SEVENTH FRAMEWORK PROGRAMME

"Ideas" Specific Programme

European Research Council

Grant agreement for Starting Grant

Annex I - "Description of Work"

Project acronym: *SKIPPERAD*

Project full title: *Simulation of the Kinetics and Inverse Problem for the Protein PolymERization in Amyloid Diseases (Prion, Alzheimer's)*

Grant agreement no.: *306321*

Duration: 60 months

Date of preparation of Annex I (latest version): August 6th, 2012

Principal Investigator: Marie Doumic

Host Institution: Inria

[Other beneficiaries (if multi-beneficiary contract)]: INRA

European Research Council

ERC Starting Grant
Research proposal (Part B section 1 (B1))
(to be evaluated in Step 1)

Simulation of the Kinetics and Inverse Problem for the Protein PolymERization in Amyloid Diseases (Prion, Alzheimer's) SKIPPER^{AD}

Name of the Principal Investigator (PI):	Marie Doumic
PI's host institution:	Institut National de Recherche en Informatique et Automatique (INRIA)
Proposal full title:	Simulation of the Kinetics and Inverse Problem for the Protein PolymERization
Proposal short name:	SKIPPER^{AD}
Duration in months:	60

Project summary

Amyloid diseases are of increasing concern in our aging society. They are a group of diseases which involve the aggregation and the deposition of misfolded proteins, called *amyloid*, which are specific for each disease (PrP for Prion, Aβ for Alzheimer's). In a healthy state, they remain monomeric, but when misfolded they propagate the abnormal configuration and aggregate to others, forming very long polymers also called *fibrils*. Elucidating the intrinsic mechanisms of these chain reactions, most probably specific for each disease, is a major challenge of molecular biology: do polymers break or do they coalesce? Do some specific sizes polymerize faster? What is the size of the so-called *nucleus*, *i.e.* the minimum stable size for polymers? Does polymerization occur by monomer, dimer, or i-mer addition? On which part of the reactions should a treatment focus to arrest the disease? Up to now, only very partial and partially justified answers have been provided. This is mainly due to the extremely high complexity of the considered processes, which may possibly involve an infinite number of species and reactions (and thus, an infinite system of equations). **The great challenge of the SKIPPER^{AD} project is to develop new mathematical methods in order to model fibrillization reactions, analyse experimental data, help the biologists to discover the key mechanisms of polymerization in these diseases and predict the effects of new therapies.**

Our approach is based on a new mathematical model which consists in the non-linear coupling of a size-structured Partial Differential Equation (PDE), of the aggregation-fragmentation type, with a small number of Ordinary Differential Equations (ODE). On the one hand, **we will solve new and broad mathematical problems, in the fields of PDE analysis, numerical analysis and statistics. These problems are mathematically challenging and have a wide field of applications:** they will directly apply to several kinds of polymerization. They solve problems that may also be relevant for to the cell cycle, for dust formation, etc. On the other hand, we will test their efficacy on real data, thanks to an already well-established collaboration with a team of biophysicists. With such a continuous system of comparative experiments, we aim at constantly aligning our mathematical problems to biological concerns.

When achieved, the project will provide biologists with an efficient platform that will: i) allow them to design their experiments optimally with respect to the specific protein they are studying ; ii) make the most of their experimental data, to extract all possible information on the mechanisms which underlie their polymerization process ; iii) help them in predicting the best strategies to prevent abnormal polymerization, with a view to patient treatment.

Section 1: *The Principal Investigator*
1(a) Scientific Leadership Potential

M. Doumic is **an applied mathematician** with a background in modelling and analysis of partial differential equations. More recently, she began an original collaboration with statisticians in order to combine several points of view on a related biophysical problem.

Scientific contributions

- **Laser wave simulation:** During her PhD, she proposed a new model and a numerical scheme for the oblique propagation of light, and analysed it theoretically, leading to the publications [P2, P10, P11]. This efficient scheme was included in the HERA platform, an AMR (adaptive mesh refinement) multiphysics hydrocode platform developed at the CEA (the French Centre for Atomic Energy).

- **Analysis of structured equations:** several notable advances were made in this field in [P1, P5, P9]. In [P9], based on self-similarity analysis, counter-intuitive behaviours of the transport-fragmentation equation were established, also a key point for the SKIPPER^{AD} project. In [P6] and [P8], the long-term behaviour of two maturity-structured models of hematopoiesis was analysed.

- **Prion and Protein polymerization modelling in collaboration with biologists:** In [P4], asymptotic analysis and compactness estimates are used to justify the PDE model for Prion, highlighting the particular role played by the boundary condition, and in [P18] a new complete model for nucleated polymerization is proposed, providing the SKIPPER^{AD} project with a strong basis.

- **Inverse problems for structured equations, deterministic and statistical viewpoint:** in a series of articles [P3, P12, P17], the problem of recovering the division/ birth/ fragmentation rate of a structured equation is explored both theoretically and numerically, using asymptotic profiles. In [P16], a new viewpoint is given, combining statistical and analytical treatment in an original way.

International recognition

Several fruitful international collaborations have been established and maintained:
- **with H.T. Banks (Raleigh, USA)**, 2 papers and an ongoing collaboration,
- **with J.P. Zubelli (Rio, Brazil)**, a well-established collaboration and a joint participation in the AmSud network (linking French and South American researchers), 3 papers,
- **with A. Marciniak (Heidelberg, Germany)**, 1 paper in the context of her ERC starting grant project BioStruct, and with **P.S. Kim (Utah, USA)**, 1 paper.

They have been developed through a number of short stays and invitations. For family reasons, no long stay abroad has yet been undertaken, but the PI has been invited to prestigious international conferences (see Vitae).

Leadership profile

The PI already has **solid supervision experience.** She fully supervised a PhD student, P. Gabriel, who obtained excellent results and defended his thesis viva voce in June 2011. She also supervised one post-doctoral student (F. Charles), and took part in the supervision of three other PhD students (A. Ballesta, T. Lepoutre, L.M. Tine). She has just begun to supervise a new PhD student, H.W. Haffaf.
From September 2009 to September 2012, M. Doumic is the scientific coordinator of an ANR (French funding agency Agence Nationale de la Recherche) **project grant, TOPPAZ**. This is a joint project between H. Rezaei's team of biologists at INRA, and has already achieved interesting results (*e.g.* [P4, P5, P9, P18]) that will serve as solid grounding for the SKIPPER^{AD} project.
Last but not least, the PI also gained experience as a team and project manager in public works, managing a team of 35 between 2003 to 2007. This leadership and practical experience has developed the PI's operational skills and taste for project management.

1(b) Curriculum Vitae

Personal data

Marie DOUMIC, born: 18.06.1976.

French citizenship

Married (married name: JAUFFRET), four children born in 2005, 2006, 2009 and 2010.

Email: Marie.Doumic@inria.fr

Web-page: http://www-rocq.inria.fr/bang/MDJ/index.html

Address: INRIA Paris-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France.

Degrees

2005 : PhD in mathematics, University of Paris VII Denis-Diderot, defended 20.05.2005,
Asymptotic Study and Numerical Simulation of Laser Wave Propagation in an Inhomogeneous Medium .
Principal Advisor: F. Golse (Professor at Ecole Polytechnique, France).
Secondary Advisor: R. Sentis (Senior Researcher and Department head at CEA).

2000-2003 : Ecole Nationale des Ponts et Chaussées (prestigious French *Grande Ecole* to which there is a special access for graduates of Ecole Normale Supérieure or Ecole Polytechnique, for senior public service)

1998 : Agrégation de Mathématiques (high-level national competitive examination for professorship)

1999 : Master of Science in Applied Non-Linear Analysis, University of Paris-Dauphine, with rank 1.

1996-2000: Ecole Normale Supérieure, Paris (one of the most prestigious French *Grandes Ecoles*)

Employment

2007–present : Research Scientist at INRIA Paris-Rocquencourt.

2003–2007 : Head of Service Technique de la voie d'eau, Service Navigation de la Seine, Voies Navigables de France.

Project management of river and navigation public engineering works (*e.g.* overall locks and dams reconstruction on the Oise river €100 Million, reconstruction of Chatou dam €35 Million) ; management of a 35-person team ; overall plan of engineering reorganization in the Seine Navigation Department.

Research Topics

- structured population models
- non-linear transport and fragmentation equations
- inverse problems (deterministic and statistical viewpoint, numerical analysis)
- asymptotic analysis
- laser-plasma interaction, absorbing boundary condition (subject of PhD)

PhD supervision

2011–present : H.W. Haffaf, PhD student. Thesis subject: *Numerical methods for fragmentation and polymerization equations*.

2008–2011 : P. Gabriel, M.Sc. and Ph.D Student. Thesis subject: *Transport-Fragmentation Equations and Applications to Prion Diseases*.

Participation in the supervision of L. M. Tine in 2010-2011 (advisors: T. Goudon and F. Lagoutière), T. Lepoutre in 2008 (advisors: S. Gaubert and B. Perthame), A. Ballesta (advisors: J. Clairambault and F. Lévi).

Professional and scientific activities

2009–present : Head of the TOPPAZ project, *Theory and Observations of protein Polymerization in Prion and Alzheimer's diseases*, funded by the French National Research Agency (ANR).
Webpage: http://www-roc.inria.fr/bang/TOPPAZ/

2008–present: reviewing activities (*e.g.* for the AMS, CMS, M2AN, JTB)

2012 : Winter School and *10th ICOR* (International Conference on Operations Research), La Habana, Cuba, co-organizer of the PDE session, with S. Mischler (Paris IX Dauphine), Mariano Rodriguez Ricard (La Habana), with M.J. Cacérès (Granada), J.C. Cañizo (Barcelona), and A. Guillin.

