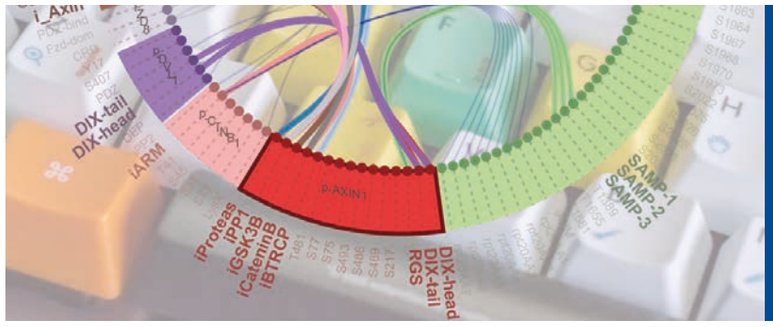


La biologie de l'information, un dialogue entre l'informatique et la biologie

Cours les mardis de 14h à 15h30
Amphithéâtre Maurice Halbwachs

Cours

29 octobre 2019	1 — La représentation de l'information biologique : Aspects statistiques du mappage des séquences aux structures ; le cas de l'ARN
05 novembre 2019	2 — L'héritage de l'information biologique : limites de transmission
12 novembre 2019	3 — Modélisation du traitement de l'information biologique I - Outils de raisonnement sur les systèmes d'interaction moléculaire : introduction Approches de la chimie et des interactions protéine-protéine basées sur des règles
19 novembre 2019	3 — Modélisation du traitement de l'information biologique II - Outils de raisonnement sur les systèmes d'interaction moléculaire : cas d'utilisation Modèles Kappa petits et grands
26 novembre 2019	3 — Modélisation du traitement de l'information biologique III - Outils de raisonnement sur les systèmes d'interaction moléculaire : La quête de renseignements utiles Analyse statique et causalité
03 décembre 2019	3 — Modélisation du traitement de l'information biologique IV - Systèmes d'assemblage combinatoire en signalisation moléculaire : échafaudage combinatoire
10 décembre 2019	3 — Modélisation du traitement de l'information biologique V - Systèmes d'assemblage combinatoire en signalisation moléculaire : Mécanique statistique
17 décembre 2019	4 — L'acquisition de l'information biologique : L'apprentissage dans les systèmes moléculaires ?

Nov 8: Eric Deeds, University of California at Los Angeles
"The evolution of cellular individuality"

Nov 15: Daniel Merkle, University of Southern Denmark
"Graph rewriting and chemistry"

Nov 22: Jean Krivine, Paris Diderot
"From molecules to systems: the problem of knowledge representation in molecular biology"

Nov 29: Eric Smith, Earth Life Sciences Institute, Tokyo
“Easy and Hard in the Origin of Life"

Dec 6: Massimiliano Esposito, University of Luxembourg
"Thermodynamics of Open Chemical Reaction Networks: Theory and Applications"

Dec 13: Yarden Katz, Harvard Medical School
"Cells as cognitive creatures"

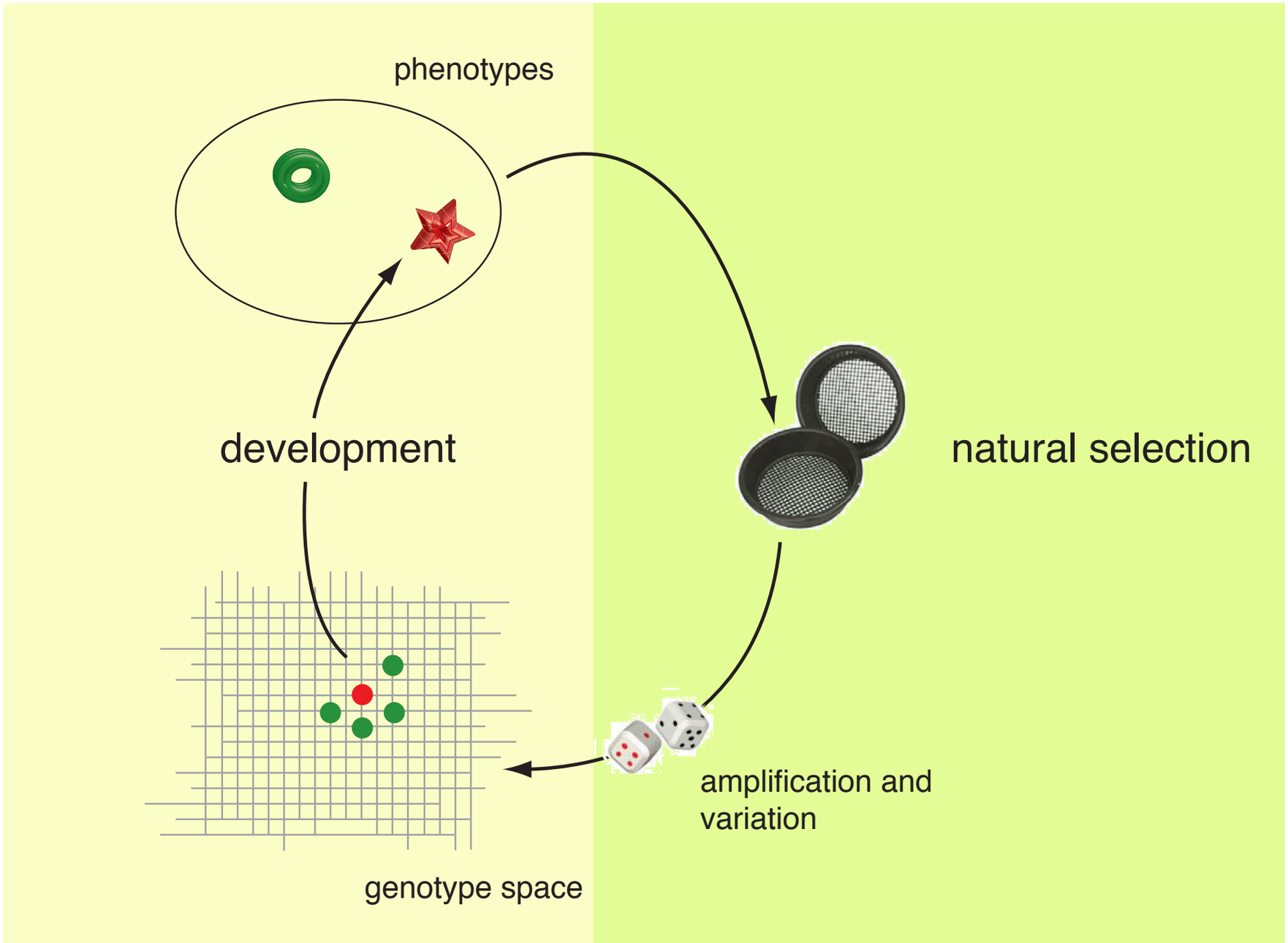
Jan 17: Aleksandra Walczak, ENS Paris
"Prediction in immune repertoires"

Jan 24: Tommy Kirchhausen, Harvard Medical School
"Imaging sub-cellular dynamics from molecules to multicellular organisms"

1.

The Topology of the Possible

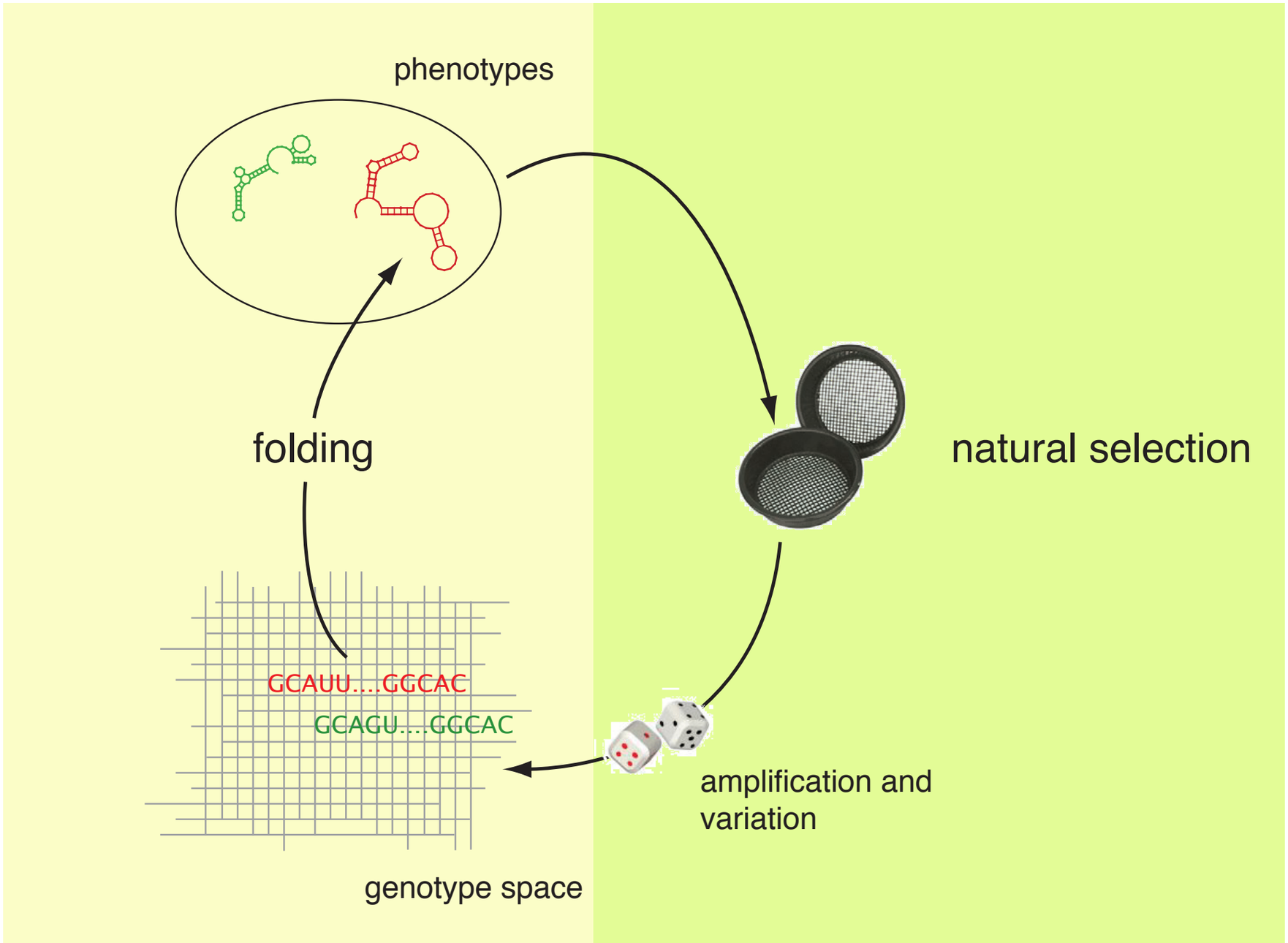
A Cartoon of Evolution



variation

selection

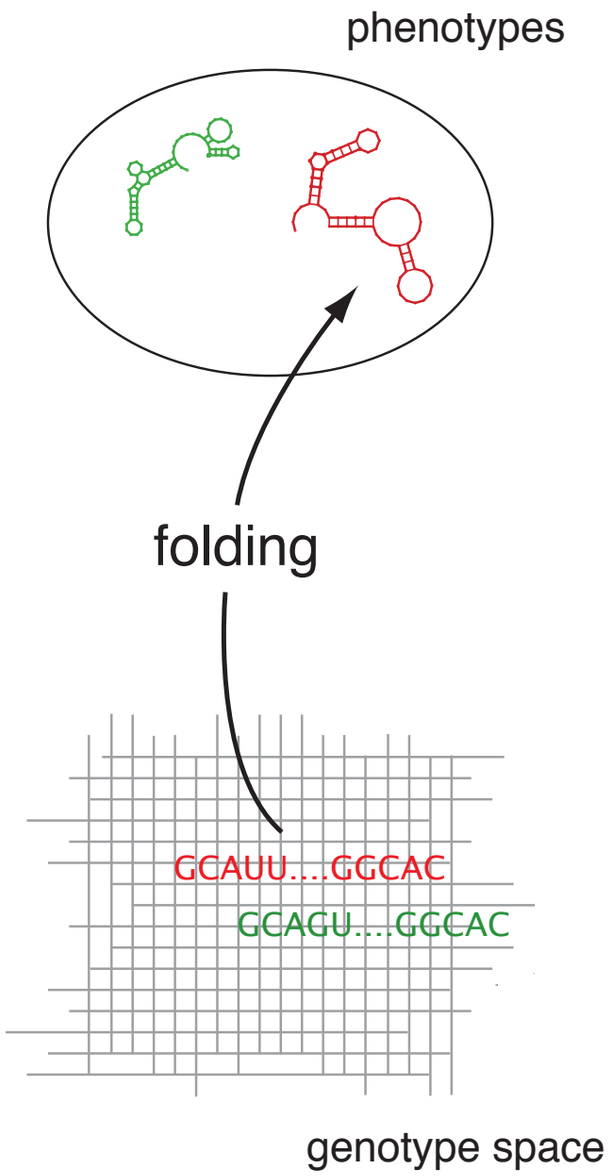
A Small Scale Model of (A Cartoon Of) Evolution



variation

selection

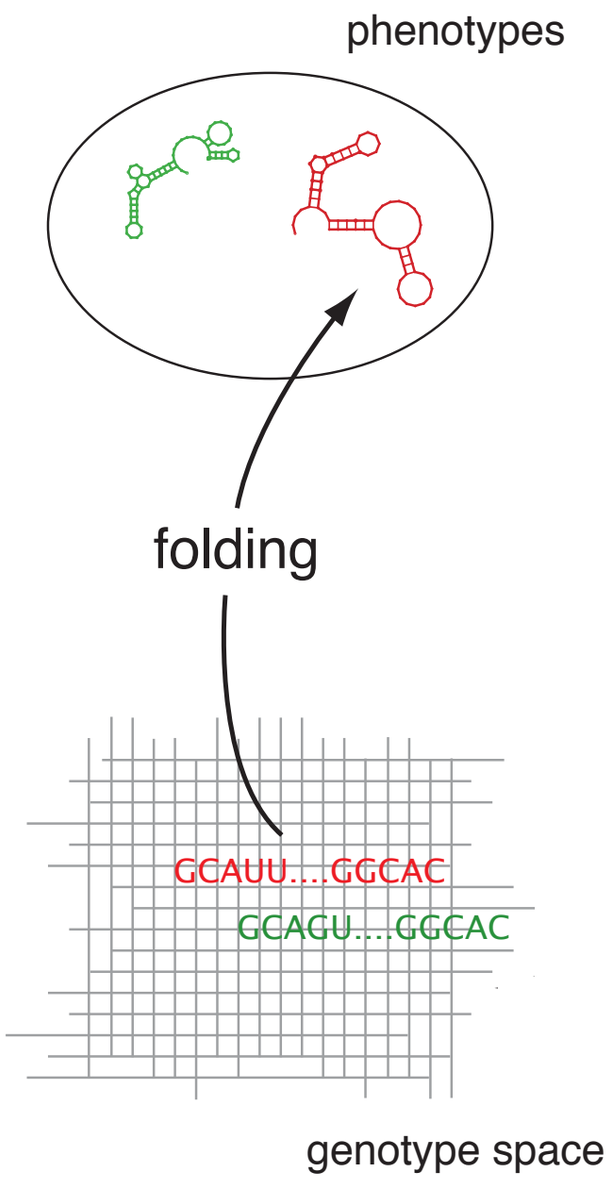
RNA Folding



conceptual focus is not on predicting the structure of a given sequence

conceptual focus is on predicting the statistical signature of the mapping from sequences to structures

RNA Folding



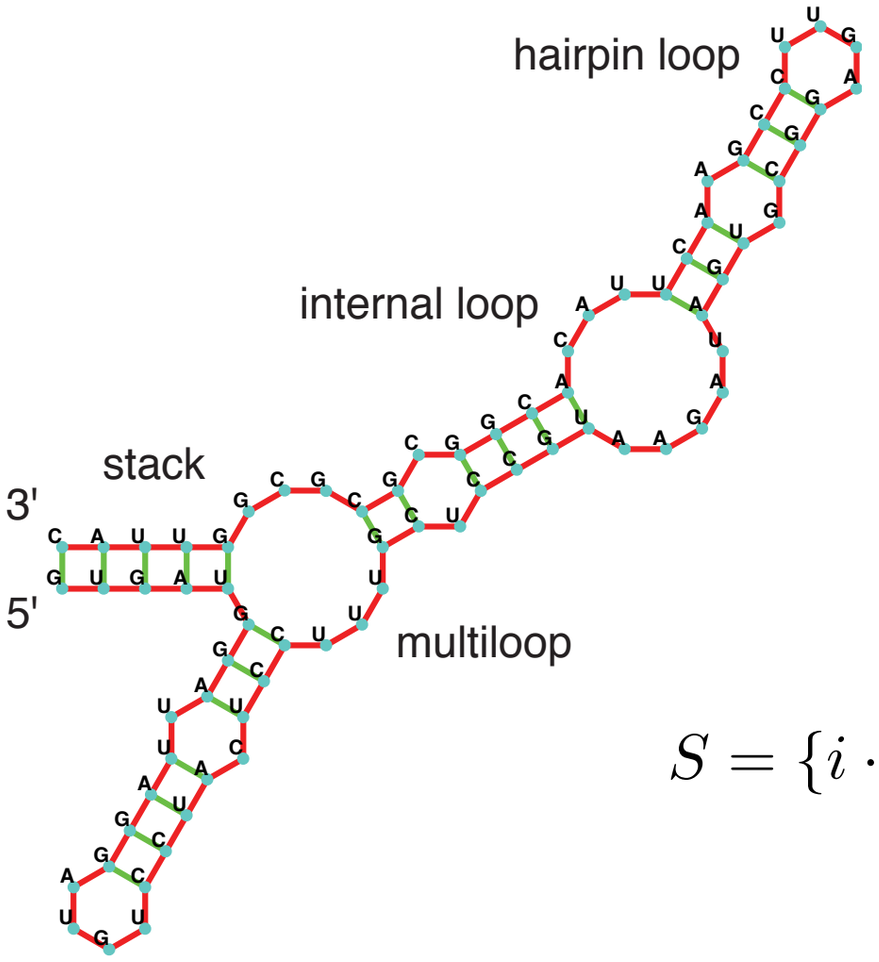
What kind of space is phenotype space?

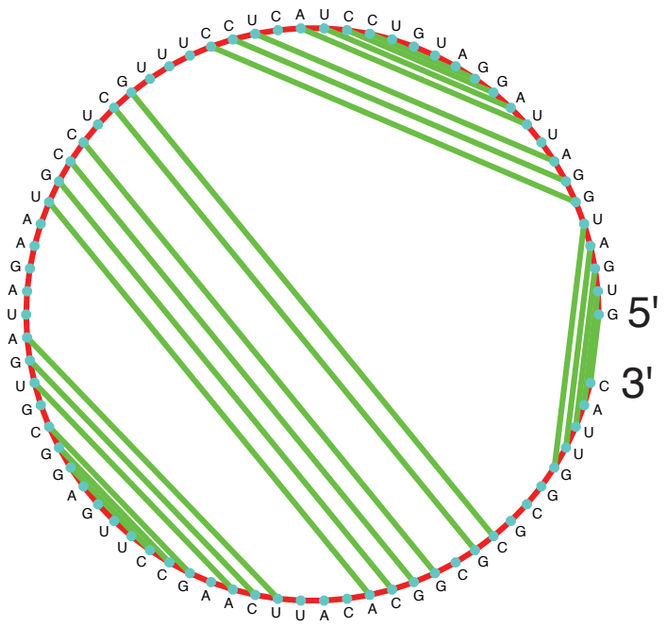
What do we mean by the "capacity to evolve"?

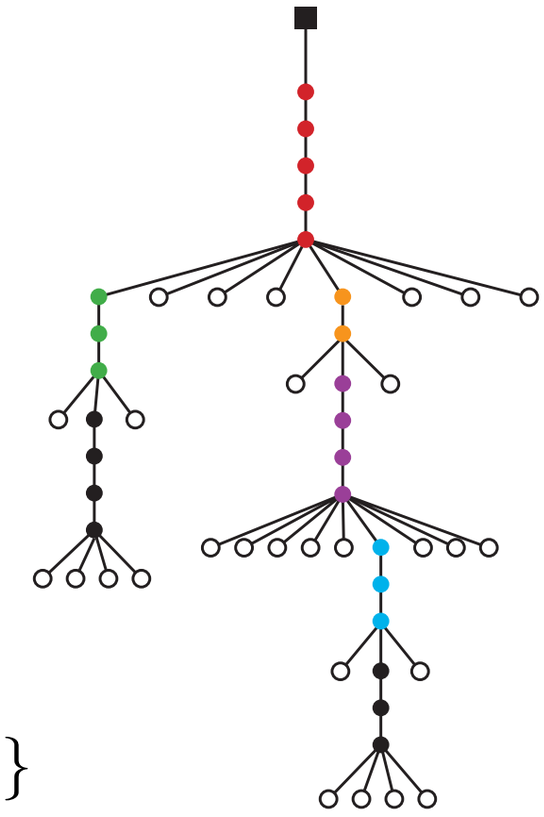
Does it depend on genotype?

Can it evolve?

