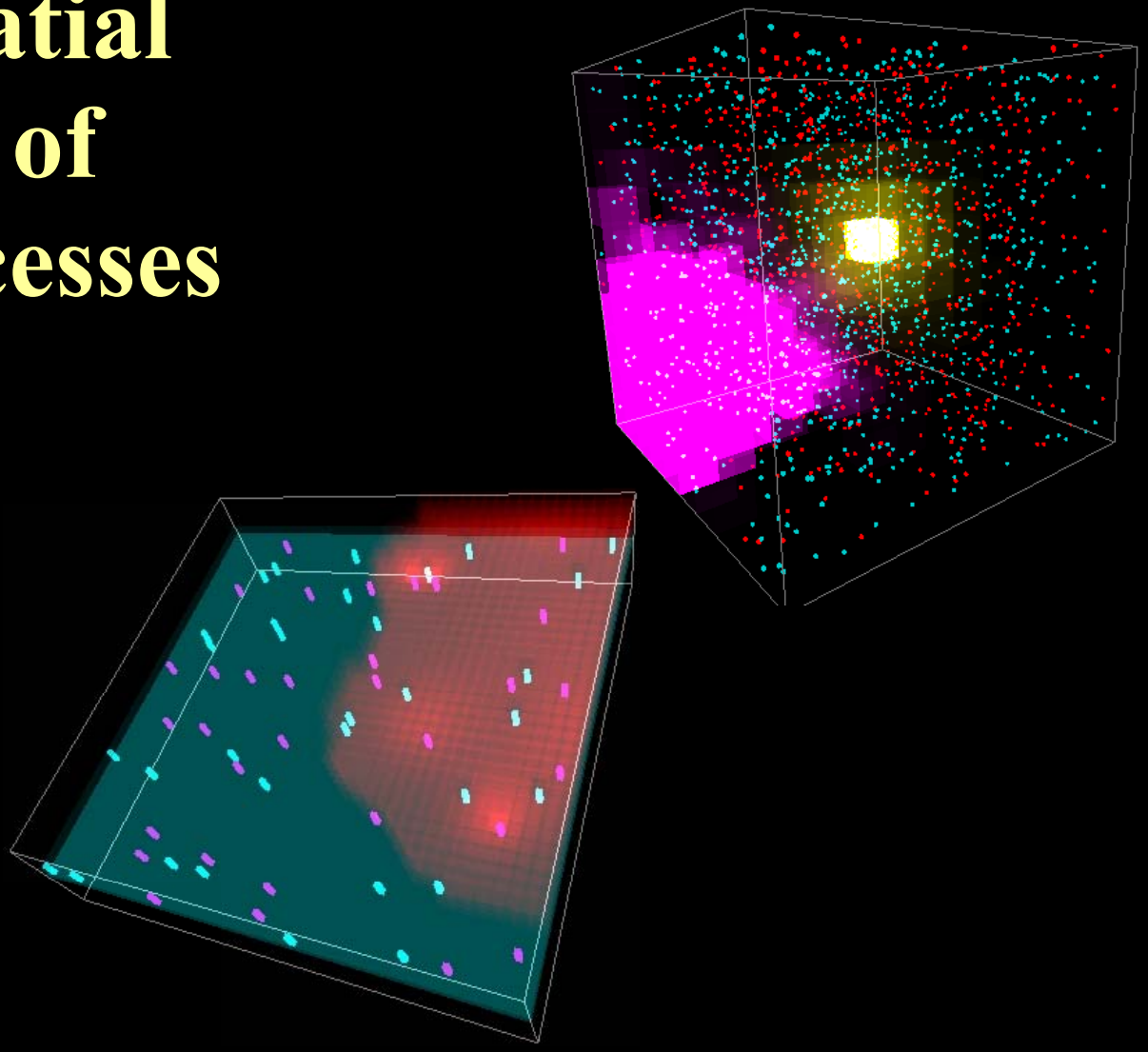
Temporal-spatial simulations of biological processes



John Parkinson
Molecular Structure and Function
Hospital for Sick Children
Department of Biochemistry
University of Toronto
Toronto
jparkin@sickkids.ca

Acknowledgements
Chris Sanford*
Matthew Yip
Carl White

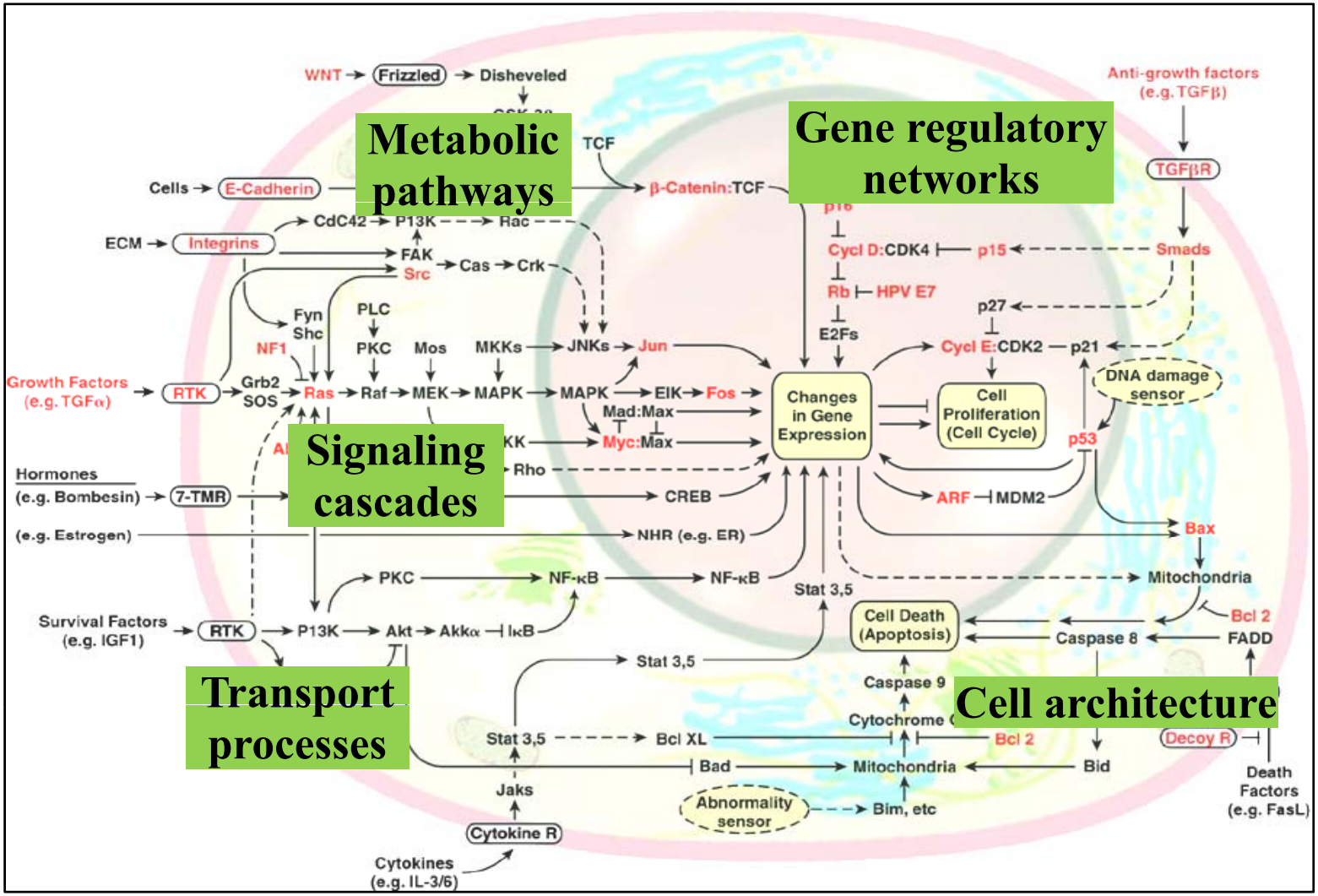
SickKids

NSERC CRSNG

Parkinson Lab
COMPUTATIONAL BIOLOGY RESEARCH AT THE HOSPITAL FOR SICK CHILDREN

The cell as a system

Cells are complex entities comprised of organized integrated biochemical pathways

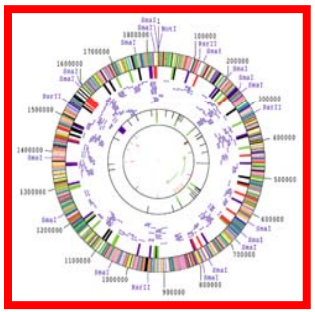


Translation
Transcription
Immune system
Hormone signaling
etc.

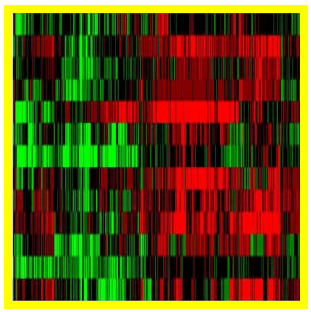
Studying how these pathways are coordinated is a central theme to understanding many aspects of biology, not least human disease

Integrating 'omics data – modelling cellular processes

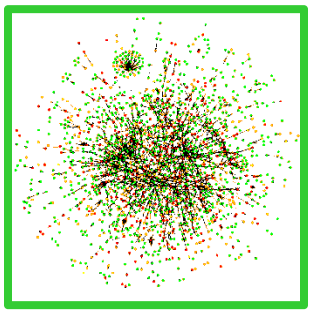
Advances in recent technologies are beginning to generate the data to help us unravel pathway organization and dynamics



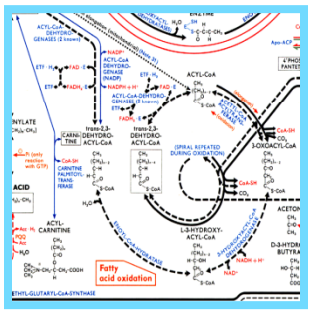
Lists of genes and gene products



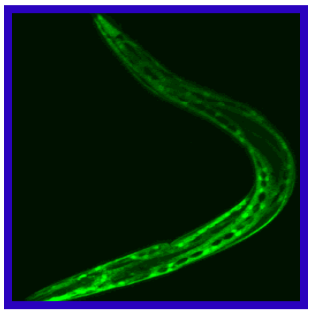
Relative abundance of transcripts / proteins



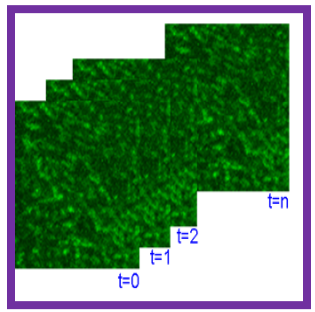
Interactions between proteins



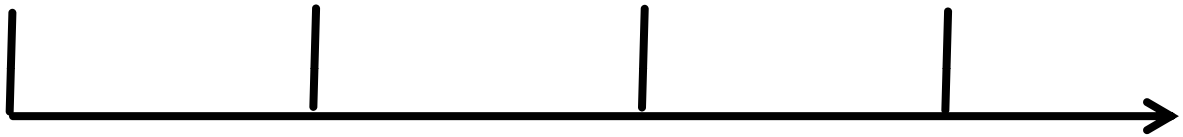
Organization of biochemical pathways



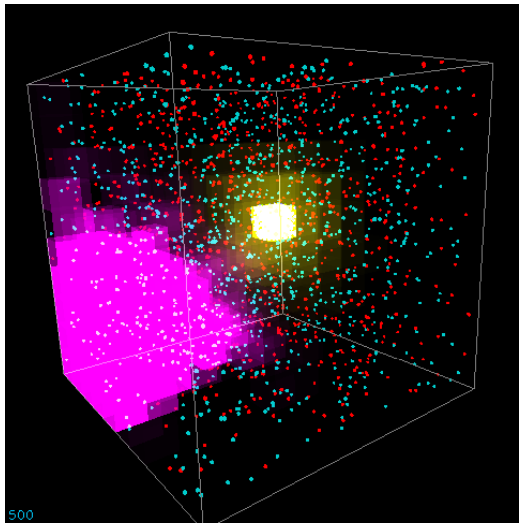
Localization of proteins in the cell



Dynamics of proteins in the cell



Need for computational simulations



The need for computational approaches

Even a system with a limited number of components can display complex - non-intuitive behavior

A classic example is the Lotka-Volterra's predator-prey model :

Prey

$$\frac{dx}{dt} = \alpha x - \beta xy$$

birth/death rate

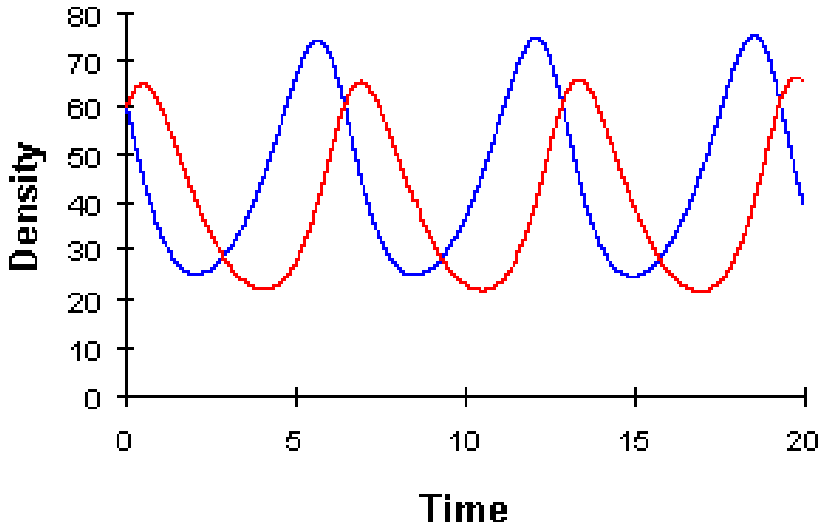
predation

Predator

$$\frac{dy}{dt} = \delta xy - \gamma y$$

competition

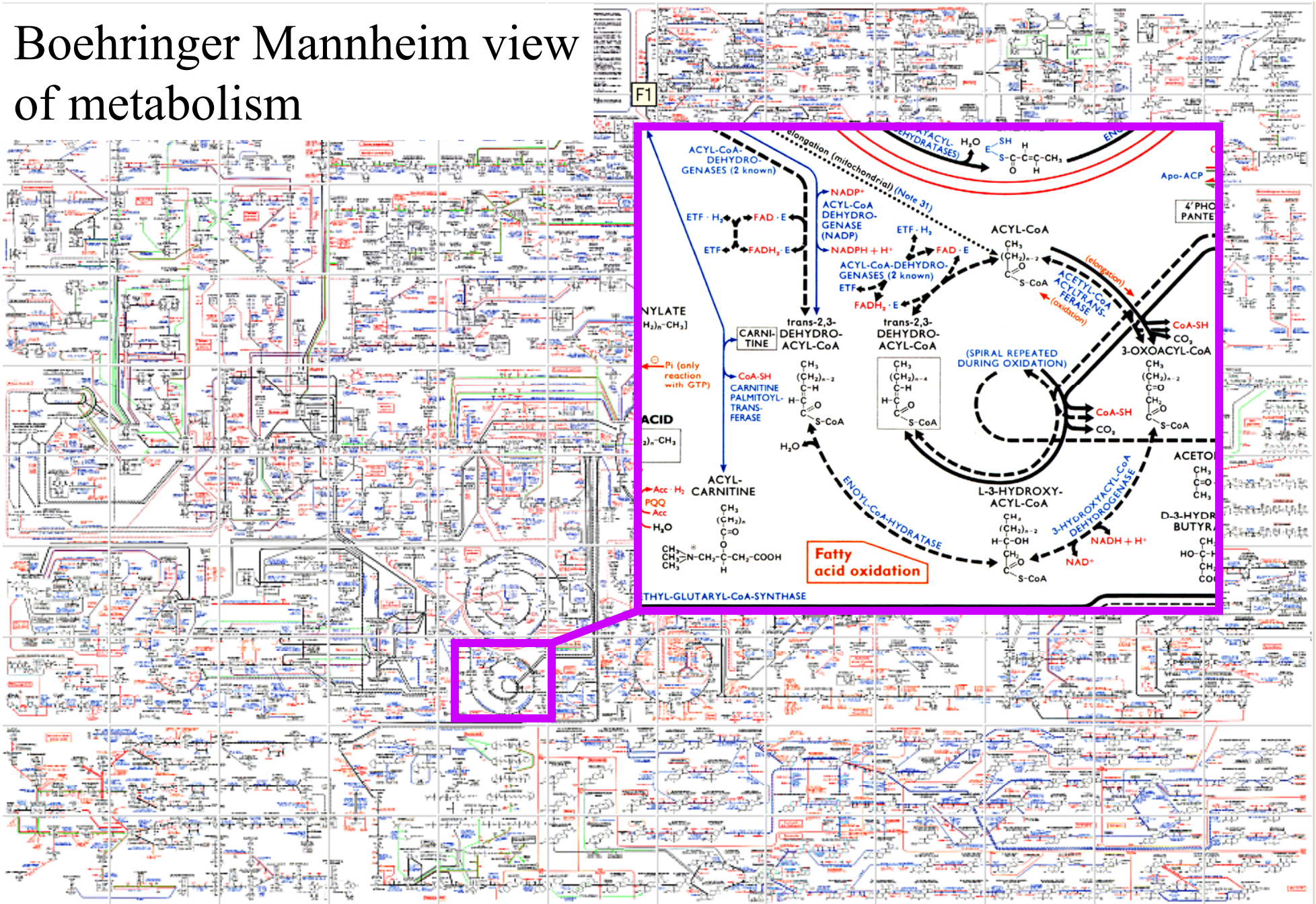
predation



In the model system, as the predator population is low the prey population increases and vice-versa. These dynamics continue in a cycle of growth and decline.

