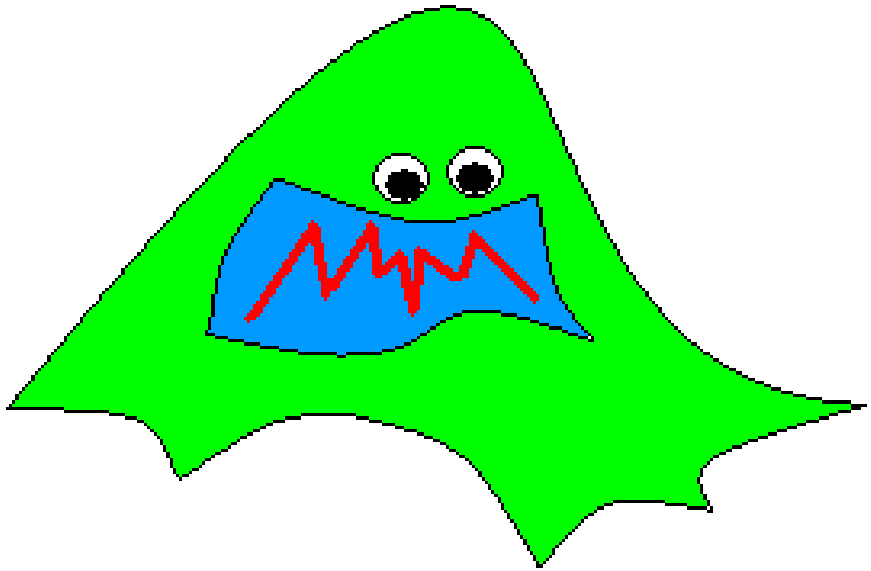


On bio-design of Argo-machine

Andrew Kuznetsov, Mark Schmitz, Kristian Müller
Freiburg University, Germany

Contents

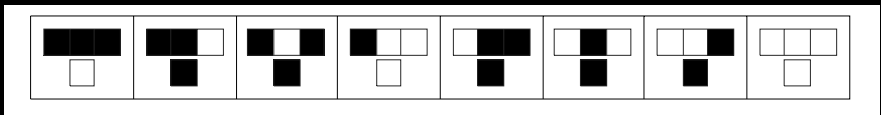
- Introduction:
  - minimal life, compositional evolution
- Theory:
  - AM description, Argonaut algorithm
- AM application:
  - IGNAF design, from monopod to bipod nuclease
- Outlook:
  - DNA synthesis, AM in a minimal cell

A minimal life

by David Deamer (University of California, Santa Cruz)

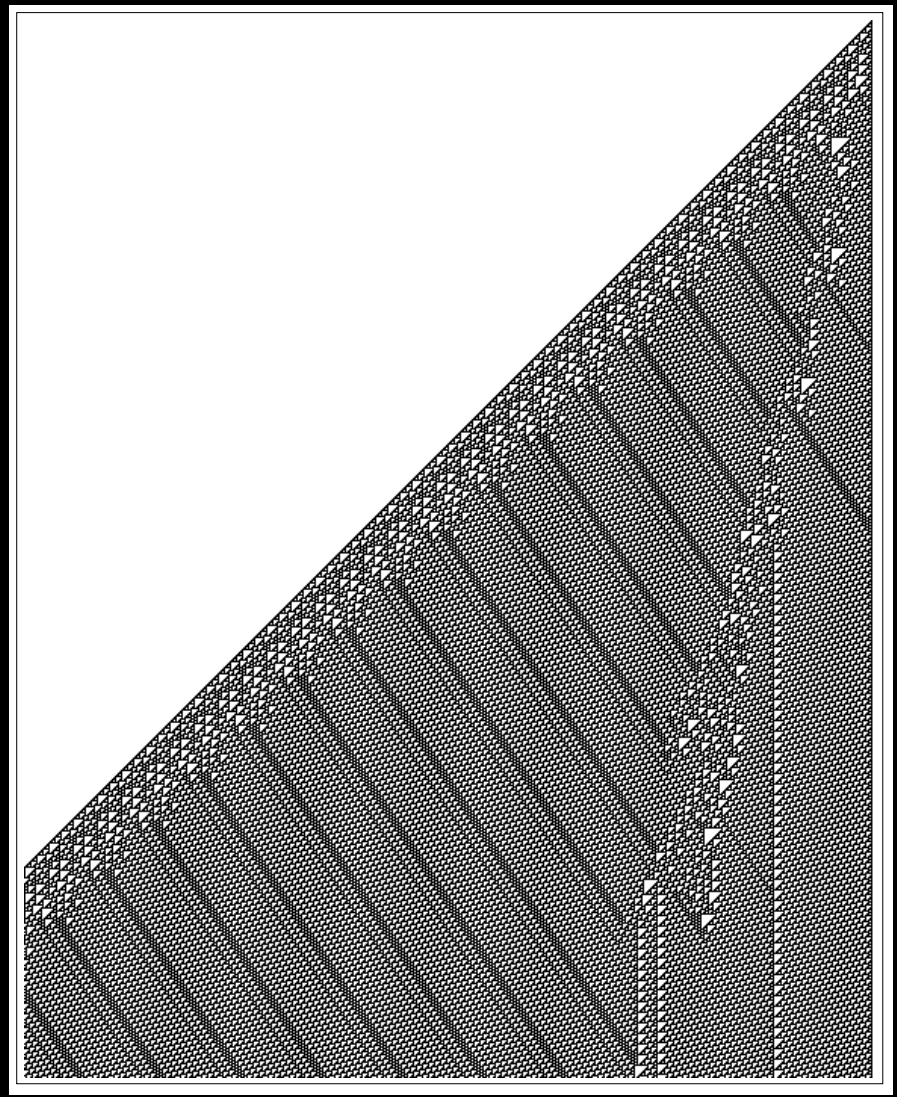
Translation system:	20 tRNAs 3 rRNAs (5S, 16S, 23S) 55 ribosomal proteins 20 aminoacyl-tRNA synthetases
Nucleic acid synthesis:	1 RNA polymerase 1 DNA polymerase
Membrane growth-phospholipid synthesis:	1 Acyltransferase
Transport:	1 α-Hemolysin
The total number of components:	**102**

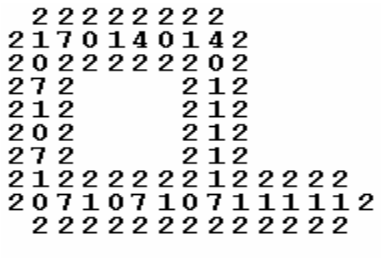
rule 110



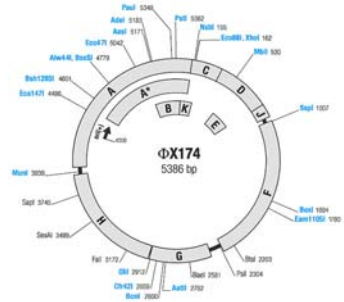
- The number 110 refers to the enumeration scheme introduced by Stephen Wolfram in 1983. Its rule outcomes are encoded in the binary representation $110=01101110_2$
- Rule 110 was investigated by Matthew Cook (1999). Amazingly, the rule 110 cellular automaton is universal

- Rule 110 if applied to a sufficiently large graph, begins to generate complex irregular structures that do not appear to be predictable from the input row – the top row of the graph





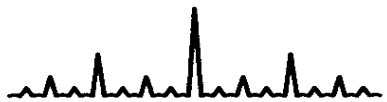
How could we engineer living organisms?



