



US010385336B2

(12) **United States Patent**
Siksnys et al.

(10) **Patent No.:** **US 10,385,336 B2**
(45) **Date of Patent:** ***Aug. 20, 2019**

(54) **PROGRAMMABLE RNA SHREDDING BY THE TYPE III-A CRISPR-CAS SYSTEM OF *STREPTOCOCCUS THERMOPHILUS***

(71) Applicant: **VILNIUS UNIVERSITY**, Vilnius (LT)

(72) Inventors: **Virginijus Siksnys**, Vilnius (LT); **Migle Kazlauskienė**, Vilnius (LT); **Gintautas Tamulaitis**, Vilnius (LT)

(73) Assignee: **VILNIUS UNIVERSITY**, Vilnius (LT)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 24 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **15/448,818**

(22) Filed: **Mar. 3, 2017**

(65) **Prior Publication Data**

US 2017/0198286 A1 Jul. 13, 2017

Related U.S. Application Data

(63) Continuation-in-part of application No. PCT/IB2015/056756, filed on Sep. 4, 2015.

(60) Provisional application No. 62/046,384, filed on Sep. 5, 2014.

(51) **Int. Cl.**

C12N 15/11 (2006.01)

C12N 15/113 (2010.01)

C12N 9/22 (2006.01)

C07K 14/315 (2006.01)

(52) **U.S. Cl.**

CPC *C12N 15/111* (2013.01); *C07K 14/315*

(2013.01); *C12N 9/22* (2013.01); *C12N 15/113*

(2013.01); *C12Y 301/27* (2013.01); *C12N*

2310/10 (2013.01); *C12N 2310/20* (2017.05)

(58) **Field of Classification Search**

None

See application file for complete search history.

(56) **References Cited**

PUBLICATIONS

Huang et al.; Germ-Line Transmission of a Myocardium-Specific GFP Transgene Reveals Critical Regulatory Elements in the Cardiac Myosin Light Chain 2 Promoter of Zebrafish; *Developmental Dynamics*; vol. 228; 2003; p. 30-40.

Tamulaitis et al.; Type III CRISPR-Cas Immunity: Major Differences Brushed Aside; *Trends in Microbiology*; vol. 25; 2016; 13 pages.

Burgess et al.; Questions about NgAgo; *Protein Cell*; vol. 7(12); 2016; p. 913-915.

Qin et al.; NgAgo-based fabp11a gene knockdown causes eye developmental defects in zebrafish; *Cell Research*; vol. 26; 2016; p. 1349-1352.

Lee et al.; Failure to detect DNA-guided genome editing using *Natronobacterium gregoryi* Argonaute; *Nature Biotechnology*; 2016; 2 pages.

Kazlauskienė et al.; Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition; *Molecular Cell*; vol. 62; 2016; 30 pages w/ Supplementary Information.

Kaya et al.; A bacterial Argonaute with noncanonical guide RNA specificity; *PNAS*; vol. 113 No. 15; Apr. 2016; p. 4057-4062 w/ Supplementary Information.

Gao et al.; DNA-guided genome editing using the *Natronobacterium gregoryi* Argonaute; *Nature Biotechnology*; 2016; 7 pages.

East-Seletsky et al.; Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection; *Nature*; 2016; 17 pages.

Abudayyeh et al.; C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector; *Science*; Jun. 2016; 17 pages.

Stainier et al.; Making Sense of Anti-Sense Data; *Developmental Cell*; vol. 32; Jan. 2015; p. 7-8.

Shmakov et al.; Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems; *Molecular Cell*; vol. 60, 2015; 14 pages.

Samai et al.; Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity; *Cell*; vol. 161; 2015; 19 pages.

Rossi et al.; Genetic compensation induced by deleterious mutations but not gene knockdowns; *Nature*; 2015; 15 pages.

Kok et al.; Reverse genetic screening reveals poor correlation between Morpholino-induced and mutant phenotypes in zebrafish; *Dev Cell*; vol. 32(1); Jan. 2015; 27 pages.

Jung et al.; Crystal Structure of the Csm1 Subunit of the Csm Complex and Its Single-Stranded DNA-Specific Nuclease Activity; *Structure*; vol. 23; 2015; 10 pages.

Tamulaitis et al.; Programmable RNA Shredding by the Type III-A CRISPR-Cas System of *Streptococcus thermophilus*; *Molecular Cell*; vol. 56; Nov. 2014; 40 pages w/ Supplemental Information.

Staals et al.; RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of *Thermus thermophilus*; *Molecular Cell*; vol. 56; Nov. 2014; 42 pages w/ Supplemental Information.

Goldberg et al.; Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting; *Nature*; vol. 514; Oct. 2014; 24 pages.

Spilman et al. Structure of an RNA Silencing Complex of the CRISPR-Cas Immune System; *Molecular Cell*; vol. 52; Oct. 2013; p. 146-152.

Witzel et al.; The LIM Protein Ajuba Restricts the Second Heart Field Progenitor Pool by Regulating Isl1 Activity; *Developmental Cell*; vol. 23; Jul. 2012; p. 58-70.

De Rienzo et al.; Efficient shRNA-Mediated Inhibition of Gene Expression in Zebrafish; *Zebrafish*; vol. 9 No. 3; 2012; p. 97-107.

Bill et al.; A Primer for Morpholino Use in Zebrafish; *Zebrafish*; vol. 6 No. 1; 2009; p. 69-77.

(Continued)

Primary Examiner — Robert M Kelly

(74) *Attorney, Agent, or Firm* — Young & Thompson

(57) **ABSTRACT**

A Type III-A CRISPR-Cas (StCsm) complex of *Streptococcus thermophilus* comprising crRNA, Csm4, and Csm3 and use for cleavage of RNA bearing a nucleotide sequence complementary to the crRNA, in vitro or in vivo. Methods for site-specific cleavage/shredding of a target RNA molecule using an RNA-guided RNA endonuclease comprising a minimal complex of crRNA, Csm4, and Csm3, and methods of RNA knock-down and RNA knock-out are disclosed.

14 Claims, 45 Drawing Sheets

Specification includes a Sequence Listing.

(56)

References Cited

PUBLICATIONS

- White et al.; Transparent Adult Zebrafish as a Tool for In Vivo Transplantation Analysis; *Cell Stem Cell*; vol. 2; Feb. 2008; p. 183-189.
- Kim et al.; Real-time imaging of mitochondria in transgenic zebrafish expressing mitochondrially targeted GFP; *Biotechniques*; vol. 45 No. 3; 2008; p. 331-334.
- Eisen et al.; Controlling morpholino experiments: don't stop making antisense; *Development*; vol. 135; 2008; p. 1735-1743.
- Liu et al.; Efficient RNA interference in zebrafish embryos using siRNA synthesized with SP6 RNA polymerase; *Development Growth Diff*; vol. 47; 2005; p. 323-331.
- Lawson et al.; In Vivo Imaging of Embryonic Vascular Development Using Transgenic Zebrafish; *Developmental Biology*; vol. 248; 2002; p. 307-318.
- Knaut et al.; An Evolutionary Conserved Region in the vasa 3'UTR Targets RNA Translation to the Germ Cells in the Zebrafish; *Current Biology*; vol. 12; Mar. 2002; p. 454-466.
- Kovel et al.; Expression of a vas: EGFP transgene in primordial germ cells of the Zebrafish; *Mechanisms of Development*; vol. 116; 2002; p. 141-150.
- Heasman; Morpholino Oligos: Making Sense of Antisense?; *Developmental Biology*; vol. 243; 2002; p. 209-214.
- Nasevicius et al.; Effective targeted gene 'knockdown' in zebrafish; *Nature Genetics*; vol. 26; Oct. 2000; p. 216-220.
- Knaut et al.; Zebrafish vasa RNA but Not Its Protein Is a Component of the Germ Plasm and Segregates Asymmetrically before Germline Specification; *The Journal of Cell Biology*; vol. 149 No. 4; May 2000; p. 875-888.
- Braat et al.; Vasa protein expression and localization in the zebrafish; *Mechanisms of Development*; vol. 95; 2000; p. 271-274.
- Neuhauss et al.; Genetic Disorders of Vision Revealed by a Behavioral Screen of 400 Essential Loci in Zebrafish; *The Journal of Neuroscience*; vol. 19(19); Oct. 1999; p. 8603-8615.
- Lister et al.; nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate; *Development*; vol. 126; 1999; p. 3757-3767.
- Huang et al.; Characterising and Predicting Haploinsufficiency in the Human Genome; *PLoS Genetics*; vol. 6 Issue 10; Oct. 2010; 11 pages.
- Braat et al.; Characterization of Zebrafish Primordial Germ Cells: Morphology and Early Distribution of vasa RNA; *Developmental Dynamics*; vol. 216; 1999; p. 153-167.
- Altschul et al.; Gapped BLAST and PSI-BLAST: a new generation of protein database search programs; *Nucleic Acids Research*; vol. 25 No. 17; 1997; p. 3389-3402.
- Chen et al.; Zebrafish tinman homolog demarcates the heart field and initiates myocardial differentiation; *Development*; vol. 122; 1996; p. 3809-3816.
- Marraffini et al.; Self versus non-self discrimination during CRISPR RNA-directed immunity; *Nature*; 2010; 18 pages w/ Supplemental Information.
- Marraffini et al.; CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea; *Nature Review Genetics*; vol. 11; Mar. 2010; p. 181-190.
- Altschul et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25 (1997) 3389-3402.
- Ashkenazy et al. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38 (2010) W529-533.
- Baker et al. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98 (2001) 10037-10041.
- Brouns et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321 (2008) 960-964.
- Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & development* 22, 3489-3496.
- Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013). A novel interference mechanism by a type IIIb CRISPR-Cmr module in *Sulfolobus*. *Molecular microbiology* 87, 1088-1099.
- Dickman, M., and Hornby, D. (2006). Enrichment and analysis of RNA centered on ion pair reverse phase methodology. *RNA* 12, 691-696.
- Fischer, H., de Oliveira Neto, M., Napolitano, H., Polikarpov, I., and Craievich, A. (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J Appl Crystallogr* 43, 101-109.
- Frickey, T., and Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20, 3702-3704.
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109, E2579-2586.
- Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945-956.
- Hatoum-Aslan, A., Maniv, I., and Marraffini, L.A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc Natl Acad Sci U S A* 108, 21218-21222.
- Hatoum-Aslan, A., Maniv, I., Samai, P., and Marraffini, L.A. (2014). Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *J Bacteriol* 196, 310-317.
- Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., and Marraffini, L.A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *The Journal of biological chemistry* 288, 27888-27897.
- Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167-170.
- Hrle, A., Su, A.A., Ebert, J., Benda, C., Randau, L., and Conti, E. (2013). Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. *RNA biology* 10, 1670-1678.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30, 3059-3066.
- Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 33, 33-41.
- Kraulis, P.J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24, 945-950.
- Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., et al. (2006). Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci U S A* 103, 15611-15616.
- Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology direct* 6, 38.
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011b). Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology* 9, 467-477.
- Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843-1845.
- Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature reviews. Genetics* 11, 181-190.
- Millen, A.M., Horvath, P., Boyaval, P., and Romero, D.A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. *PLoS One* 7, e51663.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733-740.
- Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., and

(56)

References Cited

PUBLICATIONS

- Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45, 342-350.
- Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L., and White, M.F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell* 52, 124-134.
- Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J* 32, 385-394.
- Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33, W244-248.
- Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol Cell* 52, 146-152.
- Staals, R.H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K., et al. (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* 52, 135-145.
- Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). Dna interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62-67.
- Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25, 495-503.
- Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal* 2879-2886.
- Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A* 111, 9798-9803.
- Terns, R.M., and Terns, M.P. (2014). CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends in genetics : TIG* 30, 111-118.
- Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of its ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36, 860-864.
- Waghmare, S.P., Pousinis, P., Homby, D.P., and Dickman, M.J. (2009). Studying the mechanism of RNA separations using RNA chromatography and its application in the analysis of ribosomal RNA and RNA:RNA interactions. *Journal of chromatography. A* 1216, 1377-1382.
- Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* 46, 595-605.
- Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477, 486-489.
- Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V., et al. (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 45, 303-313.
- Zheng, L., Baumann, U., and Reymond, J.L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* 32, e115.
- Merritt, E.A., and Bacon, D.J. (1997). Raster3D: Photorealistic molecular graphics. *Methods in Enzymology*, vol. 277, pp. 505-524.
- Olsthoorn, R., and van Duin, J. (2011). Bacteriophages with ssRNA. In *eLS* (John Wiley & Sons Ltd, Chichester), pp. 1-8.
- Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

Fig. 1A

DGCC8004 CRISPR2-cas

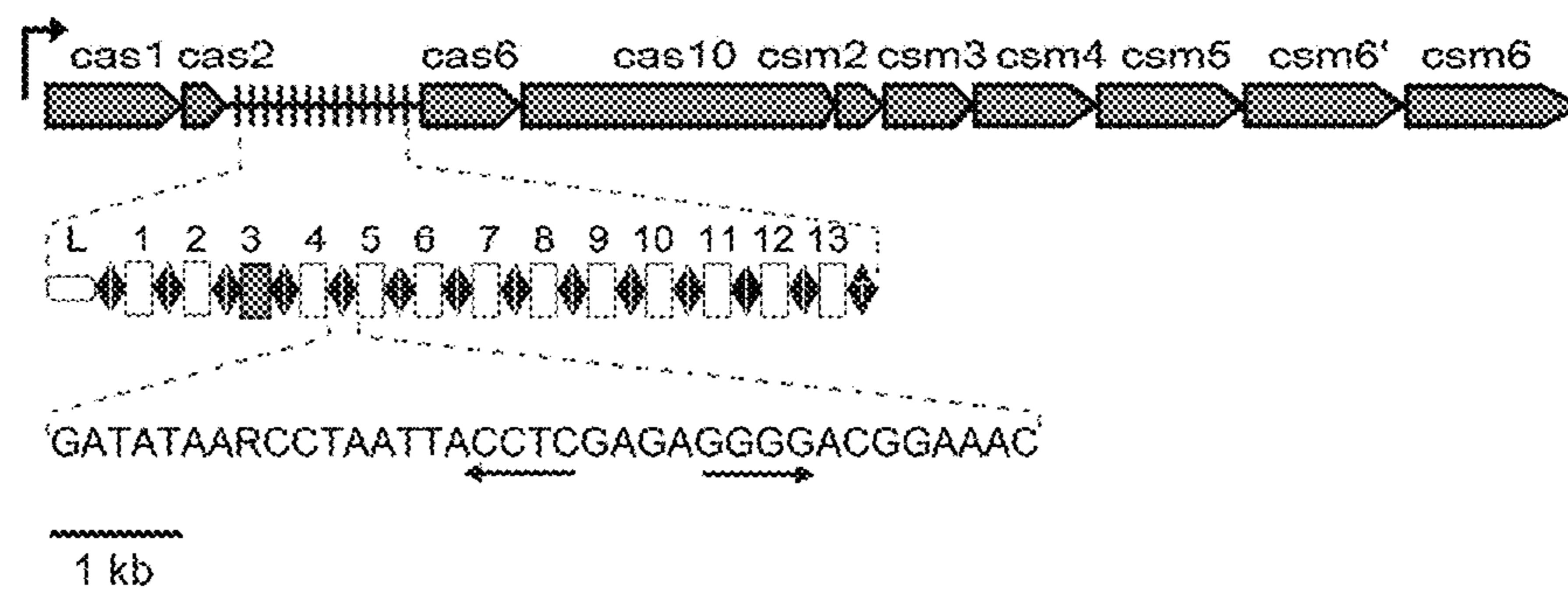


Fig. 1B

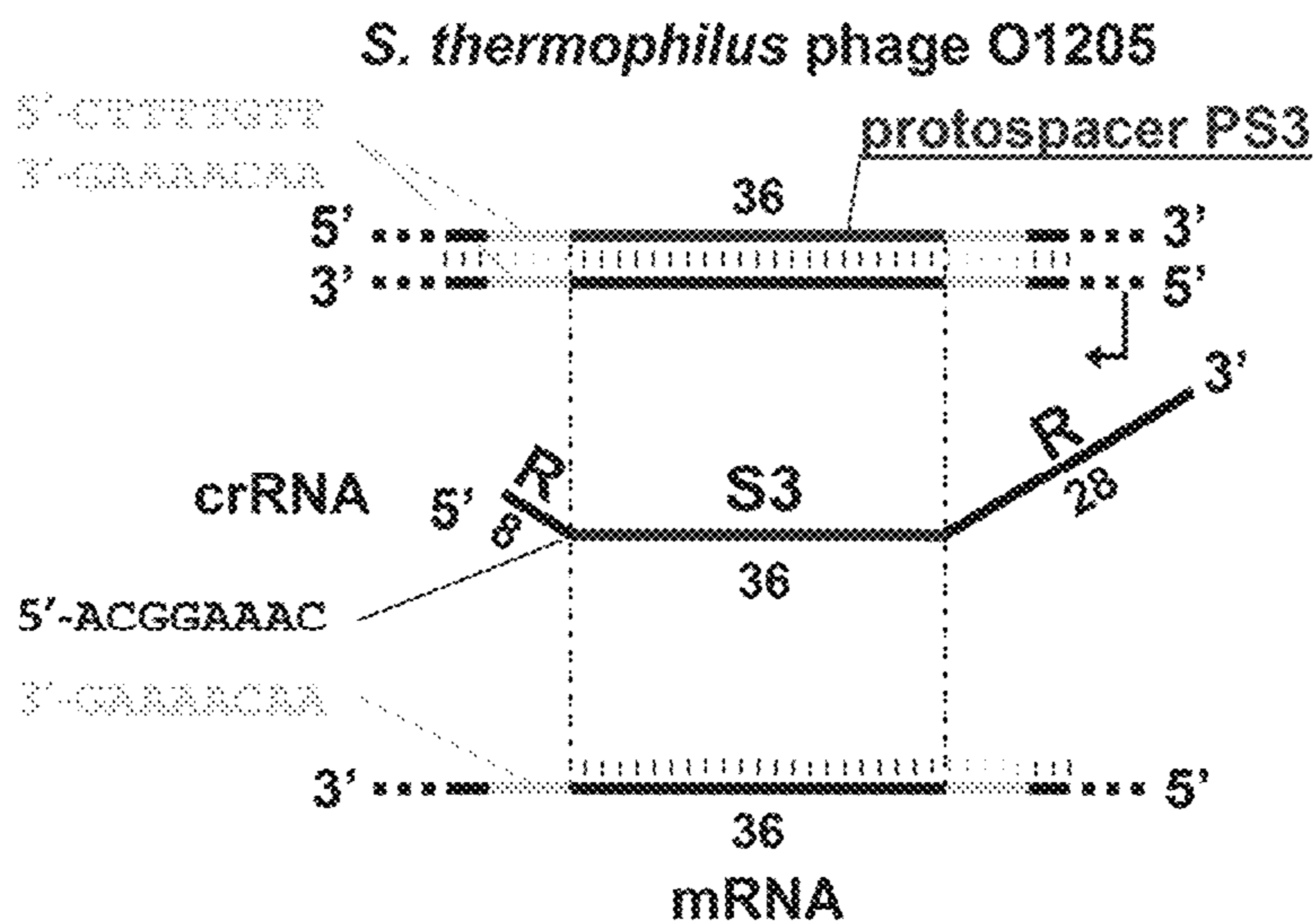


Fig. 1C

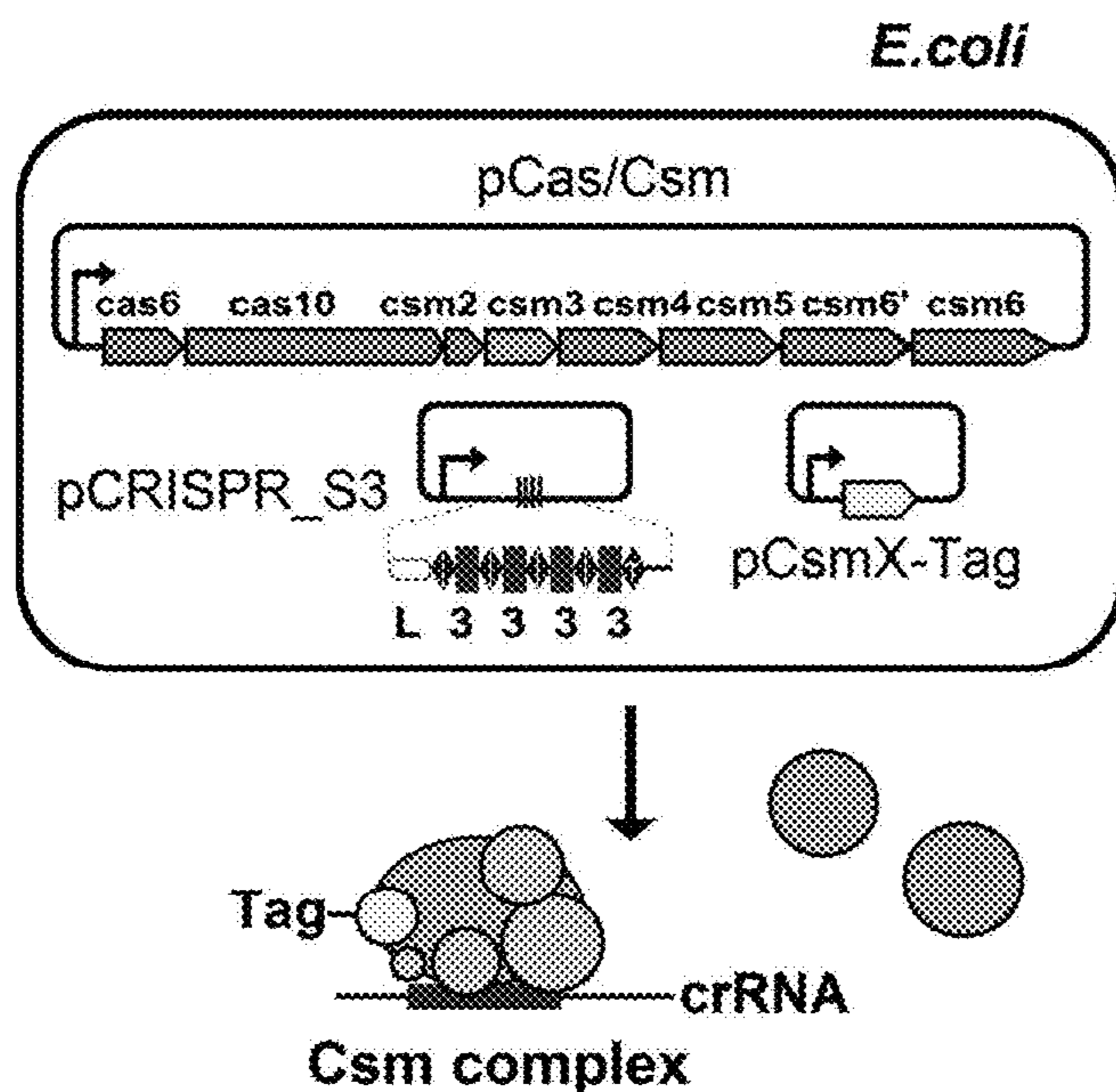


Fig. 1D

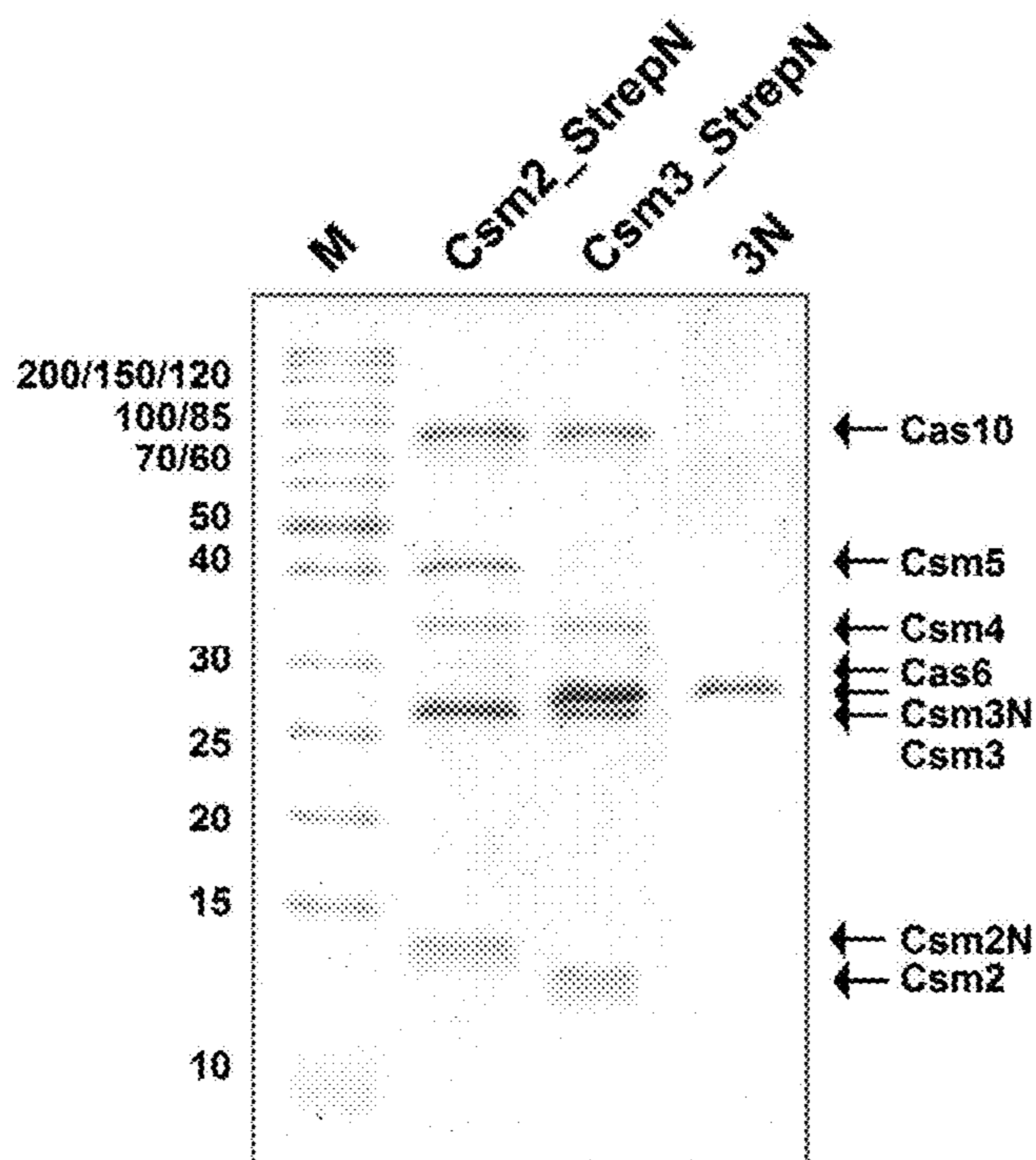
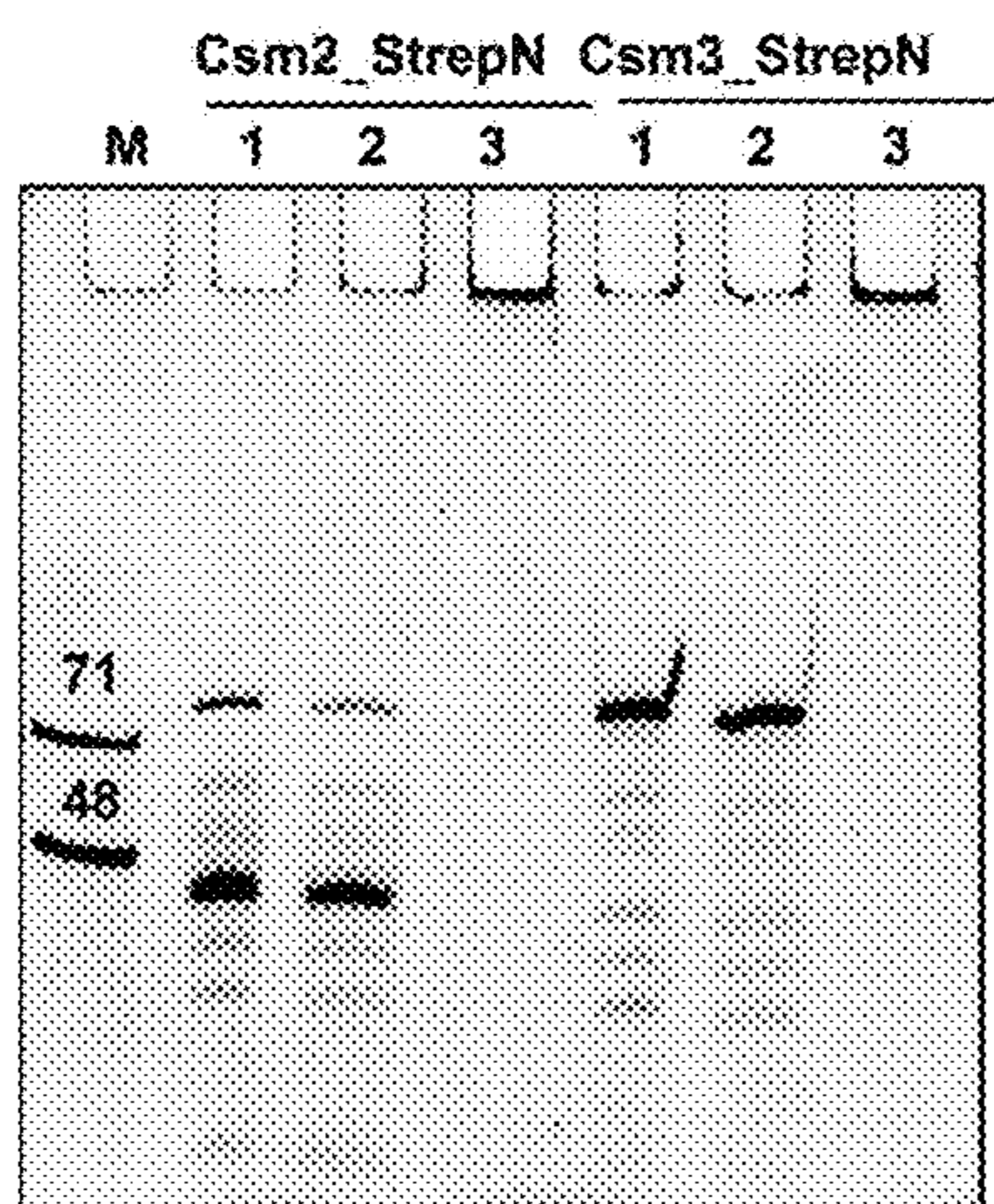
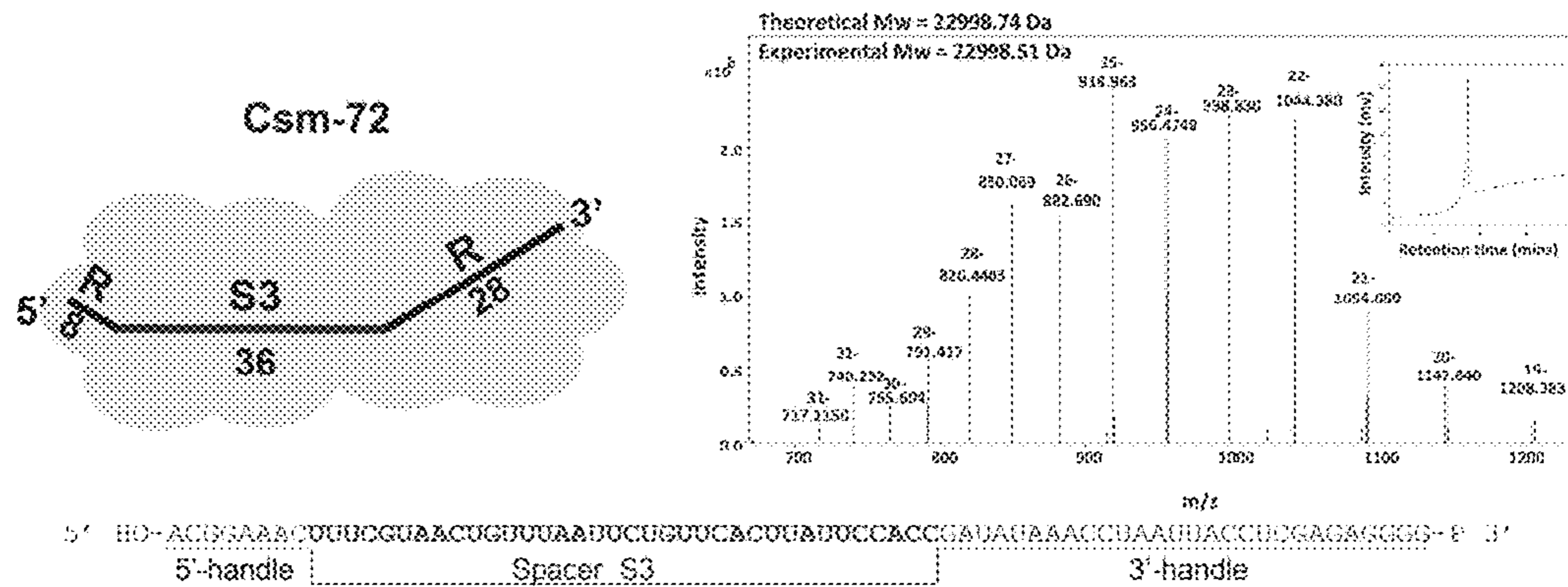


FIG. 1E



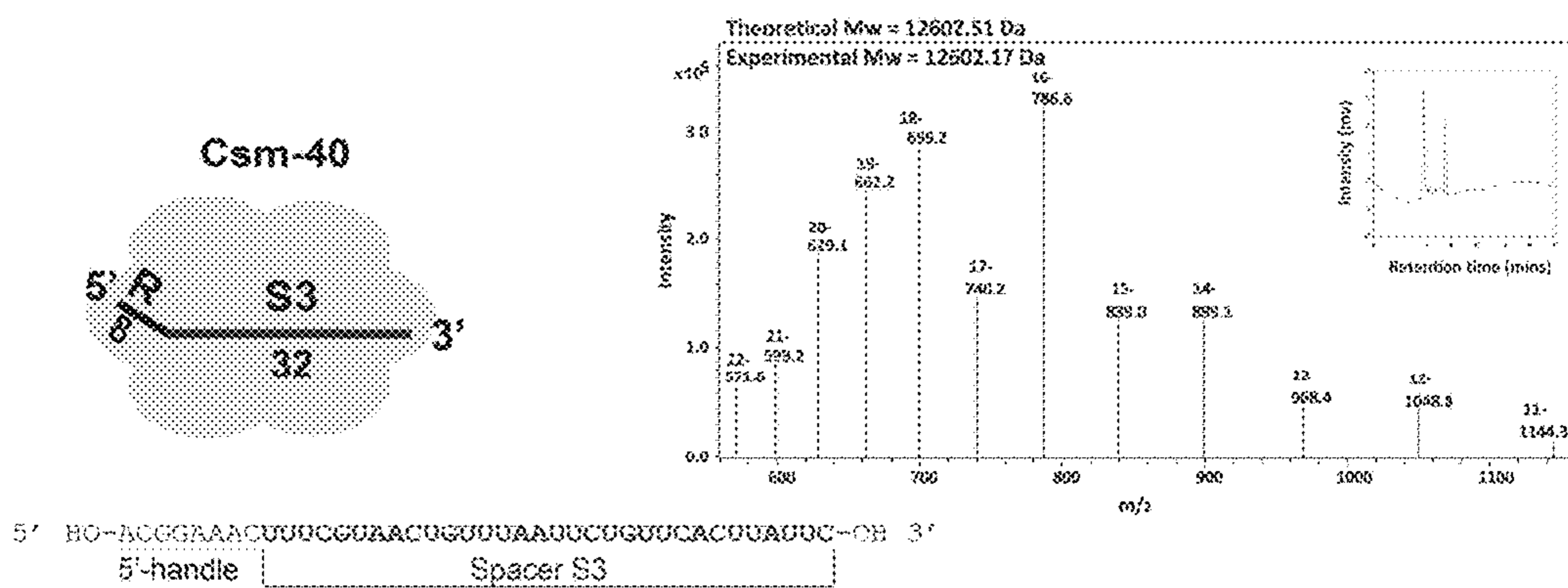
1 – NA from Csm complex
 2 – NA + DNaseI
 3 – NA + RNaseI

