The complete chemical structure of *Saccharomyces cerevisiae* rRNA: Partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9

Masato Taoka*[1,2], Yuko Nobe[1,2], Yuka Yamaki[1,2], Yoshio Yamauchi[1,2], Hideaki Ishikawa[2,3], Nobuhiro Takahashi[2,3], Hiroshi Nakayama[2,4], Toshiaki Isobe*[1,2]

[1]Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, Minami-osawa 1-1, Hachioji-shi, Tokyo 192-0397, Japan

[2]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Sanbancho 5, Chiyoda-ku, Tokyo 102-0075, Japan

[3]Department of Biotechnology, United Graduate School of Agriculture, Tokyo University of Agriculture and Technology, Saiwai-cho 3-5-8, Fuchu-shi, Tokyo 183-8509, Japan

[4]Biomolecular Characterization Unit, RIKEN Center for Sustainable Resource Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

*Corresponding authors:

Masato Taoka, Ph.D., Laboratory of Biochemistry, Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, Minami-osawa 1-1, Hachioji-shi, Tokyo 192-0397, Japan

Tel: +81 426 77 2543; Fax: +81 426 77 2525

Email: mango@tmu.ac.jp

Toshiaki Isobe, Ph.D., Laboratory of Biochemistry, Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, Minami-osawa 1-1, Hachioji-shi, Tokyo 192-0397, Japan

Tel: +81 426 77 5667; Fax: +81 426 77 2525

Email: isobe-toshiaki@tmu.ac.jp

Abbreviations: LC, liquid chromatography; MS, mass spectrometry; TEAA, Triethylammonium acetate; snoRNA, small nucleolar RNA; CID, collision-induced dissociation; *Sc*, *Saccharomyces cerevisiae*; *Sp*, *Schizosaccharomyces pombe*; Nm, 2'-O-methylnucleotide; Np, nucleoside 3'-monophosphate; N>p, 2',3'-cyclic-nucleotide; Ψ, pseudouridine; MS^2, MS/MS; RNP, ribonucleoprotein; RP, ribosomal protein; BPC, base peak chromatogram; EIC, extracted ion chromatogram.

Supplementary figure legends

Supplementary Figure S1 | Determination of a stoichiometry of RNA modification

(A) Extracted ion chromatogram of a single RNA fragment derived from natural source containing ^{12}C (black peak) or transcribed in vitro containing ^{13}C (blue peak). Note that the black peak is an unmodified, and the blue peak is a sum of unmodified and modified forms of the same RNA fragment. The stoichiometry of modification is thus determined as a ratio of the difference in signal intensities between the blue and black peaks (Δh) and the signal intensity of the blue peak (h_{full}), *i.e.* $\Delta h/h_{full}$, as indicated in the figure. Other details are given in the text. (B) Mass spectrum at the peak top in (A). Each spectrum ensures the purity of chromatographic peak in (A) and confirms the stoichiometry of modification determined.

Supplementary Figure S2 | The complete chemical structure of *Sc* rRNAs and the fragments used for the structural analysis. Black solid and black shaded bars denote the fragments produced by the RNase T1 and RNase A digestions of rRNAs, respectively. Blue arrows are RNase H–digested fragments of the rRNAs (fragment numbers correspond to the mapping data shown in Supplementary Table S1-2). Blue solid and blue shaded bars denote RNase T1–digested and RNase A–digested fragments of the RNase H fragments, respectively. All fragments were identified by Ariadne (see Methods) except for several heavily modified RNA fragments, which were identified by manual inspection of their tandem mass spectra. Modified RNAs are shown in blue letters, and their abbreviations are summarised as follows: P, pseudouridine; Λ, 2'-O-methyladenosine; B, 2'-O-methylcytidine; #, 2'-O-methylguanosine; J, 2'-O-methyluridine; Z, 2'-O-methylpseudouridine; Π, 1-methyladenosine; ζ, N6, N6-dimethyladenosine; Σ, 5-methylcytidine; 7, 7-methylguanosine; δ, 3-methyluridine; M, N4-acetylcytidine; α, 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine.

Supplementary Figure S3 | The tandem mass spectra of *Sc* 25S rRNA fragments

containing U, Um, or Ψ2345 produced by RNase T1 digestion. (A) ^{2334}UCmCCΨAUCUACUAΨCΨA^{2351}Gp, (B) ^{2334}UCmCCΨAUCUACUmAΨCΨA^{2351}Gp, and (C) ^{2334}UCmCCΨAUCUACΨAΨCΨA^{2351}Gp. The blue arrows identify the major a, c, w, and y ions. The cleavage positions of the assigned ions are mapped on each RNA sequence in the inset. Errors determined by Ariadne of MS^2 signals are plotted under each spectrum.

Supplementary Figure S4 | SILNAS-based quantification showing that U954, U2919, G1140, and G2 in *Sc* 25S rRNA are not modified. Previous reports found the modified nt, m^{5}U954, m^{5}U2919, Gm1140, and Gm2393 (Kiss-Laszlo, Z. *et al., Cell* **85**, 1077-1088 (1996); Bakin, A. *et al., Biochemistry* **33**, 13475-13483 (1994); Birkedal, U. *et al. Angew Chem Int Ed Engl* **54**, 451-455 (2014)). (A) Extracted ion monitoring of the 25S rRNA fragments containing U954 produced by RNase T1 digestion. Purified unlabelled 25S rRNA was mixed with an equal amount of the corresponding guanosine-$^{13}C_{10}$-labelled fragment (*G). The mixture was digested with RNase T1 and subjected to LC-MS. The MS signals of [^{952}UUUCCCUCA^{961}*Gp]$^{3-}$ and [^{952}UUUCCCUCA^{961}Gp]$^{3-}$ and [^{952}UUUCCCΨCA^{961}Gp]$^{3-}$ are indicated by arrows (upper and middle panels, respectively). The charges, *m/z* values, and sequences of the ion fragments are indicated. A mass window of 5 ppm was used for the extractions. The most intense signal in the top panel was set to 100%, and the peak in the middle and bottom panel scaled accordingly. A peak for the fragment containing m^{5}U954 not detected (lower panel).

(B) Extracted ion monitoring of the 25S rRNA fragments containing U2919 produced by RNase T1 digestion LC-MS of the 25S rRNA fragment was performed as in (A). The MS signals for [^{2914}UUCACCCACUAAUA^{2928}*Gp]$^{3-}$ and [^{2910}AUUmGmΨUCACCCACUAAUA^{2928}Gp]$^{4-}$ are indicated by arrows (upper and middle panels, respectively). A peak corresponding to [^{2910}AUUmGmΨm^{5}UCACCCACUAAUA^{2928}Gp]$^{4-}$ was not found (lower panel).

(C) Extracted ion monitoring of the 25S rRNA fragments containing G1140 produced by

RNase A digestion. Purified 25S rRNA was mixed with an equal amount of $^{13}C_9$-uracil- (*U) and $^{13}C_9$-cytosine-labelled 25S rRNA. After RNase H digestion of the mixture, fragment H12 (Supplementary Fig. S2) was purified by reversed-phase LC through a PLRP-S 4000 column, digested with RNase A, and subjected to LC-MS. The arrows identify the MS signals of $[GA*Up]^{3-}$ and $[GAUp]^{3-}$ in the upper and middle panels, respectively. AG*Up and AGUp are the sequence isomers located at positions 1110–1112 and 1384–1386 of Fragment H12, respectively. Note that Fragment H12 produced three GAUp fragments located at positions 1140–1142, 1347–1349, and 1352–1354, but no signals were detected for methylated GAUp (lower panel).

(D) Extracted ion monitoring of RNase A–generated fragments of 25S rRNA containing G2393. The analysis was performed as in (C). The arrows indicate the MS signals of $[^{2391}GGGGAAAGAAGA^{2403}*Cp]^{3-}$ and $[^{2391}GGGGAAAGAAGA^{2403}Cp]^{3-}$ (upper and middle panels, respectively). #, the fifth isotope of $[^{2391}GGGGAAAGAAGA^{2403}*Cp]^{3-}$. No signal was detected for $[^{2391}GGGmGAAAGAAGA^{2403}Cp]^{3-}$ (lower panel).

Supplementary Figure S5 | 3D modification maps of *Sc* rRNAs including and excluding RPs. The modified nt found in this study are assigned to the 3D structure of the *Sc* rRNAs (3U5B.pdb and 3U5D.pdb). The RPs are depicted according to the 3D source files of the *Sc* rRNAs (3U5C.pdb and 3U5E.pdb). (A) SSU and (B) LSU. The rRNA structures with and without RPs are shown in the right and left panels, respectively. The positions of the modified nt within the (A) SSU and (B) LSU are indicated by coloured balls: yellow, Ψ; red, 2'-O-methylated nt; blue, base modified nt.

Supplementary Figure S6 | Superimposed modification map of *Sc* and Sp rRNAs. The modified nt found in SSUs and LLUs of the Sc and *Sp* rRNAs (Taoka *et al*. *Nucleic Acids Res*, 2015, **43**, e115) were assigned to the 3D structure of the *Sc* rRNAs (3U5B.pdb and 3U5D.pdb). The positions of the modified nt within the (A) SSU and (B) LSU are indicated by

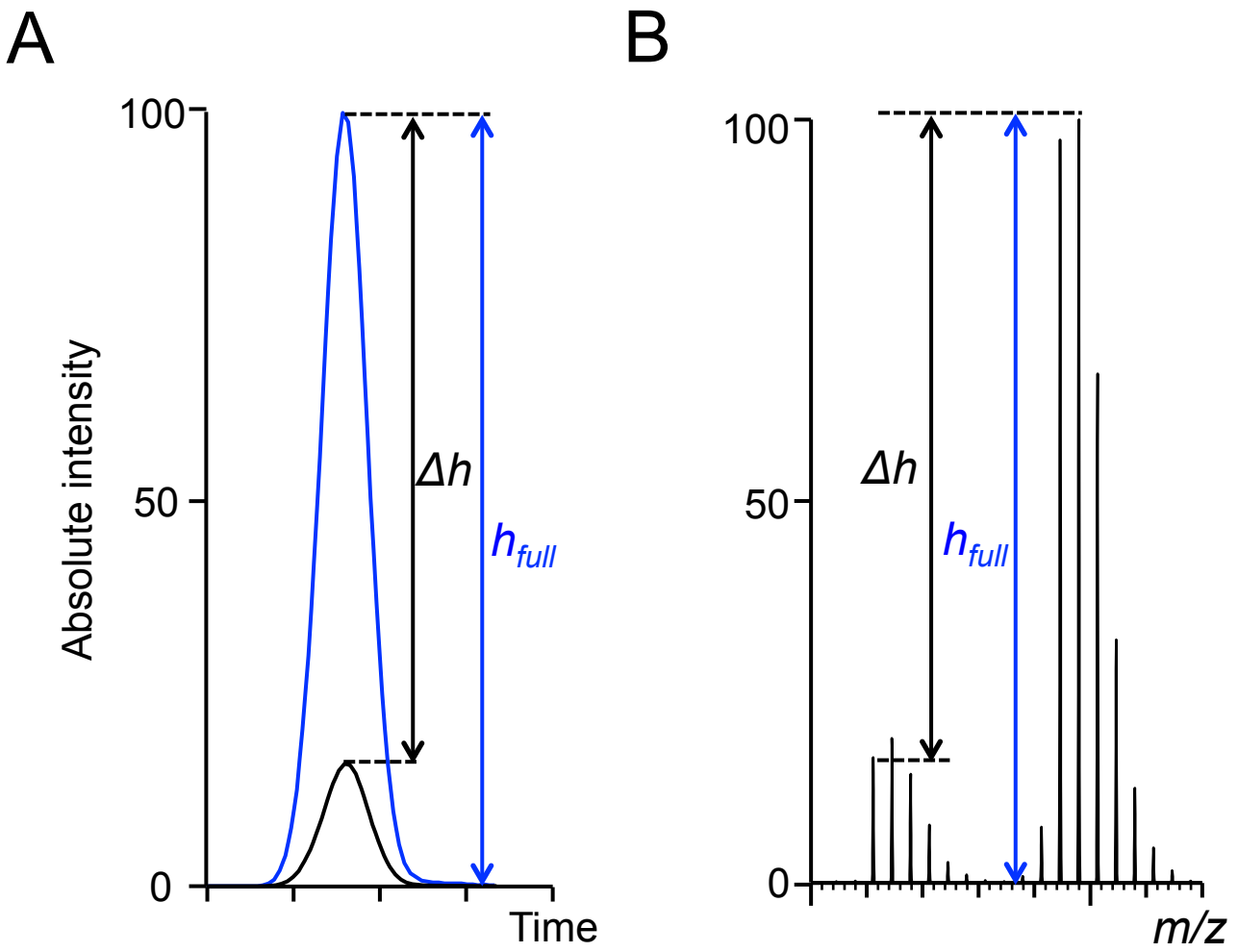
coloured balls: yellow, pseudouridine; red, 2'-O-methylated nt; blue, base modified nt; purple, modified nt unique to *Sc* rRNA; and white, modified nt unique to *Sp* rRNA.