A single cellular system can involve many 1000's of interactions

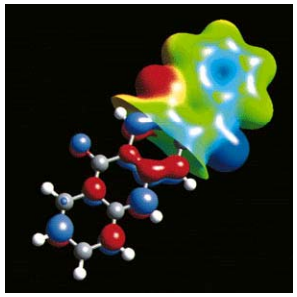
Boehringer Mannheim view of metabolism



Simulating at different levels of resolution

Å

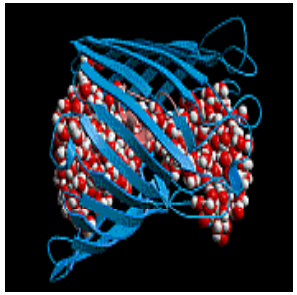
Atomic



Quantum mechanics

nm

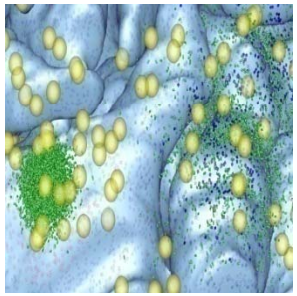
Molecular



Molecular dynamics

µm

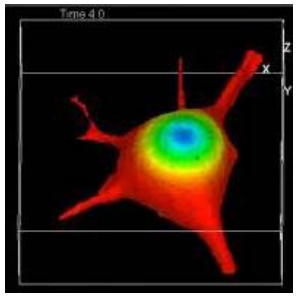
Mesoscale
Whole cells
Tissues



Mathematical models
- ODE's PDE's
- PI-calculus
- Petri Nets

mm

Organisms
Populations
Ecosystems



Stochastic simulations
- Cellular automata
- Brownian dynamics

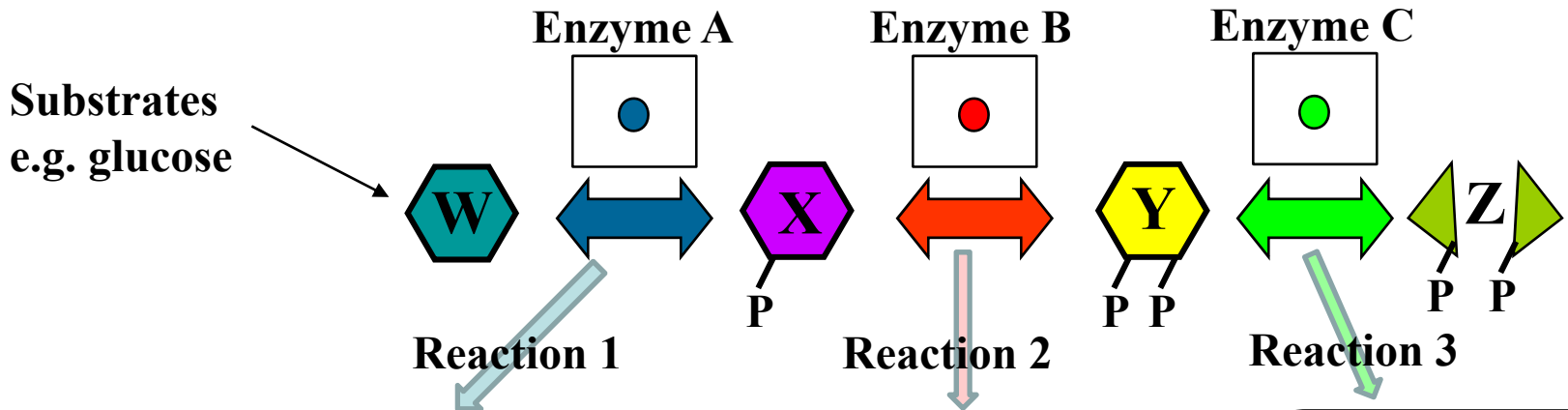
....

More accurate / More time consuming / fewer entities

Mathematical models of biochemical pathways

Several methods have been developed that attempt to model biochemical processes through integrated systems of ordinary and partial differential equations

e.g. Metabolic pathways



$$v = V_{max} \cdot \frac{\frac{[W]}{K_W}\left(1 - \frac{[X]/[W]}{K_{eq}}\right)}{1 + \frac{[W]}{K_W} + \frac{[X]}{K_X}}$$

$$v = V_{max} \cdot \frac{\frac{[X]}{K_X}\left(1 - \frac{[Y]/[X]}{K_{eq}}\right)}{1 + \frac{[X]}{K_X} + \frac{[Y]}{K_Y}}$$

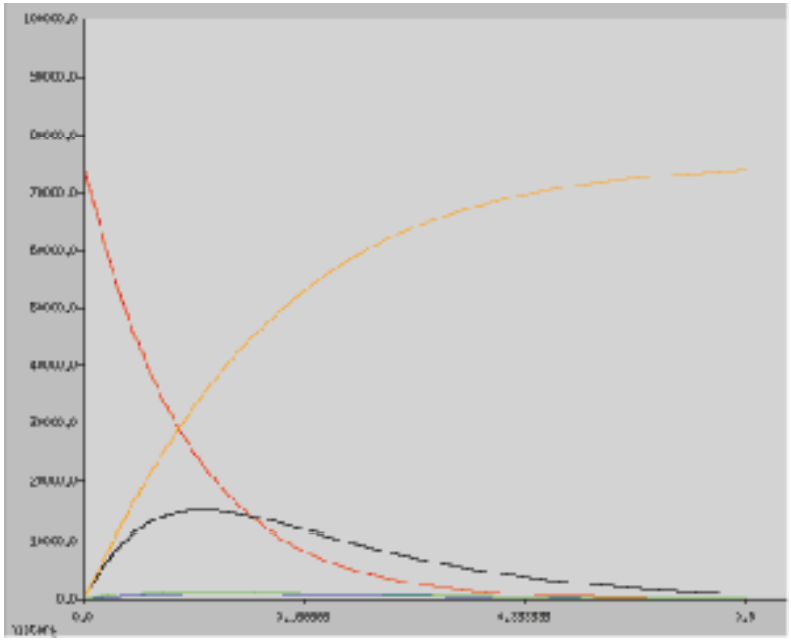
$$v = V_{max} \cdot \frac{\frac{[Y]}{K_Y}\left(1 - \frac{[Z]/[Y]}{K_{eq}}\right)}{1 + \frac{[Y]}{K_Y} + \frac{[Z]}{K_Z}}$$

Changes in substrate concentrations are then modelled through iteratively solving these equations over a series of discrete time steps

E-CELL

One of the first simulation platforms to adopt this approach is E-CELL - http://www.e-cell.org/ - now in version 3
Features a very nice GUI allowing the ready definition of pathway components (e.g. substrates and enzymes) and their relationships (e.g. enzyme kinetics, relative concentrations)

Originally applied to study mycoplasma (a minimal genome) with 127 genes and 85 small molecules. Output is in the form of graphs displaying changes in component concentration



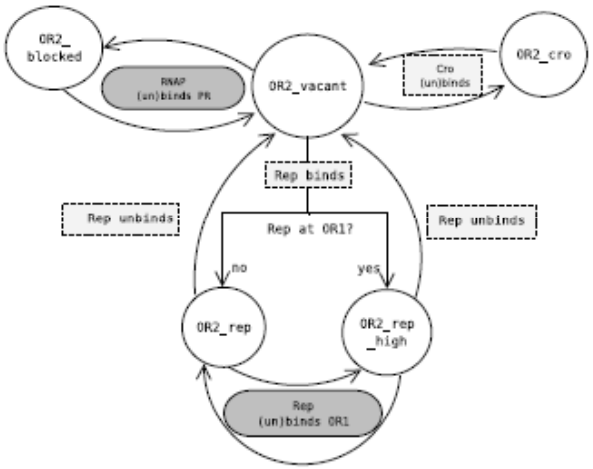
Using this system, it was possible to simulate effects such as glucose starvation to reveal conditions under which the cell could survive.

A major problem was that many enzyme kinetic parameters had to be estimated

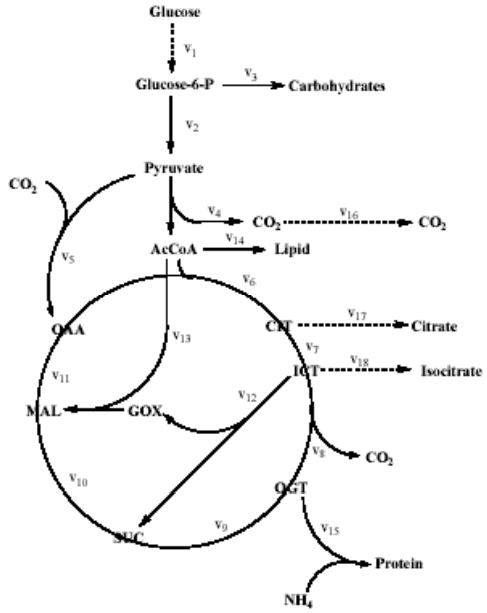
Mathematical Models

Over 100 different types of mesoscale cell simulators
While original models were mainly based on ODE's and PDE's (e.g. E-CELL), now more sophisticated mathematical models are being developed based on :

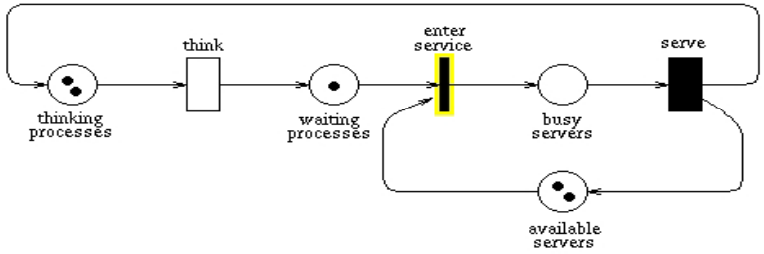
pi-calculus



Flux analysis



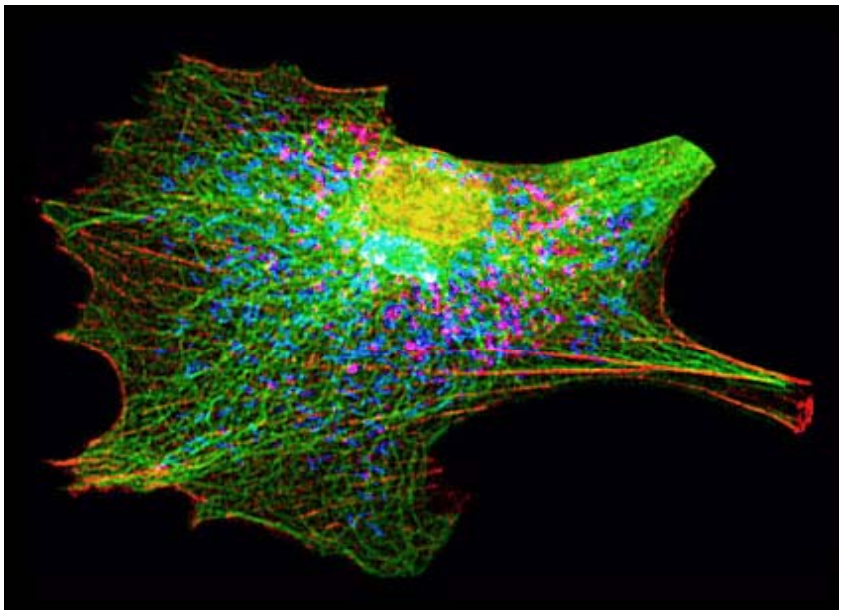
Petri Nets

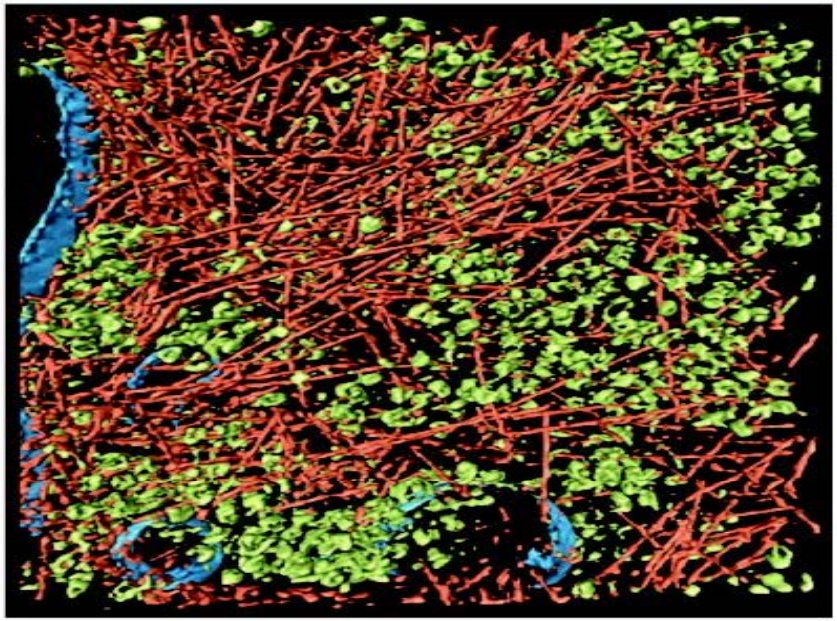


However these mathematical models treat the system as a well stirred mixture of components. As a result they fail to capture spatial and stochastic influences

Cells are not homogenously mixed fluid filled bags

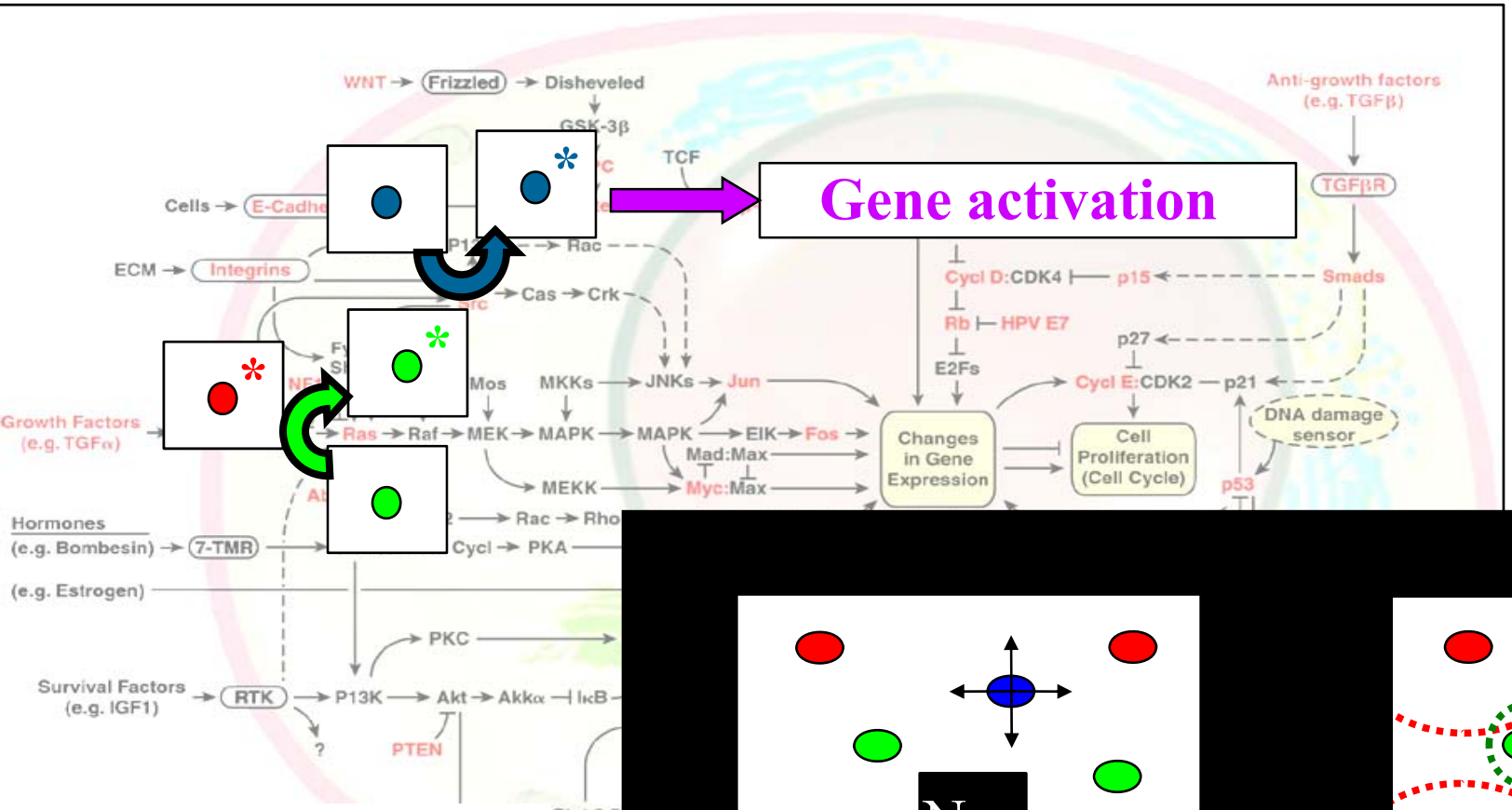
Compartmentalization
In addition to intracellular compartments such as nucleus, ER, golgi etc. The cytosol may be further subdivided into compartments by cytoskeletal elements which can impact the free diffusion of proteins



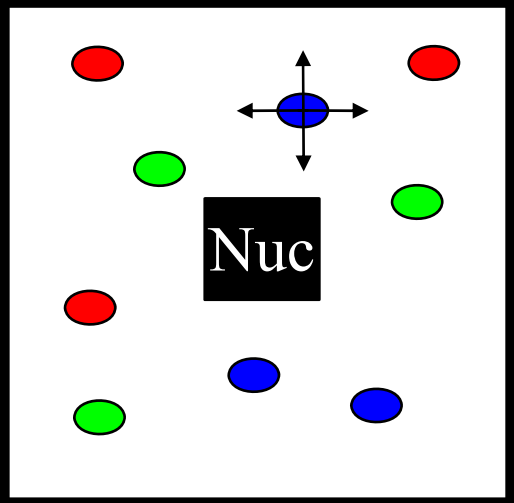


Molecular Crowding
Up to 40% of the total volume is physically occupied by macromolecules. Crowding can reduce the rate of diffusion by factors up to 10. It can also affect the stabilisation of more compact structures such as protein complexes and their ability to perform coordinated functions.