FIG. 1F



SEQ ID NO: 13

FIG. 1G



SEQ ID NO: 14

FIG. 1H

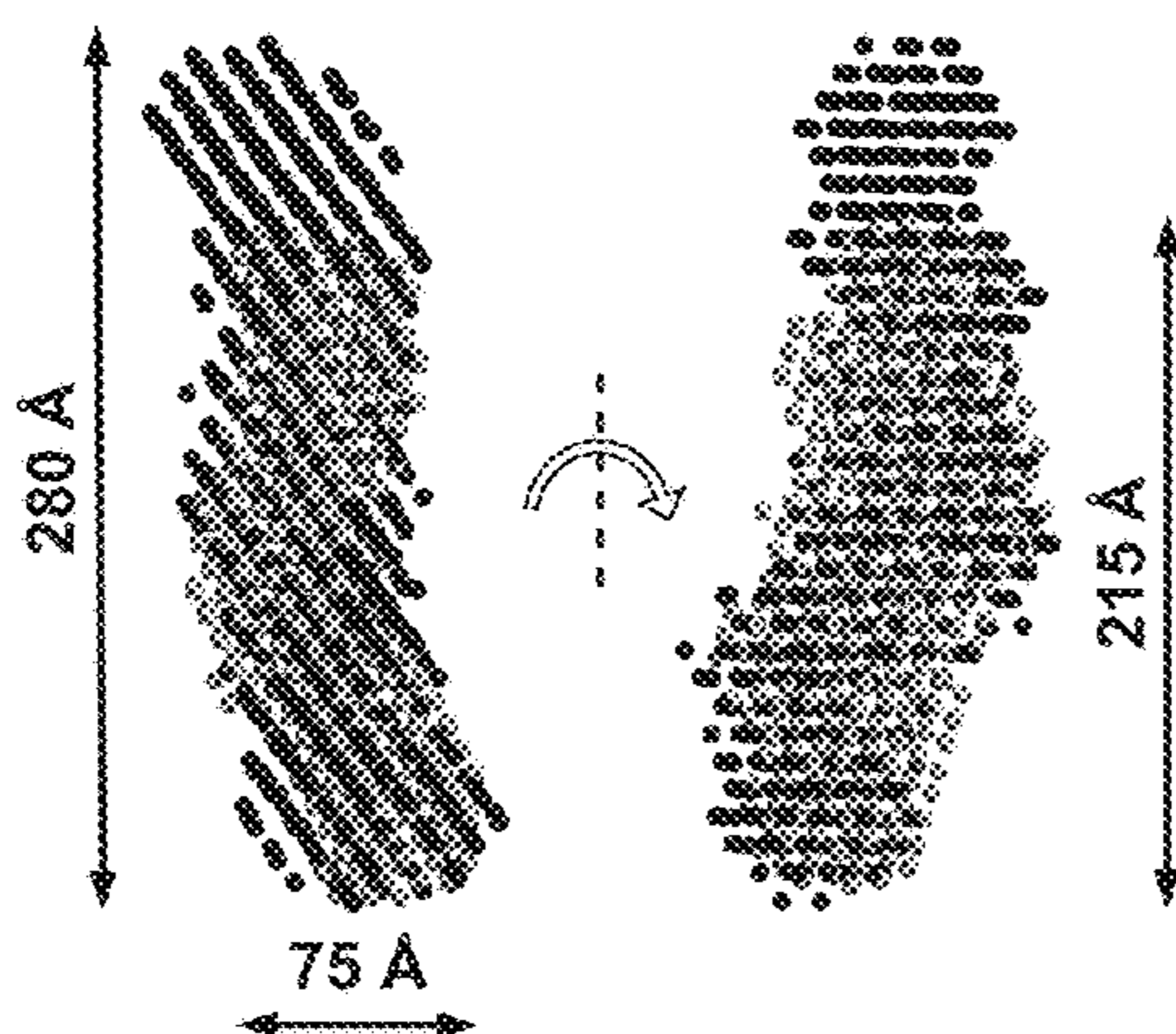


FIG. 2A

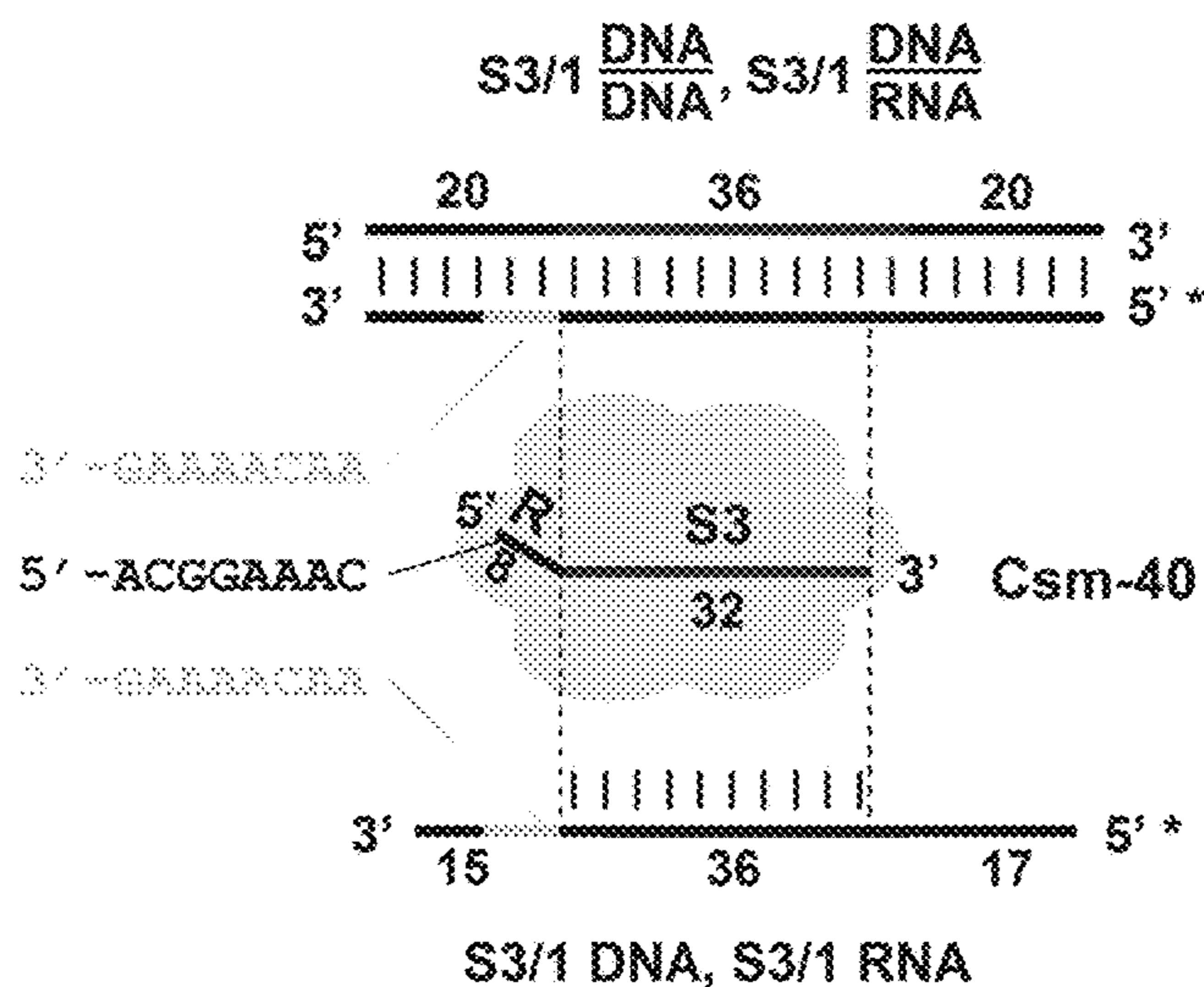


FIG. 2B

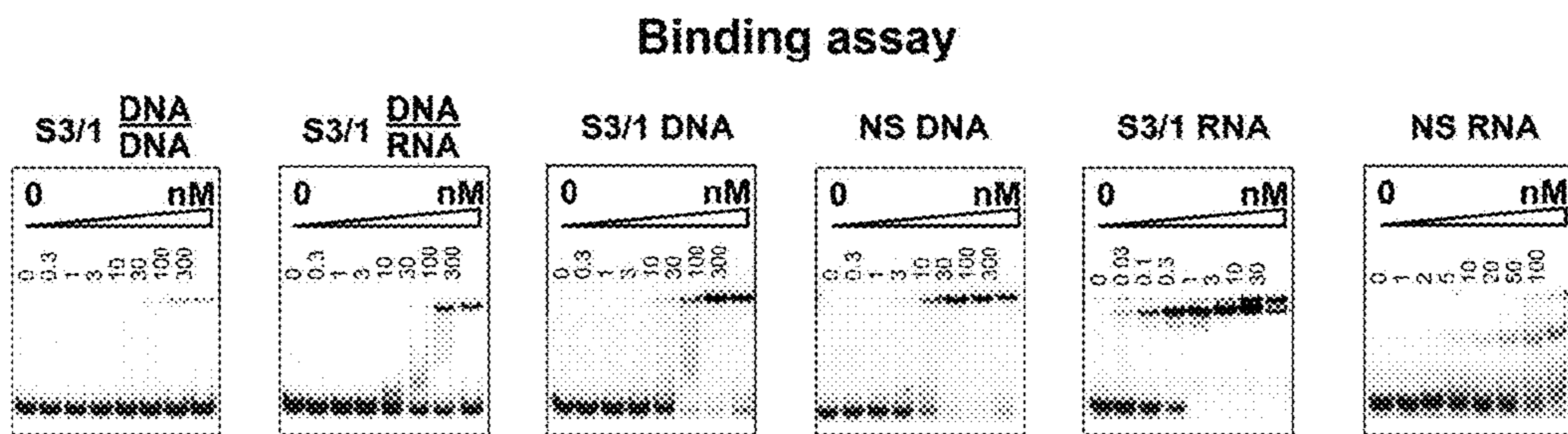


FIG. 2C

Binding competition assay

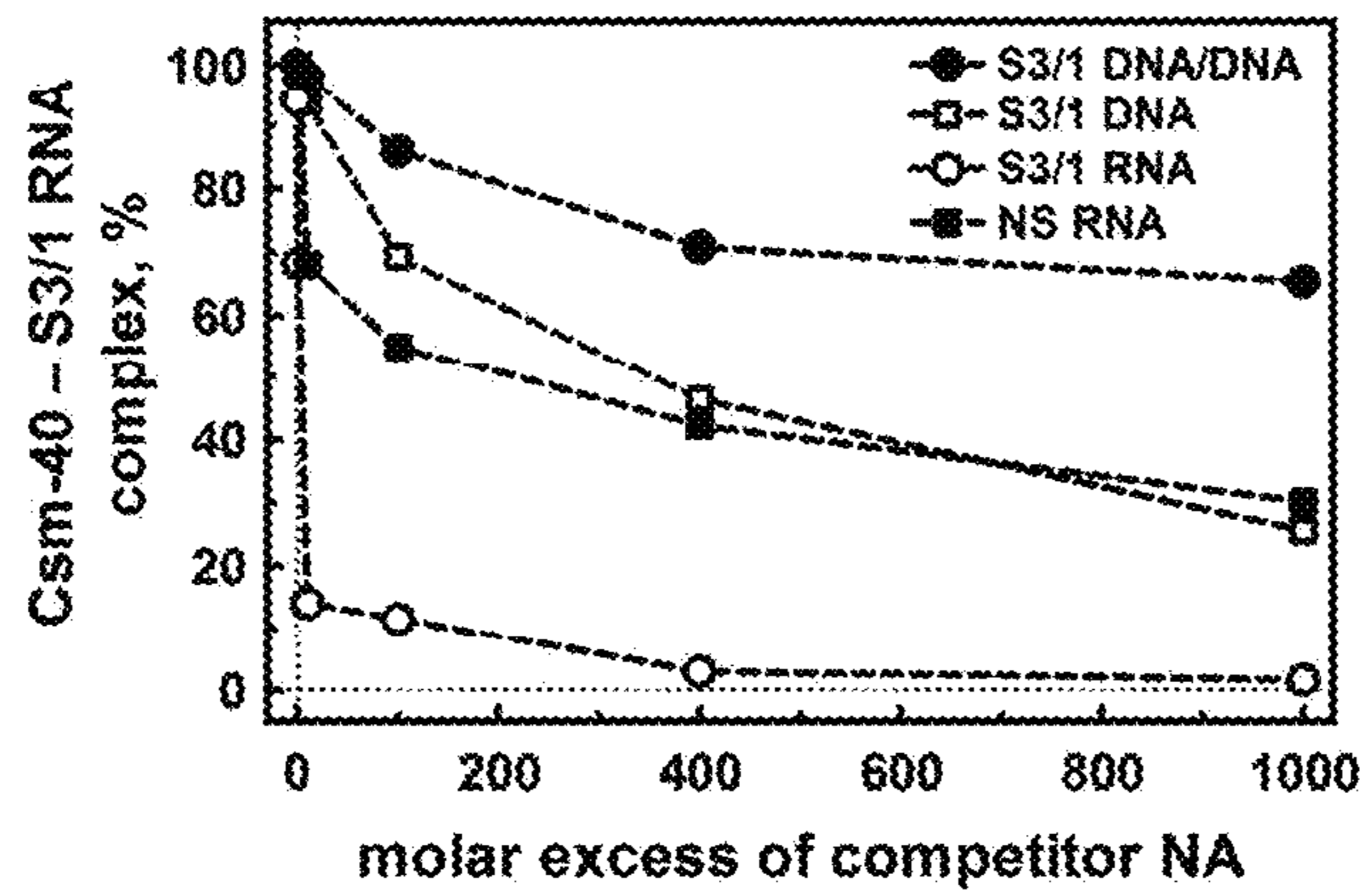


FIG. 2D

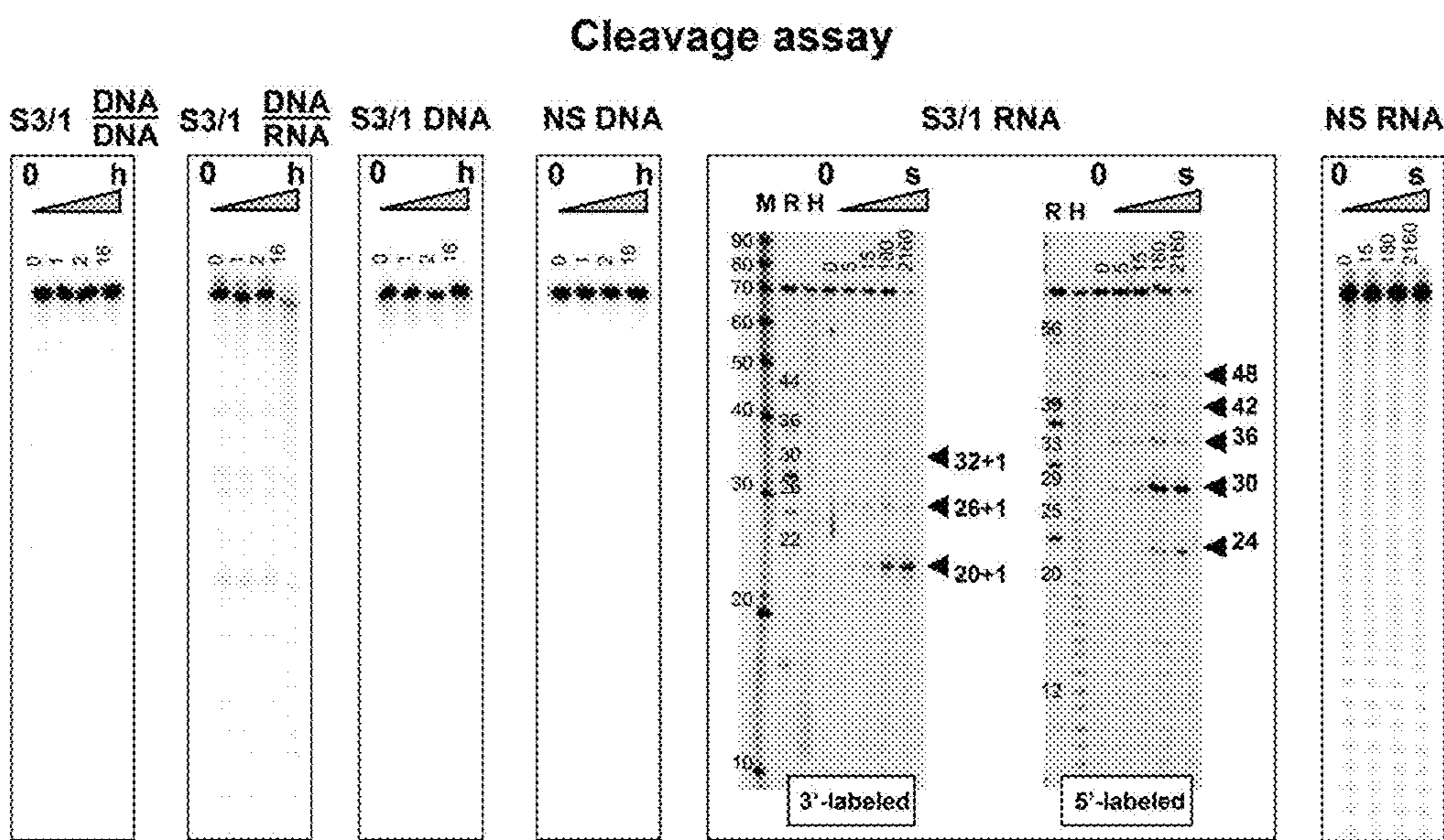


FIG. 2E

SEQ ID NO: 14 and SEQ ID NO: 22 and SEQ ID NO: 23, respectively

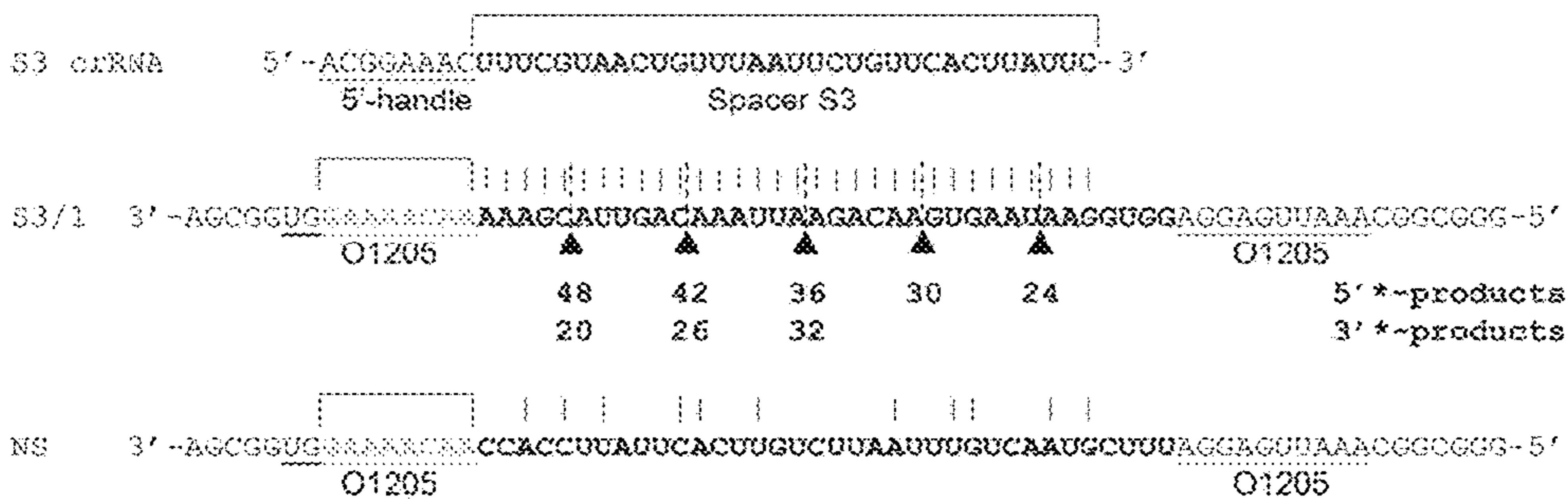


FIG. 3A

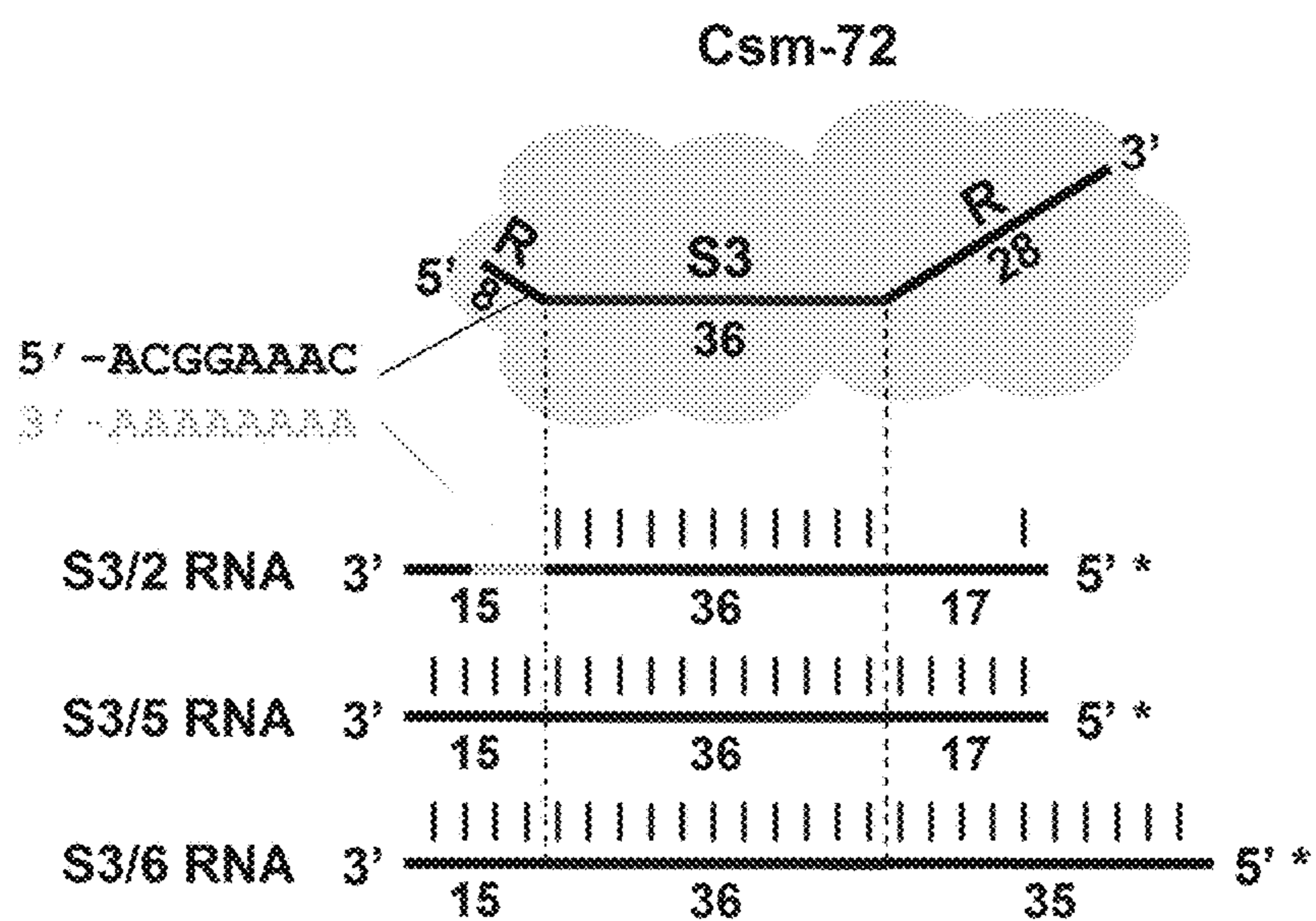


FIG. 4B

Binding assay

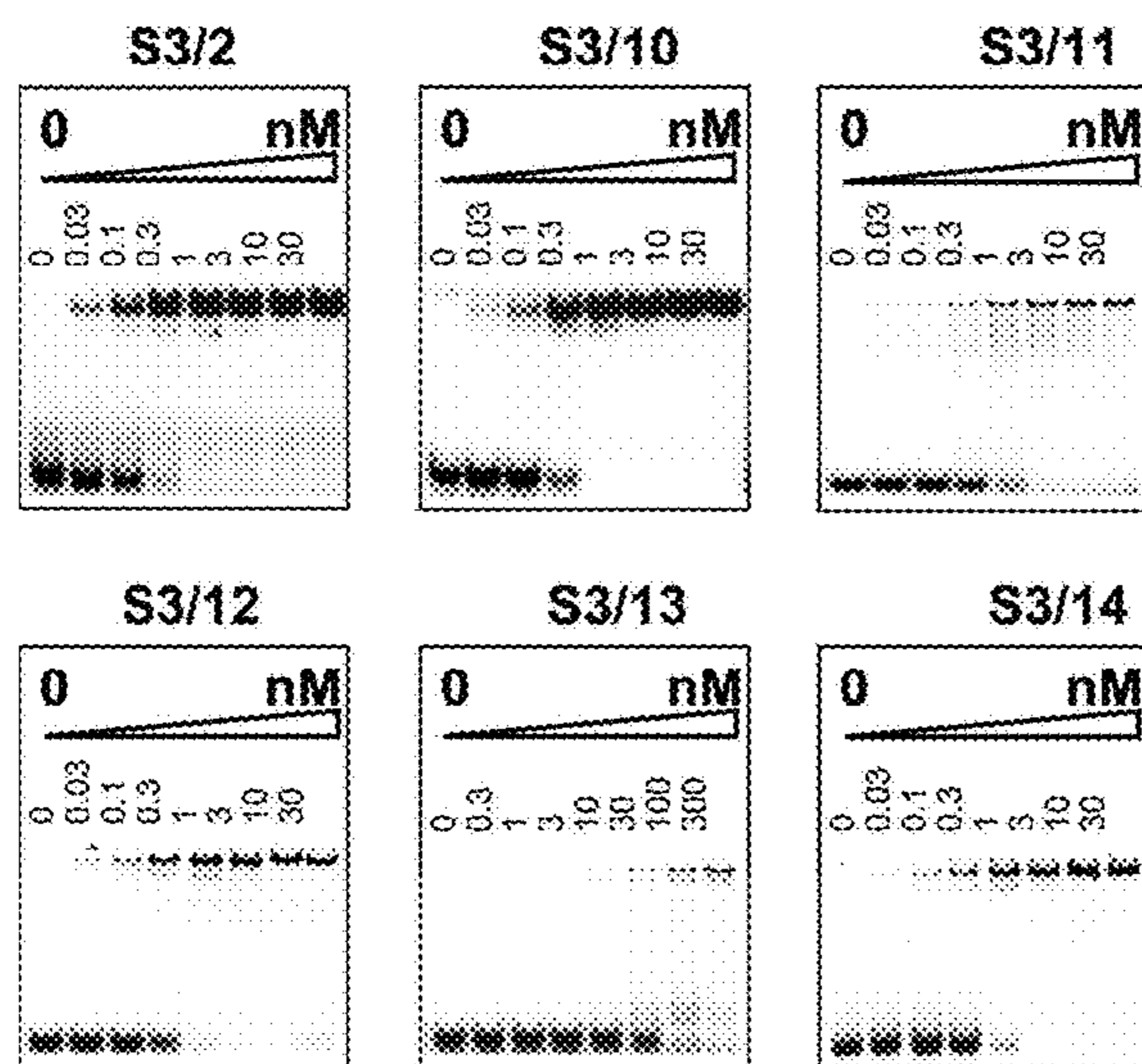


FIG. 4C

Cleavage assay

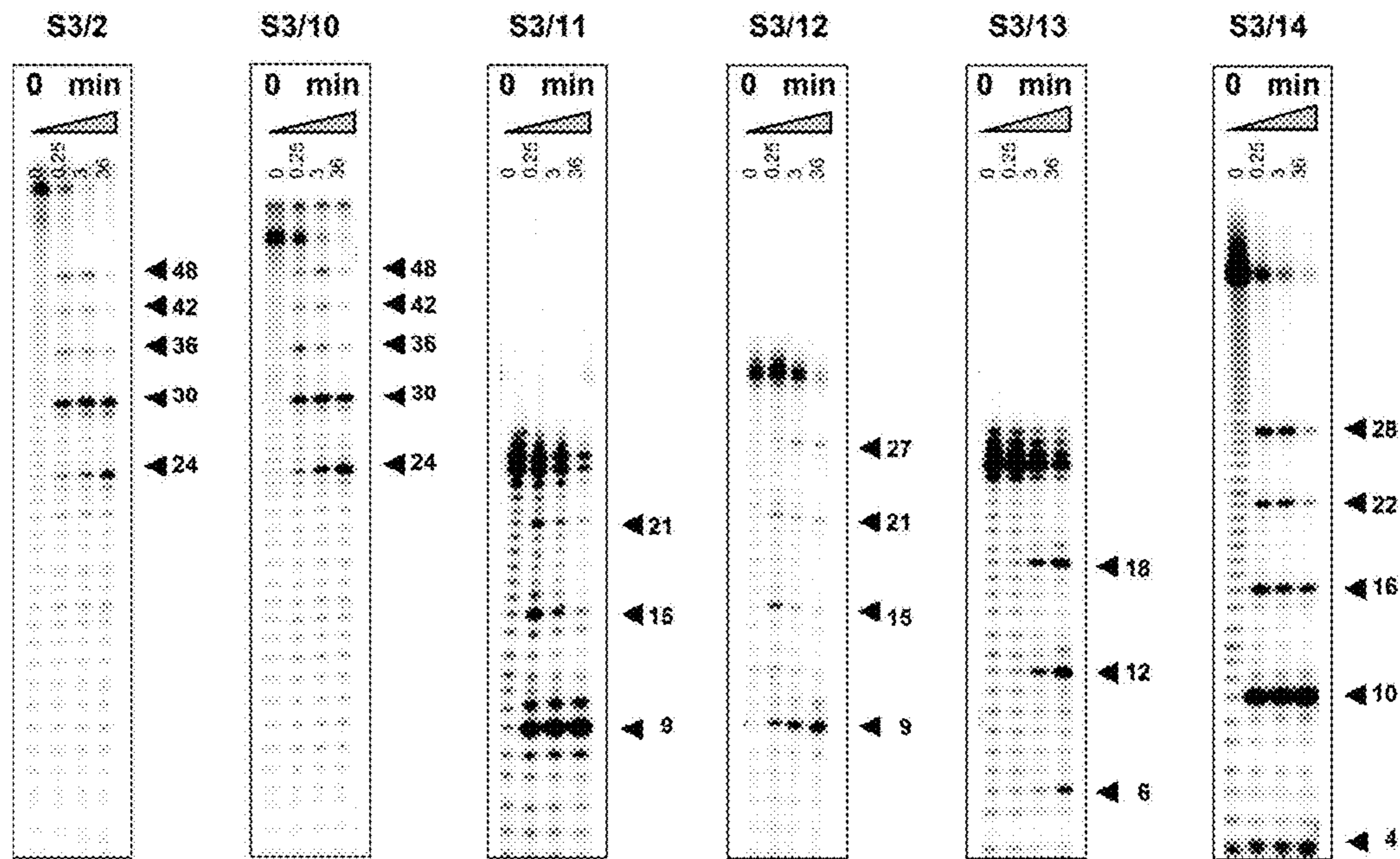


FIG. 4D SEQ ID NOS: 14, 24, 32, 33, 34, and 35

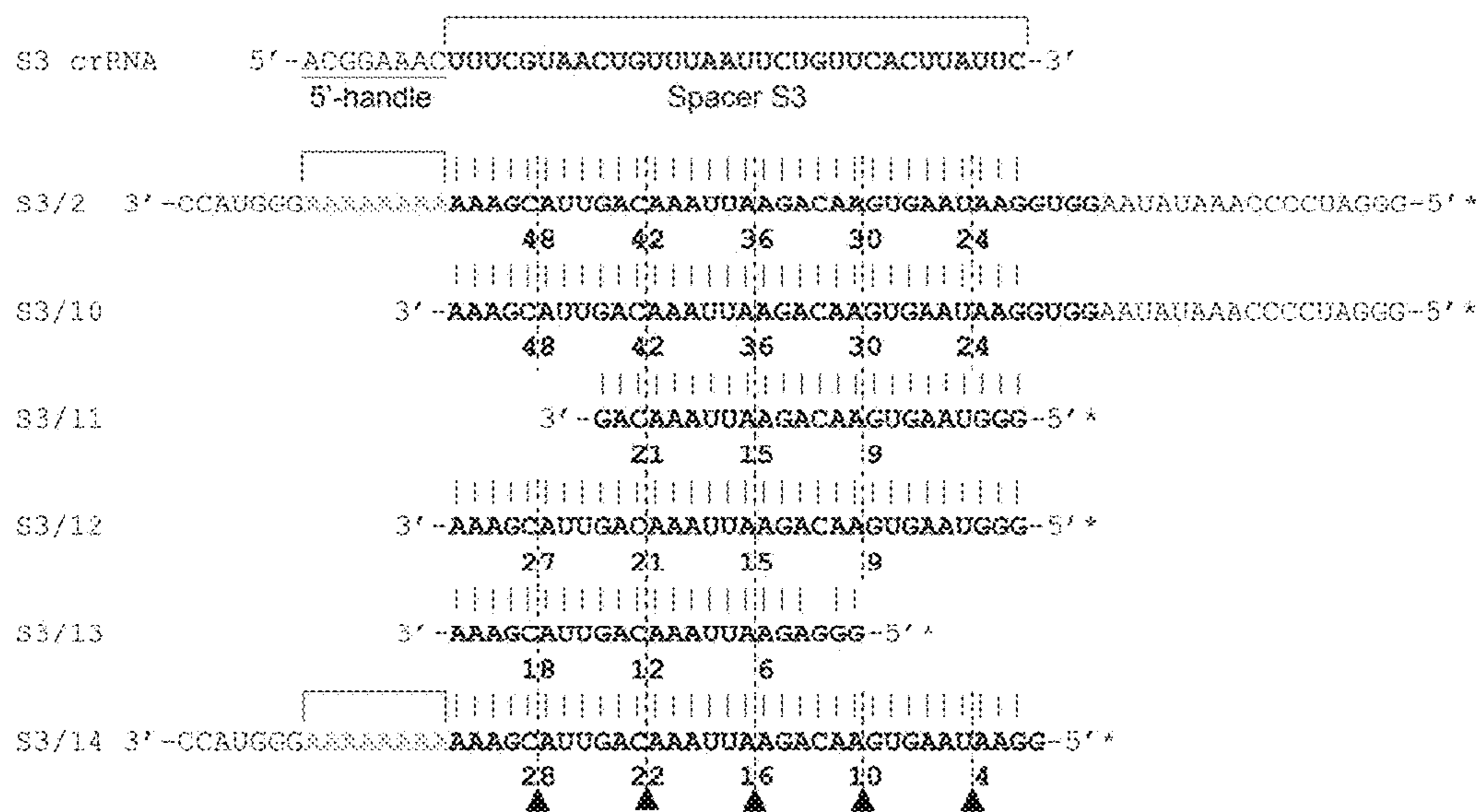


FIG. 5A-2

Csm3 S.thermophilus
 Csm3 S.epidermidis
 Csm3 L.lactis
 Csm3 M.kandleri
 Ssol425 S.solfataricus
 Ssol426 S.solfataricus
 Cmr4 T.thermophilus
 Cmr4 P.furiosus
 Cmr4 S.solfataricus

Line 1	(6) R I R D A F S N A (9) S Y U E F E N T H R I T E A --- N P R Q E R A I R N S T E D D E Y E (2 5) D Y L G C S G R C Y C K V A E E K L K A (2 0)
Line 2	(6) R Q I S D A F F S E K (1 0) A Y M E T F E N T I N R L A V A --- N P R Q E R V T R G S E E D E V F Y N (2 3) D Y L G C G G W R E N E R Q Q K D T N (1 5)
Line 3	(6) R K N D I K S N L (8) S S W E F E N T I N R K K V A --- N P R Q E R V I A G S K E D D E F Y N (2 3) D Y L G C H G W R E S E R A E N L S (1 9)
Line 4	(4 4) R A N R D A H P T Y (1 0) E P U E H E N A I N R V E A --- N P R S E R V P K G S R R G L E Y R (2 3) Q G C H S T R C Y C R V E E R I A A (7 6)
Line 5	(3) S R Y I L D A I A E K --- Y S I S Q T M A I N R V F G O M P G H I Y T Y N D P G S E E S E M I Y (3 0) I F C R K V C Y C L K L V D A K (2 6)
Line 6	(3) A R E L D S L P I S --- Y S L N T S M A I S R T E A V A R R A I V T E Y V D V G S K E S E K G Y (2 1) T Q C C H K R C F C F V K E G K V K F (5 0)
Line 7	(7) A Q V G D A K L L L (1 0 6) T A N E V A R R D E F K T V A K G A I W Y E S P A E S L L Y A L R A D (2 8) L Q F G C K A V V C R G L C R I V V G R ---
Line 8	(7) L S V G D A K L F F (1 0 9) L S N E V A R R I N A E M T V E T G G I W Y E E Y P S D L F Y S L V T (2 2) L Q C G E N E V C K C F V K V T L K E (9)
Line 9	(5) S L G S H L D A V F A I (1 3 0) R S L I R V R K H K R S K T V E E G G I W S E E Y V P V N S L F E S S F A R (1 8) L I L G C K E H I C K C I V E L R W L N D ---

