DNA-cleavage prokaryotic Argonaute prediction based on sequence analysis

基于序列分析预测DNA切割活性的原核Argonaute蛋白

PKU16F G09 张心怡 岳頔 何苑 20170114



Outline

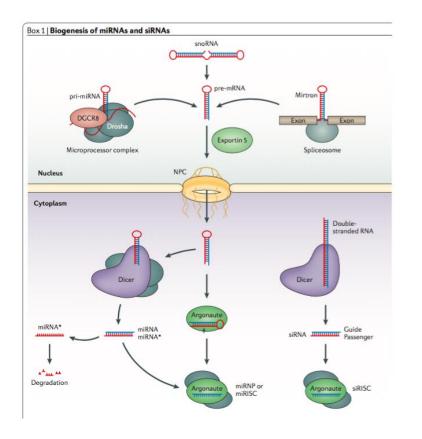
- Background
- Sequence analysis of *Tt*Ago and *Pf*Ago
- Phylogenetic analysis of Argonaute Proteins
- Structure prediction of DNA-cleavage Argonaute protein



Background

Functions of Argonaute proteins

Ago protein is the key player in eukaryotic RNA interference (RNAi) pathways



Argonaute protein is a member of RNAinduced silencing complex (RISC)

Small RNAs guide Argonaute proteins to their specific targets through sequence complementarity, which then leads to mRNA cleavage or translation inhibition.

Gunter Meister. Nature Reviews. 2013



> Functions of Argonaute proteins

Prokaryotic Argonaute proteins participate in host defense by DNA interference

DNA-guided DNA interference by a prokaryotic TtAgo Argonaute

Daan C. Swarts¹*, Matthijs M. Jore¹*, Edze R. Westra¹, Yifan Zhu¹, Jorijn H. Janssen¹, Ambrosius P. Snijders², Yanli Wang³, Dinshaw J. Patel⁴, José Berenguer⁵, Stan J. J. Brouns¹ & John van der Oost¹

Nature 2014.3

Argonaute of the archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that targets cognate DNA

Daan C. Swarts¹, Jorrit W. Hegge¹, Ismael Hinojo¹, Masami Shiimori², Michael A. Ellis², Justin Dumrongkulraksa², Rebecca M. Terns², Michael P. Terns² and John van der Oost^{1,*}

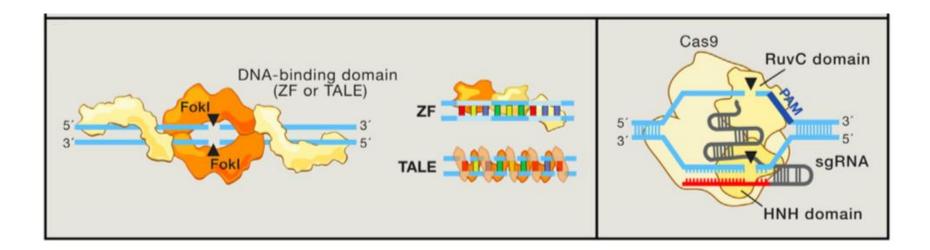
¹Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, 6703 HB Wageningen, The Netherlands and ²Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, USA

*Pf*Ago

NAR 2015.4



Several genome editing technologies



- ♦ ZFN and TALEN: Protein-DNA binding
- ♦ CRISPR/Cas9: RNA-DNA binding
- ♦ New gene editing tool?
- \diamond DNA-DNA binding?

Hsu et al., Cell, 2014

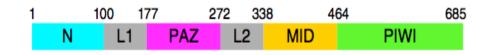


Comparison of *Tt*Ago and *Pf*Ago

	TtAgo	<i>Pf</i> Ago
Organism	Thermus thermophilus	Pyrococcus furious
Requirement	Mn ²⁺ or Mg ²⁺ ; 75°C	Mn ²⁺ or Co ²⁺ ; 50-250 mM NaCl; 87-99.9° C
Catalytic tetrad	DEDD	DEDH
Guide	5'-phosphated ssDNA (13-25 nt) 5'end:C (1 st position) A (2 nd position)	5'-phosphated ssDNA 15-31 nt
Target	ssDNA; dsDNA	ssDNA; dsDNA



Crystal structure of *Tt*Ago ternary complex



PAZ

Domain architecture of *Tt*Ago

5'-phosphorylated and 3'-hydroxyl ends of guide strands are anchored within MID and PAZ domain

The cleavage activity of Ago resides in the RNase H fold adopted by the PIWI domain

Sheng et al. PNAS. 2014

Swarts et al. Nat Struct Mol Biol. 2014



- Sequence analysis of *Tt*Ago and *Pf*Ago
- PSI-BLAST of *Tt*Ago and *Pf*Ago

Sequence	TtAgo/ <i>Pf</i> Ago
Database	Refseq protein
Expect threshold	0.00001
Word size	3
Matrix	BLOSUM45
Gap Costs	Existence 11, Extension
PSI-BLAST threshold	0.0001

