SUPREM: an engineered non-site-specific m^{6}A RNA methyltransferase with highly improved efficiency

Yoshiki Ochiai[1], Ben E. Clifton[1], Madeleine Le Coz[2], Marco Terenzio[2], Paola Laurino[1,3]*

[1] Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology Graduate University (OIST), Okinawa 904-0495, Japan.

[2] Molecular Neuroscience Unit, Okinawa Institute of Science and Technology Graduate University (OIST), Okinawa 904-0495, Japan.

[3] Institute for Protein Research, Osaka University, Osaka, 565-0871, Japan

*Corresponding author: Paola Laurino

Email: paola.laurino@oist.jp

Abstract

m^{6}A RNA methylation plays a key role in RNA processing and translational regulation, influencing both normal physiological and pathological processes. Yet, current techniques for studying RNA methylation struggle to isolate the effects of individual m^{6}A modifications. Engineering of RNA methyltransferases (RNA MTases) could enable development of improved synthetic biology tools to manipulate RNA methylation, but it is challenging due to limited understanding of structure-function relationships in RNA MTases. Herein, using ancestral sequence reconstruction we explore the sequence space of the bacterial DNA methyltransferase EcoGII (M.EcoGII), a promising target for protein engineering due to its lack of sequence specificity and its residual activity on RNA. We thereby created an efficient non-specific RNA MTase termed SUPREM, which exhibits 8-fold higher expression levels, 7 °C higher thermostability, and 12-fold greater m^{6}A RNA methylation activity compared with M.EcoGII. Immunofluorescent staining and quantitative LC/MS-MS analysis confirmed SUPREM's higher RNA methylation activity compared with M.EcoGII in mammalian cells. Additionally, Nanopore direct RNA sequencing highlighted that SUPREM is capable of methylating a larger number of RNA methylation sites than M.EcoGII. Through phylogenetic and mutational analysis, we identified a critical residue for the enhanced RNA methylation

activity of SUPREM. Collectively, our findings indicate that SUPREM holds promise as a versatile tool for *in vivo* RNA methylation and labeling.

Introduction

RNA methyltransferases (RNA MTases) are enzymes that transfer a methyl group from S-adenosyl methionine (SAM) to their RNA substrates, thus contributing to the complex biological regulation of RNAs. In eukaryotes, these enzymes have an important role in methylation of N^6-methyladenine (m^6A), which is the most prominent internal modification in eukaryotic mRNA (1). Several eukaryotic m^6A RNA MTases have been well characterized, and methylate mRNA, ribosomal RNA, and other non-coding RNA molecules at a specific site or recognition sequence motif (2–6). The m^6A modification is recognized by m^6A binding proteins, which thereby regulate RNA metabolism (4, 7, 8), pre-mRNA splicing (9–11), mRNA localization (12, 13), and translation (14–16). The m^6A modification is involved in development (17, 18) and stress response (19, 20), and is also linked to diseases such as cancer (14, 21, 22), infectious diseases (23), and type II diabetes (24). Thus, m^6A modifications are critical for both physiological and pathological processes. However, previous research on the biological functions of m^6A has largely relied on the knockout or knockdown of m^6A RNA MTases and other m^6A-associated proteins, which results in a global disruption of RNA m^6A modification states. Consequently, dissecting the biological consequences of specific m^6A modifications and distinguishing the direct and indirect effects of m^6A dysregulation has remained a challenge (25, 26). Thus, there is a need for tools that allow targeted manipulation of individual m^6A modifications in living cells to investigate their biological functions.

Few tools for site-specific m^6A methylation and demethylation have been reported previously (27–29). These tools are based on fusion constructs of the eukaryotic m^6A RNA MTase METTL3 with RNA-binding proteins, such as dead Cas proteins (dCas9 or dCas13) (27, 28) or PUF proteins (29). Dead Cas proteins can be targeted to specific sites in mRNA using guide RNA, while the RNA sequence specificity of PUF proteins is programmable by substitutions at specific residues. However, these METTL3-based tools have three important limitations. First, the range of sequences that can be methylated is restricted by the sequence specificity of METTL3, preventing the methylation of non-coding RNA and non-canonical methylation sites in mRNA. Second, since METTL3 works as part of a large methyltransferase complex (30), the isolated protein has impaired methylation efficiency (28). Third, the use of METTL3-based m^6A editing potentially introduces unwanted and confounding interactions with cellular proteins because METTL3 is an endogenous protein in eukaryotic cells. Altogether, an optimal RNA MTase for m^6A manipulation tools would be an exogenous enzyme with relaxed RNA sequence specificity that does not rely on other proteins.

The bacterial methyltransferase M.EcoGII is a promising candidate for an RNA MTase that could fulfill these requirements. M.EcoGII was identified as a DNA methyltransferase from a prophage region of a pathogenic *Escherichia coli* strain (31), and is capable of performing m^6A methylation of both DNA (~90% of adenine bases *in vitro;* ~85% *in vivo*) and RNA (~30% of adenine bases *in vitro*) (32). Unusually, M.EcoGII exhibits non-site-specific methylation activity and can potentially methylate any adenine base in DNA and RNA (31). Additionally, in contrast to mammalian methyltransferase complexes, M.EcoGII works independently of any other proteins, making it easier to engineer and orthogonal to cellular systems. Indeed, the non-specific DNA methyltransferase activity of M.EcoGII has been utilized to investigate chromatin structures in mammalian cells (33, 34).

Due to M.EcoGII's low RNA methyltransferase activity, expression level, and solubility, further protein engineering would be required to develop this enzyme as a research tool to manipulate RNA methylation states. However, protein engineering of RNA MTases, including M.EcoGII, presents several challenges. In the case of M.EcoGII, no experimental structure has been reported and the protein exhibits low homology to other protein structures, especially in the RNA recognition domain. Although structural modeling using AlphaFold2 (35) might be beneficial to infer the overall structure of the enzyme, significant conformational changes that are expected to occur upon nucleic acid binding decrease the utility of structural models for designing variants with improved function. On the other hand, screening or directed evolution of RNA MTases is also impractical due to the absence of efficient high-throughput screening techniques for measuring RNA methylation activity. Based on these limitations, we hypothesized that sequence-based engineering tools such as ancestral sequence reconstruction (ASR) would be most effective for improving the properties of M.EcoGII. ASR uses the sequences of extant proteins to statistically infer the sequences of extinct ancestral proteins, which often show higher thermostability and solubility and expanded substrate specificity compared to existing proteins (36–39). Although the underlying reasons for the improved properties of reconstructed ancestral proteins are not entirely clear (40, 41), ASR has been successfully applied to improve the properties of various proteins, including those of relatively recent evolutionary origin (42, 43). As a protein engineering technique, ASR allows for the sampling of sequence space enriched in proteins with favorable properties based on phylogenetic information (40).

Here, we used ASR to engineer ancestral M.EcoGII variants with higher expression, thermostability, and RNA methylation activity. We reconstructed five ancestral M.EcoGII variants, and one of these variants, which we named SUPer RNA EcoGII Methyltransferase

101 (SUPREM), showed a 12-fold increase in RNA methylation activity and an 8-fold increase in
102 expression compared with M.EcoGII. SUPREM activity was confirmed in an *in vivo* context
103 *via* transfection in mammalian cells, suggesting a potential use for this enzyme in biological
104 investigations. Nanopore direct RNA sequencing analysis revealed that SUPREM had less
105 biased methylation on RNA than M.EcoGII. We also investigated the amino acid residues
106 responsible for the improved RNA methylation activity of SUPREM by site-directed
107 mutagenesis. Together, our engineered M.EcoGII variant has great potential for developing
108 novel RNA biology tools to manipulate RNA methylation.
109

110 # Results

111 **Ancestral sequence reconstruction of M.EcoGII leads to variants with higher stability**

112 To reconstruct the sequences of ancestral M.EcoGII variants, we first collected a non-
113 redundant set of homologous protein sequences of β-class methyltransferases, including
114 M.EcoGII, from the Restriction Enzyme Database (REBASE) (44). A sequence similarity
115 network of the resulting protein sequences showed that M.EcoGII and highly similar sequences
116 clustered most closely with the site-specific DNA methyltransferase, M.PflMI (27.43%
117 identity), followed by two sequence clusters exemplified by M.McrMPORF631P (21.63%
118 identity) and M.Blo16BORF1919P (20.70% identity) (Supplementary Figure S1). For
119 subsequent phylogenetic analysis, we focused on the 202 sequences that clustered together with
120 these four sequences in the sequence similarity network. Analysis using the PHAST database
121 showed that all genes of these proteins are encoded in phage-derivative regions of their
122 respective host genomes (45).

123 A maximum-likelihood tree was constructed from these 202 sequences (Figure 1A:
124 schematic tree, Supplementary Figure S2: full tree). Because M.PflMI is the most closely
125 related enzyme to M.EcoGII that has confirmed site-specific methylation activity (recognition
126 sequence: CCANNNNNTGG), we postulated that non-site-specific methylation activity
127 emerged between M.EcoGII and the last common ancestor of M.EcoGII and M.PflMI, and that
128 ancestral proteins between these two nodes might retain the desired non-site-specific
129 methylation activity. We therefore selected six ancestral nodes, evenly spaced between
130 M.EcoGII and the last common ancestor of M.EcoGII and M.PflMI, for experimental
131 characterization (Figure 1A). The selected nodes were robustly supported by ultrafast bootstrap
132 values (>90%) and the Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT)

133 (>80%), and the corresponding ancestral sequences were also reconstructed with high
134 statistical confidence, with mean posterior probability values >90% (Supplementary Table S1).
135 These ancestral M.EcoGII sequences were cloned into the pET-28a(+) vector and
136 expressed in the *E. coli* SHuffle T7 strain. Five of the six ancestral proteins could be expressed
137 and purified. The ancestral proteins were expressed at higher levels than M.EcoGII (6-fold to
138 13-fold; Supplementary Figure S3A, Supplementary Table S1). In addition, the ancestral
139 proteins showed higher thermostability than M.EcoGII, as measured by differential scanning
140 fluorimetry. Whereas M.EcoGII exhibited a denaturation temperature (T_M) of 43 °C and would
141 therefore have marginal stability at physiological temperatures, the ancestral proteins showed
142 increases in T_M of 7–16 °C compared with M.EcoGII (Supplementary Figure S3B-C). Thus,
143 ASR succeeded in improving the expression level and thermostability of M.EcoGII.
144
145 ***In vitro* activity assays of ancestral M.EcoGII variants show improved catalytic efficiency**
146 **for RNA methylation**
147 To investigate the DNA and RNA methylation activity of M.EcoGII and the ancestral
148 proteins, we used the MTase-Glo assay (46), a luminescence-based assay that can be used to
149 quantify *S*-adenosyl homocysteine (SAH), a byproduct of the SAM-dependent methylation
150 reaction. PCR-amplified DNA (3569 bp, 1819 adenine bases) or *in vitro* transcribed RNA
151 (1872 nt, 544 adenine bases) was used as a substrate for *in vitro* methylation reactions. All the
152 ancestral proteins showed DNA and RNA methylation activity. There was limited variation in
153 the DNA methyltransferase activities of the ancestral proteins, with two of them, Anc291 and
154 Anc289, showing a small but significant increase in DNA methylation activity compared with
155 M.EcoGII (Anc291, 2.4-fold, P = 0.0099; Anc289, 2.4-fold, P = 0.0139) (Figure 1B).
156 Surprisingly, a much larger variation in RNA methyltransferase activity was observed, with
157 one ancestral protein, Anc284, showing a 12-fold increase in RNA methylation activity
158 compared with M.EcoGII (P < 0.0001; Figure 1C). Time-course DNA and RNA methylation
159 assays confirmed that the product formation was linear over the course of these experiments
160 and that Anc291 and Anc284 showed higher initial velocity for RNA methylation than
161 M.EcoGII (Figure 1D, Supplementary Figure S4, Supplementary Table S3).
162 To detect and quantify the m^6A modification directly, we next performed LC-MS/MS
163 analysis. A short RNA substrate (294 nt) was methylated with M.EcoGII or Anc284, digested
164 to nucleosides, and analyzed by LC-MS/MS. Treatment of RNA with M.EcoGII and Anc284
165 resulted in the formation of an additional peak at 5.8 min corresponding to m^6A, as shown by
166 comparison of peak elution volume with an m^6A standard, detection of the m^6A $[M+H]^+$ ion,

