MHV-JHM RNA, we decided to clone the smallest and most abundant viral mRNA found in JHM-infected cells (mRNA7), the messenger which represents the 1800 bases of 3' terminal viral sequence encoding the nucleocapsid (N) protein. The N protein of JHM has an apparent molecular weight, on SDS-polyacrylamide gels, of approximately 60,000[1] and has been shown to be a basic protein[2]. It is the only phosphoprotein[3] found in the virion, being phosphorylated at serine residues[3]. It can also serve as a substrate for a kinase associated with the purified virion[4]. Little is known about the role of N protein[3] in the assembly of the MHV capsid but it has been suggested[3] that the degree of phosphorylation of the protein might regulate its interaction with the viral RNA and hence the maturation of virus particles.

P. J. M. Rottier et al. (eds.), *Molecular Biology and Pathogenesis of Coronaviruses*

METHODS

Synthesis and Cloning of cDNA

Polyadenylated RNA was isolated from Sac(-) cells that had been infected with MHV-JHM, as previously described[1]. Double-stranded cDNA was prepared from 10 ug of polyadenylated RNA, using oligo dT and oligo dG to prime first and second strand synthesis, respectively, then oligo dC-tailed, double-stranded cDNA was annealed to PstI-cleaved, oligo dG-tailed pAT153, according to protocols described by Land et al[5]. Escherichia coli HB101 was transformed with the annealed plasmid/cDNA using the method of Dagert and Ehrlich[6]. Tetracycline-resistant, ampicillin-sensitive bacterial clones, from two independent cDNA cloning experiments, were screened for JHM-specific sequences by hybridization with a single-stranded cDNA probe, containing ^{32}P, that had been copied from genome RNA isolated from purified virions.

Characterization of cloned cDNA

The size of inserts in plasmids from strongly-hybridizing clones was determined by gel electrophoresis of DNA extracted by the method of Holmes and Quigley[7]. Plasmids containing the largest inserts were prepared from 1 litre cultures and were purified by equilibrium centrifugation in ethidium bromide/caesium chloride. Inserts were excised from the plasmids using PstI and were recovered from agarose gels by electroelution. The inserts were mapped by partial digestion with restriction enzymes of DNA that had been labelled using ^{32}P cordycepin triphosphate and terminal deoxynucleotidyl transferase.

Nucleotide sequencing

Fragments of two cDNA inserts were generated by a variety of restriction enzymes and cloned into the M13 vectors mp 8 and mp 9[8]. The fragments were sequenced using the chain terminator method of Sanger et al.[9]. 77 % of the cDNA was sequenced on both strands, a further 14 % on different but overlapping fragments of the same strand and the remainder was sequenced at least twice. Towards the end of the sequencing project specific clones were identified by their hybridization to a panel of characterized M13 clones. M13 hybridization probes were prepared by the method of Hu and Messing[10].

Translation in vitro, hybrid-arrested translation and polyacrylamide gel electrophoresis

Polyadenylated RNA from cells infected with MHV-JHM was translated in a rabbit reticulocyte lysate as previously described[1]. Hybrid-arrested translation experiments, using purified cDNA insert and polyadenylated RNA from cells infected with MHV-JHM, were performed according to the method of Paterson and Kuff[11], using a rabbit reticulocyte lysate. Translation products were analysed on 15 % polyacrylamide gels[12].

