Studies on the Evolution, Structure, and Function of Homing Endonuclease Containing Parasitic Genetic Elements

Alireza G Senejani

M. S. Biotechnology, Germany
B. S. Biochemistry, India

Ph. D. Defense; May 10th, 2007

Graduate Program in Genetics
Department of Molecular and Cell Biology
University of Connecticut

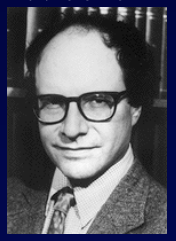
Intron: for the first time (1977)

Mosaic:

"...mosaic molecules consisting of sequences complementary to several non-contiguous segments of the viral genome..." (Sambrook, 1977)

Intron:

"...regions which will be lost from the mature messenger—which I suggest we call introns (for intragenic regions)-alternating with regions which will be expressed- exons..." (Gilbert, 1978)



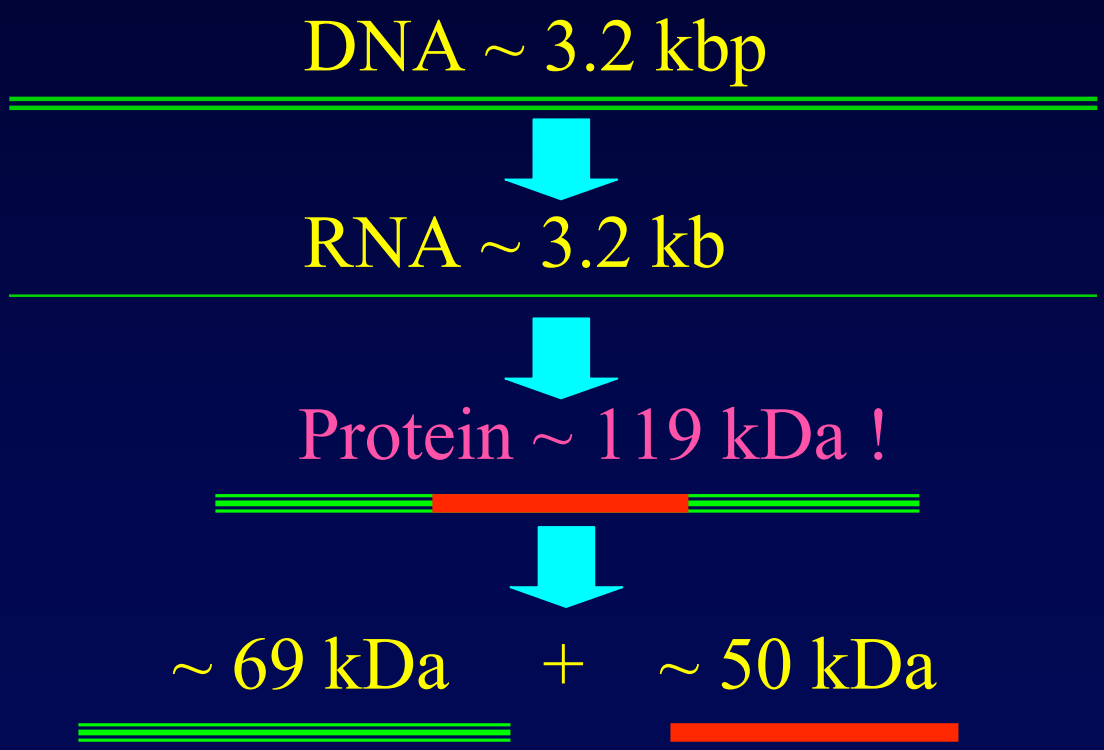
1993 Nobel Prize in Physiology or Medicine

Phillip A. Sharp and Richard J. Roberts

"Split-genes"

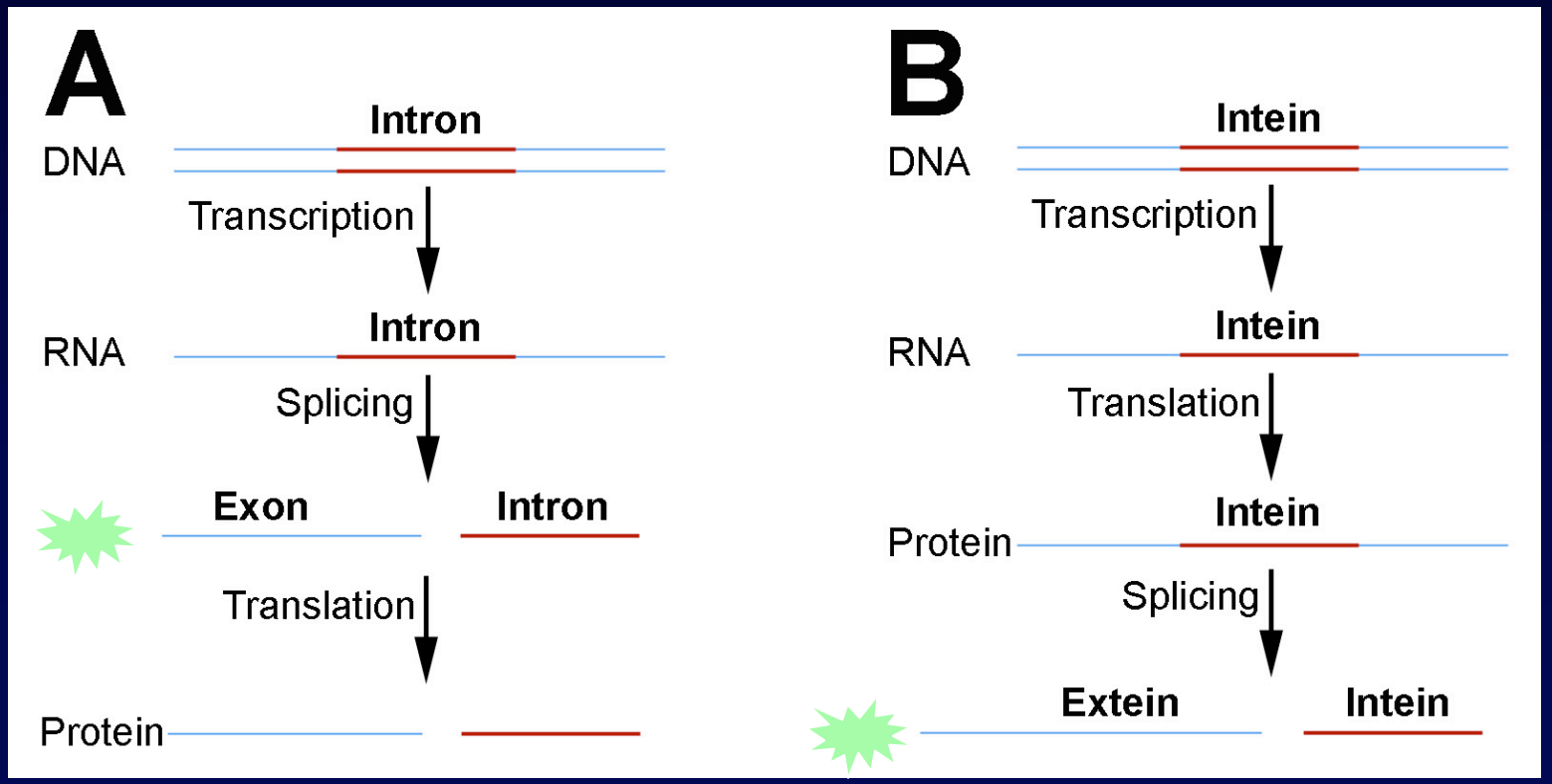
Intein: for the first time (1988)

- *Neff's group, Anraku's lab, Stevens's group* *(1st paper 1990)*
  - *Saccharomyces cerevisiae* Vacuolar ATPase, subunit A (*Sce*-VMA1)



Intervening Protein (Protein's Intron)

DNA	Protein
Exon	*Extein*
Intron	*Intein*



Distribution of Inteins

Domains of Life	Number of Species	Number of Inteins
Eukaryotes	63	80
Eubacteria	82	180
Archaea	28	134
Total	173	394

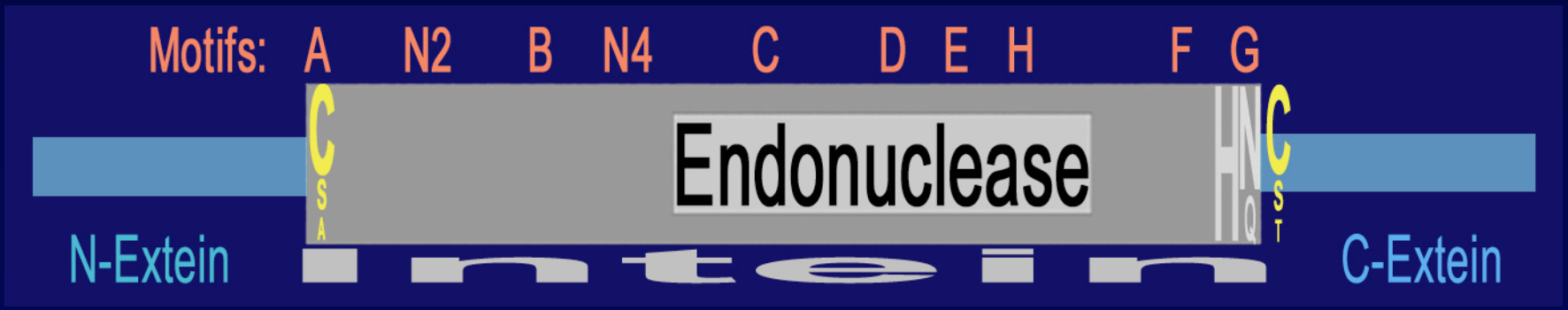
http://www.neb.com/neb/inteins.html

➔ **Different proteins with diverse functions:**

Vacuolar-type ATPase, Cell division, metabolic enzymes, DNA and RNA polymerases, proteases, ribonucleotide reductases, and more…

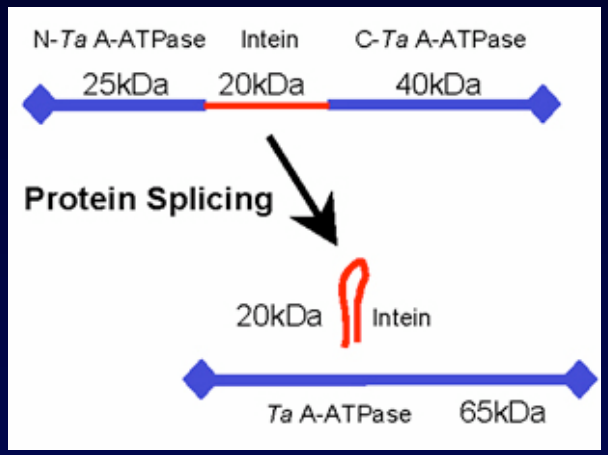
Intein Types and Functions

- Splicing activity (large and mini intein)
- Endonuclease activity (large intein)
- Other functions???



Thermoplasma mini-intein

- Euryarchaeota Archaea {pH ~ 1.5, temp ~ 59° C}
- Archaeal type-ATPase A-Subunit
- 173 a.a. (mini intein)



Protein expression problem: ???

The *Thermoplasma acidophilium* A-ATPase A-subunit **tRNA** codon usage rates compare to the host (*E.coli*)!

The intein of the Thermoplasma A-ATPase A subunit: Structure, evolution and expression in *E. coli*
Senejani AG, Hilario E, Gogarten JP; *BMC Biochemistry* 2001, **2**:13

CODON USAGE

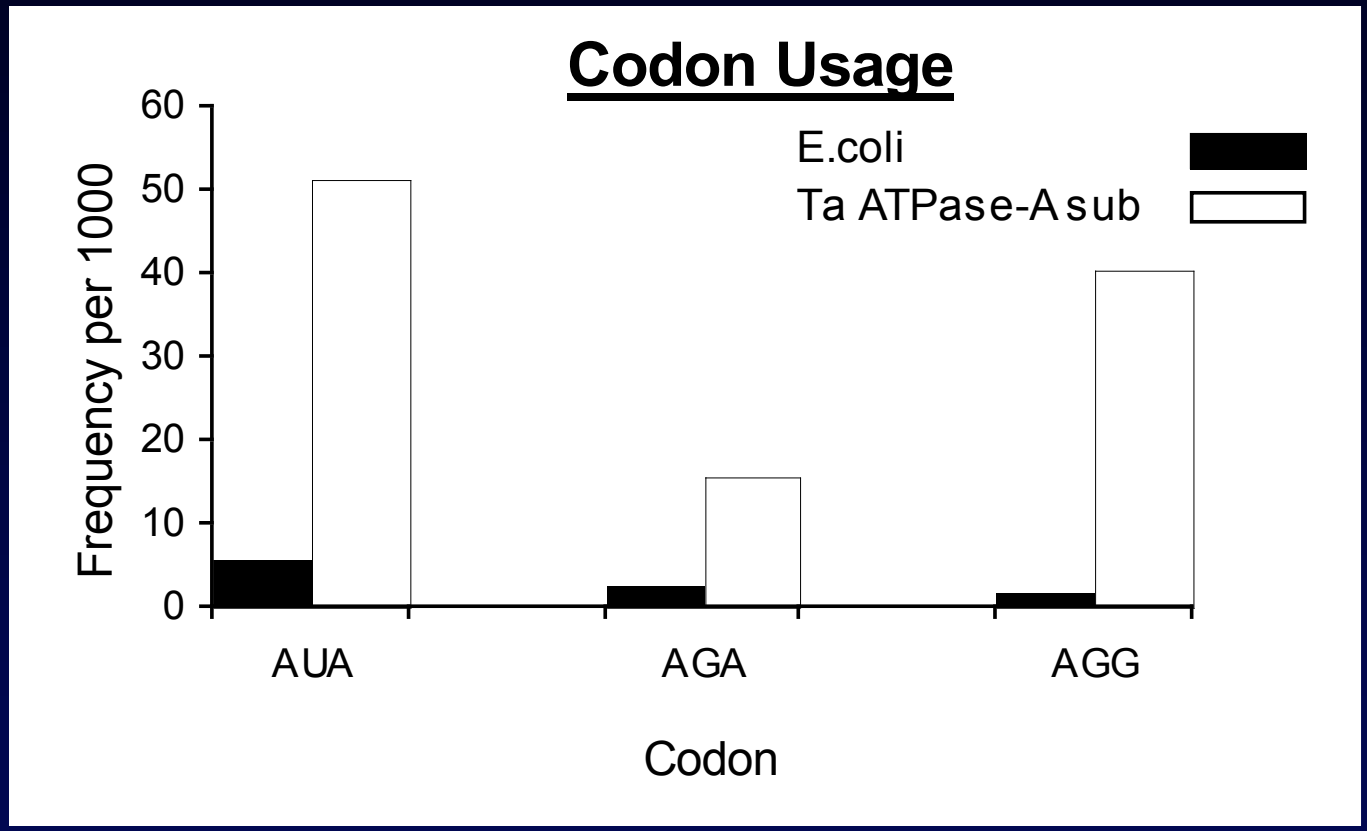
Escherichia coli:

UUU 22.3(30407)	UCU 8.5(11523)	UAU 16.2(22050)	UGU 5.2(7063)
UUC 16.6(22582)	UCC 8.6(11771)	UAC 12.2(16671)	UGC 6.5(8849)
UUA 13.9(18943)	UCA 7.2(9793)	UAA 2.0(2707)	UGA 0.9(1260)
UUG 13.7(18629)	UCG 8.9(12195)	UAG 0.2(326)	UGG 15.2(20756)
CUU 11.0(15019)	CCU 7.0(9572)	CAU 12.9(17631)	CGU 20.9(28471)
CUC 11.1(15105)	CCC 5.5(7491)	CAC 9.7(13275)	CGC 22.0(29970)
CUA 3.9(5316)	CCA 8.4(11497)	CAA 15.3(20913)	CGA 3.6(4860)
CUG 52.6(71716)	CCG 23.2(31617)	CAG 28.8(39288)	CGG 5.4(7404)
AUU 30.3(41375)	ACU 9.0(12228)	AAU 17.7(24189)	AGU 8.8(11982)
AUC 25.1(34264)	ACC 23.4(31891)	AAC 21.7(29534)	AGC 16.1(21908)
AUA 5.6(5967)	ACA 7.1(9684)	AAA 33.6(45821)	AGA 2.8(2899)
AUG 27.9(37995)	ACG 14.4(19682)	AAG 10.3(14080)	AGG 1.7(1694)
GUU 18.3(24916)	GCU 15.3(20811)	GAU 32.1(43817)	GGU 24.7(33738)
GUC 15.3(20800)	GCC 25.5(34770)	GAC 19.1(25999)	GGC 29.6(40400)
GUA 10.9(14855)	GCA 20.1(27470)	GAA 39.4(53783)	GGA 8.0(10902)
GUG 26.4(35983)	GCG 33.6(45866)	GAG 17.8(24313)	GGG 11.1(15118)

Thermoplasma acidophilium A-ATPase A-Subunit (763 codons):

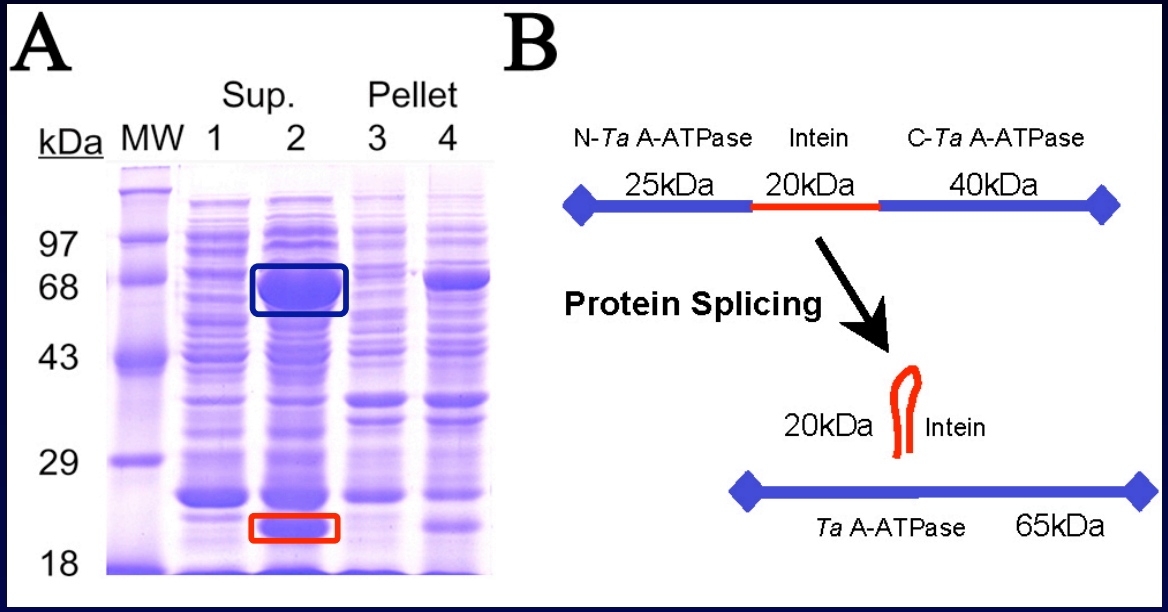
UUU 6.5(5)	UCU 10.5(8)	UAU 13.1(10)	UGU 2.6(2)
UUC 17.0(13)	UCC 15.7(12)	UAC 23.6(18)	UGC 3.9(3)
UUA 0.0(0)	UCA 15.7(12)	UAA 1.3(1)	UGA 0.0(0)
UUG 1.3(1)	UCG 11.8(9)	UAG 0.0(0)	UGG 13.1(10)
CUU 7.9(6)	CCU 7.9(6)	CAU 3.9(3)	CGU 1.3(1)
CUC 22.3(17)	CCC 9.2(7)	CAC 10.5(8)	CGC 7.9(6)
CUA 2.6(2)	CCA 13.1(10)	CAA 2.6(2)	CGA 0.0(0)
CUG 35.3(27)	CCG 14.4(11)	CAG 26.2(20)	CGG 2.6(2)
AUU 9.2(7)	ACU 5.2(4)	AAU 14.4(11)	AGU 2.6(2)
AUC 15.7(12)	ACC 17.0(13)	AAC 26.2(20)	AGC 15.7(12)
AUA 51.0(39)	ACA 13.1(10)	AAA 19.6(15)	AGA 15.7(12)
AUG 30.1(23)	ACG 14.4(11)	AAG 24.9(19)	AGG 40.6(31)
GUU 23.6(18)	GCU 13.1(10)	GAU 39.3(30)	GGU 15.7(12)
GUC 20.9(16)	GCC 22.3(17)	GAC 15.7(12)	GGC 26.2(20)
GUA 24.9(19)	GCA 18.3(14)	GAA 28.8(22)	GGA 32.7(25)
GUG 19.6(15)	GCG 15.7(12)	GAG 60.2(46)	GGG 3.9(3)

Thermoplasma mini-intein



The intein of the Thermoplasma A-ATPase A subunit: Structure, evolution and expression in *E. coli*
Senejani AG, Hilario E, Gogarten JP; *BMC Biochemistry* 2001, **2**:13

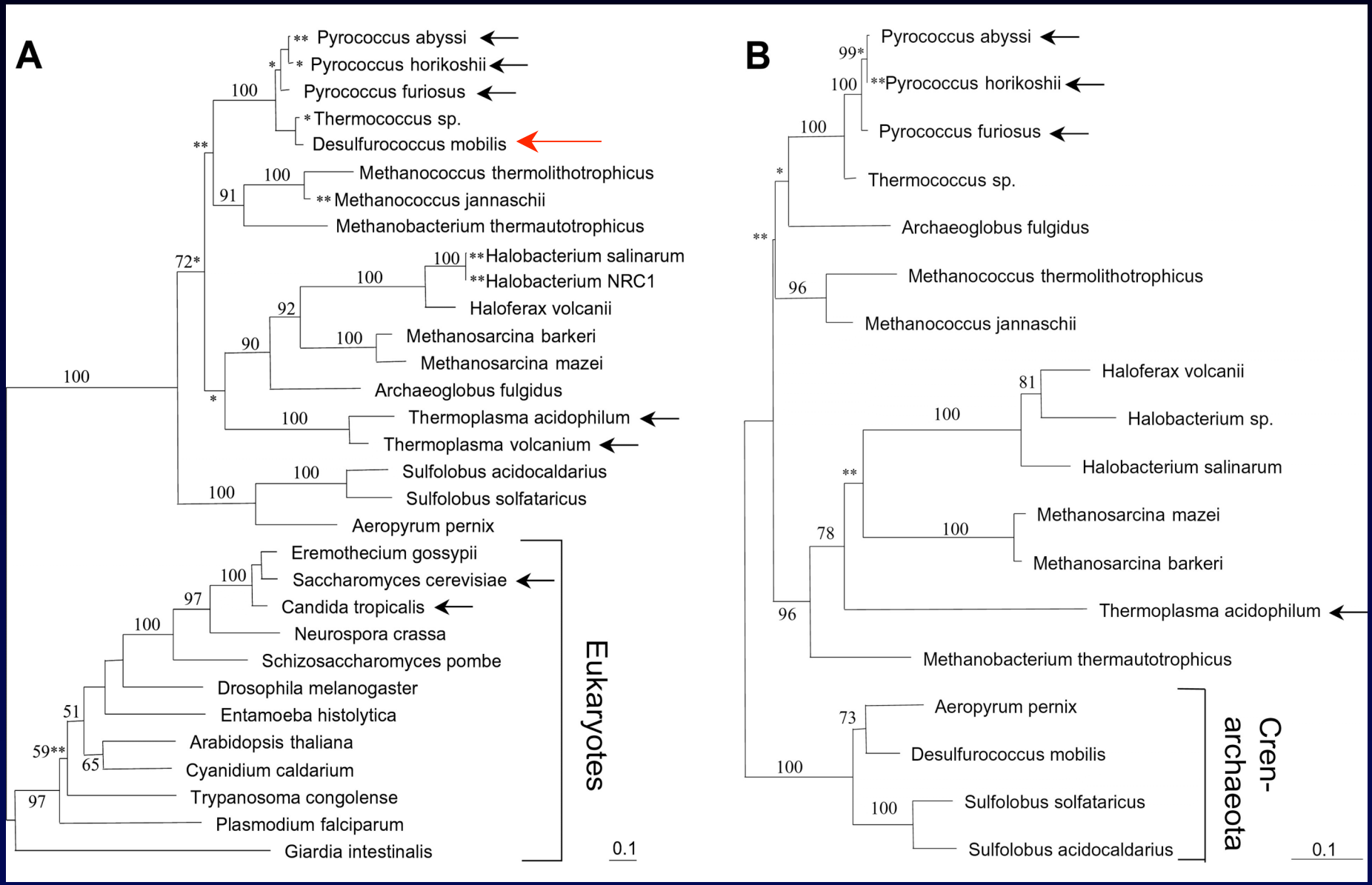
Thermoplasma mini-intein



SDS polyacrylamide gel electrophoresis of proteins from induced *E. coli.*
Panel A, Lanes 1 and 3: *E. coli* Bl21-CodonPlus(DE3)-RIL strain transformed with empty pET-11a vector (negative control); lanes 2 and 4: *E. coli* Bl21-CodonPlus(DE3)-RIL strain transformed with the *Thermoplasma* A-ATPase cloned into pET-11a

The intein of the Thermoplasma A-ATPase A subunit: Structure, evolution and expression in *E. coli*
Senejani AG, Hilario E, Gogarten JP; *BMC Biochemistry* 2001, **2**:13

Thermoplasma mini-intein



The intein of the Thermoplasma A-ATPase A subunit: Structure, evolution and expression in *E. coli*

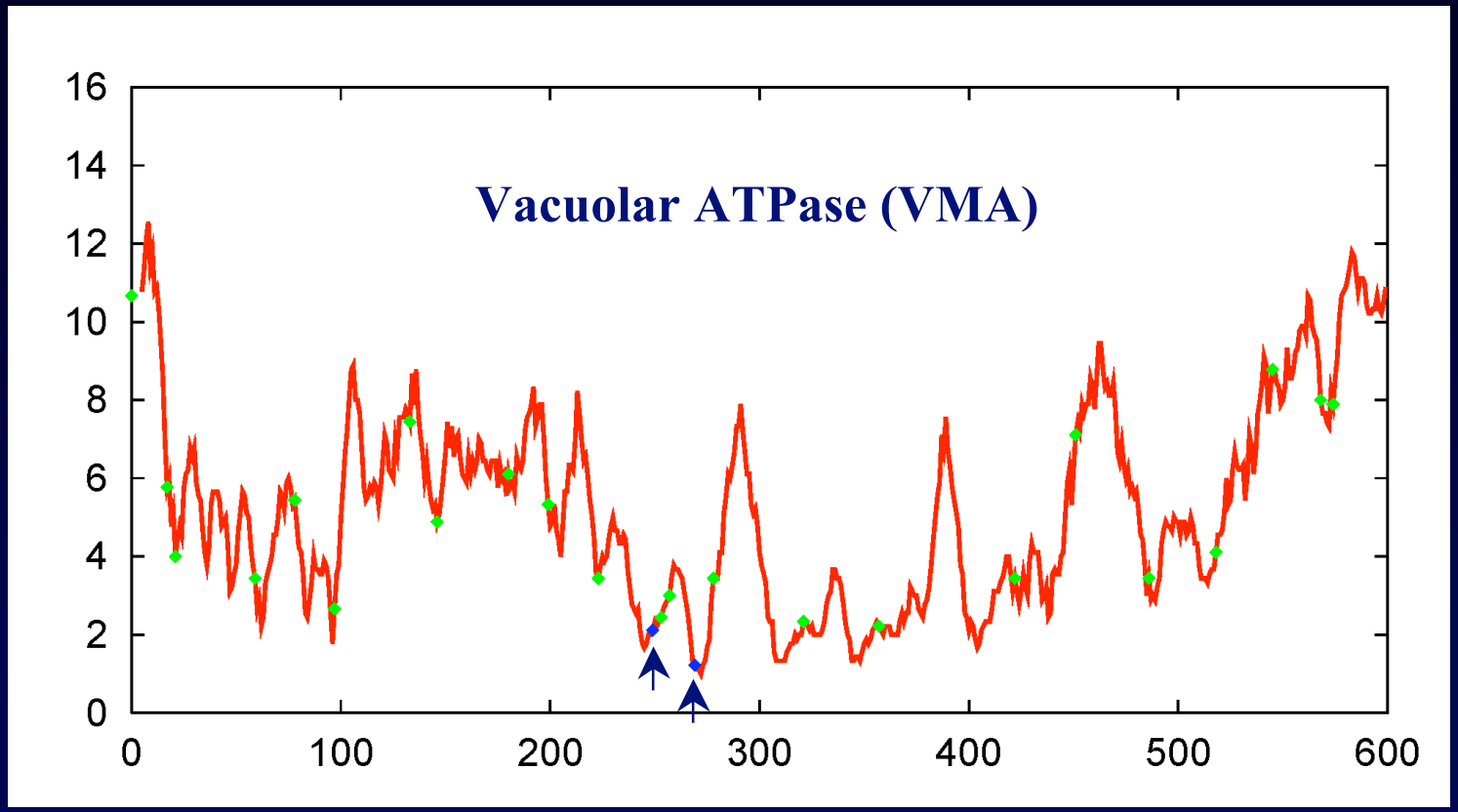
Senejani AG, Hilario E, Gogarten JP; *BMC Biochemistry* 2001, **2**:13

