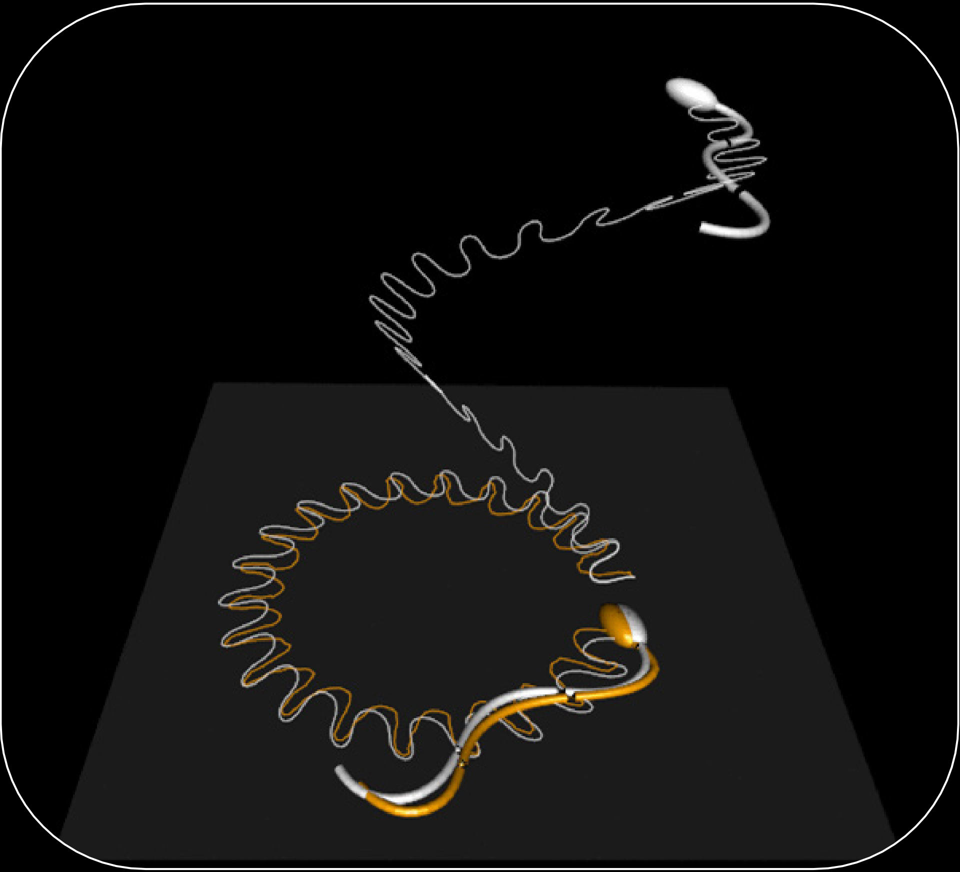
Quantitative image analysis of cells using morphodynamical models:

Sea urchin spermatozoa as case study

Pedro Ângelo Pereira da Silva



Dissertation presented to obtain the Ph.D degree in Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
April, 2017

Quantitative image analysis of cells using morphodynamical models:

Sea urchin spermatozoa as case study

Pedro Ângelo Pereira da Silva

Dissertation presented to obtain the Ph.D degree in Biology
Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras, April, 2017

Summary

Digitization and robotization of laboratory equipment has recently contributed to the generation of high content of data and its metadata. While this seems like an advantage for science's celerity, the analysis of such data became the limiting step – a very narrow bottleneck. Such is the case for imaging data acquisition and its analysis. After collecting Gigabytes of images, researchers spend several orders of magnitude of more time to determine the regions of interest (ROIs) (*e.g.* cell) and to measure relevant attributes (*e.g.* mean fluorescence intensity). This manual curation of data promotes another issue that is related with the reproducibility of the analysis, *e.g.*, the same researcher will hardly select the exact same ROIs in the same data set. Furthermore, there is also the possibility of bias in the selection of which cells to use in the analysis by biased determination of the ROIs. All of these considerations can be solved by automation of imaging data analysis. Given the same initial parameters, the analysis program will select the same ROIs, measure and process those measurements in the same, repetitive way, thus producing reproducible analysis. By checking its input parameters, one can also analyze if there is a bias in the automatic selection or in the analysis pipeline. Finally, the processing speed of current central processing units (CPUs) and graphic processing units (GPUs) allow for fast analysis, which is crucial to close the gap between data acquisition and analysis.

The idea of automation of imaging analysis is far from new and there are many commercial and open sourced programs available for either simple or complex analysis procedures. However, most of them can only be applied to specific model organism types and, sometimes, only if they are used under a specific experimental protocol. One major aspect hindering this progress is that most approaches rely on segmentation algorithms (*i.e.* classification of pixels as ROI, background or debris) and filtering (*i.e.* rejection/acceptance of ROIs according to their characteristics). In practice, the computer algorithm does not know what it is looking for in the

data and often includes different objects (including debris) or misses parts of the objects at study. This aspect will promote bias in the subsequent analysis steps, potentially creating artifactual data. We propose to use a mathematical model as *a priori* knowledge of the target model cell to be fitted directly to imaging data. Our hypothesis is that, by estimating the number and form of the particular objects of interest that best describe an image (*i.e.* maximum likelihood), we can obtain the most probable features and characteristics of those objects, thus improving their detection, tracking and characterization.

Image analysis of spermatozoa is an example where image segmentation is the primary method used, and therefore these cells were chosen as the biological case study to explore our hypothesis. Spermatozoa are highly specialised motile cells whose function is to find and fertilize the conspecific eggs during reproduction. For this reason they are cells of the utmost fundamental importance and also of major medical relevance, considering that sperm motility abnormalities are a major cause of human couple infertility. Not surprisingly Computer Assisted Sperm Analysis (CASA) systems, implementing image segmentation have been developed early and made their way into routine clinical practice. In mammals, sperm cells must swim a path thousands of times their own body length through a complex interior geometry, often filled with highly viscous liquids and potentially hostile immune cells. The overwhelming majority do not even reach the fallopian tubes, let alone the site of fertilization. In marine invertebrates, fertilization occurs during broadcast spawning events in which the sperm have to find their conspecific eggs literally in a sea of eggs of multiple species. In the laboratory studies, sea urchin spermatozoa tend to accumulate and swim confined to the liquid-solid boundary plane which made them particularly suitable for imaging motility and chemotaxis in response to molecular cues released by the egg. For this reason sea urchin sperm became an important biological paradigm. Several studies have shown that sperm respond to sperm activating peptides released from the egg jelly layer. The transduction of these signals results in series of cytosolic calcium spikes that are concomitant with transient periods of greater asymmetry of the flagellar bending waves that reorient the cellular trajectory. In sperm of some species of sea urchin, such as *Lytechinus pictus* the calcium spike trains are coordinated in space and time to produce chemotactic trajectories towards the source of the peptides, while in sperm of *Strongylocentrotus purpuratus* the calcium spikes produce disoriented behavior. However, the fact that the natural environment of the spermatozoa is a three dimensional volume has raised concerns on the relevance and generality of the knowledge derived from studies in the plane. Since spermatozoa are a class of very fast moving cells they are particularly demanding on time-lapse and three-dimensional microscopy imaging instruments. Particularly challenging is the measurement of fluorescent reporters on beating flagella in three dimensions that cannot be imaged with currently available high-performance confocal microscopes. To this end, imaging systems tailored to image sperm cells have been developed in the last two decades. One of the systems, the 2D+$Z(t)$ system (*i.e.* reported in the literature as 3D+t), uses a piezoelectric device to oscillate an objective at high frequencies, allowing us to take two-dimensional (2D) images (frames) at different depths (Z axis) as a function of time. Using this microscopy instrumentation it was shown that *S. purpuratus* sperm displayed different average path velocity and curvature when confined or when free-swimming. Unfortunately, the depth of each frame is not reported by the system and an algorithm based on image correlation and on fitting the characteristic $Z(t)$ function of the piezoelectric was used to determine it. Although this method performed adequately in various experimental data sets, it performed poorly in many others in which the real and the inferred depth functions became out-of-phase. This has been impairing comparative studies of the free-swimming motility and chemotaxis of spermatozoa from the two sea urchin species.

The objective of this thesis was two-fold. The first was to develop and apply image analysis methods based on *a priori* knowledge by fitting a mathematical model of the object to be detected or tracked directly to imaging data. The second objective was to develop and apply methods to compare the 3D motility of spermatozoa of *L. pictus* and *S. purpuratus*.

Chapter 2 is a proof-of-concept that a cellular model can be used for 2D imaging data analysis. A detailed mathematical model was developed describing the spermatozoon morphodynamics and how it swims in a fluid, given the changes of its morphology. Thus, this mechanistic model includes two modules: the shape, where the head is a revolution ellipsoid and the flagellar beating is given by defining the local curvature as traveling wave function and the local torsion as a constant value; and the mechanics, where physics are modeled by Resistive Force Theory. This model was compared to imaging data using the following procedure. First we take the current model state and render a model image corresponding to that state, then we rescale it to the imaging data spatial and temporal resolutions and lastly convolve it with the point-spread function characteristic of fluorescence microscopy, if that is the case. The model image is compared to the corresponding time-point of imaging data by correlation. It was shown that, under certain conditions, maximizing the cross-correlation is equivalent to maximizing the likelihood (*i.e.* how likely the parameters are given by the data). As our model is non-linear, we maximized the likelihood of our model parameters by a simple evolutionary Monte Carlo simulation, which selects and propagates the parameters with highest correlation coefficient (single time-point) or highest correlation sum (multiple time-points, *i.e.* the summation of the all correlation coefficients) over some iterations. The propagation step allows for small changes in the parameters to explore the optimization landscape but these changes become smaller at each iteration in order to promote exploitation.

We fitted the morphodynamical model of a sperm cell to different imaging data sets. The first imaging data set was generated using the model itself (*in silico*) to prove that the fitting algorithm is able to recover the parameters the original parameters. We show we can decrease the distance between the initial parameter set (*i.e.* randomized) and the parameter selected after fitting for three independent initial sets. A second data set was used to compare the performance of our method and that of a human. For this, an imaging data of *L. pictus* sperm was acquired with high spatial and temporal resolution. A cell in this data set was tracked semi-automatically by a collaborator reconstituting both position and flagellar conformations. The same cell was fitted by our model to show that the solutions obtained also minimize the distance between the human-determined and model-fitted conformations. Note the discrepancies can originate on the assumption of our model that the flagellar shape parameters are constant, which is more likely to break down as time elapses. To assess whether our method can track cells in images with lower spatial and temporal resolution, we fitted our model to imaging data of *S. purpuratus* in those conditions and the results were similar to the ones with high resolution data. Finally, we hypothesized that we should be able to estimate features of the cells that are modeled but not visible in the imaging data. To test it we used imaging data obtained from spermatozoa labelled with both a fluorescent membranar marker and a fluorescent Ca^{2+} indicator. The former labels the whole cell and the latter labels only the head in an unstimulated cell. On the same cell and at the same time, we can image both markers using a light splitter to produce the corresponding two images in the same frame. Fitting the model independently to the whole cell or head-only imaging data we arrived to very close parameter sets and flagellar positions. Altogether, these results show we can fit a model directly to data and even estimate features that cannot be directly measured in the data, an advantage over segmentation methods that rely on the data itself.

Chapter 3 tackles the problem of analysing and comparing the motility of the sperm of the two sea urchin species while confined to planar swimming or while moving freely in a volume. Analysing spermatozoa swimming in a volume requires special imaging systems. We used the system proposed by Corkidi et al., (2008) which implies inferring the depth of each frame from its time-point before obtaining the 3D coordinates of the cells. To deal with this inference problem we started by estimating the times at which the piezoelectric was at either a maximum or a minimum depth position. Based on the assumption that cells do not displace significantly within a piezoelectric period, meaning frames at the same depth position should be highly correlated, we used the collective information of the correlation of all frames and their subsequent images. After we fitted the canonical, empirically determined piezoelectric $Z(t)$ function between those extremes. The computational speed of the implementation of the method developed in *Chapter 2* was low and its application to data containing thousands of frames (*i.e.* as the ones produced by the 2D+$Z(t)$ microscopy system) was not computationally feasible. As the optimization of such algorithms is not the focus of this thesis, we used a mixed approach to detect cells and reconstitute their 3D trajectories from 2D+t data. To detect cells we used a more traditional 3D template of the spermatozoon composed by 2D frames, each corresponding to a different diffraction pattern resulted from the relative offset between the centroid of the cell and the objective. Each diffraction pattern was cross-correlated to each frame and the position of the maximum correlation position was saved, from which we calculated the centroid position after clustering (*i.e.* to know which saved positions belong to the same cell). These detections were then attributed to cells by clustering according to their distances in space and time. To infer the average path of the cells (neglecting the fine wiggling of the heads), we fitted the minimal set of helical segments that explained the trajectory points of each cell using Bayesian information criteria as the scoring criteria of a Dynamical Programming problem, a method we called piecewise helix fitting.

S. purpuratus spermatozoa were studied both in confined and free-swimming modes but we know nothing about the latter mode for *L. pictus*. As we also wanted to assess the performance of our method, we used it to determine the velocity, curvature and torsion of both species in both free and confined swimming. Our results for confined swimming of both species and free-swimming for *S. purpuratus* were consistent with previous studies. We show, for the first time, that the trajectories of free-swimming spermatozoa from *L. pictus* have lower velocity than those of *S. purpuratus*, contrary to when they are confined to the surface. Also, *L. pictus* sperm have lower trajectory torsion than those of *S. purpuratus*. Interestingly, the trajectory curvature *L. pictus* sperm does not change between confined and free swimming modes, while it is different for *S. purpuratus* sperm. Recent studies with *Arbacia punctulata* show their sperm trajectories are more similar to the ones of *L. pictus*. To try to gain insight into the different confining behaviors of the spermatozoa of the three species, we then used the mechanistic sperm model to simulate both free and confined swimming when flagellar torsion is non-zero. We searched the parameter sets which would produce trajectories similar to the ones of these species. We showed the data can be explained if free swimming *S. purpuratus* sperm have higher mean flagellar curvature (and possibly higher flagellar torsion) than in confined swimming. We argue that this difference results from a higher sensitivity of the sperm of this species to the higher viscosity of the liquid boundary, when compared to the sperm of the two other species.

Using the 2D+$Z(t)$ imaging analysis procedure developed in *Chapter 3*, we then addressed the free-swimming chemotactic behavior of both *L. pictus* and *S. purpuratus* sperm in *Chapter 4*. In studies of chemotaxis of these sperm in confined swimming, a gradient of a chemoattractive peptide released by the conspecific egg (Speract) was artificially created. To do this, a chemically blocked, caged form of the peptide with 1000 fold less affinity for its receptor was used in solution and a concentration gradient of the active form was produced by modulated ultraviolet (UV) irradiation (uncaging). The uncaging procedure was used to produce a chemoattractant gradient in the 3D microscopy system. Two conditions were tested: with and without caged chemoattractant, irradiating both with UV light for 2 seconds, defining time intervals before, during and after irradiation. Trajectories were reconstituted by piecewise helical fitting and time series for each of the helical path parameters obtained. To deal with the dependency of data points within the time intervals, we fitted linear mixed models to the data. Although this linear mixed model approach readily distinguished the parameters obtained for the two species, we could not find any alteration of the parameters consistent with chemotaxis elicited by uncaging of the chemoattractant. We interpret this result as a consequence of an ineffective experimental creation of the gradient which was presumably undetected by the cells. We then suggested several alterations to the protocol in order to assess whether other experimental conditions might promote chemotaxis. In another study, free-swimming chemotaxis of *A. punctulata* sperm was achieved by using sustained UV irradiation throughout the experiment. A similar strategy might be effective also with the sperm of the two species studied here. Alternatively, different gradients could be created by modulating the irradiation profile using different optical fibers.

Finally, we discuss our thesis that using *a priori* knowledge in the form of mathematical model of the target cell or organism to fit to data can increase our resolution and allow us to infer features that are not visible nor able the to directly measured from the data. It is possible to expand this framework to the target biological model by making the appropriate adaptations. In the case study of sea urchin sperm, we can envisage extending this framework to include a signaling module to analyze sperm chemotaxis. We were also able to improve the determination of the depths of the frames of the 2D+$Z(t)$ system. Furthermore, we improved the analysis of that data in order to resolve the different motility and confining behavior of two different species. The application of our morphodynamical model allowed us to suggest that the flagellar beating of *S. purpuratus* sperm has higher mean curvature on free swimming when compared to confined. Unfortunately, we could not significantly detect chemotaxis for either species in the conditions tested and different gradients should be assessed either by changing the diameter of the UV light fiber or the duration of the UV irradiation. Overall, the frameworks developed here, together with other recent techniques and methods, may prove instrumental for imaging data analysis and, more specifically, to understand spermatozoan chemotaxis.

Sumário

A digitalização e robotização do equipamento de laboratório contribuiu para a recente rápida aquisição de dados e dos seus meta-dados. Embora este facto possa parecer uma vantagem para o avanço célere da ciência, a análise destes dados tornou-se no fator limitante. Este é também o caso da aquisição e análise de dados de imagem. Após colecionar Gigabytes de imagens, os investigadores investem uma quantidade de tempo várias vezes superior ao de aquisição para determinar as regiões de interesse (ROIs) (*e.g.* célula) e para medir as suas características relevantes (*e.g.* intensidade de fluorescência média). Esta manipulação manual tem várias desvantagens como tornar a análise não-reproduzível, *e.g.* o mesmo investigador dificilmente escolherá exatamente as mesmas ROIs, mesmo usando o mesmo conjunto de imagens. Através da determinação enviesada das ROIs, é possível enviesar os resultados pela escolha de que células incluir na análise. Tudo isto pode ser evitado se a análise de imagens for digitalizada e robotizada. Definindo os mesmos parâmetros iniciais, o programa de análise irá selecionar as mesmas ROIs, medir e processar essas medições de forma repetitiva e exata, produzindo análises reproduzíveis. Observando os parâmetros iniciais, é possível determinar se existe viés na seleção automática ou na sequência de análise. Outra vantagem óbvia é que a capacidade e velocidade de processamento das unidades centrais de processamento e das unidades de processamento gráfico permitem análises rápidas, facto essencial para diminuir o tempo entre a aquisição e análise de dados.

A ideia de automatizar a análise de imagens não é nova e já existem muitos programas comerciais ou com código-fonte livre que permitem fazê-lo, seja com algoritmos simples ou complexos. No entanto, a maioria deles só podem ser aplicados a organismos específicos e, por vezes, apenas se usados em protocolos muito específicos. O principal factor que bloqueia este progresso é o uso de algoritmos de segmentação (*i.e.* classificação dos pixeis em ROIs, fundo ou detritos) e filtração

(*i.e.* rejeição/aceitação de ROIs de acordo com as suas características). Na prática, o computador não sabe o que está à procura nas imagens e frequentemente inclui objectos diferentes ou exclui partes do objecto em estudo. Isto poderá enviezar os passos seguintes da análise, potencialmente criando artefactos. Propomos o uso de modelos matemáticos como conhecimento *a priori* da célula modelo alvo para ser ajustado diretamente às imagens. A nossa hipótese é que podemos obter as características mais prováveis desses objectos, estimando o número e forma dos objetos de interesse que melhor descrevem uma imagem (*i.e.* máxima verosimilhança), e assim melhorar a sua deteção, rastreamento e caracterização.

A análise de imagens de espermatozoides é um exemplo onde a segmentação é o principal método usado. Desta forma, estas células foram escolhidas como caso-estudo biológico para testar a nossa hipótese. Os espermatozoides são células móveis altamente especializadas cuja função é encontrar e fertilizar o ovo conspecífico aquando da reprodução. Por esta razão, tratam-se de células extremamente importantes tanto do ponto de vista fundamental como do ponto de vista médico, se tomarmos em conta que as anormalias na motilidade dos espermatozoides são das principais causas de infertilizade em casais. Não surpreendentemente, cedo se desenvolveram sistemas de análise de expermatozoides assistidos por computador (CASA) que implementam segmentação de imagem e que são rotineiramente usados em análises clínicas. Em mamíferos, as células espermáticas têm que nadar um percurso milhar de vezes superior ao seu comprimento através de uma geometria interior complexa, frequentemente cheios de fluídos altamente viscosos e de células imunes potencialmetne hostis. A grande maioria delas não chegarão aos tubos de Falópio, quanto mais ao lugar de fertilização. Nos invertebrados marinhos, a fertilização ocorre durante eventos de desova onde o espermatozoide tem que encontrar o ovo conspecífico literalmente num mar de ovos de múltiplas espécies. Em laboratório, os espermatozoides de ouriço-do-mar tendem a acumular e a nadar confinados no plano de interface entre o líquido e o vidro, que fazem deles particularmente adequados para adquirir imagens da sua motilidade e da sua quimiotaxia em resposta a moléculas libertadas pelo ovo. Por esta razão, o esperma de ouriço-do-mar tornou-se num paradigma biológico interessante.

Vários estudos mostraram que os espermatozoides respondem aos péptidos activadores de espermatozoides (SAP). A tradução destes sinais resulta em sequências de picos de concentrações intracelulares de cálcio que são concomitantes com os períodos de assimetria das ondas do batimento flagelar e que reorientam a trajetória celular. Em certas espécies de ouriço-do-mar, como *Lytechinus pictus*, as sequências de picos de cálcio estão espacialmente e temporalmente coordenadas de forma a produzirem trajetórias quimiotáticas direcionadas para a fonte dos péptidos, enquanto noutras espécies, como *Strongylocentrotus purpuratus*, fá-los produzir comportamenteos desorientados. No entanto, o facto do ambiente natural destes espermatozoides ser um volume tridimensional tem levantado preocupações sobre a relevância e aplicação dos estudos gerados no plano. Sendo os espermatozoides uma classe de células extremamente rápidas, exigem muito dos instrumentos de microscopia tridimensional e de séries temporais. Particularmente desafiante é a medição tridimensional de marcadores fluorescentes nos flagelos em movimento que não podem ser feitas nos actuais microscópios confocais de alta performance. Para esse fim, nas duas últimas décadas foram desenvolvidos vários sistemas de microscopia especificamente para estudar espermatozoides. Um deles, o sistema 2D+$Z(t)$ (*i.e.* reportado na literatura como *3D+*t), usa um dispositivo de piezoelétrico para oscilar a objetiva a altas frequências, permitindo adquirir imagens bidimensionais (2D) a diferentes profundidades (eixo Z) em função do tempo. Usando este sistema, foi demonstrado que *S. purpuratus* apresenta velocidades e curvaturas da trajetória média maiores em natação livre, quando comparado com natação confinada. Infelizmente, a profundidade a que cada imagem é adquirida não é reportada pelo sistema pelo que foi desenvolvido um algoritmo baseado na correlação entre imagens e no ajuste da função $Z(t)$ característica do dispositivo piezoelétrico para a determinar. Apesar deste método conseguir fazê-lo adequadamente em muitos dados experimentais, noutros tantos não o consegue devido ao aumento do desfasamento entre a profundidade real e a estimada. Este facto tem atrasado estudos comparativos da mobilidade e quimioataxia de espermatozoides.

Esta tese teve dois objetivos. O primeiro foi desenvolver e aplicar métodos de análise de imagem baseados em conhecimento *a priori*, ajustando modelos matemáticos do objeto a ser detetado ou rastreado diretamente às imagens. O segundo foi desenvolver e aplicar métodos para comparar a mobilidade 3D de *L. pictus* e *S. purpuratus*.

O capítulo 2 é uma prova de conceito de que um modelo celular pode ser usado para analisar imagens em 2D. Construímos um modelo matemático detalhado da morfodinâmica dum espermatozoide e de como este nada num fluído através das suas alterações morfológicas. Assim, este modelo mecânico é composto por dois módulos: forma, onde assumimos que a cabeça é um elipsoide de revolução, que a curvatura local do flagelo é dada pela onda progressiva e que a torção flagelar é constante, definindo assim o batimento flagelar; e mecânica, onde a física é aproximada pela Teoria de Força Resistiva. Este modelo foi comparado diretamente com uma imagem usando o procedimento seguinte. Primeiro pegámos no estado atual do modelo e construímos a imagem-modelo correspondente a esse estado, depois reescalámos a imagem-modelo para a resolução espacial e temporal da imagem experimental e, por fim, fizémos a sua convolução com a função de propagação de ponto ótico (*point-spread function*) característica da microscopia de fluorescência, se for o caso. A imagem-modelo foi comparada com a imagem correspondente ao mesmo instante através de correlação. Foi reportado que, em certas condições, maximizar a correlação-cruzada é equivalente a maximizar a verosimilhança (*i.e.* uma medição de quanto os conjuntos de parâmetros são explicados pelos dados). Devido ao nosso modelo não ser linear, maximizámos a verosimilhança dos parâmetros do nosso modelo usando um simples modelo evolutivo baseado em simulações de Monte Carlo, selecionando e propagando os conjuntos de parâmetros com maior coeficiente de correlação (uma imagem) ou com a maior soma de correlações (várias imagens, *i.e.* a soma de todos os coeficientes de correlação) durante algumas iterações. A etapa de propagação permite pequenas alterações aos parâmetros de forma a permitir explorar o espaço de parâmetros sendo as alterações tornam-se cada vez menores a cada nova iteração.

Ajustámos o nosso modelo morfodinâmico de um espermatozoide a diferentes conjuntos de imagens. O primeiro conjunto foi criado a partir do próprio modelo (*in silico*) com o intuito de provar que conseguimos obter os parâmetros usados para gerar as imagens. Mostrámos que a distância entre o conjunto de parâmetros real e final (*i.e.* ajustado) diminui a cada iteração, após o ajuste independente de três condições iniciais aleatórias. O segundo conjunto de imagens teve como objetivo averiguar a eficácia do nosso método comparando-a à de um humano. Para este fim, obtivemos images de alta resolução espacial e temporal de *L. pictus*. Uma célula deste conjunto de dados foi rastreada de forma semi-automática por um colaborador, reconstruindo a posição e forma do flagelo. Ajustámos o nosso modelo à mesma célula e mostrámos que soluções obtidas diminuem a distância entre o modelo humano e modelo morfodinâmico. Note-se que as pequenas diferenças entre os dois modelos poderão ter origem na suposição do modelo de que os parâmetros são constantes, o que é cada vez mais improvável quanto mais tempo passa. Para determinar se o nosso método consegue rastrear células em imagens com baixa resolução espacial e temporal, ajustámos o nosso modelo a imagens de *S. purpuratus* com essas características, tendo obtido resultados semelhantes aos de alta resolução. Por último, levantámos a hipótese de ser possível estimar características modeladas mas que não são visíveis nos dados experimentais. De forma a testá-la, usámos imagens obtidas de espermatozoides tratados com marcadores de fluorescência de membrana (célula completa) e de cálcio (Ca^{2+}, apenas a cabeça durante condições não-quimiotáticas). É possível obter a informação de ambos os marcadores na mesma célula no mesmo instante usando um separador de feixes. Conseguimos obter parâmetros e posições flagelares semelhantes depois do ajuste independente aos dados onde a célula completa ou apenas a cabeça estão visíveis. No seu conjunto, os nossos resultados mostram que é possível ajustar modelos diretamente a imagens com a vantagem de conseguir estimar características que não podem ser medidas diretamente através desses dados, o que não é fazível através dos métodos tradicionais de segmentação, que dependem muito dos próprios dados.

O capítulo 3 trata do problema de analisar e comparar a motilidade dos espermatozoides de duas espécies de ouriço-do-mar quando estão confinados ou nadando livremente num volume. A análise de espermatozoides nadando num volume necessita de sistemas de aquisição de imagens especiais. Usámos o sistema proposto por Corkidi et al., (2008) que implica inferir a profundidade de cada imagem através do tempo a que foi adquirida antes de obter as coordenadas 3D das células. Para isso, começámos por estimar os instantes em que o piezoelétrico se posicionou a profundidades máximas e mínimas. Apoiados na suposição de que as células não se movem significativamente durante um período do piezoelétrico (imagens adquiridas à mesma profundidade deverão estar positivamente correlacionadas), usámos a informação coletiva da correlação entre todas as imagens e as suas vizinhas subsequentes (dentro de um período do piezoelétrico). Depois ajustámos a função canónica do movimento do piezoelétrico entre esses extremos. Para detetar as células usámos um modelo 3D do espermatozoide médio composto por imagens 2D, cada uma correspondente a um padrão de difração resultante da distância entre a objetiva e o centroide da célula. Cada padrão de difração foi usado para fazer correlação-cruzada com cada imagem, sendo gravadas as posições de correlação máxima através das quais o centroide da célula foi calculado, após o seu agrupamento (*i.e.* de forma a identificar que posições pertencem à mesma célula, dentro de um período do piezoelétrico). Estas deteções foram atribuídas a diferentes células através do agrupamento pela sua distância no espaço e no tempo. De forma a averiguar os vários comportamentos natatórios, ajustámos o menor conjunto de hélices que explicasse um conjunto de pontos de uma trajetória. Para isso, usámos o critério de informação Bayesiano como critério de pontuação de um problema de programação dinâmica, um método que apelidámos como ajuste por pedaços helicoidais (*piecewise helix fitting*).

Os comportamentos de natação em modo livre ou confinado dos espermatozoides de *S. purpuratus* foram estudados previamente, mas nada é conhecido acerca da natação livre dos espermatozoides de *L. pictus*. Querendo também saber o desempenho do nosso método de rastreio, usámo-lo para determinar a velocidade, curvatura e torção das trajetórias médias de ambas as espécies para os dois modos de natação. Os nossos resultados foram consistentes com o que está reportado na literatura. Mostrámos, pela primeira vez, que as trajetórias de natação livre dos espermatozoides de *L. pictus* têm velocidades menores do que a outra espécie, ao contrário do que acontece no modo confinado. A torção média das trajetórias espermáticas desta espécie é também menor do que a de *S. purpuratus*. Interessantemente, a curvatura da trajetória dos espermatozoides de *L. pictus* não se altera entre os modos de natação livre e confinado, enquanto esta Ãl' diferente para os de *S. purpruatus*. Estudos recentes com espermatozoides de *Arbacia punctulata* mostram que as suas trajetórias têm características semelhantes às de *L. pictus*. De forma a poder explicar todos estes dados, adaptámos o nosso modelo morfodinâmico para conseguir simular a natação confinada quando a torção flagelar é diferente de zero e procurámos os conjuntos de parâmetros que produzissem trajetórias semelhantes às destas espécies. Os nossos resultados sugerem que os espermatozoides de *S. purpuratus* têm maior curvatura flagelar média (e, provavelmente, também maior torção flagelar) quando nadam em modo livre, comparativamente à natação confinada. Este efeito poderá dever-se ao aumento de viscosidade perto da superfície que afeta mais o batimento flagelar dos espermatozoides nesta espécie do que nas outras duas.

Usando o procedimento de análise de imagens do sistema 2D+$Z(t)$ que desenvolvemos no capítulo 3, investigámos o comportamento quimiotático dos espermatozoides de *L. pictus* e de *S. purpuratus* no capítulo 4. Em estudos de quimiotaxia destes espermatozoides em natação confinada, foi criado um gradiente de um péptido quimitático libertado pelo ovo (Speract). Para esse efeito, usou-se uma forma bloqueada (enjaulada) do péptido cuja afinidade para o seu recetor é 1000 vezes menor e um gradiente de concentração da forma ativa foi modulada por irradiação ultravioleta (UV) (desenjaulamento). Este processo the desenjaulamento foi usado para produzir um gradiente quimiotático no sistema de microscopia 3D. Duas condições foram testadas: com e sem quimioatrator enjaulado, tendo irradiando ambas com UV durante dois segundos, definindo assim os intervalos temporais antes, durante e depois da irradiação. As trajectórias das células foram reconstituídas por ajuste por pedaços helicoidais e obtivémos séries temporais dos parâmetros dos segmentos helicoidais. De forma a ter em consideração a dependência dos dados dentro de cada intervalo temporal, ajustámos modelos lineares mistos aos dados. Apesar desta estratégia ter sido capaz de distinguir os parâmetros obtidos para as duas espécies, não conseguimos encontrar nenhuma alteração nos parâmetros que fosse consistente com quimiotaxia provocada pelo desenjaulamento do quimioatrator. Interpretámos este resultado como sendo consequência de uma criação deficiente do gradiente que não foi, presumidamente, detetado pelas células. Depois sugerimos várias alterações aos protocolos experimentais de forma a averiguar se existem outras condições que promovam quimiotaxia. Noutro estudo, a quimiotaxia em natação livre dos espermatozoides de *A. punctulata* foi conseguida usando irradiação UV sustentada. Uma estratégia semelhante poderá ser eficiente para os espermatozoides das duas espécies estudadas aqui. Alternativamente, poderemos criar gradientes diferentes modulando o perfil de irradiação através do uso de fibras óticas diferentes.

Finalmente, discutimos a nossa tese de que o uso do conhecimento *a priori* sob a forma de modelo matemático da célula- ou organismo-alvo para ajustar dados pode aumentar a resolução e permitir estimar características que não foram nem podem ser medidas diretamente através desses dados. É possível expandir o algoritmo desenvolvido para o modelo biológico alvo fazendo as alterações apropriadas. No caso apresentado, podemos adicionar um módulo de sinalização e estudar a natação espermatozóide em ambientes quimiotáticos. Acreditamos que isto será crucial para descobrir as causas e mecanismos dos casos de infertilidade em humanos que ainda não compreendemos. Conseguimos também melhorar a determinação da profundidade das imagens geradas pelo sistema 2D+$Z(t)$ e ainda a análise desses dados. Além disso, melhorámos a análise desses dados de forma a conseguir

diferenciar a natação e comportamento de confinamento entre as duas espécies. A aplicação do modelo morfodinâmico permitiu-nos sugerir que os espermatozoides de *S. purpuratus* têm maior torção flagelar aquando a natação livre, quando comparado com o modo confinado. Infelizmente, não conseguimos detetar comportamento quimiotático nas condições testadas, pelo que outros gradientes deverão ser testados. No geral, os métodos desenvolvidos aqui, em conjunto com outras técnicas recentes, serão essenciais para a análise de imagem e, mais especificamente, para compreender a quimiotaxia dos espermatozoides.

Acknowledgements

Having changed from molecular to computational biology I knew my path would not be an easy one. However, it was for sure far more challenging, rewarding and fun that I would have imagined. For all this, first and foremost, I would like to thank my supervisor, Jorge Carneiro, for the incredible opportunity that he gave me. With his guidance and company, I was able to grow more as a scientist and as a person. Those moments of long discussions, either frustrating or enlightening, will simultaneously haunt and motivate me as precious lessons that will hopefully guide me in the right direction, whichever it may be. Be it near a blackboard or beside a margarita, I hope to continue to take both his scientific and personal advices as a professional and a friend in the near and far future.

To the Quantitative Organism Biology group, I really appreciate the growth, guidance and even the occasional silly talks. Thank you Thiago Guzella, Tom Weber, Danesh Tarapore, Delphine Pessoa, Eleonora Tulumello and Diogo Santos (you know you are like part of our group). Special thanks go to Tiago Macêdo for the many discussions and much time spent during breaks, lunches and coding advice. I also would like to thank Nuno Sepúlveda, a former group member, for advice on statistics.

I thank all the people from the *Instituto Gulbekian de Ciência* for making every working day lighter and fun. I thank Claudine Chaouiya, Mónica Dias and Andreas Bohn for accepting to be part of my thesis committee and for the useful insights and criticism. I also thank the many PhD students committees that organized the AMeeGuS retreats which allowed me to train presentations, discuss my work and enjoy the beautiful venues. I also thank the IGC band for allowing me to play music with them.

This work would not be possible without the collaboration with the consortium *Consorcio de Fisiología del Espermatozoide* and the *Departamento de Ingeniería Celular y Biocatálisis* of the*Instituto de Biotecnología, Universidad Nacional*

Autónoma de México (UNAM). I thank Alberto Darszon for the patience, advice, insight, criticism and friendship. For all of them, I thank the amazing hospitality that they received me with and for making home to seem less distant. Specificaly, I would like to thank the following persons:

- Alberto Darszon and Cláudia Treviño, for making some of my trips to Mexico possible and for receiving me in open arms.
- Arturo Pimentel, Adán Guerrero, Tatiana Luna and Gabriel Corkidi for the discussions, criticism and for providing the imaging data, which was essential for this thesis, and for the patience in training me to operate the 2D+$Z(t)$ microscopy system.
- Omar Jose, Tatiana Luna, Alberto Vicens, Arturo Pimentel, Adán Guerrero, Ana Laura, Carmem Santana, Ana Ruth and Francisco Balderas for showing me the both the fun and the delicious sides of Mexico. You made it easier to avoid home sickness.
- Ana Ruth for sharing the house, meals, birthdays and providing a second home.
- Many other group members who, one way or another, contributed either academically or culturally.

Also from UNAM, I would like to thank Gustavo Mekkel for support and Daniel Priego for the discussions about sperm channels and company during lunches.

To the amazing CotA group, I thank André Santos, Sara Silva, Diana Macedo and João Batista for the most unproductive mornings, afternoons and nights that I could have only imagined. It was only for a short time that we managed to work together but it was fun!

On a more personal note, I would like to address my parents who supported me in every imaginable way during all different aspects of my life. Without you, I simply would not get this far. Thank you for being there... always... unweary... time after time!

Finally, I would like to acknowledge the person who sacrificed more than anyone during my PhD. Thank you Audrey Lopes for those lonely nights I could not make you company because of my work, for getting my spirits up when I thought only time could do it and for being the amazing person I can count on for a laught, a getaway and for sharing a life. No words can ever repay you.