E119 E123 E139

FIG. 5B

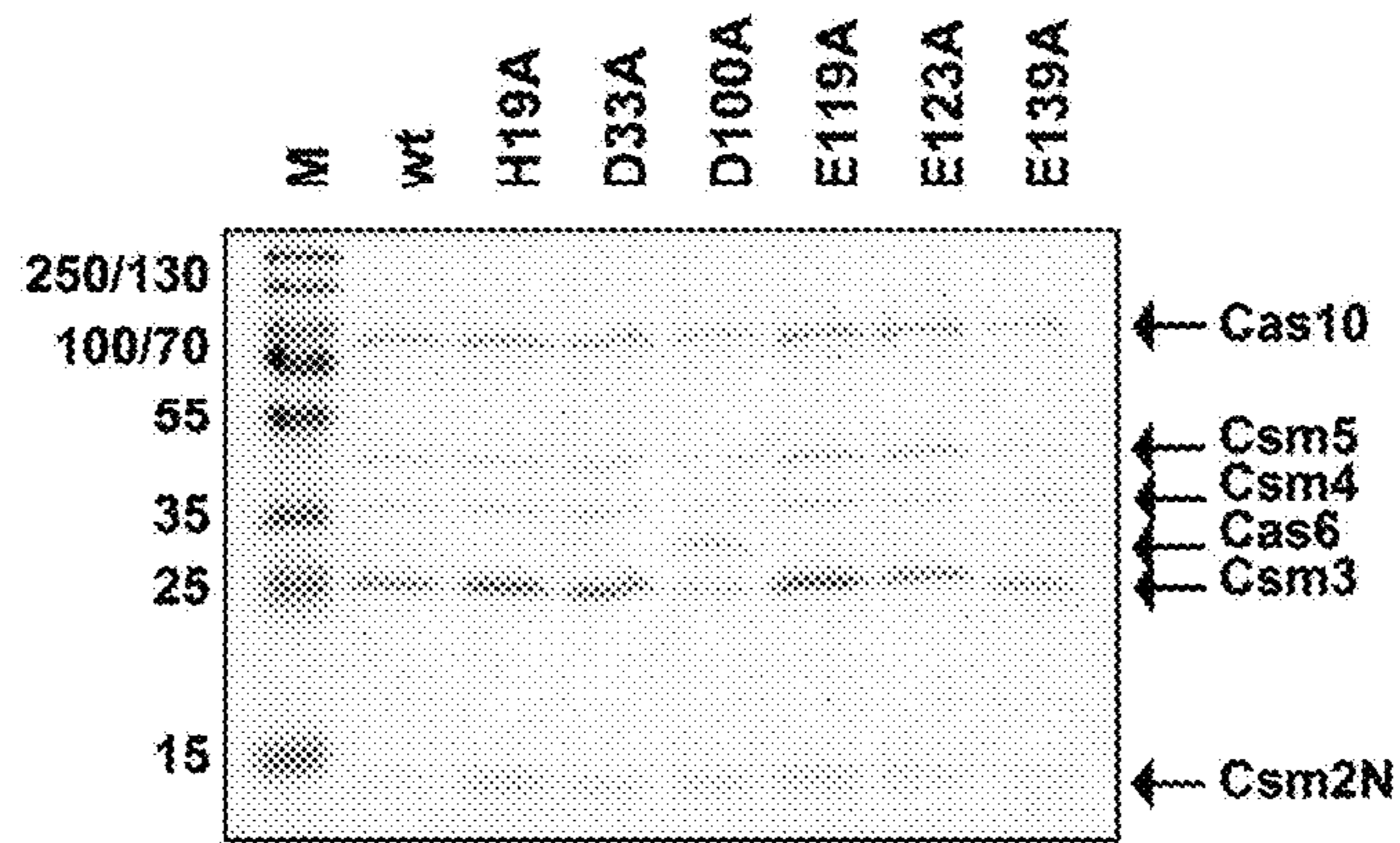


FIG. 5C

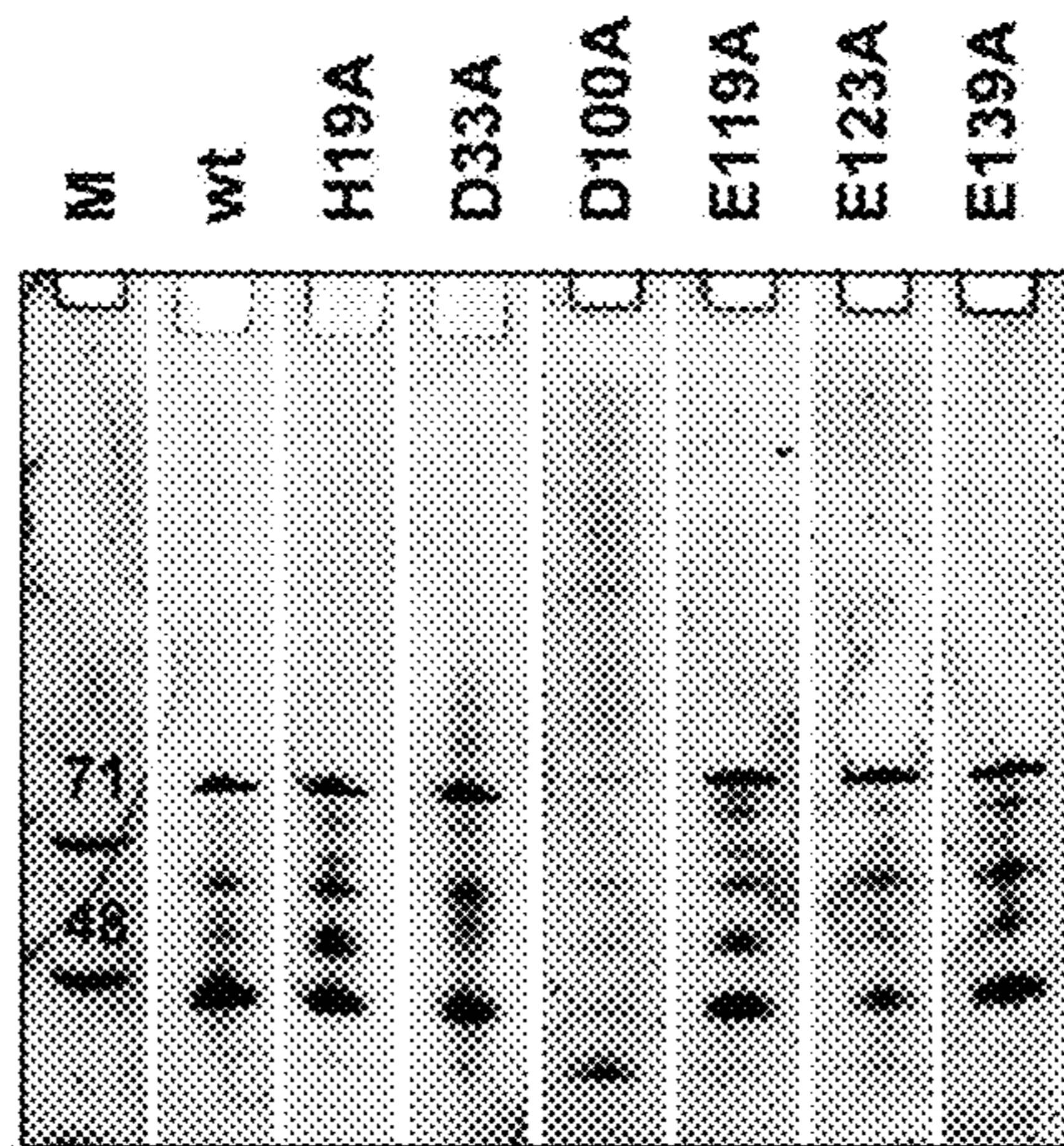


FIG. 5D

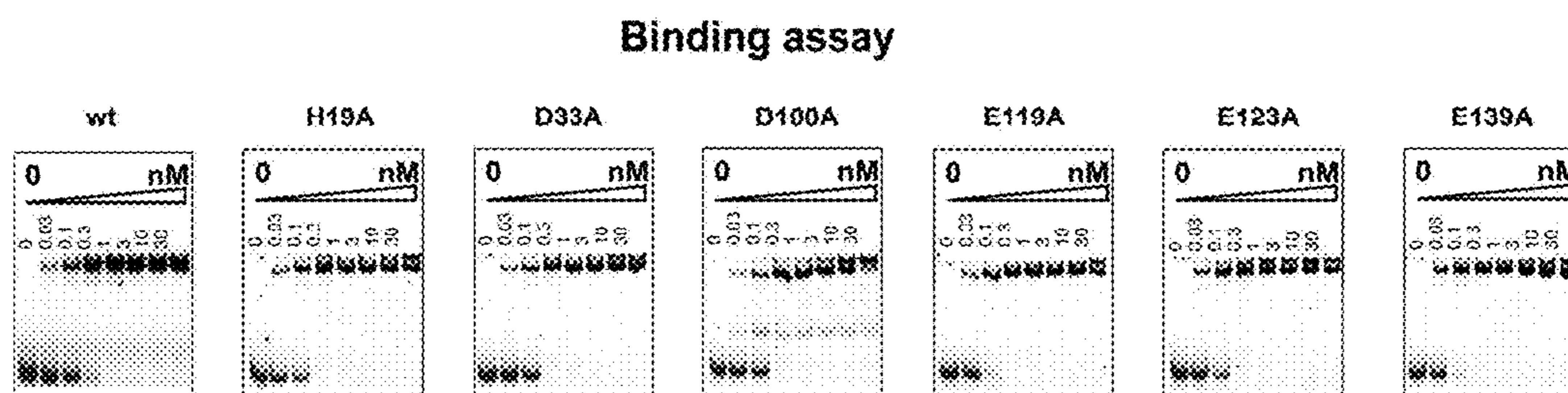


FIG. 5E

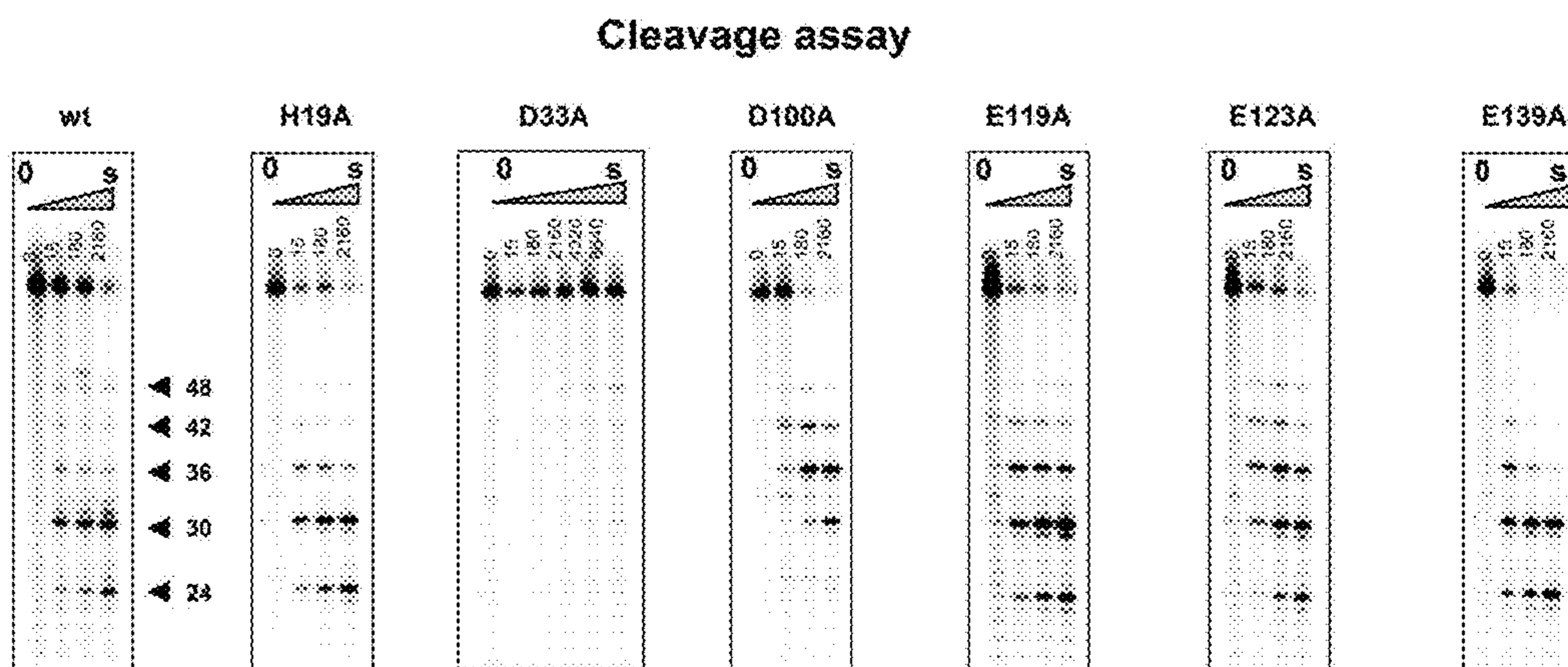


FIG. 5F

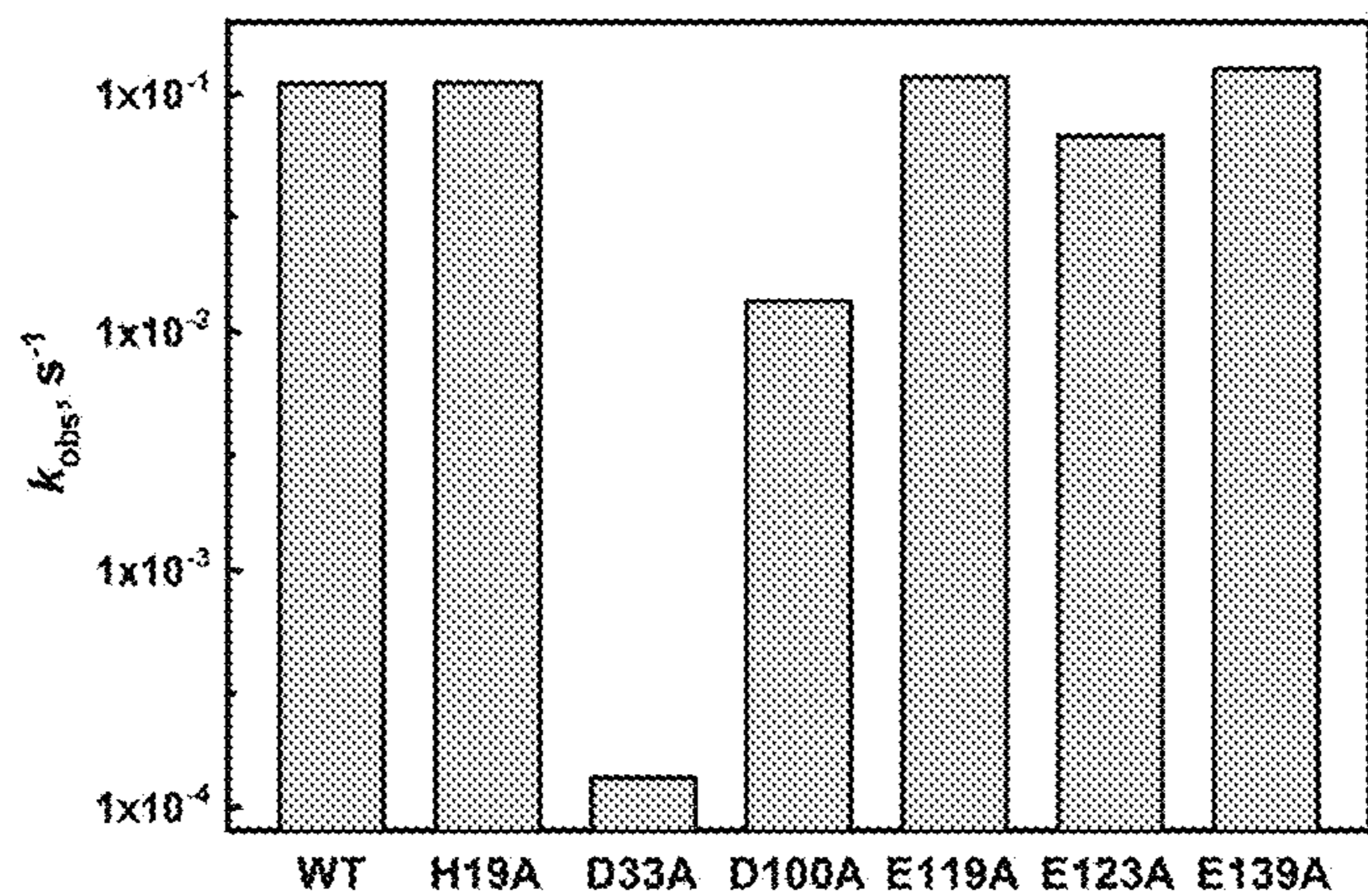


FIG. 6A

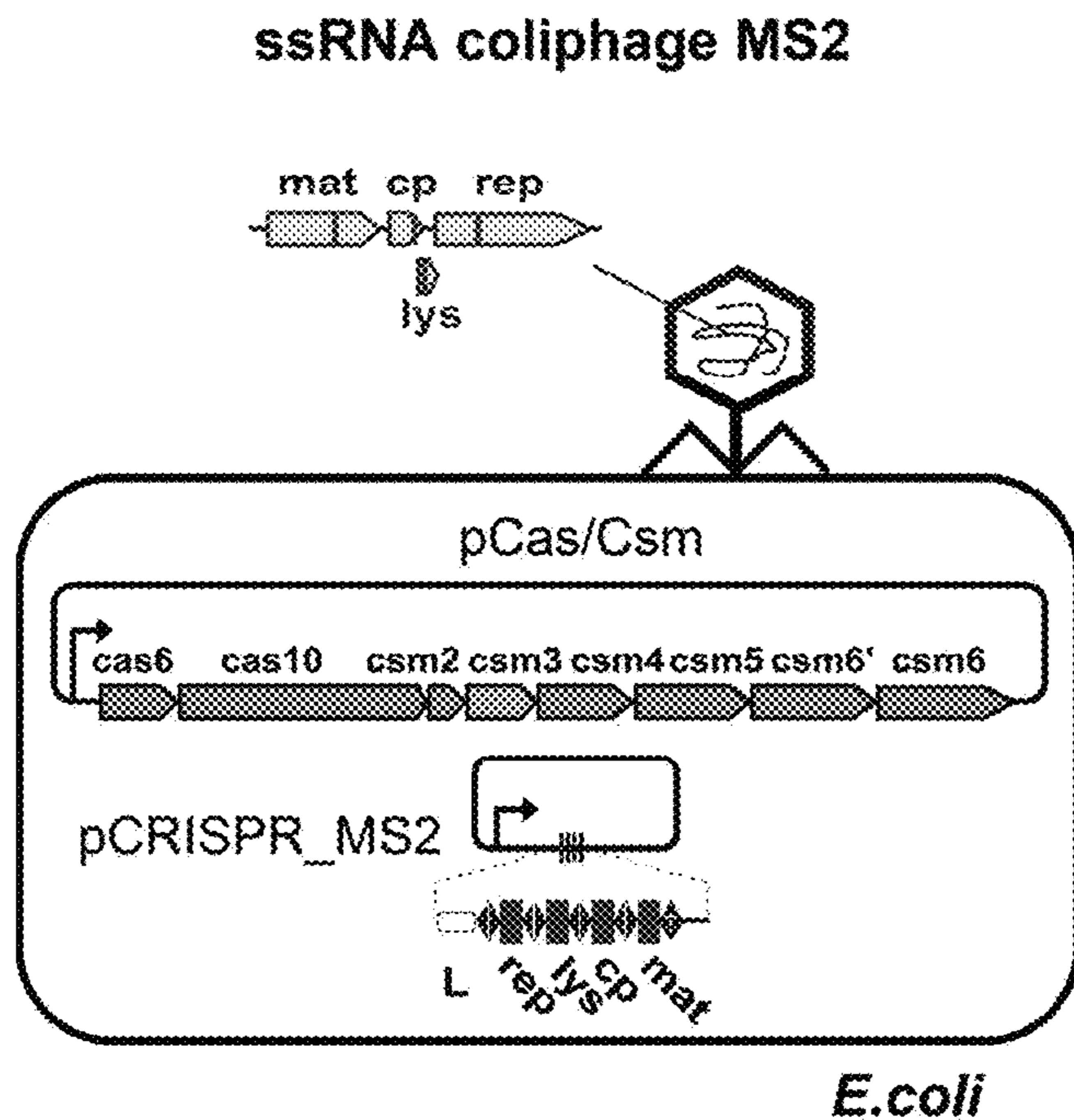


FIG. 6B

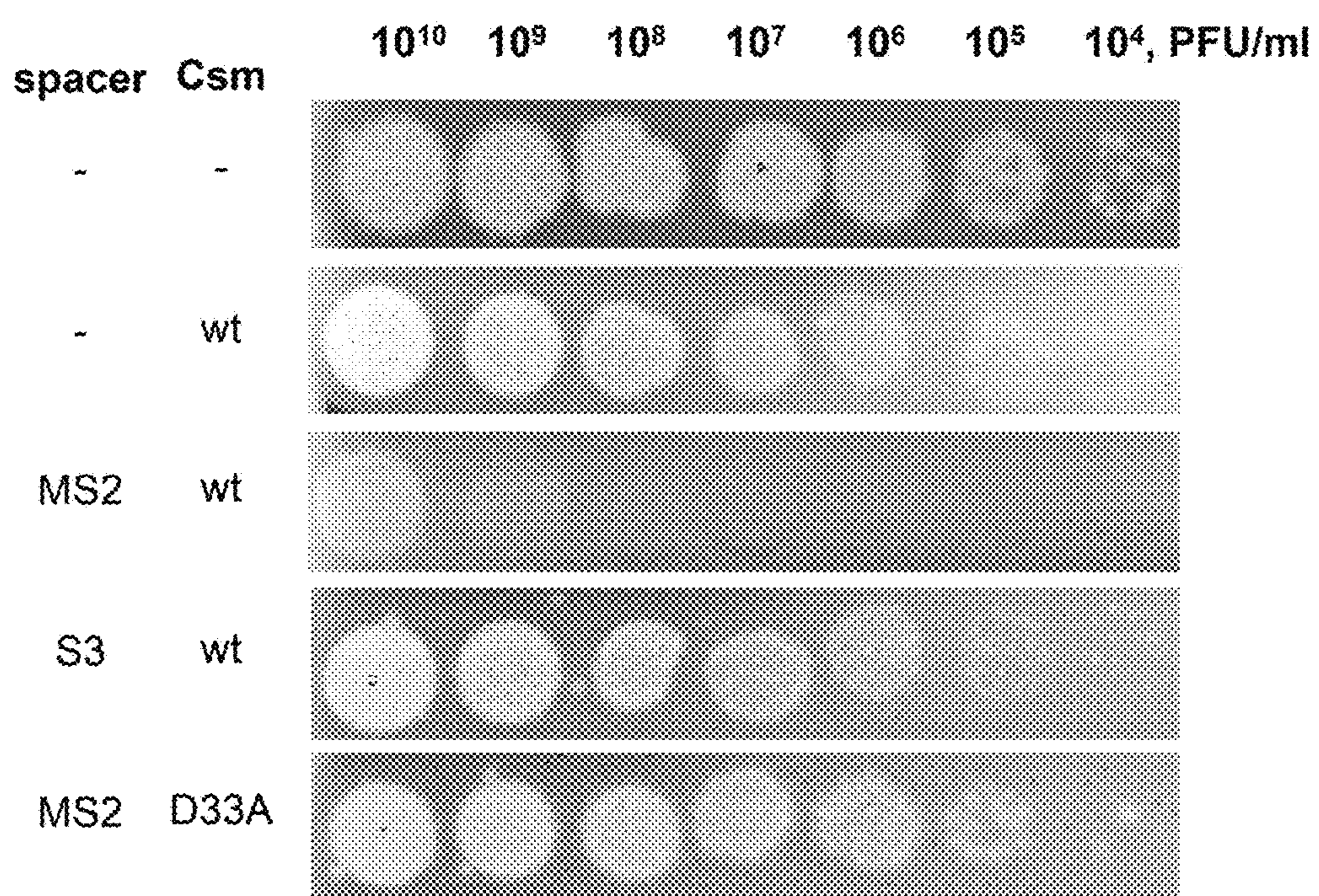
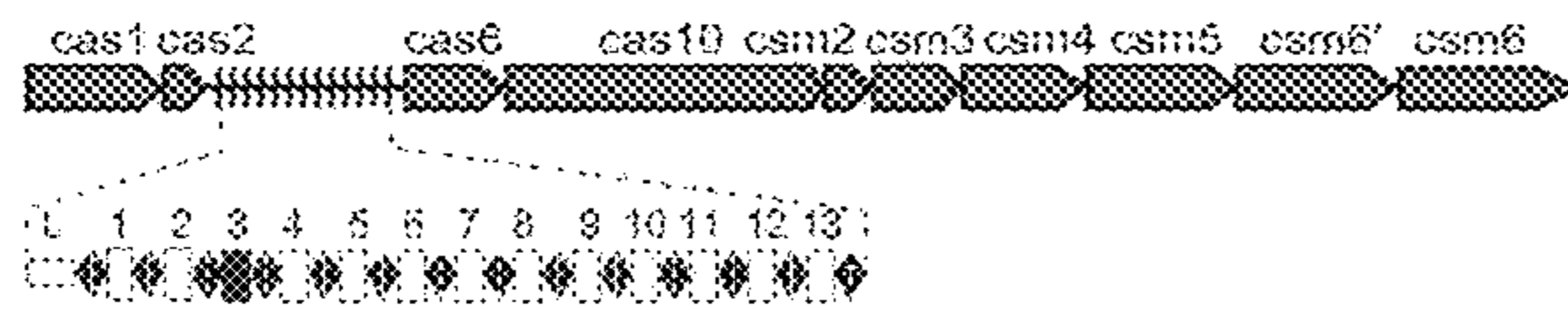
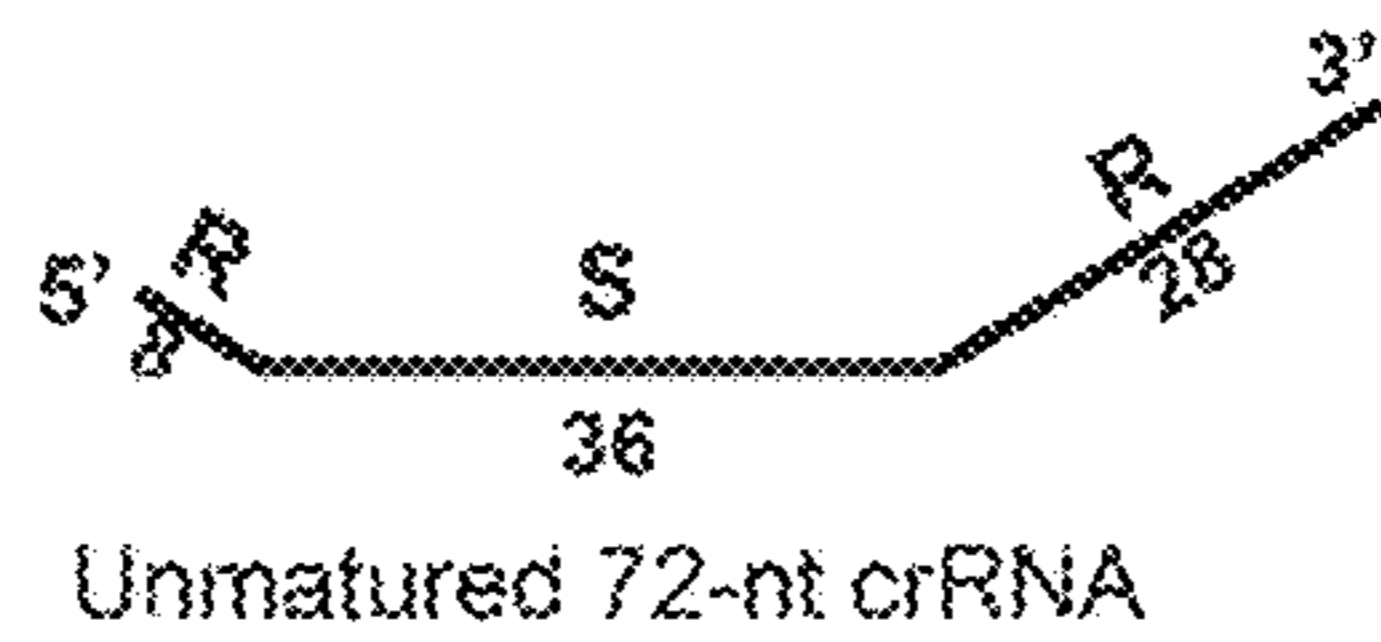


FIG. 7

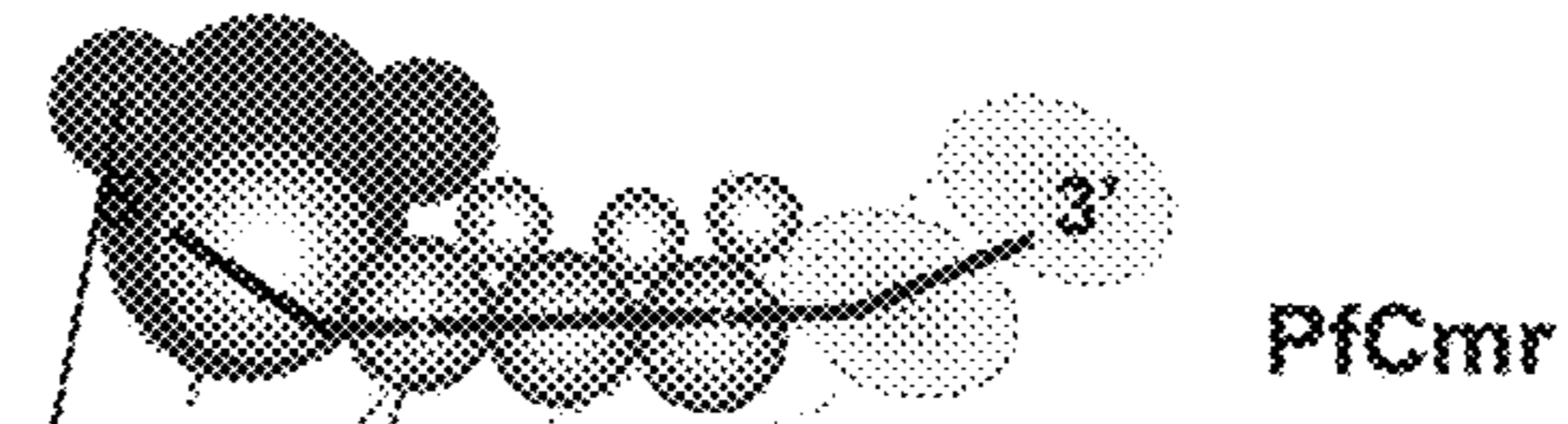
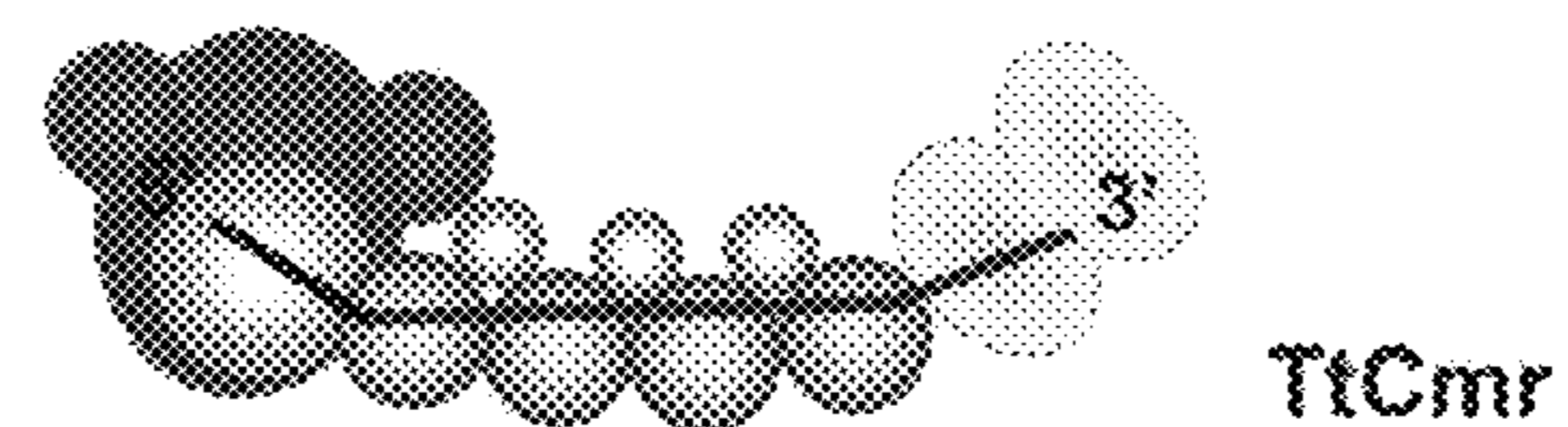
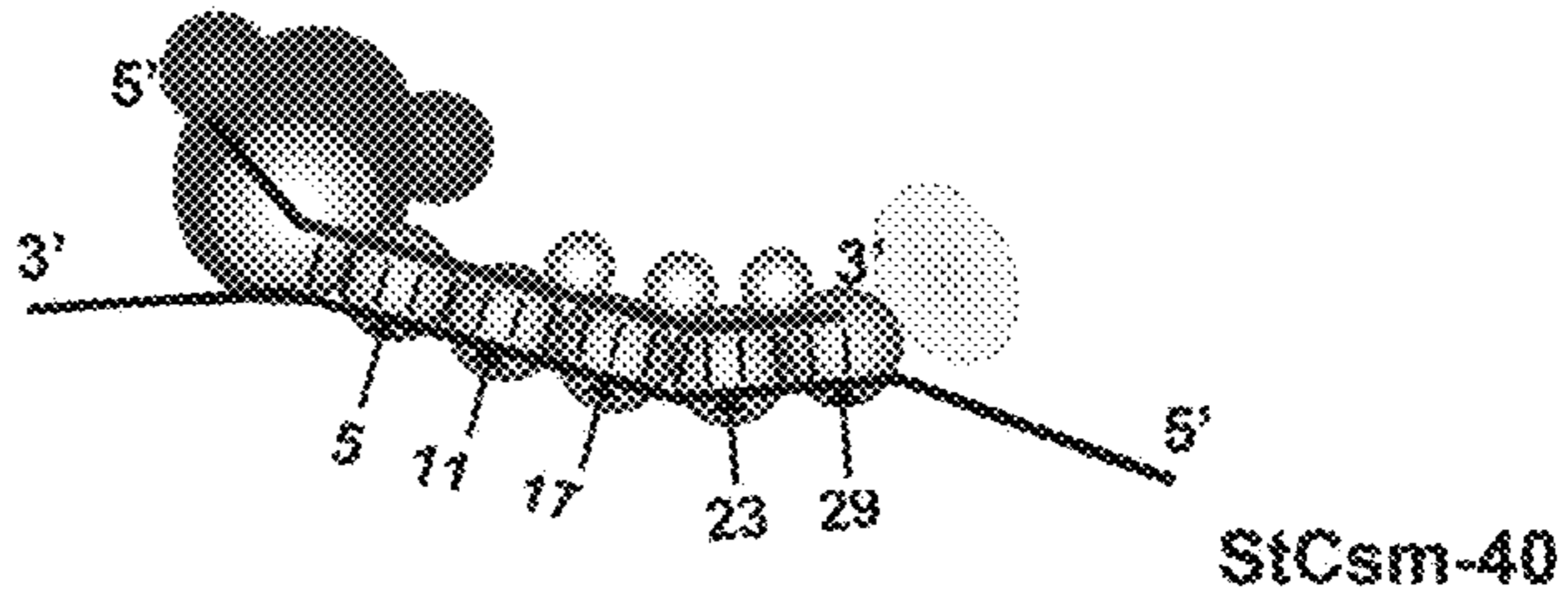
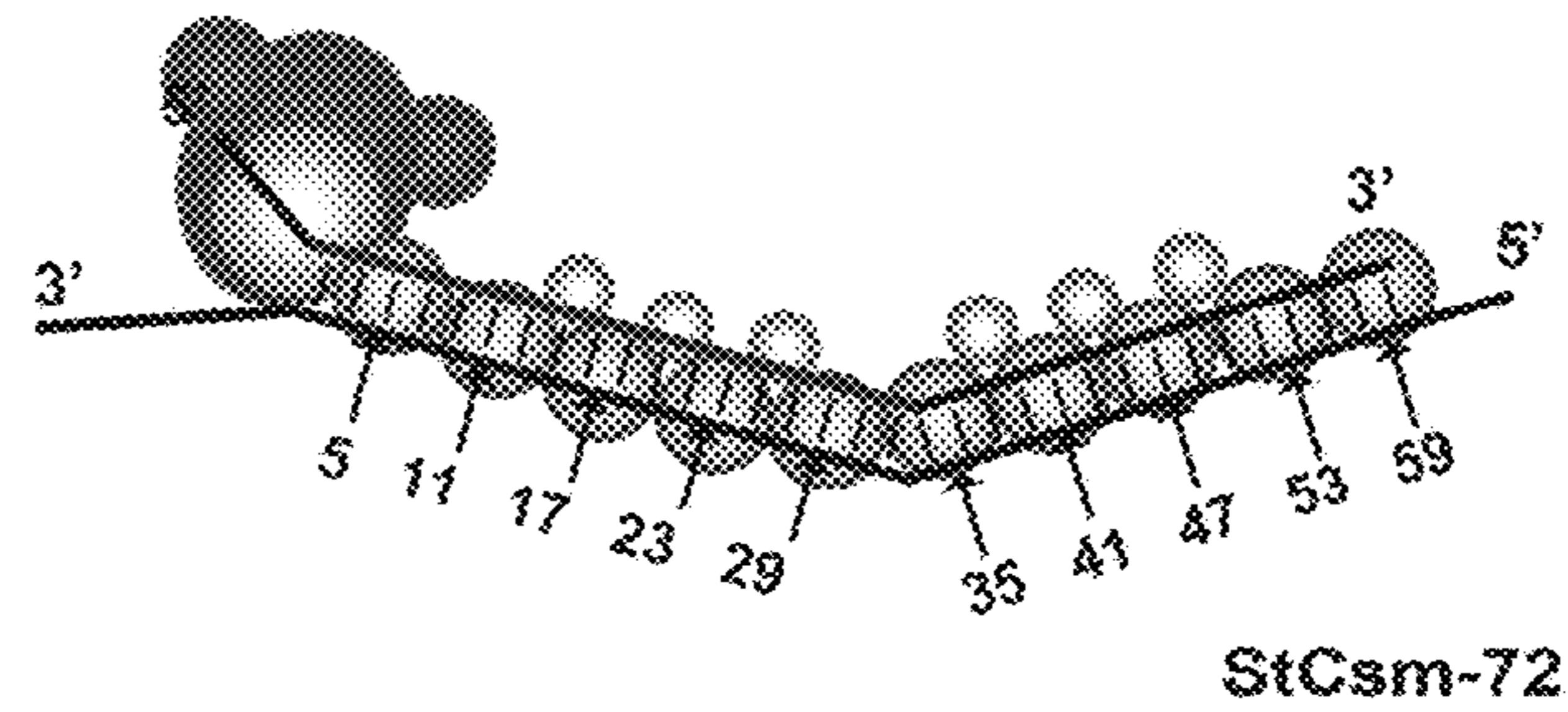
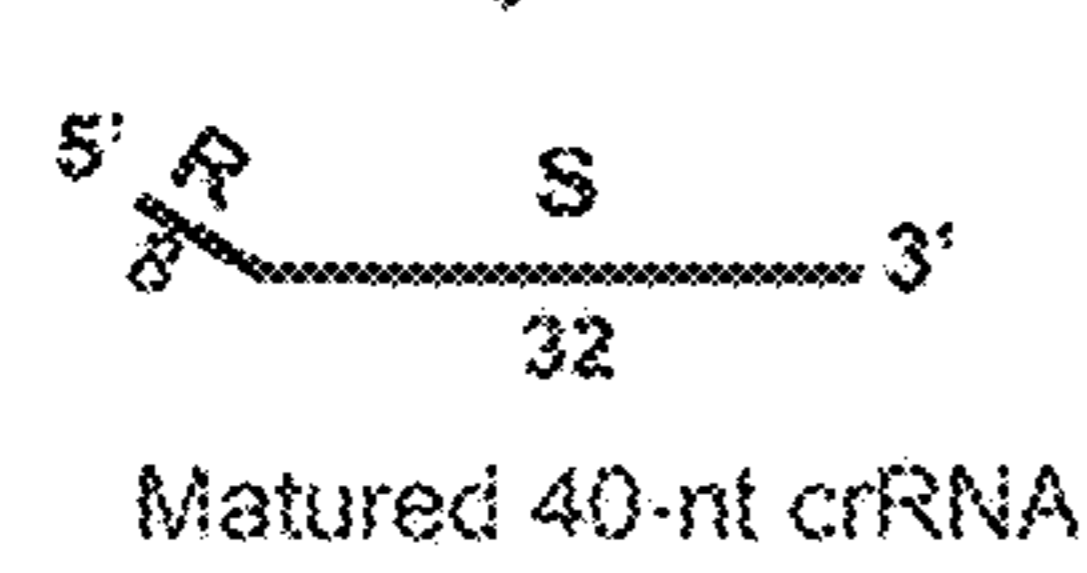
DGCC8004 CRISPR2-cas



Transcription,
Cas6 processing



Maturation



- | | | |
|-------|-------|-------|
| Csm6 | Cmr1 | |
| Csm2 | Cmr6 | |
| Csm3 | Cmr5 | |
| Csm4 | Cmr4 | |
| Cas10 | Cmr3 | |
| | Cas10 | |

FIG. 8

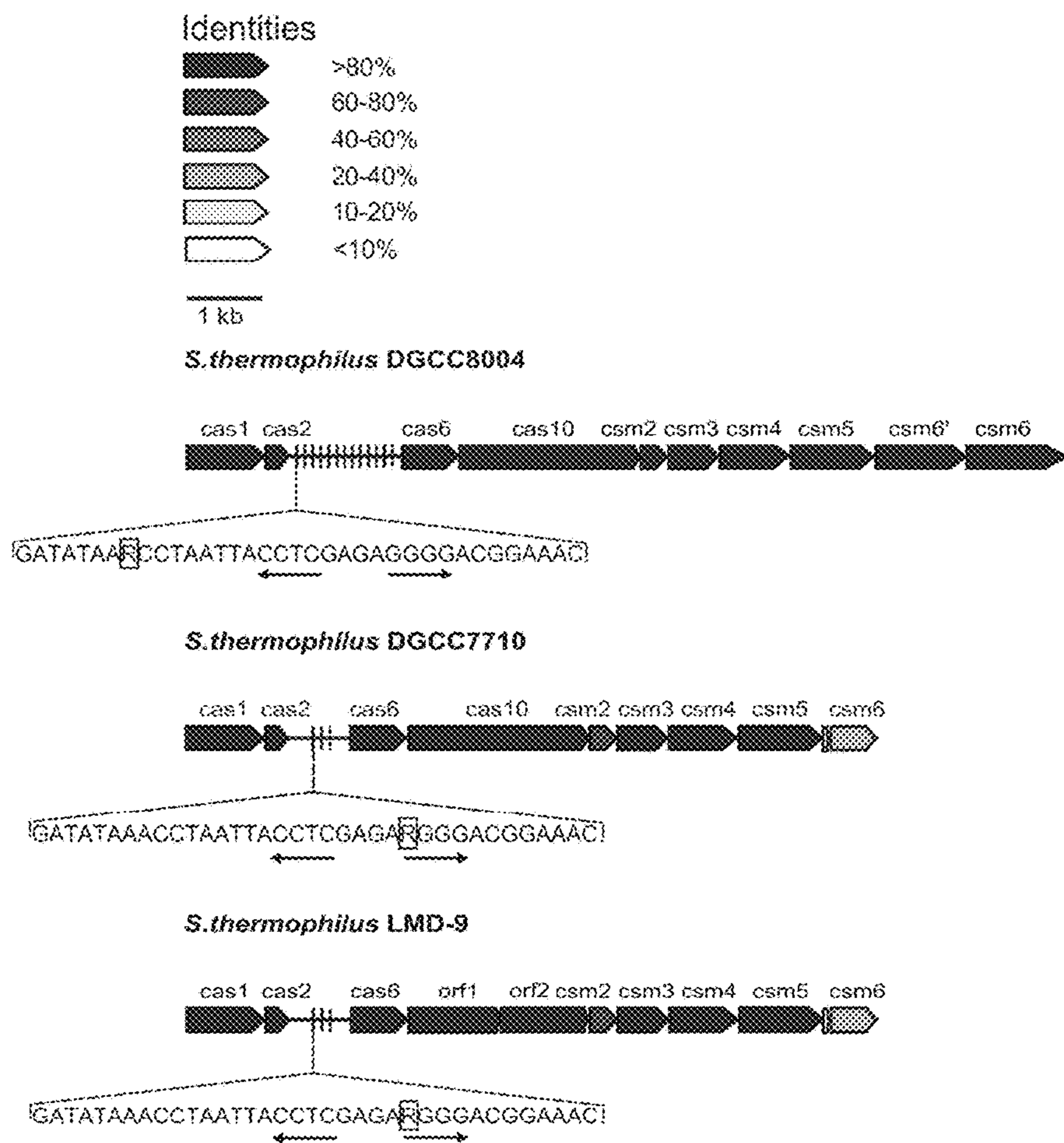


FIG. 8 con't

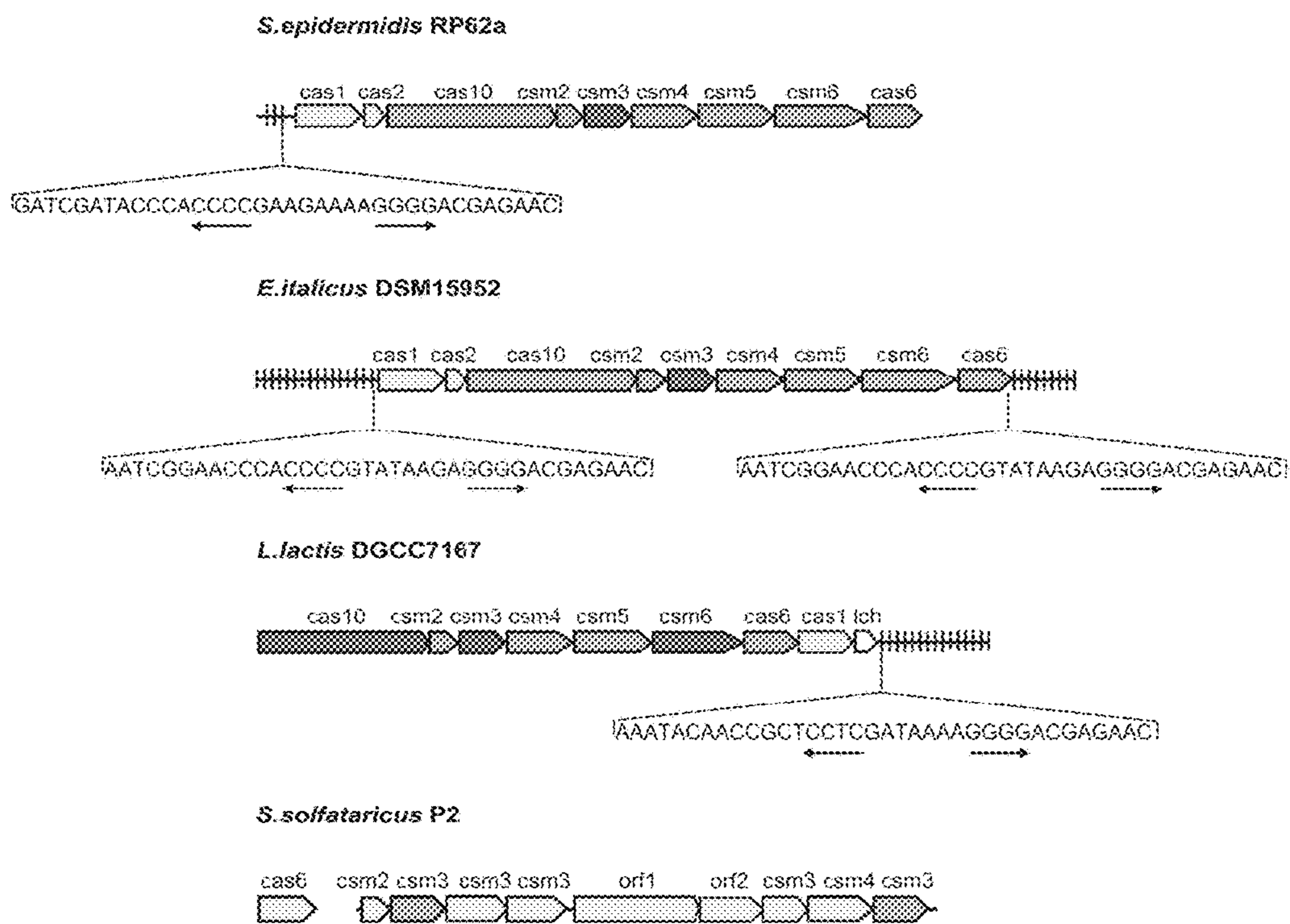
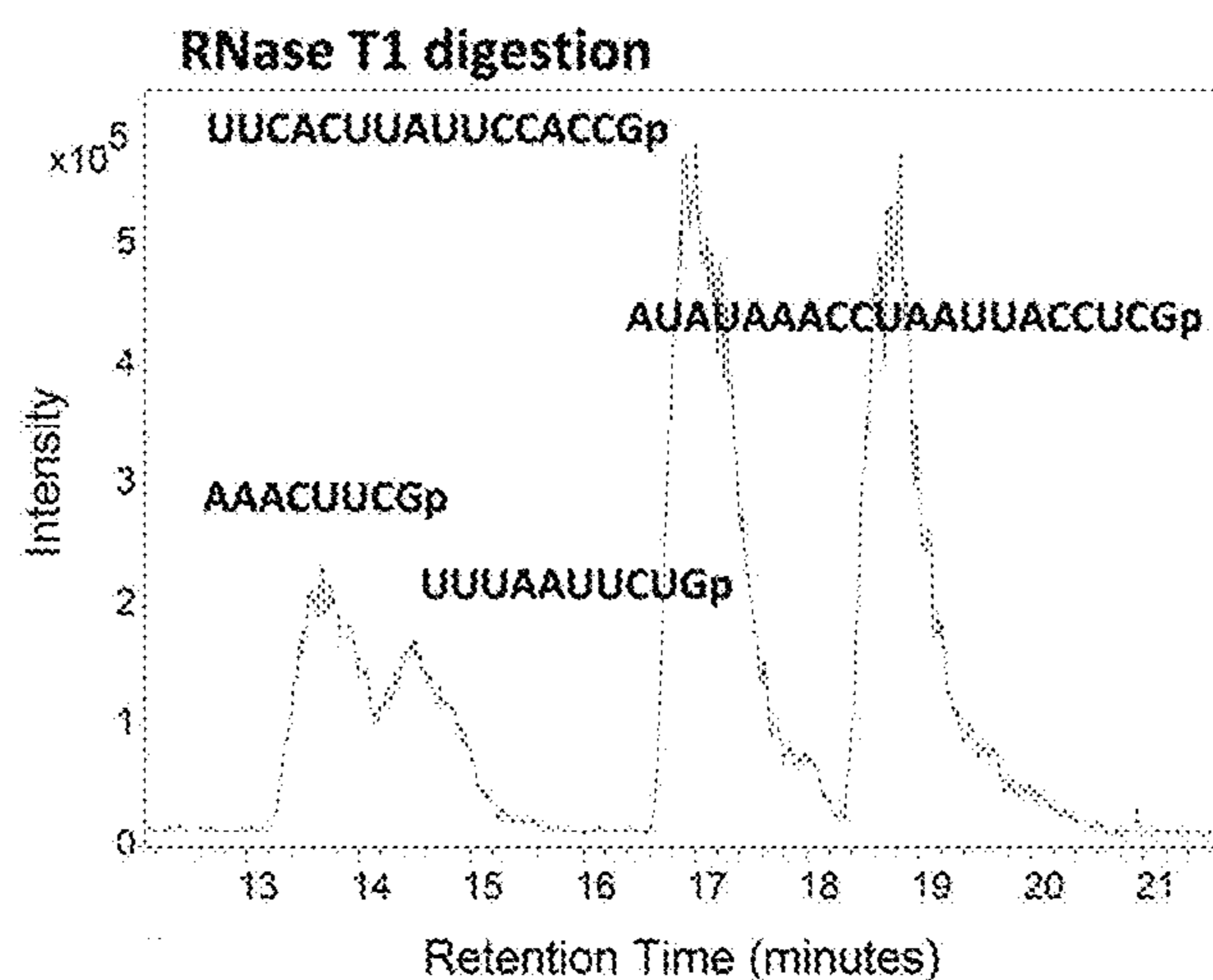
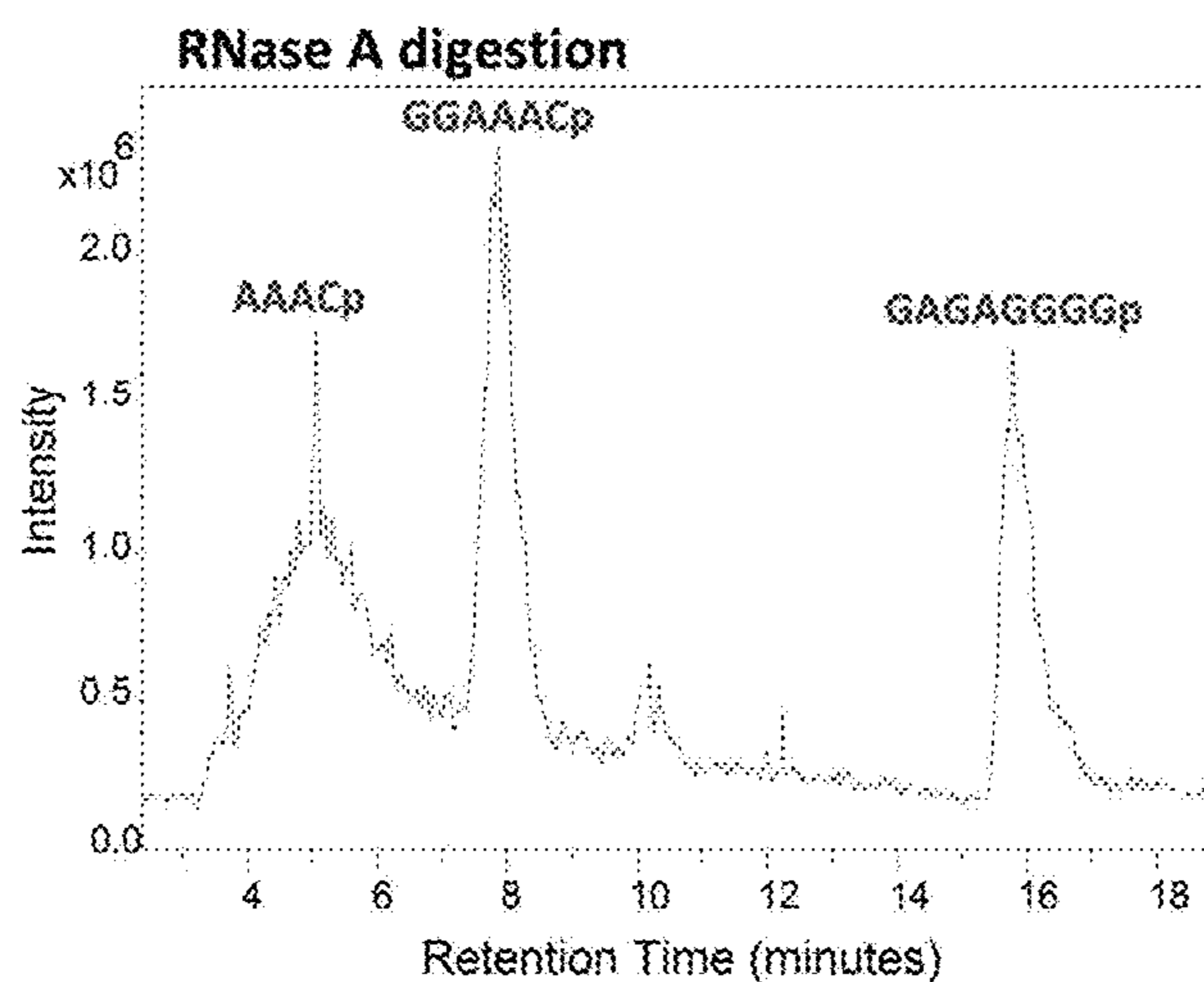


FIG. 9A



Position	Theoretical mass (Da)	Experimental Mass (Da)	Sequence
A5:G13	2879.725	2879.529	AAACUUUCGp
U20:G29	3163.839	3163.663	UUUAUUUCUGp
U30:G45	5018.969	5018.918	UUCACUUAUUC CACCGp
A46:G65	6359.830	6359.347	AUAUAAACCUAA UUACCUCGp

FIG. 9B



Position	Theoretical mass (Da)	Experimental Mass (Da)	Sequence
G3:C8	2001.244	2001.3	GGAAACp
A50:C53	1310.279	1310.3	AAACp
G65:G72	2747.306	2747.4	GAGAGGGGp

FIG. 9C

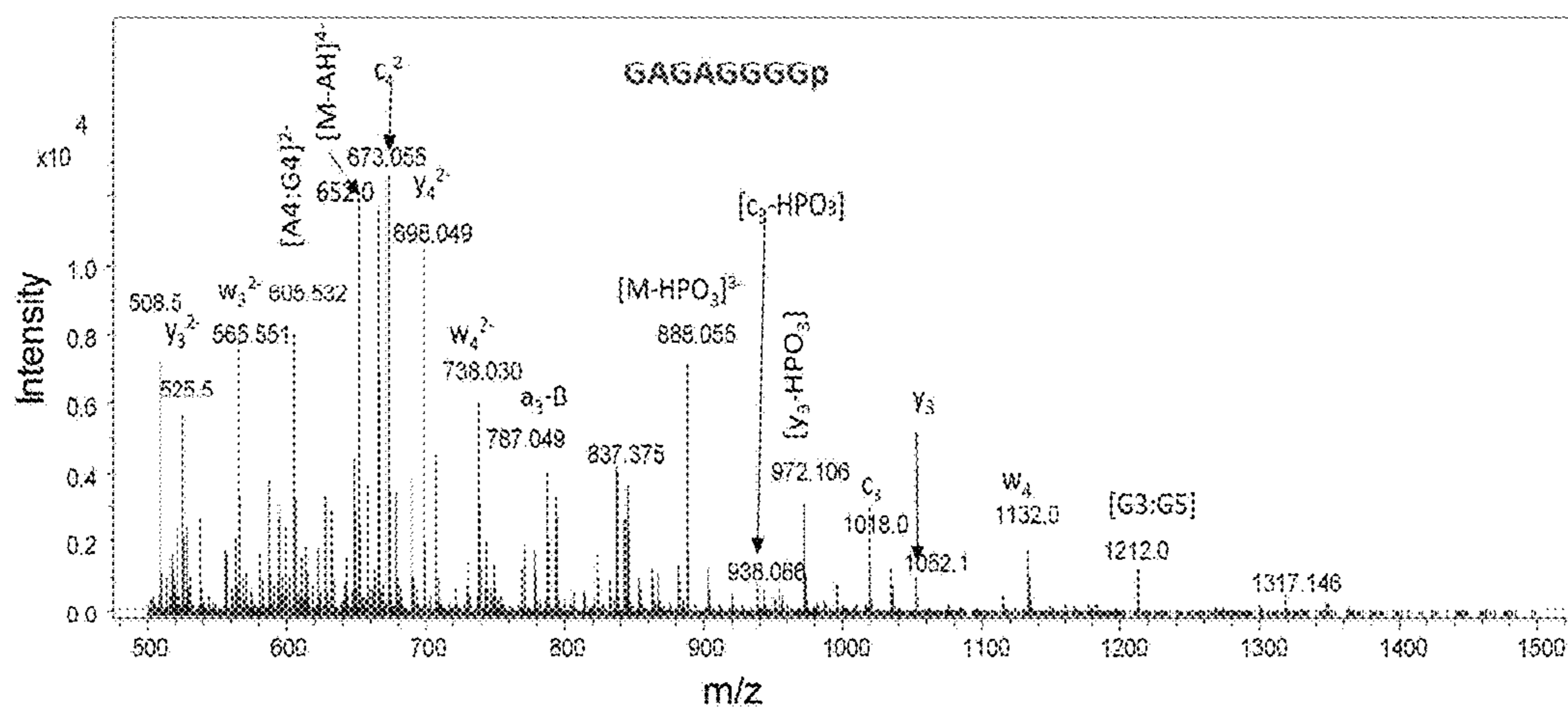
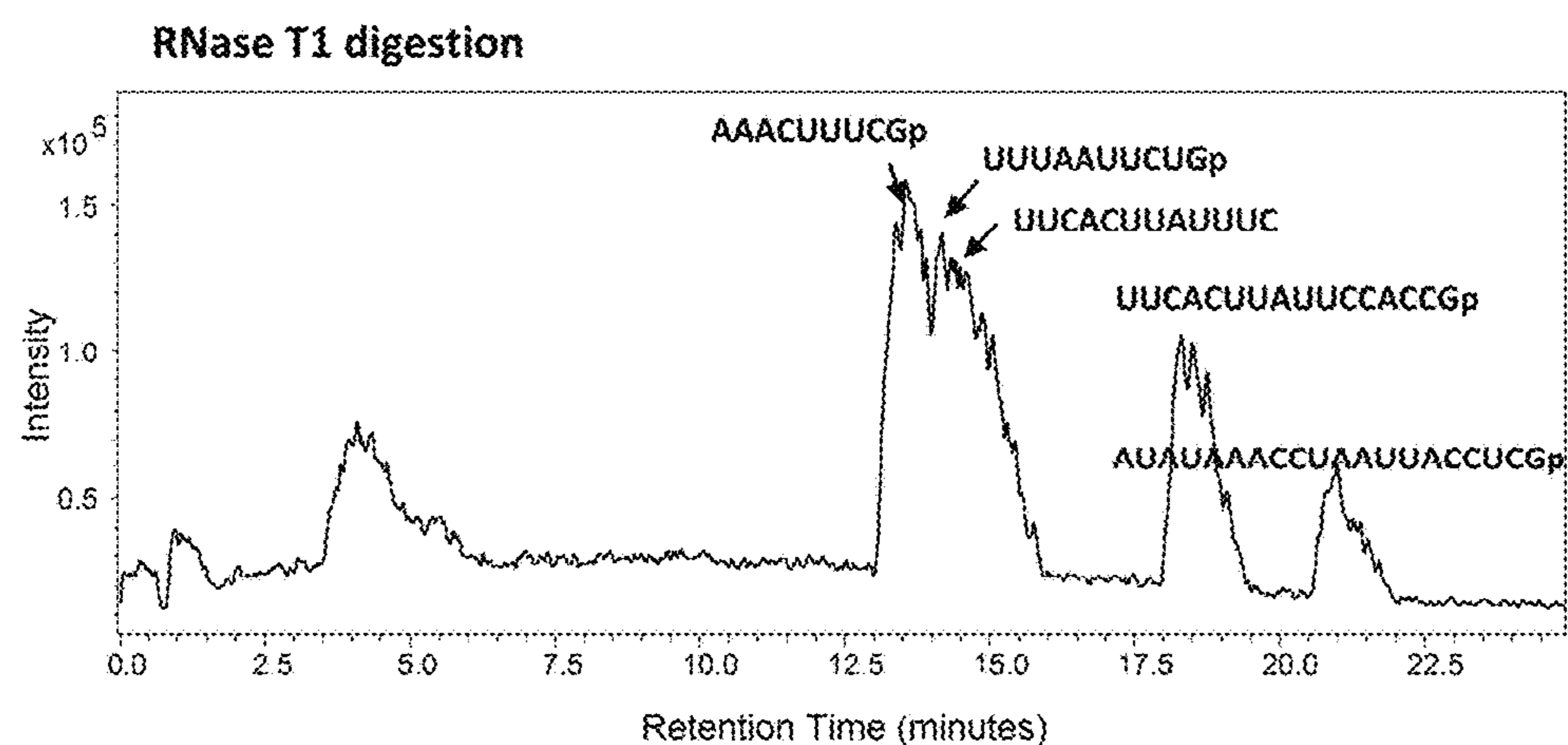


FIG. 9D



Position	Theoretical mass (Da)	Experimental Mass (Da)	Sequence
A5:G13	2879.725	2879.526	AAACUUUCGp
U20:G29	3163.839	3163.381	UUUAUUCUGp
U30:C40	3349.019	3348.906	UUCACUUAUUC
U30:G45	5018.969	5018.478	UUCACUUAUUCACCGp
A46:G65	6359.830	6359.333	AUAUAAACCUAUUACCUCGp
U30:G45	5000.954	5001.082	UUCACUUAUUCACCG>p
A46:G65	6341.815	6342.009	AUAUAAACCUAUUACCUCG>p