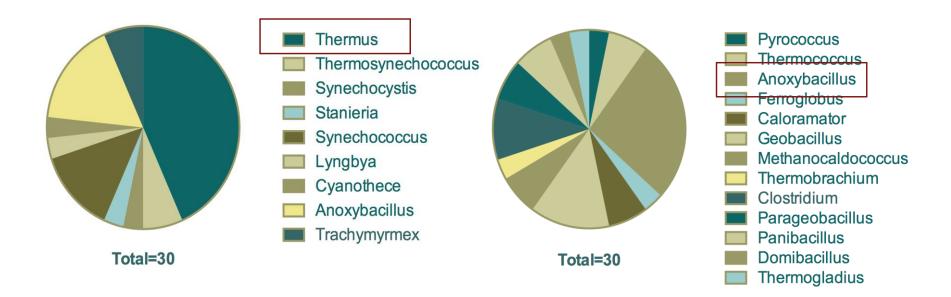
3



PSI-BLAST of *Tt*Ago and *Pf*Ago

TOP 30 hits of *Tt*Ago

TOP 30 hits of PfAgo



Thermus 栖热菌属

Anoxybacillus 厌氧芽胞杆菌属



- Phylogenetic analysis of Argonaute Proteins
- Alignment of *Tt*Ago and *Pf*Ago

Species/Abbrv	Group Name								*							*	7	1	*																\square
1. WP_011174533.1 hypothetical protein Thermus thermophilus		AGD	ΤΡL		EAL	AH	QI	FH	I L T	RL	ΥP	A S	G F	A F	PR	LF	P A F	P L F	HL,	A D	RL	<mark>v</mark> k	ΕV	G R	LG		-	RH	I L I	ΚE	V D	RE	KL	FF	V
2. WP_038030409.1 hypothetical protein Thermus sp. NMX2.A1		AGD	ΤΡL		ΕAL	AH	QI	F F	I L T	RL	ΥP	A S	GΕ	ΑF	ΡR	LF	P <mark>A</mark> F	۲L	ΗL	A D	R L	VΚ	ΕV	G R	L G		- 1	RΗ	I L I	ΚE	۷D	RE	ΚL	. F F	V
3. WP_060384876.1 hypothetical protein Thermus parvatiensis		AGD	ΤΡL		ΕAL	AH	QI	F F	IL T	R L	ΥP	A S	GΕ	ΑF	ΡR	LF	P <mark>A</mark> F	۲L	ΗL	A D	R L	<mark>v</mark> k	ΕV	G R	L G		- 1	R٢	I L I	ΚE	۷D	R E	ΚL	. F F	V
4. WP_011229221.1 hypothetical protein Thermus thermophilus YP_1		AGD	ΤΡL		ΕAL	AH	QI	FH	IL T	R L	ΥP	A S	GΕ	ΑF	ΡR	LF	P <mark>A</mark> F	P L F	ΗL	A D	R <mark>L</mark>	<mark>v</mark> k	ΕV	G R	LG		- 1	RΗ	I L I	ΚE	V D	R E	ΚL	. F F	V
5. WP_008631444.1 hypothetical protein Thermus parvatiensis		AGD	ΤΡL		ΕAL	AH	QI	F F	I L T	RL	ΥP	A S	GΕ	ΑF	ΡR	LF	P <mark>A</mark> F	L	ΗL	A D	R L	<mark>v</mark> k	ΕV	G R	LG		- 1	RΗ	I L I	ΚE	V D	R E	ΚL	. F F	V
6. WP_014630542.1 hypothetical protein Thermus thermophilus		AGD	ΤΡL		ΕAL	AH	QI	F F	I L T	RL	ΥP	A S	GΕ	ΑF	ΡR	LF	P A F	۲L	ΗL	A D	R L	<mark>ν</mark> κ	ΕV	G R	L G		- 1	RΗ	I L I	ΚE	۷D	RE	ΚL	. F F	V
7. WP_039457454.1 hypothetical protein Thermus sp. 2.9		AGD	ΤΡL		ΕAL	AH	QI	F H	I L T	RL	ΥP	A S	GΕ	ΑF	ΡR	LF	P A F	۲L	ΗL	A D	R L	νĸ	ΕV	G R	L G		-	RΗ	I L I	ΚE	۷D	RE	ΚL	. F F	V
8. WP_071678161.1 hypothetical protein Thermus brockianus		EGE	ΤΡL		ΕAL	. A E	QI	ΥF	IL T	RL	ΥP	A S	GΕ	ΑF	ΡR	LF	P <mark>A</mark> F	۲L	ΗL	A D	R L	<mark>ν</mark> κ	ΕV	G R	L G		- 1	RH	I L I	ΚE	۷D	RE	ΚL	. F F	V
9. WP_014514637.1 hypothetical protein Thermus sp. CCB_US3_UF1		AGE	ΤΡL		EΕΝ	/ A R	QV	/YF	IL T	R L	ΥP	P S	GΕ	LΡ	ΡR	LF	P <mark>A</mark> F	P L F	ΗL	A D	R <mark>L</mark>	<mark>v</mark> k	ER	A R	V G		- 8	FF	P Q I	3 R	I Q	ΡS	SL	. F F	V
10. WP_038066338.1 hypothetical protein Thermus filiformis		EGN	ΤΡL		V D L	AL	QI	YF	IL T	RL	ΥP	A S	GΕ	LΡ	P R	LF	P <mark>A</mark> F	L	ΗL	A D	R L	V R	ΕV	GR	LG		- 1	RH	I L I	ΚE	V P	R D	R L	. F F	V
11. WP_011056792.1 hypothetical protein Thermosynechococcus elo		YGN	ΤΡL		ELL	AL	Q T	ΥH	I L T	Q L	ΗP	A S	GΕ	RS	CR	LF	۷ W	/LH	ΗL	A D	RS	s K	E F	Q R	I G		QI	SL	. L	аN	V D	R E	ΚL	. I A	۷ ،
12. WP_041429921.1 hypothetical protein Thermosynechococcus sp.		YGN	ΤΡL		ELL	. A L	QT	ΓY Η	I L T	' Q L	ΗP	A S	G F	s s	CR	LF	۷ w	/ L I	ΗL	A D	R S	s K	E F	Q R	I G		QI	S L	. L I	ΩN	I D	RE	ΚL	. I A	v ،
13. WP_038044516.1 hypothetical protein Thermus scotoductus		EGP	T P V		EEL	AR	QI	ΥH	IL S	R L	ΥP	P S	GΥ	R F	P S	LF	P <mark>A</mark> F	۲L	ΗL	A D	R L	V R	ΕV	G R	V G		- L	S 8	BL I	H G	L D	RE	ΚL	. F F	V
14. WP_028947682.1 hypothetical protein Synechocystis sp. PCC 671		YGN	T S L		D L L	AL	Q T	TY G	ιL Τ	Q L	ΗP	A S	GΕ	R S	CR	LF	۷ W	/ L	ΗL	A D	RS	s K	E F	Q R	L G		QI	SI	L	Q N	I S	RE	ΚL	. I A	ίV