Contents

List of Tables

List of Figures

Acronyms

2D	two dimensions
3D	three dimensions
AIC	Akaike's information criteria
ASW	artificial sea water
ATP	adenosine triphosphate
BIC	Bayesian information criteria
BK	calcium-dependent K^+
$[Ca^{2+}]_i$	internal calcium (II) concentration
CASA	Computer Assisted Sperm Analysis
Ca_V	voltage-dependent Ca^{2+}
CCD	charged coupling device
cGMP	cyclic guanosine monophosphate
CMOS	complimentary metal oxide semiconductor
CPU	central processing unit
DP	dynamical programming
EM	electron microscopy
GPU	graphic processing unit

HCN	hyperpolarization-activated cyclic nucleotide-gated
KCNG	K^+-selective cyclic nucleotide-gated
MC	Monte Carlo
MLE	maximum likelihood estimation
NCE	Na^+-Ca^{2+}-K^+ exchanger
NHE	Na^+-H^+ exchanger
ODE	ordinary differential equation
PCA	Principal Component Analysis
PDF	probability density function
pH_i	intracellular pH
PSF	point-spread function
RFT	Resistive Force Theory
ROI	region of interest
SAP	sperm activating peptide
SBT	Slender Body Theory
SNR	signal-to-noise ratio
TIRF	total internal reflection fluorescence
UV	ultraviolet

Chapter 1

Introduction

Today biological sciences are generating imaging data at a huge pace. This is usually followed by several weeks of painstaking analysis by researchers who usually select the regions of (their) interest. As such, imaging data analysis is currently one of the major bottlenecks in scientific productivity. Also, the analysis made by a person is seldom reproducible, even if the same person were to reanalyze the same data (*e.g.* by selecting different regions of interest). This has led to an effort to automate imaging analysis but many methods cannot be generally applied to all cell types, markers, etc., because the computer does not know what it is looking for and filters like size and marker intensity are not enough to detect or track the cell efficiently. For these reasons many scientists still resort to manual annotation and analysis of their imaging data.

Sperm analysis is one such example. These fast cells require high temporal resolution from the microscopy setup if one hopes to study their motility; this is the reason why more data per second is generated in this kind of setup. If you couple this with three dimensions (3D) stack imaging, a few Gigabytes are easily generated and stored in a few seconds. Automatic image analysis is critical to deal with such problem. This aspect has hindered comparative studies on how spermatozoa from different species swim and, consequently, on how they react to different chemotactic gradients. These kind of studies can give insights on human fertility problems which causes are currently unknown.

In this thesis we explored automatic image analysis based on *a priori* knowledge to detect and track spermatozoa in both 2D+t and 2D+$Z(t)$ imaging data. We then used this methodology to make a comparative study of the motility of spermatozoa from two different sea urchin species and to assess their chemotaxis behavior in 3D. As this work involved diverse scientific fields, which made a continuous introductory flow challenging, we will first introduce some of these fields independently. They will be integrated and expanded as we introduce new ones. In more detail, we will start by making an introduction to modeling and Modeling Relation, then we will move on to current imaging analysis methods and how models have been used to do it and, finally, we will introduce our biological case study – the sea urchin spermatozoon, expanding its state-of-art in the light of modeling and image analysis. Thereafter, the objective and outline of this thesis will be presented.

1.1 Modeling and Modeling Relation

Making supported statements about the world requires evidence. This evidence is usually the result of fitting a model to data obtained somewhere and somehow. It can be as simple as the case of assuming the model of people height in a class to be Gaussian distributed and to assess if those distributions are different as we fit the model to samples of different classes, using the appropriate statistical model. However, a lot can go wrong in this simple process. Is the distribution of heights on each class correctly modeled by a Gaussian one? Is the sample of each class representative of their real distribution? Is the ruler well calibrated and suitable to measure the expected differences? What is the error or noise of the measurements? Are all the assumptions of the statistical test met, *e.g.*, a Student's t-test to test for differences of the mean height assumes independent data but are we sure there is not a student taking both classes present in both samples? A good experimental design will deal with most of these issues but complex questions, sampling methods, measuring devices or models might have characteristics which, in the particular ensemble used, are incompatible to answer the question. Many times, the incompatible characteristic is not obvious, it is difficult to identify or simply is not even noticed to exist. At the end, there can be trouble in either or both model and data.

In order to better understand whether we should believe more in the data or the model, we will first introduce the Modeling Relation developed by Rosen, (1991) (fig. 1.1). In a nutshell, this is a framework that relates the natural system, how we perceive it and how we can infer and validate the mechanisms occuring in it using formal (*i.e.* mathematical) language. Thus, we find useful to describe the different

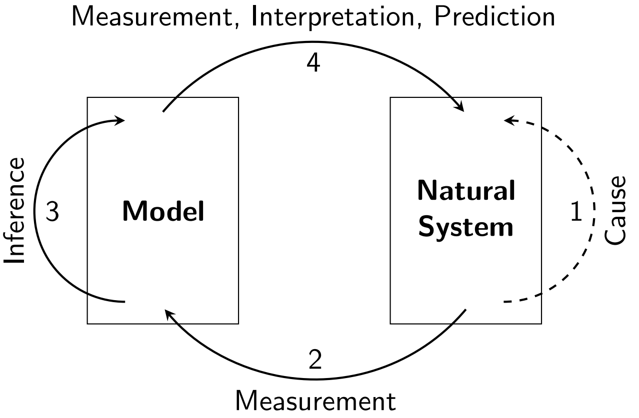


Figure 1.1: Rosen's Modeling Relation. Adapted from (Rosen, 1991).

types of inferential processes.

A natural system is a set of properties and the events that change those properties. To understand the natural system we must know and understand the cause of such properties and events (mapping 1). However, we cannot make direct assessment of the causes in the natural system. Thus, we must use our senses to perceive it so we take measurements of the properties and events – an encoded representation of a subset of the natural system (mapping 2). We must also develop a model that infers something about the encoded data (mapping 3). This, however, does not say anything about the natural system. In order to achieve correspondence between the inferential (mapping 3) and the causal (mapping 1) processes, the inferenced result must be decoded back to the natural system (mapping 4), hence making a guess or expectation of how the natural system changed. If this last step cannot be made, then our model does not describe the causal process.

In the end, we can choose to believe the data, the model or a mixture of both. Believing the data is the most popular approach. Researchers often collect data from samples and then fit different models to it until one is found, usually the one most significantly supported by the data. This can lead to data-driven models that merely describe the data but that do not provide new knowledge. New samples or more data points may need alterations to the previous model to accommodate them. Hence, this is not the case of a phenomenological model.

Believing the model is the most infamous approach as there is the case of believing the alternative hypothesis of a question and fitting different distributions and statistical models until the researcher is proved right. This is a typical case when people with poor statistics knowledge “torture the data until the data confess”. As this is considered bad scientific practice, we do not find it useful to discuss this further. There is, however, a reasonable preference for model-driven beliefs if it is impossible to perceive the external world with current technology if measurements (mapping 4) are believed to suffer from the issues referred above (or others). The former was the case of the prediction of the Higgs boson (ATLAS Collaboration, 2012; CMS Collaboration, 2012) and also of the gravitational waves proposed by Einstein’s General Relativity which could only be measured recently (Abbott et al., 2016). Similarly, in biology there are many theoretical predictions awaiting for technology to be able to measure them. Many metabolic and signaling pathways or membrane channel functions and mechanisms, proposed by available data and knowledge, are not easy to be tested *in vivo* due to lack of specific inhibitors and blockers. Also, measurement off all components of the pathway is usually impractical as the available methods can lead to artifactual behavior.

In my opinion, mixed-belief is where most scientist want to be. By formulating a hypothesis and designing an experiment to test it, which is able to join both the appropriate procedure for data collection and the correct statistical test, the scientist formalizes a model (mapping 3) to which they will feed data (mapping 2) and from which they get a prediction which is validated (mapping 4). Then (s)he is happy to say that the cause of the natural system has correspondence to the inference process of the model (mapping 3) and a new discovery (or a replicate assay) is confirmed. In this case, the scientist is both confident on the properties of the collected data characteristics and on the model details. Note that all model predictions from every meaningful input data must be verified for effective correspondence of the inferential and causal processes. For example, Newton’s law of gravity allowed Le Verier and Adams to hypothesize the existence of an unknown planet affecting the predicted orbit of Uranus, and their calculations predicted the position of Neptune, later confirmed by Galle. Both data and gravitational model were believed to be corrected and, consequentially, the solar system model was modified accordingly. Although the gravitation model was formulated and calibrated from earthly objects and correctly predicted the position and existence of many ’out-of-this-world’ objects, it was not successful in modeling Mercury’s orbit. Only with Einstein’s General Relativity was able to predict the correct amount of precession of the orbit’s eccentricity.

Taking all this into consideration, we should be looking to build a model where, independently of the measuring process, there is enough detail in the inferential process that predicts several different aspects of the natural system (*i.e.* our biological model). As an effort to confirm all predictions (mapping 4) should be made, from all meaningful measurements (mapping 2), the model should also be simple enough that the correspondence between inferential and causal processes can be attained.

1.1.1 Mathematical modeling

Looking at the Modeling Relation from the mathematics point-of-view, we define a mathematical model which is a mapping that will take some input (as the form of data) and generate some predictions (also measured in the form of data). This model will be described by a set of assumptions and their consequent parameter set ($\boldsymbol{\theta}$). Taking measurements of the natural world (*i.e.* data) is often affected by noise or other transformations (see section 1.2 for more details) which can also be described mathematically. How should we choose between two different representations of the data? If we go for a more complex model (*e.g.* the extreme case of the data itself), no or little information may be gained, and, if we go for the simpler model, the abstraction or imprecision may be so great it could be representing either this or completely different data. Both in statistics and other areas, there is the tradition of choosing the most parsimonious model, meaning the simpler one that can model the data well enough. Modeling the data well enough is usually given by the likelihood ($\mathcal{L}$), a measure that a set of parameters are supported by the data, and the complexity is usually given by the number of parameters (k). Both Frequentist and Bayesian schools derived the Akaike's information criteria (AIC) and Bayesian information criteria (BIC), respectively, which are scores that give different weights to the likelihood and to the parameter number (Hastie et al., 2009). The first is derived from information theory and both can be derived from the Bayesian framework. When to use one or the other is a debate lasting for decades but it is generally considered BIC gives more penalty to the number of parameters (Burnham and Anderson, 2002). Whatever the criteria chosen, it is its relative difference that holds some meaning – the model with lower score is the most parsimonious.

Consider now that we have two different inferential models (*i.e.* the measurement models are the same) that may or may not have the same parameter set. How should we choose between these competing models when they are applied to the same data?

A lot of effort and debate has also been put to answer this question. As our inferential model transforms measured data into predictive data, we can also use AIC or BIC to compare models (Hastie et al., 2009). In the special case where the models are nested (*i.e.* one of them is a sub-model of the other, *e.g.*, with one of the parameters fixed), we can use the likelihood ratio test: using the ratio of each model's likelihood, we can build a statistic that is chi-squared distributed and compute a confidence interval for the difference. If there is no difference, the simpler model should be chosen.

How are Model Relations and model construction affected in imaging data processing and analysis? To answer this question we must first think about what is an image and how it is formed.

1.2 Imaging data

1.2.1 Imaging data acquisition: from scene to image

In a modern optical or electron microscopy setup, a sample is irradiated with photons or electrons, respectively, and these alter their path by interacting with it. The irradiated media is then captured by a detector which translates the information to electric impulses, *e.g.* by a charged coupling device (CCD) or a complimentary metal oxide semiconductor (CMOS) device, which are in turn saved as digital information. The final imaging data ($\boldsymbol{I}$, *i.e.* the measurement) will be a distorted representation of the scene ($\boldsymbol{S}$, *i.e.* the natural system) and this distortion can be the result of four different transformations on the unstransformed representation of the scene, the image ($\boldsymbol{I}_u$) (Knill and Richards, 1996). These transformations are listed below (note '$A \rightarrow$ B' means object A is mapped to object B):

- Noise and blur ($\boldsymbol{S} \rightarrow (\boldsymbol{I_u} * \boldsymbol{\beta})_i + \epsilon_i$): the image is the result of the convolution of the scene by the blurring kernel ($\boldsymbol{\beta}$), to which background noise (b) and sampling error ($\epsilon_i \in \boldsymbol{\epsilon} \sim \mathcal{N}(b, \sigma)$) are added to the pixel i. This is the transformation most commonly addressed in imaging analysis. The noise may not be Gaussian, *i.e.* most times it is not, although we can use it as a fair approximation in many of those cases. Note the symbol '$*$' represents convolution.
- Superposition ($\boldsymbol{S} \sim (\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_n)$ or $\sigma(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_n)$): it deals with the fact that a complex scene or signal can usually be decomposed into simpler elements. Wavelet functions and Fourier analysis are examples where linear combinations

of wavelet or sinusoidal functions, respectively, are added to model a complex signal. Alternatively, other types of rules (σ) might be applied to the set of the simpler individual components. These methods can be used to decrease the number of parameters necessary to describe a scene. The use of a subset of components resulting from the Principal Component Analysis (PCA) of the signal is also a popular approach.

- Domain warping ($\boldsymbol{S} \rightarrow (\boldsymbol{I}_u \circ \psi)$): different temporal and spatial contractions or expansions (ψ) of the scene's domain might occur when acquiring the signal of the same object. An example is taking a photograph of a person's face from two different point-of-views – the object is the same but the resulting images are different. Note the symbol '$\circ$' represents function composition.
- Interruptions ($\boldsymbol{S} = \{O_1, O_2\} \rightarrow \boldsymbol{I}_u = \{\boldsymbol{I}_1 \mid D', \boldsymbol{I}_2 \mid D - D'\}$): many times a scene is a composition of several objects (O_i), from which we only observe a subdomain (D, D'), *e.g.*, due to occlusions or missing data.

Because any combination of these transformations is possible, imaging data often needs to be pre-processed and analyzed to extract the information the researcher desires. Pre-processing usually involves low-level operators, *e.g.* background correction, contrast enhancement or deconvolution, while image analysis tries to detect and measure features on the images, *e.g.* track cells and measure their mean fluorescence or size.

An important pre-processing method of fluorescent imaging data is deconvolution, which is related to the first transformation referred above. In more detail, the non-uniform generation of irradiation (*i.e.* excitation) source, its travel through the optical components until it reaches the sample, the random emission in space and time by the fluorescent marker and its caption through more optical components until it reaches one of the several units of the detector causes the final image to be a distorted representation of the original sample – a convoluted one. The distortion operator is called point-spread function (PSF) and, if we know this function, we can deconvolute the final image and obtain a sharper image which is a closer representation of the object (Agard and Sedat, 1983; Zhang et al., 2007). Both 3D fluorescence microscopy and super-resolution microscopy need to estimate accurate PSFs so the deconvolution does not create artifacts or aberrations, the reason why different methods to calculate the effective PSF of each system are currently being developed (Patwary and Preza, 2015). On the other hand, for most uses the PSF can be approximated by a Gaussian function using the information regarding the microscopy setup used (*e.g.* objective numerical aperture and the excitation and the fluorophore's emission wavelengths) (Zhang et al., 2007). After recovering the deconvoluted image, and assuming this process does not create artifacts, it is easier to analyze it, in order to estimate the scene it encodes.

1.2.2 Imaging data analysis: from image to scene

The objective of imaging data analysis is to decode the scene encoded in images. Are the imaged objects cells? Where are they located? What is their shape? What is their size and what is the concentration of a specific marker inside them? These are the kind of questions a researcher often does when analyzing imaging data and their respective answers will provide the data to verify the hypotheses at hand. Although a human can understand a scene represented by an image, a computer program needs to be coded with the operations to perform such task. We can use a model with parameters $\boldsymbol{\theta}$ to represent a scene encoded by an image ($\boldsymbol{I}$). Due to the signal transformations introduced before, specifically the random sources, there is a probability ($P(\boldsymbol{I} \mid \boldsymbol{\theta})$) that a given scene/parameters results in a particular image. When we specify a model, the set of possible images will follow a given distribution with total probability summing to one. Inversely, we also have the likelihood of the parameters given a particular image ($\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{I})$), meaning that we are measuring how likely it is for a set of parameters (*i.e.* scene) to have formed that specific imaging data (*i.e.* the sum over the parameter space can be different than one). It is not an unusual practice to estimate the model's parameters by maximizing the likelihood, usually by solving to when the partial derivatives are zero. The parameters can also be estimated by Bayesian inference. Bayes theory defines that the posterior probability ($P(\boldsymbol{\theta} \mid \boldsymbol{I})$) is the likelihood ($P(\boldsymbol{I} \mid \boldsymbol{\theta})$)[1] times the prior probability ($P(\boldsymbol{\theta})$) over the expectedness of the image ($P(\boldsymbol{I})$): $P(\boldsymbol{\theta} \mid \boldsymbol{I}) = P(\boldsymbol{I} \mid \boldsymbol{\theta})P(\boldsymbol{\theta})/P(\boldsymbol{I})$ (Knill and Richards, 1996). Assuming the same imaging data, the expectedness is constant (*i.e.* a normalizing factor) so the posterior is proportional to the likelihood times the prior ($P(\boldsymbol{\theta} \mid \boldsymbol{I}) \propto P(\boldsymbol{I} \mid \boldsymbol{\theta})P(\boldsymbol{\theta})$). While both the Frequentist and Bayesian approaches make use of a model to generate the probability mass function (or probability density function, in the case of a continuous model) only the latter takes prior knowledge into account, *i.e.* in the shape of the probability of the parameters. Either case present

---

[1] Note they define the likelihood differently than the Frequentists.

the same challenge when dealing with complex or numerical models, it is sometimes impractical to get an algebraic function of the likelihood so different optimization algorithms can be used to find the parameters which maximize $P(\boldsymbol{\theta} \mid \boldsymbol{I})$.

A typical experimental pipeline where imaging analysis is used is summarized in Dufour et al., (2015). Following a starting experimental design, microscopy data is obtained and cells are detected. Their descriptions are then extracted and fed to a machine learning algorithm which allow us to select the most relevant features to make biological inferences. Finally, these allow us to propose new hypothesis and design new experiments, closing the cycle. We will now detail some of the processes and characteristics entailed by the current implementation of this experimental pipeline.

One of the crucial steps in image analysis is its segmentation in order to determine the regions of interest (ROIs) (*e.g.* cells) in each image. We can consider two different strategies to do this. The first is image-based and tries to assign each pixel to a given class, *e.g.* cell, background, fluorescent marker or debris. This is the main method used to study cell morphology (Smith et al., 2009b), protein colocalization, remote signaling, magnetic resonance images (Ahmed and Mohamad, 2011), angiogenesis and stem cell (Rabbani and Javanmard, 2011), to name a few representative studies. There are two main approaches: pixel-based, where the multidimensional information of each pixel is used by supervised and unsupervised methods to classify it; and object-oriented classification, a bottom-up approach were neighboring pixels are sequentially clustered according to some homogeneity criteria and the different groups are then classified using, *e.g.*, pixel intensity, shape or texture features (Inglis et al., 2010). The homogeneity criteria is usually a score or an energy function that represents similarity or dissimilarity to be maximized or minimized, respectively, and its formulation is crucial to solve the problem at hand without creating artifacts. Correlation or least squared distance are popular as similarity measures (Hastie et al., 2009). Pixel-based methods are perhaps the most widely used ones and can be as simple as defining an intensity threshold. Then, every pixel which intensity is above that threshold is retained while those which do not are set to zero. A body of work was done in order to define the threshold value(s) automatically (Ren et al., 2010). The second segmentation strategy is model-based and tries to estimate the data that produced an image or to fit some kind of model directly to the image. Given an explicit or implicit parametric model it is possible to maximize a score function to *e.g.* obtain the boundary regions of an object, *e.g.* using active contours (Xu and Prince, 1998).

In these cases, objects' detection and descriptors can be obtained in a single step. As explained before, the parameters of stochastic models can be estimated or fitted using the maximum likelihood estimation (MLE) or different criteria in the case of Bayesian inference (*e.g.* maximum *a posteriori* or minimum mean squared-error, (Knill and Richards, 1996)).

Immediately after detection, the descriptors of a cell model can be as complex as the position and intensities of all pixels within the ROI. Image segmentation is then usually followed by fitting parametric lines, ellipses or other shape models to individual clusters of pixels of the same class (Brokaw, 1984; Baba and Mogami, 1985) or simply by measuring their size and mean intensity in order to reduce the complexity of the model. Prior knowledge is very often introduced at this stage in the form of filtering the detections according to the desired range of each feature. In the case where model-based procedures were used, the estimated parameters can be directly used as descriptors. On the other hand, many of these algorithms only select the ROIs, thus feature extraction from the imaging data must still follow. After reducing the models' complexity, the selected cells and their features can be used by machine learning algorithms (*e.g.* Principal Component Analysis, Linear Discriminat Analysis and Deep-learning algorithms) in order to select the relevant features that allow us to distinguish significantly different classes of cells (*e.g.* the percentage of cells affected by a treatment or the response between different treatments).

An example of the image-based approach is the super-resolution reconstitution of 3D structures using electron microscopy (EM) (Vulović et al., 2013; Kervrann et al., 2016). The main idea is to model the complex image formation of the microscopy system and, from many slices of the structure at different orientations, to estimate the position and relative orientation of every part of the object such that it would reproduce those slices. This method even allows to distinguish between different frequent conformations of the same structure. However, it cannot be applied to cells which are constantly moving or deforming and assuming shapes that are (slightly) different to other cells.

In the case of tracking, several semi- and full automatic algorithms exist which link the detection of one frame with the most similar object in the next frame in terms of position, shape and texture (Tsai et al., 2012). Consequentially, the atemporal parametric models fitted individually to an object at each frame (*i.e.* time-point) can be used to fit temporal models (Friedrich et al., 2010). Other algorithms simply use the current detection as initial condition to estimate the one on the next frame given a motility model (de La Gorce et al., 2008). This however can lead to poor fitting when the displacement between consecutive frames is significant. Different tracking methods assume both motility and statistical models to estimate the position of the object at the next time-point and correct it using the object closest to the prediction. To the best of our knowledge, there is no report in the literature about a model-based approach to characterize cells that change shape and position in time using parametric functions that explicitly define their shape and the physics governing their motility. We propose such strategy will be instrumental for objective, fast, detailed and rigorous analysis of imaging data. As such, we present a framework which uses this apporach in *Chapter 2*.

1.3 Sea urchin spermatozoa

Sea urchins have been popularly known by man since pre-historic times as a food source. Their gonads are specially appreciated by many cultures even today, generating a global market value around $>10^9$ millions dollars per year. Before the 20th century they were thought to cure several different illnesses, including treating kidney stones and helping digestion, while their shells were used as personal apothecaries, lamps, necklaces, amulets and to produce indelible ink (*i.e* after grinded up). (Harvey, 1956)

These marine invertebrates belong to the phylum *Echinodermata*, class *Echinoidea*. They are animals with 5-fold radial symmetry that live in benthic, intertidal and pelagic ecosystems, usually living in groups on rocky substrates but can also be found isolated during low tides of intertidal areas (Hernández and Russell, 2010). Adult females and males spawn their gametes almost synchronously to the sea during interspecies broadcast spawning events (Johnson et al., 2013). Spermatozoa must outcompete their intra- and interindividual male counterparts to be able to fertilize an ovule. After fertilization, the embryo develops into a larvae during the morphogenesis phase in pelagic ecosystems. Later, they suffer metamorphosis and settle in benthic habitats as juveniles, a more identifiable form resembling the adult phase. At two years of age, they maturate and become sexually active (Dupont et al., 2013). How their spermatozoa are able to locate and reach the conspecific ovule during broadcast spawning events is a challenge yet to be understood. To help us understand these challenges we will look at the spermatozoan morphology, motility and chemotaxis, preferably the ones related to sea urchin, whenever possible.

1.3.1 Sperm cell morphology

The first whole cell analysis of a matured sea urchin spermatozoon with EM was performed by Afzelius, (1955). These cells are very compact, meaning they are almost deprived of cytoplasm, and have their organelles highly organized. Similar to sperm of other taxa, this cell is divided into head and flagellum. The head contains the nucleus and the mitochondrion, and the flagellum the axoneme (fig. 1.2A).

Usually the head has a wedge or conical shape, measuring 3-4 μm long, and harbors the nucleus, the acrosome and two centrioles. The nucleus is surrounded by a double membrane and appears homogeneous, although it can present granulation. The acrosome is a Golgi-like structure with approximately 0.25 μm radius located at the tip of the head. It is filled with hydrolytic enzymes that will allow the penetration of sperm into the ovule protective layers (Ikawa et al., 2010; Li et al., 2010). There are two centrioles: a proximal, longer one which is coupled to the flagellar basal body located in a posterior cave of the nucleus named centriolar fossa; and a distal, short one which is freely located between the mitochondria and the nucleus (Afzelius, 1955; Marshall and Luykx, 1973). No ribossomal machinery nor protein translation were detected in these cells.

Although in some cases the flagella close to the head is more similar to mammalian sperm midpiece, in sea urchin sperm it is better described as an extension of the head. It is filled with mitochondrial material in a ring shape around the basal plate and the axoneme (Afzelius, 1955).

In many sperm cells, the flagellum represents the major part of the cell length (Bishop, 1962) and was later proposed to promote optimal progression of the cell (Tam, 2008). The backbone of the flagellar basal body is composed by a central pair of singlets surrounded by nine triplet filaments of microtubules (Marshall and Luykx, 1973). In the flagellar piece close to the head, the outer triplets turn into doublets by losing the outer tubule and, in the last micrometers of the flagellum, the central pair also vanishes. The standard structure of the sea urchin axoneme is depicted in figure 1.2B. The central singlets are connected to each other by a proteic bridge and they form the central pair complex in conjunction with other material. The central pair complex can be connected to each outer doublet by a radial spoke, an array

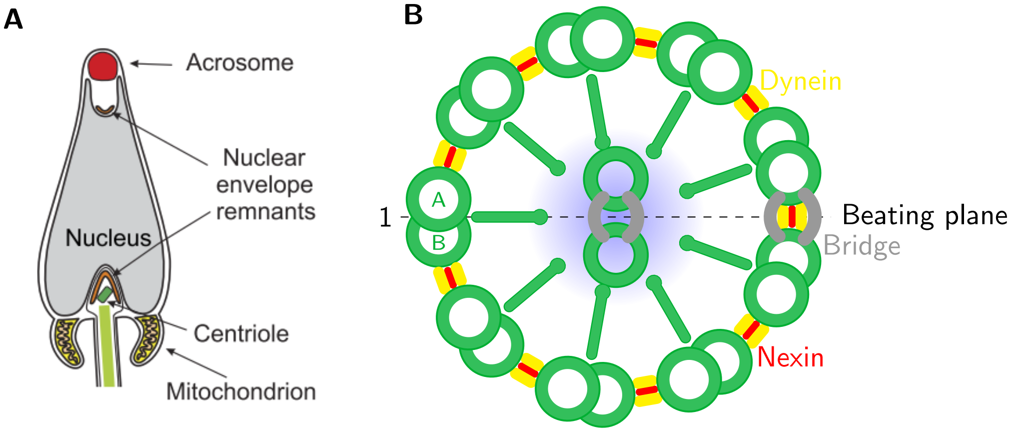


Figure 1.2: Morphology of a sea urchin sperm cell. A: head, Copyright: doi:10.1152/physrev.00028.2010, with permission from Alberto Darszon. B: flagellar axoneme.

of components that control kinases and phosphatases on the doublet microtubules (Nicastro et al., 2006). Each outer doublet is composed by a major (A) and a minor (B) subtubule and is connected to the next by nexin links and two rows of dyneins. The dynein regulatory complex connects the radial spokes to the doublets. Each doublet is numbered relatively to the plane defined by the central pair and its perpendicular plane that passes through the center of the axoneme. The doublet that crosses this plane is doublet number 1 – the order then follows clockwise (*i.e.* considering anteroposterior orientation of the flagellum). Usually, there is also a protein bridge between doublets number 5 and 6.

1.3.2 Sperm motility

Sea urchin spermatozoa are among the fastest cells of the animal kingdom. As it happens with most biological cells, they swim under a low Reynolds number regime. This means that, due to their size and velocity relative to the kinematic viscosity of the fluid in which they typically swim in, the viscous forces dominate over the inertial ones (Purcell, 1977). In practice, the cell can only be displaced in the fluid when it exerces force on it and stops swimming almost immediately when force exertion ceases. Sea urchin sperm cells can achieve velocities up to 300 μm.s^{-1} with a flagellar beating frequency of 20-30 Hz (Corkidi et al., 2008). Bishop, (1962) sates that, in 1898, motility was attributed to be the sole function of the sperm flagellum. The flagellar patterns of both sea urchin and bull sperm were identified as propagating waves which originate at the flagellar body (Bishop, 1962), contrary to the beating patterns of many bacteria and mammalian sperm which tend to be helicoidal (Su et al., 2012). The wave propagation pattern suggested the flagellum is controlled by a whiplash mechanism but several early and recent studies demonstrated it was not the case. One showed that the energy source required for motility (*i.e.* adenosine triphosphate (ATP)) is available in all the flagellum and others detected local activation of the flagellum in developing spermatids, in reactivated sperm cells and in demembranated flagella (Bishop, 1962; Chen et al., 2015).

Most sea urchin spermatozoa have a quasi-planar beating (Gray and Hancock, 1955; Cosson et al., 2003) and this was shown to optimize the velocity for cells with spherical heads up to ~ 2.5 μm of radius (Brokaw, 2003). This plane is defined perpendicularly to the one defined by the central pair of microtubules of the axoneme (fig. 1.2 B). The asymmetry in the flagellar beating, characterized by a principal wave which displaces more to one side of the cell, allows the cell to swim in circles when it is confined near a water-surface interface (Brokaw, 1965; Cosson et al., 2008; Corkidi et al., 2008; Friedrich et al., 2010) and in helices otherwise (Crenshaw et al., 2000; Corkidi et al., 2008; Jikeli et al., 2015). It was shown that several external conditions significantly affect both the beating patterns and the overall trajectory. For example, at high viscosities the beating frequency, the maximum wave amplitude, the wavelength and the cell velocity are decreased (Brokaw, 1966) and, in the case of bull sperm, the flagellar amplitude near the head decreases while it is maintained near the tip (Rikmenspoel, 1984). Also in bull sperm, the average path curvature is decreased at high viscosities (Friedrich et al., 2010). Some sperm even change their quasi-planar beating into helicoidal patterns at increased viscosity (Woolley and Vernon, 2001).

The kinesins and dyneins promote relative sliding between the axoneme filaments of opposite sides (Nicastro et al., 2006) and asymmetry is proposed to be induced by the 5-6 bridge of the axoneme or to arise due to non-linear instabilities in flagellar dynamics (Gadêlha et al., 2010). Adenosine biphosphate has been shown to influence flagellar curvature in bull sperm, supposedly by directly increasing the affinity of dyneins to the microtubule filaments (Lesich et al., 2008). Local control of dynein activity has been proposed to be regulated by local flagellar curvature (Brokaw, 1971), to simply arise from force-dependent sliding control (Brokaw, 1975; Camalet and

Jülicher, 2000) or to be a natural consequence of varying inter-doublet distance as the flagellum bends (Lindemann, 1994). The control of a quasi-planar beating pattern has also been proposed to be more complex than the helicoidal counterpart because of the radial symmetry of the axoneme, but its mechanism is not fully understood (Brokaw, 2003; Riedel-Kruse et al., 2007; Gadêlha et al., 2010; Chen et al., 2015). Although much is known about the mechanistic process, we are yet to understand how the several components act together in order for the cell to control its flagellar shape and deformation and, thus, its swimming trajectory.

1.3.3 Sperm chemotaxis

Sperm chemotaxis – the ability of a spermatozoon to detect and find the conspecific egg by decoding chemical gradients – is believed to be essential for the reproduction of many species (Kaupp, 2012). Most of our knowledge of spermatozoan chemotaxis comes from experiments with sea urchins. This is because they produce more cells than mammals, both in number and in concentration, and the media where they swim is also easy and cheap to reproduce, contrary to the highly heterogeneous and viscous fluid in the female mammalian genital tract. Also, they swim in circles when confined (*i.e.* swimming close to a water-glass or water-air interface), allowing us to follow the same cell for several minutes (Harvey, 1956; Cosson et al., 2003; Darszon et al., 2008; Gaffney et al., 2011; Guerrero et al., 2010).

The chemotaxis of animal cells was first proposed in sea urchin sperm by Lillie, (1912) but it was only after isolation of resact from *Arbacia punctulata*'s eggs and its establishment as a sperm chemoattractant that animal chemotaxis was accepted (Hansbrough and Garbers, 1981; Ward et al., 1985). It has been shown *L. pictus* sperm display chemotaxis at small distances from the gradient center when releasing the chemoattractant (Speract) from its caged form (Guerrero et al., 2010). A 2-nitrobenzyl group was inserted at a backbone amide that greatly decreases Speract's affinity for the sperm receptors and can be removed by ultraviolet (UV) light. Sperm swim up the gradient by first increasing and then decreasing the asymmetry of the flagellar beating, which promotes an acute turn followed by a straighter trajectory. Calcium (II) has been shown to affect the flagellar curvature on demembranated flagella (Brokaw, 1979) and these turn-and-run events have also been shown to be related with internal calcium (II) concentration ($[Ca^{2+}]_i$) spiketrains (Böhmer et al., 2005; Guerrero et al., 2010). Later it was established in complete spermatozoa that the rate of change of its concentration controls chemotaxis (Wood et al., 2003; Alvarez et al., 2012). The mechanism by which calcium affects the flagellar shape is still unknown.

There is a considerable body of work done to establish the signalling cascade from the sperm activating peptide (SAP) binding to guanylyl cyclase (GC, *i.e.* the receptor) to changes in $[Ca^{2+}]_i$ (Darszon et al., 2011; Kaupp, 2012; Seifert et al., 2015; González-Cota et al., 2015). Briefly (fig. 1.3), the binding elicits synthesis of cyclic guanosine monophosphate (cGMP) which will activate the K^+-selective cyclic nucleotide-gated (KCNG) channels. The exit of potassium ions will hyperpolarize the cell membrane and thus activate the Na^+-H^+ exchanger (NHE) and the hyperpolarization-activated cyclic nucleotide-gated (HCN) channels which will alkalinize the cytosol and depolarize the membrane, respectively. The depolarization activates the voltage-dependent Ca^{2+} (Ca_V) channels which will increase the $[Ca^{2+}]_i$. The increase in intracellular pH (pH_i) will presumably activate CatSper, a pH_i and mildly voltage dependent Ca^{2+} channel, thus also increasing the $[Ca^{2+}]_i$. Returning to basal $[Ca^{2+}]_i$ levels after stimulation is done by the Na^+-Ca^{2+}-K^+ exchanger (NCE) and a phospodiesterase (PDE) which hydrolyzes cGMP. Also, a calcium-dependent K^+ (BK) channel hyperpolarizes de membrane to the basal membrane potential. While some of these components have been identified in sea urchin sperm and their function has been established, others remain elusive. The identity of the Ca_V channel is still unknown and the function of NHE has not been firmly proved. There is also evidence for a different calcium-dependent K^+ channel which can modulate the calcium spiketrain (Espinal et al., 2011; Espinal-Enríquez et al., 2014). Also, the ensemble and proportions of calcium channels present in the sea urchin sperm flagellum are yet to be determined.

Contrary to *L. pictus*, for the same chemoattractant molecule and gradient, *S. purpuratus* sperm react by eliciting turn-and-run episodes but they do it in random directions (Guerrero et al., 2010), a behavior that can hardly be called chemotaxis. As these studies were performed when the cells are confined to the water-glass interface, we cannot exclude the hypothesis that far from the boundary, in 3D, cells would be able to turn in a chemotactic way. The chemotaxis of *A. punctulata* free swimming sperm has been characterized in 3D (Jikeli et al., 2015) but the question whether all sea urchin species sperm react the same way remains to be answered. Because small local changes of the trajectory curvature can have an huge impact on the global

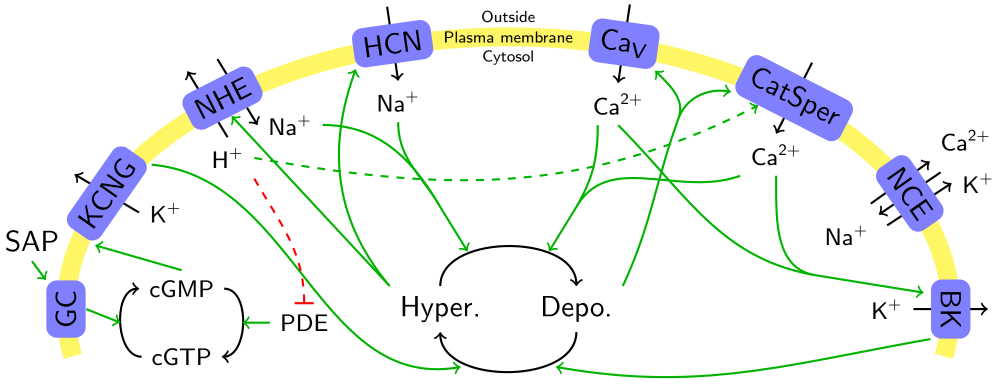


Figure 1.3: Chemotactic signaling cascade of sea urchin sperm. The plasmatic membrane can be hyperpolarized (Hyper.) or depolarized (Depo.). Blue boxes represent different membranar channels (see section 1.3.3 for details). Black arrows represent ion transport, state transferences and reactions, depending on the context. Green and red solid arrows represent activation and inhibition, respectively, of channels and processes. Dashed arrows represent inhibition by low pH_i (red) and activation by high pH_i (green).

trajectory (Guerrero et al., 2011), further studies on sperm motility are required to understand how sperm cells swim and reorient in space in order to have more insights on spermatozoan chemotaxis.