Supplementary Figure S7 | *Sc* rRNA map of partially modified nt. The partially modified nt found in this study are mapped onto the 3D structures of the *Sc* rRNAs (3U5B.pdb and 3U5D.pdb). The positions of the modified nt within the rRNAs are indicated by coloured balls: green, partially modified nt common to *Sc* and *Sp* rRNA; and pink, partially modified nt unique to *Sc* rRNA. (A) SSU and LSU complex, (B) SSU and (C) LSU.

Supplementary Figure S8 | Potential base-pairing interactions between snR9 and snR33 with the 25S rRNA fragments containing Ψ2345 or Ψ2338, respectively. The snR9/snR33 sequences are shown as the upper strands, with their hairpin domains depicted as solid lines. The rRNA sequences around Ψ2345 and U/Ψ2338 are shown in the lower strands, with an arrow pointing to the respective Ψs.

Supplementary Figure S9 | The 25S rRNA in ΔsnR9 does not contain Ψ2338 and Ψ2345. The RNase T1 digest of U/C-5D-labelled and reverse-phase LC–purified 25S rRNA (100 fmol) was subjected to LC-MS and MS^2. (A) Extracted ion chromatogram of the fragments from the RNaseT1 digest containing U2338 and U2345. The extracted ion has an *m/z* of 1422.192 within a mass window of ±10 ppm. (B) Mass spectrum of the RNA peak in (A). The most abundant isotopomer is marked by an asterisk *. The *m/z* value of the most abundant signal coincide with that of the theoretical value within 5 ppm. (C) Tandem mass spectrum of ^{2334}UCmCCUAUCUACUmAΨCΨA^{2351}Gp. The blue arrows identify the major a, c, w, and y ions. The cleavage positions of the assigned ions are mapped on the RNA sequence in the inset.

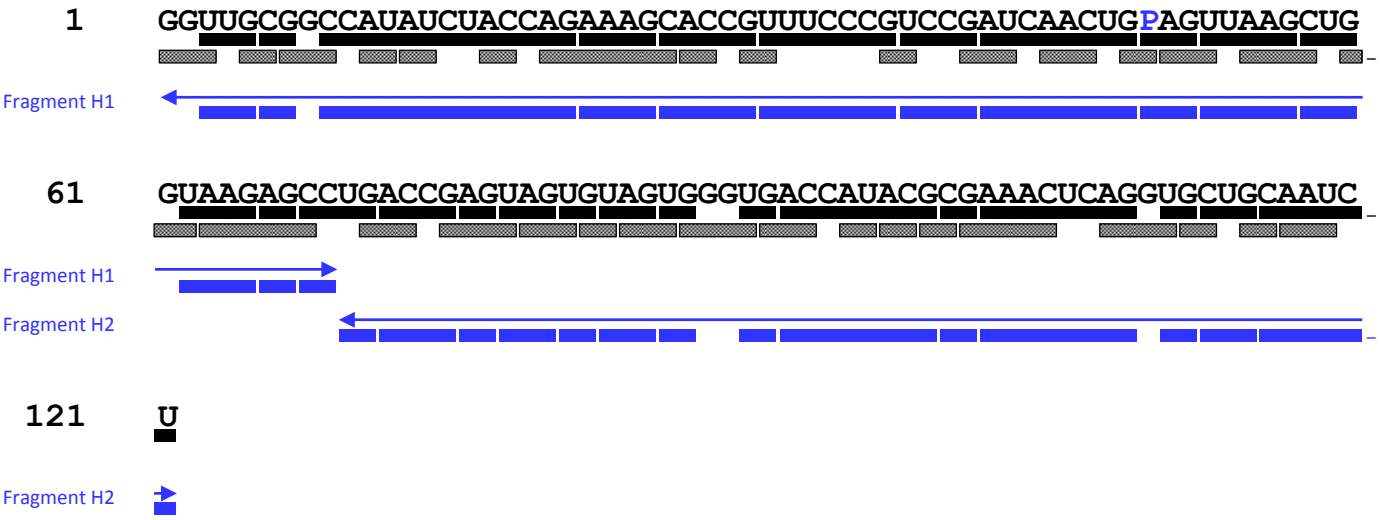
Supplementary Figure S1



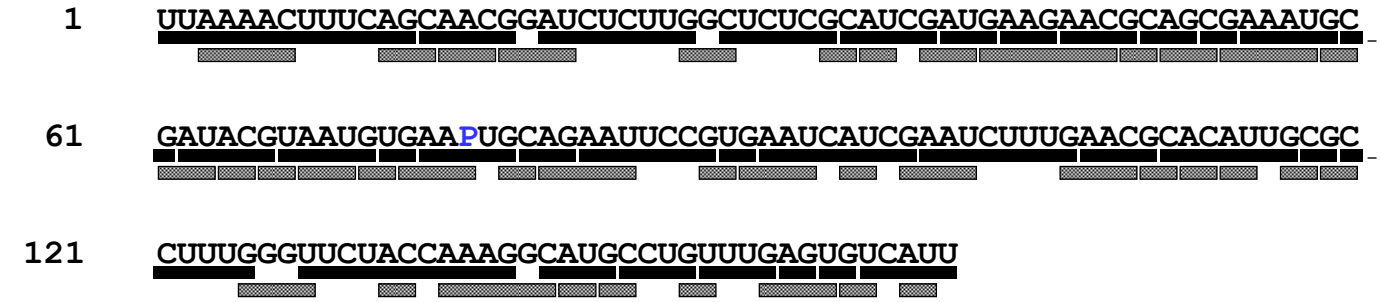
Stoichiometry of modification
$=\Delta h / h_{full}$

