No protocell is an island

Roberto Serra

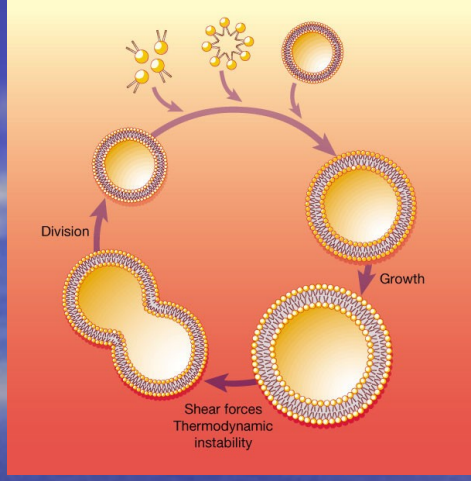
Modena and Reggio Emilia University
European Centre for Living Technology, Venice

No man is an island, entire of itself...any man's death diminishes me, because I am involved in mankind; and therefore never send to know for whom the bell tolls; it tolls for thee

John Donne (1624)

- A single protocell, even if it as able to split, is not life
  - But just a prerequisite
- “life” has to do with a population of protocells which can grow in a sustainable way
- And undergo evolution
  - Diversity in the population
  - Selection of the fastest replicating protocells
  - Inheritability of the distinguishing features

Synchronization



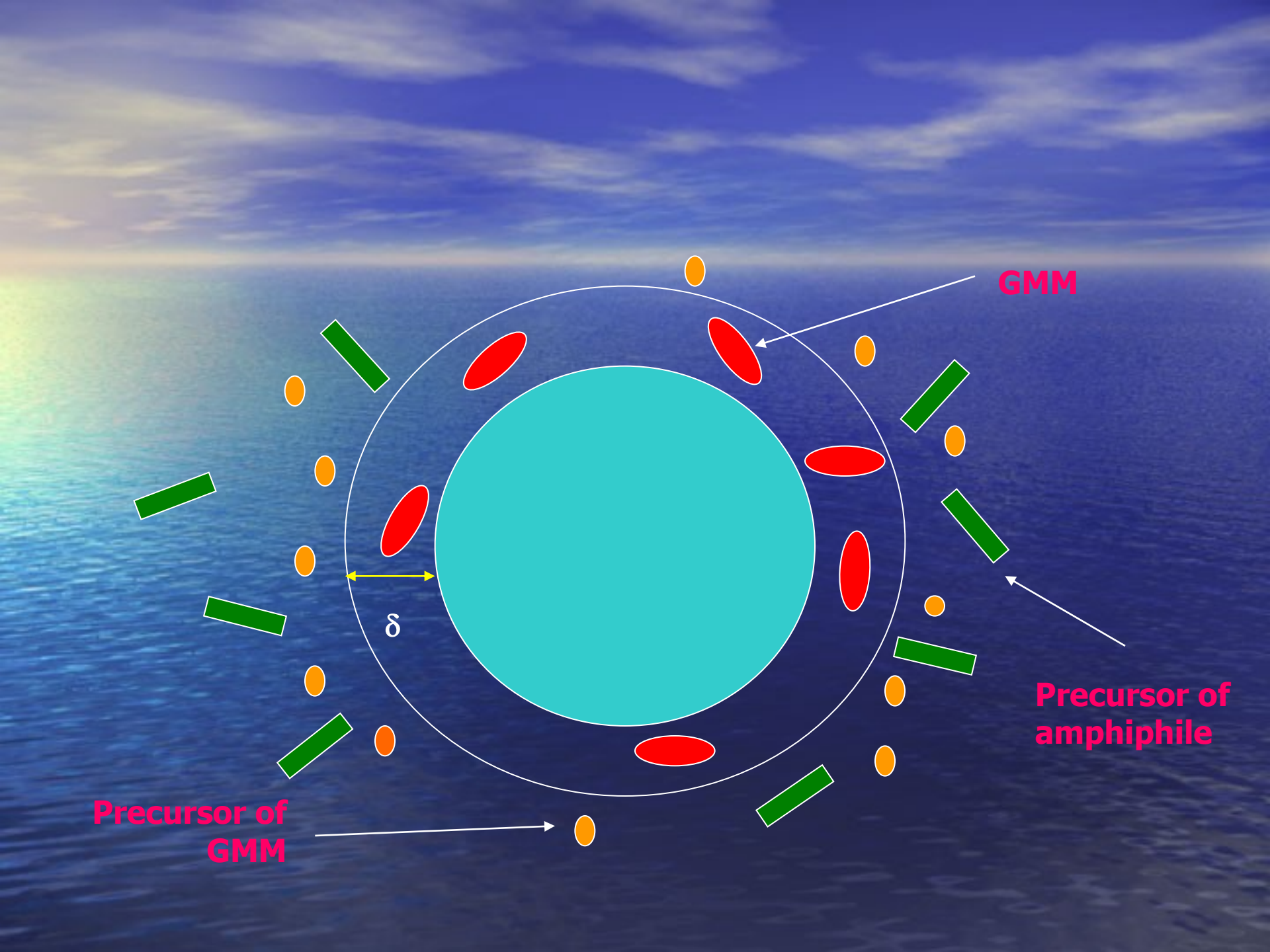
- Protocells need to contain some kind of "genetic" material which undergoes replication
- If the duplication of the genetic material took place at a pace different from that of the lipid container, sustainable growth of protocell populations could not be achieved
- Synchronization should be stable
  - A supersmart biochemist would not be sufficient

Models

- Models to understand under which conditions synchronization can be achieved
- Abstract models which capture some key features of the process
  - Emergent properties
  - Applicable to several different detailed scenarios
- Two main classes of protocell architectures:
  - The action takes place inside the cell
  - The action takes place on the surface

Surface reaction models

- An abstract model based on the Los Alamos Bug picture
- Useful to study the time behaviour of a population of such protocells
  - Turns out to be applicable also to other detailed models
- The protocell is made of
  - A lipid "container"
  - One kind (or more) of genetic memory molecules (GMMs)
  - The GMMs increases the rate of growth of the container
  - Both GMMs replication and container growth take place near the surface of the vesicle or micelle


GMM
δ
Precursor of amphiphile
Precursor of GMM

- Assume that
  - There is one kind of GMM
  - Its concentration is homogeneous in the lipid phase
  - precursors are buffered
  - surface is proportional to V^{β} (β between 1 and 2/3)
  - Spontaneous container growth is negligible

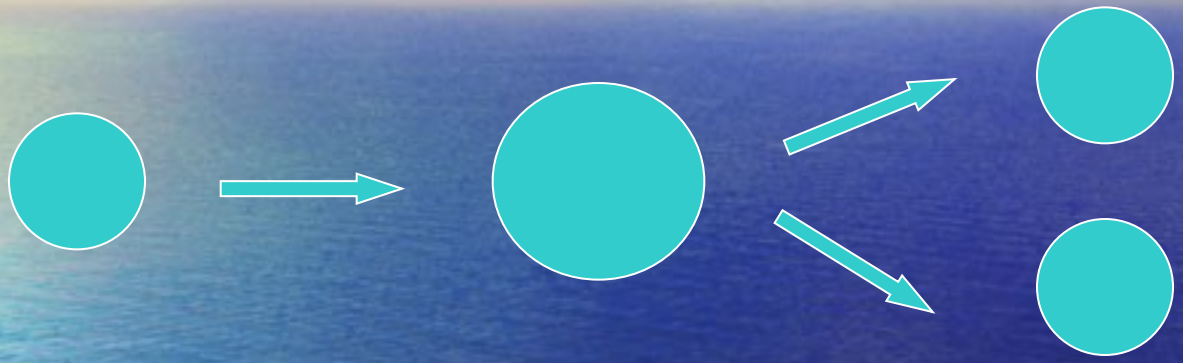
- C is the quantity of lipid container
  - Proportional to V, as density is constant)
- X is the quantity of the genetic memory molecule

$$\frac{dC}{dt} = \alpha C^{\beta-1} X$$

$$\frac{dX}{dt} = \eta C^{\beta-\nu} X^{\nu}$$

If halving takes place when saturation effects on the growth of C and X are negligible

How can we prove synchronization?



