Some first results of PhD-project: Inference of within cell protein interactions and spatial structure, using FRET

Jan-Otto Hooghoudt

Aalborg University

Avignon SSIAB, May 10 2012

Outline of the talk

- Introduction to the general problem and research questions
- Short review of theory of Fluorescence Resonance Energy Transfer
- Dependence of FRET-efficiency on point processes
  - Modeling of FRET efficiency
  - Generating the point patterns
  - Data analysis
- Discussion of some results

Introduction: Problem description

- Distribution and interaction between proteins in cells not well understood
- The interactions take place at the molecular level (1-100 nm)
  - These scales can presently not be resolved directly by available microscopic techniques.
- However, FRET-microscopy does provide indirect information regarding proximity of proteins at molecular level
  - By FRET, information available where in a cell proteins are close to each other
  - But, no information available concerning the protein distribution within a pixel

Introduction: Project Objectives

The project objectives are to:

- develop spatial models modeling the protein distribution at the molecular level
- develop likelihood based inference methods using an available FRET-efficiency model as the generating stochastic mechanism. $Y = g(X; \theta)$ with $g(\cdot)$ the stochastic mechanism which we can simulate.
- infer information concerning the parameters that define the type and strength of clustering
- infer information concerning the absolute concentrations of proteins and their complexes throughout a cell.

Theory Fluorescence Resonance Energy Transfer

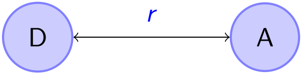
Electrodynamic phenomenon:
Donor molecule gets excited by laser light and de-excites by:

-photon emission (rate k_{rad})
-FRET (rate k_{FRET})

Where the following relationship exists:

$$\mathrm{k}_{FRET} = \mathrm{k}_{\mathrm{rad}} \left(\frac{\mathrm{R}_0}{\mathrm{r}} \right)^6$$

-r = distance between donor and acceptor
-R_0 = Forster distance, the distance r for which 50% of de-excitations due to FRET and 50% due to donor-emission.

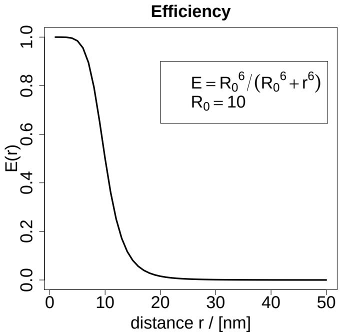


FRET efficiency

The main parameter describing FRET is the FRET efficiency:

$$E = \frac{\text{rate of de-excitations due to FRET}}{\text{de-excitation rate}} = \frac{k_{FRET}}{k_{\text{rad}} + k_{FRET}}$$

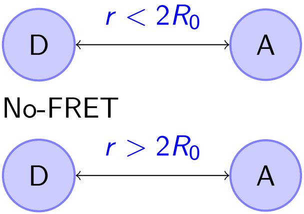
$$\mathrm{k}_{FRET} = k_{\text{rad}} \left(\frac{\mathrm{R}_0}{\mathrm{r}} \right)^6 \quad \rightarrow \quad E(r) = \frac{R_0^6}{R_0^6 + r^6}.$$



Highly sensitive to the distance due to r^{-6}:

FRET

No-FRET



FRET-efficiency multiple acceptors

When multiple acceptors surround a donor, total rate of de-excitations due to FRET becomes:

$$k_{FRET}^{tot} = k_{rad} \sum_{i=1}^{n} \left(\frac{R_0}{r_i} \right)^6$$

And total rate of de-excitation:

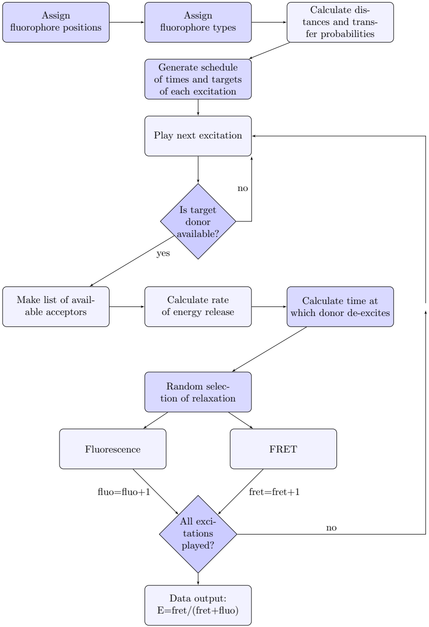
$$k_{tot} = k_{rad}(1 + \sum_{i=1}^{n} \left(\frac{R_0}{r_i} \right)^6)$$