Supplementary Figure S2

5S rRNA

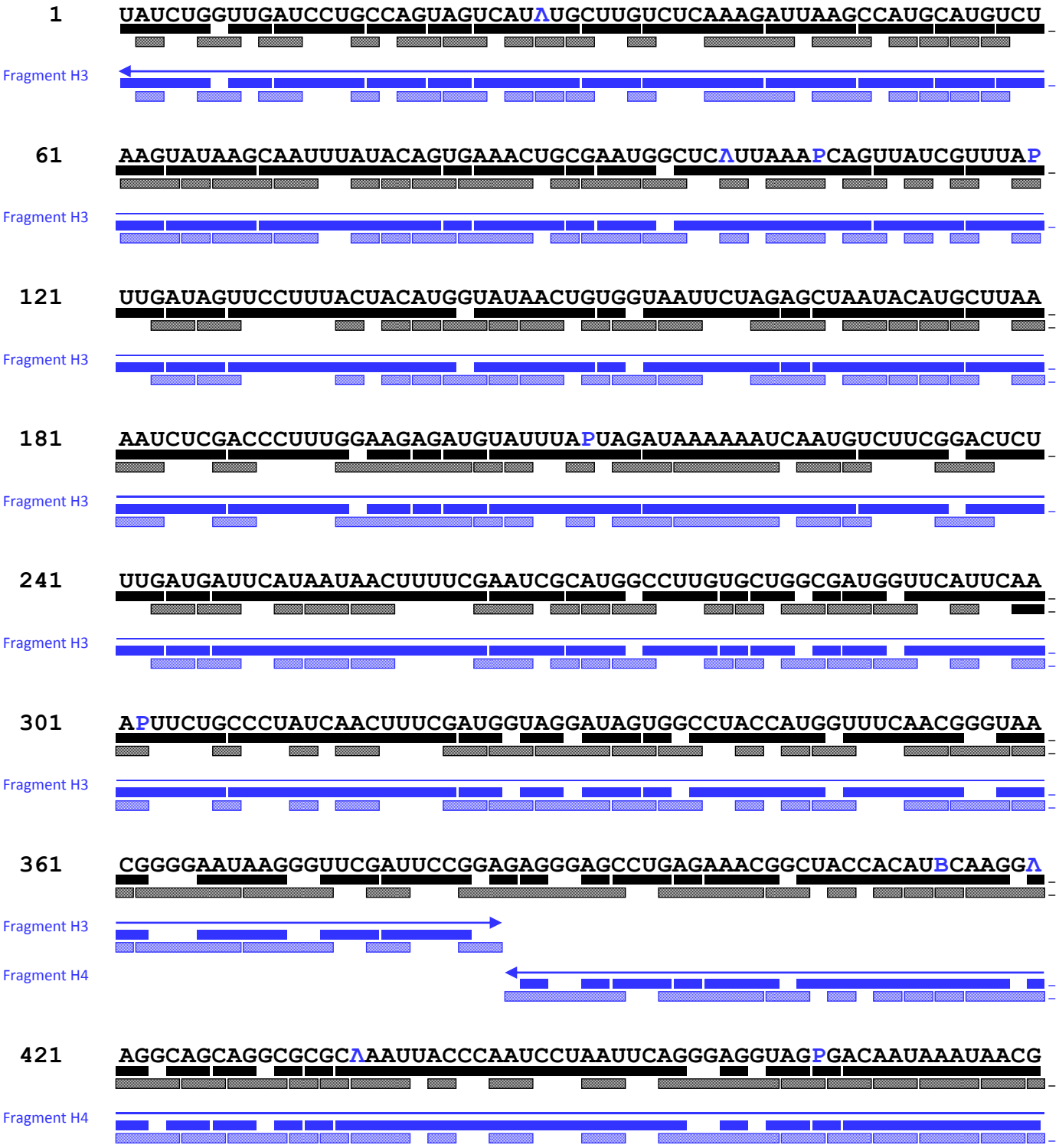


5.8S rRNA

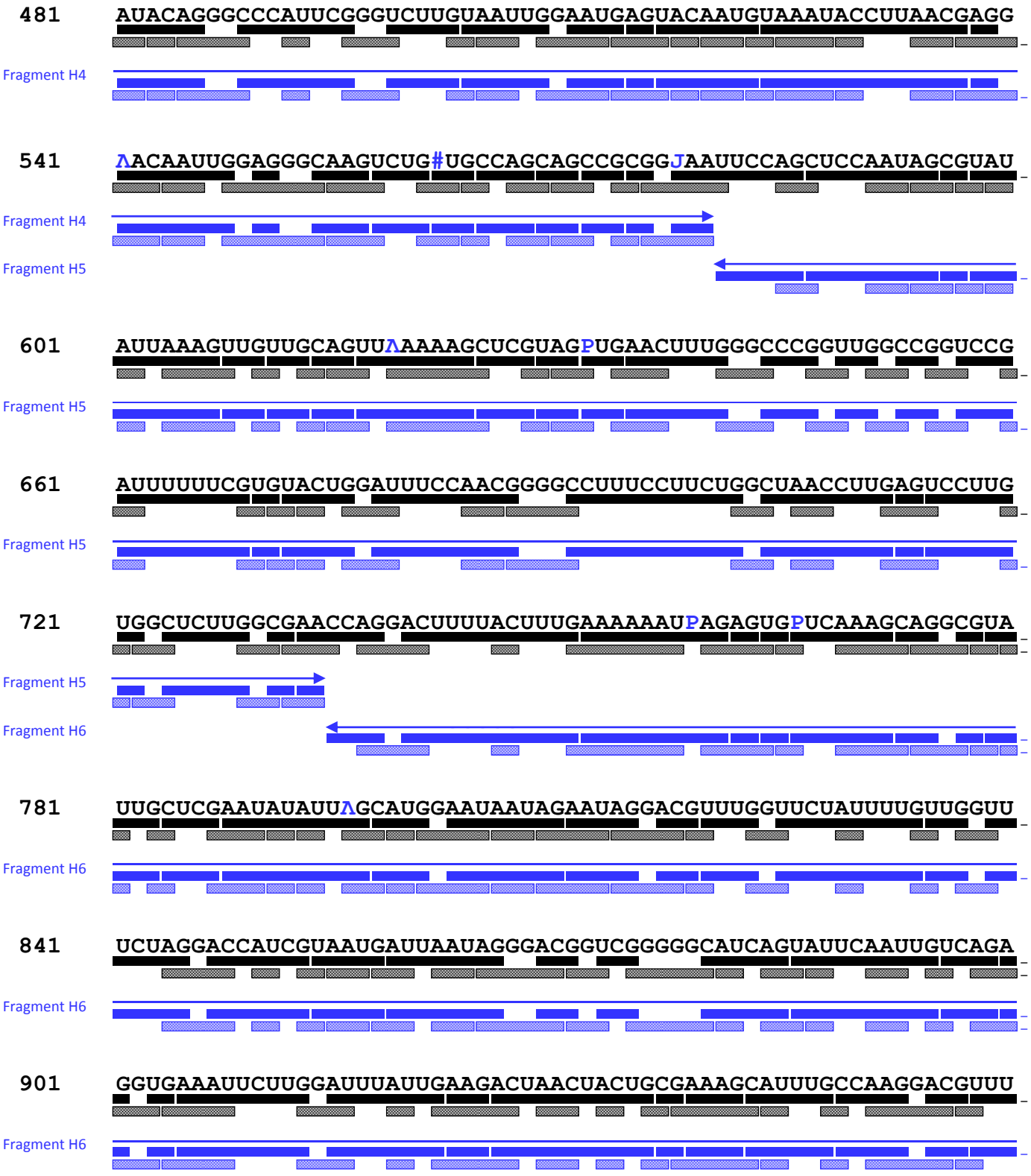


Supplementary Figure S2-*continued*

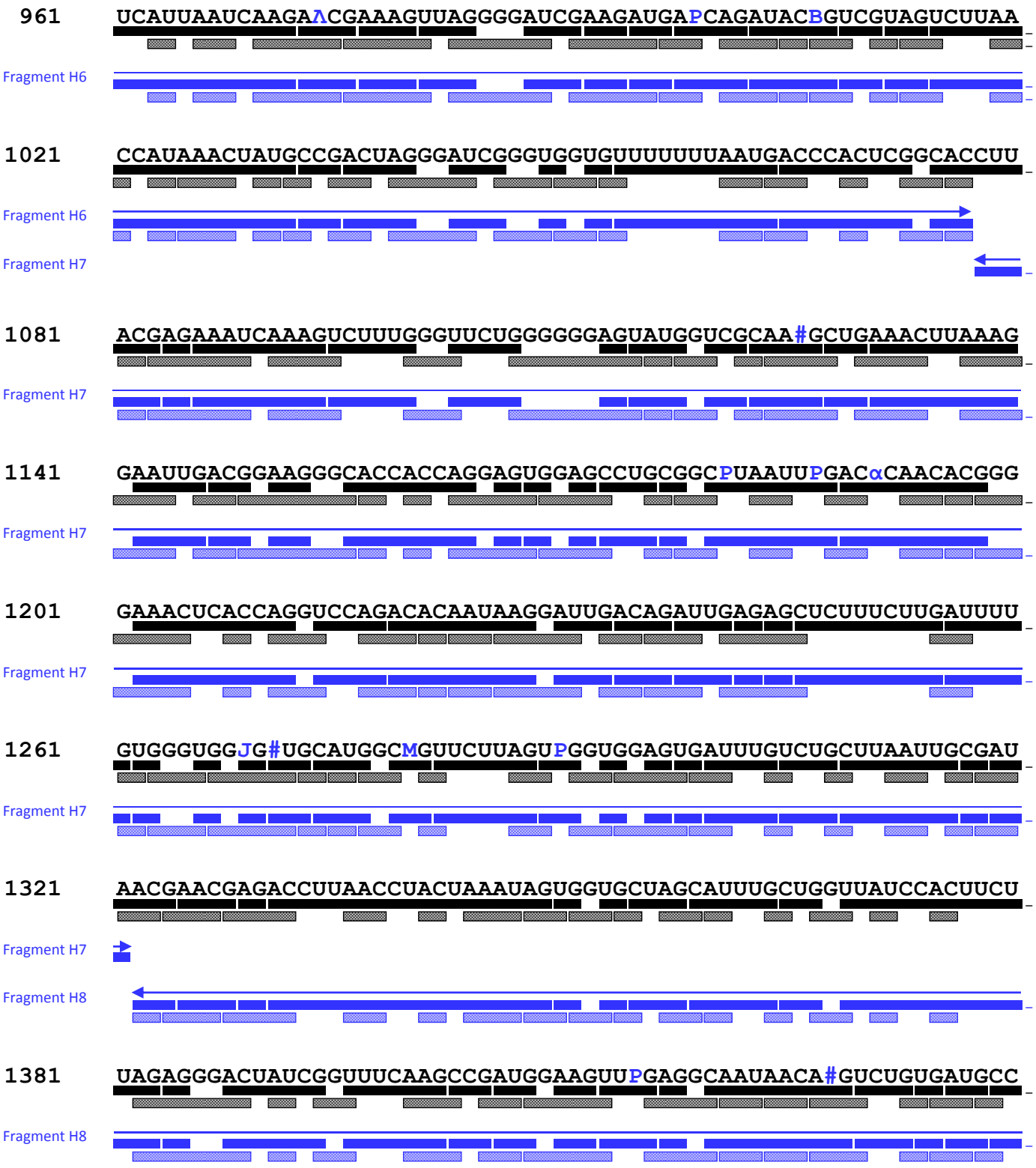
18S rRNA



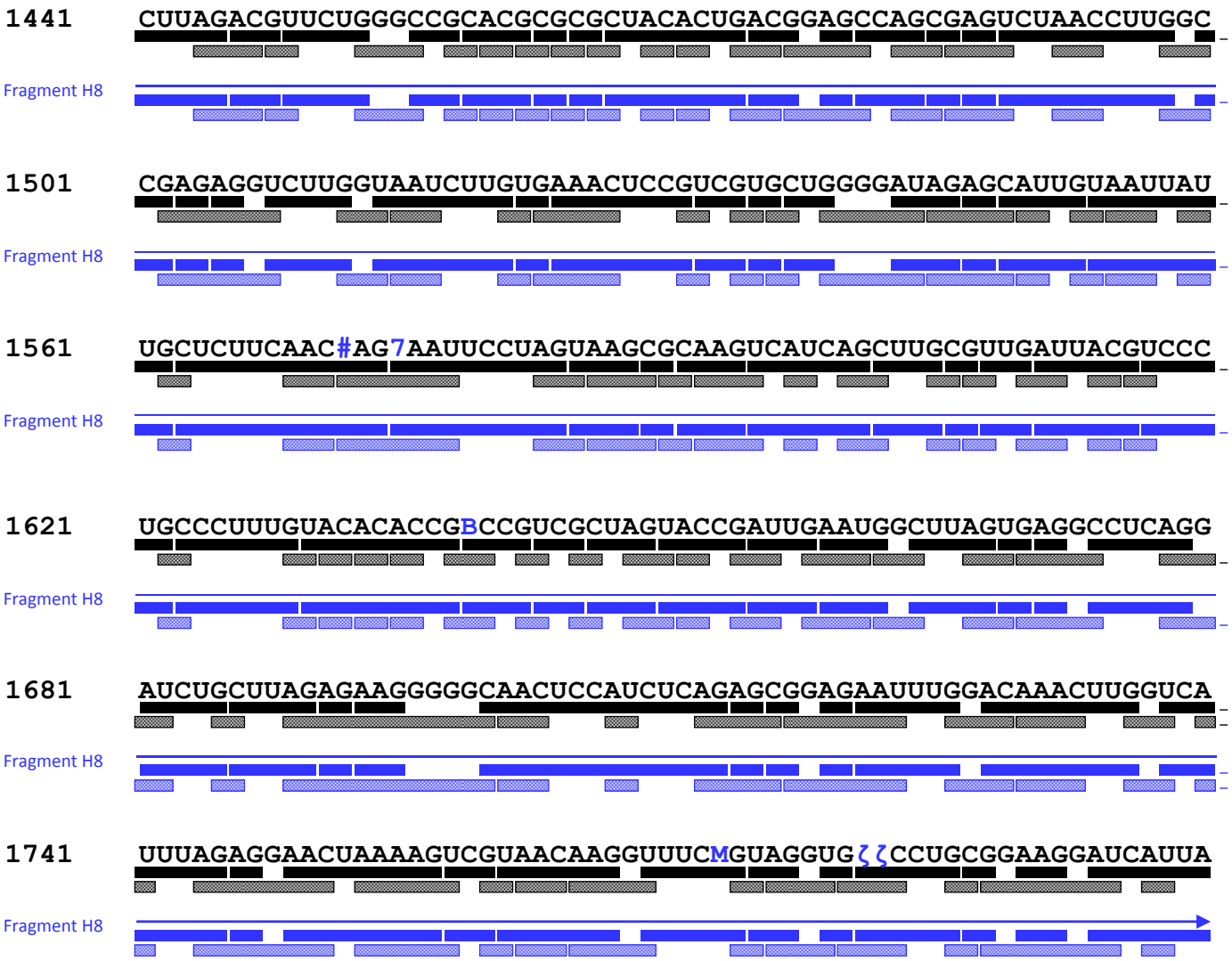
Supplementary Figure S2-*continued*



Supplementary Figure S2-*continued*

