TtAgo 和 NgAgo 蛋白为代表的AGO 家族的生物信息学结构分析及功能预测

Argonaute (Ago) Family Protein Research based on Protein Structure and Functions

G10
报告人：王博仑
成员：塞嘉 莘少红 邵昕宁
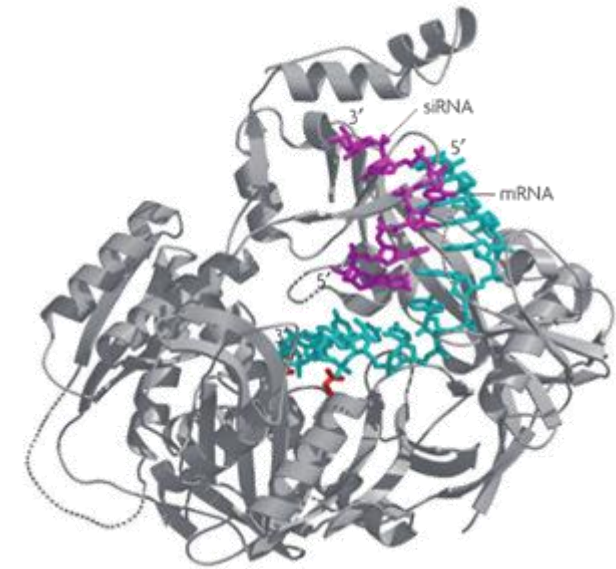
20160618

Content

Background

Sequence Analysis

Phylogeny Construction

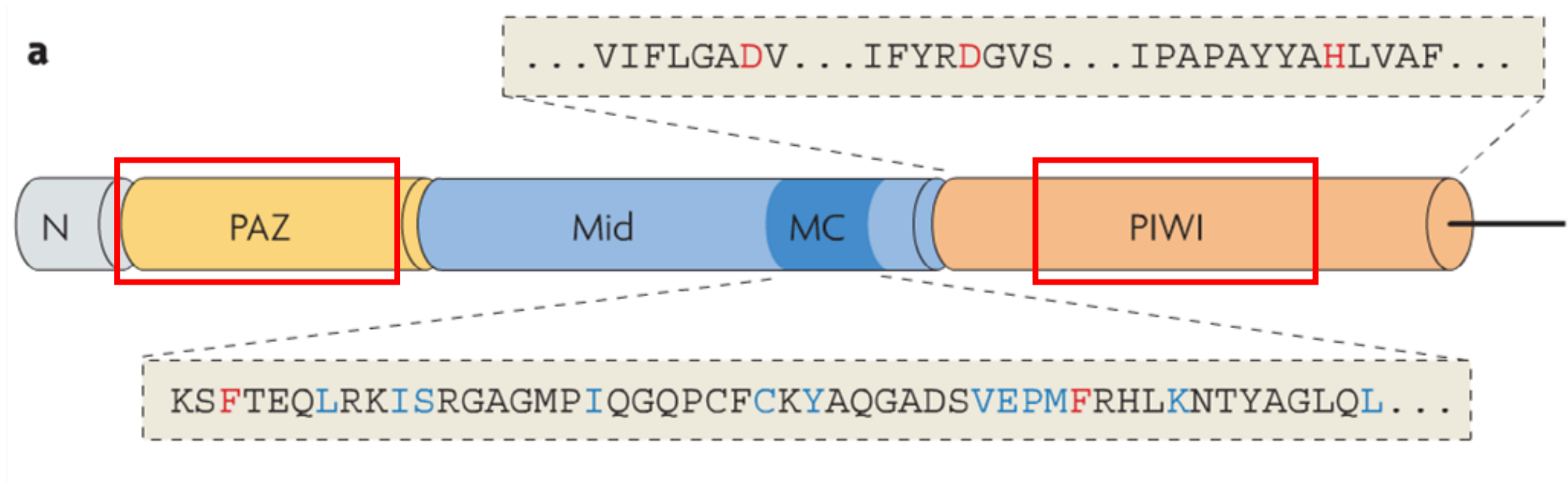3D Structure Prediction

BACKGROUND

Argonaute (Ago)

- Maintaining genome integrity

- Controlling protein synthesis and RNA stability

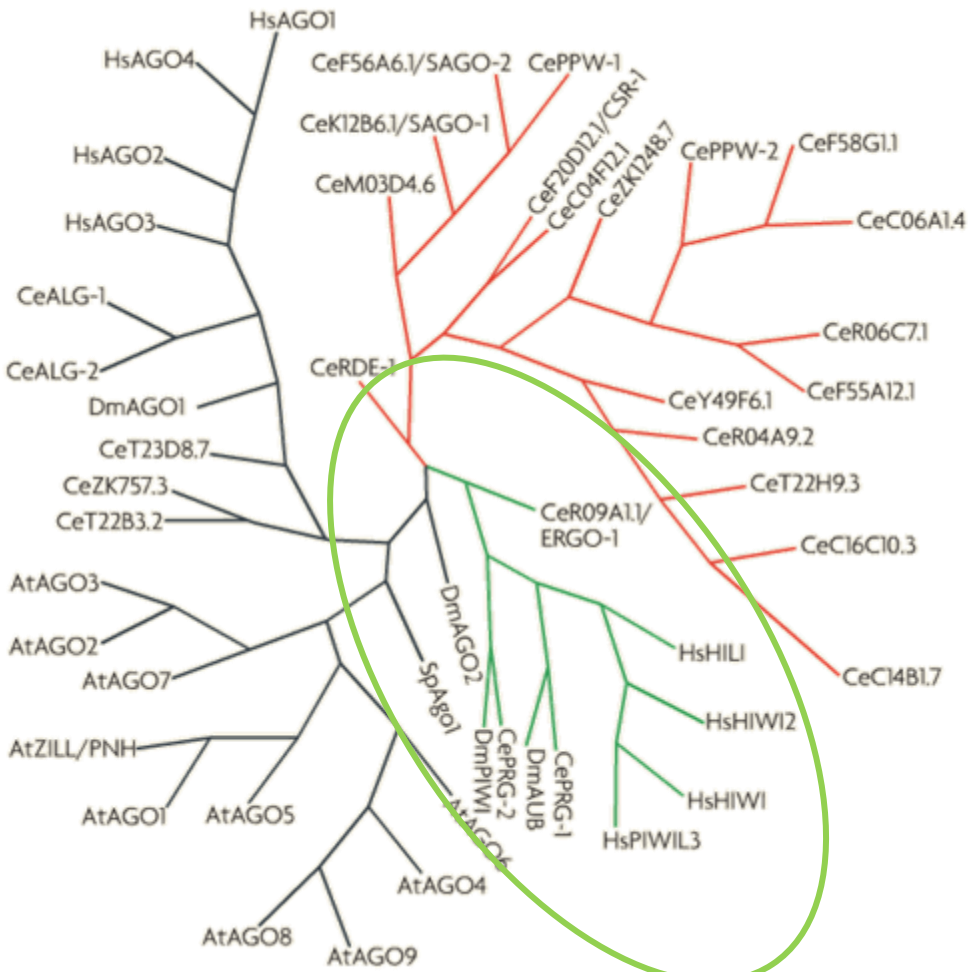- Production of a specific set of small non-coding RNAs

-The original function of Ago family was similar to that of members of the RNase H family of endonucleases.