2011 : *Mathematics and Biology: Young Investigators International Workshop*, April 4-6, University of Rouen, France. Organized with H.T. Banks and A. Blouza.

2008 : ECMTB08, Edinburgh, Scotland. Organization of a session on *Inverse problems and growth models*, with J. Zubelli.

Funding ID

There is and there will be no funding overlap with the ERC grant requested and any other source of funding for the same activities and costs that are foreseen in this project.

2009–2012 : Head of the TOPPAZ project funded by the ANR. €190 000 in total, of which €100 000 is allocated to the PI's team and €90 000 for the team of biophysicists head by H. Rezaei. End in December 2012.

2012–2015 : Participation in the CALIBRATION project funded by the ANR (on Statistical Calibration). Almost no personal funding (only some invitations, up to approx. €2.000).

2010–present: Participation in the national interdisciplinary network *Protein Misfolding and Pathologies* No funding.

2007–2008 : Participation in the MCRTN (Marie Curie Research Training Network) "TUMATHER" (Modelling, Mathematical Methods and Computer Simulation for Tumour Growth and Therapy)
webpage: http://calvino.polito.it/ mcrtn/

About career breaks.

My first professional experience was out of the research domain. After graduating from the prestigious Ecole des Ponts in 2003, I managed a 35-person team in fluvial public works, in charge of major locks and dam reconstruction projects in the basin of the river Seine. I kept in contact with science however, and defended my PhD on May, 20th, 2005. I decided to return to research in March 2007.
I am also a mother of 4 children, born in 2005, 2006, 2009 and 2010. The total duration of maternity leaves from March 2007 up to now is of one year (the first semester of year 2009 and the period from december 2010 to june 2011).
These personal and professional breaks have resulted in a total research period of less than 4 years from my PhD defense until the present.

1(c) Early Achievement Track Record

M. Doumic is the author of 15 publications in major international peer-reviewed journals or conference proceedings. The period 2003-2007 represents a break in her scientific career. **The references of the main articles are underlined, thus [P18].** Citation records were obtained via Google Scholar and MathSciNet, and exclude self-citations.

Publications in peer-reviewed journals without PhD supervisor

[P1] M. Doumic, *Analysis of a Population Model Structured by the Cells Molecular Content*, MMNP, 2007, vol. 2 **(3)**, 121–152. (12 cit.)

[P2] M. Doumic, *Boundary Value Problem for an Oblique Paraxial Model of Light Propagation*, MAA, vol. 16 **(1)**, 2009, 119–138.

[P3] M. Doumic, B. Perthame, J. Zubelli, *Numerical Solution of an Inverse problem in Size-Structured Population Dynamics*, **Inverse Problems**, vol. 25 **(4)**, 2009, 1–22. (5 cit.)

[P4] M. Doumic, T. Goudon, T. Lepoutre, *Scaling Limit of a Discrete Prion Dynamics Model*, **Comm. in Math. Sc.**, vol.7 **(4)**, 2009, 839–865. (4 cit.)

[P5] M. Doumic, P. Gabriel, *Eigenelements of a General Aggregation-Fragmentation Model*, **M3AS**, 2010, vol. 20 **(5)**, 757–783.(3 cit.)

[P6] M. Doumic, P. Kim, B. Perthame, *Stability Analysis of a Simplified Yet Complete Model for Chronic Myelogenous Leukemia*, Bull. of Math. Biol, 2010, vol. 72 **(7)**, 1732–1759. (2 cit.)

[P7] H.T. Banks, F. Charles, M. Doumic, K.L. Sutton, W.C. Thompson, *Label structured cell proliferation models.* Appl. Math. Lett. 23 **(12)**, 2010, 1412–1415.

[P8] M. Doumic, A. Marciniak, B. Perthame, J.P. Zubelli, *A Structured Population Model of Cell Differentiation*, **SIAM J. of Appl. Maths,** in press.

[P9] V. Calvez, M. Doumic, P. Gabriel, *Self-similarity in a General Aggregation-Fragmentation Problem ; Application to Fitness Analysis*, accepted in **J. Math. Pures Appl.**

Publications in peer-reviewed journals with PhD supervisor

[P10] M. Doumic, F. Golse, R. Sentis, *A paraxial model for the propagation of light: the boundary value problem for the Schrödinger-advection equation in a tilted frame*, C. R. Acad. Sci. Paris, Ser. I, 2003, 336, 23–28. (6 cit.)

[P11] M. Doumic, F. Duboc, F. Golse, R. Sentis, *Boundary value problem for an oblique paraxial model of light propagation: the advection-Schrödinger equation*, **J. Comp. Phys.**, 2009, vol. 228 **(3)**, 861–880. (2 cit.)

Publications in peer-reviewed proceedings without supervisors

[P12] M. Doumic, P. Maia, J. Zubelli, *On the Calibration of a Size-Structured Population Model from Experimental Data*, Acta Biotheoretica, vol. 58 **(4)**, 2010, 405–13.

Publications not as main author (in proceedings or journals, with or without supervisors)

[P13] H.T. Banks, K.L. Sutton, W. C. Thompson, G. Bocharov, M. Doumic, T. Schenkel, J. Arguilaguet, S. Giest, C. Peligero, A. Meyerhans, *A New Model for the Estimation of Cell Proliferation Dynamics Using CFSE Data*, to appear in J. of Immunological Methods.

[P14] R. Sentis, M. Doumic, F. Golse, *Mathematical and numerical aspects of wave propagation*, WAVES 2003, 862–867, Springer, Berlin.

[P15] V. Calvez, N. Lenuzza, M. Doumic, J-P. Deslys, F. Mouthon, B. Perthame, *Prion dynamic with size dependency - strain phenomena*, Journal of Biological Dynamics, 2010, Vol. 4 **(1)**, 28–42.

Preprints (as main or one of the main authors)

[P16] M. Doumic, M. Hoffmann, P. Reynaud, V. Rivoirard, Nonparametric estimation of the division rate of a size-structured population, submitted.

[P17] M. Doumic, L. M. Tine, A General Inverse Problem for Polymerization-Fragmentation Equations, submitted.

[P18] M. Doumic, A. Ballesta, F. Charles, N. Lenuzza, S. Prigent, P. Gabriel, L-M. Tine, H. Rezaei, *An Efficient Kinetic Model for Amyloid Fibrils Assemblies - Application to Huntington's Disease*, submitted.

Selection of some main invited presentations

I was invited to give a talk in several prestigious places, among which I can first quote:

- October 2011 : Applications of Kinetic Theory and Computation, ICERM, Providence, USA.
- September 2010: PDEs in Mathematical Biology, Bedlewo, Poland.
- June 2010 : Non-Linear Days, conference in honor of Venakidès, J-L. Lions Lab., UPMC, Paris.
- February 2010 : 9th ICOR (International Conference on Operations Research), La Havana, Cuba.
- October 2009 : math. biology conference, HSC (Hausdorff Centre for Mathematics), Bonn, Germany, .
- May 2009 : MFO, Oberwolfach, Germany, workshop on mathematical biology.

I can also quote (a more complete list is on my webpage):

- September 2011: Non-Linear PDE arising in Biology, Edinburgh, UK.
- August 2011: Conference on Mathematical Methods and Modeling in Life Sciences and Biomedicine, Turkey.
- February 2010 : applied mathematics seminar, Raleigh (invited by H.T. Banks, 1 week stay), USA.
- November 2009 : Biomathematics and biomechanics, Tozeur, Tunisia.
- July 2008 : ECMTB 2008, Edinburgh, session *Inverse problems for growth models*, co-organized with J.P. Zubelli
- May 2008: CancerSim2008, Euroconference on Modelling and Simulation of Cancer Growth and therapy, Torino, Italy.
- August 2007 : Mathematical Methods and Modeling of Biophysical Phenomena, Buzios, Brazil.
- July 2007 : DEASE European project meeting conference, Vienna, Austria.

Section 2: *The Research Project*

a. State-of-the-art and objectives

Amyloid diseases are of increasing concern in our aging society: indeed, they include *e.g.* Alzheimer's, Parkinson's, Huntington's and Prion diseases. Currently, there exists no effective treatment even to slow their progression. SKIPPER^{AD} proposes a new perspective to approach this urgent issue: it will provide the biology community with new methods to investigate the mechanisms of the polymerization involved in these diseases.

Fibrillization in amyloid diseases

Amyloid diseases are characterized by the aggregation and deposition of misfolded proteins called *amyloid*. The protein is specific for each amyloid disease (*e.g.* PrP for Prion, αsynuclein for Parkinson's, Aβ for Alzheimer's). In a healthy state, proteins remain monomeric, but when misfolded in a certain way they propagate the abnormal configuration and aggregate to others, forming very long polymers (also called *fibrils*). For example, PrP, in Prion diseases, undergoes a structural change and forms an amyloid aggregate in an unidentified kinetic pathway. The intrinsic mechanisms of these chain reactions, most probably specific for each disease, remain largely unknown: do polymers break or coalesce? Do some particular sizes polymerize faster than others? What is the size of the so-called *nucleus*, the minimum stable size for polymers? Does polymerization occur by monomer, dimer, or i-mer addition [21]? On which part of the reactions should treatment focus in order to arrest the disease ?