$X=X_k$
$C=\theta/2$

$X=X_f$
$C=\theta$

$X=X_{k+1}= X_f /2$
$C=\theta/2$

Given the model assumptions, all the continuous growth dynamics is determined by the initial value of X : therefore $X_k=X_{k+1}$ implies synchronization

The existence of a first integral would allow one to find a time-discrete map between X_k and X_{k+1}

Linear kinetics for SRM: $\nu=1$

$$\frac{dC}{dt} = \alpha C^{\beta-1} X$$
$$\frac{dX}{dt} = \eta C^{\beta-1} X$$

- It can be verified that
- $Q \equiv \eta C - \alpha X$
- is conserved during the growth period
- Let $D = \eta\theta/2\alpha$; then

$$X_{k+1} = \frac{X_k + D}{2}$$

$$X_{k+1} = \left(\frac{1}{2}\right)^{k+1} X_0 + \frac{D}{2}\sum_{m=0}^{k}\left(\frac{1}{2}\right)^m$$

$$X_k \to D$$

- Duplication time (C from $\theta/2$ to θ) depends upon the initial value of X, so it also tends to a constant value
- In the limiting case $\beta=1$ the doubling time can be explicitly computed

$$2X_{k+1} = X_k e^{\eta T_k}$$

$$T_k \underset{k\to\infty}{\to} \frac{1}{\eta}\ln 2$$

- The system tends to a state where the doubling of X is synchronized with that of the container
- Synchronization is not prescribed but it emerges in the dynamics
- It is a robust property, and it does not depend upon a supersmart chemist

Nonlinear kinetics: the case of a single self-replicating GMM

$$\begin{cases} \dfrac{dC}{dt} = \alpha X C^{\beta-1} \\ \dfrac{dX}{dt} = \eta X^{\nu} C^{\beta-\nu} \end{cases}$$

- a first integral is (β=1)

$$Q = C^{2-\nu} - \frac{\alpha}{\eta} X^{2-\nu}$$

$$p \equiv \left(\frac{1}{2}\right)^{2-\nu}$$

$$H \equiv \frac{\eta p(1-p)}{\alpha} \theta^{2-\nu}$$

$$\xi_n \equiv X_n^{2-\nu}$$

$$\xi_n = H + p\xi_{n-1}$$

In the large n limit
ξ_n -> H/(1-p)

So ξ_n and Therefore also X_n tend to a constant value: Synchronization again
Provided that ν<2

This result holds for more general nonlinear functions of X/C

Constant duplication time implies Exponential growth of the population: darwinian selection even if the kinetic equations are sublinear

Generalizations

- synchronization is not based on the linearity of the replication kinetics, it holds for all the systems like

$$\frac{dC}{dt} = \alpha f(C, X)$$
$$\frac{dX}{dt} = \eta f(C, X)$$

- The qualitative features of the asymptotic dynamics are not affected by the exponent β
  - Rescaling time
  - So one can limit to study the $\beta=1$ case to draw general conclusions
  - "micelle or vesicle doesn't matter"

Generalizations

- Synchronization is achieved also
- if we relax the $S \approx C^{\beta}$ approximation and adopt a more realistic vesicle geometry (e.g. spherical shell)
- if growth limiting terms are retained in the continuous growth phase
- If we assume that the growth of the container is a nonlinear function of X/C
- If we assume that the replicator is itself a part of the lipid container, like in GARD models
  - Concentration is not X/C, but X/(X+C)

Pure self-replicators, linear case

$$\begin{cases} \dfrac{dC}{dt} = \alpha' C^{\beta-1} X + \alpha'' C^{\beta-1} Y \\ \dfrac{dX}{dt} = \eta' C^{\beta-1} X \\ \dfrac{dY}{dt} = \eta'' C^{\beta-1} Y \end{cases}$$

- Two linear replicators,
- Without direct interaction
- Each one drives its own synthesis
- There are two first integrals
- It can be proven that the GMM which reproduces fastest prevails, while the second eventually vanishes through the generations (survival of the fittest)

- If the kinetics is nonlinear both replicators survive
  - Thir relative proportion is determined by the ratio of kinetic coefficients

Interacting molecules: linear kinetics

- The most interesting case is that where the informational molecules directly interact
- The simplest case is that of linear kinetic equations
  - The overall model is of course nonlinear
  -

$$\frac{dC}{dt} = \sum_k \alpha_k X_k$$
$$\frac{dX_i}{dt} = \sum_k M_{ik} X_k$$

$$\frac{dC}{dt} = \vec{\alpha} \cdot \vec{X}$$
$$\frac{d\vec{X}}{dt} = M\vec{X}$$

- By integrating and halving one gets

$$\vec{X}(T_{k+1}) = e^{M\Delta T_k}\,\vec{X}_k$$
$$\vec{X}_{k+1} = \frac{1}{2}e^{M\Delta T_k}\,\vec{X}_k$$

- So, if there is a stationary value

$$\vec{X}_\infty = \frac{1}{2}e^{M\Delta T_\infty}\,\vec{X}_\infty$$

- X can be stationary provided that it is an eigenvector of M

$$M\,\vec{X}_\infty = \lambda\,\vec{X}_\infty$$

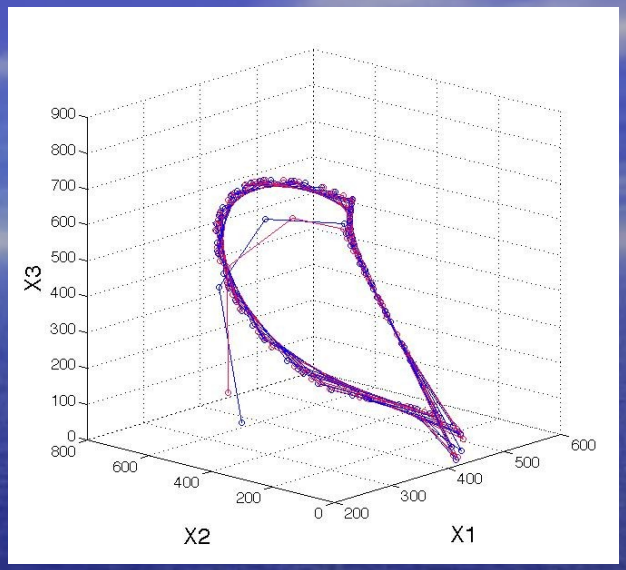
- The eigenvalue is related to the asymptotic duplication time

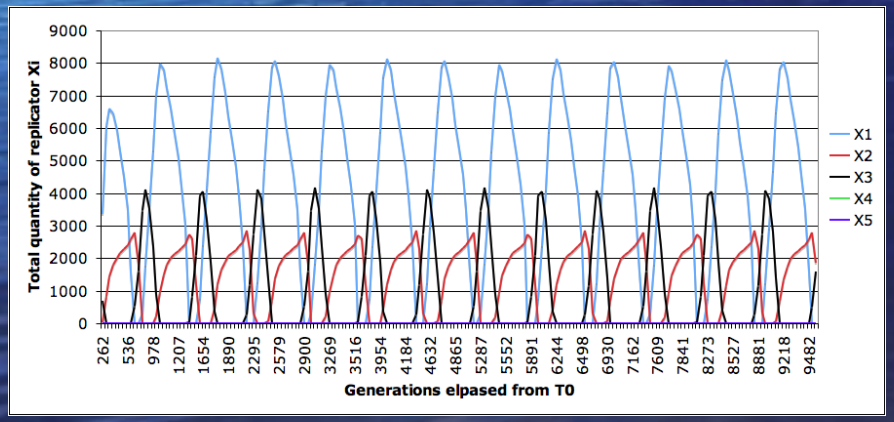
$$\Delta T_\infty = \frac{\ln 2}{\lambda}$$