Functional domains of Argonautes

...VIFLGA**D**V...IFYR**D**GVS...IPAPAYYA**H**LVAF...

KS**F**TEQ**L**RK**I**S**R**GAGMP**I**QGQPCF**C**KYAQGADS**VE**P**MF**RHL**K**NTYAGLQ**L**...

human AGO2 is shown as an example

The Origin of the Argonaute

Argonaute-like group found in plants, animals and fungi

Piwi-like group

The C. elegans-specific group 3 Argonaute

At, Arabidopsis thaliana;
Ce, Caenorhabditis elegans;
Hs, Homo sapiens;
Sp, Schizosaccharomyces pombe

PAZ & PIWI domain

- PAZ Domain

-OB-like folding (oligonuclotide binding) might bind single-stranded nucleic acids

-Recognized the 3'-ends of ssRNAs

- PIWI Domain

-RNase H-like fold: cleave RNA using a DNA template (possibly DNase activity)

-Cleavage-competent Argonaute proteins have a catalytic centre:

Asp-Asp-Asp/Glu/His/Lys

Genome Editing

- A type of genetic engineering in which DNA is inserted, deleted or replaced in the genome of an organism using engineered nucleases.

- Double-strand breaks (DSBs) at desired locations

- Repaired through non-homologous end-joining (NHEJ) or homologous recombination (HR), resulting in targeted mutations
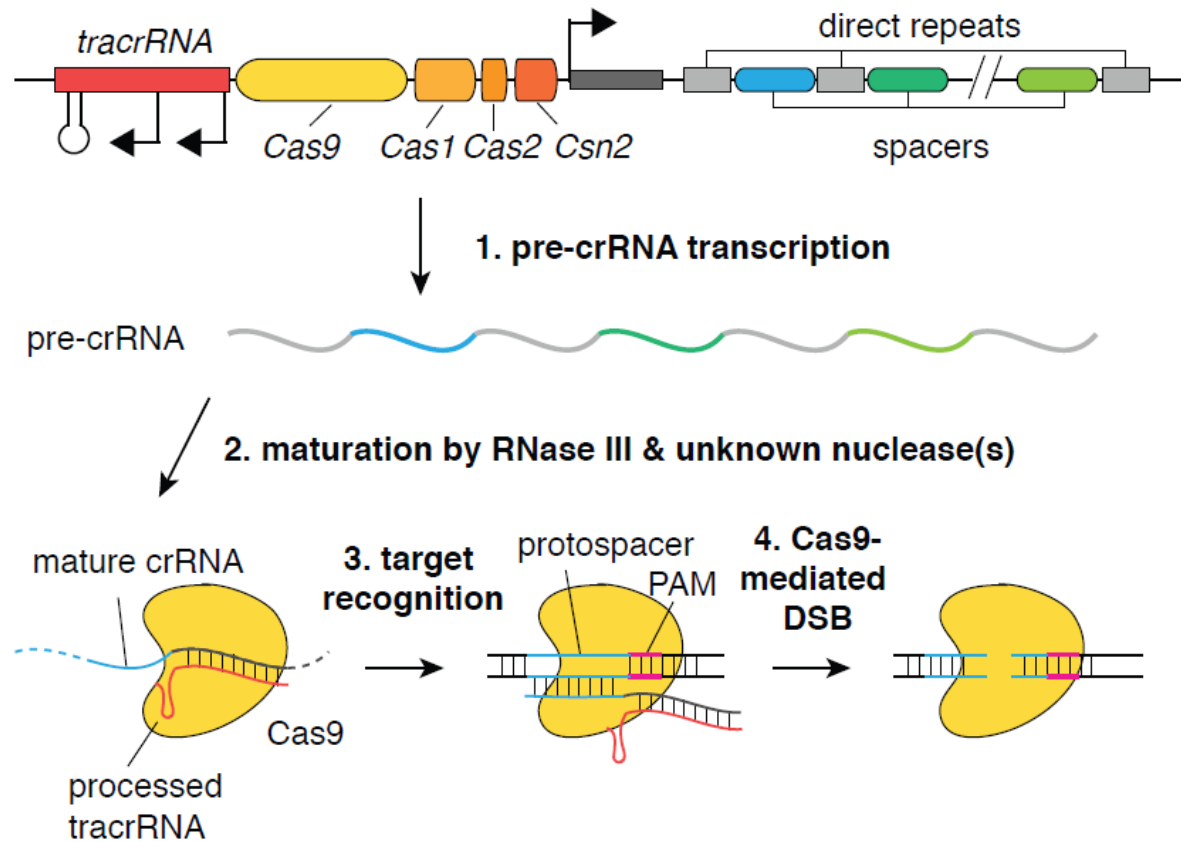
Hybrid Meganuclease

5′ AGCACAGATTCCGAATCGATATCAGAAAGCTTT 3′
3′ TCGTGTCTAAGGCTTAGGTATAGTCTTTCGAAA 5′

ZFN

5′ AGCACAGATTCCGAATCGATATCAGAAAGCTTT 3′
3′ TCGTGTCTAAGGCTTAGGTATAGTCTTTCGAAA 5′

Zinc finger domains

TALEN

5′ AGCACAGATTCCGAATCGATATCAGAAAGCTTT 3′
3′ TCGTGTCTAAGGCTTAGGTATAGTCTTTCGAAA 5′

TALE subunits

active FokI catalytic subunit heterodimer
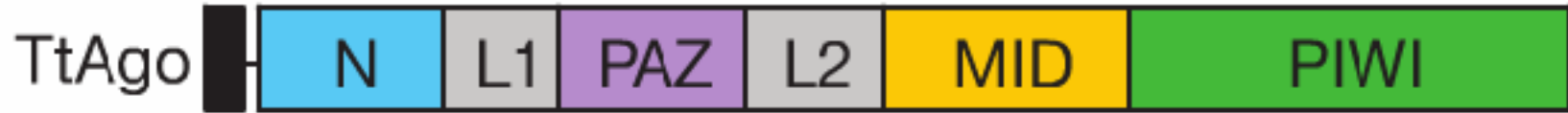
CRISPR/Cas9