Up to now, only very partial and partially justified answers have been given. This is mainly due to the extremely high complexity of the processes, which may possibly involve an infinite number of species and reactions (and thus, an infinite system of equations). This huge amount of chain reactions includes intermediate conformers formation, polymerization by monomer, dimer or oligomer addition, coalescence, depolymerization (by loss of monomers, dimers or oligomers), fragmentation (breakage into two or more polymers), and degradation.

For many years, ordinary differential equations have been used to model such reactions [41], but the number of equations is at least equal to the maximal size of polymers, which is extremely large in the case of fibrillization. Usual computational techniques leading to too time-consuming simulations, it is commonly accepted among the biology community that very simplifying assumptions are required, e.g. constant reaction rates, meaning that a polymer of any size behaves in roughly the same way [8, 32, 51, 33]. Such assumptions are difficult to justify theoretically, and have been contradicted in some cases (Alzheimer's [50], Prion [45]).

Aggregation-Fragmentation equations

Mathematically, polymerization mechanisms are described by fragmentation-coalescence type equations, which are either discrete [2] or continuous [40]. Coming originally from physics, they have been widely studied and remain an active research field, a large literature being already available - *e.g.* [30, 24, 36] and references therein. **Their mathematical study when applied to protein polymerization is much more recent however, and has raised new unexpected problems.** Various mathematical studies related to fragmentation equations have recently brought improvements in the area of our concern, and will serve as research directions to build on and develop new methods. These problems of broad interest still present many open features, such as their asymptotic behaviour when nonlinearities occur; how aggregation and fragmentation can balance each other to result in an asymptotic steady profile [34, 15], or on the contrary how dominant fragmentation can lead to concentration in size [25]; what the most efficient numerical schemes to solve them are [9, 26]; how the proliferation rate, also called the *fitness* of the population, depends on the main coefficients, especially the fragmentation and aggregation rates [35]; how to derive PDEs from ODEs or stochastic processes describing growth and division, as carried out in many different articles, what is the rate of convergence towards the asymptotic behaviour [10], etc.

Inverse problems in this type of equations are a very open field, with still few existing studies. **This last item is particularly promising and should result in break-through discoveries by allowing optimal use of the increasingly rich experimental databases.** Using a statistical non-parametric method directly calibrated on microscopic data, in [17] we have obtained a new estimation method combining the most recent kernel methods (inspired by Goldenschluger-Lepski algorithm) with the inverse problem solution developed in [43, 18]. Recent work by H.T. Banks *et al.* [6] is also promising and may be relevant here.

Modelling of Prion replication and Alzheimer's disease

Historically, the first success of mathematical modelling was to demonstrate that essential features of Prion diseases can be explained by purely physico-chemical mechanisms. The modelling of intracellular Prion infection has been dramatically improved in the past few years due to recent progress in the molecular biology of this pathology. Relevant models have been designed to investigate the conversion of PrPc (the monomer) into PrPsc (misfolded proteins) according to the auto-catalytic process in fibril aggregation. These models are based on a linear growth of PrPsc polymers [20]. In [33], Masel et al. proposed a discrete infinite-dimensional system of Ordinary Differential Equations (ODE) where the Prion population is described by its distribution with respect to the size of polymer aggregates. Greer et al. analyzed this process in a continuous setting [28, 23]. They propose a Partial Differential Equation (PDE) to render the above-mentioned polymerization/fragmentation process. Inspired from many previous studies linking such discrete and continuous models, a complete derivation from ODE is proposed in [16]. Greer and co-authors point out an interesting dichotomy in the asymptotic behavior of the system: depending on the coefficients, either the healthy steady-state (no PrPsc proliferation) is stable, or the disease steady-state (PrPsc infection) is stable. Recent studies investigated the general case of size-dependent coefficients [12].

Concerning Alzheimer's, there exist several contributions to the mathematical modelling of fibrillization of the Aβ structure in the brain [42]. There is also a highly detailed description of biomolecular aspects of this fibrillization. One difficulty for Aβ amyloid fibrils of Alzheimer's disease is their polymorphism [29]: capturing this variety will imply a system of coupled fragmentation-coalescence equations. This is a major modelling and mathematical challenge.

The foundation of this work: a new overall PDE approach

In collaboration with biologists from INRA, the PI and co-authors propose a new overall framework that generalizes the Prion model of [28] and can be adapted to most protein polymerization reactions [4]. This method relies on PDEs: instead of handling an infinite set of ODEs, we derived an equivalent model composed of a small number of ODEs coupled with a single size-structured PDE. It belongs to the family of transport and aggregation-fragmentation equations, and states:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\Big(\big(V(t)\tau(x) - d(x)\big)\,u\Big) = -\beta(x)u(t,x) + 2\int_x^\infty k(x,y)\beta(y)u(t,y)dy - \mu(x)u(t,x)$$
$$+\frac{1}{2}\int_{x_0}^{x} c(y,x-y)u(t,y)u(t,x-y)dy - \int_{x_0}^{\infty} c(x,y)u(t,x)u(t,y)dy, \tag{1}$$

$$\frac{dV}{dt} = \lambda - \mu_0 + \int_0^\infty \big(d(x) - V\tau(x)\big)u(t,x)dx, \qquad V(t=0) = V_0, \tag{2}$$

$$u(t,x=0) = \frac{k_{on}V^{i_0}}{k_{off} + \tau(0)V}, \qquad u(t=0,x) = u^{in}(x). \tag{3}$$

Here $u(t,x)$ represents the density of polymers of size x at time t, V the density of the monomers, i_0 the nucleus size at which polymers become stable, τ and d the polymerization and depolymerization rates, β the fragmentation rate, $k(x,y)$ the fragmentation kernel (polymers of size y break into two polymers of respective sizes x and $y-x$ with a rate $\beta(y)k(x,y)$, the factor 2 before the integral term comes from this splitting into 2 pieces). The coalescence terms are on the second line of Equation (1): $c(x,y)$ is the coalescence rate of two polymers of sizes x and y giving rise to a polymer of size $x+y$ (here again, the factor $1/2$ comes from this coalescence of two). A possible production rate of monomers is represented by λ, whereas degradation terms are μ_0 and $\mu(x)$. The boundary condition at $x=0$ comes from the nucleation process, where i_0 monomers give rise to i_0-mers with exchange rates k_{on} and k_{off}. When modelling an experiment of spontaneous fibrillization, with no "seed", the initial condition is $u^{in}(x) \equiv 0$. This model can be adapted, simplified or made more complex to fit particular cases, for instance to take into account a conformer/monomer exchange, or to neglect coalescence or depolymerization, etc., with much freedom and many possible variants.

We thus obtained a fully general model that is easier to handle both theoretically and numerically, and results in faster computations than for the full ODE set of equations. In [4], as a proof of concept, we ran a first set of simulations to the variant of the model which applies to PolyQ polymerization, as occurs in Huntington's disease. We want to build on this, developing new mathematical and numerical methods, and comparing simulations with well-designed experiments.

The advantages of using this new model are twofold.
First, developing new efficient schemes [27], it allows us to investigate numerically how a change in the coefficients can influence the overall reaction - and, more specifically, the size distribution of polymers. Also, inverse problem techniques, including statistical treatment [17] can be developed, as was successfully done in other contexts, in order to estimate size-dependent parameters [6].
Secondly, as a PDE problem it is a good conceptual object, modelling a wide variety of physical and biological situations. It belongs to the active research field of nonlinear transport, aggregation-fragmentation equations [9, 10, 28]. Theoretical analysis can help to understand the intrinsic mechanisms and formulating new paradigms, as shown in recent studies [12, 1, 11]. Many challenging questions remain open, since, as proved for instance in [11], the behaviour of such equations with respect to their parameter functions can prove counter-intuitive. Inverse problem analysis for such equations are promising and are still at their beginning [18, 17], especially in the biophysical community [51, 47, 7].

Main challenge of the SKIPPER^{AD} project

The great objective of this project is to quantitatively determine the intricate mechanisms that constitute protein polymerizations. To do so, we will

- solve new mathematical problems arising from the biological issues,
- design efficient numerical schemes,
- accurately define inverse problem solutions.

Our studies will provide the biophysics community with overall methods for optimally designing their experiments with respect to the specific protein under study; for extracting, from their experimental data, all the possible information on the mechanisms which underlie the polymerization process, and quantifying the quality of this information; and for predicting the best strategies to prevent abnormal polymerization, ultimately stopping the disease.
With the experimental data obtained and provided by H. Rezaei and his team (at INRA, Jouy-en-Josas), with whom collaboration is already well-established, we also want to test these new methods on the protein polymerization occurring in Prion and Alzheimer's diseases. Indeed, we believe it useless to design methods that would not be permanently aligned to up-to-date experiments: designing methods and applying them to real data are complementary. Moreover, **challenging new mathematical problems are very likely to emerge from such interdisciplinary contacts.**
A wide variety of skills are required in this project, from deep theoretical PDE analysis on the one hand, in order to investigate the (local, long-time, singular etc) behaviours of the equations, since the biologists bring challenging new problems to the mathematical community. On the other hand, especially as concerns the inverse problem solving, it is also necessary to collaborate with applied mathematicians, who have a good understanding of biological issues and experimental devices, in order to stick to the "real life" problems. Halfway, numerical analysis of such problems also presents major issues.

b. Methodology

As previously mentioned, our approach is based on a new modelling framework, proposed by the PI and co-authors in the submitted article [4]. One of its possible variants is given by Equations (1)–(3). This model consists in the coupling of a Partial Differential Equation (PDE) with a small number of Ordinary Differential Equations (ODE). On the one hand we will build general simulation and mathematical methods applicable to any polymerization pathway, and on the other hand test their efficacy on real data (produced by H. Rezaei's INRA team) for Huntington's, Prion and Alzheimer's. **The following tasks will thus be carried out at least partially in parallel, so that a possible delay in a given task could be anticipated.**

We describe below five main tasks of unequal length: first, the numerical aspects, secondly, the analytical part, thirdly, the inverse problem solving. The fourth task consists in the design and use of a numerical and predictive platform that will bring together all the previously designed methods of Tasks 1 and 3. As a result of these tasks, the fifth and final task will help the biologists to discover new therapeutic strategies.

