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]*

[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.

*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 (RNMTs) could enable development of improved synthetic biology tools to manipulate RNA methylation, but is challenging due to limited understanding of structure-function relationships in RNMTs. 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 RNMT 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 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 (RNMTs) 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 RNMTs have been well characterized, and methylate mRNA and non-coding RNA molecules at a specific recognition sequence motif (2, 3). The m^6A modification is recognized by m^6A binding proteins, which thereby regulate RNA metabolism (4–6), pre-mRNA splicing (7–9), mRNA localization (10, 11), and translation (12–14). The m^6A modification is involved in development (15, 16) and stress response (17, 18), and is also linked to diseases such as cancer (12, 19, 20), infectious diseases (21), and type II diabetes (22). 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 RNMTs 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 distinguish the direct and indirect effects of m^6A dysregulation has remained a challenge (23, 24). 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 (25–27). These tools are based on fusion constructs of the eukaryotic m^6A RNMT METTL3 with RNA-binding proteins, such as dead Cas proteins (dCas9 or dCas13) (25, 26) or PUF proteins (27). 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 (28), the isolated protein has impaired methylation efficiency (26). 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 RNMT 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 RNMT that could fulfill these requirements. M.EcoGII was identified as a DNA methyltransferase from a prophage region of a pathogenic *Escherichia coli* strain (29), 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*) (30). Unusually, M.EcoGII exhibits non-site-specific methylation activity and can potentially methylate any adenine base in DNA and RNA (29). 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 (31, 32).

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 RNMTs, 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. Significant conformational changes that are expected to occur upon nucleic acid binding decrease the reliability of structural prediction tools such as AlphaFold2 for improving the function of the protein. On the other hand, screening or directed evolution of RNMTs 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 (33–36). Although the underlying reasons for the improved properties of reconstructed ancestral proteins are not entirely clear (37, 38), ASR has been successfully applied to improve the properties of various proteins, including those of relatively recent evolutionary origin (39, 40). As a protein engineering technique, ASR allows for the sampling of sequence space enriched in proteins with favorable properties based on phylogenetic information (37).

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 (SUPREM), showed a 12-fold increase in RNA methylation activity and an 8-fold increase in

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

Results

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

To reconstruct the sequences of ancestral M.EcoGII variants, we first collected a non-redundant set of homologous protein sequences of β-class methyltransferases, including M.EcoGII, from the Restriction Enzyme Database (REBASE) (41). A sequence similarity network of the resulting protein sequences showed that M.EcoGII and highly similar sequences clustered most closely with the site-specific DNA methyltransferase, M.PflMI (27.43% identity), followed by two sequence clusters exemplified by M.McrMPORF631P (21.63% identity) and M.Blo16BORF1919P (20.70% identity) (Supplementary Figure S1). For subsequent phylogenetic analysis, we focused on the 202 sequences that clustered together with these four sequences in the sequence similarity network. Analysis using the PHAST database showed that all genes of these proteins are encoded in phage-derivative regions of their respective host genomes (42).

A maximum-likelihood tree was constructed from these 202 sequences (Figure 1A: schematic tree, Supplementary Figure S2: full tree). Because M.PflMI is the most closely related enzyme to M.EcoGII that has confirmed site-specific methylation activity (recognition sequence: CCANNNNNTGG), we postulated that non-site-specific methylation activity emerged between M.EcoGII and the last common ancestor of M.EcoGII and M.PflMI, and that ancestral proteins between these two nodes might retain the desired non-site-specific methylation activity. We therefore selected six ancestral nodes, evenly spaced between M.EcoGII and the last common ancestor of M.EcoGII and M.PflMI, for experimental characterization (Figure 1A). The selected nodes were robustly supported by ultrahigh bootstrap values (>90%), and the Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) (>80%), and the corresponding ancestral sequences were also reconstructed with high

statistical confidence, with mean posterior probability values (>90%) (Supplementary Table S1).

These ancestral M.EcoGII sequences were cloned into the pET-28a(+) vector and expressed in the *E. coli* SHuffle T7 strain. Five of the six ancestral proteins could be expressed and purified. The ancestral proteins were expressed at higher levels than M.EcoGII (6-fold to 13-fold; Supplementary Figure S3A, Supplementary Table S1). In addition, the ancestral proteins showed higher thermostability than M.EcoGII, as measured by differential scanning fluorimetry. Whereas M.EcoGII exhibited a denaturation temperature (T_M) of 43 °C and would therefore have marginal stability at physiological temperatures, the ancestral proteins showed increases in T_M of 7–16 °C compared with M.EcoGII (Supplementary Figure S3B-C). Thus, ASR succeeded in improving the expression level and thermostability of M.EcoGII.

In vitro activity assays of ancestral M.EcoGII variants show improved catalytic efficiency for RNA methylation

To investigate the DNA and RNA methylation activity of M.EcoGII and the ancestral proteins, we used the MTase-Glo assay (43), a luminescence-based assay that can be used to quantify *S*-adenosyl homocysteine (SAH), a byproduct of the SAM-dependent methylation reaction. PCR-amplified DNA (3.5 kbp, 1841 adenine bases) or *in vitro* transcribed RNA (1.8 knt, 544 adenine bases) was used as a substrate for *in vitro* methylation reactions. All the ancestral proteins showed DNA and RNA methylation activity. There was limited variation in the DNA methyltransferase activities of the ancestral proteins, with two of them, Anc291 and Anc289, showing a small but significant increase in DNA methylation activity compared with M.EcoGII (Anc291, 2.4-fold, $P = 0.0099$; Anc289, 2.4-fold, $P = 0.0139$) (Figure 1B). Surprisingly, a much larger variation in RNA methyltransferase activity was observed, with one ancestral protein, Anc284, showing a 12-fold increase in RNA methylation activity compared with M.EcoGII ($P < 0.0001$; Figure 1C). Time-course DNA and RNA methylation assays confirmed that the product formation was linear over the course of these experiments and that Anc291 and Anc284 showed higher initial velocity for RNA methylation than M.EcoGII (Figure 1D, Supplementary Figure S4, Supplementary Table S3).

To detect and quantify the m^6A modification directly, we next performed LC-MS analysis. A short RNA substrate (294 bases) was methylated with M.EcoGII or Anc284, digested to nucleosides, and analyzed by LC-MS. Treatment of RNA with M.EcoGII and Anc284 resulted in the formation of m^6A, as shown by comparison of peak elution volume with an m^6A standard and detection of the m^6A $[M+H]^+$ ion by MS (Supplementary Figure S5),

165 indicating the presence of RNA m^6A methylation activity in Anc284. Furthermore, the fraction
166 of m^6A-methylated adenosine determined by peak area integration was significantly higher for
167 Anc284, compared with M.EcoGII (35.2 ± 3.3% vs. 72.4 ± 0.1 %; $P = 0.0001$ by unpaired t-
168 test). Altogether, these results show that Anc284 has substantially higher RNA methylation
169 activity than M.EcoGII. Thus, we coined Anc284 as SUPer Rna Ecogii Methyltransferase
170 (SUPREM).
171
172 **Methylation activity of SUPREM in human cells**
173 To check that M.EcoGII and SUPREM are able to methylate RNA in human cells,
174 M.EcoGII-eGFP, SUPREM-eGFP, and eGFP control constructs were transfected in HEK 293T
175 cells. To quantify the exogenous methyltransferase activity, m^6A methylation efficiency was
176 analyzed by immunofluorescence. M.EcoGII expression in cells only showed a trend towards
177 a 2.8-fold increased m^6A fluorescence in transfected cells by compared to eGFP control, which
178 was not significant (Figure 2A&B, Supplementary Figure S6), while SUPREM expression
179 resulted in a 5.4-fold increase, which was significant compared to both eGFP control and
180 M.EcoGII (Figure 2A&B, Supplementary Figure S6). Compared to the data obtained *in vitro*,
181 while both M.EcoGII and SUPREM were able to methylate RNA, M.EcoGII activity in cells
182 was relatively low compared to SUPREM, which exhibited a more robust methylation activity.
183 Thus, we have successfully engineered an RNMT that could be exploited for biological studies.
184
185 **The non-specificity in the methylation motifs is retained in SUPREM**
186 A decrease in the site-specificity of RNMTs is advantageous for engineering RNA
187 methylation states in that more RNA sequences can be targeted. M.EcoGII showed non-site-
188 specific methylation activity for DNA (29), while the specificity for RNA has not been
189 explored so far. To investigate the presence of non-site-specific RNA methylation activity by
190 M.EcoGII and ancestral proteins, we performed Oxford Nanopore direct RNA sequencing
191 (DRS) to characterize the site specificity of RNA methylation in M.EcoGII, Anc291, and
192 SUPREM. The RNA substrate (the same substrate used for MTase-Glo methylation assays;
193 1872 nt, 1468 adenine-containing 5-mer motifs) was methylated using each methyltransferase
194 and sequenced using an Oxford Nanopore MinION device. The Nanopolish and Nanocompore
195 pipeline was then used to identify modification sites (specifically, 5-mer nucleotide sequences
196 containing at least one m^6A modification) through comparison with an unmethylated control
197 (Figure 3A) (44).