1.3.4 3D imaging of sperm cells

Multidimensional recording of biological processes is a standard approach in biological research and it can encompass spatial dimensions, time and color (*i.e.* usually by different fluorescent labels). This scaling in dimensions has been instrumental to study aspects and details (*e.g.* by correlation) that were not possible before, including in the study of sperm motility and chemotaxis. Using fluorescent probes and a confocal microscope we are now able to get 3D spatial reconstruction of cells and their labelled structures by focusing at different Z positions (slice). However, most of these assays require immobile or fixed material. To make 3D temporal screenings, different cells are fixed at different time-points *e.g.* after treatment. Imaging motile shape-changing cells in 3D plus time, however, has been challenging, specially when they swim at 200-300 μm.s^{-1}.

Some 3D imaging setups have previously been developed and can help to address this issue but they can only detect the head of the cell. Examples include using two perpendicular cameras to observe a 3D volume (Crenshaw, 1991) or using an objective coupled to a piezoelectric device in order to take XY slices at different Z positions at an enormous rate (Corkidi et al., 2008). The major disadvantage of the latter is the non-negligible error on the estimation of the Z position of the cell while scanning considerable volumes (*i.e.* determined by the Z amplitude of the piezoelectric device). Su et al., (2012) developed different system based on holography where over 1,500 cells can be tracked with submicron precision on a volume as big as 17 mm^3. The same system was later used to study the 3D chemotaxis of *Arbacia punctulata* sperm (Jikeli et al., 2015). Nevertheless, combining this system with fluorescent labeling is not straightforward (Rosen and Brooker, 2008; Nadeau et al., 2016), a feature that is essential to study spermatozoan chemotaxis, namely, how does the $[\text{Ca}^{2+}]_\text{i}$ and the pH_i change. Also, the flagella cannot be yet resolved in any of these setups and we know that the cell's trajectories are not sufficient to understand 3D chemotaxis (Crenshaw, 1989). On the other hand, the 2D+$Z(t)$ system of Corkidi et al., (2008) has already been shown to be able to segment the human sperm flagellum (Silva-Villalobos et al., 2014) and has the potential to use fluorescent markers. In that study, they scanned only 16 μm in the Z-axis with a piezoelectric frequency of 90 Hz and a framerate of 5000 Hz, corresponding to a mean spacing between slices is ~ 0.6 μm and the Z-error mentioned becomes negligible. However, this is not the case if we track multiple cells in a considerable large volume, as in Pimentel et al., (2012), so an improvement of the cell's coordinate precision is required. Although its promising potential to elucidate chemotaxis, no significant output has been produced with the 2D+$Z(t)$ system. Solving its few issues will be intrumental to add another dimension to 3D sperm chemotaxis.

1.3.5 Mathematical modeling of spermatozoa

Spermatozoa do not have the machinery necessary to synthesize proteins so it is not possible to use a genetic manipulation approach to study them. Hence, much of the research on spermatozoa have been made using different markers and drugs such as inhibitors. For the case of sperm motility, the theoretical approach has been a precious instrument. Gray and Hancock, (1955) were the first to successfully model invertebrate sperm motility. They proposed the Resistive Force Theory (RFT) where the flagellum is approximated as a set of infinitely small rods. As the axoneme motors elicit movement of the cell, each flagellar piece moves, creating a force on the fluid, which will in turn exert an opposite force on that flagellar piece (*i.e.* Newton's action-reaction law). They came to the conclusion that, in order for the cell to move forward, the flagellar drag coefficients perpendicular and parallel to the flagellar centreline need to have a ratio higher than one. Half a century later it was shown that this theory can model bull (Friedrich et al., 2010) and sea urchin (Jikeli et al., 2015) sperm with high precision, in either confined or free-swimming assays. For confined swimming, *ad hoc* increases of flagellar ratios are usually considered or measured in order to compensate the neglect of long-range hydrodynamic forces (Smith et al., 2009a; Friedrich et al., 2010).

RFT is in fact an approximation (*i.e.* neglects long-range hydrodynamic forces) of the more general framework – Slender Body Theory (SBT), which explicitly describes the fluid flows generated by the cell(s) and how the fluid also affects the spermatozoan conformation using Navier-Stokes equations (Johnson and Brokaw, 1979). Later it was established that one could implement this theory using stokelets, a faster numerical approximation (Gillies et al., 2009; Cortez, 2001). These and other similar frameworks allow to study the confining process (Smith et al., 2009a; Elgeti et al., 2010), *i.e.* how cells become trapped in the water-glass interface, as they can model the forces the boundary exerts both on the fluid and on the cell. As confined swimming is more prevalent in internal fertilizing species, most of these studies are performed assuming a mammalian spermatozoan model. Although no off-plane components of the flagellar beating were required for accumulation in surfaces, it was also possible to confine cells with helicoidal flagellar beating (Smith et al., 2009a). For the sea urchin sperm model, it was shown that attraction to the cell surface is the result of hydrodynamic interactions between the flagellum, the surface and the forward thrust of the cell. Those forces promote both a torque on the swimmer that aligns it parallel to the boundary and a force that approximates it to the wall, including a tail repulsion (thus, a head attraction) at short distances from the surface (Elgeti et al., 2010). On the other hand, it was also shown for the active flagellar model (*i.e.* eukaryotic) that the confining behavior depends on many factors and cannot be known *a priori* (Evans and Lauga, 2010).

These and other theoretical frameworks have been developed to study other perspectives of the sperm cell. In interspecies comparative studies, the optimal morphology for motility for uniflagellated cells was shown to be dependent on the ratio between head and flagellar length, rather than their absolute values (Humphries et al., 2008), and that for each head length there is a finite optimal flagellar length (Tam, 2008). The optimal flagellar stroke characteristics (displacement given expended energy) were found to be very similar to the biological cases (Tam, 2008; Spagnolie and Lauga, 2010; Lauga and Eloy, 2013). These include the existence of half-integer wavelength, which is proposed to reduce rotation and increase the translational velocity, similar wave amplitude to flagellar length ratio and the decrease in curvature along the propagating wave. Lauga and Eloy, (2013) even suggest that eukaryotic flagella are mechanically optimal.

Note there are also studies which model the sliding tubules of the axoneme explicitly (Camalet and Jülicher, 2000; Cibert, 2002; Riedel-Kruse et al., 2007) and some even explicitly include the dynein machinery (Hines and Blum, 1978; Hines and Blum, 1979; Brokaw, 2014).

Spermatozoan chemotaxis has also been the object of many theoretical studies. Simpler models for motility and chemotaxis which also neglect long-range hydrodynamical forces or the flagellum altogether. One example using ordinary differential equations (ODEs) also measured chemotaxis in population by assuming stochastic processes of agglomeration (*i.e.* as an inverse process of diffusion) with different degrees of behavioral and environmental assumptions (Keller and Segel, 1970; Horstmann, 2003). Other approaches use either the stimulus (chemoattractant) or the derivative of $[Ca^{2+}]_i$ to affect the average path curvature directly (Friedrich and Jülicher, 2007; Alvarez et al., 2012). In a consequent study where both stimulus and calcium derivative model chemotaxis by affecting the mean flagellar curvature, two different behaviors were identified: an 'on response', where the cell steadily but slowly redirects its trajectory to maximize the mean stimulation (*i.e.* towards the center of a chemoattractant source-point); and an 'off response', where the cell performs an abrupt change in direction when it swims down the gradient (Jikeli et al., 2015). Notwithstanding, we cannot give meaning to the simple adaptation module (*i.e.* signaling module) parameters in terms of the components described for the biological signaling cascade. There are some mathematical models that translate the signaling cascade to boolean networks (Espinal et al., 2011) and those that use ODEs are under development (Daniel Espinosa, manuscript under preparation, personal communication). However, we have yet to see an integrative model where shape, mechanics and signaling cascade are present (also under developement).

1.4 In this thesis

As introduced before, there is a bottleneck between imaging data acquisition and its analysis. This occurs at several levels, including speed, objectivity, reproducibility and robustness. Automatic analysis by computers can address these issues but painstaking work is generally needed to adapt an algorithm to the system in study, as they are not generaly appliable to all of them. One example where this occurs is the lack of comparative study of sea urchin sperm free-swimming motility and chemotaxis, which we will use here as a case study.

As a part of the speciatiation process, one expects sperm from different species to react differently to the chemoattractants released by the homologous eggs. However, the nature of such difference is unknown. It has been speculated that species-specific chemotaxis plays a key role in increasing the chances of fertilization during broadcast spawning. Motility and its coordination are the cornerstones of chemotaxis which justifies the demand for comparative studies of sperm of different species. Since sea urchin spermatozoa are released and presumably fertilize the eggs in a 3D setting, it is essential to understand how spermatozoa from different species swim freely in the volume. These 3D comparative studies have been hindered by the limitations of 3D imaging methods of fast cells and the lack of reliable and robust imaging analysis methods that can deal with this data type and amount. We propose that methods using *a priori* knowledge, in the form of a mathematical model that describes both the form and deformation of the cell and the mechanics that propel it forward within a fluid, will improve the detection and tracking quality, with the advantage of easy change of the form and physic descriptors. Due to its potential to study chemotaxis, we decided to use the 2D+$Z(t)$ microscopy setup (Corkidi et al., 2008) but an improvement of the accuracy and precision of the cell's position in 3D is of paramount importance to obtain reliable data. Thus, the objective of the work presented in this thesis is two-fold:

1. To develop and apply an automatic method for spermatozoan detection and tracking in imaging data using a mechanistic model of these cells as *a priori* information;
2. To develop and test tools for the analysis of spermatozoan motility and chemotaxis that enable comparative studies across species and planar and 3D modes of swimming.

To accomplish the first objective we hypothesize that we can use a mechanistic model of the spermatozoon as *a priori* knowledge to be fitted directly to imaging data by maximum likelihood. Due to its physical nature and high level of detail, the model should impose morphological and kinematic constrains that accurately describe the object(s) depicted in the imaging data. If this methodological hypothesis holds true, then these methods is expected to allow us to distinguish different treatments or conditions as they will result in disparate parameterization of the model.

For the second objective we will address several issues, mostly related to the 2D+$Z(t)$ microscopy system and the data it generates. Can we improve the detection's precision and accuracy of the cell's position? Furthermore, can we use the approach developed for the first objective to describe spermatozoan motility and chemotaxis? If so, is there a difference in 3D sperm motility between *L. pictus* and *S. purpuratus*?

In *Chapter 2* we developed a simple framework for biological model-based image analysis by comparing *in silico* imaging data produced by a morphodynamical model to microscopy images in order to estimate the position, orientation, form and physical parameters of a spermatozoon and its surrounding media. We proved our framework allows us to track shape-shifting cells quite precisely (as well as a human does it). Better yet, we show we can infer the invisible flagellar bending patterns and positions by tracking only the head. The study in this chapter is the core of a manuscript in preparation.

The 3D comparative study is described in *Chapter 3*. Using the mechanistic models approach turned out to be computationally not feasible due the dimension of potential parameter space and data sets. We had to reduce the combinatorial and computational problem using a mixed approach. In a more traditional approach, that involved the development of an accurate estimation of the Z position associated with each frame, we used prior knowledge on the defocused appearance of the cells to estimate cell centroids in space and time. Then, a simplified kinematic model of the helical swimming paths was used to reconstitute the trajectories by piecewise helical segment fitting. Using this method we were able to reconfirm the trajectory parameters of previous manual analysis (Corkidi et al., 2008; Guerrero et al., 2010) and, more importantly, to reveal a difference in confining behavior between the two species considered. We then used the morphodynamical model to infer that *S. purpuratus* sperm have higher mean flagellar curvature in free as compared to confined swimming.

In *Chapter 4* we used the detection method developed in the third chapter to assess the free-swimming chemotactic behavior of *L. pictus* and *S. purpuratus*. By fitting linear mixed models to the data, we concluded that no chemotaxis was detected in the experimental conditions used.

Finally, we make a general discussion and conclusion connecting the previous chapters, highlighting the thesis contributions and future perspectives in *Chapter 5*.

1.5 Mathematical notation

For convenience of the reader, we present a few details of the mathematical notation used throughout this thesis.

Generally, bold symbols represent tensors (either column vectors or matrices) and normal typeface represent scalars. Parametric functions are represented normally by $f(x)$ but sometimes, abusing the notation, they can be represented without their variable(s) (f). Note that, due to the high number of parameters and variables used throughout this thesis, the same symbol can have different meanings in different chapters.

References

Abbott, B. P. et al. (2016). "Observation of gravitational waves from a binary black hole merger". *Physical Review Letters* 116.(6), pp. 1–16.

Afzelius, B. A. (1955). "The fine structure of the sea urchin spermatozoa as revealed by the electron microscope." *Zeitschrift fur Zellforschung und mikroskopische Anatomie (Vienna, Austria : 1948)* 42.(1-2), pp. 134–148.

Agard, D. A. and Sedat, J. W. (1983). "Three-dimensional architecture of a polytene nucleus". *Nature* 302, pp. 676–681.

Ahmed, M. M. and Mohamad, D. B. (2011). "Segmentation of brain magnetic resonance images (MRIs): a review". *International Journal of Advances in Soft Computing and its Applications* 3.(3), pp. 1–24.

Alvarez, L., Dai, L., Friedrich, B. M., Kashikar, N. D., Gregor, I., Pascal, R., and Kaupp, U. B. (2012). "The rate of change in Ca 2+ concentration controls sperm chemotaxis". *Journal of Cell Biology* 196.(5), pp. 653–663.

ATLAS Collaboration (2012). "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Physics Letters B* 716.(1), pp. 1–29.

Baba, S. A. and Mogami, Y. (1985). "An approach to digital image analysis of bending shapes of eukaryotic flagella and cilia". *Cell Motility* 5.(6), pp. 475–489.

Bishop, D. W. (1962). "Sperm Motility". *The American Physiological Society* 42.(1), pp. 338–360.

Böhmer, M. et al. (2005). "Ca2+ spikes in the flagellum control chemotactic behavior of sperm." *The EMBO journal* 24.(15), pp. 2741–2752.

Brokaw, C. J. (1965). "Non-sinusoidal bending waves of sperm flagella." *The Journal of experimental biology* 43, pp. 155–169.

— (1966). "Effects of increased viscosity on the movements of some invertebrate spermatozoa". *Control* 45, pp. 113–139.

— (1971). "Bend propagation by a sliding filament model for flagella". *Journal of Experimental Biology* 55, pp. 289–304.

— (1975). "Molecular mechanism for oscillation in flagella and muscle". *Proceedings of the National Academy of Sciences* 72.(8), pp. 3102–3106.

— (1979). "Calcium-induced asymmetrical beating of triton-demembranated sea urchin sperm flagella." *The Journal of cell biology* 82.(2), pp. 401–11.

— (1984). "Automated methods for estimation of sperm flagellar bending parameters." *Cell motility* 4.(6), pp. 417–430.

— (2003). "Swimming with three-dimensional flagellar bending waves". *Second International Symposium on Aqua-Biomechanics, Honolulu, HI.*

— (2014). "Computer simulation of flagellar movement X: Doublet pair splitting and bend propagation modeled using stochastic dynein kinetics". *Cytoskeleton* 71.(4), pp. 273–284.

Burnham, K. P. and Anderson, D. R. (2002). "Model Selection and Multimodel Inference". Second. Springer-Verlag.

Camalet, S. and Jülicher, F. (2000). "Generic aspects of axonemal beating". *New Journal of Physics* 2, pp. 24.1–24.23.

Chen, D. T. N., Heymann, M., Fraden, S., Nicastro, D., and Dogic, Z. (2015). "ATP consumption of eukaryotic flagella measured at a single-cell level." *Biophysical journal* 109.(12), pp. 2562–73.

Cibert, C. (2002). "Axonemal activity relative to the 2D/3D-waveform conversion of the flagellum". *Cell motility and the cytoskeleton* 51.(November 2001), pp. 89–111.

CMS Collaboration (2012). "Observation of a new boson at a mass of 125GeV with the CMS experiment at the LHC". *Physics Letter B* 716.(1), pp. 30–61.

Corkidi, G., Taboada, B., Wood, C. D., Guerrero, A., and Darszon, A. (2008). "Tracking sperm in three-dimensions." *Biochemical and Biophysical Research Communications* 373.(1), pp. 125–129.

Cortez, R. (2001). "The method of regularized Stokeslets". *SIAM Journal on Scientific Computing* 23.(4), pp. 1204–1225.

Cosson, J., Groison, A.-L., Suquet, M., Fauvel, C., Dreanno, C., and Billard, R. (2008). "Marine fish spermatozoa: racing ephemeral swimmers." *Reproduction (Cambridge, England)* 136.(3), pp. 277–94.

Cosson, J., Huitorel, P., and Gagnon, C. (2003). "How spermatozoa come to be confined to surfaces." *Cell motility and the cytoskeleton* 54.(1), pp. 56–63.

Crenshaw, H. C. (1989). "Kinematics of helical motion of microorganisms capable of motion with four degrees of freedom". *Biophysical journal* 56.(November), pp. 1029–1035.

— (1991). "A technique for tracking spermatozoa in three dimensions without viscous wall effects". *Comparative Spermatology* 20, pp. 353–357.

Crenshaw, H. C., Ciampaglio, C. N., and McHenry, M. (2000). "Analysis of the three-dimensional trajectories of organisms: estimates of velocity, curvature and torsion from positional information". *Journal of Experimental Biology* 203, pp. 961–982.

Darszon, A., Guerrero, A., Galindo, B. E., Nishigaki, T., and Wood, C. D. (2008). "Sperm-activating peptides in the regulation of ion fluxes, signal transduction and motility." *The International journal of developmental biology* 52.(5-6), pp. 595–606.

Darszon, A., Nishigaki, T., Beltran, C., and Treviño, C. L. (2011). "Calcium channels in the development, maturation, and function of spermatozoa." *Physiological reviews* 91.(4), pp. 1305–55.

de La Gorce, M., Paragios, N., and Fleet, D. J. (2008). "Model-based hand tracking with texture, shading and self-occlusions". *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR* (June), pp. 1–8.

Dufour, A., Liu, T. Y., Ducroz, C., Tournemenne, R., Cummings, B., Thibeaux, R., Guillen, N., Hero, A., and Olivo-Marin, J. C. (2015). "Signal processing challenges in quantitative 3-D cell morphology: More than meets the eye". *IEEE Signal Processing Magazine* 32.(1), pp. 30–40.

Dupont, S., Dorey, N., Stumpp, M., Melzner, F., and Thorndyke, M. (2013). "Long-term and trans-life-cycle effects of exposure to ocean acidification in the green sea urchin Strongylocentrotus droebachiensis". *Marine Biology* 160.(8), pp. 1835–1843.

Elgeti, J., Kaupp, U. B., and Gompper, G. (2010). "Hydrodynamics of sperm cells near surfaces." *Biophysical journal* 99.(4), pp. 1018–1026.

Espinal, J., Aldana, M., Guerrero, A., Wood, C. D., Darszon, A., and Martínez-Mekler, G. (2011). "Discrete dynamics model for the speract-activated Ca2+ signaling network relevant to sperm motility." *PloS one* 6.(8), e22619.

Espinal-Enríquez, J., Darszon, A., Guerrero, A., and Martínez-Mekler, G. (2014). "In silico determination of the effect of multi-target drugs on calcium dynamics signaling network underlying sea urchin spermatozoa motility". *PLOS one* 9.(8), pp. 1–11.

Evans, A. and Lauga, E. (2010). "Propulsion by passive filaments and active flagella near boundaries". *Physical Review E* 82.(4), pp. 1–12.

Friedrich, B. M. and Jülicher, F. (2007). “Chemotaxis of sperm cells.” *Proceedings of the National Academy of Sciences of the United States of America* 104.(33), pp. 13256–61.

Friedrich, B. M., Riedel-Kruse, I. H., Howard, J., and Jülicher, F. (2010). “High-precision tracking of sperm swimming fine structure provides strong test of resistive force theory.” *The Journal of experimental biology* 213.(Pt 8), pp. 1226–34.

Gadêlha, H., Gaffney, E. A., Smith, D. J., and Kirkman-Brown, J. C. (2010). “Nonlinear instability in flagellar dynamics: a novel modulation mechanism in sperm migration?” *Journal of the Royal Society Interface* 7, pp. 1689–1697.

Gaffney, E. A., Gadêlha, H., Smith, D. J., Blake, J. R., and Kirkman-Brown, J. C. (2011). “Mammalian sperm motility: Observation and Theory”. *Annual Review of Fluid Mechanics* 43.(1), pp. 501–528.

Gillies, E. a. E., Cannon, R. M. R., Green, R. B. R., and Pacey, A. A. (2009). “Hydrodynamic propulsion of human sperm”. *Journal of Fluid Mechanics* 625.(April), p. 445.

González-Cota, A. L., Silva, P. Â., Carneiro, J., and Darszon, A. (2015). “Single cell imaging reveals that the motility regulator speract induces a flagellar alkalinization that precedes and is independent of Ca 2+ influx in sea urchin spermatozoa”. *FEBS letters* 589, pp. 2146–2154.

Gray, J. and Hancock, G. (1955). “The propulsion of sea-urchin spermatozoa”. *Journal of Experimental Biology* 32.(4), pp. 802–814.

Guerrero, A., Carneiro, J., Pimentel, J. A., Wood, C. D., Corkidi, G., and Darszon, A. (2011). “Strategies for locating the female gamete: the importance of measuring sperm trajectories in three spatial dimensions”. *Molecular Human Reproduction* 17.(8), pp. 511–523.

Guerrero, A., Nishigaki, T., Carneiro, J., Yoshiro Tatsu, Wood, C. D., and Darszon, A. (2010). “Tuning sperm chemotaxis by calcium burst timing.” *Developmental Biology* 344.(1), pp. 52–65.

Hansbrough, J. R. and Garbers, D. L. (1981). “Speract. Purification and characterization of a peptide associated with eggs that activates spermatozoa.” *Journal of Biological Chemistry* 256.(3), pp. 1447–1452.

Harvey, E. B. (1956). “The American arbacia and other sea urchins”. Ed. by Drukkerij, N. V. and Thieme, G. J. Princeton, New Jersey: Princeton University Press.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). “The elements of statistical learning: data mining, inference and prediction”. Second. Vol. 27. (2). Springer, pp. 1–278.

Hernández, J. C. and Russell, M. P. (2010). “Substratum cavities affect growth-plasticity, allometry, movement and feeding rates in the sea urchin Strongylocentrotus purpuratus.” *The Journal of experimental biology* 213.(3), pp. 520–525.

Hines, M. and Blum, J. J. (1978). “Bend propagation in flagella I. Derivation of equations of motion and their simulation.” *Biophysical journal* 23.(1), pp. 41–57.

— (1979). “Bend propagation in flagella II. Incorporation of dynein cross-bridge kinetics into the equations of motion.” *Biophysical journal* 25.(3), pp. 421–441.

Horstmann, D. (2003). "From 1970 until present: the Keller-Segel model in chemotaxis and its consequences". *Jahresbericht der Deutschen Mathematiker-Vereinigung* 105.(3), pp. 103–165.

Humphries, S., Evans, J. P., and Simmons, L. W. (2008). "Sperm competition: linking form to function." *BMC evolutionary biology* 8, p. 319.

Ikawa, M., Inoue, N., Benham, A. M. A., and Okabe, M. (2010). "Fertilization: a sperm's journey to and interaction with the oocyte". *The Journal of Clinical Investigation* 120.(4), pp. 984–994.

Inglis, T., De Sterck, H., Sanders, G., Djambazian, H., Sladek, R., Sundararajan, S., and Hudson, T. J. (2010). "Multilevel space-time aggregation for bright field cell microscopy segmentation and tracking." *International journal of biomedical imaging* 2010, p. 582760.

Jikeli, J. F., Alvarez, L., Friedrich, B. M., Wilson, L. G., Pascal, R., Colin, R., Pichlo, M., Rennhack, A., Brenker, C., and Kaupp, U. B. (2015). "Sperm navigation along helical paths in 3D chemoattractant landscapes." *Nature Communications* 6, pp. 1–10.

Johnson, D. W., Monro, K., and Marshall, D. J. (2013). "The maintenance of sperm variability: Context-dependent selection on sperm morphology in a broadcast spawning invertebrate". *Evolution* 67.(5), pp. 1383–1395.

Johnson, R. E. and Brokaw, C. J. (1979). "Flagellar hydrodynamics. A comparison between resistive-force theory and slender-body theory". *Biophysical Journal* 25.(1), pp. 113–127.

Kaupp, U. B. (2012). "100 Years of Sperm Chemotaxis." *The Journal of general physiology* 140.(6), pp. 583–6.

Keller, E. F. and Segel, L. A. (1970). "Initiation of slime mold aggregation viewed as an instability". *Journal of Theoretical Biology* 26.(3), pp. 399–415.

Kervrann, C., Sorzano, C., Acton, S. T., Olivo-Marin, J.-C., and Unser, M. (2016). "A guided tour of selected image processing and analysis methods for fluorescence and electron microscopy". *IEEE Journal of Selected Topics in Signal Processing* 10.(1), pp. 6–30.

Knill, D. C. and Richards, W., eds. (1996). "Perception as Bayesian inference". Cambridge University Press.

Lauga, E. and Eloy, C. (2013). "Shape of optimal active flagella". *Journal of Fluid Mechanics* 730, R1.

Lesich, K. A., Pelle, D. W., and Lindemann, C. B. (2008). "Insights into the mechanism of ADP action on flagellar motility derived from studies on bull sperm". *Biophysical Journal* 95, pp. 472–482.

Li, C.-Y., Jiang, L.-Y., Chen, W.-Y., Li, K., Sheng, H.-Q., Ni, Y., Lu, J.-X., Xu, W.-X., Zhang, S.-Y., and Shi, Q.-X. (2010). "CFTR is essential for sperm fertilizing capacity and is correlated with sperm quality in humans." *Human reproduction (Oxford, England)* 25.(2), pp. 317–27.

Lillie, F. R. (1912). "The production of sperm iso-agglutinins by ova". *Science* 36.(929), pp. 527–530.

Lindemann, C. B. (1994). “A ”Geometric Clutch” hypothesis to explain oscillations of the axoneme of cilia and flagella”. *Journal of Theoretical Biology* 168.(2), pp. 175–189.

Marshall, R. D. and Luykx, P. (1973). “Observations of the centrioles of the sea urchin spermatozoon”. *Development Growth and Differentiation* 14.(4), pp. 311–323.

Nadeau, J. L., Cho, Y. B., Kühn, J., and Liewer, K. (2016). “Improved tracking and resolution of bacteria in holographic microscopy using dye and fluorescent protein labeling”. *Frontiers in Chemistry* 4.(April).

Nicastro, D., Schwartz, C., and Pierson, J. (2006). “The molecular architecture of axonemes revealed by cryoelectron tomography”. *Science* 313, pp. 944–948.

Patwary, N. and Preza, C. (2015). “Image restoration for three-dimensional fluorescence microscopy using an orthonormal basis for efficient representation of depth- variant point-spread functions”. *Biomedical Optics Express* 6.(10), pp. 3826–3841.

Pimentel, J. A., Carneiro, J., Darszon, A., and Corkidi, G. (2012). “A segmentation algorithm for automated tracking of fast swimming unlabelled cells in three dimensions.” *Journal of Microscopy* 245.(Pt 1), pp. 72–81.

Purcell, E. M. (1977). “Life at low Reynolds number”. *American Journal of Physics* 45.(3), p. 11.

Rabbani, H. and Javanmard, S. H. (2011). “Image analysis in in vivo stem cell tracking”. 1.(4), pp. 123–142.

Ren, Y., Wen, P., Li, S., Liang, Y., and Huang, W. (2010). “An improved algorithm of rat sperm image segmentation and counting”. *Intelligent Computing and Integrated Systems (ICISS), 2010 International Conference on*. 3. IEEE, pp. 201–204.

Riedel-Kruse, I. H., Hilfinger, A., Howard, J., and Jülicher, F. (2007). “How molecular motors shape the flagellar beat.” *HFSP journal* 1.(3), pp. 192–208.

Rikmenspoel, R. (1984). “Movements and active moments of bull sperm flagella as a function of temperature and viscosity”. *Journal of Experimental Biology* 108, pp. 205–230.

Rosen, J. and Brooker, G. (2008). “Non-scanning motionless fluorescence three-dimensional holographic microscopy”. *Nature Photonics* 2, pp. 190–195.

Rosen, R. (1991). “Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life”. Columbia University Press.

Seifert, R. et al. (2015). “The CatSper channel controls chemosensation in sea urchin sperm”. *The EMBO Journal* 34.(3), pp. 379–392.

Silva-Villalobos, F., Pimentel, J. A., Darszon, A., and Corkidi, G. (2014). “Imaging of the 3D dynamics of flagellar beating in human sperm.” *Conference Proceedings IEEE Engineering in Medicine and Biology Society* 2014, pp. 190–193.

Smith, D. J., Gaffney, E. A., Blake, J. R., and Kirkman-Brown, J. C. (2009a). “Human sperm accumulation near surfaces: a simulation study”. *Journal of Fluid Mechanics* 621, p. 289.

Smith, D. J., Gaffney, E. A., Gadêlha, H., Kapur, N., and Kirkman-Brown, J. C. (2009b). “Bend propagation in the flagella of migrating muman sperm, and its modulation by viscosity”. *Cell Motility and the Cytoskeleton* 66, pp. 220–236.

Spagnolie, S. E. and Lauga, E. (2010). “The optimal elastic flagellum”. *Physics of Fluids* 22.(3), pp. 1–15.

Su, T.-W., Xue, L., and Ozcan, A. (2012). “High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories”. *Proceedings of the National Academy of Sciences* 109.(40), pp. 3–7.

Tam, D. S. W. (2008). “Motion at low Reynolds number”. PhD. Massachusetts Institute of Technology.

Tsai, D., Flagg, M., Nakazawa, A., and Rehg, J. M. (2012). “Motion coherent tracking using multi-label MRF optimization”. *International Journal of Computer Vision* 100.(2), pp. 190–202.

Vulović, M., Ravelli, R. B. G., Vliet, L. J. van, Koster, A. J., Lazić, I., Lücken, U., Rullgård, H., Öktem, O., and Rieger, B. (2013). “Image formation modeling in cryo-electron microscopy”. *Journal of Structural Biology* 183.(1), pp. 19–32.

Ward, G. E., Brokaw, C. J., Garbers, D. L., and Vacquier, V. D. (1985). “Chemotaxis of Arbacia punctulata spermatozoa to resact, a peptide from the egg jelly layer”. *Journal of Cell Biology* 101.(6), pp. 2324–2329.

Wood, C. D., Darszon, A., and Whitaker, M. (2003). “Speract induces calcium oscillations in the sperm tail”. *The Journal of Cell Biology* 161.(1), pp. 89–101.

Woolley, D. M. and Vernon, G. G. (2001). “A study of helical and planar waves on sea urchin sperm flagella, with a theory of how they are generated.” *The Journal of Experimental Biology* 204.(Pt 7), pp. 1333–45.

Xu, C. and Prince, J. L. (1998). “Snakes, shapes, and gradient vector flow”. *IEEE Transactions on Image Processing* 7.(3), pp. 359–369.

Zhang, B., Zerubia, J., and Olivo-Marin, J.-C. (2007). “Gaussian approximations of fluorescence microscope point-spread function models.” *Applied optics* 46.(10), pp. 1819–1829.

Chapter 2

Morphodynamical image analysis of spermatozoa swimming in the plane

Pedro Ângelo Silva[1], Adán Guerrero[2], Alberto Darszon and Jorge Carneiro[1].

[1] Instituto Gulbenkian de Ciência, Portugal; [2] Instituto de Biotecnología, Universidad Nacional Autónoma de México, Estados Unidos de México

Abstract

Modern imaging analysis techniques rely on *a priori* knowledge to detect and track objects in imaging data. These methods are either unsupervised, in which the resulting detections are filtered *a posteriori* by criteria based on the target object, or supervised, where training data sets are used to generate templates or calibrate them. These method often fail to find target objects or include extraneous, undesired ones, thus requiring painstaking manual curation before or after the analysis. Here we present a new framework that uses morphodynamical models as *a priori* knowledge. We will make a proof-of-concept using the imaging of spermatozoa as case study. In this case, the model will include the form and deformation of these cells, the physics that propel them forward and the imaging acquisition process. We fitted the model to both synthetic and experimental imaging data showing we can recover

the true parameters used to generate the synthetic data and that we can track a spermatozoon as good as a human can. Using this framework we were even able to infer the form and position of the flagellum by tracking only the head, *i.e.*, in the images in which the flagella where not visible. Using such morphodynamical models as *a priori* knowledge might give the necessary information that automated imaging analysis methods demand in order to become more reliable, independent and robust.

Acknowledgments

Jorge Carneiro and Pedro Ângelo Silva designed the research. Pedro Ângelo Silva performed the work and Jorge Carneiro supervised. Adán Guerrero and Alberto Darszon provided the 2D+t imaging data. Pedro Ângelo Silva, Alberto Darszon and Jorge Carneiro contributed to writing the manuscript. This work was funded by Fundação para a Ciência e Tecnologia, Portugal, (SFRH/BD/79261/2011), Instituto Gulbenkian de Ciência, Portugal, and Instituto de Biotecnología, UNAM, Mexico.
The work developed in this chapter is the object of a manuscript in preparation, to be submitted to an international peer-reviewed journal.

2.1 Introduction

The automation of microscopy systems and the emergence of multidimensional measurement of biological data has recently been producing data at a overwhelming pace, one that current imaging analysis procedures are unable to cope with. There has been an effort to automate image analysis but the specifics of each study makes it difficult to have general tools that allow to extract information from the imaging data. Therefore, scientists often wander on empirical testing of these tools parameters. This creates not only highly biased results but also irreproducible ones, as many of these parameters are rarely reported in many studies. As such, automatic methods for image analysis which do not require parameter tweaking and can be generally used in an wide array of studies are desired.

The major difference between a computer and a human when extracting information from imaging data is that the latter is rich in *a priori*, implicit knowledge. Based on past experiences, a person uses implicit models to distinguish between background, cell, nucleus, and so on. In comparison, the computer algorithms are very limited in the resources they can deploy, which are restricted to raw spectral, spatial and temporal data, and a set of statistical tools but no model to connect them all. What if the computer was endowed with *a priori* knowledge on the object to study, including its morphology and dynamics, and of the microscopy setup such as the point-spread function and lighting conditions? It should then be possible for the computer algorithms to extract information as good as any human. Better, in fact, if we consider the reproducibility of the deterministic methods. Not only that, what if the computer can extract information that a human cannot, *e.g.* a feature non-linearly correlated with morphological dynamics which is not directly measurable/noticeable in the imaging data?

When one is dealing with imaging data, there is a distinction between image processing (*i.e.* low-level image manipulation such as reduction of noise and background subtraction) and image analysis – the extraction of information from imaging data. For this end, there is a long tradition of segmenting the image or fitting parametric models to each time-point independently and only then the extracted information is used to fit morphodynamical models (Brokaw, 1984; Baba and Mogami, 1985; Friedrich et al., 2010; Su et al., 2012; Su et al., 2013; Jikeli et al., 2015). In the computer vision field, face-recognition and tracking usually resorts to the Lucas-Kanade algorithm as it allows to deform templates in order to identify them in the next frame (Baker and Matthews, 2004). Other methods use maximum likelihood to estimate the data that produced a given imaging data set of immobile structures (Vulović et al., 2013; Kervrann et al., 2016). To the best of our knowledge, methods where the form and deformation of the object, its interaction with surrounding media and transformation of the imaging process are all combined as purely mathematical and theoretical descriptors have not been implemented to estimate the morphology and position of motile cells that change their shape in time.

Our objective is to develop a framework to detect and track cells using a morphodynamical model. We will generate synthetic imaging data based on the model and compare the synthetic images rendered from this model directly to experimental imaging data by correlation, a proxy of likelihood (Zucker, 2003). We will show here that this method is able to detect and track sea urchin spermatozoa as good as a human and even outperform the human ability to recognize patterns by inferring flagellar positions and conformations when these structures are not visible in the images. As such, the use of mechanistic models within the image analysis procedures is instrumental by increasing the resolution of the analysis and by allowing to infer structures that are missing from the imaging data.

2.2 Materials and Methods

2.2.1 Morphodynamical model

Consider a mathematical model of a cell in which the morphology changes are defined within the model itself – this is a morphodynamical model. In our particular case, the spermatozoon cell model ($\boldsymbol{S}$) is defined by the shape ($\boldsymbol{\Sigma}$) and mechanics ($\boldsymbol{\Phi}$) modules.