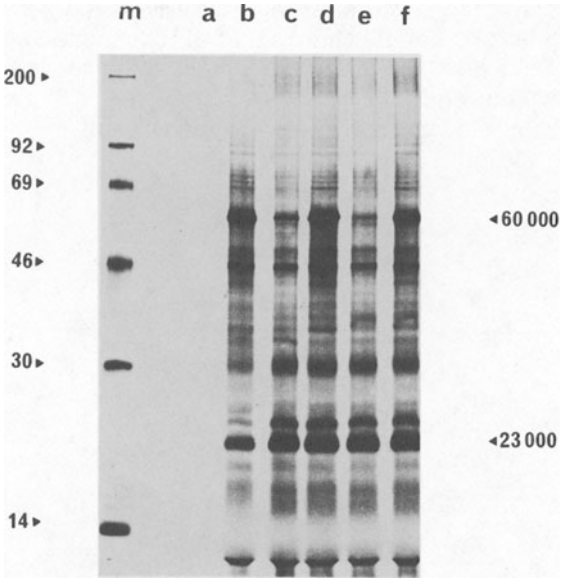

Fig. 1 Hybrid-arrested translation of MHV-JHM mRNA7

Autoradiograph of the ^{35}S-methionine-labelled products synthesized in a rabbit reticulocyte lysate and separated on a 15 % polyacrylamide-SDS gel. The samples translated were: (a) no added RNA, (b-f) 250 ng of polyadenylated, cytoplasmic RNA from cells infected with MHV-JHM, and either 500 ng of the insert DNA (c and d) or 1000 ng of the insert DNA (e and f). Samples (c) and (e) were in the hybrid conformation and samples (d) and (f) were heated to melt the hybrids. The major products of 60,000 and 23,000 mol.wt. have been identified as the nucleocapsid and matrix proteins,respectively. Sample (m) contained ^{14}C-labelled molecular weight markers (CFA626, Amersham Buchler, Braunschweig, F.R.G.).

RESULTS AND DISCUSSION

JHM-specific cDNA clones were isolated from two independent cDNA preparations. The largest clone derived from each preparation (1700 bp in pSS38 and 830 bp in pMS38) were then mapped by partial digestion with restriction endonucleases. This analysis showed that the two clones, both of which hybridized to the genomic and all the subgenomic RNAs of JHM (unpublished results) were overlapping such that the insert from pMS38 carried an extra 40 base pairs at one end. The 1700 base pair insert (in pSS38) was sufficient to account for most of mRNA7, and hybrid-arrested translation was used to confirm that this insert represented the body of mRNA7 (Fig.1). Hybridization of the cDNA insert to infected cell RNA, before translation of the RNA, resulted in specific inhibition of the translation of a 60,000 mol.wt. polypeptide that has been identified as the intracellular precursor of the virion N protein[1]. Melting of the hybrids before translation restored the synthesis of N protein.

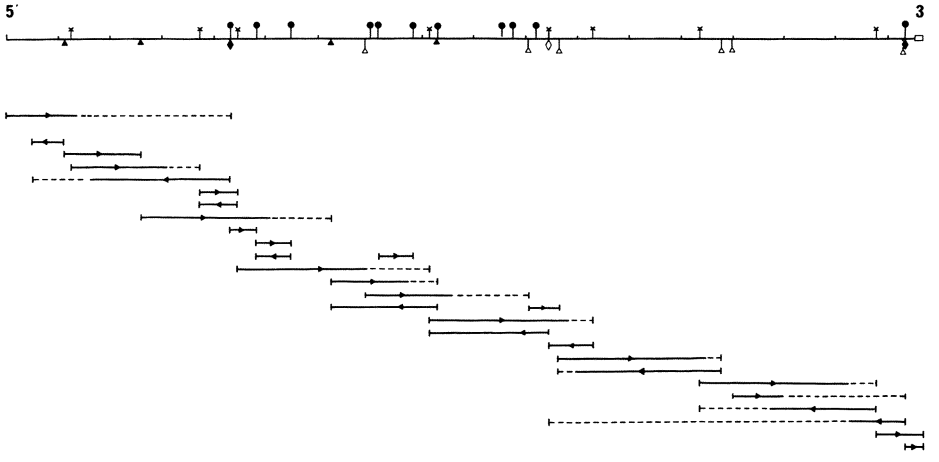

Fig. 2 Diagram showing those restriction endonuclease sites, in the DNA copy of MHV-JHM mRNA 7, used for subcloning into M13 vectors. Solid arrows show the direction and extent of sequence obtained from each clone. Broken lines indicate the probable extent of each clone. Restriction endonuclease cleavage sites are denoted by the symbols: ⏚ HaeIII, ▲ MspI, ♦ BalI, ⚸ AluI, ⅄ Sau3a and ◇ PvuII. The box at the 3' terminus represents the poly(A) tract.