198 SUPREM methylated a greater number of diverse motifs (523 motifs of adenine-
199 containing motifs / 623 motifs of total detected motifs; false positive rate (FPR): 16%) than
200 Anc291 (389 motifs / 472 motifs; FPR: 18%) and M.EcoGII (225 motifs / 247 motifs; FPR:
201 9%) (Figure 3A, B). Thirty-six percent of adenine-containing 5-mer motifs contained the m^6A
202 modification for SUPREM, compared to Anc291's 26% and M.EcoGII's 15%. Analysis of the
203 5-mer sequences modified by each methyltransferase showed that the methyltransferase
204 activity of M.EcoGII was largely non-site-specific, but displayed a slight bias against GC-rich
205 sequences (Figure 3B). On the other hand, this bias was reduced in SUPREM, with the
206 frequency of each base more similar to the expected value if all adenine bases were methylated
207 equally (Figure 3B). Thus, our Oxford Nanopore DRS analysis revealed the presence of a
208 certain degree of bias in M.EcoGII's methylation activity for RNA, while SUPREM showed
209 less bias against GC-rich motifs in its methylation activity on RNA.
210
211 **Investigation of key residues for improved RNA methyltransferase activity**
212 To investigate which amino acid residues are important for RNA methylation activity,
213 we compared multiple sequence alignments of sequences in the M.EcoGII clade (clade I) and
214 SUPREM clade (clade II) in our maximum-likelihood phylogenetic tree (Figure 4A). We chose
215 14 amino acids as candidate residues that might be responsible for the observed differences in
216 RNA methylation activity based on two conditions, being that the residues were highly
217 conserved among each clade, and that the residues differed between clades I and II. Fourteen
218 single-point mutants of SUPREM were constructed, of which 12 variants could be expressed,
219 and 11 variants could be successfully purified (the amino acid sequences of these variants are
220 given in Supplementary Table S4). Using the MTase-Glo assay to measure RNA methylation
221 activity, one substitution, E179Q, was found to significantly decrease the RNA methylation
222 activity of SUPREM (3.4-fold decrease, $P = 0.0079$) (Figure 4B), suggesting that this residue
223 is important for RNA binding and methylation.
224
225 **Structural analysis of M.EcoGII and its ancestral variants**
226 To gain a greater understanding of the structural basis for the non-site-specific
227 methylation activity of M.EcoGII and the differences in DNA and RNA methylation activity
228 between M.EcoGII and its ancestral variants, we aimed to generate structural models of
229 M.EcoGII, Anc291, and SUPREM. To date, experimental structures of M.EcoGII have not
230 been determined. We performed extensive crystallization screening of M.EcoGII, SUPREM,
231 and Anc291, but no crystals were obtained, most likely due to the conformational flexibility of

232 the long interdomain linker. Homology modeling was unfeasible because of the low sequence
233 identity (<20%) of M.EcoGII to proteins of known structure. Indeed, although the structure of
234 M.EcoGII was previously modeled using the structure of M.EcoP15I (PDB 4ZCF) as a
235 template, M.EcoP15I bears no recognizable sequence homology to M.EcoGII in the target
236 recognition domain (45). Based on these considerations, we decided to use AlphaFold2 for
237 structural modeling.

238 To this end, we first aimed to determine the oligomeric state of M.EcoGII and the
239 ancestral proteins. Although previous research suggested that M.EcoGII is a monomeric
240 protein based on native-mass spectrometry experiments (30), previously characterized β-class
241 methyltransferases have shown a dimeric structure (45–47), and monomeric structures appear
242 to be incompatible with catalytic activity (48). We used size-exclusion chromatography (SEC)
243 to estimate the molecular weight of M.EcoGII and the ancestral proteins based on a standard
244 curve and found that each protein eluted as a dimer (Figure 4C, Supplementary Figure S7).

245 Finally, structural models of M.EcoGII and ancestral protein homodimers were
246 constructed using AlphaFold2 multimer modeling (Figure 4D, Supplementary Figure S8) (49).
247 According to the structural prediction, each M.EcoGII monomer is composed of a Rossmann
248 fold catalytic domain (residues 1–134 and 260–349, mean pLDDT score 95.8) and a four-helix
249 bundle target recognition domain (residues 160–237, mean pLDDT score 94.4) containing a
250 helix-turn-helix motif. The two domains are connected by a long, flexible, and highly charged
251 linker, partly composed of a two-helix structure (mean pLDDT score 89.1 for the predicted
252 helical portion). A search of the Protein Data Bank for similar structures using the DALI server
253 showed that the predicted structure of the Rossmann fold domain of M.EcoGII shows the
254 highest similarity to other DNA methyltransferases such as M.MboIIA and M.CcrMI. In
255 contrast, the target recognition domain bears similarity to various helix-turn-helix domain
256 DNA-binding proteins (e.g., transcriptional regulators), but generally not to known
257 methyltransferase structures (Supplementary Table S5). The overall predicted structure of the
258 site-specific DNA methyltransferase M.PflMI, including its target recognition domain, is
259 similar to M.EcoGII, with the exception of the highly charged linker region, which is absent in
260 M.PflMI (Supplementary Figure S8E). The large positively charged surface of M.EcoGII
261 spanning the target recognition domain and interdomain linker is likely important for the non-
262 specific binding of DNA and RNA (Figure 4E), providing a large interaction surface that
263 effectively compensates for the lack of protein-DNA and protein-RNA interactions with
264 specific bases.

265 The ancestral proteins showed a similar overall predicted structure to M.EcoGII
266 (Supplementary Figure S8A). The E179 residue identified as being important for RNA binding
267 and/or methylation is located on the edge of the putative DNA/RNA binding pocket, based on
268 the structural models predicted by AlphaFold2 (Figure 4D). In addition, the amino acids (97-
269 110, 116-140) present on the interaction surface between the subunits are completely conserved
270 among M.EcoGII and ancestral proteins, suggesting that the dimer interface might be crucial
271 for these enzymes. Co-crystal structures of SUPREM and RNA would be required for a
272 detailed understanding of the mechanisms underlying the improved RNA methylation activity
273 and promiscuity compared with M.EcoGII.
274

275 **Discussion**

276 Growing recognition of the importance of RNA methylation, particularly m^6A
277 methylation, in a broad range of physiological and pathological processes has driven demand
278 for new tools for the manipulation and analysis of RNA methylation states (23, 24). M.EcoGII
279 was identified as a potentially useful tool to manipulate m^6A methylation in mammalian cells
280 based on its non-site-specific DNA/RNA methyltransferase activity and its non-mammalian
281 origin, but the enzyme has not yet been utilized for this purpose due to its low expression and
282 RNA methylation activity. In this work, we used ASR to engineer an M.EcoGII variant,
283 SUPREM, with a number of advantages over the wild-type enzyme, including a 12-fold
284 increase in RNA methylation activity, 8 °C increase in thermostability, an 8-fold increase in
285 expression, and highly relaxed sequence specificity, resulting in methylation of up to 72% of
286 adenine bases in single strand RNA. M.EcoGII and other RNMTs are generally challenging
287 targets for other protein engineering techniques such as rational design or directed evolution;
288 for example, the lack of experimental structural information on M.EcoGII and the complexity
289 of structure-function relationships in nucleic acid enzymes would have limited the potential of
290 rational protein engineering in this case. A computational approach for the engineering of
291 RNMT-RNA interaction would be challenging due to difficulty in predicting interactions
292 between proteins and RNA. In contrast, our work shows the power of sequence-based
293 engineering approaches such as ASR to enable rapid improvement of a broad range of protein
294 properties, including challenging cases where structural information is limited.