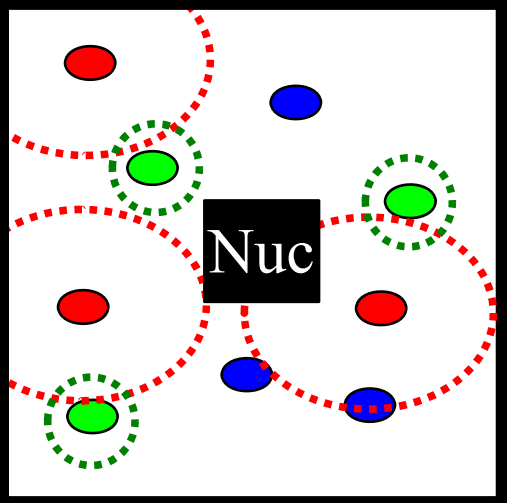
The influence of spatial structure: A simple signalling pathway



How does spatial organisation affect the behaviour of biochemical pathways ?



Unrestricted movement



Restricted movement

Discrete Physical Models

Spatial influences will give rise to local structure and potentially stochastic effects

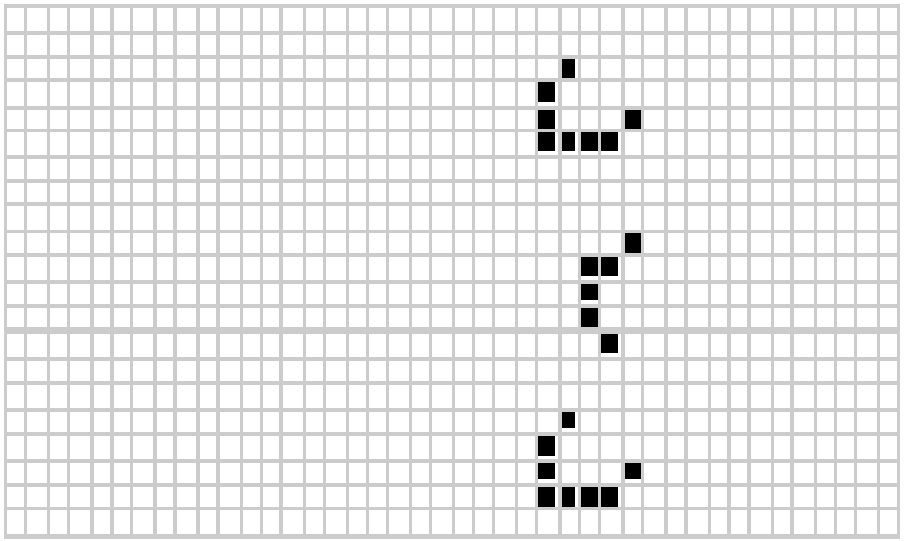
While mathematical models such as E-CELL can provide spatially defined structures, they have the following drawbacks

Deterministic: difficult to introduce stochastic behaviour

Not very flexible: very difficult to incorporate spatial factors

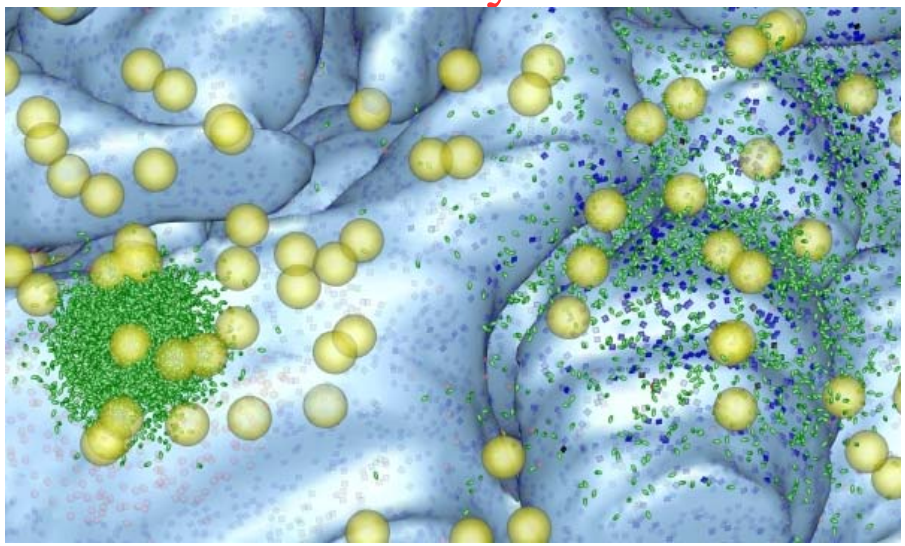
To explore spatial influences requires alternative approaches, e.g.:

Cellular Automata



Game of Life, SimCell, CyberCell

Brownian dynamics



Event driven (Stochsim, MesoRD)
Real time (MCell, Smoldyn)

Introducing Cell++

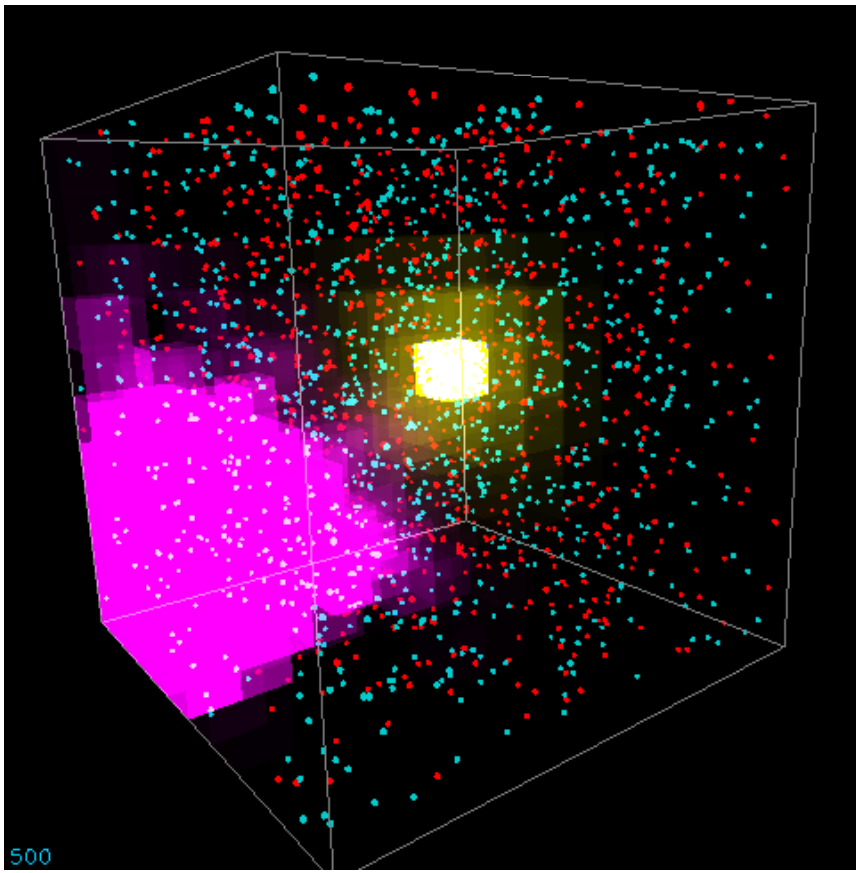
Drawbacks of current physical models methods

- Not very flexible - spatial factors not easily implemented
- Due to the restriction to a lattice - cellular automata models can lack accuracy
- Difficult to incorporate large numbers (10^6+) of small molecules in Brownian dynamics methods
- Visualisation often not be well integrated

So we built our own..Cell++

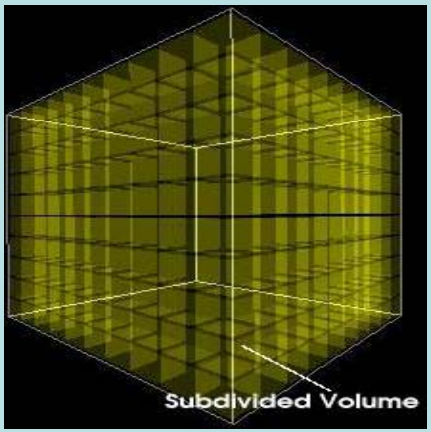
- A flexible modelling environment, written in C++, aimed at simulating basic cellular processes accounting for temporal / spatial factors
- Possesses a graphical mode for visualization (openGL) and text-only mode for batch simulations (harvesting of statistical data)

Intracellular signalling / metabolic pathways



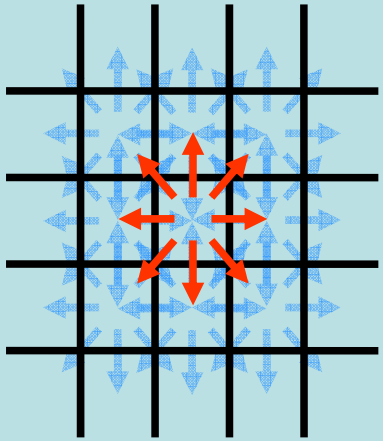
Sanford *et al.*, Bioinformatics 2006

Cell++ : A temporal spatial modelling tool



Cellular environments

- Cubic 3D lattice
- Lattice sites user defined (membrane/ nucleus etc.)

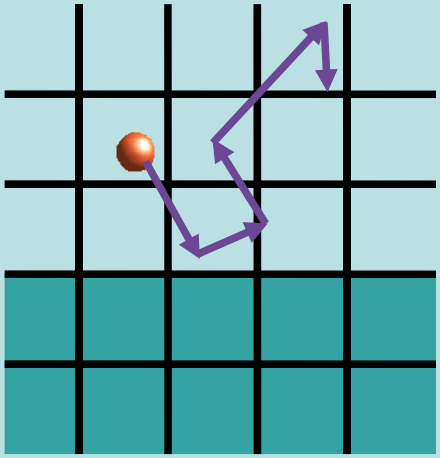


Small molecule diffusion

- ATP, Ca^{2+} etc.
- Concentrations defined at each lattice site
- Based on Euler method
- At each time step, a % of molecules move to adjacent sites

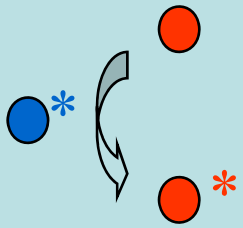
Large molecule movement

- Proteins
- Based on Brownian motion
- Off lattice random walks



Molecular interactions

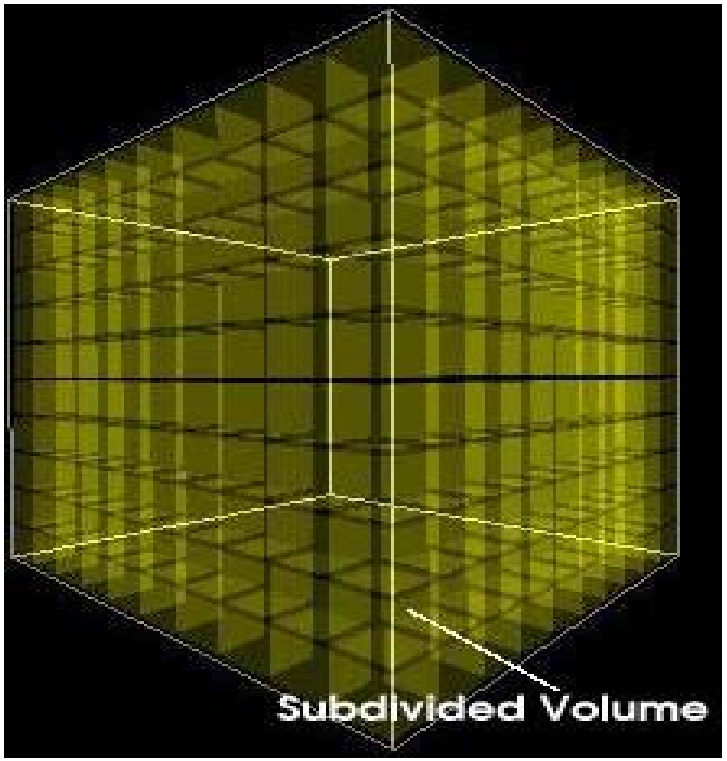
- Signal transduction, enzyme catalysis, transporters
- Simulation specific rules
- Deterministic (metabolism)
- Probabilistic (signal transduction)

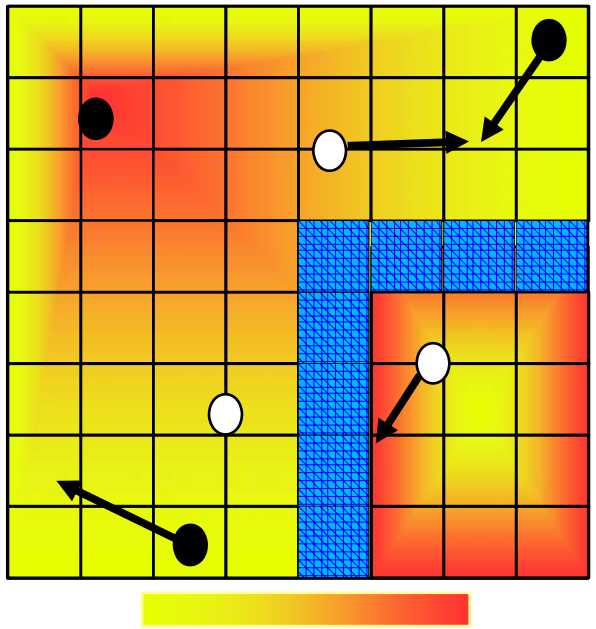


P = f (distance (● , ●*))