On the basis of the restriction enzyme map, fragments of both cDNA inserts were subcloned into M13 vectors and their sequences were determined (Fig.2). The combined sequence of 1767 base pairs (contained between the oligo dC/dG tails made during the process of cloning) is shown in Fig.3. The sequence contains an open-reading frame, beginning at the first possible AUG codon (nucleotide 84) and ending at nucleotide 1449, capable of encoding a basic protein (Table 1) of 455 amino acids with a molecular weight of 49,700. As the N protein is known to be basic and phosphorylated, such a figure is consistent with the apparent molecular weight of 60,000. The only other open-reading frame of greater than 100 nucleotides was from nucleotides 361-771, potentially able to encode 137 amino acids.

The N protein of Semliki Forest Virus has large clusters of lysine and proline residues in the 110 amino acids adjacent to the N terminal[13]. Such a feature is not seen in the predicted amino acid sequence of the JHM nucleocapsid protein (Fig.3) but some regions enriched in basic amino acids are apparent (e.g. arg-43 to lys-50, lys-101 to lys-113, arg-196 to lys-230, arg-264 to arg-290 and lys-392 to lys-405). Such regions may be responsible for binding of the N protein to the RNA. As in the N protein of SFV, the carboxy terminus is acidic. In view of the fact that phosphorylation of the N protein appears to be important in allowing it to associate with membranes (S. Stohlman, unpublished) and that N protein is phosphorylated at serine residues , it is interesting to consider the distribution of serine residues in the predicted sequence. It is striking that the region ser-194 to ser-220 contains 9 serine residues, or 24 % of the total serine

Table 1 Amino acid composition of the predicted nucleocapsid protein sequence of MHV-JHM.