295 All of the ancestral variants of M.EcoGII characterized in this work showed higher
296 thermostability and expression than the wild-type protein, with ΔT_m between 7 and 16 ºC. It is
297 noteworthy that major improvements in thermostability were achieved despite the limited

298 phylogenetic diversity of the sequences in the M.EcoGII clade, which are mainly derived from
299 prophage regions in Enterobacteria genomes. Previous studies that have used ASR for protein
300 engineering have typically focused on more phylogenetically diverse and evolutionarily
301 ancient protein families (50), based on the hypothesis that the ancient progenitors of modern
302 proteins existed in thermophilic organisms and were therefore thermostable. In the case of
303 M.EcoGII, however, there is little reason to assume *a priori* that there would be a trend in
304 thermostability from the ancestor of the M.EcoGII clade towards the modern enzyme. Other
305 recent studies have similarly shown that it is unnecessary to reconstruct evolutionarily ancient
306 proteins to achieve substantial increases in thermostability by ASR (39, 40). For example,
307 Livada *et al*. systematically characterized all 56 ancestral nodes and 57 extant nodes in a
308 phylogeny of gene reductases and found that even the shallowest nodes (i.e., the most recent
309 ancestral nodes, adjacent to the extant nodes at the tips of the tree) had a T_M up to 16 °C higher
310 than their closest extant protein (40). The reason for this general trend of reconstructed
311 ancestral proteins showing higher thermostability compared with extant proteins is still unclear,
312 but the bias in maximum-likelihood ASR towards common amino acids (38) and consensus
313 amino acids (35) at ambiguously reconstructed positions may be partially responsible. The
314 ability of ASR to preserve coevolutionary relationships between residues also provides an
315 advantage over related methods such as consensus design (38). Overall, our work contributes
316 to a growing body of evidence that ASR can be successfully applied as a protein engineering
317 tool to a diverse range of protein families, even those with limited phylogenetic diversity.
318 Substantial differences in RNA methyltransferase activity were also observed among
319 the ancestral M.EcoGII variants, which likely reflects the neutral variation of the promiscuous
320 RNA methyltransferase activity of M.EcoGII throughout its evolutionary history. Although the
321 precise biological function of M.EcoGII is unknown, its DNA methyltransferase activity is
322 more likely to be biologically relevant, given that (1) phage DNA methyltransferases have a
323 general role in the protection of phage DNA from host restriction enzymes during infection
324 (50); (2) M.EcoGII shows higher DNA methyltransferase activity than RNA methyltransferase
325 activity (30); and (3) m^6A modification of mRNA has no known function in bacteria, although
326 it is known to occur in Gram-negative bacteria (18). Thus, we hypothesize that the DNA
327 methyltransferase activity of M.EcoGII represents its native activity and that its RNA
328 methyltransferase activity represents a promiscuous activity that is not itself under selection
329 pressure. In other words, we hypothesize that the RNA methyltransferase activity is a
330 byproduct of non-specific DNA methyltransferase activity that is driven mainly by non-
331 specific ionic interactions between the positively charged surface of M.EcoGII and the

332 negatively charged phosphate groups of DNA, which also enables binding of RNA. Consistent
333 with the fact that the promiscuous activities of enzymes generally show greater variability
334 among homologous enzymes than their native activities (51), the observed variation in RNA
335 methyltransferase activity among the M.EcoGII variants (~12-fold) is greater than the variation
336 in DNA methyltransferase activity (~3-fold). The fact that there is no trend in RNA
337 methyltransferase activity from the earlier ancestral proteins towards M.EcoGII is also
338 suggestive of neutral variation. Altogether, these considerations suggest that the ~12-fold
339 increase in RNA methyltransferase activity of M.EcoGII achieved by ASR can be attributed to
340 sampling of the neutral variation of RNA methyltransferase activity that occurred throughout
341 its evolutionary history. More generally, these results suggest that ASR can be used to obtain
342 enzyme variants with improved properties and increased activity, even if the desired activity is
343 not the native enzyme activity.

344 The combination of non-site-specific and highly efficient m^6A methylation activity
345 observed in SUPREM is unique among known RNMTs and should be broadly useful in
346 applications that require unbiased and efficient labeling of RNA. For example, SUPREM could
347 potentially be applied to proximity labeling for transcriptome-wide detection of RNA binding
348 sites for RNA-binding proteins. Fusion of a specific RNA-binding protein to SUPREM would
349 lead to increased m^6A methylation adjacent to the binding site of the RNA-binding protein,
350 allowing detection of the binding site using techniques for m^6A detection, including Oxford
351 Nanopore direct RNA sequencing. A similar approach, TRIBE, relies on fusion of an RNA-
352 binding protein to the RNA-editing enzyme adenosine deaminase (ADAR) rather than an
353 RNMT for proximity labeling of the RNA binding site (52). This method provides numerous
354 advantages over the widely-used cross-linking immunoprecipitation (CLIP) method for
355 detection of RNA-protein interactions, such as the relatively low amount of input material
356 needed; however, this method may also be susceptible to bias due to the inherent substrate
357 specificity of ADAR (52), a problem that could potentially be addressed using SUPREM.
358 SUPREM could also be applied to site-specific modification of RNA through conjugation of
359 SUPREM to sequence-programmable RNA-binding proteins, such as dCas13 and PUF
360 proteins (26, 27). Unlike previous iterations that rely on fusion proteins of the human m^6A
361 writer METTL3, SUPREM would also allow site-specific methylation of sites targeted by other
362 m^6A writers such as METTL16, as well as non-canonical methylation sites and sites that are
363 not naturally methylated (for example, to modulate mRNA stability, localization or translation
364 efficiency). Finally, the improved thermostability, expression, and RNA methyltransferase
365 activity of SUPREM make it a useful starting point for further protein engineering. For

366 example, engineering of the cofactor binding pocket of SUPREM to promote the binding of
367 synthetic, functionalized *S*-adenosyl-L-methionine analogs, an approach that has been applied
368 to sequence-specific DNA methyltransferases (53), would allow bioorthogonal sequence-
369 independent functionalization and labeling of mRNA. Altogether, we expect that SUPREM
370 will be a powerful enabling tool for the development of new methods for RNA manipulation
371 and analysis.
372
373 **Materials and methods**
374 **Phylogenetic analysis**
375 Protein BLAST was used to search the collection of amino acid sequences of type II class β
376 methyltransferases, to which M.EcoGII belongs, in the REBASE database (41) (retrieved on
377 2020-09-17) using M.EcoGII (GenBank accession number EGT69837.1) as a query sequence.
378 An e-value cutoff of 10^{-7} was used. Redundant sequences were omitted by CD-HIT using a
379 90% sequence identity threshold, giving a total of 2368 sequences (54). The sequence
380 similarity network was created by all-versus-all BLAST of the resulting sequences using an e-
381 value threshold of 10^{-30} and visualized in Cytoscape software. For ancestral sequence
382 reconstruction, multiple sequence alignments of the M.EcoGII-containing cluster (158
383 sequences) and the outgroup clusters (46 sequences) were constructed separately using
384 MUSCLE software (55). The aligned protein sequences were manually edited to remove two
385 outlier sequences, large insertions in the outgroup, and N-/C-terminal extensions. Each edited
386 multiple sequence alignment was combined by profile-profile alignment in MUSCLE. An
387 unrooted maximum-likelihood tree was constructed using IQ-TREE software (56). To assess
388 the robustness of the tree, inference of the phylogenetic tree was repeated five times using the
389 sample substitution model (LG+I+G4) or two other substitution models (WAG and VT). The
390 LG+I+G4 model was chosen as the best model based on IQ-tree calculation of Bayesian
391 information criterion (BIC) score. The robustness of each node was assessed using the
392 Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) and ultrafast bootstrap
393 analysis (UFboot) with 1000 replicates. Ancestral protein sequences were reconstructed using
394 the empirical Bayes method implemented in PAML using the LG substitution model (57). The
395 posterior probability distribution at each site for each ancestral node was also calculated using
396 PAML.
397
398 **Cloning and mutagenesis**