RNA Secondary Structure





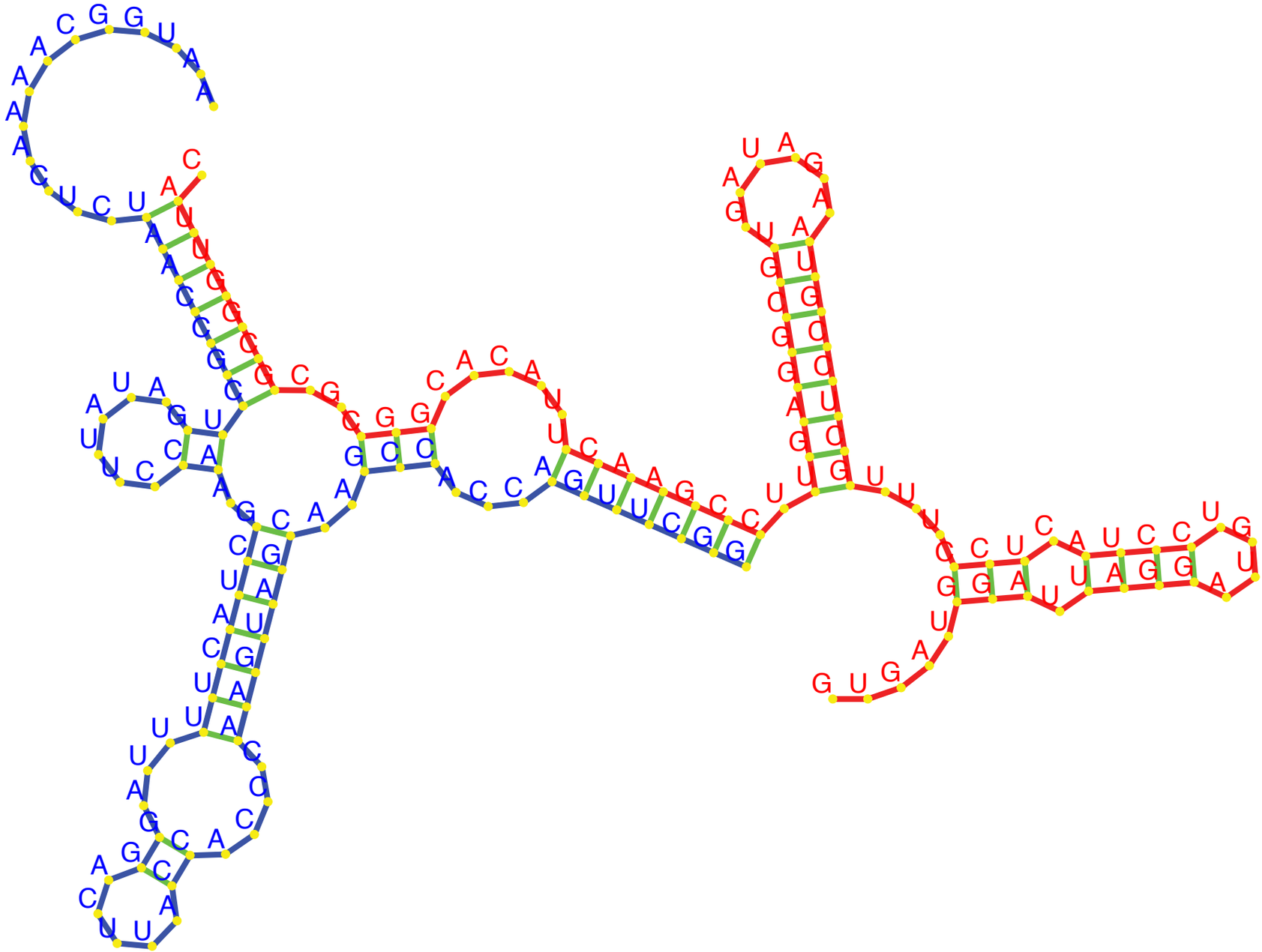


$$S = \{i \cdot j \,|\, \forall i \cdot j \;\; \text{s.t.} \;\; \nexists k \cdot l \quad i \leq k \leq j \leq l\}$$

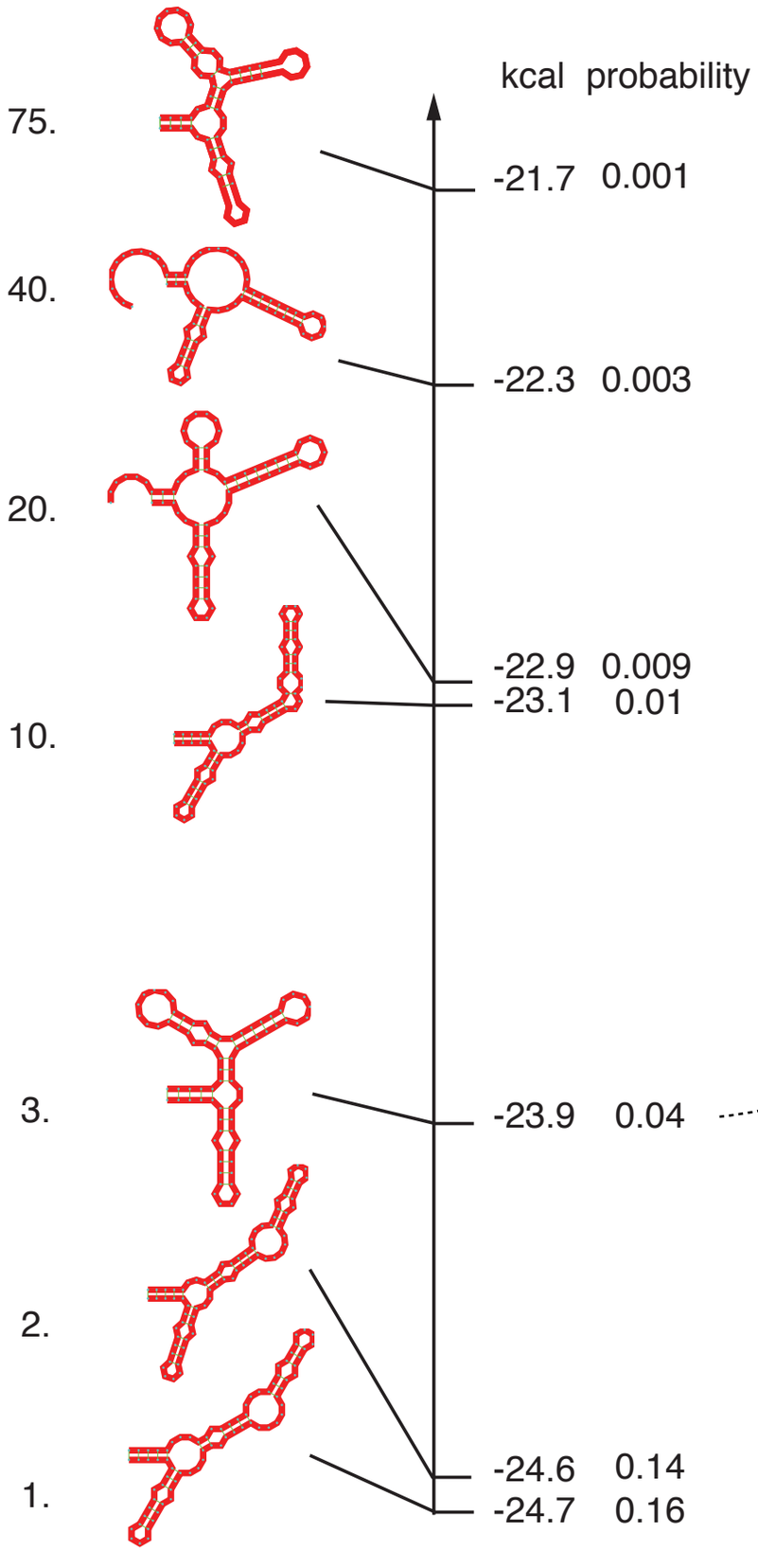
"chords don't cross"

((((((((.((((....)))).)))...((.((((.....(((.(((....))).)))...))))..))...))))

RNA Interaction



Suboptimal Structures



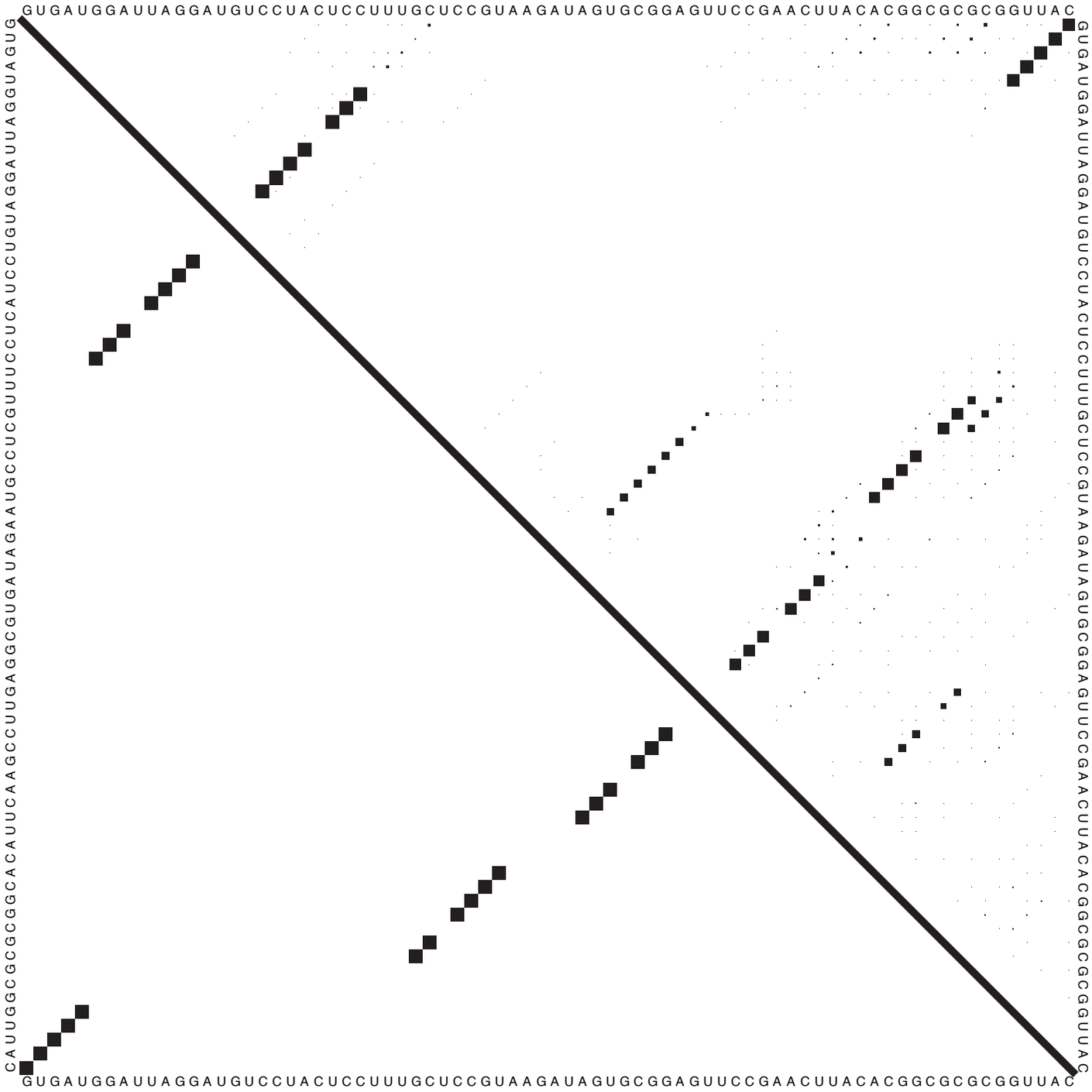
phenotypic plasticity
("microenvironmental")

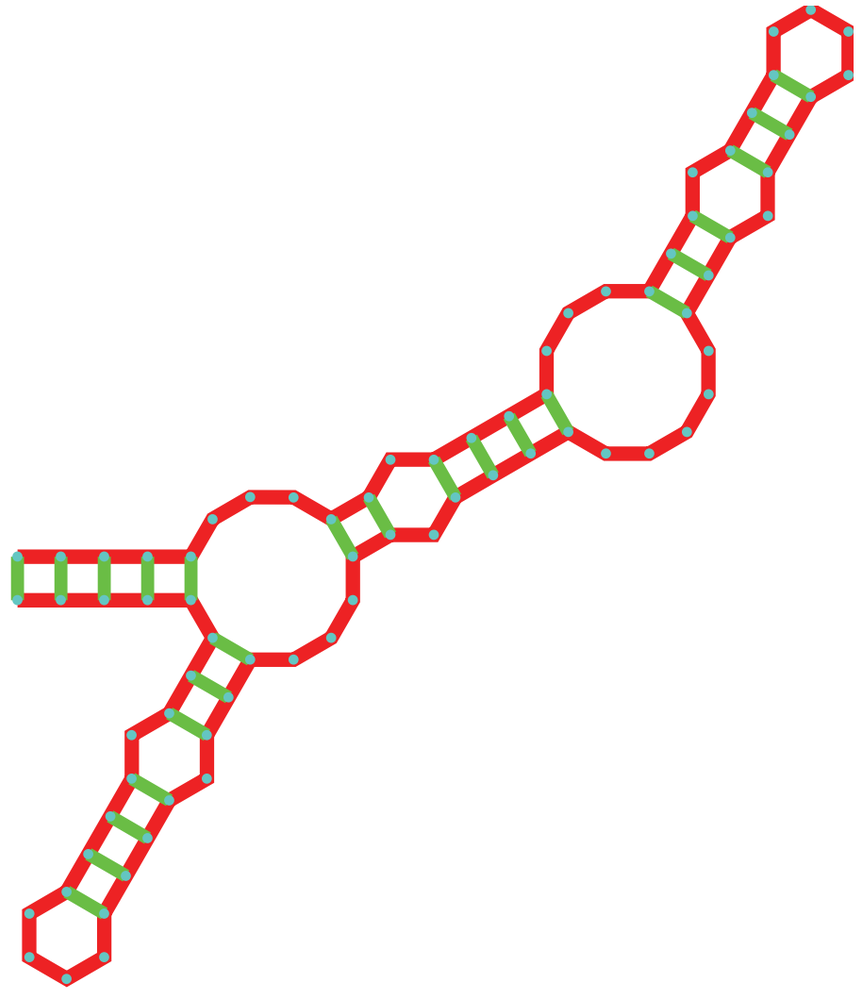
$$P(S_j) = \frac{e^{\Delta G_{S_j}/kT}}{\sum_i e^{\Delta G_{S_i}/kT}}$$

partition function Z

the probability of a structure

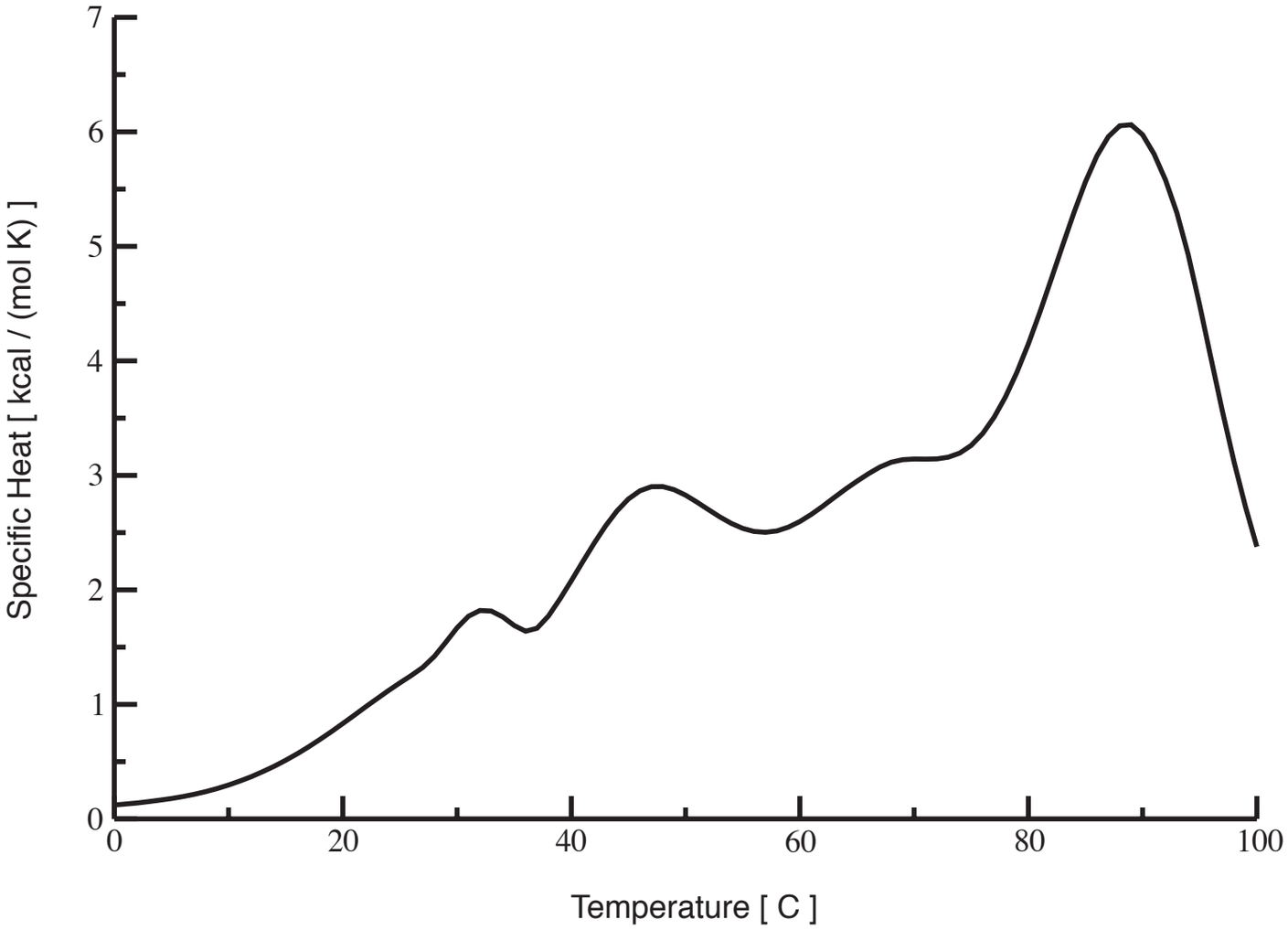
Base Pair Probabilities

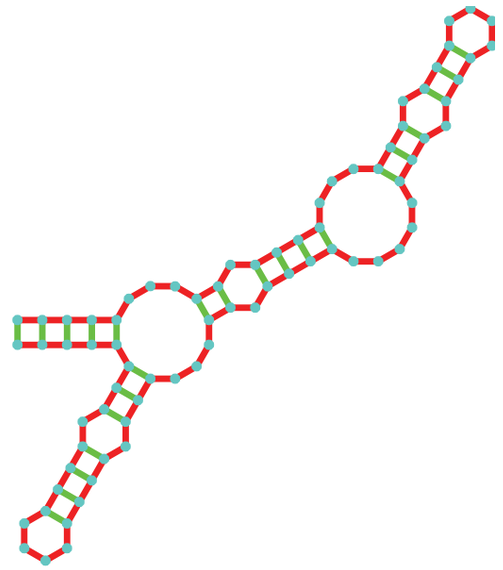




Heat Capacity

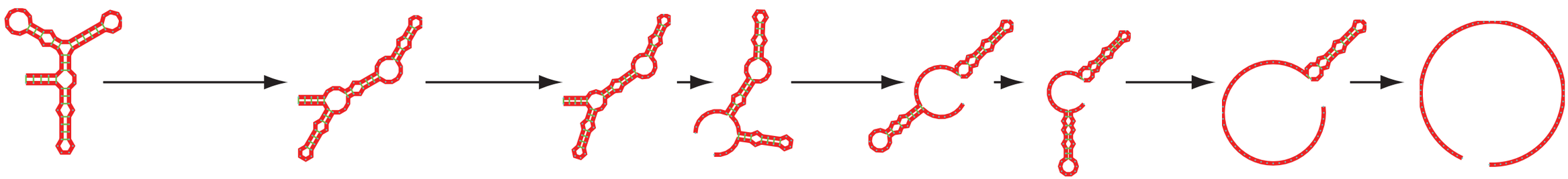
$$C_p = -T\frac{\partial^2 \Delta G}{\partial T^2} \text{ with } \Delta G = -RT \ln Z$$





phenotypic plasticity

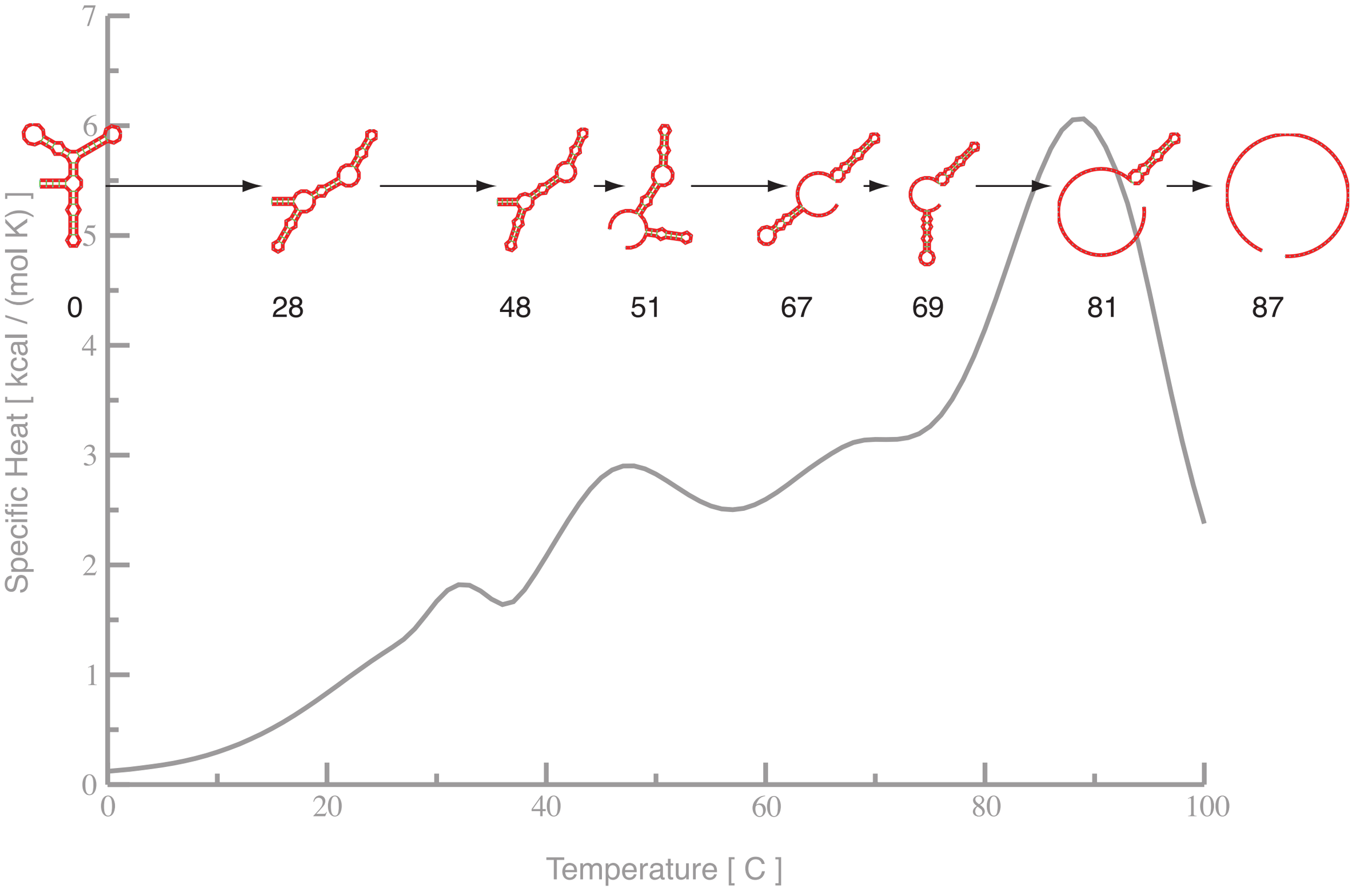
Structural Melting



minimum free energy structure as a function of temperature (C)

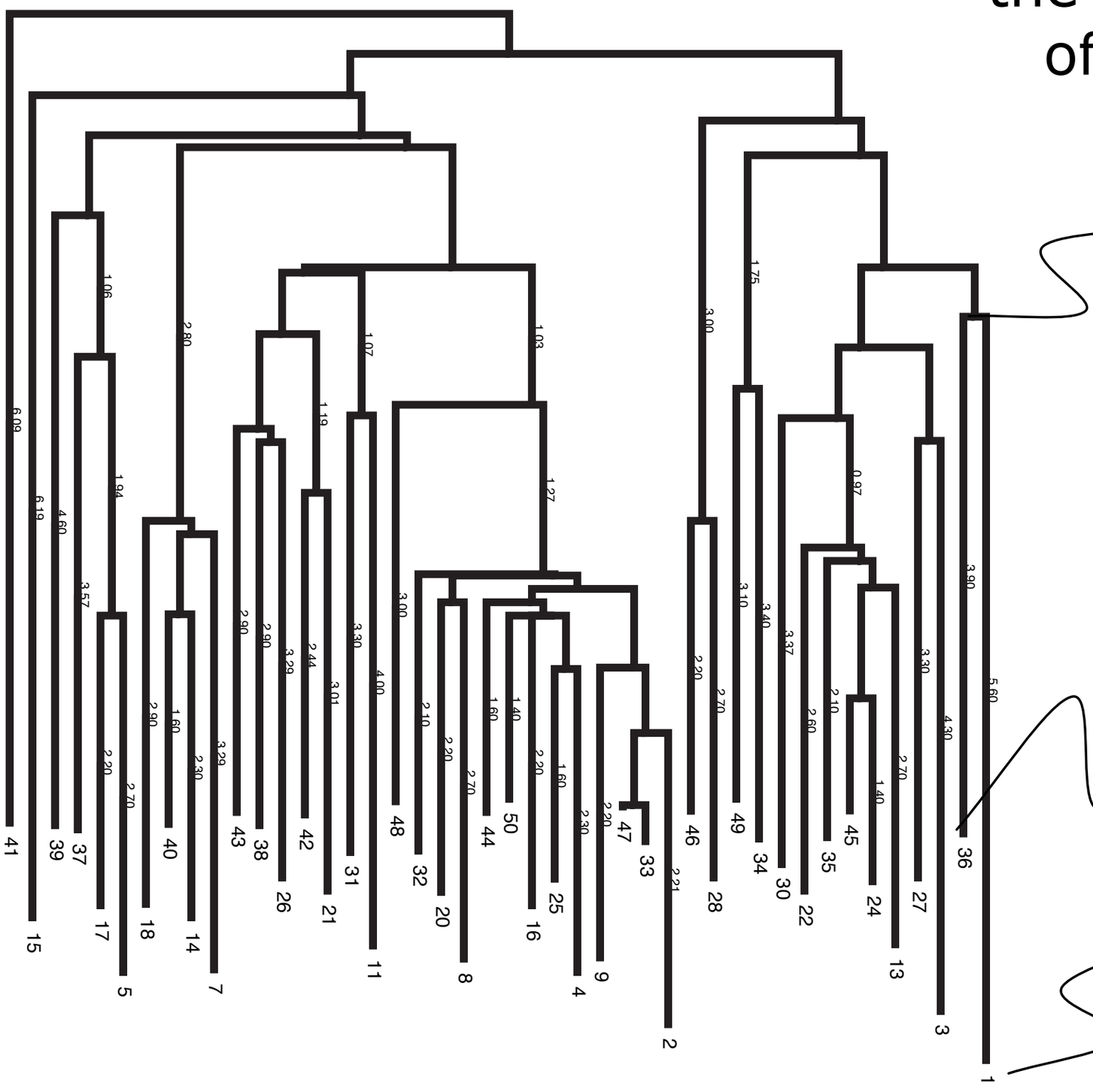
phenotypic plasticity

Heat Capacity & Melting



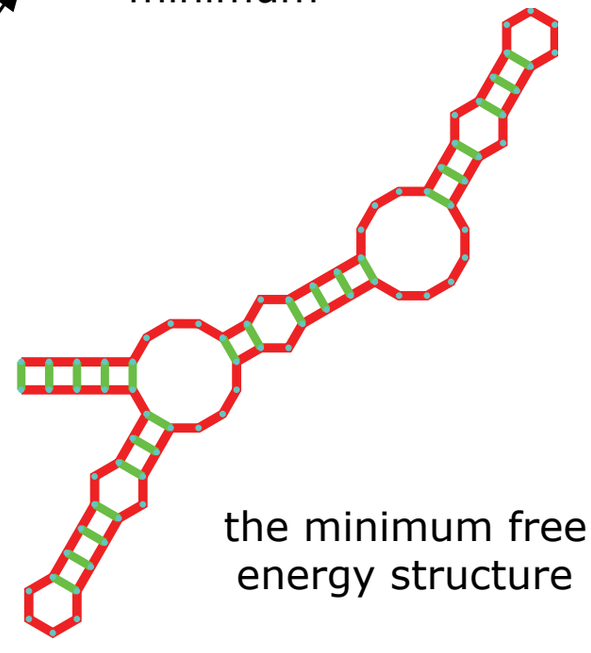
Folding (Process)

the free energy landscape of configuration space



a saddle point: the minimum of the maxima along all paths between two energy “basins”

a local free energy minimum



the minimum free energy structure

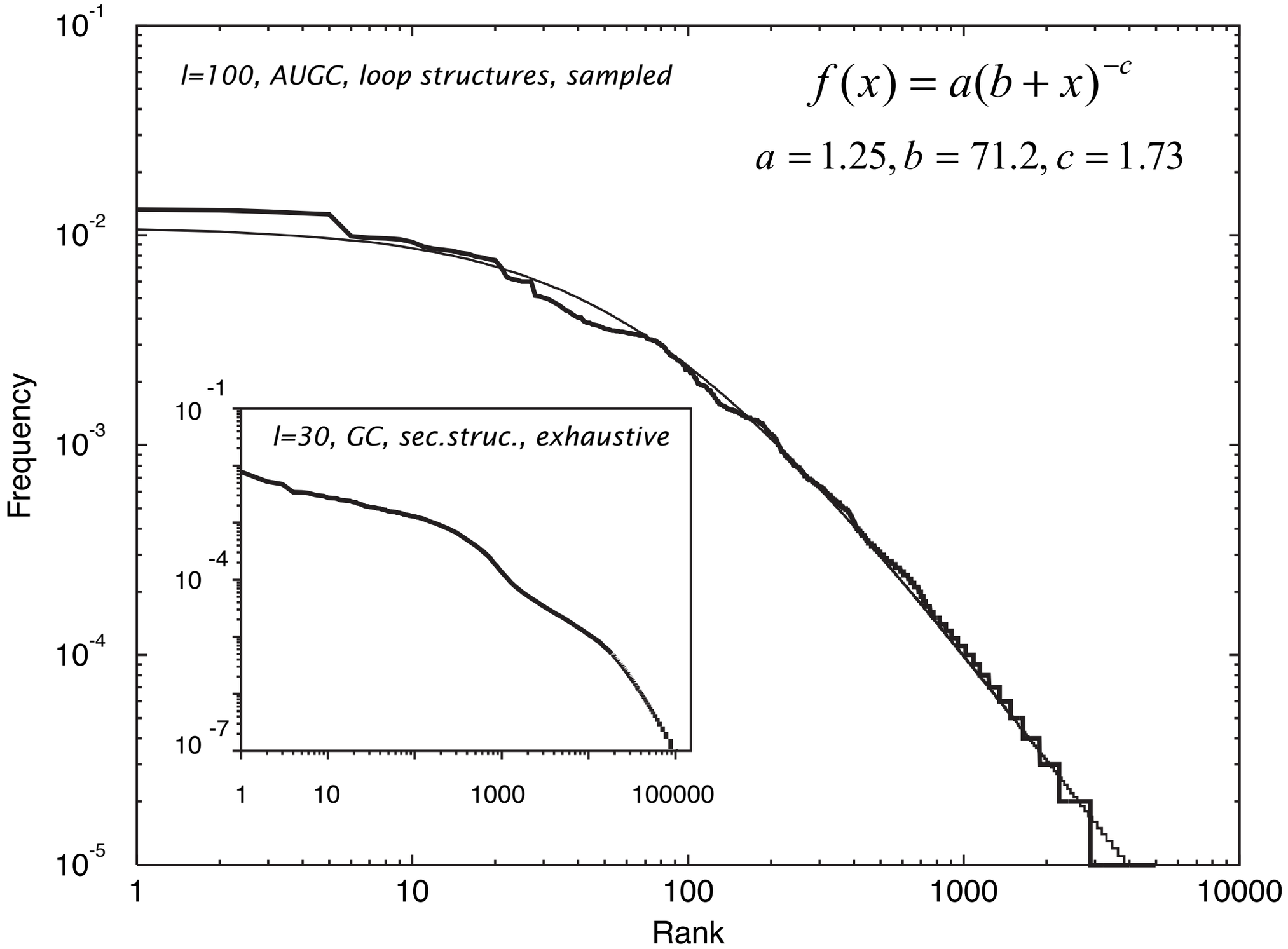
Counting Structures

4^n sequences

fold into

$\sim 2.6^n n^{-3/2}$ (unconstrained!) structures

Ranking Structures



Many-to-one

Example: exhaustive folding of GC-sequences of length 30

- 1 billion sequences
- 220,000 different shapes
- 10% of shapes are realized by more than 5000 sequences
- 93% of all sequences fold into 10% of all shapes

"Typical" Shapes

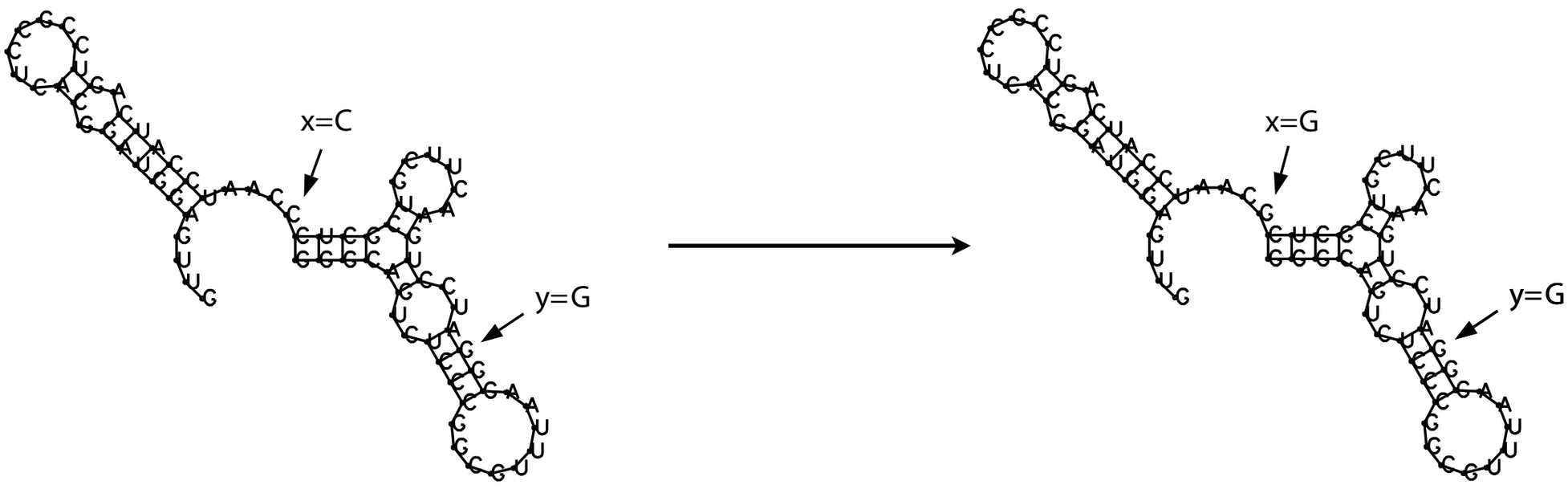
typical shapes Ω

As sequence length tends to infinity,

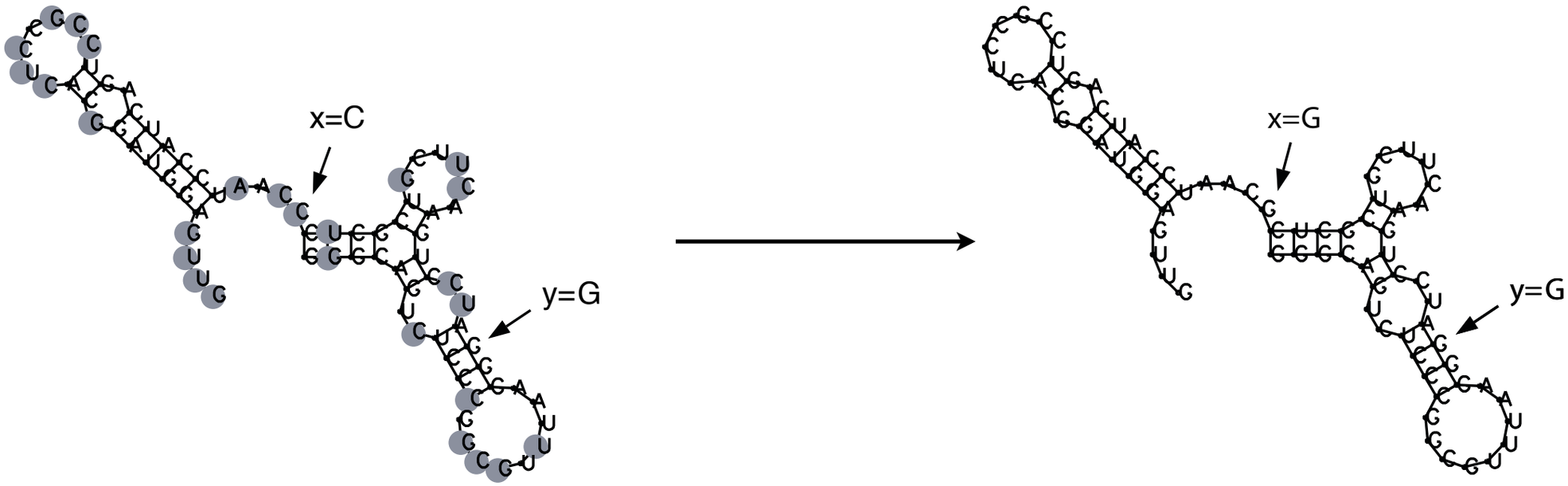
$$\frac{\#\text{ of sequences folding into } \Omega}{\text{all sequences}} \quad \text{tends to} \quad 1$$

$$\frac{\#\text{ of shapes in } \Omega}{\text{all shapes}} \quad \text{tends to} \quad 0$$

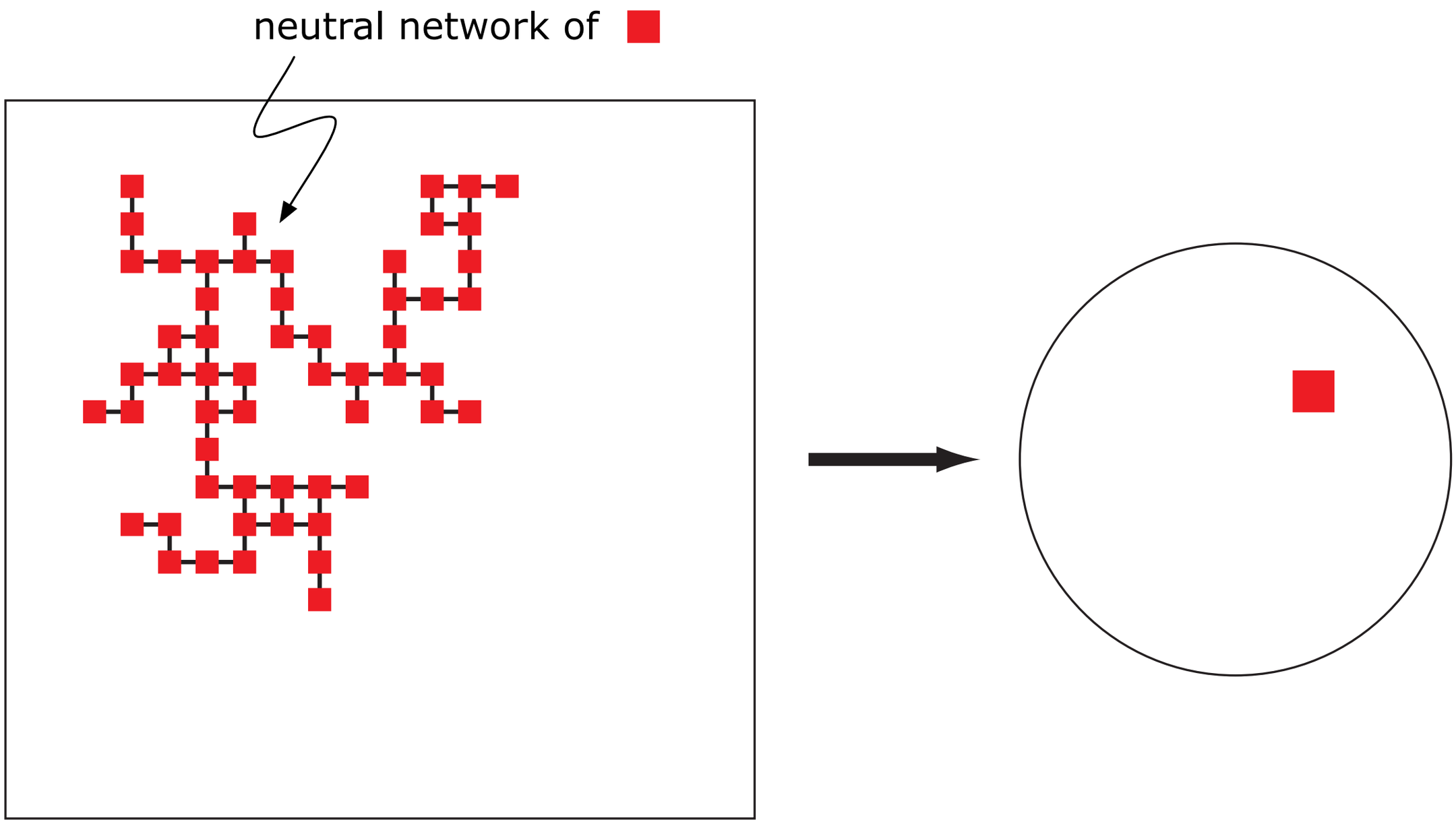
Neutrality



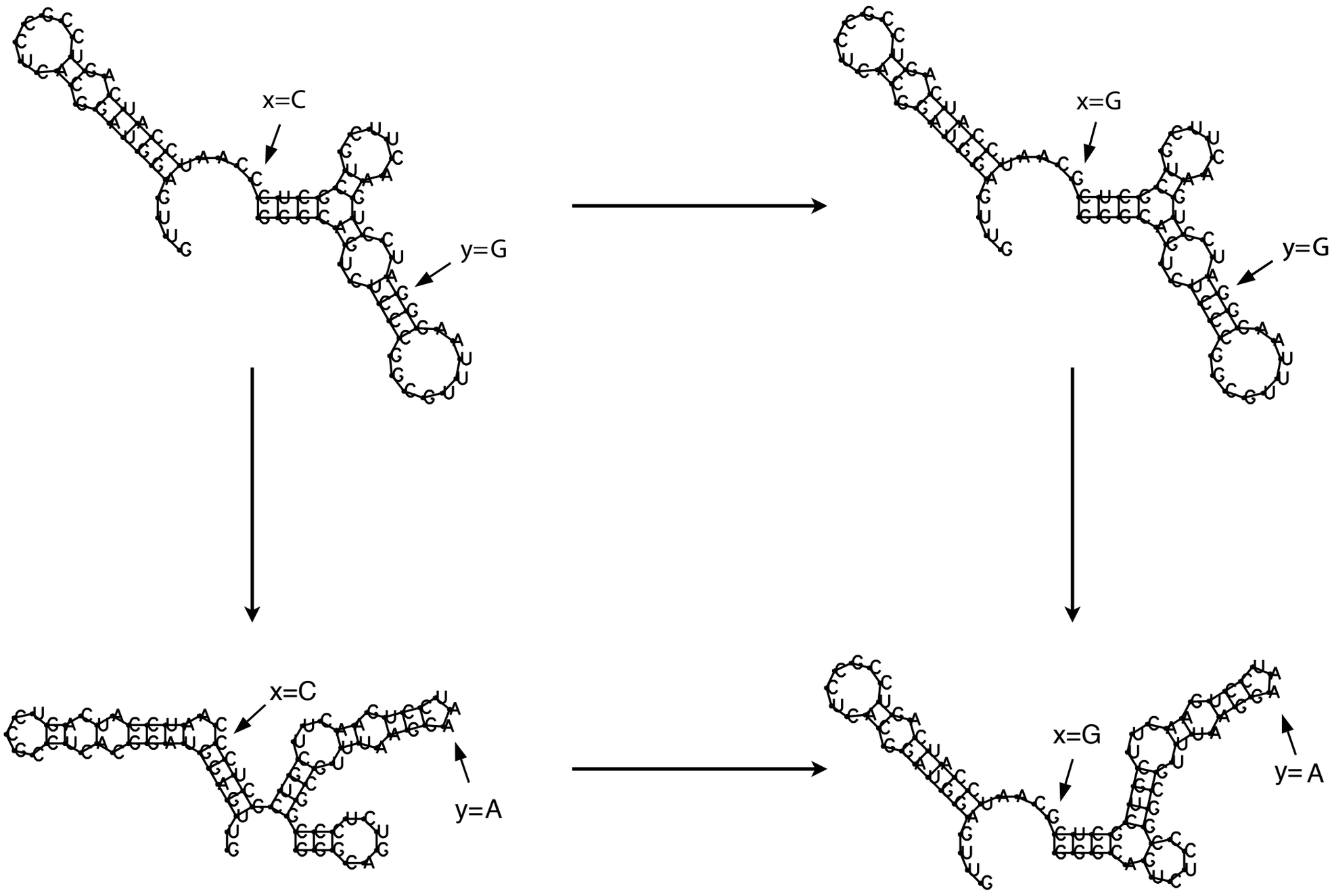
Neutrality



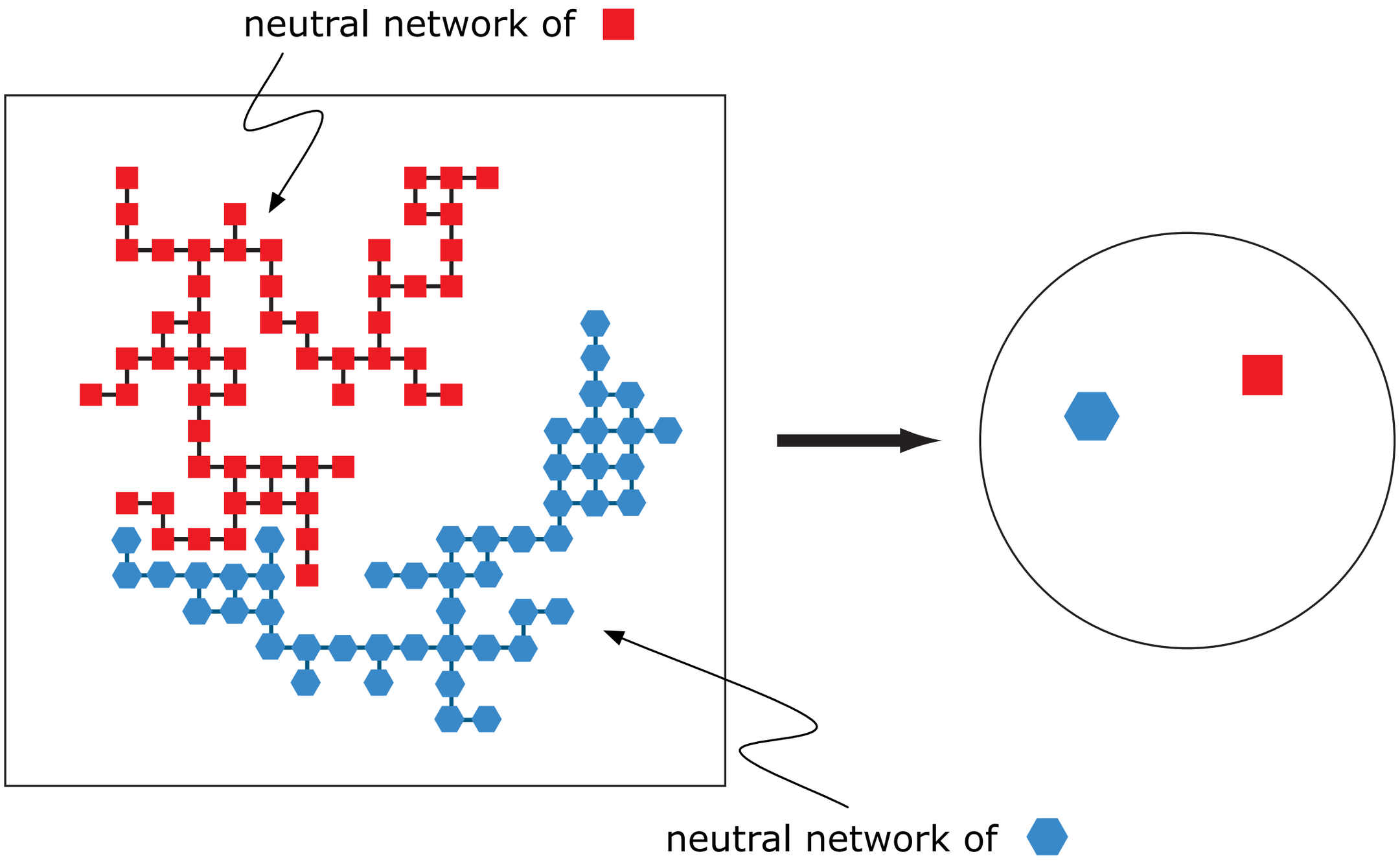
Connected Neutrality



Epistasis



Connected Neutrality



Experimental Neutrality

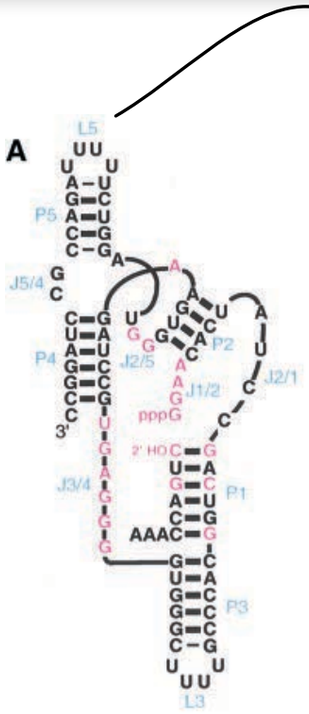
One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds

Erik A. Schultes and David P. Bartel*

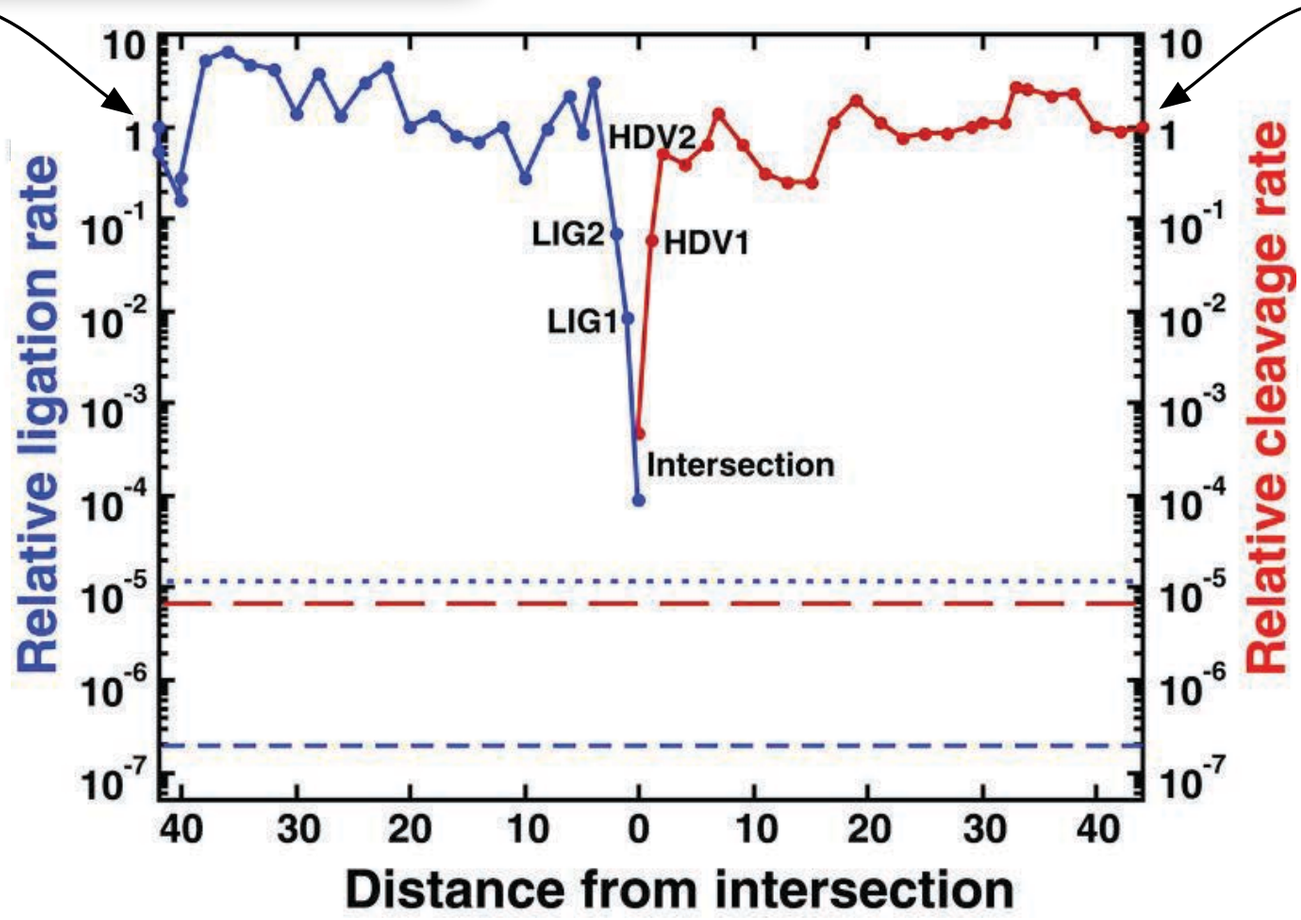
We describe a single RNA sequence that can assume either of two ribozyme folds and catalyze the two respective reactions. The two ribozyme folds share no evolutionary history and are completely different, with no base pairs (and probably no hydrogen bonds) in common. Minor variants of this sequence are highly active for one or the other reaction, and can be accessed from prototype ribozymes through a series of neutral mutations. Thus, in the course of evolution, new RNA folds could arise from preexisting folds, without the need to carry inactive intermediate sequences. This raises the possibility that biological RNAs having no structural or functional similarity might share a common ancestry. Furthermore, functional and structural divergence might, in some cases, precede rather than follow gene duplication.

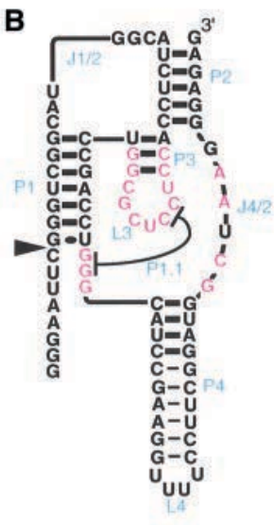
21 JULY 2000 VOL 289 SCIENCE www.sciencemag.org

the Schultes/Bartel experiment



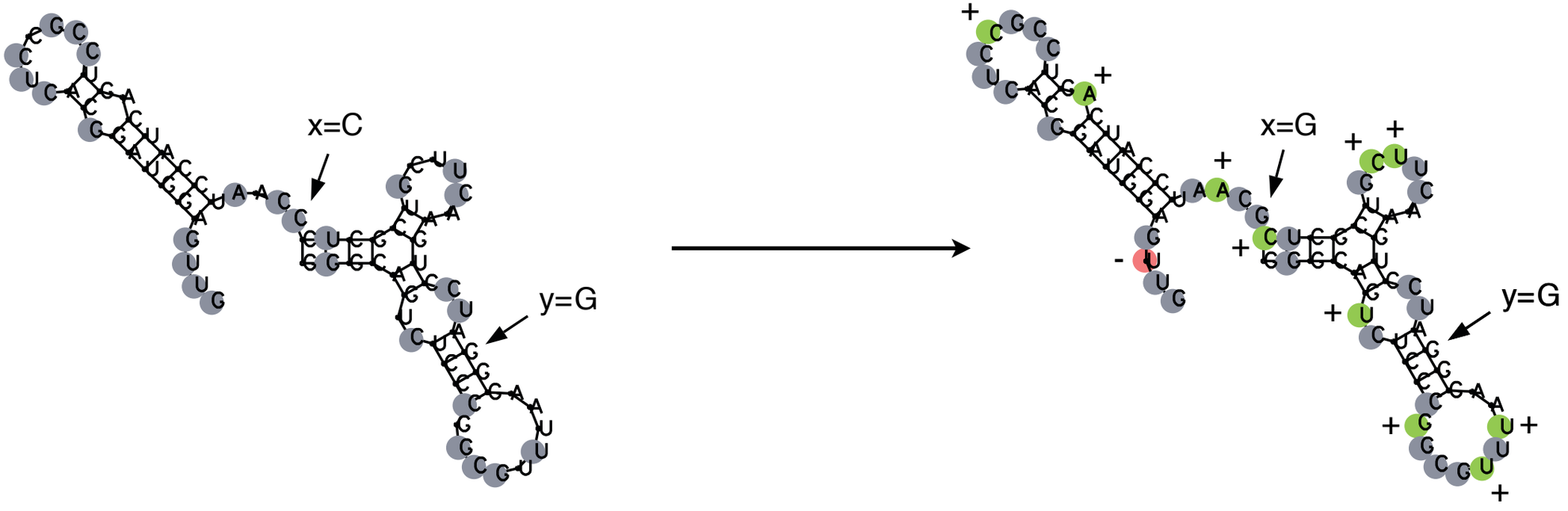
in vitro evolved ribozyme with ligation activity



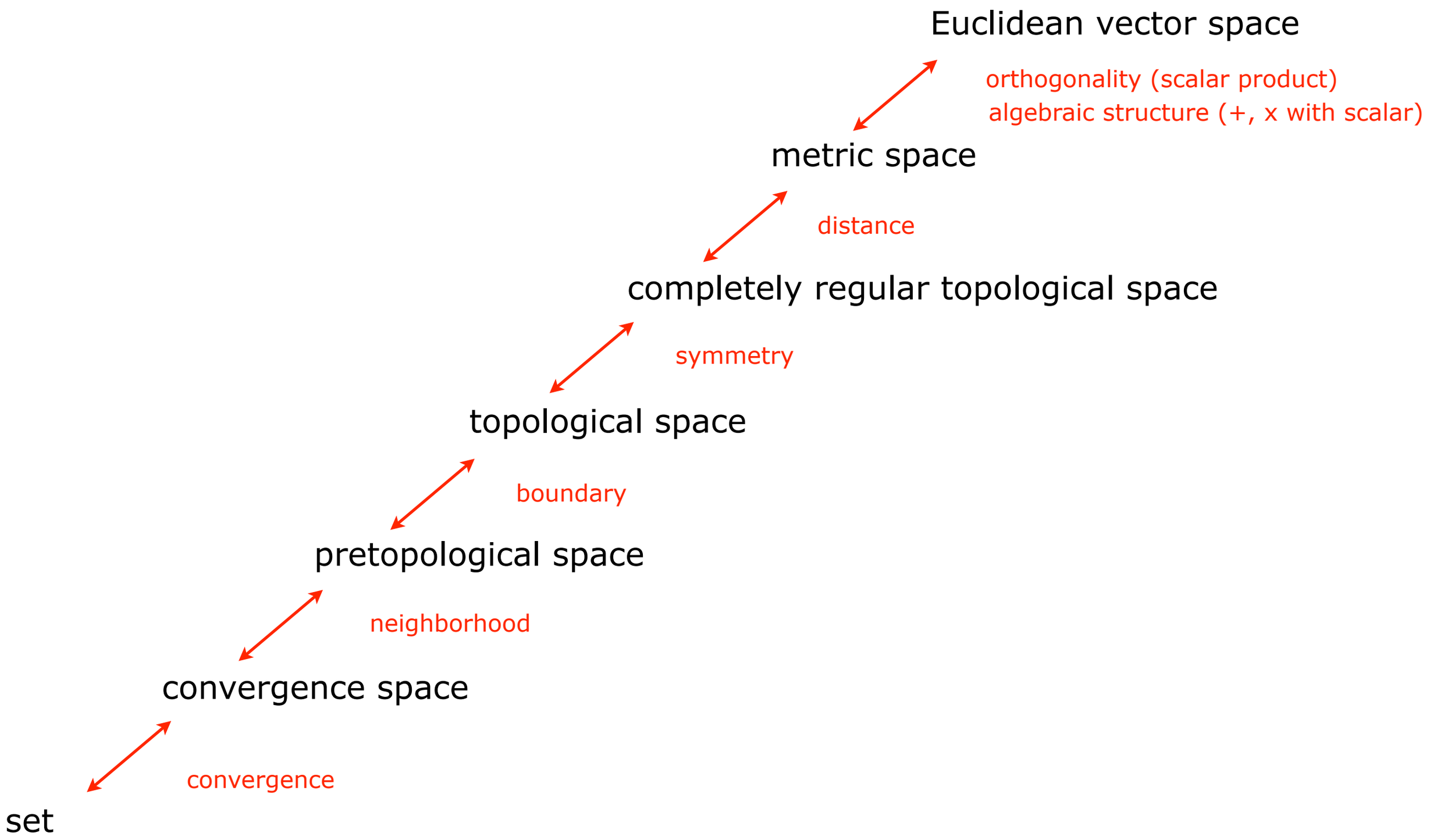


naturally evolved HDV ribozyme

Genetic Control of Neutrality



Space

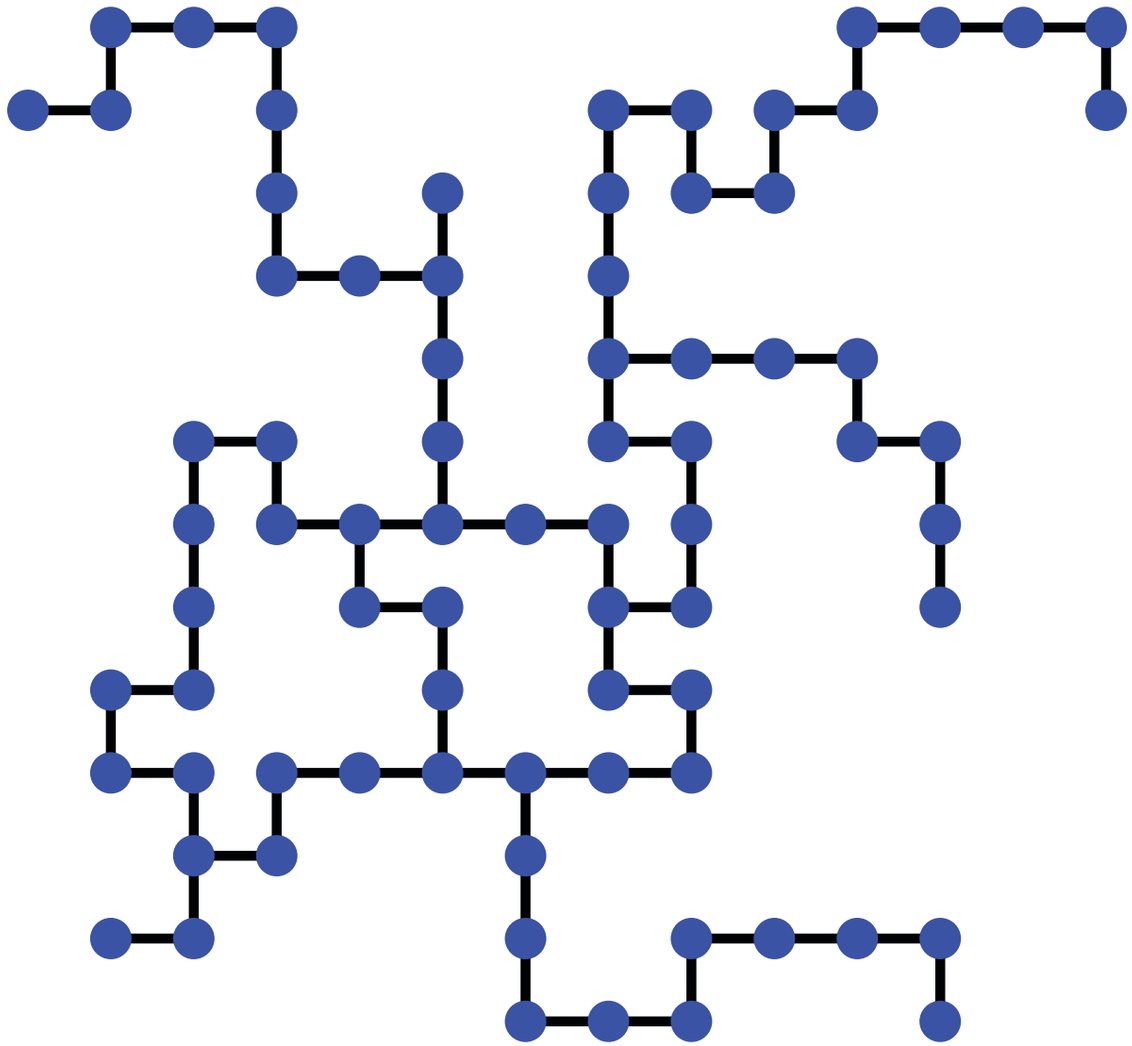

Euclidean vector space
orthogonality (scalar product)
algebraic structure (+, x with scalar)
metric space
distance
completely regular topological space
symmetry
topological space
boundary
pretopological space
neighborhood
convergence space
convergence
set

The construction of phenotype space

Construction by similarity measure on phenotypes:
independent of the genotype/phenotype map

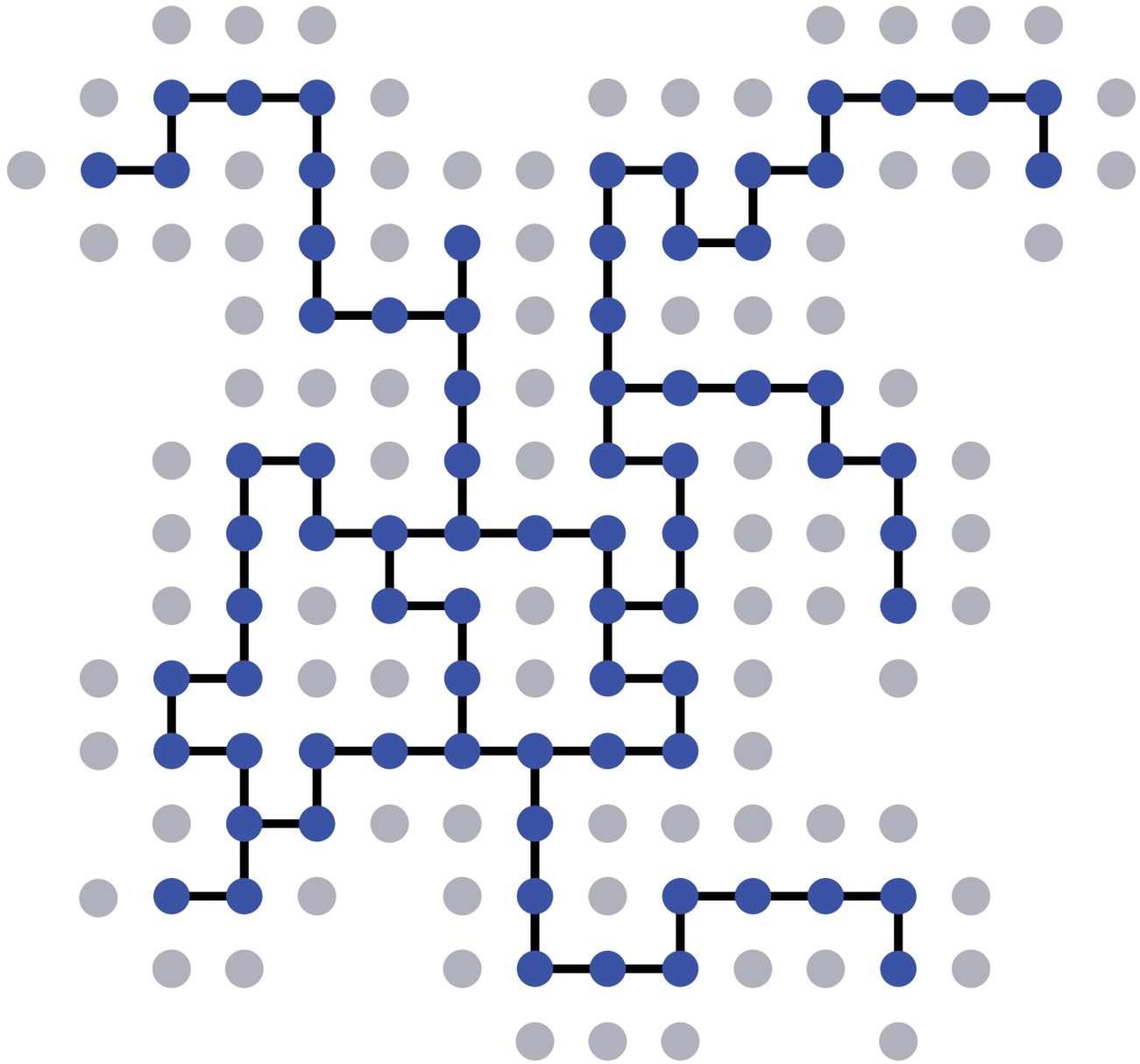
Construction by “accessibility”:
induced by the genotype/phenotype map

A Neutral Network



- $S(\alpha)$

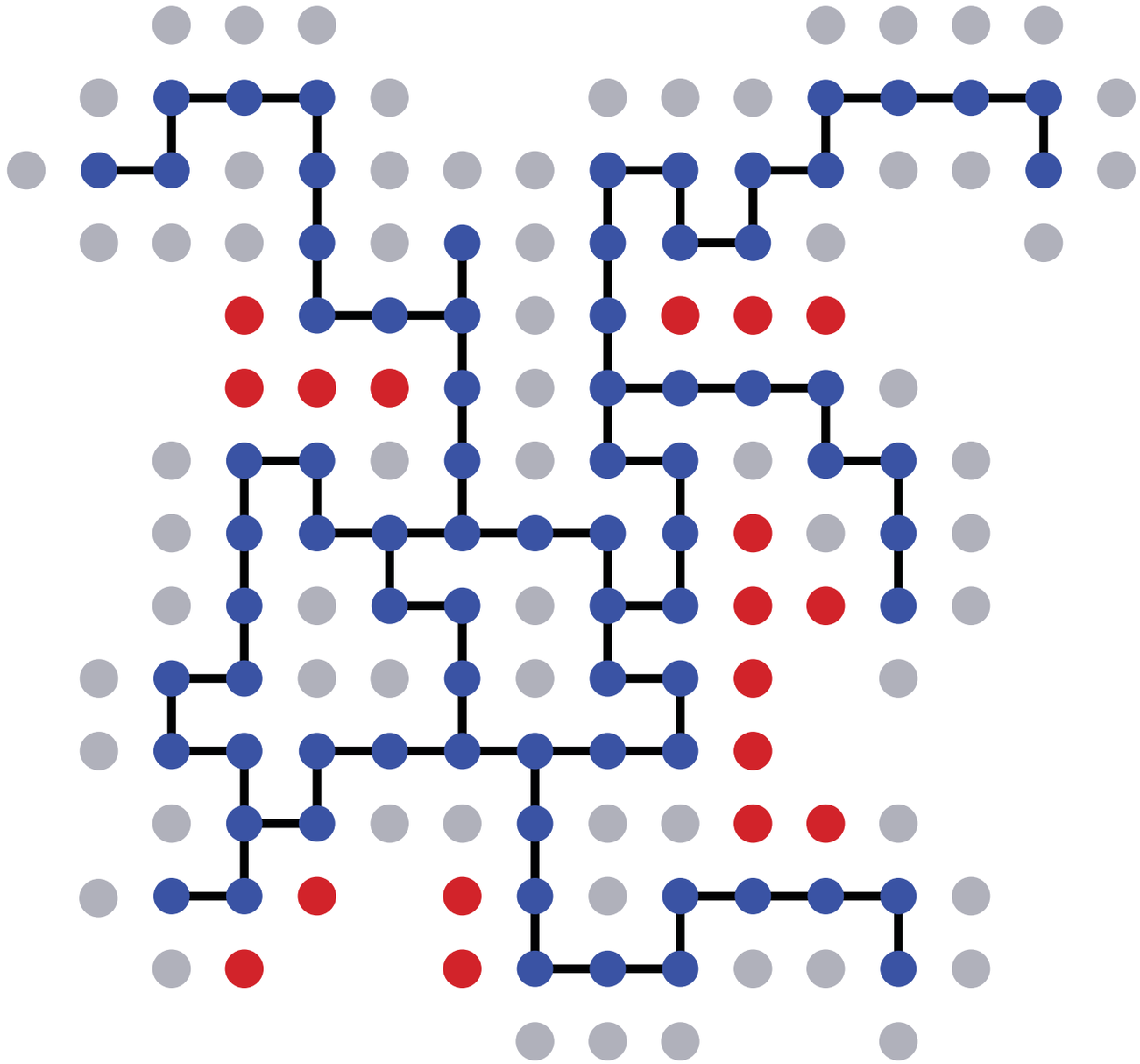
The Boundary of a Neutral Network



$S(\alpha)$

$\partial S(\alpha)$

Accessibility



$S(\alpha)$

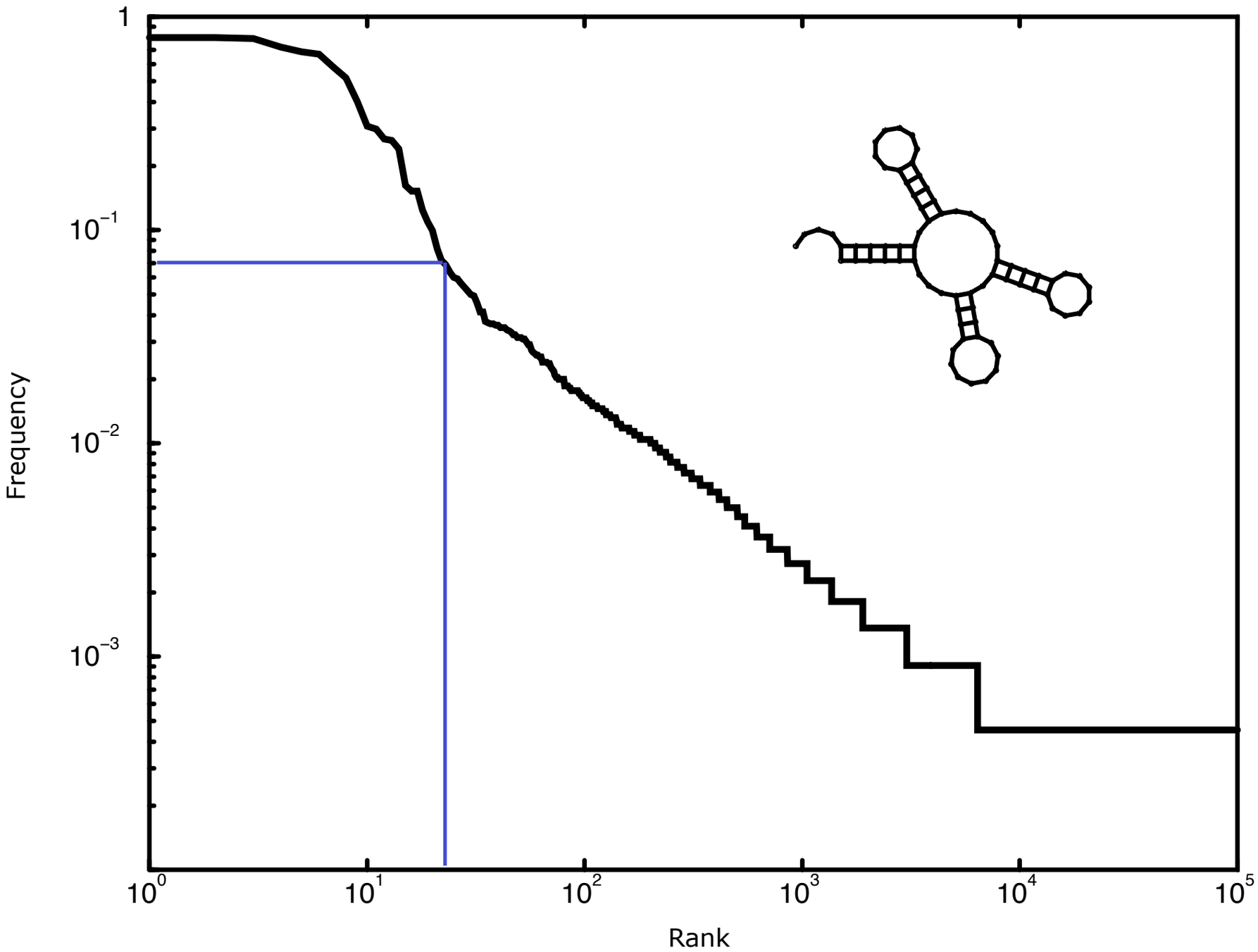
$\partial S(\alpha)$

$S(\beta) \cap \partial S(\alpha)$

$$\mathcal{A}(\beta \leftarrow \alpha) = \frac{|S(\beta) \cap \partial S(\alpha)|}{|\partial S(\alpha)|}$$

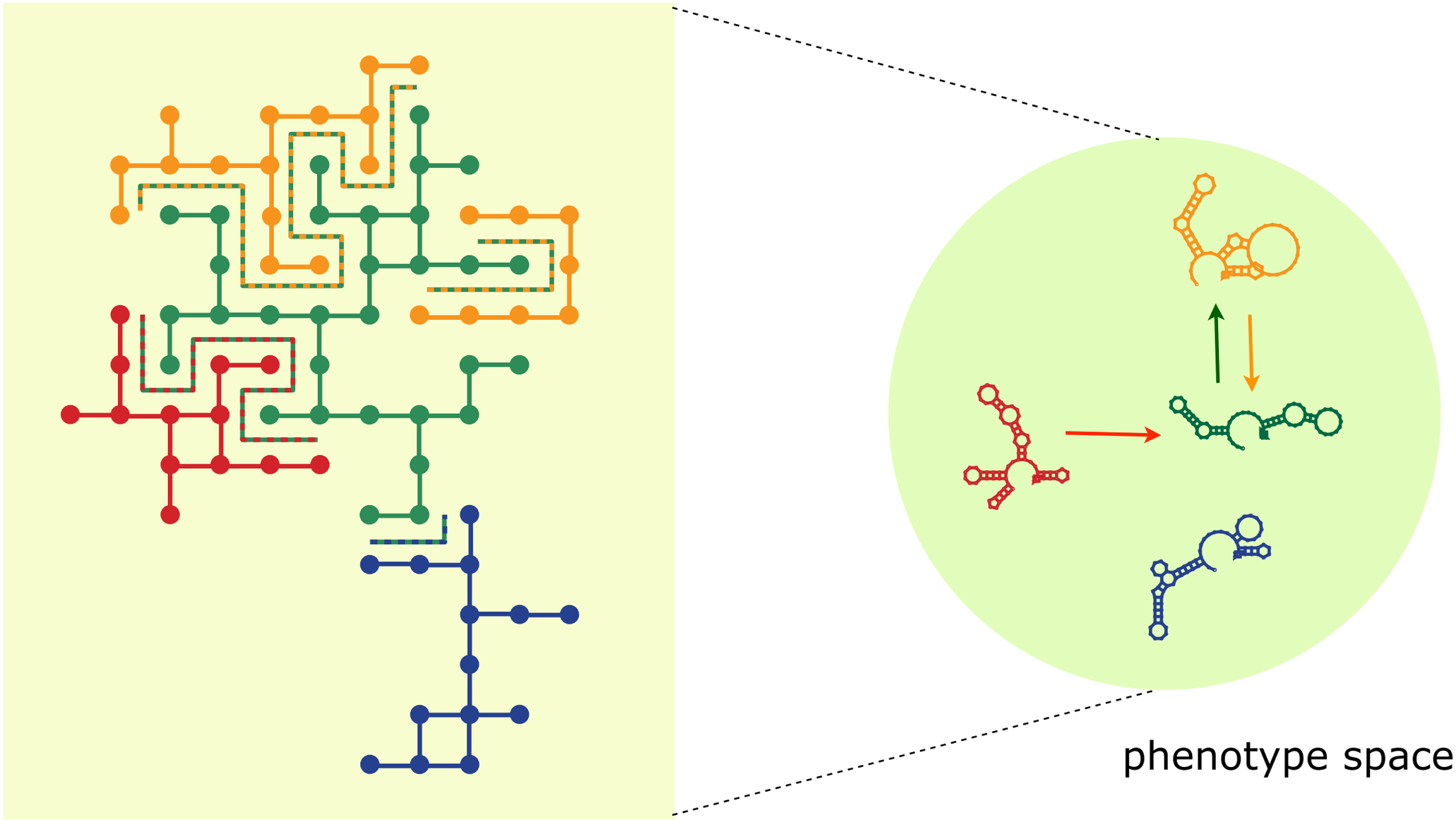
accessibility

Distribution of Structures in the Boundary



shapes near the tRNA shape

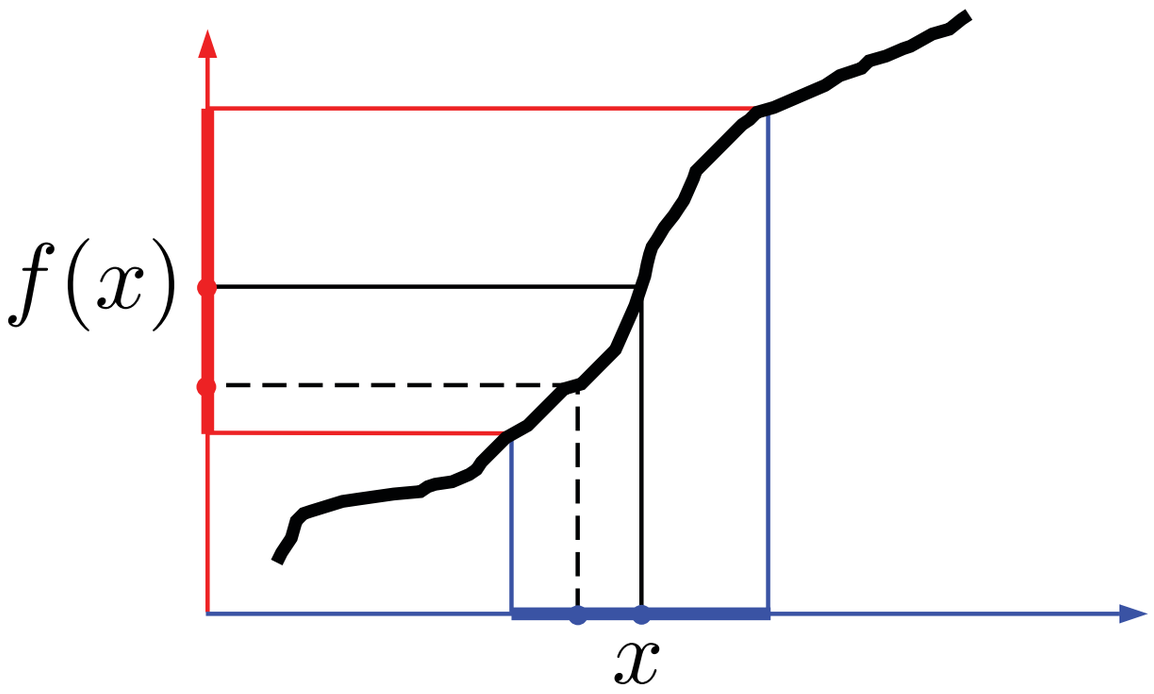
Nearness via Accessibility

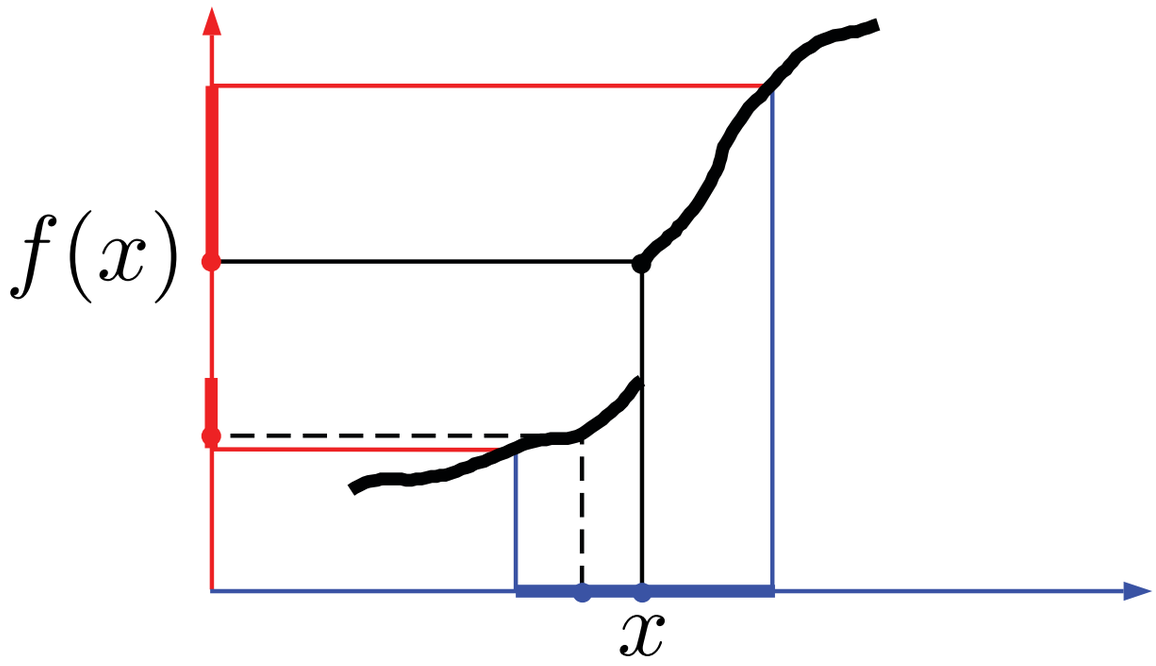


Boundary Topology

Pennsylvania is near New Jersey, but New Jersey is not near Pennsylvania

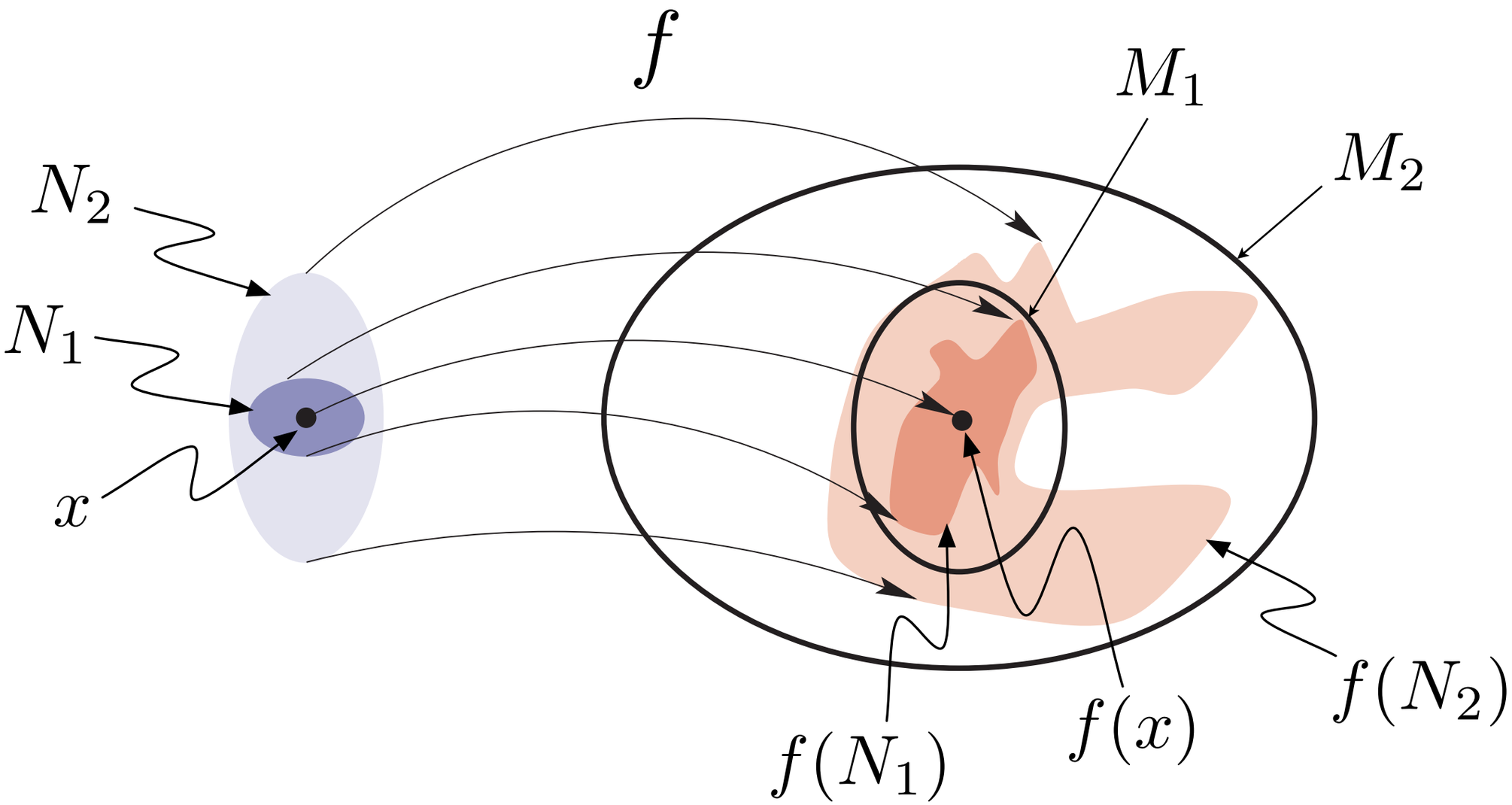
Continuity via Distance





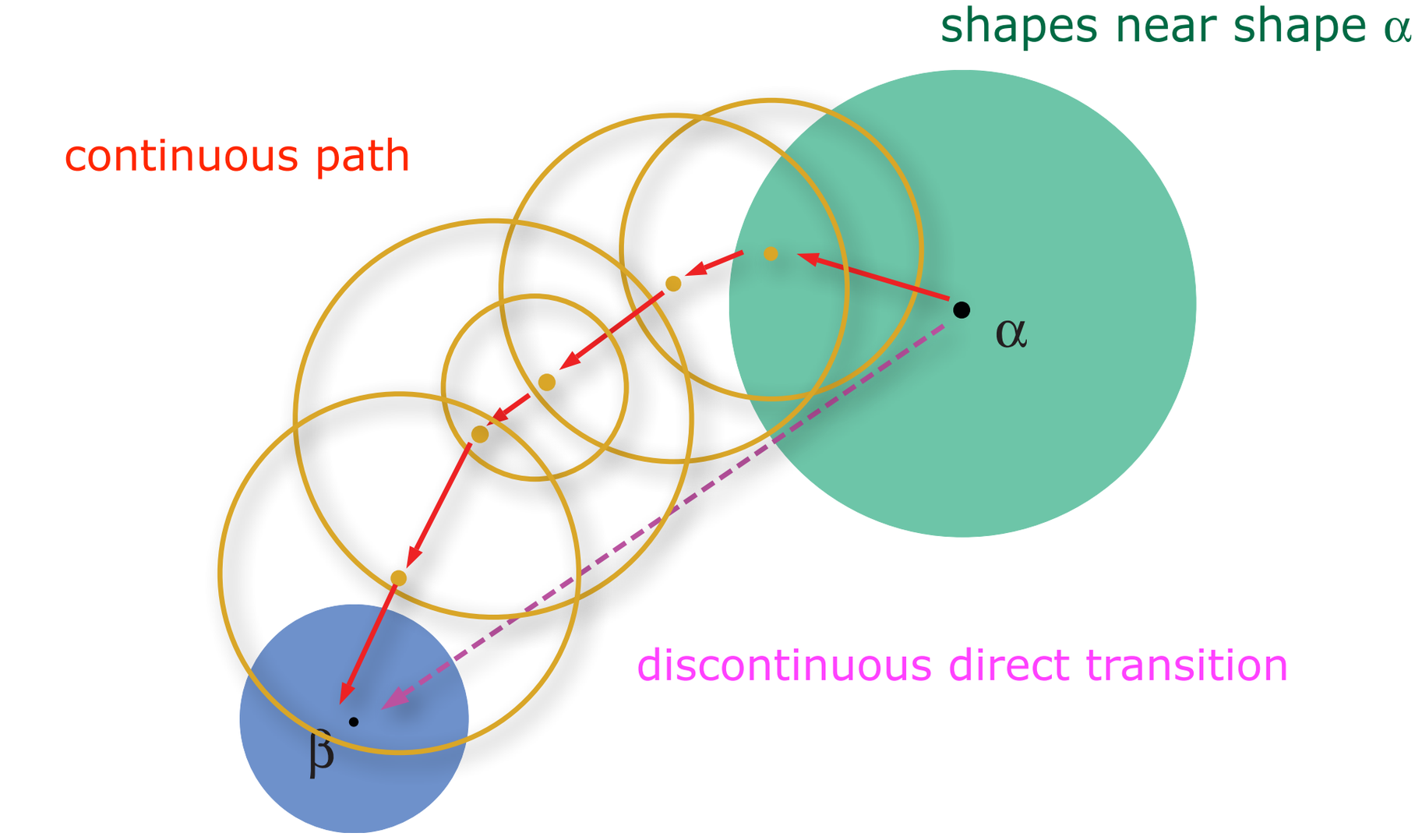
A function is continuous if you don’t have to lift the pencil when drawing it.

Continuity via Neighborhood



For each neighborhood M of f(x), there is a neighborhood N of x, such that f(N) is contained in M.

Continuity via Neighborhood



Continuity via Neighborhood

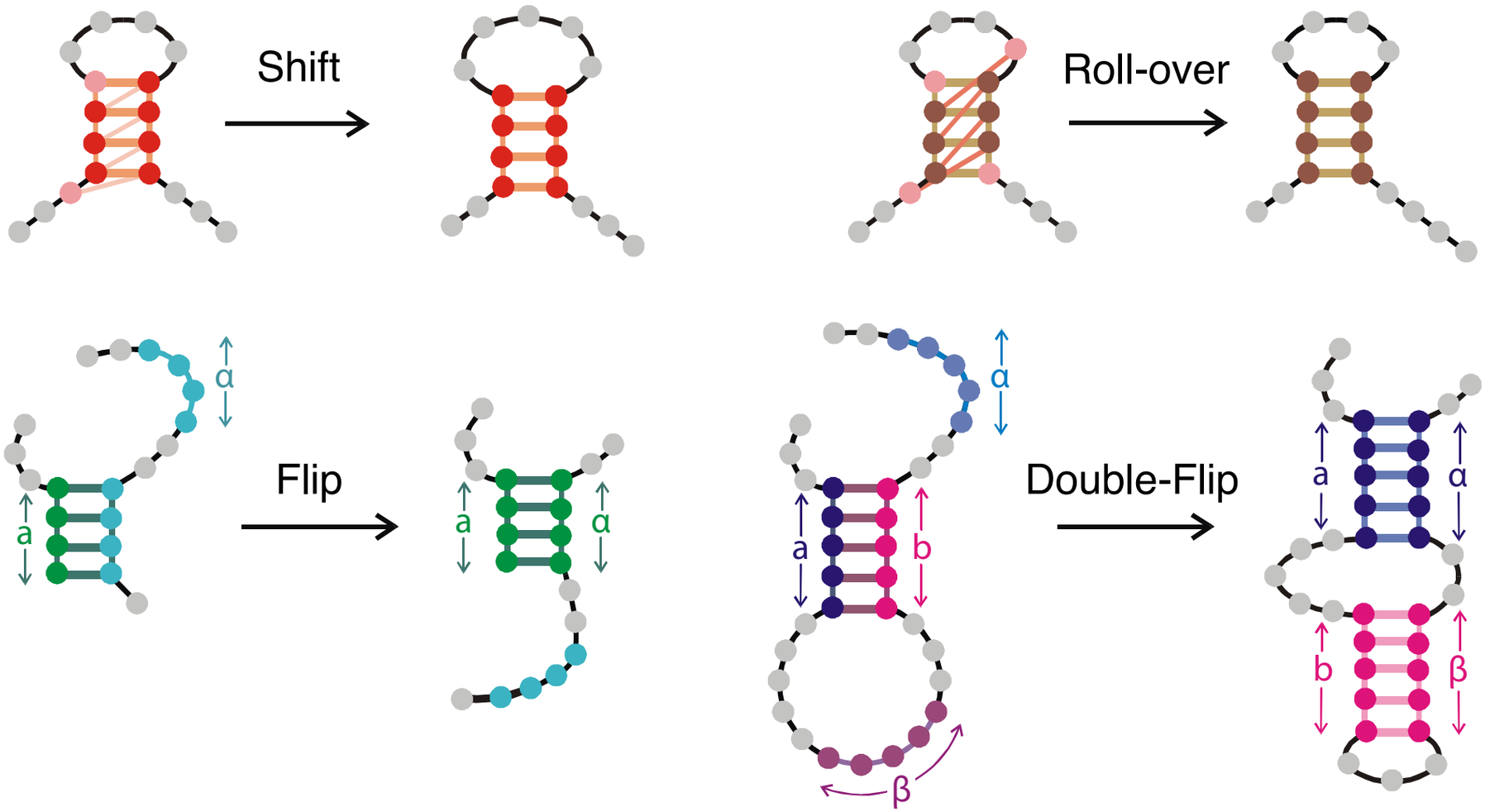
Can any two shapes be connected by a continuous path?

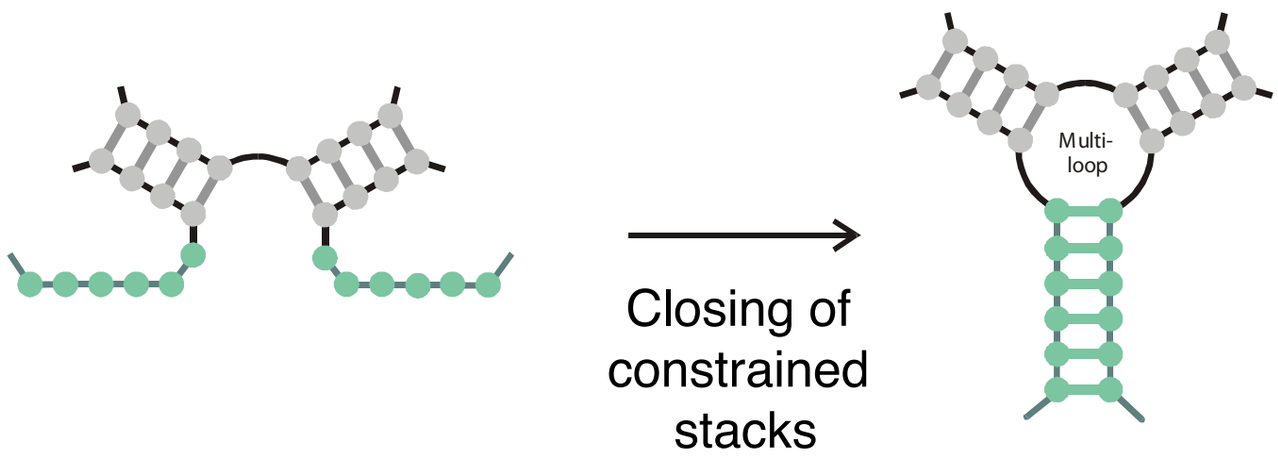
NO.

Certain shape transformations require, along any path, a constricted (*) passage from one neutral network to another.

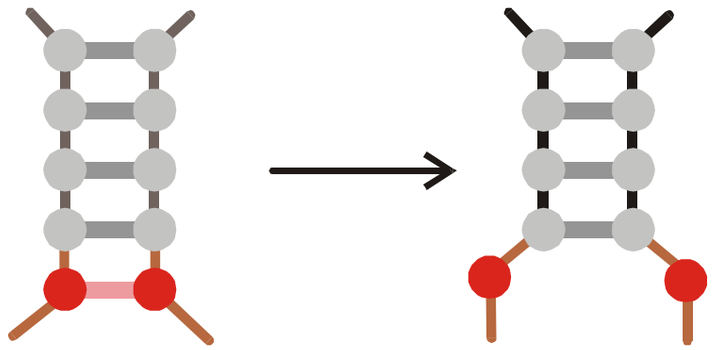
(*) small fraction of shared boundary in the direction of the passage

Discontinuous Transformations

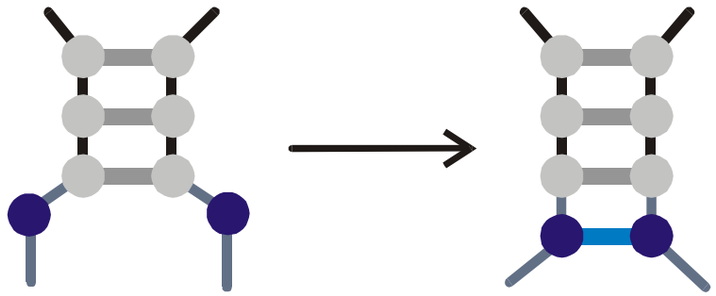




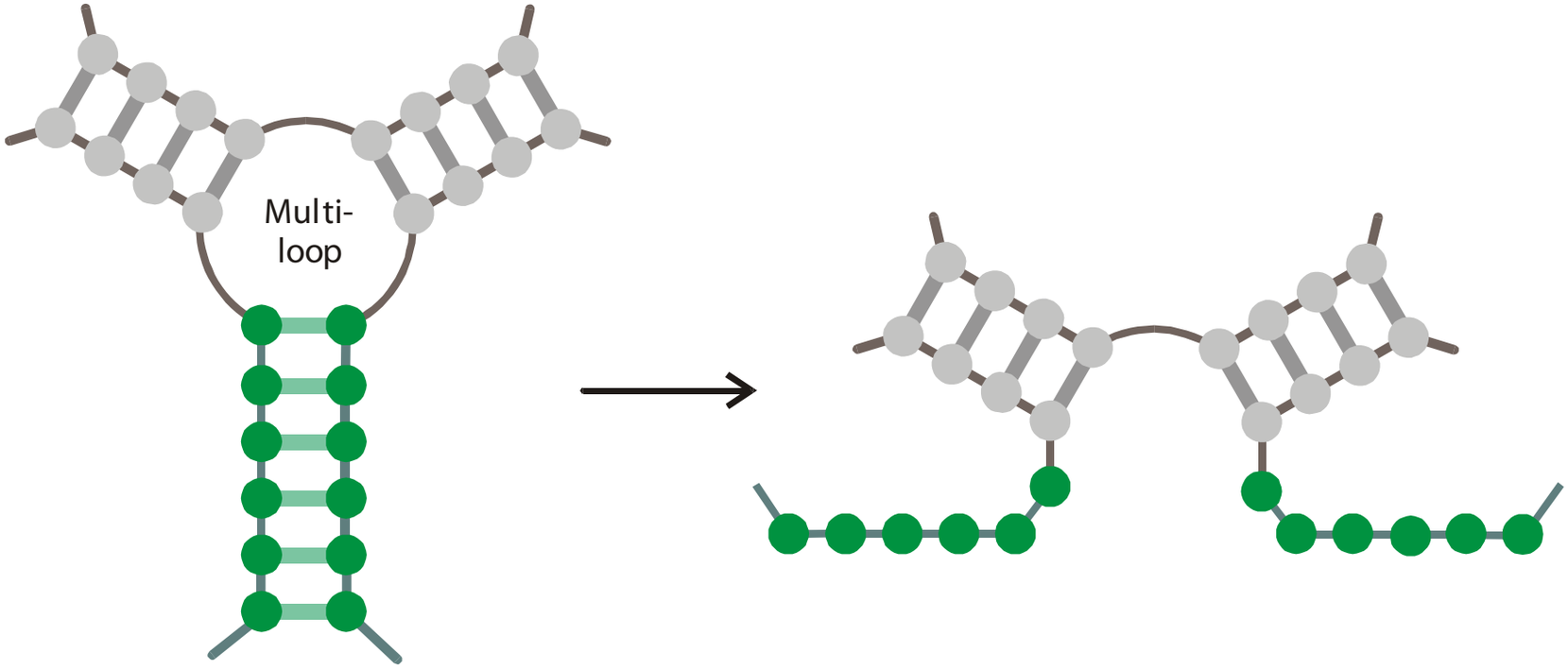
Continuous Transformations



Stack shortening

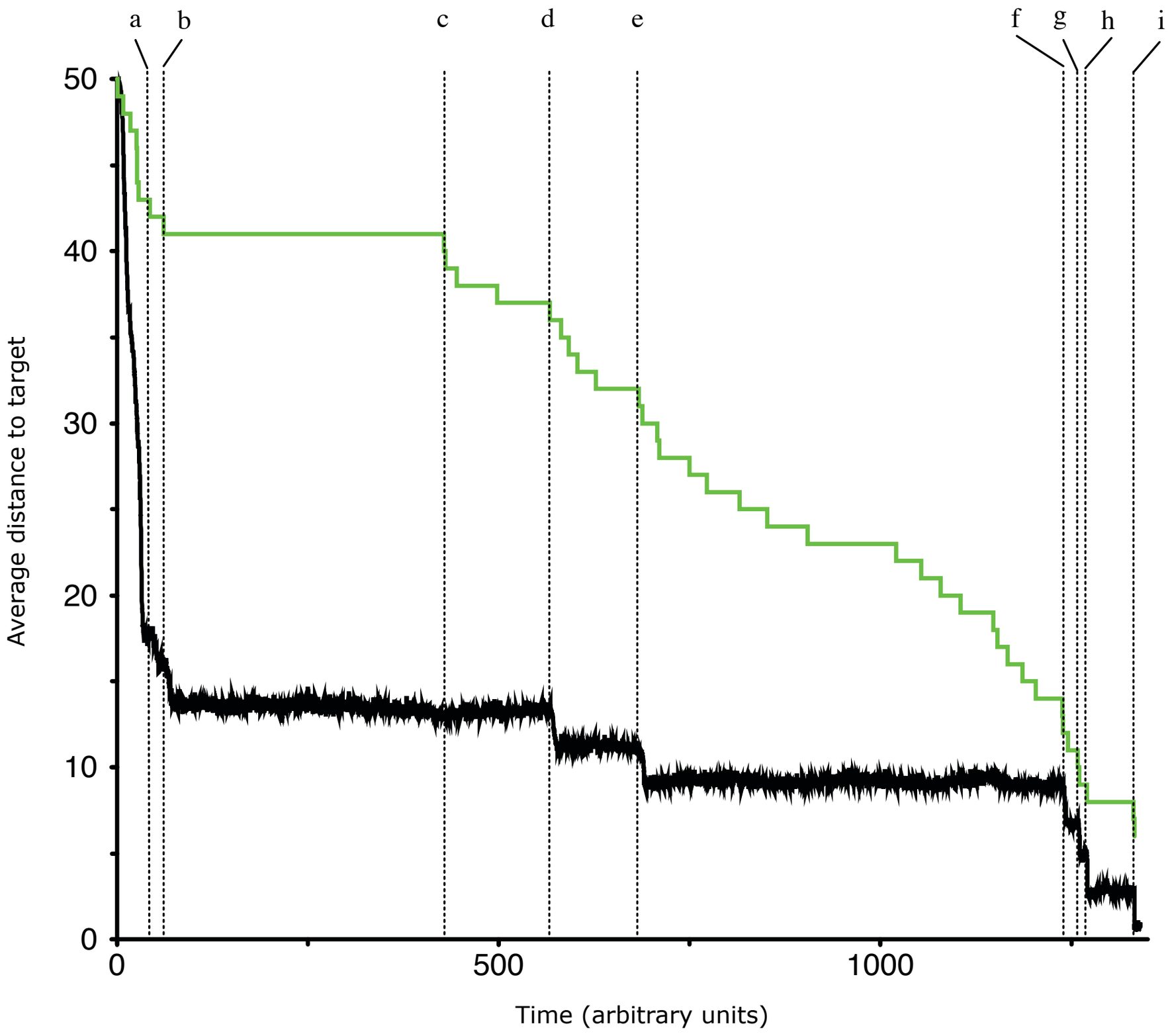


Stack elongation

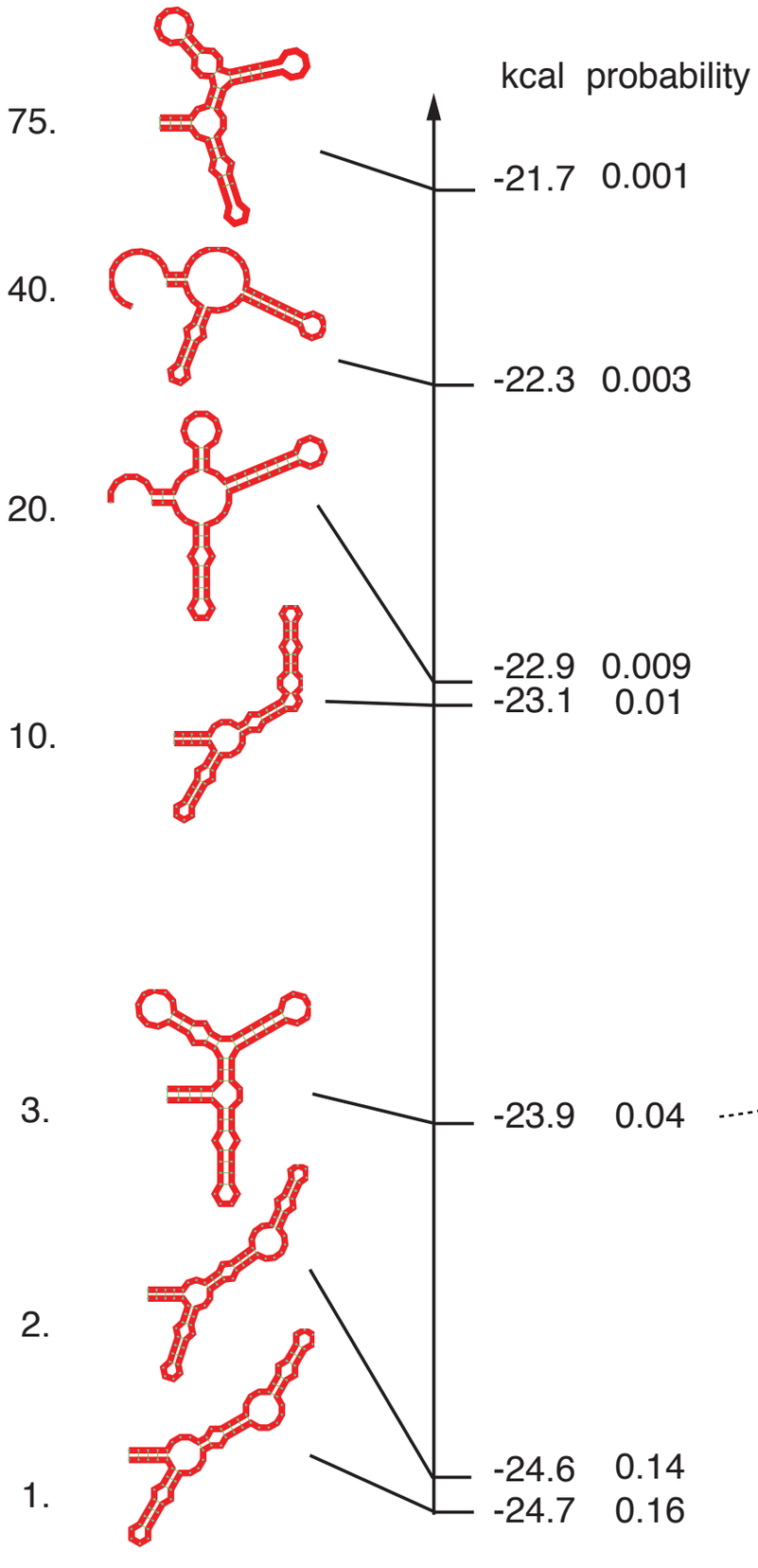


Unzipping of constrained stacks

Punctuation



Suboptimal Structures



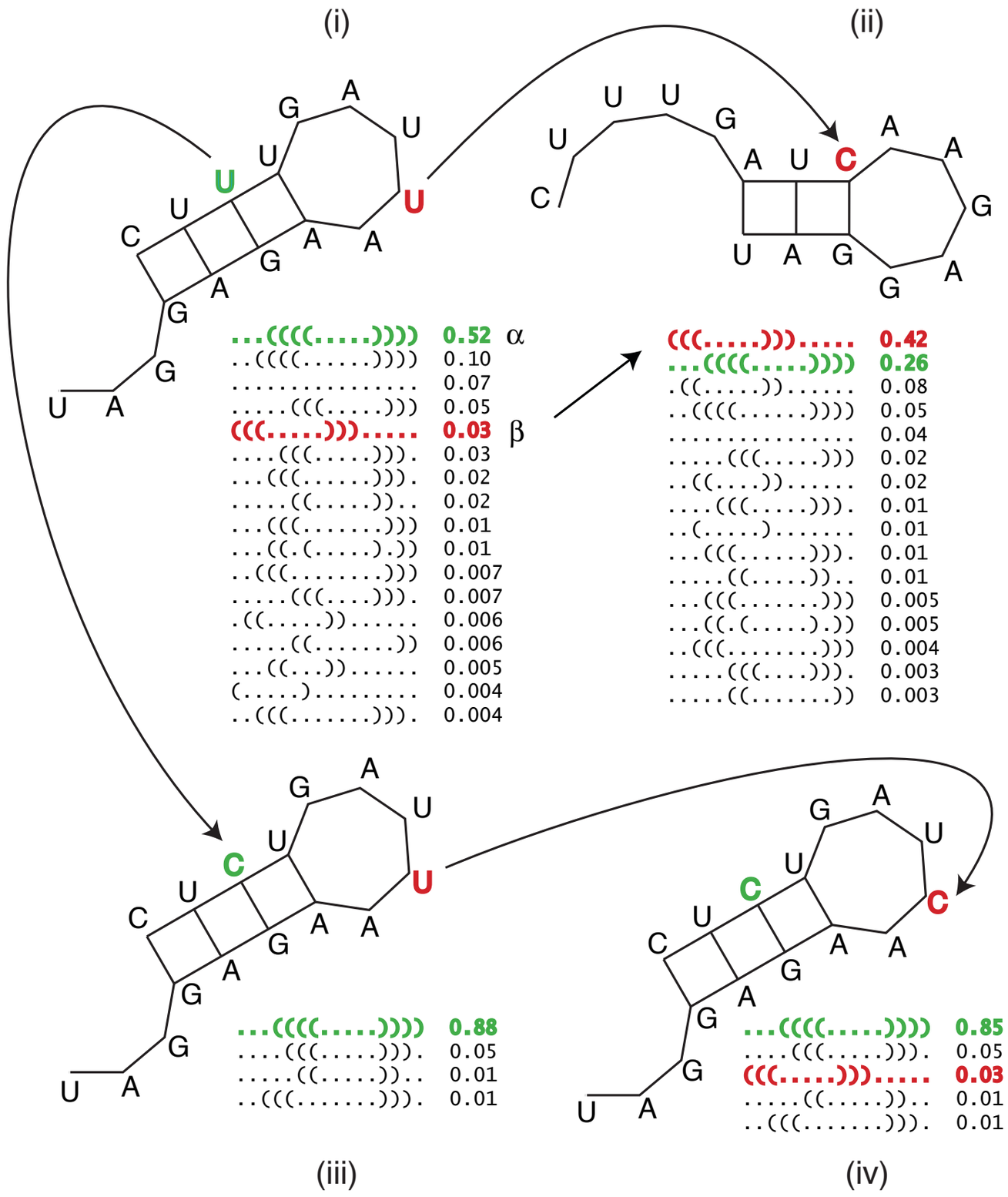
phenotypic plasticity
("microenvironmental")

$$P(S_j) = \frac{e^{\Delta G_{S_j}/kT}}{\sum_i e^{\Delta G_{S_i}/kT}}$$

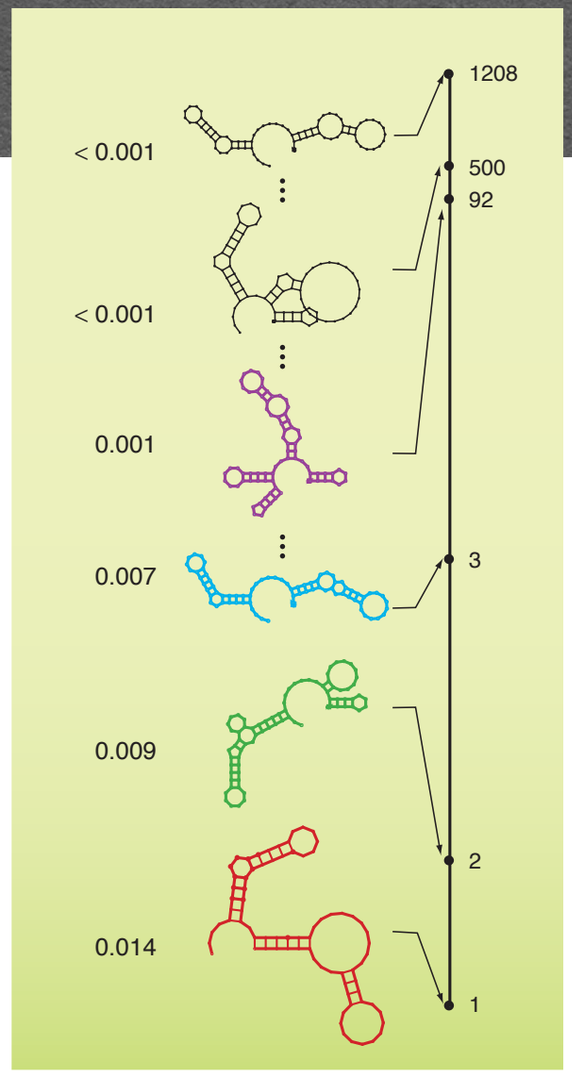
partition function Z

the probability of a structure

Suboptimal Structures And Mutations



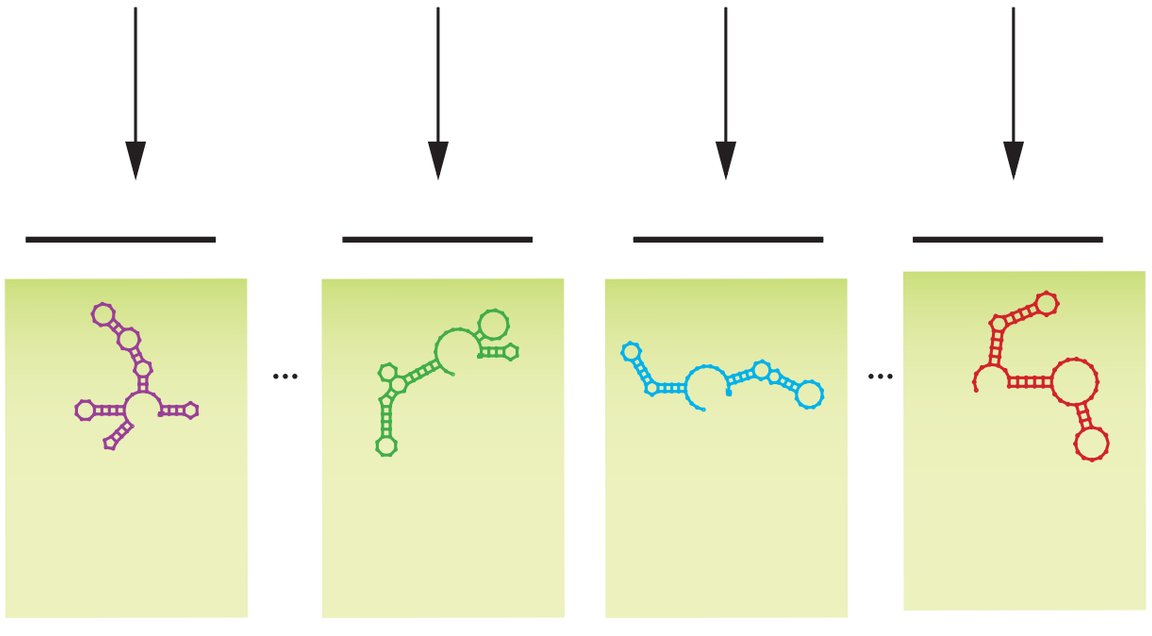
Plasticity Mirrors Variability



alternative phenotypes
at constant genotype
(plasticity)

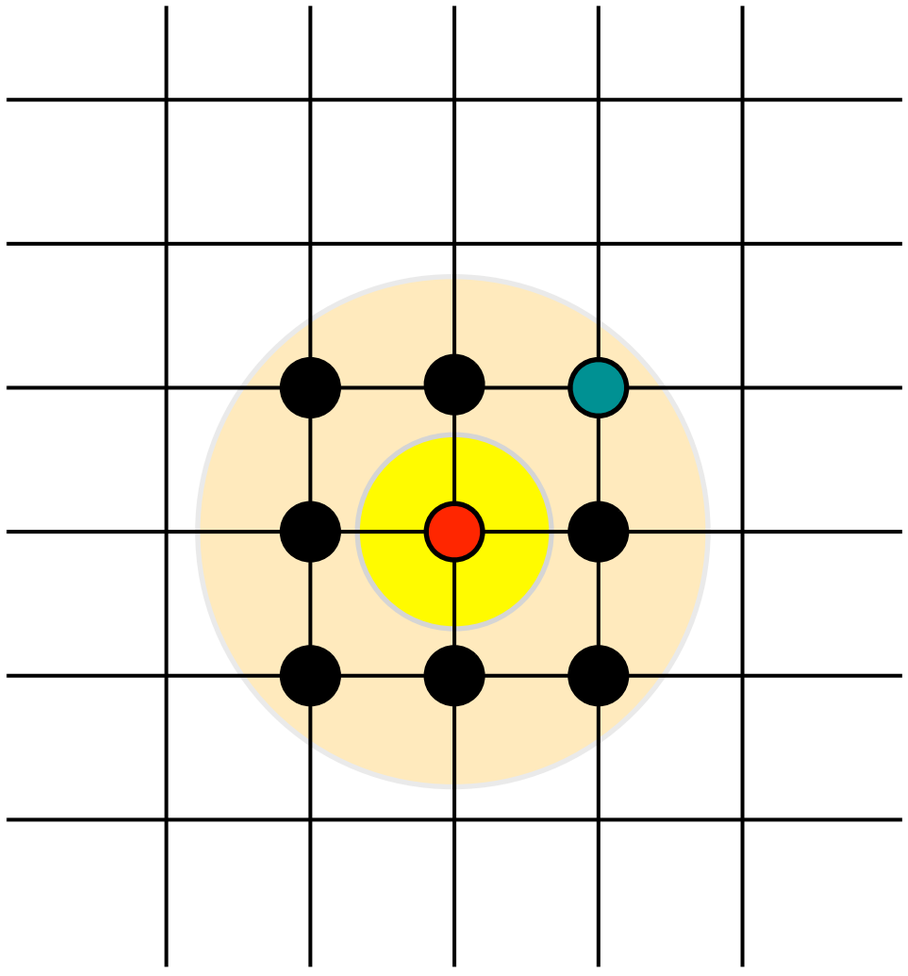
GCUGUUAUCGGCGCUCCGUACUACGCUUAAAAAACAGGACAGUUGGGAUACUUGCAAAACCAGGUUCAUCUUGUGA

min. free energy structure
of reference sequence



min. free energy structures
of neighbors of reference sequence
(variability)

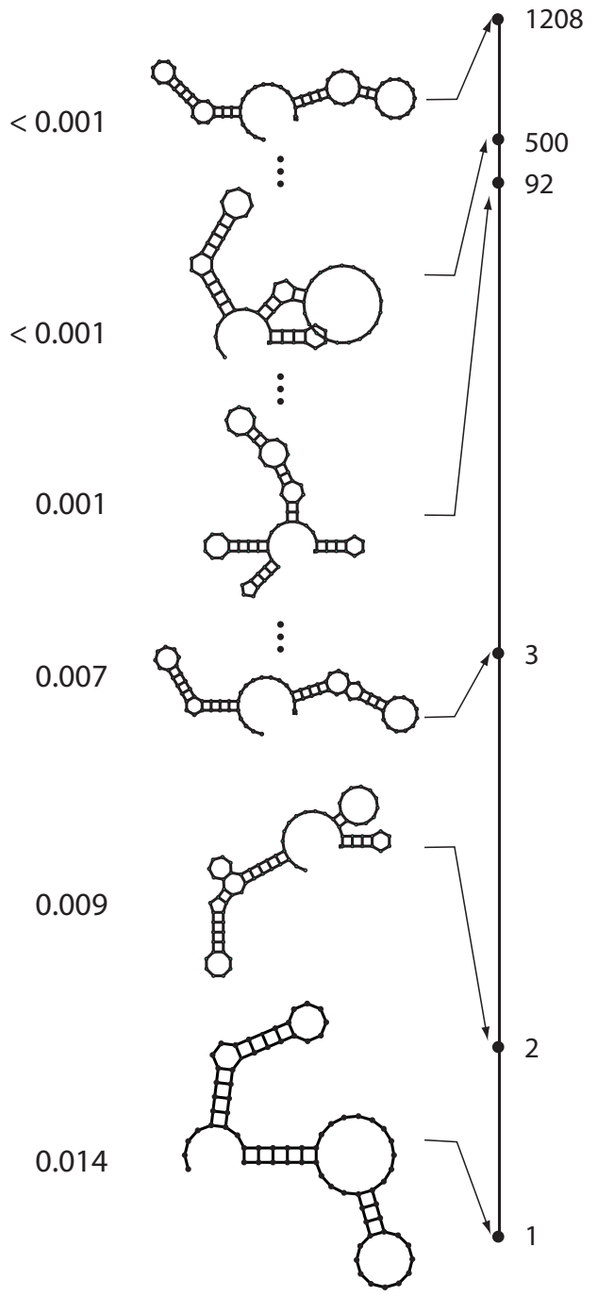
"The Future is now"

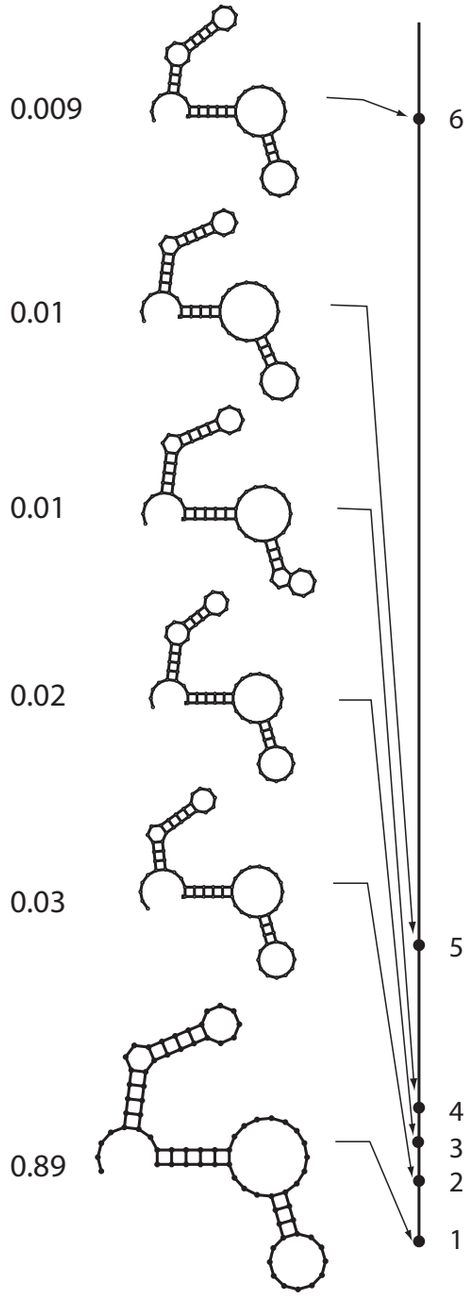


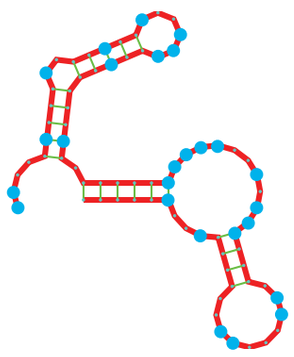
Every sequence adopts a native shape
+ a set of inducible shapes {S_1...S_n...}

A sequence ● that adopts S_{42} *inducibly*
is only one mutational step away from
a sequence ● that adopts S_{42} *natively*.

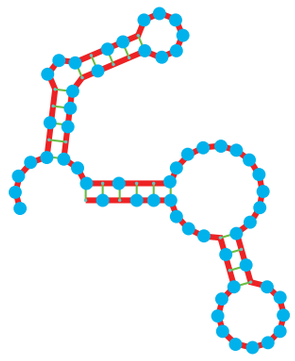
Canalization







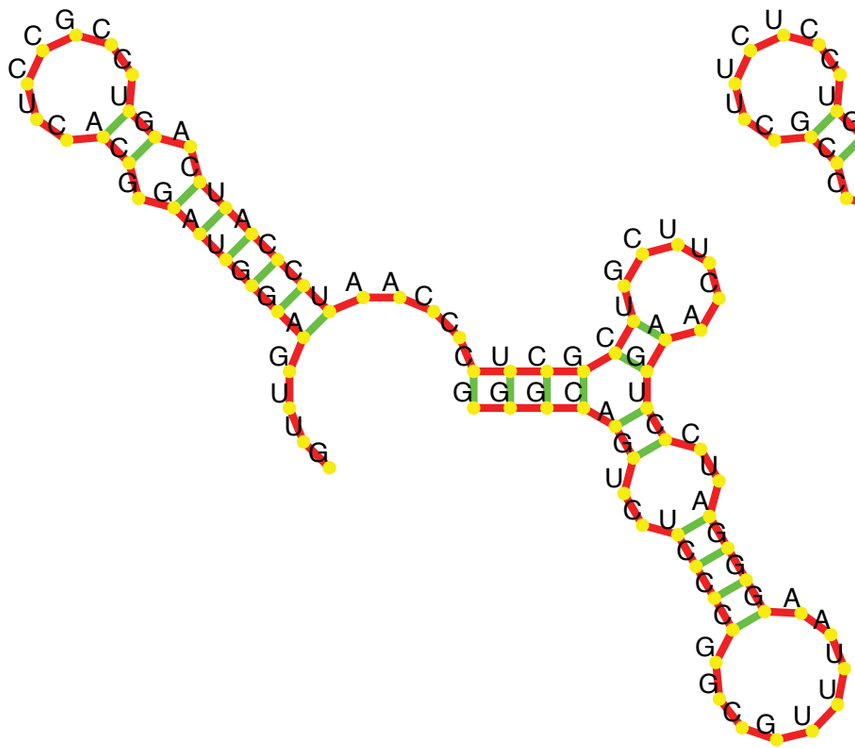
random



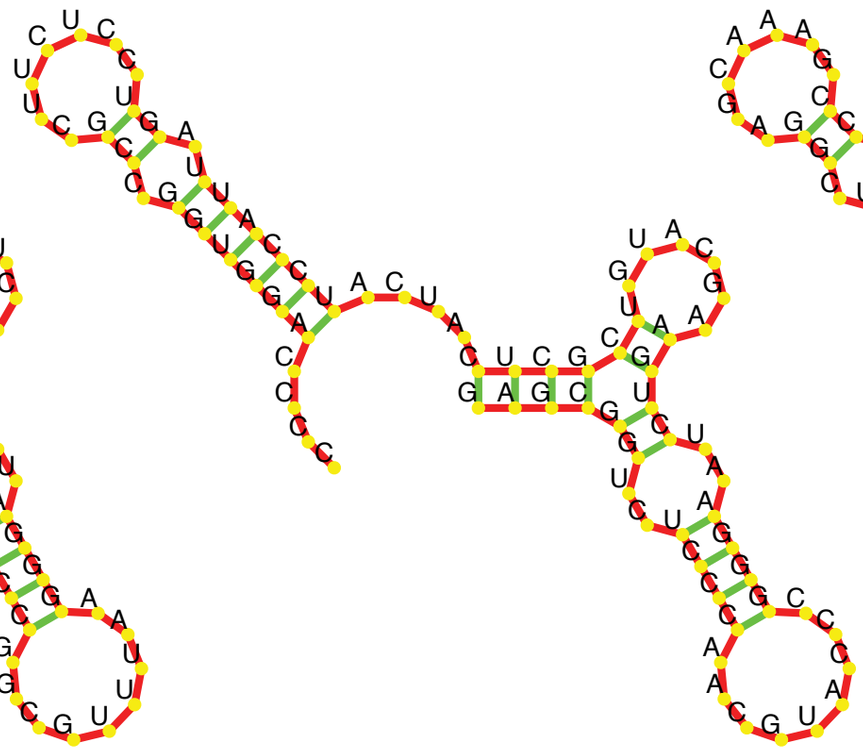
evolved

evolutionary reduction of plasticity
in a constant environment

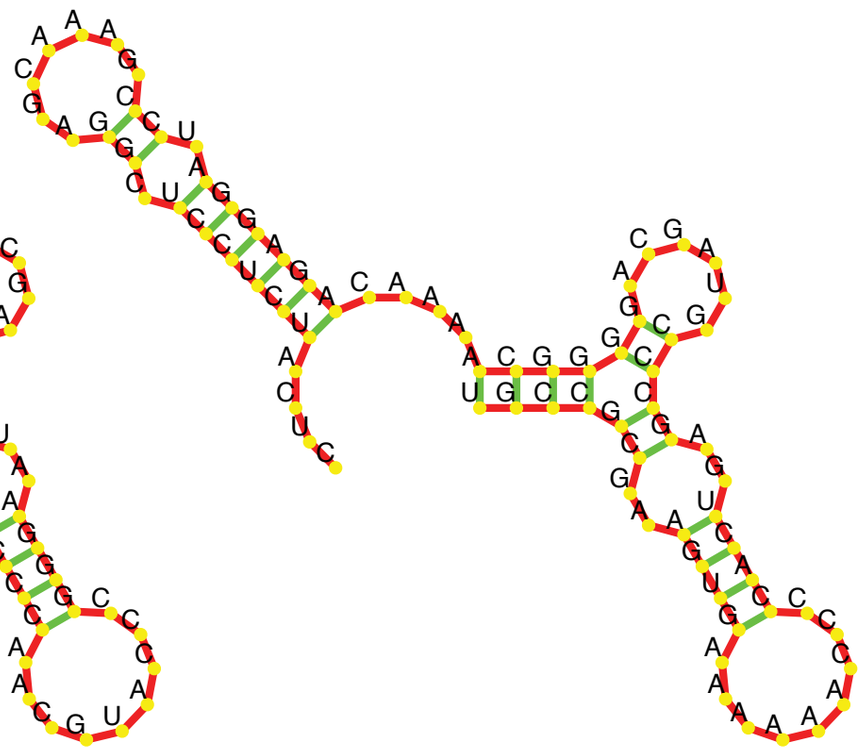
Which structure is modular?



random

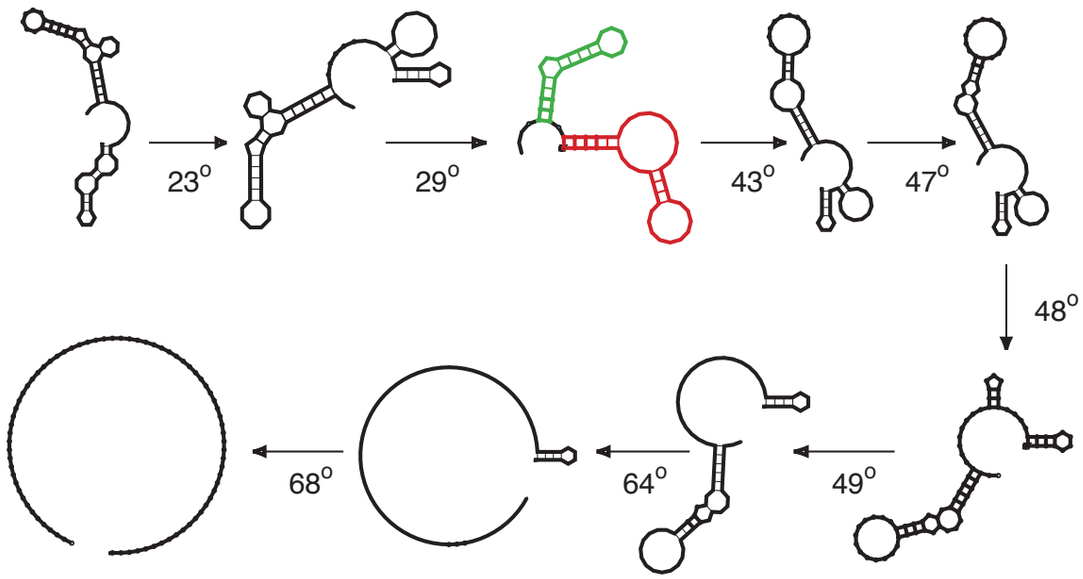


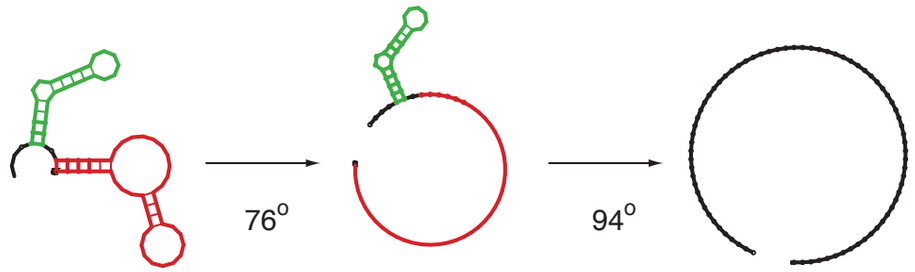
neutral evolution

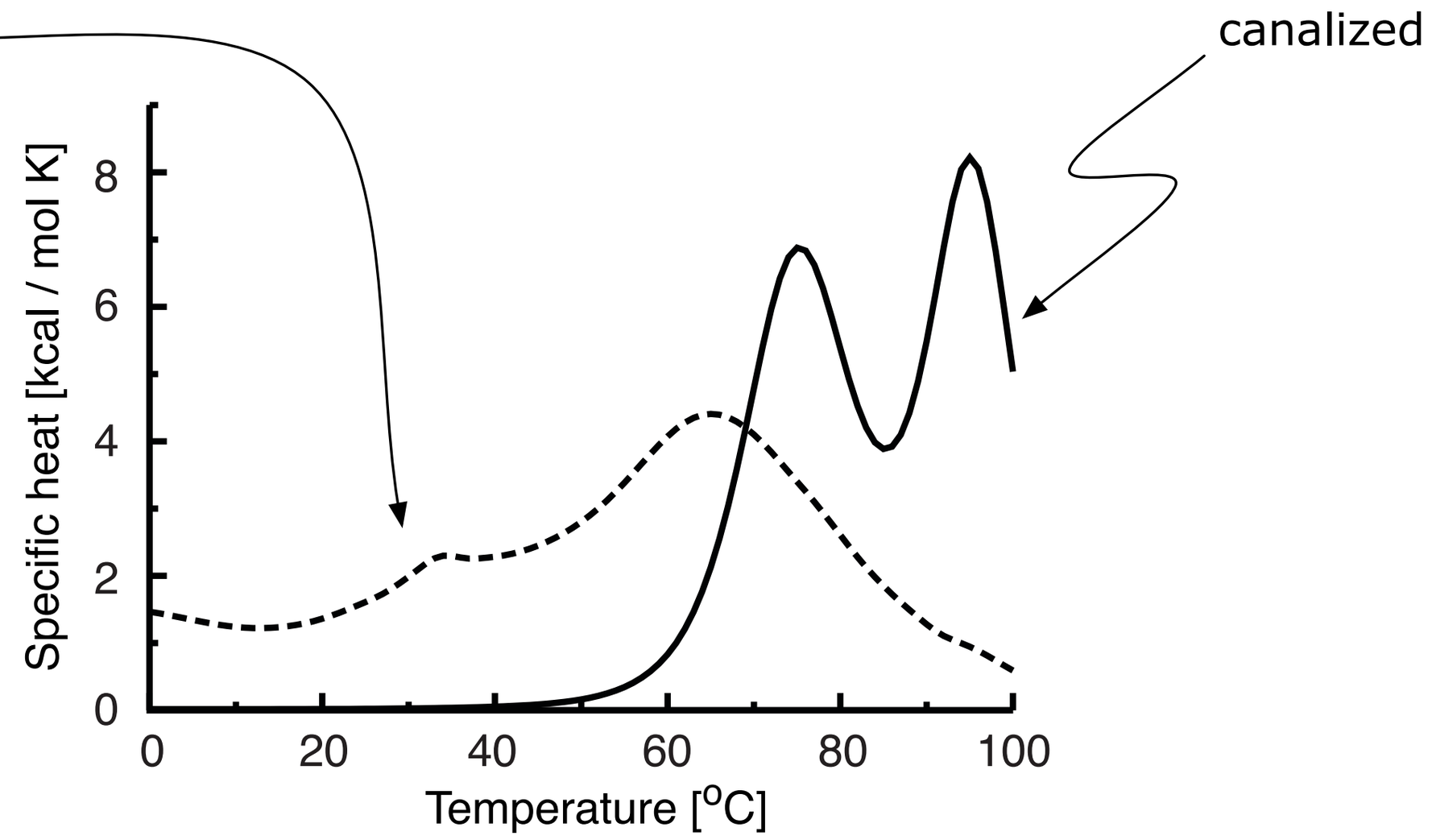


canalized

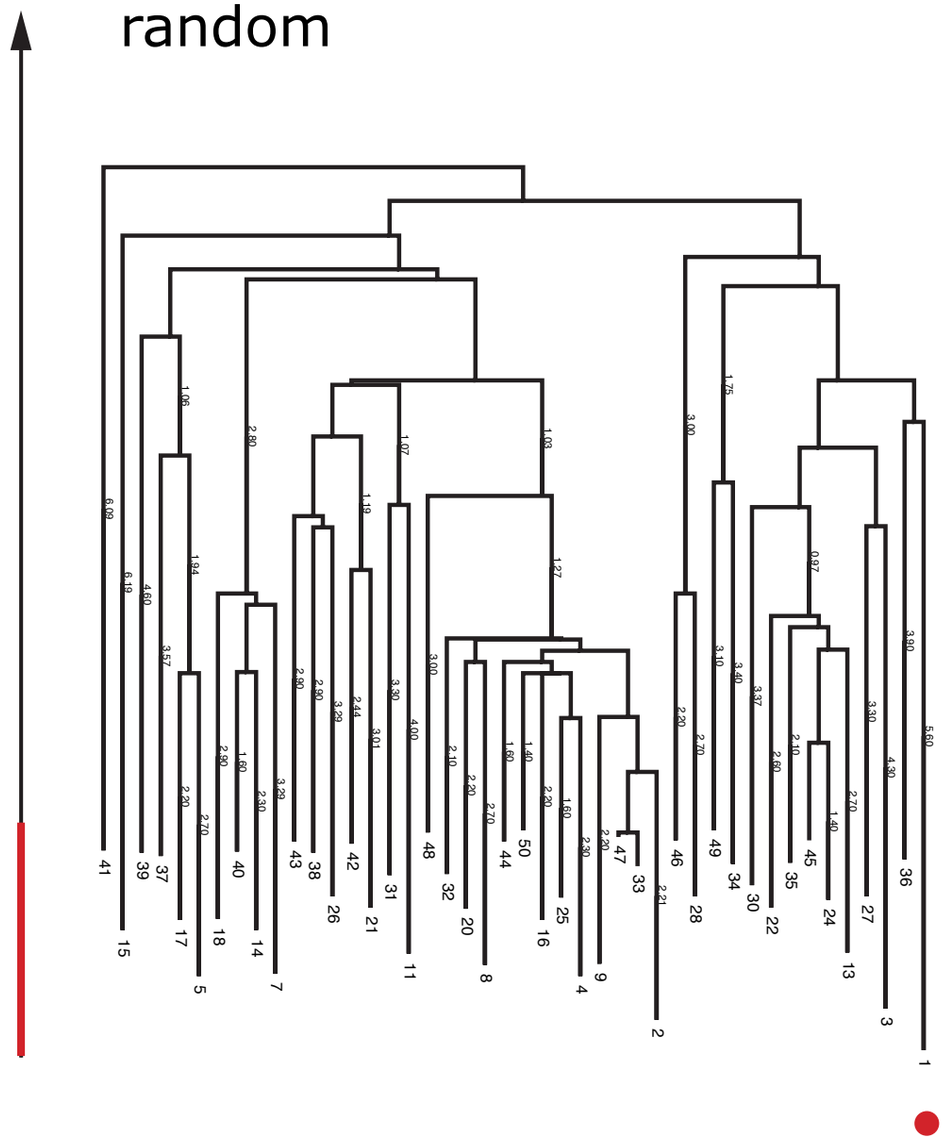
Thermo-physical Modularity

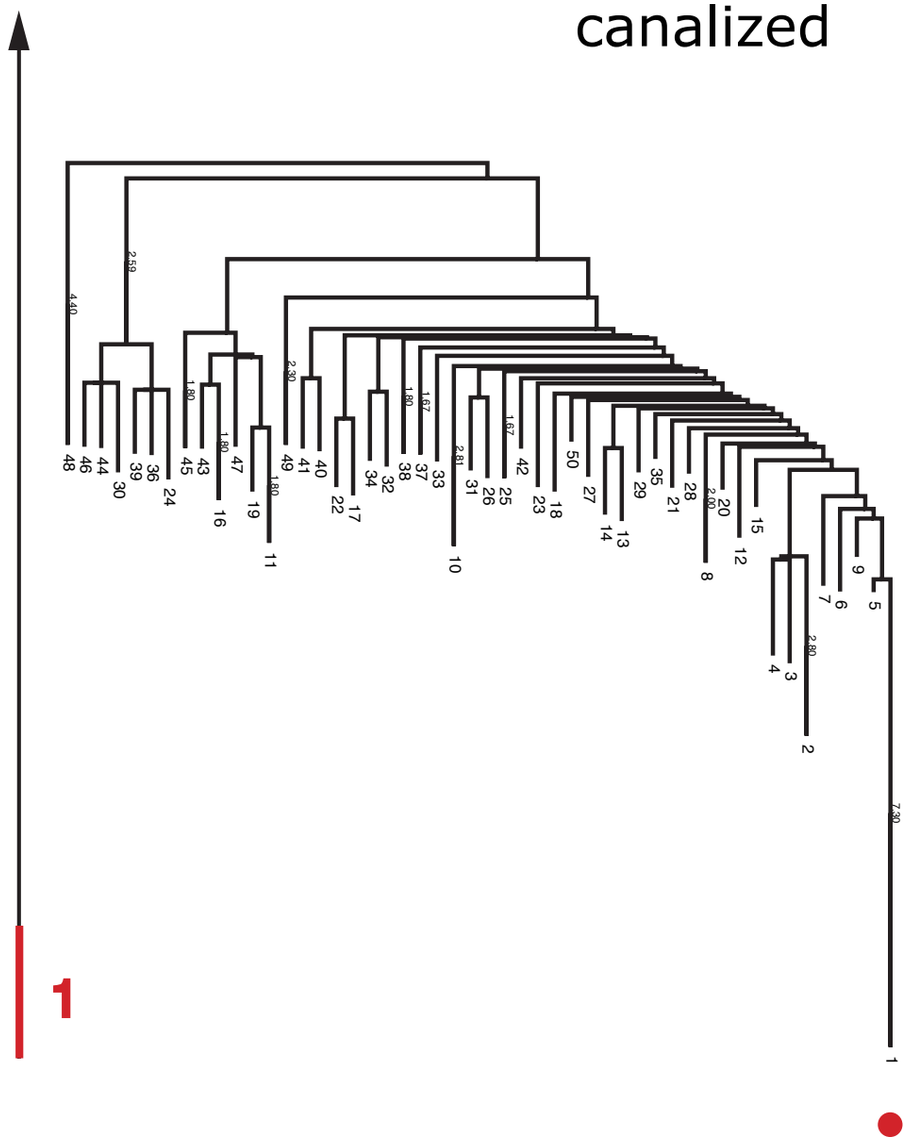




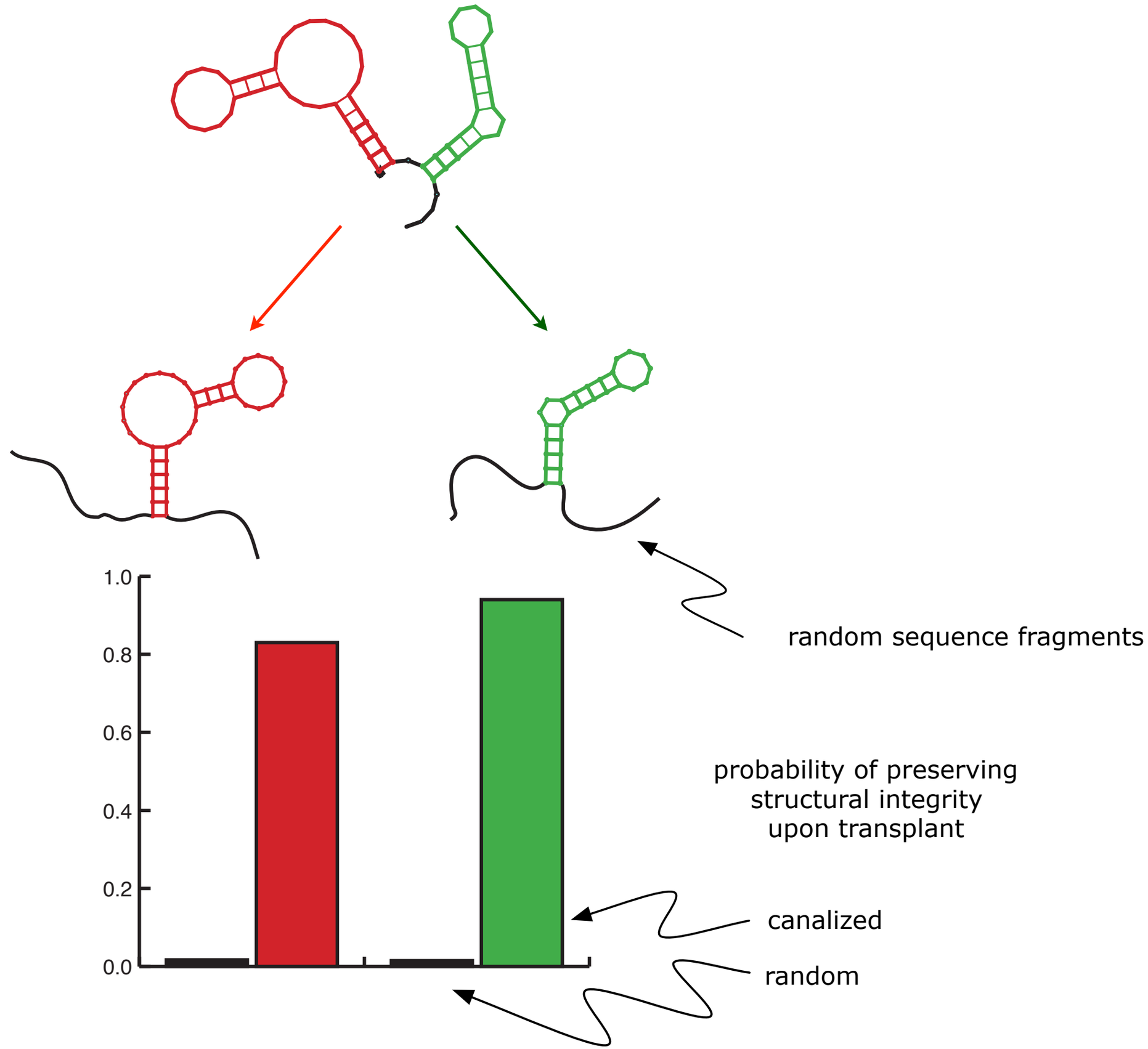


Kinetic Modularity





Context Independence (Autonomy)



Neutral Evolution of The Predecessor Function...

$$\mathbf{pred} \equiv \lambda x_1.((x_1)\underbrace{\lambda x_2.((x_2)\lambda x_3.x_3)\lambda x_4.\lambda x_5.((x_2)x_4)(x_4)x_5)}_{S}\underbrace{\lambda x_6.(x_1)\lambda x_7.\lambda x_8.\lambda x_9.x_9}_{A}$$

$$(\mathbf{pred})\mathbf{n} = (S)^n A_n$$

$$\mathbf{n} \equiv 0 \qquad A_0 \equiv \mathbf{0}$$

$$\mathbf{n} > 0 \qquad A_{n>o} \equiv A'$$

$$(S)A' \equiv \mathbf{0}$$

$$(S)^{n-1}\mathbf{0} \equiv \mathbf{n-1}$$