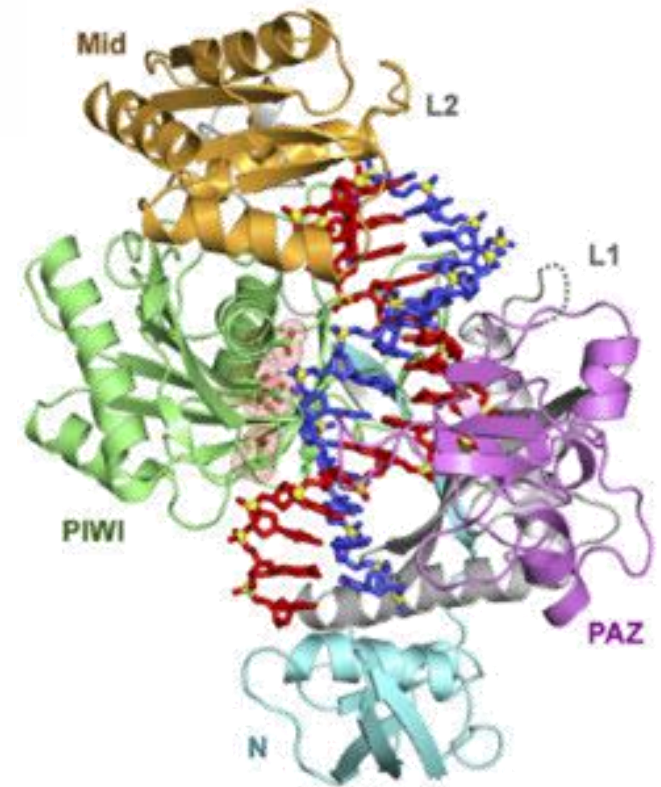
High Efficiency
Simplicity

Off-target Effect
Restriction on PAM

# TtAgo (Tt for *Thermus thermophilus*) (685aa)

- Functions in host defence by DNA-guided DNA interference

- Structure information acquired

- Problem: require 65℃ for reaction

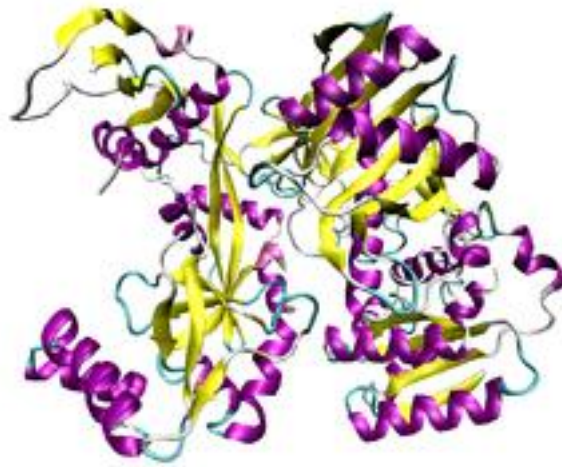# NgAgo (Ng for *Natronobacterium gregoryi*) (887aa)

- DNA-guided endonuclease suitable for genome editing in human cells

- Creates site-specific DNA double-strand breaks when loaded with gDNAs

- Perfectly cleavages DNA at 37 ℃
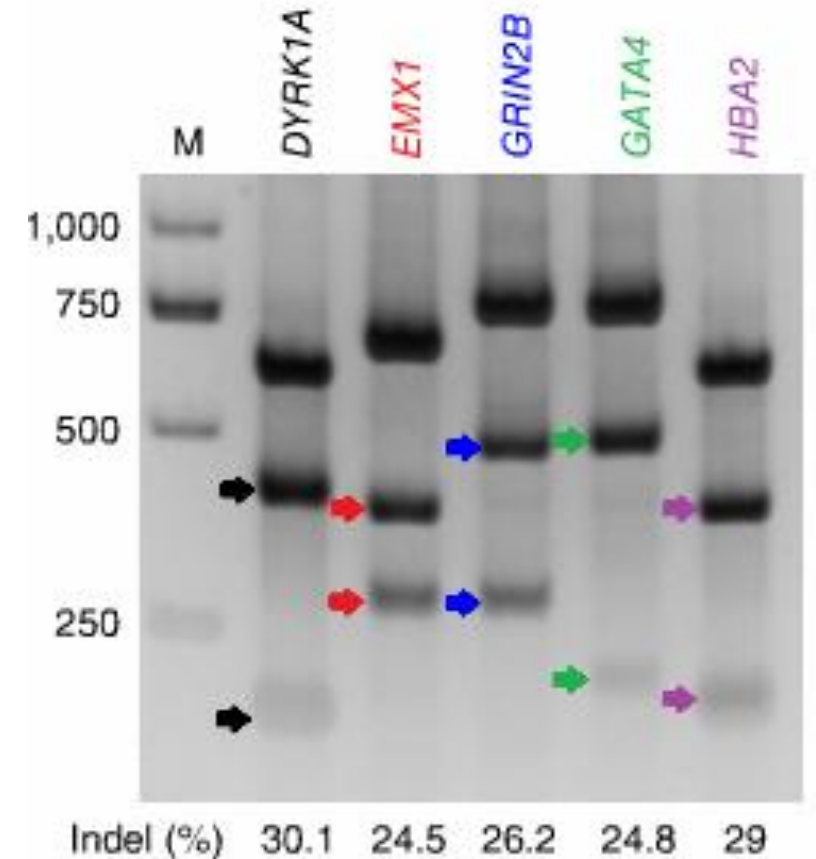
ARGONAUTE
The Next Powerful Genome Editing Tool?

An argonaute protein from Pyrococcus furiosus.
PIWI domain is on the right, PAZ domain to the left.