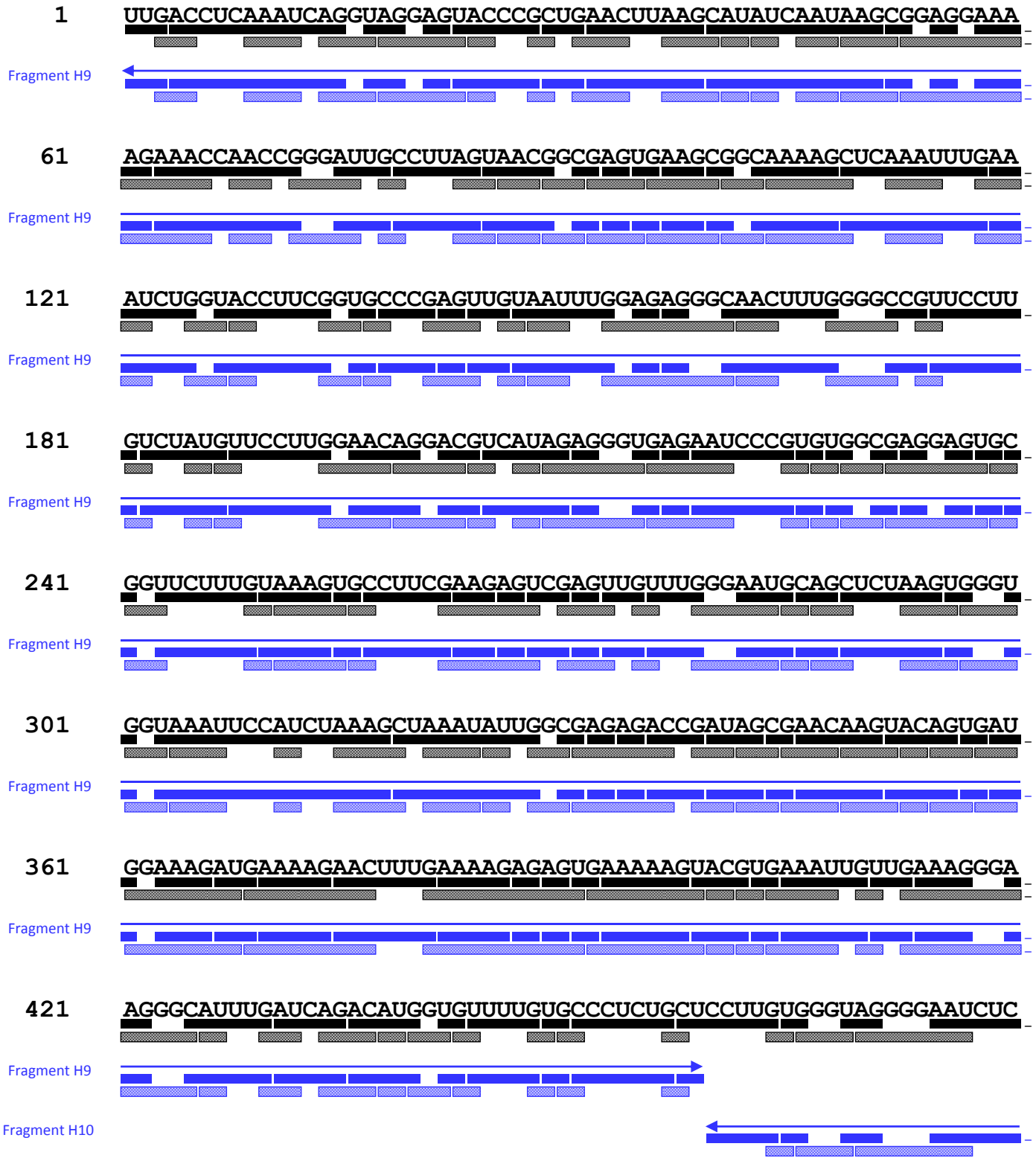


Supplementary Figure S2-*continued*

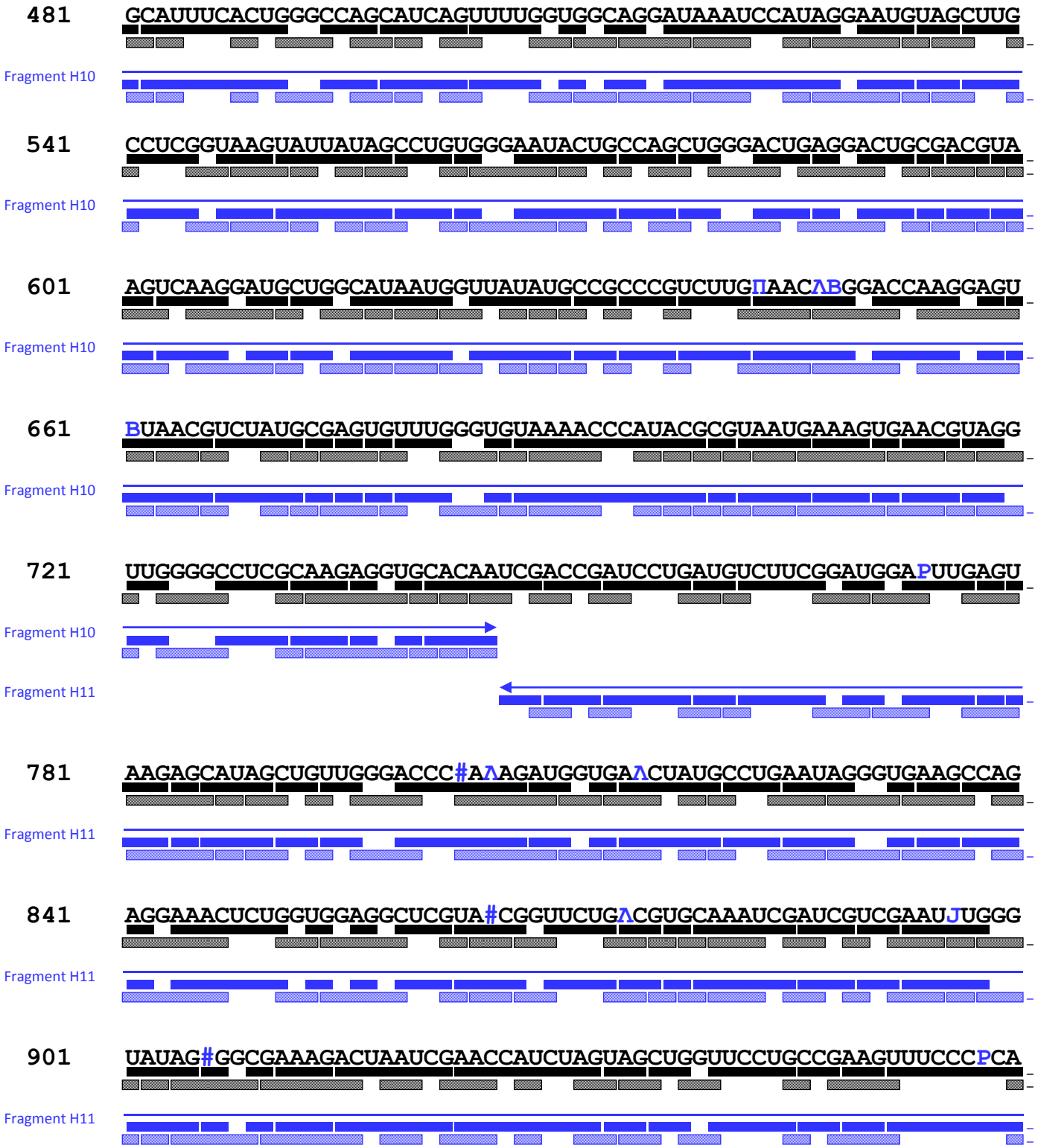


Supplementary Figure S2-*continued*

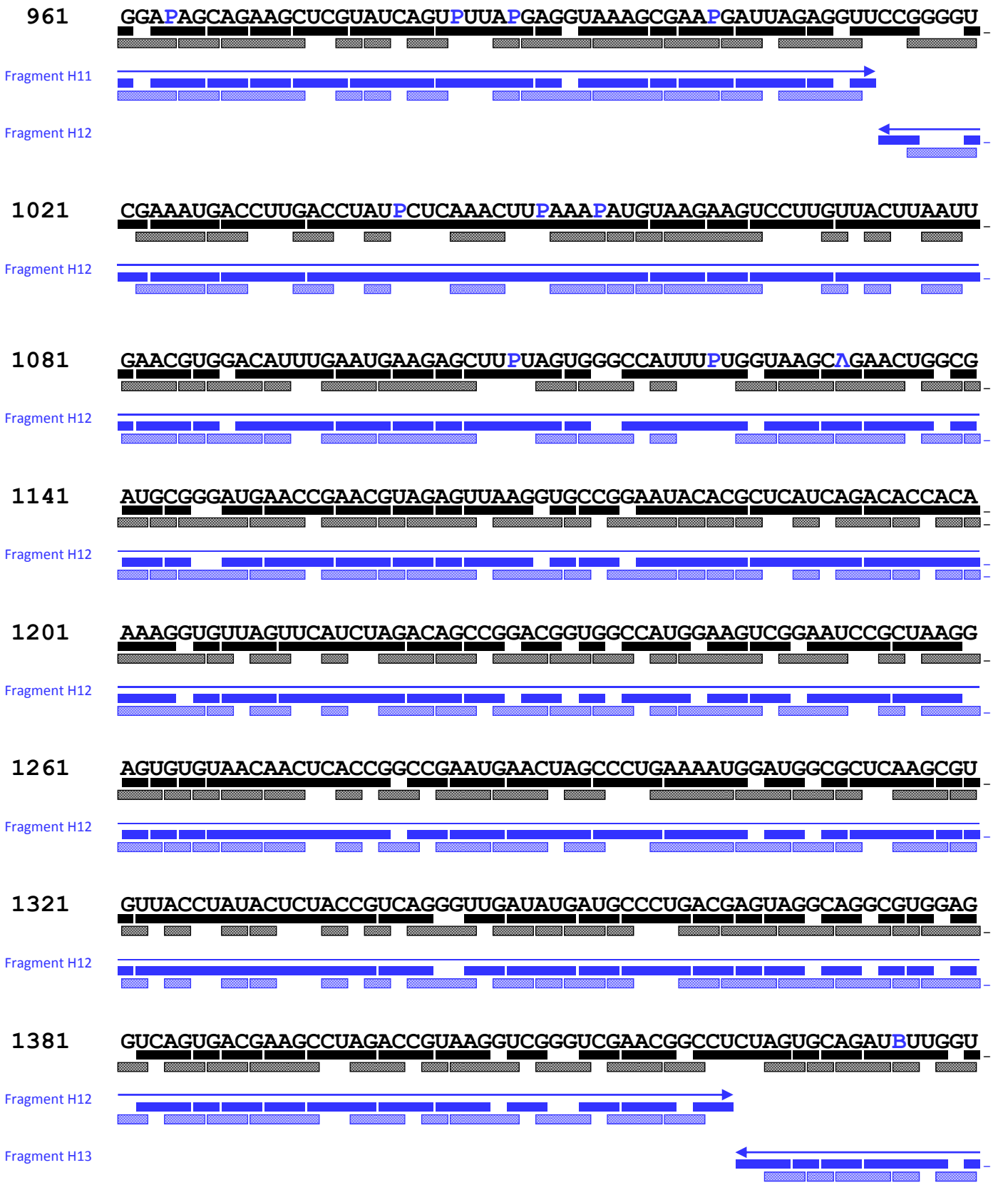
25S rRNA



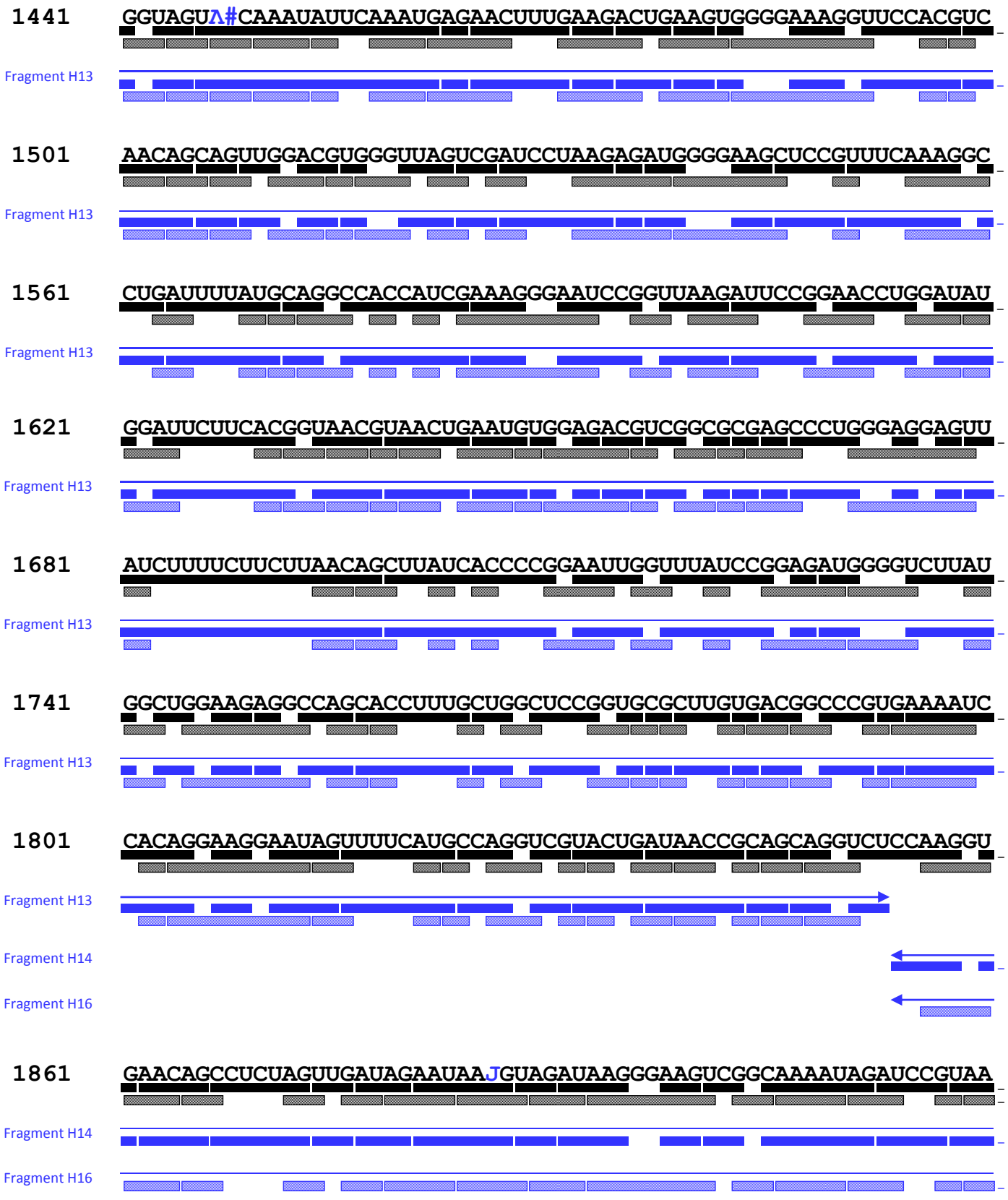
Supplementary Figure S2-*continued*



Supplementary Figure S2-*continued*