Inteins and Introns Insertion Sites

- Vacuolar ATPase Catalytic Subunit (VMA1)
- Replication Factor C (RFC)
- Cell Division Control Protein 21 (CDC21)
- DNA polymerase (POL)
- Cytochrome C Oxidase Subunit I (COX1)

Inteins and Spliceosomal Introns Insertion Sites

Inteins and Spliceosomal Introns Insertion Sites

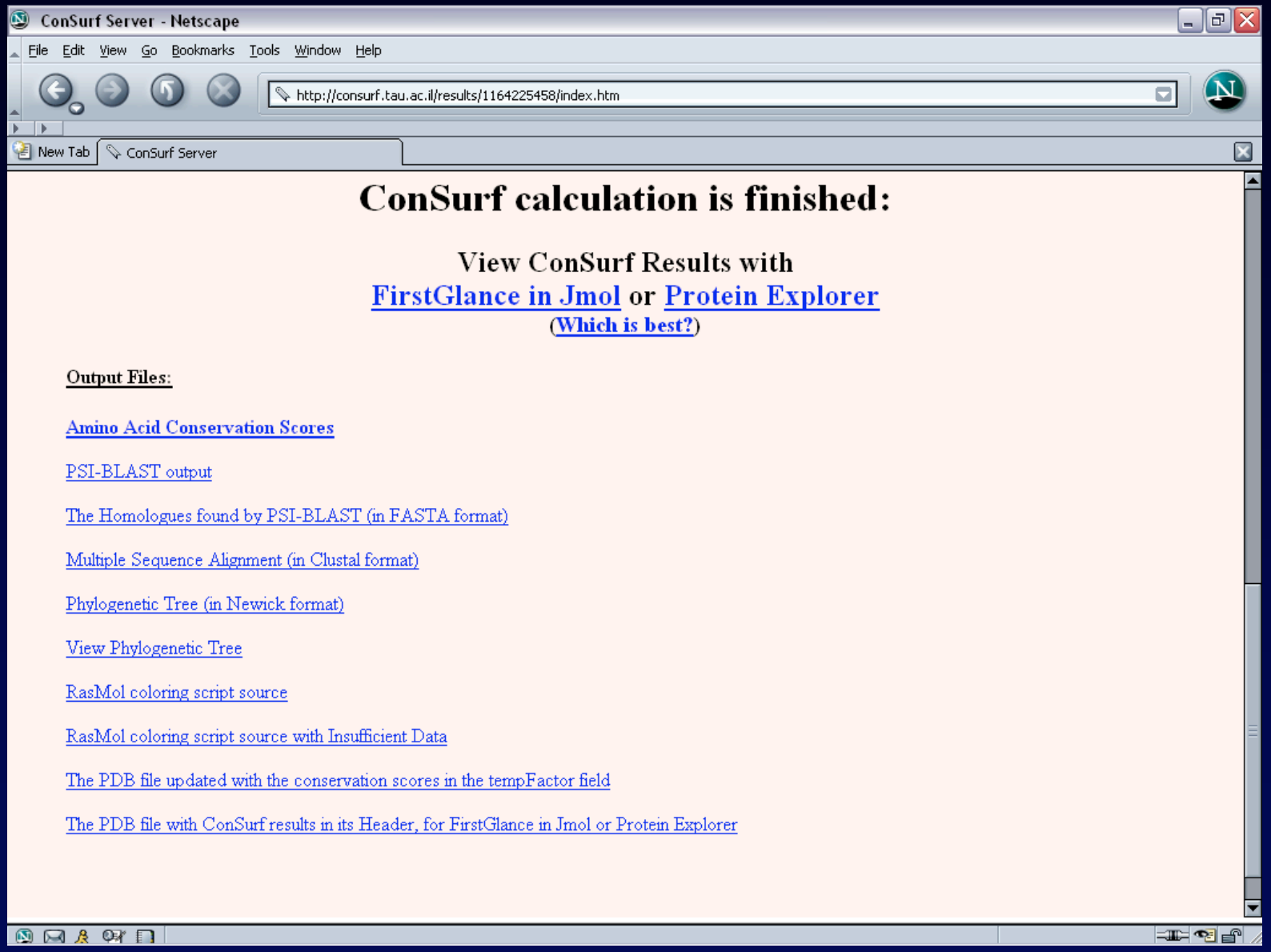


Inteins: Structure, Function, and Evolution
Gogarten, Senejani, Zhaxybayeva, Olendzenski, Hilario; Annu. Rev. Microbiol. 2002, 56:263-87

Inteins Insertion Sites

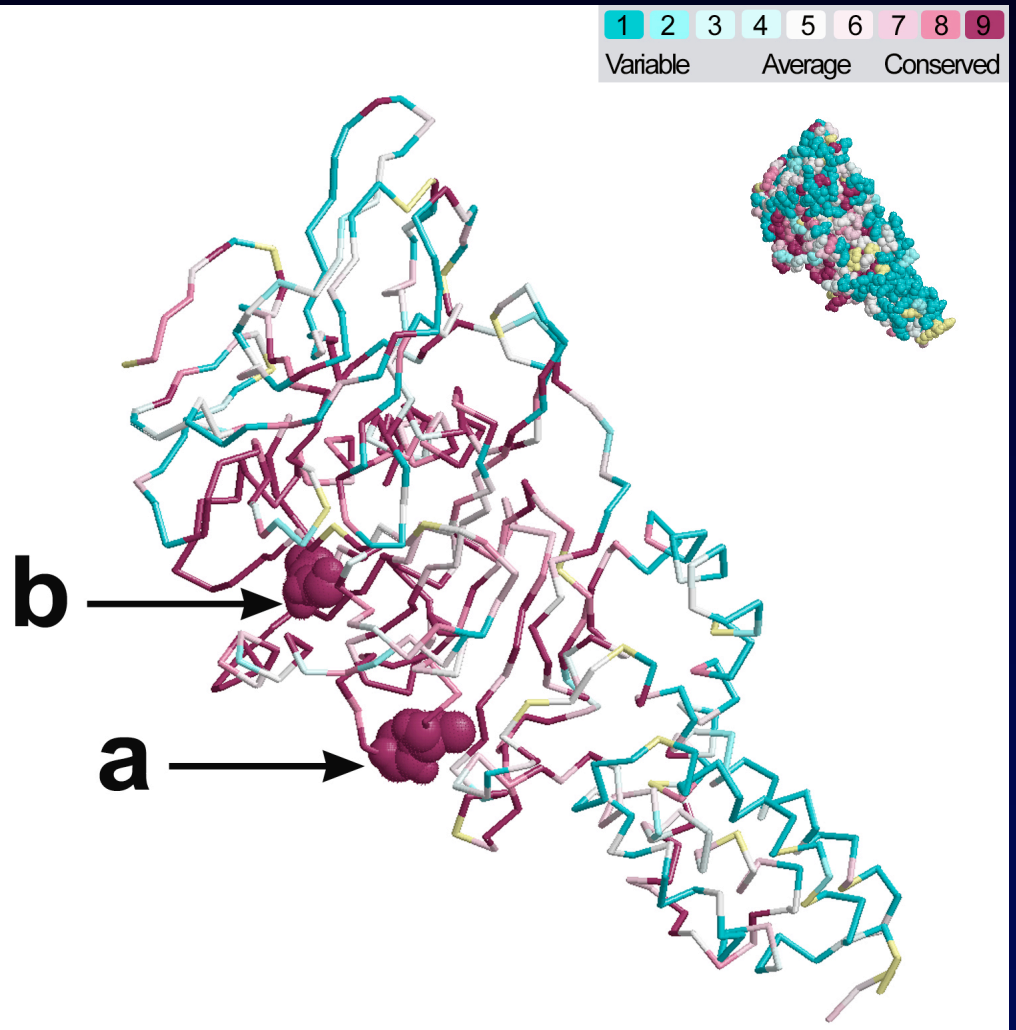
ConSurf: http://consurf.tau.ac.il/

Inteins Insertion Sites



ConSurf: http://consurf.tau.ac.il/

Inteins Insertion Sites



A-Type ATPase Catalytic Subunit A structure from *Pyrococcus Horikoshi* OT (PDB: 1vdz)

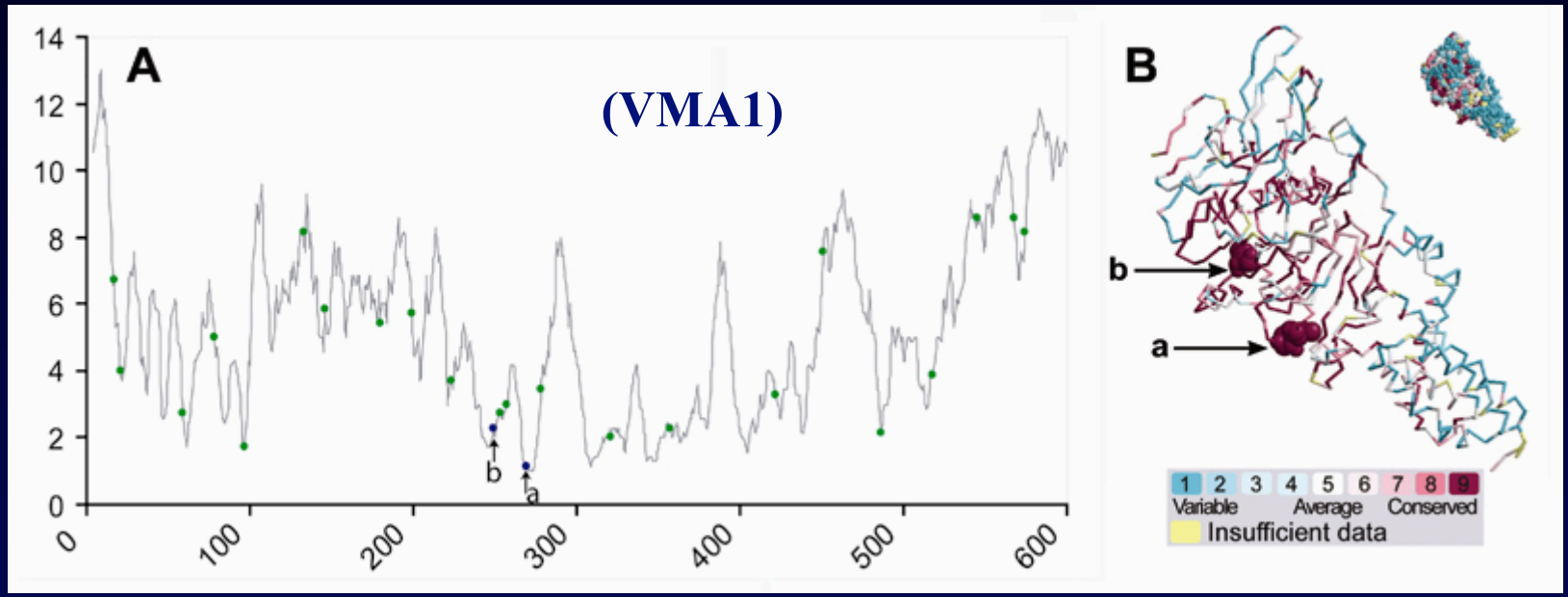
b insertion site (Archaeal):

between Lys240/Thr241

a insertion site (Eukaria):