FIG. 10A

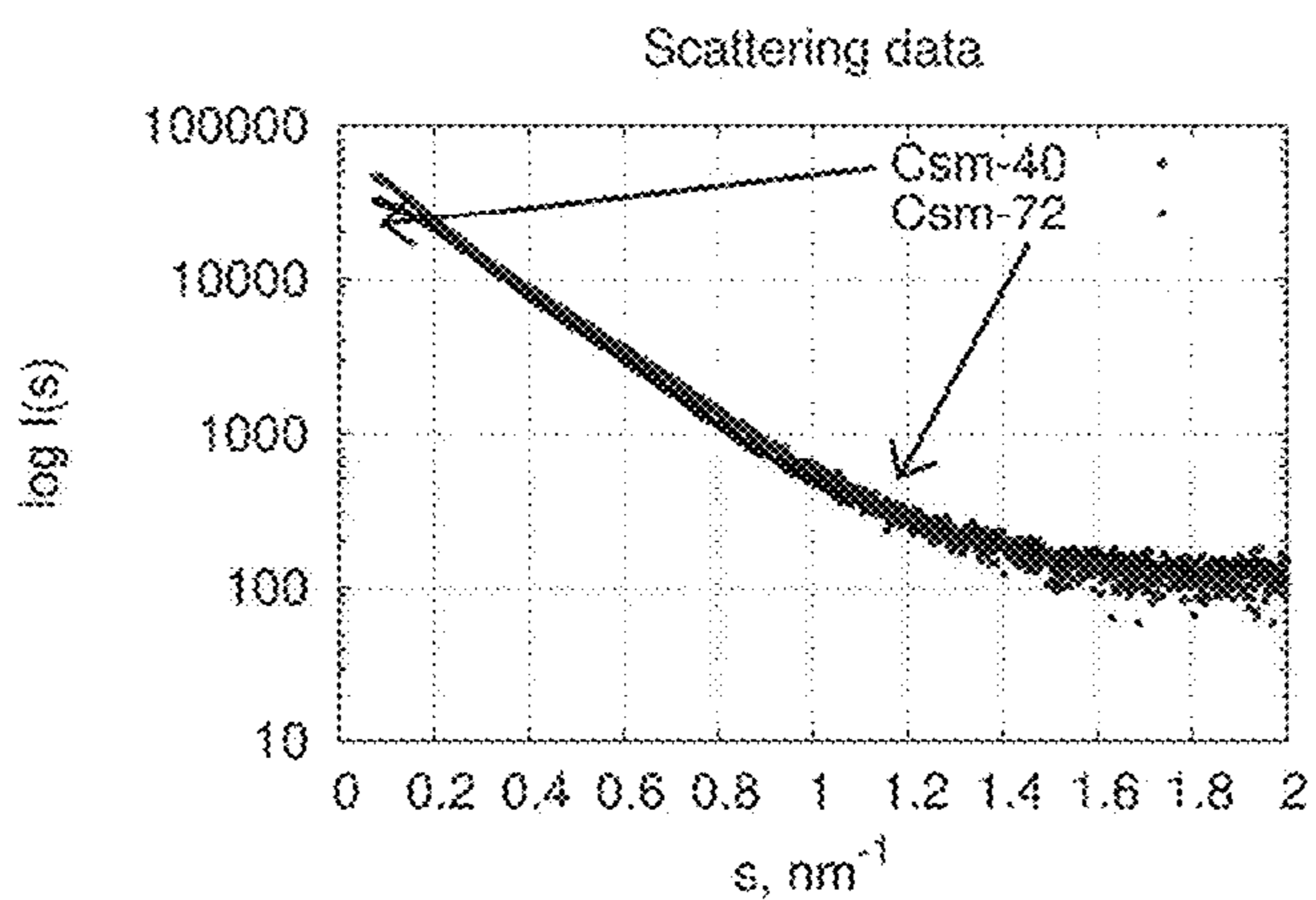


FIG. 10B

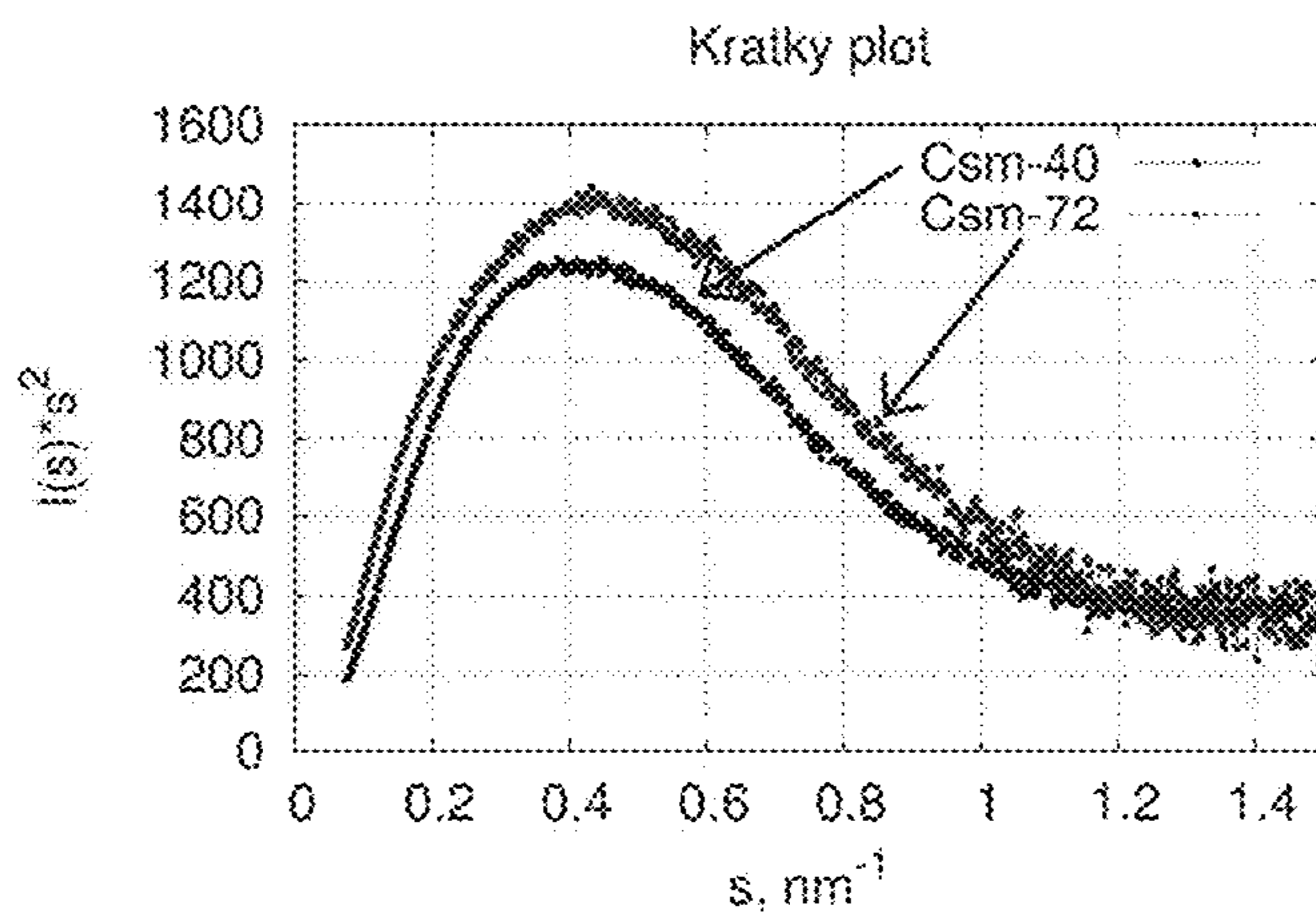


FIG. 10C

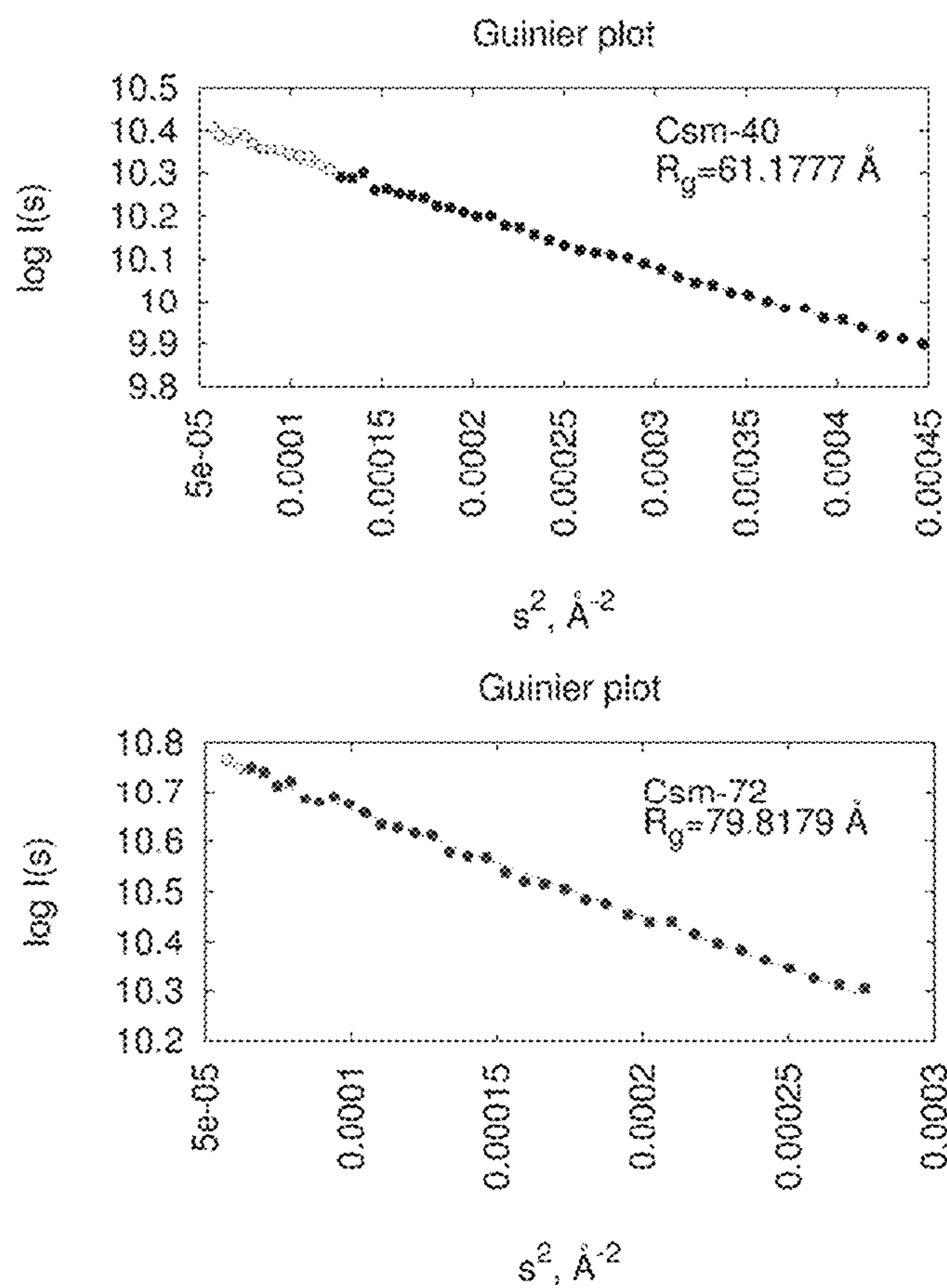


FIG. 10D

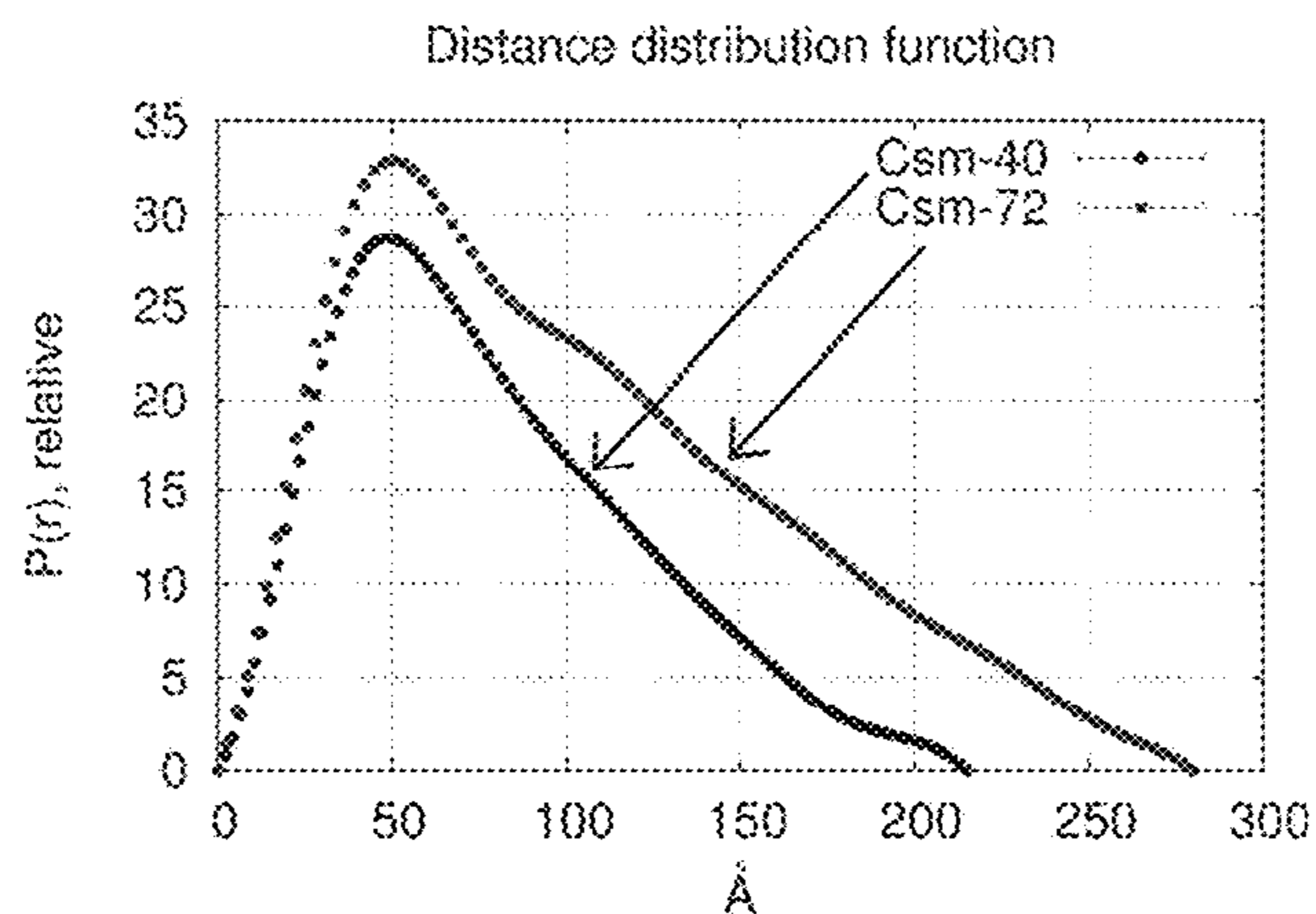


FIG. 10E

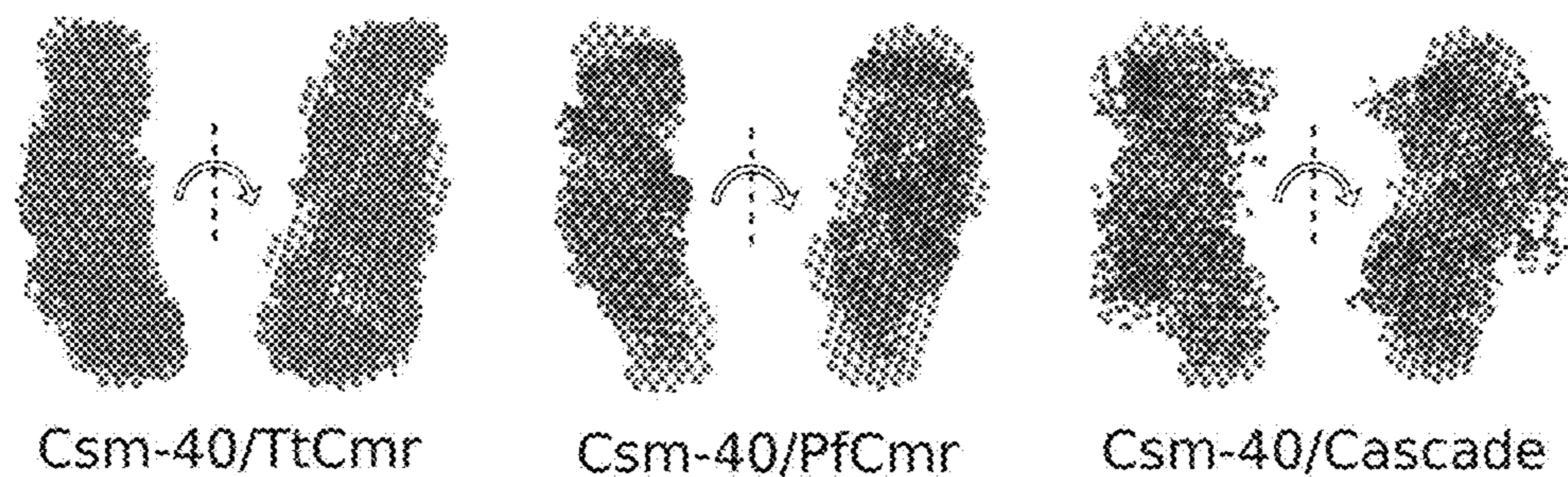


FIG. 11A

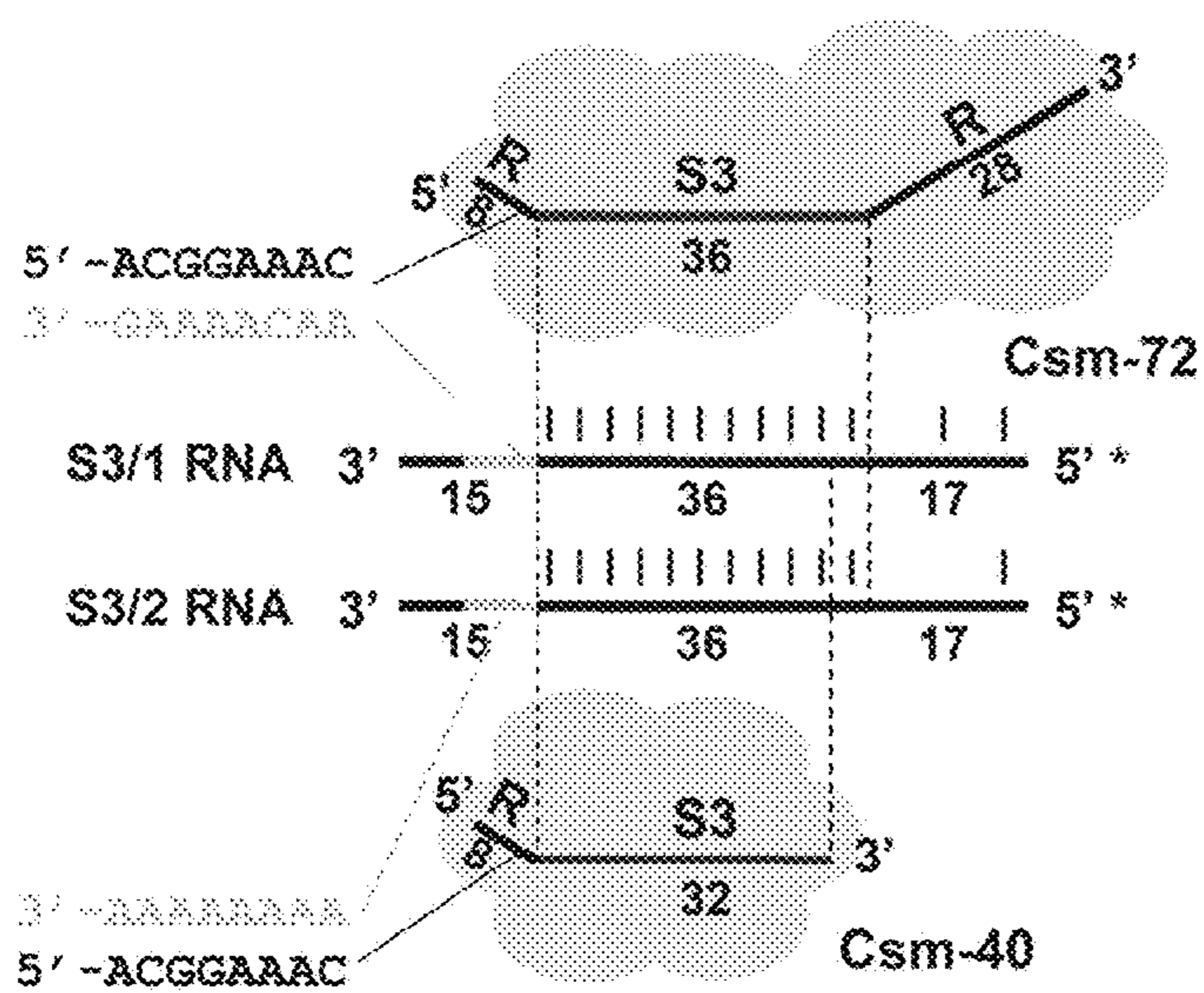


FIG. 11B

Binding assay

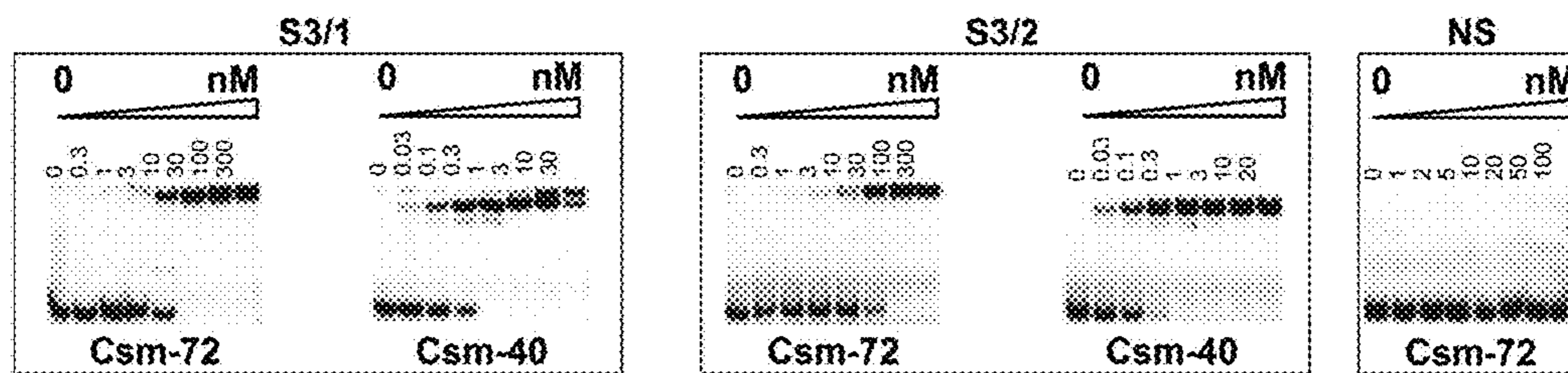


FIG. 11E

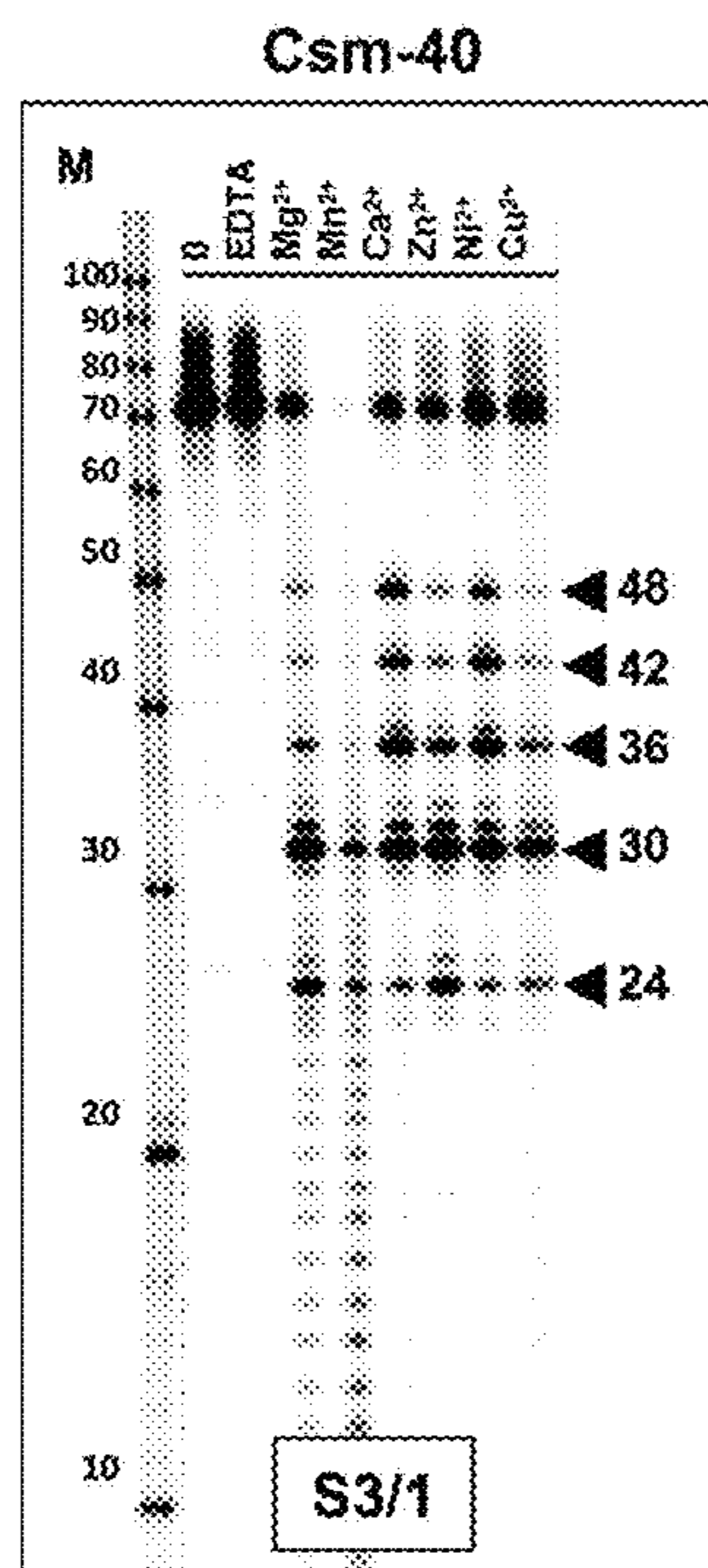


FIG. 11F

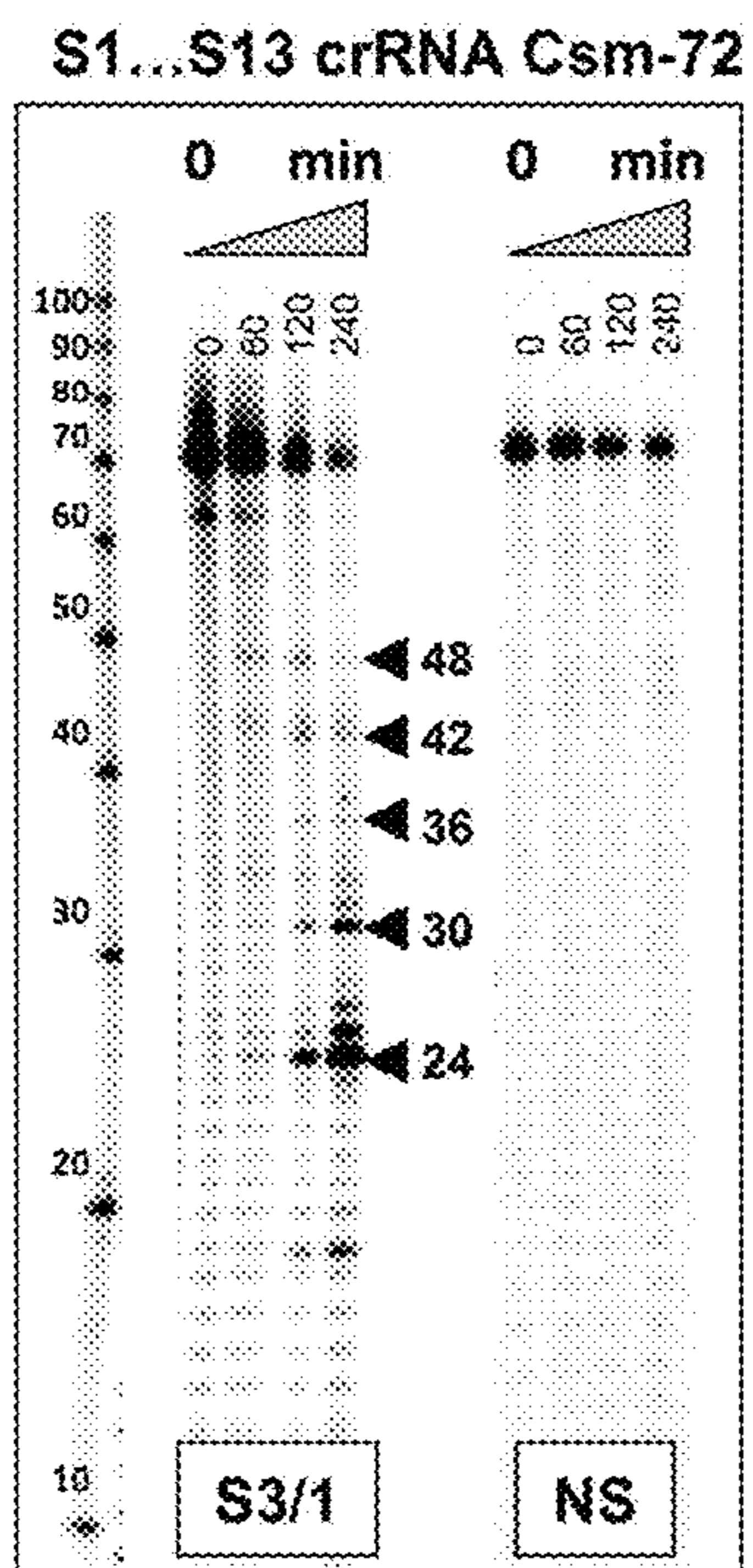


FIG. 12A

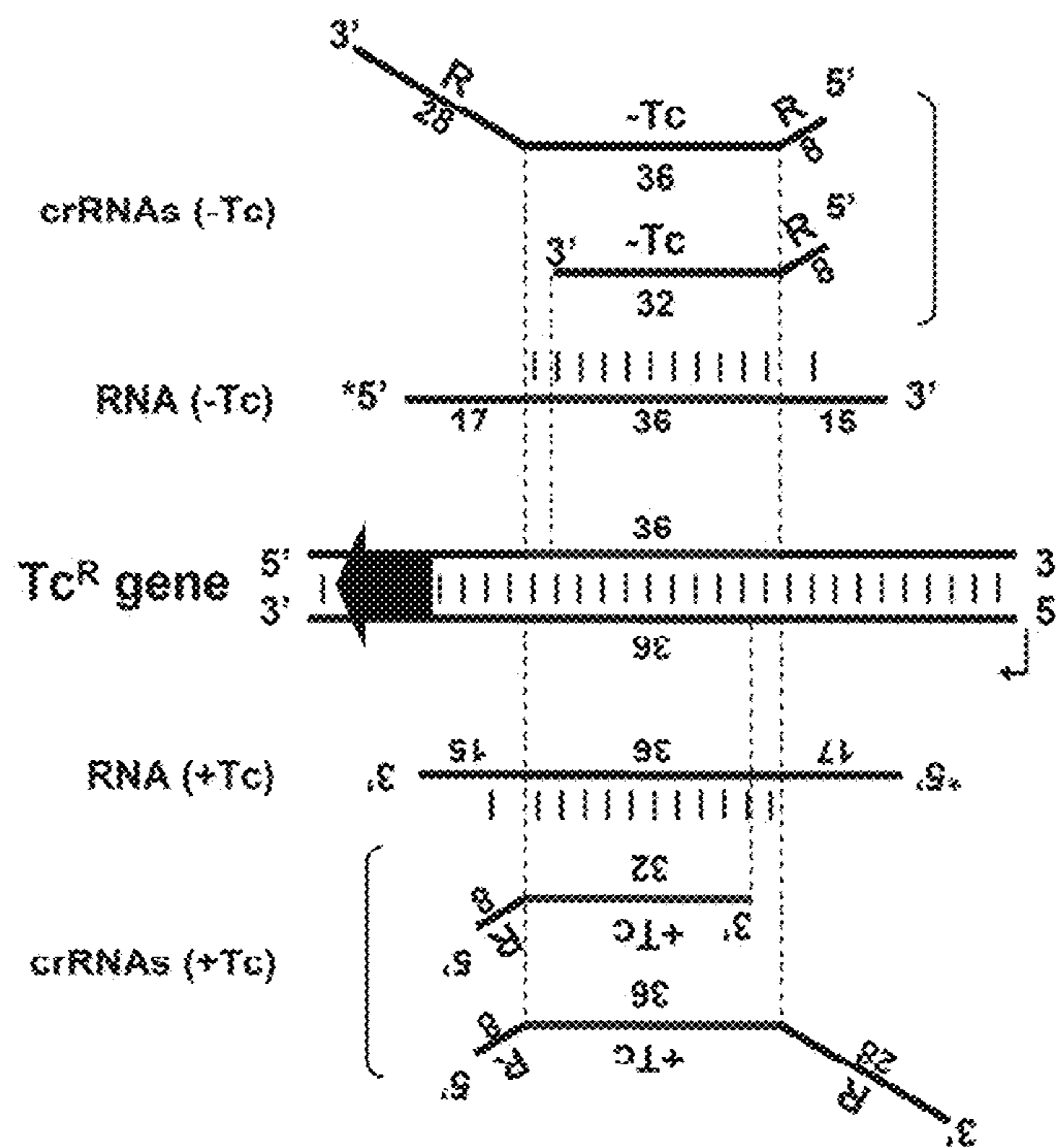


FIG. 12B

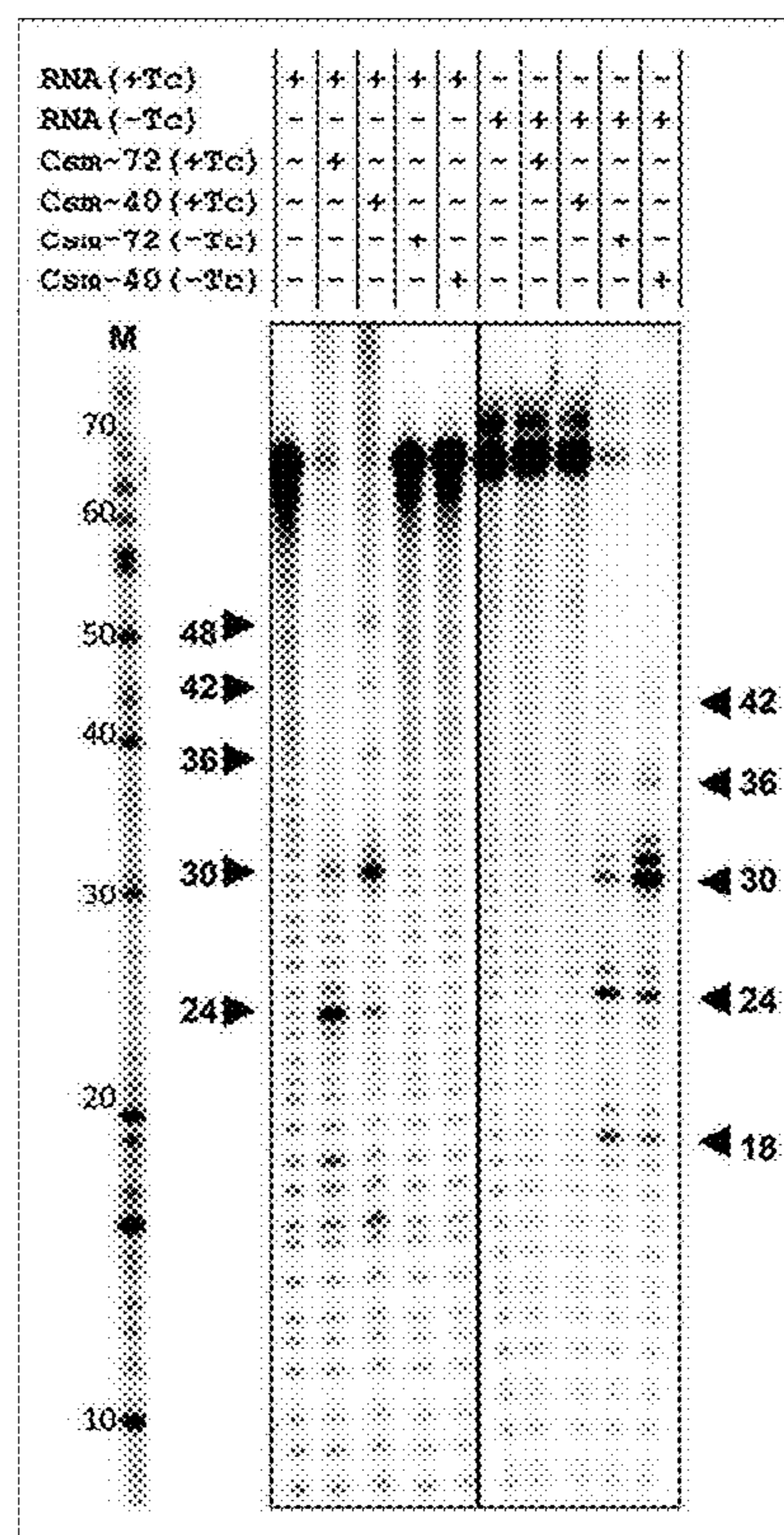


FIG. 14A

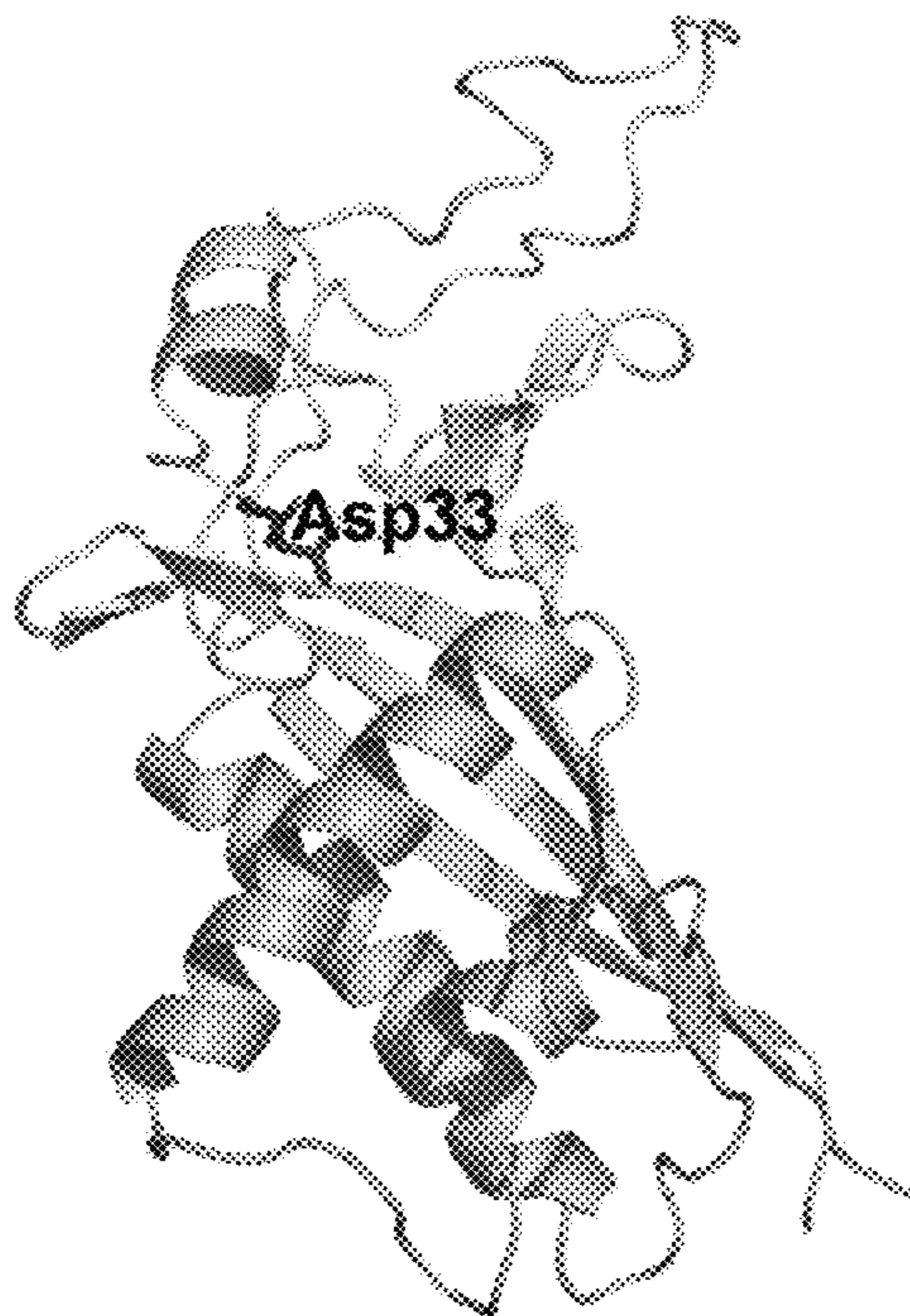


FIG. 14B

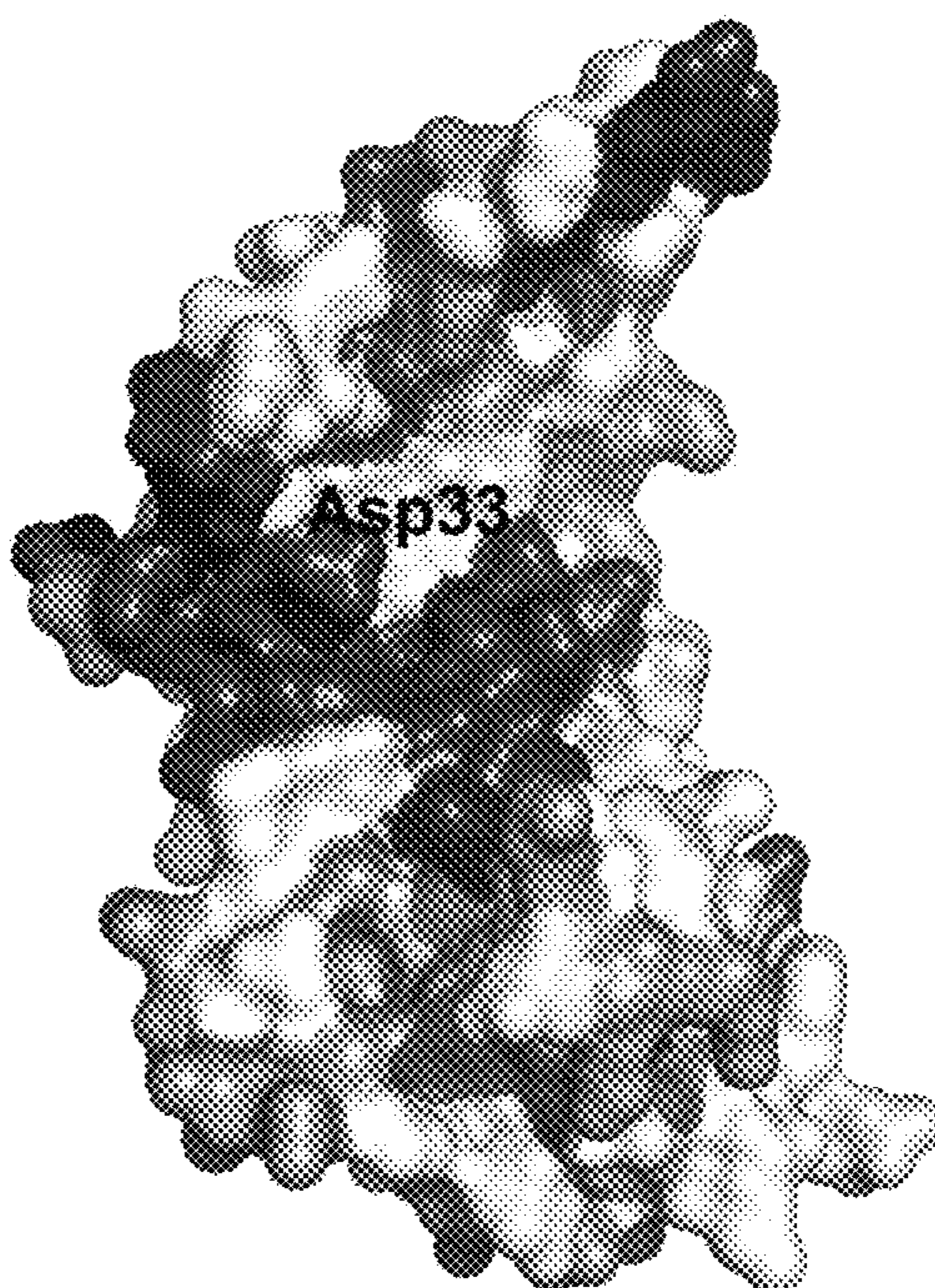


FIG. 14C

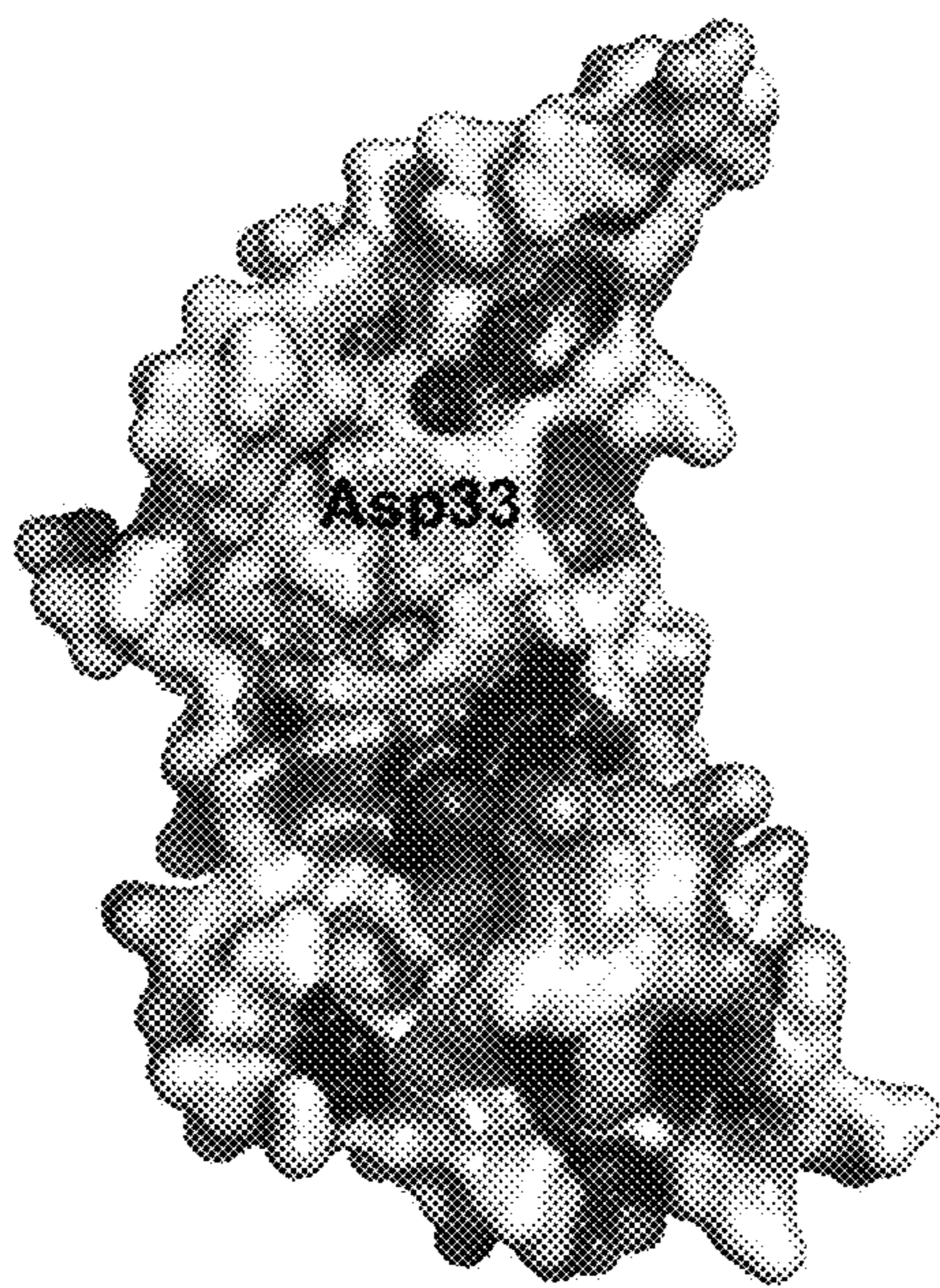


FIG. 14D

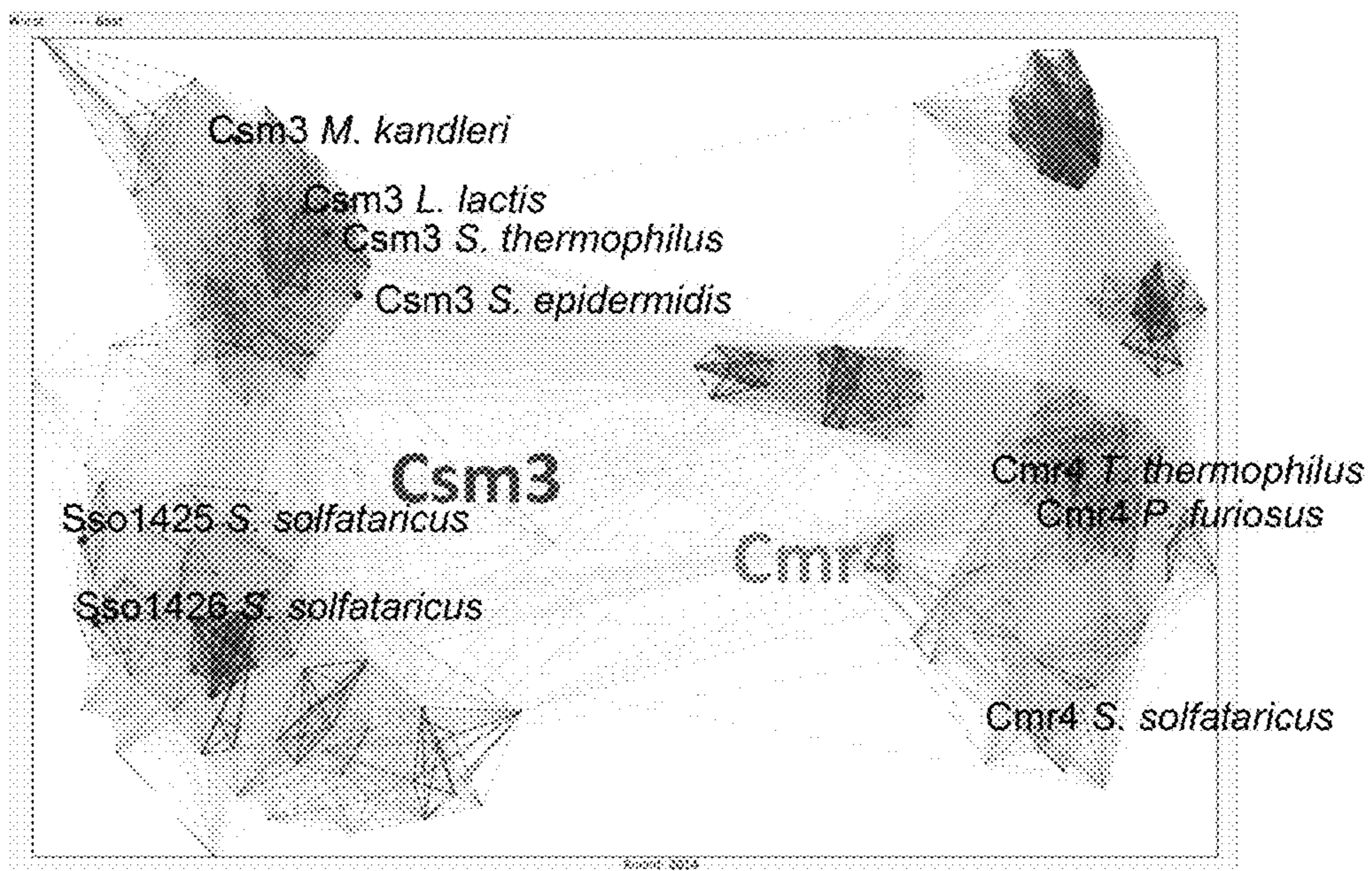


FIG. 15

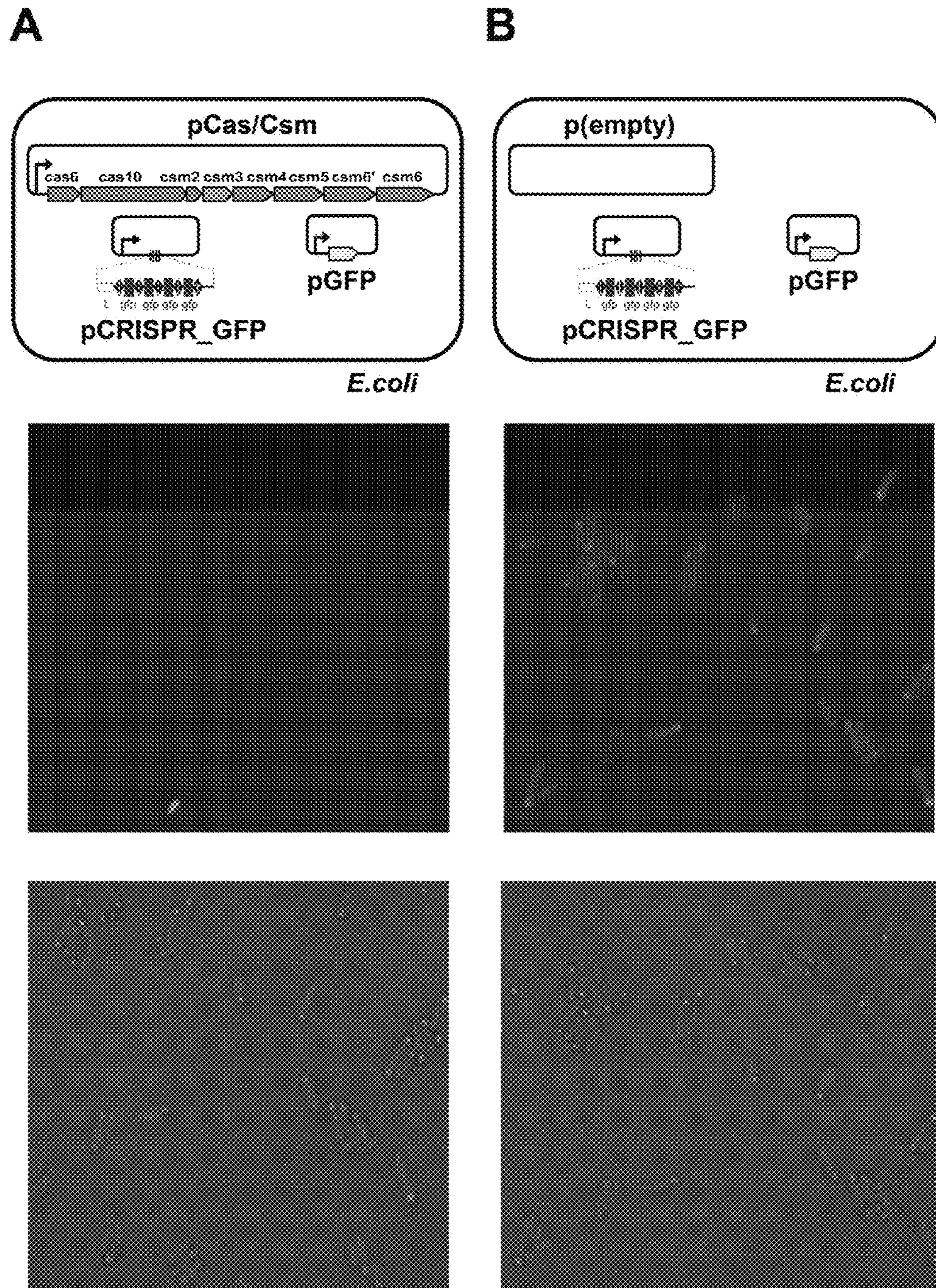


FIG. 15 con't

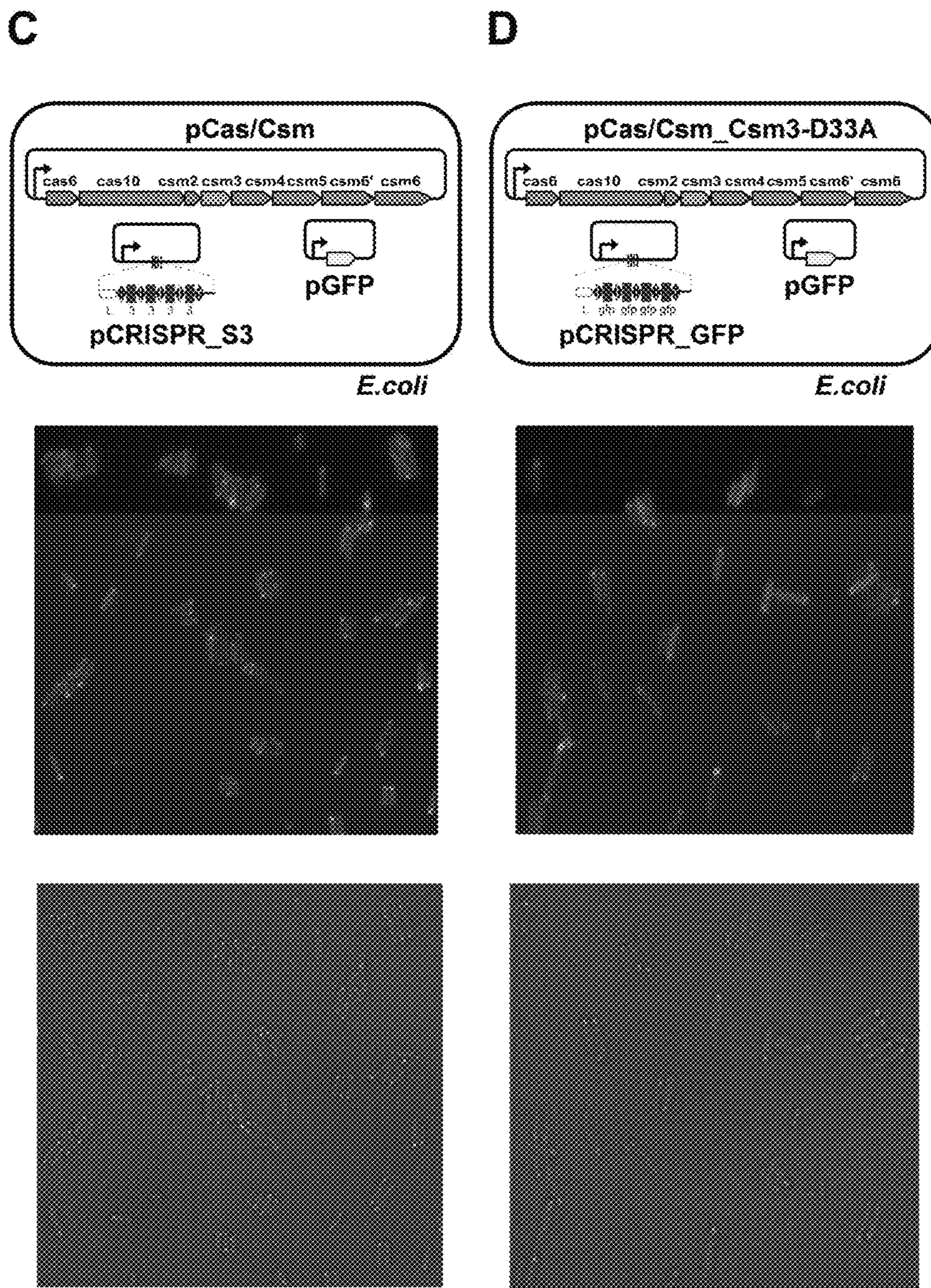