- Minimal life? Programmable artificial cell?
  - Chris Langton's Self-Reproducing Loop, 86 cells, 8 states
  - phiX174, 5386 nt, 11 genes
- Minimal cell, [~100, 265-350] genes
  - Top-down: reprogramming simple organisms
    - *Mycoplasma genitalium* G-37, 580 Kbp, 480 genes, Craig Venter
    - *Mesoplasma florum* L1, 793 Kbp, 517 genes, Tom Knight
    - Synthetic genomic Inc, 2005, Craig Venter
  - Bottom-up: creating cells from nonliving material
    - Los Alamos Bug, PNA, Steen Rasmussen
    - ProtoLife, 2005, Norman Packard, Mark Bedau
- Evolution under the control of a man or a computer?
  - Rational vs. evolution design?
  - Computation *in silico, in vitro, in vivo* or something else?

Algorithmic paradigms of evolution

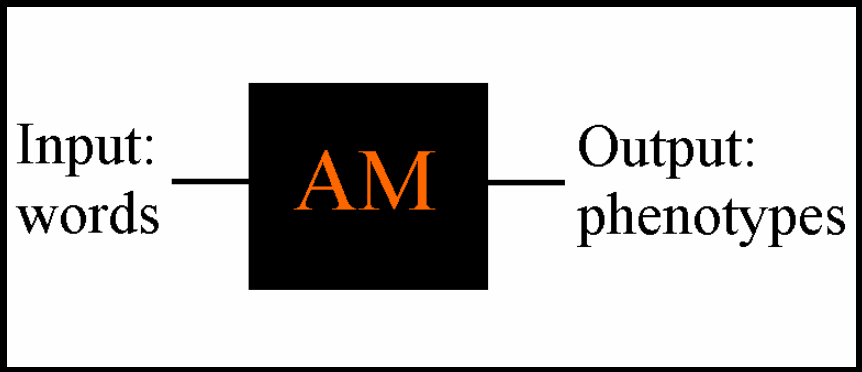
Richard Watson, 2006

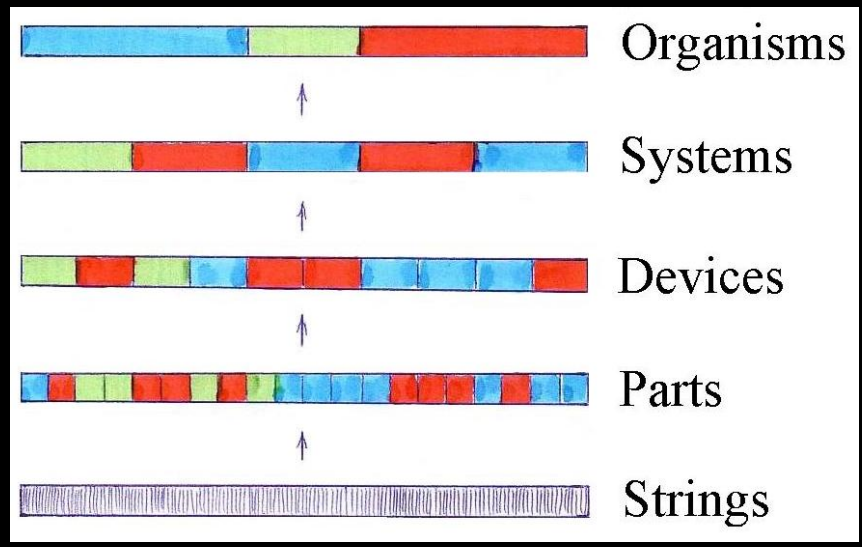
Dependency of variables	Few / weak interdependencies	Modular interdependencies	Arbitrary interdependencies
Landscape			
Algorithmic paradigm	hill-climbing – accumulation of small variations	divide-and-conquer problem decomposition	exhaustive search, random search
Complexity	KN	N^K	K^N
Evolutionary analogy	gradual evolution	compositional evolution	“impossible” / ”intelligent design”

N – # of variables, K – # of values for each variable

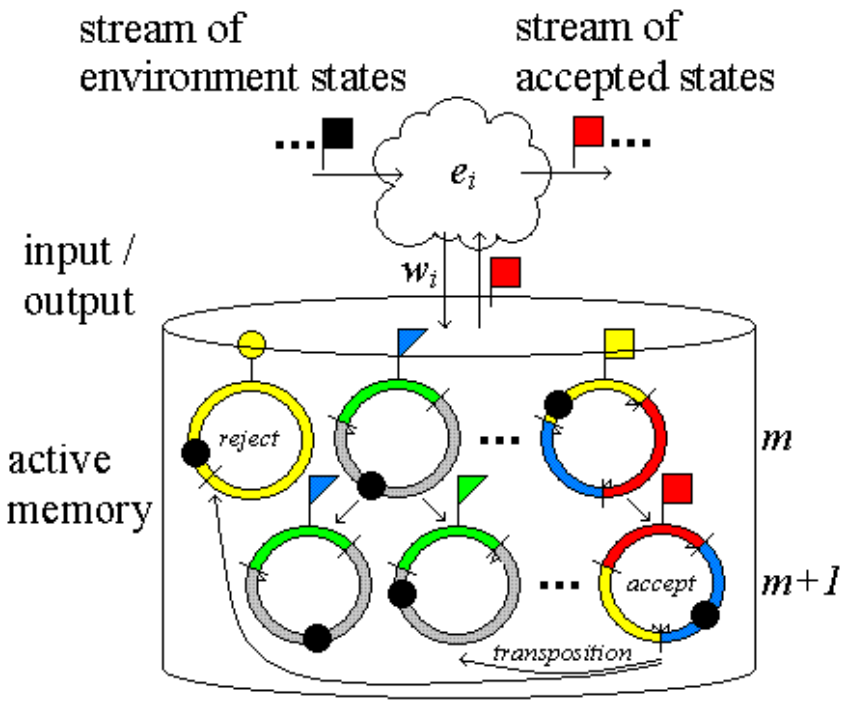
Production of LEGO set and hierarchical assembling

Consider an evolving system–an abstract machine and an environment that is continuously changing creates input words for the machine to stimulate an adaptation of this device to the surrounding…





Argo-machine



The system operates on inputs and memory, uploads the memory and yields outputs

- The *Argo-machine (**AM**)* consists of *agents*; each of these has a head, a tape and can be in different output states. The tape is a nonempty string of symbols that may be linear or circular. The head scans the tape according to an input word w_i, and cuts it at recognized sites. The agent arbitrarily pastes the tape. For each tape-configuration there is an appropriate output state of the agent that is checked by the environment. Special 'accept' and 'reject' states take immediate effect. An agent accepts, if its output state corresponds to the environment state; an agent will reject if less than two matches to the input word exist on the tape. ***AM*** can accept if at least one agent accepts, reject if all agents reject, or loop. If environment has changed, then it delivers a transposition and a new word w_{i+1}.
- The transposition means to make a copy of tape from the accepted agent to other ones and join it in head-to-tail
- ***AM*** looks for an agreement with the environment again and again

Argonaut algorithm

A* = “On word *w*:

1. Scan the tape to be sure that it contains at least two matches. If not, reject.
2. Cut at the matching sites and arbitrarily paste the tape’s fragments.
3. Take the output state according the new tape.
4. Check it with the state of environment. If satisfy, accept; otherwise loop.”

How does it work?

AM computation in winning branch

Language notations:

~,<,(– strings, cut before open brackets;
– boundary symbol

Example 1. Adaptation without transposition:

environment '<~~>', word '<'

1.	<~~>	environment
2.	<	word
3.	#~<~<~<~#	tape_tick_1
4.	#~<~~><~#	tape_tick_2
5.	<~~>	accept

Example 2. Two adaptations with one transposition:

environment_1 '<~(~>', word_1 '<',
environment_2 '<~~~>', word_2 '('

1.	<~(~>	environment_1
2.	<	word_1
3.	#~(<~<~)<~#	tape_tick_1.1
4.	#~(<~(~><~#	tape_tick_1.2
5.	<~(~>	accept_1
6.	<~~~>	environment_2
7.	#~(<~<~)<~##~(<~(~><~#	transposition
8.	(	word_2
9.	#~(<~<~)<~~(<~(~><~#	tape_tick_2.1
10.	#~(<~<~)<~~~>)(~><~#	tape_tick_2.2
11.	<~~~>	accept_2

The elongation of input words leads to the increasing of building blocks

Alphabet: {a,b,c}
Language: {**a**,**ab**,**abc**}
Tape: aababcaabacbaa

Examples:

Case 1. On input word |**a**:
a a**b** a**bc** **a** a**b** a**cb** **a** **a**

Case 2. On input word |**ab**:
a **ab** **ab**ca **ab**acbaa

Case 3. On input word |**abc**:
aab **abc**aabacbaa

Description:

Case 1. Input is a short word; enormous number of rearrangements allows an exhaustive search, but all previous results are destroyed

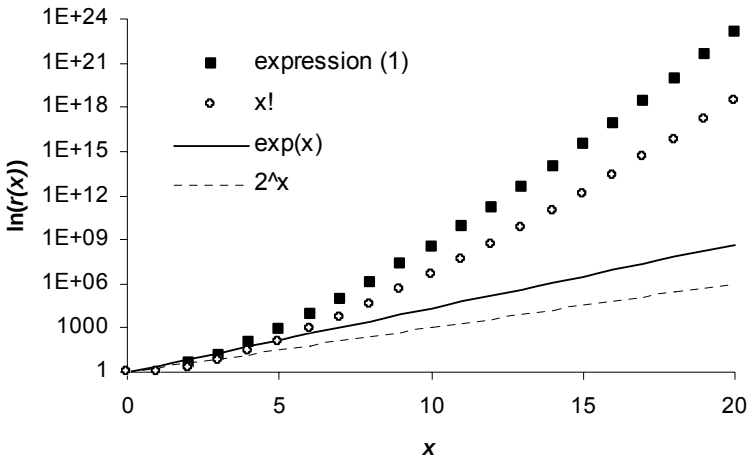
Case 2. **What language is optimal to maintain an appropriate level of diversity for a creative combinatorial design? What about the rules to form this language?**

Case 3. Input is a long word; deterministic kind of design

An analysis

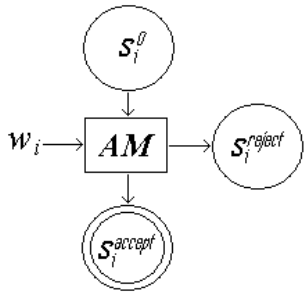
Combinatorial formula (1)

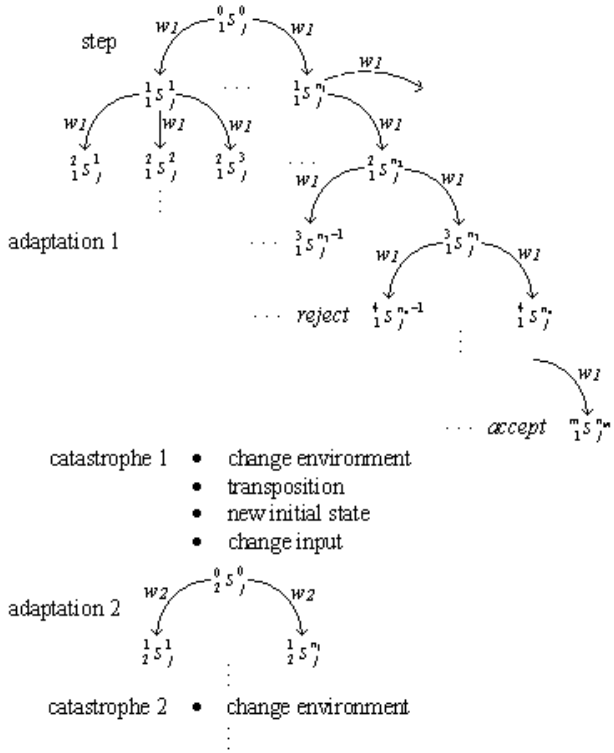
$$\begin{cases} r_0 = r_1 = 0, \\ r_x = 2^x * (x-1)!, \; x \geq 2 \end{cases} \qquad (1)$$



Combinatorial power of expression (1)

Adaptation





Nondeterministic computation

Requirements to AM

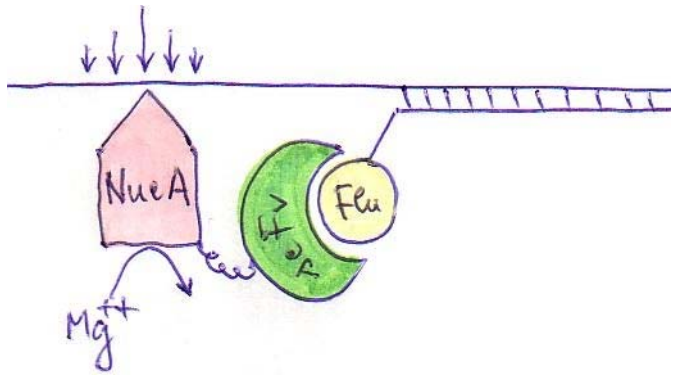
- definition, description, and refinement of AM
- investigation of AM behavior: a sample run of AM on input in the environment
- variants of AM: isomorphism, robustness
- comparison of AM with TM and others machines: decidability, halting problem
  - proof of equivalence in power
  - simulate one by the other

implementation

- conventional computer (special case)
- bio-molecules
- living/artificial cells


THE
REAL
WORLD
Key West
February 2006
PLAY
SHARE

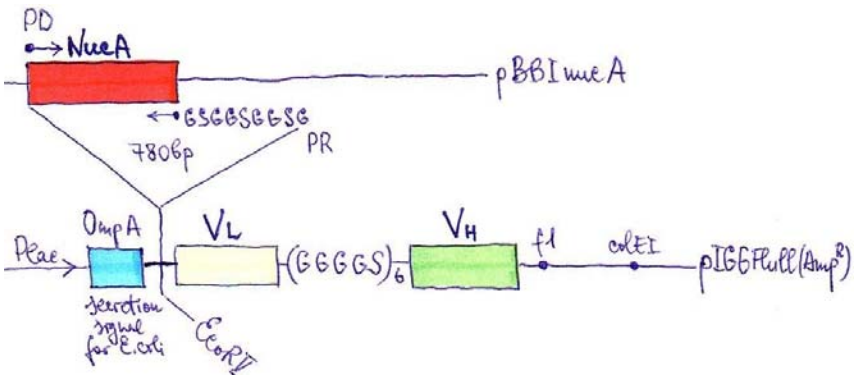
The oligonucleotide-guidable endonuclease α-IGNAF

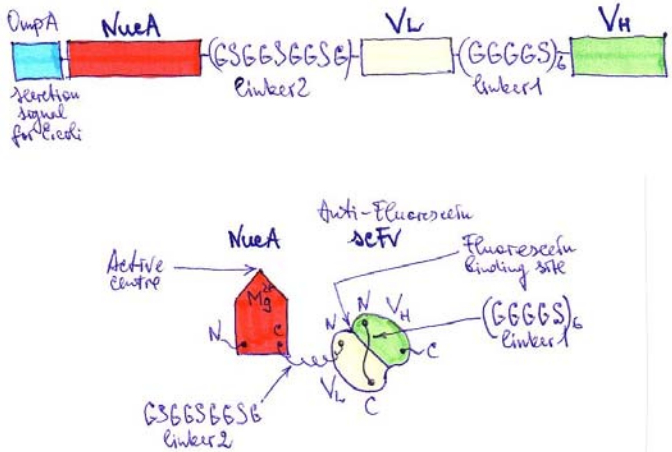


The specificity of this hybrid enzyme can be easily altered. It would be a 'programmable molecular device'. Two alternatives are considered:

1. the catalytical method - hybrid nuclease acts as enzyme with substrate turnover above Tm,
2. the robust method means carrying out repeated hybridization and cleavage reactions in a thermocycler

pIGNucAFlu
Two domains of α-IGNAF protein

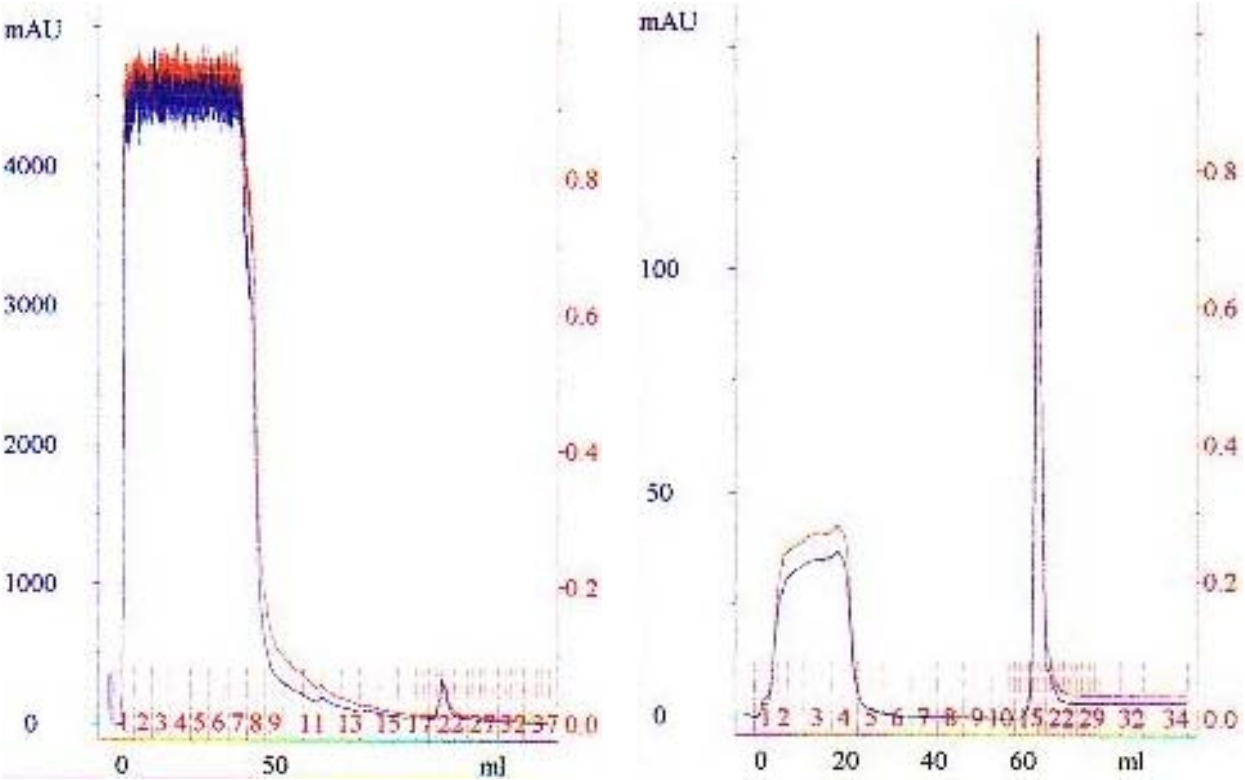




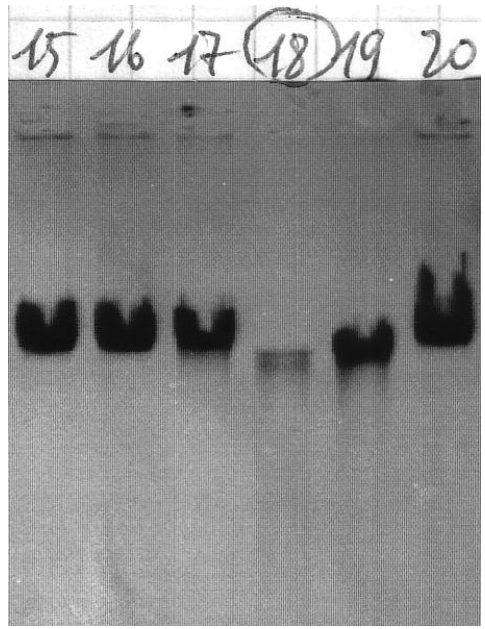
- Plasmid pIGNucAFlu consists of lacI promoter, IGNAF sequence, f1 origin, colEI origin, and *bla* gene

- Protein IGNAF with MW ~60 kD includes the ompA secretion signal, FLAG, NucA domain, GSGGSGGSG peptide tether from 9 aminoresidues, variable light-chain (V_L) domain, $(GGGGS)_6$ 30-mer linker, variable heavy-chain (V_H) domain of 4-4-20 scFv antibody to fluorescein, myc-Tag, and His-Tag

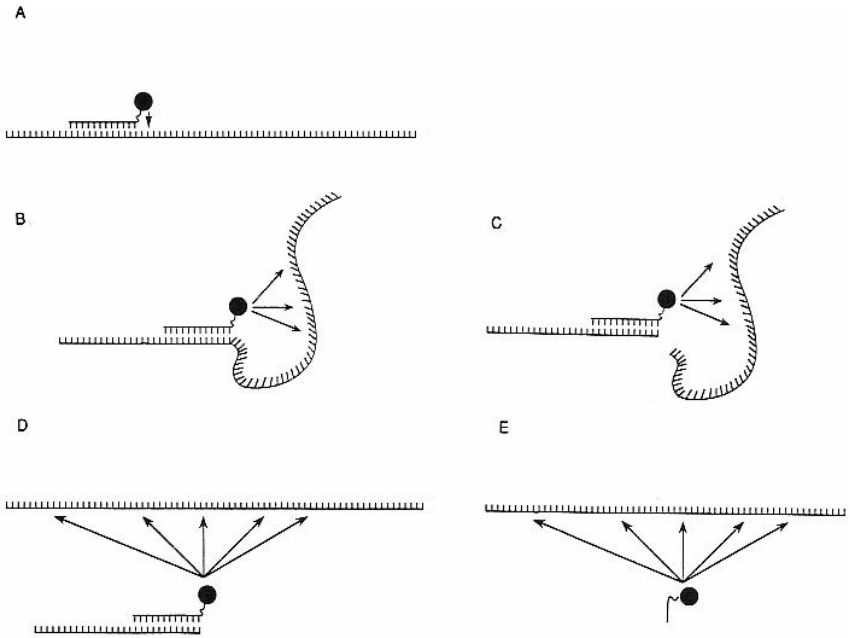
Chromatography on Ni-NTA and Heparin. DNase activity in fractions



The fraction # 18 is most active



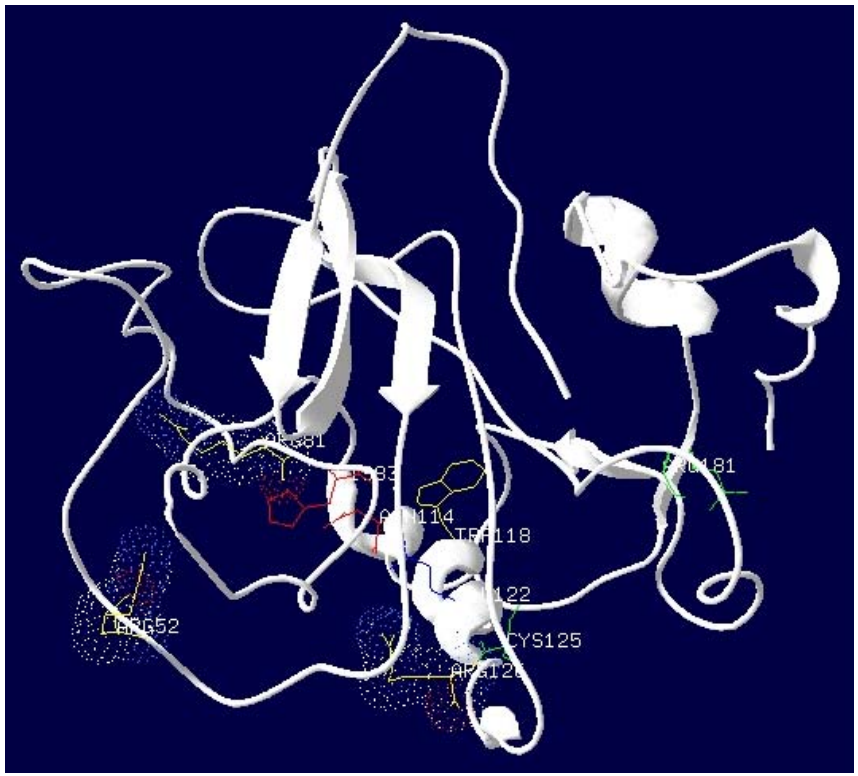
The problem is a nonspecific cleavage



Corey *et al*., 1989

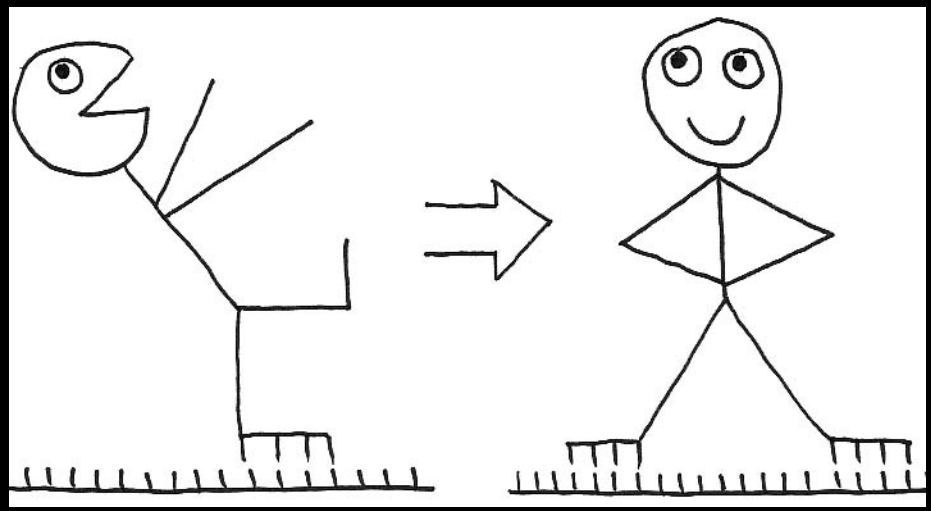
- It can occur in an intramolecular fashion, in which specific binding first localizes the nuclease at the target site, so as in an intermolecular reaction, which is independent on oligonucleotide
- Can a ‘nonspecific binding’ be decreased by mutations in the α-helix and DNA-binding loop of NucA domain?

NucA nuclease from *Anabaema sp.* with important aminoresidues (model)



- Mutations:
  - R93A and W159A
  - Unfortunately, it's not a solution of the problem, because the mechanism of reaction was not changed
- Smart IGNAF molecules have to bind at the target site, then switch on, next cleave DNA strand, and finally switch off

From monopod to bipod IGNAF



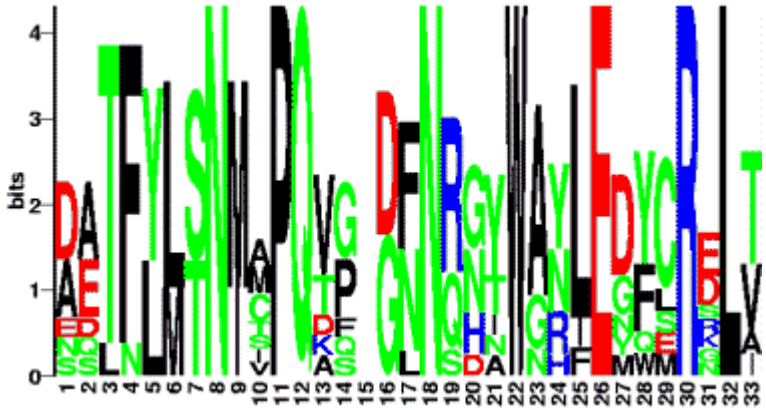

NucA split

Comparative sequence analysis

by NCBI CDD BLASTP http://www.ncbi.nlm.nih.gov/Structure/cdd and
by Structure Logo http://www.cbs.dtu.dk/~gorodkin/appl/plogo.html

Multiple alignment:

```
                                  β                   α
consensus    LDRGHLAPAA.[8].QDATFYLTNMAPQ.[3].FNQGNWAYLEDYLRDL 126
NucA query   YDRGHIAPSA.[8].NAATFLMTNMMPQ.[3].NNRNTWGNLEDYCREL 115
SM 1QL0_A    VDRGHQAPLA.[7].WESLNYLSNITPQ.[3].LNQGAWARLEDQERKL 129
gi 128831    YDRGHQAPAA.[8].MDDTFYLSNMCPQ.[4].FNRDYWAHLEYFCRGL 184
gi 585595    YDRGHIAPSA.[8].NAATFLMTNMMPQ.[3].NNRNTWGNLEDYCREL 169
gi 1723567   YDRGHQVPAA.[8].MNETFYLSNMCPQ.[4].FNRNYWAYFEDWCRRL 188
gi 3914183   FDRGHMAPAG.[8].MDQTFYLSNMSPQ.[4].FNRHYWAYLEGFCRSL 133
gi 6093589   YDRGHQAPAA.[8].MDETFLLSNMAPQ.[4].FNRHYWAYLEGFMRDL 201
gi 17233277  FDRGHMAPSA.[8].NSATFLMTNIIPQ.[3].NNQGIWANLENYSRNL 165
gi 18203628  WSRGHMAPAG.[8].MAETFYLSNIVPQ.[3].NNSGYWNRIEMYCREL 185
```



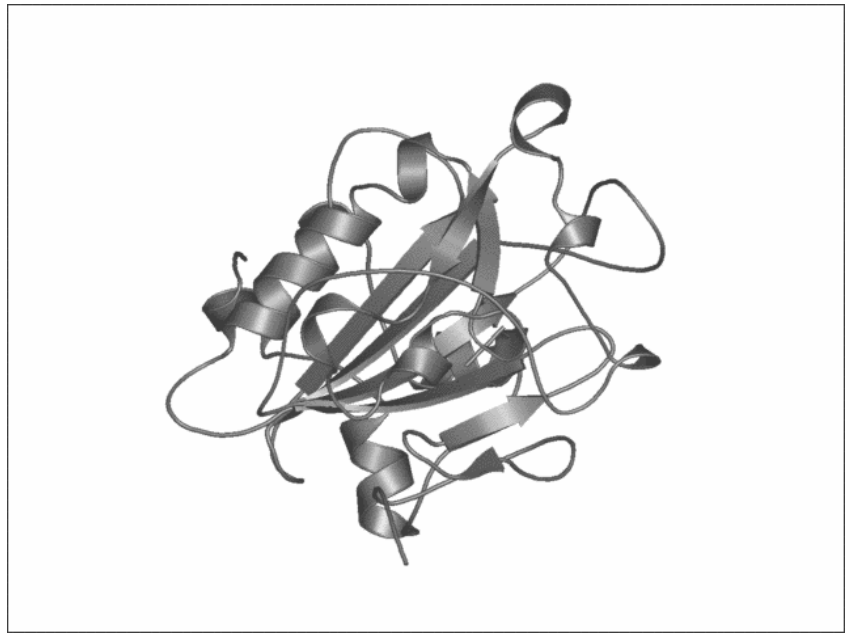
Split:

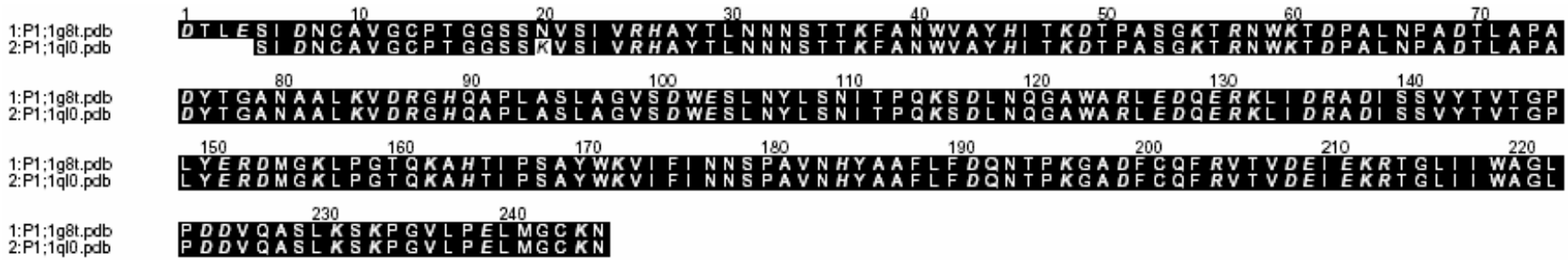
```
                  β                                α
NucA NAATFLMTNMMPQ.[T↓PD].NNRNTWGNLEDYCREL
SM   WESLNYLSNITPQ.[K↓SD].LNQGAWARLEDQERKL
```

Hinge of SM nuclease

SM → d4N-SM http://molmovdb.org

by the Yale Morph Server





Split point of NucA

N-...-Thr-|-Pro-...-C

NucANFlu:

OmpA-Flag-NucAN-GGSGGGSGGS-aFlu-His_5

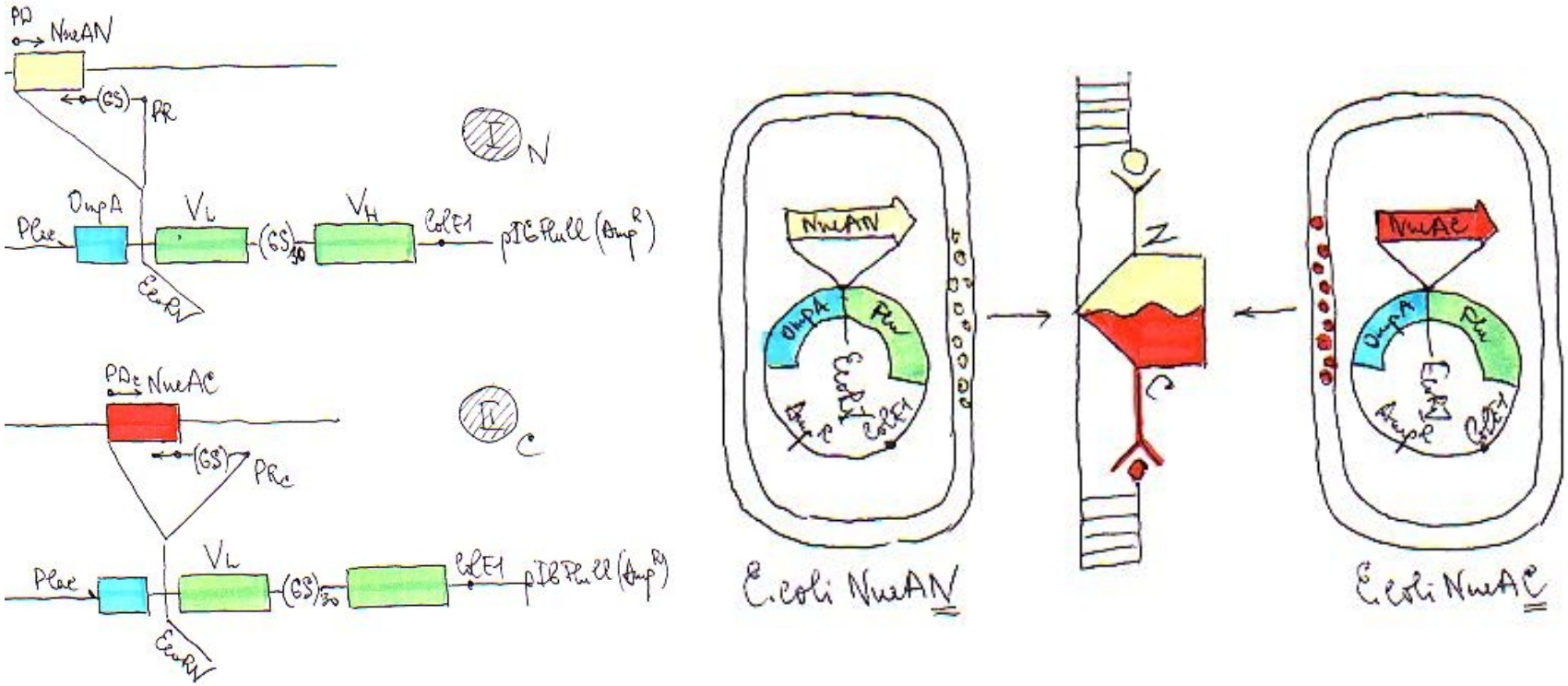
47.2kD

NucACFlu:

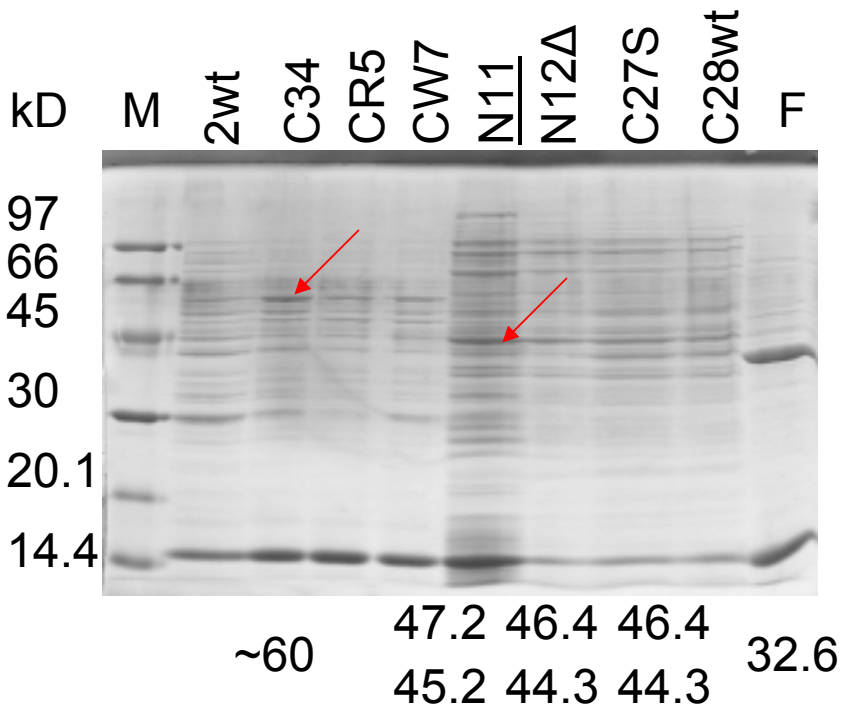
OmpA-Flag-GG-NucAC-GGSGG-aFlu-His_5

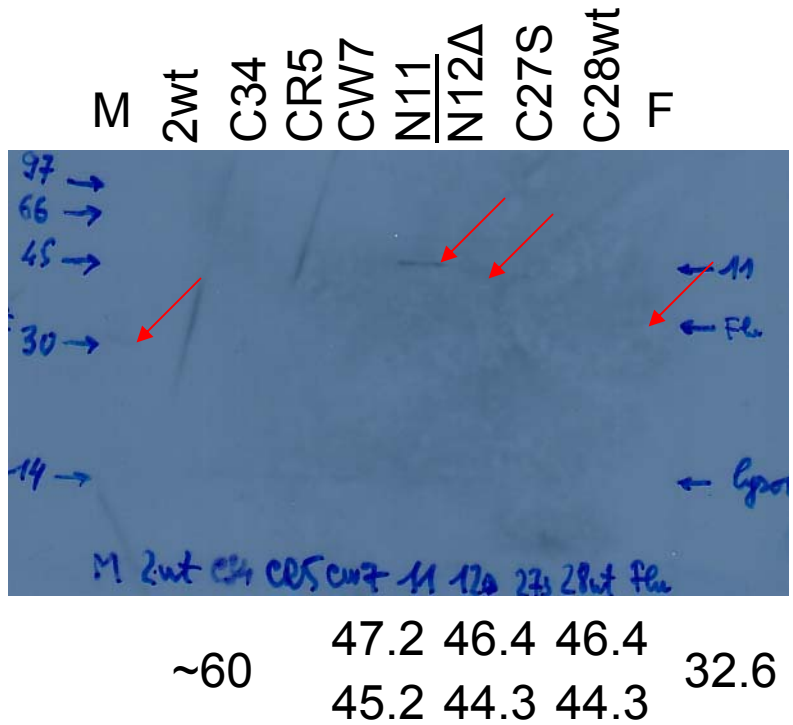
46.4 kD

Cloning, expression, and test of β-IGNAF *in vitro*



SDS electrophoresis & Western blot





The stable form of NucAN/2-Flu # N11 is detected ☺

Next problem is the low level protein expression

Comparison of α-, β-, and γ-versions

IGNAF		2xNucA/2-Flu		2xNucA/2-FluDig	
α-version (in a refrigerator)		β-version (in a refrigerator)		γ-version (yet mental)	
advantage	**disadvantage**	**advantage**	**disadvantage**	**advantage**	**disadvantage**
1 molecule			2 molecules		2 molecules
	permanent activity	regulation by selfassembling		regulation by selfassembling	
	wobbling	fixation		fixation	
			homopod - 25% activity	heteropod - 100% activity	
1	**2**	**2**	**2**	**3**	**1**
☹		😐		☺	

Outlook: codon optimization, DNA synthesis, minimal cell

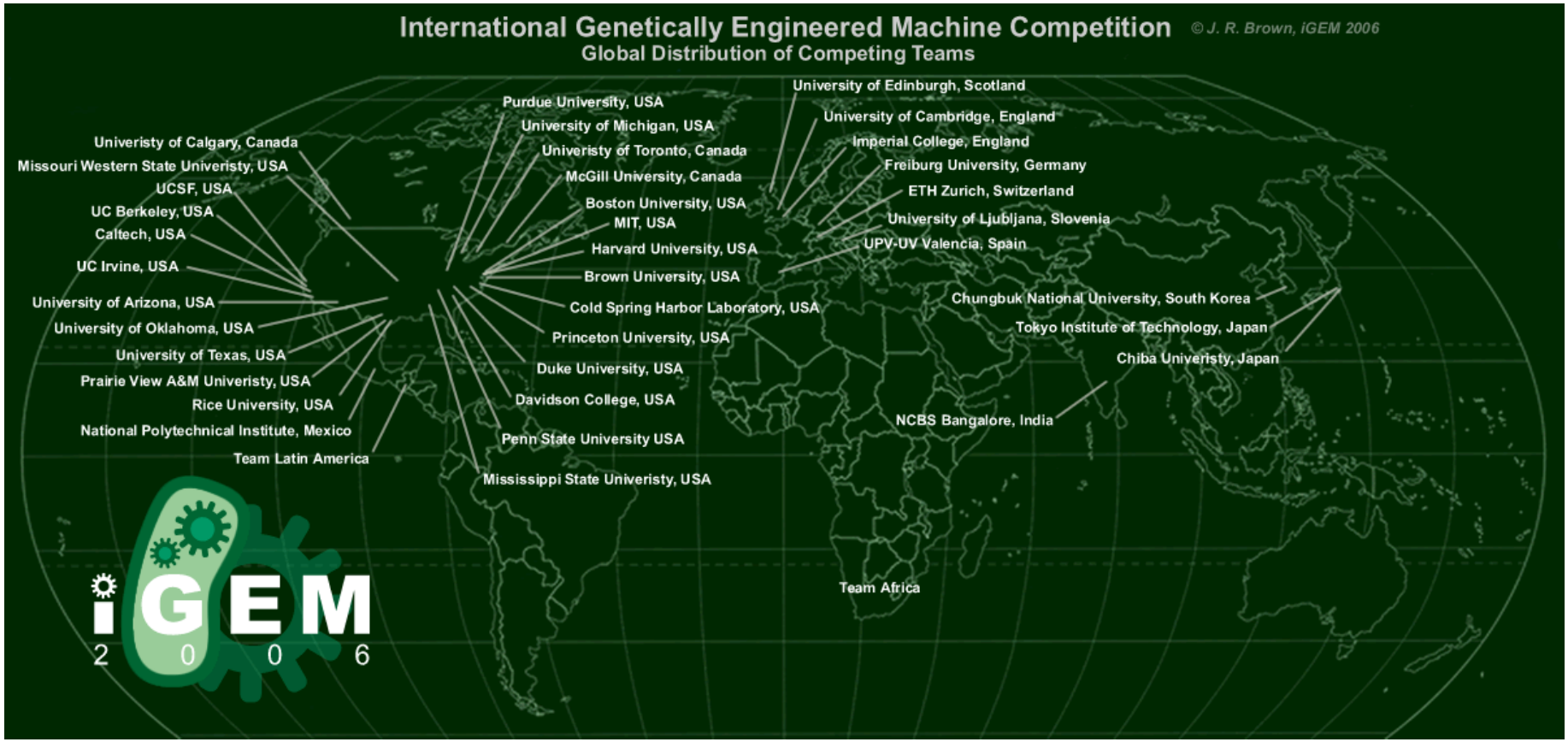
- IGNAF protein consists of two parts:
  1. NucA endonuclease from cyanobacterium *Anabaena* sp., and
  2. scFv mouse antibody to fluorescein from Eukaryote
  - This chimera expressed in the Enterobacteria *Escherichia coli*
  - Is it a challenge now?