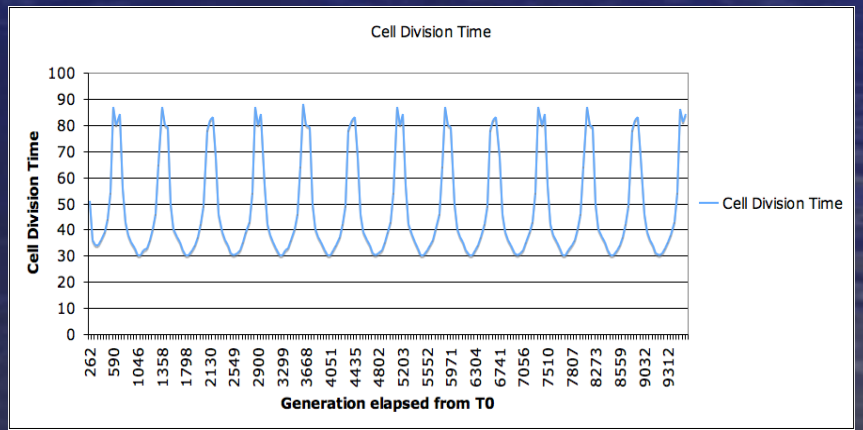
- The long term behaviour is ruled by the eigenvalue with the largest real part (λ_1)
  - If its real part is negative or null, X gets extinguished
- If $\lambda_1$is real and positive, its eigenvector v_1 correctly identifies the components which survive
  - i.e. those which are >0
- If M is nonnegative, λ_1 is real and positive
  - Perron-Frobenius theorem

A different behaviour: supersynchronization

- If M has negative entries and λ_1 has a nonvanishing imaginary part, the duplication time may oscillate in time
- This behaviour allows sustainable population growth at a variable rate







Comments on the linear case

- In the linear case the replicator equations rule the behaviour of the system
- Vesicle splitting prevents unphysical infinite growth
- The asymptotic vector of concentrations is the same as that of the ODE system for the X
- The coefficients of the container growth equation do not affect the final ratio of abundances of various types of replicators, although they influence their absolute values

Nonlinear interacting replicators

- The quasilinear case:
- Imposing an upper limit on the rate of production of GMMs
  - Using a sigmoid function
- The behaviour is similar to that of the corresponding linear cases
  - Except that in some cases linear oscillates while quasilinear synchronizes

- In the case of two replicators with quadratic kinetics

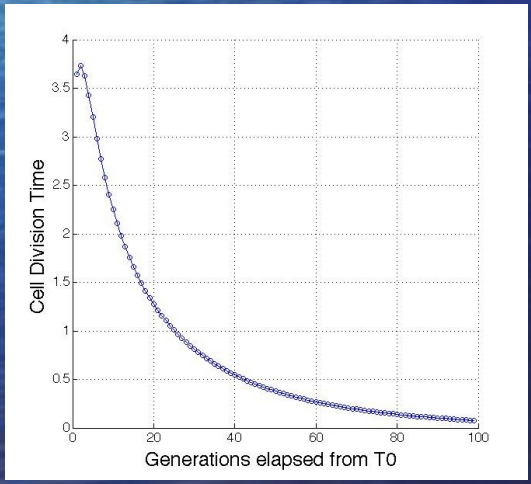
$$\frac{dC}{dt} = \alpha' C^{\beta-1} X$$
$$\frac{dX}{dt} = \eta' C^{\beta-2} XY$$
$$\frac{dY}{dt} = \eta'' C^{\beta-2} XY$$

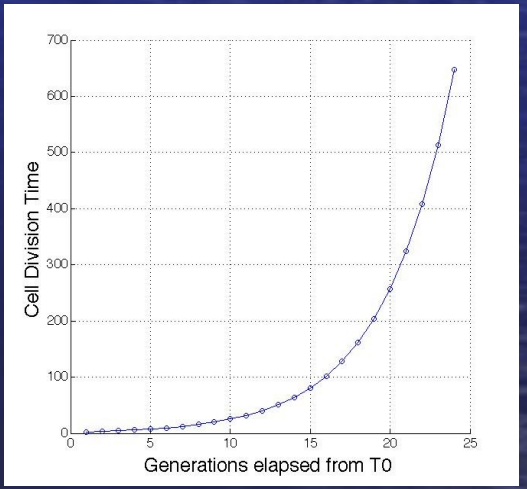
- No synchronization is observed

Several replicators

$$\frac{dX_i}{dt} = C^{\beta-2} \sum_{k=1}^{N} M_{ik} X_i X_k$$

- The results which hold for two replicators generalize to the case of several of them
- No synchronization:
- extinction or explosion

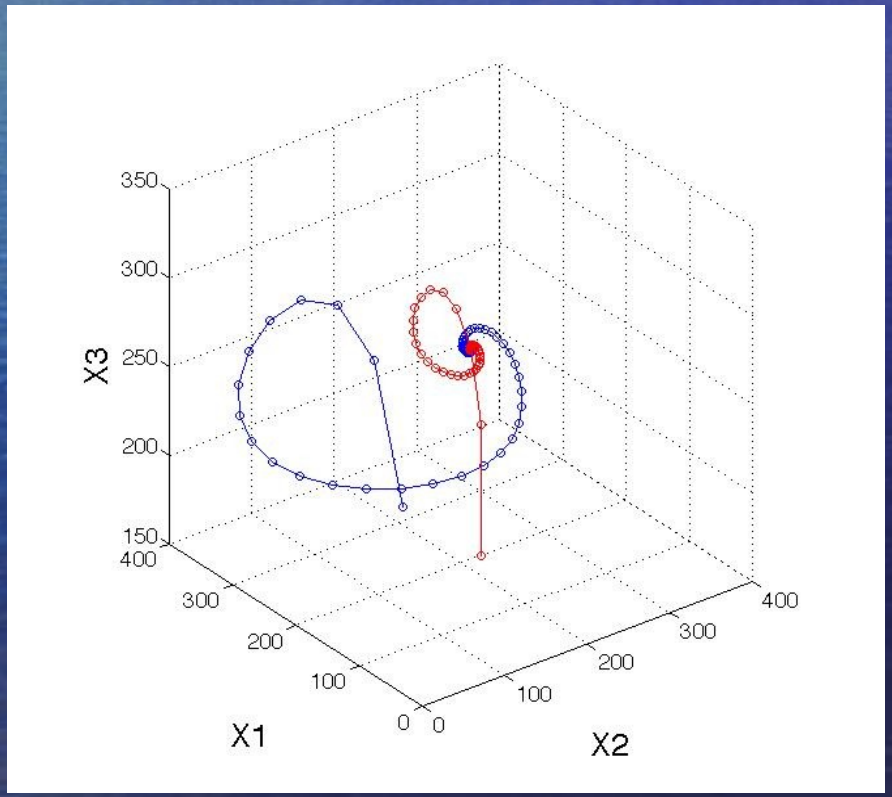




Adding a linear term

$$\frac{dX_i}{dt} = C^{\beta-2}\sum_{k=1}^{N} M_{ik}X_iX_k + C^{\beta-1}\eta_i X_i$$

- Synchronization or extinction are observed



Second order, without self-replication

$$\frac{dX_i}{dt} = C^{\beta-2} \sum_{k=1}^{N} M_{ijk} X_j X_k$$

$$M_{ijk} = \mu_{ijk} (1 - \delta_{ij})(1 - \delta_{ik})$$

- Synchronization is observed if there are enough nonvanishing matrix elements
- Otherwise extinction

The Lorenz equations

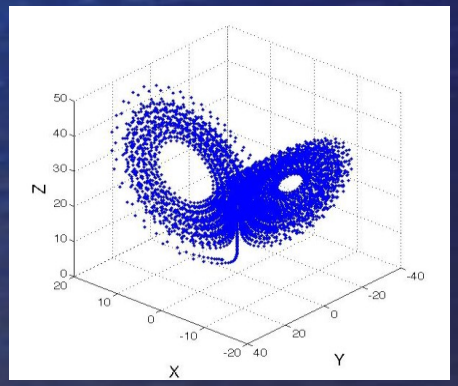
- Let us suppose that the replicator equations are those of the Lorenz model
- Without halving and growth, they give rise to the well-known strange attractor
- But in protocells they synchronize