FIG. 16A

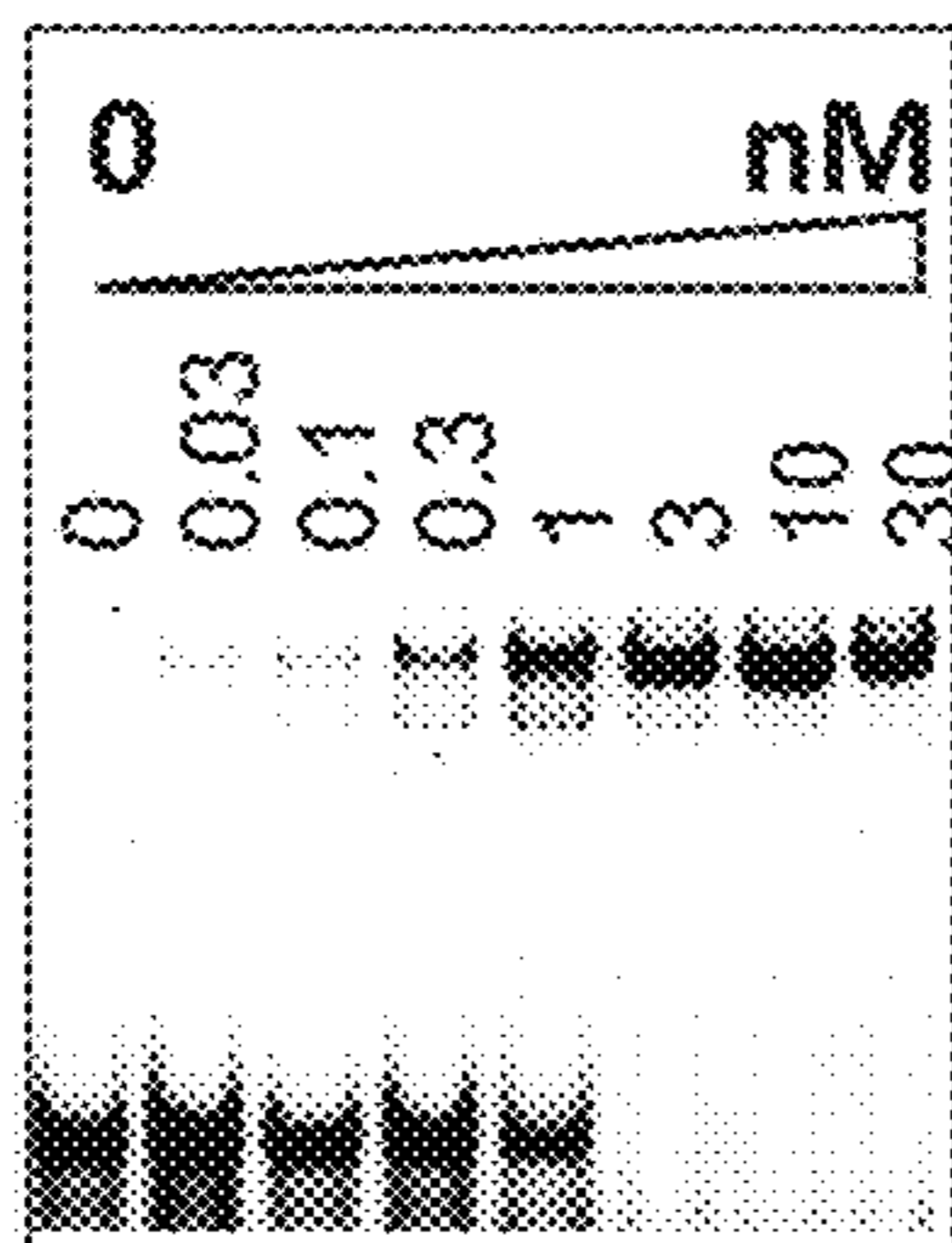
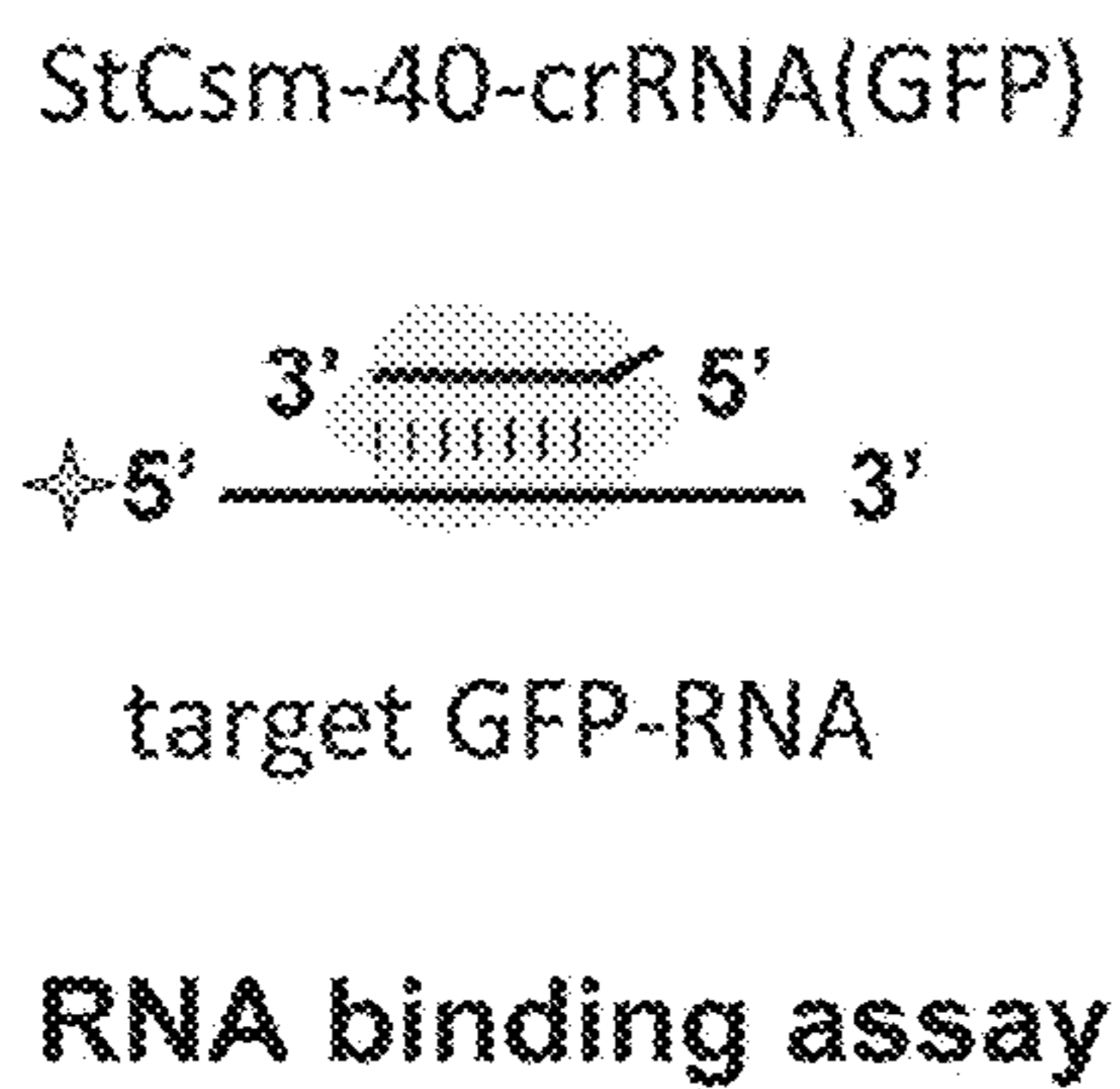


FIG. 16B

RNA cleavage assay

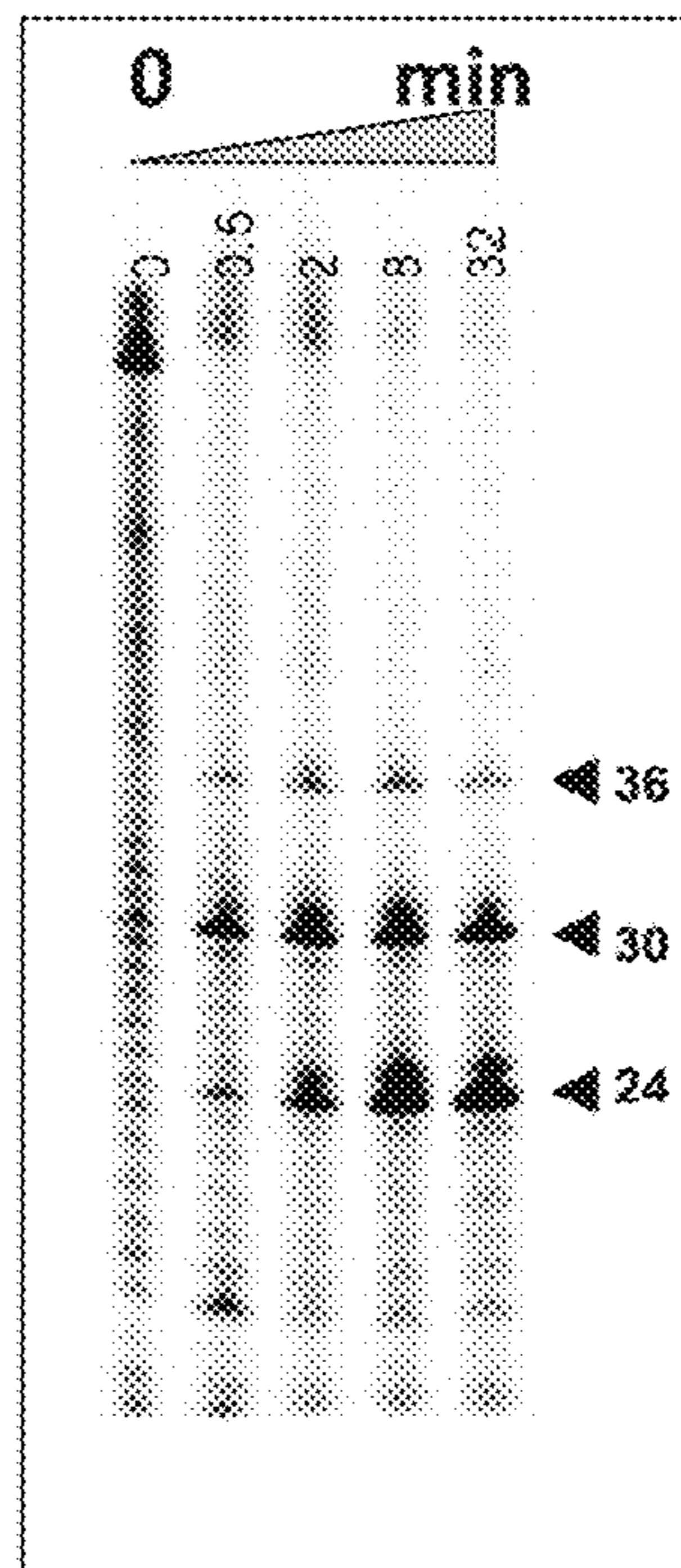


FIG. 17

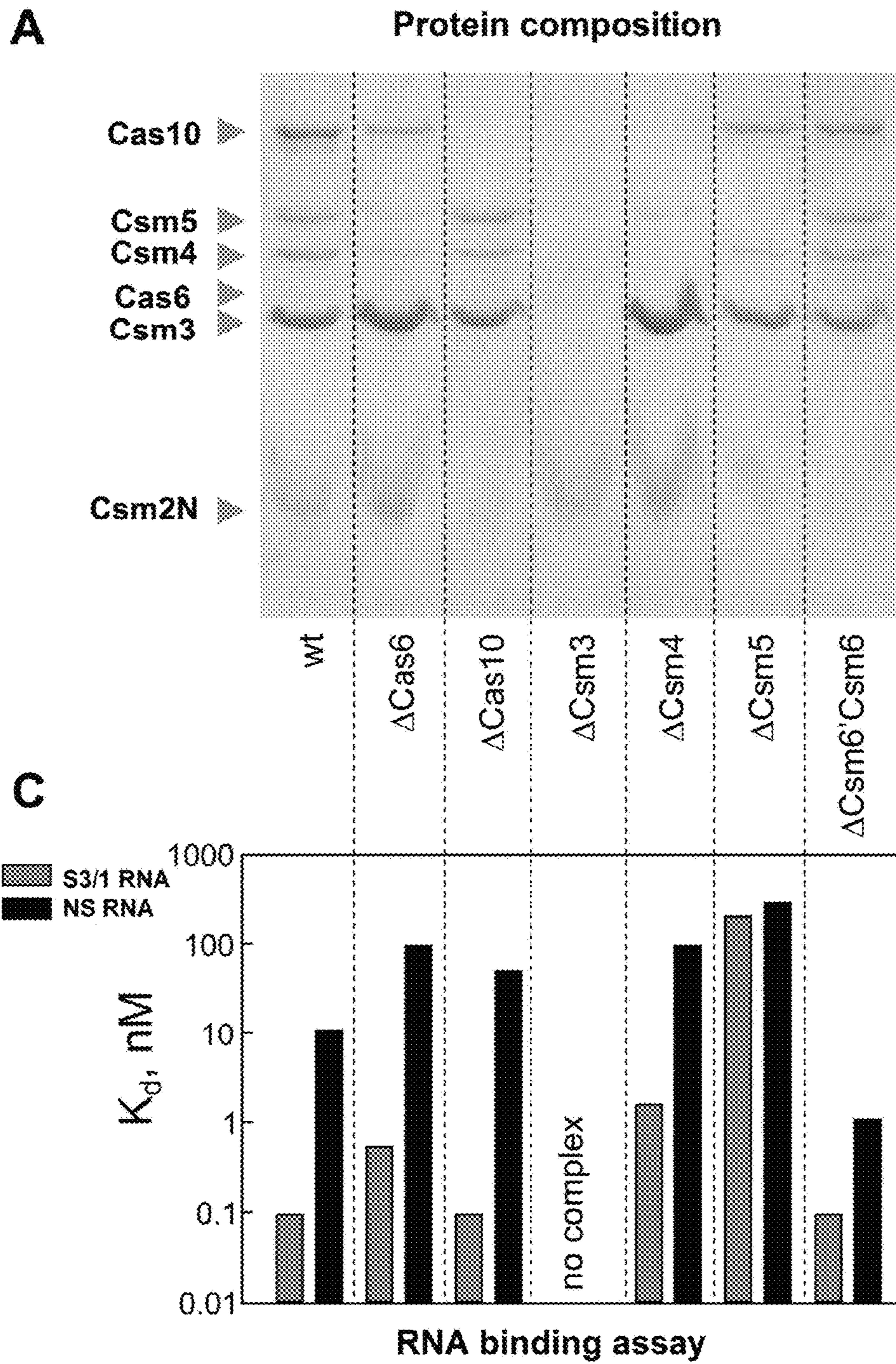


FIG. 17 con't

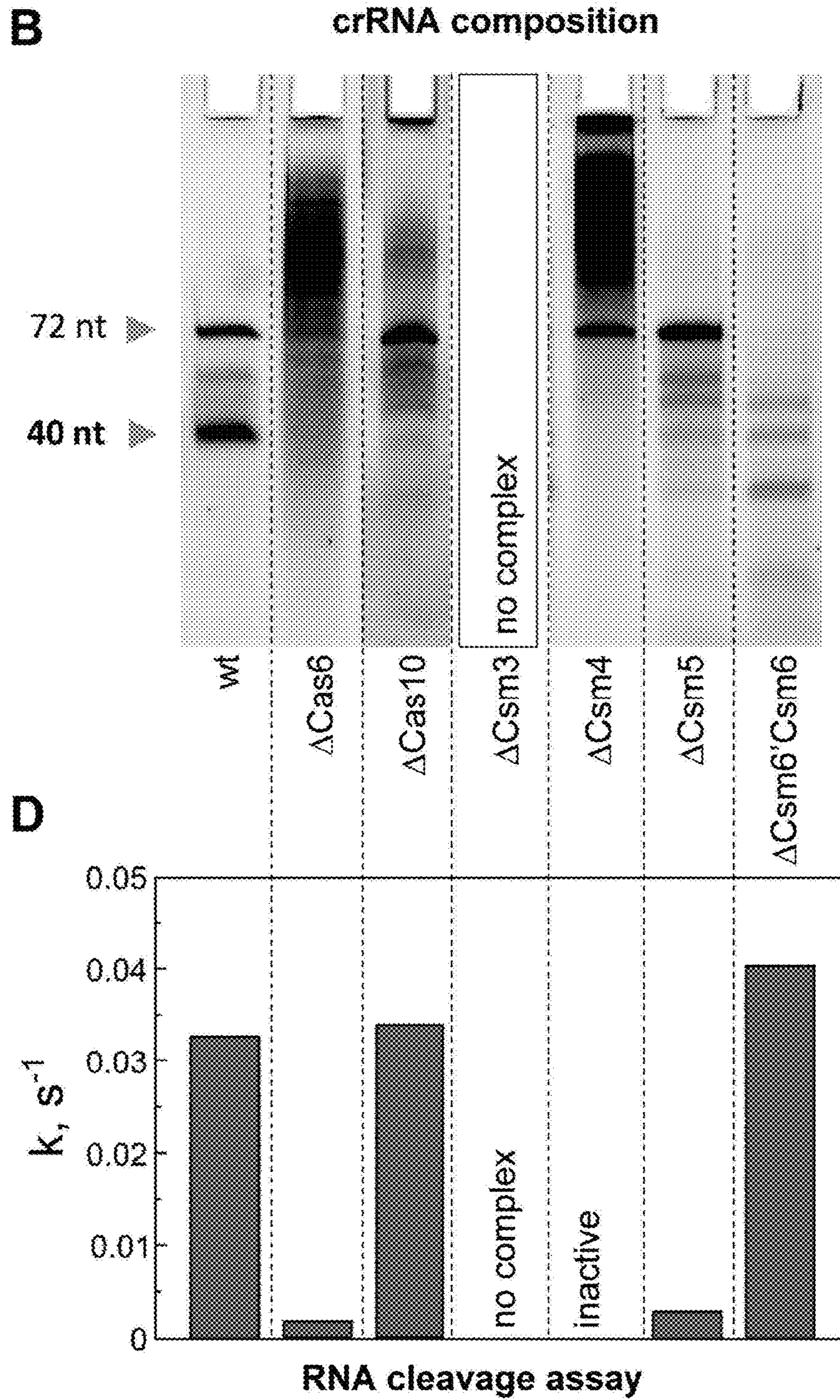


FIG. 18A

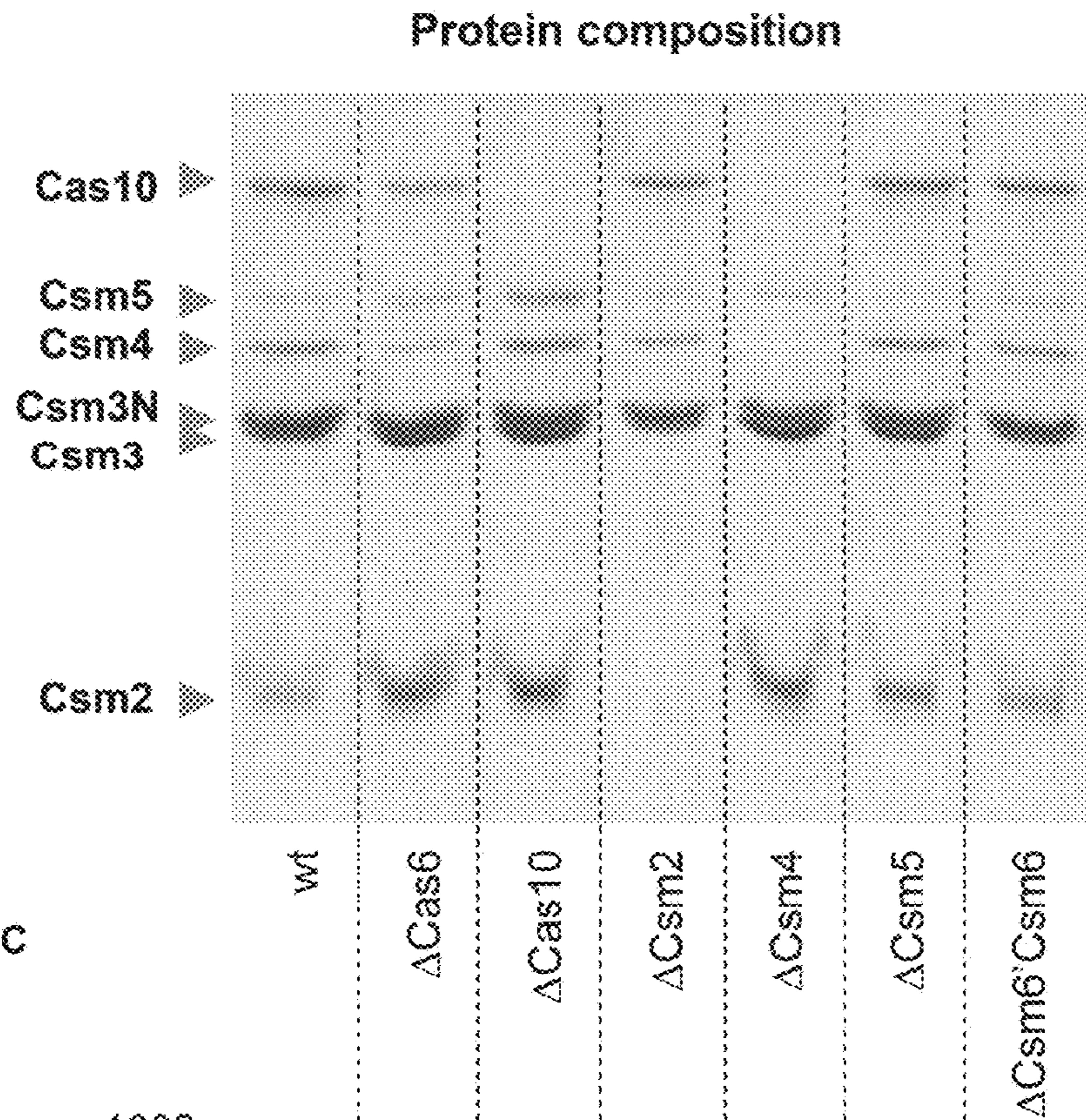


FIG. 18C

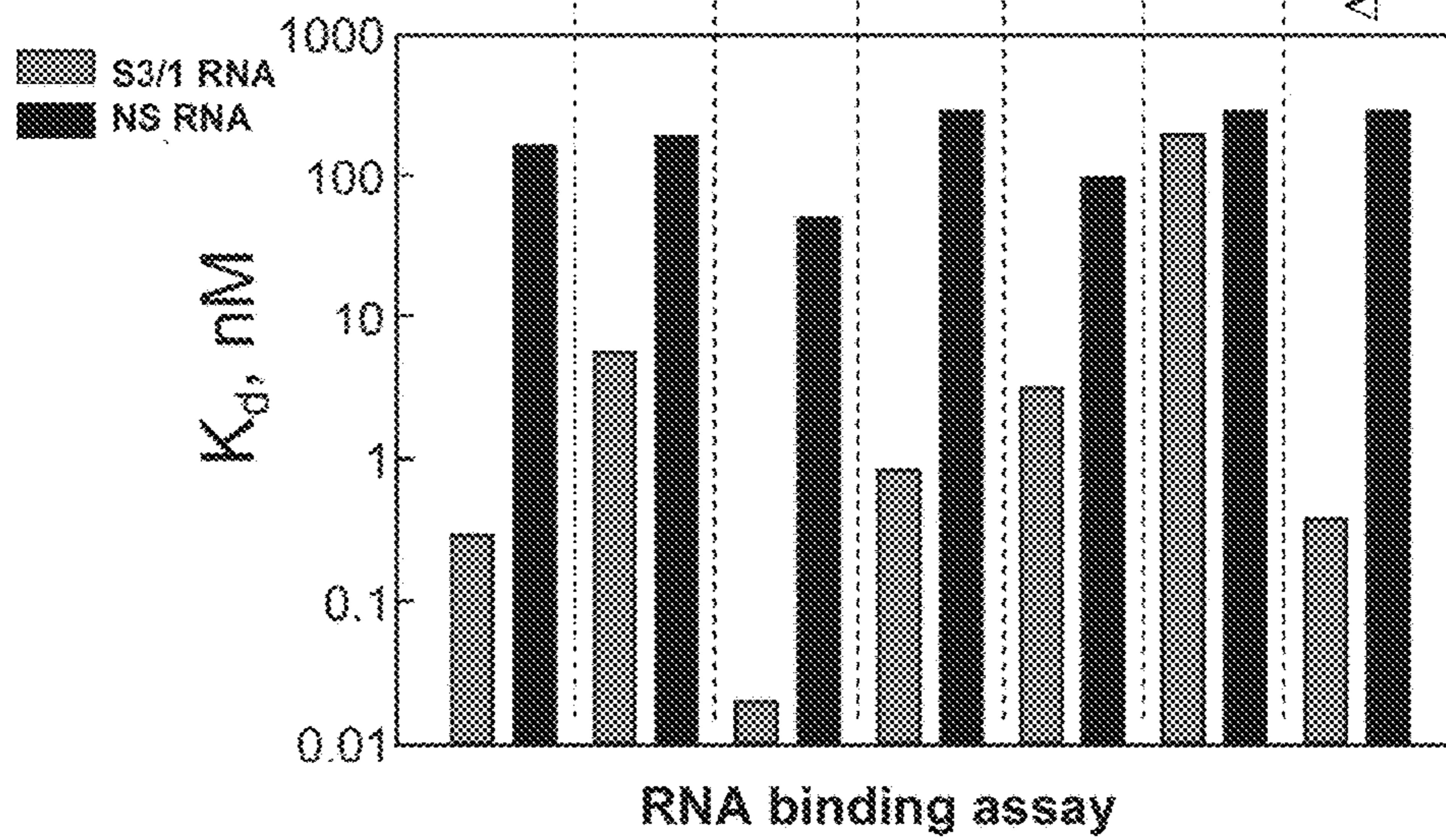


FIG. 18B

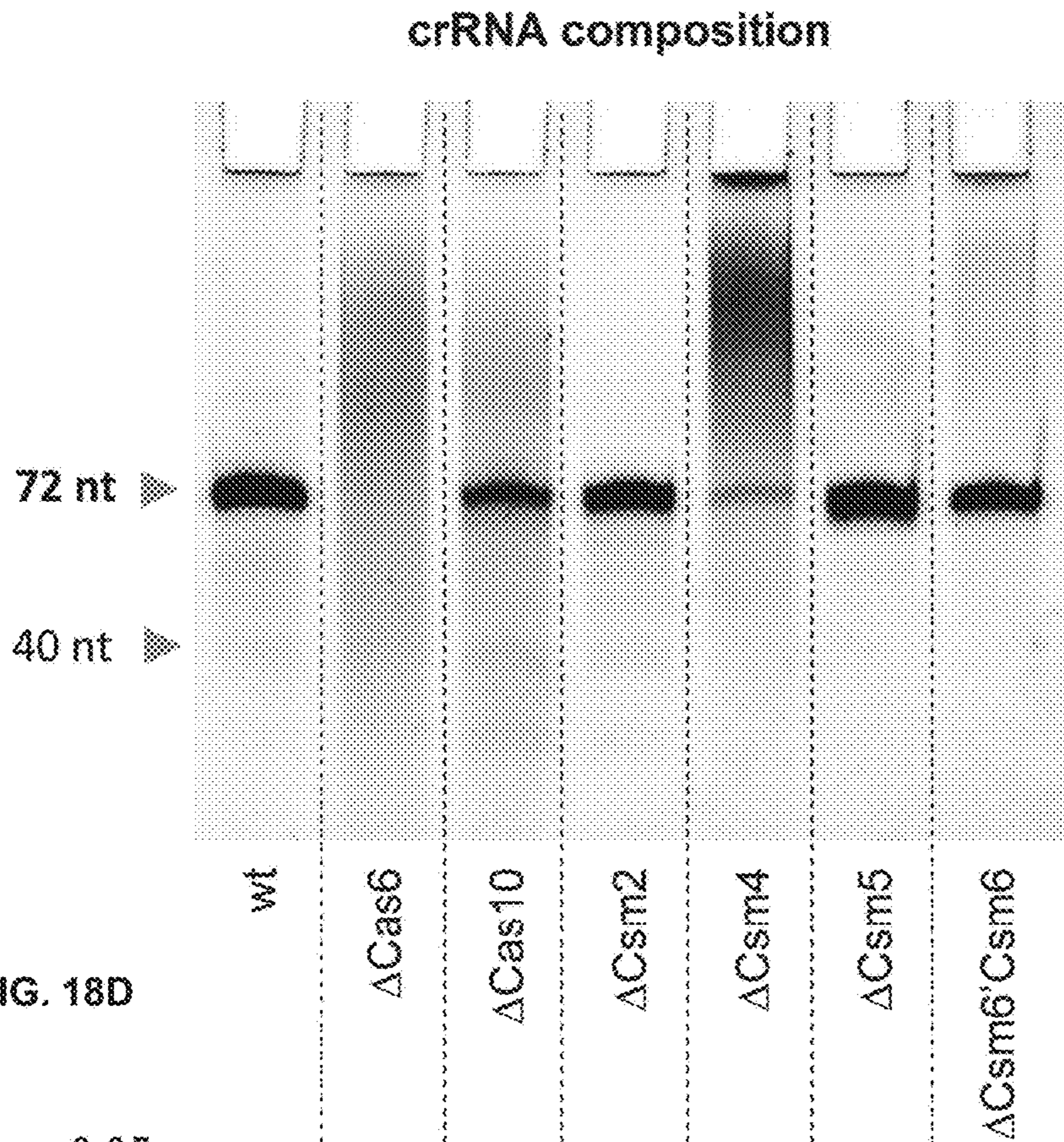


FIG. 18D

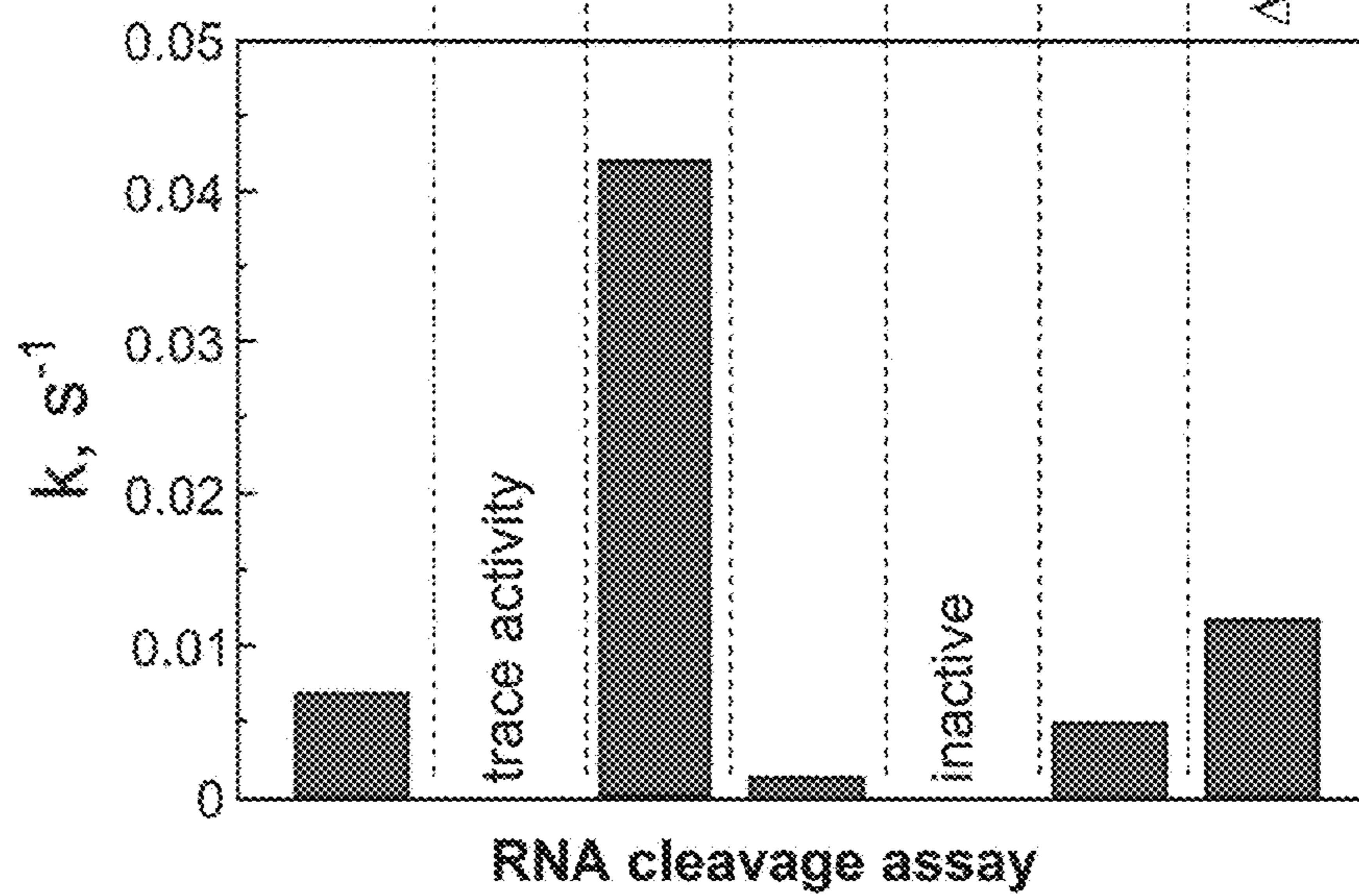


FIG. 19

RNA cleavage assay

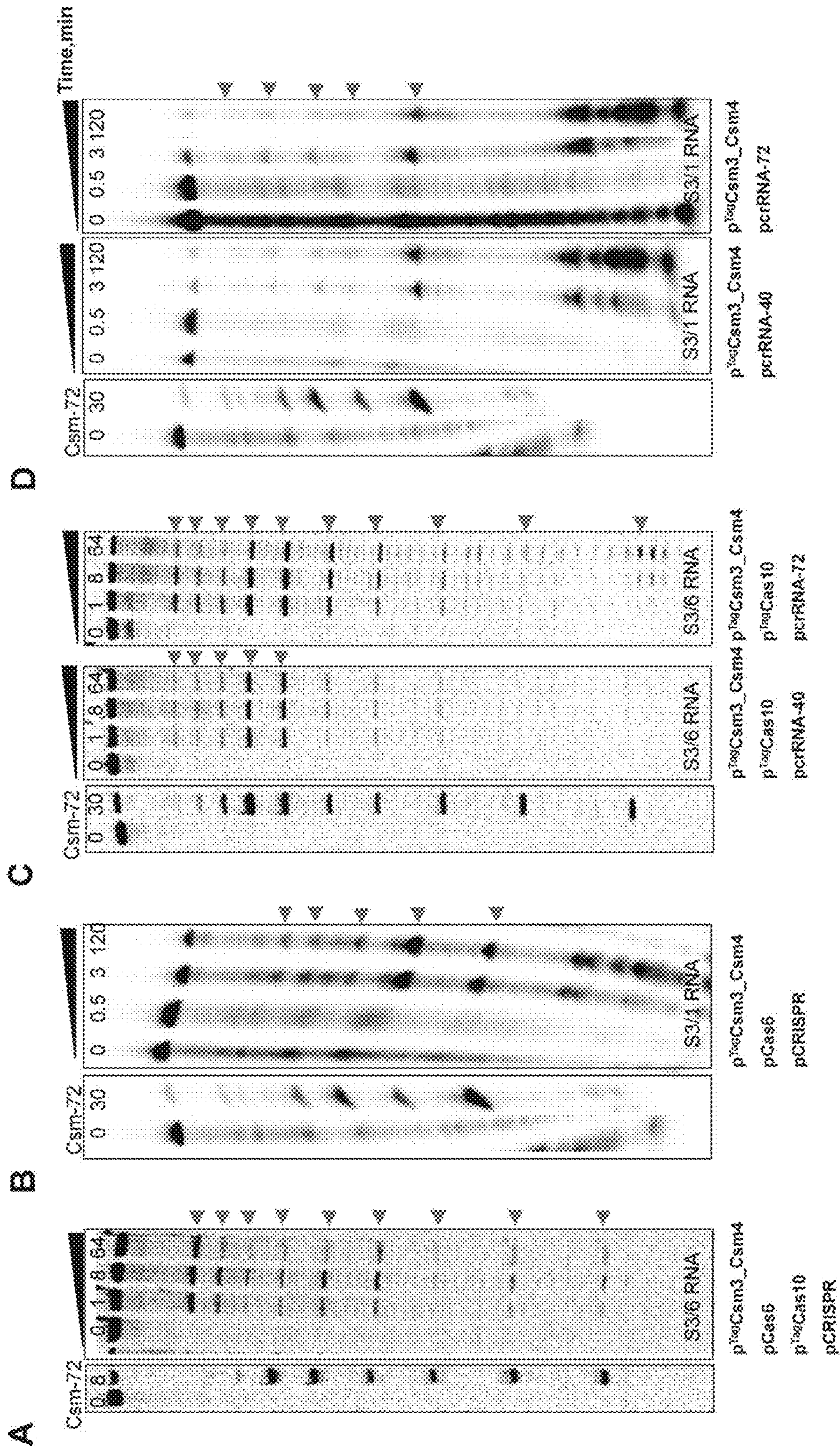
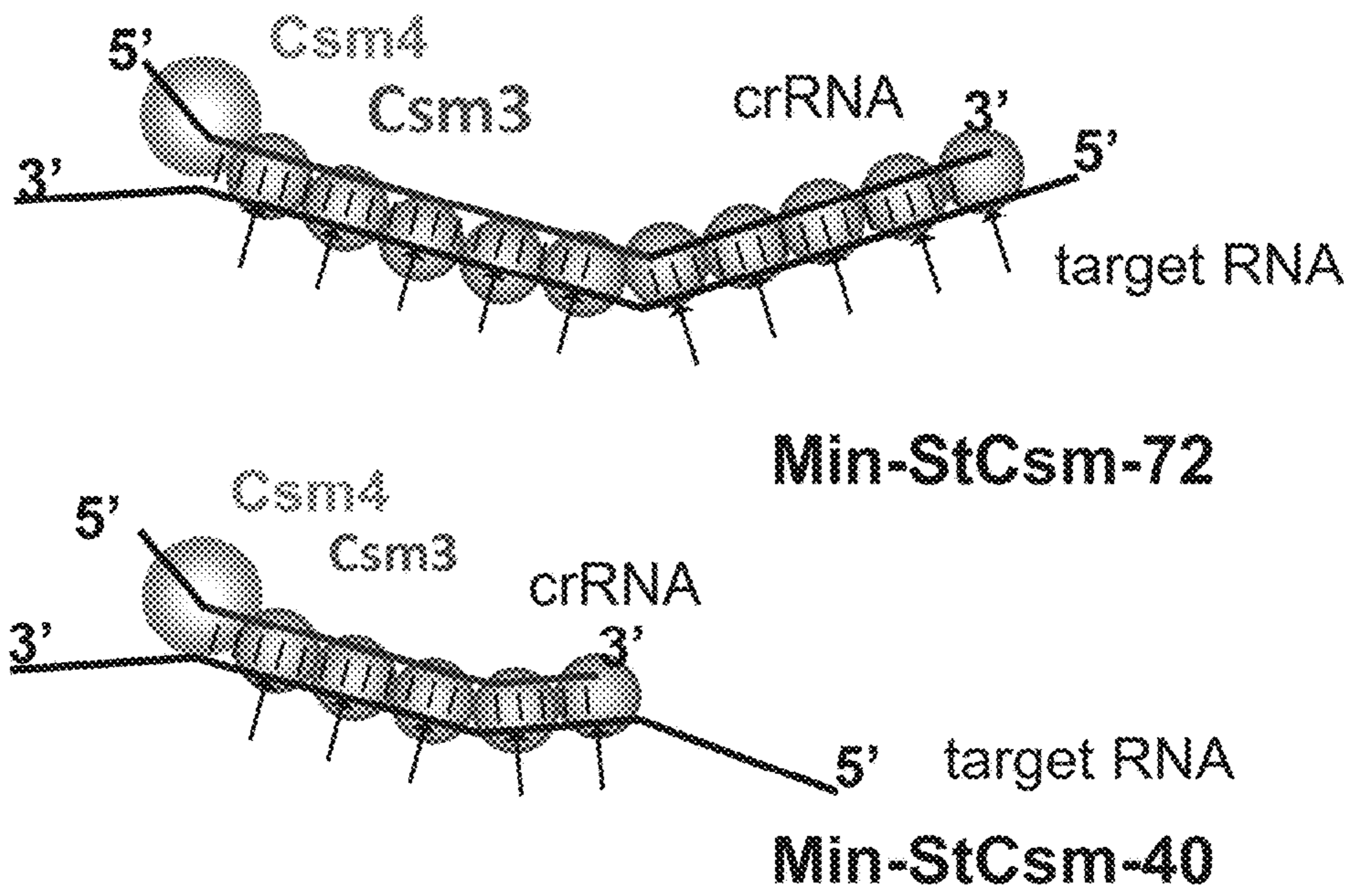


FIG. 20



999688.1

**PROGRAMMABLE RNA SHREDDING BY
THE TYPE III-A CRISPR-CAS SYSTEM OF
*STREPTOCOCCUS THERMOPHILUS***

This application claims priority to U.S. Provisional Appli- 5
cation 62/046,384 filed Sep. 5, 2014 which is expressly
incorporated by reference herein in its entirety.