399 M.EcoGII and six ancestral protein sequences were reverse-translated into nucleotide
400 sequences that were codon optimized for expression in *E. coli*. A sequence encoding a C-
401 terminal His_6 tag with a linker sequence (N-GGAAALGHHHHHH-C) was added to each gene,
402 and the resulting DNA sequences were synthesized (Twist Bioscience, California, USA) and
403 cloned into the pET-28a(+) plasmid using In-Fusion cloning kit (Takara Bio Inc., Shiga, Japan)
404 or the iVEC3 method (58). In-frame cloning into the pET-28a(+) vector was confirmed by
405 DNA sequencing. Mutagenesis of SUPREM was performed using the PrimeStar Mutagenesis
406 method and PrimeStarMax polymerase (Takara Bio Inc., Shiga, Japan).
407 The DNA sequences corresponding to M.EcoGII and SUPREM were inserted into the
408 pCDNA3.1-eGFP-GSx3 plasmid using the In-Fusion cloning kit (Takara Bio Inc., Shiga,
409 Japan), resulting in a sequence encoding eGFP with a flexible linker sequence $(GGGGS)_3$
410 linked to the N-terminus of the methyltransferase.
411
412 **Protein expression**
413 Plasmids containing M.EcoGII and all ancestral protein genes were transformed into *E.*
414 *coli* strain SHuffle T7 (New England Biolabs Inc., Massachusetts, USA) and plated on LB agar
415 containing 30 μg/mL kanamycin and 2% glucose, incubating at 37 °C overnight. Colonies were
416 inoculated into 5-20 mL LB media with 30 μg/ml kanamycin and 2% glucose and incubated at
417 37 °C with shaking at 220 rpm overnight. The preculture was transferred into 0.5-2 L LB media
418 with 30 μg/mL kanamycin and 1% glucose, and the *E. coli* cells were grown at 37 °C with
419 shaking at 220 rpm until the OD_{600} reached 0.5-0.8. Protein expression of M.EcoGII and the
420 ancestral proteins was induced by 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4
421 h at 25 °C. SUPREM variants were expressed with 1 mM IPTG for 18 h at 18 °C. Cells were
422 harvested and stored at -20 °C before protein purification.
423
424 **Protein purification**
425 Cells were resuspended with lysis buffer (20 mM HEPES pH 7.0, 250 mM NaCl, 50
426 mM imidazole with 1× Protease Inhibitor Cocktail (EDTA-free) (Nacalai Tesque Inc., Kyoto,
427 Japan) and 1 mg/ml lysozyme (Sigma-Aldrich, Missouri, USA) and lysed by sonication. The
428 lysate was centrifuged at 10000 × *g*, 4 °C for 30 min, and the supernatant was filtered through
429 a 0.45 μm filter. The initial protein purification was performed using an ÄKTA Pure 150
430 system (Cytiva Japan, Tokyo, Japan) and a 1 ml HisTrap FF column (Cytiva Japan, Tokyo,
431 Japan) with Ni wash buffer (20 mM HEPES pH 7.0, 250 mM NaCl, 20 mM imidazole),
432 followed with elution buffer (20 mM HEPES pH 7.0, 250 mM NaCl, 200 mM imidazole) as a

433 stepwise purification. Peak fractions were analyzed by SDS-PAGE. Fractions enriched in the
434 target protein were pooled on ice, filtered through a 0.22 μm filter, and purified by size
435 exclusion column chromatography (HiLoad 26/600 Superdex 75 pg, Cytiva Japan,
436 Tokyo, Japan) with working buffer (20 mM HEPES pH 7.0, 250 mM NaCl). Peak fractions
437 were analyzed by SDS-PAGE. Fractions enriched in the target protein were pooled on ice,
438 filtered through a 0.22 μm filter, and concentrated in an Amicon Ultra 10 kDa molecular weight
439 cutoff (MWCO) centrifugal concentration device (Merck Millipore) with working buffer. The
440 concentrated protein samples (> 1 mg/ml) were mixed with glycerol to reach 10% (v/v)
441 concentration, quickly frozen by liquid nitrogen, and stored at -80 ºC.
442 For SUPREM variants, cells were resuspended with lysis buffer (20 mM HEPES pH
443 7.0, 250 mM NaCl, 20 mM imidazole with 1× Protease Inhibitor Cocktail (EDTA-free)
444 (Nacalai Tesque Inc., Kyoto, Japan) and 1 mg/ml lysozyme (Sigma-Aldrich, Missouri, USA)
445 and lysed by sonication. The lysis supernatant was obtained by centrifugation at 10000 × *g*,
446 4 °C for 30 min. SUPREM variants were purified by affinity chromatography in an open
447 column using cOmplete His-Tag Purification Resin (Roche, Basel, Switzerland), eluting with
448 20 mM HEPES pH 7.0, 250 mM NaCl, and 100 mM imidazole at room temperature. Each
449 fraction was analyzed by SDS-PAGE. Fractions enriched in the target protein were pooled,
450 filtered through a 0.22 μm filter, and concentrated in an Amicon Ultra 10 kDa MWCO
451 centrifugal concentration device (Merck Millipore) with working buffer. The concentrated
452 protein samples (>1 mg/ml protein and < 1 mM imidazole) were mixed with glycerol to reach
453 10% (v/v) concentration, quickly frozen by liquid nitrogen, and stored at -80 ºC.
454
455 **MTase-Glo assay**
456 The DNA substrate encoding M.EcoGII (3569 bp, containing 1819 adenine bases) was
457 amplified by PCR and purified using a NucleoSpin Gel and PCR Clean-up Kit (MACHEREY-
458 NAGEL GmbH & Co. KG, Düren, Germany) with Milli-Q water. The RNA substrate encoding
459 the firefly luciferase gene (1872 bases, 544 adenine bases) was prepared by *in vitro*
460 transcription using HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs Inc.,
461 Massachusetts, USA) with 1 unit Rnase Inhibitor (New England Biolabs Inc., Massachusetts,
462 USA) and purified in RNase-free water using a NucleoSpin RNA Clean-up spin column kit
463 with a DNA removal column (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany). *In*
464 *vitro* methylation experiments were performed for 10 min (DNA) or 1 h (RNA) at 37 °C in 8
465 μL reaction mixtures containing 1× CutSmart buffer (New England Biolabs Inc.,
466 Massachusetts, USA) with 20 nM DNA or 50 nM RNA with 1 unit RNase Inhibitor (New

