

Supporting Information

Enzymatic characterization of three human RNA adenosine methyltransferases reveals diverse substrate affinities and reaction optima

Dan Yu¹, Gundeep Kaur¹, Robert M. Blumenthal², Xing Zhang^{1,*}, Xiaodong Cheng^{1,*}

¹Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

²Department of Medical Microbiology and Immunology, and Program in Bioinformatics, The University of Toledo College of Medicine and Life Sciences, Toledo, OH 43614, USA

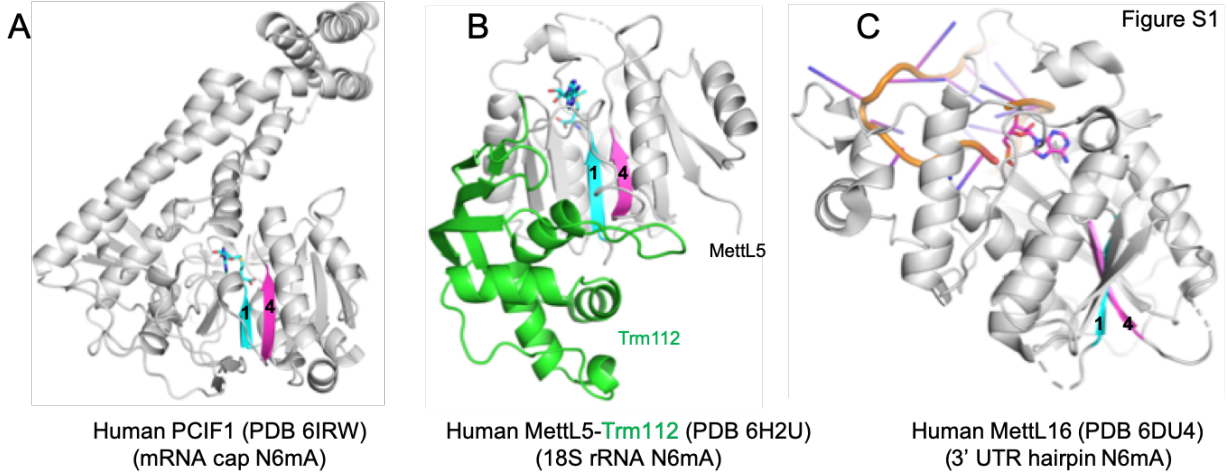
* Correspondence: XZhang21@mdanderson.org or XCheng5@mdanderson.org

Email addresses:

DY (DYu6@mdanderson.org); GK (GKaur1@mdanderson.org);

RMB (Robert.Blumenthal@utoledo.edu); XZ (XZhang21@mdanderson.org);

XC (XCheng5@mdanderson.org)



D Group α or γ MTases = NH₂ ···· β 1-2-3 ······ β 4-5-6-7 ···· COOH

	β 1	Motif I	β 4	Motif IV	substrate/product
E. Coli Dam	31	LVEP FV GAGSVFLNT 45.....175		SSVVYCD PPY APLSAT 190	DNA N6mA
MettL5	53	KVVAD LGG CGVLSI 67.....120		FDTVIM PPF GTKNNK 135	rRNA N6mA
MettL16	104	RRGID IGT GASCIYP 118.....178		YDFCMC PPF FANQLE 193	snRNA N6mA
HemK2	47	EICLE VGG SGVWSA 61.....116		VDLLV FNP PYVVTPPQ 131	protein Q/K(me1)
PCIF1/CAPAM	513	TFEC FAS PLNCYFKQ 527.....547		SGSFEAN PPF CEELMD 562	mRNA cap N6mA(m)