Immunity against viruses and plasmids provided by
CRISPR-Cas systems relies on a ribonucleoprotein effector
complex that triggers degradation of invasive nucleic acids 10
(NA). Effector complexes of Type I (Cascade) and II (Cas9-
dual RNA) target foreign DNA. Genetic evidence suggests
that Type III-A Csm complex targets DNA, whereas bio-
chemical data show that III-B Cmr complex cleaves RNA.

Disclosed is NA specificity and mechanism of CRISPR- 15
interference for the *Streptococcus thermophilus* Csm (III-A)
complex (StCsm). When expressed in *Escherichia coli*, two
complexes of different stoichiometry co-purified with 40-
and 72-nt crRNA species, respectively. Both complexes
targeted RNA and generated multiple cuts at 6 nucleotide 20
(nt) intervals. The Csm3 protein, present in multiple copies
in both Csm complexes, acts as endoribonuclease. In the
heterologous *E. coli* host StCsm restricts MS2 RNA phage
in Csm3 nuclease-dependent manner.

As subsequently disclosed in detail, the inventors deter- 25
mined that *Streptococcus thermophilus* Type III-A Csm
(StCsm) complex targets RNA. The inventors also deter-
mined that multiple cuts are introduced in the target RNA at
6 nt intervals. Target RNA that is complimentary to crRNA
is cleaved at multiple sites at regular 6 nt intervals, also 30
termed shredding. RNA cleavage is protospacer-adjacent
motif (PAM) independent. A Csm3 subunit is responsible for
endoribonuclease activity of the complex. Because multiple
copies of Csm3 subunits are present in the Csm complex,
cleavage occurs at multiple sites. By systematic deletion of 35
the genes encoding individual subunits, the minimal Csm
complex composition required for target RNA cleavage was
established.

The StCsm complex offers a novel programmable tool for 40
RNA-degradation or modification, e.g., in methods similar
to RNA Interference (RNAi) methods, and using RNAi
methods known in the art. However, different from RNA
interference based methods that rely on the RNAi binding to
the target RNA resulting in the gene product knock-down,
RNA-targeting by the Csm complex allows knock-out of the 45
gene product because the target RNA is cleaved at multiple
sites. If an RNA-cleavage deficient Csm complex
(Csm3D33A) is used, knock-downs instead of knockouts
can be achieved, which provides additional flexibility.

Clustered Regularly Interspaced Short Palindromic 50
Repeats (CRISPR) together with Cas (CRISPR-associated)
proteins provide RNA-mediated adaptive immunity against
viruses and plasmids in bacteria and archaea (Terns and
Terns, 2014). Immunity is acquired through the integration
of invader-derived nucleic acid (NA) sequences as ‘spacers’ 55
into the CRISPR locus of the host. CRISPR arrays are
further transcribed and processed into small interfering
CRISPR RNAs (crRNAs) that together with Cas proteins
assemble into a ribonucleoprotein (RNP) complex which,
using crRNA as a guide, locates and degrades the target NA. 60
CRISPR-Cas systems have been categorized into three
major Types (I-III) that differ by the structural organization
of RNPs and NA specificity (Makarova et al., 2011b).

Type I and II systems provide immunity against invading 65
DNA. In Type I-E systems, crRNAs are incorporated into a
multisubunit RNP complex called Cascade (CRISPR-asso-
ciated complex for antiviral defense) that binds to the

matching invasive DNA and triggers degradation by the
Cas3 nuclease/helicase (Brouns et al., 2008; Sinkunas et al.,
2013; Westra et al., 2012). In Type II systems, CRISPR-
mediated immunity solely relies on the Cas9 protein. It binds
a dual RNA into the RNP effector complex, which then
specifically cuts the matching target DNA, introducing a
double strand break (Gasiunas et al., 2012; Jinek et al.,
2012). In Type I and II CRISPR-Cas systems, the target site
binding and cleavage requires a short nucleotide sequence
(protospacer-adjacent motif, or PAM) in the vicinity of the
target (Mojica et al., 2009). Target DNA strand separation,
necessary for the crRNA binding, is initiated at PAM and
propagates in a directional manner through the protospacer
sequence to yield the R-loop intermediate, one strand of
which is engaged into the heteroduplex with crRNA, while
the other strand is displaced into solution (Sternberg et al.,
2014; Szczelkun et al., 2014). Thus, despite differences in
their architecture, Type I and II RNP complexes share three
major features: i) they act on the invasive double-stranded
DNA (dsDNA), e.g., viral DNA or plasmids, ii) they require
the presence of a PAM sequence in the vicinity of the target
site, and iii) they generate an R-loop as a reaction interme-
diate.

Type III CRISPR-Cas systems are believed to target either
DNA (Type III-A) or RNA (Type III-B) (Makarova et al.,
2011b). In the III-B systems Cas RAMP proteins (Cmr) and
crRNA assemble into a multisubunit RNP complex. Using
crRNA as a guide, this complex in vitro binds single-
stranded RNA (ssRNA) in a PAM-independent manner and
triggers the degradation of target RNA (Hale et al., 2009;
Staals et al., 2013; Zhang et al., 2012). The Cmr effector
complex is comprised of six Cmr proteins (Cmr1, Cas10,
Cmr3-6) that are important for target RNA cleavage; how-
ever roles of the individual Cmr proteins and the ribonu-
clease (RNase) component have yet to be identified. Cmr1,
Cmr3, Cmr4 and Cmr6 are predicted RNA-binding proteins
that share a ferredoxin-like fold and RNA-recognition motif
(RRM) identified in RNA-binding proteins (Terns and Terns,
2014).

The cas genes encoding the RNA-targeting Type III-B
(Cmr) and DNA-targeting Type III-A (Csm) effector com-
plexes share a partial synergy (Makarova et al., 2011a). In
Staphylococcus epidermidis the Csm complex (SeCsm) is
comprised of Cas10, Csm2, Csm3, Csm4, and Csm5 pro-
teins, however the function of individual Csm proteins is
unknown. The evidence that Type III-A systems target DNA
remains indirect and relies on the experimental observation
that Type III-A RNP complex from *Staphylococcus epider-
midis* (SeCsm) limits plasmid conjugation and transforma-
tion in vivo, but the DNA degradation has not been dem-
onstrated directly (Marraffini and Sontheimer, 2008, 2010).
The Csm complex from the archaeon *Sulfolobus solfataricus*
(SsCsm) binds dsDNA, however, it shows no crRNA-
dependent nuclease activity in vitro (Rouillon et al., 2013).
Thus, while the RNA cleavage activity of the Cmr complex
has been characterized in vitro, the DNA degradation activ-
ity of the Type III-A Csm complex has yet to be demon-
strated. The Csm complex so far remains the only CRISPR-
Cas effector complex, for which the function is not yet
reconstituted in vitro. The inventors established the compo-
sition and mechanism of the Csm complex for Type III-A
system *Streptococcus thermophilus* (St) and demonstrated
its RNA cleavage activity both in vitro and in the cell.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 shows cloning, isolation and characterization of
the Type III-A Csm complex of *S. thermophilus* DGCC8004.

(A) Schematic organization of the Type III-A CRISPR2-cas locus (see also FIG. 8). Repeats and spacers are indicated by diamonds and rectangles, respectively, T is for the terminal repeat, L is for the leader sequence, and the arrow indicates the promoter. (B) Protospacer PS3 and the 5'-flanking sequence in the *S. thermophilus* phage 01205 genome. (C) Strategy for expression and isolation of the StCsm complex. Four copies of the spacer S3 have been engineered into the pCRISPR_S3 plasmid to increase the yield of the Csm-crRNA complex. (D) Coomassie blue-stained SDS-PAAG of Strep-tagged Csm2 and Csm3 pull-downs. 3N—Csm3 StrepN protein, M—protein mass marker. (E) Denaturing PAGE analysis of NA co-purifying with the Csm2 StrepN and Csm3 StrepN complexes. M—synthetic DNA marker. (F) and (G) Characterization of crRNA in the isolated StCsm complexes. Cartoon models illustrate crRNA which co-purify with StCsm-72 and StCsm-40 complexes. Composition of the crRNA was determined using LC ESI MS analysis (see also FIG. 9). IPRP HPLC analysis and ESI MS spectra of IPRP HPLC purified crRNA from StCsm-40 and StCsm-72 are presented. (H) Superimposed averaged dummy atom models obtained from SAXS data of StCsm-40 (light beads) and StCsm-72 (dark beads) (see also FIG. 10).

FIG. 2 shows nucleic acid binding and cleavage by the Type III-A Csm complex of *S. thermophilus*. (A) Schematic representation of DNA and RNA substrates used for in vitro binding and cleavage assays. NAs were 5'- or 3'-end labeled with ^{32}P (indicated as *). (B) EMSA analysis of DNA or RNA binding by StCsm-40. NS stands for a non-specific RNA. (C) Binding competition assay. 0.5 nM of ^{32}P -labelled S3/1 RNA was mixed with increasing amounts of unlabelled competitor NAs and 0.3 nM StCsm-40, and analyzed by EMSA. (D) StCsm-40 cleavage assay. Gel-purified DNA or RNA were used as substrates in the NA cleavage assay. Triangles with corresponding numbers indicate cleavage product length. M—RNA Decade marker, R—RNase A digest marker, H—alkaline hydrolysis marker. (E) RNA cleavage products mapped on the S3/1 RNA substrate sequence. Triangles and dashed lines indicate cleavage positions. Short vertical lines above the sequence indicate nucleotides complementary to crRNA. crRNA (StCsm-40) sequence is depicted above the matching substrate fragments.

FIG. 3 shows the effect of the sequence complementarity outside the spacer region on the StCsm-72 cleavage pattern. (A) Schematic representation of the StCsm-72 complex and RNA substrates used in the cleavage assay. RNA substrates were 5'-end labeled with ^{32}P (indicated as *) and gel-purified. (B) StCsm-72 cleavage assay. M—RNA Decade marker. (C) RNA cleavage products mapped on the S3/2, S3/5 and S3/6 RNA substrates sequences.

FIG. 4 shows the effect of protospacer truncations on the StCsm-40 cleavage pattern. (A) Schematic representation of the StCsm-40 complex and RNA substrates used in the cleavage assay. RNA substrates were 5'-end labeled with ^{32}P (indicated as *) and gel-purified. (B) StCsm-40 cleavage assay. M—RNA Decade marker. (C) RNA cleavage products mapped on the RNA substrates sequences.

FIG. 5 shows computational and mutational analysis of Csm3. (A) Alignment of Csm3 and Cmr4 sequence representatives from experimentally characterized Type III effector complexes. Identical and similar residues in more than half of sequences are shaded in dark and light correspondingly. StCsm3 positions subjected to site-directed mutations are indicated by triangles above the sequence. (B) Coomassie blue-stained SDS-PAAG of StCsm complexes containing Csm3 mutants. M—protein marker. (C) Denaturing

PAGE analysis of NA co-purifying with mutant StCsm complexes. M—synthetic DNA marker. (D) EMSA analysis of S3/2 RNA binding by mutant StCsm complexes. (E) S3/2 RNA cleavage by the mutant StCsm complexes. (F) The cleavage rate constant k_{obs} values for Csm3 mutant variants of StCsm-40.

FIG. 6 shows restriction of ssRNA phage MS2 in *E. coli* cells expressing StCsm complex. (A) Schematic representation of the assay. The arrow indicates the promoter. (B) Phage plaque analysis. Serial 10-fold dilutions of MS2 were transferred onto lawns of *E. coli* NovaBlue (DE3, F⁺) strain expressing StCsm-crRNA complex targeting the MS2 genome, or control cells.

FIG. 7 shows structural and cleavage models of StCsm complexes. The CRISPR2 transcript is first processed into 72-nt crRNA intermediates which undergo further maturation into 40-nt crRNA. Both crRNAs are incorporated into StCsm complexes that target RNA but differ by the number of Csm3 and Csm2 subunits. The number of RNA cleavage products correlate with the number of Csm3 nuclease subunits. Schematic models of StCsm complexes were generated based on similarity to TtCmr and PfCmr. Csm analogs of Cmr proteins according to Makarova et al., 2011a are shaded the same, and as indicated.

FIG. 8 shows schematic organization of the Type III-A CRISPR-Cas systems of *Streptococcus thermophilus* DGCC8004, DGCC7710, LMD-9, *Staphylococcus epidermidis* RP62a, *Enterococcus italicus* DSM15952, *Lactococcus lactis* DGCC7167 and *Sulfolobus solfataricus* P2. Schematic organization of the Type III-A CRISPR-Cas systems of *Streptococcus thermophilus* DGCC8004 (GenBank KM222358), DGCC7710 (GenBank AWWZ01000003), LMD-9 (GenBank NC008532), *Staphylococcus epidermidis* RP62a (GenBank NC002976), *Enterococcus italicus* DSM15952 (GenBank AEPV01000074), *Lactococcus lactis* DGCC7167 (GenBank JX524189) and *Sulfolobus solfataricus* P2 (GenBank AE006641)*. Arrows are shaded according to the percentage of identical residues (Vector NTI AlignX tool) in Csm/Cas proteins in respect to the *S. thermophilus* DGCC8004. Conserved repeat sequences are shown in the inserts. Partially palindromic repeat sequences are indicated by arrows. In *L. lactis* DGCC7167 CRISPR2 system lch gene which shows a partial homology to the relE/parE toxin gene is present instead of cas2 (Millen et al., 2012). In CRISPR-Cas loci of *S. thermophilus* LMD-9 and *S. solfataricus* P2 cas10 is split in two open reading frames ORF1 and ORF2. The Type III-A system of DGCC8004 contains 10 cas genes flanking the CRISPR2 array and includes cas1, cas2, cas6, cas10, csm2, csm3, csm4, csm5, csm6 and csm6' genes. The DGCC8004 CRISPR2 locus share similar gene arrangement to that of DGCC7710 (GenBank AWWZ00000000, (Horvath and Barrangou, 2010)) and LMD-9 (GenBank NC_008532, (Makarova et al., 2006)). The major difference is an additional csm6' gene in DGCC8004. The Csm6' protein in DGCC8004 is comprised of 386 aa and shows ~34% amino acid identity to the 428 aa Csm6 protein, suggesting a possible ancient gene duplication event followed by sequence divergence. In contrast, DGCC7710 contains only a short 117-nt ORF in front of csm6. The Cas/Csm proteins associated to CRISPR2 in DGCC8004 are homologous to the corresponding proteins in DGCC7710 and LMD-9 (more than 90% aa identity, except for the Csm2 protein, which shares ~70% identity). Other experimentally characterized Type III-A systems including *S. epidermidis* RP62a (GenBank NC002976, (Marraffini and Sontheimer, 2008)), *Enterococcus italicus* DSM15952 (GenBank AEPV01000074, (Millen et al.,

5

2012)) and *Lactococcus lactis* DGCC7167 (GenBank JX524189, (Millen et al., 2012)) share with DGCC8004 a conserved arrangement of the cas10-csm2-csm3-csm4-csm5-csm6 gene cluster, while the position of cas6 and cas1/cas2 genes differ in some strains. The Type III-A signature protein Cas10 of DGCC8004 shows ~34-40% identity (-50-55% similarity) to Cas10 of *S. epidermidis*, *E. italicus* and *L. lactis*. In LMD-9, the cas10 gene is split into two ORFs which match to the N- and C-terminal fragments of Cas10 in DGCC8004 (>92% identical aa). Type III-A CRISPR-Cas locus in *S. solfataricus* P2 (GenBank AE006641) has different gene organization and shows low protein sequence similarity to Cas/Csm orthologues in DGCC8004. Noteworthy, the Csm3 protein is most conserved among the Cas/Csm proteins across different strains and 5 copies of the Csm3 paralogues are present in *S. solfataricus*. Repeat sequences in *S. epidermidis*, *E. italicus* and *L. lactis* are of the same length (36 nt), however the nucleotide conservation is limited to the palindromic parts and 3'-terminal end of the repeats. The 8-nt 3'-terminal sequence of the repeat, which may contribute to the crRNA 5'-handle, shows an ACGRRAAC consensus between *S. thermophilus*, *S. epidermidis*, *E. italicus* and *L. lactis* but differs from that of *S. solfataricus* (AUUGAAG (Rouillon et al., 2013)).

FIG. 9 shows ESI MS/MS oligoribonucleotide mapping of crRNA isolated from StCsm-72 and StCsm-40. ESI MS/MS oligoribonucleotide mapping of crRNA isolated from StCsm-72 and StCsm-40. (A-C) ESI MS/MS oligoribonucleotide mapping of crRNA isolated from StCsm-72. (A) Base peak chromatogram of RNase T1 digest. RNase T1 cleaves single-stranded RNA 3' of G residues. Predominant oligoribonucleotide peaks of the crRNA are highlighted. Masses of each oligoribonucleotide are presented in the table. The theoretical and experimental masses are shown for the oligoribonucleotides identified. (B) Base peak chromatogram of RNase A digest. RNase A cleaves single-stranded RNA 3' of C or U residues. (C) ESI MS/MS analysis of the oligoribonucleotide GAGAGGGGp. Tandem MS was used to verify the oligoribonucleotide. The predominant fragment ions are highlighted. (D) ESI MS/MS oligoribonucleotide mapping of crRNA isolated from StCsm-40. Base peak chromatogram of RNase T1 digest. The oligoribonucleotide UUCACUUAUUC was unique to the 40-nt crRNA. p=3'-phosphate, >p--2',3'-cyclic phosphate

FIG. 10 shows SAXS data for StCsm complexes. SAXS data for StCsm-40 (black dots) and StCsm-72 (gray dots) are shown. (A) Scattering profiles shown as a logarithmic plot of scattering intensity $I(s)$ vs momentum transfer $s=4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is X-ray wavelength. (B) Kratky plot of SAXS data, $I(s)*s^2$ vs s . (C) Guinier plots of SAXS data, $\ln I(s)$ vs s^2 and its linear fit. The truncated first points are shown as open circles. (D) Distance distribution functions of StCsm-40 and StCsm-72 complexes calculated using GNOM (Svergun, 1992). (E) The electron density of the TtCmr (dark beads), PfCmr (dark beads), and *E. coli* Cascade complexes (dark beads) aligned with the StCsm-40 (light beads) model.

FIG. 11 shows target RNA binding and cleavage by StCsm-72 and StCsm-40 complexes. (A) Schematic representation of S3/1 and S3/2 RNA substrates used in binding and cleavage assays. Nucleic acids were 5'-end labeled with ^{32}P (indicated as *). (B) Electrophoretic mobility shift binding assay. The binding reactions contained ^{32}P -labeled RNA (0.5 nM) and the StCsm-72 or StCsm-40 at concentrations indicated by each lane. Samples were analyzed by

6

PAGE under non-denaturing conditions. NS shows the non-specific RNA control. (C) Cleavage assay. Cleavage reactions were performed at 37° C. for Csm-72 and 25° C. for Csm-40 for indicated time intervals in the Reaction buffer supplemented with 10 mM Mg-acetate, 20 nM RNA substrate and 125 nM StCsm-72 or 62.5 nM StCsm-40. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. In control experiments RNA substrate was incubated for 64 min at 37° C. or 36 min at 25° C. in the Reaction buffer alone ("lane 0") or the storage buffer was added instead of the Csm complex ("lane B"). Triangles denote the reaction products (the sizes of cleavage fragments are given near triangles). M—RNA Decade marker, R—oligoribonucleotide fragments generated from RNase A digestion of RNA, H—alkaline hydrolysis of RNA. (D) RNA cleavage products mapped on the RNA substrates sequence. Triangles and dashed lines indicate cleavage positions. Short vertical lines above the sequence indicate nucleotides complementary to crRNA. 40-nt and 72-nt crRNAs containing spacer S3 sequences are depicted above the matching substrate fragments. NS stands for non-specific RNA. (E) Metal ion (Me^{2+}) dependency of the RNA cleavage by the StCsm complex. S3/1 RNA substrate was pre-incubated with StCsm-40 and reaction products were analyzed in denaturing polyacrylamide gels. Cleavage reactions were conducted at 25° C. for 3 min in Reaction buffer containing 20 nM ^{32}P -5'-labelled gel purified S3/1 RNA substrate, 62.5 nM Csm-40 and 1 mM EDTA, 10 mM Mg-acetate, 10 mM $MnCl_2$, 1 mM Ca-acetate, 0.1 mM $ZnSO_4$, 0.1 mM $NiCl_2$, or 1 mM $CuSO_4$. Triangles and numbers denote the reaction products and their sizes, respectively. M—RNA Decade marker. (F) S3/1 RNA cleavage pattern of the heterogeneous Csm-complex. To express and to purify heterogeneous Csm complex the wt CRISPR2 region containing 13 spacers of *S. thermophilus* DGCC8004 was cloned into the pACYC-Duet-1 vector. The heterogeneous StCsm-72 complex was expressed and purified following the same procedure described for the homogenous Csm complex targeting the S3 protospacer. Specific S3/1 RNA substrate and non-specific NS RNA were pre-incubated with heterogeneous StCsm-72. Cleavage reactions were performed at 37° C. for indicated time intervals in the Reaction buffer supplemented with 10 mM Mg-acetate, 20 nM RNA substrate and 350 nM heterogeneous StCsm-72. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. Triangles and numbers denote the reaction products and their sizes, respectively. M—RNA Decade marker.

FIG. 12 shows reprogramming of the StCsm complex to cleave a desired RNA. (A) Schematic representation (+Tc) and (-Tc) RNA substrates used in the cleavage assay. Arrows indicate Tc^R gene promoter and direction of transcription. RNA substrates were 5'-end labeled with ^{32}P (indicated as *) and gel purified. The Tc (tetracycline resistance protein) gene transcript or RNA corresponding to the non-coding strand of Tc in the pBR322 plasmid (nt 851-886) were used as RNA targets. To reprogram the StCsm complex a synthetic CRISPR locus containing five 36-nt length repeats interspaced by four identical 36-nt spacers complementary to the sense or antisense DNA strands of the Tc gene were engineered into the pACYC-Duet-1 plasmids which were independently co-expressed in *E. coli* together with plasmids pCsm/Cas and pCsmX-Tag. StCsm-40 and StCsm-72 complexes reprogrammed for the sense (+Tc) RNA or anti-sense (-Tc) RNA fragments were isolated similarly to StCsm bearing spacer S3. (B) Cleavage reactions were performed at 37° C. for 120 min in the Reaction buffer supplemented with 10 mM Mg-acetate, 20 nM gel purified

RNA substrate and 40-120 nM of Csm complex. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. Triangles with corresponding numbers indicate cleavage product length. M—RNA Decade marker. (C) RNA cleavage products mapped on the (+Tc) and (–Tc) RNA substrates sequences. The sequences of reprogrammed 40-nt and 72-nt length crRNAs are depicted above the substrates. Short vertical lines above the sequence indicate nucleotides complementary to crRNA. Triangles and dashed lines indicate cleavage positions. Translated fragment which corresponds to the tetracycline resistance protein gene RNA transcript is indicated under (+Tc) RNA substrate.

FIG. 13 shows the effect of crRNA:target RNA complementarity on the StCsm-40 cleavage pattern. (A) Schematic representation of the StCsm-40 complex and RNA substrates used in the cleavage assay. RNA substrates were 5'-end labeled with ^{32}P (indicated as *) and gel purified. (B) Cleavage reactions were performed at 25° C. for indicated time intervals in the Reaction buffer supplemented with 10 mM Mg-acetate, 20 nM RNA substrate and 62.5 nM StCsm-40. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. Triangles with corresponding numbers indicate cleavage product length. M—RNA Decade marker. (C) RNA cleavage products mapped on RNA substrates sequences. Short vertical lines above the sequence indicate nucleotides complementary to crRNA. Triangles and dashed lines indicate cleavage positions. The sequences of both 40-nt and 72-nt length crRNAs containing spacer S3 present in StCsm-40 preparation are depicted above the substrates.

FIG. 14 shows computational analysis of Csm3. (A-C) A structural model of StCsm3 in different representations. (A) Cartoon representation with the core RRM region shown in dark and the “lid” domain shown in light. Active site residue D33 is indicated. (B) Molecular surface of the Csm3 model colored according to sequence conservation (dark—conserved, light—variable). (C) Molecular surface of the Csm3 model colored according to electrostatic potential (dark—positive, light—negative). (D) Clustering of 604 Csm3 and Cmr4 sequence homologs with CLANS. Representatives of Csm3 and Cmr4 families from experimentally characterized (Hale et al., 2009; Hatoum-Aslan et al., 2013; Hrle et al., 2013; Millen et al., 2012; Rouillon et al., 2013; Staals et al., 2013; Zhang et al., 2012) Type III CRISPR-Cas systems are labeled. Each dot represents a sequence, connecting lines represent the similarity between sequences. Thicker lines and shorter distances indicate higher sequence similarity. Only connections corresponding to P-values of 1e-12 or better are shown.

FIG. 15 shows StCsm-triggered GFP transcript degradation in vivo. *E. coli* BL21 (DE3) was transformed with three compatible plasmids (see schematic representation in panels A-D): (i) pCRISPR_MS2 plasmid bearing the synthetic CRISPR array of five repeats interspaced by four 36-nt spacers targeting the GFP transcript; (ii) pCsm/Cas plasmid for the expression of Cas/Csm proteins; (iii) pGFP plasmid for GFP expression. The StCsm and GFP transcript expression was induced in *E. coli* and the GFP transcript degradation was monitored by inspecting GFP fluorescence in the cells. The cells were imaged by contrast (see bottom images in panels A-D) and fluorescence microscopy (see top images in panels A-D).

FIG. 16 shows in vitro cleavage activity of the StCsm targeted to the GFP transcript. Schematic representation of StCsm bound to the target RNA used for in vitro binding and cleavage assays is presented above the gels. RNA was 5'-end labeled with ^{33}P as indicated with asterisk. (A) Gel shift assay of RNA binding by StCsm-40. The binding reactions

contained the ^{33}P -labeled S3/1 RNA (0.5 nM) and StCsm-40 at concentrations indicated by each lane. Samples were analyzed by PAGE under non-denaturing conditions. (B) StCsm-40 cleavage assay. Cleavage reactions were performed at 25° C. for indicated time intervals in the Reaction buffer supplemented with 10 mM Mg-acetate, 8 nM GFP RNA substrate and 160 nM StCsm-40. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. Triangles with corresponding numbers indicate cleavage product length.

FIG. 17 shows protein composition and cleavage in vitro cleavage activity of deletion mutants of StCsm-40 complex. Single-gene deletion variants of pCas/Csm plasmid were generated by disrupting individual cas genes by deletions or frameshift mutations. *Escherichia coli* BL21(DE3) cells were transformed with a corresponding deletion mutant variant of pCas/Csm, pCRISPR_S3, and pCsm2-Tag plasmids. The deletion mutants of StCsm-40 complex were isolated from such cells by Strep-chelating affinity and size exclusion chromatography. (A) Protein composition of the purified StCsm-40 deletion mutant variants as revealed by SDS-PAAGE. In all cases the protein that corresponds to the disrupted cas gene is lacking. In case of ΔCsm4 mutant, the obtained Csm-complex also lacks Cas10, in addition to Csm4. In cells that are deprived of csm3 gene, no complex is pull-downed by the Csm2-Tag subunit. (B) crRNAs co-purified with StCsm-complex deletion mutant variants were extracted using phenol:chloroform:isoamylalcohol and precipitated with isopropanol. Isolated nucleic acids were separated on a denaturing 15% PAAG and visualised by means of SybrGold staining. In cases of ΔCas6 and ΔCsm4 , the purified nucleic acid samples contained a ribonucleic acid molecules of variable size. In almost all cases analysed (ΔCas6 , ΔCas10 , ΔCsm4 , and ΔCsm5 ; with the exception of $\Delta\text{Csm6}'\Delta\text{Csm6}$) the crRNA is not fully matured 40 nt species. However, a band corresponding to 72 nt crRNA was visible in cases of ΔCas10 , ΔCsm4 , and ΔCsm5 mutants. (C) RNA binding affinity of StCsm complex deletion mutant variants was analyzed using electrophoretic mobility shift assay. Two ^{33}P -5'-labeled 68 nt RNA substrates were used for this experiment: specific S3/1 (containing a sequence fully complementary to the 36 nt crRNA encoded by spacer S3; data corresponding to it is depicted in light bars) and the non-complimentary NS RNA (dark bars). Different amounts of the StCsm (0.01-300 nM) were mixed with 0.5 nM of the RNA substrate in the binding buffer containing 1 mM EDTA. Samples were analysed using native 8% PAAG. The dissociation constants (K_d) for RNA binding by StCsm-40 deletion mutants were evaluated assuming the complex concentration at which half of the substrate is bound as a rough estimate of K_d value. Notably, while target RNA binding is significantly decreased in the case of ΔCas6 and ΔCsm4 variants, deletion of Csm5 fully abolishes target RNA binding. (D) RNA cleavage assays for StCsm-40 variants were conducted using the radioactively labelled 68 nt specific S3/1 RNA substrate. Reactions, containing 4 nM S3/1 RNA substrate and 160 nM (or 320 nM) of StCsm in the reaction buffer (33 mM Tris-acetate (pH 7.9 at 25° C.), 66 mM K-acetate, 0.1 mg/ml BSA, and 1 mM Mg-acetate), were initiated by addition of Mg^{2+} ions and performed at 15° C. Consequent reaction products were separated on a denaturing 20% PAGE and depicted by autoradiography. The RNA cleavage rate constants were determined by fitting single exponentials to the substrate depletion data. The obtained constants for each of StCsm variant are depicted in the graph. The cleavage activity of both ΔCas10 and $\Delta\text{Csm6}'\Delta\text{Csm6}$ are similar to wt. In all other cases the

hydrolysis rate is significantly diminished. Deletion of Csm4 completely abolishes StCsm cleavage activity completely.

FIG. 18 shows protein composition and in vitro cleavage activity of deletion mutants of StCsm-72 complex. In order to obtain StCsm-72 deletion mutants, pCas/Csm deletion variants were co-expressed with pCRISPR_S3 and pCsm3-Tag plasmids in *Escherichia coli* BL21(DE3). StCsm-72 complexes were isolated by affinity and size exclusion chromatography. (A) Protein composition of the purified StCsm-72 deletion mutants as revealed by SDS-PAAGE. In all cases the protein that corresponds to the disrupted cas gene is lacking. Δ Csm4 variant, in addition to Csm4, lacks Cas10 subunit. (B) crRNAs that co-purify with StCsm-72 were isolated (as described in FIG. 17), separated on a denaturing 15% PAAG, and visualised by SybrGold staining. In the case of Δ Cas6 and Δ Csm4 variants, the purified nucleic acid co-purified with the Csm-complex deletion variant contained a ribonucleic acid molecules of variable size. In all other cases analysed (Δ Cas10, Δ Csm2, Δ Csm4, and Δ Csm6' Δ Csm6) a clear band, corresponding to 72 nt crRNA is present. (C) Electrophoretic mobility shift assay was employed to evaluate binding affinities of StCsm complexes to the complimentary target (light bars) and non-targeting (dark bars) RNAs. The experiment was performed as as described for StCsm-40 complexes (see FIG. 17). Deletion of Csm5 significantly decreased specific binding. (D) RNA cleavage assays for StCsm-72 variants were carried out similarly as described for StCsm-40 (see FIG. 17 legend). The graph depicts rate constants for the target RNA cleavage. The Δ Cas6 or Δ Csm4 variants of StCsm-72 display almost no activity. In all other cases cleavage products, that are characteristic to StCsm-72, are visible on the gel.

FIG. 19 shows RNA cleavage activity of the minimal StCsm complex variants. According to analysis of the StCsm deletion mutants, Csm3 and Csm4 are absolutely required for complex formation. To co-express the Csm3 and Csm4 proteins csm3 and csm4 genes were cloned into pCDFDuet-1 vector and StrepII-Tag sequence was fused to the N-terminal part encoding csm3 gene to obtain p^{Tag}Csm3_Csm4. p^{Tag}Cas10 plasmid was engineered by cloning the cas10 gene into pETDuet-1 vector. pCas6 plasmid was engineered by cloning the cas6 gene into pCOLA-Duet-1 vector. The expression of Cas6 protein together with pCRISPR (encoding the S3 CRISPR array) would generate 72 nt crRNAs. Alternatively, plasmids from which 40 nt or 72 nt crRNAs could be obtained by in vitro transcription were constructed on the basis of pACYCDuet-1 vector (with cloned BBa J23119 promoter) and named perRNA-40 and perRNA-72, respectively. p^{Tag}Csm3 Csm4 was co-expressed with perRNA-40, perRNA-72, or both pCas6 and pCRISPR plasmids in *Escherichia coli* BL21(DE3). Subsequent Strep-chelating affinity chromatography yielded ribonucleoprotein (RNP) complexes, containing Csm3 and Csm4 proteins. The RNA cleavage activity of these RNP complexes was tested on the complimentary 68 nt S3/4 RNA target or S3/6 RNA target (containing a sequence fully complementary to the 36 nt crRNA encoded by spacer S3). Reactions, containing 4 nM RNA substrate and ~15 ng/ \square of the RNP complex in the reaction buffer (33 mM Tris-acetate (pH 7.9 at 25° C.), 66 mM K-acetate, 0.1 mg/ml BSA, and 10 mM Mg-acetate), were initiated by addition of the purified RNP complex and incubated at 37° C. for the time indicated on the top of the lanes (in minutes). Reaction products corresponding to the wt StCsm-72 complex cleavage are indicated by grey triangles. The same RNA cleavage pattern, characteristic to the wt StCsm-72, is produced by (A) minimised RNP complexes, containing Csm3, Csm4,

Cas10, and crRNA, generated by Cas6, (B) minimal RNP complexes containing Csm3, Csm4, crRNA and Cas6, (C) minimised RNP complexes, containing Csm3, Csm4, Cas10, and crRNA (derived from perRNA-40 or perRNA-72 plasmids), and (D) minimal RNP complexes containing only Csm3 and Csm4 subunits and crRNA.

FIG. 20 shows structural and cleavage model of minimal StCsm complex variants. The minimal catalytically active StCsm complex contains Csm3 and Csm4 proteins and crRNA molecule. The 5'-handle of crRNA is recognized by the Csm4 subunit. Csm3 is endoribonuclease that cuts target RNA. The difference in the number of RNA cleavage positions suggests the different number of Csm3 subunits in the complexes.

In one embodiment, a method for the site-specific modification/shredding of a target RNA molecule is provided, by contacting under suitable conditions, a target RNA molecule and an RNA-guided RNA endonuclease complex comprising at least one RNA sequence and at least two different Csm protein subunits, to result in the target RNA molecule being modified/shredded in a region that is determined by the complimentary binding of the RNA sequence to the target RNA molecule. The method includes incubating under suitable conditions a composition that includes a target RNA molecule with a StCsm complex comprising a polyribonucleotide (crRNA) comprising a 5' handle, a 3' handle, and a spacer which is complementary, or substantially complementary, to a portion of the target RNA. In one embodiment, the crRNA lacks the 3' handle. In one embodiment, the minimal StCsm complex required for target RNA cleavage comprises Csm4 and (Csm3)_X (X=1-10) proteins and 40 or 72 nt crRNA. In embodiments, crRNA is produced by in vitro transcription or chemical synthesis. In embodiments, suitable conditions means conditions in vitro or in vivo where reaction might occur.

In embodiments, the disclosed engineered StCsm complex is used as an RNA Interference tool, to knock-out or knock-down a target RNA, such as mRNA. In one embodiment, Csm3 is modified to include a mutation. One such mutation is D33A, which inactivates the endonuclease activity of Csm3. In various embodiments, the Csm3 D33A may be used to knock-down mRNA expression. Target RNA knock-out results due to the RNA cleavage by the Csm3 protein in the Csm-complex. D33A mutation impairs target RNA cleavage by retains RNA binding ability of the Csm-complex that enables knock-down of the gene product.

StCsm complex might be isolated from a genetically modified microbe (for example *Escherichia coli* or *Streptococcus thermophilus*). In the genetically modified microbe, components of the complex might be encoded on the one, two or three separate plasmids containing host promoters of the genetically modified microbe or promoters from a native host genome.

In one embodiment, a composition is provided, and comprising an engineered StCsm complex comprising crRNA, Csm4, and Csm3. The crRNA of the engineered complex is programmed to guide the StCsm complex to a selected site in a target RNA molecule, wherein the StCsm complex is capable of shredding the target RNA molecule under suitable conditions.

Type III-A CRISPR-Cas loci in *S. thermophilus*

S. thermophilus strain DGCC8004 carries 13 spacers in its Type III-A CRISPR2 array (FIGS. 1A and 8). This strain also contains a Type II CRISPR1 system that is ubiquitous in the *S. thermophilus* species. In the CRISPR2 locus of DGCC8004 the 36-nt repeat sequences, that are partially palindromic, are conserved with the exception of the two

terminal repeats (FIG. 1A). An A+T rich 100-bp leader sequence is located upstream of the CRISPR2 array.

DGCC8004 CRISPR2 (Type III-A) spacers range in size between 34 and 43 nt, but 36-nt spacers are the most abundant. In total, 38 unique spacers were identified among 5 CRISPR2-positive *S. thermophilus* strains and a majority (20 out of 38) of these spacer sequences have matches (protospacers) in *S. thermophilus* DNA phage sequences, although phage interference for the *S. thermophilus* CRISPR2 locus has not yet been demonstrated. Analysis of the sequences located immediately upstream and downstream of these protospacers failed to identify any consensus sequence as a putative PAM, either due to the relatively small number of protospacers or targeting of RNA that is often PAM-independent (Hale et al., 2009). In DGCC8004, 15 although no CRISPR2 spacer gives perfect identity with currently known sequences, 6 spacers out of 13 (S3, S4, S6, S8, S12 and S13) show strong sequence similarity with *S. thermophilus* DNA phages (at least 94% identity over at least 80% of spacer length). All phage matching protospacers appear to have been selected from the template strand. For example, the 36-nt spacer S3 matches 34 nt of a protospacer in the *S. thermophilus* phage 01205 genome (FIG. 1B). A corresponding crRNA would match the template DNA strand of the protospacer S3, and would pair with the target sequence on the coding strand of phage DNA or the respective mRNA sequence. If crRNA processing in the *S. thermophilus* Type III-A locus was similar to that in *S. epidermidis* (Hatoum-Aslan et al., 2011; Hatoum-Aslan et al., 2014; Hatoum-Aslan et al., 2013), the resulting crRNA 5'-handle in the mature crRNA would be non-complementary to the protospacer S3 3'-flank in the phage DNA coding strand or mRNA (FIG. 1B). In the *S. epidermidis* Type III-A system, which limits the spreading of plasmid DNA, the crRNA/target DNA non-complementarity outside of the spacer sequence plays a key role in silencing of invading DNA and self vs non-self DNA discrimination (Marraffini and Sontheimer, 2010). Taking these elements into consideration, crRNA encoded by the spacer S3 was selected as the guide, and a complementary protospacer sequence as the NA target (DNA or RNA) (FIG. 1B).

Cloning, Expression, and Isolation of the *S. thermophilus* DGCC8004 Type III-A Effector Complex

To isolate the Type III-A RNP effector complex (StCsm) of the DGCC8004, the CRISPR2 locus was split into the three fragments and cloned them into three compatible vectors (FIG. 1C). Plasmid pCas/Csm contained a cassette including all the cas/csm genes (except cas1 and cas2), while plasmid pCRISPR_S3 carried 4 identical tandem copies of the repeat-spacer S3 unit flanked by the leader sequence and the terminal repeat. Plasmids pCsm2-Tag or pCsm3-Tag carried a StrepII-tagged variant of csm2 or csm3 genes, respectively. Next, all three plasmids were co-expressed in *E. coli* BL21(DE3) and tagged Csm2 or Csm3 proteins were isolated by subsequent Strep-chelating affinity and size 55 exclusion chromatography.

Strep-tagged Csm2 or Csm3 proteins pulled-down from *E. coli* lysates co-purified with other Csm/Cas proteins suggesting the presence of a Csm complex (FIG. 1D). Csm complexes isolated via N-terminus Strep-tagged Csm2 (Csm2 StrepN) and the N-terminus Strep-tagged Csm3 proteins (Csm3 StrepN) were subjected to further characterization. SDS-PAGE of these complexes revealed six bands that matched the individual Cas proteins Cas6, Cas10, Csm2, Csm3, Csm4 and Csm5 (FIG. 1D). The identity of proteins in these Csm complexes was confirmed by mass spectrometry (MS) analysis (Tables 1 and 2).

The Csm complexes were examined for the presence of NA using basic phenol-chloroform extraction followed by RNase I or DNase I digestion. Denaturing PAGE analysis revealed that ~70-nt and ~40-nt RNA molecules co-purified with the Csm3 StrepN and Csm2 StrepN pulled-down Csm complexes, respectively (FIG. 1E). The complex isolated via Csm2 StrepN subunit also contained ~10% of the ~70-nt RNA. When subjected to RNase I protection assay the RNA in the complexes showed no visible degradation, indicating that the RNA is tightly bound and protected along its entire length (data not shown).

Characterization of the crRNA

Denaturing RNA chromatography was used in conjunction with electrospray ionization mass spectrometry (ESI-MS) to analyse the crRNA sequence and determine the chemical nature of the 5'- and 3'-termini of crRNAs co-purified with both Csm complexes. Denaturing ion pair reverse phase chromatography was used to rapidly purify the crRNA directly from the Csm complexes. The crRNA isolated from the Csm3_StrepN pull-down complex revealed a single crRNA with a retention time consistent with an approximate length of 70 nt (FIG. 1F). The crRNA isolated from Csm2_StrepN pull-down complex revealed the presence of an additional crRNA, with a retention time consistent with an approximate length of 40 nt (FIG. 1G). Purified crRNAs were further analyzed using ESI-MS to obtain the accurate intact masses. A molecular weight of 22 998.5 Da was obtained for RNA isolated from Csm3 and 12 602.2 Da for RNA isolated from Csm2 pull-downs, respectively. Csm2 pull-down also contained a minor component, with a molecular weight of 12 907.3 Da (data not shown). In addition, ESI MS/MS was also used to analyze the oligonucleotide fragments generated from RNase A/T1 digestion of the crRNAs (FIG. 9). In conjunction with the intact mass analysis, these results revealed a 72-nt crRNA in the complex isolated via Csm3 (further termed Csm-72 according to the length of crRNA) and a 40-nt crRNA in the complex isolated via Csm2 (further termed Csm-40 complex). The MS analysis of the 72-nt crRNA is consistent with the pre-CRISPR cleavage at the base of the CRISPR RNA hairpin to yield a 8-nt 5'-handle, a 36-nt spacer and a 28-nt 3'-handle with 5'-OH and 3'-P, and could represent immature crRNA intermediate (FIG. 1F) similar to that of Type III-A and III-B CRISPR-Cas systems (Hale et al., 2009; Hatoum-Aslan et al., 2013). Further verification of the 3'-P termini was obtained upon acid treatment of the 72-nt crRNA where no change in mass was observed using ESI-MS. Likewise, the MS analysis of the 40-nt crRNA in the Csm-40 complex revealed an 8-nt 5'-handle and a 32-nt spacer with 5'-OH and 3'-OH that would correspond to the mature crRNA (FIG. 1G). The difference in the chemical nature of the 3'-end between intermediate and mature crRNAs suggests that primary processing and final maturation are achieved by distinct catalytic mechanisms as proposed by Hatoum-Aslan for the *S. epidermidis* model system (Hatoum-Aslan et al., 2011).