However, the five tasks will be continually reviewed and tuned, and this is a fundamental characteristic of this project: **even the most theoretical part will be constantly aligned with real-life biological concerns.** Let us show briefly how this will work for a given application such as the modelling of the *in vitro* fibrillization of the PrP protein responsible for Prion disease. Figure 1 illustrates this methodology. This has already been sketched, for the polyglutamine (PolyQ) expansion in the protein huntingtin (Htt) responsible for Huntington's disease, in the preprint [4], so that the methodology we want to apply is quite clear, even if what we now want to achieve is far more ambitious.

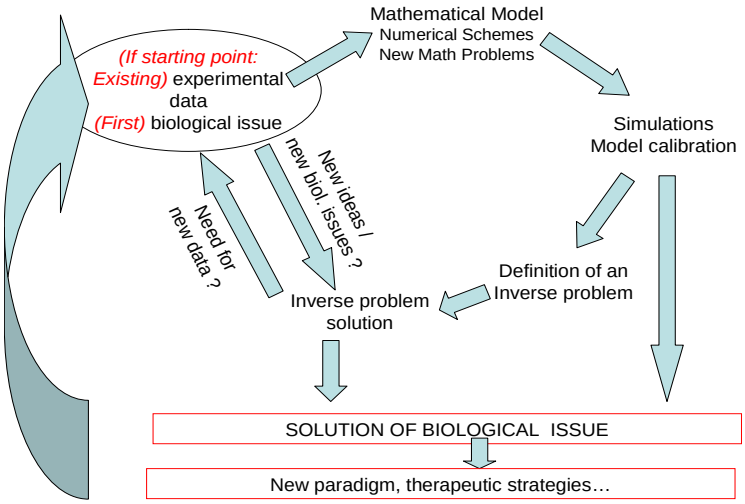

Figure 1: Research Paradigm

First, we depart from existing experimental data and biological questions. Our idea is to save the biologists from carrying out costly experiments that could be useless if this inquiry came before any mathematical analysis. Through careful discussions with H. Rezaei, we will define how much we can rely on these data, what are the origins of the noises and measurement errors, and **we will formulate a model and one (or better several) adequate numerical scheme(s).** To build such a model, we will adapt our general problem (1)-(3) to the case corresponding to the experiment under consideration. Even for the same disease and the same protein, it may not be the same part of the model that will play the main role according to the experimental conditions (sonication or not, high or low temperature, monomeric or fibril initial conditions etc.)

At this stage, either new mathematical questions will arise, or already considered problems (as listed in Task 2) will apply. As part of the work of Task 2, their study will then become a priority. Also, new numerical schemes or asymptotic analysis may be needed, *e.g.* to check that the use of the PDE model is correct, or to go back to the ODE system.

By simulations and the use of generalized sensitivity functions, and of a standard inverse problem technique (such as least-squares optimization [6]), we will obtain initial results on the influence of each process, and on the orders of magnitude of each parameter.

A specific definition of the relevant inverse problem to solve will then be possible. Such a problem can either be one of the problems defined in Task 3.1., or it mightl be a new one. It will be addressed both numerically, on simulated data, and theoretically. **At this stage, depending on our findings concerning the inverse problem solution, it may appear necessary to carry out new experiments:** this will be defined with H. Rezaei and carried out by his team.

Once such adequately defined data have been obtained, **a complete solution of the inverse problem will be possible, and the calibrated model will become predictive.** New therapeutic strategies could then be designed and tested (Task 5): for this again, numerical and theoretical analysis of Tasks 1 and 2 will be of great help.

Of course, as shown in Figure 1, such a process may need to be iterated several times to fully succeed. The consequences of such findings are potentially huge in terms of both their biological and social impact.

Task 1: Direct Problem: Design of efficient numerical schemes

Start: year 1, **Duration:** 3 years, **Risk:** low / moderate

This part has to be treated (at least in a preliminary version) at the very beginning. A major question here is the numerical strategy to adopt: which scheme should be used in which context? Conservative or not, implicit or explicit, of which order? Indeed, most accurate schemes such as WENO [27] are much more time-consuming than simpler but less accurate ones [48]. Numerical diffusion can also lead to artificial asymptotic profiles [13]. Time optimization of the direct problem is also a key point for many inverse problem techniques.

Moreover, schemes have to conserve in the best possible way two main properties of both the discrete and the continuous model: the mass balance equation, involving $\int xc(,t,x)dx$, and the number of polymers balance equation, involving $\int c(t,x)dx$. However, this may turn out to be too time-consuming when used for the inverse problem: optimizing them or combining them with non-conservative schemes may be necessary.

Finally, the numerical scheme has to be automatically adapted during the course of the experiment, since different mechanisms may dominate each period of the reaction. For an *in vitro* experiment of spontaneous fibrillization for instance, as in Figure 2, at the beginning polymerization (transport term of the equation) dominates, and at the end transport vanishes and fragmentation-coalescence dominate. In particular, numerical schemes for fragmentation-coalescence reactions [9], where either gelation or dust formation can occur, have to be adapted to such singular behaviours.

WENO schemes have been introduced successfully in [13, 27] and we will test them in our case where a new specificity are the nonlinear boundary condition (3) and the nonlinear coupling with possibly several ODE including different time-scales. Such schemes are certainly promising to treat the case of non-continuous coefficients, as is likely to occur for the polymerization and the fragmentation rates. Recent progress has also been done to reduce the numerical diffusion [48] for a model with coagulation; investigation of such methods is another possible research direction. All these studies show how diverse the approaches are, and that new ones are certainly possible.

To tackle the problem of degenerative coefficients towards zero or infinity, we will rely on self-similar solutions [30, 31, 11] to build an adapted mesh. For the original ODE system, we will also explore existing schemes such as [49] to see whether they would apply in this new context or if coupling them with PDE schemes is efficient.

To ensure the project has a perfectly solid foundation, we will use these new methods for a complete validation of our model. Numerically, this will be done by a systematic comparison between the discrete and the continuous model, in a large variety of examples, where all the possible phenomena can occur: gelation, dust, asymptotic profile, even chaotic or oscillatory behaviours.

Theoretically, we aim to prove a complete theorem, justifying our approximation of the ODE system by the PDE one (1)–(3) and in particular shedding extra light on the nonlinear coupling boundary condition. This will also allow us to identify possible difficulties or need for adaptation in limit cases. Speeds of convergence of the discrete setting towards the continuous one is also a challenging open problem. We will investigate whether powerful estimates such as those obtained in [15], which already proved efficient in various applications [10], could help.

Risk evaluation: the non-degeneracy cases for non-asymptotic times present a lower risk ! The treatment of singular behaviours and time optimization are more delicate: how to face coefficients for which oscillations, gelation, dust formation etc. occur? How to properly treat the boundary condition at $x = 0$, and is it necessary to refine the grid at this point? Initial promising studies such as [11, 14] give a good starting point and we will build different approaches in order to ensure that observations of singular behaviors are not simply due to numerical errors [13].

Task 2: Nonlinear Dynamics of systems expressing growth and decay

We will address here the theoretical investigation of the direct problem (1)–(3) and its variants, *i.e.* the qualitative and possibly quantitative behaviour of solutions with respect to the initial data and parameter functions. Mathematically, **nonlinear dynamics of systems expressing growth and decay are a challenging field where many questions remain open**; indeed, new methods are necessary to analyse them, since the usual ones (entropy dissipation [36], Hamiltonian) do not work. Moreover, as shown by pioneer studies such as [12, 11], new mathematical problems are very likely to emerge in the course of the project, highlighting unsuspected properties of aggregation-fragmentation models. Durations as well as departure dates given in this work-package are only approximate; indeed, because we want to focus on biological questions, it is highly probable that my priorities will require adjustments. However, I give here a first set of original new problems that have been already identified as crucial to

the understanding of both the mathematical equations and the biological mechanisms.

To illustrate this, we have organised the mathematical problems deriving from the biological issues that triggered them. However, since they are formulated in general mathematical terms, they could apply to a wide range of other applications. Figure 2 illustrates, for a typical *in vitro* experiment, each particular phase of the reaction. This gives the outline of the following tasks.