So probability of de-excitation by FRET to acceptor A_i and due to emission are given by:

$$P_{FRET}^{A_i} = \frac{\left(\frac{R_0}{r_i} \right)^6}{(1 + \sum_{i=1}^{n} \left(\frac{R_0}{r_i} \right)^6)} \; ; \; P_{rad} = \frac{1}{(1 + \sum_{i=1}^{n} \left(\frac{R_0}{r_i} \right)^6)}$$

For simulation compute transfer probability-matrix defining all probabilities of de-exitation of D_j to A_i or due to emission.

Modeling the FRET efficiency



Flow diagram of MC-simulation to model the FRET efficiency for:

- different types of proteins (monomer, dimer, etc)
- absolute concentrations of the proteins

Diagram by *Corry et. al.* (2005, Biophys. J.)

A FRET image

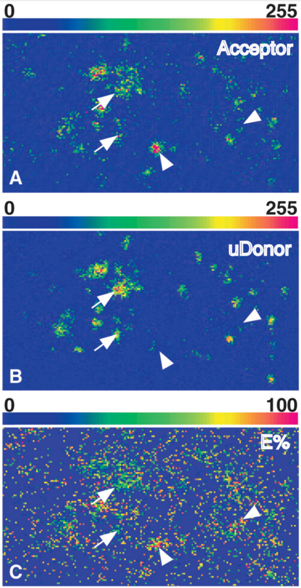


Figure: Wallrabe et.al 2003

To calculate the FRET efficiency, emission is measured in 3 channels:

-Acceptor Channel: Acceptor excitation and acceptor emission

-Donor Channel: Donor excitation and donor emission

-FRET Channel: Donor excitation, acceptor emission

Generating the point patterns (in R)

For a Strauss hardcore point process $\mathbf{X}$, the (unnormalized) density is given by:

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \gamma^{s_R(\mathbf{x})} 0^{s_{hc}(\mathbf{x})} \tag{1}$$

-$n(\mathbf{x})$ number of points in pattern $\mathbf{x}$
-$s_R(\mathbf{x})$ number of pair-of-points within distance R in pattern $\mathbf{x}$.
-$s_{hc}(\mathbf{x})$number of pair-of-points within distance hc in pattern $\mathbf{x}$.

$$s_R(\mathbf{x}) = \sum_{\{u,v\} \subseteq \mathbf{x}} \mathbf{1}[\|u - v\| \leq R] \tag{2}$$

$\beta > 0$, and γ the interaction parameter defining the behavior of the process.

- $0 < \gamma < 1$, $\mathbf{X}$ is repulsive,
- $\gamma = 1$, $\mathbf{X} \sim$ Poisson hard-core
- $\gamma > 1$, $\mathbf{X}$ is clustered, but repulsive at a small scale.

Generating the point patterns (in R)

Further we have used the Multi-Strauss hardcore process:

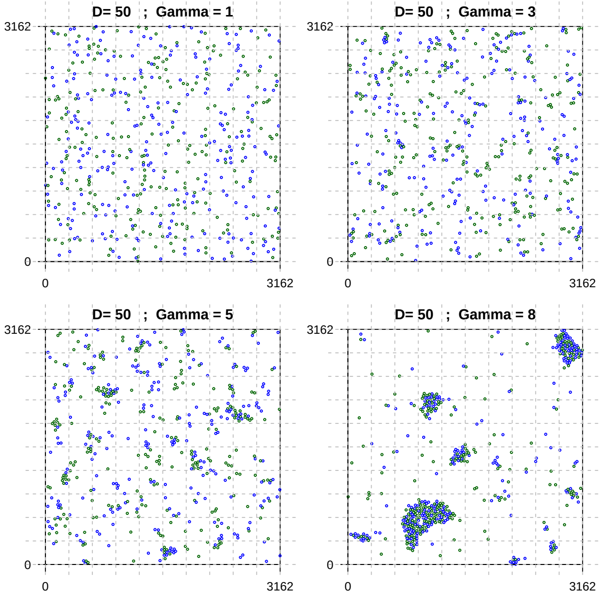
$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \, \gamma_{aa}^{s_{Raa}(\mathbf{x})} \, \gamma_{dd}^{s_{Rdd}(\mathbf{x})} \, \gamma_{da}^{s_{Rda}(\mathbf{x})} \, 0^{s_{Raa}(\mathbf{x})} \, 0^{s_{Rdd}(\mathbf{x})} \, 0^{s_{Rda}(\mathbf{x})} \tag{3}$$

Parameters and interaction radius depending on the type of point (Donor or Acceptor)

Strauss Patterns

- donor
- acceptor

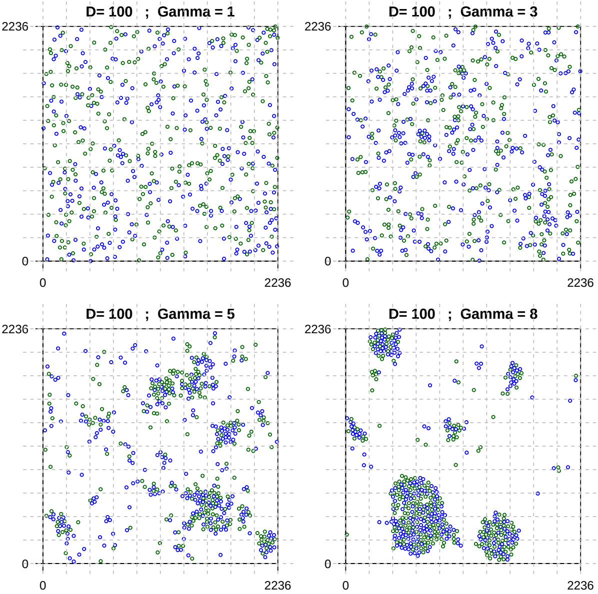
Density in #points per pixel
Pixel-size = 100x100nm



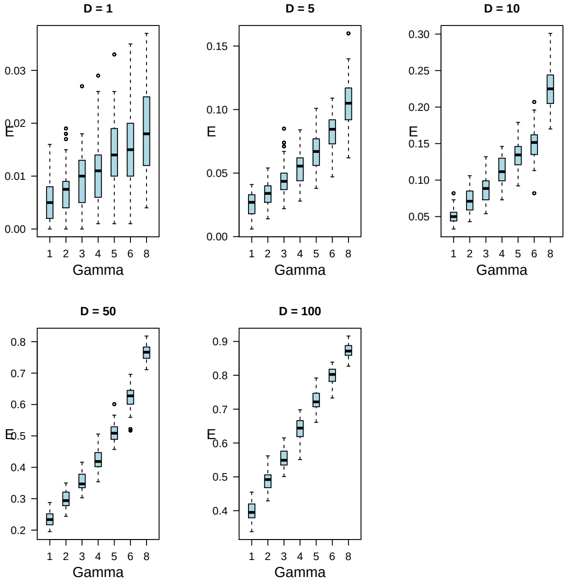
Strauss Patterns

- donor
- acceptor

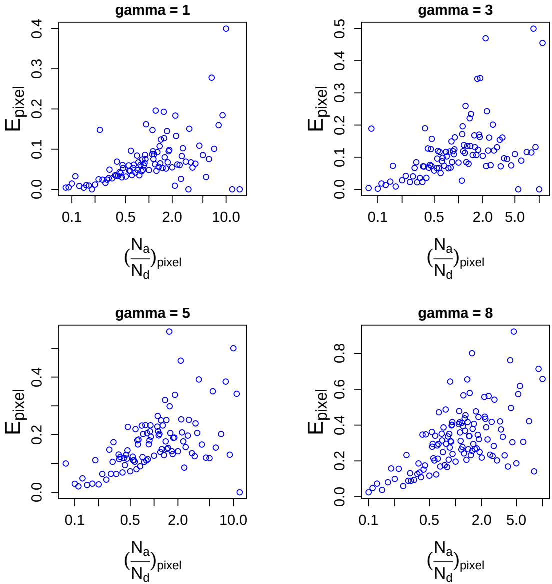
Density in #points per pixel
Pixel-size = 100x100nm



Strauss process: E versus gamma



Strauss process: Dependency E pixel on ratio #acceptors-to-#donors

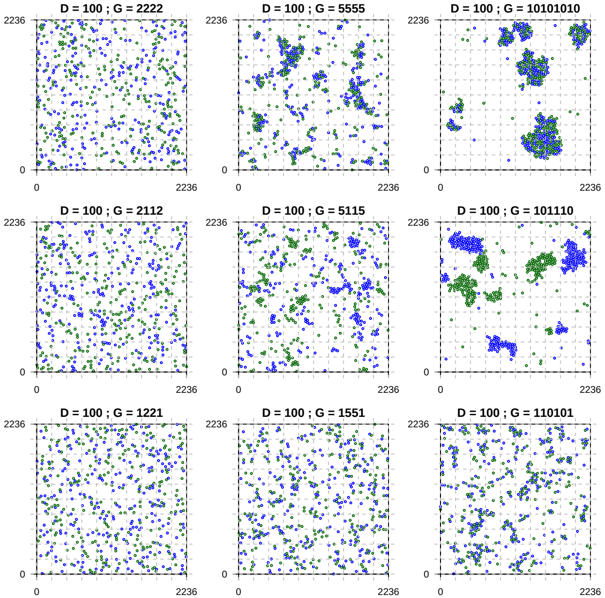


Multi-Strauss Patterns

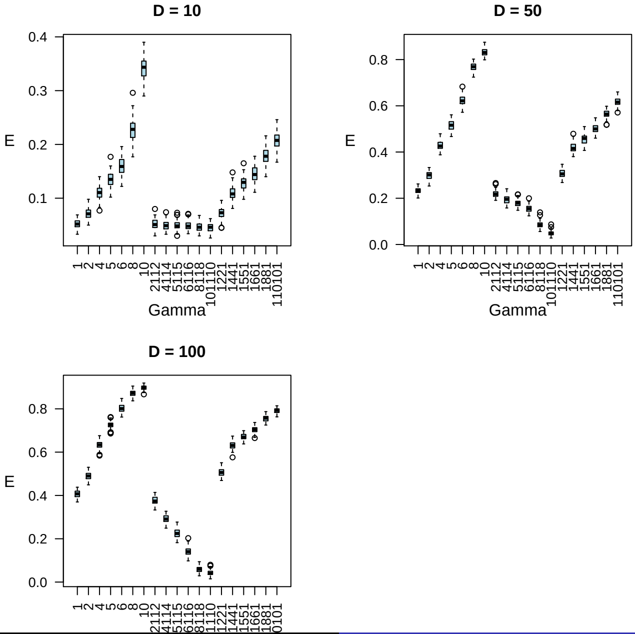
- donor
- acceptor

Notation for Multi-Strauss:

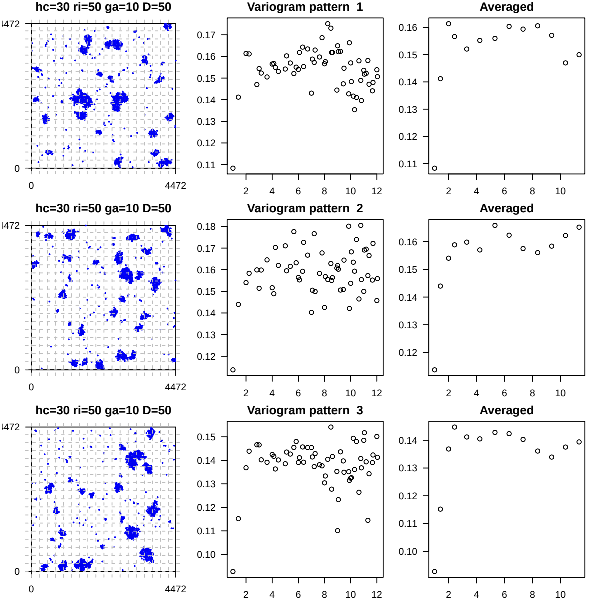
$$\textbf{Gamma} = \begin{bmatrix} g_{DD} & g_{DA} \\ g_{AD} & g_{AA} \end{bmatrix} = (g_{DD}, g_{DA}, g_{AD}, g_{AA})$$



Multi-Strauss process: E versus gamma



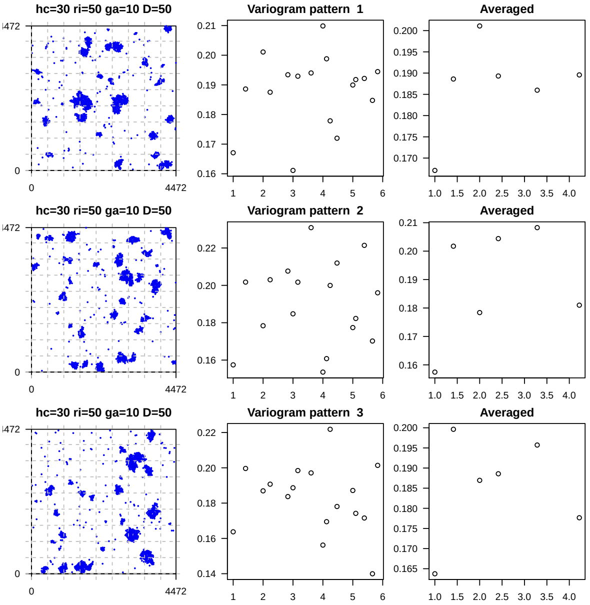
Variogram: correlation E values between pixels



pixel-resolution used
= 25 x 25 nm

empirical pixel sizes
= 100x100 nm

Variogram: correlation E values between pixels



pixel-resolution used
= 50 x 50 nm

empirical pixel sizes
= 100x100 nm

Inference

Possibilities:

- method of moments: match empirical summary statistics (pixel means, variances, variograms...) with theoretical counterparts (approximated using simulation)
- implicit likelihood: approximate likelihood function for FRET pixel intensities using simulation
- Bayesian inference (MCMC): X viewed as missing data.

Implicit likelihood

Defining θ as a multi-dimensional parameter containing; type of model, clustering strenght, absolute concentrations of proteins.

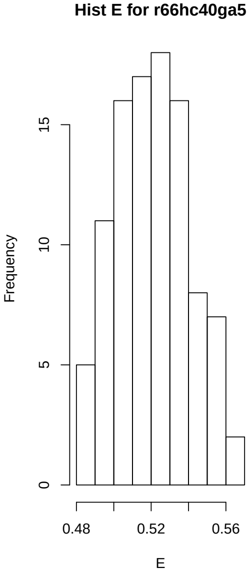
- we obtain a probability distribution function $P(X;\theta)$
- from $P(X;\theta)$ we generate $Y = g(X;\theta)$, with $g(\cdot)$ the MC-simulation, and Y the FRET efficiency

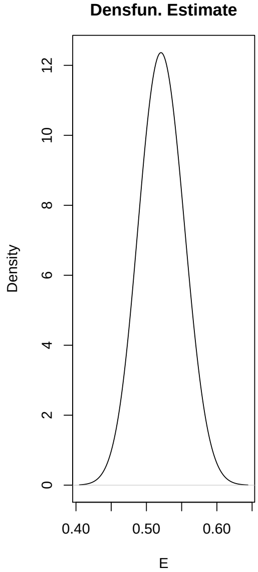
In this way we obtain the likelihood function:

$$L(\theta) = P(Y;\theta)$$

which can not be obtained explicitly.

1st trial with estimating implicit likelihood
Density function estimation: E image-to-image





$hc = 30$
$ri = 40, 50, 60, 70$

$hc = 35$
$ri = 46, 58, 70, 81$

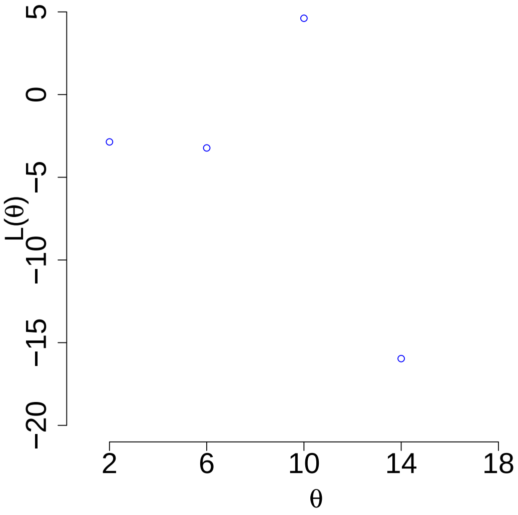
$hc = \mathbf{40}$
$ri = 53, \mathbf{66}, 80, 93$

$hc = 45$
$ri = 60, 75, 90, 105$

$hc = 50$
$ri = 66, 82, 100, 116$

$\gamma = \mathbf{5}$

Likelihood function...



Questions ... ?