467 England Biolabs Inc., Massachusetts, USA), 100 nM methyltransferase, and 20 μM SAM. The
468 methylation reaction in the 8 μL reaction was quenched by heating at 80 °C for 1 min at each
469 time point. The methyltransferase reaction product *S*-adenosylhomocysteine was quantified
470 using MTase-Glo kit (Promega, Wisconsin, USA) following the manufacturer's protocol (43).
471 To normalize the SAH production rate in DNA/RNA methylation, a calibration curve for SAH
472 was constructed in two independent experiments using from 0 to 5 μM SAH according to the
473 manufacturer's protocol of the MTase-Glo kit (Promega, Wisconsin, USA). The signal was
474 detected using a Tecan Spark microplate reader (Tecan Japan Co., Ltd, Japan).
475
476 **UPLC assay**
477 The RNA substrate, a transcribed portion of the pET-28a(+) vector (294 bases, 80
478 adenine bases), was prepared by *in vitro* transcription using HiScribe T7 High Yield RNA
479 Synthesis Kit (New England Biolabs Inc., Massachusetts, USA) with 1 unit RNase Inhibitor
480 (New England Biolabs Inc., Massachusetts, USA) and purified in RNase-free water using a
481 NucleoSpin RNA Clean-up spin column kit with a DNA removal column (MACHEREY-
482 NAGEL GmbH & Co. KG, Düren, Germany). *In vitro* RNA methylation experiments were
483 performed for 2 h at 37 °C in 20 μL reaction mixtures containing 1× CutSmart buffer (New
484 England Biolabs Inc., Massachusetts, USA) with 1 μg RNA and 1 unit RNase Inhibitor (New
485 England Biolabs Inc., Massachusetts, USA), 5 μg methyltransferase or no enzyme, and 320
486 μM SAM (New England Biolabs Inc., Massachusetts, USA). The methylation reaction was
487 quenched by heating at 80 °C for 1 min. RNA was purified using an RNA Clean &
488 Concentrator-5 kit (Zymo Research, California, USA) into 20 μL RNase-free water, and then
489 digested from RNA to nucleosides using a Nucleoside Digestion Mix kit (New England
490 Biolabs Inc., Massachusetts, USA) at 37 ºC for 1 h. The digestion reaction was quenched with
491 33% ethanol as the final concentration. The solvent was removed using a centrifugal evaporator,
492 and the sample was dissolved in 100 μL of 40 mM ammonium acetate pH 6.0. LC-MS analysis
493 was performed in triplicate on a Waters UPLC H-class system (Waters Corporation,
494 Massachusetts, USA) using an ACQUITY UPLC BEH C18 column (2.1 mm × 50 mm, 1.7
495 μm) with a buffer system consisting of 40 mM ammonium acetate pH 6.0 (A) and 40%
496 acetonitrile (B). The nucleosides were eluted at a flow rate of 0.25 ml/min with a linear gradient
497 of mobile phase of 0% B (0-1.5 min), 0-30% B (1.5-15 min), and 0% B (15-30 min) for re-
498 equilibration. The nucleoside peaks were compared between enzyme-treated samples and
499 standard compounds to identify each peak.
500

501 **Differential scanning fluorimetry assay**

502 The DSF experiments were conducted using the StepOnePlus real-time PCR instrument
503 (Applied Biosystems, Massachusetts, USA), as described previously (59). Reaction mixtures
504 were prepared in 96-well Real-Time PCR Plates, comprising 5× SYPRO Orange (Sigma-
505 Aldrich, Missouri, USA), 5 μM enzyme, and a total volume of 20 μL working buffer (20 mM
506 HEPES pH 7.0, 250 mM NaCl). The plate was sealed with optically clear sealing film and
507 centrifuged at 2000 × g for 1 min prior to loading into the real-time PCR instrument. The
508 temperature was ramped at a rate of 1% (approximately 1.33 °C/min) from 4 ºC to 95 ºC.
509 Fluorescence signals were monitored using the ROX channel. The melting temperature values
510 were determined by calculating the derivative of fluorescence intensity with respect to
511 temperature and fitting the resulting data to a quadratic equation within a 6 °C window around
512 the T_M, utilizing R software.

513

514 **Nanopore direct RNA sequencing**

515 For direct RNA sequencing by MinION sequencer (Oxford Nanopore Technologies,
516 Oxford, UK), *in vitro* RNA methylation experiments were performed for 2 h at 37 °C in 20 μL
517 reaction mixtures containing 1× CutSmart buffer with 5 μg RNA substrate (1.8 knt. Same
518 substrate of MTase-Glo experiment) and 1 unit RNase Inhibitor (New England Biolabs Inc.,
519 Massachusetts, USA), 1 μM methyltransferase or no enzyme as a negative control, and 320
520 μM SAM for two independent reaction mixture. The reaction mixture without enzyme was
521 used as a negative control. The polyadenylated tail was added by incubation with 5 U of poly-
522 A polymerase (New England Biolabs Inc., Massachusetts, USA) for 30 min at 37 °C. The
523 methylated and poly-A-tailed RNA samples were purified using an RNA purification kit
524 (Zymo Research, California, USA) with RNase-free water. The concentration and purity of
525 the RNA samples were measured using a Qubit 4 fluorometer with the Qubit RNA IQ kit
526 (Thermo Fisher Scientific, Massachusetts, USA) (RIN score > 8.5). The preparation of libraries
527 for direct RNA sequencing was conducted using the SQK-RNA002 kit from 500 ng of *in vitro*
528 transcribed RNA. The direct RNA sequencing was carried out on a flow cell R.9.4.1 (FLO-
529 MIN106, Oxford Nanopore Technologies, Oxford, UK) by the Sequencing Section of the
530 Okinawa Institute of Science and Technology, following the protocol provided by Oxford
531 Nanopore Technologies.

532

533 **Sequencing data analysis**

534 The data analysis was performed using the Nanocompore pipeline (44). Base-called
535 fastq files were created using Guppy software (version 6.4.2) and were concatenated into one
536 file. The transcript data were aligned to a reference transcript sequence by minimap2 (version
537 2.20), following sorting and indexing in samtools (version 1.15.1). The read file was indexed
538 and resquiggled using the index and eventalign commands in Nanocompore (44). The m^6A
539 peaks were detected using the eventalign_collapse and sampcomp commands in Nanocompore
540 (44). The detected peaks were visualized using ggplot2 (60), and methylated 5-mer motifs
541 containing adenine bases were extracted from the total detected motifs and visualized using an
542 R script and Prism 9 software (GraphPad Software).
543
544 **Size exclusion chromatography analysis**
545 Fifty to one hundred μg SUPREM or M.EcoGII samples were injected into a Superdex
546 200 Increase 10/300 GL column (Cytiva Japan, Tokyo, Japan), and eluted at 0.75 ml/min flow
547 rate with 1 column volume of working buffer (20 mM HEPES pH 7.0, 250 mM NaCl). A
548 calibration curve was constructed using the elution volumes of the LMW Marker Kit (Cytiva
549 Japan, Tokyo, Japan).
550
551 **Structural modeling and analysis**
552 Dimer structures of methyltransferases were modeled by Alphafold (ver.2.2.1) using
553 the multimer prediction option (49). Protein structural models were visualized using PyMOL
554 software. Structures similar to M.EcoGII and its target recognition domain were searched using
555 the DALI server (61).
556
557 **Immunofluorescence assay**
558 HEK 293T cells were grown in DMEM high glucose (Nacalai Tesque, 08458-45)
559 supplemented with 10% fetal bovine serum (FBS, Gibco, 102-70-106) and 100 μg/mL
560 primocin (InvivoGen, ant-pm-1). Cells were seeded at an appropriate confluency on polylysine
561 (Sigma, P4832) coated coverslips one day before JetPei® (Polyplus, 101000053) transfection
562 with the different constructs, following the manufacturer's recommendations. 48 hours after
563 transfection, cells were fixed for 30 minutes at room temperature (RT) in 4% paraformaldehyde
564 in PBS under a fume hood, then permeabilized in 0.3% Triton in PBS for 30 minutes at RT.
565 Cells were blocked with 1% (v/v) BSA in PBS for 1 hour at RT and incubated overnight at
566 4 °C with primary anti-m^6A antibody (Synaptic systems, 202-003) diluted in the blocking
567 solution. The next day, cells were incubated with secondary conjugated Alexa Fluor 647

antibody (Invitrogen, A21245) for 1 hour in the blocking solution, and then with NucBlue™ (Invitrogen, R37606) for 5 minutes in PBS. Slides were mounted in Fluoromount G™ (Invitrogen, 00-4958-02), and fluorescence was observed with Zeiss LSM 900 confocal microscope. m^6A immunofluorescence staining was quantified using the software ImageJ. A mask was made using the GFP fluorescence driven by the transfected construct and m^6A intensity was quantified inside the mask and averaged by the cell area. Statistical analysis was performed using the software GraphPad Prism. Statistical significance was evaluated using one-way ANOVA followed by *Tukey* posttest.

