

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

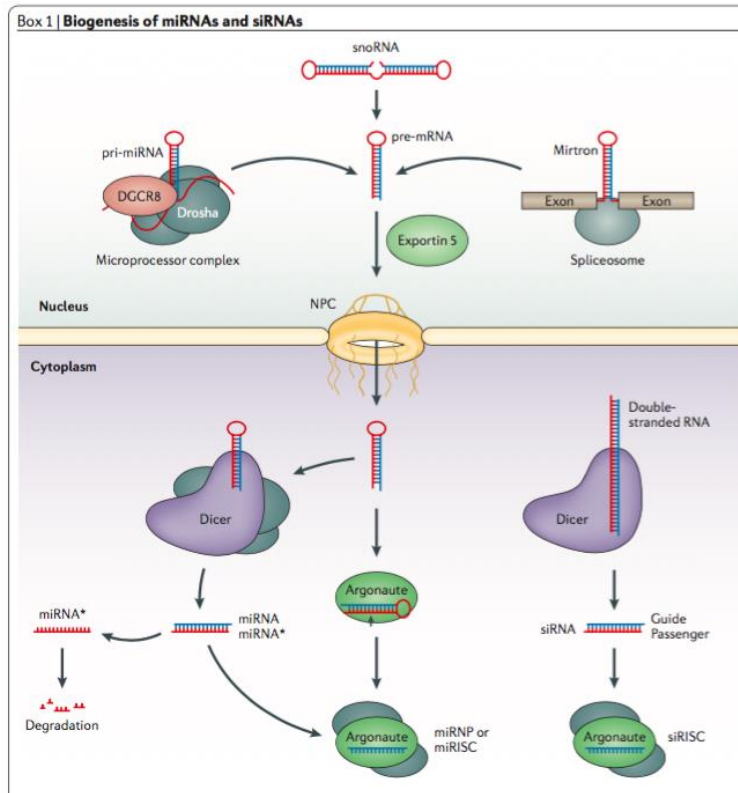
Outline

- Background
- Sequence analysis of *TtAgo* and *PfAgo*
- 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 RNA-induced 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 Argonaute

TtAgo

Daan C. Swarts^{1*}, Matthijs M. Jore^{1*}, 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

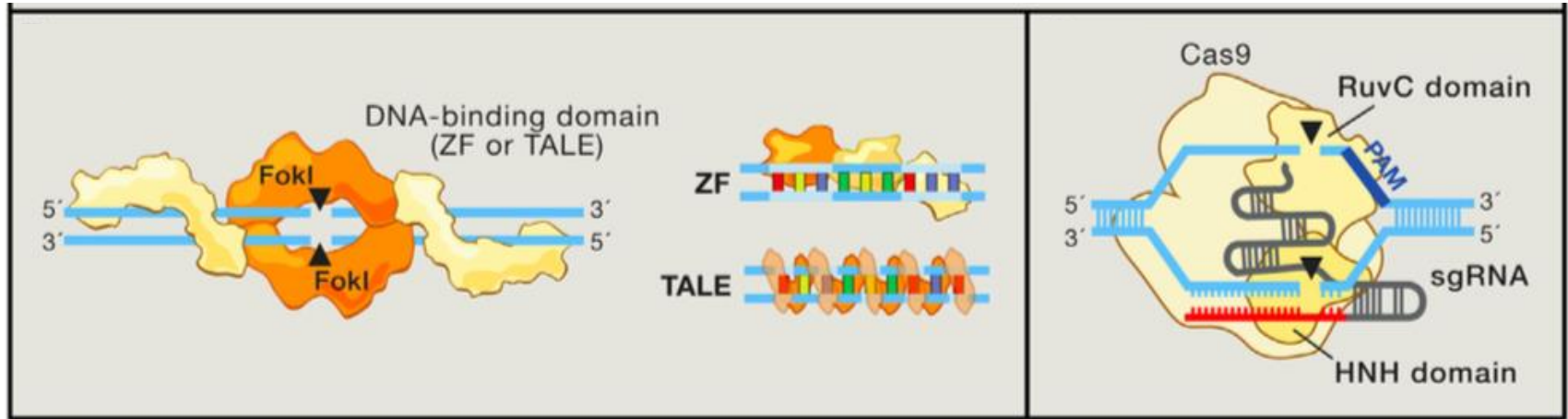
PfAgo

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,*}

NAR 2015.4

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

➤ Several genome editing technologies



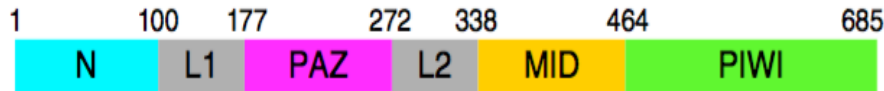
- ✧ ZFN and TALEN: Protein-DNA binding
- ✧ CRISPR/Cas9: RNA-DNA binding
- ✧ New gene editing tool?
- ✧ DNA-DNA binding?

Hsu *et al.*, *Cell*, 2014

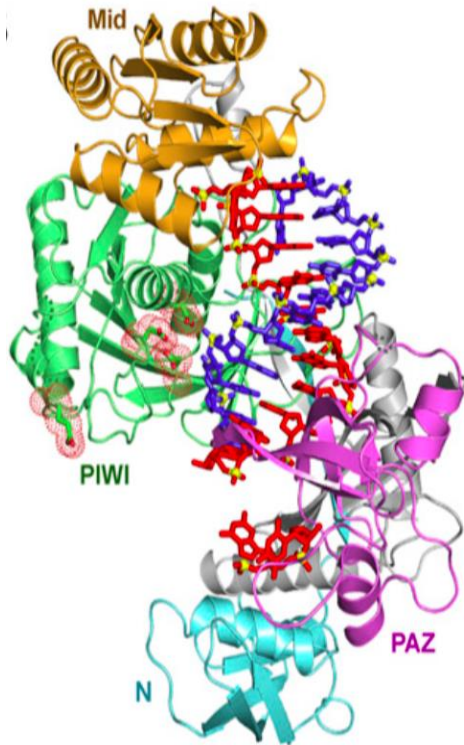
➤ Comparison of *TtAgo* and *PfAgo*

	<i>TtAgo</i>	<i>PfAgo</i>
Organism	<i>Thermus thermophilus</i>	<i>Pyrococcus furiosus</i>
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 *TtAgo* ternary complex