Shape

We can define the morphology of a spermatozoon by defining two regions, the head and the flagellum. We assume the head is a revolution ellipsoid with half axes a and $b = c$. The flagellum is composed of N rods with total length L (μm). The conformation of the flagellum at a particular time t is given by its curvature κ along flagellar arclength s, a first order traveling wave: $\kappa(s,t) = K_0 + (A_0 + A_1 e^{-A_1 s}) \cos(\omega_\kappa t - \lambda_\kappa s + \phi)$, which is defined by the mean flagellar curvature (K_0), basal curvature amplitude (A_0), exponential term of curvature amplitude (A_1), angular velocity of beating (ω_κ) and curvature wavelength (λ_κ). The phase (ϕ) near the apical part of the flagellum (first 0.5 μm) is ϕ_1 while for the remaining flagellum is ϕ_0. For a planar beating, we assume the flagellar torsion along arclength to be zero ($\tau(s,t) = T_0 = 0$ rad.μm^{-1}). The intrinsic flagellar position ($\boldsymbol{r}_f(s,t)$) is calculated by arclength integration of the Cosserat frame (Cao et al., 2006; Jikeli et al., 2015), which is defined by the orthonormal basis $\boldsymbol{e}_1(s,t)$, $\boldsymbol{e}_2(s,t)$ and $\boldsymbol{e}_3(s,t)$: $\partial \boldsymbol{r}_f(s,t)/\partial s = \boldsymbol{e}_3(s,t)$, $\partial \boldsymbol{e}_3(s,t)/\partial s = \kappa(s,t)\boldsymbol{e}_1(s,t)$, $\partial \boldsymbol{e}_1(s,t)/\partial s = -\kappa(s,t)\boldsymbol{e}_3(s,t) + \tau(s,t)\boldsymbol{e}_2(s,t)$, $\partial \boldsymbol{e}_2(s,t)/\partial s = -\tau(s,t)\boldsymbol{e}_1(s,t)$. Note $\boldsymbol{e}_3$ represents the centreline along the flagellum, $\boldsymbol{r}_f(0,t) = \{-a,0,0\}$, $\boldsymbol{e}_1(0,t) = \{0,-1,0\}$, $\boldsymbol{e}_2(0,t) = \{0,0,1\}$ and $\boldsymbol{e}_3(0,t) = \{-1,0,0\}$. At a given time t, the centroid of the head has extrinsic position $\boldsymbol{R}_h(t)$ and orientation matrix $\boldsymbol{\Theta}(t)$, which rotates from extrinsic to intrinsic coordinates. We can define the orientation matrix as combination of rotations on the X, Y and Z axes – Euler angles. In our specific case $\boldsymbol{\Theta} = \boldsymbol{R}_X \boldsymbol{R}_Y \boldsymbol{R}_Z$, where:

$$\boldsymbol{R}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\nu_X) & \sin(\nu_X) \\ 0 & -\sin(\nu_X) & \cos(\nu_X) \end{bmatrix};\boldsymbol{R}_Y = \begin{bmatrix} \cos(\nu_Y) & 0 & -\sin(\nu_Y) \\ 0 & 1 & 0 \\ \sin(\nu_Y) & 0 & \cos(\nu_Y) \end{bmatrix};\boldsymbol{R}_Z = \begin{bmatrix} \cos(\nu_Z) & \sin(\nu_Z) & 0 \\ -\sin(\nu_Z) & \cos(\nu_Z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

are rotation matrices on the X, Y and Z axes, respectively. Notice we dropped the time from the previous equations for simplicity. Thus, the orientation of a cell at a given time is specified by the parameters $\nu_X(t)$, $\nu_Y(t)$ and $\nu_Z(t)$. In the confined swimming case we assume $\nu_X(t) = 0$ and $\nu_Y(t) = 0$.

Mechanics

As the cell moves within a viscous fluid, the latter exerts drag forces which, under some conditions, propel the cell in the media. The physics behind it have been shown to be approximated by Resistive Force Theory (RFT) (Gray and Hancock, 1955; Friedrich et al., 2010; Jikeli et al., 2015), which we will use here. Briefly, the drag force density that fluid exerts on the flagellar piece is $\boldsymbol{f}(s,t) = \xi_{\|}\partial \dot{\boldsymbol{r}}_{f,\|}(s,t)/\partial t + \xi_{\perp}\partial \dot{\boldsymbol{r}}_{f,\perp}(s,t)/\partial t$, where $\dot{\boldsymbol{r}}_{f,\|}$ and $\dot{\boldsymbol{r}}_{f,\perp}$ are the tangent and normal components, respectively, of the mean flagellar piece velocity $\dot{\boldsymbol{r}}_f(s,t) = (\boldsymbol{\Psi}'(\delta\boldsymbol{\theta}(t)) \cdot \boldsymbol{r}(s,t+\delta) + \delta\boldsymbol{v}(t) - \boldsymbol{r}(s,t+\delta))/\delta$ during the time interval δ. Note $\boldsymbol{v}$ and $\boldsymbol{\theta}$ are the intrinsic head translational and rotational velocities, respectively, $\boldsymbol{\Psi}'(\boldsymbol{\nu})$ is the an approximation of Rodrigues rotation formula ($\boldsymbol{\Psi}'(\boldsymbol{\nu})$) for small angles ($\sin\beta \simeq \beta$ and $\cos\beta \simeq 1$) and $\xi_{\|}$ and $\xi_{\perp}$ are the tangent and normal drag coefficients of the force the fluid exerts on the flagellum, respectively.

We can obtain the intrinsic head velocities by solving the system of force and torque equilibria defined by $\boldsymbol{\xi_T}\boldsymbol{v}(t)\int_0^L \boldsymbol{f}(s,t)\partial s = 0 \wedge \boldsymbol{\xi_R}\boldsymbol{\theta}(t) - \int_0^L \boldsymbol{f}(s,t)\times\boldsymbol{r}_f(s,t)\partial s = 0$, where $\boldsymbol{\xi_T}$ and $\boldsymbol{\xi_R}$ are the translational and rotational drag coefficients of the head and the symbol $\times$ denotes the crossproduct operator. The intrinsic head velocities are transformed into extrinsic velocities by multiplication of the transposed orientation matrix and applied to the cell to calculate the new extrinsic position $\boldsymbol{R}_h(t+\delta) = \boldsymbol{R}_h(t) + \delta\boldsymbol{\Theta}(t)^T \cdot \boldsymbol{v}(t)$ and orientation $\boldsymbol{\Theta}(t+\delta) = \boldsymbol{\Psi}(\delta\boldsymbol{\Theta}(t)^T \cdot \boldsymbol{\theta}(t)) \cdot \boldsymbol{\Theta}(t)$. Note we used the approximated Rodrigues rotation matrix to calculate the local flagellar velocity in order to make the system linear but this approximation is not needed when we calculate the extrinsic velocities from the intrinsic ones.

Assuming a dynamic viscosity η we can calculate the translational and rotational drags for the head based on its size using Perrin's formulas (Perrin, 1936). As we used Euler integration method, we will update the system every δ time steps and then we solve it for the intrinsic translational and rotational velocities assuming they are constant within each time period δ. For spatial integration of the flagellum, we discretized it in N segments and calculated the intrinsic midpoint position $\kappa[i,t]$ where $i = L/N$ We determined $N = 149$ and $\delta = 50$ μs to produce errors below 1% when calculating the intrinsic translational and rotational velocities, comparing to $N = 499$ and $\delta = 10$ μs.

The spermatozoon model $\boldsymbol{S}(\boldsymbol{\Sigma},\boldsymbol{\Phi})$ is thus completely defined by its shape parameters $\boldsymbol{\Sigma} = (a,b,c,L,N,\boldsymbol{R_h},\boldsymbol{\Theta},K_0,A_0,A_1,\phi_0,\phi_1,\omega_\kappa,\lambda_\kappa,T_0)$ and its physical or mechanical parameters $\boldsymbol{\Phi} = (\boldsymbol{\xi_T},\boldsymbol{\xi_R},\xi_\parallel,\xi\perp)$. Although not explicit here, some of these parameters are dependent on time, as shown above.

To measure the difference between two models instances, say $\boldsymbol{S}^a$ and $\boldsymbol{S}^b$, we defined a convenient list of 14 parameters, composed of some of basic parameters and their ratios. The distance between the parameter lists, $\boldsymbol{P}^a$ and $\boldsymbol{P}^b$, of the two models is computed as:

$$\chi^2 = \sum_j \frac{(p_j^a - p_j^b)^2}{(p_j^b)}, \tag{2.1}$$

where $j \in \{1,...,14\}$ is the index of the parameter or ratio in the list $\{a, a/b, L, K_0, A_0,\ A_1, \omega_\kappa, \lambda_\kappa, \phi_0, \phi_1, \xi_{T_x}, \xi_{T_y}/\xi_{T_x}, \xi_\parallel, \xi_\perp/\xi_\parallel\}$. Note we did not include in this measure some parameters as they will not affect the cell's motility if we assume planar flagellar beating, which we did in this chapter. Also note only the ratios between some of parameters affect the swimming path of the cell, the reason why we compare the distance of those ratios and not of the parameters themselves. We will usually refer to the distance to the ground truth model or to the true parameters, meaning this set of parameters is considered as the model S_b in the formula above.

2.2.2 Comparing and fitting the model to imaging data

We generated a model instance at time t_i with a given set of parameters $\boldsymbol{S}$ and rendered it in a new image ($\boldsymbol{I}_{S;t=t_i}$) with weights (*i.e.* adimensional pixel intensity) w_h and w_f for the head and flagellum, respectively. Then, we convolved the $\boldsymbol{I}_{S;t=t_i}$ with a Gaussian filter $\mathcal{G}(0,(\sigma/6)^2)$ (kernel of size σ, down-scaled by area interpolation (refer to OpenCV function *resize*) to match the spatial and temporal resolution of the experimental imaging data ($\boldsymbol{I}_{t=t_i}$)). The correlation score of a parameter set of the model is the sum of the correlation coefficients between the imaging data ($\boldsymbol{I}_{t=t_i}$) and $\boldsymbol{I}_{S;t=t_i}$, only of the pixels that are within the ρ_r radius around the object: $\text{Score}(\boldsymbol{S}, t_i, t_j, \rho_r, \sigma, w_h, w_f) = \sum_{k=i}^{j} \text{Correlation}(\boldsymbol{I}_{t=t_k}, \boldsymbol{I}_{S;t=t_k} \mid \rho_r, \sigma, w_h, w_f)$, for $i \leq j$, $t_i \leq t_j$, $\forall i, j \in \mathbb{N}_0$, where $\mathbb{N}$ is set of the natural numbers. Note that we will also refer to the average correlation coefficient as $\text{Score}(\boldsymbol{S}, t_i, t_f, \rho_r, \sigma, w_h, w_f)/(j - i + 1)$.

The likelihood is the probability of a parameter set given the data. Thus, we searched the parameter space for the set which best describes the data by maximizing this probability. Our model has many parameters, is non-linear and the solution space is not convex, which makes it difficult to define the maximum likelihood algebraically and also to find it with optimization algorithms. To overcome this difficulty we used the result of (Zucker, 2003) who have shown that under some conditions, by maximizing the cross-correlation between a model and imaging data one is effectively maximizing the likelihood. We implemented a simple Monte Carlo (MC) evolutionary algorithm (fig. 2.2 B), where the fitness function is given by the correlation score. To understand the procedure, let us introduce the following symbol: $\boldsymbol{S}_k^g$ is the sperm model indexed k at iteration g of the evolutionary algorithm. At iteration zero ($g = 0$) we input our initial parameter set $\boldsymbol{S}^0$ and replicated it $N_c = 1000$ times while adding Gaussian noise to the parameters, except for one set which is an exact replicate. We then started the iteration 1 at time t_i and computed the individual score of each parameter set ($\text{Score}(\boldsymbol{S}_k^1)$, notice the abuse of notation by discarding all the remainder parameters). Then we selected the $N_b = 10$ fittest parameter sets (*i.e.*

the top highest Score($\boldsymbol{S}_k^1$)) and each one of these generated 100 new children ($\boldsymbol{S}_k^2$) to be evaluated and at next iteration. We repeated this for $N_i = 20$ iterations. Unless stated otherwise, we first performed preliminary optimization of only the shape parameters ($\boldsymbol{\Sigma}_0$) at the initial time ($t_i = t_f$) and subsequently performed the optimization of the all the mechanical and temporal parameters (see section 2.3.1) on the whole imaging data set ($t_i \neq t_f$). The Gaussian perturbations to the parameters were performed using a zero-centered Gaussian with standard deviation 5% of the parameter value, except for angles which were 5% of π rad. At each passing iteration $g \in \mathbb{N}_0$, the percentage dropped as $5\%/(1+g)$. Some logical constrains were applied, *e.g.* $1 < \xi_\perp/\xi_\parallel < 2$ and $a > b = c > 0$. To increase the speed of the algorithm while fitting the mechanical and temporal parameters, we aborted the computation of the correlation score for children with average correlation that was below 0.5 at time $t_\varnothing$, where $t_i \leq t_\varnothing < t_f$, and the correlation score for that model was assumed to be Score($\boldsymbol{S}_k^g, t_i, t_\varnothing, \rho_r, \sigma, w_h, w_f$). Similarly, the computation of the score was aborted for candidate solutions with an average score that is lesser or equal to half the average score of the best solution obtained until that stage in the execution of MC evolutionary algorithm.

2.2.3 Implementation details

Unless stated otherwise, all the algorithms for image manipulation, model definition and optimization were developed and encoded in C/C++ using the free computer vision library OpenCV 3 (Intel, Santa Clara, United States of America). Simulations were run in a Intel® i7-6700HQ CPU @ 2.60GHz×8 processor (Intel, Santa Clara, United States of America) in Ubuntu 16.04 (Canonical, London, United Kingdom) using the *parallel* (Tange, 2011) to run several simulations at the same time. Plots, statistics and figures were performed using R v.3.0.3 (R Foundation for Statistical Computing, Vienna, Austria) and LaTeXusing the *tikZ* package, respectively.

2.2.4 Imaging data

L. pictus and *S. purpuratus*

All *in vivo* imaging data was obatined as described in Guerrero et al., (2010), with a few alterations. A brief description follows.

Materials. *L. pictus* and *S. purpuratus* spermatozoa (Marinus Inc., Long Beach, CA, USA; Pamanes S. A. de C.V., Ensenada, Mexico) were extracted undiluted by intracoelomic injection of 0.5 M KCl, stored on ice and used within 24 hours. artificial sea water (ASW) was prepared with 486 mM NaCl, 10 mM KCl, 10 mM $CaCl_2$, 26 mM $MgCl_2$, 30 mM $MgSO_4$, 2.5 mM $NaHCO_3$, 10 mM HEPES and 1 mM EDTA, up to 950-1000 mOsm. Final pH was 8.0 and 7.4 for *S. purpuratus* and *L. pictus*, respectively. Low Ca^{2+} ASW was prepared similarly to ASW but using 1 mM $CaCl_2$ and setting the pH to 7.0. Fluo-4-AM and pluronic F-127 were from Molecular Probes, Inc. (Eugene, OR, USA). All other reagents were from Sigma-Aldrich (Toluca, Edo de Mexico, Mexico), unless stated otherwise.

Labeling spermatozoa and loading to incubator chamber. Ten volumes of low Ca^{2+} ASW containing 0.2% wt/vol pluronic F-127 and 20 μM of Fluo-4 AM were used to suspend undiluted spermatozoa of *S. purpuratus*. After incubation for two hours at 14 °C, spermatozoa were stored in the dark and on ice. To prevent cells to adhere to the glass, all coverslips were coated in PolyHEME (poly(2-hydroxyethyl methacrylate)). Sperm from either species were diluted in ASW into a reusable chamber and maintained at 15 °C throughout the experiment.

Fluorescence imaging of *S. purpuratus* spermatozoa. Images were acquired with Nikon Plan Fluor 40× 1.3 NA objective using a Chroma filter set (ex, HQ470/ 40×; DC, 505DCXRU; em, HQ510LP) and recorded on a EMCCD Andor camera (DV887, Andor iXon). Stroboscopic lighting was used such that 2 ms of flash was synchronized with the camera exposure (also 2 ms). Images were collected with Andor iQ 1.8 software (Andor Bioimaging, NC) with framerate 200 Hz in cropped-chip mode (window = 60×60 μm). Pixel resolution is 1.56 μm.pixel^{-1}.

A light splitter was used to produce side-by-side image frames in which the whole spermatozoa or only the sperm heads were visible. We noticed that the light splitter produced misaligned images. The image side where sperm head where labelled and visible was misaligned $(-3.15, 2.36)$ and -0.0605 rad relative to the side of the image where the whole-cell label was visible. The misalignment was corrected by estimating of the affine transformation matrix that maximizes the Enhanced Correlation Coefficient (Evangelidis and Psarakis, 2008) of the projections of the two imaging data sets. The projection was performed using the bitwise OR operator of the first fifteen images at ten by ten intervals of each data set (*i.e.* frames number

$\{0, 10, ..., 140\}$). The algorithm was implemented using the OpenCV and NumPy libraries in Python 2.7.12. The affine transformation was applied to the initial position and orientation of the model fitted to the image with invisible flagella of *S. purpuratus* to project the coordinates onto the image where the flagella are visible.

Imaging of *L. pictus* spermatozoa *L. pictus* images were acquired with Optronics CR500X2 camera at a framerate of 500 Hz in full chip in a bright-field Olympus inverted microscope (IX71) with a 60 × 1.6 0.7 NA long working distance objective. Pixel resolution is 0.33 μm.pixel^{-1}. Later these were processed in ImageJ as follows: (1) Image type 16 bits, (2) Smooth, (3) Subtract background (Rolling bar radius $= 13px$, Light background), (4) Create an average time projection (1000 frames) (5) Subtract the resultant image to each frame of the stack (32 bit result), (6) Enhance contrast (Normalize, Use stack histogram), (7) Invert and (8) Image type 16 bits. Trajectories and flagella were tracked using BohBoh software v3.29 (BohBohSoft, Tokyo, Japan).

In silico

The model coordinates were transformed to image coordinates taking into account the image origin, spatial resolution and time. With the parameters used, there can be an error up to 50 μs between image and model, due to their temporal resolutions. The head was rendered as an ellipse with the appropriate parameters (see section *Comparing imaging data to model*) with basal intensity multiplied by w_h and the discrete flagellum was rendered as linear segments with width w and basal intensity multiplied by w_f ($w_f = 1$ to draw the flagellum and $w_f = 0$ to not draw it). The point-spread function (PSF) function was approximated by a Gaussian filter with $\sigma/6$ standard deviation (*i.e.* approximated as an odd integer in image discrete dimensions) that was convolved to the rendered model image. The mask to be used to compute the correlation (see section 2.2.2) was produced by the same procedure that renders the model sperm image using a radius r that was added to both head half axis and flagellar width. Parameters used: $a = 2.50$ μm, $b = c = 1.25$ μm, $L = 45$ μm, $N = 150$, $\boldsymbol{R}_h(0) = \{45, 45, 0\}$ μm, $\boldsymbol{\Theta}(0) =$ Identity Matrix, $\boldsymbol{\xi_T} = \{3.06, 3.51, 3.51\}$ pN.s.μm^{-1}, $\boldsymbol{\xi_R} = \{8.55, 15.96, 15.96\}$ pN.s.μm, $\xi_{\|} = 0.300$ pN.s.μm^{-2}, $\xi_{\perp} = 0.525$ pN.s.μm^{-2}, $K_0 = 0.035$ rad.μm^{-1}, $A_0 = 0.144$ rad.μm^{-1}, $A_1 = 0.100$, $\omega_\kappa = 182.485$ rad.s^{-1}, $\lambda_\kappa = 0.184$ rad.μm^{-1}, $\phi_1 = 1.571$ rad and the remainder are zero.

2.3 Results

2.3.1 Correlation scores sensitivity to RFT parameters

We built a morphodynamical model of a sea urchin spermatozoon where the shape of the head is given by an ellipse, the flagellum is parameterized by the curvature along its length and the mechanics of the interactions with the fluid is described by Resistive Force Theory (RFT). Although an approximation of the physical reality (more realistically described by Slender Body Theory (SBT) (Johnson and Brokaw, 1979)), this framework has been shown to model the swimming behavior of sperm cells quite accurately (Gray and Hancock, 1955; Friedrich et al., 2010; Jikeli et al., 2015). The state of the model at any given time is a specification of the position, orientation and form of the spermatozoon. To compare the model to imaging data, the state was used to render a synthetic image, this image was convolved with PSF expected for the microscopy setup used and then correlated it with the corresponding frame of the experimental imaging data. The maximization of correlation between the image rendered from the model and the experimental image allows to maximize the likelihood of the model parameters given the latter image. To accelerate computation and reduce effects of spurious noise the correlation is only performed up to a maximum radius around the modeled cell (fig. 2.1 A).

To assess whether this method allows to make precise and accurate estimates of the parameters of spermatozoon model, we generated a synthetic imaging data set (in which the *real* parameter set by which it was generated are known by definition, the ground truth) and we explored how changing the different parameters independently or in groups affected the correlation. The first question was how does the procees to render the model image affects the correlation coefficient? To address this, we took the initial frame and correlated it with the ground truth model but changing the width of the flagellum (w), the standard deviation ($\sigma/6$) of the Gaussian function that approximates the PSF and the radius (ρ_r) around the cell that defines the area used for correlation (fig. 2.1B). We confirmed that the correlation is maximal for the expected standard deviation and flagellar width ($\sigma = 5$ μm, $w = 0.5$ μm). As we increase the area of correlation by increasing ρ_r, the peak of every flagellar width used converges to the expected standard deviation. For this reason we decided to use higher radius values but not so high it would be affected by other sperm cells or debris on the experimental data so we set $r = 15$ μm for all subsequent analysis.

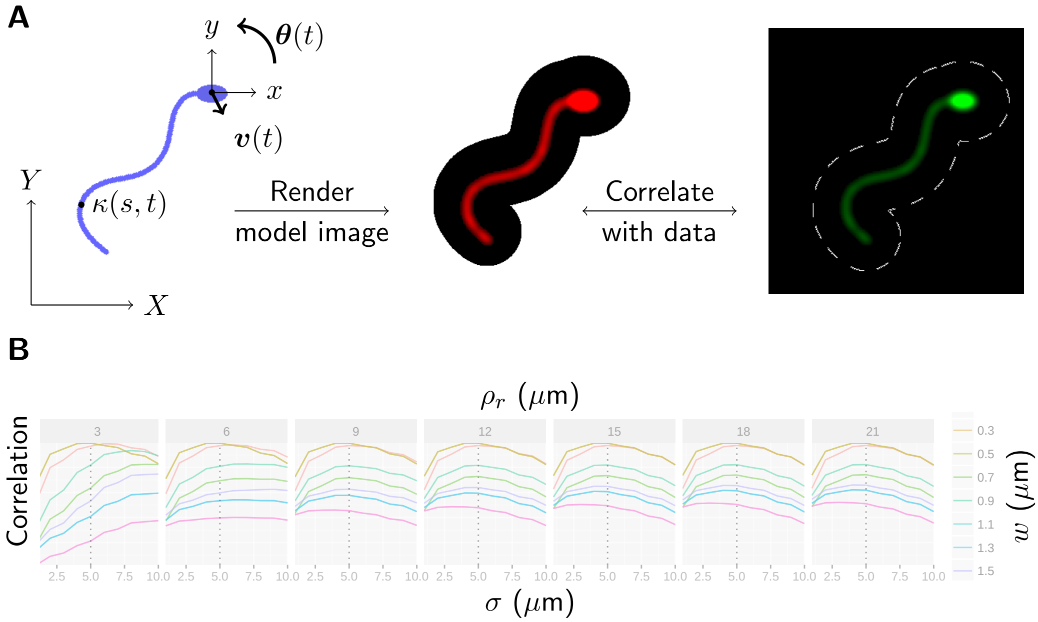


Figure 2.1: Comparing model to imaging data. **A** – the model state (blue) specifying the position, orientation and conformation of the spermatozoon at a given time is used to render the model image (red), which is then correlated to the imaging data (green). **B** – Correlation radius (ρ_r), flagellar width (w) and Gaussian kernel size of the PSF (σ) effect on the correlation; Dashed line limits the correlation area.

The second question was: how do the shape and mechanic parameters affect the correlation score and how sensitive this score is to changes in these parameters? Again, taking the ground truth model, we perturbed each parameter independently and checked how the correlation coefficient changed along time. We observed different types of sensitivities (fig. 2.2A) and decided to categorize them as *Low sensitivity* (A_1, ϕ_1 and η), *High sensitivity* (X and Y) and *Temporal sensitivity* (ν_Z, K_0, A_0, ϕ_0, λ_κ, ω_κ, $\xi_{||}$ and $\xi_\perp$) (please refer to section 2.2.1 for parameter description). As expected, the correlation is very sensitive to most mechanical parameters as time progresses and this also occurs for some shape descriptors (*i.e.* flagellar shape and initial head orientation) as they affect how and where the cell swims. The initial position slightly affects the correlation score but this is rather insensitive to the exponential term of the curvature amplitude, the initial apical flagellar phase and the media viscosity. By closer examination of the initial time-point, it is possible to see that the correlation score is sensitive to shape parameters, suggesting that it should be possible to estimate these parameters using the information of a single frame.

We then needed to choose the method to fit the model to the imaging data. Although many different optimization algorithms exist, it was not within the scope of the present work to compare them and select the one that performs better. For this reason, we used a simple MC by tournament (fig. 2.2 B) – the best scoring parameter sets of the current iteration are mutated and give rise to the candidate parameter sets in the next iteration, whereas the remainder sets do not propagate. Moreover, the mutations lose strength with each passing iteration, to ensure local exploitation. Knowing that rendering the model image and correlating it to the experimental imaging data is the most expensive computational step and taking the previous results into account, we decided to first fit the shape parameters ($\mathbf{\Sigma}$) to the initial time-point for some iterations and only then fit all the mechanical and flagellar shape parameters ($\mathbf{\Sigma}$ and $\mathbf{\Phi}$) to the whole imaging data set. Also note that, for the same computational reasons, parameter sets that lead to average correlation scores across the images that were lower than half of the current best maximum correlation coefficient were not further iterated and only the cells at the last iterated time-point were considered for the next iteration (section 2.2.2 for more details). With a model, a comparison criteria and an optimization strategy we were prepared to characterize spermatozoa in imaging data.

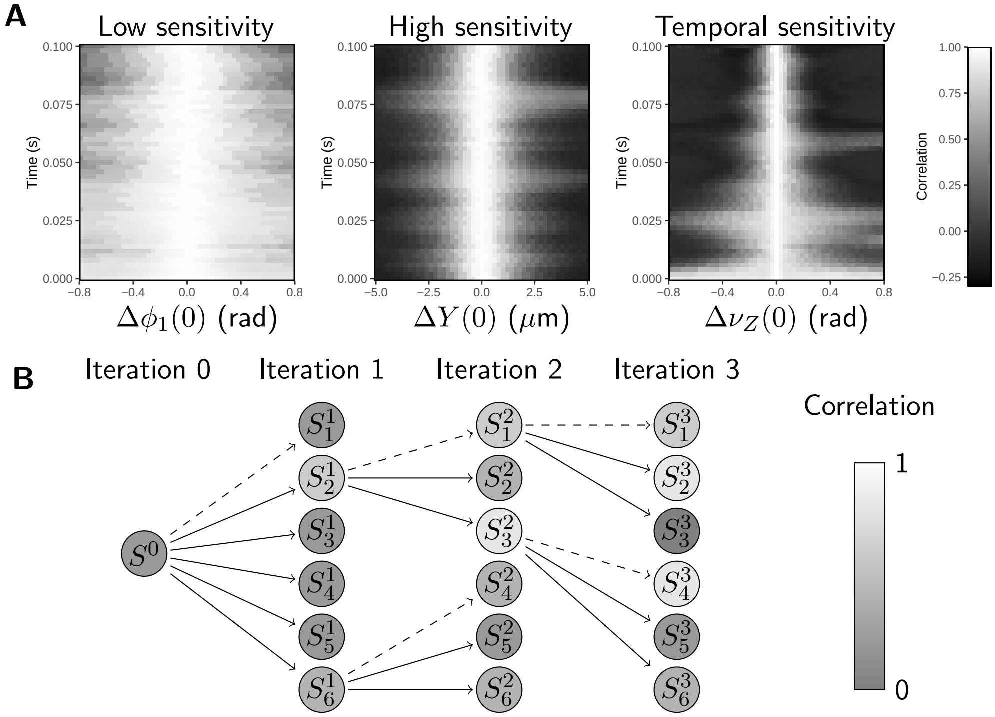


Figure 2.2: Parameter sensitivity and optimization strategy. **A** – Examples of independent score sensitivity types to each model parameter, where Δ represents the difference between the initial values of the simulated parameter and the ground truth. **B** – Evolutionary algorithm used to fit the model to data where the parent parameter sets (*i.e.* the ones with highest Scores) are both maintained (dashed arrow) and propagated with Gaussian noise to the next iteration (solid arrows). In this example, a maximum population of six parameter sets is assumed ($N_c = 6$) and only the best two are selected at each iteration ($N_b = 2$).

Table 2.1: Chi-squared distance of the estimated parameters to the true parameters.

Species	Condition	Sample		χ^2
In silico	Whole cell	#1	Initial	7.28
			Final	0.13
		#2	Initial	0.47
			Final	0.67
		#3	Initial	1.05
			Final	0.14
	Head-only	#4	Initial	7.71
			Final	0.81
		#5	Initial	9.95
			Final	*2.43
		#6	Initial	3.23
			Final	1.50
S. purpuratus	Head-only	#8	Initial	0.37
			Final	0.66

(*) Parameter set does not describe the data (MSE distance to true parameters greater than 5 μm).

2.3.2 Precise estimation of model parameters using either *in silico* or *L. pictus* data

Contrary to the most standard approaches to image analyses where segmented shape, texture, moments and many other descriptors of both template and target objects are compared, our premiss is that we can use a mechanistic model of the biological object of interest and compare the imaging data it would generate to the imaging data obtained experimentally. It is worth emphasizing here that the mechanistic model of the cell, given a set of parameters, will generate the form and deformation dynamics of the cell independently of the experimental image data. The task is to find the parameters that lead to rendered images that match as closely as possible the experimental data. To validate this approach we need to assess the efficacy of our method to: a) correctly detect and track a spermatozoon and b) to obtain unbiased estimations of the cellular and physical parameters. To do this we resorted again to the synthetic imaging data set, as ground truth, to quantify the mean squared distance of the flagellum (MSE) and the relative error of the parameters (χ^2, eq. 2.1) of three randomly perturbed models, relatively to the ground truth parameters (fig. 2.3). We can see that the mean correlation of the best ten parameter sets increase at each iteration when searching for the shape parameters (up to iteration 20) an also when the mechanical parameters are fitted (iterations 21-40). Comparing the flagellar positions obtained with our method to the ground truth we can see the MSE decreases to submicron accuracy as the algorithm progresses.

Encouraged by the results obtained with synthetic images, we use the method with experimental images of *L. pictus* sperm. A single spermatozoon was tracked semi-automatically for comparison with our method – this was taken as our pseudo-ground truth for this imaging data. For the initial estimation of the shape parameters and flagellar beating frequency we performed Fourier analysis of the curvature along arclength in intrinsic coordinates. The initial mechanical and temporal parameters where first estimated by manual *ad hoc* fitting. After MC fitting, the results were qualitatively similar to the previous ones (fig. 2.4): the correlation increased during the fitting of the shape parameters to the initial frame while the distance to the pseudo-ground truth decreased. For the fitting of the mechanical and temporal parameters the distance to the pseudo-ground truth also decreased. The mean correlation, however, remained near the threshold throughout the temporal fitting, an indication that either there can be better parameters sets that explain the data, which could be attained with different initial parameters, or that our model is not adequate to fit spermatozoa of this species.

2.3.3 Inference of flagellar conformations by tracking only the head

Our model describes the form and deformation of the whole spermatozoon, its swimming mechanics and, once rendered into synthetic images, how it would be detected in a fluorescence microscope. Assuming that a cell is swimming by deforming and that its trajectory is a direct consequence of the deformations, there should not be many flagellar conformational changes that cause a given translation and rotation from one instant in time to the next. The translation and rotation at subsequent time-points should constrain the possible initial and subsequent conformations and also the physical parameters. Based on this line of thought we hypothesized that it should be possible to infer the flagellar conformation by fitting the model of a spermatozoon to imaging data where only the head is visible. To test this hypothesis we compared the fitting of the model to two different imaging data sets of the same cell, one in which only the head is visible and another in which the head and the flagellum are visible. We used two fluorescent markers on the same cell, one that is distributed

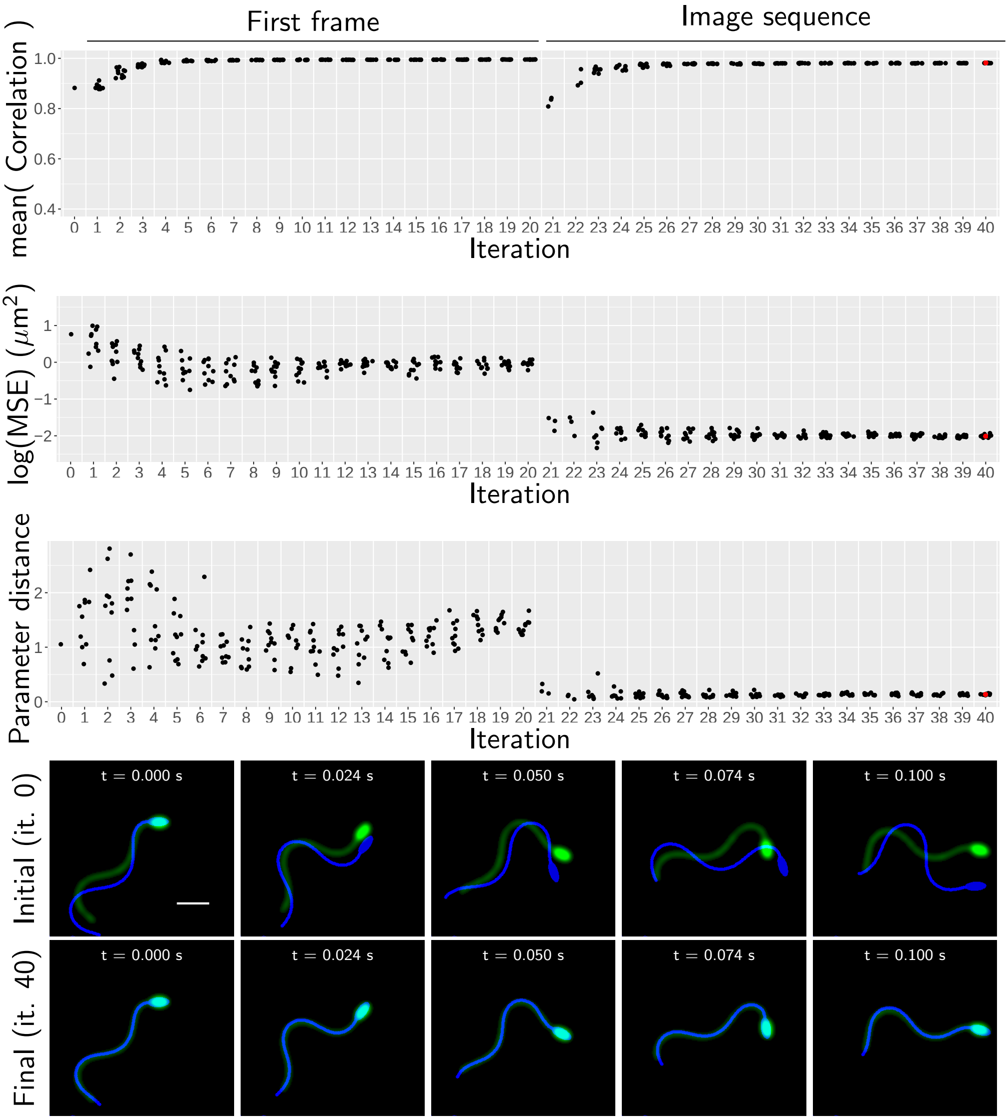


Figure 2.3: Fitting a model to *in silico* data. Points in the plots represent the cells that propagate to the next iteration. Iterations 0 (top image sequence) and 40 (red dot, bottom image sequence) of a representative run (whole cell, #3, table 2.1) are depicted for comparison. Colors of the image sequence are the same as in figure 2.1A. The green channel intensity was rescaled for better depiction of the cell. Scale bar is 10 μm.

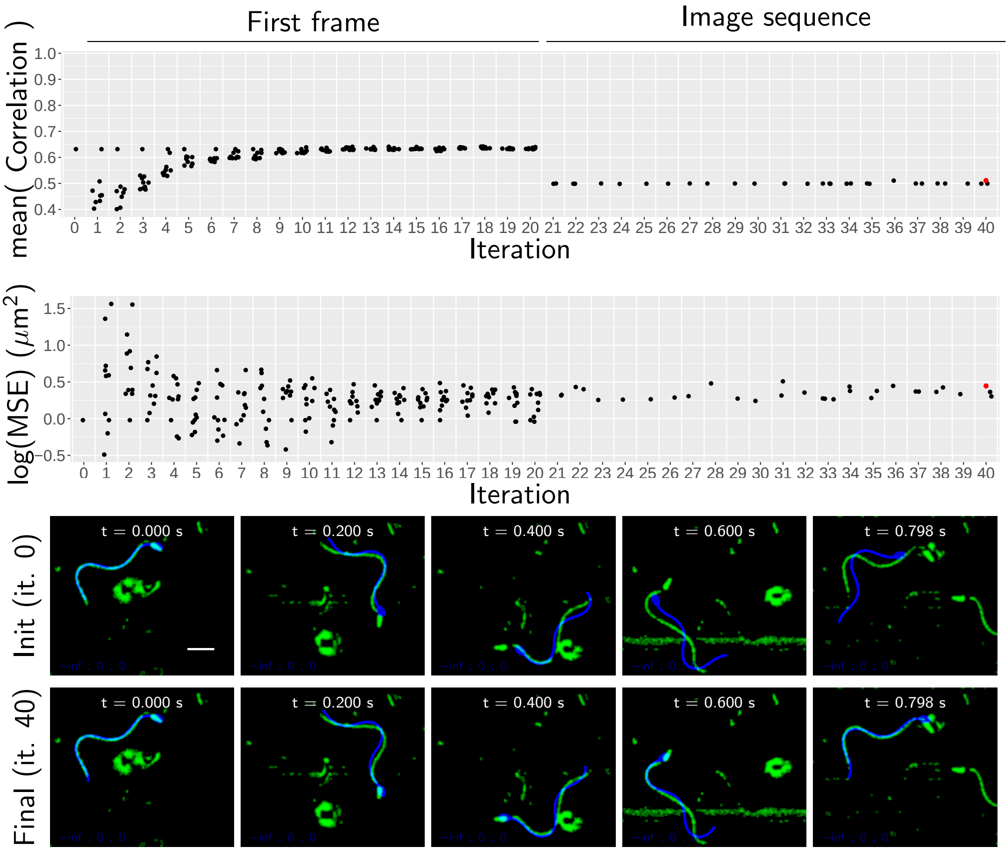


Figure 2.4: Fitting a model to *L.pictus* data. Please refer to figure 2.3 legend for details. Scale bar is 10 μm.

on the whole cell and other that marks only the head. Using a light splitter, the photons with wavelengths of the two markers were separated and captured by the same charged coupling device (CCD) camera, appearing side by side on each frame. One side that can be used to fit the model when the information on flagellar position and form is missing and the other can be used to confirm the quality of the predictions of the fitting.