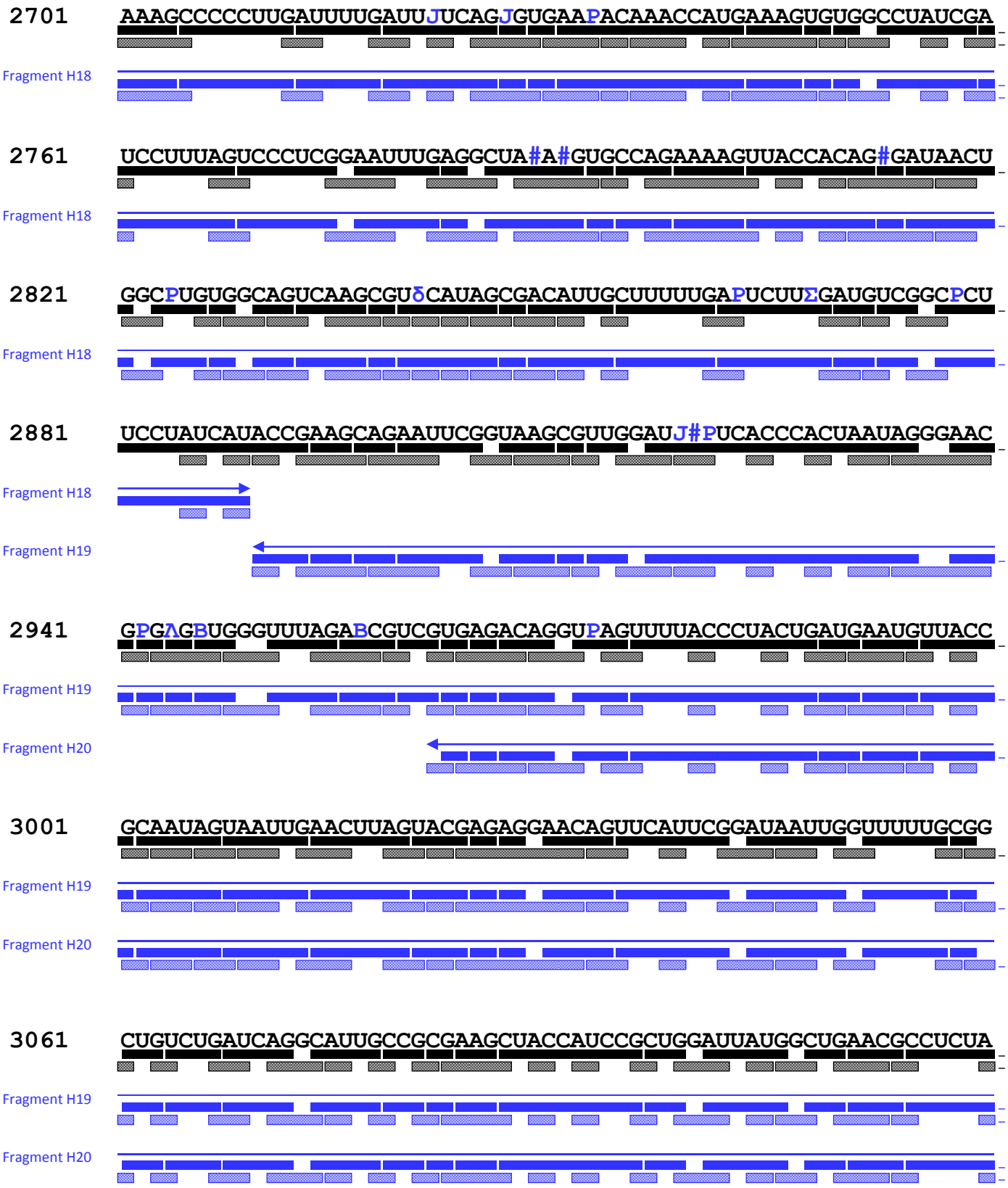
Supplementary Figure S2-*continued*



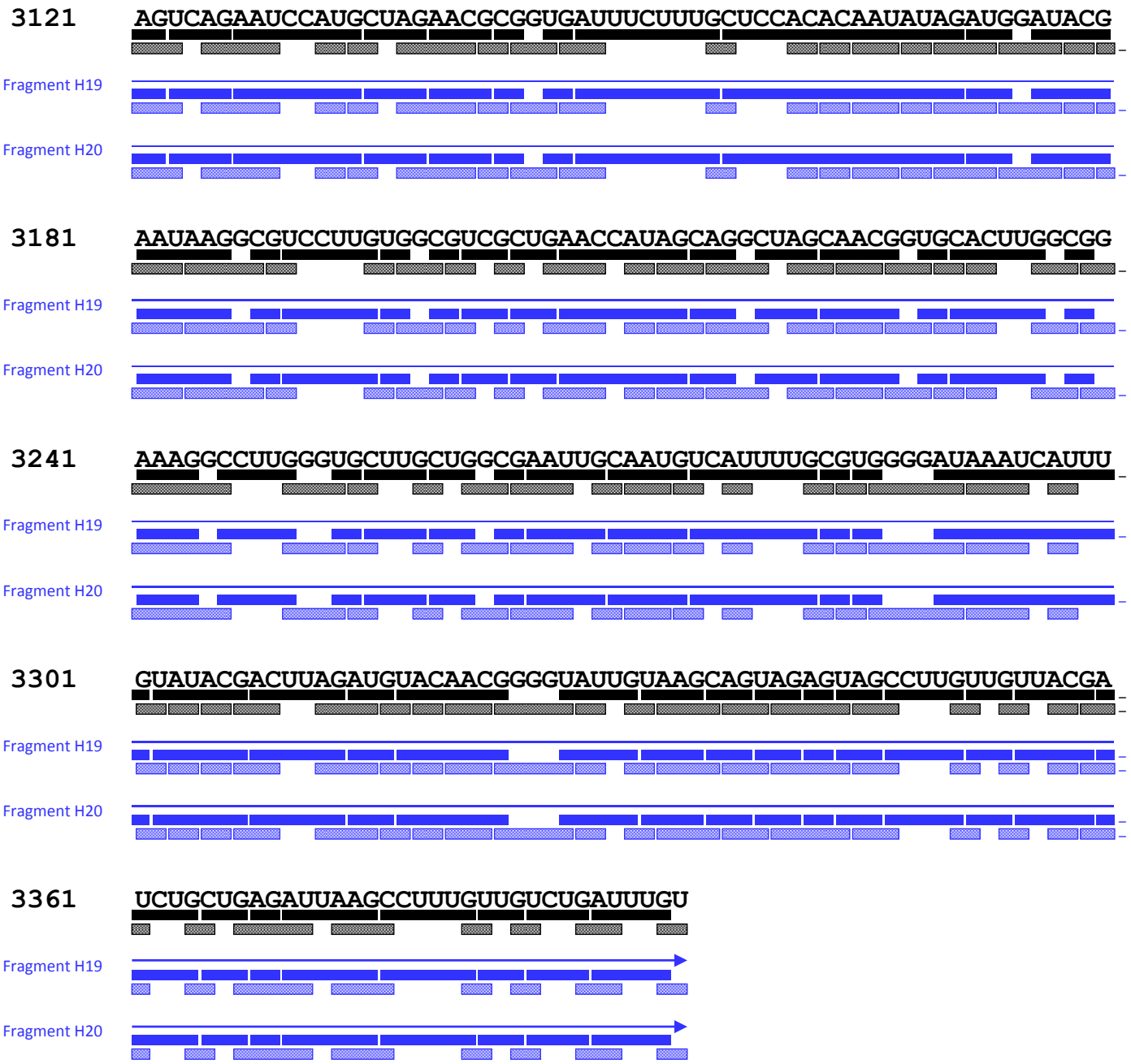
Supplementary Figure S2-*continued*

Supplementary Figure S2-*continued*

Supplementary Figure S2-*continued*

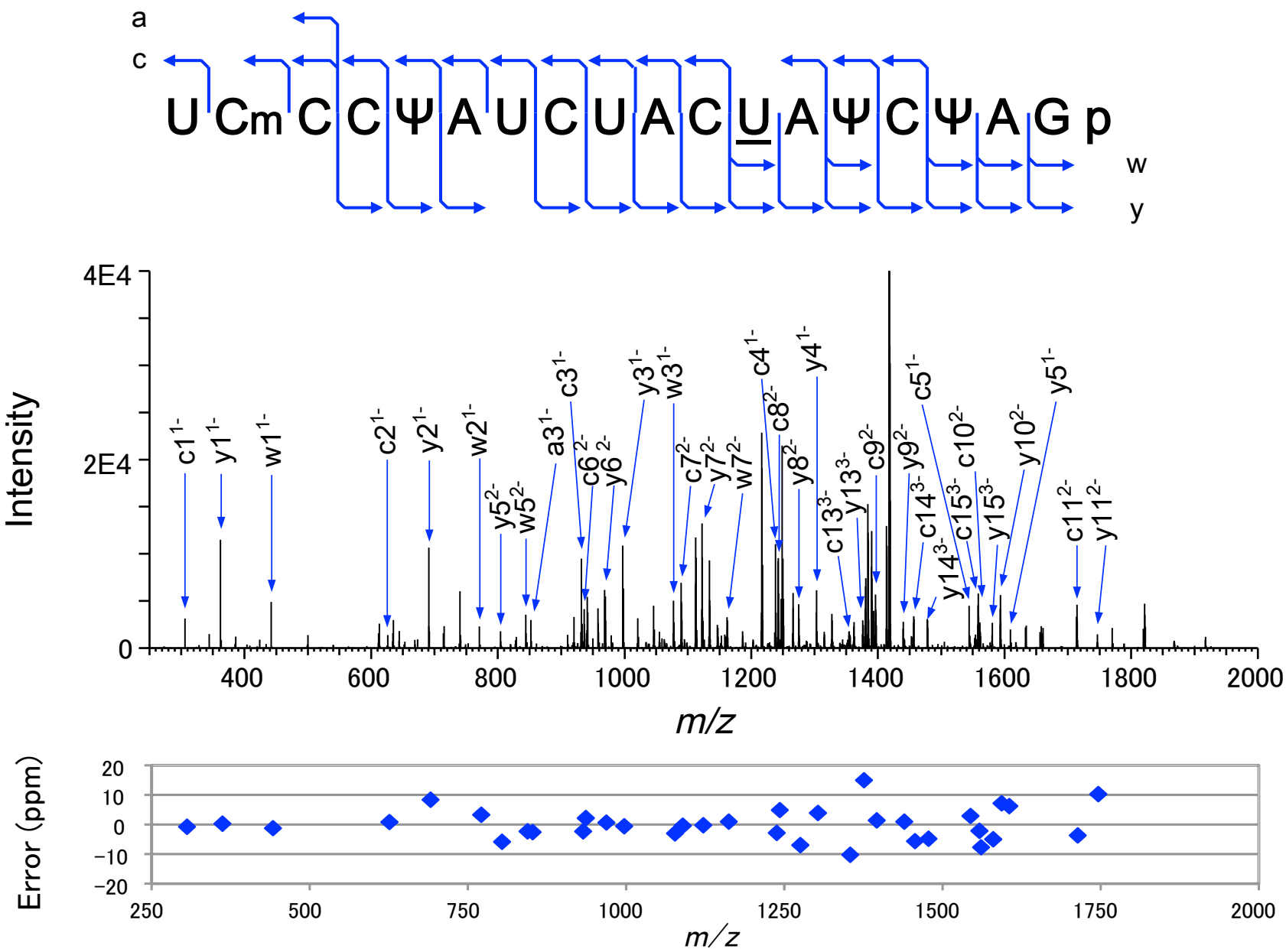


Supplementary Figure S2-*continued*

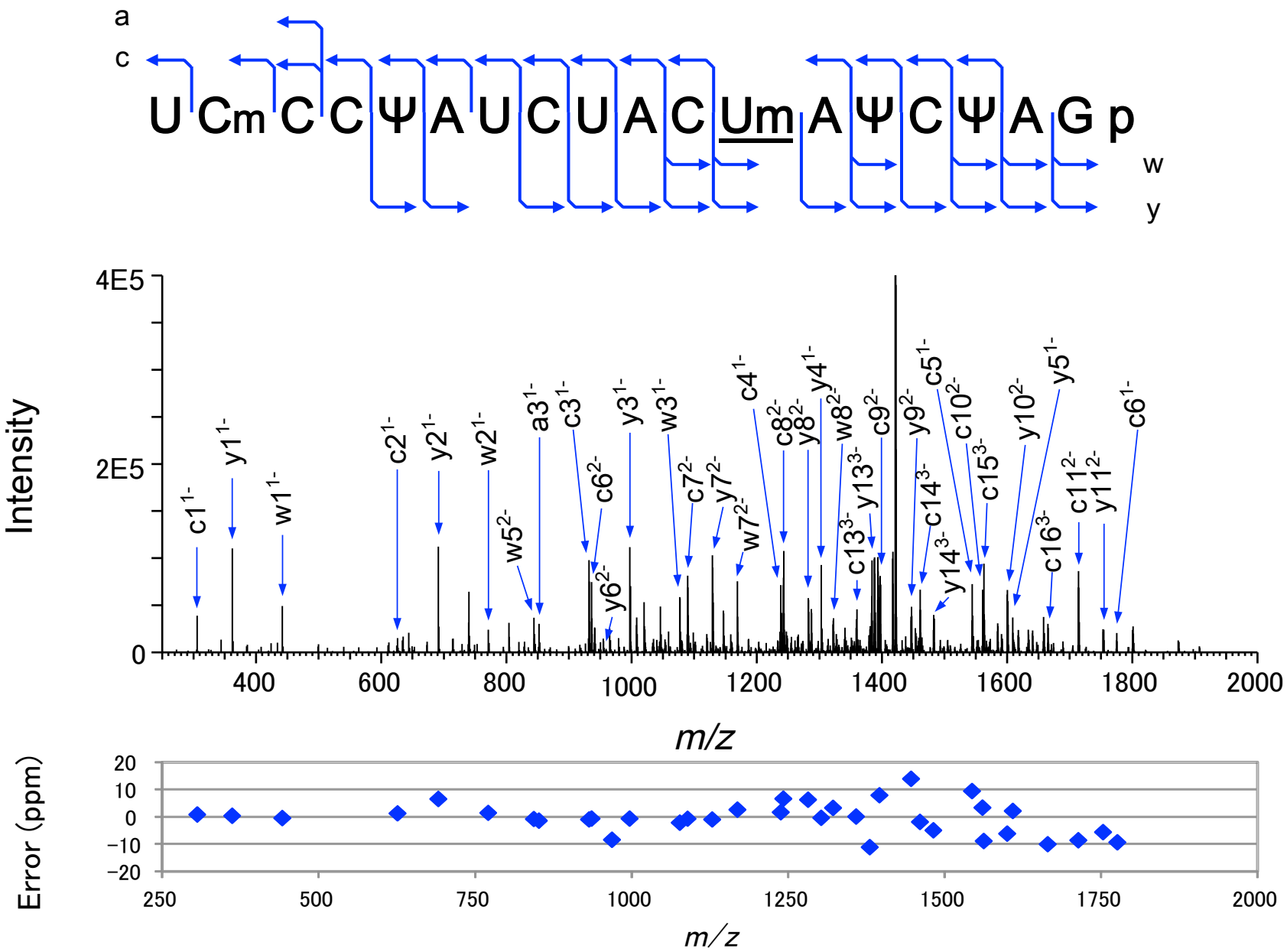