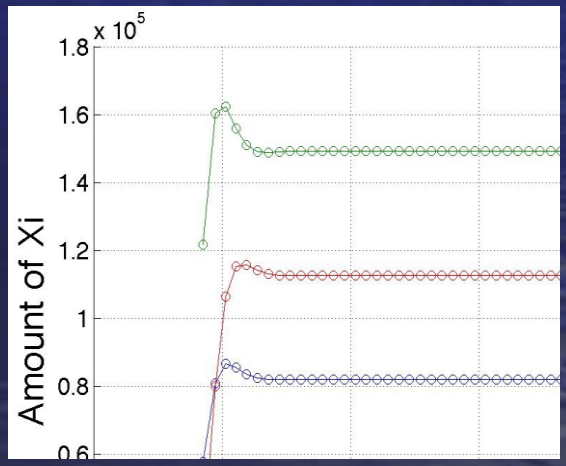
$$\frac{dC}{dt} = \alpha X$$

$$\frac{dX}{dt} = \sigma(Y - X)$$

$$\frac{dY}{dt} = rX - Y - \frac{1}{C} XZ$$

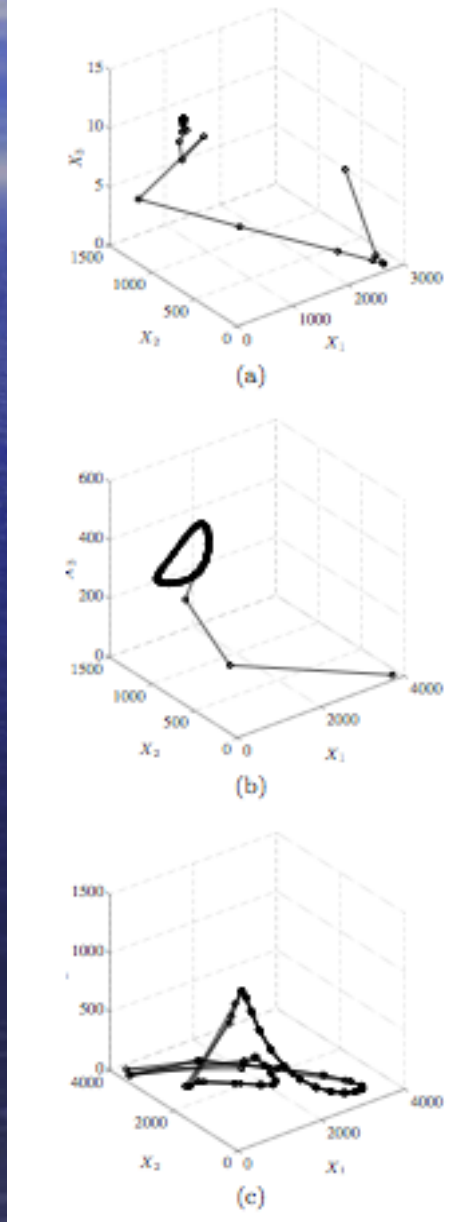
$$\frac{dZ}{dt} = \frac{1}{C} XY - bZ$$





A kinetic model with chaotic behaviour

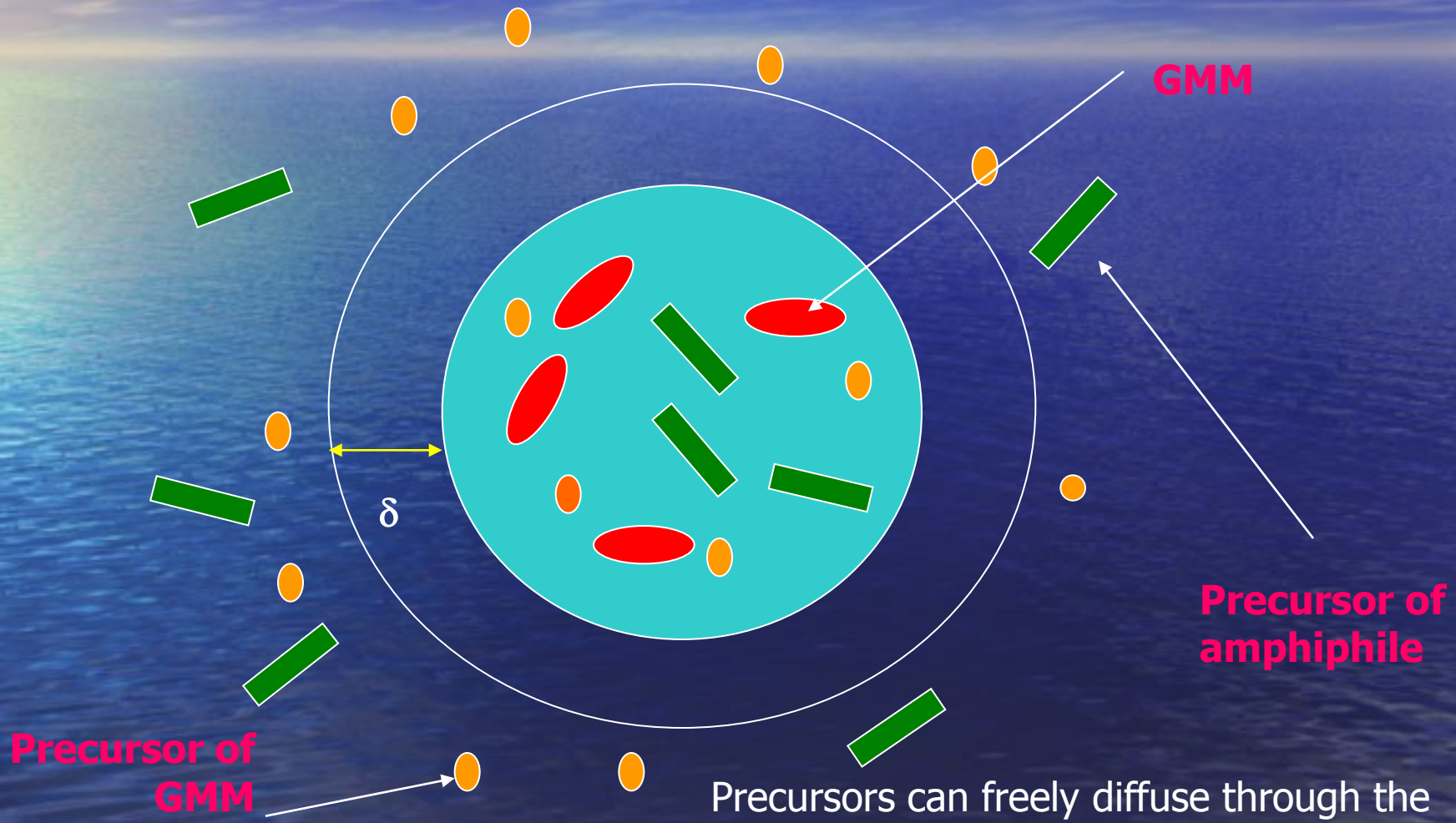
- Williamowsy-Roessler
- Adapted to the protocell case
- Coupled to the container growth
- dC/dt=αX
- A bifurcation as a function of α
- Small α: chaotic behaviour is conserved
- increasing α one finds a periodic orbit
- As α is further increased a fixed point
- is found



Internal reaction models

- In these models Both GMMs and amphiphiles are formed in the interior of the protocell
- Using precursors
- which can freely diffuse through the membrane (“fast IRMs”)

Internal reaction models



Precursors can freely diffuse through the membrane ("fast IRMs")

IRMs behave much like SRMs

- If the diffusion of precursors is fast the equations turn out to have the same form as those of surface reaction models
- Synchronization in the linear case
- Lack of synchronization in the purely quadratic case
- The behaviour of nonlinear models is similar to that of SRMs

Finite diffusion rate

- Let us now consider that the membrane limits the diffusion of precursors
- We will suppose that the external concentration of precursors of amphiphiles and that of precursors of replicators are buffered
- But the internal concentration is ruled by Fick's law
- The rate of production of new X is proportional to
- $V_i[X][P_X]=V_i^{-1}XP_X$

$$\frac{dC}{dt} = \alpha' h_C V_i^{-1} X P_C$$

$$\frac{dX}{dt} = \eta' h_X V_i^{-1} X P_X$$

$$\frac{dP_X}{dt} = SD_X\left(E_X - \frac{P_X}{V_i}\right) - \eta' h_X V_i^{-1} X P_X$$

$$\frac{dP_C}{dt} = SD_C\left(E_C - \frac{P_C}{V_i}\right) - h_C \alpha' V_i^{-1} X P_C$$