167 and analysis of the fragment ion spectrum (Supplementary Figure S6-8), indicating the
168 presence of RNA m^6A methylation activity in Anc284. Observation of a major fragment at *m/z*
169 150, corresponding to loss of the ribose moiety from m^6A, provided further evidence that
170 methylation occurs on the adenine ring rather than the ribose (Supplementary Figure S8).
171 Spiking experiments were also performed to provide further evidence that the peak at 5.8 min
172 corresponds to m^6A (Supplementary Figure S9). These experiments conclusively establish that
173 the product of methylation by Anc284 is m^6A, consistent with the fact that M.EcoGII has
174 already been established to be a m^6A methyltransferase (31, 32) and that Anc284 was
175 engineered based on sequence data of M.EcoGII and other m^6A methyltransferases. The
176 fraction of m^6A-methylated adenosine determined by peak area integration was significantly
177 higher for Anc284, compared with M.EcoGII (32.5 ± 6.1% vs. 69.3 ± 0.7%; $P = 0.0006$ by
178 unpaired t-test). Altogether, these results show that Anc284 has substantially higher RNA
179 methylation activity than M.EcoGII. Thus, we coined Anc284 as SUPer Rna Ecogii
180 Methyltransferase (SUPREM).
181
182 **Methylation activity of SUPREM in human cells**
183 To check that M.EcoGII and SUPREM are able to methylate RNA in human cells,
184 M.EcoGII-eGFP, SUPREM-eGFP, and eGFP control constructs were transfected in HEK 293T
185 cells. To quantify the exogenous methyltransferase activity, m^6A methylation efficiency was
186 analyzed by immunofluorescence. M.EcoGII expression in cells only showed a trend towards
187 a 2.8-fold increased m^6A fluorescence in transfected cells by comparison to eGFP control,
188 which was not significant (Figure 2A&B, Supplementary Figure S10), while SUPREM
189 expression resulted in a 5.4-fold increase, which was significant compared to both eGFP
190 control and M.EcoGII (Figure 2A&B, Supplementary Figure S10). Compared to the data
191 obtained from *in vitro* experiments with the purified enzyme, while both M.EcoGII and
192 SUPREM were able to methylate RNA, M.EcoGII activity in cells was relatively low
193 compared to SUPREM, which exhibited a more robust methylation activity. These results were
194 confirmed by LC/MS-MS analysis of nucleoside content in purified cellular mRNA, which
195 showed that mRNA from SUPREM-expressing cells contained a significantly higher fraction
196 of m^6A (26.0% of total adenine, i.e., $m^6A + A$) than mRNA from eGFP- (3.48%) or M.EcoGII-
197 expressing cells (6.68%) (Figure 2C&D). Thus, we have successfully engineered an RNA
198 MTase that could be exploited for biological studies.
199 As a proof of principle of the potential for biological applications of SUPREM, we
200 decided to target this enzyme to the nucleus, to restrict RNA methylation to a subcellular

201 compartment. This approach might pave the way for a deeper analysis of RNA methylation,
202 which takes into the account different areas of the cell. Nuclear targeting was achieved by
203 fusing each construct, eGFP, M.EcoGII and SUPREM, with the SV40 Nuclear localization tag
204 (NLS3x3). The constructs were then transfected in HEK293T cells and the m^6A methylation
205 efficiency was quantified by immunofluorescence (Supplementary Figure S12). While both
206 M.EcoGII and SUPREM showed clear nuclear localization (Supplementary Figure S12A), and
207 a significant increase in RNA methylation compared to eGFP, neither enzyme showed an
208 enrichment in nuclear RNA labelling (Supplementary Figure S12B). We hypothesize that the
209 lack of preferential nuclear labelling might be a consequence of the rapid nuclear export of the
210 labelled mRNA.
211
212 **The non-specificity in the methylation motifs is retained in SUPREM**
213 A decrease in the site-specificity of RNA MTases is advantageous for engineering RNA
214 methylation states since more RNA sequences can be targeted. M.EcoGII has been shown to
215 have non-site-specific methylation activity on DNA (31), while the specificity for RNA has
216 not been explored so far. To investigate the presence of non-site-specific RNA methylation
217 activity in M.EcoGII and the ancestral proteins, we used Oxford Nanopore direct RNA
218 sequencing (DRS) together with the Nanocompore pipeline (47, 48) that in addition identifies
219 modification sites in RNA methylated by M.EcoGII, Anc291, and SUPREM. Briefly,
220 Nanocompore identifies potential modification sites in RNA based on the raw electrical signal
221 derived from Nanopore DRS by identifying position-specific differences in dwell time and
222 signal intensity between a modified RNA sample and an unmodified control. For each position
223 in a given RNA sequence, reads from the modified and unmodified samples are grouped into
224 two clusters based on the extracted $\log_{10}$(dwell time) and signal intensity values using a
225 Gaussian mixture model (GMM). A logistic regression test (logit) is then used to determine
226 whether reads derived from the modified sample are significantly overrepresented in one
227 cluster, indicating that the signal at that position is significantly affected by the presence of a
228 nearby modification (usually within the 5-mer occupying the pore). Unlike other commonly
229 used NGS-based methods for analysis of m^6A modification sites (1, 49–53), the Nanocompore
230 pipeline is appropriate for detection of m^6A in RNA methylated by M.EcoGII and its variants
231 because it enables detection of m^6A in any sequence context, even in highly modified RNA,
232 with high spatial resolution.
233 To analyze the methylation site specificity of M.EcoGII, Anc291, and SUPREM, an
234 RNA substrate produced by *in vitro* transcription (the same substrate used for MTase-Glo

methylation assays; 1872 nt, 1456 adenine-containing 5-mers) (Supplementary Table S2) was methylated using each enzyme and sequenced using an Oxford Nanopore MinION device. Modified 5-mers in each RNA sample were identified using the Nanocompore pipeline with two different thresholds: a standard threshold (logit *p*-value < 0.01, corrected for multiple comparisons using the Benjamini-Hochberg method) to identify all potential modification sites, and a more stringent threshold (corrected logit *p*-value < 0.01 and absolute log odds ratio > 0.5) to identify the positions with the strongest evidence for modification (54). Based on either threshold, SUPREM showed significant modification of a larger number of 5-mers than M.EcoGII or Anc291 (Figure 3A, Supplementary Figure S13). According to the standard threshold, 632 (43.4%) of the 1456 adenine-containing 5-mers showed significant modification for SUPREM, compared with 524 (36.0%) for Anc291 and 332 (22.8%) for M.EcoGII. A similar trend was observed with the stringent threshold, with 95 (6.5%) of adenine-containing 5-mers showing significant modification for SUPREM, compared with 44 (3.0%) for Anc291 and 32 (2.2%) for M.EcoGII.

Sequence analysis of the modified 5-mers showed that the RNA methyltransferase activity of M.EcoGII and its variants was largely non-site-specific. No discernable sequence motif was observed among 5-mers modified by M.EcoGII, Anc291, or SUPREM according to the standard threshold (Figure 3B), although a slight bias against GC-rich sequences was observed for M.EcoGII, which was reduced for SUPREM and Anc291 (Figure 3B). On the other hand, the top-ranking modification sites identified using the stringent threshold showed a tendency to have multiple adenine bases, including an adenine at the center of the 5-mer (Figure 3C). This trend may reflect an intrinsic sequence preference of M.EcoGII and its variants, but may also reflect the fact that 5-mers containing multiple adenine bases can have multiple modifications and therefore exert a larger effect on the nanopore signal. The bias towards adenine bases observed in the top-ranked sequences was substantially reduced for SUPREM and Anc291, compared with M.EcoGII (Figure 3C). Altogether, the Nanopore DRS results indicate that M.EcoGII and its variants are largely non-sequence-specific, capable of methylating adenine bases in a wide variety of sequence contexts, and that any potential biases in M.EcoGII are reduced in SUPREM.

Investigation of key residues for improved RNA methyltransferase activity

To investigate which amino acid residues are important for RNA methylation activity, we compared multiple sequence alignments of sequences in the M.EcoGII clade (clade I) and SUPREM clade (clade II) in our maximum-likelihood phylogenetic tree (Figure 4A). We chose

269 14 amino acids as candidate residues that might be responsible for the observed differences in
270 RNA methylation activity based on two conditions, being that the residues were highly
271 conserved among each clade, and that the residues differed between clades I and II. Fourteen
272 single-point mutants of SUPREM were constructed, of which 12 variants could be expressed,
273 and 11 variants could be successfully purified (the amino acid sequences of these variants are
274 given in Supplementary Table S4). Using the MTase-Glo assay to measure RNA methylation
275 activity, one substitution, E179Q, was found to significantly decrease the RNA methylation
276 activity of SUPREM (3.4-fold decrease, $P = 0.0079$) (Figure 4B), suggesting that this residue
277 is important for RNA binding and methylation.
278
279 **Structural analysis of M.EcoGII and its ancestral variants**
280 To gain a greater understanding of the structural basis for the non-site-specific
281 methylation activity of M.EcoGII and the differences in DNA and RNA methylation activity
282 between M.EcoGII and its ancestral variants, we aimed to generate structural models of
283 M.EcoGII, Anc291, and SUPREM. To date, experimental structures of M.EcoGII have not
284 been determined. We performed extensive crystallization screening of M.EcoGII, SUPREM,
285 and Anc291, but no crystals were obtained, most likely due to the conformational flexibility of
286 the long interdomain linker. Homology modeling was unfeasible because of the low sequence
287 identity (<20%) of M.EcoGII to proteins of known structure. Indeed, although the structure of
288 M.EcoGII was previously modeled using the structure of M.EcoP15I (PDB 4ZCF) as a
289 template, M.EcoP15I bears no recognizable sequence homology to M.EcoGII in the target
290 recognition domain (55). Based on these considerations, we decided to use AlphaFold2 for
291 structural modeling.
292 To this end, we first aimed to determine the oligomeric state of M.EcoGII and the
293 ancestral proteins. Although previous research suggested that M.EcoGII is a monomeric
294 protein based on native-mass spectrometry experiments (32), previously characterized β-class
295 methyltransferases have shown a dimeric structure (56–58), and monomeric structures appear
296 to be incompatible with catalytic activity (56). We used size-exclusion chromatography (SEC)
297 to estimate the molecular weight of M.EcoGII and the ancestral proteins based on a standard
298 curve and found that each protein eluted as a dimer (Figure 4C, Supplementary Figure S14).
299 Finally, structural models of M.EcoGII and ancestral protein homodimers were
300 constructed using AlphaFold2 multimer modeling (Figure 4D, Supplementary Figure S15) (59).
301 According to the structural prediction, each M.EcoGII monomer is composed of a Rossmann
302 fold catalytic domain (residues 1–134 and 260–349, mean pLDDT score 95.8) and a four-helix

303 bundle target recognition domain (residues 160–237, mean pLDDT score 94.4) containing a
304 helix-turn-helix motif. The two domains are connected by a long, flexible, and highly charged
305 linker, partly composed of a two-helix structure (mean pLDDT score 89.1 for the predicted
306 helical portion). A search of the Protein Data Bank for similar structures using the DALI server
307 showed that the predicted structure of the Rossmann fold domain of M.EcoGII shows the
308 highest similarity to other DNA methyltransferases such as M.MboIIA and M.CcrMI. In
309 contrast, the target recognition domain bears similarity to various helix-turn-helix domain
310 DNA-binding proteins (e.g., transcriptional regulators), but generally not to known
311 methyltransferase structures (Supplementary Table S5). The overall predicted structure of the
312 site-specific DNA methyltransferase M.PflMI, including its target recognition domain, is
313 similar to M.EcoGII, with the exception of the highly charged linker region, which is absent in
314 M.PflMI (Supplementary Figure S15E). The large positively charged surface of M.EcoGII
315 spanning the target recognition domain and interdomain linker is likely important for the non-
316 specific binding of DNA and RNA (Figure 4E), providing a large interaction surface that
317 effectively compensates for the lack of protein-DNA and protein-RNA interactions with
318 specific bases.

