Systems-wide Identification of Cystic Fibrosis Disease Map

Objectives: To develop a computational model of molecular mechanisms which are affected in the genetic disease cystic fibrosis, to quantitatively elucidate the underlying molecular and cellular events.

Methodology: Cystic fibrosis (CF) is the most frequent of fatal monogenic disorders, affecting ~30,000 solely in Europe. CF causes major respiratory symptoms and ultimately lung failure. The causative gene (CFTR) encodes a chloride (Cl^-)/ bicarbonate(HCO_3^-) channel expressed at the apical membrane of epithelial cells. CFTR was identified ~25 years ago and extensive research on its pathophysiology has been pursued ever since. Despite the apparent "simplicity" of being a monogenic disorder, CF is indeed a complex disease. There are more than 2,000 CF-causing mutations whose effect on the severity of the lung phenotype is not easy to predict. CFTR is also a key regulator of other channels such as the epithelial sodium (Na^+) channel (ENaC). Consequently, the mutations in the CFTR protein also cause Na^+ hyperabsorption. Moreover, to compensate for the absence of CFTR, the epithelium upregulates the calcium (Ca^{2+})-activated Cl^- channels (CaCCs, now identified as the ANO1 protein), but this response is not sustained due to a not yet understood mechanism. Altogether, these complex synergistic effects trigger a cascade of events [1]: dehydration of the airway surface liquid (ASL), mucus thickening, recurrent bacterial infections and ultimately irreversible lung damage, making CF a devastating disease. The failure in understanding CF pathophysiology stems from the difficulty in integrating extensive data from the molecular, functional and physiological levels. CF, however, presents a set of features, which may lead this disease to become a fully understood process via a Systems Biology approach:

1. A single genetic defect (mutation F508del) accounts for the disease in almost 90% of CF patients, thus constituting a single system perturbation identified at the molecular level.
2. The defect of F508del is well characterized at the cellular level as a failure in trafficking to the plasma membrane due to premature degradation at multiple checkpoints of the endoplasmic reticulum (ER) quality control [2]. The trafficking pathways can be studied at the cellular level;
3. At the tissue level, there is a clear electrophysiological quantitative output, defined by the measurable transport rate of ions (Na^+ and Cl^-) across the epithelia. Quantitative methods are also in place to assess these outputs at the organ and individual level, i.e. in patients;
4. There are known (but weak) corrector compounds of the basic defect (i.e. the F508del-CFTR processing defect is druggable) [1,3]. These compounds are "tools" to investigate how the pathological network can be restored to a more physiological state [4], while revealing their mode of action;
5. A large body of 'Omics data' are available on CF, including those generated by ourselves [5-9], which will be cast into an draft multi-scale computational model right at the beginning of the project.

The purpose of this project is threefold.
In the first place, to integrate all the available 'Omics resources for a broader view of CF allowing for an integrated perspective of all the available knowledge which will include the relevant molecular interactions and all the relevant regulatory pathways. This will involve modeling the available data in a flexible information system that is able to encompass all the relevant information and metadata gathered from the available resources. Information from literature on disease mechanisms will be integrated into a large-scale model. Pathway information from available databases (for example Reactome, PANTHER) will be contextualised and reused where possible. The network will be presented in the standard Systems Biology Graphical Notation (SBGN) languages. A community of

domain experts will be involved in order to ensure that all the relevant disease hallmarks are adequately represented. The resulting network will be made available online as a reference resource for the molecular mechanisms involved in CF.
Secondly a logical model network will be constructed from the available data; this model may encompass a Bayesian or Markov model.
The third part will involve testing and stressing the developed model so that it is possible to make inference by changing and modifying the model inputs. Results will be verified against available experimental data and new essays will be proposed to validate the model consistency.

[1] Amaral MD, Kunzelmann K (2007) Molecular targeting of CFTR as a therapeutic approach to cystic fibrosis. *Trends Pharmacol Sci* **28**: 334-341. [PMID: 17573123].
[2] Farinha CM & Amaral MD (2005) Most F508del-CFTR is targeted to degradation at an early folding checkpoint and independently of calnexin. *Mol Cell Biol* **25**: 5242-5252. [PMID: 15923638].
[3] De Boeck K, Amaral MD (2016) Highlights of progress in therapies for cystic fibrosis. *Lancet Respir Med*. Epub April 1. [PMID: 27053340]
[4] Amaral MD, Balch WE (2015) Hallmarks of therapeutic management of the cystic fibrosis functional landscape. *J Cyst Fibros* **14**: 687-99. [PMID: 26526359].
[5] Clarke LA, Botelho HM, Sousa L, Falcao AO, Amaral MD (2015) Transcriptome meta-analysis reveals common differential and global gene expression profiles in cystic fibrosis and other respiratory disorders and identifies CFTR regulators. *Genomics* **106**:268-77. [PMID: 26225835].
[6] Botelho HM, Uliyakina I, Awatade NT, Proença MC, Tischer C, Sirianant L, Kunzelmann K, Pepperkok P, Amaral MD (2015) Protein Traffic Disorders: an Effective High-Throughput Fluorescence Microscopy Pipeline for Drug Discovery. *Sci Rep* **5**: 9038. [PMID: 25762484].
[7] Almaça J, Faria D, Sousa M, Uliyakina I, Conrad C, Sirianant L, Clarke LA, Martins JP, Santos M, Hériché JK, Huber W, Schreiber R, Pepperkok R, Kunzelmann K, Amaral MD (2013) High-content siRNA screen reveals global ENaC regulators and potential cystic fibrosis therapy targets. *Cell* **154**: 1390-1400. [PMID: 24034256].
[8] Clarke LA, Sousa L, Amaral MD (2013) Changes in transcriptome of native nasal epithelium expressing F508del-CFTR and intersecting data from comparable studies. *Resp Res* **14**: 38. [PMID: 23537407].
[9] Roxo-Rosa M, da Costa G, Luider TM, Scholte BJ, Coelho AV, Amaral MD, Penque D (2006) Proteomic analysis of nasal airway cells from cystic fibrosis (CF) Patients and non-CF control individuals: search for novel biomarkers of lung disease. *Proteomics* **6**: 2314-2325. [PMID: 16518875].

Key words: Cystic Fibrosis, Omics data, systems biology, data integration, disease maps

Supervisor: André Falcão (FCUL-BioISI) **Co-Supervisors:** Margarida D. Amaral, BioFIG/FCUL;
Alexander Mazein (SGBN, European Institute for Systems Biology and Medicine)

Type of fellowship:
Mixed (Portugal and abroad): FCUL/(LaSIGE and BioISI) and SGBN (http://www.sbgn.org)