Available at http://www.compsysbio.org/CellSim

Cell++ - Underlying principles





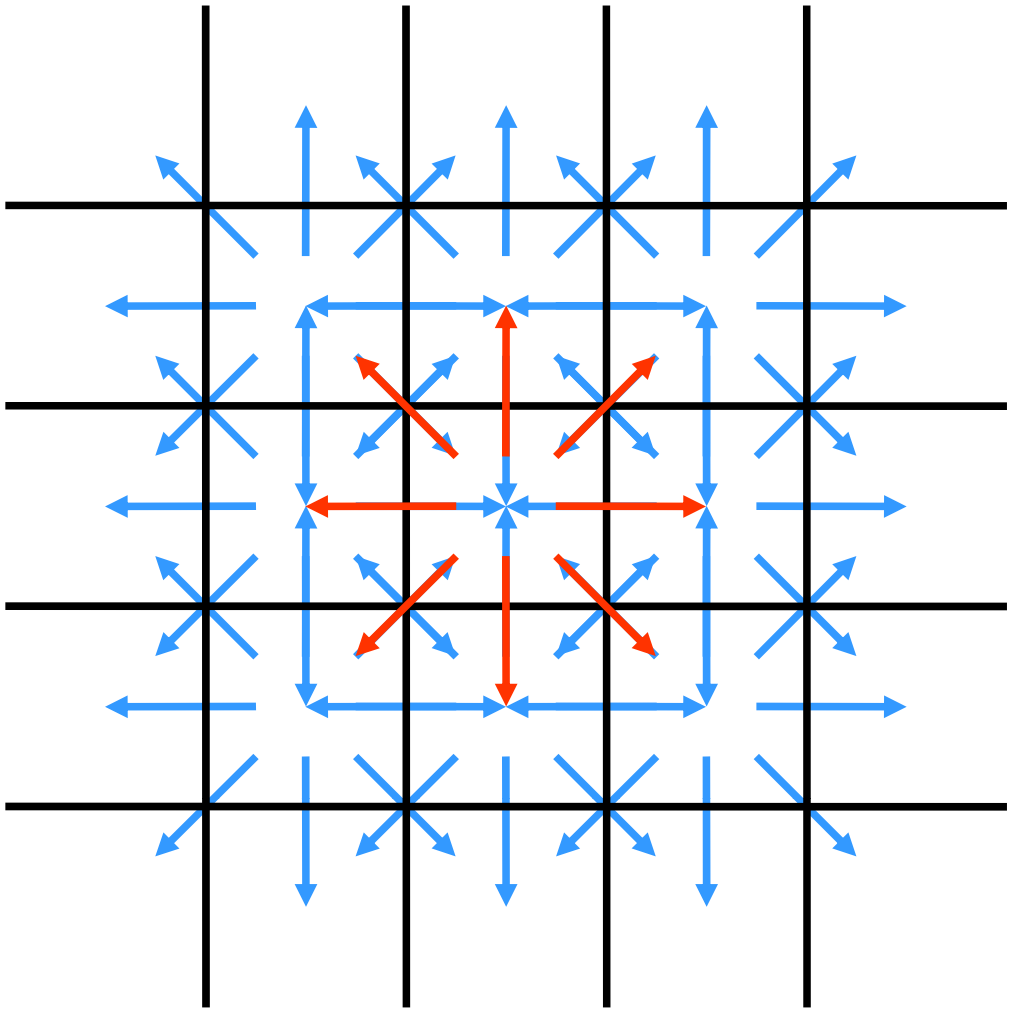
Relative concentration of e.g. Ca^{2+}

In Cell++, larger molecules such as enzymes are individually represented and move within a continuum.

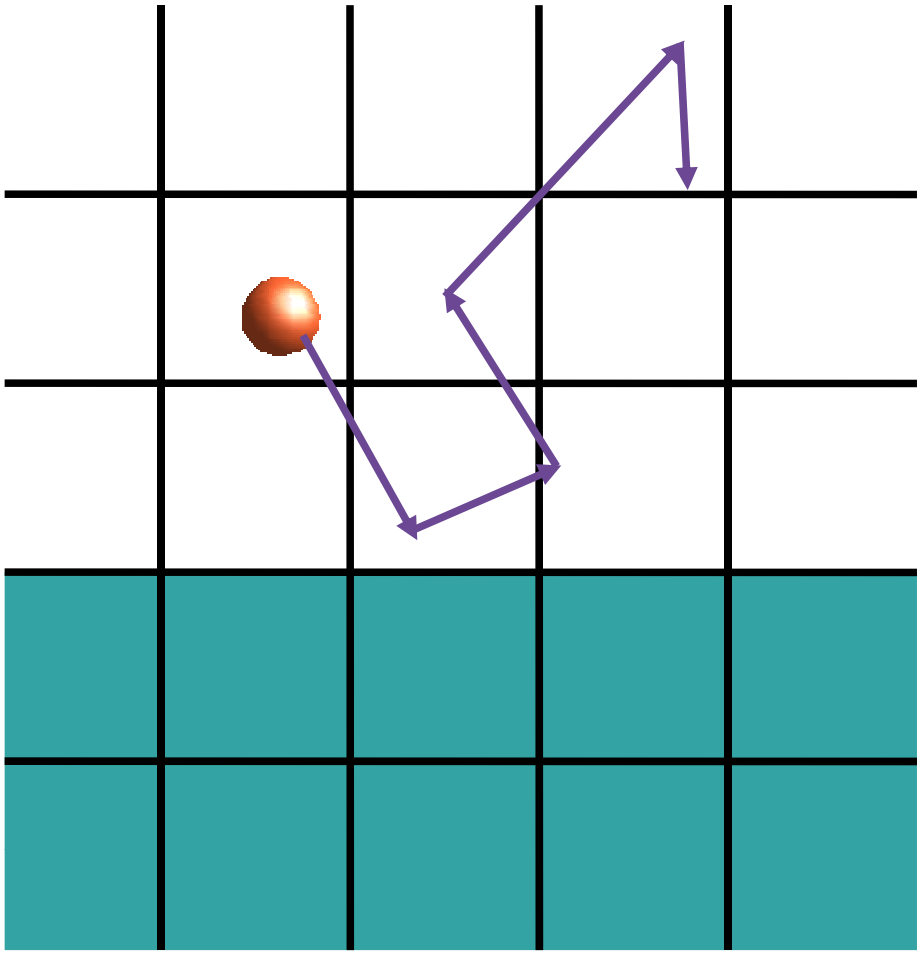
The continuum is superimposed upon a lattice, within which physical structure (membranes, for example) and concentrations of small molecules (such as calcium ions) are defined.

Cell++ - diffusion of small molecules

- 3D cubic lattice
- Concentrations of small molecules (substrates, Ca^{2+} etc.) represented within each lattice element
- 26 nearest neighbours
- Diffusion based on the Euler scheme
- At each time step a proportion of molecules at each lattice site diffuses to neighbouring sites
- While not the most accurate or fastest – provides a reasonable approximation
- Investigating alternative methods – Crank-Nicholson



Cell++ - diffusion of large molecules

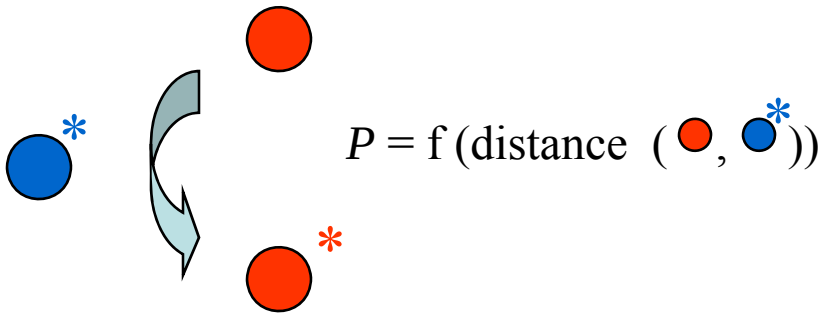


- Random walks on the lattice used to approximate Brownian motion
- Cell++ uses an off-lattice model with direction being determined by a randomly selected vector
- Distance determined according to a linear distribution reflecting the molecules diffusion coefficient
- Environments may affect behaviour
  - Slow
  - Inaccessible

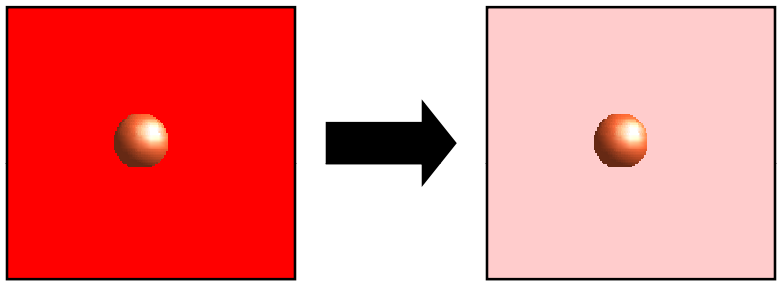
Cell++ - Molecular interactions

Interactions may be deterministic or probabilistic (based on e.g. distance and nature of interacting components)

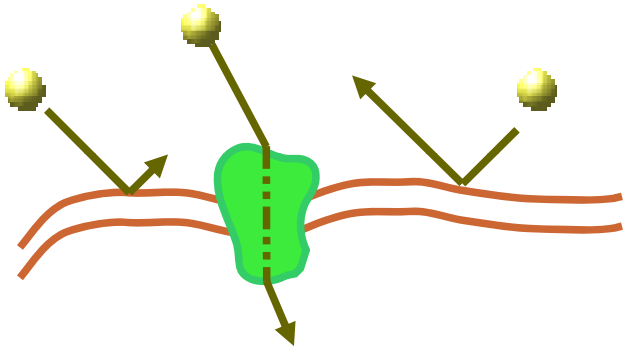
Signal transduction



Enzyme activity



Membrane transport

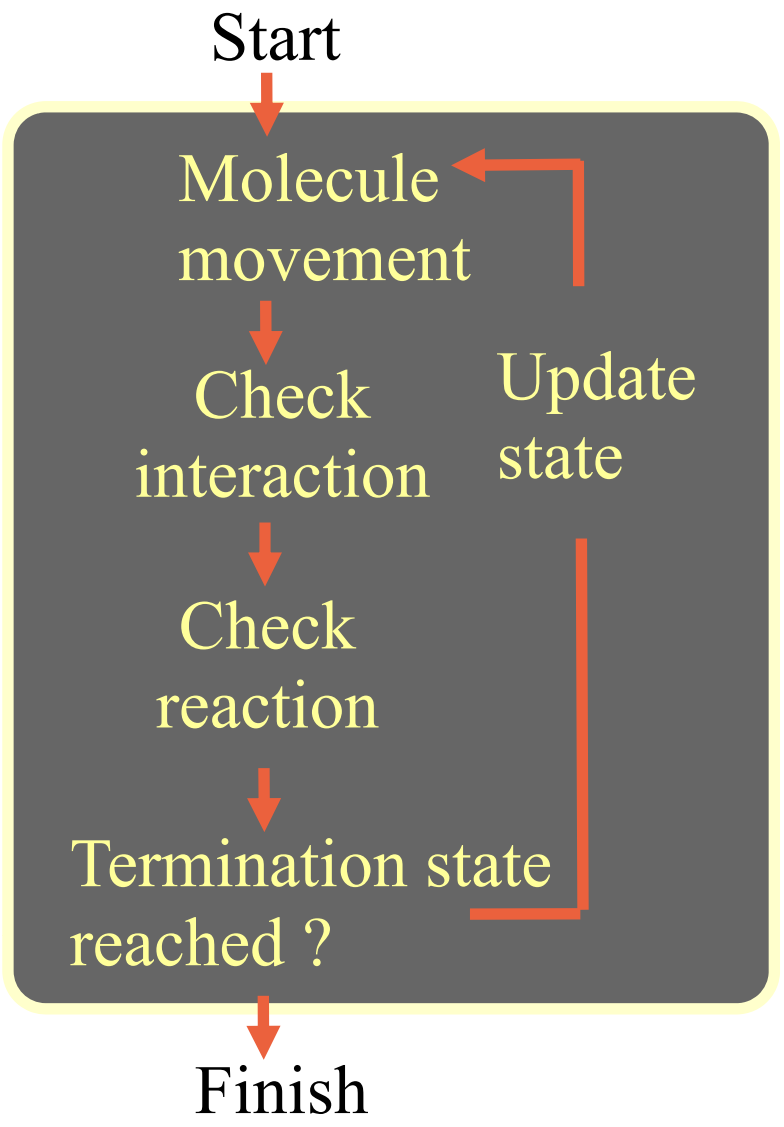


Cell++ - Methods overview

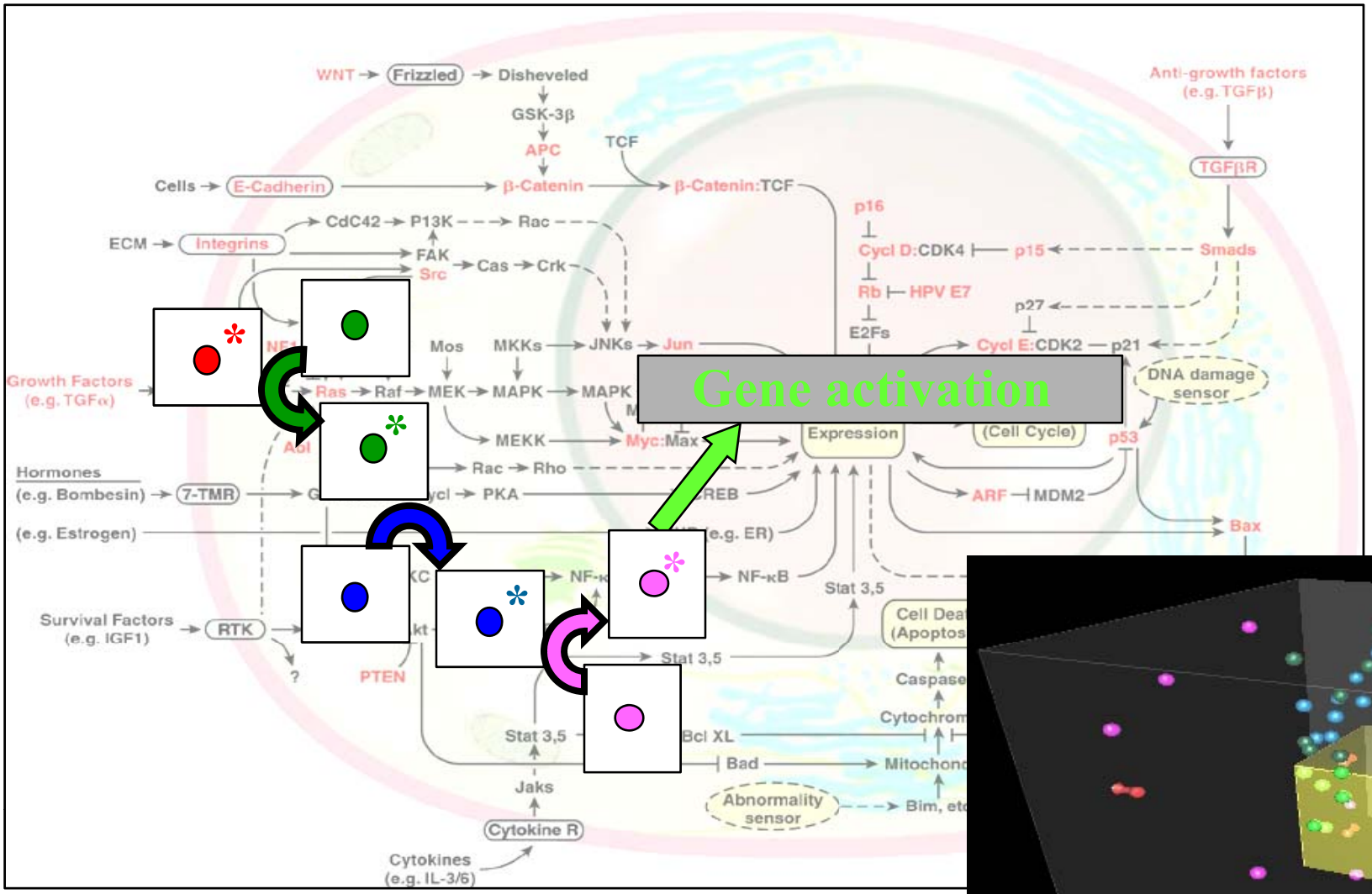
Initially the system is defined by a set of configuration files characterising:

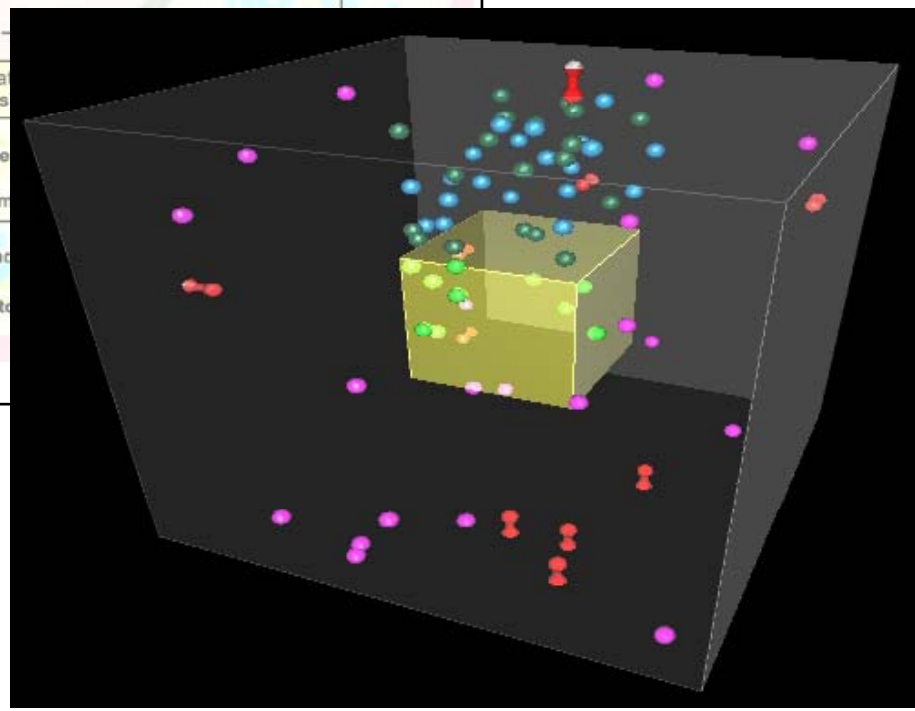
- environment (cell geography)
- number and types of discrete components
- concentration of small molecules
- rules of movement and interaction