SEQUENCE ANALYSIS

Align TtAgo with NgAgo

```
L0AJX6 L0AJX6_NATGS    373 -----PI--NTDLLDAIEAADRRVVETRRQGHGDDAVSFPQELLAVEPNTHQIKQFASDG    425
Q746M7 Q746M7_THET2    248 KDPRKPIPHLTGLLVPVL-----TLEDLHEEEGSLALSLPWEER--RRRTREIASWIGRR    300
                           **    * **  :      .:*  :: .*. *:*:* *    . .*::* .: .

L0AJX6 L0AJX6_NATGS    426 FHQQARSKTRLSASRCSEKAQAFAERLDPVRLNGSTVEFSSEFFTGNNEQQLRLLYENGE    485
Q746M7 Q746M7_THET2    301 LGLG-----------TPEAVRAQAYRLSIPKLMGRR------------------------    325
                           :              * .:* * **. :* *

L0AJX6 L0AJX6_NATGS    486 SVLTFRDGARGAHPDETFSKGIVNPPESFEVAVVLPEQQADTCKAQWDTMA--DLLNQAG    543
Q746M7 Q746M7_THET2    326 -------AVSKPADALRVGFYRAQETALALLRL------DGAQGWPEFLRRALLRAFG    370
                                  ::* ::: *:. *: . :*       *  :   **. *

L0AJX6 L0AJX6_NATGS    544 APPTRSETVQYDAFSSPESISLNVAGAI---DPSEVDAAFVVLPPDQEGFADLASPTETY    600
Q746M7 Q746M7_THET2    371 ASGA---SLRLHTLHAHPSQGLAFREALRKAKEEGVQAVLVLTPPMA---------WEDR    418
                           *  :    :::  .:: :  *  .*.  *:   . . *:*.:*: **          *

L0AJX6 L0AJX6_NATGS    601 DELKKALANMGIYSQMAYFDRFRDAKIFYTRNVALGLLAAAGGVAFTTEHAMPGDADMFI    660
Q746M7 Q746M7_THET2    419 NRLKALLLREGLPSQILNVP-LREEERHRWENALLGLLAKAGLQVVALSGAYPAEL--AV    475
                           :.**  *  . *: **:    .  :*: : .  .*. ***** **  ..: . * *.:   :

L0AJX6 L0AJX6_NATGS    661 GIDVSRSYPEDGASGQINIAATATAVYKDGTILGHSSTRPQLGEKLQSTDVRDIMKNAIL    720
Q746M7 Q746M7_THET2    476 GFDA-------GGRESFRFGGAACAVGGDGGHLLWTLPEAQAGERIPQEVVWDLLEETLW    528
                           *:*.         *.  .:.:..:* ** **  ** .* **:: . * * *:::::

L0AJX6 L0AJX6_NATGS    721 GYQQVTGESPTHIVIHRDGFMNE-DLDPATEFLNEQGVEYDIVEIRKQPQTRLLAVSDVQ    779
Q746M7 Q746M7_THET2    529 AFRRKAGRLPSRVLLLRDGRVPQDEFALALEALAREGIAYDLVSVRKSGGGRVYPVQGRL    588
                           .:::.:*. *::::: *** : : :  * * * .:*: **:*.:**.  *:  *.

L0AJX6 L0AJX6_NATGS    780 YDTPVKSIAAINQNEPRATVATFGAPEYLATRDGGGLPRPIQIERVAGETDIETLTRQVY    839
Q746M7 Q746M7_THET2    589 ADGLYVPLE------------DKTFLLLTVHRDFRGTPRPLKLVHEAGDTPLEALAHQIF    636
                           *    :             . ** * ***::: : **:* :*:*::*::

L0AJX6 L0AJX6_NATGS    840 LLSQSHIQVHNSTARLPITTAYADQASTHATKGYLVQTGAFESNVGFL-    887
Q746M7 Q746M7_THET2    637 HLTRLYPASGFAFPRLPAPLHLADRLVKEVGRLGIRHLKEVDREKLFFV    685
                           *:: :      :  ***      **:  ...: : :   .: :  *:
```
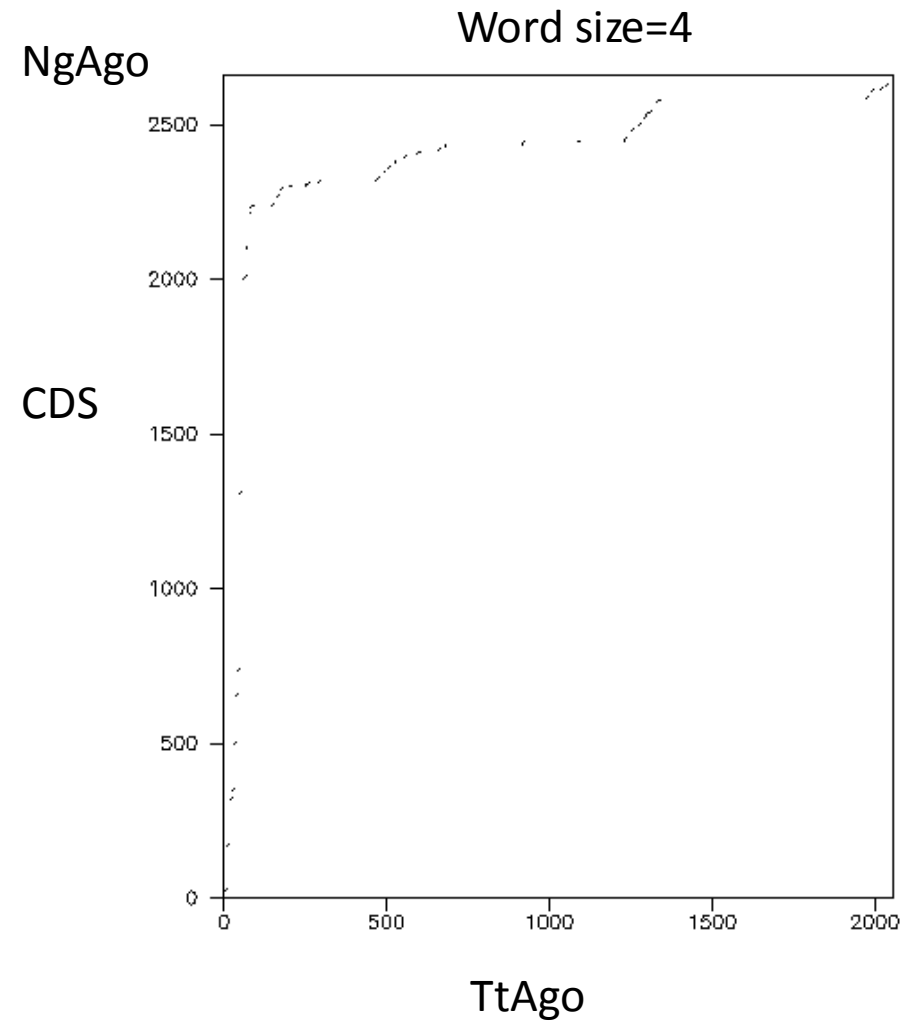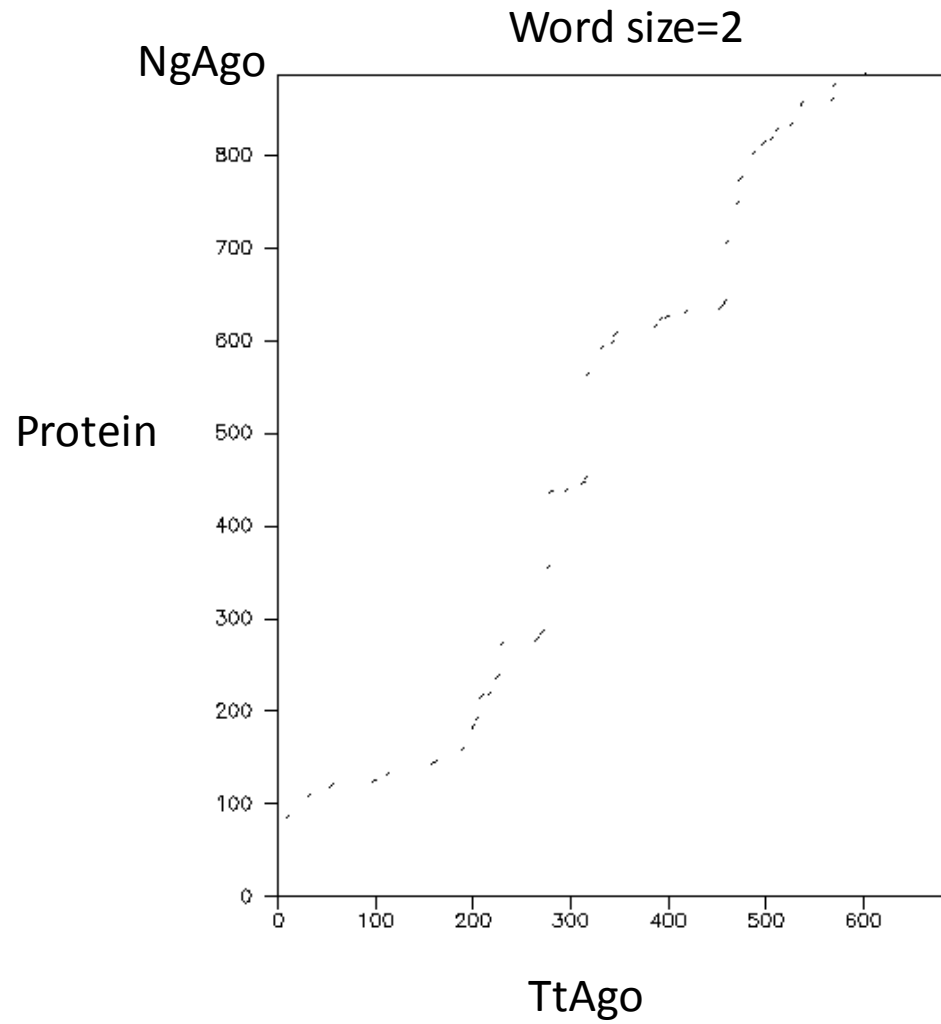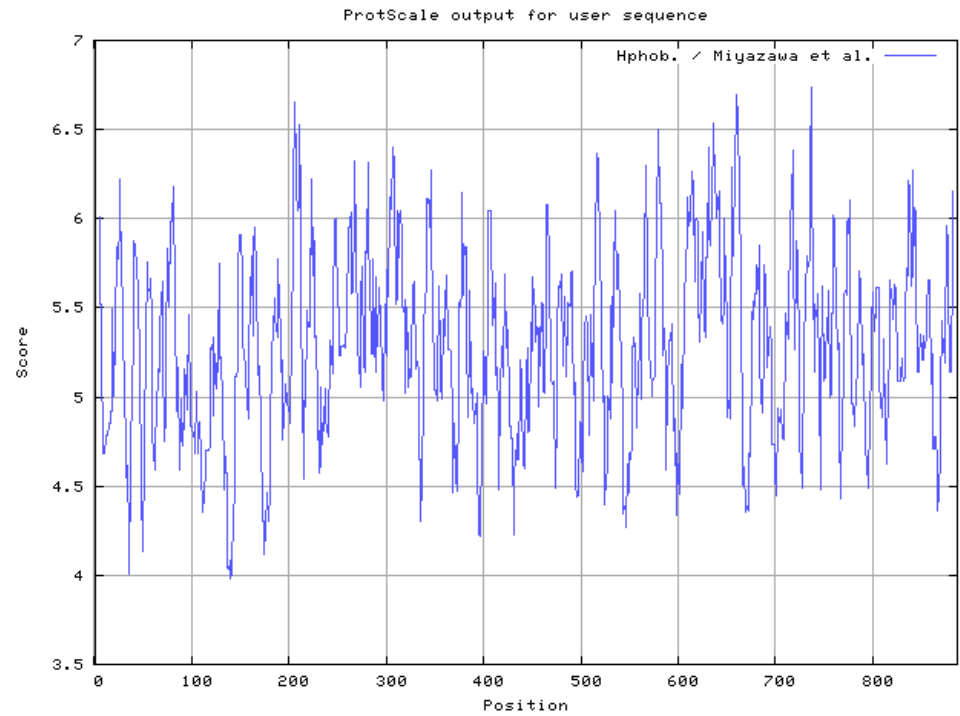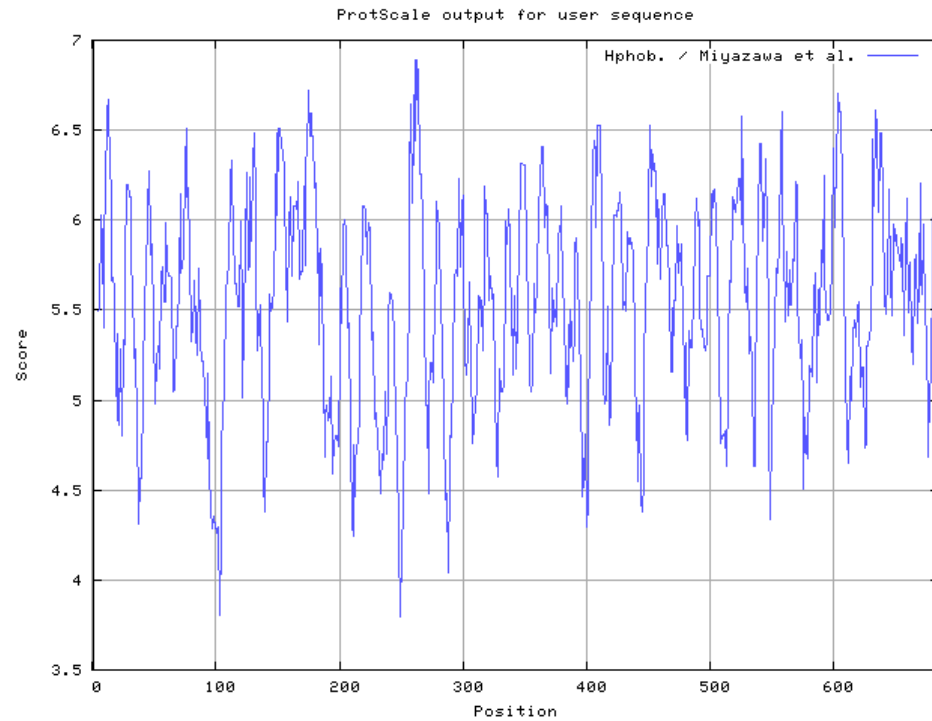
- Matrix: BLOSUM62
- Gap penalty: 10.0
- Extend penalty: 0.5
- Length: 1015

- Identity: 194/1015 (19.1%)
- Similarity: 280/1015 (27.6%)
- Gaps: 458/1015 (45.1%)
- Score: 250.0

Alignment result of TtAgo and NgAgo (PIWI domain alignment shown)
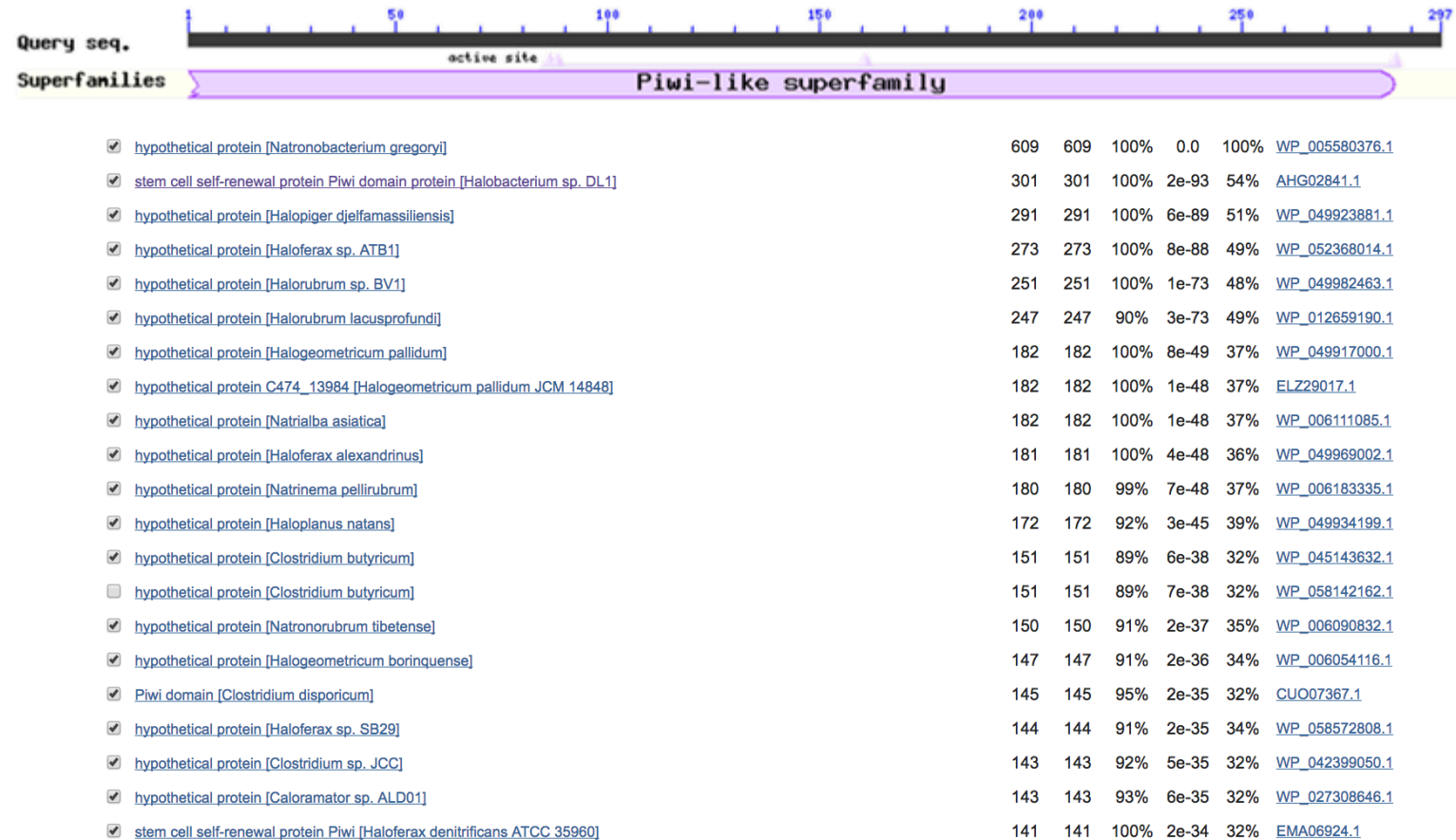
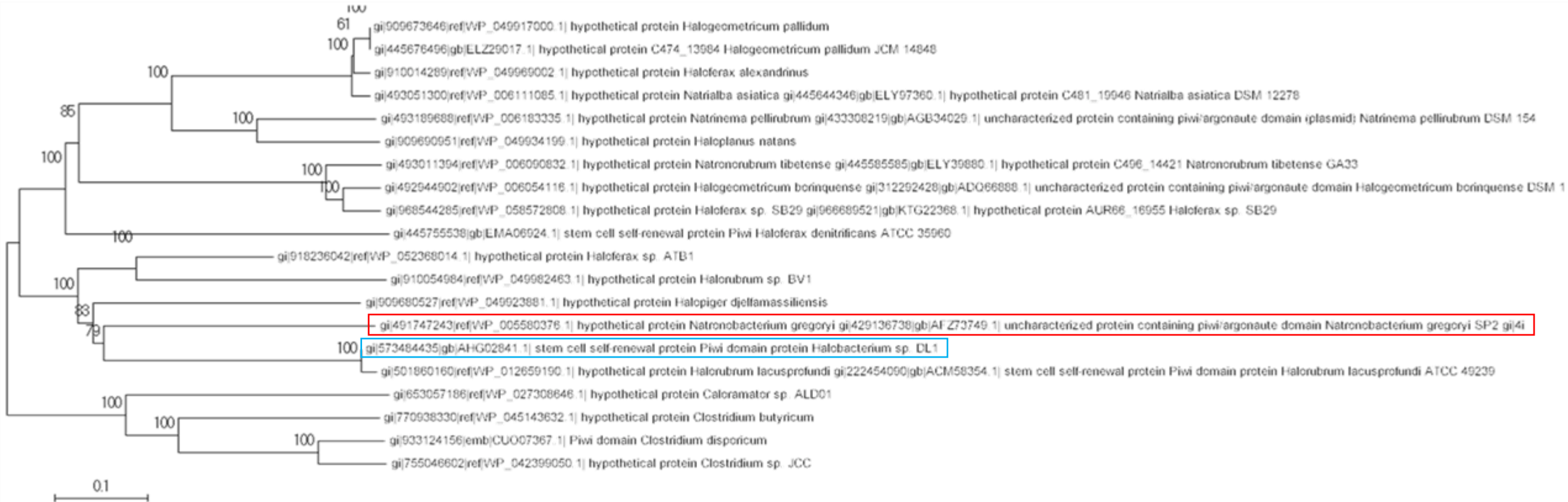# DotPath Results

ProtScale- Hydrophobicity

PHYLOGENY CONSTRUCTION

NgAgo PIWI Domain PSI-BLAST

☑	hypothetical protein [Natronobacterium gregoryi]	609	609	100%	0.0	100%	WP_005580376.1
☑	stem cell self-renewal protein Piwi domain protein [Halobacterium sp. DL1]	301	301	100%	2e-93	54%	AHG02841.1
☑	hypothetical protein [Halopiger djelfamassiliensis]	291	291	100%	6e-89	51%	WP_049923881.1
☑	hypothetical protein [Haloferax sp. ATB1]	273	273	100%	8e-88	49%	WP_052368014.1
☑	hypothetical protein [Halorubrum sp. BV1]	251	251	100%	1e-73	48%	WP_049982463.1
☑	hypothetical protein [Halorubrum lacusprofundi]	247	247	90%	3e-73	49%	WP_012659190.1
☑	hypothetical protein [Halogeometricum pallidum]	182	182	100%	8e-49	37%	WP_049917000.1
☑	hypothetical protein C474_13984 [Halogeometricum pallidum JCM 14848]	182	182	100%	1e-48	37%	ELZ29017.1
☑	hypothetical protein [Natrialba asiatica]	182	182	100%	1e-48	37%	WP_006111085.1
☑	hypothetical protein [Haloferax alexandrinus]	181	181	100%	4e-48	36%	WP_049969002.1
☑	hypothetical protein [Natrinema pellirubrum]	180	180	99%	7e-48	37%	WP_006183335.1
☑	hypothetical protein [Haloplanus natans]	172	172	92%	3e-45	39%	WP_049934199.1
☑	hypothetical protein [Clostridium butyricum]	151	151	89%	6e-38	32%	WP_045143632.1
☐	hypothetical protein [Clostridium butyricum]	151	151	89%	7e-38	32%	WP_058142162.1
☑	hypothetical protein [Natronorubrum tibetense]	150	150	91%	2e-37	35%	WP_006090832.1
☑	hypothetical protein [Halogeometricum borinquense]	147	147	91%	2e-36	34%	WP_006054116.1
☑	Piwi domain [Clostridium disporicum]	145	145	95%	2e-35	32%	CUO07367.1
☑	hypothetical protein [Haloferax sp. SB29]	144	144	91%	2e-35	34%	WP_058572808.1
☑	hypothetical protein [Clostridium sp. JCC]	143	143	92%	5e-35	32%	WP_042399050.1
☑	hypothetical protein [Caloramator sp. ALD01]	143	143	93%	6e-35	32%	WP_027308646.1
☑	stem cell self-renewal protein Piwi [Haloferax denitrificans ATCC 35960]	141	141	100%	2e-34	32%	EMA06924.1

Search NgAgo PIWI domain (578-874aa) with PSI-BLAST

NgAgo PIWI Domain PSI-BLAST

Top 20 of the NgAgo PIWI domain PSI-BLAST resulting proteins (NgAgo included)
(PIWI domain shown)

Phylogenetic Tree Built by 20 Proteins

NgAgo   Top 1 Result (HdAgo)