Arg	29	Glu	22
Lys	33	Asp	22
His	3		
		Val	29
Ser	38	Ile	12
Thr	23	Leu	25
Tyr	12	Phe	17
Gln	40	Ala	35
Asn	25	Pro	34
Gly	45	Trp	5
Cys	2	Met	4

```
                TATAAGAGTGATTGGCGTCCGTACGTACCCTCTCTACTCTAAAACTCTTGTAGTTTAAATCTAATCTAATCTAAACTTTAAGG

      Met Ser Phe Val Pro Gly Gln Glu Asn Ala Gly Ser Arg Ser Ser Ser Gly Asn Arg Ala Gly Asn Gly Ile Leu (25)
(84)  ATG TCT TTT GTT CCT GGG CAA GAA AAT GCC GGT AGC AGA AGC TCC TCT GGA AAC CGC GCT GGT AAT GGA ATC CTC

      Lys Lys Thr Thr Trp Ala Asp Gln Thr Glu Arg Gly Leu Asn Asn Gln Asn Arg Gly Arg Lys Asn Gln Pro Lys (50)
(159) AAG AAG ACC ACT TGG GCT GAC CAA ACC GAG CGC GGG TTA AAT AAT CAA AAT AGA GGC AGA AAG AAT CAG CCC AAG

      Gln Thr Ala Thr Thr Gln Pro Asn Ser Gly Ser Val Val Pro His Tyr Ser Trp Phe Ser Gly Ile Thr Gln Phe (75)
(234) CAG ACT GCA ACT ACT CAA CCC AAT TCC GGG AGT GTG GTT CCC CAT TAC TCT TGG TTT TCG GGC ATT ACC CAA TTC

      Gln Lys Gly Lys Glu Phe Gln Phe Ala Gln Gly Gln Gly Val Pro Ile Ala Asn Gly Ile Pro Ala Ser Gln Gln (100)
(309) CAG AAG GGA AAA GAG TTT CAG TTT GCA CAA GGA CAA GGA GTG CCT ATT GCC AAT GGA ATC CCA GCT TCA CAG CAA

      Lys Gly Tyr Trp Tyr Arg His Asn Arg Arg Ser Phe Lys Thr Pro Asp Gly Gln Gln Lys Gln Leu Leu Pro Arg (125)
(384) AAG GGA TAT TGG TAC AGA CAC AAC CGA CGT TCC TTT AAA ACA CCT GAT GGC CAG CAG AAG CAG CTA CTG CCC AGA

      Trp Tyr Phe Tyr Tyr Leu Gly Thr Gly Pro Tyr Ala Gly Ala Glu Tyr Gly Asp Asp Ile Glu Gly Val Val Trp (150)
(459) TGG TAT TTT TAC TAT CTT GGA ACA GGG CCC TAT GCT GGC GCA GAG TAT GGC GAC GAT ATC GAA GGA GTT GTC TGG

      Val Ala Ser Gln Gln Ala Glu Thr Arg Thr Ser Ala Asp Ile Val Glu Arg Asp Pro Ser Ser His Glu Ala Ile (175)
(534) GTC GCA AGC CAA CAG GCC GAG ACT AGG ACC TCT GCC GAT ATT GTT GAA AGG GAC CCA AGT AGC CAT GAG GCT ATT

      Pro Thr Arg Phe Ala Pro Gly Thr Val Leu Pro Gln Gly Phe Tyr Val Glu Gly Ser Gly Arg Ser Ala Pro Ala (200)
(609) CCT ACT AGG TTT GCG CCC GGT ACG GTA TTG CCT CAA GGT TTT TAT GTT GAA GGC TCA GGA AGG TCT GCA CCT GCT

      Ser Arg Ser Gly Ser Arg Pro Gln Ser Arg Gly Pro Asn Asn Arg Ala Arg Ser Ser Ser Asn Gln Arg Gln Pro (225)
(684) AGT CGA TCT GGT TCG CGG CCA CAA TCC CGT GGG CCA AAT AAT CGC GCT AGA AGC AGT TCC AAC CAG CGC CAG CCT

      Ala Ser Thr Val Lys Pro Asp Met Ala Glu Glu Ile Ala Ala Leu Val Leu Ala Lys Leu Gly Lys Asp Ala Gly (250)
(759) GCC TCT ACT GTA AAA CCT GAT ATG GCC GAA GAA ATT GCT GCT CTT GTT TTG GCT AAG CTC GGT AAA GAT GCC GGC

      Gln Pro Lys Gln Val Thr Lys Gln Ser Ala Lys Glu Val Arg Gln Lys Ile Leu Asn Lys Pro Arg Gln Lys Arg (275)
(834) CAG CCT AAG CAA GTA ACA AAG CAA AGT GCC AAA GAA GTC AGG CAG AAA ATT TTA AAC AAG CCT CGT CAA AAG AGG

      Thr Pro Asn Lys Gln Cys Pro Val Gln Gln Cys Phe Gly Lys Arg Gly Pro Asn Gln Asn Phe Gly Gly Pro Glu (300)
(909) ACT CCA AAC AAG CAG TGC CCA GTG CAG CAG TGT TTT GGA AAG AGA GGC CCC AAT CAG AAT TTT GGA GGC CCT GAA

      Met Leu Lys Leu Gly Thr Ser Asp Pro Gln Phe Pro Ile Leu Ala Glu Leu Ala Pro Thr Ala Gly Ala Phe Phe (325)
(984) ATG TTA AAA CTT GGA ACT AGT GAT CCA CAG TTC CCC ATT CTT GCA GAG TTG GCC CCA ACA GCT GGT GCC TTC TTC

      Phe Gly Ser Lys Leu Glu Leu Val Lys Lys Asn Ser Gly Gly Ala Asp Gly Pro Thr Lys Asp Val Tyr Glu Leu (350)
(1059) TTT GGA TCT AAA TTA GAA TTG GTC AAA AAG AAC TCT GGT GGT GCT GAT GGA CCC ACC AAA GAT GTG TAT GAG CTG

      Gln Tyr Ser Gly Ala Val Arg Phe Asp Ser Thr Leu Pro Gly Phe Glu Thr Ile Met Lys Val Leu Asn Glu Asn (375)
(1134) CAA TAT TCA GGT GCA GTT AGA TTT GAT AGT ACT CTA CCT GGT TTT GAG ACT ATC ATG AAA GTG TTG AAT GAG AAT

      Leu Asn Ala Tyr Gln Asn Gln Asp Gly Gly Ala Asp Val Val Ser Pro Lys Pro Gln Arg Lys Arg Gly Thr Lys (400)
(1209) TTG AAT GCC TAC CAG AAT CAA GAT GGT GGT GCA GAT GTA GTG AGC CCT AAG CCT CAG AGA AAG AGA GGG ACA AAG

      Gln Lys Ala Gln Lys Asp Glu Val Asp Asn Val Ser Val Ala Lys Pro Lys Ser Ser Val Gln Arg Asn Val Ser (425)
(1284) CAA AAG GCT CAG AAA GAT GAA GTA GAT AAT GTA AGC GTT GCA AAG CCC AAA AGC TCT GTG CAG CGA AAT GTA AGT

      Arg Glu Leu Thr Pro Glu Asp Arg Ser Leu Leu Ala Gln Ile Leu Asp Asp Gly Val Val Pro Asp Gly Leu Glu (450)
(1359) AGA GAG TTA ACC CCT GAG GAT CGC AGC CTT CTG GCT CAG ATC CTA GAT GAT GGC GTA GTG CCA GAT GGG TTA GAA

      Asp Asp Ser Asn Val
(1434) GAT GAC TCT AAT GTG TAAAGAGAATGAATCCTATGTCGGCACTCGGTGGTAACCCCTCGCGAGAAAGTCGGGGATAGGACACTCTCTATCAGAAT

(1528) GGATGTCTTGCTGTCATAACAGATAGAGAAGGTTGTGGCAGACCCTGTATCAATTAGTTGAAAGAGATTGCAAAATAGAGAATGTGTGAGAGAAGTTAG

(1627) CAAGGTCCTACGTCTAACCATAAGAACGGCGATAGGCGCCCCCTGGGAAGAGCTCACATCAGGGTACTATTCCTGCAATGCCCTAGTAAATGAATGAAG

(1726) TTGATCATGGCCAATTGGAAGAATCACAAAAAAAAAAAAAAA
```