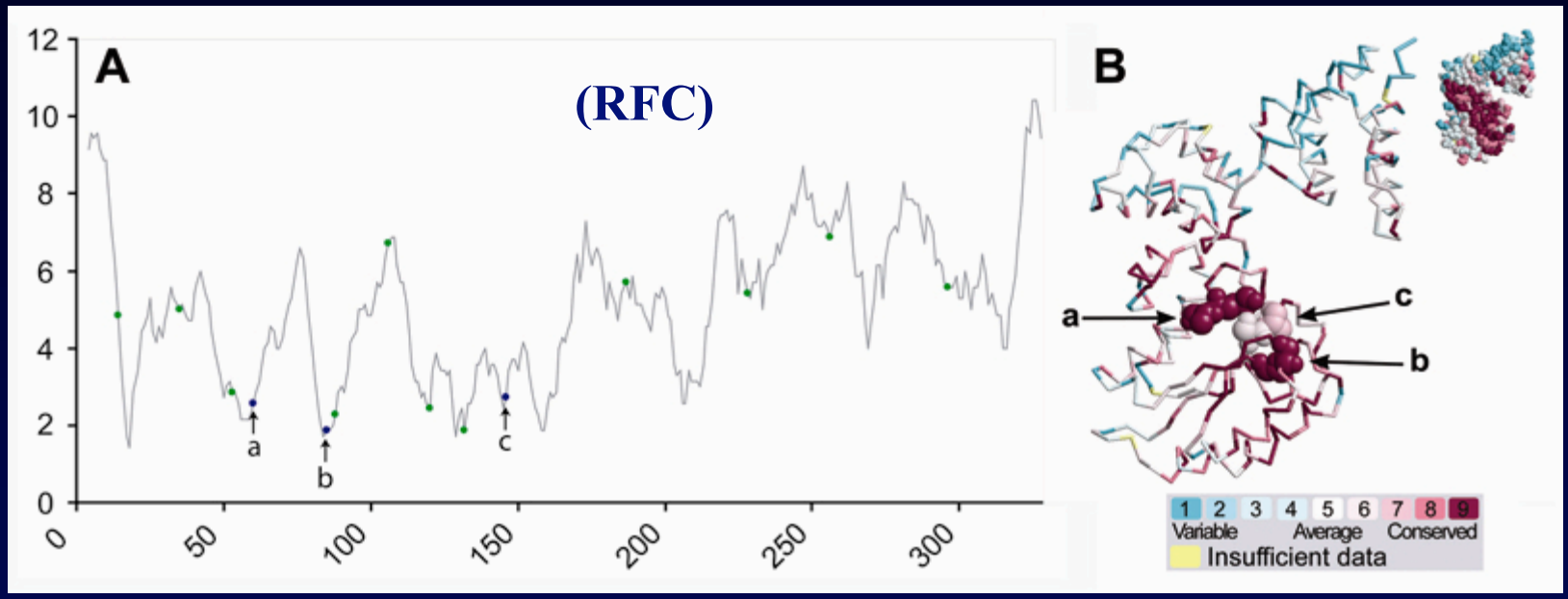
between Gly260/Cys261

Inteins and Spliceosomal Introns Insertion Sites



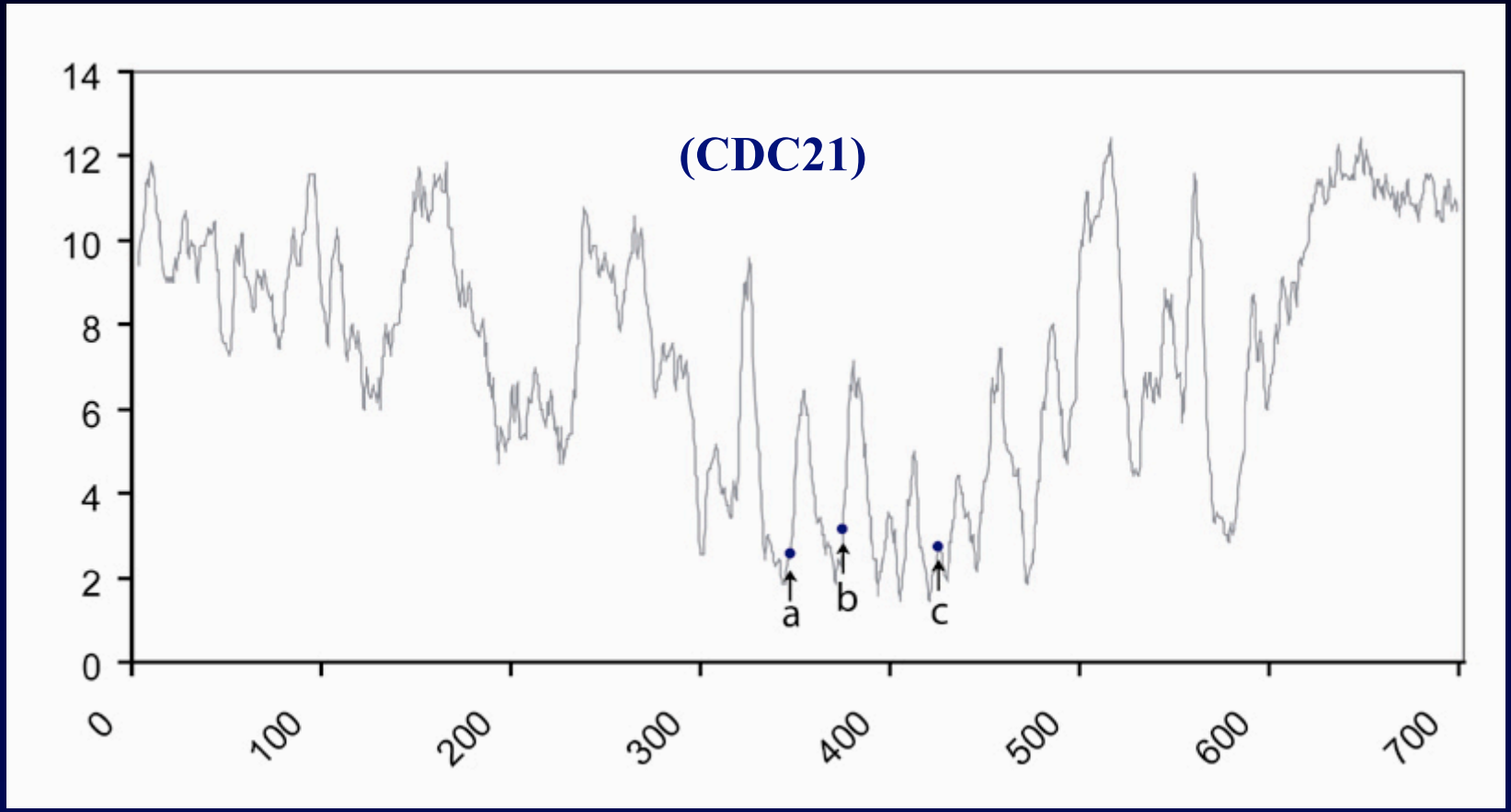
Positions of inteins (dots with arrows) and spliceosomal introns (dots) along the coding sequence (panel A) and in the structure of ATPase catalytic subunit A structure from *Pyrococcus horikoshii* OT3 (PDB ID: 1VDZ).

Inteins and Spliceosomal Introns Insertion Sites



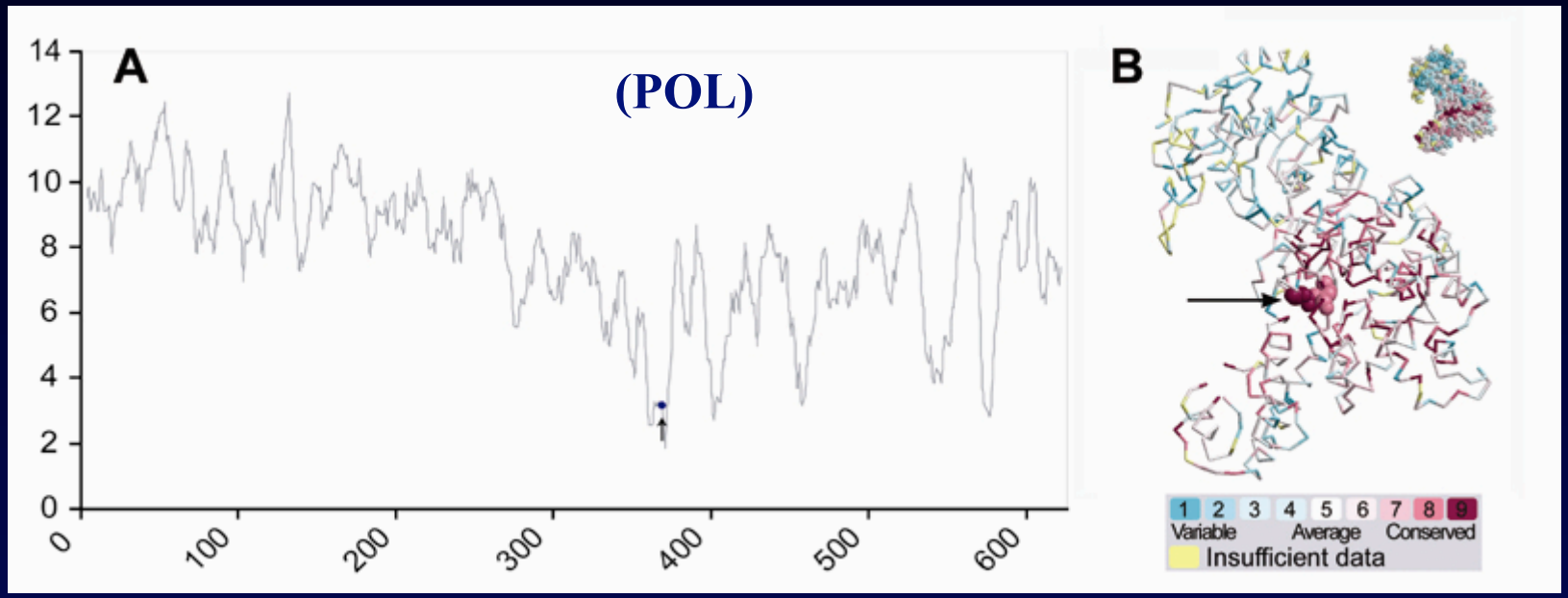
Positions of inteins (dots with arrows) and spliceosomal introns (dots) along the coding sequence (*panel A*) and in the structure of the *Archaeoglobus fulgidus* Replication Factor C (*panel B;* PDB ID: 2CHV).

Inteins Insertion Sites



Group I and Group II Introns Insertion Sites

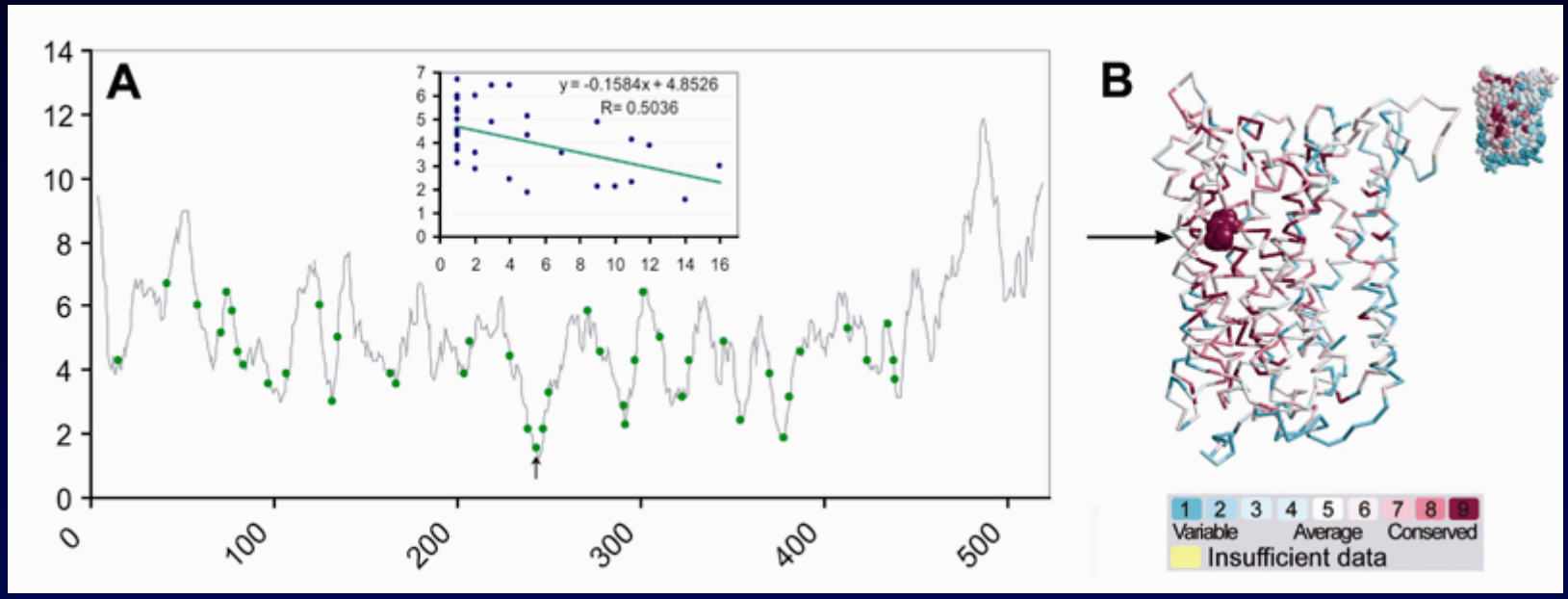
Group I and II Introns Insertion Sites



DNA pol *B. subtilis* phage SPO1 intron (with homing endonuclease)

insertion site: between Pro674/Asn675

Group I and II Introns Insertion Sites



Cytochrome C Oxidase Subunit I (COX1)

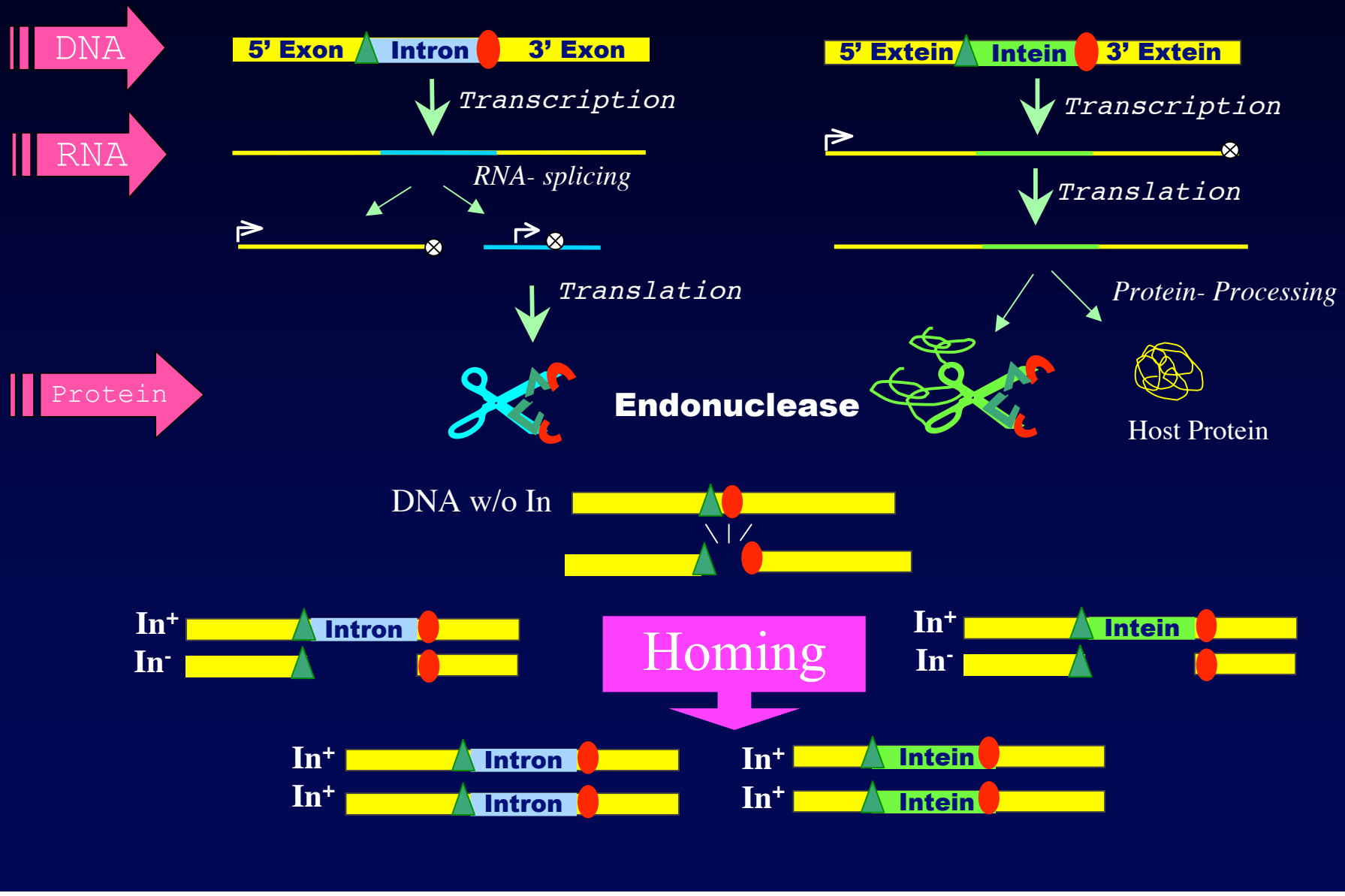
Fungi, vascular plants (dot with arrow), the green algae, the liveworts, the soil-living amoeba, and the single-cell protist

Homing Endonuclease Function

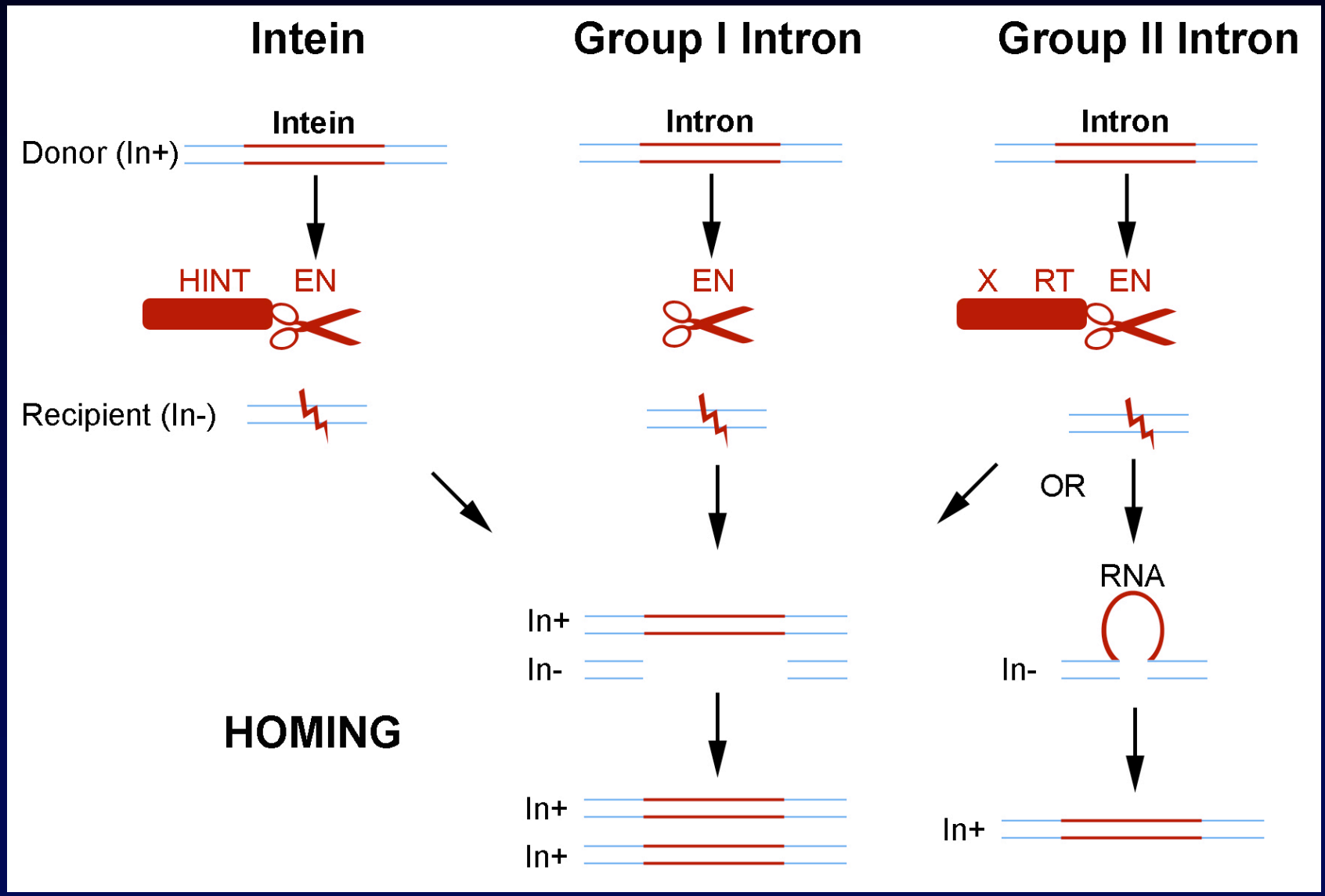
Homing endonuclease are site specific, rare-cutting restriction enzymes that recognize a long DNA sequence between 12-40 bps creating a double strand break.

The role of the endonuclease is to enable intein/intron to horizontally transfer to unoccupied intein/intron integration-sites via a process termed *'Homing'.*

Homing Endonuclease (Group I Intron & Intein)



Homing Endonuclease (Intein, Group I and II Intron)



Homing endonuclease *vs* Intron/Intein

- Different origin

- Joined for mutual benefit

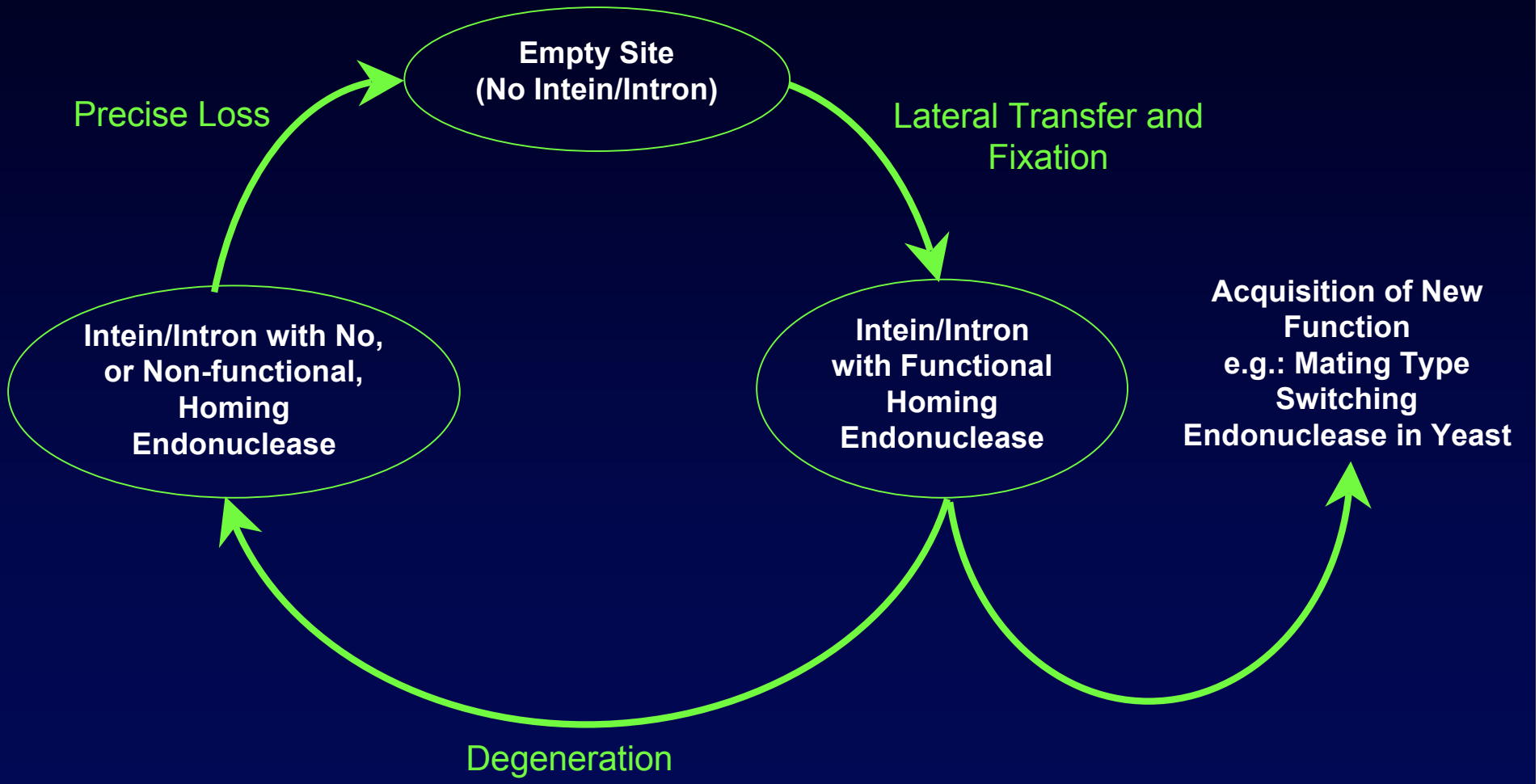
  - Homing endonuclease:

    The ferry to transfer

  - Intron/Intein:

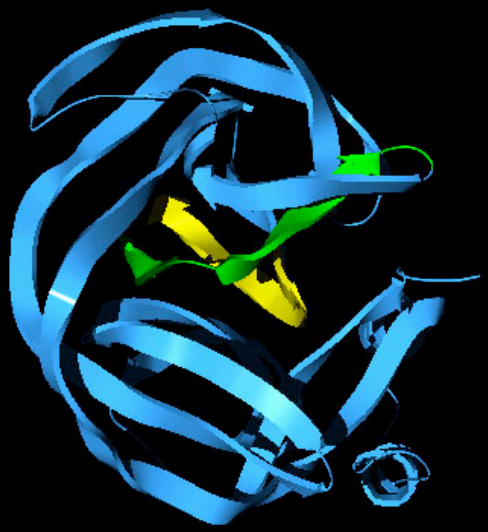
    Splicing to maintain their survival

Homing Cycle

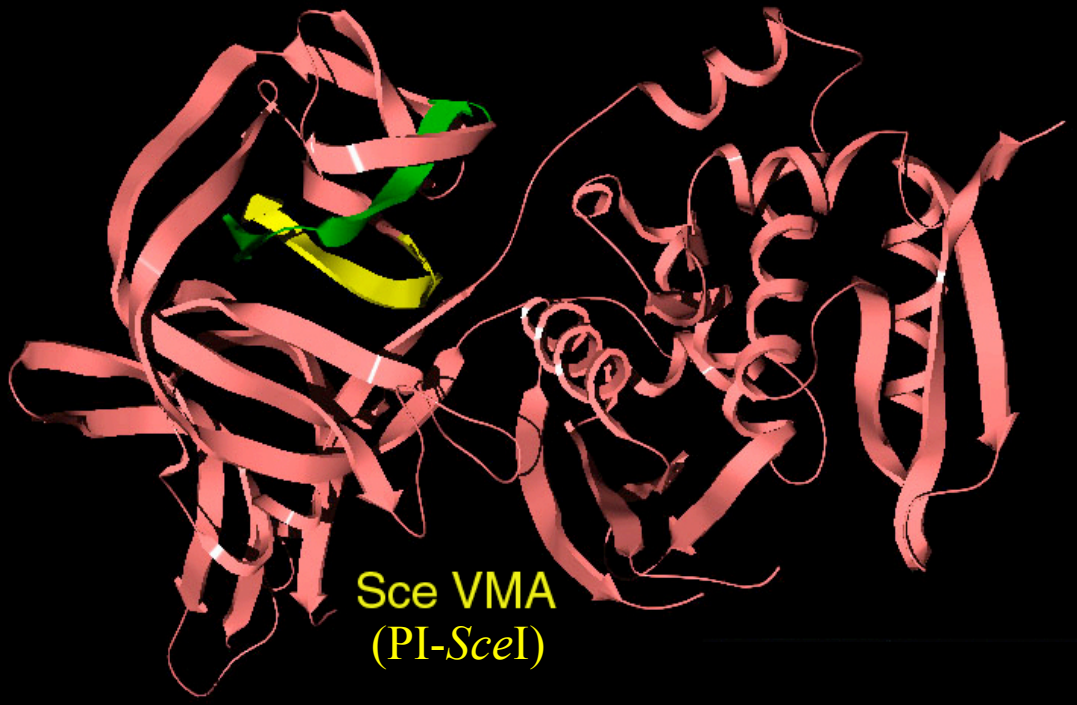


Inteins: Structure, Function, and Evolution
Gogarten, Senejani, Zhaxybayeva, Olendzenski, Hilario; Annu. Rev. Microbiol. 2002, 56:263-87

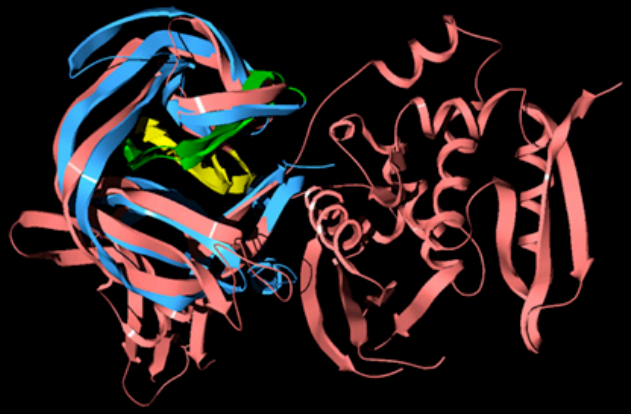
Intein Structure



Mxe Gyr



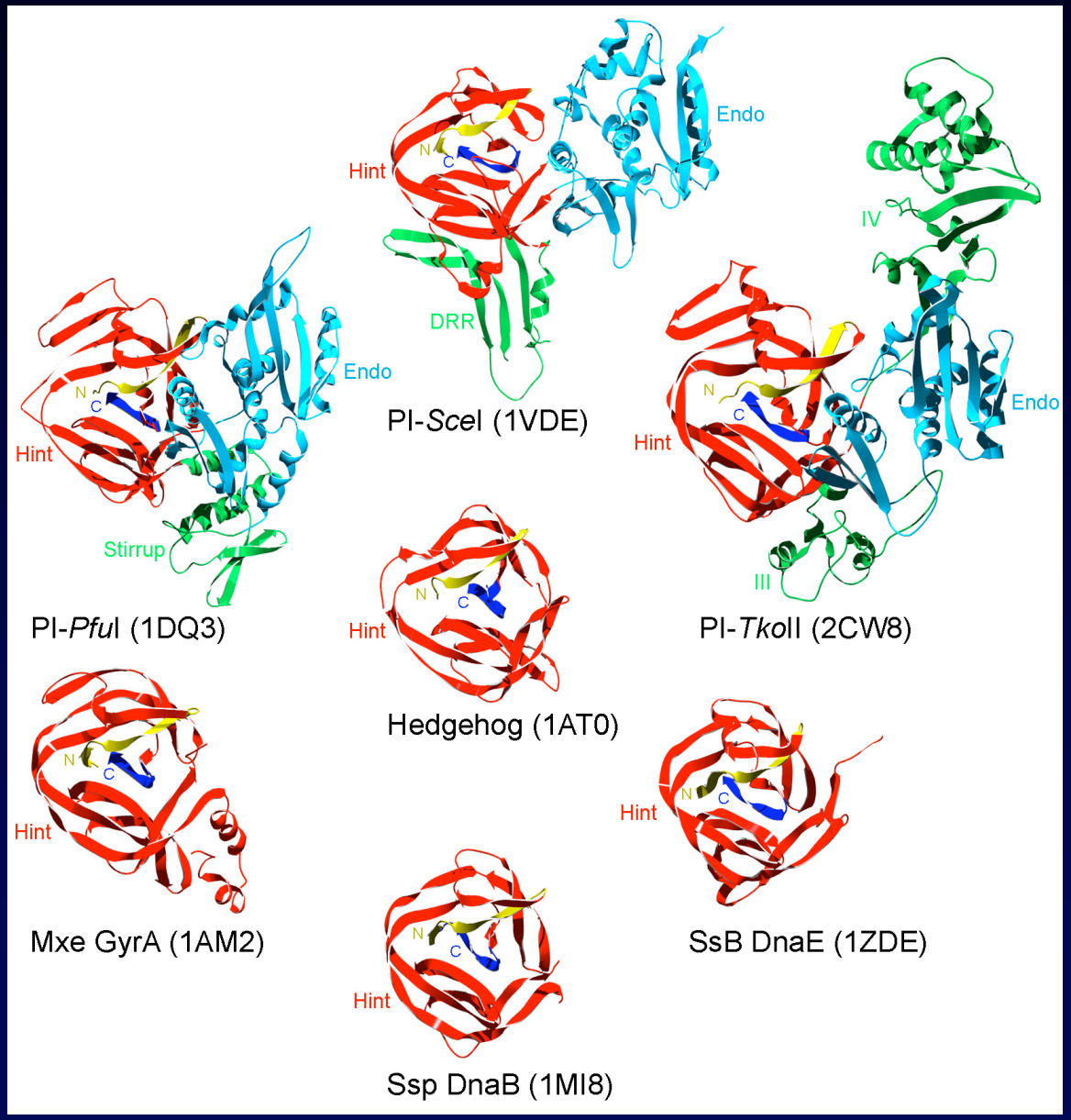
Sce VMA
(PI-*Sce*I)



Mxe Gyr + Sce VMA

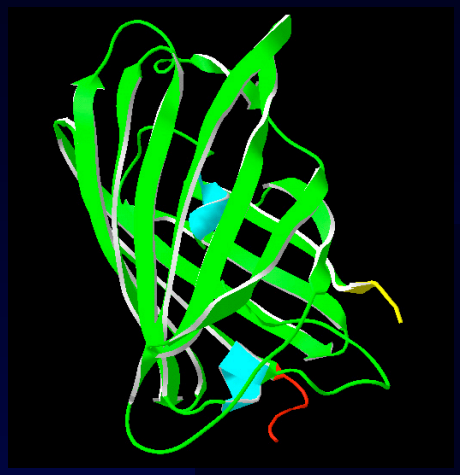
- Duan, et al (1997) Cell 89:555-564
- Klabunde, T et al (1998) Nature Struct. Biol. 5:31-36
- Gogarten et al (2002) Annu. Rev. Microbiol. 2002, 56:263-87