Neighbour-joining, Bootstrap 100

Top 1 Resulted Protein (HdAgo)

stem cell self-renewal protein Piwi domain protein [Halobacterium sp. DL1]
Sequence ID: gb|AHG02841.1| Length: 759 Number of Matches: 1

Range 1: 454 to 746 GenPept Graphics   ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
301 bits(770)	2e-93	Compositional matrix adjust.	159/297(54%)	205/297(69%)	4/297(1%)

```
Query  1     AAFVVLPPDQEGFADLASPTETYDELKKALANMGIYSQMAYFDRFRDAKIFYTRNVALGL  60
             AA V+LP       + +S ++ Y E+KKAL    + SQMA+ D   A +    NVALGL
Sbjct  454   AACVILPKADFSMGE-SSASDIYHEMKKALRQRRVDSQMAHIDTL--ATSYALPNVALGL  510

Query  61    LAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQINIAATATAVYKDGTILGHSS  120
             +AAAGG+ FTTE AMPG+ D+FIGIDVS   YP D    +++IAA+ T++Y DGTILG++S
Sbjct  511   VAAAGGIPFTTEDAMPGETDLFIGIDVSHRYPRD-TDERVHIAASTTSIYGDGTILGYTS  569

Query  121   TRPQLGEKLQSTDVRDIMKNAILGYQQVTGESPTHIVIHRDGFMNEDLDPATEFLNEQGV  180
             +PQ GEK+    +++++ + +I GY+Q  GE P  IVIHRDGFM EDLD   E L    +
Sbjct  570   AKPQTGEKVPPKELKNLTRQSIAGYKQEHGEYPDRIVIHRDGFMREDLDQVEEMLESMDI  629

Query  181   EYDIVEIRKQPQTRLLAVSDVQYDTPVKSIAAINQNEPRATVATFGAPEYLATRDGGGLP  240
              YD+VEIRKQ    R+L ++D     P K IAA+N+ E RA +ATFG PE  AT    GLP
Sbjct  630   NYDVVEIRKQSPARVLNLADGVAKIPDKGIAALNREENRAILATFGDPESQATSSNTGLP  689

Query  241   RPIQIERVAGETDIETLTRQVYLLSQSHIQVHNSTARLPITTAYADQASTHATKGYL  297
             +PIQ+ER AG+TDI+TLT QVYLLSQSH+   NSTARLPITT YAD+AS  A +GYL
Sbjct  690   QPIQVERKAGDTDIKTLTAQVYLLSQSHVGAMNSTARLPITTYYADRASEAAAEGYL  746
```
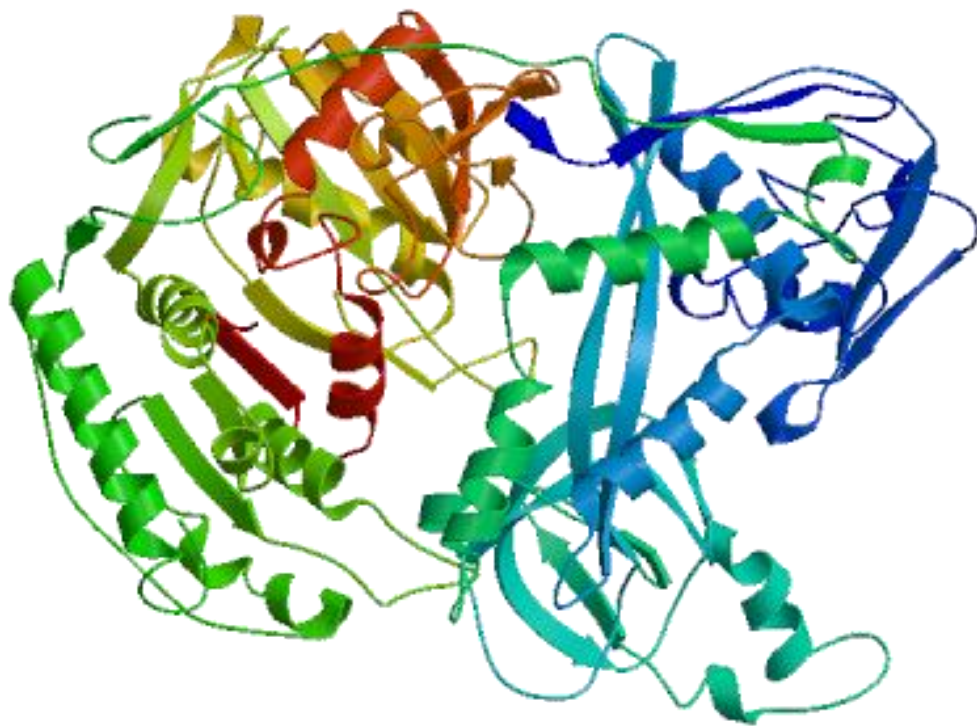
Argonaute Protein found in *Halobacterium sp. DL1* (PIWI domain: 454-746aa)