Fig. 3 Complete nucleotide sequence of the DNA copy of MHV-JHM mRNA7 (1767 nucleotides), including a 15 base long, terminal poly (A) tract. The predicted sequence of the encoded protein is also depicted. Underlined sequences show the RNase T_1-resistant oligonucleotides, found in the 5' non-coding sequence, that are considered in the discussion. The triple repeat of the pentamer AATCT is also indicated.

content within just 6 % of the coding sequence. Moreover, this is a basic region of the protein (Fig. 4). Another 4 serine residues occur in the five residues from ser-12 to ser-16.

The sequence of the N gene of JHM can also be compared with the N gene from MHV-A59 (J. Armstrong, this volume). The nucleotide sequences show 94 % overall homology within the coding sequences and this is reflected within the predicted amino acid sequences (93 % homology). Although the homology is high overall, it is not constant throughout the length of the coding sequence. Between nucleotides 497-569 the sequence homology falls to 63 % (16 out of the 24 amino acids are conserved) and in a sequence of 23 bases (nucleotides 1271-1293) only 9 bases are common to both strains (39 % homology). Although none of the 8 amino acids are conserved the region retains its basic nature in both strains. In addition to these differences, the JHM sequence has an extra glutamine codon at position 1227.

The 5' non-coding region of the JHM sequence has a terminal RNase T_1-resistant oligonucleotide TATAAG, which is very similar to the sequence found at the 5' end of MHV-A59 RNAs by Lai et al.[14] (cap-NUAAG). The 301 bases of 3' non-coding sequence are followed by a 15 base poly (A) tract. Thus it is probable that the sequence represents a full copy of JHM mRNA7.

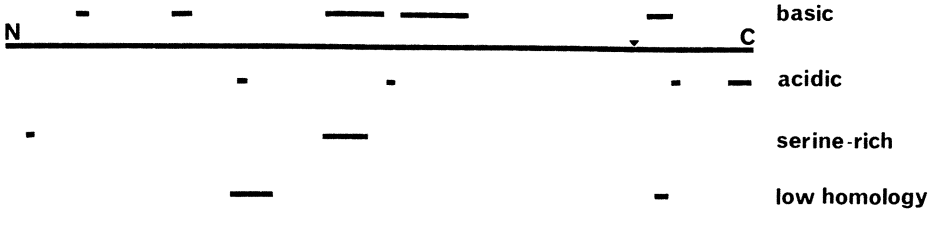

Fig. 4 Illustration to show features of the coding sequence of JHM mRNA7: basic and acidic regions, regions of high serine content and the two regions where the homology with the A59 sequence (J. Armstrong, this volume) is notably reduced. The arrowhead indicates the position where the extra glutamine codon is found in the JHM sequence.