Domain architecture of *TtAgo*



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

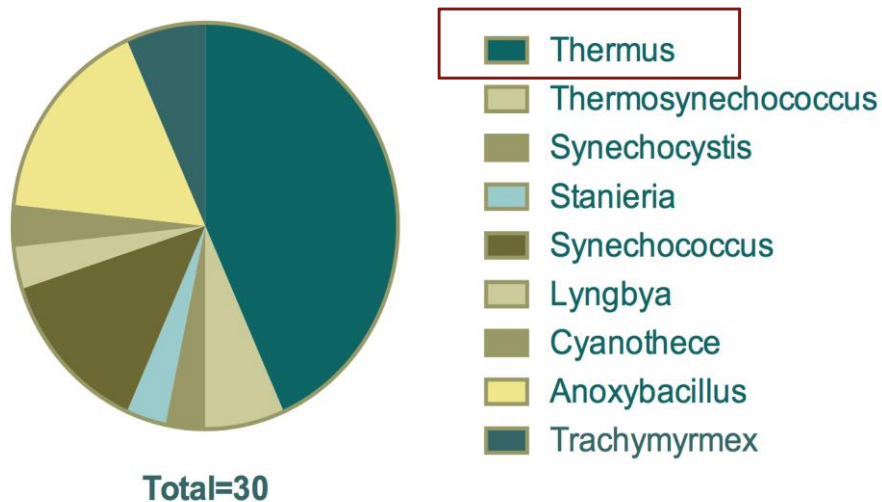
- **Sequence analysis of *TtAgo* and *PfAgo***

- PSI-BLAST of *TtAgo* and *PfAgo*

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

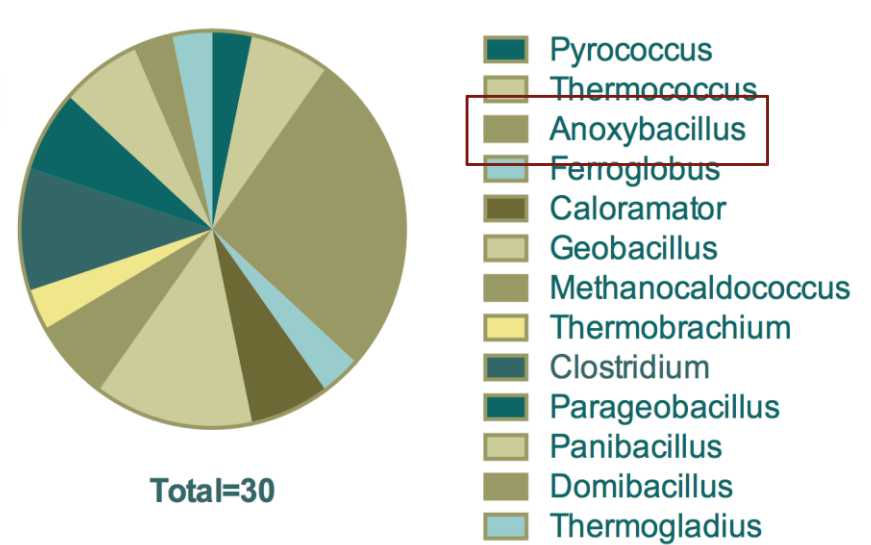
➤ PSI-BLAST of *TtAgo* and *PfAgo*

TOP 30 hits of *TtAgo*



Thermus 栖热菌属

TOP 30 hits of *PfAgo*



Anoxybacillus 厌氧芽胞杆菌属

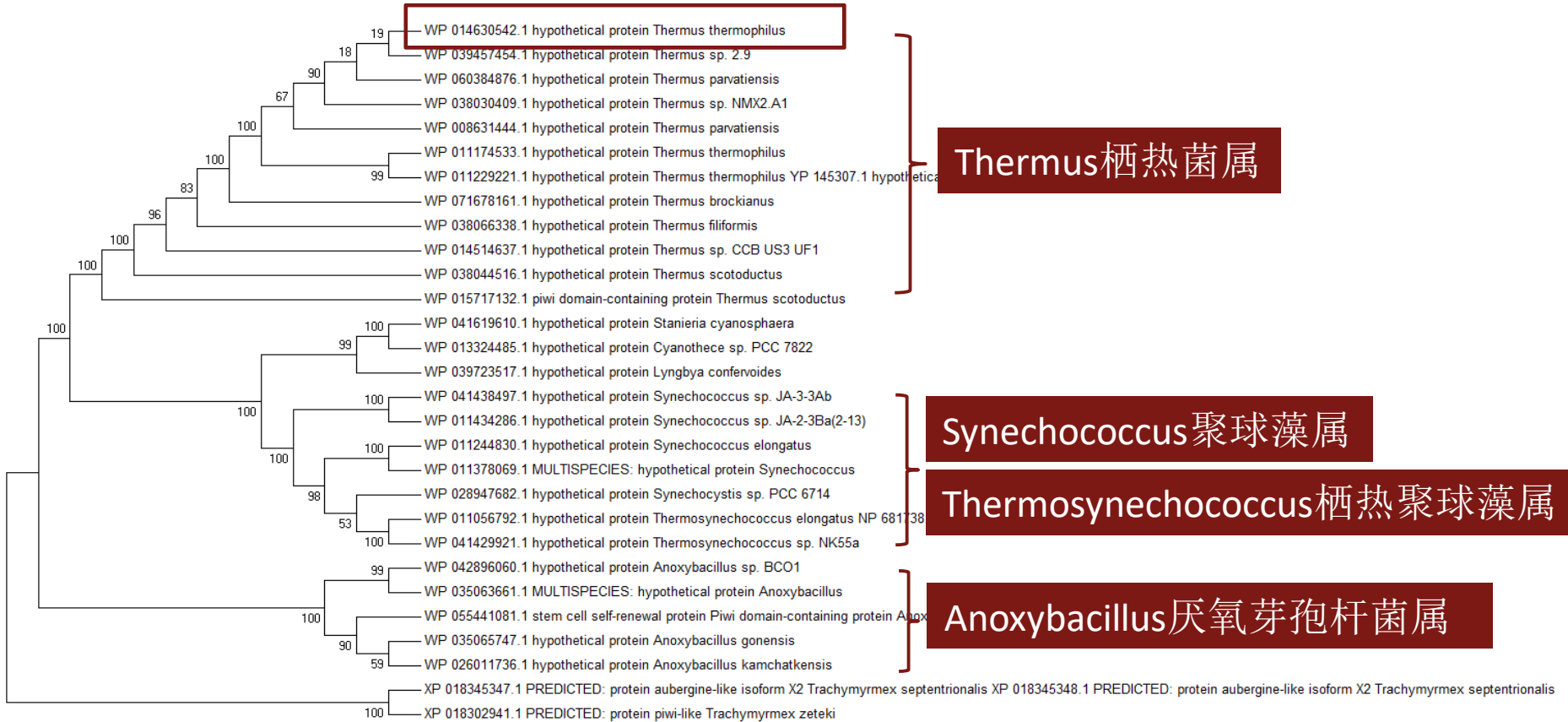
• Phylogenetic analysis of Argonaute Proteins