Simulations proceed through a series of iterations, representing discrete time steps



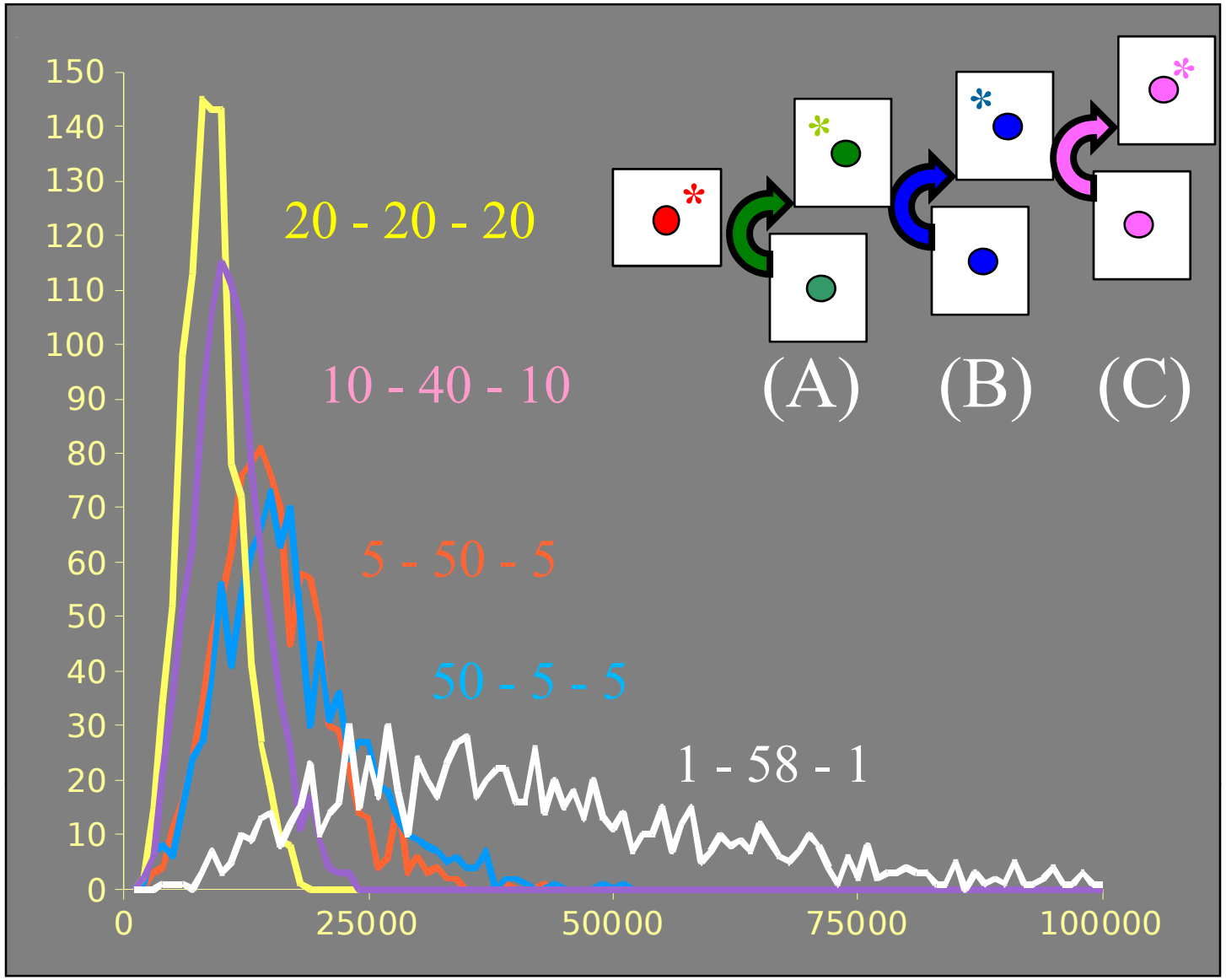
Exploring signalling pathways with Cell++





Effect of altering relative component concentrations

Frequency of simulations (1000 simulations per condition)

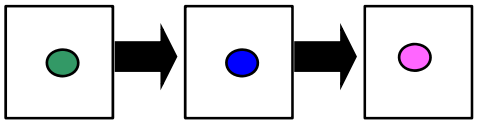


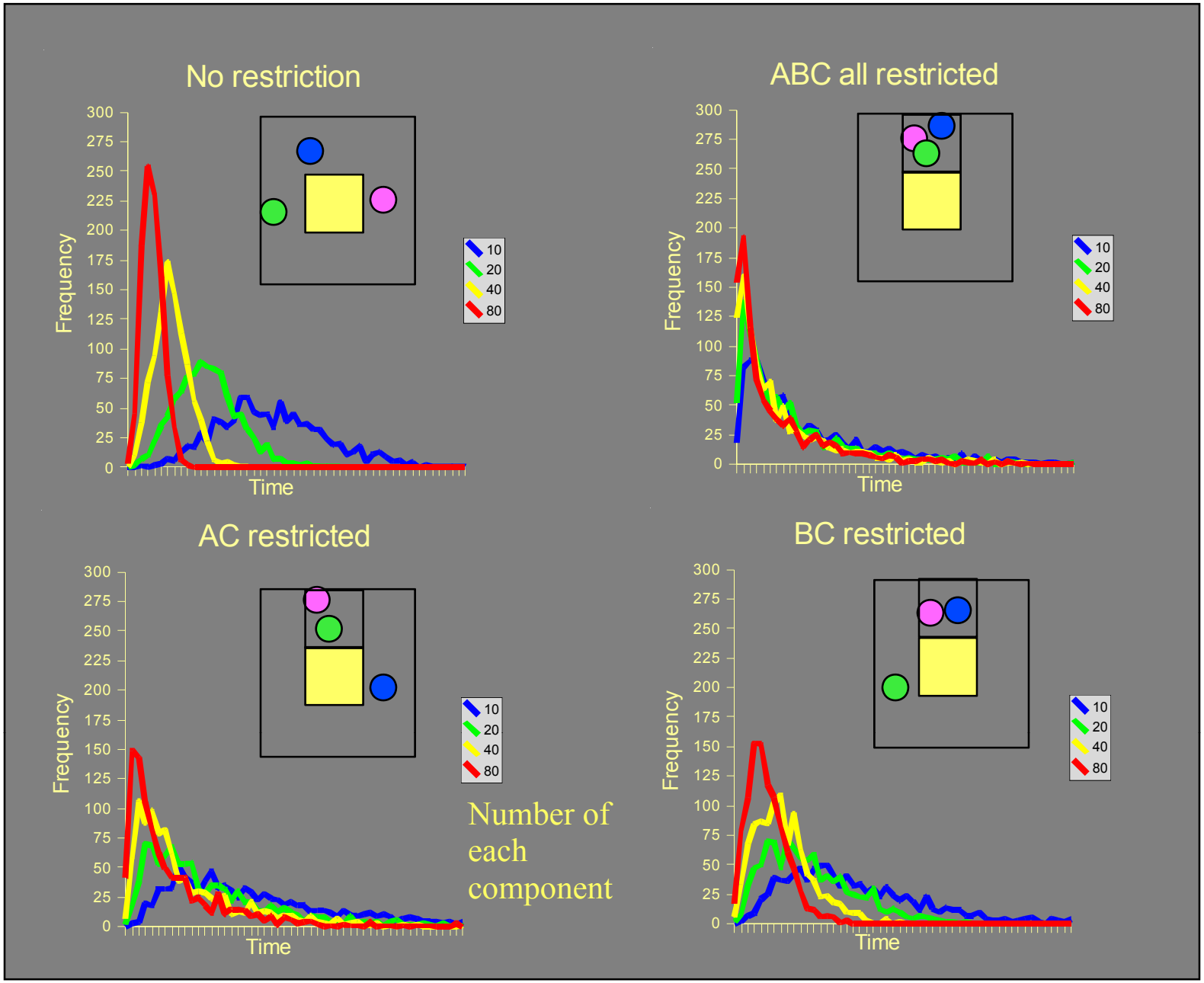
Time of signal transduction (iterations)

Effect of restricting component movement on time of signal transduction

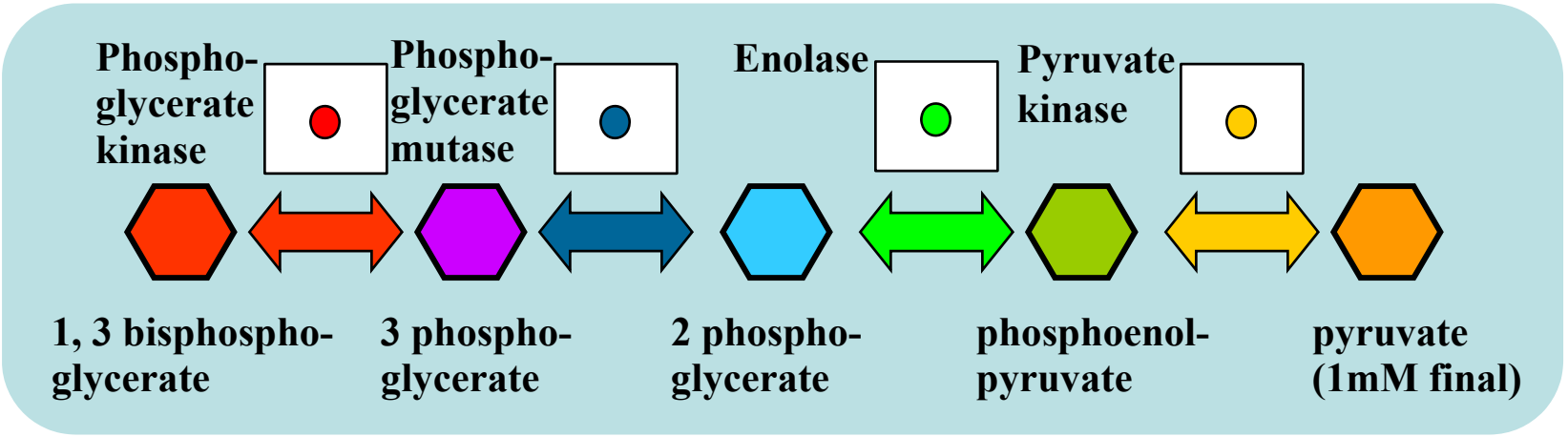
1000 simulations performed for each condition

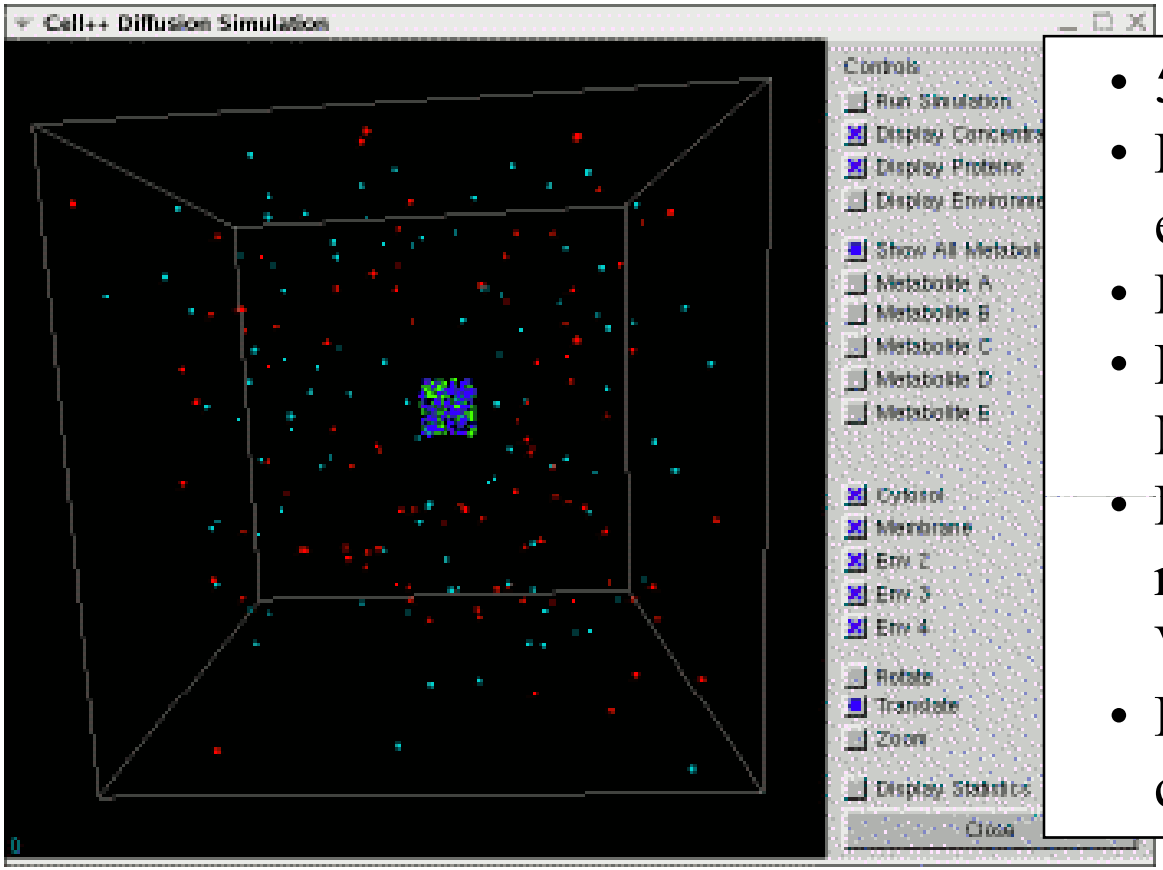
Colocalization leads to faster signal transduction, and also depends on pathway architecture





Modelling metabolic pathways: Glycolysis

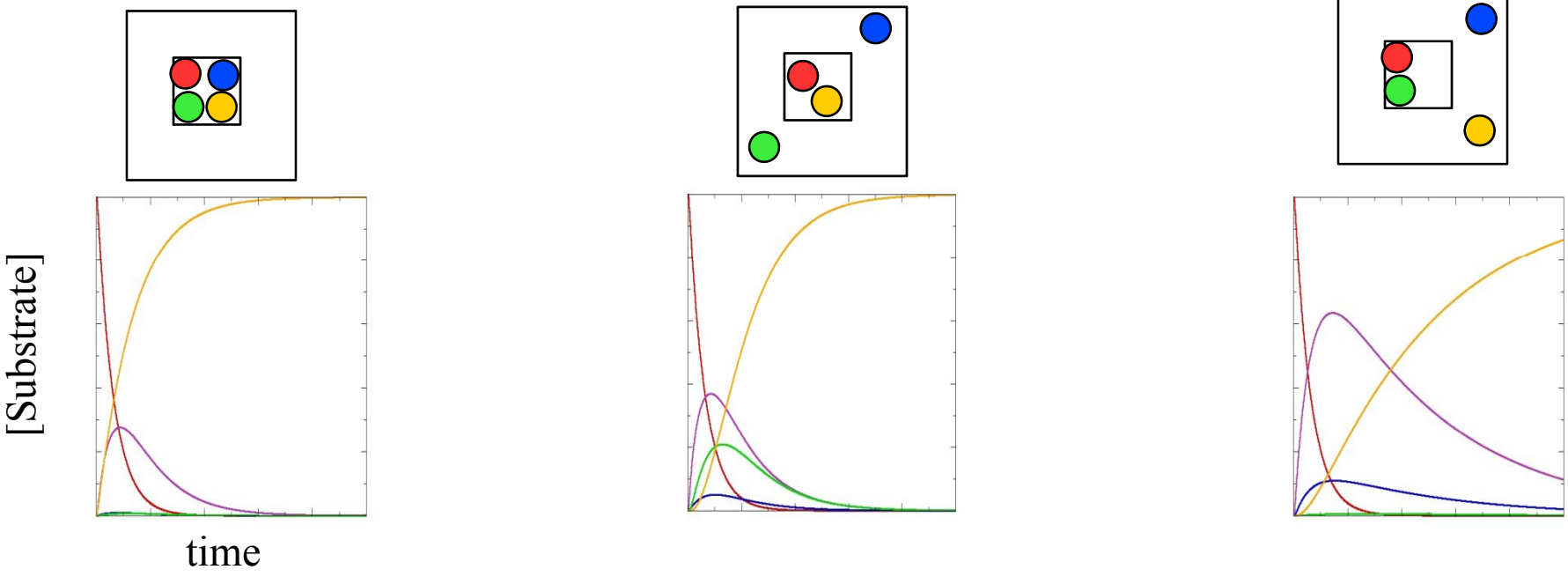




- 500 nm cubic lattice
- Four enzyme species (1000 of each)
- Five metabolites
- Reactions follow Michaelis-Menten kinetics
- Parameters obtained with reference to literature (K_m, V_{max}, diffusion coefficients)
- Different localization conditions

Spatial localisation reduces the build up of substrate intermediates

Enzyme localization

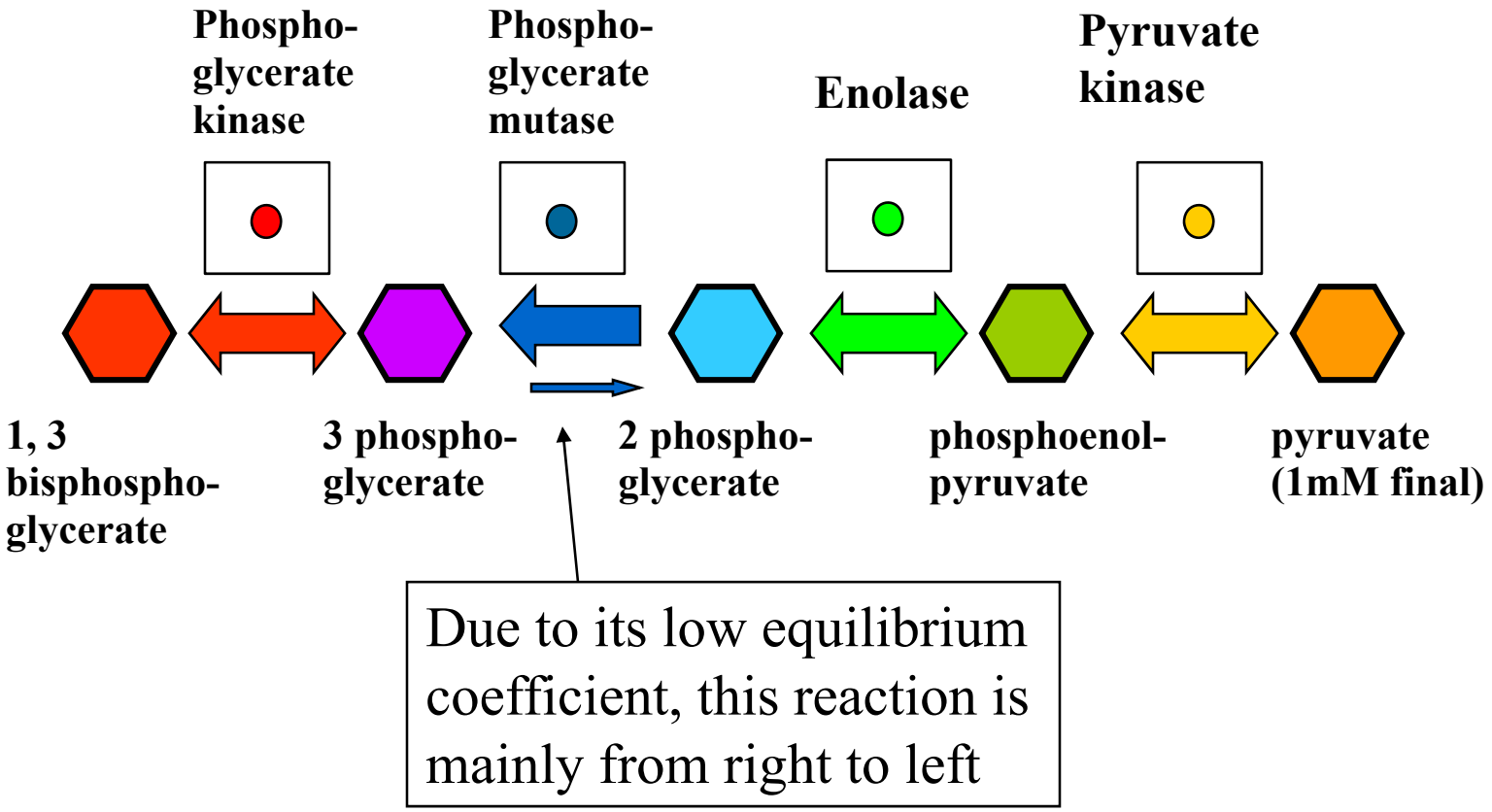


Co-localizing all four enzymes leads to the fast production of the final metabolite

The different locales of the orange and green enzymes result in accumulation of green substrate

The different locales of the blue and green enzymes result in accumulation of blue substrate AND the low equilibrium coefficient associated with the blue enzyme results in a slower reaction

Spatial localisation improves pathway efficiency

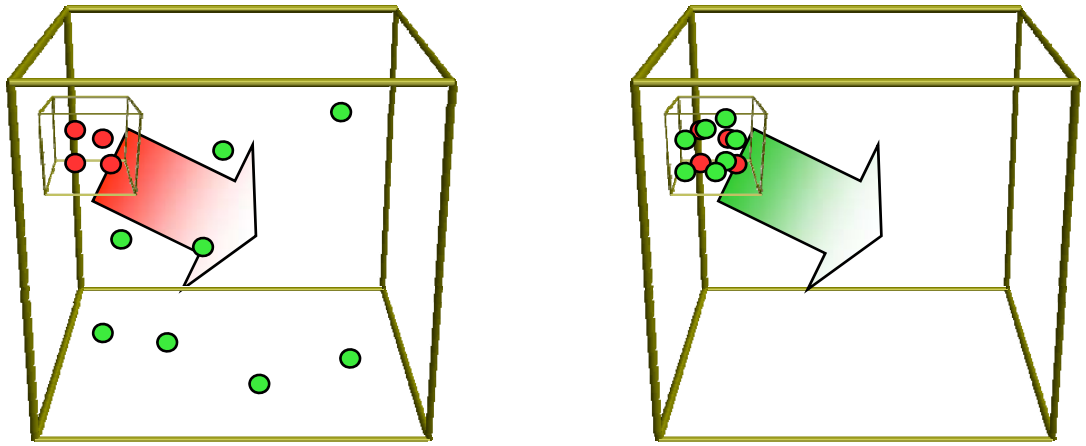


Co-localizing enolase ● with phosphoglycerate mutase ● allows the rapid removal of 2 phosphoglycerate ⬢ and production of phosphoenolpyruvate ⬢ ensuring the reaction proceeds rapidly from left to right

Spatial localisation permits metabolic channeling

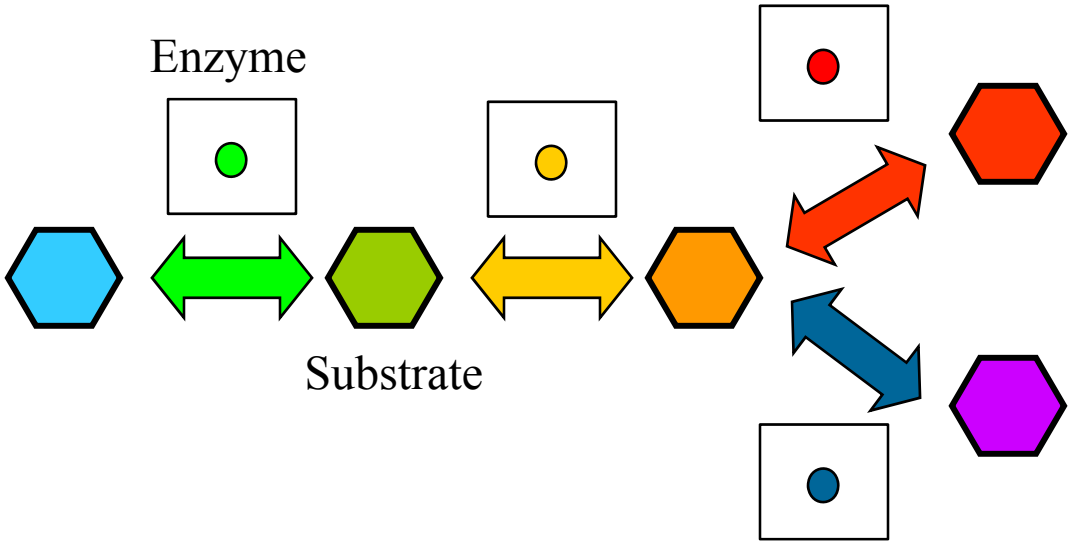
Metabolic channelling originally observed in plants where enzymes form complexes

May be exploited in industrial contexts / metabolic engineering
Increases pathway efficiency – esp. for reactions with low equilibrium coefficients
Prevents accumulation of possibly toxic intermediates

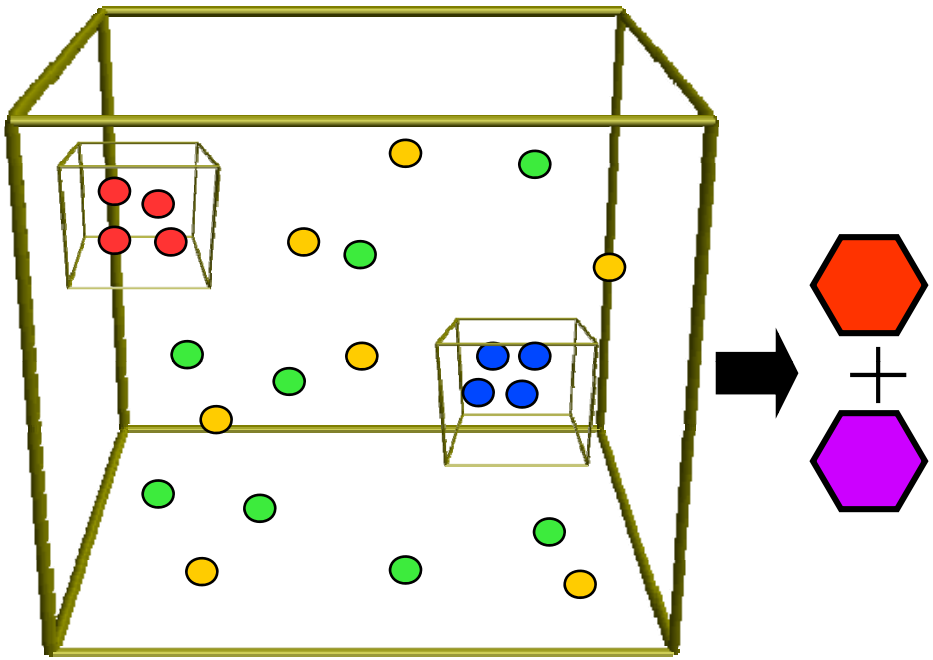


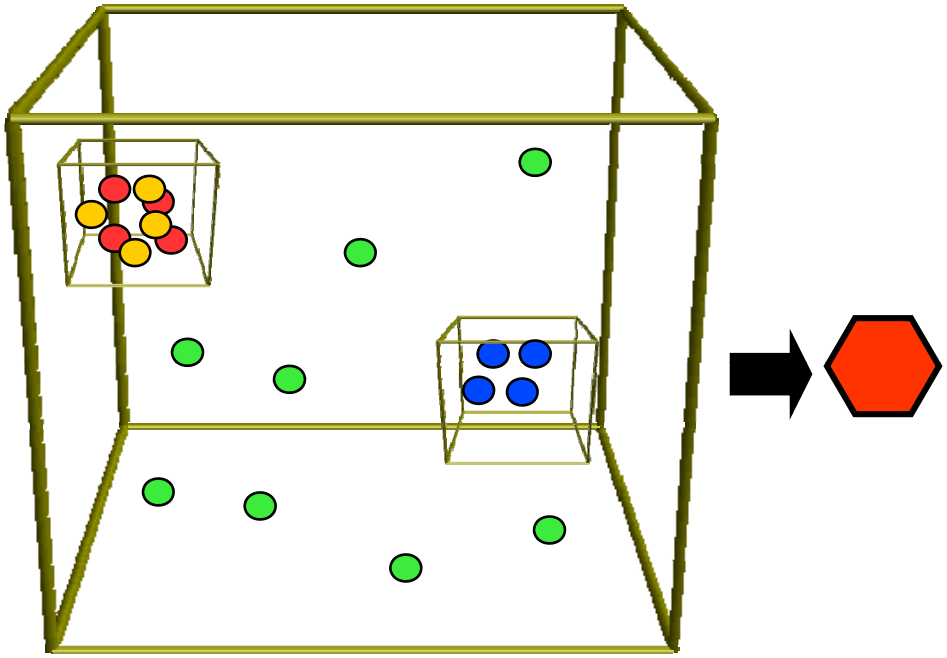
May control pathway crosstalk / re-direction of metabolism

Spatial localization as a means of controlling metabolic flux

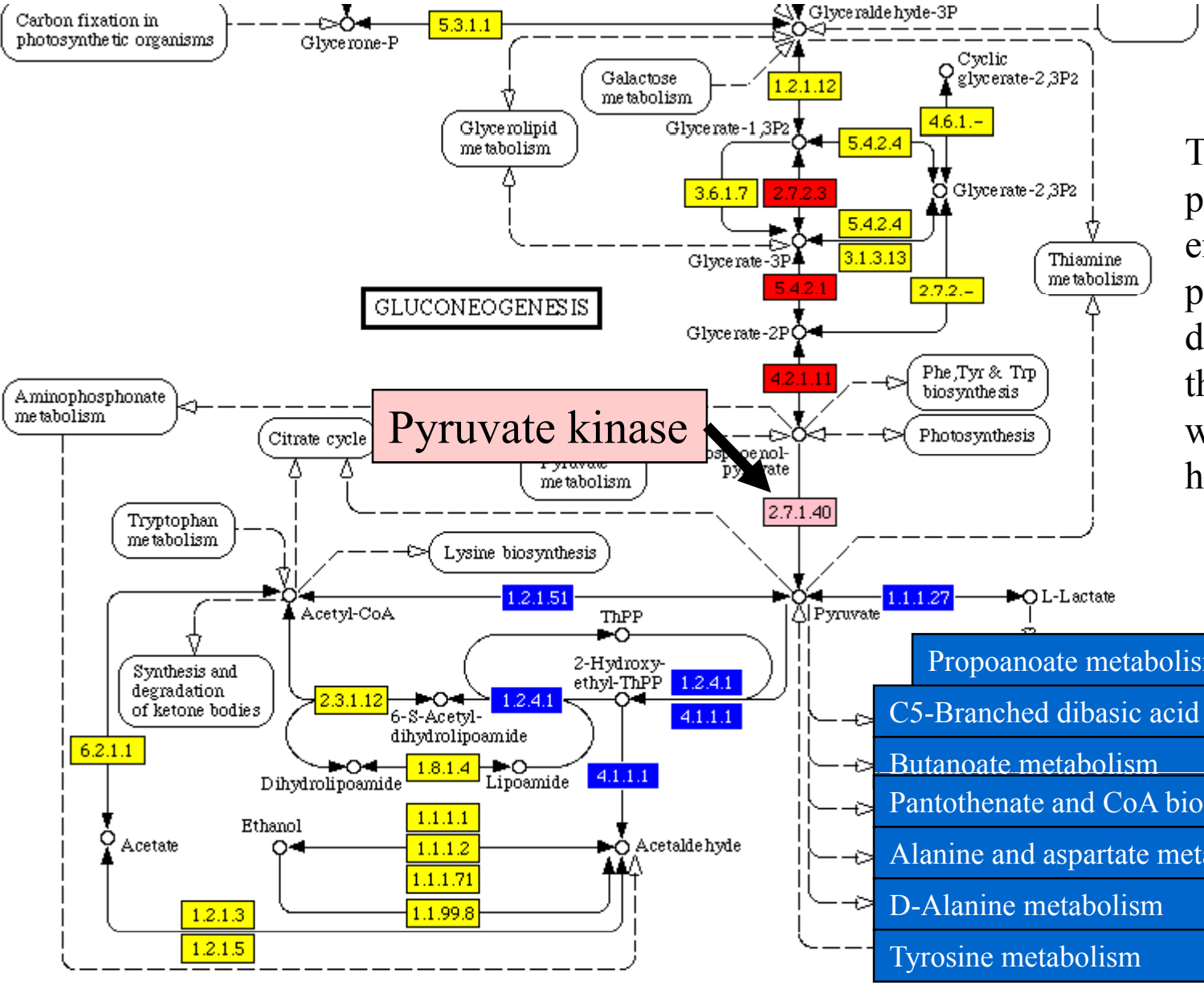


Depending upon the localization of a key enzyme (e.g. through binding to another protein), a cell may be able to rapidly switch its metabolism between pathways





Localization could control pathways



The spatial relationship of pyruvate kinase together with enzymes that directly act on pyruvate (in addition to the diffusability of pyruvate and the various enzyme kinetics) will determine which pathway has access to pyruvate

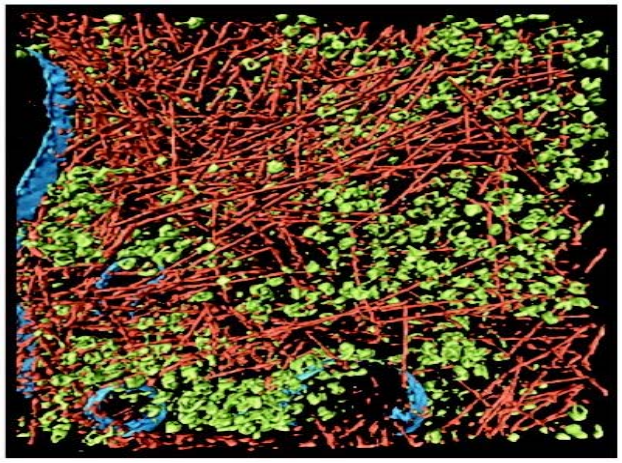
Advantages of Cell++

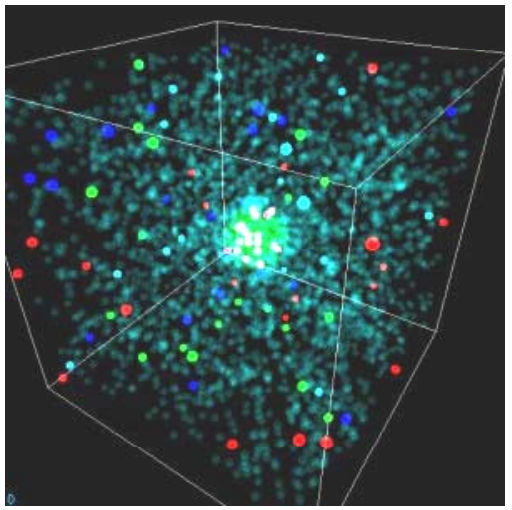
- Combines the power of cellular-automata and Brownian dynamics so that both large quantities of small molecules may be simulated along with a more discrete representation of larger molecules
- Three dimensional environment allows investigations into the influence of cellular environments on biochemical processes
- Visualization interface to view dynamic evolution of system under study. Allows users to gain insights that may otherwise not be intuitive
- Relatively fast – simulations can be performed in real time and may be readily parallelized allowing i) assessment of how stochastic a simulation is and ii) parameter scans
- Flexible – Cell++ has been applied to study kinase cascades, metabolic pathways, calcium waves and lipid raft mediated signalling

BUT.... Need more *in vivo* data: localization, diffusion, kinetic parameters **AND...**

Current limitations of Cell++ - I

Inability to accurately detect collisions (currently have a probability of interaction not collision). Phenomena such as molecular crowding can currently only be implemented through the occupation of lattice sites





Restricts accuracy of simulations. A better method would allow molecules to be treated as hard spheres (or other shapes) and identify collisions

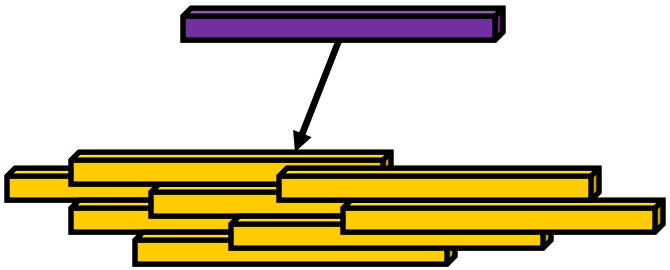
Problem: how to do this efficiently off-lattice for 10,000's of molecules ?

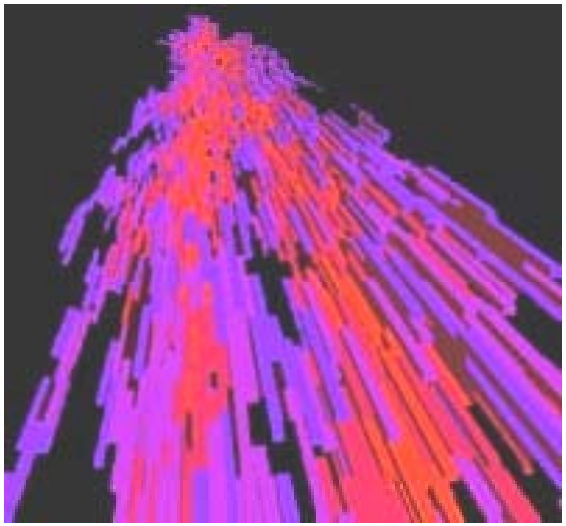
Current limitations of Cell++ - II

Accurate modelling of aggregation phenomena (e.g. microtubule assembly; lipid rafts)

While Cell++ allows the formation of complexes, again since all units are treated as point particles – the morphology of the complexes are ignored

Providing molecules with shapes allows the formation of aggregates that reflect their molecular architecture (e.g. treating collagen molecules as rods leads to the formation of fibrils)

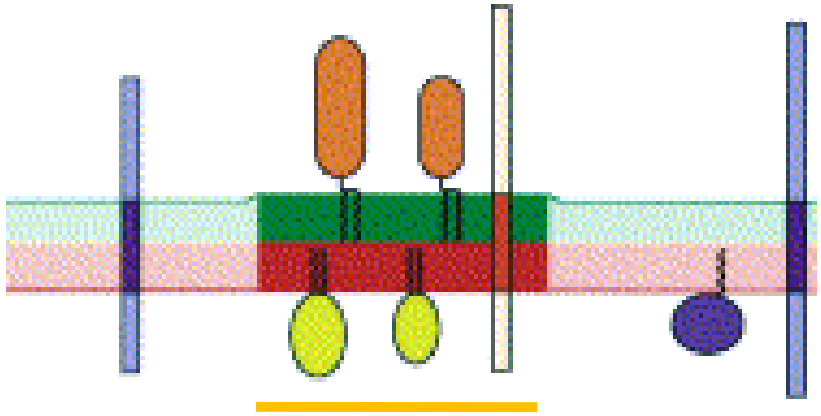




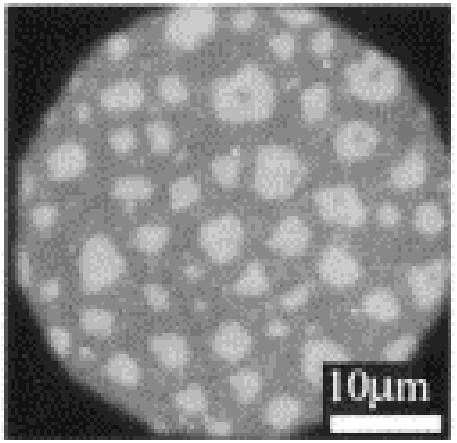
Problem: how to implement potentially complex rules of aggregation ?

Lipid-rafts

Lipid rafts are subdomains of the plasma membrane rich in cholestrol and glycosphingolipids first proposed in 1988 and still a contentious issue



Lipid raft rich in
sphingolipids/cholesterol

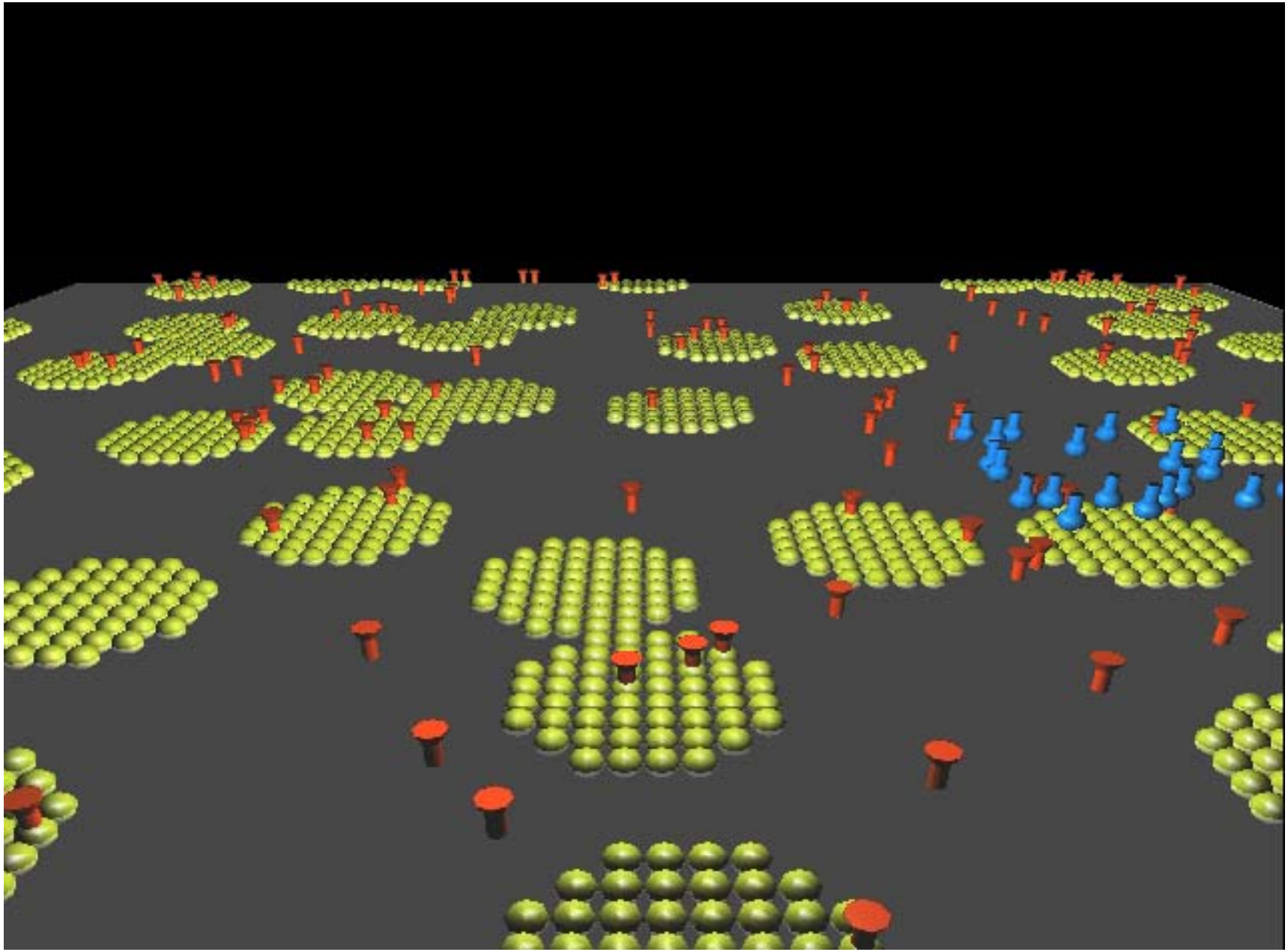


Certain proteins are thought to partition with lipid rafts:

glucosylphosphatidylinositol (GPI)-anchored proteins
doubly-acylated tyrosine kinases of the Src family

Lipid rafts may therefore help regulate signalling pathways

Cell++ and lipid-rafts

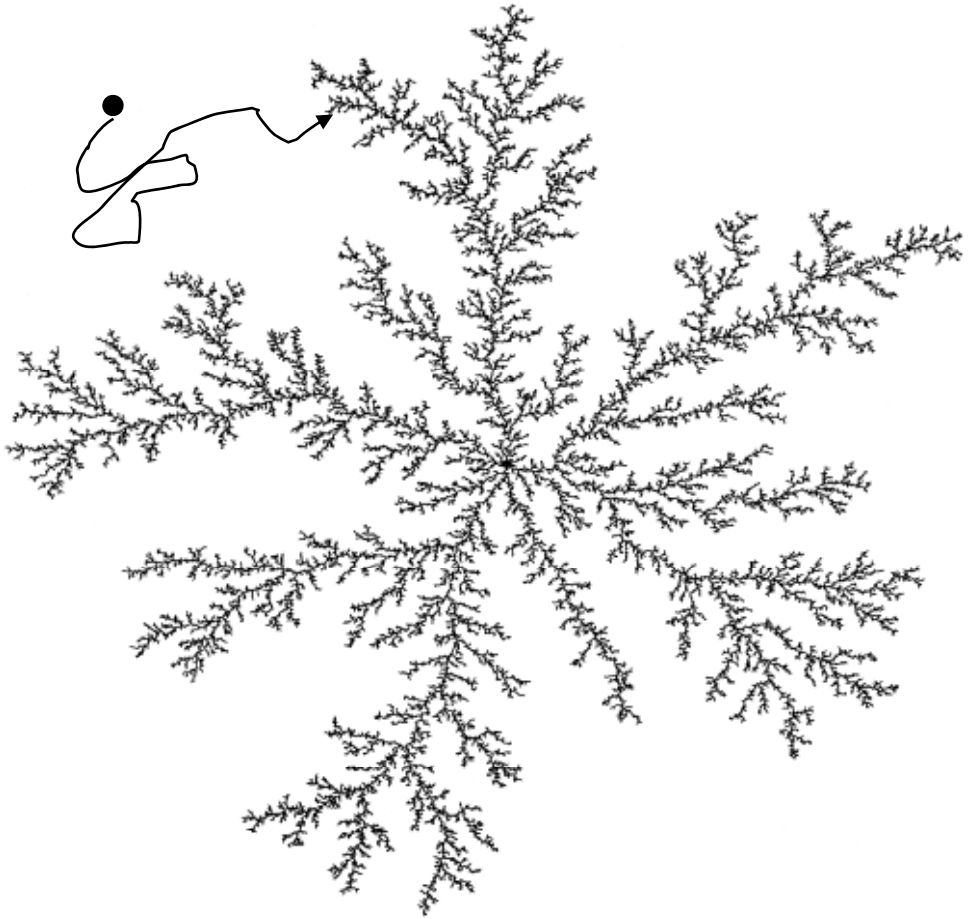


Modelling aggregation

What drives aggregation ? – for many systems (e.g. assembly of microtubules and actin filaments) the rules are relatively straightforward due to limited sites of interaction and precise rules of assembly.

Other systems are more stochastic and may not have such straightforward rules or may require other components

One useful model is diffusion limited aggregations which results in the formation of highly branched and fractal-like structures

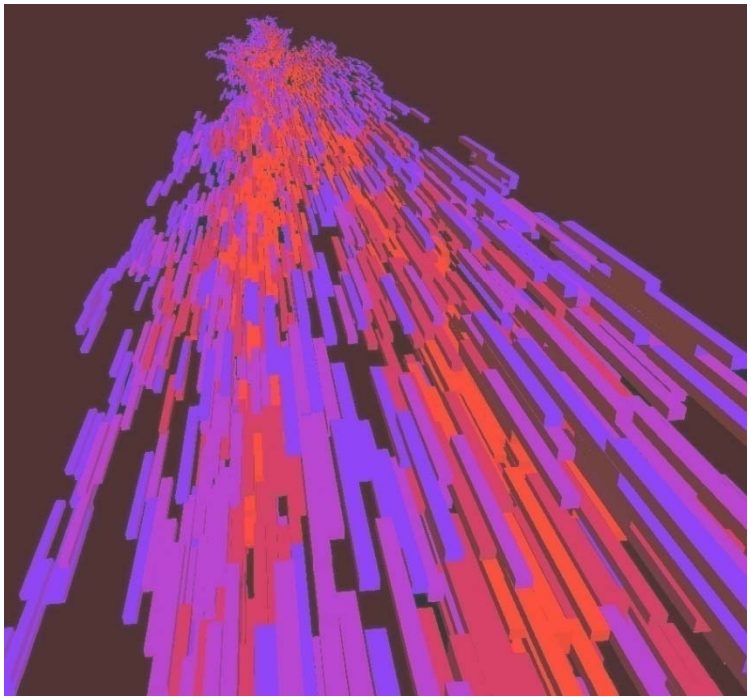


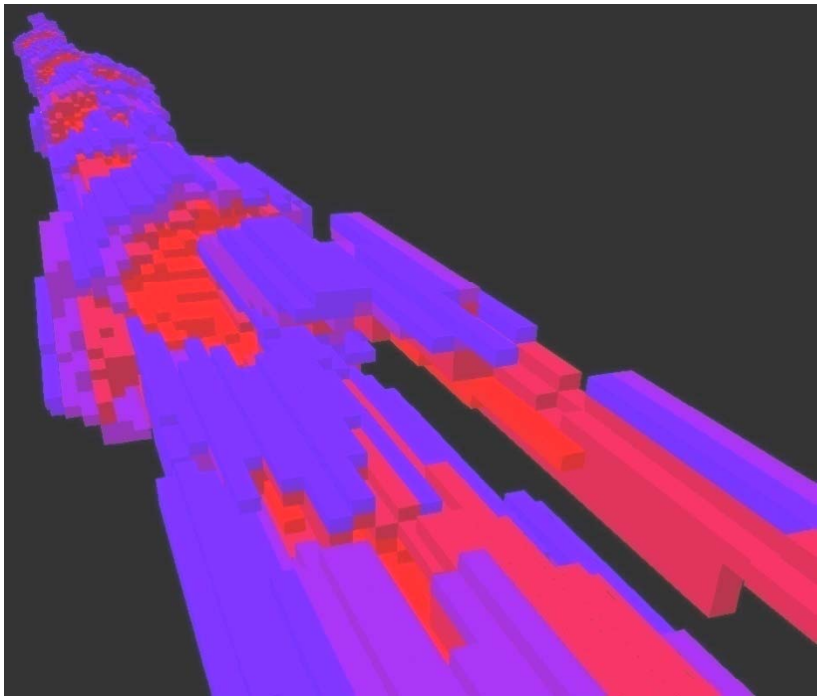
As a new particle is diffusing, it is difficult for it to penetrate deep within the aggregate before finding a suitable site for aggregation

Modelling aggregation

Many biological aggregates do not possess such highly branched structure and additional constraints control their morphology

e.g. self-assembly of collagen is dependant on entropic effects that attempt to minimize the surface area of an aggregate exposed to the surrounding solvent to form a compact fibril

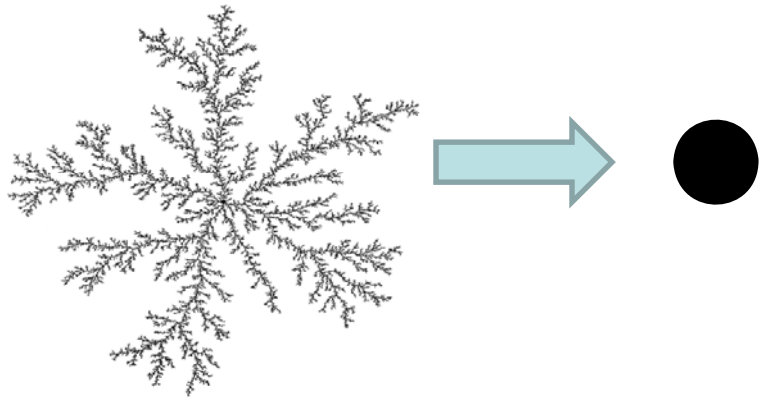




How can mesoscopic models account for these entropic effects ?

Modelling aggregation – minimization of exposed surface

In accounting for entropic effects, we also need to account for the reversibility of the aggregation events. This can be achieved using Monte-Carlo methods. During aggregation (or after aggregation) particles may aggregate (or change positions) with a probability that reflects the relative change in energy using the Boltzmann function



$$p = \min (1, e^{-\Delta E/kT}) \qquad [1]$$

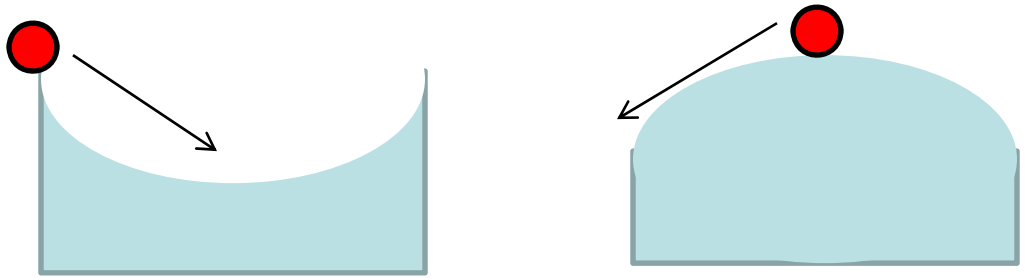
This energy term can include both local terms (e.g. number of direct interactions) and a non-local term reflecting solvation - surface curvature can be used as a good approximation (related to Gibbs-Thomson effect)

$$\Delta E = \Delta E_l + \Delta E_c \qquad [2]$$

$$\Delta E_c = \Gamma(1/R_i - 1/R_d) \qquad [3]$$

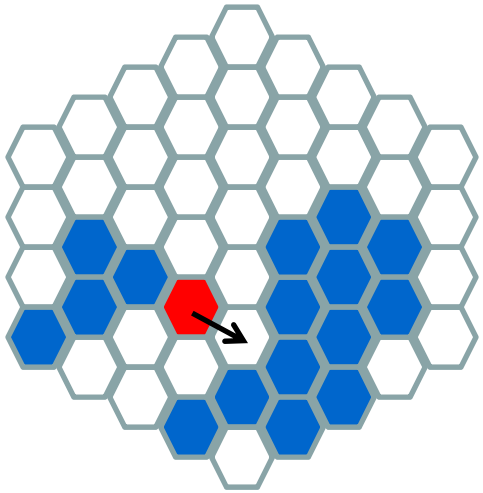
where R_i and R_d are terms reflecting the surface curvature before and after the particle moves

eq. [3] provides a method for minimizing solvent exposed surface area

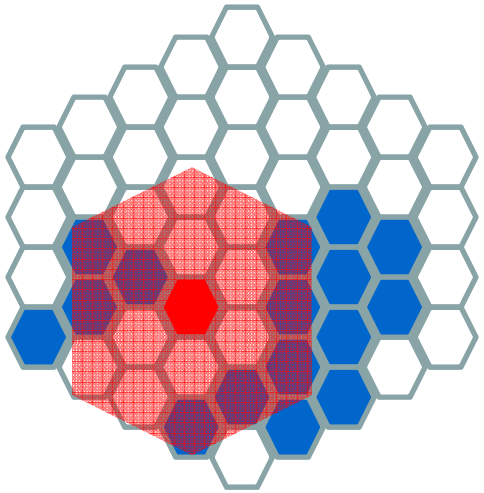


Modelling aggregation – minimization of exposed surface

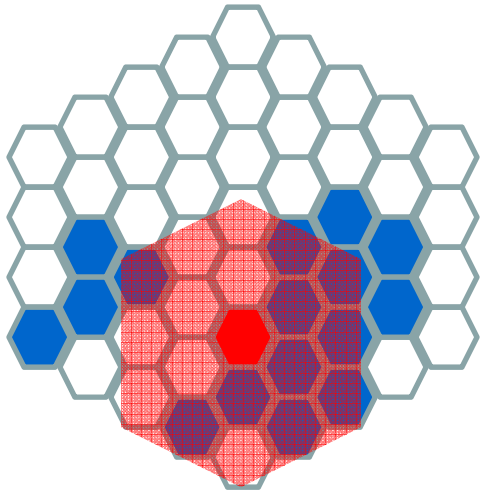
In a simple two dimensional lattice model, we can determine the surface curvature term through examining the proportion of sites within a certain distance from the particle.



What is the probability that the red particle moves to the designated unoccupied lattice site ?



'Curvature term' for the old site = 8/18



'Curvature term' of the new site = 10/18

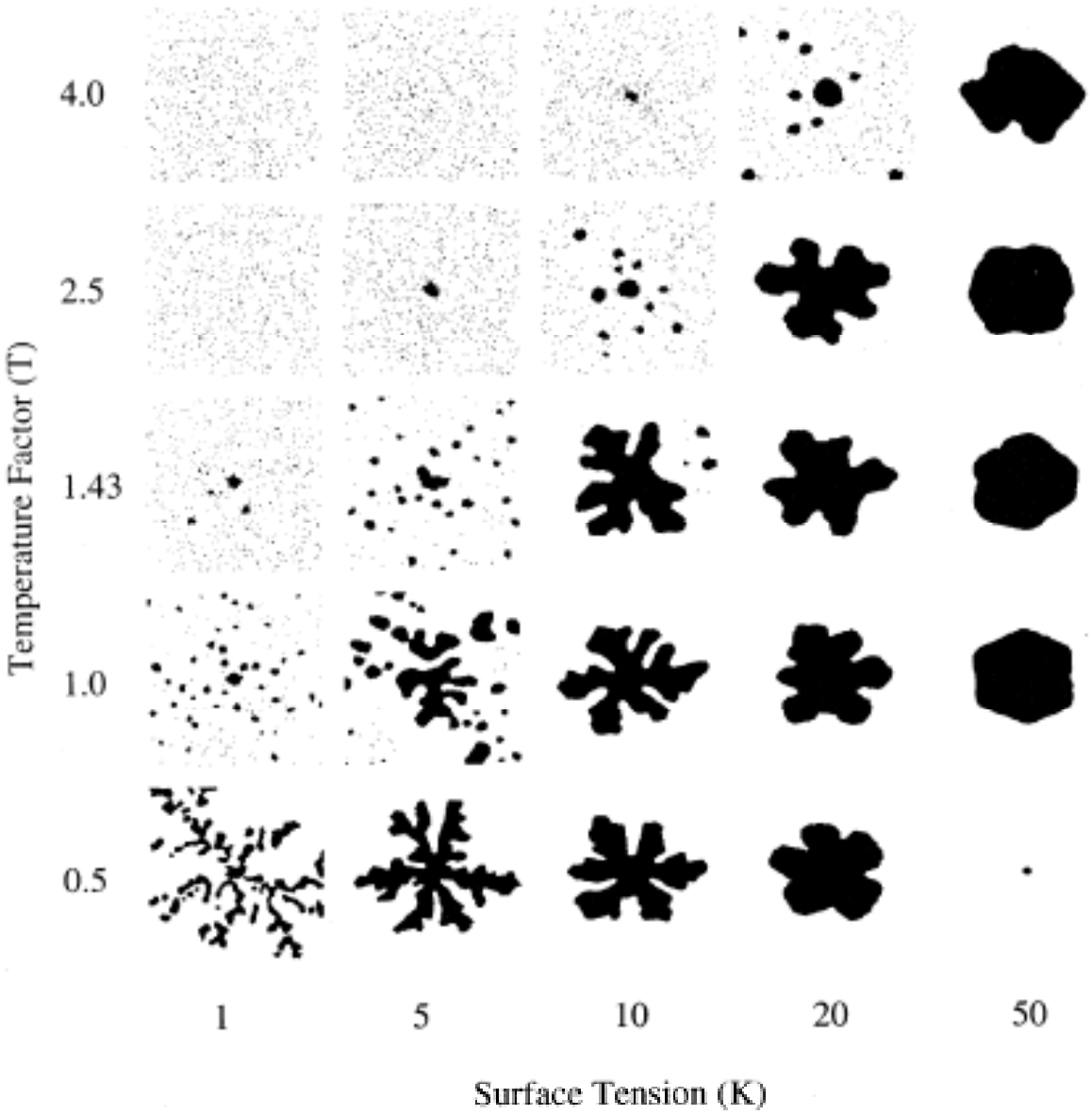
$$\Delta E_c = \Gamma(10/18 - 8/18)$$

$$\Delta E_b = 3 - 1 \text{ (number of interactions)}$$

$$p = \min\left(1, e^{-(2+\Gamma/9)/kT}\right)$$

Current limitations of Cell++

Varying temperature and surface tension terms can result in a spectrum of aggregates being created, representing crystals, droplets and vapour

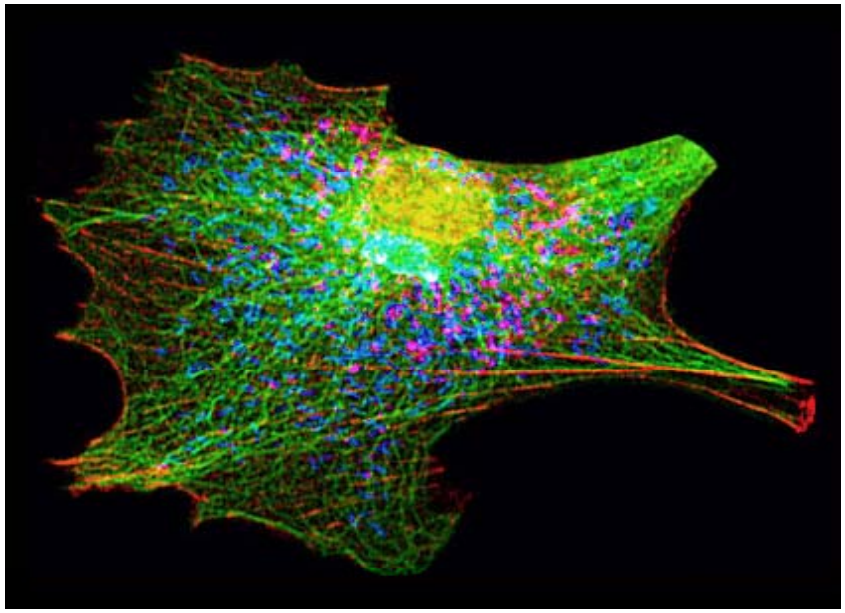


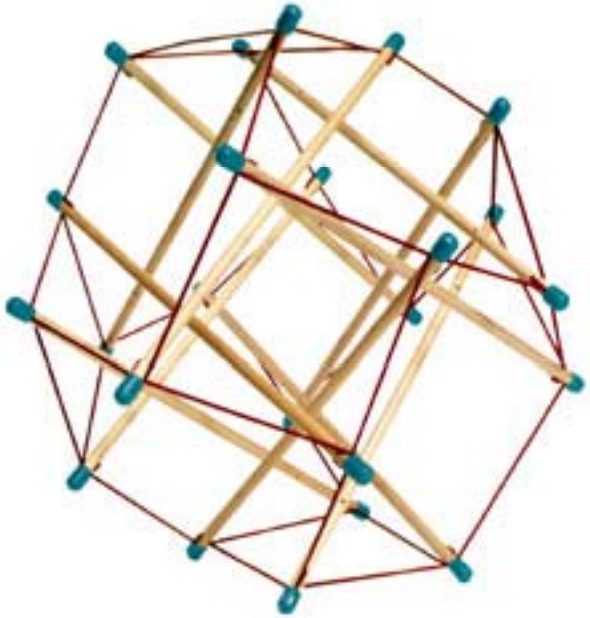
Problem: how do we do this in three dimensions !

Current limitations of Cell++ - III

Accurate modelling of cellular mechanics – models of tensegrity

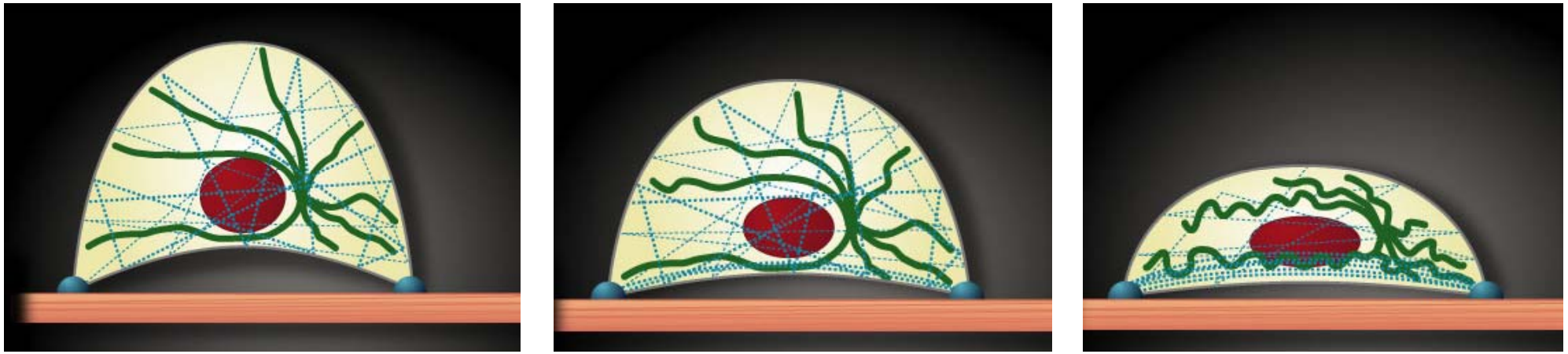
Recent models of the cell cytoskeleton suggests that it behaves as a tensegrity structure with microtubules providing resistance to compression and other microfilaments providing tension





Current limitations of Cell++ - III

Cellular tensegrity – championed by Donald Inger 1996 who performed a series of experiments to show that distorting the morphology of a cell can impact biochemistry and gene exprssion
-Mechanobiology



http://www.childrenshospital.org/research/ingber/Tensegrity.html

As tension is applied to the microfilaments, the microtubules start to buckle and the cell acquires a less rounded more flattened morphology

Problem: how do we simulate the changing morphology of the simulation environment ?