This strategy was used on a preparation of *S. purpuratus* swimming confined to a plane in the absence of any stimulus. We fitted the model to the side of the image data set containing information on the whole *S. purpuratus* cell following the approach described above for *L. pictus* spermatozoon images. This fitted spermatozoon became the pseudo-ground truth for this experiment (fig. 2.5). We can see reasonable agreement between imaging data and fitted model in all imaging data set, with the exception of the earlier time-points, presumably because the targeted cell is not truly stationary. We then proceeded to fit the model to the side of the images where the information on the flagella is missing. To set unbiased initial values of the parameters we did spatial and temporal rescaling of the parameters of the reference *L. pictus*, according the flagellar size and beating frequency described for the *S. purpuratus*. Note this rescaling was enough to obtain a small parameter distance from the start ($\chi^2 \sim 0.37$). Because the flagellum is not visible in the images, we skipped the initial step of fitting flagellar conformations to the initial frames and proceeded to fit all parameters to the whole data. The final fitting to the head marker data was very similar to the one obtained with the whole cell marker, both in terms of model parameters and cellular position and conformations (fig. 2.5). The parameters sets resulting from fitting the whole cell or the head have a $\chi^2 \sim 0.66$. Although the parameter distance of the final fitting was higher than the one the inital guess, it is clearly visible the former provides better fitting.

2.4 Discussion

In this article we explored the possibility that a mechanistic model of a spermatozoon can be successfully deployed for quantitative image analyses. We have shown that one can use a biologically meaningful model to track cells and to estimate the different characteristics of the cells as detailed in the model. Given a parameter set we rendered synthetic images which can be directly compared to the experimental imaging data.

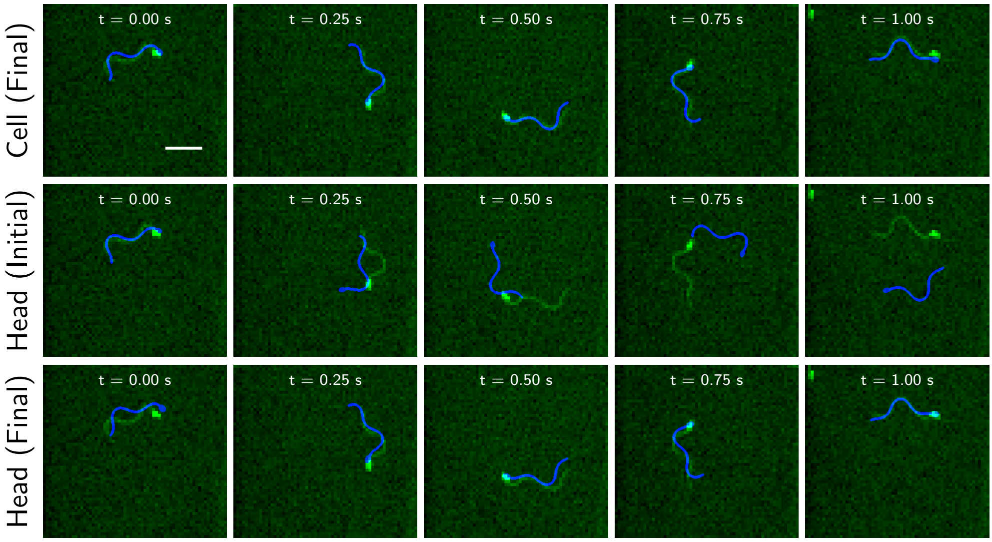


Figure 2.5: Fitting a model to noisy, low-resolution *S. purpuratus* data by fitting data where the whole spermatozoon is present (top) or by fitting to data where only the head is visible (middle and bottom). The middle sequence is the initial condition upon rescaling of the *L. pictus* mechanistic model and the top and bottom represent the best fitted models. Note only the imaging data with the flagellum visible is shown for better comparison of the experimental and model flagellar form and deformation. Please refer to figure 2.3 legend for details. The green channel intensity was rescaled for better depiction of the cell. Scale bar is 20 μm.

Better yet, we were able to infer invisible structures and their dynamics by making good predictions of the flagellar positions and forms on series of images where only the head was visible. Although fitting parametric models to statistical data or to a set of points extracted from imaging data has been done for a long time, to our knowledge, this is the first time a mechanistic model, able to realistically reproduce the behavior of cells, was fitted directly to imaging data in order to track and measure such cells. The power of this approach is evidenced by the measurement of cellular components which are not present nor can be directly measured in the imaging data.

Visual examination of the observed and fitted sperm conformations indicates that there are small deviations. One can interpret this pessimistically as indicating that the image analysis based on model-fitting is not performing well enough. Models have a purpose and if our purpose was to estimate average parameters of the swimming trajectory or of the flagellar bending waves the quality of the fitting would be sufficient. In contrast, the fitting would not be good enough if one would be interesting in using the model predictions as a fine resolution mask to make further measurements on the image. This leads to the other, perhaps more interesting interpretation of the small discrepancies between modeled and observed conformations: the image analysis method proposed here allows to infer that model is oversimplifying the mechanics or degrees of freedom of the flagellar bending waves. As a model inference tool these results are rather promising.

There is no general and definitive conclusion on what is the best scoring criteria to compare two images, from which correlation-based and l_p norm-based are the most commonly used (Evangelidis and Psarakis, 2008). While using the sum of per pixel intensity distance of the two images can be used to calculate the likelihood of the parameters given the data directly, this measurement is more sensitive to differences in contrast and brightness levels. On the other hand, maximizing the cross-correlation is equivalent to maximizing the likelihood (Zucker, 2003) and correlation normalizes the intensity distances by the mean of each image, which increases the robustness of the method to those photometric effects.

We decided to use the sum of correlation as the criteria to optimize in the time-lapse imaging data. For computational efficiency we aborted the calculation of the correlation score for parameters sets in which the average correlation dropped below 0.5 or below half the maximal average correlation coefficient found by the optimization procedure until that iteration. This implies that at least one correlation coefficient at a given frame was below this threshold. By doing this we might have missed parameter sets which have higher correlation sum over the series of images but poor correlation in a subset of the images. As the final solutions obtained do not have any time-point with correlation coefficient below this threshold (data not shown) we are confident that the solution found is better than all the candidates solutions that were discarded.

In the optimization procedures we used fixed flagellar width, PSF standard deviation and radius of correlation area. We could have fed the optimizer with these parameters and fit them along the others. Doing so might improve the correlation coefficients obtained when comparing rendered images and experimental images. This should have higher impact when fitting a model to imaging data without pre-processing. Note however that correlation assumes the constant average background so phenomena such as uneven lighting should be addressed before fitting the model. Another possible approach would be to include such effects in the model itself at the expense of increasing the number of parameters to fit and to estimate.

To perform an unbiased test when inferring the flagellum by fitting the model to time-lapse images where only the head is visible, we used the spatial and temporal rescaling of *L. pictus* high resolution model as initial guess of the parameters to model *S. purpuratus* cell. We remained within the spirit of our premiss which is to use *a priori* information to fit a model to data. To use the mean flagellar length or a measure of the sort and the flagellar beating frequency, which can be estimated by the Fourier analysis of the head orientation, is thus acceptable. The point is that we were able to obtain similar parameters, shapes, positions and orientations when tracking the whole cell in images where on the heads of visible and in images where the flagellum was not visible.

We did not explore the effect of noise on the ability of our method to estimate the correct parameters. To test both the robustness and limit of this framework we can generate *in silico* imaging data with different signal-to-noise ratios (SNRs). However, we do not expect this to be different than from a ordinary statistics correlation where a lower bound of SNR exists from which the method cannot differentiate the cell from the background.

2.4.1 Expanding our knowledge-based model

The objective of this work was to merely do a proof-of-concept and not to develop, test or compare different algorithms for parameter initialization and optimization. For this reason we implemented a simple MC evolutionary algorithm and some *ad hoc* guesstimates were used as initial parameters. It is possible to further automate the image analysis processing with our framework and to increase both the speed of convergence and the goodness-of-fit of the final model by taking advantage of a great body of work regarding the estimation of parameters from imaging data (Cootes et al., 1998; Baker and Matthews, 2004). These methods include both the issue of the initial estimation (Wu et al., 2013), on the search method and some of them also deal with effects such as uneven illumination or background. Also, parallelization using graphics processing units for producing the model image, to correlate them with the imaging data and even for the parameter optimization are expected to increase the computational speed and will help to deploy our method in tracking multiple cells in a useful time frame.

Our model assumes that a cell has constant behavior, meaning it will swim in perfect circles with the same periodic oscillations in translation and rotation. However, the trajectory of a cell can be altered if it is perturbed by local changes in physicochemical properties of the environment (media or surface), if it bumps into other cells, or even by endogenous changes of the cellular state . Also, circular drifting of trajectories have been reported under chemoattractant effect and our model does not account for it (Böhmer et al., 2005; Friedrich and Jülicher, 2007; Guerrero et al., 2010). For all these reasons, the best possible fit to experimental data with our current model will eventually accumulate errors due to trajectory perturbations. *A priori*, there is nothing preventing us to extend the mechanistic model of the cell to include time-dependent variables that adapt the swimming behavior. The additional complexity would render the data fitting more challenging. Another alternative is to fit the same model in a piecewise manner, and choose the most parsimonious combination of time-dependent parameters sets that best fits the data. For example, a cell which changed its beating frequency from a constant value to another can be described by no less and no more than two single model parameter sets. This global model can be stated as a dynamical programming (DP) problem and its parsimony could be achieved using Akaike's information criteria (AIC) or Bayesian information criteria (BIC) as scoring criteria (see *Chapter 3* for an combined implementation of these frameworks).

Another challenge for future research is to fit the cellular model to a set of images containing several spermatozoa. This was not implemented in this work because our objective was to test a new framework where a parametric model of a cell generates the observed data. While tracking multiple spermatozoa was not essential for this seminal proof-of-concept it is important to address if the comparison criteria chosen (*i.e.* correlation coefficient) is suitable to track different spermatozoa in more complex scenarios. A highly correlated model image suggests most pixels are explained by that number of cells at those positions, orientations, shapes and with those physical properties. Using our current parameter optimization we envisage that choosing the correct right number of cells can be a challenge. This is a problem similar to the finding the appropriate number of clusters in cluster analysis, and strategies devised for unsupervised cluster analysis may serve as inspiration. This notwithstanding, there are some specifics to the problem at hand that are worth mentioning. While the need for one more cell can be assessed by different parsimony criteria (*e.g.* AIC or BIC), including one more cell with the wrong parameters can decrease either the correlation coefficient or the parsimony criteria. To avoid this, the new added cell should be allowed to converge before calculating its contribution to the global picture. Similarly to the case of a cell which changes its behavior we can use DP to choose between models with different number of cells at different time-points.

There is no limit to the biological detail that can be introduced in a morphodynamical model of a cell. We can apply it to others cells by changing the shape module accordingly and we can further expand it to *e.g.* grow according to cell cycle rules or to use a different mechanics module for *e.g.* locomotion by filopodia or even the signaling module for *e.g.* regulation of immune cells. In fact the mathematical and computational biology field is plenty of models of the cell capable of generating meaningful behaviors. In the case of sperm cells, the mechanics module can be implemented as SBT, which models the long range hydrodynamic forces and how these affect the flagellar shape, or to finite elements (Yang et al., 2008), which allow to model confinement and direct interaction with other sperm. The shape function can also be modified. One can also increase the complexity of the shape module by modeling the axoneme directly (Riedel-Kruse et al., 2007). Finally, the signaling cascades from receptor binding to internal calcium (II) concentration ($[\mathrm{Ca}^{2+}]_i$) have been modeled (Espinal et al., 2011) and we know that the flagellar curvature can be estimated by the derivative of $[Ca^{2+}]_i$ (Alvarez et al., 2012). This aspect can be of utmost value because spermatozoa agglomerate at the center of the chemoattractant gradient and it becomes extremely difficult for either computer or human to track cells in these conditions. Using an adequate model for chemotaxis should allow our method to track cells even in these conditions and to distinguish how different treatments or conditions are affecting the cells. Note that extending the cell model can increase its number of parameters and their estimation from the imaging data might be more difficult. The difficulties can be mitigated by noticing that once a mechanistic model is successfully fitted to time-lapse imaging we have more than just tracked the cells, we have inferred all the biological meaningful parameters contained in the model.

2.5 Conclusion

In this chapter, we were able to prove a new concept: we generated a morphodynamical model, based on biological and mathematical principles, and fitted it directly to timelapsed two dimensional imaging data in order to detect and track spermatozoa. Using the *a priori* knowledge of a sperm morphology, flagellar bending waves and the physics of their motility, we could expand the capabilities of the microscope by being able to estimate the position of the flagellum in imaging data where this structure was not labeled and therefore not visible. This feature of our approach can be generally applied to other systems, provided the appropriate model is developed, and will allow to estimate what was is currently difficult to be measured directly. Overall, this technological achievement opens the door to new avenues for imaging analysis in general.

References

Alvarez, L., Dai, L., Friedrich, B. M., Kashikar, N. D., Gregor, I., Pascal, R., and Kaupp, U. B. (2012). "The rate of change in Ca 2+ concentration controls sperm chemotaxis". *Journal of Cell Biology* 196.(5), pp. 653–663.

Baba, S. A. and Mogami, Y. (1985). "An approach to digital image analysis of bending shapes of eukaryotic flagella and cilia". *Cell Motility* 5.(6), pp. 475–489.

Baker, S. and Matthews, I. (2004). "Lucas-Kanade 20 Years On: A Unifying Framework". *International Journal of Computer Vision* 56.(3), pp. 221–255.

Böhmer, M. et al. (2005). "Ca2+ spikes in the flagellum control chemotactic behavior of sperm." *The EMBO journal* 24.(15), pp. 2741–2752.

Brokaw, C. J. (1984). "Automated methods for estimation of sperm flagellar bending parameters." *Cell motility* 4.(6), pp. 417–430.

Cao, D., Liu, D., and Wang, C. H.-T. (2006). "Three-dimensional nonlinear dynamics of slender structures: Cosserat rod element approach". *International Journal of Solids and Structures* 43.(3-4), pp. 760–783.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (1998). "Active Appearance Models". *Proceedings of European Conference on Computer Visison* 2, pp. 484–498.

Espinal, J., Aldana, M., Guerrero, A., Wood, C. D., Darszon, A., and Martínez-Mekler, G. (2011). "Discrete dynamics model for the speract-activated Ca2+ signaling network relevant to sperm motility." *PloS one* 6.(8), e22619.

Evangelidis, G. D. and Psarakis, E. Z. (2008). "Parametric Image Alignment Using Enhanced Correlation Coefficient Maximization". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.(10), pp. 1–8.

Friedrich, B. M. and Jülicher, F. (2007). "Chemotaxis of sperm cells." *Proceedings of the National Academy of Sciences of the United States of America* 104.(33), pp. 13256–61.

Friedrich, B. M., Riedel-Kruse, I. H., Howard, J., and Jülicher, F. (2010). "High-precision tracking of sperm swimming fine structure provides strong test of resistive force theory." *The Journal of experimental biology* 213.(Pt 8), pp. 1226–34.

Gray, J. and Hancock, G. (1955). "The propulsion of sea-urchin spermatozoa". *Journal of Experimental Biology* 32.(4), pp. 802–814.

Guerrero, A., Nishigaki, T., Carneiro, J., Yoshiro Tatsu, Wood, C. D., and Darszon, A. (2010). "Tuning sperm chemotaxis by calcium burst timing." *Developmental Biology* 344.(1), pp. 52–65.

Jikeli, J. F., Alvarez, L., Friedrich, B. M., Wilson, L. G., Pascal, R., Colin, R., Pichlo, M., Rennhack, A., Brenker, C., and Kaupp, U. B. (2015). "Sperm navigation along helical paths in 3D chemoattractant landscapes." *Nature Communications* 6, pp. 1–10.

Johnson, R. E. and Brokaw, C. J. (1979). "Flagellar hydrodynamics. A comparison between resistive-force theory and slender-body theory". *Biophysical Journal* 25.(1), pp. 113–127.

Kervrann, C., Sorzano, C., Acton, S. T., Olivo-Marin, J.-C., and Unser, M. (2016). "A guided tour of selected image processing and analysis methods for fluorescence and electron microscopy". *IEEE Journal of Selected Topics in Signal Processing* 10.(1), pp. 6–30.

Perrin, F. (1936). "Mouvement Brownien d'un ellipsoide (II). Rotation libre et dépolarisation des fluorescences. Translation et diffusion de molécules ellipsoidales". *Le Journal de Physique et le Radium* Série VII.(1), pp. 1–11.

Riedel-Kruse, I. H., Hilfinger, A., Howard, J., and Jülicher, F. (2007). "How molecular motors shape the flagellar beat." *HFSP journal* 1.(3), pp. 192–208.

Su, T.-W., Choi, I., Feng, J., Huang, K., McLeod, E., and Ozcan, A. (2013). “Sperm trajectories form chiral ribbons.” *Scientific reports* 3.(1664).

Su, T.-W., Xue, L., and Ozcan, A. (2012). “High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories”. *Proceedings of the National Academy of Sciences* 109.(40), pp. 3–7.

Tange, O. (2011). “GNU Parallel: The Command-Line Power Tool”. *;login:* 36.(1), pp. 42–47.

Vulović, M., Ravelli, R. B. G., Vliet, L. J. van, Koster, A. J., Lazić, I., Lücken, U., Rullgård, H., Öktem, O., and Rieger, B. (2013). “Image formation modeling in cryo-electron microscopy”. *Journal of Structural Biology* 183.(1), pp. 19–32.

Wu, Y., Lim, J., and Yang, M.-h. (2013). “Online Object Tracking: A Benchmark”. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2411–2418.

Yang, Y., Elgeti, J., and Gompper, G. (2008). “Cooperation of sperm in two dimensions: Synchronization, attraction, and aggregation through hydrodynamic interactions”. *Physical Review E* 78.(6), pp. 1–9.

Zucker, S. (2003). “Cross-correlation and maximum-likelihood analysis: A new approach to combining cross-correlation functions”. *Monthly Notices of the Royal Astronomical Society* 342.(4), pp. 1291–1298.

Chapter 3

Comparative study of sea urchin sperm motility – confined and free swimming

Pedro Ângelo Silva[1], Arturo Pimentel[2], Tatiana Luna[2], Gabriel Corkidi[2], Alberto Darszon[2] and Jorge Carneiro[1].

[1] Instituto Gulbenkian de Ciência, Portugal; [2] Instituto de Biotecnología, Universidad Nacional Autónoma de México, Estados Unidos de México

Abstract

How sea urchin spermatozoa swim and orient in space to find their conspecific egg during broadcast spawning events is a fascinating problem. Most of our current understanding of this problem was derived from quantitative imaging studies of spermatozoa of different species swimming confined to the liquid-solid boundary plane. Whether these quantitative studies in the plane can be extrapolated to the natural free swimming in three dimensions is an open question that has motivated the development of special imaging systems able to track these fast swimming cells in a volume. Several studies have reported the three dimensional trajectories of some species of sperm but a comparative analysis across species and swimming environments is hitherto lacking. The objective of this article is to provide a

quantitative comparison of *L. pictus* and *S. purpuratus* spermatozoa swimming trajectories when confined to a plane and when swimming freely in a volume.

We adopted the three-dimensional imaging system proposed by Corkidi and collaborators that uses a piezoelectric device to oscillate an objective at high frequencies, allowing to acquire images at different depths as a function of time. In this system the oscillations of the focal plane and the image acquisition are independent and concurrent leading to time-stamped bidimensional images (2D+t) from which the vertical $Z(t)$ coordinate must be inferred before cells can be tracked. Earlier approaches to this inference problem proved to be inaccurate leading to occasional systematic phase shift of the estimated $Z(t)$ function that compromised the reconstitution of the sperm trajectories. We report here a full revision of the method to derive sperm trajectories from the 2D+t data. The first methodological result is the development of a new algorithm able to generate accurate $Z(t)$ from the image correlogram. The second methodological contribution is the use of the defocused patterns of the cell to obtain Z-coordinates of cells that have a precision that is higher than the nominal precision associated with the image resolution and acquisition rate. The third result is the introduction of piecewise helical trajectory fitting to reconstitute the trajectories of spermatozoa and identify changes in swimming path.

Using these algorithms we reconstituted the trajectories of spermatozoa of the two species of sea urchin and compared these trajectories with those observed in planar swimming. We identified a trend in the way spermatozoa alter their swimming path when interacting with a surface. We found that, during free swimming, the progressive speed of *L. pictus* sperm is higher than that *S. purpuratus* but the latter has higher trajectory torsion. Also, the trajectory curvature of *S. purpuratus* sperm is reduced when confined to the surface but not for those of *L. pictus*. Intrigued by this finding we modeled the flagellar morphodynamics and sperm trajectories using Resistive Force Theory in three dimensions. By fitting the model to data in free and confined swimming we identified the potential causes of the observed differences. The species-specific sensitivity to the planar confinement is likely related to the viscosity at the liquid-solid interface, suggesting that the smaller *S purpuratus* sperm cells swim closer to the interface that those of *L. pictus* and therefore face an effectively more structured and viscous environment.

Acknowledgments

Jorge Carneiro and Pedro Ângelo Silva designed the research. Pedro Ângelo Silva performed the work and Jorge Carneiro supervised. Tatiana Luna, Arturo Pimentel and Gabriel Corkidi provided the 2D+$Z(t)$ imaging data. Adán Guerrero provided the 2D+t imaging data. Arturo Pimentel provided the canonical piezoelectric function and the diffraction patterned cell template; Pedro Ângelo Silva, Alberto Darszon and Jorge Carneiro wrote the manuscript. All authors read and contributed to this chapter. Fundação para a Ciência e Tecnologia, Portugal, (SFRH/BD/79261/2011), Instituto Gulbenkian de Ciência, Portugal, and Instituto de Biotecnología, UNAM, Mexico, provided funding for this work.

The work developed in this chapter is the object of a manuscript in preparation, to be submitted to an international peer-reviewed journal.

3.1 Introduction

Sperm are essential for sexual reproduction of multicellular organisms and thus one of the most important cells in their life cycle. In many animal species, there will be no progeny for an individual whose spermatozoa are unable to either locate, reach or fertilize a conspecific egg. Most of our current knowledge on sperm motility comes from sea urchin models confined to planar swimming (Guerrero et al., 2011). However, the knowledge obtained from such convenient setup might not apply to the real three-dimensional (3D) world where these cells swim to find and fertilize their conspecific egg. To deal with this problem, efforts have been made to image these fast cells when swimming freely in 3D (Crenshaw, 1991; Corkidi et al., 2008; Su et al., 2012). However, a full analysis comparing 2D and 3D swimming patterns is lacking, as well as more interesting interspecies comparisons between different sea urchin species.

Although many studies tried to establish the differences between confined and free swimming spermatozoa (Cosson et al., 2003; Riedel-Kruse et al., 2007; Friedrich et al., 2010), these studies were mostly based on models of the three-dimensional behavior that were not directly compared to experimental free swimming data. Several studies describe experimental 3D trajectories of sea urchin (Crenshaw et al., 2000; Corkidi et al., 2008; Jikeli et al., 2015), human (Su et al., 2012) and horse (Su et al., 2013) spermatozoa have been reported but an exhaustive comparison to those obtained when the sperm cells are confined was lacking. Furthermore, both theoretical and experimental studies suggested differences in spermatozoa swimming close or far from a surface boundary (Corkidi et al., 2008; Guerrero et al., 2010; Jikeli et al., 2015). Here we present the first comparative study of two different sea urchin species – *Lytechinus pictus* and *Strongylocentrotus purpuratus* – in both free and confined swimming.

An algorithm to detect and track spermatozoan cells in 3D using the microscopy setup developed by Corkidi et al., (2008) (Pimentel and Corkidi, 2010; Pimentel et al., 2012) has been described. However, the resulting trajectories have non-negligible noise in the Z component that made the estimation of trajectory parameters virtually impossible based on standard filtering and smoothing methods (e.g. smoothing and regression splines, discrete geometry (Crenshaw et al., 2000; An et al., 2011)). Some methods allow estimation of such parameters by fitting helical segments to data with only two dimensions (Gurarie et al., 2011; Cherin et al., 2014), notwithstanding, these cannot be applied to our data as we do not meet the minimum requirements to apply them (*e.g.* our data has different noise in the Z component, compared to the other two dimensions). Thus, it is necessary not only to improve both detection accuracy and precision, but also to develop a method able to estimate the trajectory parameters out of our data. Here we describe a new and improved method to detect and track cells using the referred microscopy setup, which is a full revision and improvement of all steps necessary to transit from 2D+t imaging data to 3D+t tracking of individual cells, and trajectory characterization.

To gain insight into the flagellar beating differences between constrained (i.e. confined to planar swimming by a planar boundary) and unconstrained swimming, we followed an approach similar to the one used by Friedrich et al., (2010) and Jikeli et al., (2015). We applied a 3D version of Resistive Force Theory (Gray and Hancock, 1955) to model a flagellum attached to an elliptical head swimming either confined to a planar boundary or freely in three dimensions (3D).

In this study, we were able to improve the method to detect and track spermatozoa in 3D+t from 2D+t data, to develop a new method based in Dynamical Programming and Bayesian Information Criteria to fit helical segments to data with asymmetric noise in order to assess the differences in swimming behavior of two sea urchin species and to bridge the confined and free swimming models. The mathematical model allowed us propose that *S. purpuratus* sperm have significantly different flagellar beating when they confine, while the others from *L. pictus* do not.

3.2 Materials and Methods

3.2.1 Biological materials and image acquisition

Undiluted spermatozoa from *S. purpuratus* or *L. pictus* (Marinus Inc., Long Beach, CA, USA; Pamanes S. A. de C.V., Ensenada, Mexico) were obtained by intracoelomic injection of 0.5 M KCl and stored on ice until used within a day. Artificial seawater (ASW) was 950-1000 mOsm and contained (mM): 486 NaCl, 10 KCl, 10 $CaCl_2$, 26 $MgCl_2$, 30 $MgSO_4$, 2.5 $NaHCO_3$, 10 HEPES and 1 EDTA (pH 8.0) for *S. purpuratus*. To reduce the number of spontaneous acrosome reaction, *L. pictus* spermatozoa were suspended in slightly acidified artificial sea water (pH 7.4). [Ser5; nitrobenzyl-Gly6]speract, referred to throughout the text as "caged speract", was prepared as previously described (Tatsu et al., 2002). PolyHEME (poly(2-53 hydroxyethyl methacrylate)) and other reagents, unless indicated, were from Sigma-Aldrich (Toluca, Edo de Mexico, Mexico). A dilution of cells was prepared with ASW stored on ice and 2μl were added to 200 μl of ASW at 14°C in a imaging camera previously coated with PolyHEME.

Acquisition of 2D+$Z(t)$ imaging data of both sea urchin species were obtained using the microscopy setup developed by Corkidi et al., (2008), using the settings defined in Pimentel et al., (2012). Briefly, a piezoelectric device coupled to the objective (Olympus 40X/0.60 N.A., Olympus America Inc. U.S.A.) installed in an inverted optical microscope Olympus IX71, oscillating at 30 Hz with $\sim$250 μm amplitude was used. Framerate acquisition was at 2000 Hz with a high-speed camera (Optronics CR5000x2, Optronics GmbH, Germany). Images were pre-processed by subtracting the background (average intensity of the first 100 frames), transforming to 8-bit and horizontal line filtering by applying a horizontal low-pass filter of 95%.

3.2.2 3D trajectories of free swimming spermatozoa

To characterize the trajectories of free swimming spermatozoa we processed the 2D+t imaging data as illustrated on figure 3.2. Briefly, the data are a series of time-stamped bidimensional image frames $\boldsymbol{I}_{t_i}$ (with $i \in \{1, ..., f\}$) captured by a high speed camera while the piezoelectric oscillations scan a volume in an independent and concurrent manner. Since the microscope system does not return the Z position at which each imaging frame was acquired this needs to be inferred from the image data itself. Once the function $Z(t_i)$ is reconstructed one has to identify and measure the spatial and temporal coordinates of the centroid of the head of each sperm cells index s. This implies estimating the positions of the centroids $\{X, Y, Z\}_s[t_i]$ of each cell s at each time t_i and then tracking the cell during the period of the analysis. The methods used in each of these steps are described in the following sections. They were inspired on the previously proposed methods Pimentel et al., (2012) but have been fully revised at each step.

Inference of the depth function $\mathbf{Z(t)}$

To accurately estimate the $Z(t_i)$ of each frame i, we built on the seminal idea by Pimentel et al., (2012). Sperm cells have an average path velocity of 250 μm/s and the frequency of the piezoelectric and the high acquisition frame rate ensure sperm cells are not significantly displaced between two consecutive frames at the same Z

position, *i.e.*, within a piezoelectric period debris are virtually static and cells do not displace more than two times the average head size. Under these conditions, frames at the same Z position should be highly correlated whereas frames with different Z should not, and therefore the Z positions can be inferred by correlation. This notwithstanding, there is a threefold challenge: first, neither the position $Z(t_i)$ nor the velocity are produced by the microscopy setup; second, the piezoelectric moves non-linearly due to the mass of the objective such that consecutive images are not at the same distance in the Z-dimension; and third, the vibrations of the piezoelectric produces horizontal wobbling of the objective translating the referential for the (X, Y) coordinates that can spoil the expected image correlations.

The new algorithm to reconstitute the function $Z(t_i)$ implements a strategy to overcome these challenges. First, we constructed a correlogram in which the correlation coefficient $\rho_{i,i+k}$ is the maximum of the cross-correlation between the normalized ith image frame and a normalized inset of the $(i+k)$th image, computed in Fourier space. Using the inset of 7.8μm (or 10 pixels) allowed to accommodate and correct for the translations in the horizontal plane introduced by the small but not negligible vibrations of the piezoelectric. The bidimensional correlogram obtained shows a conspicuous periodic structure (fig. 3.2B). The next step involves realizing that if at some instant t the focal plane is at the minimum or maximum Z position then the images acquired before and after a fixed time lapse are a the same depth and therefore should be highly correlated (if the function $Z(t)$ is approximately symmetric). Based on this realization we defined a new image index $m = i + k/2$ such that the local maxima of the function $\rho_m =$ Average$(\rho_{i,i+k})$ correspond to the extremes of the vertical positions of the focal plane. The dots following slanted lines on top of the correlogram figure 3.2B, correspond to the indicated m values. As m increases the slanted line slides to the right and the average correlation under the line is maximal at the extremes. The figure also explains the choice of $k \in \{3, ..., 64\}$ as the expected period of 30 Hz corresponds to about 66.7 frames. The average ρ_m is plotted in figure 3.2C as a function of m. Since by construction the first extreme is a minimum then the series of minima and maxima is obtained in straightforward way. Finally, to reconstruct the function $Z(t)$ a linear rescaling of each half of characteristic piezoelectric periodic function (Pimentel and Corkidi, 2009) was applied in time to fit the series of extremes in a piecewise manner. An example of the reconstructed function $Z(t)$ is depicted in figure 3.2D.

Cell detection and tracking

The previously proposed algorithm (Pimentel et al., 2012) segmented regions containing concentric rings of dark and light intensity, characteristic of the out-of-focus patterns of the sperm cells, selected those in which the pattern appeared inverted in consecutive neighboring image frames and averaged the Z position. It turned out that this method yields Z estimates of the sperm head centroids with significant uncertainty. To improve the performance of the method we explored quantitatively the relationship between the diffraction patterns characteristic of the out-of-focus cells and the offset between the plane of the sperm cell and the focal plane (fig. 3.2E). Note that the asymmetry of the patterns when the focal plane of the objective is imaging above and below the focal plane of the cell (a feature that was incorrectly misinterpreted by Pimentel et al., (2012)). Averaging over hundreds of images sperm heads, we constructed an ordered template set, $(\boldsymbol{P}_k)$ with $k = 1, ..., 9$, where patterns are ordered according to the offset from the vertical position of the focus plane of the cell to the focal plane of the objective (fig. 3.2E). It is important to emphasize that these patterns are ordered but the exact offset values are unknown. Each pattern k in the template set is used to perform cross-correlation with every image frame i in the downward movement of the piezoelectric (to reduce the amount of data to be processed), and the detection events are collected every time the coefficient is higher than 0.7. In this way, one obtains a collection of detected objects $\{X, Y, Z(t_i), k, r\}$, where X and Y are the coordinates of the center pattern on the image frame, $Z(t_i)$ is the vertical coordinate or depth of the frame, k is the index of the pattern in the ordered template set and r is the correlation coefficient between the template and the frame. The objects are clustered in space (X,Y and Z) and time (t) by a friends of friends algorithm using a link-threshold of 5μm and 1.5 ms. The number of objects in each cluster is reduced by retaining the object with highest correlation coefficient r for each (t_i, k) pair and eliminating the remaining ones. Associated to each sperm head we obtain a set of up to 9 ordered objects indexed k corresponding to the highest correlation for each template (cells moving close to the limits of the imaged volume, inside or outside, will be represented by less than 9 objects). We call this the diffraction set of a cell (fig. 3.2E). After this selection, we regressed the Z coordinate over the template index k with weights $w = (r - 0.7)/0.3$ within each diffraction set of a cell. The regression slope is averaged over all the diffraction sets obtaining an average slope β. We recalculate the regression of Z

over k imposing the average slope β across the diffraction set and the Z-coordinate corresponding to $Z(k = 5)$ is obtained and taken as the vertical coordinate of the cell. The bidimensional coordinate (X, Y) is obtained by weighted average with the weights w, and t is obtained by linear interpolation. It is important to emphasize that the coordinates of the cell $\{X, Y, Z, t\}$ will have higher precision than nominal precision associated with image resolution and temporal acquisition rate.

Tracking of the individual sperm heads in the whole data set (fig. 3.2H) was performed using two clustering steps. In the first step, we clustered the points by normalized $\{X, Y, Z, t\}$ Euclidean distances with the friends-of-friends agglomeration method, using a normalized maximal distance of 0.238. To resolve distinct cells that happen to cross near each other, we applied space-wise friends-of-friends clustering (i.e. agglomerating points by increasing spatial distance) ensuring a maximal spatial (X,Y and Z) euclidean distance of 25 μm between all cluster members if their temporal distance was equal to or below 0.1 s. This method breaks the trajectories of two crossing cells into exactly five different clusters. The region where two cells meet in space and time were discarded and the clusters belonging to the same cell were determined as such by clustering in normalized trajectories parameter space (see below). We now have collections of distinct sets of points, each corresponding to time series of the centroid of the head of distinct spermatozoa.

Piecewise helix fitting

The final step in tracking the spermatozoa in the imaging data set was to estimate the individual trajectories and their parameters. To this end, we performed piecewise helical segment fitting to the sets of points obtained in the previous step using Dynamical Programming and Bayesian information criteria (BIC) (fig. 3.2I). This method chooses the most parsimonious model that best fits the data, meaning the minimal number of helical paths necessary to approximate the points in a trajectory. For single helix fitting, the coordinate-based Point Distance Method (Liu and Wang, 2008) was applied to measure the distance between model and data points. Note the set of helices which describe a trajectory are non-continuous in space. Also, our method allows for different weights for different dimensions. In this case, we used weight 1 for either X and Y and 0.2 for Z axis, to account for the higher uncertainty in the Z estimates.

Consider a parametric curve $\boldsymbol{r}(\boldsymbol{p}, t)$ defined by its parameters $\boldsymbol{p} = \{p_1, ..., p_m\}$,

$m \in \mathbb{N}$ and $p, t \in \mathbb{R}$. We will consider a curve in Euclidean $\mathbb{R}^3$ space. In this study, every curve is an helix described as follows:

$$\boldsymbol{r}(\boldsymbol{p}, t) = \begin{bmatrix} X(t) \\ X(t) \\ Z(t) \end{bmatrix} = \boldsymbol{R}_X \boldsymbol{R}_Y \boldsymbol{R}_Z \begin{bmatrix} r\cos(2\pi t/T) \\ r\sin(2\pi t/T) \\ \|\boldsymbol{a}\| t \end{bmatrix} + \boldsymbol{c}_1 \tag{3.1}$$

where the parameter vector is $\boldsymbol{p} = \{\|\boldsymbol{a}\|, r, T, \nu_X,\ \nu_Y, \nu_Z, c_X, c_Y, c_Z\}^T$, $\boldsymbol{R}_i$ are reference frame rotation matrices on i axis, r is the helix radius, T is the helix revolution period and $\|\boldsymbol{a}\|$ is the progressive speed (i.e. helix pitch over revolution period). The rotation matrices are also a function of a subset the parameter vector $\boldsymbol{p}$:

$$\boldsymbol{R}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\nu_X) & \sin(\nu_X) \\ 0 & -\sin(\nu_X) & \cos(\nu_X) \end{bmatrix}; \boldsymbol{R}_Y = \begin{bmatrix} \cos(\nu_Y) & 0 & -\sin(\nu_Y) \\ 0 & 1 & 0 \\ \sin(\nu_Y) & 0 & \cos(\nu_Y) \end{bmatrix}; \boldsymbol{R}_Z = \begin{bmatrix} \cos(\nu_Z) & \sin(\nu_Z) & 0 \\ -\sin(\nu_Z) & \cos(\nu_Z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and the translation vector is $\boldsymbol{c}_1 = \{c_X, c_Y, c_Z\}^T$.