Within the 5' non-coding sequence, two large RNase T_1-resistant oligonucleotides can be found. Examination reveals that their base compositions are similar to those of oligonucleotides 10 and 19 of A59[14]. It has been suggested[15] that oligonucleotide 10 (equivalent to the JHM oligonucleotide at position 26-50) is contained within a leader sequence derived from the 5' end of the genome and that oligonucleotide 19 (equivalent to the JHM oligonucleotide at position 54-82) is formed by fusion of the leader to the body of mRNA7[14,15]. The relative positions of the two oligonucleotides in the 5' non-coding sequence of JHM mRNA7 are consistent with such a proposal. This matter is considered in more detail by Spaan et al., in this volume. It should be noted, however, that the JHM equivalent of oligonucleotide 19 contains a triple repeat of the pentamer AATCT (or ATCTA), a striking feature which probably plays a significant role in the production of subgenomic mRNAs. The triple repeat was seen in both cDNA clones and so is unlikely to represent a sequence found in only a minor population of mRNA7. Comparison of the 3' non-coding sequences of JHM and A59 reveals that the two are highly conserved in this region (98 % homology). Such a high conservation of sequence in a non-coding region is possibly due to the fact that it must contain the binding site for the RNA polymerase function which synthesises negative-strand template.

Consideration of the protein sequence of the JHM nucleocapsid protein has shown several interesting structural features. It is to be hoped that studies on the interaction of N protein with genomic RNA and with other viral proteins will elucidate the functional role of these features. As might have been expected, the sequences encoding the N protein of MHV have been well conserved even between two strains that show differences in pathogenicity. It will be interesting to determine the level of homology for genes more directly involved in the interaction between the virus and the host.

ACKNOWLEDGEMENTS

We thank Helga Kriesinger for typing the manuscript. This work was supported by the Deutsche Forschungsgmeinschaft.

REFERENCES

1. Siddell, S.G., Wege, H., Barthel, A. and ter Meulen, V. (1980). J. Virol. 33, 10-17.
2. Siddell, S., Wege, H., Barthel, A. and ter Meulen, V. (1981). J. Gen. Virol. 53, 145-155.
3. Stohlman, S.A. and Lai, M.M.C. (1979). J. Virol. 32, 672-675.
4. Siddell, S.G., Barthel, A. and ter Meulen, V. (1981). J. Gen. Virol. 52, 235-243.

5. Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schütz, G., (1983). In Methods in Enzymology, eds. Wu, R., Grossman, L. and Moldave, K. (Academic Press, New York), Vol. 100, in press.
6. Dagert, M. and Ehrlich, S.D. (1979). Gene 6, 23-28.
7. Holmes, D.S. and Quigley, M. (1981). Anal. Biochem. 114, 193-197.
8. Messing, J. and Vierira, J. (1982). Gene 19, 269-276.
9. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA 74, 5463-5467.
10. Hu, N. and Messing, J. (1982). Gene 17, 271-277.
11. Paterson, B.M., Roberts, B.E. and Kuff, E.L. (1977). Proc. Nat. Acad. Sci. USA 74, 4370-4374.
12. Laemmli, U.K. (1970). Nature (London) 227, 680-685.
13. Garoff, H., Frischauf, A., Simons, K., Lehrach, H. and Delius, H. (1980). Proc. Natl. Acad. Sci. USA 77, 6376-6380.
14. Lai, M.M.C., Patton, C.D. and Stohlman, S.A. (1982). J. Virol. 41, 557-565.
15. Spaan, W.J.M., Rottier, P.J.M., Horzinek, M.C. and van der Zeijst, B.A.M. (1982). J. Virol. 42, 432-439.