Supplementary Figure S3

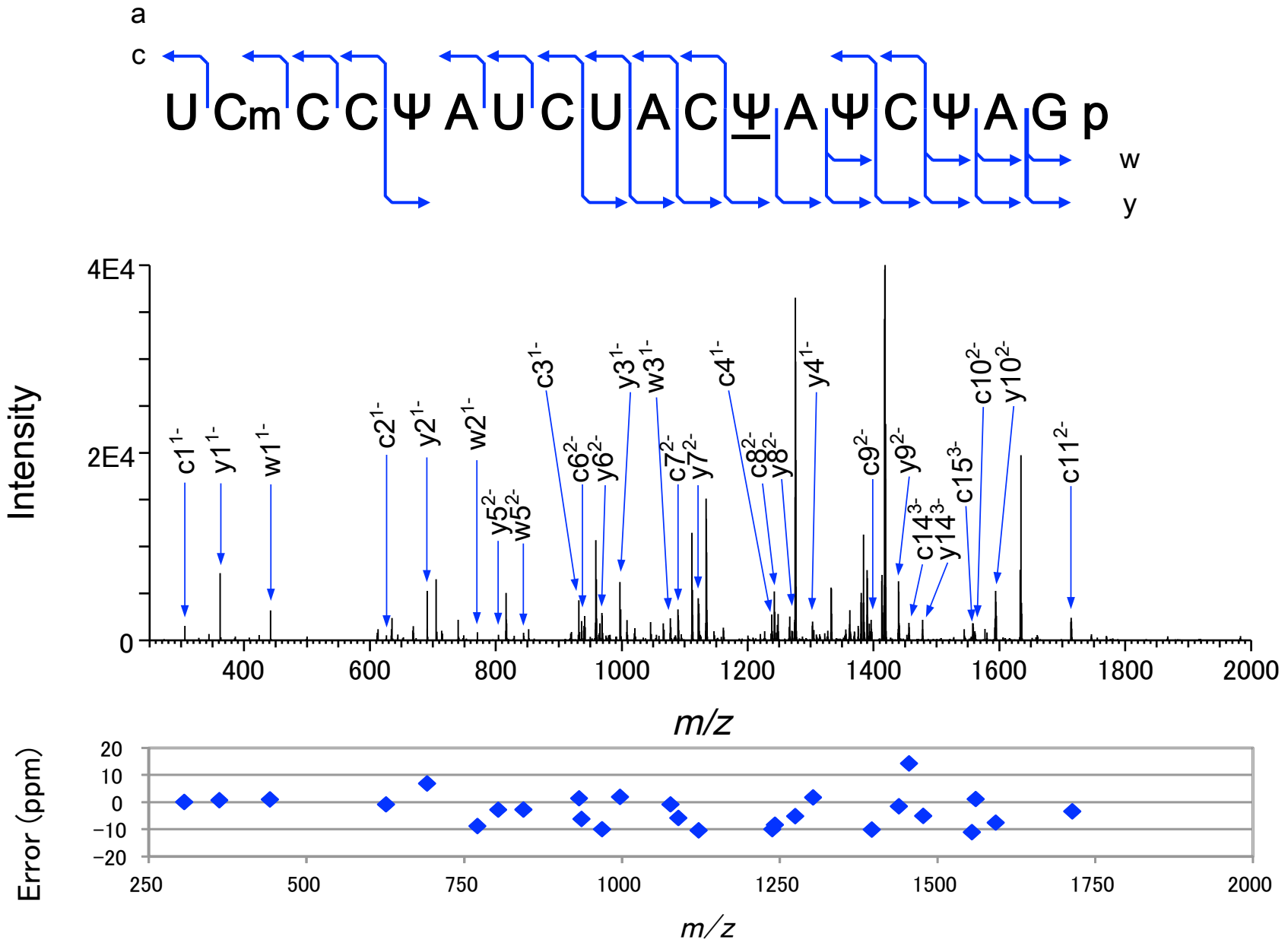
A



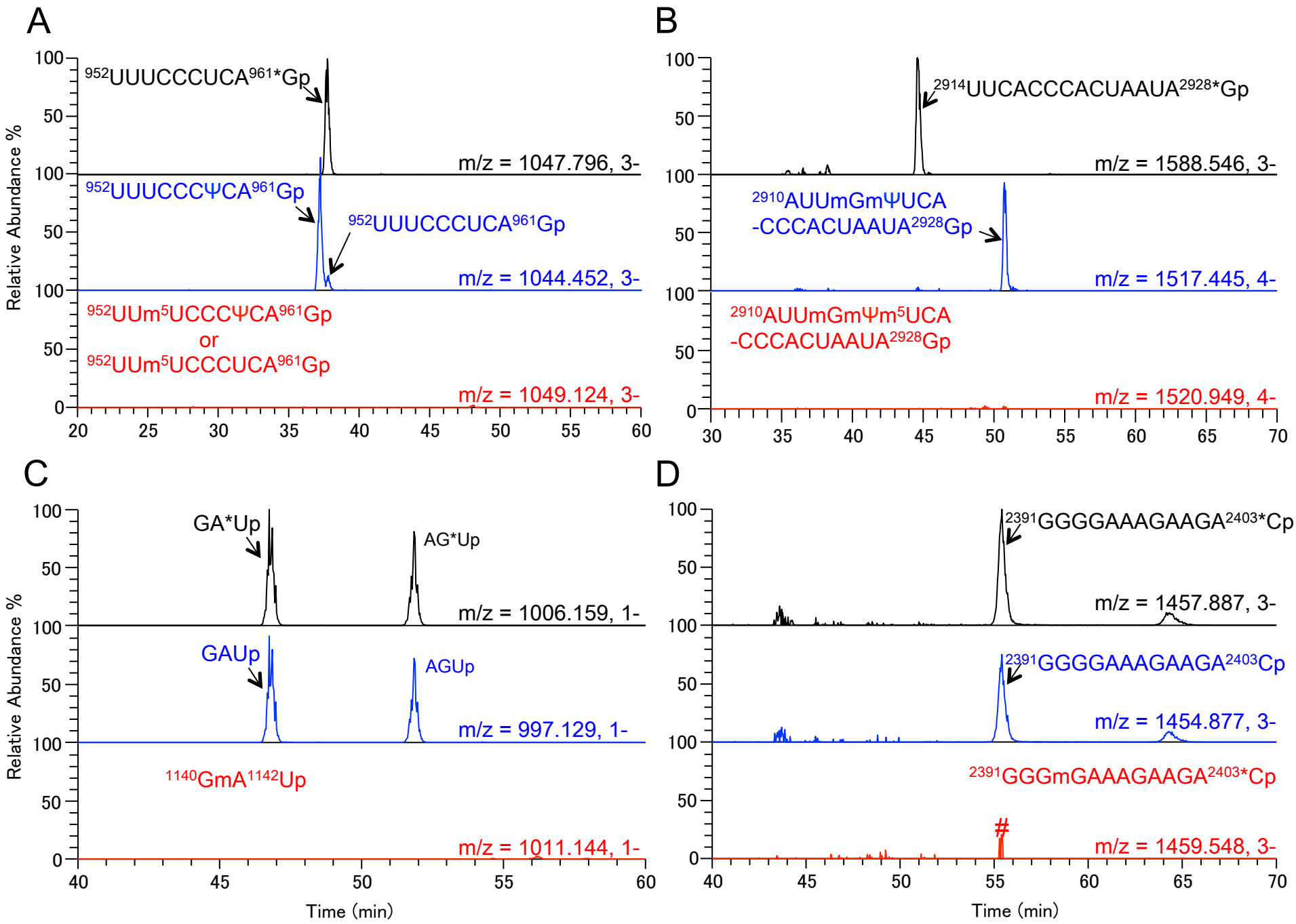
B


a
c
U Cm C C Ψ A U C U A C Um A Ψ C Ψ A G p
w
y
Intensity
4E5
2E5
0
400
600
800
1000
1200
1400
1600
1800
2000
m/z
Error (ppm)
20
10
0
−10
−20
250
500
750
1000
1250
1500
1750
2000
m/z

C

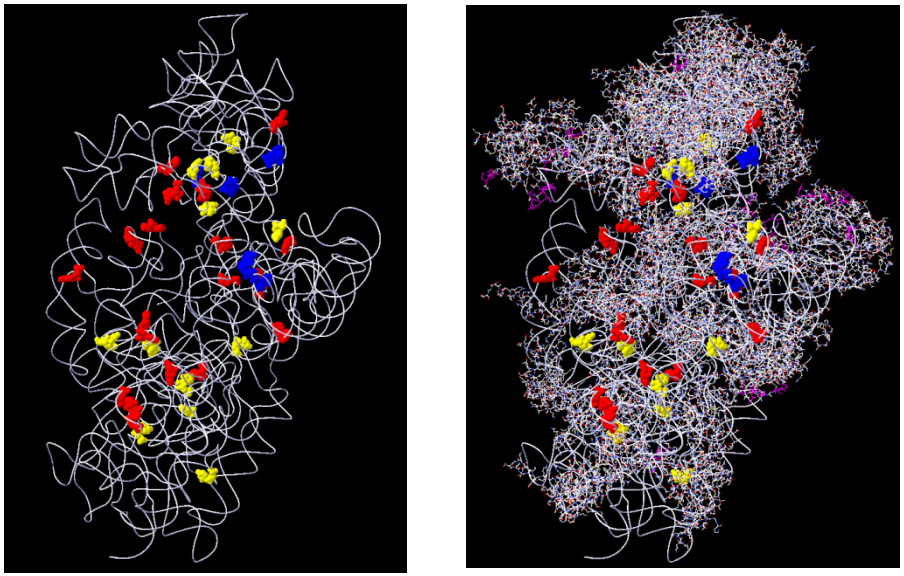

a
c
UCmCCΨAUCUACΨAΨCΨAGp
w
y
4E4
2E4
0
Intensity
c1 1-
y1 1-
w1 1-
c2 1-
y2 1-
w2 1-
y5 2-
w5 2-
c3 1-
c6 2-
y6 2-
y3 1-
w3 1-
c7 2-
y7 2-
c4 1-
c8 2-
y8 2-
y4 1-
c9 2-
y9 2-
c14 3-
y14 3-
c15 3-
c10 2-
y10 2-
c11 2-
400
600
800
1000
1200
1400
1600
1800
2000
m/z
Error (ppm)
20
10
0
−10
−20
250
500
750
1000
1250
1500
1750
2000
m/z