Consider the experimental data column vector $\dot{\boldsymbol{e}}^T = \{\boldsymbol{e}_i\}_{i=1}^n$, where $\boldsymbol{e}_i$ is the position defined experimentally at time t_i ($t \in \{t_1, \ldots, t_n\}$), $n \in \mathbb{N}$. The sum of squared errors distance (coordinate-based (Ahn, 2004)) between experimental and modeled state is given by the sum of squared errors

$$SSE = (\dot{\boldsymbol{m}} - \dot{\boldsymbol{e}})^T \boldsymbol{W}^T \boldsymbol{W} (\dot{\boldsymbol{m}} - \dot{\boldsymbol{e}}) \tag{3.2}$$

where $\dot{\boldsymbol{m}}^T = \{\boldsymbol{r}(t_i)\}_{i=1}^n$ and $\boldsymbol{W} = diag(\{\boldsymbol{w}\}_{i=1}^n)$ is a weight (non-singular) matrix defined by $\boldsymbol{w} = \{w_1, \ldots, w_m\}$. As this error measure depends on the model and these depend on the parameters used, the error also depends on the model parameters ($SSE(\boldsymbol{p}(t))$).

Suppose you have experimental data $\dot{\boldsymbol{e}}^T$ as defined above and that $\boldsymbol{p}(t)$ is constant ($\boldsymbol{p}$). It is possible to calculate which parameters of your model best fit the data. Using equation 3.2 as our distance measure between model and experimental data, we want to find which parameters minimize this distance:

$$\underset{\boldsymbol{p}}{\operatorname{argmin}}\, SSE(\boldsymbol{p}) \tag{3.3}$$

By using initial conditions of $\boldsymbol{p}$ close to the solution we can (hopefully) find the global minima using hill-climbing algorithms like the quasi-newton methods. In particular we applied the `optim` function with argument `method="BFGS"` in R statistical software.

The initial conditions were set as follows, assuming the data defines a helix with several revolutions and radius smaller than the helix height. We started by estimating the timescale of the line $\hat{a} = \|\boldsymbol{r}_n - \boldsymbol{r}_1\|/\Delta t_{1,n}$. Principal Component Analysis gives us the principal axis vector of the helix ($\boldsymbol{z'}$) and the center of mass of the point cloud ($\boldsymbol{c}_c$). Then, $\boldsymbol{c}_1$ is estimated by the displacement between the closest point to the center of mass ($\boldsymbol{r}_c$) along the principal axis: $\hat{\boldsymbol{c}}_1 = \boldsymbol{c}_c - t_c \hat{a} \boldsymbol{z'}$. The revolution period T can be estimated as the low frequency ($1/\hat{T}$ less than 10% of the calculated Nyquist frequency) with higher magnitude of the angles between each sampled point and $PC2$ as a function of time ($\angle_t(\boldsymbol{r}_t - \boldsymbol{c}_c, PC2)$). $PC2$ gives us a radial vector (e.g. passes through the center of the circle) and the radius can be estimated using three points as in Coeurjolly and Svensson, (2003): let $d_{i,j} = \|\boldsymbol{r}_{t_j} - \boldsymbol{r}_{t_i}\|$ and $s = (d_{i-k,i} d_{i-k,i+k} d_{i,i+k})/2$ then $\hat{r} = s/(2\text{Area}(\triangle_{\boldsymbol{r}_{t_{i-k}}, \boldsymbol{r}_{t_i}, \boldsymbol{r}_{t_{i+k}}}))$. In this work, i was chosen to be the closest point to $\boldsymbol{c}_c$ ($\boldsymbol{r}_c$) and k is a random point under the restriction $|t_{i \pm k} - t_i| \leq \hat{T}/4$. If we solve $\boldsymbol{z'} = \boldsymbol{R}_X(\nu_X)\boldsymbol{R}_Y(\nu_Y)\boldsymbol{R}_Z(0)\{0, 0, 1\}^T$ we obtain $\nu_Y = \arcsin(-z'_X)$ and $\nu_X = \arctan\{(-z'_Y/\cos(\nu_Y))/(z'_Z/\cos(\nu_Y))\}$. Note we should calculate ν_X taking the quadrants into account, so we must use the function *atan2*. Also note that Euler angles systems (i.e. the rotation system used here) have two solutions and here we consider the one where $-\pi/2 \leq \nu_Y \leq \pi/2$. Finally, we estimate ν_Z as the angle between $\boldsymbol{r}(t_i, \hat{\boldsymbol{p}})$ and $\boldsymbol{r}_i$.

Consider the possible time uniform knot set $\Omega = \{\dot{t}_1, \ldots, \dot{t}_o\}$, $\dot{t}_i = t_1 + (t_n - t_1)(i-1)/(o-1)$, $i \leq o$, $o > 1$ and $i, o \in \mathbb{N}$, we will determine the subset $\dot{\Omega} \subseteq \Omega$, $\dot{t}_1, \dot{t}_o \in \dot{\Omega}$ that minimizes the BIC. This is needed to make a balance between number of parameters (i.e. number of helices fitted) and the error of that fit. In detail, we want to find $\dot{\Omega}$ that minimizes the cost function $d(\dot{t}_o) = d(\dot{t}_1, \dot{t}_o)$. As we are not assuming C^0 nor G^0 continuities, the general problem have independent smaller sub-problems problems that are part of the general problem solution. Thus, we can use solve it using Dynamical Programming. Formally, the general problem $d(\dot{t}_o)$ has sub-problems

$$d(\dot{t}_j) = \underset{i,j \in \mathbb{N}}{min}(n_i + n_{i,j}) \ln\left(\frac{SSE_i + SSE_{i,j}}{n_i + n_{i,j}}\right) + (K_i + K_{i,j} + 1) \ln(n_i + n_{i,j}) \quad (3.4)$$

where $1 \leq i < j \leq o$, $n_{i,j}$ is the number of sample points where $\dot{t}_i \leq t_k \leq \dot{t}_j$,

$SSE_{i,j} = SSE(\boldsymbol{p_{i,j}})$ and $K_{i,j}$ is the number of parameters of the fitting between knots $\dot{t}_i$ and $\dot{t}_j$, respectively. Also, $n_i = n_{1,i}$ and this is similar in SSE_i and K_i. Finally, $d(\dot{t}_1) = 0$. Note the BIC formula in the equation above.

The solution to $\underset{\dot{\Omega}}{\operatorname{argmin}}\, d(\dot{t}_o)$ is obtained iteratively by orderly solving sub-problems from $d(\dot{t}_1)$ to $d(\dot{t}_o)$. We then extract the parameters vectors $\boldsymbol{p_{i,j}}$ where $\dot{t}_i, \dot{t}_j \in \Omega'$ to attain our piecewise helix fitting.

After obtaining the first fittings with this method, we fitted all trajectories with three initial conditions: the one estimated and the median parameters of either sea urchin species. Finally, cells were uniformly resampled using their respective piecewise helix model at the piezoelectric frequency.

3.2.3 2D imaging data of spermatozoa and trajectories

L. pictus and *S. purpuratus* imaging data was kindly provided by Guerrero et al., (2010). Trajectories were manually obtained using the MtrakJ plugin (Meijering et al., 2012) of ImageJ v1.4. As caged Speract was present, we only analyzed data in the 0-3 s interval, before UV irradiation in order to study sperm motility in non-chemotactic conditions.

Trajectory parameters were estimated as in the 3D case, but constraining progressive speed to zero, so we fitted circle arcs instead of helices. Note that this method calculates the velocity, curvature and torsion along the average path.

Trajectory parameters were estimated as in the 3D case, but constraining progressive velocity to zero, so we fitted circle arcs instead of helices. Note that this method calculates the velocity, curvature and torsion along the average path.

3.2.4 Data

All piecewise fitted trajectories included in this study were manually inspected and discarded if appropriate, keeping trajectories with one and only one cell with reasonable speed ($50 \leq \|v\| \leq 300$ μm.s^{-1}), curvature ($0 \leq |\kappa| \leq 1$ rad.μm^{-1}) and torsion ($0 \leq |\tau| \leq 1$ rad.μm^{-1}). Furthermore, only trajectories spanning more than one second were considered.

Data was analyzed using R statistical program v.3.0.3 (R Foundation for Statistical Computing, Vienna, Austria). Comparison of empirical cumulative distributions was performed by two-sample Kolmogorov-Smirnov while the median was compared using the Mann-Whitney test. For all tests we assumed a type I error of $\alpha = 0.05$. When stated, Bonferroni correction was applied with a α/α' factor, where is α' the type I error considered for that specific statistical test.

3.2.5 Morphodynamical model

Consider a mathematical model of a cell in which the morphology changes are defined within the model itself – this is a morphodynamical model. In our particular case, the spermatozoon cell model ($\boldsymbol{S}$) is defined by the shape ($\boldsymbol{\Sigma}$) and mechanics ($\boldsymbol{\Phi}$) modules.

Shape

We can define the morphology of a spermatozoon by defining two regions, the head and the flagellum. We assume the head is a revolution ellipsoid with half axes a and $b = c$. The flagellum is composed of N rods with total length L. The conformation of the flagellum at a particular time (t) is given by its flagellar curvature (κ) along flagellar arclength (s), a first order traveling wave: $\kappa(s,t) = K_0 + A_0 \cos(\omega_\kappa t - \lambda_\kappa s + \phi)$, which is defined by the mean flagellar curvature (K_0), basal curvature amplitude (A_0), angular velocity of beating (ω_κ), curvature wavelength (λ_κ) and phase (ϕ). We assume the flagellar torsion ($\tau(s,t) = T_0$) along arclength to be constant. The intrinsic flagellar position ($\boldsymbol{r}_f(s,t)$) is calculated by arclength integration of the Cosserat frame (Cao et al., 2006; Jikeli et al., 2015), which is defined by the orthonormal basis $\boldsymbol{e}_1(s,t)$, $\boldsymbol{e}_2(s,t)$ and $\boldsymbol{e}_3(s,t)$: $\partial \boldsymbol{r}_f(s,t)/\partial s = \boldsymbol{e}_3(s,t)$, $\partial \boldsymbol{e}_3(s,t)/\partial s = \kappa(s,t)\boldsymbol{e}_1(s,t)$, $\partial \boldsymbol{e}_1(s,t)/\partial s = -\kappa(s,t)\boldsymbol{e}_3(s,t) + \tau(s,t)\boldsymbol{e}_2(s,t)$, $\partial \boldsymbol{e}_2(s,t)/\partial s = -\tau(s,t)\boldsymbol{e}_1(s,t)$. Note $\boldsymbol{e}_3$ represents the centreline along the flagellum, $\boldsymbol{r}_f(0,t) = \{-a, 0, 0\}$, $\boldsymbol{e}_1(0,t) = \{0, -1, 0\}$, $\boldsymbol{e}_2(0,t) = \{0, 0, 1\}$ and $\boldsymbol{e}_3(0,t) = \{-1, 0, 0\}$. At a given time t, the centroid of the head has extrinsic position $\boldsymbol{R}_h(t)$ and orientation matrix $\boldsymbol{\Theta}(t)$, which rotates from extrinsic to intrinsic coordinates.

Mechanics

As the cell moves within a viscous fluid, the latter exerts drag forces which, under some conditions, propel the cell in the media. The physics behind it have been shown to be approximated by Resistive Force Theory (RFT) (Gray and Hancock, 1955; Friedrich et al., 2010; Jikeli et al., 2015), which we will use here. Briefly, the drag force density that fluid exerts on the flagellar piece is $\boldsymbol{f}(s,t) = \xi_{\|}\partial\dot{\boldsymbol{r}}_{f,\|}(s,t)/\partial t + \xi_{\perp}\partial\dot{\boldsymbol{r}}_{f,\perp}(s,t)/\partial t$, where $\dot{\boldsymbol{r}}_{f,\|}$ and $\dot{\boldsymbol{r}}_{f,\perp}$ are the tangent and normal components, respectively, of the mean flagellar piece velocity $\dot{\boldsymbol{r}}_f(s,t) = (\boldsymbol{\Psi}'(\delta\boldsymbol{\theta}(t))\cdot\boldsymbol{r}(s,t+\delta) + \delta\boldsymbol{v}(t) - \boldsymbol{r}(s,t+\delta))/\delta$ during the time interval δ. Note $\boldsymbol{v}$ and $\boldsymbol{\theta}$ are the intrinsic head translational and rotational velocities, respectively, $\boldsymbol{\Psi}'(\boldsymbol{\nu})$ is the an approximation of Rodrigues rotation formula $(\boldsymbol{\Psi}'(\boldsymbol{\nu}))$ for small angles $(\sin\beta \simeq \beta$ and $\cos\beta = 1)$ and $\xi_{\|}$ and $\xi_{\perp}$ are the tangent and normal drags exerted on the flagellum, respectively. We can obtain the intrinsic head velocities by solving the system of force and torque equilibria defined by $\boldsymbol{\xi_T}\boldsymbol{v}(t)\int_0^L \boldsymbol{f}(s,t)\partial s = 0 \wedge \boldsymbol{\xi_R}\boldsymbol{\theta}(t) - \int_0^L \boldsymbol{f}(s,t)\times\boldsymbol{r}_f(s,t)\partial s = 0$, where $\boldsymbol{\xi_T}$ and $\boldsymbol{\xi_R}$ are the translational and rotational drag coefficients of the head and the symbol $\times$ denotes the crossproduct operator. The intrinsic head velocities are transformed into extrinsic velocities by multiplication of the transposed orientation matrix and applied to the cell to calculate the new extrinsic position $\boldsymbol{R}_h(t+\delta) = \boldsymbol{R}_h(t) + \delta\boldsymbol{\Theta}(t)^T\cdot\boldsymbol{v}(t)$ and orientation $\boldsymbol{\Theta}(t+\delta) = \boldsymbol{\Psi}(\delta\boldsymbol{\Theta}(t)^T\cdot\boldsymbol{\theta}(t))\cdot\boldsymbol{\Theta}(t)$. Note we used the approximated Rodrigues rotation matrix to calculate the local flagellar velocity in order to make the system linear but this approximation is not needed when we calculate the extrinsic velocities from the intrinsic ones.

Assuming a dynamic viscosity η we can calculate the translational and rotational drags for the head based on its size using Perrin's formulas (Perrin, 1936). As we used Euler integration method, we will update the system every δ time steps and then we solve it for the intrinsic translational and rotational velocities assuming they are constant within each time period δ.For spatial integration of the flagellum, we discretized it in N segments and calculated the intrinsic midpoint position $\kappa[i,t]$, where $i = L/N$. We determined $N = 49$ and $\delta = 50$ μs to produce errors $(\sum(\text{obs} - \text{exp})/\text{exp}\times 100)$ below 5% when calculating the intrinsic translational and rotational velocities, comparing to $N = 499$ and $\delta = 10$ μs. Note the simulations were performed with N=49 in order for the exploration of the parameter space to be accomplished in feasible time.

The spermatozoon model $\boldsymbol{S}(\boldsymbol{\Sigma},\boldsymbol{\Phi})$ is thus completely defined by its shape parameters $\boldsymbol{\Sigma} = (a,b,c,L,N,\boldsymbol{R}_h,\boldsymbol{\Theta},K_0,A_0,\phi,\omega_\kappa,\lambda_\kappa,T_0)$ and its physical or mechanical parameters $\boldsymbol{\Phi} = (\boldsymbol{\xi_T},\boldsymbol{\xi_R},\xi_{\|},\xi\perp)$. Although not explicit here, some of these parameters are dependent on time, as shown above.

Table 3.1: Parameters used in the RFT model.

Parameter	Units	Lp	Sp	Ap*
a, b	μm	3.25,1.625	2.88,1.44	2.50,1.25
L	μm	44	38	41
N	–	49	49	49
η	Pa.s	0.00900	0.01115	0.01080
$\xi_{\parallel}$	pN.s.μm^{-2}	0.01427867	0.01041734	0.010692
$\xi_{\perp}$	pN.s.μm^{-2}	0.02855734	0.02073051	0.01935252
$\boldsymbol{\xi_T}$	pN.s.μm^{-1}	(0.332,0.380,0.380)	(0.364,0.417,0.417)	(0.306,0.351,0.351)
$\boldsymbol{\xi_R}$	pN.s.μm	(1.567,2.924,2.924)	(1.353,2.522,2.522)	(0.855,1.596,1.596)
s	μm	0.898	0.776	0.837
K_0	rad.μm^{-1}	0.03483757	0.04110697	0.0351
A_0	rad.μm^{-1}	0.17597653	0.20163048	0.160
A_1	rad.μm^{-1}	0	0	0
λ_κ	rad.μm^{-1}	0.1927098	0.22425032	0.2122698
ω_κ	rad.s^{-1}	180.72	195.42	273.32
T_0	rad.μm^{-1}	0	0	0.00477

Parameter description and fitted values to a confined spermatozoon of *L. pictus* (*Lp*) and *S. purpuratus* (*Sp*). (*) Free swimming parameters of *A. punctulata* (*Ap*) as reported on other work (Jikeli et al., 2015).

One of the proposed mechanisms for the confined swimming mode near the water-glass interface is that the boundary and the hydrodynamic forces therein constrain the (quasi-)planar beating waves of the flagellum to be parallel to that surface (Cosson et al., 2008). This suggested the following simplification to the modeling of the confined swimming. We projected the 3D free swimming translational velocity on the plane defined by the first two principal components of all flagellar positions within a beating period. The angular velocity was projected into the third principal component. An Euler integration with the same timestep as in the free swimming mode was used to calculate the confined position and orientation of the cell along time using the projected velocity vectors.

3.2.6 Comparing and fitting the model to trajectories

To fit the model to experimental trajectories we first estimated the flagellar shape parameters by random walk and then by gradient descent optimization using high temporal and spatial resolution imaging data of each species sperm in confined swimming (Table 3.1). The cost function to minimize was the mean flagellar distance between data and model as returned by an alignment by the Kabsch method (see R

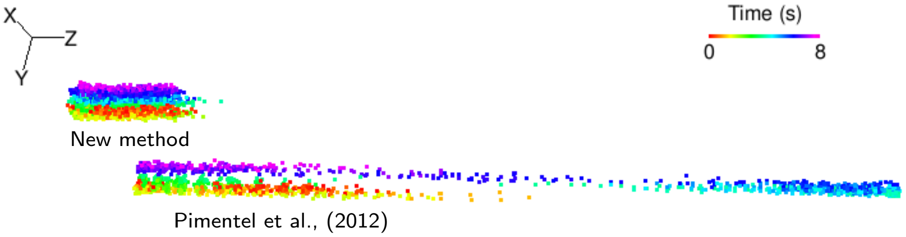


Figure 3.1: Determining the Z position of each imaging frame. The same stationary cell as detected by the new method (top) and by Pimentel and Corkidi, (2010) (bottom). Notice the different Z variance at 4-7 s. Axis length is 10 μm.

function `pracma::kabsh`). Using the resulting fitting, we searched the parameter space (effective viscosity, mean flagellar curvature, flagellar curvature amplitude and flagellar torsion) by random walk, in order to obtain the sets which generate free and confined trajectories within the first and third quartiles of the trajectory speed, curvature and torsion observed for each species. For each species and each swimming type, we randomly chose $\sim$1000 independent parameter sets which are uniformly distributed within the experimental trajectory quantiles and computed the ratios between free and confined swimming. For the test between the parameter medians between the two swimming modes a Bonferroni corrected type I error was used ($\alpha' = 0.0042$). The representative trajectories of each species were chosen by selecting the parameter set which generates the trajectory closest to the median values observed for free swimming and then by multiplying the parameters by the median parameter ratio to generate the confined swimming trajectory. These parameter ratios were calculated as all possible ratios between the sampled parameter sets of the swimming modes, for each species.

3.3 Results

3.3.1 Accurate and precise reconstitution of 3D sperm trajectories

To study sperm motility and to establish the species-specific differences between sperm swimming confined to a boundary plane or free from a boundary, we need accurate and precise trajectory data. The microscopy setup developed by Corkidi et al., (2008) was used in this study to obtain the 2D+t imaging data but the previously proposed algorithms to derive the 3D sperm trajectories from these data (Pimentel et al., 2012) proved to be insufficiently accurate and precise or were insufficiently expedite to be deployed in automatic and objective analysis. These limitations of the available algorithms motivated a systematic revision of the different steps of the imaging analysis and trajectory reconstitution algorithms.

The inference of the vertical coordinate associated of each image frame $\boldsymbol{I}_{t_i}$ is the first challenge for image analysis. The $Z(t)$ position of each frame as generated by the methods reported by Pimentel et al., (2012) did not perform reliably for some imaging data sets, often producing a phase shift of the estimated $Z(t)$ (fig. 3.1). The method was improved by producing a correlogram based on the correlation coefficients of each frame and its neighboring frames within the theoretical piezoelectric period. This correlogram shows a regular periodic structure with maxima that correspond to the instants when the piezoelectric device inverts the direction movement of the focal plane (fig. 3.2B and C). These correlation maxima occur either when the $Z(t)$ is at its upper or lower extremes. An empirical characteristic piezoelectric function (Pimentel et al., 2012) was then rescaled in time to fit the pairs of consecutive upper and lower extremes in a piecewise manner, obtaining the full series of $Z(t)$ position. This method increased the accuracy in determining the Z position of each frame, reducing some spurious variance in $Z(t)$ and avoiding the progressive phase-shift in the function $Z(t)$ produced by the previous method (fig. 3.1).

The method to detect and measure the position of cells in 3D proposed by Pimentel et al., (2012) generated Z values that were very noisy. The resulting high variance in the measurements impaired the resolution of different cells that swam within Z distance of $\sim 50\mu$m, thus also precluding the estimation of the trajectory parameters such as speed, curvature and torsion. As the diffraction patterns are asymmetric when the focal plane of the objective is above or below the plane of the center of the cell, we used the offset characteristic of each pattern for the same imaging settings (fig. 3.2E) to improve the precision of the measurements of the center of the cell (fig. 3.2F).

3.3.2 Piecewise helix fitting allows discrimination of both species by their trajectory parameters in free swimming

Sea urchin spermatozoa swim in helical trajectories in 3D (Crenshaw, 1996), and, in the limit, any curve can be locally approximated by an helical segment. To

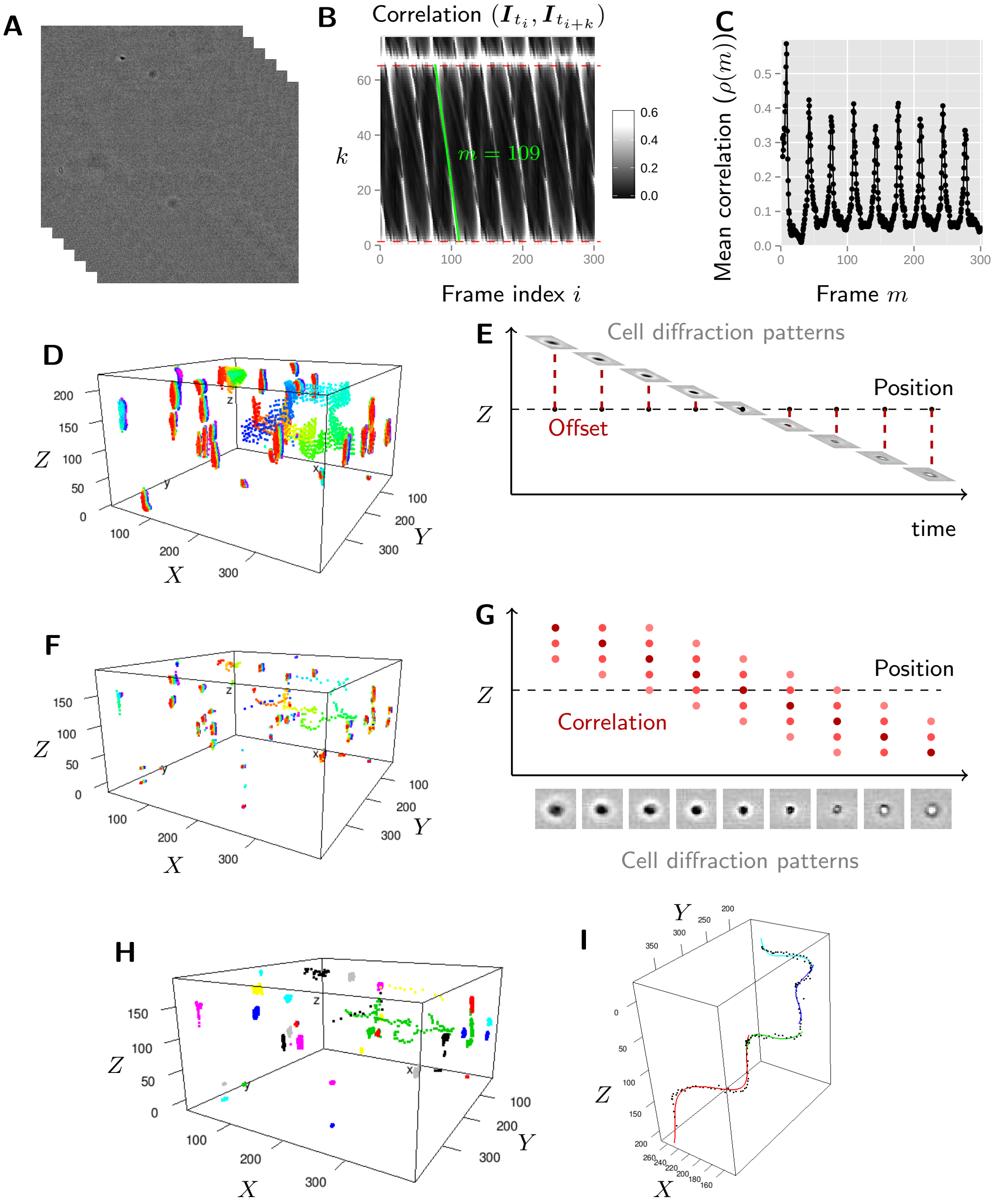


Figure 3.2: Diagram of the algorithm to estimate speed, curvature and torsion of spermatozoan trajectories from 2D+$Z(t)$ imaging data. Estimation of the $Z(t)$ position of each time-stamped frame (A) using the cross-correlation matrix (B) and estimated piezoelectric extreme position (C). Then, detection of the position of the center of each cell (D) using templates of average 2D diffraction patterns of a cell (E), each corresponding to a given expected Z offset, and consequent precise estimation of the cell Z position (F) using the estimated offset of each diffraction pattern (G). Tracking was performed by two-step clustering in space and time (H) and trajectories parameters are estimated by piecewise helix fitting (I). Scale of three-dimensional plots is in micrometers.

estimate the speed, curvature and torsion of the sperm cell, helical segments were fitted in a piecewise manner to the trajectory data. Constant helical trajectories (*i.e.* characterized by constant speed, curvature and torsion during the time under analysis) should be fitted with only one helix path, while trajectories where either one of the parameters changes or has directional changes during the analysis should be fitted with multiple different helical segments, each being a good local approximation of the trajectory. To define where and when a helical segments stops and another starts we used Dynamical Programming to screen all possible solutions under a given temporal resolution, using BIC as a cost function. The BIC allows to choose the modeling solution with the lowest number of helical segments which describes the trajectoriy appropriately. Because the data coordinates X and Y have higher accuracy and precision than those of Z, different weights were given to each spatial dimension, during the piecewise fitting procedure. A smaller weight for the Z axis is especially critical when cells are detected near the piezoelectric extremes, because only the upper or lower diffraction patterns are imaged when the cell is near the minimum or maximum, respectively, of the piezoelectric displacement amplitude. This biased information introduces a potential artifact such that cells appear to swim parallel to the upper or lower limits of the imaged volume. The smaller weight on Z minimizes this artifact. In fact, depending on the orientation of a true helical trajectory, it is possible to estimate its true parameters from the projection of the plane XY plane if we assume the weight on the Z axis to be zero. Furthermore, the application of the linear offset function while determining the cellular centroids allows to infer positions beyond the piezoelectric movement extremes. Thus, using the detection and tracking methods developed here, it is possible to infer cellular positions that are outside the imaged volume.

Using these improved methods, we were able to obtain and characterize the average trajectories of both free and confined swimming spermatozoa with little human intervention (fig. 3.3 and table 3.2). In order to validate our approach, we compared our estimates of the trajectory patterns with those obtained in previous studies and found the available values to be consistent with those we obtained (table 3.2). Contrary to confined swimming, both speed and curvature are significantly higher for *S. purpuratus*. The novelty is that torsion is also higher for this species. Furthermore, we see more variability between the different quantiles of the helices fitted to *L. pictus* free swimmers than to the other species. Note these helices were defined as the closest experimental helix to the first, second and third quartiles of speed, curvature and torsion for that species. Finally, the distribution of progressive speed is significantly different between species, and the ratio (*S. purpuratus* over *L. pictus*) of their mean and median values are 0.83 and 0.93, respectively.

3.3.3 *S. purpuratus* radius of osculating circle is different between free and confined swimming

It is worth examining the distributions of the trajectory parameters with more detail. Both speed and radius of the osculating circle (i.e. reciprocal of curvature) of the free and confined trajectories of *L. pictus* sperm were not found to be significant different (p-value>0.05 for both t-Student and Wilcoxon test). In contrast, the speed and curvature of the *S. purpuratus* sperm free trajectories were both significantly higher than those of the confined trajectories (p-value<0.001). The ratio between the mean free-swimming helix radius and the mean confined radius of the osculating circle was 0.50 and 0.42 for *L. pictus* and *S. purpuratus*, respectively. Also, the revolution period was significantly different only in the case of *S. purpuratus* (p-value<0.001). These species-specific differences might be a consequence of the factors that promote higher torsion on the free trajectory, which are expected in turn to result from a higher asymmetry in the the z component of the flagellar beating.

3.3.4 Higher asymmetry of flagellar beating accounts for the observed curvature ratio

We speculated whether the difference in radius of curvature between confined and free swimming observed in the case of *S. purpuratus* sperm is the result of greater z asymmetry in the flagellar bending, provoking both higher torsion when the cell is free and greater interaction with the boundary than the other species when it confines to the interface. To test this hypothesis, we simulated spermatozoa *in silico* using RFT in 3D, not only for the two species referred before but also for *Arbacia punctulata*. For parameterization of the model of *L. pictus* and *S. purpuratus* spermatozoa, we fitted the model to high temporal and spatial resolution imaging data of a confined spermatozoon from each species. The sizes of the head and flagellum were measured directly in the imaging data while the remainder parameters (including the functions that model the flagellar shape) were obtained by performing non-linear least-squares fitting of the simulated trajectories to the experimental ones after an initial search

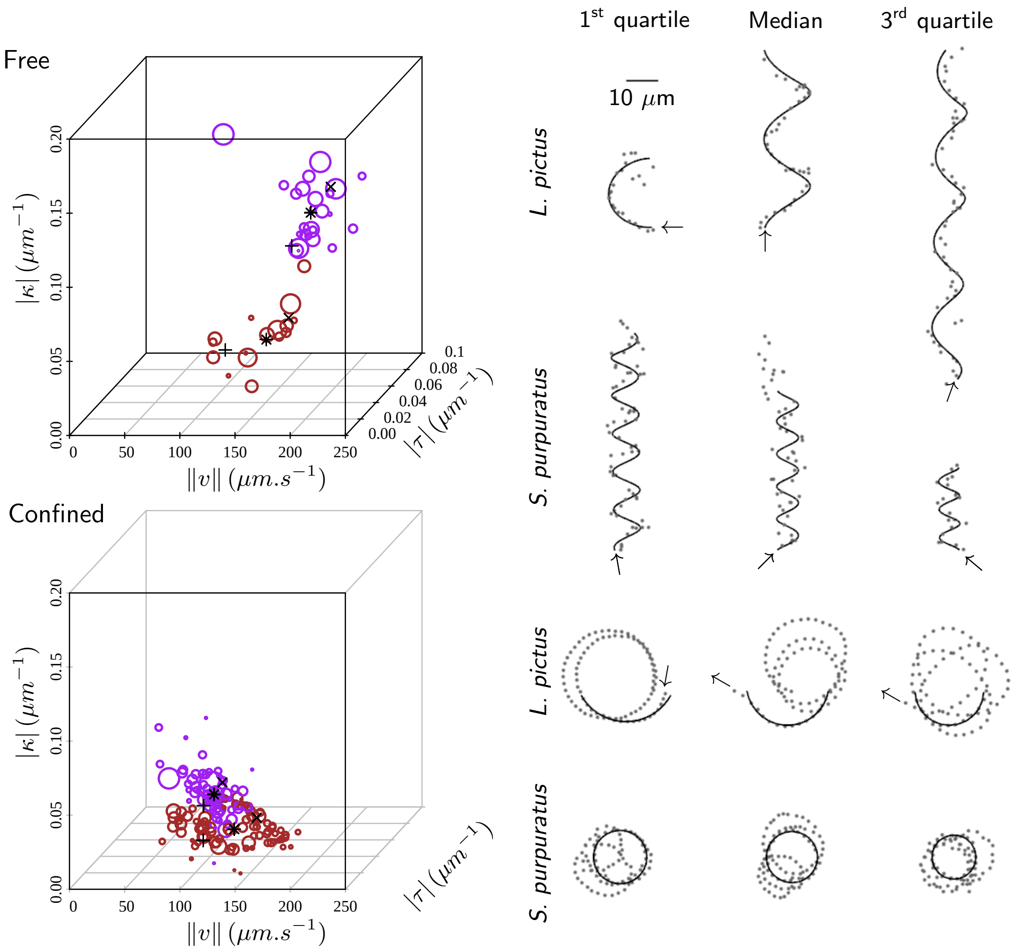


Figure 3.3: The helical trajectories of the spermatozoa of the two species are more distinct during free swimming than when confined to the plane. Left: Distribution of speed, curvature and torsion of the helical pieces fitted for *L. pictus* (brown) and *S. purpuratus* (purple) in both free (top) and confined (bottom) swimming. Size of circle is proportional to the time a cell spent swimming with such parameters. Right: Helices and circle arcs (black lines) closest (i.e. scaled Euclidean distance) to the first (+), second (*) and third (×) quartile sets of speed, curvature and torsion are displayed on the right, for each condition. Grey points represent the experimental data points used to fit the helices. Arrows represent direction of swimming. Total swimming time varies. All helices are right-handed and all confined turn clockwise.

Table 3.2: Trajectory parameters in the literature for the species studied here.

Species	Swim	Parameter	(mean±sd)[1]	(mean±se)[2]	(mean±sd)*
L. pictus	Confined	$\|r\|$ (μm)	n.a.	24.9±1.0	25.8±7.7
		$\|a\|$ (μm/s)	–	–	–
		$1/\|T\|$ (s^{-1})	n.a.	1.67±n.a.	0.9±0.3
		v (μm/s)	n.a.	(261.3)	143.6±30.4
		$1/\kappa$ (μm)	n.a.	24.9±1.0	25.8±7.7
		τ (μm^{-1})	–	–	–
	Free	r (μm)	n.a.	n.a.	13±3.7
		a (μm/s)	n.a.	n.a.	89.4±25.5
		$1/T$ (s^{-1})	n.a.	n.a.	1.5±0.5
		v (μm/s)	n.a.	n.a.	149.4±24.4
		$1/\kappa$ (μm)	n.a.	n.a.	24.6±12.3
		τ (μm^{-1})	n.a.	n.a.	0.04±0.01
S. purpuratus	Confined	r (μm)	16.3±0.3	17.8±1.0	16.1±3.8
		a (μm/s)	–	–	–
		$1/T$ (s^{-1})	1.3±0	1.45±n.a.	1.3±0.2
		v (μm/s)	134.7±9.3	(162.2)	128.6±17.2
		$1/\kappa$ (μm)	16.3±0.3	17.8±1.0	16.1±3.8
		τ (μm^{-1})	–	–	–
	Free	r (μm)	6.8±0.6	n.a.	6.7±0.9
		a (μm/s)	(55.5)	n.a.	74.6±15
		$1/T$ (s^{-1})	4±0.5	n.a.	3.7±0.7
		v (μm/s)	179.7±11.4	n.a.	175.5±30.9
		$1/\kappa$ (μm)	(7.5)	n.a.	8.3±1.1
		τ (μm^{-1})	(0.04)	n.a.	0.06±0.01

Mean spermatozoon radius (r), progressive velocity (a, pitch/period), frequency ($1/T$), velocity (v), radius of curvature ($1/\kappa$) and torsion (τ). Values in parenthesis were not available in the study but were estimated using the experimental data that was available for that experiment. For this we assumed helical trajectories. ([1]) Corkidi et al., (2008); ([2]) Guerrero et al., (2010); (*) Present study; (n.a.) not available; (–) not applicable.