Composition and Shape of the Csm Complex

Evaluation of the complex composition by densitometric analysis of the SDS gels suggests the Cas₁₀₁:Csm₂₆:Csm₃₁₀:Csm₄₁:Csm₅_{0.14} stoichiometry for Csm-72, and the Cas₆_{0.10}:Cas₁₀₁:Csm₂₃:Csm₃₅:Csm₄₁:Csm₅₁ stoichiometry for Csm-40. Fraction numbers for Cas6 and Csm5 proteins are presumably due to the weak transient interactions of these proteins in the respective complexes. Protein subunits that are involved in pre-crRNA processing, e.g. Cas6, would not necessarily occur in stoichiometric amounts in the purified effector complex.

Small angle X-ray scattering (SAXS) measurements was also performed in order to characterize the molecular mass/shape of both Csm-40 and Csm-72 effector complexes in solution. M_w values obtained using SAXS are in agreement both with DLS and gel-filtration data (Table 3). Taken together these data are consistent with the stoichiometry Cas10₁:Csm2₆:Csm3₁₀:Csm4₁:crRNA₁ (calculated Mw 486.2 kDa including the 72-nt crRNA) for Csm-72 and Cas10₁:Csm2₃:Csm3₅:Csm4₁:Csm5₁:crRNA₁ (calculated Mw 344.8 kDa including 40-nt crRNA) for Csm-40.

SAXS measurements revealed that the Csm-40 complex in solution has elongated and slightly twisted shape. The maximal interatomic distance (D_{max}) of the complex estimated from SAXS data is 215 Å, whereas its diameter is 75-80 Å (Table 4). The shape of this effector complex (FIG. 1H) is very similar to the electron microscope structure of Cmr complexes from *Thermus thermophilus* (Staals et al., 2013), *Pyrococcus furiosus* (Spilman et al., 2013) and Cascade from *E. coli* (Wiedenheft et al., 2011) (FIG. 10E). The Csm-72 complex with D_{max} of 280 Å (Table 4) is significantly more elongated than the Csm-40 complex (FIG. 1H). The lowest normalized spatial discrepancy was obtained for the end-to-end superimposition of the Csm-40 and Csm-72 models (FIG. 1H).

Nucleic Acid Specificity of the Type III-A StCsm Complex

In the CRISPR2 locus of DGCC8004, 34 out of 36 nt of the spacer S3 match a sequence present in the genome of *S. thermophilus* phage 01205. Thus, to probe the functional activity of the Csm-40 complex, DNA and RNA substrates were first designed that are fully complementary to the 32-nt crRNA encoded by spacer S3 and that carry phage 01205-flanking sequence. These flanking sequences lack complementarity to the 8-nt 5'-handle of the crRNA identified in the Csm-40/Csm-72 complexes (FIG. 2A and Table 5). For binding analysis DNA or RNA substrates were 5'-end radioactively labeled and the Csm-40 complex binding was evaluated by an electrophoretic mobility shift assay (EMSA) in the absence of any divalent metal (Me^{2+}) ions. Csm-40 showed weak affinity for oligoduplex S3/1 DNA/DNA and DNA/RNA substrates since binding was observed only at high (100-300 nM) complex concentrations. Single-stranded S3/1 DNA (ssDNA) was bound to Csm-40 with an intermediate affinity ($K_d \square 30$ nM), whereas a single-stranded S3/1 RNA (ssRNA) showed high affinity binding ($K_d \square 0.3$ nM) (FIG. 2B). Binding competition experiments with various nucleic acids further supported the single-stranded RNA specificity for the Csm-40 complex (FIG. 2C). Cleavage data correlated with the binding affinity: S3/1 DNA/DNA, DNA/RNA and ssDNA are refractory to cleavage, whereas S3/1 ssRNA complementary to the crRNA is cut by Csm-40 in the presence of Mg^{2+} ions (FIG. 2D). RNase activity of Csm-40 complex requires Mg^{2+} or other divalent metal ions (Mn^{2+} , Ca^{2+} , Zn^{2+} , Ni^{2+} or Cu^{2+}) and is inhibited by EDTA (FIG. 11E).

Csm-40 cuts the S3/1 RNA target at 5 sites regularly spaced by 6-nt intervals to produce 48-, 42-, 36-, 30- and 24-nt products, respectively (FIGS. 2D, 2E). The sequence complementarity between the crRNA in the complex and the RNA target is a key pre-requisite for the cleavage: a non-specific RNA (FIG. 2E, bottom) was resistant to Csm-40. The Csm-40 cleavage pattern of the 3'-labeled S3/1 RNA substrate differs from that of the 5'-labeled variant. While the 5'-labeled substrate cleavage produces 48-, 42-, 36-, 30- and 24-nt products, short degradation products of 21, 27, and 33 nt (1 nt shift is due to an additional nucleotide added during the 3'-labeling) are visible on the gel (FIGS. 2D, 2E). Taken together, cleavage data for the 5'- and 3'-end labeled RNA

substrates suggest that Csm-40 cuts the RNA molecule initially at its 3'-end and endonucleolytic degradation is further extended towards the 5'-end with 6-nt increments.

The Csm-72 complex carrying a 72-nt crRNA (8-nt 5'-handle plus 36 nt of the spacer S3 and 28 nt of the 3'-handle, FIG. 11A) showed ~30-fold weaker binding affinity ($K_d \square 10$ nM) to S3/1 RNA in comparison to the Csm-40 (FIG. 11B). Nevertheless, similar to the Csm-40 complex, in the presence of Mg^{2+} ions Csm-72 cleaved S3/1 RNA, albeit at a decreased rate which may correlate with its weaker binding affinity (FIG. 11C). The 5'- and 3'-labeled S3/1 RNA cleavage pattern is identical to that of Csm-40 (FIGS. 11C, 11D and data not shown). Like the Csm-40 complex, Csm-72 showed no cleavage of S3/1 ssDNA, DNA/DNA or DNA/RNA substrates (data not shown). The heterogeneous Csm complex isolated from the *E. coli* host carrying the wt CRISPR array containing 13 spacers produces RNA cleavage products identical to those of the homogenous StCsm (FIG. 11F). Taken together, these data unambiguously demonstrate that Csm-40 and Csm-72 complexes in vitro target RNA but not DNA, and cut RNA at multiple sites regularly spaced by 6-nt intervals.

Reprogramming of the StCsm Complex

To demonstrate that the Type III-A StCsm complex can be reprogrammed to cut a desired RNA sequence in vitro, we designed and isolated Csm complexes loaded with crRNA (+Tc) or crRNA(-Tc) targeting, respectively, the 68-nt sense (+) and anti-sense(-) mRNA fragments obtained by in vitro transcription of the tetracycline (Tc) resistance gene in the pBR322 plasmid (nts 851-886) (FIG. 12A and Table 5). Both Csm-40 and Csm-72 complexes guided by the crRNA(+Tc) sliced the complementary sense RNA fragment but not the antisense RNA sequence (FIG. 12B). In contrast, Csm-40 and Csm-72 complexes guided by the crRNA(-Tc) cleaved antisense RNA but not a sense Tc mRNA fragment (FIG. 12B). In both cases target RNA was cleaved at multiple sites regularly spaced by 6-nt intervals (FIG. 12C).

To demonstrate that StCsm complex can be reprogrammed to cut the desired RNA target and silence gene expression in vivo, we designed Csm complexes targeting the GFP gene transcript in the heterologous *E. coli* host. *E. coli* BL21 (DE3) was transformed with three compatible plasmids: (i) pCRISPR_MS2 plasmid bearing the synthetic CRISPR array of five repeats interspaced by four 36-nt spacers targeting the GFP transcript; (ii) pCsm/Cas plasmid for the expression of Cas/Csm proteins; (iii) pGFP plasmid for the GFP expression, and the GFP transcript degradation was monitored by inspecting GFP fluorescence in *E. coli* cells (FIG. 15). No GFP fluorescence is detected when wt StCsm targeting the GFP transcript is expressed in *E. coli* (FIG. 15A). On the other hand, GFP fluorescence is observed in *E. coli* cells lacking the StCsm complex (FIG. 15B), or bearing the RNA-cleavage deficient Csm3-D33A mutant complex (FIG. 15C), or containing StCsm complex with CRISPR2 S3 crRNA (FIG. 15D). In a separate set of the experiments we show that isolated StCsm-40 loaded with GFP mRNA targeting crRNA (GFP) specifically binds and cuts 68 nt length GFP RNA in vitro (FIG. 16).

Target RNA Determinants for Cleavage by crRNA-Guided Csm Complex

Whether the nucleotide context downstream or upstream of the protospacer sequence modulates RNA cleavage by the Csm complexes was further examined. To this end, the S3/2 RNA substrate was designed in which the flanking regions originating from O1205 phage DNA in the S3/1 substrate are replaced by different nucleotide stretches that are non-complementary to the 5'-handle of crRNA in the Csm-40 and

Csm-72 complexes, and to the 3'-handle in the Csm-72 complex. RNA binding and cleavage data showed that despite differences in the nucleotide context of flanking sequences in the S3/1 and S3/2 substrates, cleavage patterns for the Csm-40 and Csm-72 complexes are nearly identical, except for an extra 18-nt product for the Csm-72 (FIG. 11C).

Whether the base-pairing between the flanking sequences of the RNA target and 5'- and 3'-handles of crRNA in the Csm-40 and Csm-72 complexes affect either the cleavage efficiency or pattern was examined. We designed S3/3, S3/4, and S3/5 RNA substrates that contain flanking sequences complementary to the 5'-handle (40- or 72-nt crRNA), 3'-handle (72-nt crRNA) or both 5'- and 3'-handles in 72 nt-crRNA, respectively (FIG. 13A and Table 5). The cleavage analysis revealed that base-pairing between the 8-nt 5'-handle of crRNA and the 3'-flanking sequence had no effect on the cleavage pattern of the Csm-40 and Csm-72 complexes. The S3/3 substrate is cleaved with the same 6-nt step by Csm-40, suggesting that the non-complementarity of the flanking sequences is not a necessary pre-requisite for cleavage by the Csm complex (FIG. 13). For the Csm-72 complex, extension of the base-pairing between the 3'-handle of the 72-nt crRNA and the protospacer 5'-flanking sequence in S3/5 RNA substrate results in target RNA cleavage outside the protospacer, yielding 12- and 6-nt cleavage products (FIG. 3). Moreover, the S3/6 substrate, which has extended complementarity between crRNA 3'-handle and 5'-flanking sequence was cleaved at multiple positions along the full length of RNA duplex, except for the region complementary to the crRNA 5'-handle (FIG. 3). The cleavage at 18 and 12 nt outside the protospacer was also detected for the Csm-40 complex on S3/4 and S3/5 RNA substrates (FIG. 13). The 40-nt crRNA present in the Csm-40 complex lacks the 3'-handle and therefore cannot form RNA duplex with the 5'-flanking sequence in the S3/5 and S3/6 RNA substrates. However, the Csm-40 complex preparation still contains ~10% of unmatured 72-nt crRNA, and this heterogeneity results in the extra cleavage outside the protospacer (FIG. 13C).

To interrogate the importance of base-pairing within the protospacer region for target RNA cleavage, a set of RNA substrates was designed harboring two adjacent nucleotide mutations in the spacer region (substrates S3/7, S3/8 and S3/9, see FIG. 13A and Table 5). Two nucleotide mismatches in these substrates did not compromise RNA cleavage by the Csm-40 (FIG. 6) and Csm-72 complexes (data not shown), suggesting that the StCsm complex tolerates at least two contiguous mismatches in the protospacer region homologous to the crRNA.

To explore whether 3'- or 5'-ends of the target RNA are important for cleavage by the Csm-40 complex, a set of truncated RNA substrates was designed. In S3/10, S3/14 and S3/12 RNA substrates unpaired flaps at the 3'-, 5'- or both ends of the target RNA were truncated, while in S3/11 and S3/13 substrates the truncations extend into the region complementary to crRNA (FIG. 4A). Binding affinity for most of the truncated substrates was not compromised (FIG. 4B) and target RNA cleavage occurred at multiple sites spaced by 6-nt intervals at conserved protospacer positions (FIG. 4C). Truncations extending into the protospacer region (S3/11 and S3/13) showed decreased binding and reduced cleavage rates. This could be a result of the decreased duplex stability; however, the role of the "seed" sequence cannot be excluded. For all RNA substrates the cleavage sites were located at a fixed distance with respect to the conserved 5'-handle of crRNA (FIG. 4C).

Identification of the Ribonuclease Subunit in the StCsm Complex

Regularly spaced cleavage pattern of the RNA target (FIGS. 2-4, 11-13) implies the presence of multiple cleavage modules in the Csm complex. According to the densitometric analysis, 3 Csm2 and 5 Csm3 subunits are identified in the Csm-40 complex, while 6 Csm2 and 10 Csm3 subunits are present in the Csm-72 complex. Multiple copies of the Csm2 and Csm3 proteins in the Csm complexes make them prime candidates for catalytic subunits. StCsm2 is a small (121 aa) α -helical protein of unknown structure. StCsm3 (220 aa) contain a conserved RRM core and is fairly closely related (~35% sequence identity) to *Methanopyrus kandleri* Csm3, whose crystal structure has been solved recently (Hrle et al., 2013). It was reasoned that, since the catalytic activity of the StCsm complex requires the presence of Me^{2+} ions, the active site is likely to contain one or more acidic residues. Multiple sequence alignments of both Csm2 and Csm3 protein families were inspected for conserved aspartic or glutamic residues. No promising candidates in StCsm2 were found but several, including D33, D100, E119, E123, and E139 were identified in StCsm3 (FIG. 5A). To probe the role of these conserved negatively charged Csm3 residues, single residue alanine replacement mutants were constructed. The H19A mutant was also constructed, since it was shown that the corresponding mutation (R21A) in *M. kandleri* Csm3 abolished binding of single-stranded RNA (Hrle et al., 2013). Each mutant was expressed in the context of other StCsm/Cas proteins and analyzed the cleavage activity of the StCsm-40 complex containing mutant Csm3 subunits. StCsm3 H19A, D100A, E119A, E123A, and E139A mutants did not compromise the formation, RNA binding or cleavage activity of Csm-40 complex (FIG. 5B-5F). However, the D33A mutant impaired Csm-40 RNA cleavage (FIG. 5E-5F) without affecting RNA binding (FIG. 5D) or complex assembly. Taken together, these data demonstrate that Csm3 is an RNase, producing multiple cleavage patterns spaced by regular 6-nt intervals, and that the D33 residue is part of the catalytic/metal-chelating site. StCsm3 structural model based on the homologous structure of *M. kandleri* Csm3 is in good agreement with the identified role for this residue (FIG. 14A). D33 belongs to the highly conserved surface patch that extends from the RRM core into the "lid" subdomain (FIG. 14B). Part of this surface patch is positively charged, supporting the idea that it represents an RNA-binding site (FIG. 14C).

In Vivo RNA Targeting by the StCsm Complex

To test whether the StCsm complex can target RNA in vivo, the MS2 phage restriction assay was employed. MS2 is a lytic single-stranded RNA coliphage which infects *E. coli* via the fertility (F) pilus. The MS2 phage is a preferable model to investigate RNA targeting by the CRISPR-Cas system in vivo as no DNA intermediate is formed during the life cycle of this phage (Olsthoorn and van Duin, 2011). For in vivo RNA-targeting experiment, the *E. coli* NovaBlue (DE3, F⁺) strain was transformed with two compatible plasmids: i) pCRISPR_MS2 plasmid bearing the synthetic CRISPR array of five repeats interspaced by four 36-nt spacers targeting correspondingly the mat, lys, cp, and rep MS2 RNA sequences, and ii) pCsm/Cas plasmid for the expression of Cas/Csm proteins (FIG. 6A). The phage-targeting and control *E. coli* strains were plated and infected with series of dilutions of MS2 using the drop plaque assay. The assay revealed that the *E. coli* strain expressing wt Csm and crRNAs that target MS2 induces a 3 to 4 Log reduction of the plaquing efficiency with respect to the control cells (FIG. 6). No resistance to the MS2 phage infection was

observed in the strain expressing either the non-targeting crRNA or the cleavage-deficient (D33A) Csm3 mutant. Taken together these data demonstrate that the StCsm complex conveys *in vivo* resistance to RNA phage in the heterologous *E. coli* host.

We established the NA specificity and mechanism for the Type III-A CRISPR-Cas system of *Streptococcus thermophilus*. In sharp contrast to other CRISPR-Cas subtypes, the functional activity of Type III-A system so far has not been reconstituted *in vitro*. Cas/Csm proteins in the Type III-A CRISPR locus of the *S. thermophilus* DGCC8004 are homologous to those of *S. thermophilus* DGCC7710 and LMD-9. They also show more distant but significant similarities to Cas/Csm proteins of *L. lactis*, *E. italicus* and *S. epidermidis* (Marraffini and Sontheimer, 2008; Millen et al., 2012) (FIG. 8).

Csm Complexes of *S. thermophilus*

The Type III-A CRISPR-Cas locus of the DGCC8004 was expressed in *E. coli* and two RNP complexes, termed Csm-40 and Csm-72, were isolated. Both complexes share a conserved set of Cas10, Csm2, Csm3 and Csm4 proteins. In addition to this core, the Csm-40 also contains the Csm5 protein. Two distinct crRNAs of 72- and 40-nt co-purify with Csm-40 and Csm-72 complexes isolated from the heterologous *E. coli* host. The 72-nt crRNA comprised of an 8-nt 5'-handle, a 36-nt spacer and a 28-nt 3'-handle would result from the pre-crRNA cleavage between 28 and 29 nt within the conserved repeat region presumably by the Cas6 nuclease, similar to the III-B CRISPR-Cas system (Carte et al., 2008). The shorter 40-nt crRNA co-purified with the Csm-40 complex of *S. thermophilus* contains the conserved 8-nt 5'-handle and 32-nt spacer indicating that the 72-nt crRNA intermediate undergoes further 3'-end processing to produce a mature 40-nt crRNA that lacks the 3'-handle and 4 nt within the spacer region (FIG. 7). The RNase involved in the maturation of 72 nt crRNA intermediate remains to be identified, however the Csm5 protein which is absent in Csm-72 but is present in Csm-40 could be a possible candidate. Indeed, *csm5* gene deletion in DGCC8004 produces only unmaturing Csm-72 complexes (data not shown).

The crRNA processing and maturation pathway in the *S. thermophilus* Type III-A system (FIG. 7) shows striking similarity to that in *S. epidermidis*. First, the SeCsm complex includes the same set of Cas10, Csm2, Csm3, Csm4 and Csm5 proteins as the StCsm-40. Furthermore, in *S. epidermidis*, the primary processing by Cas6 produces a 71-nt crRNA intermediate, that is subjected to further endonucleolytic processing at the 3' end (Hatoum-Aslan et al., 2011; Hatoum-Aslan et al., 2014).

StCsm Complex Cuts RNA Producing a Regular Cleavage Pattern

The Csm complexes of *S. epidermidis* and *S. solfataricus* have been reconstituted and isolated, however the NA cleavage activity has not been reported so far. *In vivo* studies in *S. epidermidis* suggested that the Type III-A SeCsm RNP complex targets DNA (Marraffini and Sontheimer, 2008) in a PAM-independent manner and prevents autoimmunity by checking the complementarity between the crRNA 5'-handle and the 3'-flanking sequence in the vicinity of the protospacer (Marraffini and Sontheimer, 2010). In contrast to these data we found that the StCsm-40 and StCsm-72 complexes bind ssRNA with high affinity and cut a ssRNA target in a PAM-independent manner in the presence of Me^{2+} ions, producing a regular 6-nt cleavage pattern in the protospacer region (FIGS. 2D and 11C-11E). In this respect the Type III-A StCsm complex resembles the RNA-targeting Type III-B Cmr-complexes PfCmr, SsCmr and TtCmr (Hale

et al., 2009; Staals et al., 2013; Zhang et al., 2012) (FIG. 7) rather than DNA targeting Type I and II complexes. By targeting RNA rather than DNA, the StCsm complex avoids autoimmunity. It was demonstrated that the nucleotide context and non-complementarity outside the protospacer have no effect on the target RNA cleavage, demonstrating that PAM or unpaired flanking sequences of the protospacer are not required for cleavage by the StCsm (FIG. 13). The complementarity of the protospacer is the only pre-requisite for the StCsm cleavage: non-matching RNA is not cleaved; however, either two contiguous mismatches or end truncations in the complementary protospacer S3 are tolerated (FIG. 13). The differences in the cleavage patterns of the 5'- and 3'-labeled RNAs (FIG. 2D) imply that cleavage first occurs at 3'-end of the target RNA. It remains to be established whether the observed cleavage pattern is dictated by the "seed" sequence (eg. directionality of base pairing process between the crRNA and target RNA) or by nucleotide context-dependent differences of cleavage rate.

It was found that for the Csm-72 complex the target RNA is being cleaved at regular 6-nt intervals outside the protospacer if it retains base complementarity to the crRNA 3'-handle. Such regularly spaced cleavage pattern of the RNA target (FIGS. 2-4, 11 and 13) implies the presence of multiple cleavage modules in the Csm complex. The major difference between the Csm-40 and Csm-72 complexes is the number of Csm2 and Csm3 subunits. The Csm-40 contains 3 Csm2 and 5 Csm3 subunits while Csm-72 contains 6 Csm2 and 10 Csm3 subunits (FIG. 1D). The size of the complexes determined by SAXS correlates with the different stoichiometry of Csm-40 and Csm-72. Both complexes show a slightly twisted elongated shape but the Csm-72 is significantly more elongated than Csm-40 complex (FIG. 1H). Taken together these data suggest that the longer unmaturing 72-nt crRNA intermediate in the Csm-72 complex binds additional copies of Csm2 and Csm3 subunits into a RNP filament (FIG. 7).

Csm3 is a RNase Subunit in the StCsm Complex

Computational analysis revealed that StCsm3 has a conserved RRM core and is fairly closely related (~35% sequence identity) to *M. kandleri* Csm3 (Hrle et al., 2013). StCsm3 displays close structural similarity to MkCsm3, in particular the RRM-core and insertions into RRM-core that form the "lid" subdomain (FIG. 14A). In contrast, StCsm3 lacks both the N-terminal zinc binding domain and the C-terminal helical domain, making its structure more compact compared to that of MkCsm3. Thus, StCsm3 may be considered as a trimmed-down version of MkCsm3. Guided by the multiple sequence alignment and homology model of StCsm3, candidate active site/metal chelating residues of Csm3 were selected and subjected to alanine mutagenesis. The highly conserved D33 residue of the StCsm3 was critical for the RNA cleavage activity of the Csm complex, demonstrating that Csm3 is an RNase in the StCsm and other Type III-A CRISPR-Cas systems (FIG. 5).

Implications for Other RNA-Targeting CRISPR Systems

The StCsm complex was specific for RNA and cuts it in a PAM-independent manner producing a regular 6-nt cleavage pattern. The Csm3 protein, which is present in Csm-40 and Csm-72 complexes in multiple copies, was demonstrated to act as an RNase responsible for the target RNA cleavage. In this respect the Type III-A Csm complex of *S. thermophilus* closely resembles the RNA targeting Type III-B Cmr complex of *T. thermophilus* (TtCmr complex) that also produces a regular 6-nt cleavage pattern (Staals et al., 2013). The RNA degrading subunit in the Type III-B Cmr-module remains to be identified. Although there is currently

no experimental evidence, Staals et al. suggested that Cmr4 could fulfill this role (Staals et al., 2013). Clustering of Csm3 and Cmr4 homologs by sequence similarity revealed that they form two related but separate groups (FIG. 14D). On the other hand, neither Csm3 nor Cmr4 families are homogenous. They are comprised of sequence clusters of various sizes. StCsm3 is a member of a large representative group of Csm3 homologs that includes those from *S. epidermidis*, *L. lactis* and *M. kandleri*. Another large, but more loosely connected group does not have proteins from experimentally characterized systems, except for the Csm complex from *S. solfataricus*. Sso1425 and Sso1426, two of its Csm3-like proteins (Makarova et al., 2011a), are members of this group albeit they are non-typical. The Cmr4 family appears even more heterogeneous than Csm3. Cmr4 proteins of experimentally characterized III-B systems from *T. thermophilus* and *P. furiosus* represent one of the larger clusters, while Cmr4 from *S. solfataricus* is a non-typical outlier. Biochemical characterization revealed that PfCmr and TtCmr RNA cleavage mechanism are similar and follow a 3'- or 5'- ruler mechanism, respectively (Hale et al., 2009; Staals et al., 2013). Meanwhile, SsCmr endonucleolytically cleaves both target RNA and crRNA at UA dinucleotides (Zhang et al., 2012). It thus would not be surprising if members of other, so far experimentally uncharacterized groups were part of Cmr complexes with somewhat different properties.

We questioned if Csm3 and Cmr4 proteins may have similarly organized active site. The aligned sequences of Csm3 and Cmr4 subunits from characterized systems revealed that sequences of both families have Asp in the corresponding positions, suggesting similar active sites (FIG. 5A). The exception is Sso1426. This is surprising, considering the composition of the *S. solfataricus* Csm complex. Four copies of Sso1426 were found to be present within the complex suggesting that this subunit might play a role of the Csm3 (Rouillon et al., 2013). In contrast, another Csm3-like protein Sso1425 does have the D33 counterpart suggesting it can cleave ssRNA. However, only a single copy of Sso1425 was found in the *S. solfataricus* complex. Taken together, these data suggest that Csm-modules in *S. thermophilus* and *S. solfataricus* have different architectures and RNA cleavage mechanisms.

It is demonstrated for the first time that the Csm effector complex of the *S. thermophilus* Type III-A system targets RNA and establish the mechanism of RNA cleavage. It is demonstrated that in the Type III-A effector complex Cas/Csm proteins assemble into an RNP filament (FIG. 7) that contains multiple copies of Csm2 and Csm3 proteins. The inventors provided evidence that the Csm3 subunit acts as an RNase that cleaves target RNA at multiple sites spaced by regular 6-nt intervals (FIG. 7). The number of cleavage sites correlates with the number of Csm3 subunits in the Csm effector complex. Easy programmability of the Type III-A StCsm complex by custom crRNAs (FIG. 12), paves the way for the development of novel molecular tools for RNA interference.

RNA cleavage specificity established here for the StCsm complex in vitro is supported by in vivo experiments of MS2 RNA phage interference in the heterologous *E. coli* host (FIG. 6). It remains to be established whether RNA silencing by the StCsm complex can contribute to the DNA phage interference in the *S. thermophilus* host. Transcription-dependent DNA targeting mechanism has been proposed recently for the Type III-B CRISPR-Cmr system (Deng et al., 2013); however, it has yet to be demonstrated for *S. thermophilus* and other Type III-A systems.

Deletion Analysis of StCsm Complex

Csm-complexes are composed of several Cas proteins (Cas10, Csm2, Csm3, Csm4, Csm5) and contain traces of Cas6. In the StCsm complexes Csm3 acts as the ribonuclease that cuts target RNA. Cas6 is responsible for the pre-crRNA maturation into 72 nt crRNAs. To establish a functional role of other Csm proteins, we engineered pCas/Csm plasmid variants with disrupted individual cas/csm genes and isolated a set of StCsm deletion mutant complexes. These StCsm deletion mutant complexes were then subjected to biochemical analyses to determine the role of each individual protein in the StCsm complex assembly and RNA cleavage.

First we examined the composition of proteins and crRNAs in the StCsm complex deletion mutants. SDS-PAGE analysis of protein composition of the purified StCsm-40 and StCsm-72 deletion mutants confirmed that in all cases protein corresponding to the disrupted cas gene is missing in the complex (FIG. 17A, FIG. 18A). Csm3 makes a backbone of StCsm complex since no complex is detected when csm3 gene is deleted. Cas10 seems to be associated to the Csm4 in the Csm-complex since Cas10 subunit is missing in the StCsm-40 Δ Csm4 and StCsm-72 Δ Csm4 samples.

crRNAs co-purified with deletion mutant complexes are distinct (FIG. 17B, FIG. 18B). Nucleic acids purified from StCsm-40 Δ Cas6 and StCsm-72 Δ Cas6 complexes predominantly contain long pre-crRNAs, that support Cas6 function in crRNA maturation. In the case of StCsm Δ Csm4 variant 72 nt crRNA co-purifies together with long pre-crRNA molecules implying that crRNA binding specificity is compromised. Wt StCsm-40 and StCsm-40 Δ Csm6' Δ Csm6 predominantly contains 40 nt crRNAs, while in all other cases 72 nt prevails in the Csm-complex. Taken together, available data suggest that Cas10, Csm5 and possibly Csm2 and Csm4 proteins are important for crRNA maturation from 72 to 40 nt species.

Next we explored the impact of single protein deletions on the StCsm complex capability to bind and cleave RNA. StCsm complex deletion mutants were probed on target substrates complementary to the crRNA encoded by the spacer S3 and non-targeting RNA substrates (FIG. 17C, FIG. 18C). Csm5 deletion dramatically impacts specific RNA binding: StCsm Δ Csm5 complexes bind target and non-target RNAs with nearly the same affinity while wt StCsm complexes show ~100-fold tighter binding to the target RNA. To compare the RNA cleavage capabilities of the StCsm complex variants, we performed cleavage assays and determined cleavage rate constants (FIG. 17D, FIG. 18D). Surprisingly, most of the StCsm complex deletion mutants retained at least partial RNA cleavage activity. Only for StCsm Δ Cas6 and StCsm Δ Csm4 complexes RNA cleavage activity was nearly fully abolished. In all other cases, the reaction products (and hence cleavage positions) were identical. The cleavage activity of StCsm Δ Csm6' Δ Csm6 variants is similar to that of wt. This is not surprising since Csm6' and Csm6 proteins are not present in the StCsm complex. Cas10 significantly impacted only the yield of the complex but had no effect on the cleavage rate. The RNA cleavage assay data suggests that only Csm3, Csm4, and crRNA-generating Cas6 are required for the target RNA cleavage. Csm2 is completely dispensable in respect to RNA cleavage.

Minimal StCsm Complex Assembly

Deletion analysis shows that Csm3 and Csm4 proteins are critical for Csm-complex assembly/activity. Therefore, we explored the possibility to assemble a minimal Csm-complex arranged of three components including Csm3 and Csm4 subunits, and crRNA. Such minimal engineered vari-

ant StCsm would be a convenient tool for specific RNA targeting both in vitro and in vivo. We cloned the csm3 and csm4 genes into pCDFDuet-1 vector and added a StrepII-Tag sequence to the N-terminal part of csm3 to obtain p^{Tag}Csm3 Csm4 plasmid. p^{Tag}Cas10 plasmid was constructed by cloning cas10 gene into pETDuet-1 vector. pCas6 plasmid was constructed by cloning cas6 gene into pCOLADuet-1 vector. The expression of Cas6 protein together with pCRISPR_S3 (encoding the S3 CRISPR region) would generate unmaturing 72-nt crRNA which would be incorporated in Csm ribonucleoprotein complex. Alternatively, to omit Cas6-mediated pre-crRNA maturation, plasmids perRNA-40 and perRNA-72 were constructed. Transcription of these plasmids in *E. coli* will produce 40-nt or 72-nt crRNA species in the absence of Cas6. These plasmids were engineered on basis of pACYCDuet-1 vector with under a control of the BBa J23119 promoter. p^{Tag}Csm3 Csm4 was co-expressed in *E. coli* BL21(DE3) either with pCas6 and pCRISPR plasmids or with perRNA-40/perRNA-72 plasmids. Since omitting Cas10 proved to significantly reduce yields of the complex during the deletion analysis, we also tested how the presence or absence of p^{Tag}Cas10 in the expression system would affect the minimal Csm-complexes. Affinity purification on the Strep-chelating column yielded minimal Csm complexes, containing Csm3 and Csm4 protein subunits and crRNA. When p^{Tag}Cas10 plasmid was present in the expression system, the complexes also contained Cas10 protein and the total yield of the complexes was significantly increased. When pCas6 plasmid was present in the expression system, Cas6 protein co-purified with the Csm-complex. The RNA cleavage activity of these RNP complexes was assayed on the 68-nt S3/4 or 86-nt S3/6 RNA target substrate (FIG. 19). Minimal Csm-complexes containing only Csm3 and Csm4 subunits, as well as other minimized Csm-complex variants, show RNA cleavage pattern, characteristic to the wt StCsm-72 complex. Taken together, data provided here show that minimal Csm complex assembled using only Csm3, Csm4, and crRNA cleaves target RNA. In this respect, it provides a versatile tool for RNA knock-outs in the cell. Cleavage deficient variant of the minimal complex could be used for RNA knock-downs or pull-down of the desired target RNA from cells.

Experimental Procedures

Expression and Isolation of Csm Complexes

The sequence of CRISPR2-cas locus of *S. thermophilus* DGCC8004 was deposited in GenBank (accession number KM222358). Heterologous *E. coli* BL21(DE3) cells producing the Strep-tagged Csm complexes were engineered and cultivated as described. Csm-40 and Csm-72 complexes were isolated by subsequent Strep-chelating affinity and size exclusion chromatography steps.

Streptococcus thermophilus DGCC8004 was cultivated at 42° C. in M1 7 broth (Oxoid) supplemented with 0.5% (w/v) lactose. Chromosomal DNA was extracted and purified using GeneJET Genomic DNA Purification Kit (Thermo Scientific). CRISPR2-Cas region was amplified by polymerase chain reaction (PCR) and sequenced using primers designed by genomic comparison with *S. thermophilus* DGCC7710 (GenBank accession number AWVZ01000003). Annotation of the predicted ORFs was performed using BLASTP at NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). CRISPR region was identified through repeat sequence similarity to that of *S. thermophilus* DGCC7710. Multiple sequence alignments of cas/csm genes, spacers and repeats sequences were carried out with ClustalW2 (<http://www.ebi.ac.uk>). Genomic DNA isolated

from *S. thermophilus* DGCC8004 strain was used as the template for PCR amplification of the cas/csm genes. DNA fragment covering the 8.5 kb cas6-cas10-csm2-csm3-csm4-csm5-csm6-csm6' gene cassette was cloned into pCDFDuet-1 expression vector via NcoI and AvrII restriction sites in two separate subcloning steps to generate plasmid pCas/Csm. Individual cas/csm genes were cloned into pETDuet-1_N-StrepII and pETDuet-1_C-StrepII expression vectors, except of cas10 (which was cloned into pBAD24 C-His-StrepII-His) and csm6 or csm6' (that were cloned into pBAD24 N-His-StrepII-His) to generate pCsmX-Tag and pCasY-Tag plasmids, where X=2,3,4,5,6,6' and Y=6,10. A synthetic 445-nt CRISPR locus containing five 36-nt length repeats interspaced by four identical 36-nt spacers S3 of the *S. thermophilus* DGCC8004 CRISPR2 system was obtained from Invitrogen and cloned into the pACYC-Duet-1 vector to generate a plasmid pCRISPR_S3. Four copies of the spacer S3 have been engineered into the pCRISPR_S3 plasmid to increase the yield of the Csm-crRNA complex. Full sequencing of cloned DNA fragments confirmed their identity to the original sequences.

All three plasmids were co-expressed in *Escherichia coli* BL21 (DE3) grown at 37° C. in LB medium supplemented with streptomycin (25 µg/µl), ampicillin (50 µg/µl), and chloramphenicol (30 µg/µl). The fresh LB medium was inoculated with an overnight culture (1/20 (v/v)), and bacteria were grown to the mid-log phase (OD_{600nm} 0.5 to 0.7), then 1 mM IPTG (and 0.2% (w/v) L-(+)-arabinose in case of Cast 0, Csm6 and Csm6') was added and cell suspension was further cultured for another 4 h. Harvested cells were resuspended in a Chromatography buffer (20 mM Tris-HCl (pH 8.5), 0.5 M NaCl, 7 mM 2-mercaptoethanol, 1 mM EDTA) supplemented with 0.1 mM phenylmethylsulfonyl fluoride (PMSF), and disrupted by sonication. Cell debris was removed by centrifugation. Csm complexes were captured on the StrepTrap affinity column (GE Healthcare) and further subjected to the Superdex 200 size exclusion chromatography (prep grade XK 16/60; GE Healthcare). SDS-PAGE of individual Strep-tagged Csm2, Csm3, Csm4, CsmS, Cas6 and Cas10 proteins isolated by affinity chromatography from *E. coli* lysates revealed co-purification of other Csm/Cas proteins suggesting the presence of a Csm complex. The abundance of the Csm complex co-purified via the Csm4-, CsmS-, Cas6- and Cas10-Strep tagged subunits was very low, and no complex was pull-downed via Csm6 or Csm6' subunits (data not shown). Therefore, Csm complexes isolated via N-terminus Strep-tagged Csm2 (Csm2 StrepN) and the N-terminus Strep-tagged Csm3 proteins (Csm3 StrepN) were subjected to further characterization. Individual Csm3-N-Strep protein was purified using StrepTrap affinity column. Csm3-N-Strep and Csm complexes eluted from the columns were dialysed against 10 mM Tris-HCl (pH 8.5) buffer containing 300 mM NaCl, 1 mM DTT, 0.1 mM EDTA, and 50% (v/v) glycerol, and stored at -20° C.

The composition of the isolated Csm-40 and Csm-72 complexes was analysed by SDS-PAGE and the sequence of Csm proteins was further confirmed by the mass spectrometry of tryptic digests. In order to estimate the stoichiometry of Csm complexes, protein bands in SDS-PAGE were quantified by densitometric analysis taking a count the different staining of Cas/Csm proteins. The molecular weights of the Csm complexes were estimated by dynamic light scattering (DLS) using Zetasizer pV (Malvern) and respective software. For DLS analysis Csm-40 and Csm-72 samples were analysed in a Chromatography buffer at 0.36 mg/ml and 0.6 mg/ml concentrations, respectively. Csm complex concen-

trations were estimated by Pierce 660 nm Protein Assay (Thermo Scientific) using bovine serum albumin (BSA) as a reference protein. Conversion to molar concentration was performed assuming that the Csm-72 stoichiometry is Cas10₁:Csm2₆:Csm3₁₀:Csm4₁:crRNA72₁ and the Csm-40 stoichiometry is Cas10₁:Csm2₃:Csm3₅:Csm4₁:Csm5₁:crRNA40₁.

Bioinformatic Analysis and Mutagenesis of Csm3

Putative active site residues of Csm3 were identified from multiple alignment of Csm3/Cmr4. Csm3 mutants were constructed using quick change mutagenesis and purified as described.

Mutagenesis of Csm3

The Csm3 mutants H19A, D33A, D100A, E119A, E123A and E139A were obtained by the Quick Change Mutagenesis (QCM) Protocol (Zheng et al., 2004). First, a 3.0 kb DNA fragment containing csm2 and csm3 genes was subcloned from pCas/Csm plasmid into the pUC18 vector pre-cleaved with SphI and KpnI. The resulting plasmid pUC18_Csm2_Csm3 was used for Csm3 QCM mutagenesis. After QCM, the same fragment containing mutated versions of the Csm3 gene was transferred back into the pCas/Csm vector using NdeI and SpeI sites, reconstituting the gene cassette. Sequencing of the entire cloned DNA fragment for each mutant confirmed that only the designed mutation had been introduced. Csm-40 complexes containing Csm3 mutants were isolated following the procedures described for the wt StCsm-40 (see above). D100A mutant StCsm-40 was purified only using the affinity chromatography.

Extraction, HPLC Purification and ESI-MS Analysis of crRNA

NAs co-purified with Csm-40 and Csm-72 were isolated using phenol:chloroform:isoamylalcohol (25:24:1, v/v/v) extraction and precipitated with isopropanol. Purified NAs were incubated with 0.8 U DNase I or 8 U RNase I (Thermo Scientific) for 30 min at 37° C. NAs were separated on a denaturing 15% polyacrylamide gel (PAAG) and visualized by SybrGold (Invitrogen) staining.

Ion-pair reversed-phased-HPLC purified crRNA architecture was determined using denaturing RNA chromatography in conjunction with electrospray ionization mass spectrometry (ESI-MS) as described in (Sinkunas et al., 2013).

All samples were analyzed by ion-pair reversed-phased-HPLC (Dickman and Homby, 2006; Waghmare et al., 2009) on an Agilent 1100 HPLC with UV260 nm detector (Agilent) using a DNasep column 50 mm×4.6 mm I. D. (Transgenomic). The chromatographic analysis was performed using the following buffer conditions: A) 0.1 M triethylammonium acetate (TEAA) (pH 7.0) (Fluka); B) buffer A with 25% LC MS grade acetonitrile (v/v) (Fisher). The crRNA was obtained by injecting purified intact Csm-40 or Csm-72 at 75° C. using a linear gradient starting at 15% buffer B and extending to 60% B in 12.5 mM, followed by a linear extension to 100% B over 2 mM at a flow rate of 1.0 ml/min. Analysis of the 3' terminus was performed by incubating the HPLC-purified crRNA in a final concentration of 0.1 M HCl at 4° C. for 1 hour. The samples were concentrated to 10-20 µl on a vacuum concentrator (Eppendorf) prior to ESI-MS analysis.

ESI-MS Analysis of crRNA

Electrospray Ionization Mass spectrometry (ESI-MS) was performed in negative mode using an Amazon Ion Trap mass spectrometer (Bruker Daltonics), coupled to an online capillary liquid chromatography system (Ultimate 3000, Dionex, UK). RNA separations were performed using a monolithic (PS-DVB) capillary column (50 mm×0.2 mm

I.D., Dionex, UK). The chromatography was performed using the following buffer conditions: C) 0.4 M 1,1,1,3,3,3,-Hexafluoro-2-propanol (HFIP, Sigma-Aldrich) adjusted with triethylamine (TEA) to pH 7.0 and 0.1 mM TEAA, and D) buffer C with 50% methanol (v/v) (Fisher). RNA analysis was performed at 50° C. with 20% buffer D, extending to 40% D in 5 min followed by a linear extension to 60% D over 8 min at a flow rate of 2 µl/min, 250 ng crRNA was digested with 1 U RNase A/T1 (Applied Biosystems). The reaction was incubated at 37° C. for 4 h. The oligoribonucleotide mixture was separated on a PepMap C-18 RP capillary column (150 mm×0.3 mm I.D., Dionex, UK) at 50° C. using gradient conditions starting at 20% buffer C and extending to 35% D in 3 mins, followed by a linear extension to 60% D over 40 mins at a flow rate of 2 µl/min. The mass spectrometer was operated in negative mode, a capillary voltage was set at -2500 V to maintain capillary current between 30-50 nA, temperature of nitrogen 120° C. at a flow rate of 4.0 L/h and N2 nebuliser gas pressure at 0.4 bar. A mass range of 500-2500 m/z was set. Oligoribonucleotides with -2 to -4 charge states were selected for tandem mass spectrometry using collision induced dissociation.

Small Angle X-Ray Scattering (SAXS) Experiments

SAXS data for Csm-40 and Csm-72 were collected at P12 EMBL beam-line at PETRAIII storage ring of DESY synchrotron in Hamburg (Germany). Csm-40 and Csm-72 complexes were measured in 3 different concentrations in buffer containing 20 mM Tris-HCl (pH 8.5 at 25° C.), 0.5 M NaCl, 1 mM EDTA and 7 mM 2-mercaptoethanol. Data collection, processing and ab initio shape modeling details are presented in Table 4 and FIG. 10.

Ab initio shape modeling of both complexes was performed with the samples having highest concentration (1.3 mg/ml for Csm-40 and 2.0 mg/ml for Csm-72). Unprocessed scattering data with subtracted buffer scattering, Guinier plots of the low s region of the scattering curves used for the shape determination and P(r) functions of the highest concentration samples of Csm-40 and Csm-72 are presented in FIG. 10. Two-dimensional scattering curves were transformed and distance distribution functions P(r) were calculated using GNOM (Svergun, 1992). At this stage data were truncated to s values 0.15-0.1 Å⁻¹ and calculated distance distribution function was used for following ab initio modeling. 10 independent bead models for both complexes were generated using DAMMIN (Svergun, 1999). These models were aligned, filtered and averaged based on occupancy using DAMAVER (Volkov and Svergun, 2003). The averaged NSD of superposition of DAMMIN models of Csm-40 complex was 0.563±0.028 (for Csm-72 models averaged NSD is 0.575±0.019), no model was rejected in both cases.

The inertia tensor was calculated for averaged models of both complexes and models were aligned along the largest principal axis so as the end points of both models coincided. After that the protruding part of the longer Csm-72 complex was truncated. Csm-40 model was aligned with truncated Csm-72 models by automatic procedure SUPCOMB (Kozin and Svergun, 2001) producing an NSD value. Then Csm-40 model was shifted along the principal axis of Csm-72 model by the fixed step (5 or 10 Å) and again Csm-40 model was aligned by SUPCOMB with the Csm-72 model after truncation of protruding parts. Thus the Csm-40 model was sequentially shifted along the principal axis of Csm-72 model and the best superposition showed the lower NSD value (S. Grazulis, personal communication). MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Bacon, 1997) programs were used for SAXS models presented in FIGS. 1 and 10.

DNA and RNA Substrates

Synthetic oligodeoxynucleotides were purchased from Metabion. All RNA substrates were obtained by in vitro transcription using TranscriptAid T7 High Yield Transcription Kit (Thermo Scientific). A full description of all the DNA and RNA substrates is provided in the Table 5. DNA and RNA substrates were either 5'-labeled with [^{32}P] ATP and PNK or 3'-labeled with [^{32}P] cordycepin-5'-triphosphate (PerkinElmer) and poly(A) polymerase (Life Technologies) followed by denaturing gel purification.

To assemble DNA oligoduplexes, complementary oligodeoxynucleotides were mixed at 1:1 molar ratio in the Reaction buffer (33 mM Tris-acetate (pH 7.9 at 25° C.), 66 mM potassium acetate), heated to 90° C. and slowly let to cool to room temperature.

For generation of S3/1-10, S3/14 RNA substrates, first pUC18 plasmids pUC18_S3/1 and pUC18_S3/2, bearing S3/1 or S3/2 sequences were constructed. For this purpose, annealed synthetic DNA oligoduplexes S3/1 or S3/2 were ligated into pUC18 plasmid pre-cleaved with SmaI. Engineered plasmids pUC18_S3/1 and pUC18_S3/2 were sequenced to persuade that only copy of DNA duplex was ligated into the vector. Further these plasmids were used as a template to produce different DNA fragments by PCR using appropriate primers containing a T7 promoter in front of the desired RNA sequence. Purified PCR products were used in the in vitro transcription reaction to obtain RNA substrates. S3/11-13 RNAs were prepared by hybridizing two complementary DNA oligonucleotides, containing a T7 promoter in front of the desired RNA sequence followed by in vitro transcription.

DNA/RNA hybrids were assembled in similar manner annealing complementary oligodeoxynucleotide to RNA obtained by in vitro transcription.

pBR322 plasmid bearing the Tc gene, encoding tetracycline (Tc) resistance protein, was used to produce Tc RNA and ncTc RNA substrates using the same in vitro transcription reaction as described above for S3/1-10, S3/14. Prior to ^{32}P 5'-labeling RNA substrates were dephosphorylated using FastAP thermosensitive alkaline phosphatase (Thermo Scientific).