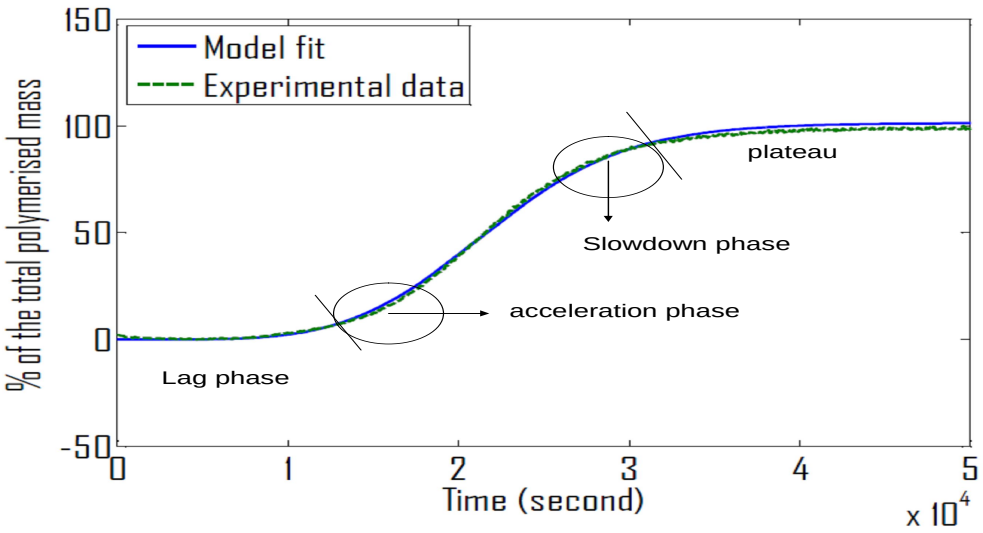

Figure 2: PolyQ polymerization with $V(0) = 100\mu Mol^{-1}$ and no monomers at initial time.

2.1 Probabilistic model and application to the *lag-phase* investigation: initiating the chain reaction

Start: year 2, **Duration:** 1-2 years, **Risk:** moderate / high

What biologists call *the lag phase* is the period during which the polymerization process slowly initiates [1]. It is characterized by a slow increase in the total polymerized mass. Asymptotic developments were proposed by [8] and complemented in more recent research such as [44], but all these cases were considered in the context of constant parameters (*i.e.* a constant polymerization rate or linear fragmentation). We want to build on such studies, looking at more general cases such as fast-varying polymerization rates and exploring very high or very low concentration effects.

To explore low concentrations, where experimental variations prove to be very large [51], **a microscopic statistical model will be built in collaboration with M. Hoffmann and N. Krell,** in order to quantify the observed variability between experiments, and to see how far we can rely on the deterministic ODE model. This way, we will extend our method to take into account stochastic interaction between individuals, linked to possibly intrinsic variations. This needs to develop probabilistic models for fragmentation. We will study how the probabilistic model at the microscopic level converges at the macroscopic level towards an ODE-controlled dynamics. By a statistical scales convergence analysis, we will also measure the modelling error of our models. This will lead us to define new statistical hypothesis tests capable of quantitatively answering questions such as the variability in the polymerization rate. For this part, we will also rely on studies concerning the cell division cycle for E. Coli, since it follows the same size-structured equation but is much easier to observe experimentally.

Risk evaluation: the risk of the asymptotic analysis is low, and although the risk of the statistical model is rather high it is attenuated thanks to the expertise of M. Hoffmann and N. Krell. This will have the notable effect of completing the existing theory of polymerization equations and will bridge stochastic and deterministic approaches of fragmentation processes.

2.2. Exponential phases investigation: the balance between growth, fragmentation and coalescence, and parameters dependence

Start: year 2, **Duration:** 3 years, **Risk:** moderate to high

These phases follow the lag phase, and consist, first, in an exponential-like increase of the total polymerized mass, and secondly, to a slow-down when the plateau is reached, all monomers becoming consumed. They raise several difficult and important mathematical problems. The first is to study under which conditions the solution is approximately equal to the first eigenvector of the related linearized problem (maybe at least for large concentrations); otherwise, under which conditions it is described by a self-similar profile. For the first phase, if we can ignore coalescence, degradation and depolymerization (which we can most probably do at the early stage of the reaction) the problem with an intermediate conformer V^* becomes:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\Big(\big(V^*(t)\tau(x)\big)\,u\Big) = -\beta(x)u(t,x) + 2\int_x^\infty k(x,y)\beta(y)u(t,y)dy, \tag{4}$$

$$\frac{dV^*}{dt} = k_+V - k_-V^* - \int_0^\infty \big(-V^*\tau(x)\big)u(t,x)dx, \qquad \frac{dV}{dt} = -k_+V + k_-V^*, \tag{5}$$

$$u(t,x=0) = \frac{k_{on}V^{i_0}}{k_{off}+\tau(0)V}. \tag{6}$$

On the one hand, we will investigate the nonlinear problem, in the spirit of the articles [28, 12]. In our recent study [15], we found optimal -or almost optimal- assumptions on τ, β and κ for a steady profile to occur in the linear transport-fragmentation equation (4). We will investigate the complementary cases, *i.e.* on which minimal assumptions mass transport towards infinity or yet dust formation occur in the linear equation. Here, refine compactness estimates or self-similar transforms as in [11] can reveal interesting methods to attack the problem.

On the other hand, when the reaction reaches the plateau, depolymerization cannot be neglected and polymerization vanishes. We will study whether the linearized eigenvalue problem around zero can help to understand the long-term solution.

If an asymptotic behaviour proves to be an accurate approximation in one or other of the situations, it would be of major interest to apply methods developed in [18, 19] for solving the inverse problem (we refer to Task 3). We will address this question both numerically, to guide the intuition, and theoretically, by a stability analysis in the spirit of [12].

Risk evaluation: whereas the more qualitative part of this task seems of moderate risk, its more theoretical part is more delicate, but is of major mathematical interest, in order to perform a complete theory of the linear transport-fragmentation equation.

2.3. Final phase investigation: nonlinear asymptotics, Lifshitz-Slyozov revisited

Start: year 2, **Duration:** 1-2 year, **Risk:** moderate

Departing now from fibrils as an initial condition at time $t = 0$ and with no monomers, the behaviour of the population follows either the Becker-Döring system, in a discrete way [3], or the Lifshitz-Slyozov equations [40], namely:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\Big(\big(V(t)\tau(x) - d(x)\big)\,u\Big) = 0 \tag{7}$$

$$\frac{dV}{dt} = \int_0^\infty \big(d(x) - V\tau(x)\big)u(t,x)dx, \tag{8}$$

$$u(t,x=0) = \frac{k_{on}V^{i_0}}{k_{off}+\tau(0)V}. \tag{9}$$

Although this system (except the boundary condition (9)) and its asymptotic limits have been widely studied, our biological focus sheds new light on this problem since our assumptions regarding the coefficients are far from the original ones as stated in the seminal paper of Lifshitz and Slyozov (1961). We will investigate the following questions.

1. Is it possible to obtain (as experimental observations seem to show) permanent oscillations? We will study under what conditions on the polymerization and depolymerization coefficients we can observe multiple

steady states for a given initial condition. Hopf bifurcation and the studies carried out for related equations such as Prion provide us with different research directions.

2. Is the boundary condition (9) satisfactory? Else, we will derive a new one, either with deterministic methods inspired by numerous previous studies or from a probabilistic model derived in Task 2.1.
3. Can we deduce the polymerization and depolymerization size-structured rates from the time observation of the total polymerized mass? This last question corresponds to the problem described in 3.1.1.

Task 3: Inverse Problem Solving

Inverse problem solving for size-structured models is a recent and active field of research [18, 17], but in our case many questions remain largely open. Indeed, to our knowledge, so little is known for certain about protein aggregation in amyloid diseases that we need to estimate almost all the types of parameters from experimental data: polymerization rate, fragmentation rate *and* fragmentation kernel, coalescence, etc. All the above steps, together with Task 3.0. described below, should allow us to break this large problem down into a number of smaller ones: first, determine the polymerization pathway (*i.e.* to know whether the dominant reaction occurs *via* monomer, dimer or $i-$mer addition). Secondly, determine the fragmentation kernel, for instance by isolating specific sizes of polymers at sufficiently low concentration so that coalescence does not interfere. Thirdly, determine fragmentation and coalescence rates - coalescence being favoured by high concentrations of polymers and low concentration of monomers.

3.0.: Definition and First study of the Inverse Problems to be Solved

Start: year 1, **Duration:** 3 years, **Risk:** High

This third work package is at the same time the key to our project and the most difficult part to anticipate exactly, since it most needs to align with the biological concerns and data. For these reasons, it also presents the highest risk, and could possibly have the most important impact on mathematical, biological and even social fields.

Due to these difficulties, **a preliminary step consists in defining the exact contours of the inverse problems to tackle.** This will be done mainly by the PI and H. Rezaei in close collaboration.

As described in the research paradigm of Figure 1, once the inverse problem is defined, we will first test simple existing inverse problem techniques such as least-squares methods [6] or existing software such as PREDICI [49], in order to have a better view of the problems we face and a first idea of parameter estimations. Such optimization methods being time-consuming, the more efficient the numerical schemes of Task 1, the further this part will go.

Another research direction for this overall study is the generalized sensitivity functions derived by H.T. Banks and co-authors [5]. It will allow us to gain a first insight into which part of the reaction which type of process influences. This method could also help us to design better experiments, in order to obtain experimental data at the right moment where it depends most on the parameters being sought.

Once these exploratory studies are done, the tasks are organized as follows. Subsection 3.1. below describes some of the specific problems that will be addressed, whereas possible research directions and methods are investigated in Task 3.2. For each problem of Task 3.1., one or several of the methods of Task 3.2. (or even new ones) will be applied according to what appears to be the most suitable at the time. It is probable that variational methods, such as adjoint-based variational strategy in data assimilation, could prove efficient for the fibrils depolymerization problem 3.1.1. and the partial size-structured data of Problem 3.1.2., whereas the eigenvalue method could probably apply to the size-structured data of Problem 3.1.2 and the hidden variables problem 3.1.3.