Supplementary Figure S4

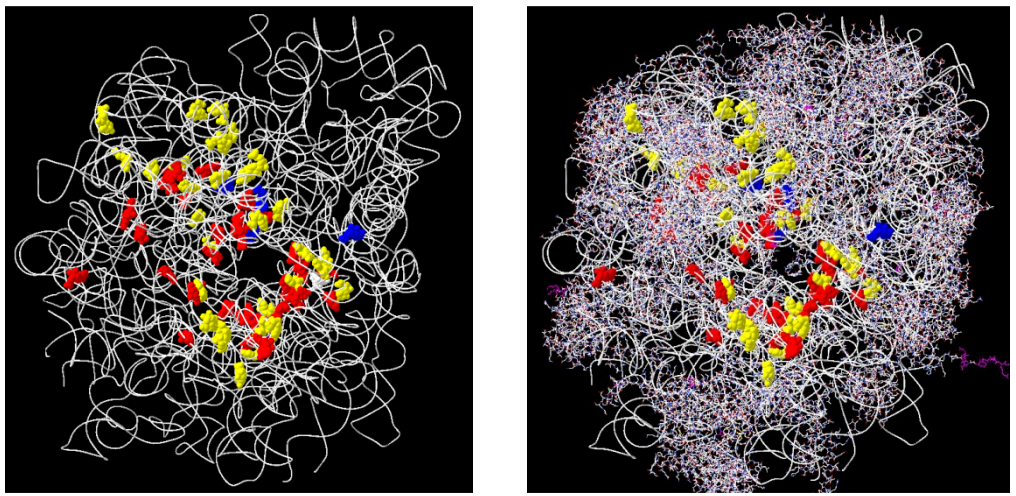


Supplementary Figure S5

A

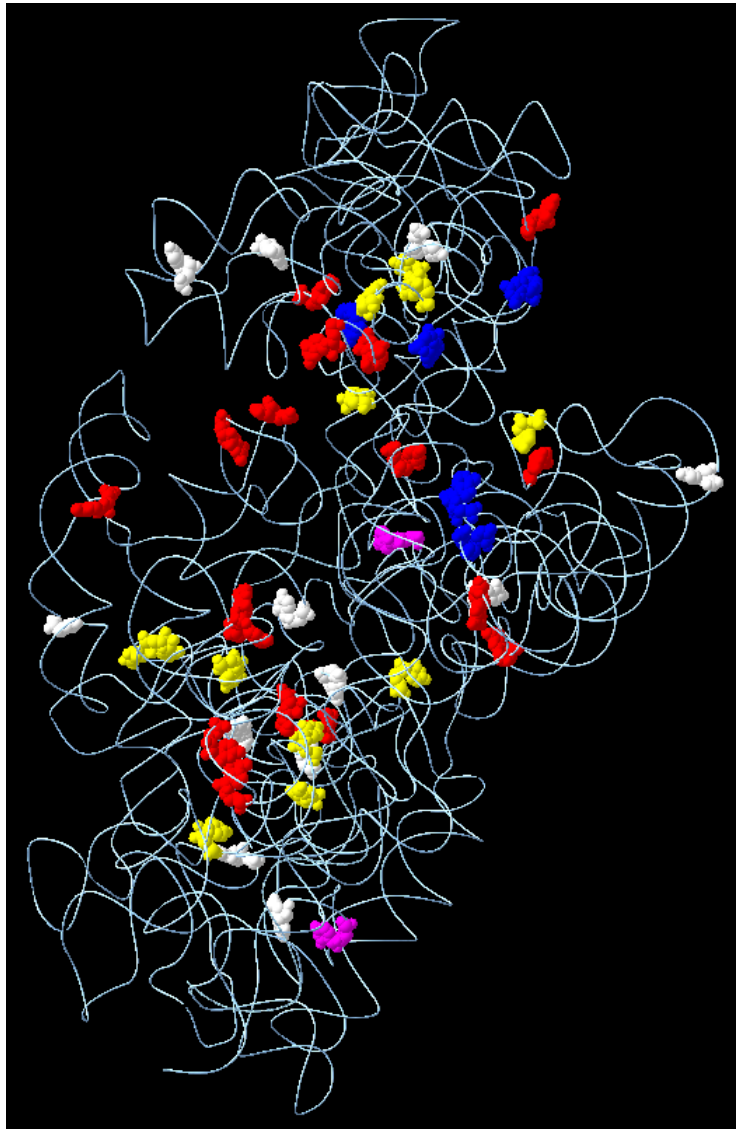


B

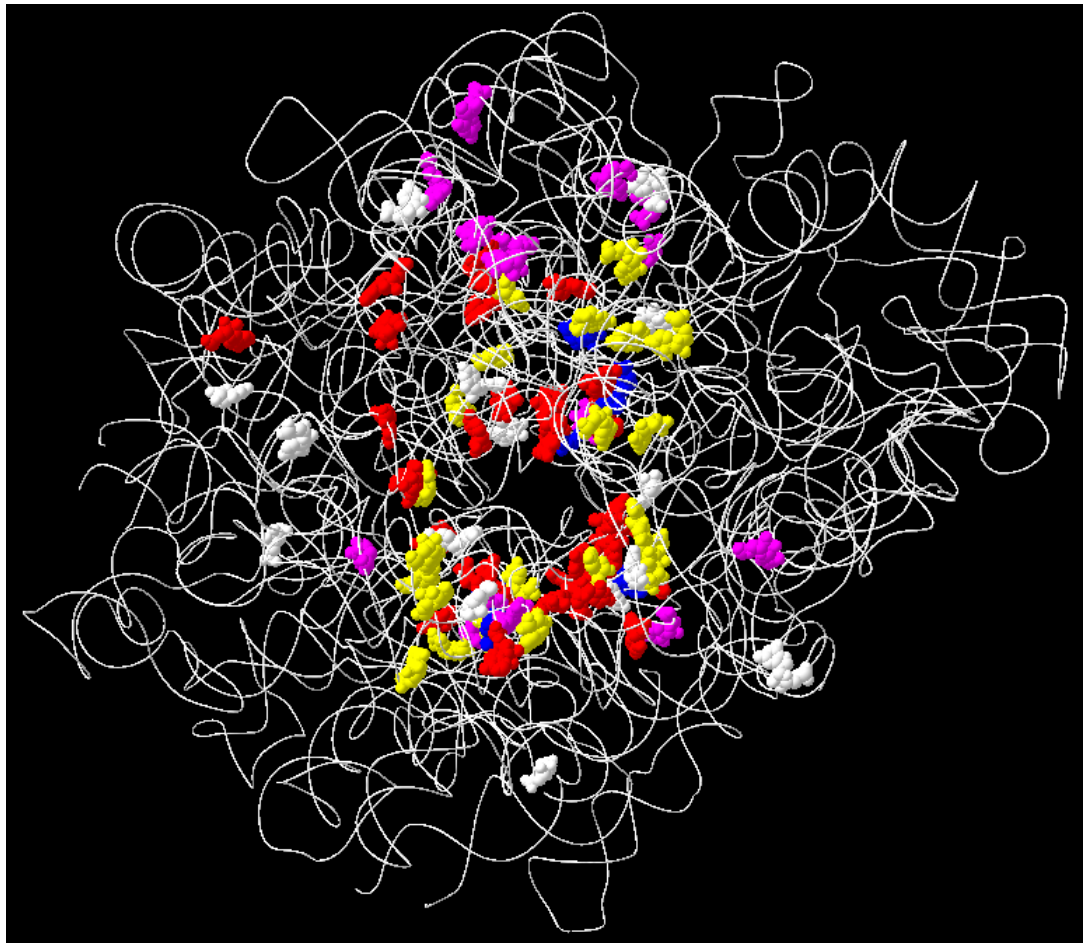


Supplementary Figure S6

A

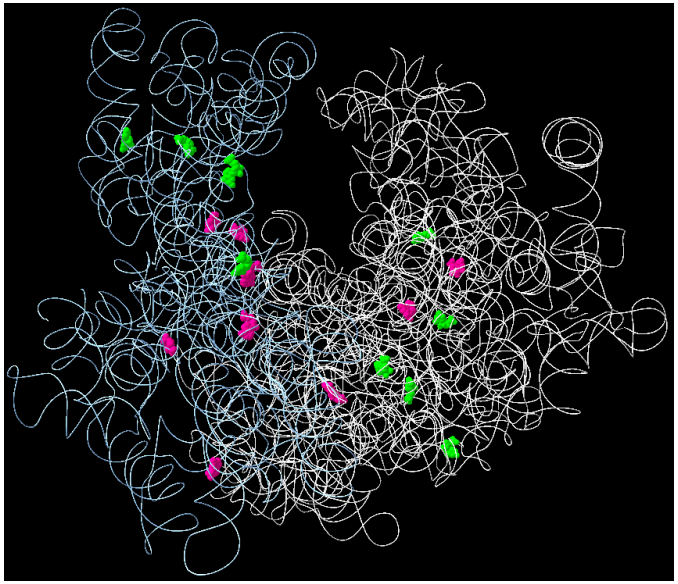


B

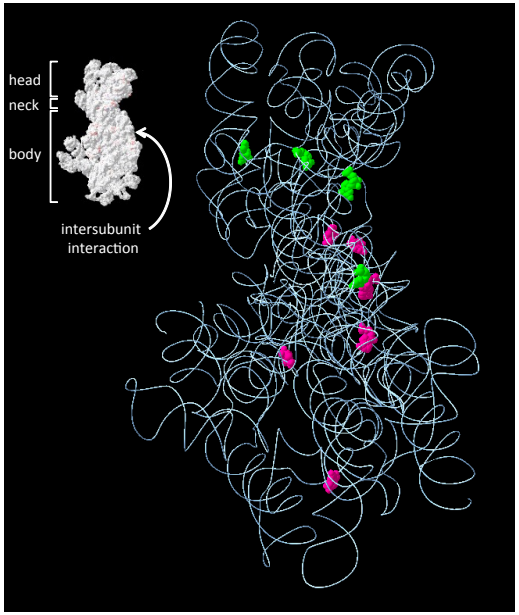


Supplementary Figure S7

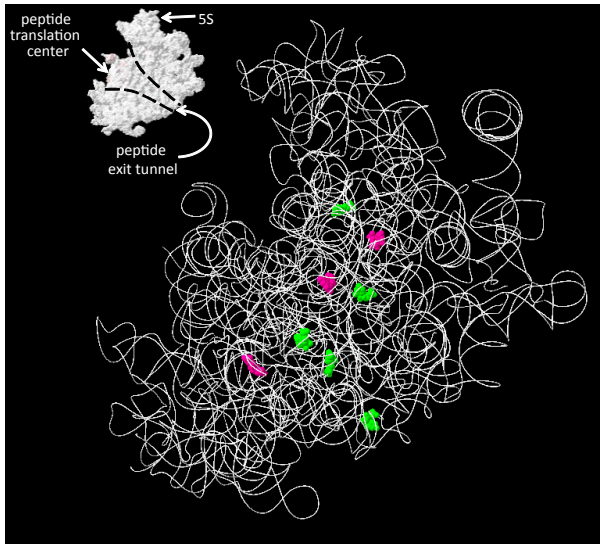
A



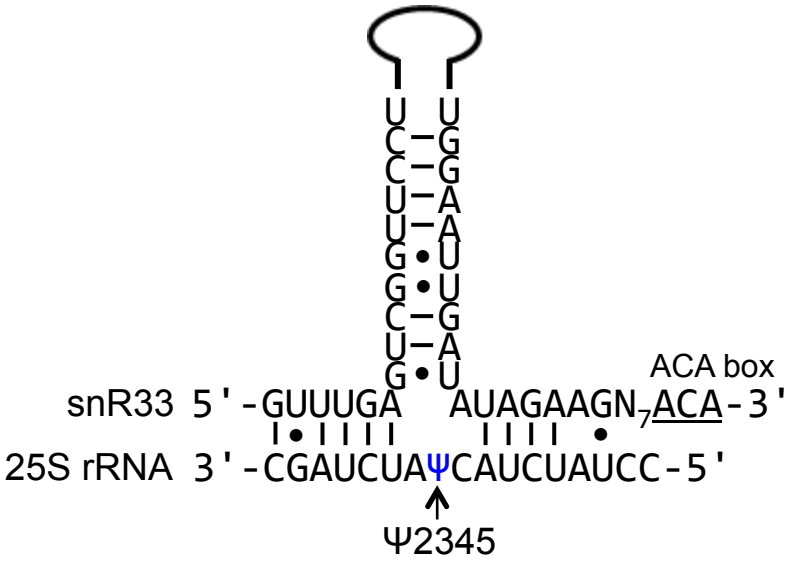
B

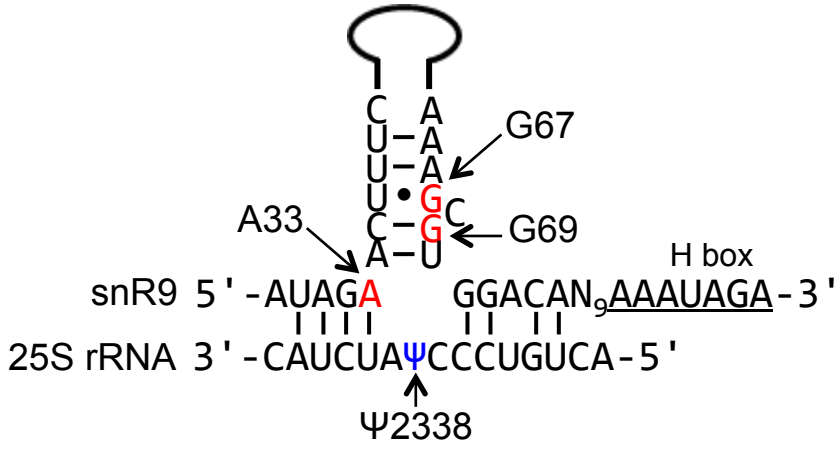


C



Supplementary Figure S8





Supplementary Figure S9