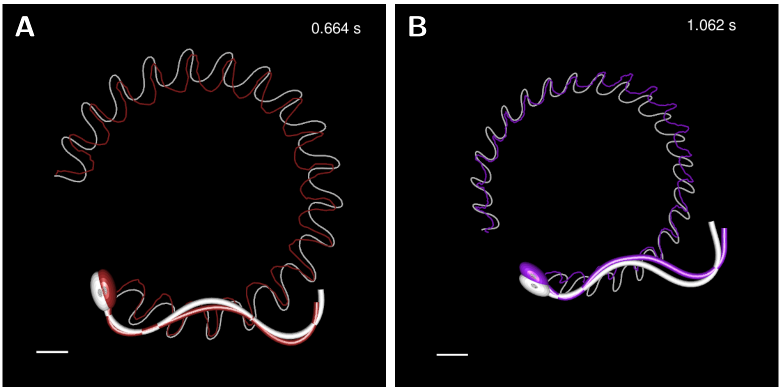


Figure 3.4: Fitting the 3DRFT model to confined swimming experimental data. Model fitting (white) to a spermatozoon of *L. pictus* (A, brown) and *S. purpuratus* (B, purple) in confined swimming. Scale bar is 5 μm.

by random walk (fig. 3.4). The parameterization of *A. punctulata* was previously described by others (Jikeli et al., 2015).

A previous study analyzed the free swimming motility of spermatozoa by assuming constant torsion along the flagellum (Jikeli et al., 2015). Another one discussed and suggested different models of 3D quasi-planar flagellar shape (Cosson et al., 2003). Here, we increased flagellar asymmetry in Z by assuming constant torsion along the flagellum $\tau(s,t) = T_0$. To model the boundary interaction, we assumed the forces at play are such that the plane defined by the two first principal components of the flagellar positions within a beating period are always parallel to the boundary surface (see section 3.2.5 for details). The drag coefficients are known to increase as an object approaches the boundary, relative to its size (Ramia et al., 1993). Because sperm cells have a quasi-planar beating pattern when confined and the discretization of the flagellum used makes a flagellar piece very small relative to the head radius, we assume the drag ratio between head and flagellum can increase when the cell is confined. We model this by changing the effective dynamic viscosity used to calculate the drag coefficients of the head. Also, it has been described that both flagellar mean curvature and flagellar curvature amplitude decrease with increased viscosity (Brokaw, 1966; Friedrich et al., 2010; Chen et al., 2015). Finally, we also assume the mechanism by which increased viscosity reduces the flagellar curvature can affect the flagellar torsion in a similar manner.

We took the fitted models for confined *L. pictus*, *S. purpuratus* and *A. punctulata*

Table 3.3: Quantiles of parameter ratios between free and confined swimming.

Ratios	*L. pictus*			*A. punctulata**			*S. purpuratus*		
η	0.21	1.00	4.95	-0.10	0.79	3.7	0.22	0.70	1.40
K_0	0.68	1.01	1.59	0.21	1.00	1.70	0.98	1.56	2.05
A_0	0.68	0.99	1.45	0.58	1.22	2.26	0.61	1.03	1.62
T_0	0.11	0.98	657.74	0.10	0.53	535.52	0.31	4.62	2592.69

Minimum, median and maximum quantiles for the effective viscosity (η), mean flagellar curvature (K_0), flagellar curvature mean amplitude (A_0) and mean flagellar torsion (T_0) ratios obtained by the model. (*) Parameterized as reported on other work (Jikeli et al., 2015).

and searched the parameter space (*i.e.* effective viscosity, mean flagellar curvature, flagellar curvature amplitude and flagellar torsion) that generated trajectories that lied within the first and third quartiles of the experimentally observed speed, curvature and torsion of the trajectories (fig. 3.5A). We then computed the ratio between free and confined parameters for each species (fig. 3.6 and table 3.3), which we expect to be one for the parameters that do not change significantly between these two swimming modes and depart from the unit for those that change. The parameter ratios obtained when *L. pictus* sperm are confined do not change significantly as compared to when they swim freely (probability of parameter ratios being equal or below 1 is 0.50, 0.47, 0.53 and 0.52 for drag, mean curvature, curvature amplitude and torsion, respectively). For *A. punctulata* sperm, both mean curvature and curvature amplitude ratios have lower probability but these are not statistically significant (probabilities 0.71, 0.51, 0.15 and 0.79). Both drag and curvature amplitude ratios are similar between the two swimming modes for *S. purpuratus* (probabilities of 0.98 and 0.42, respectively). Although not significant, the flagellar torsion is higher for free swimming sperm this species, relative to confined (probability of torsion ratio is 0.066). Also, the mean curvature is significantly higher when *S. purpuratus* sperm is free swimming (probability of <0.001). These results suggest the flagellar bending of *S. purpuratus* sperm is more affected when they confine, when compared to both *L. pictus* and *A. punctulata* sperm. Using the median parameter ratios (table 3.3) we were able to recover both free and confined median trajectories for all three species (fig. 3.5B).

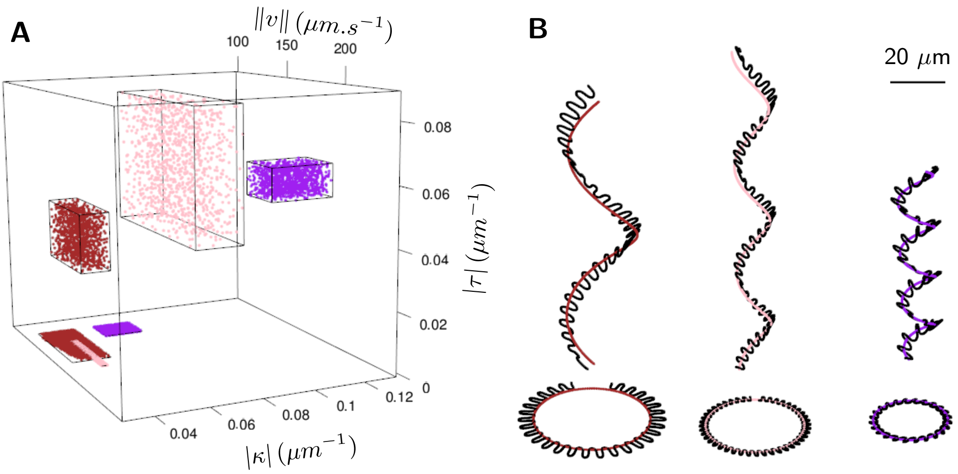


Figure 3.5: 3DRFT model recovers the experimental median trajectories of *L. pictus* (brown), *S. purpuratus* (purple) and *A. punctulata* (pink). (A) Parameters that generate trajectories which properties are within the first and third quartiles of the experimental values observed. (B) Using the median parameterization ratios we generated trajectories (black) that fit the median trajectory (color).

3.4 Discussion

Spermatozoa must deliver their genetic material to the female gamete. In most species these cells have evolved elaborate mechanisms to regulate their flagellar beating and swimming behavior to find the oocyte and achieve fertilization. The environment where these cells swim is a 3D one but historicaly they have been studied mostly in two dimensions (2D), when the cell is confined to the water-glass interface. The question whether these studies were representative of the free swimming mode was raised, as small transient changes in the path curvature result in small changes in the overall confined trajectory but significant changes in 3D helical trajectories (Guerrero et al., 2011).

Here we proposed a new method to address detection and tracking of spermatozoa imaged in 2D at oscillating Z depths which was used to measure the different confining behavior of *L. pictus* and *S. purpuratus* sperm. Then, to address the mechanism by which this difference arises, we used a mathematical approach based on Resistive Force Theory to model confined and free swimming sperm of these two species and also of *A. punctulata*. Our model suggests mean flagellar curvature is significantly diminished when *S. purpuratus* sperm confines, while flagellar torsion is

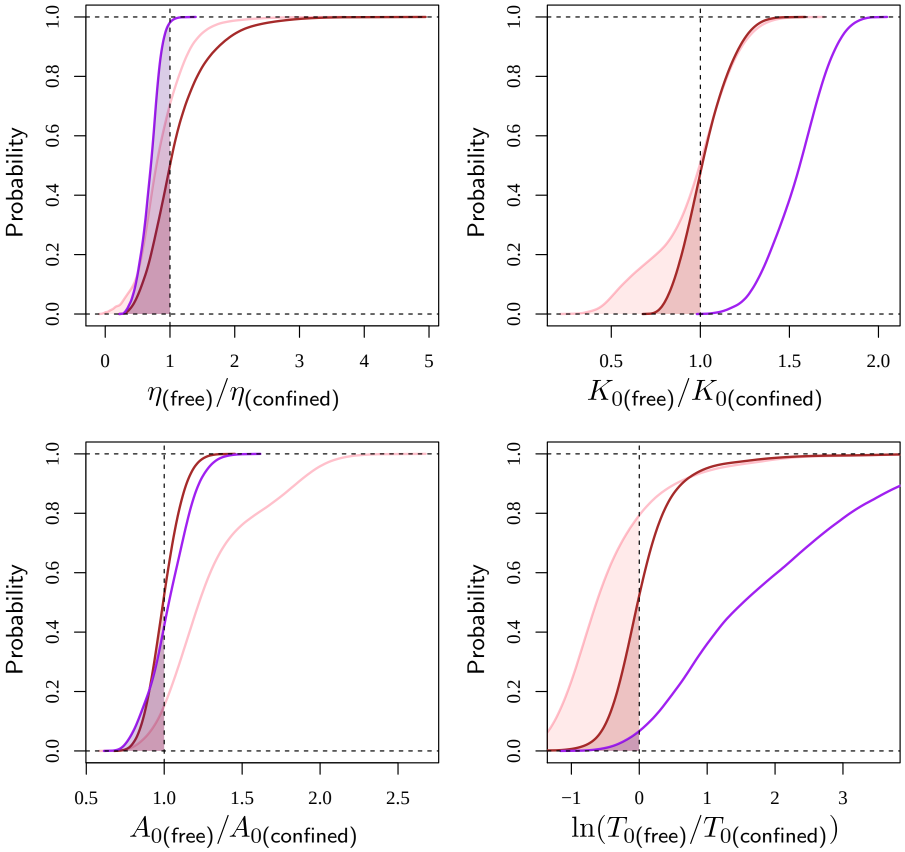


Figure 3.6: Cumulative histograms of parameter ratios between free and confined swimming obtained to *L. pictus* (brown), *S. purpuratus* (purple) and *A. punctulata* (pink). The parameters represented are the effective viscosity (η), the mean flagellar curvature (K_0), the flagellar curvature mean amplitude (A_0) and the mean flagellar torsion (T_0) ratios obtained by the model.

not.

3.4.1 From 2D+$\mathrm{Z}(\mathrm{t})$ to 3D trajectories

The use of priors are known to improve statistical inference. It is a classical issue how much we believe our priors and how much we believe the data. In the present case we made use of the knowledge that spermatozoa display helical trajectories when swimming freely in three dimensions. Piecewise fitting of helical path segments to sets of points (X, Y, Z, t) allowed us to parameterize and resolve the paths of two species of spermatozoa. It is interesting to note that attempts to use moving averaged data and/or splines fitting did generate trajectories and path parameters with local noise that failed to resolve the two species (data not shown; implicit in figure 3.1). The piecewise fitting of helical path segments allows to reach the most parsimonious account of a full trajectory, identifying in an objective way when and where sperm changed their behavior, as the discrete transitions between two segments. These discrete transitions may reflect responses of the cell to external cues on intrinsic dynamics of the cell. The heuristic potential of these new methods in scenarios of chemotaxis remains to be explored. A criticism that can be made to the method is that it will tend to present as discrete events smooth continuous changes in the parameters of a path. Such scenarios should leave a signature in the residuals of the fitting. Defining whether swimming cells undergo continuous changes or more discrete transitions in behavior remains to be clarified.

Here we proposed a new method to address detection and tracking of spermatozoa imaged in 2D at oscillating Z positions. Our method required little human intervention, primarily for rechecking tracking efficiency. Many trajectories were discarded due to poor fitting, most with a small helix radius. This was because under this situation the noise becomes significant and the fitting ended up approximating the helix axis instead of the helix itself. This was particularly relevant for *S. purpuratus*, as its helical radius is substantially smaller. For this reason, we might have a biased estimation of the trajectory parameters. Although we cannot exclude the possibility that other non-helical swimming patterns do exist for these species and that simpler models (*i.e.* with less parameters) would characterize the trajectories adequately, the piecewise helix fitting method should able to describe those trajectories by segments of helical arcs.

Other work focused on estimating swimming parameters by helical fitting or by obtaining helical parameters from 2D information only (Crenshaw et al., 2000; Gurarie et al., 2011; Cherin et al., 2014). Notwithstanding, these methods cannot be applied to our study as some of their assumptions do not apply to our data. A number of methods are unable to resolve helices where the helical axis is parallel to Z, some need more data points per trajectory than what we have and others assume equal noise in all dimensions. There are other alternatives methods to study changes in animal behavior (i.e. movement) (Gurarie et al., 2016). These, however, need many parameters or assume data independece, which does not occur for the same cell. Although having the disadvantage of only fitting helices, the only parameter of the piecewise helix fitting method developed here is the desired time resolution, making it easier and straightforward to be applied to helicoidal data than the other methods.

The strategy we have used is discontinuous on transition points, for example time-points where a helix ends and another begins in the same trajectory; caution must be advised when interpreting swimming behavior near these points. This should not affect the conclusions at which this work arrives, as we are studying the main swimming behaviors that these species display in two different boundary conditions. In the study of taxis, our method should be able to detect changes in speed, curvature, torsion or direction. However, this method has limitations when studying fine events such as acute turning, as these occur near the transition points.

We can apply this method since our study deals with the average path (i.e. the maximum temporal resolution we have for this volume in this microscopy system). If we take the flagellar beating into account, the trajectory should be a chiral ribbon as described for some human and horse spermatozoa (Su et al., 2012; Su et al., 2013). This can be implemented by changing the helix function to be fitted but one must be aware that more information might be needed to fit the extra parameters of more complex functions.

Finally, it is worth noting that new microscopy systems based on holography are emerging, being able to track fast cells in 3D with submicron resolution (Su et al., 2012; Jikeli et al., 2015). Although our system has less resolution, it has the potential to be coupled to fluorescent measurements of the flagellum. Also, it is possible to increase the spatial and temporal resolution of our system by scanning a smaller volume (*i.e.* lower Z amplitude), which allowed us to image the flagellum in 3D (Silva-Villalobos et al., 2014). Together with the improvements made in this work, the 2D+$Z(t)$ microscopy system remains an up-to-date, powerful tool to study sea urchin chemotaxis.

3.4.2 Free swimming trajectories

A few free swimming trajectories of sea urchin sperm have been characterized, namely for *A. punctulata* (Crenshaw et al., 2000; Jikeli et al., 2015) and *S. purpuratus* (Corkidi et al., 2008) species. Although the first focuses on the speed, curvature and torsion of a trajectory, the second focuses on the helical path parameters like radius and revolution speed, but does not analyze progressive speed. In spite of the fact that these two definitions of helical trajectories are redundant, as one can be calculated from the other, the six parameters here described are important to establish differences and similarities between cellular trajectories.

3.4.3 Confined *vs* free swimming

A discrepancy between free swimming helix radius and confined osculating circle radius was been previously described for *S. purpuratus* sperm. In the present work we confirmed that this difference is also observed in *L. pictus* spermatozoa. However, the free swimming osculating circle radius had not been described for these species until now. Interestingly, the trajectory curvature was comparable in free and confined swimming for *L. pictus* but markedly different for *S. purpuratus* spermatozoa. According to the parameters of the trajectories of *A. punctulata* sperm reported in the literature these sperm seem to be an interpolation between the two.

We hypothesized and then proved that higher asymmetry in the z component of the flagellar beating (*i.e.* in this case provided by constant flagellar torsion on an increasingly asymmetrical curved flagellum) is sufficient to increase the free swimming trajectory torsion, but not to increase the trajectory's velocity and curvature. As previously reported, we cannot discard the possibility that hydrodynamic interactions near the boundary affect the swimming behavior (Fauci and Dillon, 2006; Smith et al., 2009) but here we simplified such interactions by forcing the plane of flagellar beating to be parallel to the bounding surface. Using this rationale we were able to quantitatively explain the experimental data of all three sperm species, assuming that average flagellar curvature and average torsion are lower when the *S. purpuratus* sperm are confined. Note that by construction, the framework used in this study assumes the boundary exerts forces on the flagellum but these do not deform it if the flagellar shape parameter are the same (*e.g.* flagellar parameter ratio between free and confined are equal to one). Therefore, one does not need to evoke any change in the spermatozoon internal machinery to explain confinement and planar swimming of *L. pictus* and *A. punctulata* sperm but that such alterations are required to explain the observations on those of *S. purpuratus*. Instead, different flagellar stiffness might account for the differences observed. Another interpretation would be that higher flagellar torsion of *S. purpuratus* sperm provokes higher off-plane beating asymmetry which propels the cell closer to the boundary. As the effective drag exerted on the cell increases exponentially as the cell gets closer to the surface, the flagellar beating quickly becomes planar if the internal axonemal forces remain the same for both free and confined swimming. Nosrati et al., (2015) studied human and bull sperm swimming close to surfaces using total internal reflection fluorescence (TIRF) microscopy and concluded that bull spermatozoa swim closer to the boundary and have more marked changes in their trajectory. Also, these sperm experience more flattening of the flagellar beating waves. Likewise, we speculate that *S. purpuratus* sperm swim so close to the boundary that its flagellar beating becomes more planar than that of *L. pictus*, even though the former have a higher flagellar torsion in free swimming.

The comparison of the parameters of the swimming trajectories and the modeling identified a trend in the way spermatozoa alter their swimming path when interacting with a surface. The speed and the curvature of the swimming path tends to decrease when the sperm get confined, but in the case of the *S. purpuratus* this change is very marked. Mathematical modeling of the morphodynamics of these cells led us to hypothesize that the confinement brings the cells to a more viscous environment at the liquid-solid interface that changes both the drag ratios and the torsion and curvature of the flagellar bending waves. The smaller *S. purpuratus* spermatozoa would be more sensitive to these changes in viscosity and/or may swim closer to the solid surface being subjected to a higher viscosity than *L. pictus* and *A. punctulata* spermatozoa.

We reproduced the observed free swimming trajectories using a model parameterized based on the data from confined spermatozoa and setting flagellar torsion to a positive values. This suggests that is possible to infer *A. punctulata* or *L. pictus* free swimming trajectories from the studies on confined settings, or at least define a shorter family of possibilities. However, the properties of planar motility cannot be extrapolated for 3D in the case of *S. purpuratus* sperm. This raises a cautionary note when trying to extrapolate swimming behavior from experiments where natural free-swimmers are studied under planar confinement, at least for some species. We cannot discard the possibility that in previous chemotaxis studies using *S. purpuratus* (Guerrero et al., 2010) the cells responded in such a way that would bring them closer to the gradient center if they were swimming far from the surface but the proximity to the boundary impaired such response by cancelling either or both flagellar curvature and torsion.

3.4.4 Conclusion and future work

Here we present the first characterization of *L. pictus* sperm free swimming trajectories in the absence of any nominal stimulus and complemented this with a meta-analysis of the available data on *S. purpuratus* and *A. punctulata*. For the first time, sea urchin species were compared regarding their swimming behavior in unconstrained and constrained environments. This comparative study allowed us to to better understand how spermatozoa swim and is an example of the importance of interspecies studies. Our results indicate that the conclusions on the motility and chemotaxis studies performed on sperm confined to planar movement might be convenient but not directly extrapolated to the behavior in natural conditions. This emphasizes the need to characterize the strategies used by sperm to steer the swimming path in 3D.

References

Ahn, S. J. (2004). "Least squares orthogonal distance fitting of curves and surfaces in space". PhD thesis. Berlin Heidelberg: University of Stuttgart, Germany.

An, Y., Shao, C., Wang, X., and Li, Z. (2011). "Geometric properties estimation from discrete curves using discrete derivatives". *Computers & Graphics* 35.(4), pp. 916–930.

Brokaw, C. J. (1966). "Effects of increased viscosity on the movements of some invertebrate spermatozoa". *Control* 45, pp. 113–139.

Cao, D., Liu, D., and Wang, C. H.-T. (2006). "Three-dimensional nonlinear dynamics of slender structures: Cosserat rod element approach". *International Journal of Solids and Structures* 43.(3-4), pp. 760–783.

Chen, D. T. N., Heymann, M., Fraden, S., Nicastro, D., and Dogic, Z. (2015). "ATP consumption of eukaryotic flagella measured at a single-cell level." *Biophysical journal* 109.(12), pp. 2562–73.

Cherin, N., Cordier, F., and Melkemi, M. (2014). "Modeling piecewise helix curves from 2D sketches". *Computer-Aided Design* 46, pp. 258–262.

Coeurjolly, D. and Svensson, S. (2003). "Estimation of curvature along curves with application to fibres in 3d images of paper". *Proceedings of the 13th scandinavian conference on image analysis. Lecture notes in computer science*. Ed. by Bigun, J. and Gustavsson, T. Vol. 2749. Berlin Heidelberg: Springer-Verlag, pp. 247–54.

Corkidi, G., Taboada, B., Wood, C. D., Guerrero, A., and Darszon, A. (2008). "Tracking sperm in three-dimensions." *Biochemical and Biophysical Research Communications* 373.(1), pp. 125–129.

Cosson, J., Groison, A.-L., Suquet, M., Fauvel, C., Dreanno, C., and Billard, R. (2008). "Marine fish spermatozoa: racing ephemeral swimmers." *Reproduction (Cambridge, England)* 136.(3), pp. 277–94.

Cosson, J., Huitorel, P., and Gagnon, C. (2003). "How spermatozoa come to be confined to surfaces." *Cell motility and the cytoskeleton* 54.(1), pp. 56–63.

Crenshaw, H. C. (1991). "A technique for tracking spermatozoa in three dimensions without viscous wall effects". *Comparative Spermatology* 20, pp. 353–357.

— (1996). "A new look at locomotion in microorganisms: rotating and translating". *American Zoologist* 36.(6), pp. 608–618.

Crenshaw, H. C., Ciampaglio, C. N., and McHenry, M. (2000). "Analysis of the three-dimensional trajectories of organisms: estimates of velocity, curvature and torsion from positional information". *Journal of Experimental Biology* 203, pp. 961–982.

Fauci, L. J. and Dillon, R. (2006). "Biofluidmechanics of Reproduction". *Annual Review of Fluid Mechanics* 38.(1), pp. 371–394.

Friedrich, B. M., Riedel-Kruse, I. H., Howard, J., and Jülicher, F. (2010). "High-precision tracking of sperm swimming fine structure provides strong test of resistive force theory." *The Journal of experimental biology* 213.(Pt 8), pp. 1226–34.

Gray, J. and Hancock, G. (1955). "The propulsion of sea-urchin spermatozoa". *Journal of Experimental Biology* 32.(4), pp. 802–814.

Guerrero, A., Carneiro, J., Pimentel, J. A., Wood, C. D., Corkidi, G., and Darszon, A. (2011). "Strategies for locating the female gamete: the importance of measuring sperm trajectories in three spatial dimensions". *Molecular Human Reproduction* 17.(8), pp. 511–523.

Guerrero, A., Nishigaki, T., Carneiro, J., Yoshiro Tatsu, Wood, C. D., and Darszon, A. (2010). "Tuning sperm chemotaxis by calcium burst timing." *Developmental Biology* 344.(1), pp. 52–65.

Gurarie, E., Bracis, C., Delgado, M., Meckley, T. D., Kojola, I., and Wagner, C. M. (2016). "What is the animal doing? Tools for exploring behavioural structure in animal movements." *The Journal of animal ecology* 85.(1), pp. 69–84.

Gurarie, E., Grünbaum, D., and Nishizaki, M. T. (2011). "Estimating 3D movements from 2D observations using a continuous model of helical swimming." *Bulletin of mathematical biology* 73.(6), pp. 1358–77.

Jikeli, J. F., Alvarez, L., Friedrich, B. M., Wilson, L. G., Pascal, R., Colin, R., Pichlo, M., Rennhack, A., Brenker, C., and Kaupp, U. B. (2015). "Sperm navigation along helical paths in 3D chemoattractant landscapes." *Nature Communications* 6, pp. 1–10.

Liu, Y. and Wang, W. (2008). "A revisit to least squares orthogonal distance fitting of parametric curves and surfaces". *Proceedings of the 5th Internati onal Conference on Advances in Geometric Modeling and Processing. Springer-Verlag, Hangzhou, China*, pp. 384–397.

Meijering, E., Dzyubachyk, O., and Smal, I. (2012). "Methods for cell and particle tracking". *Imaging and Spectroscopic Analysis of Living Cells*. Ed. by Conn, P. M. Vol. 504. Elsevier. Chap. 9, pp. 183–200.

Nosrati, R., Driouchi, A., Yip, C. M., and Sinton, D. (2015). "Two-dimensional slither swimming of sperm within a micrometre of a surface." *Nature communications* 6, p. 8703.

Perrin, F. (1936). "Mouvement Brownien d'un ellipsoide (II). Rotation libre et dépolarisation des fluorescences. Translation et diffusion de molécules ellipsoidales". *Le Journal de Physique et le Radium* Série VII.(1), pp. 1–11.

Pimentel, J. A., Carneiro, J., Darszon, A., and Corkidi, G. (2012). "A segmentation algorithm for automated tracking of fast swimming unlabelled cells in three dimensions." *Journal of Microscopy* 245.(Pt 1), pp. 72–81.

Pimentel, J. A. and Corkidi, G. (2009). "Mechanical vibration compensation method for 3D+t multi-particle tracking in microscopic volumes." *Conference Proceedings IEEE Engineering in Medicine and Biology Society* 2009, pp. 1429–32.

— (2010). "Three dimensional template matching segmentation method for motile cells in 3D+t video sequences." *Conference Proceedings IEEE Engineering in Medicine and Biology Society* 2010, pp. 4777–80.

Ramia, M., Tullock, D. L., and Phan-Thien, N. (1993). "The role of hydrodynamic interaction in the locomotion of microorganisms". *Biophysical Journal* 65.(August), pp. 755–778.

Riedel-Kruse, I. H., Hilfinger, A., Howard, J., and Jülicher, F. (2007). "How molecular motors shape the flagellar beat." *HFSP journal* 1.(3), pp. 192–208.

Silva-Villalobos, F., Pimentel, J. A., Darszon, A., and Corkidi, G. (2014). "Imaging of the 3D dynamics of flagellar beating in human sperm." *Conference Proceedings IEEE Engineering in Medicine and Biology Society* 2014, pp. 190–193.

Smith, D. J., Gaffney, E. A., Blake, J. R., and Kirkman-Brown, J. C. (2009). "Human sperm accumulation near surfaces: a simulation study". *Journal of Fluid Mechanics* 621, p. 289.

Su, T.-W., Choi, I., Feng, J., Huang, K., McLeod, E., and Ozcan, A. (2013). "Sperm trajectories form chiral ribbons." *Scientific reports* 3.(1664).

Su, T.-W., Xue, L., and Ozcan, A. (2012). “High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories”. *Proceedings of the National Academy of Sciences* 109.(40), pp. 3–7.

Tatsu, Y., Nishigaki, T., Darszon, A., and Yumoto, N. (2002). “A caged sperm-activating peptide that has a photocleavable protecting group on the backbone amide”. *FEBS letters* 525, pp. 20–24.

Chapter 4

Analysis of sperm chemotaxis in three dimensions

Pedro Ângelo Silva[1], Arturo Pimentel[2], Tatiana Luna[2], Gabriel Corkidi[2], Alberto Darszon and Jorge Carneiro[1].

[1] Instituto Gulbenkian de Ciência, Portugal; [2] Instituto de Biotecnología, Universidad Nacional Autónoma de México, Estados Unidos de México

Abstract

Sperm chemotaxis is a fundamental biological process by which male gametes climb a gradient of concentration of a chemoattractant released by the conspecific egg. It was shown that spermatozoa of *Lytechinus pictus* and *Strongylocentrotus purpuratus* respond differently to the sperm activating peptide Speract when studied in conditions in which the motility is confined to the plane of the water-glass interface. While sperm of the former species display chemotactic responses the latter do not under the same conditions. Since the natural environment where sea urchin sperm swim is the three dimensional (3D) volume and not the convenient interface plane, it is important to ascertain whether the chemotactic responses are also observed when the sperm are swimming freely in 3D, and whether the reported species-specific differences are retained in this setting. Attempts to assess the 3D chemotaxis for the sperm of these two species based on the analysis of radial cell densities relative to the

center of sperm activating peptide (SAP) gradient were inconclusive due to several confounding factors, including cells entering and leaving the observation domain at different times. In the present study we tested the significance of chemotactic effects in data generated by the 2D+$Z(t)$ microscopy setup. The trajectories of spermatozoa were reconstituted by piecewise fitting of helical segments. The analysis of the distributions of the parameters of the helical trajectories displayed by sperm cells in the presence and in the absence of an artificial Speract gradient did not detect any effect of this chemoattractant on the swimming trajectories. We concluded that chemotaxis was not revealed under the conditions of the experiment.

Acknowledgements

Tatiana Luna, Arturo Pimentel, Gabriel Corkidi and Alberto Darszon produced the experimental 2D+$Z(t)$ imaging data; Pedro Ângelo Silva and Jorge Carneiro planned the analysis and modeling work. Pedro Ângelo Silva performed the work. Pedro Ângelo Silva, Alberto Darszon and Jorge Carneiro wrote the manuscript. Fundação para a Ciência e Tecnologia, Portugal, (SFRH/BD/79261/2011), Instituto Gulbenkian de Ciência, Portugal, and Instituto de Biotecnología, UNAM, Mexico, provided funding for this work.

4.1 Introduction

Chemotaxis refers to the ability of a cell to move up the gradient of chemical attractants. It is believed that chemotaxis is necessary to allow the sperm cell to find the egg and subsequently fertilize it. Evidence for chemotaxis was first described for sea urchin sperm by showing that *A. punctulata* sperm accumulate at the center of a gradient of the sperm activating peptide (SAP) Resact purified from conspecific eggs (Lillie, 1912). These seminal population studies were complemented by time-lapse imaging and tracking of individual cells in the presence of an artificial controlled gradient of species-specific SAP. The original studies on *A. punctulata* sperm responses to Resact (Ward et al., 1985) were extended to spermatozoa of *Lytechinus pictus* spermatozoa and *Strongylocentrotus purpuratus* in response to Speract (Guerrero et al., 2010). Intriguingly, while *L. pictus* spermatozoa were shown to redirect their circular trajectories preferentially towards the center of the chemoattractant gradient, the *S. purpuratus* sperm, despite being able to sense the same environmental cues, reoriented their swimming in a random way, unrelated to the Speract concentration.

All these studies on sperm chemotaxis were based on two dimensional analysis of sea urchin sperm when the cells are confined to the water-glass interface. It is relevant to ask if these chemotactic responses observed in two dimensions (2D) are relevant for the 3D volume where these cells swim in nature (Guerrero et al., 2011). Recently, the 3D chemotactic motility of *A. punctulata* was studied and the authors of the report suggest two different swimming modes under chemotactic conditions (Jikeli et al., 2015). The contribution of these swimming modes may be different whether the cell is confined or free-swimming. Studying the free-swimming chemotactic behavior of other species sperm is mandatory to assess the generality of these findings. Furthermore, it is unclear whether the lack of chemotactic responses previously described for *S. purpuratus* sperm is also observed in 3D, and conversely if *L. pictus* (Guerrero et al., 2010; Guerrero et al., 2011) retains its chemotactic capabilities when swimming freely. The sperm of these two species were studied in 3D using the 2D+$Z(t)$ microscope setup developed by Corkidi et al., (2008) and their chemotactic behavior was assessed using population densities (Pimentel, 2013). However, the analysis of this data led to no definitive conclusions presumably due to confounding factors associated with cells leaving and entering the observation volume and because the statistical procedures assumed independence of the observations, which did not hold (A. Pimentel, personal communication).

The objective of this chapter is to reassess the 3D chemotactic behavior of both *L. pictus* and *S. purpuratus* sperm. We will still use the same data analysed before produced by the 2D+$Z(t)$ microscope setup. We took advantage of the piecewise helix fitting method developed in *Chapter 3* to estimate the trajectory parameters and used linear mixed models to address the dependence of the data points. We applied this method to data on free (3D) swimming sperm in the presence and absence of a Speract gradient and we could not detect chemotaxis for either species for the conditions tested. We show that the chemoattractant gradient does not affect the average trajectory speed, curvature, torsion and speed to gradient center.

4.2 Methods

4.2.1 Sperm imaging data

Sperm of either species were collected after intracoelomic injection of 0.5 M of KCl stored on ice and used within 24 hours. Immediately before imaging, cells were transferred to two different solutions of artificial sea water, one with and another without 0.1 μM of caged speract (see Guerrero et al., (2010) for further details). The 2D+$Z(t)$ imaging system was previously described in Corkidi et al., (2008) and Pimentel et al., (2012). Briefly, a piezoelectric device moves the objective up and down allowing to take different Z sections at a fast rate. This is needed to scan a considerable volume where cells swim very fast (spermatozoa swim at 200-300 μm.s^{-1}). A subset of the imaged field-of-view was irradiated with ultraviolet (UV) light within the time interval 2-4 s. Please refer to section 3.2.1 for further details on sperm preparation, microscopy settings and image acquisition.

4.2.2 Imaging data analysis

The 2D+$Z(t)$ imaging data was analyzed according to section 3.2.2. Briefly, we inferred the depth of each frame using a correlogram of the imaging data, detected the position of cells using a 3D cell template of the diffraction patterns, tracked the cells by clustering and estimated the parameters of the trajectories by piecewise helix fitting. Here we also define the velocity to the gradient center as follows. Consider the tangent vector of the helical trajectory $\boldsymbol{v}(t)$ and the unit vector $\boldsymbol{u}(t) =$

$(\boldsymbol{g} - \boldsymbol{r}(t))/\|\boldsymbol{g} - \boldsymbol{r}(t)\|$ defined by current cell position ($\boldsymbol{r}(t)$) and the fixed gradient center $\boldsymbol{g}$. Then, the velocity component relative to the gradient center is the dot product of the two: $\boldsymbol{v}_g(t) = \boldsymbol{v}(t) \cdot \boldsymbol{u}(t)$.

4.2.3 Statistical analysis

All piecewise fitted trajectories included in this study were manually inspected and discarded if inadequate, keeping trajectories with one and only one cell with realistic speed ($50 \leq \|v\| \leq 300$ μm.s^{-1}), curvature ($0 \leq |\kappa| \leq 1$ rad.μm^{-1}) and torsions ($0 \leq |\tau| \leq 1$ rad.μm^{-1}). Furthermore, only trajectories spanning more than one second were considered.

Data was analyzed using R statistical program v.3.0.3 (R Foundation for Statistical Computing, Vienna, Austria). The data can be categorized into the following factors: $Species$ ($L.pictus$ or $S.purpuratus$), $Treatment$ (with (CS) or without ($None$) caged Speract), UV irradiation (according to time intervals of UV irradiation: 0-2 s ($Before$), 2-4 s ($During$), 4-6 s ($AfterI$) and 6-8 s ($AfterII$)) and $Sperm$ (unique cell identification number). Linear mixed models were fitted using the R function `lmer4::lmer` assuming the random effect of Sperm factor within the UV Irradiation factor ($UV|Sperm$), taking into account the dependence of the observations on each individual Sperm within the time intervals relative of UV uncaging. Models spanning all possible combinations of Species, Treatment and UV Irradiation as main effects and their interactions were fitted to the data. The model that best fits the data was selected by log-likelihood ratio.

Note the following notation used in the definition of the several linear mixed models: symbol ($+$) means an added effect (*e.g.*, $p \sim A + B$, p is modeled by the main effects of A and B), symbol (:) represents and interaction (correlation) between factors (*e.g.*, $p \sim A + B + C + A : B$, p is modeled by the main effects of A, B and C and also the interaction of A and B) and the symbol ($|$) represents a random effect (*e.g.*, $p \sim A + (B|C)$, p is modeled by the main effects of A, given the random effects of C within B).

4.3 Results

4.3.1 3D experiments with caged Speract

To investigate chemotaxis in 3D, sperm of *S. purpuratus* and *L. pictus* where imaged using the 2D+$Z(t)$ microscope setup (Corkidi et al., 2008). UV irradiation of caged Speract was used to produce a gradient of this SAP under conditions previously shown to elicit responses by sperm of these two species (Guerrero et al., 2010; Guerrero et al., 2011). Speract is a small peptide extracted from the eggs of *S. purpuratus* and its caged form contains a 2-nitrobenzyl group at a backbone amide, which lowers its affinity to its receptors by several orders of magnitude. This caging-group can be released under UV irradiation making the Speract active and detectable by the cell. The sperm of the two species were imaged under two different conditions, one where caged Speract is present (CS) and other where it is not (None). Within the time interval 2-4 s, all samples were irradiated with UV, thus releasing Speract and creating a gradient of this SAP in the CS treatment but not in the None treatment. Cells that detect and react to the SAP should alter the average trajectory parameters such as velocity, curvature or torsion. If, additionally, the cells display chemotactic behavior, the magnitude of the velocity component relative to the vector between the position of the cell and the center of the gradient should increase and become more positive (Pimentel, 2013) (see section 4.2.2).

We used the methodology developed in *Chapter 3* to fit piecewise helical segments to the data. We aggregated the parameter estimates of these trajectories by time intervals relative to UV irradiation (UV): 0-2s ($Before$), 2-4 s ($During$), 4-6 s ($AfterI$) and 6-8 s ($AfterII$). Note we only considered cells swimming at speeds above 50 μm.s^{-1}, neglecting all slower objects. As the same cell is being measured several times within and across UV irradiation periods, the observations are not independent. To deal with such cases we can use linear mixed models (Henderson, 1982). The interesting feature of these models is that they account for dependence in the residuals of repeated measurements (Bates et al., 2014). In this case, we expect to have a random effect per cell within each time interval of UV irradiation ($UV|Sperm$). To detect what are the main factors that have an effect on the trajectory parameters we tested different models by log-likelihood ratio. These different models were a combination of the possible main effects ($Treatment \in \{None, CS\}$; $Species \in \{L.pictus, S.purpuratus\}$ and $UV \in \{Before, During, AfterI, AfterII\}$) and the interactions between them, assuming the random effect of cell within each UV time period ($UV|Sperm$). Choosing one model over the other by log-likelihood ratio means the data supports that model more than it supports the competing model. In the case of a chemotactic response, we expect that both the main factors $Treatment$ (relative to the presence or absence of caged Speract) and UV (relative to the time interval of Speract uncaging and its subsequent diffusion: $During$, $AfterI$ and $AfterII$) to be significant, as well as their interaction ($Treatment : UV$). In fact, this interaction is essential for a chemotactic model as it suggests the combination of the presence of caged Speract and of the time of UV irradiation have an effect, which could be due to Speract being uncaged and diffused. The null hypothesis would be that there is no effect of any of the factors or their interactions.