319 The ancestral proteins showed a similar overall predicted structure to M.EcoGII
320 (Supplementary Figure S15). The E179 residue identified as being important for RNA binding
321 and/or methylation is located on the edge of the putative DNA/RNA binding pocket, based on
322 the structural models predicted by AlphaFold2 (Figure 4D). In addition, the amino acids (97-
323 110, 116-140) present on the interaction surface between the subunits are completely conserved
324 among M.EcoGII and ancestral proteins, suggesting that the dimer interface might be crucial
325 for these enzymes. Co-crystal structures of SUPREM and RNA would be required for a
326 detailed understanding of the mechanisms underlying the improved RNA methylation activity
327 and promiscuity compared with M.EcoGII.

328

329 **Discussion**

330 Growing recognition of the importance of RNA methylation, particularly m^6A
331 methylation, in a broad range of physiological and pathological processes has driven demand
332 for new tools for the manipulation and analysis of RNA methylation states (25, 26). M.EcoGII
333 was identified as a potentially useful tool to manipulate m^6A methylation in mammalian cells
334 based on its non-site-specific DNA/RNA methyltransferase activity and its non-mammalian
335 origin, but the enzyme has not yet been utilized for this purpose due to its low expression and

336 RNA methylation activity. In this work, we used ASR to engineer an M.EcoGII variant,
337 SUPREM, with a number of advantages over the wild-type enzyme, including a 12-fold
338 increase in RNA methylation activity, 8 °C increase in thermostability, an 8-fold increase in
339 expression, and highly relaxed sequence specificity, resulting in methylation of up to 69% of
340 adenine bases in single strand RNA. M.EcoGII and other RNA MTases are generally
341 challenging targets for other protein engineering techniques such as rational design or directed
342 evolution; for example, the lack of experimental structural information on M.EcoGII and the
343 complexity of structure-function relationships in nucleic acid enzymes would have limited the
344 potential of rational protein engineering in this case. A computational approach for the
345 engineering of RNA MTase-RNA interactions would be challenging due to difficulty in
346 predicting interactions between proteins and RNA. In contrast, our work shows the power of
347 sequence-based engineering approaches such as ASR to enable rapid improvement of a broad
348 range of protein properties, including challenging cases where structural information is limited.
349 All of the ancestral variants of M.EcoGII characterized in this work showed higher
350 thermostability and expression than the wild-type protein, with ΔT_m between 7 and 16 ºC. It is
351 noteworthy that major improvements in thermostability were achieved despite the limited
352 phylogenetic diversity of the sequences in the M.EcoGII clade, which are mainly derived from
353 prophage regions in Enterobacteria genomes. Previous studies that have used ASR for protein
354 engineering have typically focused on more phylogenetically diverse and evolutionarily
355 ancient protein families (60), based on the hypothesis that the ancient progenitors of modern
356 proteins existed in thermophilic organisms and were therefore thermostable. In the case of
357 M.EcoGII, however, there is little reason to assume *a priori* that there would be a trend in
358 thermostability from the ancestor of the M.EcoGII clade towards the modern enzyme. Other
359 recent studies have similarly shown that it is unnecessary to reconstruct evolutionarily ancient
360 proteins to achieve substantial increases in thermostability by ASR (42, 43). For example,
361 Livada *et al*. systematically characterized all 56 ancestral nodes and 57 extant nodes in a
362 phylogeny of gene reductases and found that even the shallowest nodes (i.e., the most recent
363 ancestral nodes, adjacent to the extant nodes at the tips of the tree) had a T_M up to 16 °C higher
364 than their closest extant protein (43). The reason for this general trend of reconstructed
365 ancestral proteins showing higher thermostability compared with extant proteins is still unclear,
366 but the bias in maximum-likelihood ASR towards common amino acids (41) and consensus
367 amino acids (38) at ambiguously reconstructed positions may be partially responsible. The
368 ability of ASR to preserve coevolutionary relationships between residues also provides an
369 advantage over related methods such as consensus design (41). Overall, our work contributes

370 to a growing body of evidence that ASR can be successfully applied as a protein engineering
371 tool to a diverse range of protein families, even those with limited phylogenetic diversity.
372 Substantial differences in RNA methyltransferase activity were also observed among
373 the ancestral M.EcoGII variants, which likely reflects the neutral variation of the promiscuous
374 RNA methyltransferase activity of M.EcoGII throughout its evolutionary history. Although the
375 precise biological function of M.EcoGII is unknown, its DNA methyltransferase activity is
376 more likely to be biologically relevant, given that (1) phage DNA methyltransferases have a
377 general role in the protection of phage DNA from host restriction enzymes during infection
378 (50); (2) M.EcoGII shows higher DNA methyltransferase activity than RNA methyltransferase
379 activity (32); and (3) m^6A modification of mRNA has no known function in bacteria, although
380 it is known to occur in Gram-negative bacteria (20). Thus, we hypothesize that the DNA
381 methyltransferase activity of M.EcoGII represents its native activity and that its RNA
382 methyltransferase activity represents a promiscuous activity that is not itself under selection
383 pressure. In other words, we hypothesize that the RNA methyltransferase activity is a
384 byproduct of non-specific DNA methyltransferase activity that is driven mainly by non-
385 specific ionic interactions between the positively charged surface of M.EcoGII and the
386 negatively charged phosphate groups of DNA, which also enables binding of RNA. Consistent
387 with the fact that the promiscuous activities of enzymes generally show greater variability
388 among homologous enzymes than their native activities (61), the observed variation in RNA
389 methyltransferase activity among the M.EcoGII variants (~12-fold) is greater than the variation
390 in DNA methyltransferase activity (~3-fold). The fact that there is no trend in RNA
391 methyltransferase activity from the earlier ancestral proteins towards M.EcoGII is also
392 suggestive of neutral variation. Altogether, these considerations suggest that the ~12-fold
393 increase in RNA methyltransferase activity of M.EcoGII achieved by ASR can be attributed to
394 sampling of the neutral variation of RNA methyltransferase activity that occurred throughout
395 its evolutionary history. More generally, these results suggest that ASR can be used to obtain
396 enzyme variants with improved properties and increased activity, even if the desired activity is
397 not the native enzyme activity.
398 The combination of non-site-specific and highly efficient m^6A methylation activity
399 observed in SUPREM is unique among known RNA MTases and should be broadly useful in
400 applications that require unbiased and efficient labeling of RNA when combined with other
401 components such as RNA-binding proteins by fusion or conjugation. For example, SUPREM
402 could potentially be applied to proximity labeling for transcriptome-wide detection of RNA
403 binding sites for RNA-binding proteins. Fusion of a specific RNA-binding protein to SUPREM

404 would lead to increased m^6A methylation adjacent to the binding site of the RNA-binding
405 protein, allowing detection of the binding site using techniques for m^6A detection, including
406 Oxford Nanopore direct RNA sequencing. A similar approach, TRIBE, relies on fusion of an
407 RNA-binding protein to the RNA-editing enzyme adenosine deaminase (ADAR) rather than
408 an RNA MTase for proximity labeling of the RNA binding site (62). This method provides
409 numerous advantages over the widely-used cross-linking immunoprecipitation (CLIP) method
410 for detection of RNA-protein interactions, such as the relatively low amount of input material
411 needed; however, this method may also be susceptible to bias due to the inherent substrate
412 specificity of ADAR (62), a problem that could potentially be addressed using SUPREM.
413 SUPREM could also be applied to site-specific modification of RNA through conjugation of
414 SUPREM to sequence-programmable RNA-binding proteins, such as dCas13 and PUF
415 proteins (28, 29). Unlike previous iterations that rely on fusion proteins of the human m^6A
416 writer METTL3, SUPREM would also allow site-specific methylation of sites targeted by other
417 m^6A writers such as METTL16, as well as non-canonical methylation sites and sites that are
418 not naturally methylated (for example, to modulate mRNA stability, localization or translation
419 efficiency).
420 SUPREM also has potential applications for analyzing the spatial distribution of mRNA
421 in specific organelles, liquid-liquid phase separated compartments, and asymmetric structured
422 cells such as neurons, as demonstrated by proof-of-principle experiments targeting SUPREM
423 to the nucleus of HEK293T cells. Site-specific localization of SUPREM would result in
424 selective m^6A labeling of region-specific mRNAs, which could be useful to understand which
425 types of transcripts are localized in specific subcellular compartments. Finally, the improved
426 thermostability, expression, and RNA methyltransferase activity of SUPREM make it a useful
427 starting point for further protein engineering. For example, engineering of the cofactor binding
428 pocket of SUPREM to promote the binding of synthetic, functionalized *S*-adenosyl-L-
429 methionine analogs, an approach that has been applied to sequence-specific DNA
430 methyltransferases (63), would allow bioorthogonal sequence-independent functionalization
431 and labeling of mRNA. Altogether, we expect that SUPREM will be a powerful enabling tool
432 for the development of new methods for RNA manipulation and analysis.
433
434 **Materials and methods**
435 **Phylogenetic analysis**
436 Protein BLAST on local environment (ncbi-blast-2.10.0+) was used to search the collection of
437 amino acid sequences of type II class β methyltransferases, to which M.EcoGII belongs, in the

438 REBASE database (44) (retrieved on 2020-09-17) using M.EcoGII (GenBank accession
439 number EGT69837.1) as a query sequence (64). An e-value cutoff of 10^{-7} was used. Redundant
440 sequences were omitted by CD-HIT using a 90% sequence identity threshold, giving a total of
441 2368 sequences (65). The sequence similarity network was created by all-versus-all BLAST of
442 the resulting sequences using an e-value threshold of 10^{-30} and visualized in Cytoscape
443 software (66). For ancestral sequence reconstruction, multiple sequence alignments of the
444 M.EcoGII-containing cluster (158 sequences) and the outgroup clusters (46 sequences) were
445 constructed separately using MUSCLE software (67). The aligned protein sequences were
446 manually edited to remove two outlier sequences, large insertions in the outgroup, and N-/C-
447 terminal extensions. Each edited multiple sequence alignment was combined by profile-profile
448 alignment in MUSCLE. An unrooted maximum-likelihood tree was constructed using IQ-
449 TREE software (68). To assess the robustness of the tree, inference of the phylogenetic tree
450 was repeated five times using the sample substitution model (LG+I+G4) or two other
451 substitution models (WAG and VT). The LG+I+G4 model was chosen as the best model based
452 on IQ-tree calculation of Bayesian information criterion (BIC) score. The robustness of each
453 node was assessed using the Shimodaira-Hasegawa approximate likelihood ratio test (SH-
454 aLRT) and ultrafast bootstrap analysis (UFboot) with 1000 replicates. Ancestral protein
455 sequences were reconstructed using the empirical Bayes method implemented in PAML using
456 the LG substitution model (69). The posterior probability distribution at each site for each
457 ancestral node was also calculated using PAML.
458
459 **Cloning and mutagenesis**
460 M.EcoGII and six ancestral protein sequences were reverse-translated into nucleotide
461 sequences that were codon optimized for expression in *E. coli*. A sequence encoding a C-
462 terminal His_6 tag with a linker sequence (N-GGAAALGHHHHHH-C) was added to each gene,
463 and the resulting DNA sequences were synthesized (Twist Bioscience, California, USA) and
464 cloned into the pET-28a(+) plasmid using In-Fusion HD cloning kit (Takara Bio Inc., Shiga,
465 Japan, 639648) or the iVEC3 method (70). In-frame cloning into the pET-28a (+) vector was
466 confirmed by DNA sequencing. Mutagenesis of SUPREM was performed using the PrimeStar
467 Mutagenesis method and PrimeStarMax polymerase (Takara Bio Inc., R045A).
468 The DNA sequences corresponding to M.EcoGII and SUPREM were inserted into the
469 pCDNA3.1-eGFP-GSx3 plasmid using the In-Fusion HD cloning kit (Takara Bio Inc., 639648),
470 resulting in a sequence encoding eGFP with a flexible linker sequence $(GGGGS)_3$ linked to the
471 N-terminus of the methyltransferase. The 3x nuclear localization tag