- Synchronization!
  - Proven by simulations

conclusions

- Synchronization is widespread
  - It is robust with respect to major changes in the models
- Fine tuning of the rate constants is not necessary
- These results are based on a deterministic model
- But the behaviour is robust with respect to
  - fluctuations in splitting threshold
  - uneven division

However

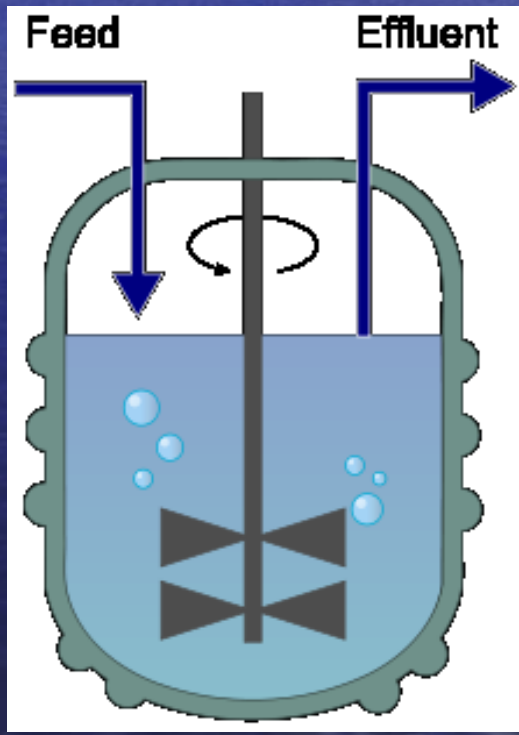
- The model discussed so far ignores creation of new molecular species
  - There can be only selection in the initial population of protocells
- Stochastic dynamics is necessary
  - In particular when the number of molecules is small

Zooming into the dynamics of the genetic memory molecules

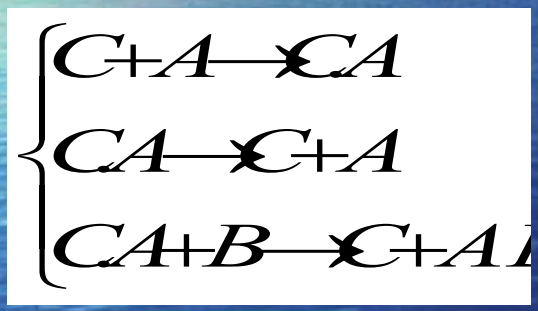
- They can be supposed to be largely random
- And so are their interactions
  - A random reaction network
- Kauffman model: the molecules are linear chains from an alphabet
  - E.g. {A,B}
- Each molecule has a certain probability of catalyzing a given reaction
  - Chosen at random
- Only catalyzed reactions are taken into account

The scenario

- A set of randomly generated molecules
- In an open-flow well-stirred reactor
  - In the future, in a protocell
- Catalysts are themselves reaction products



Catalyzed reactions



-Two types of reaction:
- Condensation
- Cleavage

- A particular species is assumed to catalyse a particular reaction with independent probability p

- It is assumed that these reactions occur at a negligible rate unless they are catalysed

- trimolecular reactions are neglected, so condenzation takes place in two steps

Sets of reactions

$$R = \sum_{i=1}^{N} (L(s_i) - 1) + N^2$$

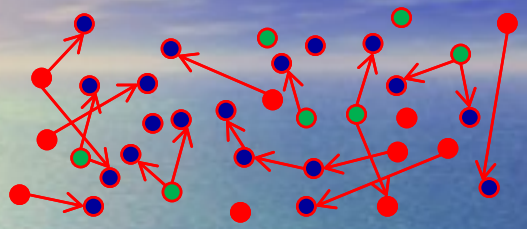
- R is the number of conceivable reactions among all the existing species
- In a given network realization only a subset could possibly be catalyzed (admissible reactions)
  - Their expected value is R*p
- In a given story, at a certain time, only a subset of the admissible reactions (the possible reactions) will have a chance to take place
  - i.e. those whose catalysts and reactants do exist at that time
- The actual reactions which take place are those which correspond to encounters which do happen

Which graph describes the reaction network?

- Asynchronicity poses a new problem
- There are at least three types of reaction graphs
- That of all the reactions which are possible, given the existing molecular species
- That of the reactions which actually have taken place
- That of the reactions which take place exactly at time t
  - In the Gillespie algorithm there is only one reaction per step, therefore the last graph is useless

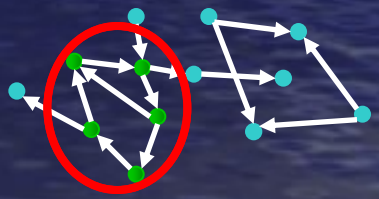
- The actual reaction graph better describes the trajectory of a system
- If a given species becomes extinct all its reactions are cancelled
  - However its properties are retained for possible future reappearance
- One has however to define a time window for forgetting
  - If the memory were infinitely long then one would continue to take into account reactions which are extremely rare

Autocatalytic sets



	N1	N2	N3	N4	N5	N6	N…	N…	N…	N…	N…	N…	N…	N…	N…
N1	0	0	0	0	0	0	…	0	0	1	0	0	0	1	0
N2	0	0	1	0	0	0	…	0	0	0	0	0	0	0	0
N3	0	0	0	0	0	0	…	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	…	0	0	0	0	0	0	0	0
N5	0	0	0	0	0	0	…	0	0	0	0	0	1	0	0
N6	0	0	0	0	1	0	…	0	0	1	0	0	0	0	0
N…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
N…	0	0	0	0	0	0	…	0	0	0	0	1	0	0	0
N…	0	0	1	0	0	0	…	0	0	0	0	0	0	0	0
N…	0	0	0	0	0	0	…	0	0	0	0	0	0	0	0
N…	0	0	0	0	1	0	…	0	0	0	0	0	0	0	0
N…	0	0	0	0	0	0	…	0	0	0	0	1	0	0	0
N…	0	0	0	0	0	0	…	0	1	0	0	0	0	0	0
N…	0	0	1	0	0	0	…	0	0	0	0	0	0	0	0
N…	0	0	0	0	0	0	…	0	0	0	0	0	0	0	0

- They are identified by the study of the eigenvalues of the adjacency matrix
  - When $\lambda_1 \geq 1$ an ACS appears
- Or by searching the Strongly Connected Components of the graph



Changing the influx

changing influx diversity (up to "tetramers")

(A, B, AA, AB, BA, BB
AAB, ABA, …, AAAA,
AAAB, …, BBBB)

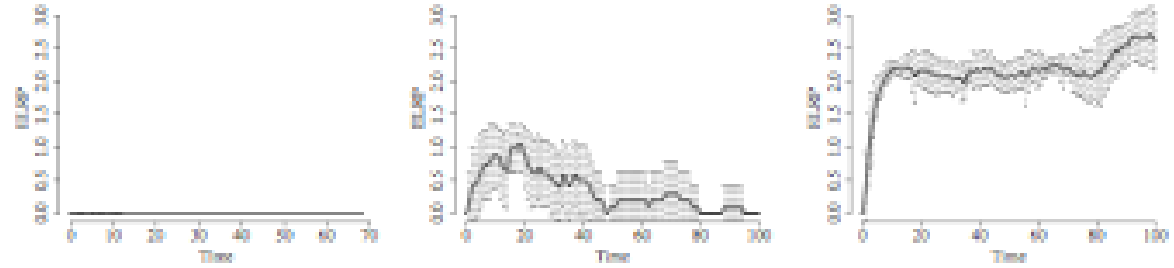


Fig. 2 Graphs show the average time behaviour of the ELRP with respect to the heterogeneity of the influx (10 different runs, the error bars represent the standard deviation). From left to right: Influx composed by all the species up to length 2, all the species up to length 3 and all the species up to length 4.