We can see that the only statistically significant effect on the average speed, curvature and torsion of the trajectories is the factor $Species$ (fig. 4.1A, B and C). This agrees with the results of the previous chapter that these parameters are different for sperm of *S. purpuratus* and *L. pictus*. More importantly, it indicates that there is no statistically significant effect of the combination of caged Speract, CS, with any of the intervals $During$ and $After$ UV irradiation on the averaged trajectory speed, curvature or torsion. In other words, no chemotactic responses were detectable under the conditions of the experiment. As a curiosity, there is a statistical effect of the interaction between species and of some conditions for the velocity to gradient center $\boldsymbol{v_g}$ (fig. 4.1D). However, a closer examination at the data indicates that the effect is observed in *L. pictus* sperm in the condition $During$ UV irradiation, when no caged Speract is present, meaning this difference cannot be accounted for by a chemotactic response, and most likely represents a type I error.

4.4 Discussion

To assess if the motility of spermatozoa swimming freely is affected by the presence of Speract, we used piecewise helix fitting to obtain the average trajectory speed, curvature, torsion and velocity component relative to the center of the gradient. Using linear mixed models we found that none of these sperm trajectory parameters were significantly affected by the presence of the chemoattractant in any of the species studied. These results do not confirm the previous report on *L. pictus* sperm chemotaxis under the same conditions (Pimentel, 2013). The previous conclusion

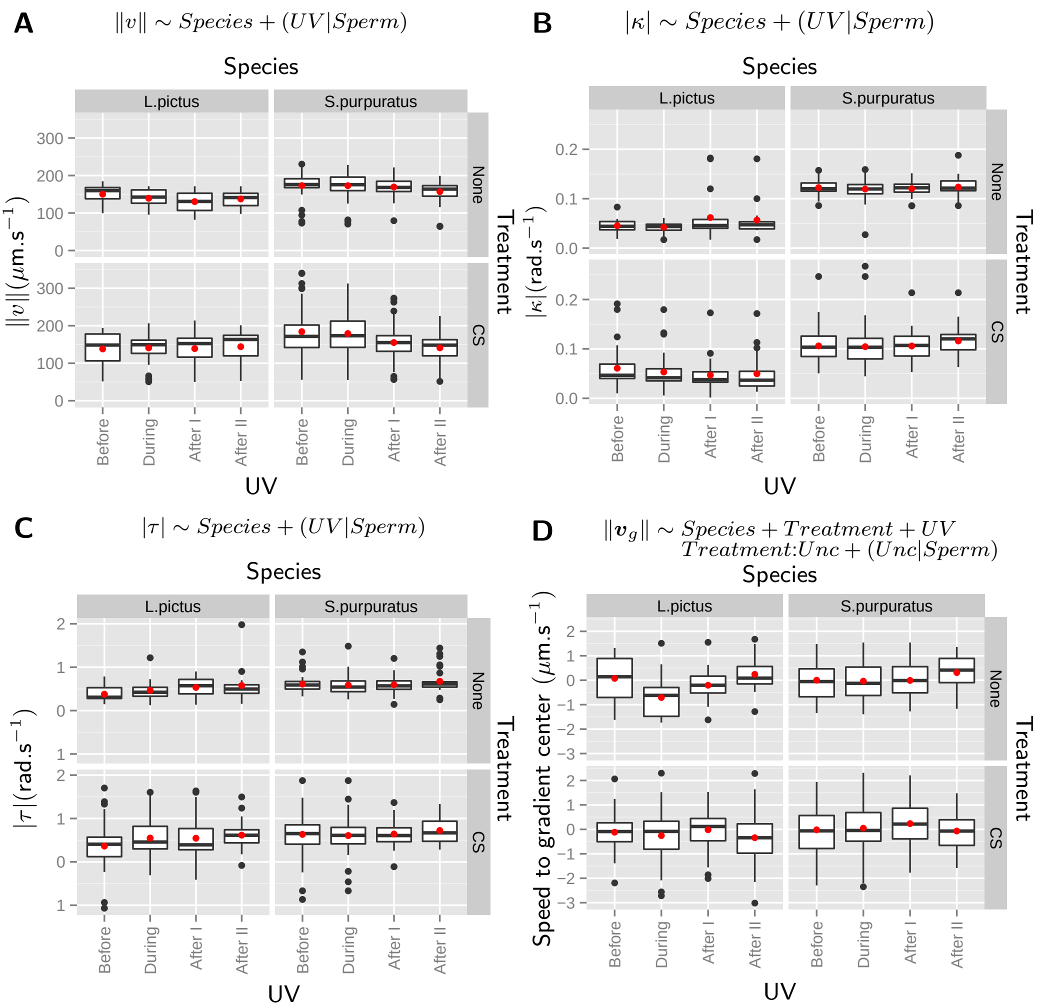


Figure 4.1: Linear mixed models of chemotaxis for either average path speed (A), curvature (B), torsion (C) and speed to center (D) for different $Treatments$ – with (CS) and without ($None$) caged Speract – at different times after UV irradiation (UV): 0-2s ($Before$), 2-4 s ($During$), 4-6 s ($After I$) and 6-8 s ($After II$). The data is represented as box-and-whiskers subplots with the mean value (red dot). On top of each plot is the model chosen by log-likelihood ratio.

might have been mislead by type one errors arising from the violation of the assumption of independence of data points and the confounding effects generated by cells entering and leaving the volume. We cannot discard the unlikely hypothesis that the independence of data points between the time intervals (factor UV) could have also affected our results. This could be addressed by assuming an additional random effect only dependent on cell ($1|Sperm$, according to the notation of the R statistical software) but the available sample size was too small to fit models based on this assumption. The fact that individual cells rarely swam in the imaged volume during the eight seconds of the experiment also impaired the possibility of this assumption (*i.e.* unbalanced data).

Overall these results suggest that sperm of *L. pictus* and *S. purpuratus* do not exhibit chemotaxis in under the conditions of the experiment. As *L. pictus* sperm were previously shown to be chemotactic in confining studies (Guerrero et al., 2010), we suggest the exploration of different chemoattractant gradients (*i.e.* using different caged Speract concentrations or UV irradiation times and conditions). Jikeli et al., (2015) showed 3D chemotaxis for *A. punctulata* spermatozoa by uncaging the chemoattractant throughout the experiment. This should also be tested with the two species used in this study. Also note that the approach taken here only allowed to test the significance of a behavior similar to the 'off-response' chemotactic mode suggested for free-swimming sperm of *A. punctulata*. In the present study, the UV irradiation was performed using the objective, while it was moving up and down driven by the piezoelectric device. Under these conditions, the uncaging forms a hour-glass shaped gradient in three dimensions (Pimentel, 2013). It is an intriguing possibility that such artificial hour-glass gradient may not be decodable by the sensorimotor system of the spermatozoa that has been evolved to locate an approximate radial gradient around the egg.

In this work, we tested whether the motility parameters of sperm differed between different conditions in the experiment, including those presumably containing meaningful concentrations of Speract. There are different statistical frameworks that test the aggregation or clustering of cells which have been previously developed (*i.e.* positional information). For example, Ripley's K-function measures the expected number of neighbors within a certain radius of a cell and compares to the null hypothesis that this number is given by a Poisson process (*i.e.* the points are uniformly distributed) (Ripley, 1977). The estimate of K can summarize the aspects like interpoint dependence and clustering. This kink of summary statistics are potentially good candidates to assess chemotaxis as cells tend to accumulate in the gradient center, thus increasing the expected occurrence of cells for smaller radius near the gradient center. Notwithstanding, those methods do not report the distance to gradient center where cells are aggregating and the distances from where cells are moving from, nor allow to quantify the chemoattractant power under different concentrations or gradients. A statistical framework which considers these aspects is important.

Chemotaxis models such as Keller and Segel, (1970) or Jikeli et al., (2015) could be used to address these issues. The first is an ordinary differential equation (ODE) system that models bacterial populational chemotaxis assuming biased random walk. By fitting such model to whole data set (including time), we could estimate the chemotactic parameters and compare them between treatments, using the simplified parameterization for non-chemotaxis as null model. The models would then be compared by log-likelihood ratio. However, this approach requires initial parameter estimation that should not be easy to get for a given condition. Also the assumption of random walk of the Keller-Segel model might be too stringent for cells that swim in circles on in helices in non-chemotactic conditions.

The other method is also a system of ODEs based on Resistive Force Theory (Gray and Hancock, 1955) and a simple chemotactic system. Again, we first need to estimate some parameters to fed to the algorithm, *e.g.* the gradient being generated in order to apply this framework. These methods can be impractical in many situations as they may also require individual cell tracking and even need increased sample size. Note both approaches are feasible and here we only mention the expected difficulties, should one wish to implement either approach.

4.5 Conclusion

In this chapter, we applied linear mixed models to assess the chemotactic behavior free-swimming sperm of *L. pictus* and *S. purpuratus*. We did not find evidence that the motility parameters were affected in the presence of the chemoattractant. Should we have found such an evidence, it would be interesting to add a signaling module to the sperm model presented in the previous chapters. This model would encompass from Speract binding to the receptor, membranar channels hyper- and depolarization up to changes in internal calcium (II) concentration ($[\mathrm{Ca}^{2+}]_\mathrm{i}$) and in flagellar conformation. Recently, 3D chemotaxis was verified in *A. punctulata* and its data can be used in conjunction with the expanded morphodynamical models to provide further knowledge on spermatozoan motility.

References

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). "Fitting linear mixed-effects models using lme4". *Journal of Statistical Software* 67.(1).

Corkidi, G., Taboada, B., Wood, C. D., Guerrero, A., and Darszon, A. (2008). "Tracking sperm in three-dimensions." *Biochemical and Biophysical Research Communications* 373.(1), pp. 125–129.

Gray, J. and Hancock, G. (1955). "The propulsion of sea-urchin spermatozoa". *Journal of Experimental Biology* 32.(4), pp. 802–814.

Guerrero, A., Carneiro, J., Pimentel, J. A., Wood, C. D., Corkidi, G., and Darszon, A. (2011). "Strategies for locating the female gamete: the importance of measuring sperm trajectories in three spatial dimensions". *Molecular Human Reproduction* 17.(8), pp. 511–523.

Guerrero, A., Nishigaki, T., Carneiro, J., Yoshiro Tatsu, Wood, C. D., and Darszon, A. (2010). "Tuning sperm chemotaxis by calcium burst timing." *Developmental Biology* 344.(1), pp. 52–65.

Henderson, C. R. J. (1982). "Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions". *Biometrics* 38.(3), pp. 623–640.

Jikeli, J. F., Alvarez, L., Friedrich, B. M., Wilson, L. G., Pascal, R., Colin, R., Pichlo, M., Rennhack, A., Brenker, C., and Kaupp, U. B. (2015). "Sperm navigation along helical paths in 3D chemoattractant landscapes." *Nature Communications* 6, pp. 1–10.

Keller, E. F. and Segel, L. A. (1970). "Initiation of slime mold aggregation viewed as an instability". *Journal of Theoretical Biology* 26.(3), pp. 399–415.

Lillie, F. R. (1912). "The production of sperm iso-agglutinins by ova". *Science* 36.(929), pp. 527–530.

Pimentel, J. A. (2013). "Sistema de rastreo tridimensional de micro-partículas: aplicación en el rastreo de espermatozoides al nado libre". PhD thesis. Universidad Nacional Autónoma de México, pp. 1–120.

Pimentel, J. A., Carneiro, J., Darszon, A., and Corkidi, G. (2012). "A segmentation algorithm for automated tracking of fast swimming unlabelled cells in three dimensions." *Journal of Microscopy* 245.(Pt 1), pp. 72–81.

Ripley, B. D. (1977). "Modelling spatial patterns". *Journal of the Royal Statistical Society. Series B (Methodological)* 39.(2), pp. 172–212.

Ward, G. E., Brokaw, C. J., Garbers, D. L., and Vacquier, V. D. (1985). “Chemotaxis of Arbacia punctulata spermatozoa to resact, a peptide from the egg jelly layer”. *Journal of Cell Biology* 101.(6), pp. 2324–2329.

Chapter 5

General Discussion

In this thesis, we set out to develop automatic imaging analysis methods enriched with priors based on biological models to cope with the large imaging data sets. For this we used a mechanistic model of a spermatozoon in order to describe these cells in the synthetic and experimental imaging data of sea urchin sperm. We fitted the model parameters by maximizing their likelihood and showed we were able to obtain estimates indistinguishable from the true parameters used to generate the synthetic imaging data. Using the same procedure on experimental data led to model parameters that generated cellular positions and flagellar conformations as good as those generated by human-assisted software. Furthermore, the morphological and physical constrains imposed by the model allowed us to track the invisible flagellum using only the information of the head, showing we can estimate features absent from the imaging data using models with the sufficient detail. Taking this altogether, we validated our hypothesis that we can use a morphodynamical model of a cell and fit it directly to the imaging data by maximum likelihood estimation (MLE). The optimization process, *i.e.* to find the likelihood maximum, is the major bottleneck in our approach and probably the greatest challenge to its deployment for routine data analysis. Reducing the time the model-based image analysis method takes to produce results to values close to the time it takes to acquire the data is a desirable yet reachable objective. The optimization process involved in parameter estimation can be improved in several ways. These can include the usage of more efficient algorithms for function maximization and the usage of graphic processing units (GPUs) to accelerate both Resistive Force Theory (RFT) calculus and comparison of model and biological imaging data. In chapter 3, we fitted the mechanistic model of the spermatozoon to the reconstituted trajectories of *L. pictus* and *S. purpuratus* sperm. The parameters sets inferred in this way were distinct for different species and different swimming modes. It is tempting to argue, although not demonstrated in this thesis, that similar parameter sets would have been inferred if the model was fitted directly to the imaging data with appropriate image rendering procedures (e.g. using the three dimensions (3D) diffraction patterns instead of the point-spread function). Confirming this argument will require further studies.

An important aspect that was not addressed in this thesis was to track several cells present in the same imaging data. How to assess the correct number of cells in a time efficient way is the most challenging issue. Similarly to the piecewise helix fitting procedure developed in *Chapter 3*, we can have many model instances generated from different initial conditions (*e.g.* multiple cells in different spatial coordinates at the same time or the same cell at different initial time-points) and then select the minimum set of instances combinations that explains the data most parsimoniously using Dynamic Programming and Bayesian information criteria (BIC). Direct and indirect interaction between spermatozoa has been reported previously (Yang et al., 2008) and RFT does not account for long-range hydrodynamic forces. Thus, a physical module that take these interactions into account might be necessary to describe spermatozoa swimming in areas of higher density. In fact, failing of a model based on RFT to fit crowded image data sets might be indicative of the presence of long-range forces and the need to extend the models. In this way, such simpler models can be used as a null-hypothesis that must be rejected to support more complex scenarios.

If one would like to alter the framework presented here for different cells types, we predict the major challenge to be adaptation of the morphological module (*e.g.* rigid body cells with flagella or cilia). On another hand, cells with plastic membranar shapes (*e.g.* cells that move by filopodia) may require a different physical motility framework. It is also possible that the approach is used for data produced with other microscopy techniques, *e.g.*, bright-field or even electron microscopy, if the procedure to render the virtual image predicted by the model is modified accordingly. As these rendering steps are part of the likelihood function that has to be maximized, their quick computation is essential, as noted above. Overall, the method presented and validated in this thesis is general and is potentially adaptable to any system.

The second objective of this thesis was to develop methods to analyze sperm motility and chemotaxis in the 3D volume, specifically using the challenging Corkidi et al., (2008). One issue of this microscopy system was the high variance observed in the depth (Z) position of both frames and cells after processing the raw data. We showed on *Chapter 3* that one of the sources of variance was an occasional phase-shift of the inferred depth $Z(t)$ function, and improved the accuracy of the frames' depths by using a method based on the correlogram of the frames within a piezoelectric period. Consequently, this allowed us to obtain better estimates of the cell's 3D coordinates. We further improved the accuracy and precision of measurements of the cells' centroids by using a 3D template of the averaged spermatozoan diffraction patterns at many focal planes that are offset from the plane of the head. We had hypothesize we could use the tracking and cell characterization method developed in *Chapter 2* to study spermatozoan free-swimming motility and chemotaxis. In theory, we could have adapted the model image rendering procedure to produce the expected diffraction pattern at the focal plane at given time. If one would like to expand this method in such way, we propose modeling these patterns using Fourier or wavelets decomposition. By including a chemotactic module (*i.e.* a module that would control the deformation of the flagellum as a function of the chemoattractant concentration) to the mechanistic module the method should be able to detect and track spermatozoa in chemotactic conditions. Furthermore, it should be able to address whether certain conditions are chemotactic or not. Thus, we were able to develop imaging analysis methods that can assess motility and potentially chemotaxis in both two dimensions (2D) and 3D data.

We tracked spermatozoa from different species and characterized their trajectory using piecewise helix fitting. We revealed *L. pictus* spermatozoan trajectories have similar curvature, confined when compared to free swimming, while *S. purpuratus* sperm trajectories do not. The osculating circle of the trajectory of *S purpuratus* sperm confined to the plane is much larger than what would be expected from the free path. We proposed this effect to result from a change in the off-plane component of the flagellar beating in these sperm as compared to that of other species analyzed, which was supported by fitting the mechanistic spermatozoan model to the data. Overall, we successfully achieved the second objective.

5.1 Modeling of spermatozoa

We used models of the spermatozoan morphodynamics to estimate different parameters from either 2D and 3D imaging data. These included the position, orientation, flagellar shape, cellular drags coefficients and trajectory characteristics of a spermatozoon at a given time-point. Both methods implemented involve a mixture of model- and data-driven approaches as presented in the introduction under the light of our interpretation of Rosen's Modeling Relation (Rosen, 1991). It is worth making explicit here the elements of the Modeling Relation in the context of the modeling efforts of the previous chapters.

In *Chapter 2* we built a mathematical model based on principles that were previously shown to predict the displacement of sperm cells given the intrinsic conformation and velocity of the flagellum (Gray and Hancock, 1955; Friedrich et al., 2010). Fitting this model to imaging data we were able to estimate the shape and mechanical parameters that describe the cell. At this stage, the Rosen's Modeling Relation is still incomplete (fig. 1.1) as we only took data (mapping 2) and parameterized our model with it (mapping 3). The cycle was closed by predicting the flagellar position of a cell by our model parameterized with imaging data where only the head was marked with fluorescence (mapping 4).

In *Chapter 3* we used a motility model of sea urchin spermatozoa based on discrete helices fitting. This model was parameterized using either 2D+t or 2D+$Z(t)$ data (mapping 2) and used (mapping 3) to predict the mean values for the average paths velocity, curvature and torsion of these species when their cells were confined or in free-swimming. These values were then compared to other methods of measuring the same data (table 3.2), which agreed with our results for confined 2D+t data (*L. pictus* and *S. purpuratus*, (Guerrero et al., 2010)) and for both confined and free-swimming 2D+$Z(t)$ data (*S. purpuratus*, (Corkidi et al., 2008)). Although it allowed us to validate our method to characterize sperm motility, the model does not corroborate to independent data. Using our detailed morphodynamical model developed in *Chapter 2* to search for the conditions in which we can obtain the empirical trajectory parameters, we predicted that *S. purpuratus* spermatozoa have higher asymmetry of the flagellar off-plane component than those of *L. pictus* and *A. punctulata* when free-swimming. Measuring the flagellar torsion of these species is thus essential to close the Modeling Relation (mapping 4). Note that the non-linearity of the Cosserat frame used to model the form and deformation of the flagellum does not necessarily increase the average trajectory torsion as the flagellar one increases. Also note that, similarly, the trajectory's curvature changes non-linearly as a function of flagellar torsion. Thus, the conclusions taken here could only be obtained using a morphodynamical model. Another prediction of the model is that *S. purpuratus* has higher drag coefficients relative to other species resulting from higher interaction with boundary and/or the hydrodynamical forces generated near the interface. As this effect could be generated by higher proximity to to the boundary we can test this prediction by measuring the confinement distance from the boundary. This experiment can be easily performed with the 2D+$Z(t)$ microscopy setup using a smaller piezoelectric Z amplitude (20-50 μm) and frequency (both to reduce the Z estimation error) and ensuring that water-glass interface is within the lower limit of that interval.

It has been reported (Nosrati et al., 2015) that bull sperm show at least two different modes of confinement. Some sperm swim closest to the boundary with planar flagellar beating, while others swim more distally and their flagellar bend waves show higher off-plane components. It would be interesting to see if these different modes of confinement are also observed in sea urchin sperm and if it is correlated with the ability to display chemotactic responses. In *Chapter 3*, it was argued that *S. purpuratus* sperm would swim closer to the boundary enduring higher viscosity then the sperm of other species. If the swimmers closer to the boundary tend to be less chemotactic this would offer an potential explanation for why *S. purpuratus* sperm do not appear to be chemotactic Guerrero et al., (2010).

Although we do not expect to be the case for the marine invertebrates participating of broadcast spawning events, it is also worth to mention the existence of cooperation studies in internally fertilizing species (*i.e.* sperm conjugation). It was shown that hydrodynamic interaction promotes synchronization and attraction between sperm, also increasing the groups' velocity thus giving competitive advantage relative to lone spermatozoa (Yang et al., 2008; Elgeti et al., 2015). In the case of sea urchin, it is important to note that interaction between spermatozoa are likely to occur within dense populations, specially when performing chemotactic assays. In this particular setting, those effects should be accounted in the physical module of the morphodynamical model to take in addition these additional complications neglected in the RFT.

5.2 Advances in Imaging analysis

We introduced the different imaging analysis schemes used in research as well as commercial products, which range from simple segmentation methods to model-based methods, in which the models are mostly data-driven and based on training data. The imaging analysis methods developed in *Chapter 2* and *Chapter 3* are also model-based but of a different kind. The first method is purely based on *a priori* knowledge as it generates the expected data based on a morphodynamical model of the sperm cell, which was developed independently of the experimental images. Note that the spermatozoon is completely defined and described using a set of mathematical functions that capture the known biology and physics of the system. This difference in strategy is extremely important as using prior mechanistic knowledge allows to circumvent the need to obtain and process (likely manually) a training set; instead, we can simply define the mathematical description of the biological object in terms (variables, parameters, relationships) that are mechanistically interpretable. The second approach was more traditional in the sense that we first transformed the imaging data to other type of data (*i.e.* 3D Cartesian coordinates) to which a motility model was fitted. Note, however, that similarly to facial and hand recognition algorithms where 2D or 3D discrete templates are used (Turk and Pentland, 1991), in our work this transformation was performed using the *a priori* knowledge of the cell aspect according to the distance of the in-focus cell plane to the objective. Contrary to the first approach, the diffraction patterned templates used to detect the cells were constructed using imaging data of objects similar to the target ones. On theoretical grounds, this diffraction patterns could be computed independently of the images taking into account the geometry and physical properties of head of the sperm. Notwithstanding the meaning behind the physicists' saying *Assuming the cow is spherical...*, caution should be used to ensure that the prior descriptors are a good approximation of the target object, i.e. the model of the biological object is in fact a good model!

Recently, deep-learning algorithms have been used that implement complex thick neural networks to detect objects in images or even to describe the scene represented by an image after being trained with a set of template images. These have a enormous potential to be used to analyze images in an reproducible and unbiased way. However, these methods represent a *black box* between input and output data, meaning that the criteria they use to analyze the image after being trained is often not known to the user and have no biological interpretation per se and that might hinder critical analysis of the results (*i.e.* the mapping 3 of the Rosen's Modeling Relation is not known). Also, these supervised algorithms cannot do better than what they are trained for, thus they are not robust to different scenarios that are qualitatively different. Notwithstanding, we could potentially use these methods to measure the features of interest in the image and then feed these to a more mechanistic model in which details and inferential structure are known to us and have a biological interpretation.

The perspective of using such artificial intelligence-based methods to create software application that can develop and select the appropriate model for some data is utterly appealing. One can envisage a future where the researcher provides the image analysis software with (or directs it to) the *a priori* knowledge of interest, e.g. cell types and the imaging technique used to acquire the data. The algorithms implemented in the software would then select the best model to fit and describe the imaging data. The selection criteria must be based on statistical and mathematical developments and involve a robust pipeline. This would greatly increase the speed and volume of imaging data analysis. All chosen parameters (of the model, statistical analysis, etc.) should be properly documented in the analysis report, so it can be reproduced anytime. This will prevent less cautious researchers from using the inadequate method for well established analysis pipelines. Of course, the key feature that is determinant for scientific progress is the ability to add to the analysis new *a priori* knowledge such as a new signalling pathway, a physics model, or better shape descriptions. There is a major drawback though. Comparable to some researchers which use a DNA extraction kit without fully understanding the protocol behind the extraction of those molecules, users may resort to these high-end image analysis software without the sufficient knowledge to make critical analysis of the quality of its output . The advantage of higher analysis speed and volume might then become a curse because the experts who can assess the quality may not be sufficient to keep up with it and a wave of misleading reports may spread undetected in the scientific community.

How far are we from this integrated software? The emergence of international standards for both image analysis methods and statistical procedures (International Organization for Standardization, Geneva, Switzerland) plus the scientific articles reporting different methods or data sources provide great source of material, methods, tests and standards of imaging data analysis pipelines. However, these are not usually integrated together, only the more common (and usually outdated) procedures are. For a particular data source, a plugin. Also, many international standards and algorithms are not open-sourced or need a paid licence to be used, the reason why their integration should be commercially nonviable. Thus, a project of this magnitude can only be possible by public funding. Taking into account how its use will promote good practices in data analysis, we expect it to boost both scientific output speed and quality. Overall, we already have the components (algorithms), a motive and the urgency to do it. I think we can achieve it in the following decade.

5.3 Bringing it together

The power of mechanistic models. Using a mechanistic model that describes structures that are missing in data allowed us to expand the information given by that same data. In *Chapter 2* we were able to increase the spatial and temporal resolutions of sea urchin spermatozoa labelled with fluorescent dyes by fitting a morphodynamical model with arbitrary temporal and spacial resolutions. More notably, we were able to infer the flagellar positions and conformations by fitting to imaging data where only the head was labelled and visible. In other words, the information in the time-lapse images of the head was sufficient to allow the inferential structure of the model to decode the missing information on the flagellum. This result can be instrumental to measure internal calcium (II) concentration ($[\mathrm{Ca}^{2+}]_\mathrm{i}$) along the flagellum. Consider the following example. In order to study how $[\mathrm{Ca}^{2+}]_\mathrm{i}$ is coupled to curvature generation on the flagellum, Guerrero et al., (2013) used a light splitter in conjunction with whole cell and calcium markers (*i.e.* one of the assays used in *Chapter 2*) to image spermatozoa under different drugs (*e.g.* control and niflumic acid). This strategy allowed to track the flagellum in the series of images and then overlay it with the signal of the Ca^{2+} reporter to quantify intensity of the influx of this cation. Since the sperm heads are constitutively labeled by the calcium indicator (Guerrero et al., 2010; Guerrero et al., 2013) it would be possible to the morphodynamical model (or any of the suggested extensions) to infer the flagellar positions and conformations directly from this data.

Approximately one of every six couples in subfertile, making human fertility a subject of growing medical and economic importance (Gaffney et al., 2011). Male factors account for around half of the cases and sperm motility is a major player (Barratt et al., 2009). A significant fraction of the infertile cases have unknown cause. It is an intriguing possibility that these can be explained by aberrations at higher-level control of sperm motility and orientation by environmental cues, such as chemotaxis, which cannot be diagnosed by current Computer Assisted Sperm Analysis (CASA) systems that measure low level properties such as progressive rate and beating frequencies. In fact, one of the most remarkable aspects of mammalian fertilization process is the journey of the sperm to the egg. Sperm cells must swim a path thousands of times their own body length through a complex interior geometry, often filled with highly viscous and potentially hostile immune cells. Initially over hundreds of millions, the overwhelming majority do not even reach the Fallopian tubes, let alone the site of fertilization (Gaffney et al., 2011). Any aberration in the spatial-temporal coordination of flagellar beating dynamics with spatial cues will prevent the sperm to find its way and therefore decrease the fertility rate. The extension of the methods developed in this thesis to model the human sperm is rather straightforward. Several adaptations can be foreseen. The shape module should account for the different rigidity, length and width of the flagellar midpiece. As these cells swim within an intricate landscape, the physical module should model the boundary explicitly as should the effect of viscosity and fluid interaction with the flagellar shape. Finally, a module describing the signalling transduction of the chemoattractant signals down to changes in $[Ca^{2+}]_i$ should also be added if one wishes to account for chemotaxis. By applying such *humanized* model as *a priori* knowledge and performing the appropriate assays, one could eventually measure features that have been omitted in sperm motility and chemotaxis analysis. Eventually, these would provide some insights on some of the unknown causes of fertility and allow for future treatments to be developed.

2D+$Z(t)$ system for chemotaxis analysis. Sperm chemotaxis is an essential process for the life cycle of many species. As per definition, chemotaxis depends on how cells reorient themselves and spermatozoa do it by modulating the asymmetry in the flagellar bending wave curvature (and perhaps torsion), which is correlated with the derivative of whole cell $[Ca^{2+}]_i$ (Wood et al., 2003; Alvarez et al., 2012). In *Chapter 1* we referred part of an extensive literature on how and why different membranar channels (*i.e.* either at the plasma or at mitochondrial membranes) operate in order to modulate $[Ca^{2+}]_i$ after chemotactic stimulus. It is clear from these studies that it is crucial to understand calcium dynamics in spermatozoa in order to understand their motility and, thus, chemotaxis. A microscopy system which is able to image calcium marker-loaded cells in 3D becomes the obvious choice to study chemotaxis and the 2D+$Z(t)$ system (Corkidi et al., 2008) has potentially such capabilities. The major challenge that must be overcome to use this system to that purpose is to improve the fluorescence detection as the acquisition exposure time must be very short, due to beating frequency of the flagella that requires a rapid movement of the piezoelectric device. Using small amplitudes of the device to capture a single individual will greatly decrease the difficulty. The computational analysis methods developed here demand minor adjustment to be able to deal with 4D+t imaging data that is expected to be generated from such system in the coming years. In particular, similar to the analysis of the planar motility in *Chapter 2*, tracking of the cell in 3D using our mechanistic model would allow to measure the calcium without extra information of the flagellar position. This would benefit, obviously, if the model is calibrated and validated with measurement of the 3D flagellar beating conformations before deploying it to the task of 4D+t image analysis.

Holographic microscopy systems based on coherent light (Su et al., 2012) are not able to measure fluorescence, which is an incoherent light source. Incoherent holography has been developed since the late 00's and has been used to acquire biological data (Rosen and Brooker, 2008). However, because of the scattering of photons emitted from the fluorescence source point, the signal-to-noise ratio (SNR) of the reconstructed image is low and therefore this technique requires immobile or fixed material. There is active development of reconstruction algorithms that improve this SNR but the best they achieve is a SNR$\sim$2 for a total exposure time of 1.5 ms (Jang et al., 2016). As fluorescent imaging of flagellar $[Ca^{2+}]_i$ and intracellular pH signals requires photomultipliers (Guerrero et al., 2010; González-Cota et al., 2015; Jansen et al., 2015), the SNR is effectively lower, implying that these signals cannot be resolved. This means holographic fluorescent microscopy will have to mature before it can be applied to the 3D study of sperm motility and chemotaxis. Until these limitations of incoherent holography are overcome, the Corkidi's piezoelectric-driven microscopy system or similar instrumentation seems to be safest bet on the near future. This 3D microscopy system is also able to image spermatozoa at high spatial resolution, inclusive to obtain the position of the flagellum (Silva-Villalobos et al., 2014). Being able to use calcium indicators with high spatial and temporal resolution of the flagellum is an important step towards comprehension of chemotaxis. However, we'll loose spatial resolution if we take the approach of using the light splitter to obtain the whole cell and the calcium markers in the same detector (*i.e.* two light channels are scaled down so they can be recorded simultaneously by the same hardware piece). We can use the alternative that we already mentioned of using a mechanistic model to fit the 3D morphodynamical model to the head data and discard the need of the whole cell marker, allowing us to use the full chip for the calcium indicator thus increasing the resolution.

The present thesis, by overcoming the earlier shortcomings of the 2D+$Z(t)$ data analysis methods and by making the proof-of-principle that mechanistic models of the spermatozoon can be deployed for more better image analysis, set the ground to enable this new experimental avenue into sperm chemotaxis.

References

Alvarez, L., Dai, L., Friedrich, B. M., Kashikar, N. D., Gregor, I., Pascal, R., and Kaupp, U. B. (2012). "The rate of change in Ca 2+ concentration controls sperm chemotaxis". *Journal of Cell Biology* 196.(5), pp. 653–663.

Barratt, C. L. R., Kay, V., and Oxenham, S. K. (2009). "The human spermatozoon - a stripped down but refined machine." *Journal of biology* 8.(7), p. 63.

Corkidi, G., Taboada, B., Wood, C. D., Guerrero, A., and Darszon, A. (2008). "Tracking sperm in three-dimensions." *Biochemical and Biophysical Research Communications* 373.(1), pp. 125–129.

Elgeti, J., Winkler, R. G., and Gompper, G. (2015). "Physics of microswimmers - single particle motion and collective behavior: a review". *Reports on Progress in Physics* 78.(056601), pp. 1–50.

Friedrich, B. M., Riedel-Kruse, I. H., Howard, J., and Jülicher, F. (2010). "High-precision tracking of sperm swimming fine structure provides strong test of resistive force theory." *The Journal of experimental biology* 213.(Pt 8), pp. 1226–34.

Gaffney, E. A., Gadêlha, H., Smith, D. J., Blake, J. R., and Kirkman-Brown, J. C. (2011). "Mammalian sperm motility: Observation and Theory". *Annual Review of Fluid Mechanics* 43.(1), pp. 501–528.

González-Cota, A. L., Silva, P. Â., Carneiro, J., and Darszon, A. (2015). "Single cell imaging reveals that the motility regulator speract induces a flagellar alkalinization that precedes and is independent of Ca 2+ influx in sea urchin spermatozoa". *FEBS letters* 589, pp. 2146–2154.

Gray, J. and Hancock, G. (1955). “The propulsion of sea-urchin spermatozoa”. *Journal of Experimental Biology* 32.(4), pp. 802–814.

Guerrero, A., Espinal, J., Wood, C. D., Rendón, J. M., Carneiro, J., Martínez-Mekler, G., and Darszon, A. (2013). “Niflumic acid disrupts marine spermatozoan chemotaxis without impairing the spatiotemporal detection of chemoattractant gradients.” *Journal of cell science* 126.(Pt 6), pp. 1477–87.

Guerrero, A., Nishigaki, T., Carneiro, J., Yoshiro Tatsu, Wood, C. D., and Darszon, A. (2010). “Tuning sperm chemotaxis by calcium burst timing.” *Developmental Biology* 344.(1), pp. 52–65.

Jang, C., Clark, D. C., Kim, J., Lee, B., and Kim, M. K. (2016). “Signal enhanced holographic fluorescence microscopy with guide-star reconstruction”. *Biomedical Optics Express* 7.(4), pp. 1271–1283.

Jansen, V., Alvarez, L., Balbach, M., Strünker, T., Hegemann, P., Kaupp, U. B., and Wachten, D. (2015). “Controlling fertilization and cAMP signaling in sperm by optogenetics”. *eLife* 2015.(4), pp. 1–15.

Nosrati, R., Driouchi, A., Yip, C. M., and Sinton, D. (2015). “Two-dimensional slither swimming of sperm within a micrometre of a surface.” *Nature communications* 6, p. 8703.

Rosen, J. and Brooker, G. (2008). “Non-scanning motionless fluorescence three-dimensional holographic microscopy”. *Nature Photonics* 2, pp. 190–195.

Rosen, R. (1991). “Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life”. Columbia University Press.

Silva-Villalobos, F., Pimentel, J. A., Darszon, A., and Corkidi, G. (2014). “Imaging of the 3D dynamics of flagellar beating in human sperm.” *Conference Proceedings IEEE Engineering in Medicine and Biology Society* 2014, pp. 190–193.

Su, T.-W., Xue, L., and Ozcan, A. (2012). “High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories”. *Proceedings of the National Academy of Sciences* 109.(40), pp. 3–7.

Turk, M. and Pentland, A. (1991). “Eigenfaces for Recognition”. *Journal of Cognitive Neuroscience* 3.(1), pp. 71–86.

Wood, C. D., Darszon, A., and Whitaker, M. (2003). “Speract induces calcium oscillations in the sperm tail”. *The Journal of Cell Biology* 161.(1), pp. 89–101.

Yang, Y., Elgeti, J., and Gompper, G. (2008). “Cooperation of sperm in two dimensions: Synchronization, attraction, and aggregation through hydrodynamic interactions”. *Physical Review E* 78.(6), pp. 1–9.

Apoio financeiro da FCT e do FSE no âmbito do Quadro Comunitário de Apoio, Bolsa nº SFRH/BD/79261/2011.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt