

# The Relay Series

## A Model for Artificial RNA Evolution

Ulrike Langhammer

Institute for Theoretical Chemistry and Structural Biology

University of Vienna, Austria

*Bled 2004, February 23<sup>rd</sup> till 27<sup>th</sup>*

# Outline

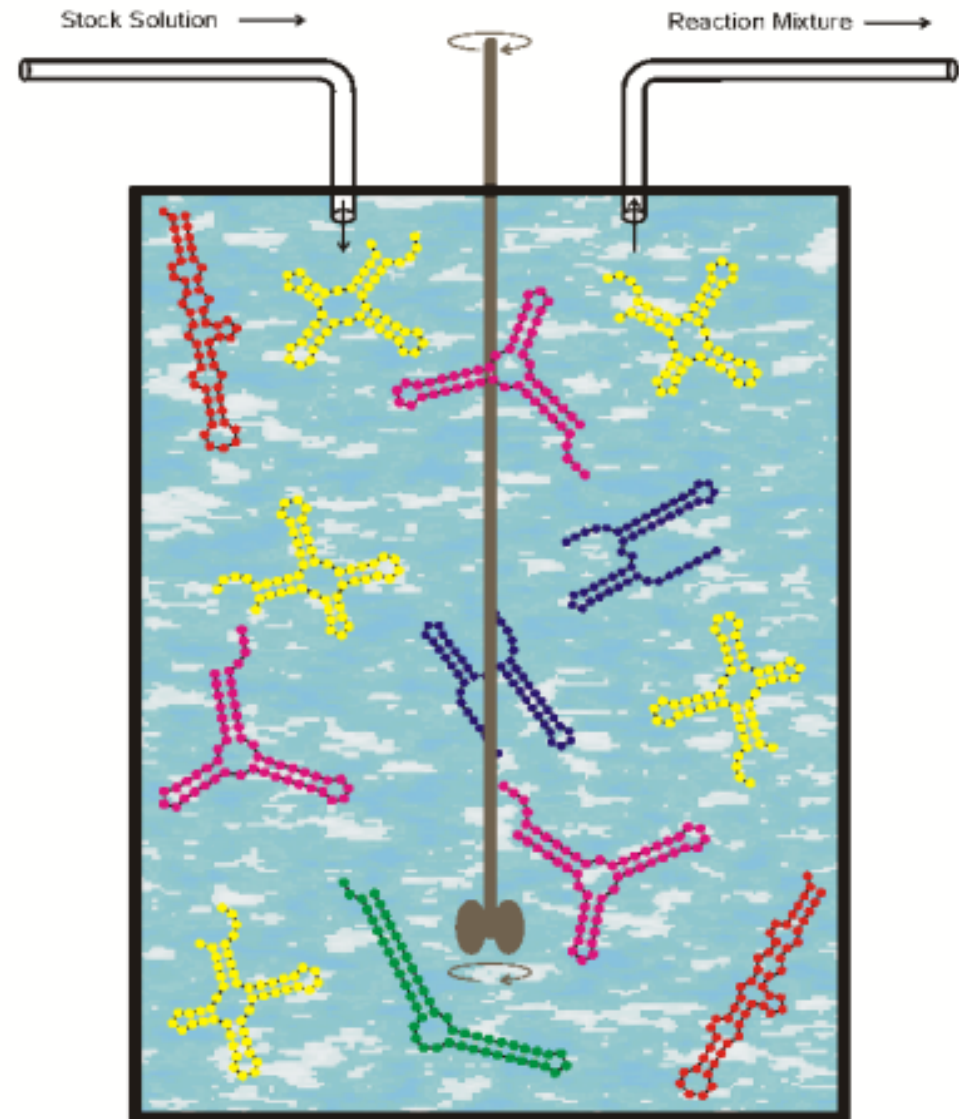
- What are Relay Series?
- The Flowreactor
- Characteristics of a Relay Series
- Continuous and Discontinuous Transitions
- Relay Series with Different Alphabets and Population Size
- Relay Series illustrated as Coarse Grained Structures
- Conclusion

## What are Relay Series?

- A relay serie is a way to reconstruct the succession of structures in an evolutionary process simulated in a flowreactor.
- It is a list of structures, beginning with the target structure and ending with the first structure.
- The relay series is reconstructed by searching for the shape  $\alpha_{n-1}$  which gave birth to the target shape  $\alpha_n$  until the shape  $\alpha_0$  is found.

# The Flow Reactor

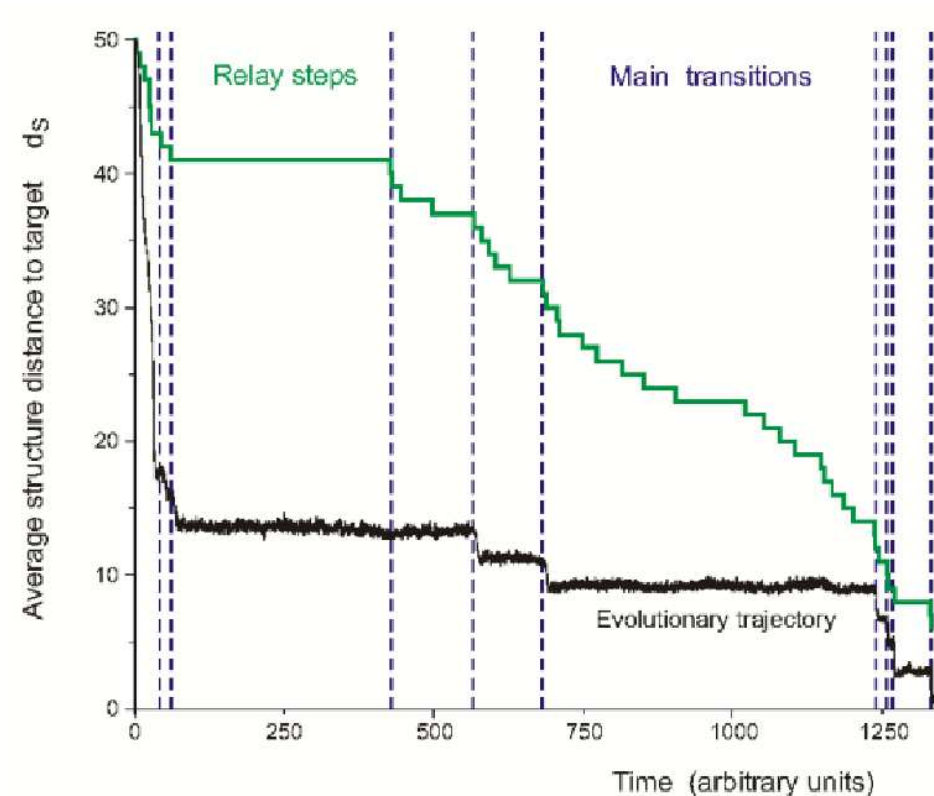
- Evolution of populations is studied under replication and mutation
- influx of stock solution and an unspecific outflux of the reaction mix
- Parameters: population size  $N$ , the chain length  $n$ , mutation rate per nucleotide and replication  $p$
- The fitness of a structure in the reactor is defined by various distances from the shape in the reactor to the target shape.



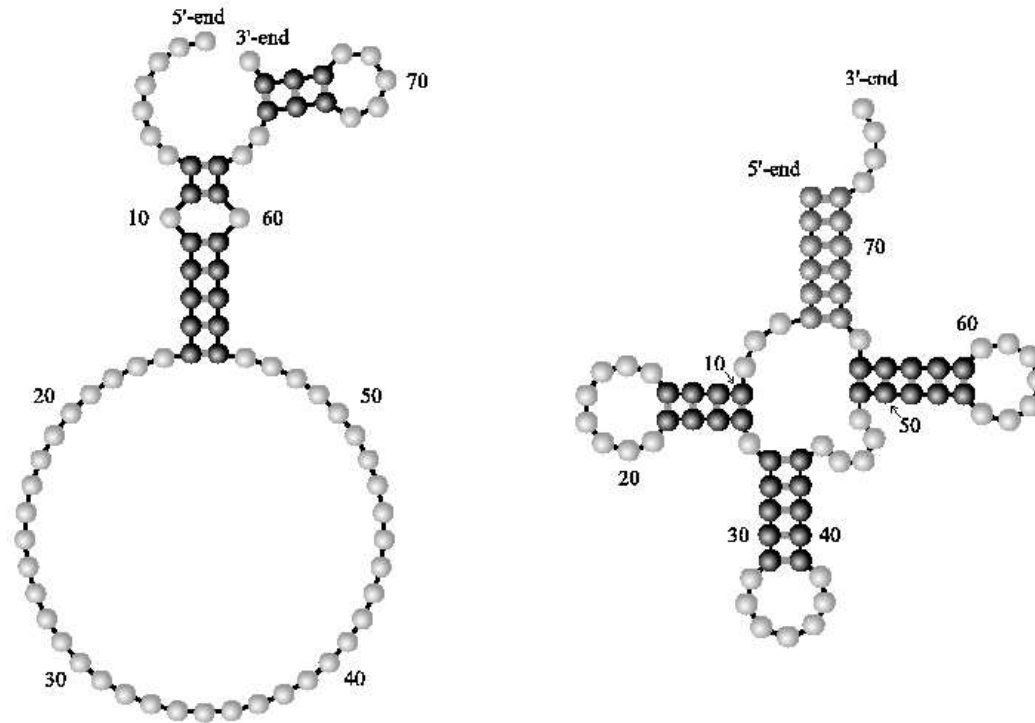
# Characteristics of a Relay Serie

In every relay series there are two phases:

- The initial period with fast increase of mean fitness, many structural changes and small genotypic changes.
- The second period with long epochs of stasis and short but strong improvements.



# Initial and Target Structure of the Relayseries



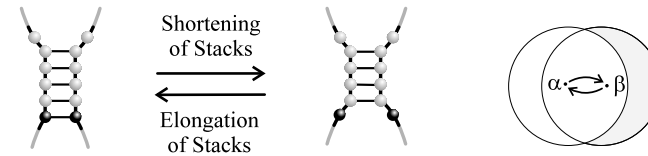
A .....((.((((.....))))).)..(((.....))).  
 B ((((((...(((.....))))).((((.....))))). ....((((.....))))).))))). ....

*Left, A: Initial structure, Right, B: Target structure*

# Reconstructed Relay Series

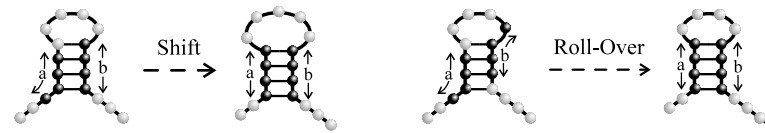
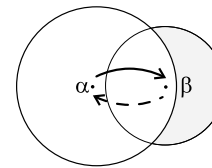
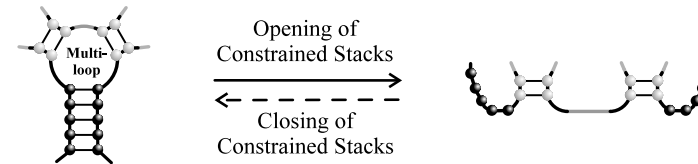
No.	Hamming distance to target	Shape	c or d
25	0	(((((...(((.....))))).(((.....))))....(((.....))))..)))))....	d
24	2	(((.(((.....))))).(((.....))))....(((.....))))..)))))....	c
23	2	(((.(((.....))))).(((.....))))....(((.....))))..)))))....	d
22	2	(((.(((.....))))).(((.....))))....(((.....))))..)))))....	d
21	2	(((((...(((.....))))).(((.....))))....(((.....))))..)))))....	c
20	2	.(((.....))))).(((.....))))....(((.....))))..)))))....	d
19	4	.(((.....))))).(((.....))))....(((.....))))..)))))....	c
18	4	.(((.....))))).(((.....))))....(((.....))))..)))))....	c
17	6	.(((.....))))).(((.....))))....(((.....))))..)))))....	d
16	10	.(((.....))))).(((.....))))....(((.....))))..)))))....	c
15	8	.(((.....))))).(((.....))))....(((.....))))..)))))....	d
14	6	.(((.....))))).(((.....))))....(((.....))))..)))))....	c
13	6	(((((...(((.....))))).(((.....))))....(((.....))))..)))))....	d
12	8	(((((...(((.....))))).(((.....))))....(((.....))))..)))))....	d
11	10	(((((...(((.....))))).(((.....))))....(((.....))))..)))))....	c
10	10	(((((...(((.....))))).(((.....))))....(((.....))))..)))))....	c
9	10	(((((...(((.....))))).(((.....))))....(((.....))))..)))))....	c
8	10	(((((...(((.....))))).(((.....))))....(((.....))))..)))))....	c
7	12	(((((...(((.....))))).(((.....))))....(((.....))))..)))))....	d
6	18	.....(((.....))))).(((.....))))....(((.....))))..)))))....	c
5	20	.....(((.....))))).(((.....))))....(((.....))))..)))))....	d
4	27	.....(((.....))))).(((.....))))....(((.....))))..)))))....	d
3	29	.....(((.....))))).(((.....))))....(((.....))))..)))))....	d
2	39	.....(((.....))))).(((.....))))....(((.....))))..)))))....	d
1	48	.....(((.....))))).(((.....))))....(((.....))))..)))))....	-

# Transitions



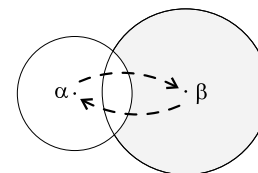
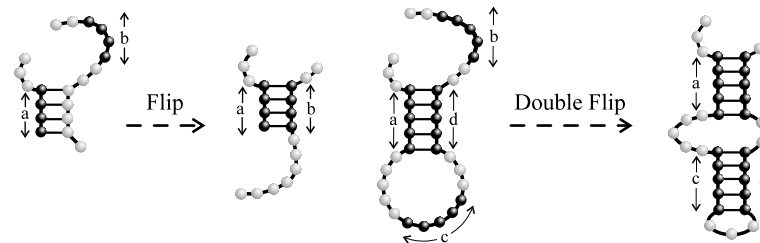
## Continuous Transitions

Minor changes like opening or closing of a base pair at one end of a stack.



## Discontinuous Transitions

Major structural changes like a shift of a whole base pair stack or the creation of a long stack in one step. They do affect structures which are not near each other.





# Continuous and Discontinuous RNA Transitions

To distinguish between minor (= continuous transitions) and major (=discontinuous transitions) the algorithm uses two distance and one relation measurements:

- The *Hamming distance*  $d_{ij}^s$  counts the number of positions in which the two sequences differ.
- The *base pair distance*  $d_{ij}^{bp}$  counts the minimal number of openings and closing of base pairs to transform one structure into the other.
- The asymmetric *base pair preserve relation*  $d_{ij}^{bpp}$  counts the number of base pairs, that can not be found in the other structure.

The transitions from one structure to another ist discontinuous, when

$$2 \cdot d_{ij}^{bp} > d_{ij}^s \text{ or} \\ 2 \cdot d_{ij}^{bp} = d_{ij}^s \geq 4 \text{ and } d_{ij}^{bpp} = 0 \text{ is true.}$$

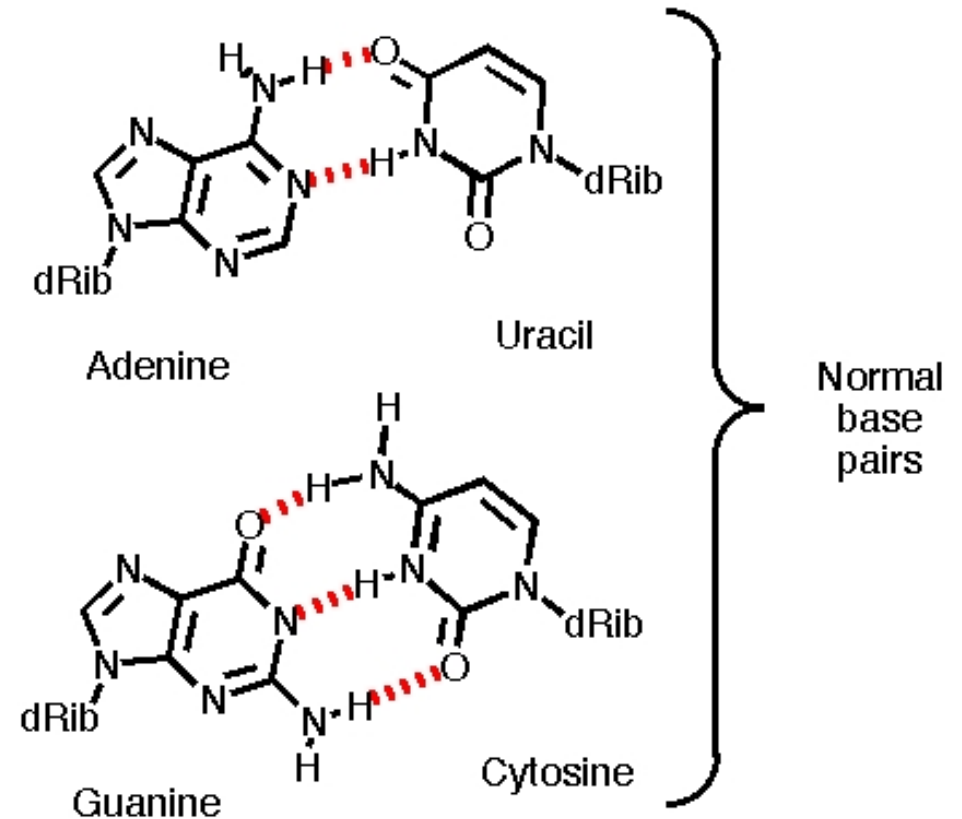
# Relay Series with Different Alphabets and Different Population Sizes

- Relay series with AUGC, GUC and GC with a population size of 3000 were analyzed.
- Relay series with 15, 1000, 3000 and 10000 individuals with an alphabet of AUGC were analyzed.
- Now we are also running the flow reactor with individuals of 5, 13 - 22, 25, 30 with the alphabet AUGC.

# Relay Series with Different Alphabets

Aims:

- Why are all four base pairs involved in RNA and DNA?
- Population of 3000 Individuals were chosen in the flow reactor.



## Relay Series with Different Alphabets

- 212 Relay series were analyzed with each alphabet.
- As expected, there are more transitions to reach the target with the alphabet GUC and GC than in AUGC.

	transitions	continuous transitions	discontinuous transitions
AUGC	25 ( $\pm 9.9\%$ )	12 ( $\pm 7.4\%$ )	14 ( $\pm 3.9\%$ )
GUC	31 ( $\pm 5.7\%$ )	14 ( $\pm 4.0\%$ )	17 ( $\pm 4.1\%$ )
GC	40 ( $\pm 6.8\%$ )	19 ( $\pm 5.5\%$ )	20 ( $\pm 3.6\%$ )

- More structures are formed during this process.

	amount of transitions	structures appearing once	structures 2-5	structures > 50
AUGC	3362	85%	12.4%	2.6% (=89)
GUC	5284	92.6%	6.47%	0.9% (=48)
GC	6735	92.9%	6.1%	0.95% (=64)

- Many structures appear only once.

# Relay Series with Different Population Sizes

Aim:

- How do the relay series change with the population size in the reactor?
- Is there any consequence in decreasing the population size?

# Relay Series with Different Population Sizes

- 120 relay series with the alphabet AUGC where analyzed with 15, 1000, 3000, 10000 individuals in the reactor
- The transition steps decrease with higher population number in the flow reactor.

population size	transitions	continuous transitions	discontinuous transitions
15	4801.2 ( $\pm 4446.3\%$ )	1391.7 ( $\pm 1268.5\%$ )	3409.5 ( $\pm 3181.5\%$ )
1000	42 ( $\pm 27.2\%$ )	26 ( $\pm 24.5\%$ )	15.8 ( $\pm 4.4\%$ )
3000	24 ( $\pm 9.0\%$ )	11 ( $\pm 7.1\%$ )	13 ( $\pm 3.1\%$ )
10000	19 ( $\pm 4.6\%$ )	7 ( $\pm 3.0\%$ )	11 ( $\pm 2.7\%$ )

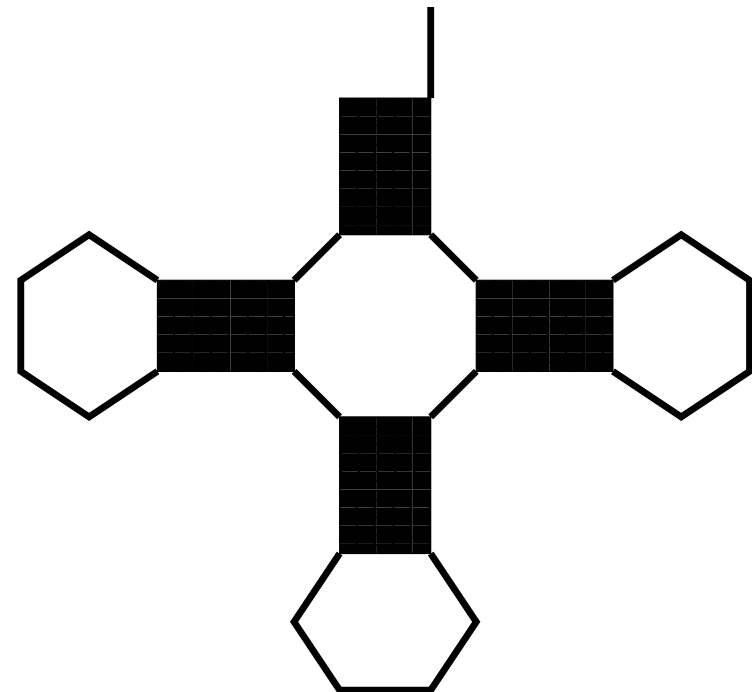
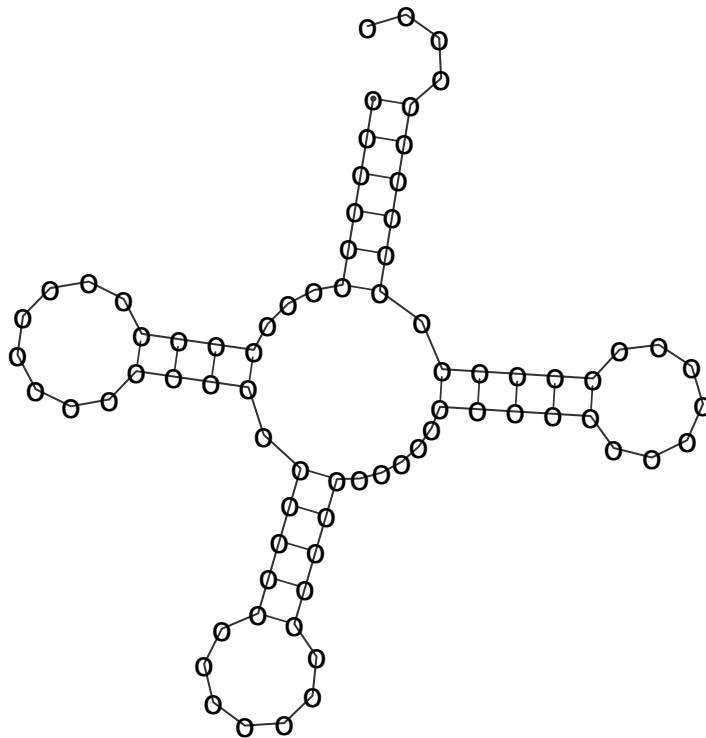
- The percentage of how many structures appear only once stays almost constant.

population size	amount of transitions	structures appearing once	structures 2-5	structures > 50
15	552945	96.9%	3.0%	0.039% (= 219)
1000	2809	84.3%	12.4%	3.3% (=92)
3000	2004	86.4%	11.9%	1.6% (=33)
10000	1508	85.4%	12.4%	2.1% (=31)

- With decreasing population size in the flow reactor the possibility that the population never reaches the target structure increases, e. g. for 120 populations which reach the target, 380 populations die out.

# Relay Series illustrated as Coarse Grained Structures

- Information about the size of stacks and loops is lost.
- Only the information about the structural elements are kept.
- 120 relay series with the alphabet AUGC where analyzed with 15, 1000, 3000, 10000 individuals in the reactor.



# Coarse Grained Structures

Reducing the relay series to coarse grained structures

- reduces the transition steps.

population size	transitions	continuous transitions	discontinuous transitions
15	3702 ( $\pm 3444.6\%$ )	673 ( $\pm 619.0\%$ )	3028 ( $\pm 2827.2\%$ )
1000	20 ( $\pm 10.1\%$ )	8 ( $\pm 8.1\%$ )	12 ( $\pm 3.5\%$ )
3000	14 ( $\pm 4.4\%$ )	3 ( $\pm 2.2\%$ )	10 ( $\pm 2.9\%$ )
10000	11 ( $\pm 3.7\%$ )	1 ( $\pm 1.9\%$ )	8 ( $\pm 2.5\%$ )

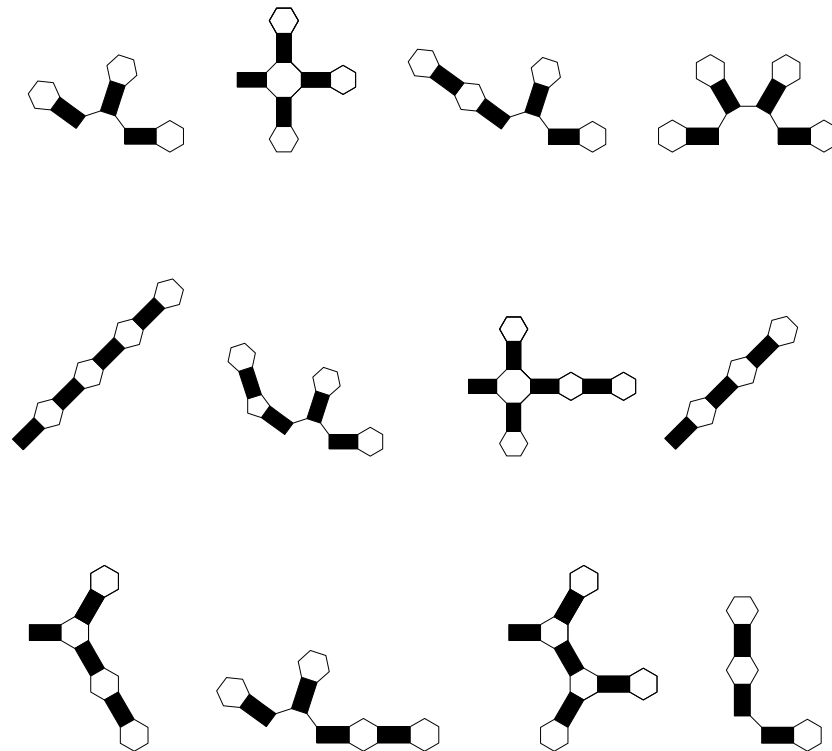
- There are fewer discontinuous transitions, because shifts are removed from the relay series.

population size	amount of transitions	structures appearing once	structures 2-5	structures > 50
15	36843	50.4%	29.3%	20.3% (= 7467)
1000	374	55.5%	27.7%	16.5% (=62)
3000	543	54.2%	26.3%	19.3% (=105)
10000	250	54.6%	27.9%	17.1% (=43)

- There are not so many structures that appear only once.



# Common Coarse Grained Structures



In every population appear the same common coarse grained structures independly on the population size or the used alphabet.

# Summary1

- The relay series consists of continuous and discontinuous transitions.
- Properties of discontinuous transitions are:
  - They involve major structural changes.
  - They are infrequent and have a small probability to occur.
- The number of relay steps decreases with increasing population, but the number of discontinuous transitions stays almost constant.

## Summary 2

- Smaller flow reactors provide too small space to hold many different shapes, that could connect two discontinuous transitions in a relay series, they have to be recreated which leads to more continuous relay steps.
- When there are only a few individuals in the flow reactor the population can even die out.
- Compared to the AUGC alphabet, GUC and GC relay series are longer and need more discontinuous transitions to get to the target shape and there are more structures that appear only once..
- When using coarse grained structures instead of the normal structures, the number of transitions decreases even more. The same coarse grained transitions appear in every population, no matter which alphabet and population size is used.
- The number of transitions can decrease until the structures can only be generated by discontinuous transitions.

Thank you