➤ Alignment of *TtAgo* and *PfAgo*

Species/Abbrv	Group Name					*					*	*																																																												
1. WP_011174533.1 hypothetical protein Thermus thermophilus		A	G	D	T	P	L	-	-	E	A	L	A	H	Q	I	F	H	L	T	R	L	Y	F	A	S	G	F	A	F	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	V	G	R	L	G	-	-	I	R	H	L	K	E	V	D	R	E	K	L	F	F	V				
2. WP_038030409.1 hypothetical protein Thermus sp. NMX2 A1		A	G	D	T	P	L	-	-	E	A	L	A	H	Q	I	F	H	L	T	R	L	Y	F	A	S	G	F	A	F	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	V	G	R	L	G	-	-	I	R	H	L	K	E	V	D	R	E	K	L	F	F	V				
3. WP_060384876.1 hypothetical protein Thermus parvatiensis		A	G	D	T	P	L	-	-	E	A	L	A	H	Q	I	F	H	L	T	R	L	Y	F	A	S	G	F	A	F	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	V	G	R	L	G	-	-	I	R	H	L	K	E	V	D	R	E	K	L	F	F	V				
4. WP_011229221.1 hypothetical protein Thermus thermophilus YP_1		A	G	D	T	P	L	-	-	E	A	L	A	H	Q	I	F	H	L	T	R	L	Y	F	A	S	G	F	A	F	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	V	G	R	L	G	-	-	I	R	H	L	K	E	V	D	R	E	K	L	F	F	V				
5. WP_008631444.1 hypothetical protein Thermus parvatiensis		A	G	D	T	P	L	-	-	E	A	L	A	H	Q	I	F	H	L	T	R	L	Y	F	A	S	G	F	A	F	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	V	G	R	L	G	-	-	I	R	H	L	K	E	V	D	R	E	K	L	F	F	V				
6. WP_014630542.1 hypothetical protein Thermus thermophilus		A	G	D	T	P	L	-	-	E	A	L	A	H	Q	I	F	H	L	T	R	L	Y	F	A	S	G	F	A	F	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	V	G	R	L	G	-	-	I	R	H	L	K	E	V	D	R	E	K	L	F	F	V				
7. WP_039457454.1 hypothetical protein Thermus sp. 2.9		A	G	D	T	P	L	-	-	E	A	L	A	H	Q	I	F	H	L	T	R	L	Y	F	A	S	G	F	A	F	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	V	G	R	L	G	-	-	I	R	H	L	K	E	V	D	R	E	K	L	F	F	V				
8. WP_071678161.1 hypothetical protein Thermus brockianus		E	G	E	T	P	L	-	-	E	A	L	A	E	Q	I	Y	H	L	T	R	L	Y	F	A	S	G	F	A	F	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	V	G	R	L	G	-	-	V	R	H	L	K	E	V	D	R	E	K	L	F	F	V				
9. WP_014514637.1 hypothetical protein Thermus sp. CCB_US3_UF1		A	G	E	T	P	L	-	-	E	E	V	A	R	O	V	Y	H	L	T	R	L	Y	F	P	S	G	F	L	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	R	A	R	V	G	-	-	S	F	P	O	G	R	I	O	P	S	L	F	F	V						
10. WP_038066338.1 hypothetical protein Thermus filiformis		E	G	N	T	P	L	-	-	V	D	L	A	L	O	I	Y	H	L	T	R	L	Y	F	A	S	G	F	L	P	R	L	P	A	P	L	H	L	A	D	R	L	V	K	E	V	G	R	L	G	-	-	V	R	H	L	K	E	V	P	R	D	L	F	F	V						
11. WP_011056792.1 hypothetical protein Thermosynechococcus elon		Y	G	N	T	P	L	-	-	E	L	L	A	L	Q	T	Y	H	L	T	Q	L	H	F	A	S	G	F	R	S	C	R	L	P	W	V	L	H	L	A	D	R	S	S	K	E	F	O	R	I	G	-	-	Q	I	S	L	L	Q	N	V	D	R	E	K	L	I	A	V			
12. WP_041429921.1 hypothetical protein Thermosynechococcus sp.		Y	G	N	T	P	L	-	-	E	L	L	A	L	Q	T	Y	H	L	T	Q	L	H	F	A	S	G	F	S	S	O	R	L	P	W	V	L	H	L	A	D	R	S	S	K	E	F	O	R	I	G	-	-	Q	I	S	L	L	Q	N	I	D	R	E	K	L	I	A	V			
13. WP_038044516.1 hypothetical protein Thermus scotoductus		E	G	P	T	P	V	-	-	E	E	L	A	R	O	I	Y	H	L	S	R	L	Y	P	S	G	Y	R	F	P	S	L	P	A	P	L	H	L	A	D	R	L	V	R	E	V	G	R	V	G	-	-	L	S	S	L	H	G	L	D	R	E	K	L	F	F	V					
14. WP_028947682.1 hypothetical protein Synechocystis sp. PCC 671		Y	G	N	T	S	L	-	-	D	L	L	A	L	Q	T	Y	Q	L	T	Q	L	H	F	A	S	G	F	R	S	C	R	L	P	W	V	L	H	L	A	D	R	S	S	K	E	F	O	R	I	G	-	-	Q	I	S	L	L	Q	N	I	S	R	E	K	L	I	A	V			