15. WP_041619610.1 hypothetical protein Stanieria cyanosphaera		YGD	A P L		EIL	AR	QI	DF	<mark>≀ L</mark> C	LL	N P	A S	GΥ	sv	SR	LF	۷Y	/ 1	ΗF	A D	ΚM	<mark>a</mark> k	ΤV	Q R	I G		ΕV	G ۱	/ L	Q G	I D	RQ	ΚI	FF	А
16. WP_041438497.1 hypothetical protein Synechococcus sp. JA-3-3A		HGK	T P L		E V L	. A L	Q T	ΥH	IL S	B Q L	ΗP	A S	GΕ	Q A	CR	LF	۷ W	/L	ΗF	A D	ĸs	G K	E F	Q R	LG	DN	IIF	SI	L	Q N	I D	RQ	ΚL	- I A	١V
17. WP_011244830.1 hypothetical protein Synechococcus elongatus		YGS	T D L		ELL	. A L	Q T	Y H	I L T	' Q <mark>L</mark>	ΗP	A S	GΕ	R S	CR	LF	۶W V	/L	ΗL	A D	R S	s K	E F	Q R	I G		QI	S١	/ L	ΩN	I S	R D	ΚL	. I A	ίV
18. WP_011378069.1 MULTISPECIES: hypothetical protein Synechoco		YGS	TDL		ELL	. A L	Q T	ΓY Η	I L T	' Q <mark>L</mark>	ΗP	A S	GΕ	R S	CR	LF	۰ W V	/L	ΗL	A D	RS	s K	E F	Q R	I G		QI	S١	/ L	ΩN	I S	R D	ΚL	. I A	ίV
19. WP_011434286.1 hypothetical protein Synechococcus sp. JA-2-3E		HGN	ΤΡL		E V L	- A 1	Q T	ΓY Η	IL S	6 Q L	ΗP	A S	G F	Q A	CR	LF	۰ w	/ L H	ΗF	A D	ĸs	s K	E F	Q R	I G	ΝS	I F	S١	/ L	Q N	ΙN	RE	ΚL	. I A	١V
20. WP_039723517.1 hypothetical protein Lyngbya confervoides		HGD	A S L		R L L	_ A N	ΕI	F A	L S	G Q F	ΗP	<mark>A</mark> S	ΑF	R S	S R	LF	P <mark>M</mark> F	۲L	ΗY	A D	R <mark>M</mark>	ΙK	ΕV	Q R	L G		QL	. S I	L	H G	I D	RQ	ΚL	. F F	A
21. WP_013324485.1 hypothetical protein Cyanothece sp. PCC 7822		TGD	A P L		E V L	_ A Q	QI	DF	L T	ML	N P	A S	GΥ	FΥ	SR	LF	^P M V	/L	Η F /	A D	ΚM	<mark>a</mark> k	ΕV	Q R	LG		Q	A F	Ľ	аĸ	V E	R E	ΚI	FF	А
22. WP_015717132.1 piwi domain-containing protein Thermus scoto		A G E	V D L	N L	E E A	A A R	Q L	.Y⊦	IL S	R I	ΥP	GS	GΥ	R F	ΡR	LF	P <mark>A</mark> F	۲L	H M I	V D	R <mark>M</mark>	V R	ΕV	A R	L G		- 0	SH	I N I	L R	L K	ΕE	QL	FF	L
23. WP_042896060.1 hypothetical protein Anoxybacillus sp. BCO1		ТКС	LSI	R -	- D \	/ V S	DA	<mark>v y</mark> k	< L S	F M	ΗI	нs	- M	LΚ	TR	LF	P I T	I V F	ΗY	A D	- L	s s	ΤF	ΗN	RG		- L	. I H	I P I	R S	VН	E Q	ΑL	P F	۷
24. WP_035063661.1 MULTISPECIES: hypothetical protein Anoxybacil		ткс	LSI	R -	- D \	/ V S	DA	<mark>x</mark> Y k	< L S	F M	ΗI	нs	- M	LΚ	TR	LF	P I T	I V F	ΗY	A D	- <mark>L</mark>	s s	ΤF	ΗN	R <mark>G</mark>		- <mark>L</mark>	. I H	I P I	RS	VН	E Q	ΑL	. P F	۷
25. WP_055441081.1 stem cell self-renewal protein Piwi domain-cont		тас	LSI	R -	- D I	I V S	DA	<mark>v Y</mark> k	< L S	FM	ΗI	нs	- M	L K	TR	LF	P I T	I V F	ΗY	A D	- <mark>L</mark>	s s	ΤF	ΗN	R <mark>G</mark>		- <mark>L</mark>	. I H	I P I	RS	VН	E Q	ΑL	P F	۷
26. WP_035065747.1 hypothetical protein Anoxybacillus gonensis		тас	LSI	R -	- D I	I V S	DA	<mark>v Y</mark> k	< L S	FM	ΗI	нs	- M	L K	TR	LF	P I T	V	ΗY	A D	- <mark>L</mark>	s s	ΤF	ΗN	R G		- <mark>L</mark>	. I F	I P I	RS	VН	E Q	ΑL	P F	V
27. XP_018345347.1 PREDICTED: protein aubergine-like isoform X2		FDN	T G L	EV	D I N	/ Q V	ΥT	T <mark>Y</mark> k	< L T	H M	ΥY	NC	SТ	ΤV	R -	VF	P <mark>A</mark> F	P C F	ΗY	A H	K L	S F	LΥ	G R	FL		- F	I Q F	PP	N S	QL	EK	QL	FF	L
28. XP_018302941.1 PREDICTED: protein piwi-like Trachymyrmex zet		FDN	T G L	E A	EII	I Q V	V T	T Y k	< L T	H M	ΥY	NC	SТ	τV	R -	VF	P <mark>A</mark> F	P C F	ΗY	A H	K L	S F	LΥ	G R	FL		- H	QF	PP	٩S	QL	ER	QL	FF	L
29. WP_026011736.1 hypothetical protein Anoxybacillus kamchatkens		TQC	LSI	R -	- D I	I V S	DA	<mark>v y</mark> k	< L S	FM	ΗI	H S	- M	LK	TR	LF	^р I Т	V	ΗY	A D	- L	s s	ΤF	ΗN	R <mark>G</mark>		- <mark>L</mark>	.	I P I	R S	VН	EQ	ΑL	P F	۷