3.1. Inverse Problem Formulation

As stated at the beginning of Task 3, to investigate the polymerization mechanisms of a specific protein, we will use Task 3.0. to delineate the inverse problems to solve and break this large problem into a number of smaller ones. Below, we give three examples of the inverse problems that will be investigated. The research timetable may need to be adapted according to a specific working plan defined by Task 3.0., however these problems will definitely be studied at one stage or another, and moreover, they are of mathematical interest *per se*.

3.1.1. Inverse Problem for Lifshitz-Slyozov equation

Start: year 2, **Duration:** 2 years, **Risk:** Moderate

As explained in Task 2.3., based on a set of experiments already carried out by H. Rezaei on PrP fibrils depolymerization, an initial question is to investigate whether it is possible to estimate size-structured coefficients $d(x)$, $\tau(x)$ from time-dependent observations of the total polymerized mass $\int xu(t,x)dx$, in the framework of System (4)–(6).

This crucial issue belongs to the general problem of identifiability or observability of our system through the data provided. The overall view is provided by numerical simulations performed under Task 1 and by sensitivity functions under Task 3.0. Ideally, an observability inequality involving the parameters to identify and the given data will be obtained, justifying the well-posedness of the inverse problem. Data assimilation methods (Task 3.2.1.) will suggest research directions from numerical investigations.

The overall view is provided by numerical simulations performed under Task 1 and by sensitivity functions under Task 3.0. To go further, a well-adapted direction is provided by data assimilation methods (Task 3.2.1.)

3.1.2. Inverse Problem for nucleation process with partial size-structured data

Start: year 3, **Duration:** 2 years, **Risk:** moderate / high

Experimental data already available in H. Rezaei's team concern PrP protein, which is involved in Prion diseases. For this protein, size-structured data for relatively small polymers (up to 100-mers) are already available; but fibrils go to up to 10^4-mers and for such fibrils we only have rough size-structured distributions, obtained using a completely different experimental device. Merging these two kinds of data raises interesting statistical questions. The questions we will investigate are thus the following.

- To what extent can we deduce full size-structured coefficients from data concerning the smallest one? Here again, the observability condition must be investigated to validate any further inverse problem strategy implementation. Data assimilation methods (Task 3.2.1.) will bring us tools to numerically understand this question.
- If we observe a size-structured asymptotic profile, under which conditions will it also imply a steady profile for larger sizes? In such a case, we will apply the eigenvalue method of Task 3.2.2.

Risk evaluation: at the date of this proposal, the difficulty and risk of this problem are difficult to evaluate, due to the lack of information concerning the experimental data.

3.1.3. The general issue of hidden intrinsic variables

Start: year 3, **Duration:** 2 years, **Risk:** Moderate / High

More generally, we also want to investigate how we can obtain information on some hidden or un-measurable quantity from an indirect measurement. In a different but similar context of label-structured cells in collaboration with H.T. Banks [6], the same question is expressed as follows. In a structured population model, we observe a non-structuring but easily measurable quantity such as a fluorescence label (or here the radius of a polymer instead of its mass; or the total polymerized mass). The question is to understand under which conditions it can provide information on the intrinsic hidden variable that really governs the overall growth and fragmentation process.

This issue can be of great interest in many fields of application of structured population models. For instance, the age of cells is often difficult to measure, so that a method to measure it indirectly would be a major advance. For this general problem, both of the following methods could be investigated and tested and could bring ideas to build new and relevant approaches. We will also merge these two different methods, for instance by using the asymptotic profile in the direction of the intrinsic variable while building on a data assimilation method to estimate the parameters related to the measurable quantity.

3.2. Methods

We have seen above some of the problems that will be considered; let us now review which methods we want to build on. **3.2.1. Data assimilation: variational approach and filtering methods**

Start: year 2, **Duration:** 2 years, **Risk:** Low / Moderate

Data assimilation strategies [39], initially defined for evolution equations in weather forecasting and geophysics, are now widely used for evolution problems in chemical systems, biomechanical systems etc. We plan to investigate how such strategies can be applied to our problems.

In essence, data assimilation consists in merging model and data in order to circumvent the initial model uncertainties and improve the computed prediction. In the data assimilation context, we distinguish two different approaches: first, the variational approach which estimates the uncertainties by minimizing a least square criterion

involving the discrepancy between the data and the corresponding outputs from a model simulation [39]; secondly, the sequential approach which filters the uncertainties over time to stabilize the computed numerical system on the actual partially observed system – see [46] and references therein. The classic examples of this approach are Kalman based filtering methods. These two strategies are different in their practical use but both rely on the same fundamental *observability condition* which expresses the fact that observing the system even partially – through a time sequence of observed mean quantities in our case – is sufficient to compensate the lack of initial knowledge about the system [46]. Moreover, the two approaches can be proved to be equivalent in various cases. Despite their easier implementation in comparison to variational strategies, most sequential approaches are known to suffer from a "curse of dimensionality" as recalled by Bellman, and therefore are not widely used for systems derived from the discretization of partial differential equations.

However, some recent studies [38, 37] have paved the way to adapt filtering strategies to the joint state-parameter estimation of large dimensional problems. P. Moireau and co-authors have demonstrated the efficiency of some adapted filtering methods from a theoretical point of view [38] and also in a clinical context. In collaboration with this group, we expect to generalize these results to our polymerization-fragmentation problems as stated in Task 3.1. The benefits will be threefold.

- It seems very likely that our inverse problems, in their most general formulation, fall within the scope of data assimilation methods which will give a well identified methodological framework for solving them in practice;
- we can then rely on a generic data assimilation library called Verdandi (http://verdandi.gforge.inria.fr/) already available and developed by P. Moireau and co-authors, where their methods are available, alongside state-of-the-art classical data assimilation algorithms;
- the numerical results obtained will help us identify observability issues, for example by analyzing uncertainty covariances resulting from the application of well-adapted methods.

Risk evaluation: once the inverse problem has been clearly defined in Task 3.0., the Verdandi library and collaboration with the research group of D. Chapelle and P. Moireau will give a strong basis to this study and lower the risk. Its impact can be very high, especially for the biophysical community. The theoretical study of observability conditions is mathematically challenging and presents a higher risk.

3.2.2. Inverse problem method based on the Eigenvalue Problem

Start: year 3, **Duration:** 2 years, **Risk:** Moderate / High

As explained in Task 2.2., an alignment with asymptotic profiles governed by the linearized eigenvalue problem is likely to occur, even if temporarily, at several moments of a reaction process (*e.g.* during acceleration and slowdown, cf. Figure 2). In such a case, the inverse problem solution first developed in [43, 18] for the size-structured cell division equation, then extended to a more general case in [19], consists in considering the time-independent equation satisfied by this steady profile. This radically simplifies the inverse problem. Several open questions will be investigated, among which are the following.

- A study for general fragmentation kernels is provided in [19], but some numerical instabilities remain, probably linked to the intrinsic ill-posedness of the inverse problem even after proper regularization. To select the correct solution and avoid instabilities when reaching the simulation domain boundaries. This problem is linked to the general study of the following "dilation" equation: for $f \in V$ (V being an adequate function space), find $u \in V$ such that

$$u(x) + \int_x^\infty \kappa(x,y)u(y)dy = f. \tag{10}$$

  Since this equation has an infinite number of solutions in the distribution space, the numerical issue is to select the right one. A combination of different strategies coming from [18, 19] will be investigated.
- How to estimate not only the fragmentation rate β, but also the fragmentation kernel $k(x,y)$ and even the other parameters? Such problems will be also addressed in the context of filtering methods of Task 3.2.2., and are linked to identifiability problems.

There are broad applications of such new methods, from physical fragmentation processes to cell division. In the framework of protein polymerization, this is of major interest since it could not only give quantitative information on the parameters but also provide a biological interpretation of size-structured data [12].

3.2.3. Statistical setting: statistical tests and Goldenschluger and Lepski's method

Start: year 3, **Duration:** 2 years, **Risk:** Moderate / High

As shown by what we have discussed above, unknowns concerning polymerization processes are so numerous that it is necessary to circumscribe them with several efficient methods. In the biophysical literature, the first attempts have already been made to build statistical tests [51, 47]. The questions we address here are the following.

- To reject some reactions processes with certainty, the probabilistic models built in Task 2.1 will allow us to define new and well-adapted statistical hypothesis tests.
- Following [17], we will adapt the previously seen eigenvalue method of Task 3.2.2. in a statistical setting to the framework of protein polymerization.
- In the spirit of Goldenschlugher and Lepski's method, we will define the best regularization parameter (or bandwidth) of the inverse problem. We will compare such methods with deterministic error-free methods such as L-curve [22].

This task being halfway between statistics and PDE analysis will be carried out directly by the PI and statistician collaborators. The probabilistic model built in Task 2.1. will provide us with a statistical interpretation of the experimental data, which will allow us to address one of the above deterministic methods (either optimal filtering methods or the eigenvalue problem) in a mixed numerical/statistical way.

Task 5. Proposing Therapeutic Strategies and opening-up new research horizons.

Start: year 5 (possibly before), **Duration:** 1 year, **Risk:** High

This is the last and, from the point of view of society, the most ambitious challenge of this project: once the main mechanisms for a given disease will have been identified thanks to the inverse problem solving of Task 3, we will use our previous studies, especially the qualitative analysis of Task 2, to propose new therapeutic strategies, validate them numerically, and assist H. Rezaei's team to test them experimentally *in vitro*.