472 (PKKKRKVDPKKKRKVDPKKKRKVLE) was inserted into the N-terminus of eGFP in
473 pCDNA3.1-eGFP-GSx3, pCDNA3.1-eGFP-GSx3-M.EcoGII, or pCDNA3.1-eGFP-GSx3-
474 SUPREM using digestion by XhoI (New England Biolabs Inc. (NEB), Massachusetts, USA,
475 R0146S) and ligation by T4 DNA Ligase (NEB, M0202S).
476
477 **Protein expression**
478 Plasmids containing M.EcoGII and all ancestral protein genes were transformed into *E.*
479 *coli* strain SHuffle T7 (NEB, C3026J) and plated on LB agar containing 30 μg/mL kanamycin
480 and 2% glucose, incubating at 37 °C overnight. Colonies were inoculated into 5-20 mL LB
481 media with 30 μg/ml kanamycin and 2% glucose and incubated at 37 °C with shaking at 220
482 rpm overnight. The preculture was transferred into 0.5-2 L LB media with 30 μg/mL
483 kanamycin and 1% glucose, and the *E. coli* cells were grown at 37 °C with shaking at 220 rpm
484 until the OD_{600} reached 0.5-0.8. Protein expression of M.EcoGII and the ancestral proteins was
485 induced by 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h at 25 °C. SUPREM
486 variants were expressed with 1 mM IPTG for 18 h at 18 °C. Cells were harvested and stored at
487 -20 °C before protein purification.
488
489 **Protein purification**
490 Cells were resuspended with lysis buffer (20 mM HEPES pH 7.0, 250 mM NaCl, 50
491 mM imidazole with 1× Protease Inhibitor Cocktail (EDTA-free) (Nacalai Tesque Inc., Kyoto,
492 Japan, 03969-21) and 1 mg/ml lysozyme (Sigma-Aldrich, Missouri, USA, 62971) and lysed
493 by sonication. The lysate was centrifuged at 10000 × *g*, 4 °C for 30 min, and the supernatant
494 was filtered through a 0.45 μm filter. The initial protein purification was performed using an
495 ÄKTA Pure 150 system (Cytiva Japan, Tokyo, Japan) and a 1 ml HisTrap FF column (Cytiva
496 Japan, 17531901) with Ni wash buffer (20 mM HEPES pH 7.0, 250 mM NaCl, 20 mM
497 imidazole), followed with elution buffer (20 mM HEPES pH 7.0, 250 mM NaCl, 200 mM
498 imidazole) as a stepwise purification. Peak fractions were analyzed by SDS-PAGE. Fractions
499 enriched in the target protein were pooled on ice, filtered through a 0.22 μm filter, and purified
500 by size exclusion column chromatography (HiLoad 26/600 Superdex 75 pg, Cytiva Japan, 28-
501 9893-34) with working buffer (20 mM HEPES pH 7.0, 250 mM NaCl). Peak fractions were
502 analyzed by SDS-PAGE. Fractions enriched in the target protein were pooled on ice, filtered
503 through a 0.22 μm filter, and concentrated in an Amicon Ultra 10 kDa molecular weight cutoff
504 (MWCO) centrifugal concentration device (Merck Millipore, UFC9010) with working buffer.

505 The concentrated protein samples (> 1 mg/ml) were mixed with glycerol to reach 10% (v/v)
506 concentration, quickly frozen by liquid nitrogen, and stored at -80 ºC.
507 For SUPREM variants, cells were resuspended with lysis buffer (20 mM HEPES pH
508 7.0, 250 mM NaCl, 20 mM imidazole with 1× Protease Inhibitor Cocktail (EDTA-free)
509 (Nacalai Tesque Inc., 03969-21) and 1 mg/ml lysozyme (Sigma-Aldrich, 62971) and lysed by
510 sonication. The lysis supernatant was obtained by centrifugation at 10000 × *g*, 4 °C for 30 min.
511 SUPREM variants were purified by affinity chromatography in an open column using
512 cOmplete His-Tag Purification Resin (Roche, Basel, Switzerland, 05893682001), eluting with
513 20 mM HEPES pH 7.0, 250 mM NaCl, and 100 mM imidazole at room temperature. Each
514 fraction was analyzed by SDS-PAGE. Fractions enriched in the target protein were pooled,
515 filtered through a 0.22 μm filter, and concentrated in an Amicon Ultra 10 kDa MWCO
516 centrifugal concentration device (Merck Millipore, UFC9010) with working buffer. The
517 concentrated protein samples (>1 mg/ml protein and < 1 mM imidazole) were mixed with
518 glycerol to reach 10% (v/v) concentration, quickly frozen by liquid nitrogen, and stored at -80
519 ºC.
520
521 **MTase-Glo assay**
522 The DNA substrate encoding M.EcoGII (3569 bp, containing 1819 adenine bases) was
523 amplified by PCR and purified using a NucleoSpin Gel and PCR Clean-up Kit (MACHEREY-
524 NAGEL GmbH & Co. KG, Düren, Germany, 740609) with Milli-Q water. The RNA substrate
525 encoding the firefly luciferase gene (1872 nt, 544 adenine bases) was prepared by *in vitro*
526 transcription using HiScribe T7 High Yield RNA Synthesis Kit (NEB, E2040S) with 1 unit
527 Rnase Inhibitor (NEB, M0314S) and purified in RNase-free water using a NucleoSpin RNA
528 Clean-up spin column kit with a DNA removal column (MACHEREY-NAGEL GmbH & Co.
529 KG, 740948). *In vitro* methylation experiments were performed for 10 min (DNA) or 1 h
530 (RNA) at 37 °C in 8 μL reaction mixtures containing 1× CutSmart buffer (NEB, R0141) with
531 20 nM DNA or 50 nM RNA with 1 unit RNase Inhibitor (NEB, M0314S), 100 nM
532 methyltransferase, and 20 μM SAM, attached to MTase-Glo kit (Promega, Wisconsin, USA,
533 V7601). The methylation reaction in the 8 μL reaction was quenched by heating at 80 °C for 1
534 min at each time point. The methyltransferase reaction product *S*-adenosylhomocysteine was
535 quantified using MTase-Glo kit (Promega, V7601) following the manufacturer's protocol (46).
536 To normalize the SAH production rate in DNA/RNA methylation, a calibration curve for SAH
537 was constructed in two independent experiments using from 0 to 5 μM SAH according to the

538 manufacturer's protocol of the MTase-Glo kit (Promega, V7601). The signal was detected
539 using a Tecan Spark microplate reader (Tecan Japan Co., Ltd, Japan).
540
541 **LC-MS/MS assay**
542 The RNA substrate, a transcribed portion of the pET-28a(+) vector (294 nt, 80 adenine bases),
543 was prepared by *in vitro* transcription using HiScribe T7 High Yield RNA Synthesis Kit (NEB,
544 E2040S) with 1 unit RNase Inhibitor (NEB, M0314S) and purified in RNase-free water using
545 a NucleoSpin RNA Clean-up spin column kit with a DNA removal column (MACHEREY-
546 NAGEL GmbH & Co. KG, 740948). *In vitro* RNA methylation experiments were performed
547 for 2 h at 37 °C in 20 μL reaction mixtures containing 1× CutSmart buffer (NEB, R0141) with
548 1 μg RNA and 1 unit RNase Inhibitor (NEB, M0314S), 5 μg methyltransferase or no enzyme,
549 and 320 μM SAM (NEB, B9003S). The methylation reaction was quenched by heating at 80 °C
550 for 1 min. RNA was purified using an RNA Clean & Concentrator-5 kit (Zymo Research,
551 California, USA, R1015) into 20 μL RNase-free water, and then digested from RNA to
552 nucleosides using a Nucleoside Digestion Mix kit (NEB, M0649S) at 37 ºC for 1 h. The
553 digestion reaction was quenched with 33% ethanol as the final concentration. The solvent was
554 removed using a centrifugal evaporator, and the sample was dissolved in 50 μL of 40 mM
555 ammonium acetate pH 6.0.
556 LC-MS analysis was performed in analytical duplicate on a Bruker compact ESI-Q-
557 TOF MS (Bruker, Massachusetts, USA) and a Shimadzu HPLC system (LC-40D, SPD-M40,
558 and CBM-40) using an Ascentis Express C18, 2.7 μm HPLC column (15 cm × 2.1 mm, 2.7
559 μm, Sigma Aldrich, 53829-U) with a solvent system consisting of 1% acetonitrile in H_2O with
560 0.1% formic acid (A) and 95% acetonitrile/5% H_2O with 0.1% formic acid (B). The
561 nucleosides were eluted at a flow rate of 0.30 ml/min with a linear gradient of mobile phase of
562 0% B (0-2 min), 0-15% B (2-6 min), 15% B (6-8 min), 90% B (8-10 min) for washing column,
563 and 0% B (10-13 min) for re-equilibration. The sample was injected in a volume of 3 μL. Each
564 nucleoside peak was identified by comparison between enzyme-treated samples and standard
565 compounds. The retention time of A and m^6A was 3.6 and 5.8 min, respectively. For ESI, the
566 full mass scan range was *m/z* 50-1000 in positive mode. The identities of the A and m^6A peaks
567 were validated based on their MS fragmentation patterns: A *m/z* 268 > 136; m^6A *m/z* 282 >
568 150 (Supplementary Figure S8). The m^6A fraction was calculated using the peak area of the
569 UV chromatogram at 254 nm and the following formula: Peak area of m^6A / (Peak area of m^6A
570 + Peak area of A).

571 To obtain total cellular RNA, HEK 293T cells were grown in DMEM high glucose
572 (Nacalai Tesque, 08458-45) supplemented with 10% fetal bovine serum (FBS, Gibco, 102-70-
573 106) and 100 µg/mL primocin (InvivoGen, ant-pm-1). HEK 293T cells were transfected with
574 NLSx3-eGFP, NLSx3-M.EcoGII-eGFP or NLSx3-SUPREM-eGFP constructs using JetPei®
575 (Polyplus, 101000053) transfection reagent, following manufacturer's recommendations. 48
576 hours after transfection, cells were lysed, and RNA was extracted with the NucleoSpin® RNA
577 Plus kit (MACHEREY-NAGEL GmbH & Co. KG, 740984) according to manufacturer's
578 instructions.
579 mRNA was purified from 50 µg total RNA samples using a NEBNext® High Input
580 Poly(A) mRNA Isolation Module (NEB, E3370S). Removal of rRNA was validated by agarose
581 gel electrophoresis (1.5% gel) and staining with SYBR Gold Nucleic Acid Gel Stain
582 (Invitrogen, S11494). 400 ng mRNA was digested to nucleosides using a Nucleoside Digestion
583 Mix Kit (NEB, M0649S) at 37 ºC overnight. The digestion reaction was quenched with 33%
584 ethanol as the final concentration. The solvent was removed using a centrifugal evaporator,
585 and the sample was dissolved in 20 µL of 40 mM ammonium acetate pH 6.0.
586 LC-MS/MS analysis for nucleoside samples derived from cellular mRNA was
587 performed in analytical duplicate on a Thermo ESI-Q-Exactive HF (Thermo Fisher Scientific,
588 Massachusetts, USA) and a Waters ACQUITY UPLC M-Class (Waters Corporation,
589 Massachusetts, USA) using an ACQUITY UPLC HSS T3 1.8 µm column (150 mm × 1.0 mm,
590 1.8 µm, Waters Corporation, 186003539) with a solvent system consisting of 99.9% H_2O with
591 0.1% formic acid (A) and 99.9% acetonitrile with 0.1% formic acid (B). The samples were
592 eluted at a flow rate of 50 µl/min with a linear mobile phase gradient of 2% B (0-2 min), 2-
593 35% B (2-7 min), 35% B (7-9 min), 35%-98% B (9-10 min), 98% B (10-13 min) for washing
594 column, and 2% B (13-20 min) for re-equilibration. The full mass scan range was *m/z* 150 to
595 350 with positive mode. The retention time of A and m^6A was 5.6 and 6.4 min, respectively.
596 Nucleoside peaks were compared between mRNA nucleoside samples and standards to identify
597 each peak. The A and m^6A were validated in positive mode as MS fragmentation patterns: A
598 268 > 136; m^6A 282 > 150, as same as *in vitro* sample. The amount of m^6A or A was calculated
599 using the calibration curve of the EIC area of m^6A (6.4 min) or A (5.6 min) standards. The m^6A
600 fraction was calculated by the following formula: amount of m^6A / (amount of m^6A + amount
601 of A).
602
603 **Differential scanning fluorimetry assay**

604 The DSF experiments were conducted using the StepOnePlus real-time PCR instrument
605 (Applied Biosystems, Massachusetts, USA), as described previously (71). Reaction mixtures
606 were prepared in 96-well Real-Time PCR Plates, comprising 5× SYPRO Orange (Sigma-
607 Aldrich, S5692), 5 μM enzyme, and a total volume of 20 μL working buffer (20 mM HEPES
608 pH 7.0, 250 mM NaCl). The plate was sealed with optically clear sealing film and centrifuged
609 at 2000 × g for 1 min prior to loading into the real-time PCR instrument. The temperature was
610 ramped at a rate of 1% (approximately 1.33 °C/min) from 4 ºC to 95 ºC. Fluorescence signals
611 were monitored using the ROX channel. The melting temperature values were determined by
612 calculating the derivative of fluorescence intensity with respect to temperature and fitting the
613 resulting data to a quadratic equation within a 6 °C window around the T_M, utilizing R software.
614
615 **Immunofluorescence assay**
616 HEK 293T cells were grown in DMEM high glucose (Nacalai Tesque, 08458-45)
617 supplemented with 10% fetal bovine serum (FBS, Gibco, 102-70-106) and 100 μg/mL
618 primocin (InvivoGen, ant-pm-1). Cells were seeded at an appropriate confluency on polylysine
619 (Sigma, P4832) coated coverslips one day before JetPei® (Polyplus, 101000053) transfection
620 with the different constructs, following the manufacturer's recommendations. 48 hours after
621 transfection, cells were fixed for 30 minutes at room temperature (RT) in 4% paraformaldehyde
622 in PBS under a fume hood, then permeabilized in 0.3% Triton in PBS for 30 minutes at RT.
623 Cells were blocked with 1% (v/v) BSA in PBS for 1 hour at RT and incubated overnight at
624 4 °C with primary anti-m^6A antibody (Synaptic systems, 202-003) diluted in the blocking
625 solution. The next day, cells were incubated with secondary conjugated Alexa Fluor 647
626 antibody (Invitrogen, A21245) for 1 hour in the blocking solution, and then with NucBlue™
627 (Invitrogen, R37606) for 5 minutes in PBS. Slides were mounted in Fluoromount G™
628 (Invitrogen, 00-4958-02), and fluorescence was observed with Zeiss LSM 900 confocal
629 microscope. m^6A immunofluorescence staining was quantified using the software ImageJ. A
630 mask was made using the GFP fluorescence driven by the transfected construct and m^6A
631 intensity was quantified inside the mask and averaged by the cell area. Statistical analysis was
632 performed using the software GraphPad Prism. Statistical significance was evaluated using
633 one-way ANOVA followed by *Tukey* posttest.
634
635 **Nanopore direct RNA sequencing**
636 For direct RNA sequencing by MinION sequencer (Oxford Nanopore Technologies,
637 Oxford, UK), *in vitro* RNA methylation experiments were performed for 2 h at 37 °C in 20 μL

638 reaction mixtures containing 1× CutSmart buffer with 5 μg RNA substrate (1872 nt, same
639 substrate as for MTase-Glo experiment, Supplementary Table S2) and 1 unit RNase Inhibitor
640 (NEB, M0314S), 1 μM methyltransferase or no enzyme as a negative control, and 320 μM
641 SAM (NEB, B9003S) for two independent reaction mixture. The reaction mixture without
642 enzyme was used as a negative control. The polyadenylated tail was added by incubation with
643 5 U of *E. coli* poly-A polymerase (NEB, M0276S) for 30 min at 37 °C. The methylated and
644 poly-A-tailed RNA samples were purified using an RNA Clean & Concentrator-5 (Zymo
645 Research, R1013) with RNase-free water. The concentration and purity of the RNA samples
646 were measured using a Qubit 4 fluorometer with the Qubit RNA IQ kit (Thermo Fisher
647 Scientific, Q33221) (RIN score > 8.5). The preparation of libraries for direct RNA sequencing
648 was conducted using the SQK-RNA002 kit from 500 ng of *in vitro* transcribed RNA. The direct
649 RNA sequencing was carried out on a flow cell R.9.4.1 (FLO-MIN106, Oxford Nanopore
650 Technologies) by the Sequencing Section of the Okinawa Institute of Science and Technology,
651 following the protocol provided by Oxford Nanopore Technologies.
652
653 **Sequencing data analysis**
654 The data analysis was performed using the Nanocompore pipeline (version 1.0.4) (47).
655 Base-called fastq files were created using Guppy software (version 6.4.2) and were
656 concatenated into one file. The transcript data were aligned to a reference transcript sequence
657 by minimap2 (version 2.20), following sorting and indexing in samtools (version 1.15.1). The
658 read file was indexed and resquiggled using the index and eventalign commands in
659 Nanocompore (47). Modification sites were detected using the eventalign_collapse and
660 sampcomp commands in Nanocompore (47). Data were visualized using ggplot2 (72) and
661 Prism 9 software (GraphPad Software).
662
663 **Size exclusion chromatography analysis**
664 Fifty to one hundred μg SUPREM or M.EcoGII samples were injected into a Superdex
665 200 Increase 10/300 GL column (Cytiva Japan, 28990944), and eluted at 0.75 ml/min flow rate
666 with 1 column volume of working buffer (20 mM HEPES pH 7.0, 250 mM NaCl). A calibration
667 curve was constructed using the elution volumes of the LMW Marker Kit (Cytiva Japan,
668 17044601).
669
670 **Structural modeling and analysis**

671 Dimer structures of methyltransferases were modeled by Alphafold (ver.2.2.1) using
672 the multimer prediction option (59). Protein structural models were visualized using the
673 PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC. The APBS
674 Electrostatics Plugin in PyMOL was used for electrostatic surface visualization (73).
675 Structures similar to M.EcoGII and its target recognition domain were searched using the DALI
676 server (74).
677

678 **DATA AVAILABILITY**

679 Raw fastq files of the Nanopore direct RNA sequencing were deposited in the National
680 Center of Biotechnology Information (BioProject PRJNA990656). The pET28a-SUPREM (ID
681 205023), pET28a-AncMEcoGII291 (Anc291; ID 205024), and pET28a-AncMEcoGII250
682 (Anc250; ID 205025) plasmids will be made available via Addgene upon publication. We
683 deposited the intermediate files for ASR including sequence list, alignment, tree, ASR output
684 and final sequences in Zenodo (https://zenodo.org/) (doi: 10.5281/zenodo.11056204).

685

686 **Author Contribution**

687 Conceptualization: PL; Supervision: BEC, PL; Investigation: YO; Analysis: BEC, YO, PL;
688 Cell Experiments and analysis: MLC and MT; Writing – original draft: YO, BEC, PL; Writing
689 – review and editing: BEC, YO, PL, MLC, MT.

690

691 **Acknowledgements**

692 We thank the Okinawa Institute of Science and Technology Sequencing Section for preparing
693 the library and sequencing Oxford Nanopore MinION Sequencer. We are grateful for the
694 help and support of LC-MS/MS data collection by Dr. Michael C. Roy from the Instrumental
695 Analysis Section of Core Facilities at Okinawa Institute of Science and Technology Graduate
696 University and thank Dr. Gen-ichiro Uechi for assistance with construction of NLS-fused
697 vectors. We are also grateful for the help and support provided by the Scientific Computing
698 and Data Analysis section of the Research Support Division at OIST.

699

700 **Funding**

701 Funding for the research was supported by Okinawa Institute of Science and Technology
702 (OIST). YO was supported by the Sasakawa Scientific Research Grant from The Japan

Science Society (No. 2023-4053). PL acknowledges Kakenhi Grant (No. 90812256). MT acknowledges Kakenhi Grant (No. 23K27107), BEC was supported by a JSPS Postdoctoral Fellowship for Overseas Researchers from the Japan Society for the Promotion of Science. This project was supported by an OIST Kick start-up grant.

CONFLICT OF INTEREST

None declared.

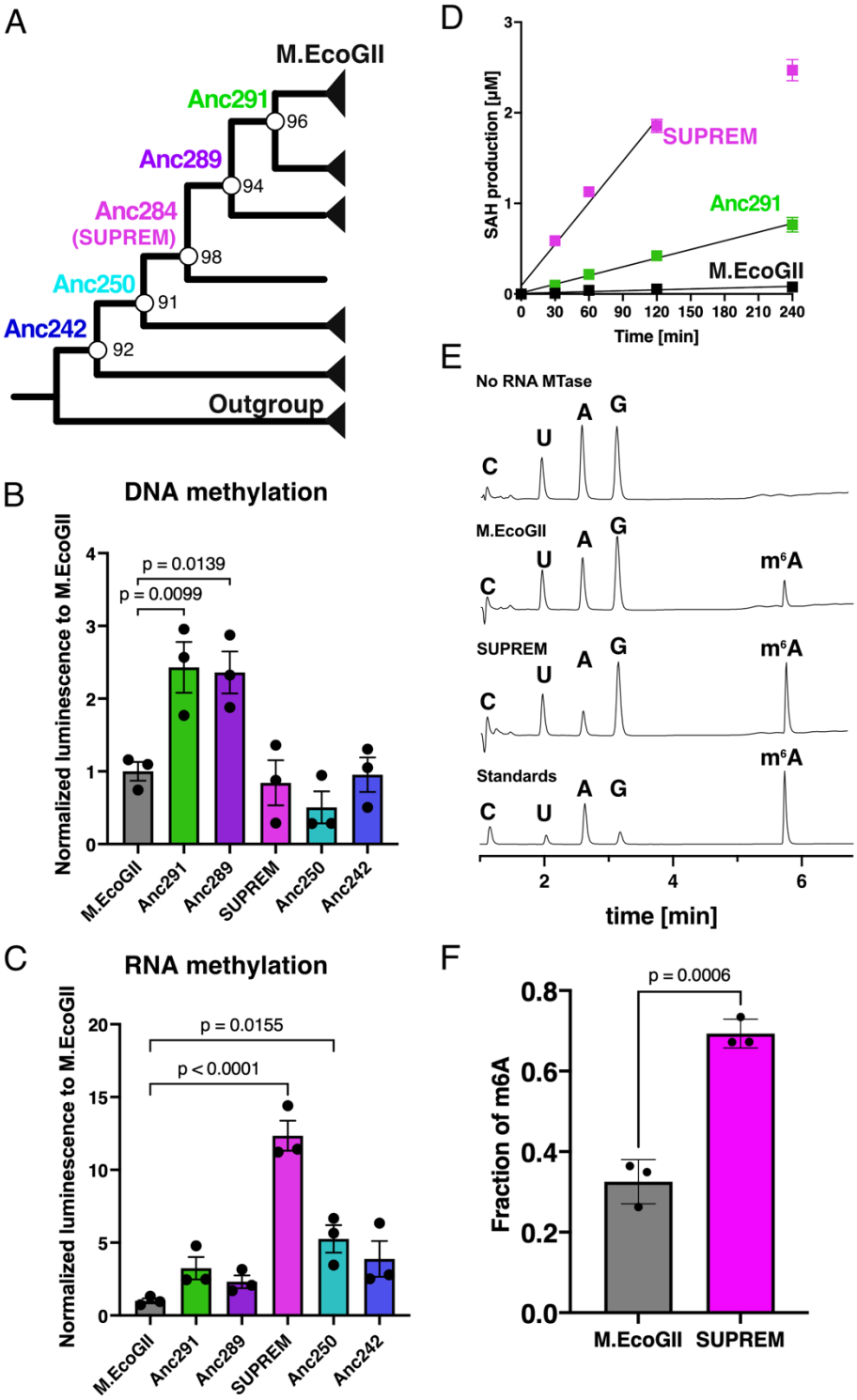

Figure 1. Ancestral sequence reconstruction of M.EcoGII. (A) Schematic maximum likelihood tree. The node values showed ultrafast bootstrap values calculated by IQ-TREE. **(B-C)** MTase-Glo assay of M.EcoGII and its ancestral proteins for (B) DNA and (C) RNA at a single time point. **(B)** The DNA methylation reaction was performed using 100 nM of methyltransferases, 50 nM DNA substrate (3569 bp containing 1819 adenine bases, Supplementary Table S2) and 10 μM SAM in 1× CutSmart buffer (NEB) at 37 °C for 10 min.

719 **(C)** The RNA methylation reaction was performed using 100 nM of the corresponding
720 methyltransferases, 25 nM RNA substrate (1872 nt, containing 544 adenine bases, sequence in
721 Supplementary Table S2), and 10 μM SAM in 1× CutSmart buffer (NEB) at 37 °C for 60 min.
722 Three independent experiments were performed, and each point showed an average of
723 technical duplicates. *P* values were determined by the one-way ANOVA test with the Dunnett
724 test. **(D)** Time course experiments for RNA substrates. The condition of RNA methylation
725 reaction was the same as (C) except for reaction time. Three independent experiments were
726 performed, and each point showed an average of technical duplicates. SAH production amount
727 was calculated by SAH standard curve. The lines represent simple linear regression analysis
728 on the data from 0 to 120 min (SUPREM) or to 240 min (M.EcoGII, Anc291). **(E)** Normalized
729 UV chromatograms of nucleoside digests of RNA treated with M.EcoGII or SUPREM at 254
730 nm. The full normalized UV chromatograms are shown in Supplementary Figure S5. **(F)**
731 Fraction of m^6A methylation in M.EcoGII- and SUPREM-treated samples (i.e., the amount of
732 m^6A as a fraction of total adenine, m^6A + A, as determined by peak area at 254 nm), determined
733 by LC-MS analysis. The RNA methylation reaction was performed using 5 μg of the
734 corresponding methyltransferases, 1 μg RNA substrate (294 nt containing 80 adenine bases,
735 sequence in Supplementary Table S2) and 320 μM SAM (NEB) in 1× CutSmart buffer (NEB)
736 at 37 °C for 120 min. Each point showed three independent methylation reactions. Error bars
737 represent standard error of the mean (SEM).

738

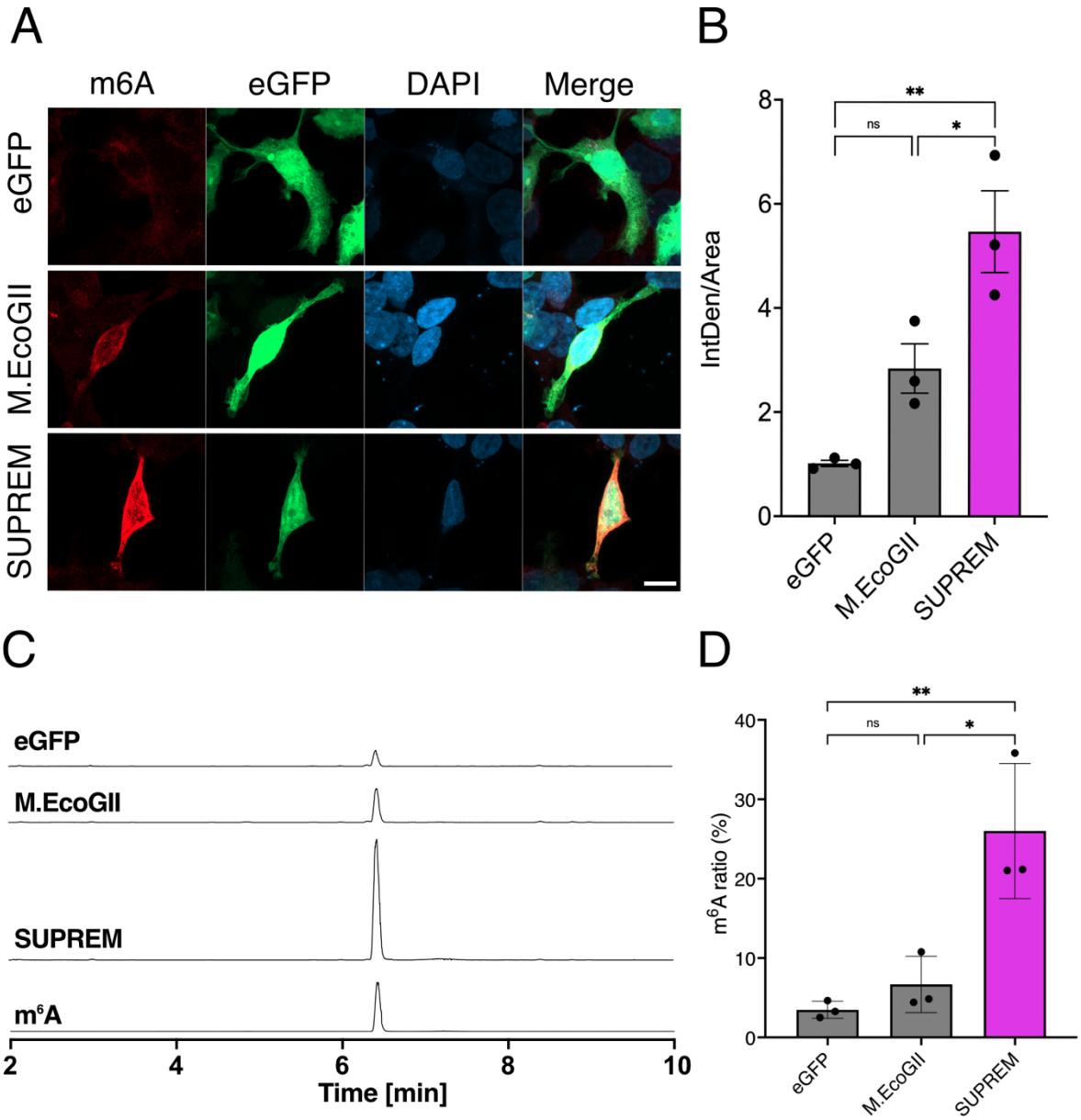

Figure 2. Methylation activity of M.EcoGII variants in HEK293T cells. (A) Representative images (63x objective with a 2x digital zoom) of HEK293T cells transfected with either eGFP, M.EcoGII-eGFP or SUPREM-eGFP constructs (reporter visible in green). m^6A modifications were detected via immunofluorescence (in red) and the nucleus was stained with NucBlue™ (in blue). The scale bar is 10 µm. (**B**) Quantification of m^6A fluorescence intensity in HEK293T transfected as described in (A). The bar chart indicates the normalized intensity by cell area. Statistical analysis was performed by one-way ANOVA followed by *Tukey* posttest (ns: not significant; *: <0.0332; **: <0.0021). (**C-D**) Analysis of m^6A content in digested mRNA from cells transfected with either eGFP, M.EcoGII-eGFP, or SUPREM-eGFP by LC-MS/MS. (**C**) Extracted ion chromatogram at m/z = 282.11 corresponding to m^6A. (**D**) Fraction of m^6A methylation in eGFP-, M.EcoGII-, and SUPREM-expressing HEK293T cells (i.e., the amount

751 of m^6A as a fraction of total adenine, $m^6A + A$, as determined by calibration curve of m^6A and
752 A standards), determined by LC-MS analysis. Each point represents a different biological
753 replicate ($n = 3$), which was analyzed in analytical duplicate by LC-MS. Statistical analysis
754 was performed by one-way ANOVA followed by Tukey posttest (ns: not significant; **:
755 <0.0221; ***: <0.0002).
756

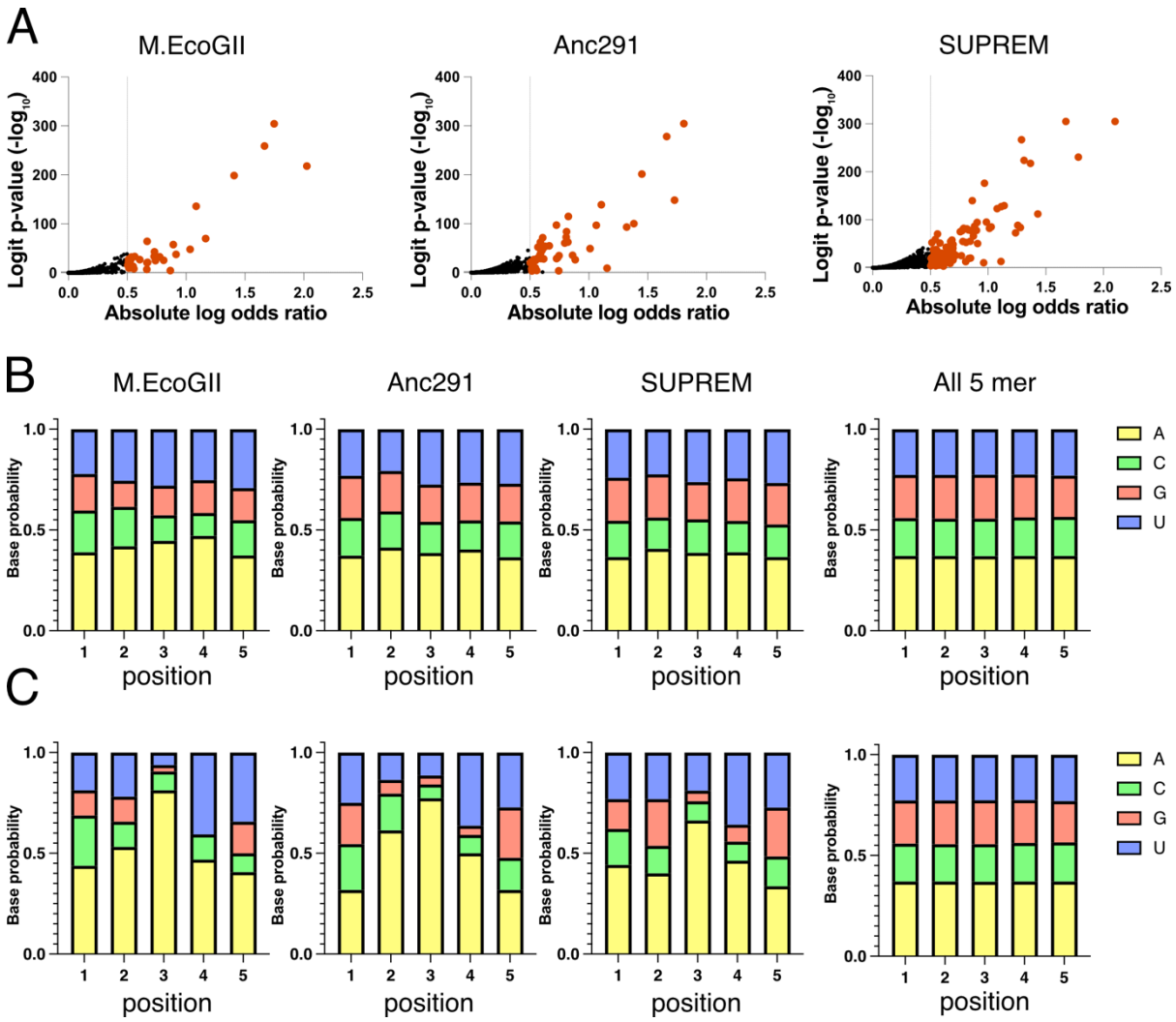

Figure 3. Analysis of RNA methylation site specificity of M.EcoGII and its variants by Nanopore direct RNA sequencing. (A) Sharkfin plots showing the logit *p*-value and absolute value of the logistic regression odds ratio for each 5-mer calculated by Nanocompore ($n = 1456$ adenine-containing 5-mers). Orange points indicate 5-mers that are significantly modified according to a stringent threshold (logit *p*-value < 0.01 and absolute log odds ratio > 0.5). (**B, C**) Frequency of each base in modified 5-mers detected by Nanocompore using (**B**) a standard threshold (logit *p*-value < 0.01) and (**C**) a stringent threshold (logit *p*-value < 0.01 and absolute log odds ratio > 0.5). For all three enzymes, a 5-mer was more likely to be methylated (according to the stringent threshold) if it contained multiple adenine bases (M.EcoGII, 3.6% of 5-mers containing >1 adenine were modified vs. 0.3% of 5-mers containing 1 adenine; Anc291, 4.7% vs 0.8%; SUPREM, 8.7% vs 3.6%; $P < 0.0001$ by Fisher's exact test for all enzymes).