Mammal	PCIF1	Motif I	Motif IV
Platypus – <i>Ornithorhynchus anatinus</i>	XP_028904763	HRLFVGSFECF FAS PLNCYFRQYCSA	F SPLSGSFEAN PPF CEELMDAMVS
Opossum – <i>Monodelphis domestica</i>	XP_007474986	HRLFVGSFECF FAS PLNCYFRQYCSA	F YPLSGSFEAN PPF CEELMDAMVS
Elephant – <i>Loxodonta africana</i>	XP_023406075	HRLFVGSFECF FAS PLNCYFRQYCSA	F SPLSGSFEAN PPF CEELMD T VS
Armadillo – <i>Dasybus novemcinctus</i>	XP_012385012	HRLF S VSFECF FAS PLNCYFRQYCSA	F SPLSGSFEAN PPF CEELMDAMVS
Human – <i>Homo sapiens</i>	NP_071387	HRLFVGSFECF FAS PLNCYFRQYCSA	F APLSGSFEAN PPF CEELMDAMVS
Mouse – <i>Mus musculus</i>	NP_001361055	HRLFVGSFECF FAS PLNCYFRQYCSA	F TPLSGSFEAN PPF CEELMDAMVS
Mole – <i>Condylura cristata</i>	XP_004687490	HRLFVGSFECF FAS PLNCYFRQYCSA	F SPLSGSFEAN PPF CEELMDAMVS
Bat – <i>Pteropus vampyrus</i>	XP_011367609	HRLFVGSFECF FAS PLNCYFRQYCSA	F SPLSGSFEAN PPF CEELMDAMVS
Goat – <i>Capra aegagrus / C. hircus</i>	XP_017913438	HRLFVGSFECF FAS PLNCYFRQYCSA	F SPLSGSFEAN PPF CEELMDAMVS
Horse – <i>Equus caballus</i>	XP_023482399	HRLFVGSFECF FAS PLNCYFRQYCSA	F SPLSGSFEAN PPF CEELMDAMVS
Pangolin – <i>Manis pentadactyla</i>	XP_036758026	HRLF S VSFECF FAS PLNCYFRQYCSA	F SPLSGSFEAN PPF CEELMDAMVS
Wolf – <i>Canis lupus</i>	XP_005635184	HRLF S VSFECF FAS PLNCYFRQYCSA	F SPLSGSFEAN PPF CEELMDAMVS

Mammal	MettL5	Motif I	Motif IV
Platypus – <i>Ornithorhynchus anatinus</i>	XP_028927764	YDDIED KVVADLGG CGVLSIGSAM	MSKSFDTVIM PPF GTKHNKGMDM
Opossum – <i>Monodelphis domestica</i>	XP_001375717	YDDIENK VIGDLGG CGMLSIGAA	I SKSFDT I IM PPF GTKHNKGMDM
Elephant – <i>Loxodonta africana</i>	XP_010601144	YDDIENK VVADLGG CGVLSIGTAM	MSKSFDTVIM PPF GTKNNKGTDM
Armadillo – <i>Dasybus novemcinctus</i>	XP_004460848	YDDIENK VVADLGG CGVLSIGTAM	MSKSFDTVIM PPF GTKNNK I DI
Human – <i>Homo sapiens</i>	NP_001280115	YDDIENK VVADLGG CGVLSIGTAM	MSKSFDTVIM PPF GTKNNKGTDM
Mouse – <i>Mus musculus</i>	NP_083556	YDDIENK VVADLGG CGVLSIGAA	MSK L FDTVIM PPF GTKNNKGTDM
Mole – <i>Condylura cristata</i>	XP_004674616	YDDIENK VADLGG CGVLSIGTAM	I SKSFDTVIM PPF GTKNNKGTDM
Bat – <i>Pteropus vampyrus</i>	XP_023385363	YDDIENK VADLGG CGVLSIGTAM	MSKSFDTVIM PPF GTKNNKGTDM
Goat – <i>Capra aegagrus / C. hircus</i>	XP_005676044	YDDIENK VVADLGG CGVLSIGTAM	I SKSFDTVIM PPF GTKNNKGTDM
Horse – <i>Equus caballus</i>	XP_001498013	YDDIENK VVADLGG CGVLSIGTAM	MSKSFDTVIM PPF GTKNNKGTDM
Pangolin – <i>Manis pentadactyla</i>	XP_036743976	YDDIENK VVADLGG CGVLSIGTAM	MSKSFDTVIM PPF GTKN K GTDM
Wolf – <i>Canis lupus</i>	XP_005640349	YDDIENK VVADLGG CGVLSIGTAM	MSKSFDTVIM PPF GTKNNKGTDM

Mammal	MettL16	Motif I	Motif IV
Platypus – <i>Ornithorhynchus anatinus</i>	XP_028938135	S LKNTLRRGID IGT GASCIYPLLGA	SEI IYDFCMC PPF FANQLEAKGV
Opossum – <i>Monodelphis domestica</i>	XP_016284792	P ARSALRRGID IGT GASCIYPLLGA	SEI IYDFCMC PPF FANQLEAKGV
Elephant – <i>Loxodonta africana</i>	XP_003417034	S DRSTLRRGID IGT GASCIYPLLGA	SEI IYDFCMC PPF FANQLEAKGV
Human – <i>Homo sapiens</i>	NP_076991	SDKSTLRRGID IGT GASCIYPLLGA	SEI IYDFCMC PPF FANQLEAKGV
Mouse – <i>Mus musculus</i>	NP_080473	SDK T TLRRGID IGT GASCIYPLLGA	SEI IYDFCMC PPF FANQLEAKGV
Mole – <i>Condylura cristata</i>	XP_004684839	SDKSTLRRGID IGT GASCIYPLL G T	SEI IYDFCMC PPF FANQLEAKGV
Bat – <i>Pteropus vampyrus</i>	XP_011358139	SDK T TLRRGID IGT GASCIYPLL G S	SEI V YDFCMC PPF FANQLEAKGV
Goat – <i>Capra aegagrus / C. hircus</i>	XP_005693382	SDKSTLRRGID IGT GASCIYPLL G T	SEI IYDFCMC PPF FANQLEAKGV
Horse – <i>Equus caballus</i>	XP_014583405	SDKSTLRRGID IGT GASCIYPLL G T	SEI IYDFCMC PPF FANQLEAKGV
Pangolin – <i>Manis pentadactyla</i>	XP_036762024	SDKSTLRRGID IGT GASCIYPLL G T	SEI IYDFCMC PPF FANQLEAKGV
Wolf – <i>Canis lupus</i>	XP_022279183	SDKSTLRRGID IGT GASCIYPLL G T	SEI IYDFCMC PPF FANQLEAKGV

Figure S1. The seven-stranded class-I methyltransferase domain in (A) PCIF1 (PDB 6IRW), (B) MettL5-Trm112 (PDB 6H2U), and (C) MettL16 (6DU4), showing the standard central

topological switch-point, seen in this family of MTases, between strands $\beta 1$ (cyan) and $\beta 4$ (magenta). **(D)** The relative sequence positions, of conserved motif I (located in the carboxy end of strand $\beta 1$) and motif IV (located in carboxy end of strand $\beta 4$), are similar to those of *E. coli* DNA adenine methyltransferase (Dam) and human HemK2, a protein glutamine and lysine MTase. The common feature of the respective substrates is the amino group (NH_2) of DNA/RNA adenine or protein glutamine or lysine (in the deprotonated state), at which all three methylation reactions are catalyzed by a NPPF/Y version of motif IV. **(E)** Structure-guided alignment shows that PCIF1 has a somewhat unusual version of motif I (SAM binding, in this case lacking the typical GxG element) that is highly conserved among mammals. The MettL5 and MettL16 motifs I (DxGxG) and IV are well conserved across the Mammalia. Not shown is *D. novemcinctus* (armadillo) MettL16 homolog (XP_023441943), which may start from the wrong (downstream) ATG; the rest of the putative amino-terminal sequence is not available. Variations from the human sequence are highlighted by cyan background.