In order to optimally design these new experiments, we go back to Task 3.0 to find definite answers to the question of which parameters most influence the reaction at which moment. For this again, we will use sensitivity functions as defined in Task 3.0., with the expertise of H.T. Banks' team [5]. Control methods could also be investigated, through contacts with J-M. Coron's team.

This will be also the time for opening-up new research directions: new fields of application through contact with other teams of biologists; setting new challenges for Prion or Alzheimer's disease, such as the modelling of *in vivo* mechanisms, coupling with a cell division model, etc.

c. Resources

If this project is approved, I will be able to build a new junior research group within INRIA, located at the Paris-Rocquencourt Research Centre.

The research environment of INRIA is not only of outstanding quality, it is also particularly well-adapted to such a research project as SKIPPER^{AD} . Indeed, the goals of SKIPPER^{AD} are perfectly in line with one of the main objective and research direction of INRIA, which is to apply numerical and mathematical methods to medical and biological issues. INRIA will not fund PhD or post-doctoral positions but it could ensure some extra support if this came to be seen as necessary during the course of the project. It also makes it possible to collaborate with renowned experts in various fields and offers me excellent working conditions. Moreover, it will help me with technical assistance for software development, which is another of its fields of expertise.

The personnel of the proposal is the following.

- **the Principal Investigator (PI) will devote 57 % of her working time to the project.** Besides overall supervision and her role in identifying and initiating the mathematical challenges, she will pay particular attention

to the most delicate aspects which present higher risks: modelling, defining inverse problems (Task 3.0), interaction with probability and statistics (Tasks 2.1. and 3.2.3.), helping the biologists propose new therapeutic strategies (Task 4.2.). She will direct the external collaborations.

- **Human Rezaei is a key associate member** and a senior researcher at INRA, the French National Institute for Agricultural Research. A biophysicist and the leader of a research group on the biochemical and biophysical properties of Prion, he is internationally renowned for his findings on the structural properties of PrP (the Prion protein), as published in PNAS in 2004 (99 cit.), JMB in 2005 (48 cit.), Plos Pathogens in 2007 (40 cit.), etc. A specialist in the oligomerization pathways of Prion protein, he played a great part in the design and planning of the SKIPPER^{AD} project, and has high expectations of our findings for fibrillization. Moreover, he appreciates the added value of mathematical and modelling approaches to his speciality, which could bring new ideas to the field. **He will devote 30 $\%$ of his working time to the project.** This is essential since he will participate in the supervision of the project, highlighting the biological problems, designing the experiments, and validating our findings.

- **2 post-doctoral fellows (PF1 and PF2) will be recruited**, the first in year 2, the second in year 3, each of them being recruited for 2 years. Due to the interdisciplinary character of the project, it will be essential that at least one of them has experience in the mathematical modelling of biological processes. The best set-up would be one student having gained his/her PhD in biomathematics, and one student either in biomathematics or in experimental biology on proteopathies but with an MSc in mathematics or informatics. They will mainly work on Task 3 (inverse problem solving), at the interface with biologists.

- **2 PhD students (PhD1 and PhD2) will also be recruited.** The first one will be recruited in year 2, and will focus on numerical analysis (Task 1). The second one, recruited in year 3, will focus on theoretical aspects (Task 2). They will be selected in an international competition.

The work will mainly be carried out by the members of this focused team, who will provide the project with strong cohesion: a task distribution scheme is proposed in the table of Figure 4. The collaboration with H. Rezaei is fundamental, otherwise the project would become purely theoretical and there would be a high risk of drifting away from applications.

Other collaborators are contacts with internationally-renowned specialists who have agreed to take part to this project. Each of them possesses his or her own field of specialization and research interest, and will contribute expertise in this specific field, through short-stays invitations. The main experts are:

- Prof. H.T. Banks (USA, Raleigh), an internationally-renowned specialist in inverse problems
- Prof. M. Hoffmann (Paris), an internationally-renowned statistician.

Other identified collaborators are Pr. J.P. Zubelli (IMPA, Brazil) on inverse problems and Dr P. Moireau (INRIA), on optimal filtering; Dr. P. Reynaud-Bouret (Nice, France) on model selection and Goldenschluger and Lepski's method; Pr. T. Goudon on numerical analysis. Of course, the list is not exhaustive and new collaborations are likely to emerge, either with mathematicians or with biologists specialising in other proteopathies: one of the goals of this project is precisely to encourage and develop new collaborations and become a leading group on polymerization modelling.

Project costs

The costs of the project are summurised in the table of Figure 3. The total amount of the proposal reaches €1 203 569, since INRIA has agreed to fund 10 months of a PhD student. The personnel costs cover the members of the team as described above. The other costs concern laptops and software (evaluated slightly higher for the two last years due to the development of the platform), invited specialists (1 month per year), the organization of 2 workshops and travel costs.

References

[1] Alvarez-Martinez MT, et al. (2011) Dynamics of polymerization shed light on the mechanisms that lead to multiple amyloid structures of the Prion protein. *Biochim. et Biophys. Acta - Pr. and Pr.* 1814:1305 – 1317.

	Cost Category	month 1 to 18	month 19 to 36	month 37 to 54	month 55 to 60	Total
INRIA Direct Costs:	**Personnel:**					
	P.I.[2]	101 523	101 523	101 523	33 841	**338 409**
	Post docs	50 000	75 000	75 000	-	**200 000**
	PhD Students	75 120	112 680	56 765		**244 565**
	Other					**0**
	Total Personnel:	**226 643**	**289 203**	**233 288**	**33 841**	**782 974**
	Other Direct Costs:					
	Equipment	-	-	-	-	**-**
	Consumables (laptops)	2 500	2 500	2 500	0	**7 500**
	Travel	11 000	11 000	11 000	5 000	**38 000**
	Invited specialist	4 000	4 000	4 000		**12 000**
	Workshops	-	10 000	10 000	0	**20 000**
	Total Other Direct Costs:	**17 500**	**27 500**	**27 500**	**5 000**	**77 500**
	Total Direct Costs:	**244 143**	**316 703**	**260 788**	**38 841**	**860 474**
Indirect Costs (overheads):	20% of Direct Costs	48 829	63 341	52 158	7 768	**172 095**
Subcontracting Costs:	(No overheads)	-	-	-	-	**-**
Total Requested - INRIA :	**(by reporting period and total)**	**292 971**	**380 043**	**312 945**	**46 609**	**1 032 569**

	Cost Category	month 1 to 18	month 19 to 36	month 37 to 54	month 55 to 60	Total
INRA Direct Costs:	**Personnel:**					
	Senior Staff (H. Rezaei 30% of working time) INRA	42 750	42 750	42 750	14 250	**142 500**
	Total Personnel:	**42 750**	**42 750**	**42 750**	**14 250**	**142 500**
	Total Direct Costs:	**42 750**	**42 750**	**42 750**	**14 250**	**142 500**
Indirect Costs (overheads):	20% of Direct Costs	8 550	8 550	8 550	2 850	**28 500**
Subcontracting Costs:	(No overheads)	-	-	-	-	**-**
Total Requested - INRA :	**(by reporting period and total)**	**51 300**	**51 300**	**51 300**	**17 100**	**171 000**

TOTAL REQUESTED GRANT	**INRIA + INRA**					**1 203 569**

Figure 3: Budget Breakdown

[2] Ball JM, Carr J (1990) The discrete coagulation-fragmentation equations: existence, uniqueness, and density conservation. *J. Statist. Phys.* 61:203–234.

[3] Ball JM, Carr J, Penrose O (1986) The Becker-Döring cluster equations: basic properties and asymptotic behaviour of solutions. *Comm. Math. Phys.* 104:657–692.

[4] Ballesta A, Charles F, Rezaei H, Doumic M (2011) An Efficient Kinetic Model for Amyloid Fibrils Assemblies - Application to Huntington's Disease.

[5] Banks HT, Ernstberger SL, Grove SL (2007) Standard errors and confidence intervals in inverse problems: sensitivity and associated pitfalls. *J. Inverse Ill-Posed Probl.* 15:1–18.

[6] Banks H, et al. (2010) Estimation of cell proliferation dynamics using cfse data. *Bull. of Math. Biol.*

[7] Bernacki J, Murphy R (2009) Model discrimination and mechanistic interpretation of kinetic data in protein aggregation studies. *Biophysical Journal* 96:2871 – 2887.

[8] Bishop M, Ferrone F (1984) Kinetics of nucleation-controlled polymerization. a perturbation treatment for use with a secondary pathway. *Biophysical Journal* 46:631 – 644.

[9] Bourgade JP, Filbet F (2008) Convergence of a finite volume scheme for coagulation-fragmentation equations. *Math. Comp.* 77:851–882.

[10] Cáceres M, Cañizo J, Mischler S (2010) Rate of convergence to an asymptotic profile for the self-similar fragmentation and growth-fragmentation equations. Preprint, arXiv:1010.5461.

[11] Calvez V, Doumic M, Gabriel P (2011) Self-similarity in a general aggregation-fragmentation problem; application to fitness analysis. *JMPA* accepted.