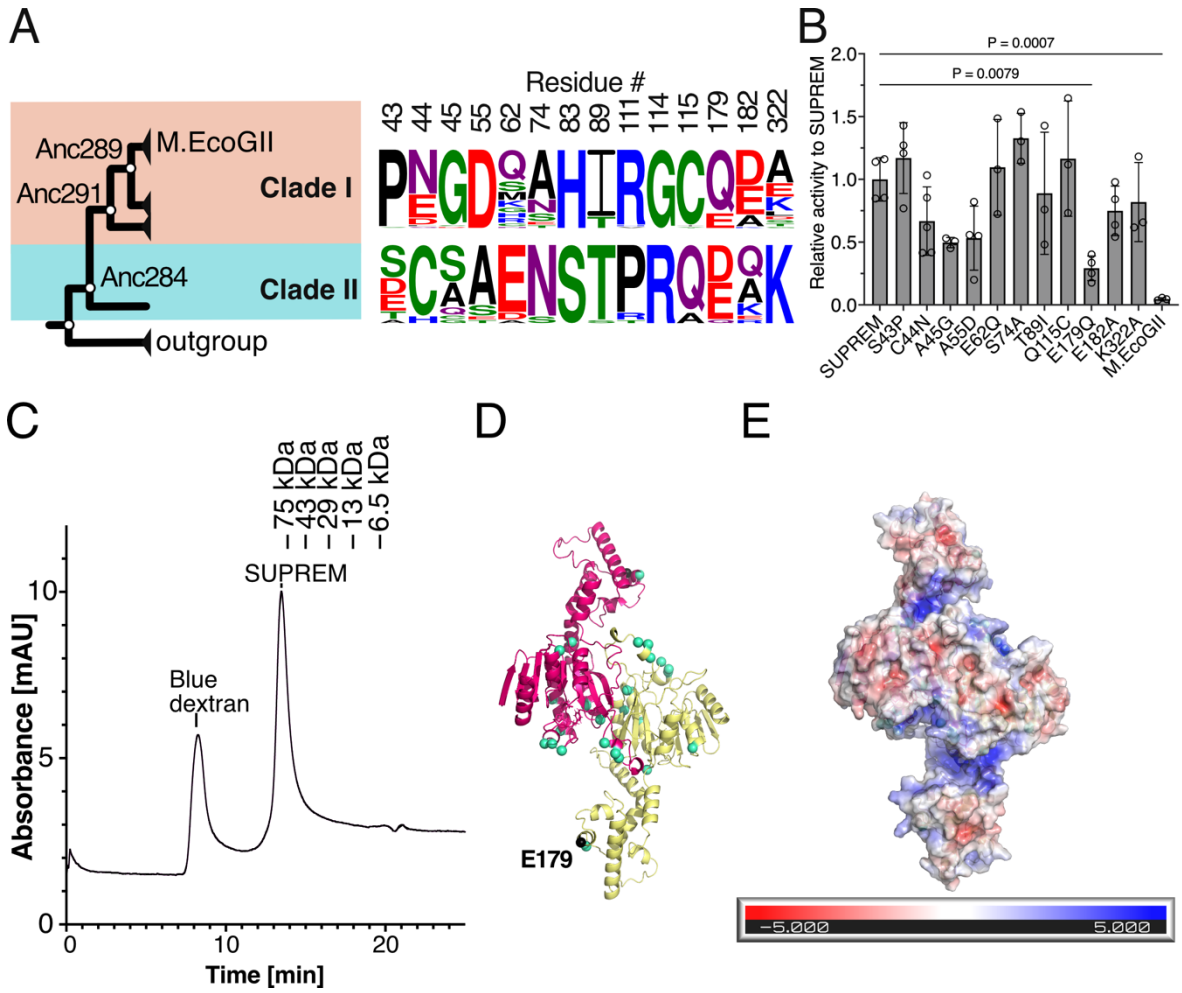

Figure 4. Mutational and structural analysis of SUPREM. (A) Candidate amino acid residues for mutational analysis. (Left) M.EcoGII clade (clade I, 46 sequences) and SUPREM clade (clade II, 19 sequences). (Right) Comparison of amino acid sequence logo of mutational candidate sites based on multiple sequence alignment of clade I and clade II. The height of the logo represents the frequency of amino acid residues at each position. **(B)** RNA methylation activity of SUPREM variants. The RNA methylation reaction was performed using 100 nM methyltransferase and 25 nM RNA substrate (1872 nt, 544 adenine bases) with 10 μM SAM in 1× CutSmart buffer (NEB) at 37 ºC for 60 min. The luminescence was detected by MTase-Glo assay. Three or four independent experiments were performed, and each point shows an average of technical duplicates. *P* values were determined by the one-way ANOVA test with the Dunnett test. **(C)** Size exclusion chromatogram of SUPREM. SUPREM with blue dextran (for estimation of void volume) was loaded into a Superdex 200 Increase 10/300 GL column, and the samples were eluted with 20 mM HEPES pH 7.4 and 250 mM NaCl. The peak positions of low molecular weight markers were shown above. The two independent measurements were performed. **(D)** Dimer structural model of SUPREM generated by AlphaFold2 multimer

786 program (version 2.2.1). Green spheres show residues selected for mutational analysis. **(E)**
787 APBS calculation of the electrostatic surface potential of SUPREM (red: negative charge; blue:
788 positive charge).
789

References

1. Dominissini,D., Moshitch-Moshkovitz,S., Schwartz,S., Salmon-Divon,M., Ungar,L., Osenberg,S., Cesarkas,K., Jacob-Hirsch,J., Amariglio,N., Kupiec,M., *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, **485**, 201–206.

2. Liu,J., Yue,Y., Han,D., Wang,X., Fu,Y., Zhang,L., Jia,G., Yu,M., Lu,Z., Deng,X., *et al.* (2014) A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol*, **10**, 93–95.

3. Warda,A.S., Kretschmer,J., Hackert,P., Lenz,C., Urlaub,H., Höbartner,C., Sloan,K.E. and Bohnsack,M.T. (2017) Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO reports*, **18**, 2004–2014.

4. Wang,X., Lu,Z., Gomez,A., Hon,G.C., Yue,Y., Han,D., Fu,Y., Parisien,M., Dai,Q., Jia,G., *et al.* (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, **505**, 117–120.

5. van Tran,N., Ernst,F.G.M., Hawley,B.R., Zorbas,C., Ulryck,N., Hackert,P., Bohnsack,K.E., Bohnsack,M.T., Jaffrey,S.R., Graille,M., *et al.* (2019) The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Research*, **47**, 7719–7733.

6. Akichika,S., Hirano,S., Shichino,Y., Suzuki,T., Nishimasu,H., Ishitani,R., Sugita,A., Hirose,Y., Iwasaki,S., Nureki,O., *et al.* (2019) Cap-specific terminal N6-methylation of RNA by an RNA polymerase II–associated methyltransferase. *Science*, **363**, eaav0080.

7. Geula,S., Moshitch-Moshkovitz,S., Dominissini,D., Mansour,A.A., Kol,N., Salmon-Divon,M., Hershkovitz,V., Peer,E., Mor,N., Manor,Y.S., *et al.* (2015) m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science*, **347**, 1002–1006.

8. Edupuganti,R.R., Geiger,S., Lindeboom,R.G.H., Shi,H., Hsu,P.J., Lu,Z., Wang,S.-Y., Baltissen,M.P.A., Jansen,P.W.T.C., Rossa,M., *et al.* (2017) N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat Struct Mol Biol*, **24**, 870–878.

9. Pendleton,K.E., Chen,B., Liu,K., Hunter,O.V., Xie,Y., Tu,B.P. and Conrad,N.K. (2017) The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell*, **169**, 824-835.e14.

10. Mendel,M., Delaney,K., Pandey,R.R., Chen,K.-M., Wenda,J.M., Vågbø,C.B., Steiner,F.A., Homolka,D. and Pillai,R.S. (2021) Splice site m6A methylation prevents binding of U2AF35 to inhibit RNA splicing. *Cell*, **184**, 3125-3142.e25.

11. Price,A.M., Hayer,K.E., McIntyre,A.B.R., Gokhale,N.S., Abebe,J.S., Della Fera,A.N., Mason,C.E., Horner,S.M., Wilson,A.C., Depledge,D.P., *et al.* (2020) Direct RNA sequencing reveals m6A modifications on adenovirus RNA are necessary for efficient splicing. *Nat Commun*, **11**, 6016.

12. Yu,J., Chen,M., Huang,H., Zhu,J., Song,H., Zhu,J., Park,J. and Ji,S.-J. (2018) Dynamic m6A modification regulates local translation of mRNA in axons. *Nucleic Acids Research*, **46**, 1412–1423.

13. Di Timoteo,G., Dattilo,D., Centrón-Broco,A., Colantoni,A., Guarnacci,M., Rossi,F., Incarnato,D., Oliviero,S., Fatica,A., Morlando,M., *et al.* (2020) Modulation of circRNA Metabolism by m6A Modification. *Cell Reports*, **31**, 107641.

14. Lin,S., Choe,J., Du,P., Triboulet,R. and Gregory,R.I. (2016) The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Molecular Cell*, **62**, 335–345.

15. Meyer,K.D., Patil,D.P., Zhou,J., Zinoviev,A., Skabkin,M.A., Elemento,O., Pestova,T.V., Qian,S.-B. and Jaffrey,S.R. (2015) 5′ UTR m6A Promotes Cap-Independent Translation. *Cell*, **163**, 999–1010.

16. Wang,X., Zhao,B.S., Roundtree,I.A., Lu,Z., Han,D., Ma,H., Weng,X., Chen,K., Shi,H. and He,C. (2015) N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell*, **161**, 1388–1399.

17. Zheng,G., Dahl,J.A., Niu,Y., Fedorcsak,P., Huang,C.-M., Li,C.J., Vågbø,C.B., Shi,Y., Wang,W.-L., Song,S.-H., *et al.* (2013) ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Molecular Cell*, **49**, 18–29.

18. Yoshinaga,M., Han,K., Morgens,D.W., Horii,T., Kobayashi,R., Tsuruyama,T., Hia,F., Yasukura,S., Kajiya,A., Cai,T., *et al.* (2022) The N6-methyladenosine methyltransferase METTL16 enables erythropoiesis through safeguarding genome integrity. *Nat Commun*, **13**, 6435.

19. Zhou,J., Wan,J., Gao,X., Zhang,X., Jaffrey,S.R. and Qian,S.-B. (2015) Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature*, **526**, 591–594.

20. Deng,X., Chen,K., Luo,G.-Z., Weng,X., Ji,Q., Zhou,T. and He,C. (2015) Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Research*, **43**, 6557–6567.

21. Liu,J., Eckert,M.A., Harada,B.T., Liu,S.-M., Lu,Z., Yu,K., Tienda,S.M., Chryplewicz,A., Zhu,A.C., Yang,Y., *et al.* (2018) m6A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. *Nat Cell Biol*, **20**, 1074–1083.

22. Zhang,C., Tunes,L., Hsieh,M.-H., Wang,P., Kumar,A., Khadgi,B.B., Yang,Y.-Y., Doxtader,K.A., Herrell,E., Koczy,O., *et al.* (2023) Cancer mutations rewire the RNA methylation specificity of METTL3-METTL14. *bioRxiv*, 10.1101/2023.03.16.532618. 16 March 2023, pre-print: not peer-reviewed.

23. Winkler,R., Gillis,E., Lasman,L., Safra,M., Geula,S., Soyris,C., Nachshon,A., Tai-Schmiedel,J., Friedman,N., Le-Trilling,V.T.K., *et al.* (2019) m6A modification controls the innate immune response to infection by targeting type I interferons. *Nat Immunol*, **20**, 173–182.

24. De Jesus,D.F., Zhang,Z., Kahraman,S., Brown,N.K., Chen,M., Hu,J., Gupta,M.K., He,C. and Kulkarni,R.N. (2019) m6A mRNA methylation regulates human β-cell biology in physiological states and in type 2 diabetes. *Nat Metab*, **1**, 765–774.

25. Boulias,K. and Greer,E.L. (2023) Biological roles of adenine methylation in RNA. *Nat Rev Genet*, **24**, 143–160.

26. Sendinc,E. and Shi,Y. (2023) RNA m6A methylation across the transcriptome. *Molecular Cell*, **83**, 428–441.

27. Liu,X.-M., Zhou,J., Mao,Y., Ji,Q. and Qian,S.-B. (2019) Programmable RNA N6-methyladenosine editing by CRISPR-Cas9 conjugates. *Nat Chem Biol*, **15**, 865–871.

28. Wilson,C., Chen,P.J., Miao,Z. and Liu,D.R. (2020) Programmable m6A modification of cellular RNAs with a Cas13-directed methyltransferase. *Nat Biotechnol*, **38**, 1431–1440.

29. Shinoda,K., Suda,A., Otonari,K., Futaki,S. and Imanishi,M. (2020) Programmable RNA methylation and demethylation using PUF RNA binding proteins. *Chem. Commun.*, **56**, 1365–1368.

30. Su,S., Li,S., Deng,T., Gao,M., Yin,Y., Wu,B., Peng,C., Liu,J., Ma,J. and Zhang,K. (2022) Cryo-EM structures of human m6A writer complexes. *Cell Res*, **32**, 982–994.

31. Fang,G., Munera,D., Friedman,D.I., Mandlik,A., Chao,M.C., Banerjee,O., Feng,Z., Losic,B., Mahajan,M.C., Jabado,O.J., *et al.* (2012) Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. *Nat Biotechnol*, **30**, 1232–1239.

32. Murray,I.A., Morgan,R.D., Luyten,Y., Fomenkov,A., Corrêa,I.R., Dai,N., Allaw,M.B., Zhang,X., Cheng,X. and Roberts,R.J. (2018) The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Research*, **46**, 840–848.

33. Sobecki,M., Souaid,C., Boulay,J., Guerineau,V., Noordermeer,D. and Crabbe,L. (2018) MadID, a Versatile Approach to Map Protein-DNA Interactions, Highlights Telomere-Nuclear Envelope Contact Sites in Human Cells. *Cell Reports*, **25**, 2891-2903.e5.

34. Shipony,Z., Marinov,G.K., Swaffer,M.P., Sinnott-Armstrong,N.A., Skotheim,J.M., Kundaje,A. and Greenleaf,W.J. (2020) Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat Methods*, **17**, 319–327.

35. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

36. Zakas,P.M., Brown,H.C., Knight,K., Meeks,S.L., Spencer,H.T., Gaucher,E.A. and Doering,C.B. (2017) Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat Biotechnol*, **35**, 35–37.

37. Nakano,S., Niwa,M., Asano,Y. and Ito,S. (2019) Following the Evolutionary Track of a Highly Specific l-Arginine Oxidase by Reconstruction and Biochemical Analysis of

910 Ancestral and Native Enzymes. *Applied and Environmental Microbiology*, **85**,
911 e00459-19.

912 38. Trudeau,D.L., Kaltenbach,M. and Tawfik,D.S. (2016) On the Potential Origins of the
913 High Stability of Reconstructed Ancestral Proteins. *Molecular Biology and Evolution*,
914 **33**, 2633–2641.

915 39. Babkova,P., Sebestova,E., Brezovsky,J., Chaloupkova,R. and Damborsky,J. (2017)
916 Ancestral Haloalkane Dehalogenases Show Robustness and Unique Substrate
917 Specificity. *ChemBioChem*, **18**, 1448–1456.

918 40. Clifton,B.E., Kozome,D. and Laurino,P. (2022) Efficient Exploration of Sequence Space
919 by Sequence-Guided Protein Engineering and Design. *Biochemistry*, **62**, 210–220.

920 41. Spence,M.A., Kaczmarski,J.A., Saunders,J.W. and Jackson,C.J. (2021) Ancestral
921 sequence reconstruction for protein engineers. *Current Opinion in Structural Biology*,
922 **60**, 131–141.

923 42. Joho,Y., Vongsouthi,V., Spence,M.A., Ton,J., Gomez,C., Tan,L.L., Kaczmarski,J.A.,
924 Caputo,A.T., Royan,S., Jackson,C.J., *et al.* (2023) Ancestral Sequence Reconstruction
925 Identifies Structural Changes Underlying the Evolution of Ideonella sakaiensis
926 PETase and Variants with Improved Stability and Activity. *Biochemistry*, **62**, 437–
927 450.

928 43. Livada,J., Vargas,A.M., Martinez,C.A. and Lewis,R.D. (2023) Ancestral Sequence
929 Reconstruction Enhances Gene Mining Efforts for Industrial Ene Reductases by
930 Expanding Enzyme Panels with Thermostable Catalysts. *ACS Catal.*, **13**, 2576–2585.

931 44. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2023) REBASE: a database for DNA
932 restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research*,
933 **51**, D629–D630.

934 45. Arndt,D., Marcu,A., Liang,Y. and Wishart,D.S. (2019) PHAST, PHASTER and
935 PHASTEST: Tools for finding prophage in bacterial genomes. *Briefings in*
936 *Bioinformatics*, **20**, 1560–1567.

937 46. Hsiao,K., Zegzouti,H. and Goueli,S.A. (2016) Methyltransferase-Glo: a universal,
938 bioluminescent and homogenous assay for monitoring all classes of
939 methyltransferases. *Epigenomics*, **8**, 321–339.

940 47. Leger,A., Amaral,P.P., Pandolfini,L., Capitanchik,C., Capraro,F., Miano,V., Migliori,V.,
941 Toolan-Kerr,P., Sideri,T., Enright,A.J., *et al.* (2021) RNA modifications detection by
942 comparative Nanopore direct RNA sequencing. *Nat Commun*, **12**, 7198.

943 48. Mulroney,L., Birney,E., Leonardi,T. and Nicassio,F. (2023) Using Nanocompore to
944 Identify RNA Modifications from Direct RNA Nanopore Sequencing Data. *Current*
945 *Protocols*, **3**, e683.

946 49. Dominissini,D., Moshitch-Moshkovitz,S., Salmon-Divon,M., Amariglio,N. and
947 Rechavi,G. (2013) Transcriptome-wide mapping of N6-methyladenosine by m6A-seq
948 based on immunocapturing and massively parallel sequencing. *Nat Protoc*, **8**, 176–
949 189.

950 50. Meyer,K.D. (2019) DART-seq: an antibody-free method for global m6A detection. *Nat*
951 *Methods*, **16**, 1275–1280.

952 51. Garcia-Campos,M.A., Edelheit,S., Toth,U., Safra,M., Shachar,R., Viukov,S., Winkler,R.,
953 Nir,R., Lasman,L., Brandis,A., *et al.* (2019) Deciphering the "m6A Code" via
954 Antibody-Independent Quantitative Profiling. *Cell*, **178**, 731-747.e16.

955 52. Linder,B., Grozhik,A.V., Olarerin-George,A.O., Meydan,C., Mason,C.E. and Jaffrey,S.R.
956 (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the
957 transcriptome. *Nat Methods*, **12**, 767–772.

958 53. Meyer,K.D., Saletore,Y., Zumbo,P., Elemento,O., Mason,C.E. and Jaffrey,S.R. (2012)
959 Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and
960 near Stop Codons. *Cell*, **149**, 1635–1646.

961 54. Chen,Y., Davidson,N.M., Wan,Y.K., Patel,H., Yao,F., Low,H.M., Hendra,C., Watten,L.,
962 Sim,A., Sawyer,C., *et al.* (2021) A systematic benchmark of Nanopore long read
963 RNA sequencing for transcript level analysis in human cell lines. *bioRxiv*,
964 10.1101/2021.04.21.440736. 21 April 2021, pre-print: not peer-reviewed.

965 55. Gupta,Y.K., Chan,S.-H., Xu,S. and Aggarwal,A.K. (2015) Structural basis of asymmetric
966 DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat Commun*,
967 **6**, 7363.

968 56. Woodcock,C.B., Horton,J.R., Zhang,X., Blumenthal,R.M. and Cheng,X. (2020) Beta
969 class amino methyltransferases from bacteria to humans: evolution and structural
970 consequences. *Nucleic Acids Research*, **48**, 10034–10044.

971 57. Horton,J.R., Woodcock,C.B., Opot,S.B., Reich,N.O., Zhang,X. and Cheng,X. (2019) The
972 cell cycle-regulated DNA adenine methyltransferase CcrM opens a bubble at its DNA
973 recognition site. *Nat Commun*, **10**, 4600.

974 58. Osipiuk,J., Walsh,M.A. and Joachimiak,A. (2003) Crystal structure of MboIIA
975 methyltransferase. *Nucleic Acids Research*, **31**, 5440–5448.

976 59. Evans,R., O'Neill,M., Pritzel,A., Antropova,N., Senior,A., Green,T., Žídek,A., Bates,R.,
977 Blackwell,S., Yim,J., *et al.* (2022) Protein complex prediction with AlphaFold-
978 Multimer. *bioRxiv*, 10.1101/2021.10.04.463034. 4 October 2021, pre-print: not peer-
979 reviewed.

980 60. Thomson,R.E.S., Carrera-Pacheco,S.E. and Gillam,E.M.J. (2022) Engineering functional
981 thermostable proteins using ancestral sequence reconstruction. *Journal of Biological*
982 *Chemistry*, **298**, 102435.

983 61. Khersonsky,O. and Tawfik,D.S. (2010) Enzyme Promiscuity: A Mechanistic and
984 Evolutionary Perspective. *Annual Review of Biochemistry*, **79**, 471–505.

985 62. McMahon,A.C., Rahman,R., Jin,H., Shen,J.L., Fieldsend,A., Luo,W. and Rosbash,M.
986 (2016) TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets
987 of RNA-Binding Proteins. *Cell*, **165**, 742–753.

988 63. Stankevičius,V., Gibas,P., Masiulionytė,B., Gasiulė,L., Masevičius,V., Klimašauskas,S.
989 and Vilkaitis,G. (2022) Selective chemical tracking of Dnmt1 catalytic activity in live
990 cells. *Molecular Cell*, **82**, 1053-1065.e8.

991 64. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and
992 Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*,
993 **10**, 421.

994 65. Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large
995 sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

996 66. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N.,
997 Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for
998 integrated models of biomolecular interaction networks. *Genome Res*, **13**, 2498–2504.

999 67. Madeira,F., Park,Y.M., Lee,J., Buso,N., Gur,T., Madhusoodanan,N., Basutkar,P.,
1000 Tivey,A.R.N., Potter,S.C., Finn,R.D., *et al.* (2019) The EMBL-EBI search and
1001 sequence analysis tools APIs in 2019. *Nucleic Acids Res*, **47**, W636–W641.

1002 68. Nguyen,L.-T., Schmidt,H.A., von Haeseler,A. and Minh,B.Q. (2015) IQ-TREE: A Fast
1003 and Effective Stochastic Algorithm for Estimating Maximum-Likelihood
1004 Phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.

1005 69. Yang,Z. (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular*
1006 *Biology and Evolution*, **24**, 1586–1591.

1007 70. Nozaki,S. and Niki,H. (2019) Exonuclease III (XthA) Enforces In Vivo DNA Cloning of
1008 Escherichia coli To Create Cohesive Ends. *J Bacteriol*, **201**, e00660-18.

1009 71. Clifton,B.E., Alcolombri,U., Jackson,C.J. and Laurino,P. (2023) Ultrahigh-affinity
1010 transport proteins from ubiquitous marine bacteria reveal mechanisms and global
1011 patterns of nutrient uptake. *bioRxiv*, 10.1101/2023.02.16.528805. 16 February 2023,
1012 pre-print: not peer-reviewed.

1013 72. Wickham,H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New
1014 York.

1015 73. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of
1016 nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.*
1017 *U.S.A.*, **98**, 10037–10041.

1018 74. Holm,L. (2022) Dali server: structural unification of protein families. *Nucleic Acids*
1019 *Research*, **50**, W210–W215.

1020