- Codon optimization by DNA2.0, Gene Composer™, or GeneDesign

- An order of 10 Kbp DNA fragment over the web with low cost $0.85 to $1.60 per bp
- It is possible to build more than 100 Kbp DNA fragments
- Throughput of DNA synthesis by different firms:
  - 8Kb Atactic, Invitrogen
  - 44Kb Agilent
  - 48Kb febit
  - 100Kb Metigen
  - 760Kb Nimblegen
  - ~Mb Blue Heron, Codon Devices (BioFAB™ platform)
- Some researchers expect that a ~1 Mbp bacterial genome will be constructed within 1-2 years

Mutants of all species, recombine!

Martin Schneider



Freiburg University 2006 - IGEM - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address http://parts.mit.edu/wiki/index.php/Freiburg_University_2006 Go Google

Kouznetsov my talk preferences my watchlist my contributions log out

article discussion edit history move watch

navigation

- Main Page
- About iGEM
- School Pages
- Ambassadors
- Teaching Resources
- OpenWetWare
- Help!

the registry

- Registry
- About the Registry

past/present/future years

- Registration FAQ
- 2006 New team FAQ
- 2005
- 2004
- 2004 (IAP)
- 2003 (IAP)

Freiburg University 2006

Welcome! We are Alife Mutants.

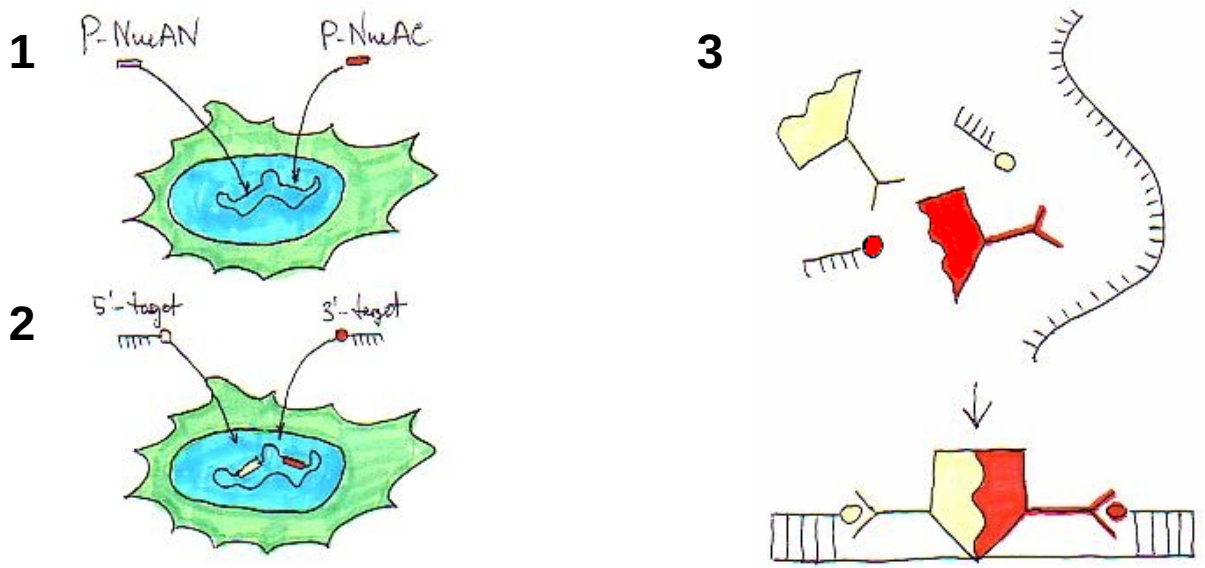
This term was invented by Martin Schneider on the Rule 110 Winter Workshop in 2004 [1]. We play without rules. We discover the rules that govern life, the universe and everything to exploit these rules and to create Artificial Life. Our short-time aim is the trip to Boston in October 2006 to take a prize in the iGEM.

Contents [hide]

[edit]

Internet

Target activation of an installed 2xNucA/2 *in vivo* or in A-cell



NucAN + NucAC = NucA

1. Preinstallation of transgenes
2. Introduction of oligonucleotides (input)
3. Target activation by <u>selfassembling</u>

Theoretically, no any background activity!!!

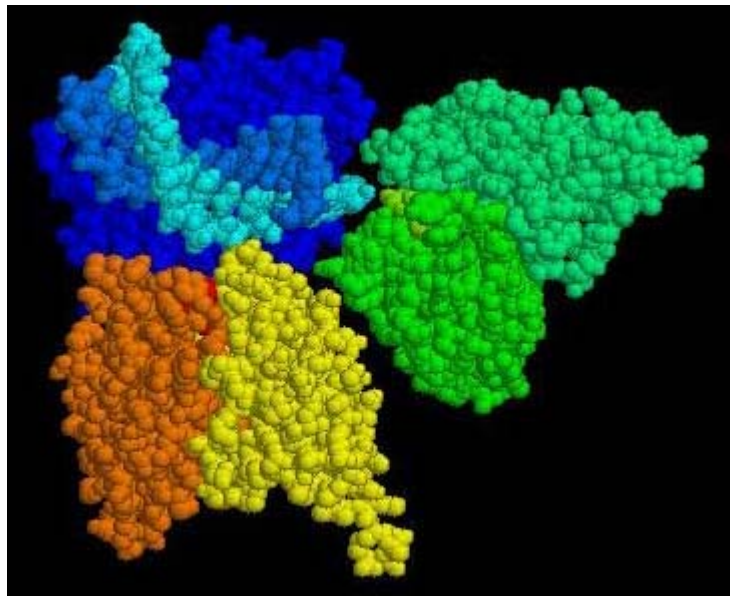
Conclusion

- *Cut-paste-select-and-transpose* model is a kind of constructive mutagenesis
- ***AM*** is a set of stochastic cut-paste agents, which act in parallel on their own tapes accordingly the instructions (input words), communicate with each other by transpositions of the tapes and interact with the environment to compare the output states. Based on the comparison it accepts or runs in a loop to fit the environment
- A computation power of ***AM*** depends on the number of agents and the number of output states for each agent
- The elongation of input words leads to the increasing of building blocks and to the hierarchical assembling

- Two different 'legs' are more preferred to achieve the particular orientation of guided nuclease on DNA; the input comprising two half-words should be studied in the Argo-machine
- Transpositions and a compartmentalization of reactions could be implemented in the frame of 'minimal cell' project
- 'Argonauts' may be seen as a part of living/artificial cells to generate a diversity in order to search for solutions

Thank's to

Dry lab:
Mikhail Kats
Andreas Karwath
Genaro Martinez
Elena Losseva
Marian Gheorghe
Paul Rothemund
Matthew Cook
George Paun



Wet lab:
Thomas Willemsen
Jody Mason
Andrew Hessel
Randy Rettberg
Drew Endy
Alfred Pingoud
Andreas Pluckthun
Albrecht Sippel

and all A-life Mutants