„key intermediate goal", as defined in section 2.b.	Estimated % of total requested grant	Expected to be completed on month :	Comment
Task 1. Numerical schemes	15%	54	PhD 1 will be recruited on the beginning of Year 2. Total completion of the work will be long and continued throughout the project.
Task 2.1. Probabilistic model	10%	36	PhD2 will devote mostly on Task 2.
Task 2.2. Parameters dependency	5%	30	PhD2 will devote mostly on Task 2.
Task 2.3. Nonlinear studies - Lifshitz-Slyozov revisited	10%	18	PhD2 will devote mostly on Task 2.
Task 3.0. First Inverse Problem Investigation	5%	18	Mainly M. Doumic and H. Rezaei
Task 3.1.1. Inverse Problem for Lifshitz-Slyozov	5%	36	PF1.
Task 3.1.2. Inverse Problem for size-structured	5%	18	PhD and PF
Task 3.1.3. Hidden variables	5%	36	with external collaborators
Task 3.2.1. Kalman Filter method	15%	54	PhD and PF
Task 3.2.2. Eigenvalue method for Inverse Problems	5%	36	a first study should be already completed on month 18
Task 3.2.3. Model Selection - Goldenschluger & Lepski's method	5%	54	with external collaborators
Task 5. Therapeutic optimization	15%	60	Task 4 has been removed to the project
Total	**100%**		

Figure 4: Table on Resources Allocation. The key intermediate goals consist in each task and subtask described in Section 2.b.

[12] Calvez V, et al. (2009) Size distribution dependence of Prion aggregates infectivity. *Math. Biosci.* 1:88–99.

[13] Carrillo JA, Goudon T (2004) A numerical study on large-time asymptotics of the Lifshitz-Slyozov system. *J. Sci. Comput.* 20:69–113.

[14] Clairambault J, Gaubert S, Lepoutre T (2009) Comparison of perron and floquet eigenvalues in age structured cell division cycle models. *Math. Model. Nat. Phenom.* p in press.

[15] Doumic Jauffret M, Gabriel P (2010) Eigenelements of a general aggregation-fragmentation model. *Math. Models Methods Appl. Sci.* 20:757–783.

[16] Doumic M, Goudon T, Lepoutre T (2009) Scaling limit of a discrete Prion dynamics model. *Communications in Mathematical Sciences* 7:839–865.

[17] Doumic M, Hoffmann M, Reynaud P, Rivoirard V (2011) Nonparametric estimation of the division rate of a size-structured population. submitted.

[18] Doumic M, Perthame B, Zubelli J (2009) Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems* 25:045008.

[19] Doumic M, Tine L (2011) A general inverse problem for aggregation-fragmentation equation. submitted.

[20] Eigen M (1996) Prionics or the kinetic basis of Prion diseases. *Biophys. Chem.* 63:A1–A18.

[21] Eghiaian F, et al. (2007) Diversity in prion protein oligomerization pathways results from domain expansion as revealed by hydrogen/deuterium exchange and disulfide linkage. *PNAS* 104:7414–7419.

[22] Engl H, Hanke M, Neubauer A (1996) *Regularization of Inverse Problems* (Springer Verlag).

[23] Engler H, Pruss J, Webb G (2006) Analysis of a model for the dynamics of Prions ii. *J. of Math. Anal. and App.* 324:98–117.

[24] Escobedo M, Laurençot P, Mischler S, Perthame B (2003) Gelation and mass conservation in coagulation-fragmentation models. *J. Differential Equations* 195:143–174.

[25] Escobedo M, Mischler S, Rodriguez Ricard M (2005) On self-similarity and stationary problem for fragmentation and coagulation models. *Ann. Inst. H. Poincaré Anal. Non Linéaire* 22:99–125.

[26] Filbet F, Laurençot P (2004) Numerical simulation of the Smoluchowski coagulation equation. *SIAM J. Sci. Comput.* 25:2004–2028 (electronic).

[27] Gabriel P, Tine L (2010) High-order WENO scheme for polymerization-type equations. *ESAIM Proc.* 30:54–70.
[28] Greer M, Pujo-Menjouet L, Webb G (2006) A mathematical analysis of the dynamics of Prion proliferation. *J. Theoret. Biol.* 242:598–606.
[29] Haupt C, et al. (2011) Pattern recognition with a fibril-specific antibody fragment reveals the surface variability of natural amyloid fibrils. *Journal of Molecular Biology* 408:529 – 540.
[30] Herrmann M, Laurençot P, Niethammer B (2009) Self-similar solutions with fat tails for a coagulation equation with nonlocal drift. *Comptes Rendus Mathematique* 347:909 – 914.
[31] Herrmann M, Niethammer B, Velázquez J (2009) Self-similar solutions for the LSW model with encounters. *Journal of Differential Equations* 247:2282 – 2309.
[32] Knowles T, et al. (2009) An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* 326:1533–1537.
[33] Masel J, Jansen V, Nowak M (1999) Quantifying the kinetic parameters of Prion replication. *Biophysical Chemistry* 77:139 – 152.
[34] Michel P (2006) Existence of a solution to the cell division eigenproblem. *Math. Models Methods Appl. Sci.* 16:1125–1153.
[35] Michel P (2006) Optimal proliferation rate in a cell division model. *Math. Model. Nat. Phenom.* 1:23–44.
[36] Michel P, Mischler S, Perthame B (2005) General relative entropy inequality: an illustration on growth models. *J. Math. Pures Appl. (9)* 84:1235–1260.
[37] Moireau P, Chapelle D (2011) Reduced-order Unscented Kalman Filtering with application to parameter identification in large-dimensional systems. *Cont. Optim. and Calc. Variat.* 17:380–405.
[38] Moireau P, Chapelle D, Le Tallec P (2008) Joint state and parameter estimation for distributed mechanical systems. *Computer Methods in Applied Mechanics and Engineering* 197:659–677.
[39] Navon I (2009) in *Data assimilation for atmospheric, oceanic, hydrologic applications* (Springer).
[40] Niethammer B, Pego R (2000) On the initial-value problem in the Lifshitz-Slyozov-Wagner theory of Ostwald ripening. *SIAM J. Math. Anal.* 31:467–485 (electronic).
[41] Oosawa F, Asakura S (1975) *Thermodynamics of the polymerization of protein* (Academic Press).
[42] Pallitto M, Murphy R (2001) A mathematical model of the kinetics of β-amyloid fibril growth from the denatured state. *Biophysical Journal* 81:1805 – 1822.
[43] Perthame B, Zubelli J (2007) On the inverse problem for a size-structured population model. *Inverse Problems* 23:1037–1052.
[44] Powers E, Powers D (2006) The kinetics of nucleated polymerizations at high concentrations: Amyloid fibril formation near and above the supercritical concentration. *Biophysical Journal* 91:122 – 132.
[45] Silveira J, et al. (2005) The most infectious Prion protein particles. *Nature* 437:257–261.
[46] Simon D (2006) *Optimal State Estimation: Kalman, H^∞, and Nonlinear Approaches* (Wiley-Interscience).
[47] Skanda D, Lebiedz D (2010) An optimal experimental design approach to model discrimination in dynamic biochemical systems. *Bioinformatics* 26:939–945.
[48] L.M. Tine, T. Goudon FL (2011) Simulations of the lifshitz-slyozov equations: the role of coagulation terms in the asymptotic behavior. *submitted.*
[49] Wulkow M (1996) The simulation of molecular weight distributions in polyreaction kinetics by discrete galerkin methods. *Macromol. Theory Simul.* 5:396–416.
[50] Xu Z, Paparcone R, Buehler M (2010) Alzheimer's aβ(1-40) amyloid fibrils feature size-dependent mechanical properties. *Biophysical Journal*.
[51] Xue WF, Homans S, Radford S (2008) Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *PNAS* 105:8926–8931.

d. Ethical and Security sensitivity Issues

ETHICS ISSUES TABLE

	Research on Human Embryo/ Foetus	YES	Page
	Does the proposed research involve human Embryos?		
	Does the proposed research involve human Foetal Tissues/ Cells?		
	Does the proposed research involve human Embryonic Stem Cells (hESCs)?		
	Does the proposed research on human Embryonic Stem Cells involve cells in culture?		
	Does the proposed research on Human Embryonic Stem Cells involve the derivation of cells from Embryos?		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

	Research on Humans	YES	Page
	Does the proposed research involve children?		
	Does the proposed research involve patients?		
	Does the proposed research involve persons not able to give consent?		
	Does the proposed research involve adult healthy volunteers?		
	Does the proposed research involve Human genetic material?		
	Does the proposed research involve Human biological samples?		
	Does the proposed research involve Human data collection?		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

	Privacy	YES	Page
	Does the proposed research involve processing of genetic information or personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?		
	Does the proposed research involve tracking the location or observation of people?		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

	Research on Animals[4]	YES	Page
	Does the proposed research involve research on animals?		
	Are those animals transgenic small laboratory animals?		
	Are those animals transgenic farm animals?		
	Are those animals non-human primates?		
	Are those animals cloned farm animals?		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

[4] The type of animals involved in the research that fall under the scope of the Commission's Ethical Scrutiny procedures are defined in the Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes Official Journal L 358 , 18/12/1986 p. 0001 - 0028

	Research Involving non-EU Countries (ICPC Countries[5]) [6]	YES	Page
	Is the proposed research (or parts of it) going to take place in one or more of the ICPC Countries?		
	Is any material used in the research (e.g. personal data, animal and/or human tissue samples, genetic material, live animals, etc) : a) Collected in any of the ICPC countries?		
	b) Exported to any other country (including ICPC and EU Member States)?		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

	Dual Use	YES	Page
	Research having direct military use		
	Research having the potential for terrorist abuse		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

[5] In accordance with Article 12(1) of the Rules for Participation in FP7, 'International Cooperation Partner Country (ICPC) means a third country which the Commission classifies as a low-income (L), lower-middle-income (LM) or upper-middle-income (UM) country. Countries associated to the Seventh EC Framework Programme do not qualify as ICP Countries and therefore do not appear in this list.

[6]