DATA AVAILABILITY

Raw fastq files of the Nanopore direct RNA sequencing were deposited in the National Center of Biotechnology Information (BioProject PRJNA990656). The pET28a-SUPREM (ID 205023), pET28a-AncMEcoGII291 (Anc291; ID 205024), and pET28a-AncMEcoGII250 (Anc250; ID 205025) plasmids will be made available via Addgene upon publication.

Author Contribution

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

Acknowledgements

We thank the Okinawa Institute of Science and Technology Sequencing Section for preparing the library and sequencing Oxford Nanopore MinION Sequencer. We are grateful for the help and support provided by the Scientific Computing and Data Analysis section of the Research Support Division at OIST.

Funding

Funding for the research was supported by Okinawa Institute of Science and Technology (OIST). PL acknowledges Kakenhi Grant (No. 90812256). 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.

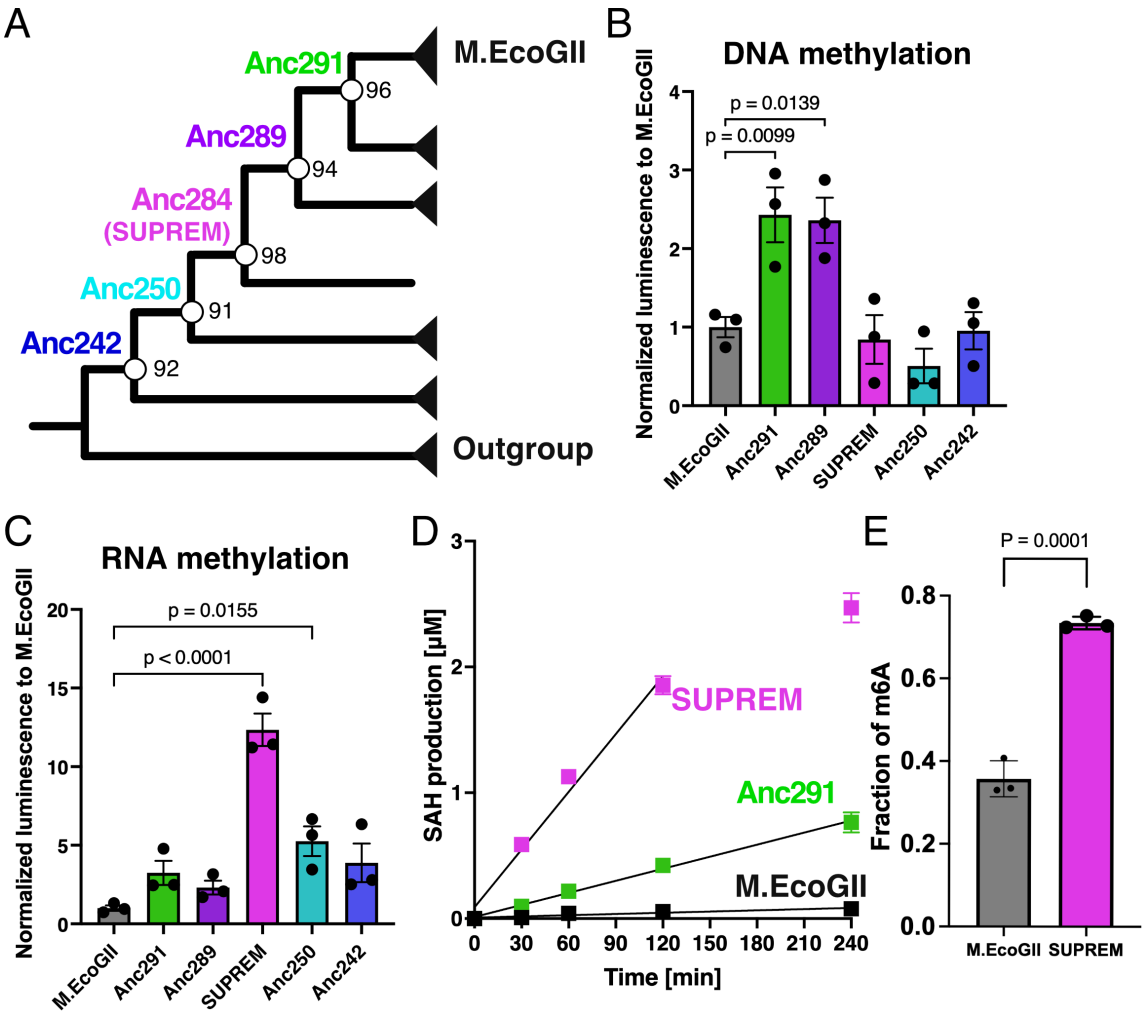

Figure 1. Ancestral sequence reconstruction of M.EcoGII. (A) Schematic maximum likelihood tree. The node values showed ultrafast bootstrap values calculated by IQ-TREE. **(B-C)** MTase-Glo assay of M.EcoGII and its ancestral proteins for (B) DNA and (C) RNA at a single time point. **(B)** The DNA methylation reaction was performed using 100 nM of methyltransferases, 50 nM DNA substrate (3.5 kbp containing 1844 adenine bases, Supplementary Table S2) and 10 μM SAM in 1× CutSmart buffer (NEB) at 37 °C for 10 min. **(C)** The RNA methylation reaction was performed using 100 nM of the corresponding methyltransferases, 25 nM RNA substrate (1.7 kb containing 544 adenine bases, sequence in Supplementary Table S2), and 10 μM SAM in 1× CutSmart buffer (NEB) at 37 °C for 60 min. Three independent experiments were performed, and each point showed an average of technical duplicates. *P* values were determined by the one-way ANOVA test with the Dunnett test. **(D)** Time course experiments for RNA substrates. The condition of RNA methylation reaction was the same as (C) except for reaction time. Three independent experiments were performed, and each point showed an average of technical duplicates. SAH production amount was calculated by SAH standard curve. The lines represent simple linear regression analysis

620 on the data from 0 to 120 min (SUPREM) or to 240 min (M.EcoGII, Anc291). **(E)** The fraction
621 of m^6A methylation in total adenine by UPLC analysis. The RNA methylation reaction was
622 performed using 5 μg of the corresponding methyltransferases, 1 μg RNA substrate (294 nt
623 containing 80 adenine bases, sequence in Supplementary Table S2) and 320 μM SAM (NEB)
624 in 1× CutSmart buffer (NEB) at 37 °C for 120 min. Each point showed three independent
625 methylation reactions. Error bars represent standard error of the mean (SEM).

626

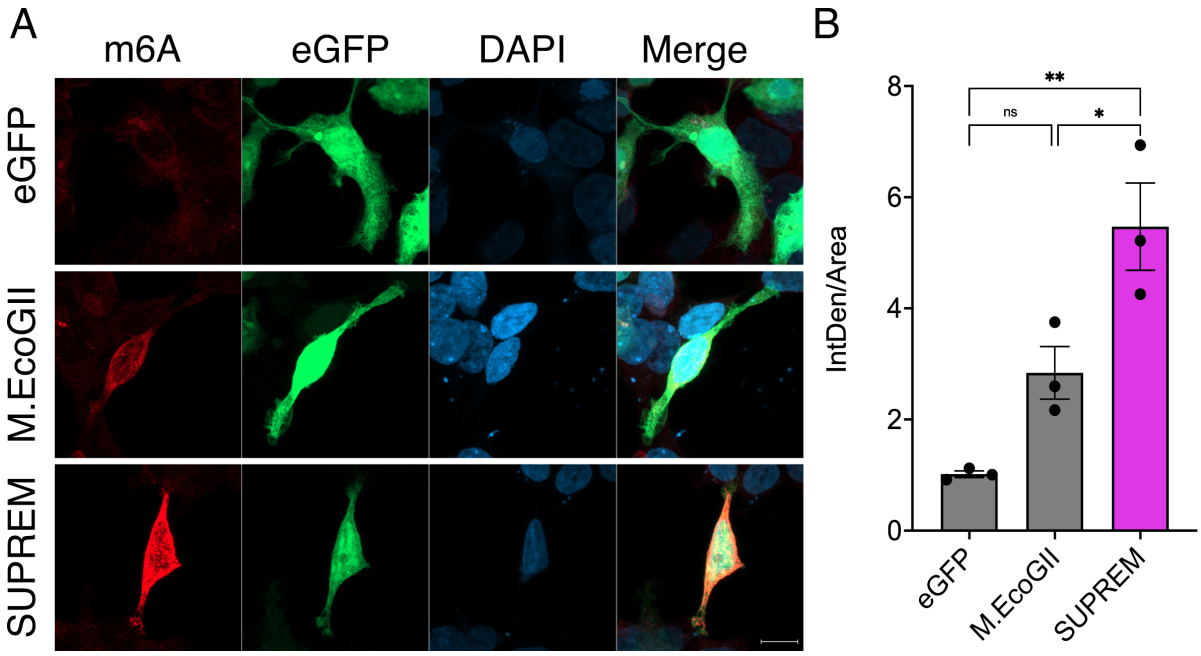

Figure 2. Methylation activity in HEK293 cell. (A) Representative images (63x objective with a 2x digital zoom) of HEK293T cells transfected with either eGFP, M.EcoGII-eGFP or SUPREM-eGFP constructs (reporter visible in green). m^6A modification were detected via immunofluorescence (in red) and the nucleus was stained with NucBlue™ (in blue). The scale bar is 10 µm. (B) Quantification of m^6A fluorescence intensity in HEK293 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 *(*: <0.0332; **: <0.0021).*

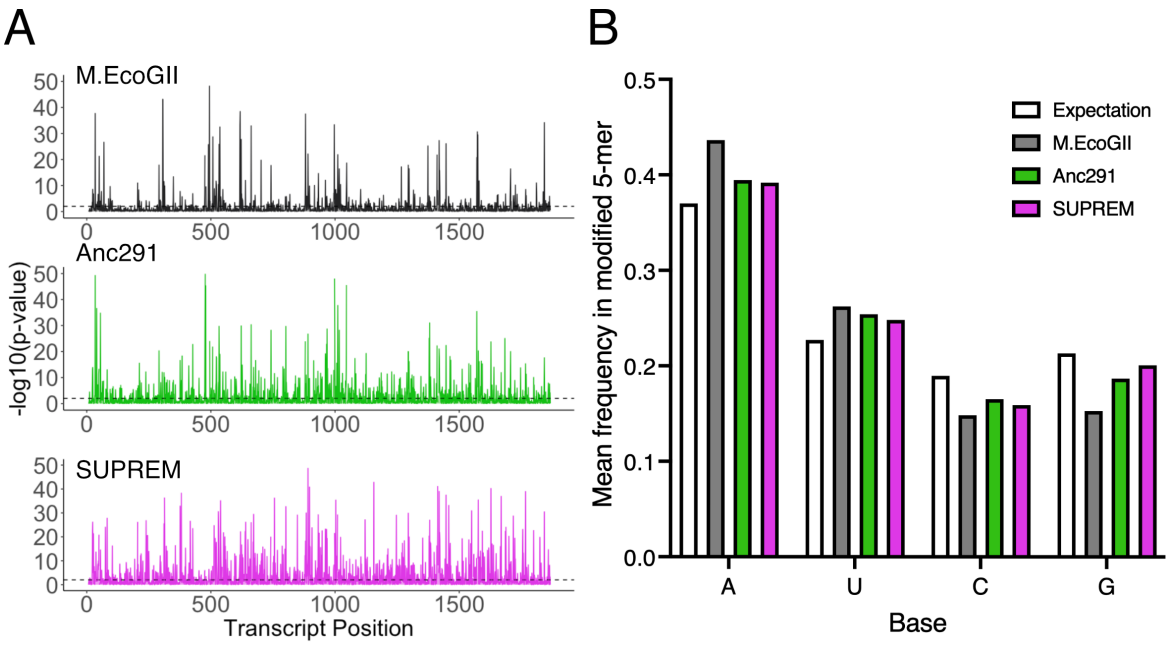

Figure 3. RNA methylation motif analysis. (A) Mapping detected RNA methylation sites on the 18 knt Fluc RNA sequence by MinION DRS and Nanocompore pipeline. **(B)** Mean RNA base frequency in modified 5-mer motifs detected by DRS. The "Expectation" represents the mean frequency of each base in 5-mers that contain at least one adenine within the RNA substrate.