Electrophoretic Mobility Shift Assay

Binding assays were performed by incubating different amounts of Csm complexes with 0.5 nM of ^{32}P -5'-labeled NA in the Binding buffer (40 mM Tris, 20 mM acetic acid (pH 8.4 at 25° C.), 1 mM EDTA, 0.1 mg/ml BSA, 10% (v/v) glycerol). All reactions were incubated for 15 min at room temperature prior to electrophoresis on native 8% (w/v) PAAG. Electrophoresis was carried out at room temperature for 3 h at 6 V/cm using 40 mM Tris, 20 mM acetic acid (pH 8.4 at 25° C.), 0.1 mM EDTA as the running buffer. Gels were dried and visualized using a FLA-5100 phosphorimager (Fujifilm). The K_d for NA binding by Csm-72 and Csm-40 was evaluated assuming the complex concentration at which half of the substrate is bound as a rough estimate of K_d value. For binding competition assay 0.5 nM ^{32}P -labelled S3/1 RNA was mixed with 0.5-5000 nM of unlabelled competitor NA and 0.3 nM StCsm-40, and analyzed by EMSA.

Cleavage Assay

The Csm-40 reactions were performed at 25° C. and contained 20 nM of 5'- or 3'-radiolabeled NA (Table 5) and 62.5 nM (unless stated otherwise) complex in the Reaction buffer (33 mM Tris-acetate (pH 7.9 at 25° C.), 66 mM K-acetate, 0.1 mg/ml BSA and 10 mM Mg-acetate). Csm-72 reactions were performed in the same Reaction buffer at 37° C. and contained 20 nM of radiolabeled NA and 125 nM of

complex unless stated otherwise. Cleavage reactions using minimal StCsm were performed in the same Reaction buffer at 37° C. and contained 4 nM of radiolabeled RNA and ~15 ng/ μl of the RNP complex. Reactions were initiated by addition of the Csm complex. The samples were collected at timed intervals and quenched by mixing 10 μl of reaction mixture with 2 \times RNA loading buffer (Thermo Scientific) followed by incubation for 10 min at 85° C. The reaction products were separated on a denaturing 20% PAAG and visualized by autoradiography. ^{32}P -5'-labeled RNA Decade marker (Ambion) was used as size marker. To map the cleavage products oligoribonucleotide markers were generated by RNase A (Thermo Scientific, final concentration 10 ng/ml) treatment of RNA substrates for 8 min at 22° C. or by alkaline hydrolysis in 50 mM NaHCO_3 (pH 9.5) at 95° C. for 5 min.

Fluorescent Microscopy

Transformed *E. coli* cells producing GFP and StCsm were diluted 1:40 from an overnight culture in fresh LB medium and cells were further grown at 37° C. for 2 h in the presence of 1% IPTG to induce Cas/Csm, GFP and crRNA expression. The GFP transcript degradation was monitored by inspecting GFP fluorescence in *E. coli* cells. For this purpose, an aliquot of bacteria (2 μl) was immediately mounted on a thin film of 1.2% agarose (ThermoFisher Scientific) on microscope slides and then overlaid with a coverslip (Roth). The cells were immediately imaged by contrast and fluorescence microscopy. Acquisition of contrast and fluorescence images was performed using a Nikon Elipse Ti-U microscope coupled to a Nikon DS-Qi1 camera. The digital images were analyzed with NIS Element v.4.00.00 (Nikon) software. No electronic enhancement or manipulation was applied to the images.

Phage Drop Plaque Assay

Phage drop plaque assay was conducted using LGC Standards recommendations. Phage drop plaque assay was conducted using LGC Standards recommendations. Briefly, *E. coli* NovaBlue(DE3) [(endA 1 hsdR17(r_{k12-} m_{K12+}) supE44 thi-1 recA1 gyrA96 relA1 lac (DE3) FlproA⁺B⁺ lad q ZAM15::Tn10] (Tet^R) was transformed with wt pCas/Csm (Str^R) or D33A Csm3 pCas/Csm (Str^R) and pCRISP-R_MS2 (Cm^R), pCRISPR_S3 (Cm^R), or pACYC-Duet-1 (Cm^R). *E. coli* cells bearing different sets of plasmids were grown in LB medium with appropriate antibiotics at 37° C. to an OD 600 of 0.9 and a 0.4 ml aliquot of bacterial culture was mixed with melted 0.5% soft nutrient agar (45° C.). This mixture was poured onto 1.5% solid agar to make double layer agar plates. Both layers of agar contained appropriate antibiotics, 0.1 mM IPTG, 0.1% glucose, 2 mM CaCl_2 and 0.01 mg/ml thiamine. When the top agar hardened, phage stock (5 μl) from a dilution series was delivered on each plate with the bacteria. The plates were examined for cell lysis after overnight incubations at 37° C. NovaBlue(DE3) was used as the indicator for determining the phage titer. pCRISP-R_MS2 plasmid bearing the synthetic CRISPR array of five repeats interspaced by four 36-nt spacers targeting the mat, lys, cp, and rep MS2 RNA sequences (GenBank accession number NC001417) was constructed similarly to pCRISPR_S3 (see above).

Computational Sequence and Structure Analysis

Sequence searches were performed with PSI-BLAST (Altschul et al., 1997) against the nr80 sequence database (the NCBI 'nr' database filtered to 80% identity) using E-value=1 e-03 or a more stringent inclusion threshold. Clustering of homologous sequences according to their mutual similarity was done using CLANS (Frickey and Lupas, 2004). Multiple sequence alignments were con-

structured with MAFFT (Katoh et al., 2002) using the accuracy-oriented mode (L-INS-i). Homology model for StCsm3 was constructed with HHpred (Söding et al., 2005) using the related structure of *M. kandleri* Csm3 (PDB code 4NOL) as a template. The analysis of surface residue conservation was performed using the ConSurf server (Ashkenazy et al., 2010). Electrostatic map of the structure surface was calculated with the APBS (Baker et al., 2001) plugin in PyMol (Schrodinger, 2010). Pictures were prepared with PyMol (Schrodinger, 2010).

Engineering of Single-Gene Deletion Mutants

pCas/Csm plasmid was used as a template to generate the following single-gene deletion mutant variants: pCas/Csm Δ Cas6, pCas/Csm Δ Cas10, pCas/Csm Δ Csm4, and pCas/Csm Δ Csm6' Δ Csm6. To obtain the pCas/Csm Δ Cas6 variant, pCas/Csm plasmid was cleaved with Bsp1407I, the remaining sticky ends were blunted, phosphorylated (using "Fast DNA End Repair Kit" from Thermo Scientific), and ligated. This resulted into the cas6 gene truncation to 67 codons. To obtain pCas/Csm Δ Cas10, a Bsp119I fragment was excised from the pCas/Csm plasmid. The re-ligated plasmid resulted in the cas10 gene truncation to 185 codons. To obtain pCas/Csm Δ Csm4, pCas/Csm was cleaved with SpeI and Eco31I, blunt-ended and re-ligated. This resulted in Csm4 ORF truncation to 41 codons. To obtain pCas/Csm Δ Csm6' Δ Csm6, pCas/Csm was cleaved with PpiI and XmaJI, and resulting larger DNA fragment blunt-ended and subjected to ligation. This resulted in the Csm6' ORF truncation to 324 codons and elimination of Csm6 ORF. To obtain pCas/Csm Δ Csm5, pUC18_Csm5_Csm6'_Csm6 plasmid was constructed by subcloning a 2.7 kb DNA fragment containing csm5, csm6', and csm6 genes from pCas/Csm plasmid into pUC18 vector, pre-cleaved with SphI and KpnI. pUC18_Csm5_Csm6'_Csm6 was cleaved with SwaI and BsaAI, the resulting larger DNA fragment was ligated to yield pUC18_ Δ Csm5_Csm6'_Csm6 plasmid, containing a frameshift mutation at the start of csm5 gene. The SphI and Pad fragment containing Δ csm5, csm6', and csm6 was subcloned into the pCas/Csm plasmid to yield pCas/Csm Δ Csm5.

pCas/Csm Δ Csm2 and pCas/Csm Δ Csm3 were engineered using pUC18_Csm2_Csm3 plasmid (see section Mutagenesis of Csm3). To obtain the pCas/Csm Δ Csm2, pUC18_Csm2_Csm3 was cleaved with BspMI and AflIII, while to obtain pCas/Csm Δ Csm3, pUC18_Csm2_Csm3 was cleaved with ClaI and XhoI. The resulting large DNA fragments were then blunted, phosphorylated, and ligated and subcloned into pCas/Csm via NdeI and SpeI sites. This resulted in the Csm2 ORF truncation to 70 codons, and Csm3 ORF truncation to 57 codons. Full sequencing of cloned DNA fragments confirmed their identity to the expected sequences. In all cases the deletions were executed in such a way that ribosome binding sites for other genes would not be disrupted. StCsm-40 and StCsm-72 complexes lacking single deleted protein were isolated following the procedures described for the wt StCsm-40 (see above).

Each of the following references are expressly incorporated by reference herein in its entirety:

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-3402.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38, W529-533.

- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98, 10037-10041.
- Dickman, M. J., and Hornby, D. P. (2006). Enrichment and analysis of RNA centered on ion pair reverse phase methodology. *RNA* 12, 691-696.
- Fischer, H., de Oliveira Neto, M., Napolitano, H., Polikarpov, I., and Craievich, A. (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J Appl Crystallogr* 43, 101-109.
- Frickey, T., and Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20, 3702-3704.
- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., and Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945-956.
- Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., and Marraffini, L. A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *The Journal of biological chemistry* 288, 27888-27897.
- Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167-170.
- Hrle, A., Su, A. A., Ebert, J., Benda, C., Randau, L., and Conti, E. (2013). Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. *RNA biology* 10, 1670-1678.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30, 3059-3066.
- Kozin, M. B., and Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 33, 33-41.
- Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24, 945-950.
- Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., et al. (2006). Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci USA* 103, 15611-15616.
- Marraffini, L. A., and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843-1845.
- Merritt, E. A., and Bacon, D. J. (1997). Raster3D: Photo-realistic molecular graphics. *Methn Enzymol* 277, 505-524.
- Millen, A. M., Horvath, P., Boyaval, P., and Romero, D. A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. *PLoS One* 7, e51663.
- Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., and Svergun, D. I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45, 342-350.
- Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilstein-Edmands, V., Cannone, G., Graham, S., Robinson, C. V., Spagnolo, L., and White, M. F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell* 52, 124-134.
- Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

- Söding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33, W244-248.
- Staals, R. H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D. W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J. J., Sakamoto, K., et al. (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* 52, 135-145.
- Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25, 495-503
- Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal* 2879-2886.
- Volkov, V. V., and Svergun, D. I. (2003). Uniqueness of it ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36, 860-864.
- Waghmare, S. P., Pousinis, P., Homby, D. P., and Dickman, M. J. (2009). Studying the mechanism of RNA separations using RNA chromatography and its application in the analysis of ribosomal RNA and RNA:RNA interactions. *Journal of chromatography. A* 1216, 1377-1382.
- Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S. V., et al. (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 45, 303-313.
- Zheng, L., Baumann, U., and Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* 32, el 15.
- Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960-964.
- Carte, J., Wang, R., Li, H., Terns, R. M., and Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & development* 22, 3489-3496.
- Deng, L., Garrett, R. A., Shah, S. A., Peng, X., and She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Molecular microbiology* 87, 1088-1099.
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 109, E2579-2586.
- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., and Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945-956.
- Hatoum-Aslan, A., Maniv, I., and Marraffini, L. A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc Natl Acad Sci USA* 108, 21218-21222.
- Hatoum-Aslan, A., Maniv, I., Samai, P., and Marraffini, L. A. (2014). Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *J Bacteriol* 196, 310-317.
- Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., and Marraffini, L. A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *The Journal of biological chemistry* 288, 27888-27897.

- Hrle, A., Su, A. A., Ebert, J., Benda, C., Randau, L., and Conti, E. (2013). Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. *RNA biology* 10, 1670-1678.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821.
- Makarova, K. S., Aravind, L., Wolf, Y. I., and Koonin, E. V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology direct* 6, 38.
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., et al. (2011b). Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology* 9, 467-477.
- Marraffini, L. A., and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843-1845.
- Marraffini, L. A., and Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature reviews. Genetics* 11, 181-190.
- Millen, A. M., Horvath, P., Boyaval, P., and Romero, D. A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. *PLoS One* 7, e51663.
- Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733-740.
- Olsthoorn, R., and van Duin, J. (2011). Bacteriophages with ssRNA. In *eLS* (John Wiley & Sons Ltd, Chichester).
- Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C. V., Spagnolo, L., and White, M. F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell* 52, 124-134.
- Sinkunas, T., Gasiunas, G., Waghmare, S. P., Dickman, M. J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J* 32, 385-394.
- Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol Cell* 52, 146-152.
- Staals, R. H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D. W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J. J., Sakamoto, K., et al. (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* 52, 135-145.
- Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62-67.
- Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci USA* 111, 9798-9803.
- Terns, R. M., and Terns, M. P. (2014). CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends in genetics: TIG* 30, 111-118.
- Westra, E. R., van Erp, P. B., Kunne, T., Wong, S. P., Staals, R. H., Seegers, C. L., Bollen, S., Jore, M. M., Semenova, E., Severinov, K., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* 46, 595-605.

31

Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S. J., van der Oost, J., Doudna, J. A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477, 486-489.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S. V., et al. (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 45, 303-313.

The embodiments shown and described in the specification are only specific embodiments of inventors who are skilled in the art and are not limiting in any way. Therefore, various changes, modifications, or alterations to those embodiments may be made without departing from the spirit of the invention in the scope of the following claims. The references cited are expressly incorporated by reference herein in their entirety.

TABLE 1

(related to FIG. 1). Protein identified following mass spectrometry analysis of StCsm-72.				
Protein	Mass (Da)	Score	Coverage (%)	Peptides
Cas10	86891	1076	36	LAYYLTR GDYAAIATR VYINQFASDK TVETLVQFEK YFKPTVLNLK YHMANYQSDK HNYKEDLFTK LYVAFGWGSFAAK DSISLFSSDYTFK DIMSELNSPESYR (SEQ ID NO: 1)
Csm3	24541	768	46	ITAEANPR FENTIDR TLNELLTAEV AT11/FGNYDVK LLELDYLGSGSR LKATTVFGNYDVK (SEQ ID NO: 2)
Csm4	33727	584	33	KQDLYK IFSALVLESLK DGNLYQVATTR HDQIDQSDVK (SEQ ID NO: 3)
Cas6	28240	197	16	LVFTFK LIFQSLMQK (SEQ ID NO: 4)
Csm2	14817	186	21	AQILEALK VQFVYQAGR (SEQ ID NO: 5)
Csm5	41013	138	12	LISFLNDNR NHESFYEMGK (SEQ ID NO: 6)

32

TABLE 2

(related to FIG. 1): Proteins identified following mass spectrometry analysis of StCsm-40.

Protein	Mass (Da)	Score	Coverage (%)	Peptides
Cas10	86891	1149	30	LAYYLTR GDYAAIATR VYINQFASDK YFKPTVLNLK YFFNHQDER YHMANYQSDK HNYKEDLFTK LYVAFGWGSFAAK DSISLFSSDYTFK DIMSELNSPESYR IDLFYGALLHDIGK DFNQFLANFQTR FITNVYDDKLEQIR EKIDLFYGALLHDIGK GNEKDSISLFSSDYTFK IWDITYTNQADIFNVFQAQTDK SKPNFASATYEPFSKGDYAAIATR IWDITYTNQADIFNVFQAQTDKR HALVGADWFDEIADNQVISDQIR (SEQ ID NO: 7)
Csm3	24541	801	57	ITAEANPR FENTIDR TLNELLTAEV ATTVFGNYDVK LLELDYLGSGSR LKATTVFGNYDVK VAEKPSDDSDILSR DPITNLPIIPGSSLK SYTEVKFENTIDR DAFLSNADELDSLGV FENTIDRITAEANPR NSTFDLFELIYEITDENENQVEEDFK (SEQ ID NO: 8)
Csm4	33727	554	33	KQDLYK IFSALVLESLK DGNLYQVATTR HDQIDQSDVK SSGFATSHATNENYR FELDIQNIPLLELSDR FELDIQNIPLLELSDR NQPHKDGNYQVATTR SSGFGEFELDIQNIPLLELSDR (SEQ ID NO: 9)
Csm6	28240	171	16	LVFTFK LIFQSLMQK RIDHPAQDLAVK SQGSYVIFPSMR (SEQ ID NO: 10)
Csm2	14817	110		AQILEALK VQFVYQAGR (SEQ ID NO: 11)
Csm5	41013	965	50	WDYSAK QADGILQR EFIYENK FYFPDMGK TILMNTTPK KFYFPDMGK

33

TABLE 2-continued

(related to FIG. 1): Proteins identified following mass spectrometry analysis of StCsm-40.				
Protein	Mass (Da)	Coverage Score (%)	Peptides	
			VSDSKPFDNK	
			LISFLNDNR	
			NHESFYEMGK	
			EYDDLFAIR	
			WNNENAVNDFGR	
			GKEYDDLFAIR	
			KGKEYDDLFAIR	
			IEFEITTTTDEAGR	
			LSLLTLAPIHIGNEK	
			DAFGNPYIPGSSLK	
			LAEKFEAFLIQTRPNAR	
			(SEQ ID NO: 12)	

34

TABLE 3

(related to FIG. 1): Mw estimations for StCsm-40 and StCsm-72 by different methods.					
	SDS-PAGE, kDa*	DLS, kDa**	Mo W server, kDa***	Porod volume, kDa***	DAMMIN models, kDa****
5					
10	Csm-40	344.8	305 ± 75	302 ± 9	282 ± 15
	Csm-72	486.2	523 ± 128	425 ± 15	350 ± 9
15					

*Molecular mass calculated from evaluation of the complex composition by densitometric analysis of the SDS-PAGE gels.

**Molecular mass calculated from dynamic light scattering (DLS) analysis.

***Molecular mass calculated from the SAXS data by the method described in (Fischer et al., 2010) using the SAXS Mo W program run on the server <http://www.if.sc.usp.br/~saxs/saxsmow.html>.

****Molecular mass was estimated using the Porod volumes calculated from SAXS data and excluded volumes of DAMMIN models as described in (Petoukhov et al., 2012).

TABLE 4

(related to FIG. 1): SAXS data collection details and structural parameters of StCsm-40 and StCsm-72 complexes.						
Data collection parameters						
Beam line				P12		
Wavelength, nm				0.124		
Sample to detector distance, m				3.1		
Detector				Pilatus 2M		
s range, nm ⁻¹				0.975786-4.665330		
exposure time of each frame, s				0.05		
Frames collected				20		
Sample storage temperature, ° C.				10		
Cell temperature, ° C.				20		
Structural parameters						
	Csm-40			Csm-72		
Sample concentrations, mg/ml	0.13	0.52	1.34	0.20	0.65	2.00
Guinier range (first-last point) as calculated by AUTORG	14-53	26-55	19-52	8-35	21-39	11-34
P(r) calculation range, Å ⁻¹	0.0114-0.2006	0.0114-0.2006	0.0117-0.1739	0.0089-0.1076	0.0108-0.1076	0.0084-0.1049
Real space R _g , calculated by GNOM, °Å	63.59 ± 0.414	62.80 ± 0.329	63.20 ± 0.163	83.82 ± 0.545	81.40 ± 0.333	83.14 ± 0.287
Real space R _g calculated by DATGNOM, °Å	64.02	62.35	63.26	84.15	81.69	84.51
Reciprocal space R _g calculated by DATGNOM, °Å	68.08	58.04	61.34	81.51	79.71	83.79
Dmax as parameter for GNOM, °Å	210	208	215	275	265	280
Dmax calculated by DATGNOM, °Å	233.2	203.1	214.7	279.2	267.0	293.3
Porod volume estimated by DATPOROD, Å ³	452186	501468	485803	611618	589997	581121
Excluded volume of DAMMIN models, Å ³ (10 models averaged)				590770 ± 5209		791440 ± 11366

TABLE 5

(related to FIGS 2, 3, 4, 5, and 6). Nucleic acid substrates used in this study*.		
Substrates	Length, nt	Sequence
S3 crRNA in Csm-72 SEQ ID NO: 13	76/76	5'-handle 5'-ACGGAAACUUUCGUAACUGUUUAAUUCUGUUCACUUAUUCACCGGAUUAUAAACCUAAUUUACCUCCGAGAGGGG-3' Spacer S3 3'-handle
S3 crRNA in Csm-40 SEQ ID NO: 14	76/76	5'-handle 5'-ACGGAAACUUUCGUAACUGUUUAAUUCUGUUCACUUAUUC-3' Spacer S3
S3/1 DNA/ DNA	76/76	5'-GGAGCTCGCCACCTTTTGTTCGTAACCTGTTAATTCTGTTCACTTATTCCACCTCCTCAATTTGCCGGCTTGG-3' (SEQ ID NO: 15) 3'-CCTCGAGCGGTGGAAAACAAAAGCATTGACAAATTAAGACAAGTGAATAAGGTGGAGGAGTTAAACGGCCGAACC-5'

TABLE 5-continued

(related to FIGS 2, 3, 4, 5, and 6). Nucleic acid substrates used in this study*.

Substrates	Length, nt	Sequence
NS DNA/DNA	73/73	5'-GACCACCCCTTTTGGATATAATATACCTATATCAATGGCCTCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3' (SEQ ID NO: 16) 3'-CTGGTGGGAAAACTATATTATATGGATATAGTTACCGGAGGGTGCCTATTTCGCGTCTATGCAAGACTCCCTT-5'
S3/2 DNA/DNA	52/52	5'-TTTTTTTTTTTCGTAAGTGTAAATTCTGTTCACTTATTCCACCTTATATTT-3' (SEQ ID NO: 17) 3'-AAAAAAAAAAGCATTGACAAATTAAGACAAGTGAATAAGGTGGAATATAAA-5'
S3/1 DNA/RNA	76/68	5'-GGAGCTCGCCACCTTTTGTTCGTAAGTGTAAATTCTGTTCACTTATTCCACCTCCTCAATTTGCCGGCTTGG-3' (SEQ ID NO: 18) 3'-AGCGGUGGAAAAACAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGGUGGAGGAGUUAACGGCGGG-5' O1205 O1205
S3/2 DNA/RNA	52/68	5'-TTTTTTTTTTTCGTAAGTGTAAATTCTGTTCACTTATTCCACCTTATATTT-3' SEQ ID NO: 17 (same as above) 3'-CCAUGGGAAAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGGUGGAAUUAACCCCUAGGG-5'
S3/1 DNA	76	3'-CCTCGAGCGGTGGAAAACAAAAGCATTGACAAATTAAGACAAGTGAATAAGGTGGAGGAGTTAAACGGCCGAACC-5' O1205 O1205 SEQ ID NO: 19
NS DNA	73	3'-CTGGTGGGAAAACTATATTATATGGATATAGTTACCGGAGGGTGCCTATTTCGCGTCTATGCAAGACTCCCTT-5'
S3/2 DNA	52	3'-AAAAAAAAAAGCATTGACAAATTAAGACAAGTGAATAAGGTGGAATATAAA-5' SEQ ID NO: 21
S3/1 RNA	68	3'-AGCGGUGGAAAAACAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGGUGGAGGAGUUAACGGCGGG-5' O1205 O1205 SEQ ID NO: 22
NS RNA	68	3'-AGCGGUGGAAAAACACCACCUUAUUCACUUGUCUUAUUUGUCAUUGCUUUAAGGAGUUAACGGCGGG-5' O1205 O1205 SEQ ID NO: 23
S3/2 RNA	68	3'-CCAUGGGAAAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGGUGGAAUUAACCCCUAGGG-5' SEQ ID NO: 24
S3/3 RNA	68	3'-CCAUGGGUGCCUUUGAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGGUGGAAUUAACCCCUAGGG-5' SEQ ID NO: 25
S3/4 RNA	68	3'-CCAUGGGAAAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGGUGGAAUUAACCCCUAGGG-5' SEQ ID NO: 26
S3/5 RNA	68	3'-CCAUGGGUGCCUUUGAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGGUGGCUAUUUUGGAUUAGGG-5' SEQ ID NO: 27
S3/6 RNA	86	3'- CCAUGGGUGCCUUUGAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGGUGGCUAUUUUGGAUUUAUGGAGCUCUCCCGCCGGG-5' SEQ ID NO: 28
S3/7 RNA	68	3'-CCAUGGGAAAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGGUGGAAUUAACCCCUAGGG-5' SEQ ID NO: 29

TABLE 5-continued

		(related to FIGS 2, 3, 4, 5, and 6). Nucleic acid substrates used in this study*.	
S3/8	68		3'-CCAUGGGAAAAAAAAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGCAGGAAUAUAAACCCCUAGGG-5'
RNA			SEQ ID NO: 30
S3/9	68		3'-CCAUGGGAAAAAAAAAAAGCAUUGACAAAUAUCACAAGUGAAUAAGGUGGAAUAUAAACCCCUAGGG-5'
RNA			SEQ ID NO: 31
S3/11	24		3'-GACAAAUAAGACAAGUGAAUGGG-5'
RNA			SEQ ID NO: 32
S3/12	32		3'-AAAGCAUUGACAAAUAAGACAAGUGAAUGGG-5'
RNA			SEQ ID NO: 33
S3/13	23		3'-AAAGCAUUGACAAAUAAGAGGG-5'
RNA			SEQ ID NO: 34
S3/14	48		3'-CCAUGGGAAAAAAAAAAAGCAUUGACAAAUAAGACAAGUGAAUAAGG-5'
RNA			SEQ ID NO: 35
(+Tc) crRNA in Csm-72		5'-handle	Spacer (+Tc) 3'-handle
		5'-ACGGAACACGCCAGCAAGACGUAGCCCAGCGCGUCGGCCGCGCAUAUAAACCUAAUUACCUAGAGGGG-3'	
			SEQ ID NO: 36
(+Tc) crRNA in Csm-40		5'-handle	Spacer (+Tc)
		5'-ACGGAACACGCCAGCAAGACGUAGCCCAGCGCGUCGGCC-3'	
			SEQ ID NO: 37
Sub-	Length,	Sequence	
strates	nt		
(+Tc)	68		3'-UCGGAGCGCAGCGCUUUGCGGUCGUUCUGCAUCGGGUCGCGCAGCCGGCGGUACGGCCGCUAAUUACGGG-5'
RNA			SEQ ID NO: 38
(-Tc) crRNA in Csm-72		5'-handle	Spacer (-Tc) 3'-handle
		5'-ACGGAACUGGGCGCCGACGCGUGGGCUACGUCUUGCUGGCGUGAUUAAACCUAAUUACCUAGAGGGG-3'	
			SEQ ID NO: 39
(-Tc) crRNA in Csm-40		5'-handle	Spacer (-Tc)
		5'-ACGGAACUGGGCGCCGACGCGUGGGCUACGUCUUGCUG-3'	
			SEQ ID NO: 40
Sub-	Length,	Sequence	
strates	nt		
(-Tc)	68		3'-GGUAAUAGCGCCGUACCGCCGGCUGCGCGACCCGAUGCAGAACCAGCCGAAGCGCUGCGCUCCGGG-5'
RNA			SEQ ID NO: 41
GFP crRNA in Csm-40		5'-handle	Spacer GFP
		5'-ACGGAACAUAAACCUUCGGCAUGGCACUCUUGAAAAAGU-3'	
			SEQ ID NO: 42
Sub-	Length,	Sequence	
strates	nt		
GFP	68		3'-UCAAGAAAGGACAUGUAUUGGAAGCCCGUACCGUGAGAACUUUUUCAGUACGACAAAGUAUACUAGGG-5'
RNA			SEQ ID NO: 43
Rep crRNA in Csm-72		5'-handle	Spacer Rep 3'-handle
		5'-ACGGAACAACGAGCCUAAAUAUGACUCGUUAUAGCGGAGAUUAAACCUAAUUACCUAGAGGGG-3'	
			SEQ ID NO: 44
Rep crRNA in Csm-40		5'-handle	Spacer Rep
		5'-ACGGAACAACGAGCCUAAAUAUGACUCGUUAUAG-3'	
			SEQ ID NO: 45
Sub-	Length,	Sequence	
strates	nt		

-continued

35	40	45	
Met Ala Asn Tyr Gln Ser Asp Lys His Asn Tyr Lys Glu Asp Leu Phe			
50	55	60	
Thr Lys Leu Tyr Val Ala Phe Gly Trp Gly Ser Phe Ala Ala Lys Asp			
65	70	75	80
Ser Ile Ser Leu Phe Ser Ser Asp Tyr Thr Phe Lys Asp Ile Met Ser			
	85	90	95
Glu Leu Asn Ser Pro Glu Ser Tyr Arg			
100		105	

<210> SEQ ID NO 2
 <211> LENGTH: 60
 <212> TYPE: PRT
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 2

Ile Thr Ala Glu Ala Asn Pro Arg Phe Glu Asn Thr Ile Asp Arg Thr			
1	5	10	15
Leu Asn Glu Leu Leu Thr Ala Glu Val Ala Thr Phe Gly Asn Tyr Asp			
20	25	30	
Val Lys Leu Leu Glu Leu Asp Tyr Leu Gly Gly Ser Gly Ser Arg Leu			
35	40	45	
Lys Ala Thr Thr Val Phe Gly Asn Tyr Asp Val Lys			
50	55	60	

<210> SEQ ID NO 3
 <211> LENGTH: 39
 <212> TYPE: PRT
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 3

Lys Gln Asp Leu Tyr Lys Ile Phe Ser Ala Leu Val Leu Glu Ser Leu			
1	5	10	15
Lys Asp Gly Asn Leu Tyr Gln Val Ala Thr Thr Arg His Asp Gln Ile			
20	25	30	
Asp Gln Ser Val Asp Val Lys			
35			

<210> SEQ ID NO 4
 <211> LENGTH: 15
 <212> TYPE: PRT
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 4

Leu Val Phe Thr Phe Lys Leu Ile Phe Gln Ser Leu Met Gln Lys			
1	5	10	15

<210> SEQ ID NO 5
 <211> LENGTH: 17
 <212> TYPE: PRT
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 5

Ala Gln Ile Leu Glu Ala Leu Lys Val Gln Phe Val Tyr Gln Ala Gly			
1	5	10	15

Arg

<210> SEQ ID NO 6
 <211> LENGTH: 19
 <212> TYPE: PRT

-continued

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 6

Leu Ile Ser Phe Leu Asn Asp Asn Arg Asn His Glu Ser Phe Tyr Glu
 1 5 10 15

Met Gly Lys

<210> SEQ ID NO 7

<211> LENGTH: 268

<212> TYPE: PRT

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 7

Leu Ala Tyr Tyr Leu Thr Arg Gly Asp Tyr Ala Ala Ile Ala Thr Arg
 1 5 10 15

Val Tyr Ile Asn Gln Phe Ala Ser Asp Lys Tyr Phe Lys Pro Thr Val
 20 25 30

Leu Asn Leu Lys Tyr Phe Phe Asn His Gln Asp Glu Arg Tyr His Met
 35 40 45

Ala Asn Tyr Gln Ser Asp Lys His Asn Tyr Lys Glu Asp Leu Phe Thr
 50 55 60

Lys Leu Tyr Val Ala Phe Gly Trp Gly Ser Phe Ala Ala Lys Asp Ser
 65 70 75 80

Ile Ser Leu Phe Ser Ser Asp Tyr Thr Phe Lys Asp Ile Met Ser Glu
 85 90 95

Leu Asn Ser Pro Glu Ser Tyr Arg Ile Asp Leu Phe Tyr Gly Ala Leu
 100 105 110

Leu His Asp Ile Gly Lys Asp Phe Asn Gln Phe Leu Leu Ala Asn Phe
 115 120 125

Gln Thr Arg Phe Ile Thr Asn Val Tyr Asp Asp Lys Leu Glu Gln Ile
 130 135 140

Arg Glu Lys Ile Asp Leu Phe Tyr Gly Ala Leu Leu His Asp Ile Gly
 145 150 155 160

Lys Gly Asn Glu Lys Asp Ser Ile Ser Leu Phe Ser Ser Asp Tyr Thr
 165 170 175

Phe Lys Ile Trp Asp Thr Tyr Thr Asn Gln Ala Asp Ile Phe Asn Val
 180 185 190

Phe Gly Ala Gln Thr Asp Lys Ser Lys Pro Asn Phe Ala Ser Ala Thr
 195 200 205

Tyr Glu Pro Phe Ser Lys Gly Asp Tyr Ala Ala Ile Ala Thr Arg Ile
 210 215 220

Trp Asp Thr Tyr Thr Asn Gln Ala Asp Ile Phe Asn Val Phe Gly Ala
 225 230 235 240

Gln Thr Asp Lys Arg His Ala Leu Val Gly Ala Asp Trp Phe Asp Glu
 245 250 255

Ile Ala Asp Asn Gln Val Ile Ser Asp Gln Ile Arg
 260 265

<210> SEQ ID NO 8

<211> LENGTH: 160

<212> TYPE: PRT

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 8

Ile Thr Ala Glu Ala Asn Pro Arg Phe Glu Asn Thr Ile Asp Arg Thr
 1 5 10 15

-continued

Leu Asn Glu Leu Leu Thr Ala Glu Val Ala Thr Thr Val Phe Gly Asn
 20 25 30
 Tyr Asp Val Lys Leu Leu Glu Leu Asp Tyr Leu Gly Gly Ser Gly Ser
 35 40 45
 Arg Leu Lys Ala Thr Thr Val Phe Gly Asn Tyr Asp Val Lys Val Ala
 50 55 60
 Glu Lys Pro Ser Asp Asp Ser Asp Ile Leu Ser Arg Asp Pro Ile Thr
 65 70 75 80
 Asn Leu Pro Ile Ile Pro Gly Ser Ser Leu Lys Ser Tyr Thr Glu Val
 85 90 95
 Lys Phe Glu Asn Thr Ile Asp Arg Asp Ala Phe Leu Ser Asn Ala Asp
 100 105 110
 Glu Leu Asp Ser Leu Gly Val Arg Phe Glu Asn Thr Ile Asp Arg Ile
 115 120 125
 Thr Ala Glu Ala Asn Pro Arg Asn Ser Thr Phe Asp Phe Glu Leu Ile
 130 135 140
 Tyr Glu Ile Thr Asp Glu Asn Glu Asn Gln Val Glu Glu Asp Phe Lys
 145 150 155 160

<210> SEQ ID NO 9
 <211> LENGTH: 124
 <212> TYPE: PRT
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 9

Lys Gln Asp Leu Tyr Lys Ile Phe Ser Ala Leu Val Leu Glu Ser Leu
 1 5 10 15
 Lys Asp Gly Asn Leu Tyr Gln Val Ala Thr Thr Arg His Asp Gln Ile
 20 25 30
 Asp Gln Ser Val Asp Val Lys Ser Ser Gly Phe Ala Phe Ser His Ala
 35 40 45
 Thr Asn Glu Asn Tyr Arg Phe Glu Leu Asp Ile Gln Asn Ile Pro Leu
 50 55 60
 Glu Leu Ser Asp Arg Phe Glu Leu Asp Ile Gln Asn Ile Pro Leu Glu
 65 70 75 80
 Leu Ser Asp Arg Leu Thr Lys Asn Gln Pro His Lys Asp Gly Asn Leu
 85 90 95
 Tyr Gln Val Ala Thr Thr Arg Ser Ser Gly Phe Gly Glu Phe Glu Leu
 100 105 110
 Asp Ile Gln Asn Ile Pro Leu Glu Leu Ser Asp Arg
 115 120

<210> SEQ ID NO 10
 <211> LENGTH: 39
 <212> TYPE: PRT
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 10

Leu Val Phe Thr Phe Lys Leu Ile Phe Gln Ser Leu Met Gln Lys Arg
 1 5 10 15
 Ile Asp His Pro Ala Gln Asp Leu Ala Val Lys Ser Gln Gly Ser Tyr
 20 25 30
 Val Ile Phe Pro Ser Met Arg
 35

<210> SEQ ID NO 11
 <211> LENGTH: 17

-continued

<212> TYPE: DNA
 <213> ORGANISM: Streptococcus thermophilus
 <400> SEQUENCE: 15
 ggagctcgcc accttttgtt tttcgtaact gtttaattct gttcacttat tccacctcct 60
 caatttgccg gcttgg 76

<210> SEQ ID NO 16
 <211> LENGTH: 73
 <212> TYPE: DNA
 <213> ORGANISM: Streptococcus thermophilus
 <400> SEQUENCE: 16
 gaccaccctt tttgatataa tataacctata tcaatggcct cccacgcata agcgcagata 60
 cgttctgagg gaa 73

<210> SEQ ID NO 17
 <211> LENGTH: 52
 <212> TYPE: DNA
 <213> ORGANISM: Streptococcus thermophilus
 <400> SEQUENCE: 17
 tttttttttt tcgtaactgt ttaattctgt tcacttattc caccttatat tt 52

<210> SEQ ID NO 18
 <211> LENGTH: 52
 <212> TYPE: DNA
 <213> ORGANISM: Streptococcus thermophilus
 <400> SEQUENCE: 18
 tttttttttt tcgtaactgt ttaattctgt tcacttattc caccttatat tt 52

<210> SEQ ID NO 19
 <211> LENGTH: 76
 <212> TYPE: DNA
 <213> ORGANISM: Streptococcus thermophilus
 <400> SEQUENCE: 19
 cctcgagcgg tggaaaacaa aaagcattga caaattaaga caagtgaata aggtggagga 60
 gttaaaccggc cgaacc 76

<210> SEQ ID NO 20
 <211> LENGTH: 73
 <212> TYPE: DNA
 <213> ORGANISM: Streptococcus thermophilus
 <400> SEQUENCE: 20
 ctgggtgggaa aaactatatt atatggatat agttaccgga ggggtgcgtat tcgcgtctat 60
 gcaagactcc ctt 73

<210> SEQ ID NO 21
 <211> LENGTH: 52
 <212> TYPE: DNA
 <213> ORGANISM: Streptococcus thermophilus
 <400> SEQUENCE: 21
 aaaaaaaaaa agcattgaca aattaagaca agtgaataag gtggaatata aa 52

<210> SEQ ID NO 22
 <211> LENGTH: 68
 <212> TYPE: RNA

-continued

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 22

agcgguggaa aacaaaaagc auugacaaau uaagacaagu gaauaaggug gaggaguuaa 60

acggcggg 68

<210> SEQ ID NO 23

<211> LENGTH: 68

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 23

agcgguggaa aacaaccacc uuauucacuu gucuuaauuu gucaaugcuu uaggaguuaa 60

acggcggg 68

<210> SEQ ID NO 24

<211> LENGTH: 68

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 24

ccaugggaaa aaaaaaagc auugacaaau uaagacaagu gaauaaggug gaauuaaac 60

cccuaggg 68

<210> SEQ ID NO 25

<211> LENGTH: 68

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 25

ccaugggugc cuuugaaagc auugacaaau uaagacaagu gaauaaggug gaauuaaac 60

cccuaggg 68

<210> SEQ ID NO 26

<211> LENGTH: 68

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 26

ccaugggaaa aaaaaaagc auugacaaau uaagacaagu gaauaaggug gcuauuuug 60

gauuaggg 68

<210> SEQ ID NO 27

<211> LENGTH: 68

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 27

ccaugggugc cuuugaaagc auugacaaau uaagacaagu gaauaaggug gcuauuuug 60

gauuaggg 68

<210> SEQ ID NO 28

<211> LENGTH: 86

<212> TYPE: RNA

<213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 28

ccaugggugc cuuugaaagc auugacaaau uaagacaagu gaauaaggug gcuauuuug 60

gauuaaugga gcucucccu gccggg 86

-continued

<210> SEQ ID NO 29
 <211> LENGTH: 68
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 29

 ccaugggaaa aaaaaaaucc auugacaaa uagacaagu gaauaaggug gaauuaaac 60
 ccuaggg 68

<210> SEQ ID NO 30
 <211> LENGTH: 68
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 30

 ccaugggaaa aaaaaaagc auugacaaa uagacaagu gaauaagcag gaauuaaac 60
 ccuaggg 68

<210> SEQ ID NO 31
 <211> LENGTH: 68
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 31

 ccaugggaaa aaaaaaagc auugacaaa uacacaagu gaauaaggug gaauuaaac 60
 ccuaggg 68

<210> SEQ ID NO 32
 <211> LENGTH: 24
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 32

 gacaaauuaa gacaagugaa uggg 24

<210> SEQ ID NO 33
 <211> LENGTH: 32
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 33

 aaagcauuga caauuaaga caagugaaug gg 32

<210> SEQ ID NO 34
 <211> LENGTH: 23
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 34

 aaagcauuga caauuaaga ggg 23

<210> SEQ ID NO 35
 <211> LENGTH: 48
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 35

 ccaugggaaa aaaaaaagc auugacaaa uagacaagu gaauaagg 48

<210> SEQ ID NO 36

-continued

<211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 36

 acggaacac gccagcaaga cguagcccag cgcgucggcc gccagauaua aaccuaauua 60
 ccucgagagg gg 72

 <210> SEQ ID NO 37
 <211> LENGTH: 40
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 37

 acggaacac gccagcaaga cguagcccag cgcgucggcc 40

 <210> SEQ ID NO 38
 <211> LENGTH: 68
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 38

 ucggagcgca ggcguugcgg ucguucugca ucgggucgcg cagccggcgg uacggccgcu 60
 auuacggg 68

 <210> SEQ ID NO 39
 <211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 39

 acggaacug gcgcccgacg cgcugggcua cgucuugcug gcgugauaua aaccuaauua 60
 ccucgagagg gg 72

 <210> SEQ ID NO 40
 <211> LENGTH: 40
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 40

 acggaacug gcgcccgacg cgcugggcua cgucuugcug 40

 <210> SEQ ID NO 41
 <211> LENGTH: 68
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

 <400> SEQUENCE: 41

 gguaauagcg gccguaccgc cggcugcgcg acccgaugca gaacgaccgc aagcgcugcg 60
 cuccgggg 68

 <210> SEQ ID NO 42
 <211> LENGTH: 40
 <212> TYPE: RNA
 <213> ORGANISM: Aequorea victoria

 <400> SEQUENCE: 42

 acggaacau aaccuucggg cauggcacuc uugaaaaagu 40

 <210> SEQ ID NO 43
 <211> LENGTH: 68

-continued

<212> TYPE: RNA
 <213> ORGANISM: *Aequorea victoria*

 <400> SEQUENCE: 43

 ucaagaaagg acauguauug gaagcccgua ccgugagaac uuuuucagua cgacaaagua 60
 uacuaggg 68

 <210> SEQ ID NO 44
 <211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: *Streptococcus thermophilus*

 <400> SEQUENCE: 44

 acggaaacca acgagccuaa auucauuga cucguuauag cggagauua aaccuaaua 60
 ccucgagagg gg 72

 <210> SEQ ID NO 45
 <211> LENGTH: 40
 <212> TYPE: RNA
 <213> ORGANISM: *Streptococcus thermophilus*

 <400> SEQUENCE: 45

 acggaaacca acgagccuaa auucauuga cucguuauag 40

 <210> SEQ ID NO 46
 <211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: *Streptococcus thermophilus*

 <400> SEQUENCE: 46

 caagggaugu ugcucggauu uaaguauacu gagcaauauc gccuggcgca cagacuaggu 60
 gccgcgugua ac 72

 <210> SEQ ID NO 47
 <211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: *Streptococcus thermophilus*

 <400> SEQUENCE: 47

 acggaaacug ucuucgacau ggguaauccu cauguuugaa uggcgauua aaccuaaua 60
 ccucgagagg gg 72

 <210> SEQ ID NO 48
 <211> LENGTH: 40
 <212> TYPE: RNA
 <213> ORGANISM: *Streptococcus thermophilus*

 <400> SEQUENCE: 48

 acggaaacug ucuucgacau ggguaauccu cauguuugaa 40

 <210> SEQ ID NO 49
 <211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: *Streptococcus thermophilus*

 <400> SEQUENCE: 49

-continued

 aagaaacaac agaagcugua cccauuagga guacaaacuu accggccgca gauaaucauc 60

uacggccuca aa 72

<210> SEQ ID NO 50
 <211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 50

acggaaacuc uuuuaggaga ccuugcaug ccuaaacaau aagcgauua aaccuaauua 60

ccucgagagg gg 72

<210> SEQ ID NO 51
 <211> LENGTH: 40
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 51

acggaaacuc uuuuaggaga ccuugcaug ccuaaacaau 40

<210> SEQ ID NO 52
 <211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 52

ccaaagguag aaaauccucu ggaacguaac ggaauuguua uucgagcgc agccuaagc 60

aucgcuuuua ac 72

<210> SEQ ID NO 53
 <211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 53

acggaaacag uuugcagcug gauacgacag acggccaucu aacugauua aaccuaauua 60

ccucgagagg gg 72

<210> SEQ ID NO 54
 <211> LENGTH: 38
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 54

acggaaacag uuugcagcug gauacgacag acggccau 38

<210> SEQ ID NO 55
 <211> LENGTH: 72
 <212> TYPE: RNA
 <213> ORGANISM: Streptococcus thermophilus

<400> SEQUENCE: 55

acagaccuuc aaacgucgac cuaugcuguc ugccguaga uugaacuaca aucauggcug 60

gacugcaugc cg 72

61

What is claimed is:

1. A method of programmable RNA shredding comprising:

providing a Type III-A CRISPR-Cas (Csm) complex comprising at least crRNA, Csm4, and Csm3, or at least crRNA, Csm4, Csm3, and Cas10, and any other subunits;

identifying target RNA;

exposing the complex to the target RNA, thereby generating a plurality of cleaved fragments of the target RNA;

wherein the cleavage sites are at 6 nucleotide intervals.

2. The method of claim 1 wherein 1 to 10 Csm3 subunits are present in the complex.

3. The method of claim 1 wherein the crRNA comprises a 5' handle and a spacer, and optionally a 3' handle, wherein the spacer is complementary or substantially complementary to a region of a target RNA.

4. The method of claim 1 wherein the programmable RNA shredding occurs in vitro.

5. The method of claim 1 wherein the programmable RNA shredding occurs in vivo.

6. A method of programmable RNA knock-out or knock-down comprising:

identifying a target RNA;

exposing the target RNA to a Type III-A Csm (III-A) (Csm) complex comprising at least crRNA, Csm4, and Csm3 and any other subunits;

62

providing conditions facilitating the cutting of the target RNA by the complex;

wherein the expression of a gene complementary to the target RNA is knocked down or knocked out.

7. The method of claim 6 wherein the Csm3 contains a mutation which inactivates the endoribonuclease activity, and the method results in RNA knock-down.

8. The method of claim 6 wherein the complex further comprises Cas 10.

9. The method of claim 6 wherein the programmable RNA knock-out or knock-down occurs in vitro.

10. The method of claim 6 wherein the programmable RNA knock-out or knock-down occurs in vivo.

11. A composition comprising an engineered Csm complex comprising only crRNA, Csm4, and Csm3, or only crRNA, Csm4, Csm3, and Cas10, where the crRNA of the engineered complex is configured for complementary binding to a selected site in a target RNA molecule, and wherein the Csm complex is capable of cutting the target RNA molecule under suitable conditions.

12. The composition of claim 11 wherein 1 to 10 Csm3 subunits are present in the complex.

13. The composition of claim 11 wherein the RNA cutting occurs in vitro.

14. The composition of claim 11 wherein the RNA cutting occurs in vivo.

* * * * *