15. WP_041619610.1 hypothetical protein Stanieria cyanosphaera		Y	G	D	A	P	L	-	-	E	I	L	A	R	O	I	D	R	L	T	C	L	L	N	F	A	S	G	F	Y	S	V	S	R	L	P	Y	V	I	H	L	A	D	K	M	A	K	I	V	O	R	I	G	-	-	E	V	Q	I	S	V	L	Q	I	D	R	K	L	I	F	F	A
16. WP_041438497.1 hypothetical protein Synechococcus sp. JA-3-3A		H	G	K	T	P	L	-	-	E	V	L	A	L	Q	T	Y	H	L	S	Q	L	H	F	A	S	G	F	Q	A	C	R	L	P	W	V	L	H	F	A	D	K	S	G	K	E	F	O	R	L	G	D	N	T	F	S	L	Q	N	I	D	R	E	K	L	I	A	V				
17. WP_011244830.1 hypothetical protein Synechococcus elongatus		Y	G	S	T	D	L	-	-	E	L	L	A	L	Q	T	Y	H	L	T	Q	L	H	F	A	S	G	F	R	S	C	R	L	P	W	V	L	H	L	A	D	R	S	S	K	E	F	O	R	I	G	-	-	Q	I	S	L	L	Q	N	I	S	R	E	K	L	I	A	V			
18. WP_011378069.1 MULTISPECIES: hypothetical protein Synechococcus sp.		Y	G	S	T	D	L	-	-	E	L	L	A	L	Q	T	Y	H	L	T	Q	L	H	F	A	S	G	F	R	S	C	R	L	P	W	V	L	H	L	A	D	R	S	S	K	E	F	O	R	I	G	-	-	Q	I	S	V	L	Q	N	I	S	R	D	K	L	I	A	V			
19. WP_011434286.1 hypothetical protein Synechococcus sp. JA-2-3B		H	G	N	T	P	L	-	-	E	V	L	A	I	O	T	Y	H	L	S	Q	L	H	F	A	S	G	F	Q	A	C	R	L	P	W	V	L	H	F	A	D	K	S	G	K	E	F	O	R	I	G	N	S	T	F	S	V	L	Q	N	I	N	R	E	K	L	I	A	V			
20. WP_039723517.1 hypothetical protein Lyngbya confervoides		H	G	D	A	S	L	-	-	R	L	L	A	N	E	I	F	A	L	S	O	F	H	A	S	A	F	R	S	S	R	L	P	M	P	L	H	Y	A	D	R	M	I	K	E	V	G	R	L	G	-	-	Q	L	S	I	L	H	G	I	D	R	K	L	F	F	A					
21. WP_013324485.1 hypothetical protein Cyanothece sp. PCC 7822		H	G	D	A	P	L	-	-	E	V	L	A	R	O	I	D	R	L	T	M	L	N	F	A	S	G	F	Y	S	R	L	P	M	V	L	H	F	A	D	K	M	A	K	I	V	O	R	L	G	-	-	Q	I	A	F	L	Q	K	V	E	R	E	K	L	I	F	F	A			
22. WP_015717132.1 piwi domain-containing protein Thermus scotoc		A	G	E	V	D	L	N	L	E	E	A	R	Q	L	Y	H	L	S	R	I	Y	P	S	G	Y	R	F	P	R	L	P	A	P	L	H	M	V	D	R	M	V	R	E	V	A	R	L	G	-	-	G	S	H	N	L	R	L	K	E	E	L	F	F	L							
23. WP_042896060.1 hypothetical protein Anoxybacillus sp. BC01		T	K	C	L	S	I	R	-	-	D	V	V	S	D	A	Y	K	L	S	F	M	H	I	S	-	M	L	K	T	R	L	P	I	T	V	H	Y	A	D	-	L	S	S	T	F	H	N	R	G	-	-	L	I	H	P	R	S	V	H	E	G	A	L	P	F	V					
24. WP_035063861.1 MULTISPECIES: hypothetical protein Anoxybacill		T	K	C	L	S	I	R	-	-	D	V	V	S	D	A	Y	K	L	S	F	M	H	I	S	-	M	L	K	T	R	L	P	I	T	V	H	Y	A	D	-	L	S	S	T	F	H	N	R	G	-	-	L	I	H	P	R	S	V	H	E	G	A	L	P	F	V					
25. WP_055441081.1 stem cell self-renewal protein Piwi domain-cont		T	K	C	L	S	I	R	-	-	D	I	V	S	D	A	Y	K	L	S	F	M	H	I	S	-	M	L	K	T	R	L	P	I	T	V	H	Y	A	D	-	L	S	S	T	F	H	N	R	G	-	-	L	I	H	P	R	S	V	H	E	G	A	L	P	F	V					
26. WP_035065747.1 hypothetical protein Anoxybacillus gonensis		T	K	C	L	S	I	R	-	-	D	I	V	S	D	A	Y	K	L	S	F	M	H	I	S	-	M	L	K	T	R	L	P	I	T	V	H	Y	A	D	-	L	S	S	T	F	H	N	R	G	-	-	L	I	H	P	R	S	V	H	E	G	A	L	P	F	V					
27. XP_018345347.1 PREDICTED: protein aubergine-like isoform X2		F	D	N	T	G	L	E	V	D	I	V	S	D	A	Y	K	L	S	F	M	H	I	S	-	C	S	T	I	T	V	R	-	V	P	A	P	C	H	Y	A	H	K	L	S	F	L	V	G	R	F	L	-	-	H	Q	P	P	N	S	Q	L	E	R	K	L	F	F	L			
28. XP_018302941.1 PREDICTED: protein piwi-like Trachymyrmex zet		F	D	N	T	G	L	E	A	E	I	V	S	D	A	Y	K	L	S	F	M	H	I	S	-	C	S	T	I	T	V	R	-	V	P	A	P	C	H	Y	A	H	K	L	S	F	L	V	G	R	F	L	-	-	H	Q	P	P	N	S	Q	L	E	R	K	L	F	F	L			
29. WP_026011736.1 hypothetical protein Anoxybacillus kamchatkensis		T	K	C	L	S	I	R	-	-	D	I	V	S	D	A	Y	K	L	S	F	M	H	I	S	-	M	L	K	T	R	L	P	I	T	V	H	Y	A	D	-	L	S	S	T	F	H	N	R	G	-	-	L	I	H	P	R	S	V	H	E	G	A	L	P	F	V					