DotPath Result

Word size=4
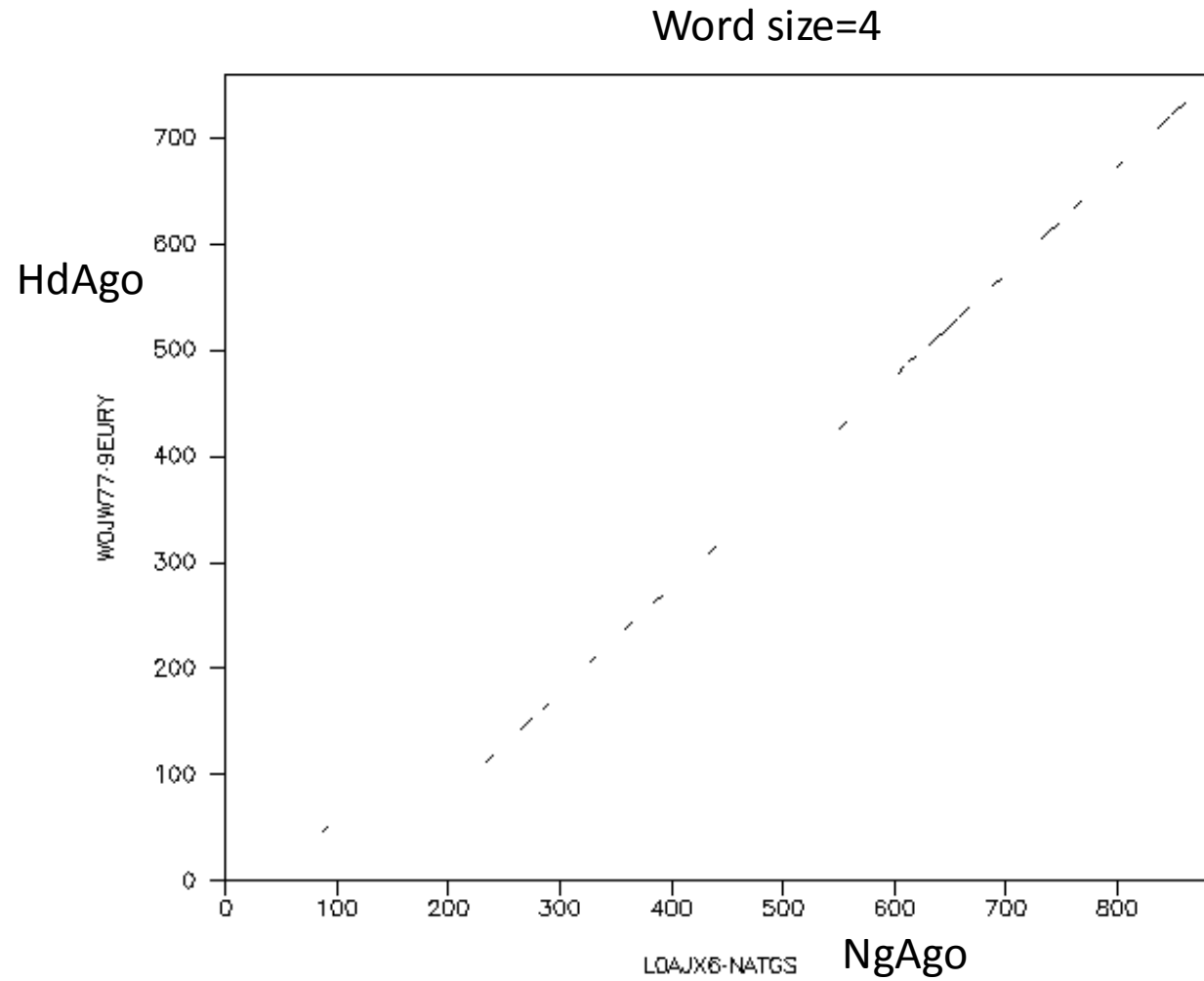
3D STRUCTURE PREDICTION

NgAgo Structure

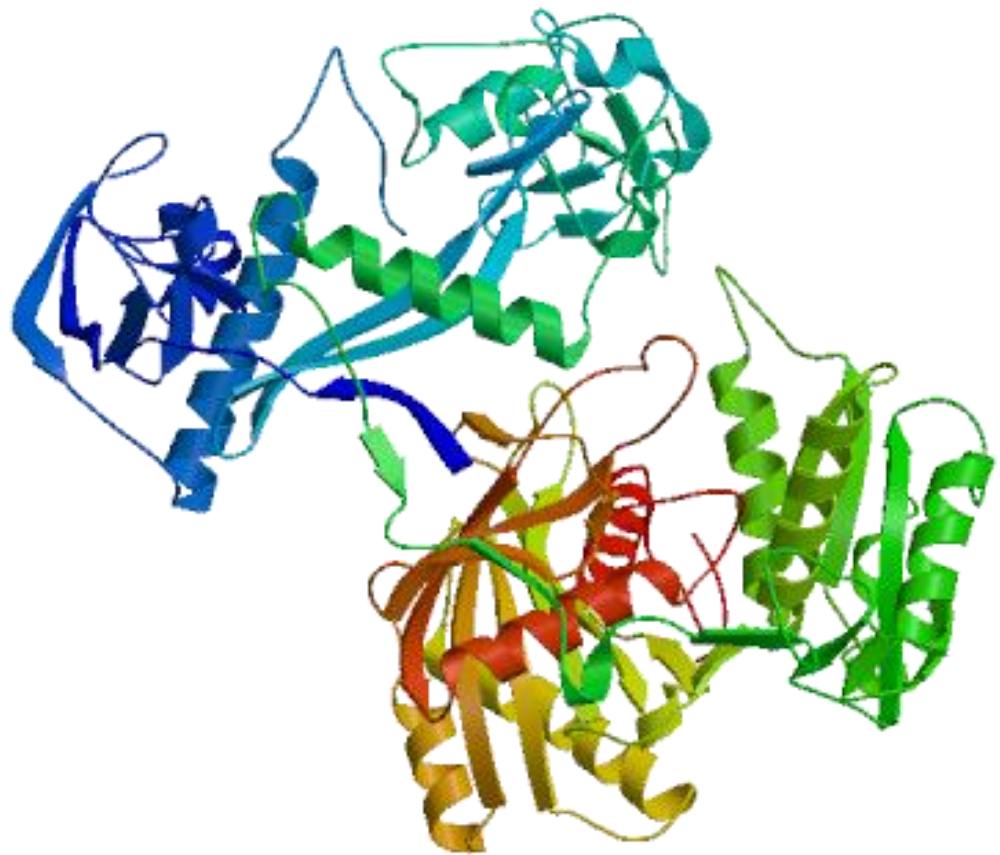| Model #02	File	Built with	Oligo-State	Ligands	GMQE	QMEAN4
	PDB	ProMod Version 3.70.	MONOMER	None	0.41	-9.77

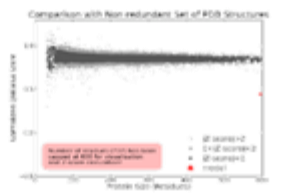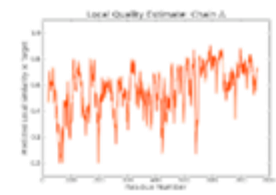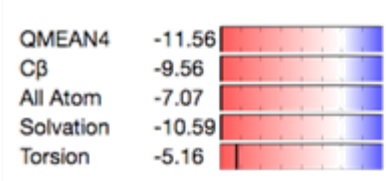|	
QMEAN4	-9.77
Cβ	-8.06
All Atom	-6.32
Solvation	-6.85
Torsion	-5.36
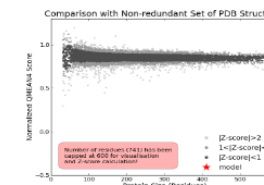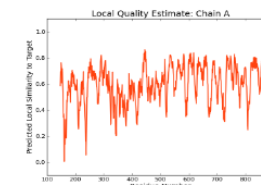
Among 3 possible models, we chose Model #02 as its QMEAN4 was the highest.

HdAgo (No.1 Candidate) Structure

Model #01	File	Built with	Oligo-State	Ligands	GMQE	QMEAN4
	PDB	ProMod Version 3.70.	MONOMER	None	0.47	-11.56

QMEAN4	-11.56	
Cβ	-9.56	
All Atom	-7.07	
Solvation	-10.59	
Torsion	-5.16	

Among 3 possible models, we chose Model #01 as its QMEAN4 was the highest.

Protein-Fitting

Fit NgAgo with HdAgo

Fit NgAgo with HdAgo (PIWI domain only)

Conclusion

- TtAgo-NgAgo Similarity: Relatively Low

- Newly recognized HdAgo : PIWI Domain Similarity With NgAgo– Acceptable (69%)

- HdAgo: Possibly Powerful Nucleic acid-guided Genome Editing Tool

Tools

-Pubmed, Webofscience and ReadCube for Paper Search

-UniProt, NCBI, ExPASy and EMBOSS as Databases

-MEGA as Phylogenetic Analysis Tool

-Phyre2, SwissModel and SPDBv as Protein 3D Model Tool

References

1. Hutvagner, G. & Simard, M. J. Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* **9,** 22–32 (2008).

2. Swarts, D. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507,** 258–261 (2014).

3. Sheng, G. *et al.* Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proceedings of the National Academy of Sciences* **111,** 652–657 (2014).

4. Gao, F., Shen, X. Z., Jiang, F., Wu, Y. & Han, C. DNA-guided genome editing using the Natronobacterium gregoryi Argonaute.*Nat. Biotechnol.* (2016). doi:10.1038/nbt.3547

Acknowledgements