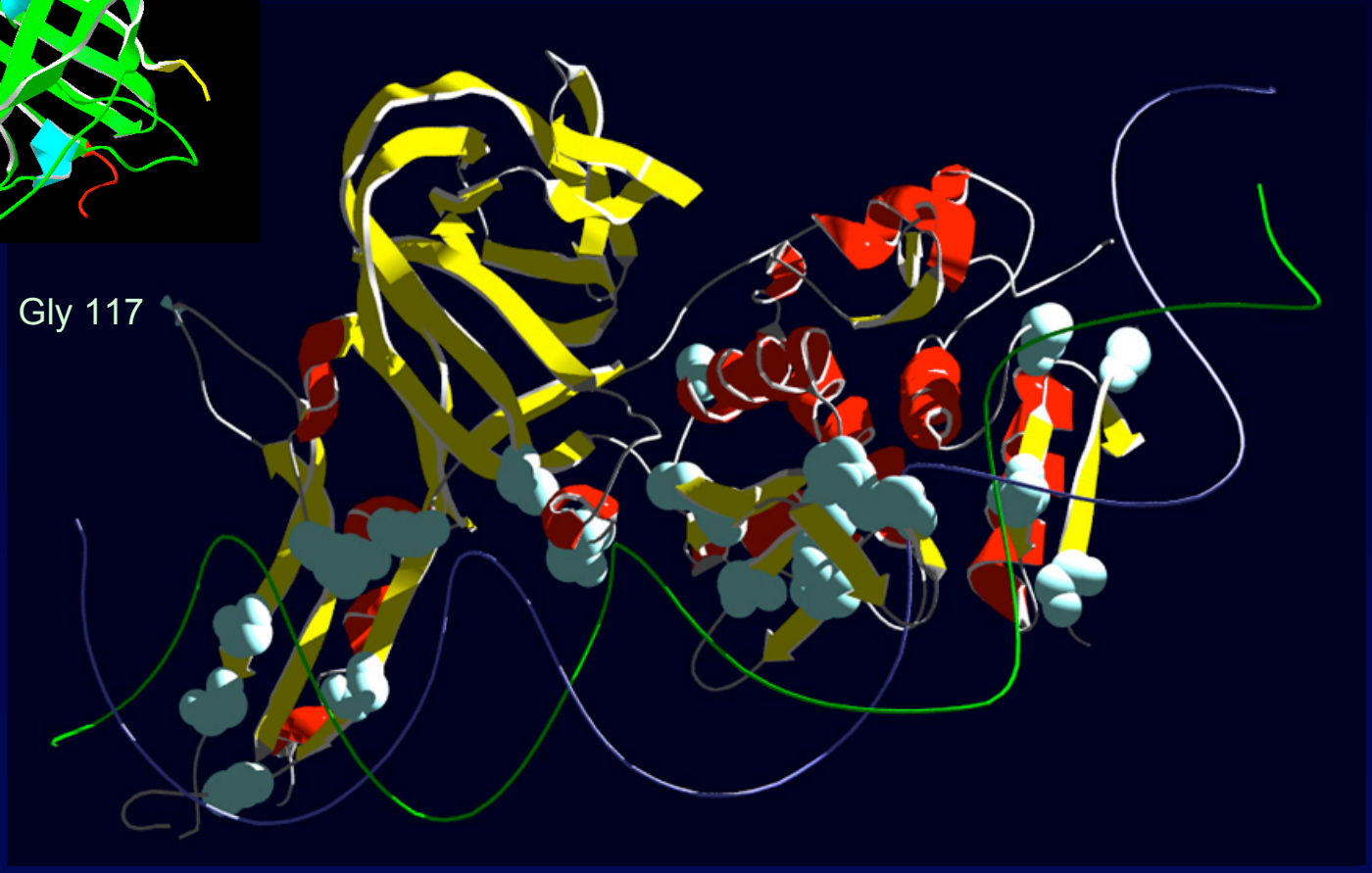
Intein Structure



Structural stability and endonuclease activity of a PI-SceI GFP-fusion protein

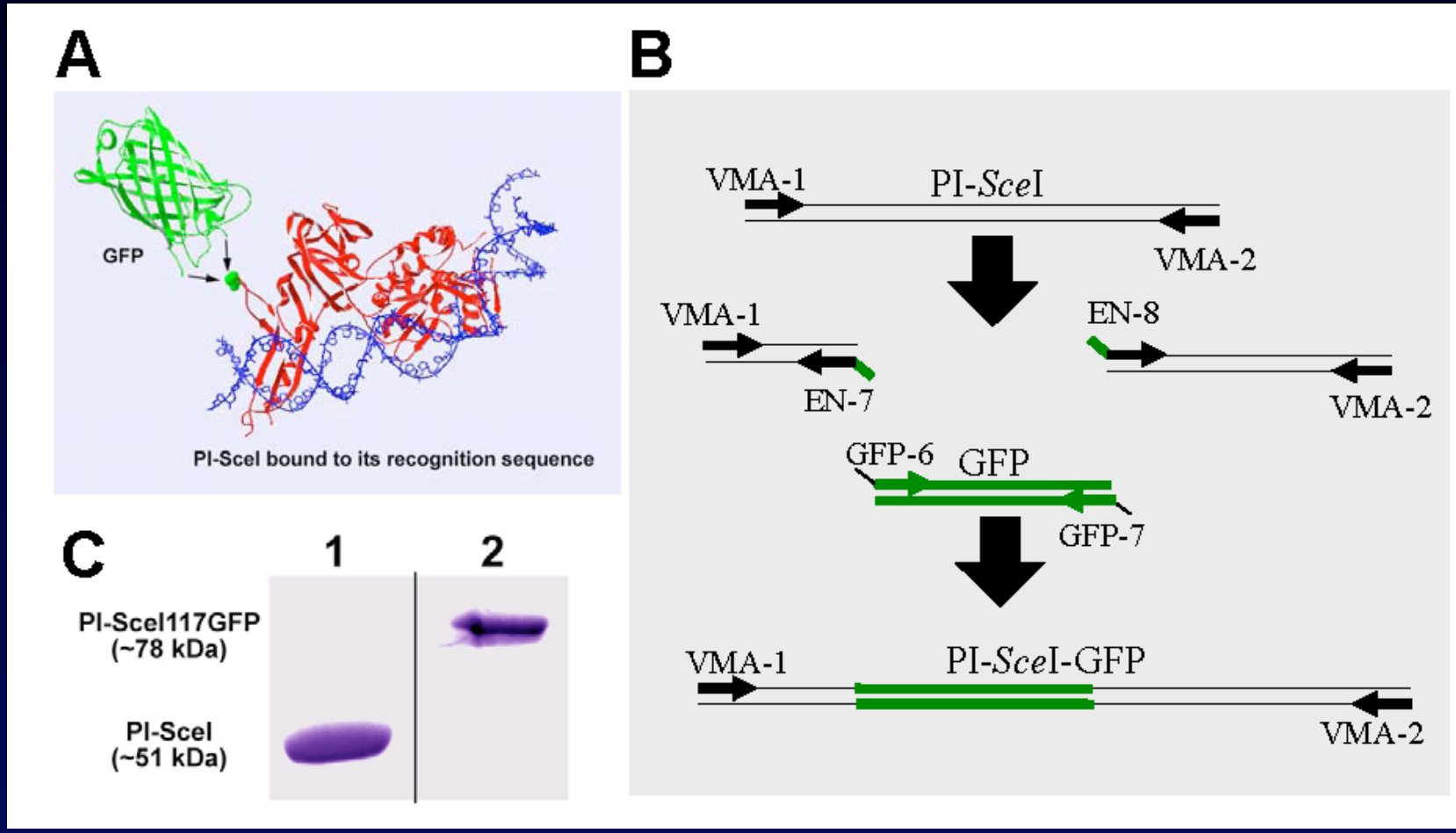
PI-SceI with Fluorescent Marker





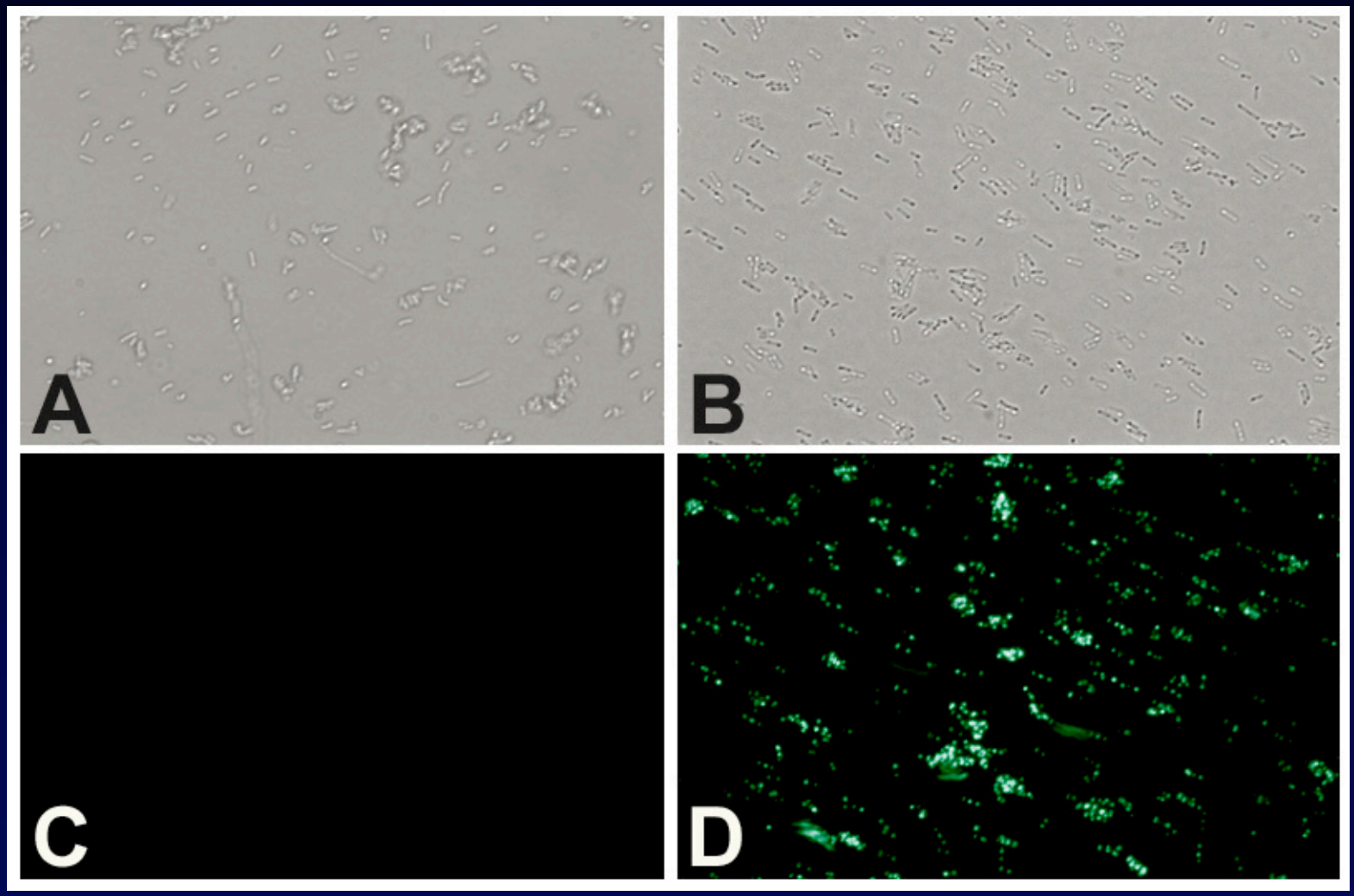
Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence
Moure, C.M., F.S. Gimble, and F.A. Quiocho, Nat Struct Biol, 2002. 9(10): p. 764-70.

PI-SceI with Fluorescent Marker



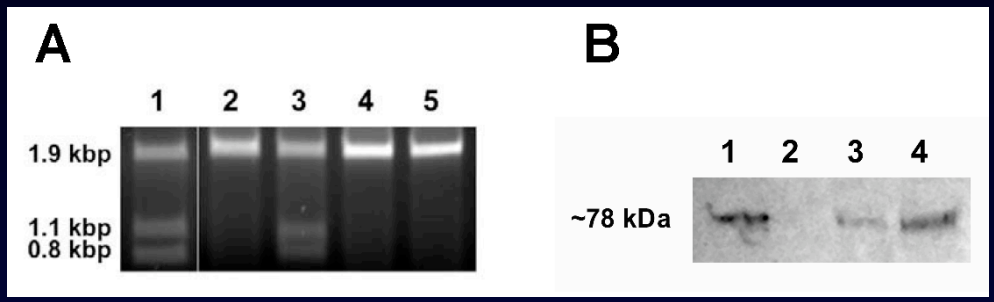
Structural stability and endonuclease activity of a PI-SceI GFP-fusion protein
Alireza G. Senejani and J. Peter Gogarten; Int. J. Biol. Sci. 2007, 3: 205-211

PI-SceI with Fluorescent Marker



- **A and C: *E. coli* expressing PI-*Sce*I (-ve control)**
- **B and D: *E. coli* expressing PI-*Sce*I + GFP**

PI-*Sce*I with Fluorescent Marker



Endonuclease activity and Protein Splicing?

A: Digestion of the *S. cerevisiae* V-ATPase catalytic subunit gene (*vma*1) without intein.

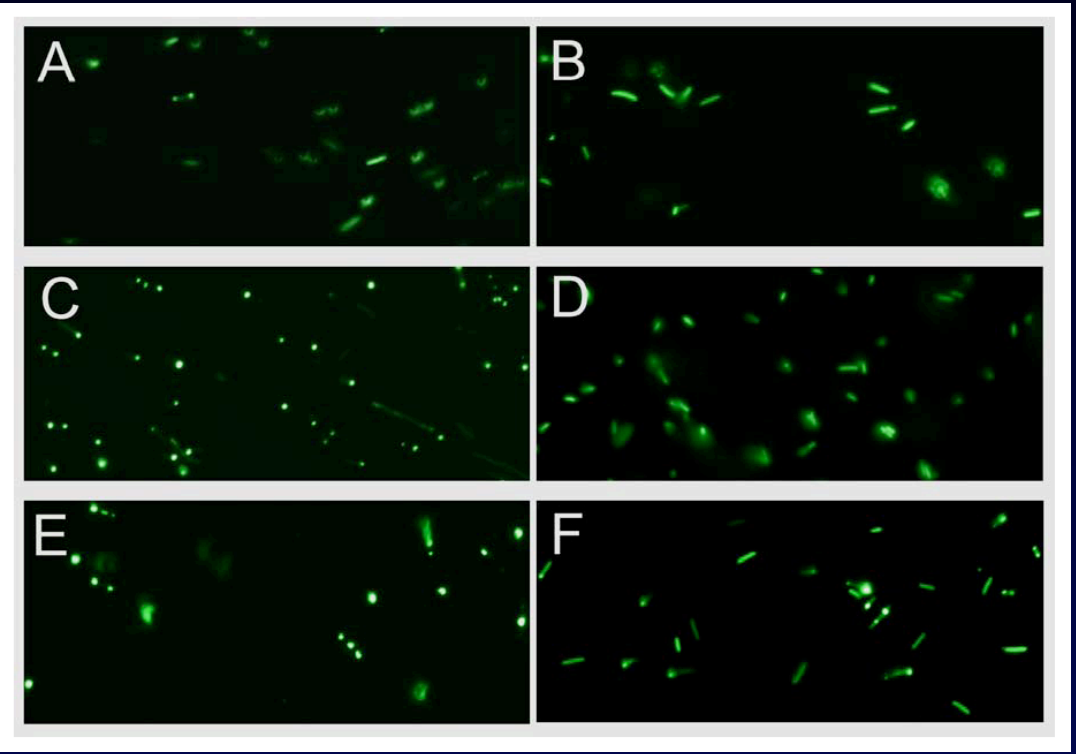
1. PI-SceI (wt) enzyme; 2.5mM MgCl2,
2-4. PI-SceI_117GFP enzyme; 2.5mM MgCl2, MnCl2, ZnCl2, or CaCl2 respectively

B: Immunoblot assay using commercial anti-*Sce* VMA1 intein antibodies.

1. Purified PI-*Sce*I117GFP protein (positive control).
2. Negative control.
3-4. Protein extracts from XL1B (DE3) with plasmids expressing VMA1_PI-*Sce*I_ 117GFP (3.9kb) and truncated VMA1_PI-*Sce*I_117GFP (3kb) respectively.

Structural stability and endonuclease activity of a PI-SceI GFP-fusion protein
Alireza G. Senejani and J. Peter Gogarten; Int. J. Biol. Sci. 2007, 3: 205-211

PI-*Sce*I with Fluorescent Marker



A and B: pET-28a_ PI-*Sce*I 117GFP.
C and D: pET-28a_*vma*1_PI-*Sce*I 117GFP3.9kb.
E and F: pET-28a_*vma*1_PI-*Sce*I 117GFP3kb.

A, C, and E expression was conducted in LB media at 10oC for 72 hours
B, D, and F expression was performed in minimal media at 10oC for 72 hours.

Structural stability and endonuclease activity of a PI-SceI GFP-fusion protein
Alireza G. Senejani and J. Peter Gogarten; Int. J. Biol. Sci. 2007, 3: 205-211

Development of a novel gene therapy method!

How to evolve homing endonucleases with novel specificity!

Objective:

Develop an in-vivo competition system that utilizes endonuclease activity to generate selection pressure in favor of a more efficient enzyme.

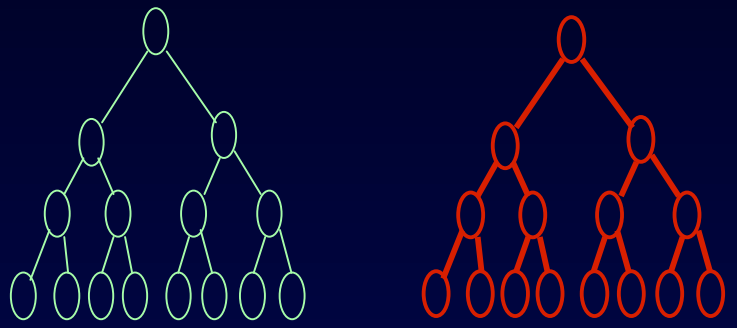
Phage Growth Competition

Target Vector

Expression Phage

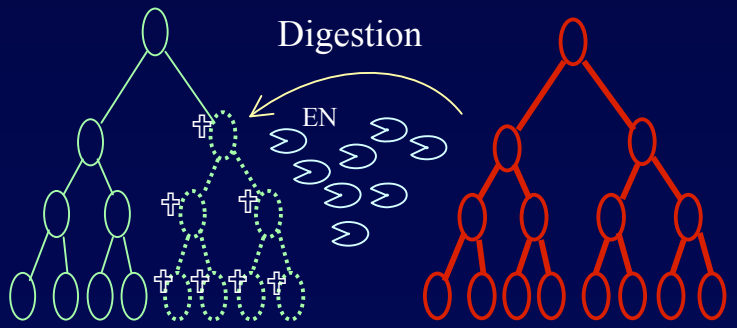
A

No Endonuclease Activity



B

Endonuclease Activity

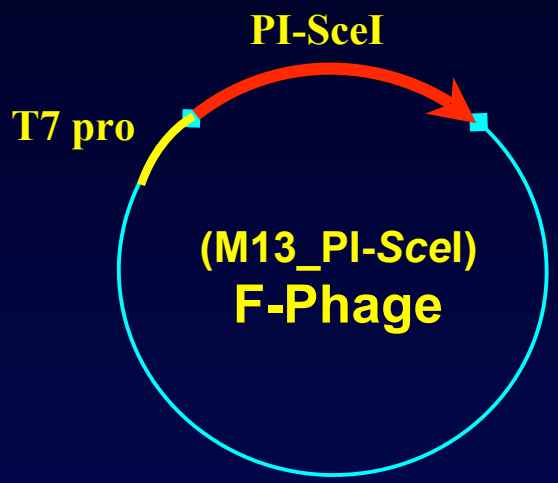


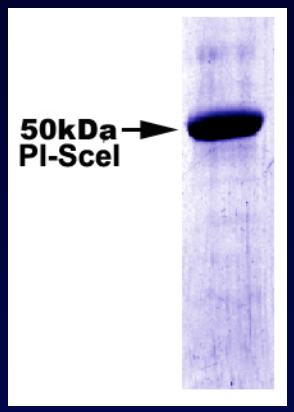
more active enzyme = phage grow faster

Host Strain and Vectors

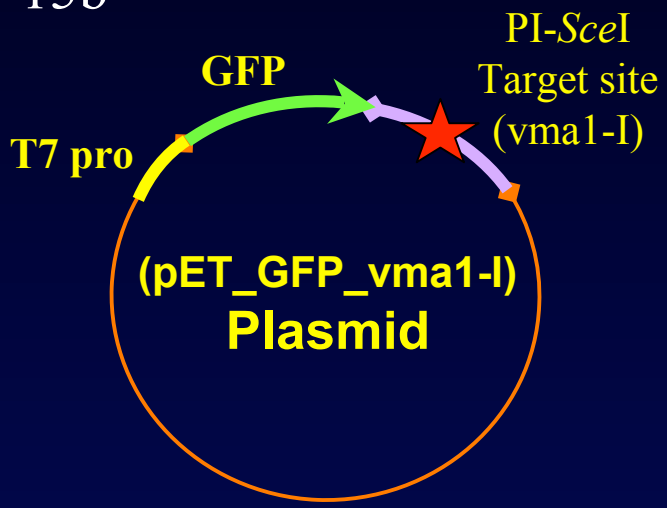
Vectors:

M13mp18 Phage



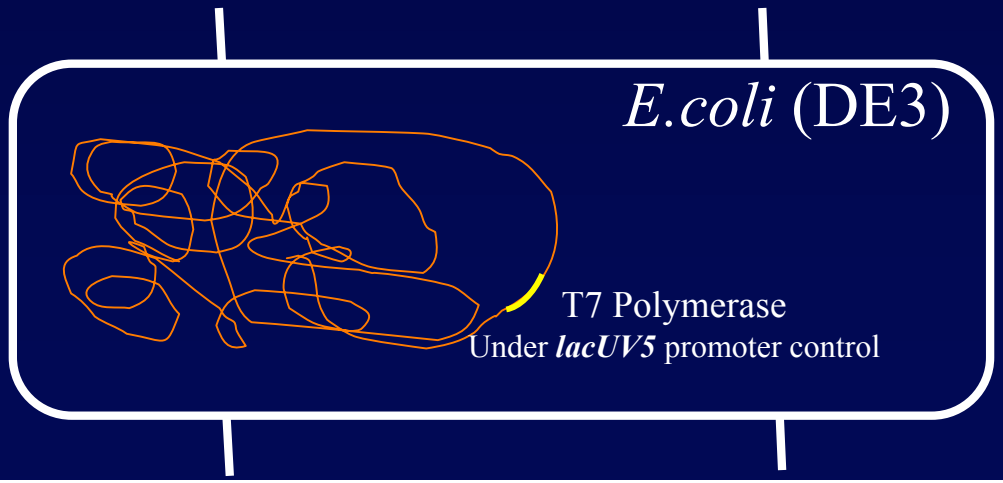


pET-15b

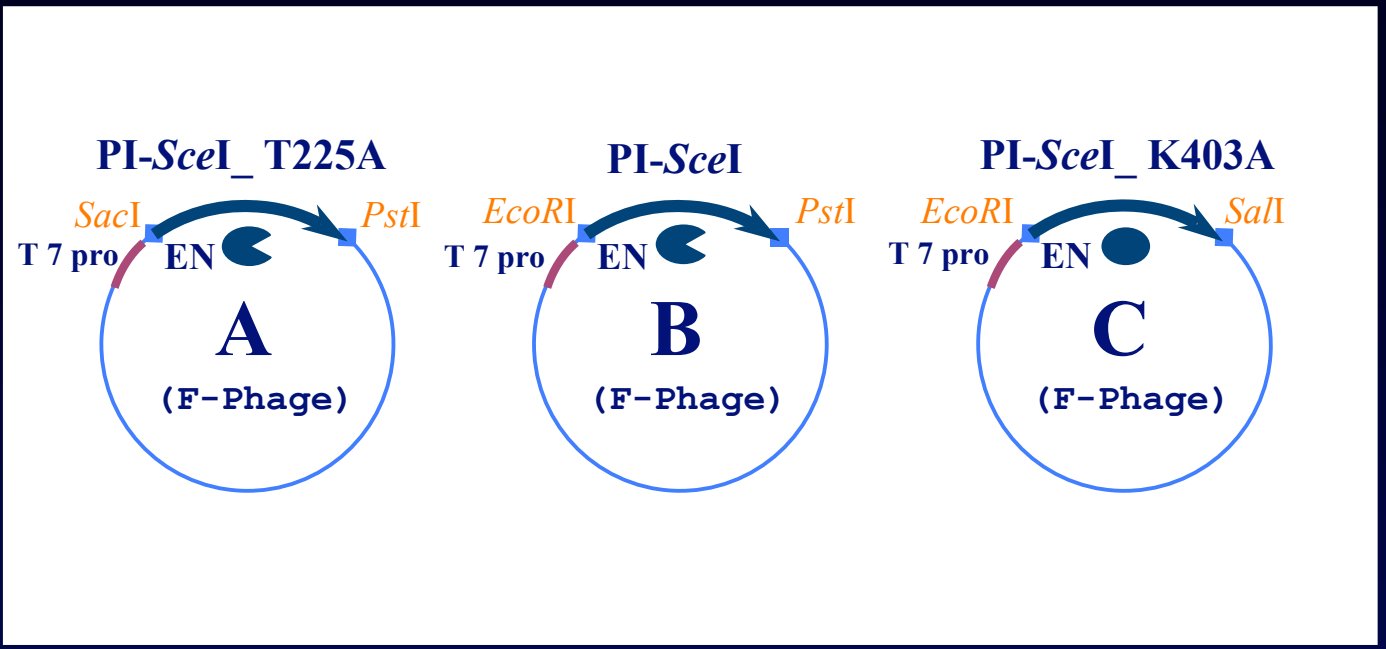


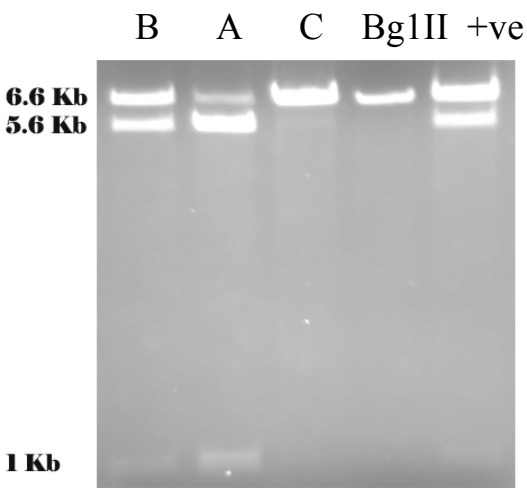
E.coli:

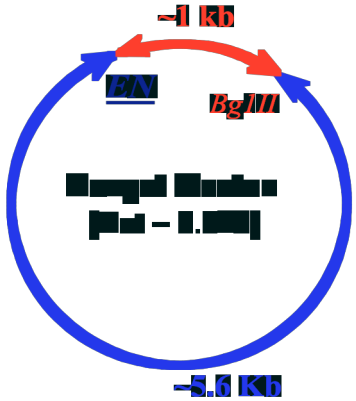
XL1Blue-MRF' (DE3)



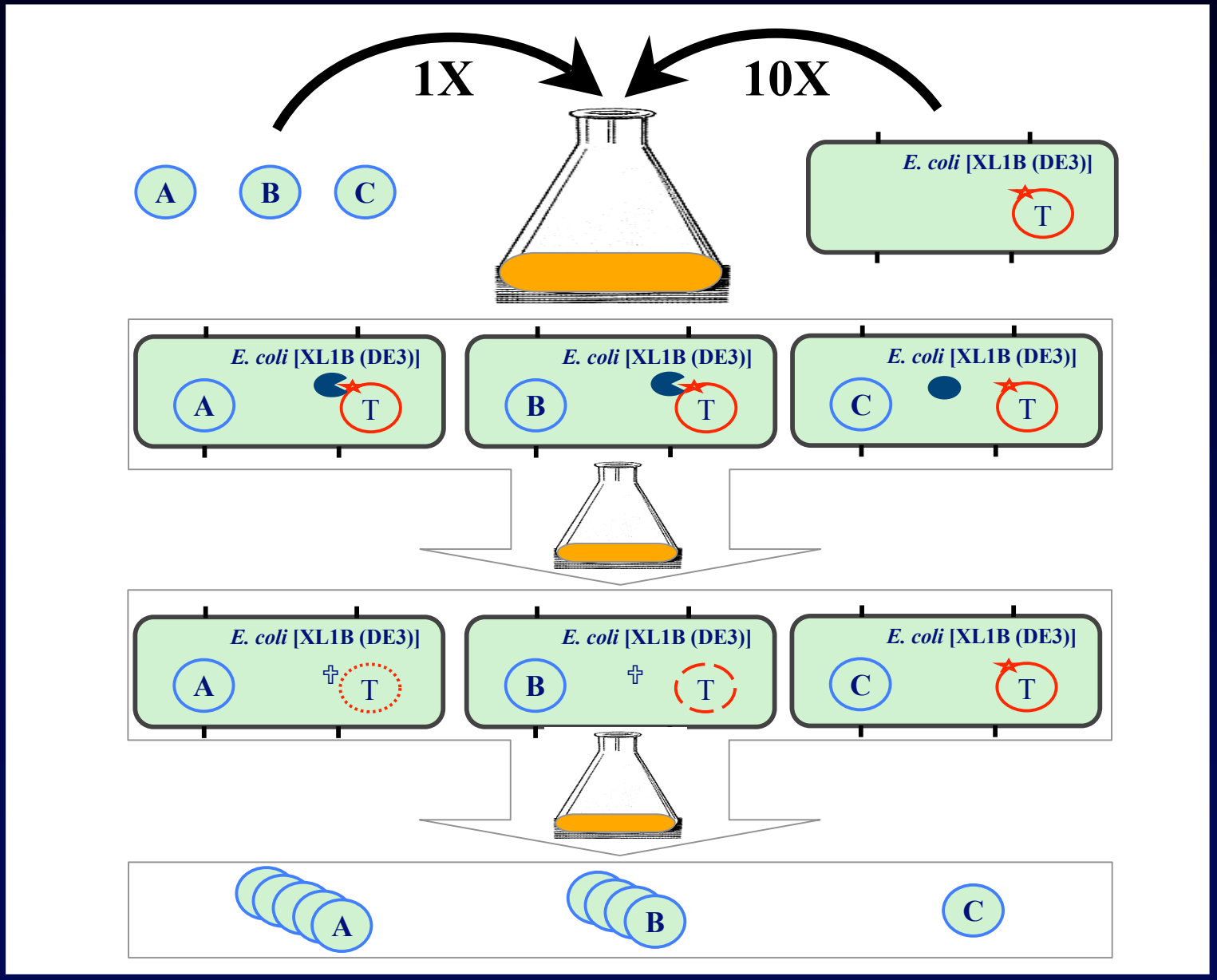
Homing endonuclease activity assay



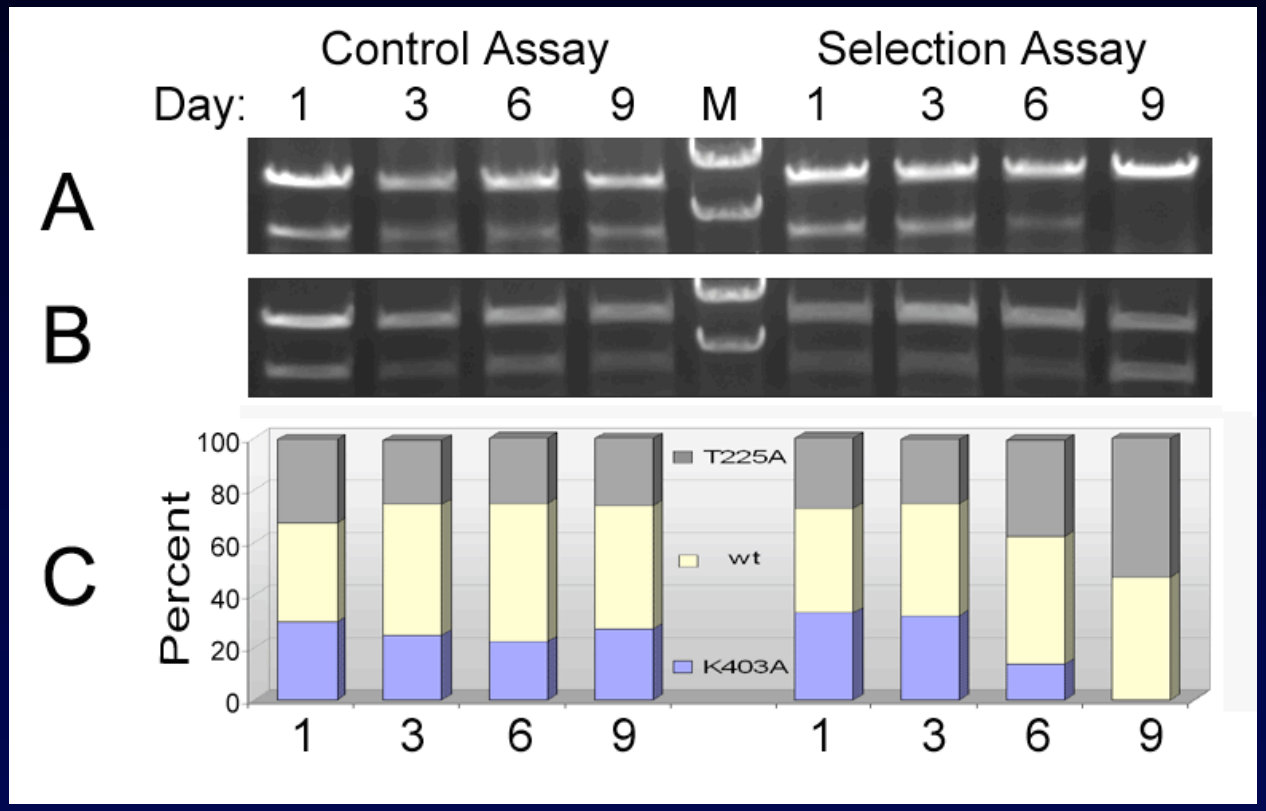




Phage Competition Assay



Results from Selection Assay



Ratio of the phage expressing PI-*Sce*I enzyme with none or a small amount of activity (K403A) decreases over time and after nine series of serial infection and phage growth only phages which express active enzyme are found. While in the control all the phages are maintained.

Conclusion 1/2

- Despite differences in pH and temperature between the *E. coli* and the *T. acidophilum* cytoplasms, the *T. acidophilum* intein retains efficient self-splicing activity when expressed in *E. coli.*

  The small intein in the *Thermoplasma* A-ATPase is closely related to the endonuclease containing intein in the *Pyrococcus* A-ATPase. Phylogenetic analyses suggest that this intein was horizontally transferred between *Pyrococcus* and *Thermoplasma.*

- Inteins are situated in the most conserved parts of the proteins. Introns that encode endonuclease (EN) are also found in the highly conserved parts of the host; whereas this does not hold true for intron w/o EN locations in general.

Conclusion 2/2

- Insertion of the Green Fluorescence Protein (GFP) into a loop which is located between the endonuclease and splicing domains of the *Sce* VMA1 intein did not interrupt the three functions of the multi domain fused proteins. However, the endonuclease activity of the newly engineered protein was different from the wild-type protein in that it required the presence of Mn2+ and not Mg2+ metal cations for activity.

- The developed *in-vivo* selection system can be used as tool to study and select homing endonucleases by linking their activity to phage growth. Once a homing endonuclease with desired target site specificity is evolved the homing mechanism can be explored and this can potentially be a novel gene therapy method to target and replace any mutated genes with a healthy copy through homing or destroy DNA of pathogens by cleavage.

Acknowledgements

•**Dr. J. Peter Gogarten**

•Dr. Strausbaugh
•Dr. Knox
•Dr. Noll
•Dr. Giardina

•Dr. Graf
•Dr. Gage
•Dr. Benson
•Dr. Nelson
•Dr. Ovchinnikov
•Dr. Gimble (Purdue Univ.)
•Dr. Perler (NEB)
•Dr. Zhaxybayeva
•Dr. Olendzenski
•Dr. Magnotta
•Dr. Hilario
•Dr. Mercier
•Pascal
•Kristen
•Greg
•Dr. Poptsova
•Holly
•Tim
•Dr. Graf lab
•Dr. Noll lab
•Dr. Benson lab
•Dr. Gage lab
•Dr. Giardina lab
•Dr. Verma
•Dr. Andalib
•Aida and Mr. Ghiaei family
•Haleh and Arman
•Mehrtash
•Dr. Maria Gogarten
•MCB-200 TA Gang
•Esmaeil
•Family and all other friends in Iran, U.S., Canada, Germany, and India
•Maryam

NASA, NSF, UCONN Foundation

Thank you for your attention!