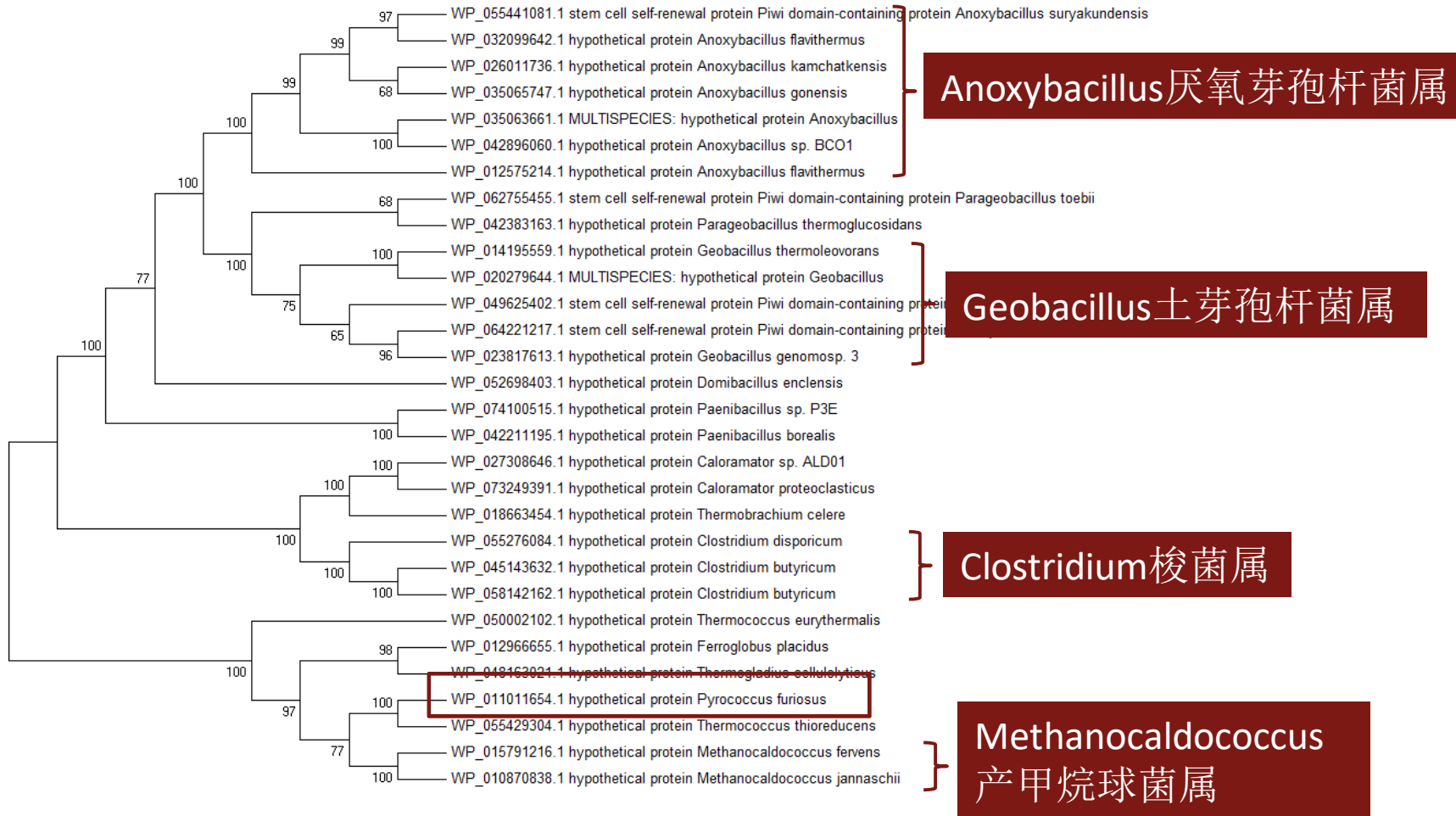
Alignment of TOP 30 hits of *TtAgo* by MEGA7 ClustalW

➤ Phylogenetic tree of *TtAgo* and *PfAgo*



TtAgo:Neighbor joining,bootstrap 500 by MEGA7

➤ Phylogenetic tree of *TtAgo* and *PfAgo*



PfAgo:Neighbor joining,bootstrap 500 by MEGA7

- **Structure prediction of DNA-cleavage Argonaute protein**

➤ Common hits in both *TtAgo* and *PfAgo*

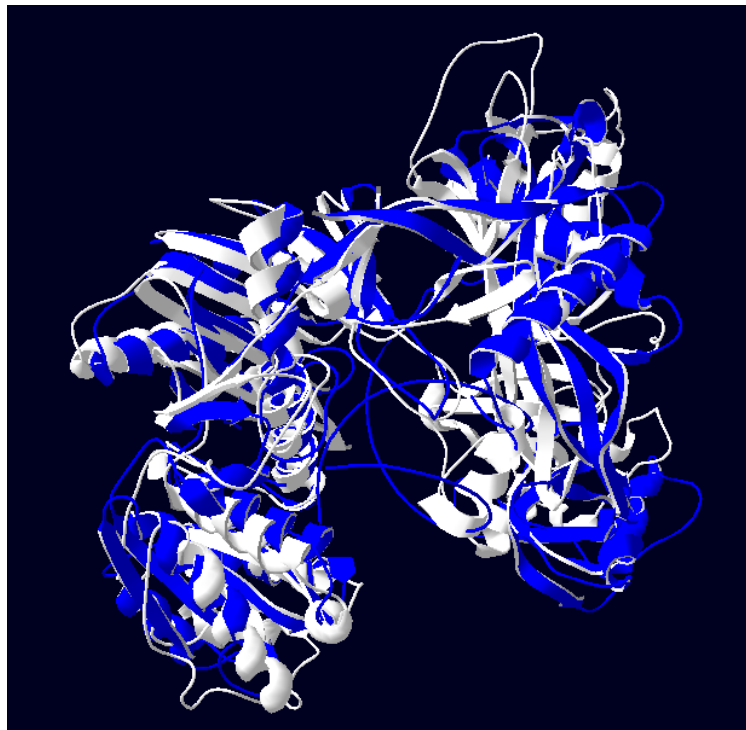
PSI-BLAST hits	Rank in <i>TtAgo</i>	Rank in <i>PfAgo</i>
<i>Anoxybacillus suryakundensis</i>	26	3
<i>Anoxybacillus kamchatkensis</i>	30	4
<i>Anoxybacillus gonensis</i>	27	6
<i>Anoxybacillus sp. BCO1</i>	24	9

➤ AsAgo structure prediction on SWISS-MODEL

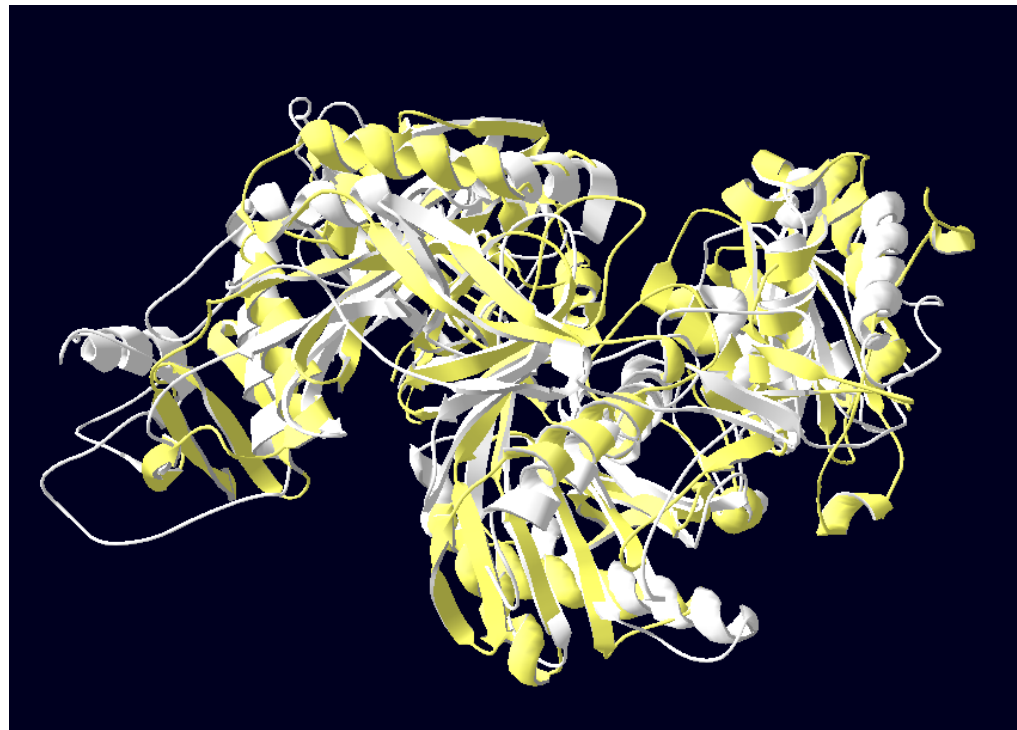


➤ *AsAgo* structure fitting with *TtAgo* and *PfAgo*

TtAgo in blue



PfAgo in yellow

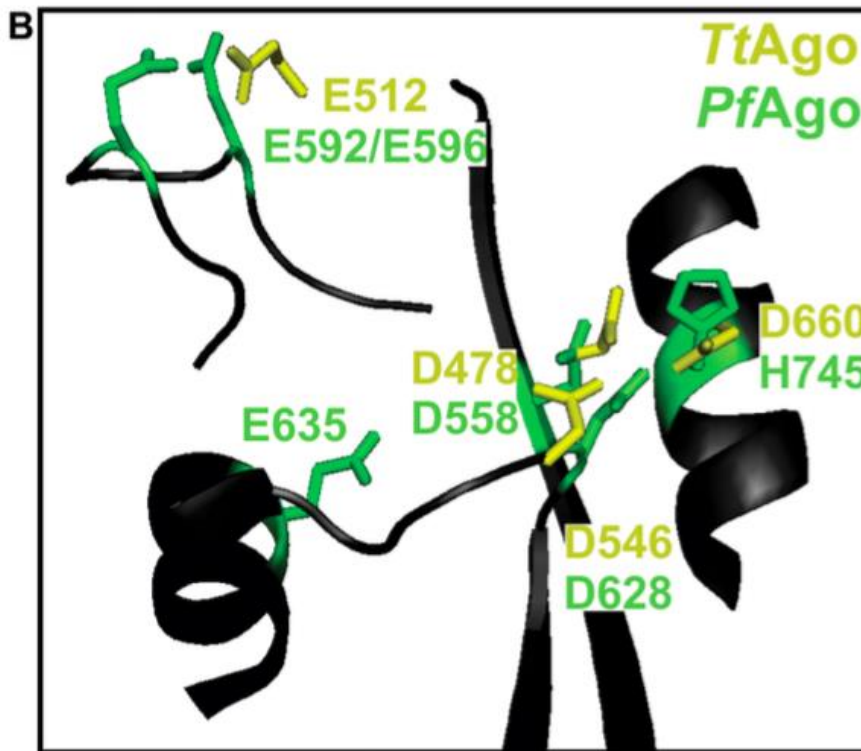


Fit by Swiss-PDBViewer

➤ AsAgo catalytic site prediction based on *TtAgo* and *PfAgo*

A

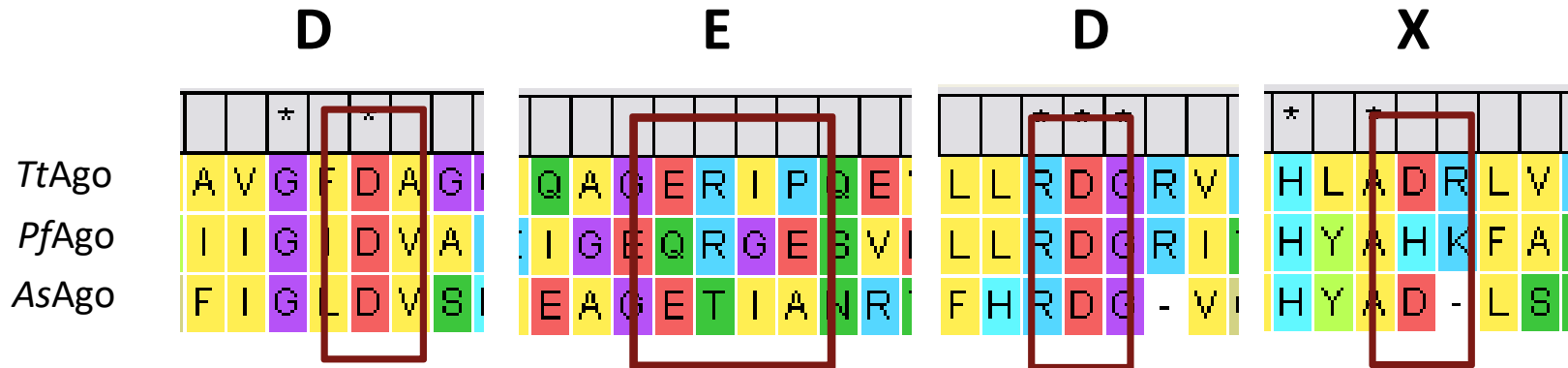
	D	E	D	X
hAGO2	IFLGADVTHPP	QHRQ E IIQDL	IIFYR D GVSEG	PAYY A HLVAFR
<i>TtAgo</i>	LAVGF D AGGRE	- G ERIPQEVV	VLLLR D GRV--	PLHL A DRLVKE
<i>MjAgo</i>	YIMGL D TGLGI	- G ERLHLPYV	ILFLR D GFI--	PIHY A DKFVKA
<i>PfAgo</i>	YIIGID V APMK	G EQRG E SVDM	IILLR D GRI--	PVHY A HKFANA



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'.

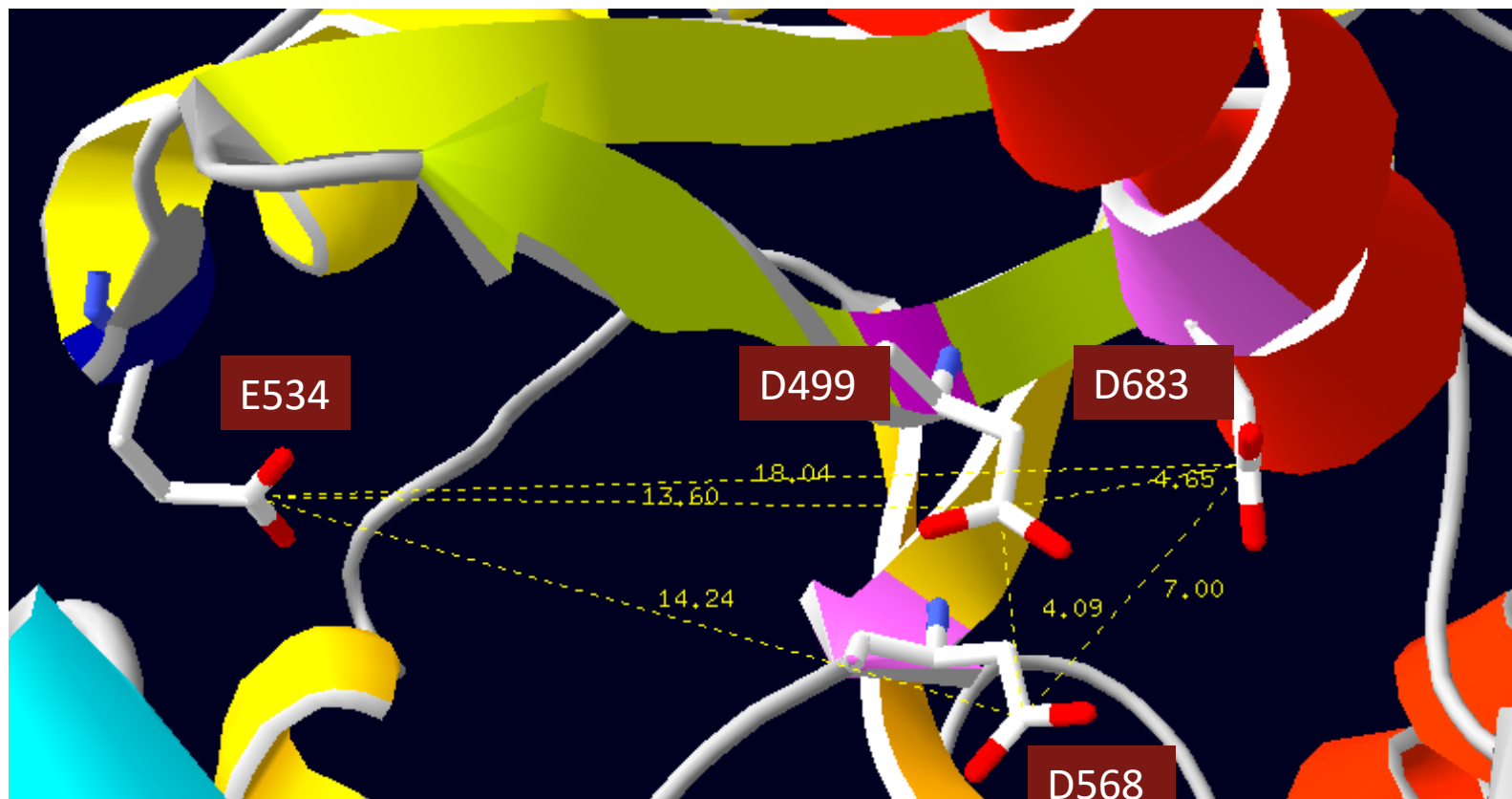
Daan C. *et al.* NAR. 2014

➤ *AsAgo* catalytic site prediction based on *TtAgo* and *PfAgo*



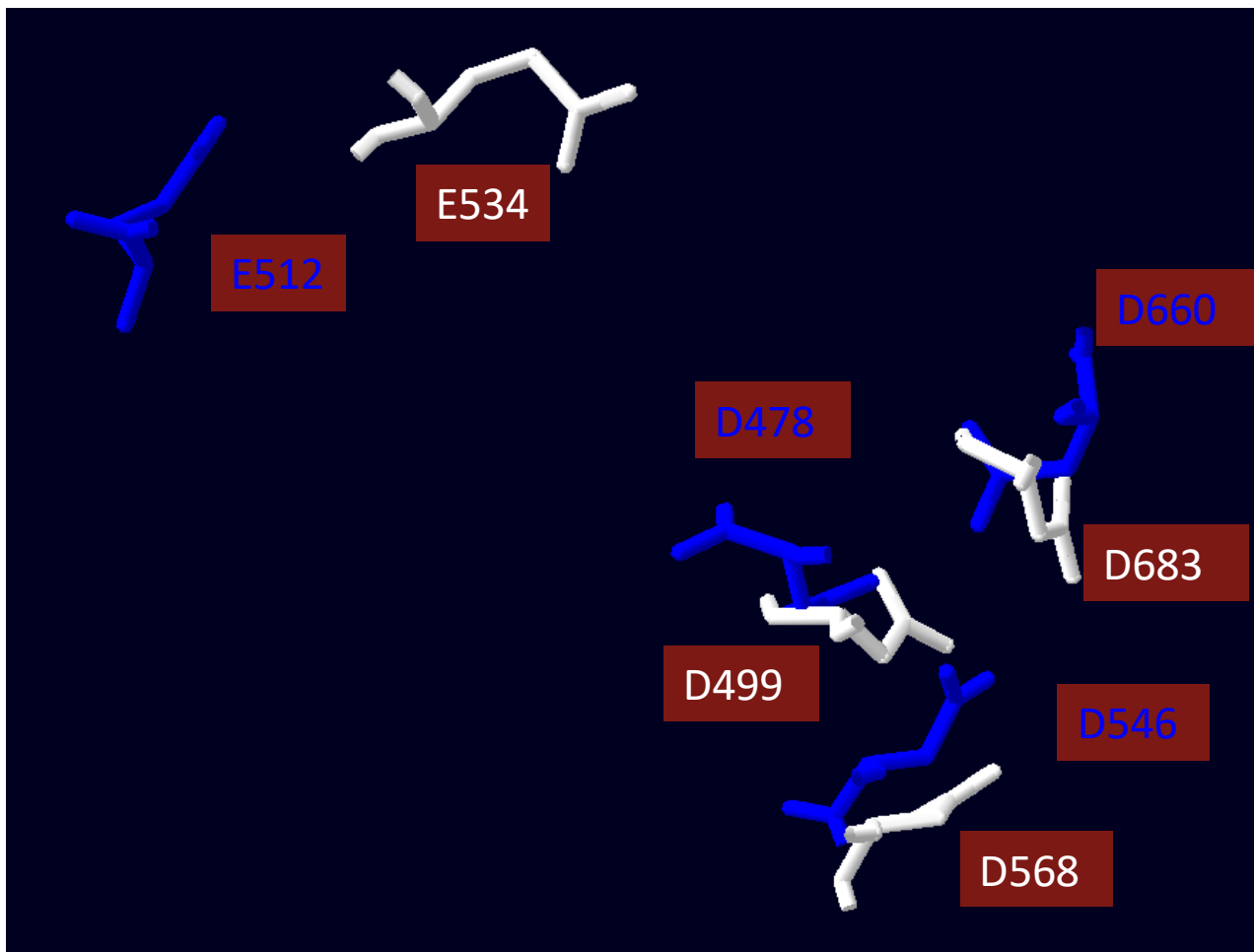
Catalytic site	D	E	D	X
<i>TtAgo</i>	478	512	546	D660
<i>PfAgo</i>	558	596	628	H745
<i>AsAgo</i>	499	534	568	D683

➤ *AsAgo* catalytic site prediction based on *TtAgo* and *PfAgo*



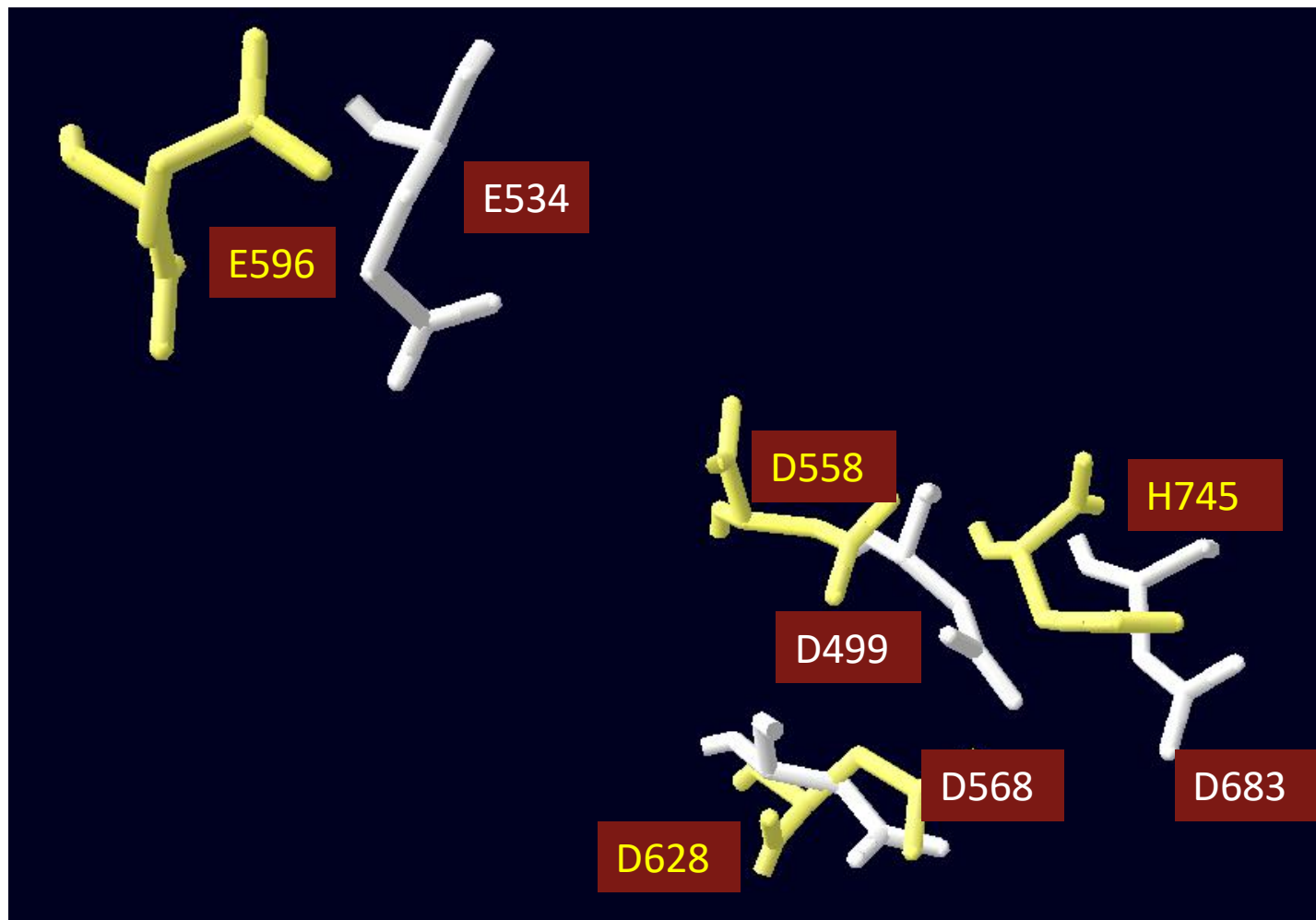
Calculate by Swiss-PDBViewer

➤ *AsAgo* catalytic site with *TtAgo*



TtAgo in blue

➤ *AsAgo* catalytic site with *PfAgo*



PfAgo 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*

Thanks !