Alignment of TOP 30 hits of *Tt*Ago by MEGA7 ClustalW



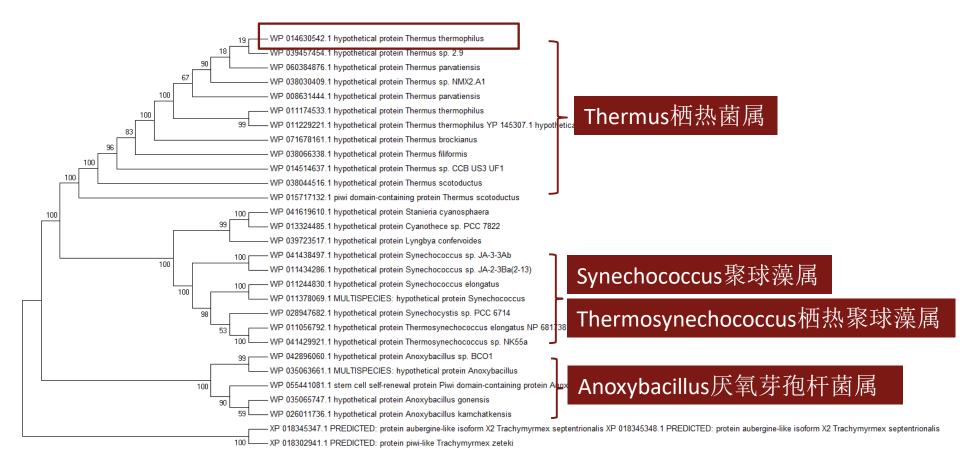
Alignment of *Tt*Ago and *Pf*Ago

Species/Abbrv	Group Name									* *	*											*					
1. WP_011011654.1 hypothetical protein Pyrococcus furiosus		ΚF	ΚE	FNI	I K L	D N	<mark>K</mark>	ΚI	LLL	R D	G R	I T N	I N E	ΕE	GL	< Y T	SEM	F D	ΙEV	V T I	ΜDΝ	/ I K	ΝH	P <mark>V</mark> R	A F	A N -	MKN
2. WP_055429304.1 hypothetical protein Thermococcus thioreducens		Q F	GΚ	FGI	I D L	EG	<mark>K</mark>	S I	LIL	. R D	GK	I T K	CD E	ΕE	GLA	V Y T	SKV	FG	IKI	ТΤ	FN	I V K	RН	LLR	I F	A N -	RK
3. WP_055441081.1 stem cell self-renewal protein Piwi domain-containing protein		S I	FS	YEK	< <mark>m</mark> Y	GΜ	k p <mark>a</mark>	нι	TFH	I R D	G - '	VCF	E D	LD	H L 1	R <mark>Y</mark>	M Q S	FQ	I P F	DF	V E I	ιк	RS	R R R	M A	V Y T	NG
4. WP_026011736.1 hypothetical protein Anoxybacillus kamchatkensis		S I	FS	YE	< L Y	GΜ	k p <mark>a</mark>	нι	TFH	I R D	G - 1	VCF	E D	LD	H L 1	R <mark>Y</mark>	M Q E	FQ	I P F	DF	V E I	I I K	RΡ	R R R	M A	V Y T	NG
5. WP_032099642.1 hypothetical protein Anoxybacillus flavithermus		S I	FS	YE	< M Y	GΜ	K P <mark>A</mark>	ΗI	T F F	I R D	G - 1	V C F	E D	L D	H L 1	R <mark>Y</mark>	<mark>M</mark> Q S	FQ	I P F	DF	V E I	I I K	RS	R R R	MA	V Y T	NG
6. WP_035065747.1 hypothetical protein Anoxybacillus gonensis		S I	FS	YE	< M Y	GΜ	K P <mark>A</mark>	ΗI	T F F	I R D	G - 1	V C F	E D	L D	H L 1	R <mark>Y</mark>	<mark>M</mark> Q S	FQ	I P F	DF	V E I	I I K	RΡ	R R R	M A	V Y T	NG
7. WP_012575214.1 hypothetical protein Anoxybacillus flavithermus		S I	FS	YEk	< M Y	GΜ	K P <mark>S</mark>	ΗI	T F F	I R D	G - 1	VCF	ED	LΑ	H L 1	° Q Y	M Q S	FQ	I P F	DF	V E I	I I K	ΚP	R R R	MA	V Y K	(NK)
8. WP_035063661.1 MULTISPECIES: hypothetical protein Anoxybacillus		S I	FS	YE	< M Y	GΜ	k p <mark>a</mark>	нι	T F F	I R D	G - 1	V C F	E D	L D	H L 1	R Y	<mark>M</mark> Q S	FQ	I P F	DF	ΙEΙ	I I K	RΡ	R R R	MA	V Y T	NG
9. WP_042896060.1 hypothetical protein Anoxybacillus sp. BC01		S I	FS	YE	< M Y	GΜ	K P <mark>A</mark>	ΗI	T F F	IR D	G - 1	VCF	ED	LD	H L 1	R <mark>Y</mark>	M Q S	FQ	I P F	DF	ΙEΙ	I I K	RΡ	R R R	MA	V Y T	NG
10. WP_012966655.1 hypothetical protein Ferroglobus placidus		ΚK	(K <mark>T</mark>	Y N I	I D F	EN	<mark>K</mark>	S I	LIL	R D	<mark>G</mark> R	I N K	EE	I N	QLN	1 E F	S E E	R N	C <mark>R I</mark>	ΤY	ΙE	I R K	NI	V H Q	FL	V N S	SQ
11. WP_027308646.1 hypothetical protein Caloramator sp. ALD01		11	LS	Y D F	R K F	KQ	КРК	ΗI	VIF	I R D	G -	FAR	E N	I N	ΝYΙ	ΝY	FSG	KN	IKF	DI	VE	/ K K	QG	A V K	FA	IKD	K E
12. WP_073249391.1 hypothetical protein Caloramator proteoclasticus		1.1	LS	YEF	r k <mark>f</mark>	NQ	КРК	ΗI	VIE	I R D	G -	FAR	R E N	I N	ΝYΙ	ΝY	FSG	ΚN	IKF	DI	۷E	/ K K	QG	A V K	FA	IKD	K D -
13. WP_014195559.1 hypothetical protein Geobacillus thermoleovorans		S I	ΥS	YEG	2 I Y	GΑ	K P R	нι	ΤFΗ	I R D	G -	I C F	E D	L D	FLC	A Y	L R S	FQ	I P F	DF	V E	ГК	ΚP	RRR	MA	I Y S	NK
14. WP_020279644.1 MULTISPECIES: hypothetical protein Geobacillus		S I	ΥS	YEG	a I Y	G A	K P R	нι	ΤFΗ	I R D	G -	I C F	E D	L D	FLG	A Y	L R S	FQ	I P F	DF	V E I	ГК	ΚP	R R R	MA	I Y S	N K
15. WP_015791216.1 hypothetical protein Methanocaldococcus fervens			- K	T D I	I N M	E N	K	ΝL	LFL	R D	GF	VQN	I S E	R E	EL	K K L	SKE	LN	S N I	ΕV	IS	I R K	ΝN	к <mark>ү</mark> к	V F	T S D	YG
16. WP_018663454.1 hypothetical protein Thermobrachium celere		1.1	LS	F E 1	ΓKF	NQ	ΡΡK	нι	VIE	I R D	G -	FAF	E D	I N	ΝY	< Q Y	FQS	ΚN	IKF	DI	VΕV	/ ĸ ĸ	Q G .	A V K	FA	ткк	G N -
17. WP_049625402.1 stem cell self-renewal protein Piwi domain-containing protei		S I	ΥS	YEG	2 I Y	GΑ	K P R	ΗI	ΤFΗ	I R D	G -	IWF	E D	L D	FLG	A Y	L R S	FQ	I P F	DF	V E	ГК	ΚP	RRR	MA	I Y S	N K
18. WP_064221217.1 stem cell self-renewal protein Piwi domain-containing protei		S I	ΥS	YEG	2 I Y	GΑ	K P R	ΗI	ΤFΗ	I R D	G -	ICF	E D	L D	FLG	A Y	L R S	FQ	I P F	DF	VE	IIK	ΚP	R R R	MA	I Y S	N K
19. WP_023817613.1 hypothetical protein Geobacillus genomosp. 3		S I	ΥS	YEC	ΣTΥ	GΑ	KPR	ΗI	ΤFΗ	I R D	G -	ICF	E D	LD	FLG	A Y	L R S	FQ	I P F	DF	V E I	ГК	ΚP	RRR	MA	I Y S	NK
20. WP_010870838.1 hypothetical protein Methanocaldococcus jannaschii			- K	A N I	I D M	EN	K	ΝT	LFL	R D	GF	I Q N	IS E	R N	DL	KE I	SKE	LΝ	S N I	ΕV	IS	I R K	ΝN	к <mark>ү</mark> к	V F	T S D	YR
21. WP_045143632.1 hypothetical protein Clostridium butyricum		V L	. I S	YK	E E N	GΕ	Y P K	ΝL	VIE	I R D	G -	FSF	E N	I D	ΝY	K E Y	FDK	KG	IKF	NI	ΙEΝ	/ ĸ ĸ	NI	P <mark>V</mark> K	(L A	κvv	GS
22. WP_050002102.1 hypothetical protein Thermococcus eurythermalis		ΕL	D Q	I V E	ТК	S N	<mark>V</mark>	DI	LIL	R D	GR	I P S	RE	ΙE	QLK	< R L	SLR	RΡ	ΥTF	TL	VG	IIK	RΡ	LVR	FF	K <mark>G</mark> T	DRI
23. WP_062755455.1 stem cell self-renewal protein Piwi domain-containing protei		S I	ΥS	YEC	ΣIΥ	GΑ	K P R	нι	TFF	I R D	G -	ICF	E D	LD	FLG	A Y	L R S	FQ	I P F	DF	V E I	ГК	ΚP	RRR	MA	I Y S	NK
24. WP_042383163.1 hypothetical protein Parageobacillus thermoglucosidans		S I	ΥG	YEG	2 I Y	GΑ	K P R	нι	ΤFΗ	I R D	G - 1	VCF	E D	LΕ	FLG	A Y	L R S	FQ	I P F	GF	V E	ГК	ΚP	RRR	MA	I Y S	N K
25. WP_074100515.1 hypothetical protein Paenibacillus sp. P3E		A I	DQ	YQG	<mark>P H</mark> Y	ΝE	КРК	ΗI	TFH	I R D	G -	FCF	E D	L N	SL	ΡEV	M N S	LG	L Q Y	DM	V E	ГК	RT	N R R	MA	LTV	GK
26. WP_058142162.1 hypothetical protein Clostridium butyricum		V L	. I S	YK	EEN	GE	Y P K	NI	VIE	I R D	G -	FSF	R E N	I D	ΝY	K E Y	F D k	KG	IKF	NI	ΙEN	/ ĸ ĸ	NI	P <mark>V</mark> K	(L A	κvv	GS
27. WP_055276084.1 hypothetical protein Clostridium disporicum		V L	LS	YEE	E E N	GQ	Y P R	NI	VIF	I R D	G -	FSF	ED	LE	NY	< <mark>N</mark> Y	FLK	ΚN	I E F	SI	VΕV	/ R K	NF	A T R	LV	N N F	ND
28. WP_052698403.1 hypothetical protein Domibacillus enclensis		V L	Y S	YQK	< Q F	GΗ	FΡK	ΗV	ТІН	I R D	G -	KWF	ES	SE	LVC	RL	FKC	RN	V A Y	DI	VΕV	/ I K	ΚP	N R R	MA	FYN	NE
29. WP_042211195.1 hypothetical protein Paenibacillus borealis		A I	DQ	YQC	ЯΗΥ	NE	RPK	ΗV	TFF	I R D	G -	FCF	ED	LL	SL	ΡEV	M N S	LD	V Q Y	DM	V E	ιк	КΤ	N R R	MA	LTV	GK
30. WP_048163021.1 hypothetical protein Thermogladius cellulolyticus		EN	Y A	ELK	<		- <mark>G</mark> K	RI	LVL	RD	G R	L T K	E E	VТ	QLA	D I	SRT	КD	СТІ	ΕV	INV	/ ĸ ĸ	RТ	P F Q	IQ L	S D S	RV
		_																									

Alignment of TOP 30 hits of *Pf*Ago by MEGA7 ClustalW



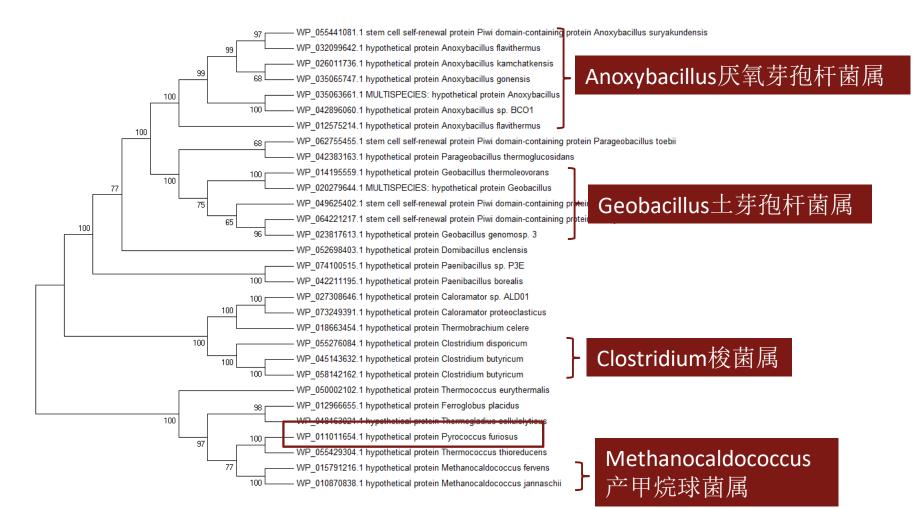
Phylogenetic tree of *Tt*Ago and *Pf*Ago



TtAgo:Neighbor joining, bootstrap 500 by MEGA7



Phylogenetic tree of *Tt*Ago and *Pf*Ago



*Pf*Ago:Neighbor joining,bootstrap 500 by MEGA7



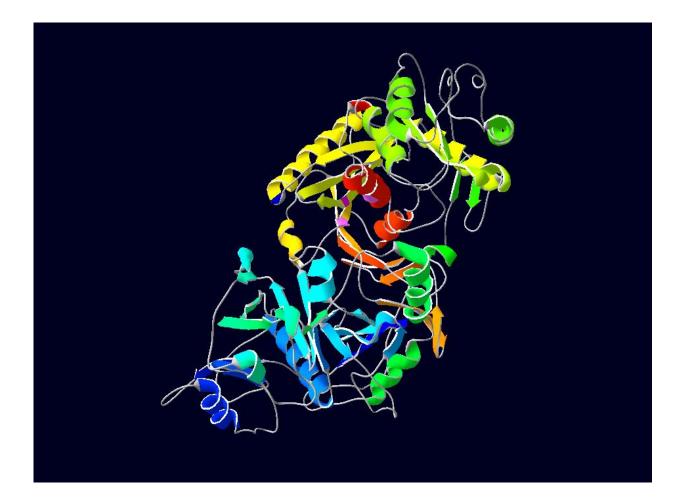
• Structure prediction of DNA-cleavage Argonaute protein

Common hits in both *Tt*Ago and *Pf*Ago

PSI-BLAST hits	Rank in <i>Tt</i> Ago	Rank in <i>Pf</i> Ago
Anoxybacillus suryakundensis	26	3
Anoxybacillus kamchatkensis	30	4
Anoxybacillus gonensis	27	6
Anoxybacillus sp. BCO1	24	9



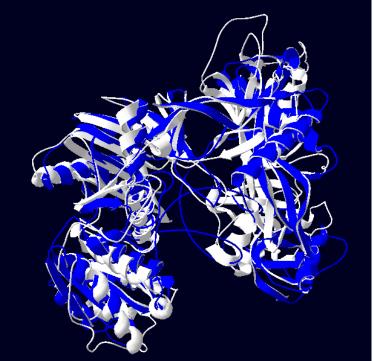
> AsAgo structure prediction on SWISS-MODEL



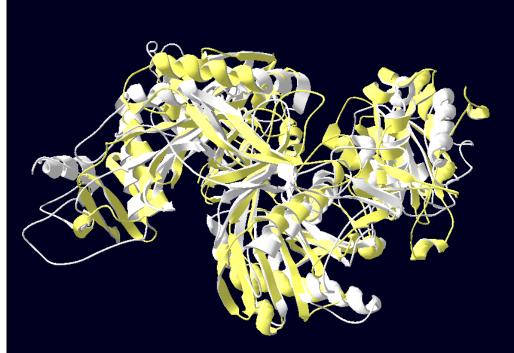


> AsAgo structure fitting with TtAgo and PfAgo

TtAgo in blue



*Pf*Ago in yellow

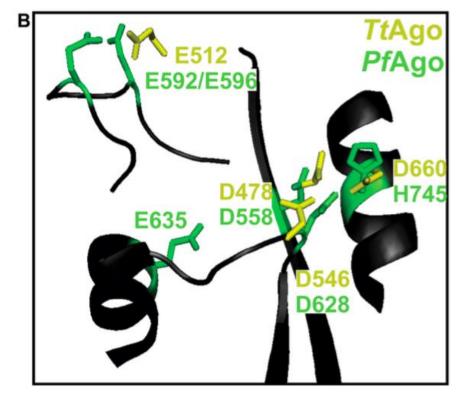


Fit by Swiss-PDBViewer



> AsAgo catalytic site prediction based on TtAgo and PfAgo

Α	D		Е		D		x
hAGO2	IFLGADVTHPP	I	QHRQEIIQDL	1	IIFYRDGVSEG	I	PAYYAHLVAFR
TtAgo	LAVGFDAGGRE	1	-GERIPQEVV	I	VLLLRDGRV	I	PLHLADRLVKE
MjAgo	YIMGLDTGLGI	I	-GERLHLPYV	I	ILFLRDGFI	T	PIHYADKFVKA
PfAgo	YIIGIDVAPMK	١	GEQRGESVDM	I	ILLLRDGRI	١	PVHYAHKFANA

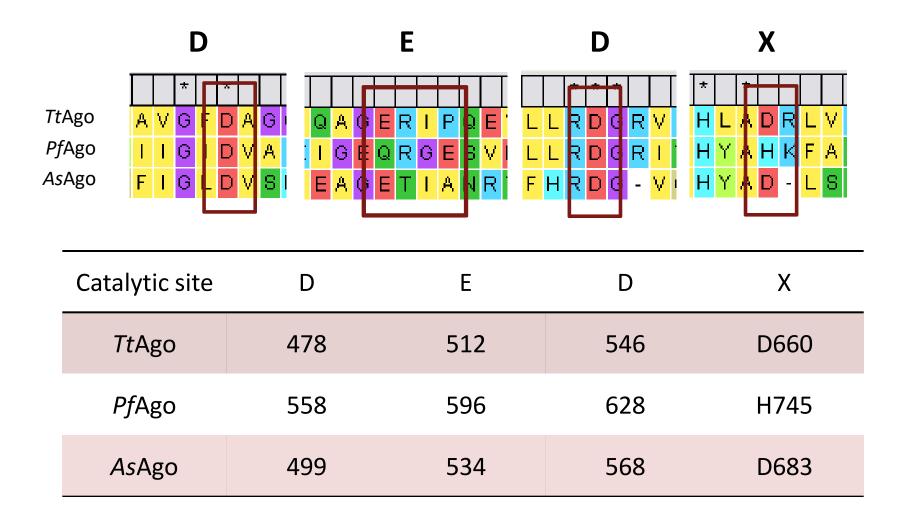


The DDX residues are positioned close together in the available pAgo structures, the glutamic acid (E) is located on a structural sub-domain termed the 'glutamate finger'.

Daan C. et al. NAR. 2014

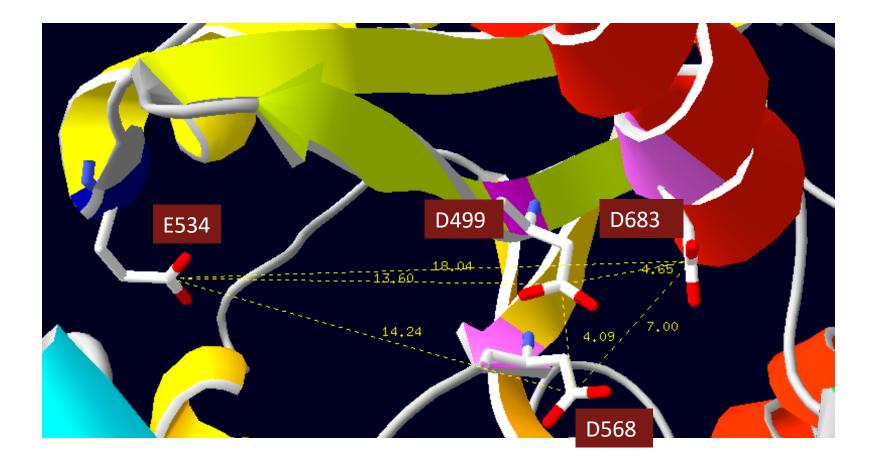


> AsAgo catalytic site prediction based on TtAgo and PfAgo





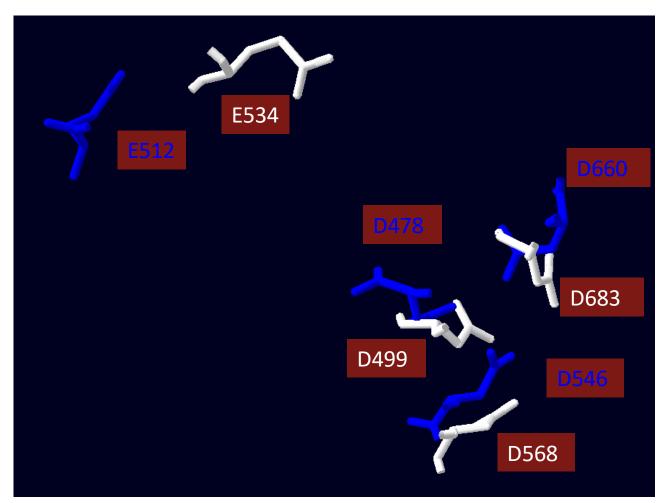
> AsAgo catalytic site prediction based on TtAgo and PfAgo



Calculate by Swiss-PDBViewer



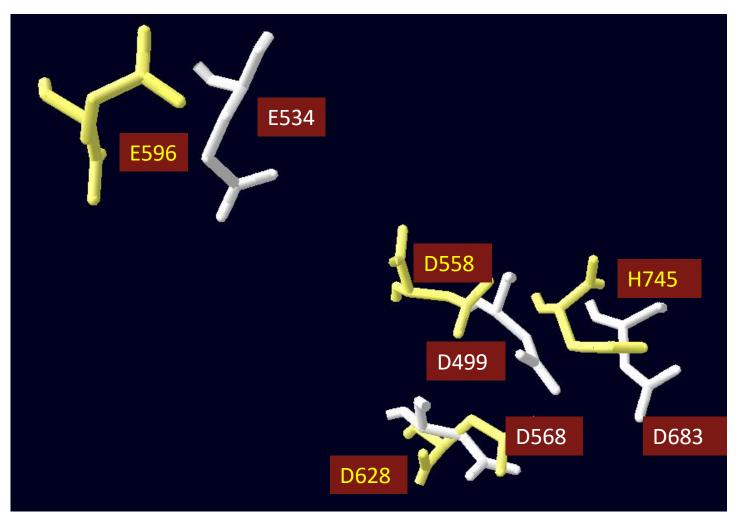
> AsAgo catalytic site with TtAgo



TtAgo in blue



> AsAgo catalytic site with PfAgo



*Pf*Ago in yellow



Summary

- AsAgo is the top hit in bost TtAgo and PfAgo PSI-BLAST result
- AsAgo catalytic residues: D499 E534 D568 D683
- AsAgo may be a potential DNA-cleavage Argonaute protein



Acknowledgement

- Professor Luo
- Teaching assistant Ke Lan
- Liu Zhiheng
- All member in Group 9