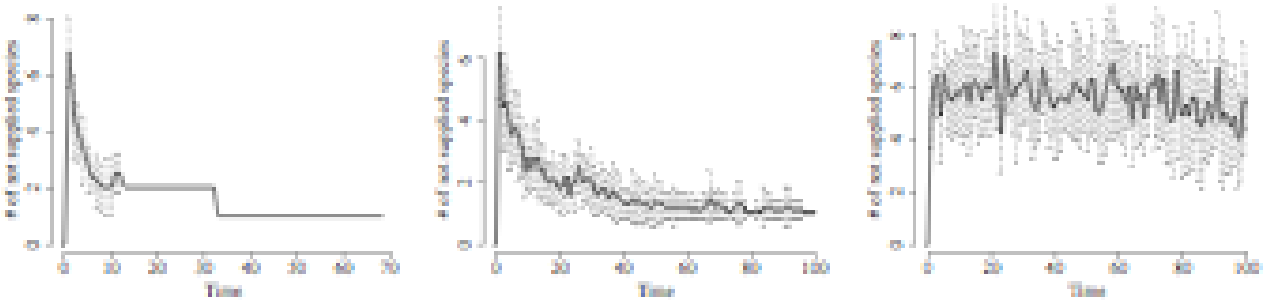


Fig. 3 Graphs show the average amount of catalysts not belonging to the influx with respect to the different compositions the influx (10 different runs, the error bars represent the standard deviation). From left to right: Influx composed by all the species up to length 2, all the species up to length 3 and all the species up to length 4.

Residence time

- The relative importance of influx vs initial internal composition depends of course upon the residence time
- The figure shows that the number of surviving species grows with residence time

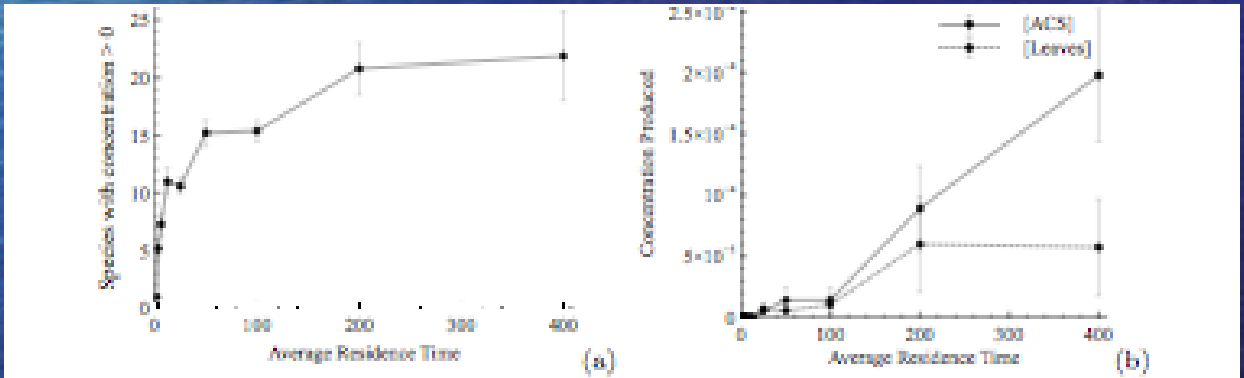
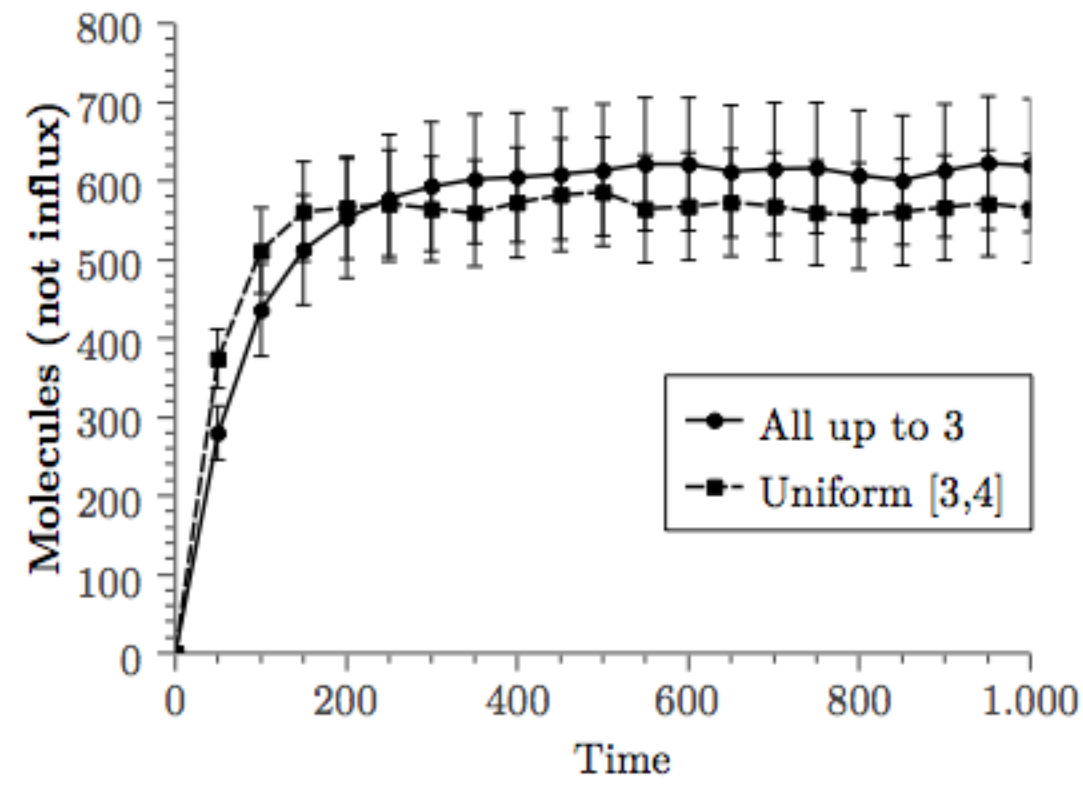


Fig. 1 The graphs show: (a) the average number of species with concentration greater than 0 not belonging to the species forming the incoming flux in function of 10 different $< \psi >$, (b) The overall molecules concentration produced within the ACSs and by the ACS leaves in

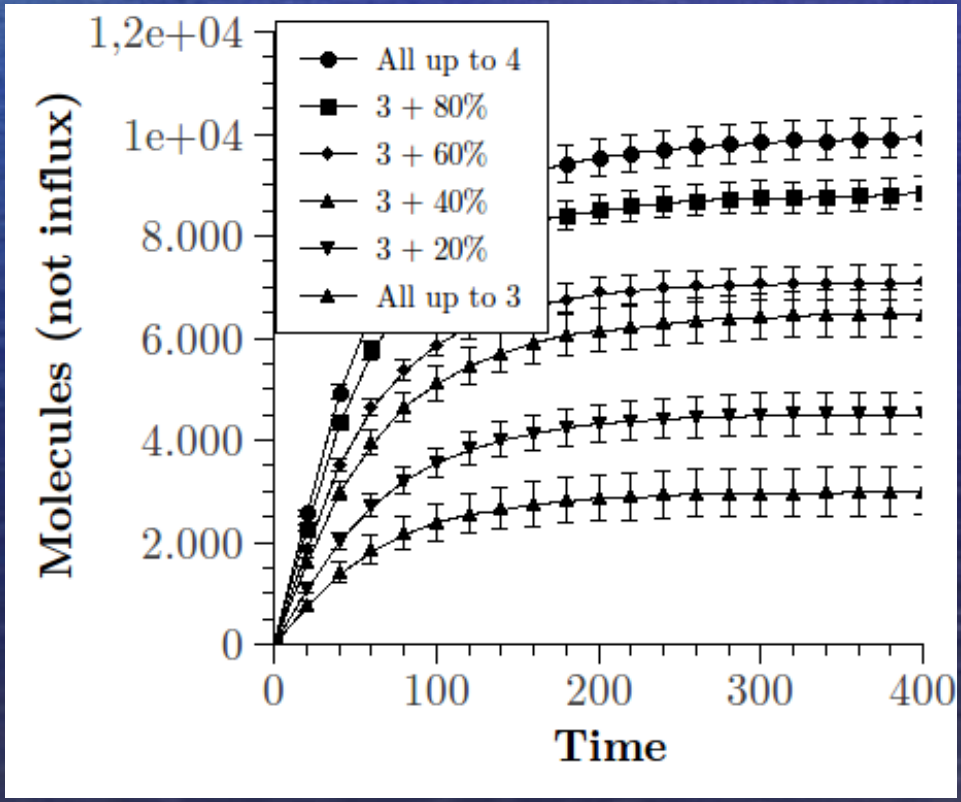
Diversity or maximum length?