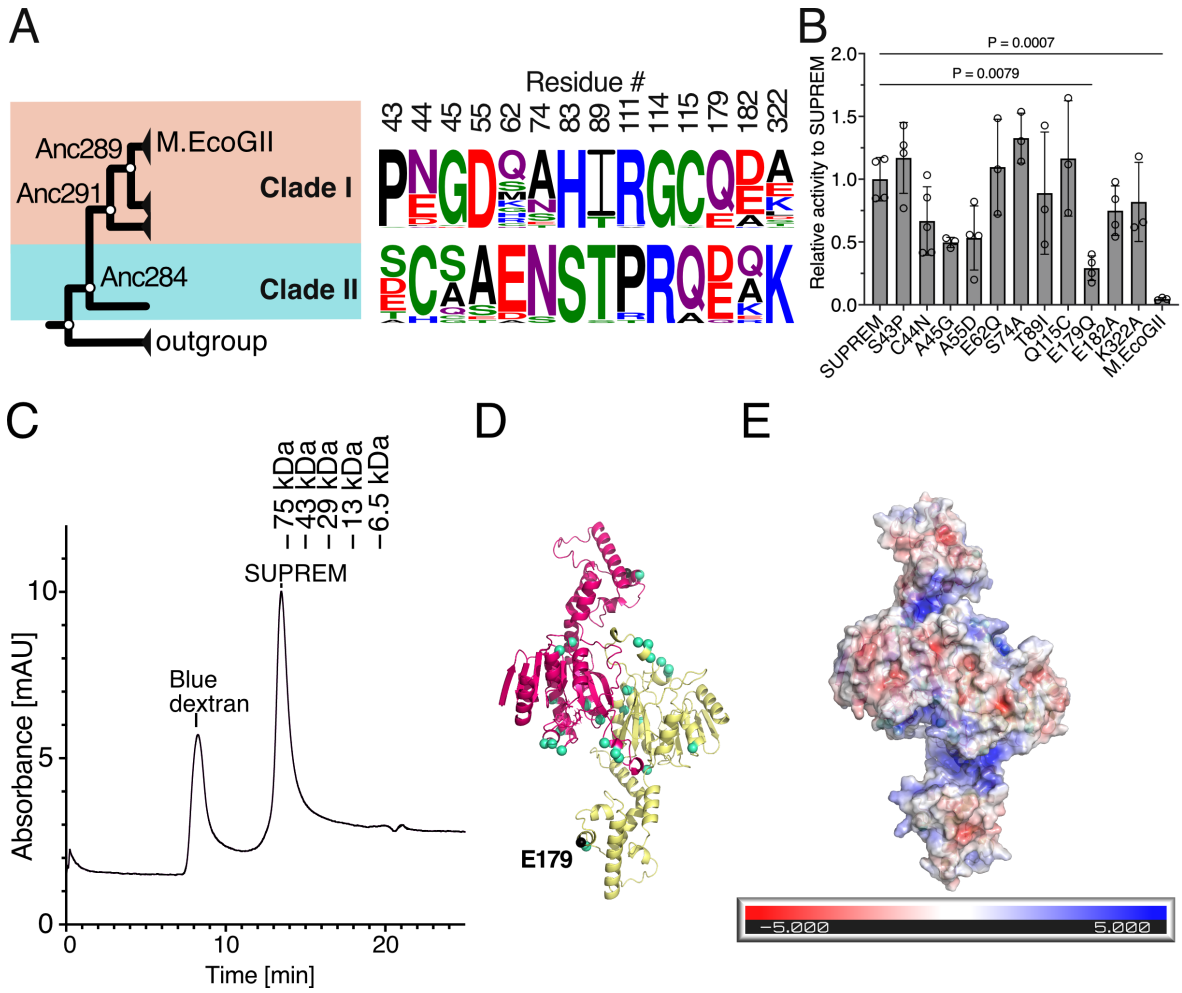

Figure 4. Mutational and structural analysis of SUPREM. (A) Candidate amino acid residues for mutational analysis. (Left) M.EcoGII clade (clade I, 46 sequences) and SUPREM clade (clade II, 19 sequences). (Right) Comparison of amino acid sequence logo of mutational candidate sites based on multiple sequence 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 (1.7 kb. 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

program (version 2.2.1). Green spheres show residues selected for mutational analysis (Fig. 3). **(E)** APBS calculation of the electrostatic surface potential of SUPREM (red: negative charge; blue: positive charge).

663 **References**

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

668 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
669 METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine
670 methylation. *Nat Chem Biol*, **10**, 93–95.

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

675 4. Wang,X., Lu,Z., Gomez,A., Hon,G.C., Yue,Y., Han,D., Fu,Y., Parisien,M., Dai,Q., Jia,G., *et al.*
676 (2014) N6-methyladenosine-dependent regulation of messenger RNA stability.
677 *Nature*, **505**, 117–120.

678 5. Geula,S., Moshitch-Moshkovitz,S., Dominissini,D., Mansour,A.A., Kol,N., Salmon-Divon,M.,
679 Hershkovitz,V., Peer,E., Mor,N., Manor,Y.S., *et al.* (2015) m6A mRNA methylation
680 facilitates resolution of naïve pluripotency toward differentiation. *Science*, **347**,
681 1002–1006.

682 6. Edupuganti,R.R., Geiger,S., Lindeboom,R.G.H., Shi,H., Hsu,P.J., Lu,Z., Wang,S.-Y.,
683 Baltissen,M.P.A., Jansen,P.W.T.C., Rossa,M., *et al.* (2017) N6-methyladenosine (m6A)
684 recruits and repels proteins to regulate mRNA homeostasis. *Nat Struct Mol Biol*, **24**,
685 870–878.

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

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

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

696 10. Yu,J., Chen,M., Huang,H., Zhu,J., Song,H., Zhu,J., Park,J. and Ji,S.-J. (2018) Dynamic m6A
697 modification regulates local translation of mRNA in axons. *Nucleic Acids Research*,
698 **46**, 1412–1423.

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

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

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

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

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

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

717 17. Zhou,J., Wan,J., Gao,X., Zhang,X., Jaffrey,S.R. and Qian,S.-B. (2015) Dynamic m6A mRNA
718 methylation directs translational control of heat shock response. *Nature*, **526**, 591–
719 594.

720 18. Deng,X., Chen,K., Luo,G.-Z., Weng,X., Ji,Q., Zhou,T. and He,C. (2015) Widespread
721 occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Research*, **43**,
722 6557–6567.

723 19. Liu,J., Eckert,M.A., Harada,B.T., Liu,S.-M., Lu,Z., Yu,K., Tienda,S.M., Chryplewicz,A.,
724 Zhu,A.C., Yang,Y., *et al.* (2018) m6A mRNA methylation regulates AKT activity to
725 promote the proliferation and tumorigenicity of endometrial cancer. *Nat Cell Biol*,
726 **20**, 1074–1083.

727 20. Zhang,C., Tunes,L., Hsieh,M.-H., Wang,P., Kumar,A., Khadgi,B.B., Yang,Y.-Y.,
728 Doxtader,K.A., Herrell,E., Koczy,O., *et al.* (2023) Cancer mutations rewire the RNA
729 methylation specificity of METTL3-METTL14. 10.1101/2023.03.16.532618.

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

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

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

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

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

743 26. Wilson,C., Chen,P.J., Miao,Z. and Liu,D.R. (2020) Programmable m6A modification of
744 cellular RNAs with a Cas13-directed methyltransferase. *Nat Biotechnol*, **38**, 1431–
745 1440.

746 27. Shinoda,K., Suda,A., Otonari,K., Futaki,S. and Imanishi,M. (2020) Programmable RNA
747 methylation and demethylation using PUF RNA binding proteins. *Chem. Commun.*,
748 **56**, 1365–1368.

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

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

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

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

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

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

767 34. Nakano,S., Niwa,M., Asano,Y. and Ito,S. (2019) Following the Evolutionary Track of a
768 Highly Specific l-Arginine Oxidase by Reconstruction and Biochemical Analysis of
769 Ancestral and Native Enzymes. *Applied and Environmental Microbiology*, **85**, e00459-
770 19.

771 35. Trudeau,D.L., Kaltenbach,M. and Tawfik,D.S. (2016) On the Potential Origins of the High
772 Stability of Reconstructed Ancestral Proteins. *Molecular Biology and Evolution*, **33**,
773 2633–2641.

774 36. Babkova,P., Sebestova,E., Brezovsky,J., Chaloupkova,R. and Damborsky,J. (2017)
775 Ancestral Haloalkane Dehalogenases Show Robustness and Unique Substrate
776 Specificity. *ChemBioChem*, **18**, 1448–1456.

777 37. Clifton,B.E., Kozome,D. and Laurino,P. (2022) Efficient Exploration of Sequence Space by
778 Sequence-Guided Protein Engineering and Design. *Biochemistry*,
779 10.1021/acs.biochem.1c00757.

780 38. Spence,M.A. (2021) Ancestral sequence reconstruction for protein engineers. *Current*
781 *Opinion in Structural Biology*.

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

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

789 41. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2022) REBASE: a database for DNA
790 restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research*,
791 10.1093/nar/gkac975.

792 42. Arndt,D., Marcu,A., Liang,Y. and Wishart,D.S. (2019) PHAST, PHASTER and PHASTEST:
793 Tools for finding prophage in bacterial genomes. *Briefings in Bioinformatics*, **20**,
794 1560–1567.

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

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

801 45. Gupta,Y.K., Chan,S.-H., Xu,S. and Aggarwal,A.K. (2015) Structural basis of asymmetric
802 DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat Commun*,
803 **6**, 7363.

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

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

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

812 49. Evans,R., O'Neill,M., Pritzel,A., Antropova,N., Senior,A., Green,T., Žídek,A., Bates,R.,
813 Blackwell,S., Yim,J., *et al.* (2022) Protein complex prediction with AlphaFold-
814 Multimer. 10.1101/2021.10.04.463034.

815 50. Thomson,R.E.S., Carrera-Pacheco,S.E. and Gillam,E.M.J. (2022) Engineering functional
816 thermostable proteins using ancestral sequence reconstruction. *J Biol Chem*,
817 10.1016/j.jbc.2022.102435.

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

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

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

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

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

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

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

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

838 59. Clifton,B.E., Alcolombri,U., Jackson,C.J. and Laurino,P. (2023) Ultrahigh-affinity transport
839 proteins from ubiquitous marine bacteria reveal mechanisms and global patterns of
840 nutrient uptake. 10.1101/2023.02.16.528805.

841 60. Wickham,H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New
842 York.

843 61. Holm,L. (2022) Dali server: structural unification of protein families. *Nucleic Acids*
844 *Research*, **50**, W210–W215.

845