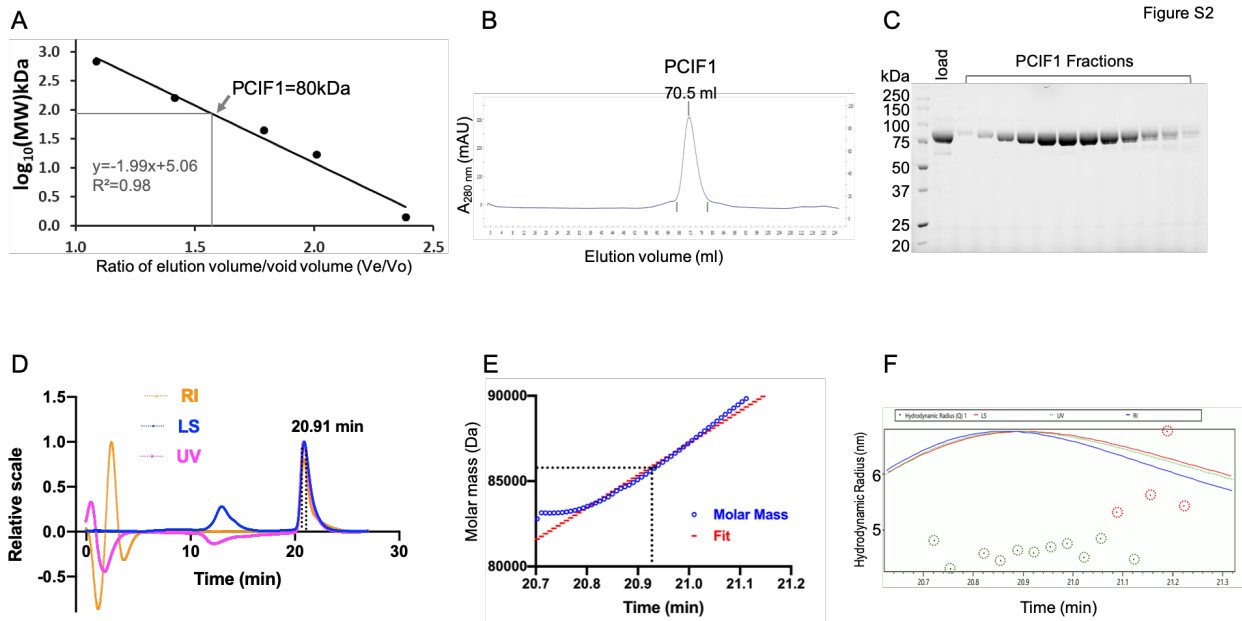


Figure S2. Purified recombinant PCIF1 used in this study. **(A)** A calibration curve for a Superdex 200 16/60 column based on five standards. The estimated apparent molecular weight of PCIF1 is 80 kDa. **(B)** Elution profile of PCIF1 and **(C)** peak fractions (shown in 12% SDS PAGE) from gel filtration column (Superdex S200 16/60). **(D)** SEC-MALS profile of PCIF1. Elution profile from the Shodex KW803 column. Peak at 20.91 min corresponding to PCIF1. **(E)** The observed molecular weight of PCIF1 is 86 kDa (Y-axis) at 20.91 min (X-axis). **(F)** Plot of hydrodynamic radius in circles (Y-axis) versus time of elution (X-axis). The observed hydrodynamic radius is 4.8 ± 0.4 nm.

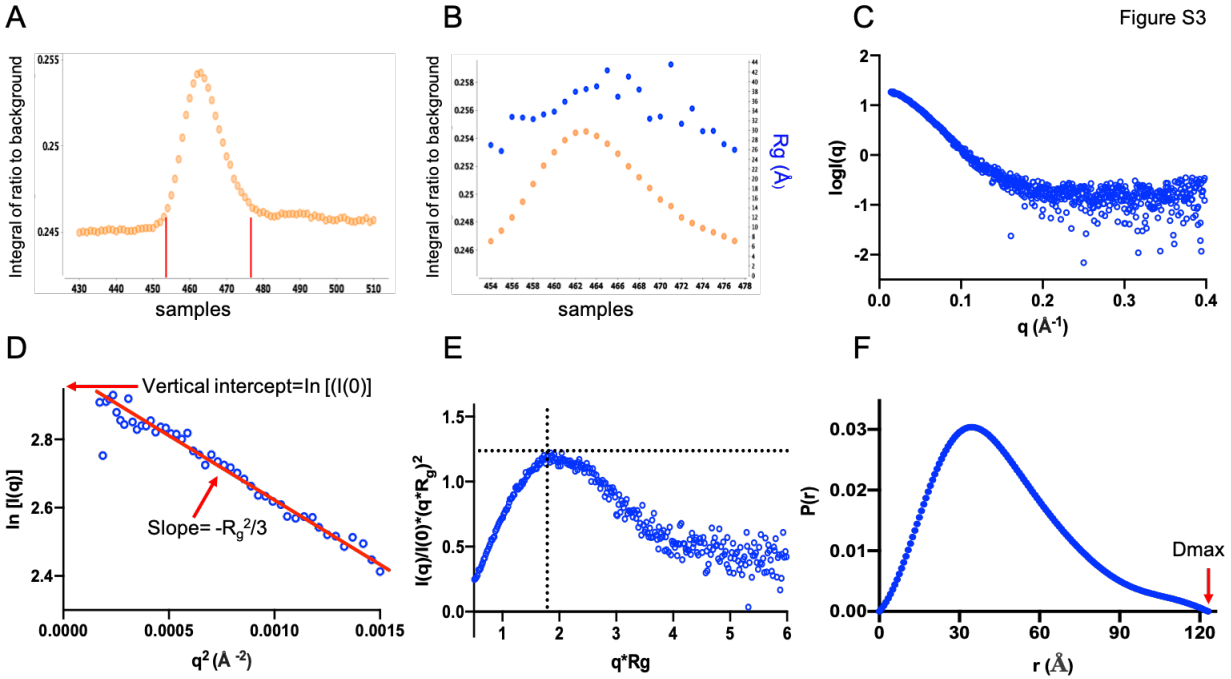


Figure S3. SEC-SAXS analysis of PCIF1. **(A)** In line SEC trace showing PCIF1 elutes as a single peak with no apparent signs of aggregation. **(B)** The blue dots show corresponding values of radius of gyration (R_g) across the elution peak. **(C)** Experimental scattering curve plotted as $\log I(q)$ versus q in units of inverse angstroms after background-subtracted from the buffer scattering ($q=4\pi\sin\theta/\lambda$ where 2θ is defined as the scattering angle and λ is the wavelength in \AA of the incoming X-ray beam). **(D)** Guinier plot as a linear fit between $\ln(I)$ versus q^2 . The R_g value ($\sim 36 \text{ \AA}$) is derived from the slope ($= -R_g^2/3$). The zero-angle scattering intensity $I(0)$ value (20.81) is derived from the vertical intercept. **(E)** Dimensionless Kratky plot is presented as $(qR_g)^2 I(q)/I(0)$ versus qR_g . The bell-shaped peak at low q suggests that PCIF1 is well-folded and globular in solution. The Porod volume is $\sim 125,875 \text{ \AA}^3$ and molecular weight is $\sim 86 \text{ kDa}$. **(F)** Normalized pair distance distribution function $P(r)$, generated from the Fourier transform of $I(q)$, represents a continuous electron-pair distances (r) in the protein. The maximum dimension of the protein D_{max} is estimated to be 121 \AA .

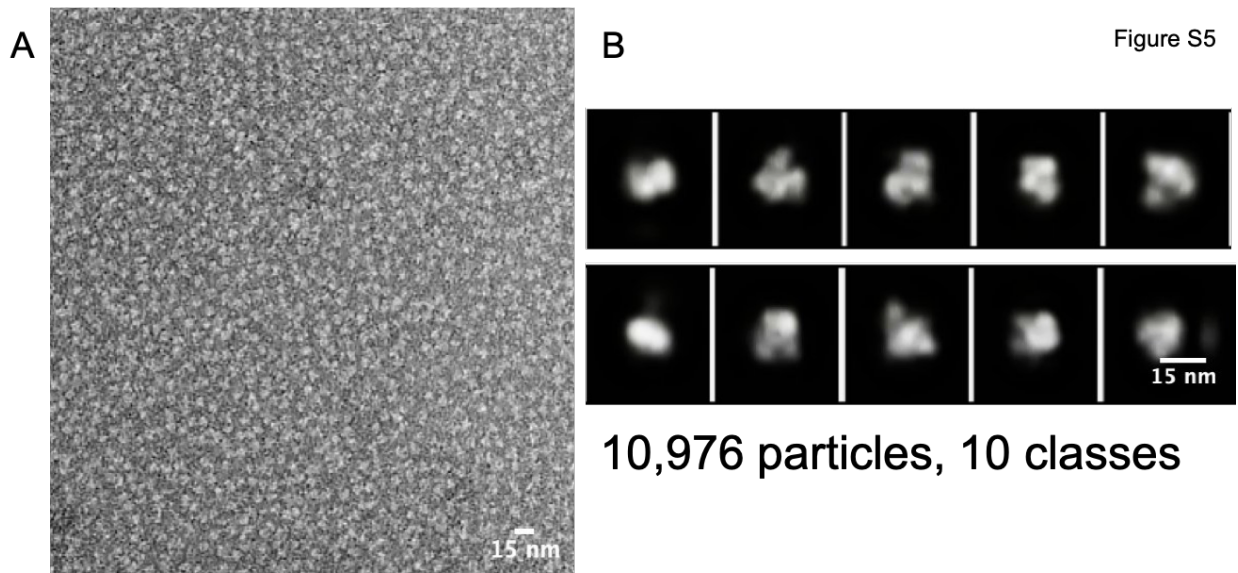


Figure S4. EM analysis of full length PCIF1. **(A)** A representative electron micrograph of PCIF1 negatively stained with 2% uranyl acetate and recorded at 40,000x. The scale bar is 15 nm. **(B)** Ten different 2D class averages generated from 10,976 particles.