- Changing the maximum length while keeping the total number fixed has a limited effect

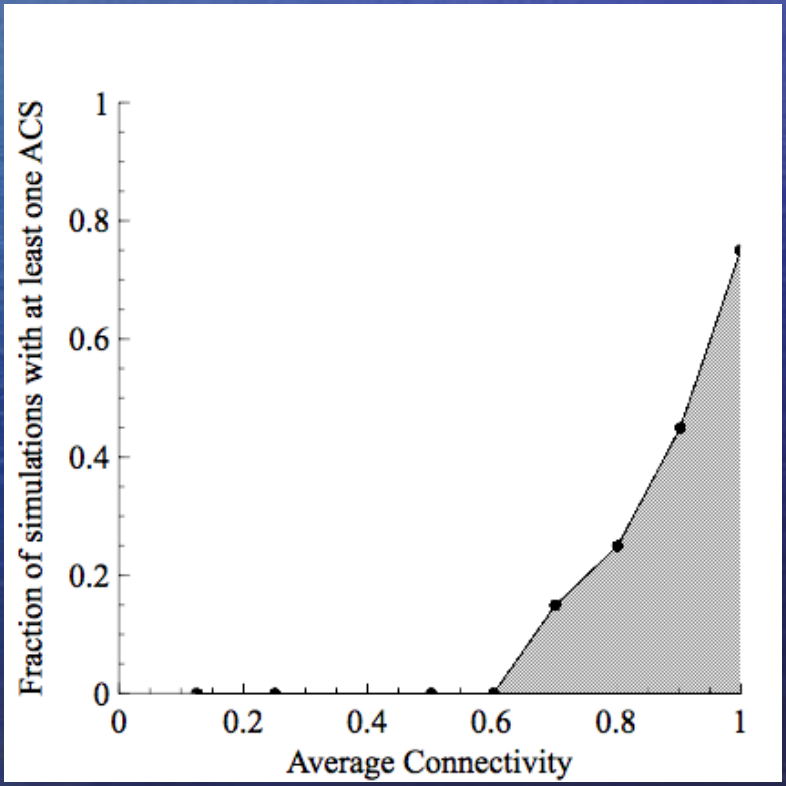


Diversity or maximum length?

- Changing the diversity of the incoming flow affects the outcome



Dependence of ACSs upon reaction probability

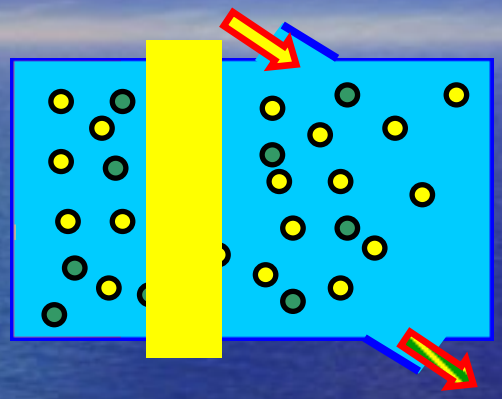


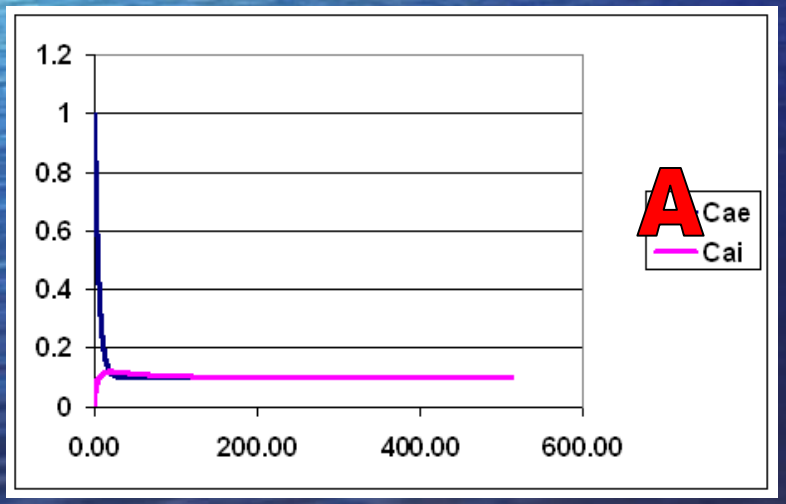
fragility

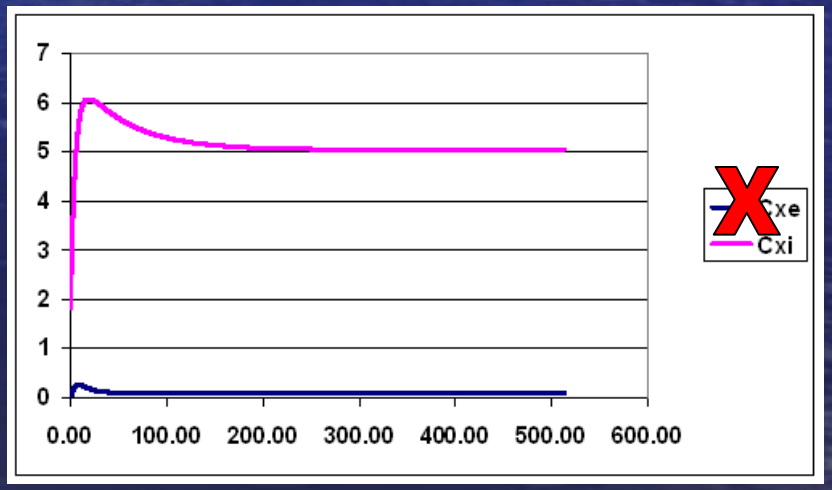
- In several simulations ACSs are fragile although the reaction probability is set to the critical value
- The influx is chosen in such a way as to not include ACSs
- Those which are formed recruit some new molecules which are formed in the growth process
- Some of them are present in low numbers and take place rarely
- The presence of ACSs does not usually produce dramatic effects on molecule number

- It would be helpful to be able to increase the concentrations of chemicals where they are needed
  - E.g. in protocells
- A possible (hypothetical) mechanism could be based on the reactions which take place on the two sides of a surface
- Which separates two compartments with different volumes
  - A large one and a small one

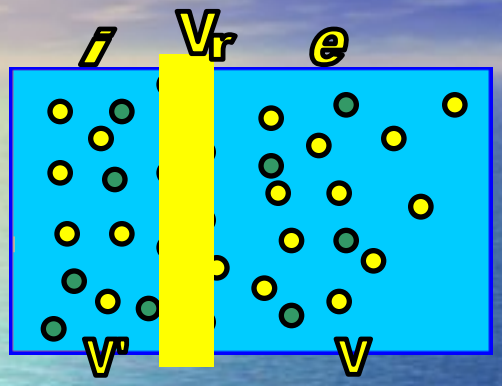
CSTR system

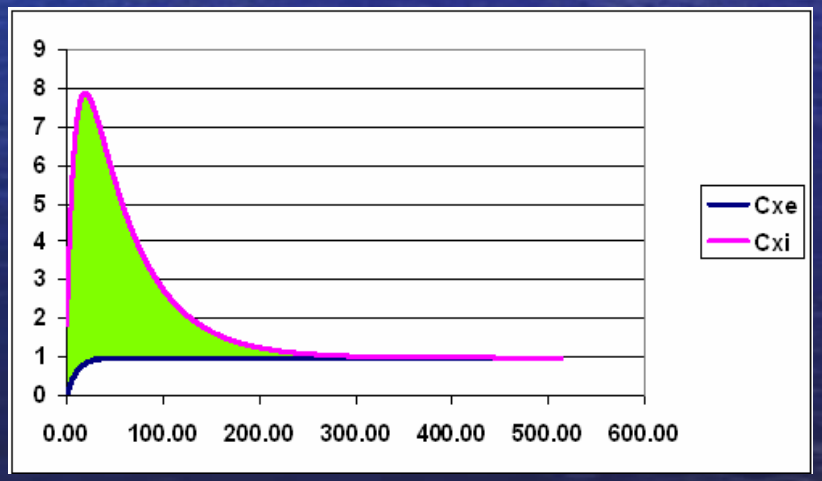



1.2
1
0.8
0.6
0.4
0.2
0
0.00
200.00
400.00
600.00
A
Cae
Cai


7
6
5
4
3
2
1
0
0.00
100.00
200.00
300.00
400.00
500.00
600.00
X
Cxe
Cxi

Isolated system



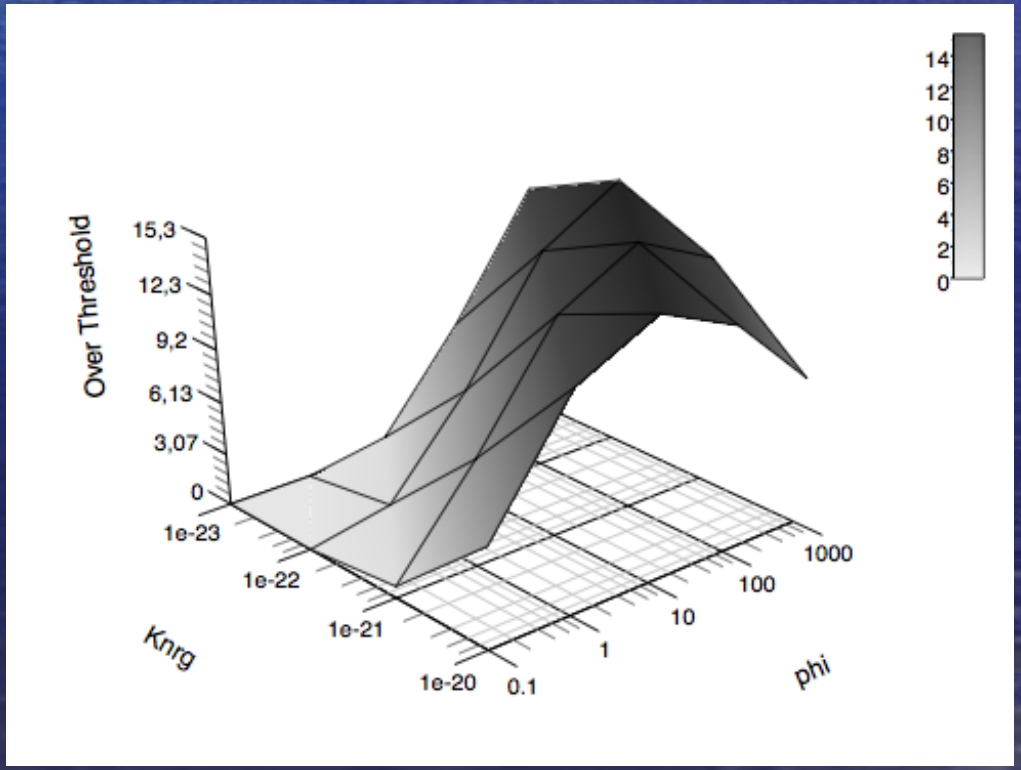


Information, matter and energy!

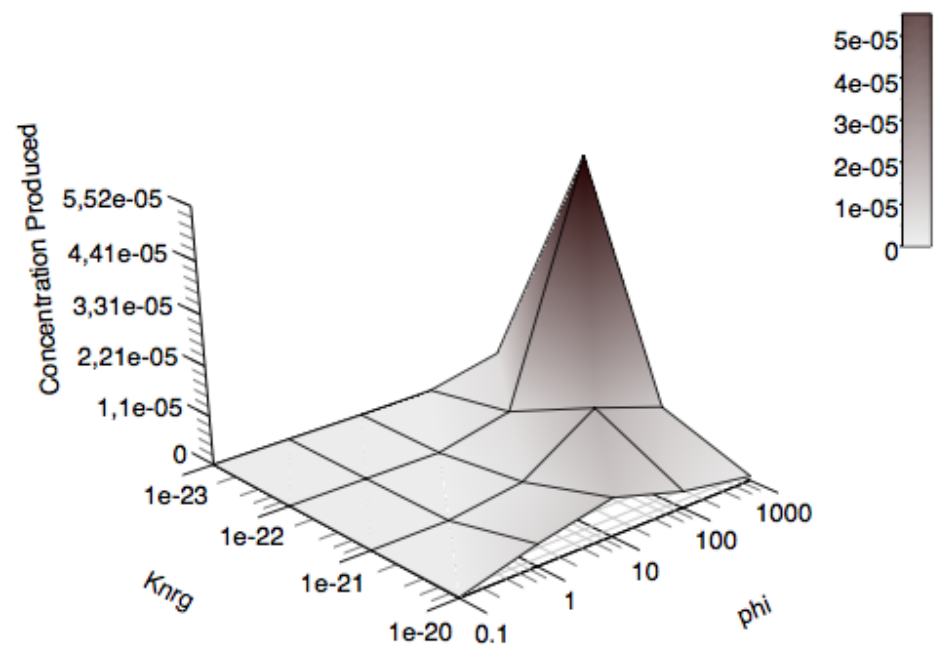
- We consider that some reactions may require activation
  - in reactants or catalysts
  - Endoergonic reactions
  - There are also exoergonic and neutral reactions
- Energy is provided from outside (e.g. as influx of energy carriers)
- In the following example
- Condensations are endoergonic
  - They require that at least one of the reactants is energized, while the catalyst is not
- Cleavages are energetically neutral

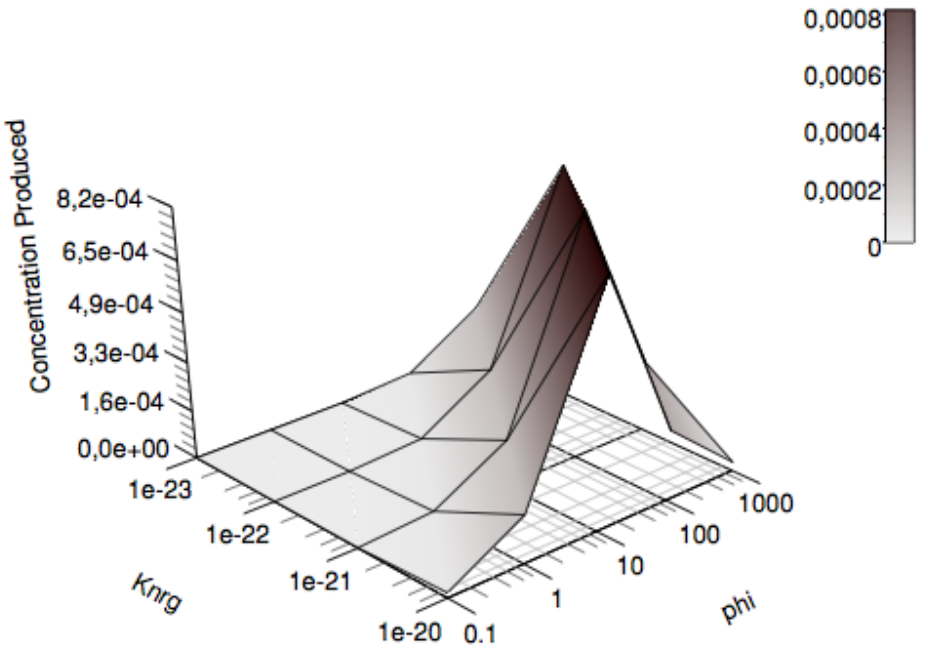
There is an optimal value of the energy flow

Number of species above
Threshold vs energy carrier
influx rate and kinetic
Constant of energization reaction



- Most molecules are produced in chains
- (left, in ACS and leaves; right, in chains)





Perspectives

- A major goal: simulate a population of protocells each one hosting a random reaction network
  - Structure –dependent probability of catalysis
  - Coupling between exergonic and endoergonic reactions
  - …
- Does heterogenity give rise to unexpected phenomena?
- A possible topic for a concerted action