Eugen Czeizler | Andrzej Mizera | Elena Czeizler | Ralph-Johan Back | John E. Eriksson | Ion Petre

Quantitative analysis of the self-assembly strategies of intermediate filaments from tetrameric vimentin

Turku Centre *for* Computer Science

TUCS Technical Report
No 963, December 2009

Quantitative analysis of the self-assembly strategies of intermediate filaments from tetrameric vimentin

Eugen Czeizler
Department of Information Technologies, Åbo Akademi University,
(eczeizle@abo.fi)
Andrzej Mizera
Department of Information Technologies, Åbo Akademi University,
(amizera@abo.fi)
Elena Czeizler
Department of Information Technologies, Åbo Akademi University,
(elena.czeizler@abo.fi)
Ralph-Johan Back
Department of Information Technologies, Åbo Akademi University,
(backrj@abo.fi)
John E. Eriksson
Turku Centre for Biotechnology and Department of Biochemistry,
Åbo Akademi University,
(john.eriksson@btk.fi)
Ion Petre
Department of Information Technologies, Åbo Akademi University,
(ipetre@abo.fi)

TUCS Technical Report

No 963, December 2009

Abstract

In vitro assembly of intermediate filaments from tetrameric vimentin consists of a very rapid phase of tetramers laterally associating into unit-length filaments and a slow phase of filament elongation. We focus in this paper on a systematic quantitative investigation of two molecular models for filament assembly, recently proposed in (Kirmse et al *J. Biol. Chem. 282*, 52 (2007), 18563–18572), through mathematical modeling, model fitting, and model validation. We analyze the quantitative contribution of each filament elongation strategy: with tetramers, with unit-length filaments, with longer filaments, or combinations thereof. In each case, we discuss the numerical fitting of the model with respect to one set of data, and its separate validation with respect to a second, different set of data. We introduce a high-resolution model for vimentin filament self-assembly, able to capture the detailed dynamics of filaments of arbitrary length. This provides much more predictive power for the model, in comparison to previous models where only the mean length of all filaments in the solution could be analyzed. We show how kinetic observations on low-resolution models can be extrapolated to the high-resolution model and used for lowering its complexity.

Keywords: Mathematical modeling — Protein self-assembly — Quantitative self-assembly strategies — Model resolution — Sensitivity analysis — Filament length distribution.

TUCS Laboratory
Computational Biomodelling

1 Introduction

The cytoskeleton of eukaryotic cells is an intricate network of protein filaments that extends throughout the cytoplasm. There are three types of protein filaments: *intermediate filaments* (IFs), *microtubules*, and *actin filaments*, [24]. Together with other proteins that attach to them, they form a system of girders, ropes, and motors that gives the cell its mechanical strength, controls its shape, and drives and guides its movements, see [17]. Compared with microtubules and actin filaments, IFs are more stable, tough and durable; in particular, IFs are the most insoluble part of the cell, see [8]. IFs have an important structural function in reinforcing the cells, organize cells into tissues, and most importantly, distribute the tensile forces across the cells in a tissue, see [17]. Major degenerative diseases of skin, muscle, and neurons are caused by disruptions of the IF cytoskeleton or its connections to other cell structures. Currently, around 80 diseases have been associated with the IF gene family, including various skin fragility disorders, as well as *laminopathies*, a family of afflictions caused by point mutations in the lamin A genes, [4, 5, 26]. A thorough understanding of the assembling principles of IFs can provide new insights on comprehending these abnormal conditions, as well as a better basis for diagnostic and possible treatment.

Contrary to the other protein filaments which are assembled from globular proteins, see [11, 25, 22], IFs subunits are α-helical rods that assemble into rope-like filaments [8]. Their assembly proceeds through a series of intermediate structures, which associate by lateral and end-to-end interactions. However, unlike in the case of microtubules and actin filaments where rich literature is available, the assembly principles of IFs are still poorly understood. We focus in this paper on the quantitative kinetic strategies for the *in vitro* assembly of IFs from human vimentin proteins (several other IF proteins exist, see [10]). On a first level of their assembly, vimentin proteins rapidly associate parallelly into dimers and then form anti-parallel, half-staggered tetramers, see [9] and Figure 1 (a)-(e). Tetramers then rapidly associate laterally to yield short filaments called *unit-length filaments* (ULFs) of the same length as the tetramers, see [8] and Figure 1 (f). On a second level of the assembly, the ULFs and the emerging longer filaments elongate longitudinally with tetramers, with ULFs, and with other filaments, [8] and Figure 2. On a third level, filaments undergo a radial compaction from an ULF diameter of about 15 nm to a filament diameter of about 11 nm, see [8] for details.

We investigate in this paper two molecular models (the so-called *simple* and *extended* models) introduced in [15] for the *in vitro* assembly of intermediate filaments from tetrameric vimentin. We perform a quantitative analysis of the predictive capabilities of these models. We construct two mass action-based mathematical models corresponding to the two molecular models. For each of them we consider several different knockdown mutant model variants where various combinations of assembly mechanisms are analyzed separately. We use COPASI [12] as a computational environment for the experimental data fitting (based on data

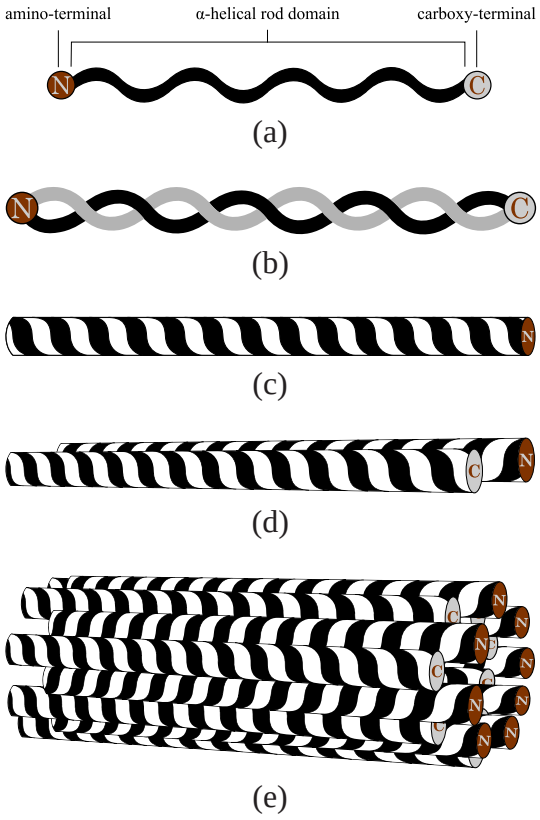


Figure 1: The first stage in the assembly of human vimentin proteins. Intermediate filament subunits are α-helical rods, that associate parallelly into coiled-coil dimers, which in turn form anti-parallel, half-staggered tetramers. Tetramers rapidly associate laterally to yield the shortest filaments called *unit-length filaments* (ULFs) of the same length as the tetramers. (a) α-helical rods, (b) coiled-coil dimer, (c) another representation of a coiled-coil dimer, (d) tetramer, (e) ULF.

of [15] and [14]), the model validation, and the sensitivity analysis. Our approach for the numerical analysis of the models differs markedly from that of [15], see Section 4 for a discussion.

Our study provides several conclusions regarding the kinetics of the *in vitro* assembly of human vimentin. On one hand, we show that the filament elongation process requires the end-to-end annealing of filaments as one of its features, which is in agreement with the results of [15]. Indeed, in all of our models where this reaction was missing, either the model did not fit the experimental data or the model was rejected in the validation round. Moreover, in almost all cases where the reaction modeling the end-to-end annealing of filaments is present, its rate constant is estimated to roughly the same value, although the other kinetic constants differ from model to model. On the other hand, the quantitative contribution of the filament elongation with tetramers depends on the turnover rate of tetramers into unit length filaments. If tetramers are quickly depleted from the system, e.g., through

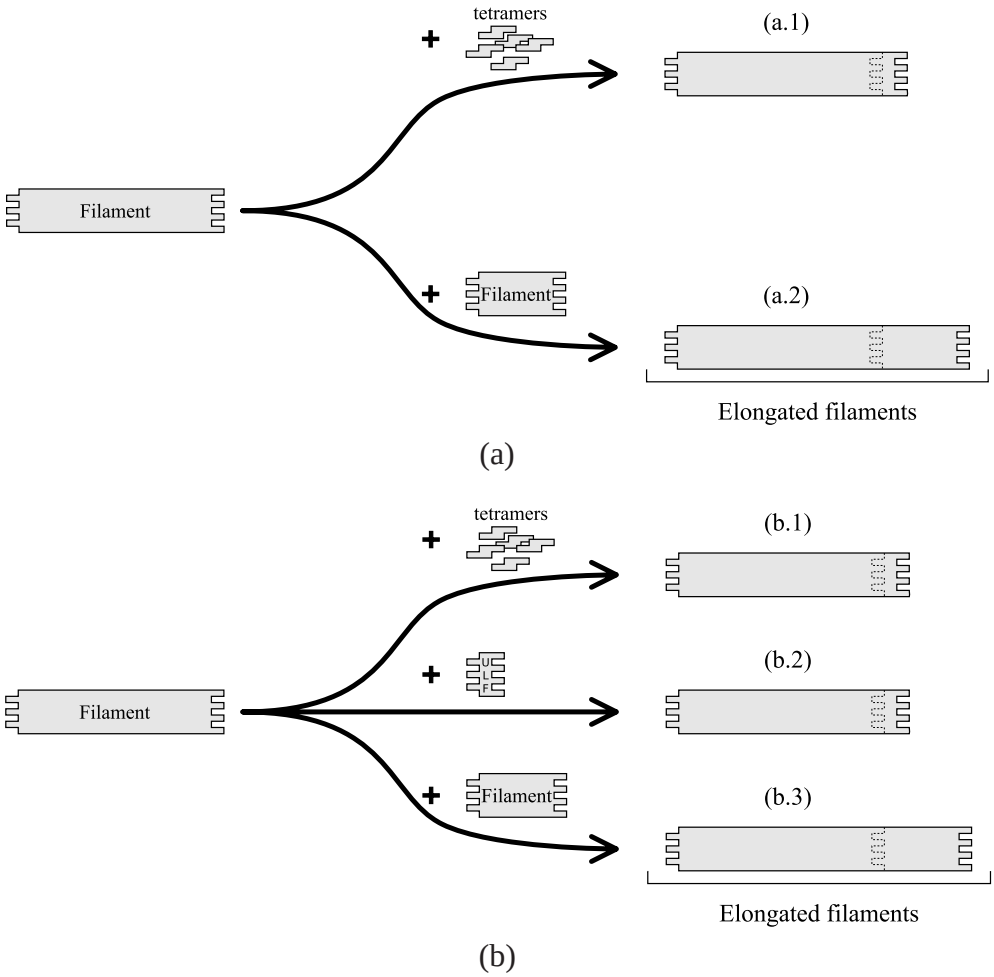


Figure 2: The two molecular models of the *in vitro* assembly of vimentin IF tetramers. (a) In the *simple model* filaments undergo elongation either by (a.1) longitudinal association of tetramers or (a.2) by end-to-end annealing of another filament. (b) The *extended model* adds a distinction between minimal-length filaments (ULFs) and longer filaments (consisting of at least 2 ULFs). In this case, there is one extra possibility for filament elongation: (b.1) by tetramer, (b.2) by the longitudinal association of a ULF, and (b.3) by another filament.

a high tetramer-to-ULF turnover rate as documented in *in vitro* experiments of [15], then only one of eight possible assembly strategies correlates well with the available experimental data, in agreement with conclusions of [15]. If free tetramers are however available throughout the assembly, then we show that several different assembly strategies correlate similarly well with the experimental data.

One of the modeling challenges identified in [15] was to increase the resolution of the model: instead of collecting all filaments into a single variable, regardless of their length, one should describe separately the dynamics of filaments of various lengths, at least up to a certain fixed, but arbitrarily high length, that we

call the resolution of the model. Indeed, the quantitative experimental data of [15] captures the levels of filaments of various lengths, but the data is only used in [15] to calculate the mean length of all filaments in the solution. We provide in this paper a generic solution to this problem, demonstrating how to enhance the existing filament assembly models with the dynamics of the filament length distribution. Our enhanced model can have arbitrarily high resolution, being able to capture the dynamics of filaments of arbitrarily high length. The size of this detailed model is considerably higher than that of the basic model, both in terms of molecular species, as well as in terms of molecular reactions. Based on kinetic observations on the basic model, we show however how the size of the high-resolution model can be drastically reduced. Our approach towards high-resolution models for protein self-assembly is independent of the particulars of vimentin filaments and can be applied to other instances of protein-protein interactions and protein assemblies.

2 Models and methodology

2.1 Two molecular models for the assembly of vimentin IFs

The *in vitro* assembly of vimentin IF proteins consists of three major phases, see [10]: (i) formation of the unit-length filaments (ULF) structures; (ii) longitudinal annealing of ULFs and growing filaments; (iii) radial compaction of immature filaments into mature IFs. We consider here two molecular models for this process, originally introduced in [15]. Both of them focus on the first two phases of the assembly, ignoring the third.

The *simple model* of [15] treats ULFs as ordinary filaments and describes the assembly process through a sequence of molecular events as follows, see also Figure 2 (a):

(i) two tetramers (denoted T) associate laterally into an octamer (denoted O):

$$2\,\mathrm{T} \to \mathrm{O}; \tag{1}$$

(ii) two octamers associate laterally to yield a hexadecamer (denoted H):

$$2\,\mathrm{O} \to \mathrm{H} \tag{2}$$

(iii) two hexadecamers associate laterally to form a (unit length) filament (denoted F):

$$2\,\mathrm{H} \to \mathrm{F} \tag{3}$$

(iv) a tetramer associates longitudinally to a filament to yield an elongated filament:

$$\mathrm{F} + \mathrm{T} \to \mathrm{F}; \tag{4}$$

(v) two filaments associate longitudinally to yield an elongated filament:

$$\mathrm{F} + \mathrm{F} \rightarrow \mathrm{F}\,. \tag{5}$$

The *extended model* of [15] adds a distinction between minimal-length filaments (ULFs, denoted U) and longer filaments (consisting of at least two ULFs), treating them as distinct species in the model, see Figure 2 (b). In terms of molecular events, the extended model consists of the following reactions:

(i') two tetramers (denoted T) associate laterally into an octamer (denoted O):

$$2\,\mathrm{T} \rightarrow \mathrm{O}; \tag{6}$$

(ii') two octamers associate laterally to yield a hexadecamer (denoted H):

$$2\,\mathrm{O} \rightarrow \mathrm{H} \tag{7}$$

(iii') two hexadecamers associate laterally to form a unit length filament (denoted U):

$$2\,\mathrm{H} \rightarrow \mathrm{U} \tag{8}$$

(iv') two unit length filaments associate longitudinally to form an elongated filament (denoted F):

$$2\,\mathrm{U} \rightarrow \mathrm{F} \tag{9}$$

(v') a filament is elongated longitudinally with a tetramer:

$$\mathrm{F} + \mathrm{T} \rightarrow \mathrm{F} \tag{10}$$

(vi') a filament is elongated longitudinally with a unit length filament:

$$\mathrm{F} + \mathrm{U} \rightarrow \mathrm{F} \tag{11}$$

(vii') two filaments associate longitudinally to yield an elongated filament:

$$\mathrm{F} + \mathrm{F} \rightarrow \mathrm{F} \tag{12}$$

2.2 Mathematical models

We consider a mathematical formulation of the simple and the extended models for IF assembly based on the mass-action law, where each molecular species is represented by a continuous non-negative real function denoting its concentration

in time. The system of differential equations corresponding to the simple model is the following:

$$d[\mathrm{T}]/dt = -2k_1^s[\mathrm{T}]^2 - k_t^s[\mathrm{T}][\mathrm{F}] \tag{13}$$
$$d[\mathrm{O}]/dt = k_1^s[\mathrm{T}]^2 - 2k_2^s[\mathrm{O}]^2 \tag{14}$$
$$d[\mathrm{H}]/dt = k_2^s[\mathrm{O}]^2 - 2k_3^s[\mathrm{H}]^2 \tag{15}$$
$$d[\mathrm{F}]/dt = k_3^s[\mathrm{H}]^2 - k_f^s[\mathrm{F}]^2 \tag{16}$$

where $k_1^s, k_2^s, k_3^s, k_t^s, k_f^s$ are the kinetic rate constants of reactions (1)-(5), respectively.

The mathematical model corresponding to the extended model consists of the following system of differential equations:

$$d[\mathrm{T}]/dt = -2k_1^e[\mathrm{T}]^2 - k_t^e[\mathrm{T}][\mathrm{F}] \tag{17}$$
$$d[\mathrm{O}]/dt = k_1^e[\mathrm{T}]^2 - 2k_2^e[\mathrm{O}]^2 \tag{18}$$
$$d[\mathrm{H}]/dt = k_2^e[\mathrm{O}]^2 - 2k_3^e[\mathrm{H}]^2 \tag{19}$$
$$d[\mathrm{U}]/dt = k_3^e[\mathrm{H}]^2 - 2k_4^e[\mathrm{U}]^2 - k_u^e[\mathrm{U}][\mathrm{F}] \tag{20}$$
$$d[\mathrm{F}]/dt = k_4^e[\mathrm{U}]^2 - k_f^e[\mathrm{F}]^2 \tag{21}$$

where $k_1^e, k_2^e, k_3^e, k_4^e, k_t^e, k_u^e, k_f^e$ are the kinetic rate constants of reactions (6)–(12), respectively.

An interesting aspect here is that the mass conservation relation on the total number of tetramers in the model is evident in the molecular models (since there is no synthesis and no degradation in the model), whereas it cannot be deduced as a property of either of the two corresponding mathematical models. This is a consequence of how, for example, the longitudinal association of two filaments is modeled: the information about the lengths of the two input filaments is not explicitly reproduced in a property of the two filaments. One can however calculate the number of tetramers integrated in the assembled filaments, as we do in Section 2.3, and then use this quantity to reason about the time-dependant dynamics of the mean filament length (MFL). We relate MFL to the experimental data of [15] and discuss the numerical fit of the models in Section 3.

2.3 Calculating the mean filament length

Relating the models proposed in the previous section for IF assembly to the quantitative data on the dynamics of the filament length is non-trivial because the two models do not represent explicitly the information about the length of the emerging filaments. Indeed, both models collect all filaments into a single variable (F), regardless of their length. We show however in this section that the dynamics of the mean filament length can in fact be deduced based on the variables of the two models.

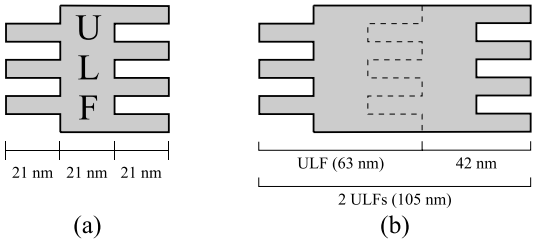


Figure 3: (a) The unit-length filament is approximately 63 nm long ([2]). (b) However, each ULF associated longitudinally at the end of an existing filament (or ULF) elongates it by approximately 42 nm ([2]). This is due to the interdigitation by which two ULFs anneal longitudinally.

During the process of ULFs aggregation atomic force microscopy (AFM) shows that each ULF associated longitudinally at the end of an existing filament adds to the length of that filament less than the stand-alone length of a ULF, see [2]. In the model for vimentin assembly of [2] this is due to interdigitation of the ULF and the filament to each other, see Figure 3. The stand-alone unit-length filament is approximately 63 nm long ([2]), while each additional ULF elongates a filament by approximately 42 nm ([2]).

We denote by $L_m(t)$ the time-dependent expression for the mean filament length (MFL) at time t. We also denote by $\#T_F(t)$ the total number of all tetramers integrated in the assembled filaments at time t. Since we consider two categories of filaments, U and F, we obtain that

$$L_m(t) = \frac{l_F(t) + l_U(t)}{\#F(t) + \#U(t)}, \tag{22}$$

where $l_F(t)$ and $l_U(t)$ denote the total length of all filaments and the total-length of all ULFs at time t, while $\#F(t)$ and $\#U(t)$ denote the total number of all filaments and that of all ULFs, respectively. Since in each filament the first ULF accounts for $l_{ULF} \simeq 63$ nm of the total length of that filament and all the additional ULFs elongate the filament by $l_{addULF} \simeq 42$ nm, we have that

$$\begin{aligned} l_F(t) &= (\#U_F(t) - \#F(t)) \cdot l_{addULF} + \#F(t) \cdot l_{ULF} \\ &= \#U_F(t) \cdot l_{addULF} + \#F(t) \cdot (l_{ULF} - l_{addULF}), \end{aligned}$$

where $\#U_F(t)$ denotes the the total number of all ULFs in all filaments, in time. Since ULFs consist on average of eight tetramers, we have that

$$\#U_F(t) = \frac{\#T_F(t)}{8},$$

where $\#T_F(t)$ is the number of tetramers already assembled into filaments.

We denote by c_0 the initial molar concentration of all tetramers in the system (occurring in any of the molecular species of the model: tetramers, octamers,

hexadecamers, ULFs, or filaments). Then, in the case of the extended model we obtain

$$\begin{aligned}\#T_F(t) = (c_0 - [T](t) - 2\,[O](t) - 4\,[H](t)\\ - 8\,[U](t)) \cdot N_A \cdot V,\end{aligned}$$

where N_A is the Avogadro constant and V is the volume of the system. Thus, (22) becomes

$$\begin{aligned}L_m(t) =& \frac{\frac{c_0 - [T](t) - 2\,[O](t) - 4\,[H](t) - 8\,[U](t)}{8} \cdot l_{addULF}}{([F](t) + [U](t))}\\ &+ \frac{[F](t) \cdot (l_{ULF} - l_{addULF}) + l_{ULF} \cdot [U](t)}{([F](t) + [U](t))}.\end{aligned}$$

In the case of the simple model, we obtain that

$$\#T_F(t) = (c_0 - [T](t) - 2\,[O](t) - 4\,[H](t)) \cdot N_A \cdot V.$$

Thus, (22) becomes

$$\begin{aligned}L_m(t) =& \frac{\frac{c_0 - [T](t) - 2\,[O](t) - 4\,[H](t)}{8} \cdot l_{addULF}}{[F](t)}\\ &+ (l_{ULF} - l_{addULF}).\end{aligned}$$

Since the volume V of the considered system does not change, the molar concentrations are expressed simply in terms of micromoles (without reciprocal of the volume unit) in the continuation.

2.3.1 Experimental data and model fitting

For the parameter estimations and model validations we used the experimental data from [14] on the *in vitro* assembly process of recombinant vimentin at $37\,^{\circ}\mathrm{C}$. The data consists of two sets, each containing the length distributions of growing filaments at distinct time points up to 20 min. The data sets were obtained by adsorption of the filaments onto carbon-coated copper grids and measurements of the filament lengths from images recorded with electron microscopy (EM) in two cases: when the initial amount of tetramers was $0.45\ \mu$M and $0.9\ \mu$M. For each set the time-dependent mean filament length (MFL) was calculated. The MFL values together with the 0.95 confidence intervals are presented in Table 1. For detailed description of experimental procedures and discussion on the independence of the measured MFLs from the support medium we refer to [15].

For fitting our mathematical models, we used the MFL data obtained for an initial tetramer concentration of $0.45\ \mu$M. For model validation, we then compared the numerical prediction for the mean filament length with the experimental data in Table 1 for an initial tetramer concentration of $0.9\ \mu$M.

Table 1: Measurements on the mean filament length of vimentin protein IFs, based on EM data of [14] (data in [nm]); a preliminary version of the data (containing a few minor errors) is in [15].

Time [s]	Initial molar concentration of all tetramers (c_0)	
	0.45 μM	0.9 μM
10	65.1±1.4	62.8±2.1
20	68.2±2.0	
30	76.5±2.1	84.1±2.0
60	112.9±4.0	131.4±5.2
180	172.6±8.4	
300	233.0±10.0	289.1±15.8
600	320.7±18.5	418.6±24.7
900		544.1±34.8
1200	474.9±37.2	821.3±41.5

We set the initial molar concentrations of all molecular species other than tetramers to 0, based on the setup of the experimental assays. Thus, there remained to be estimated five independent parameters (rate constants k_1^s, k_2^s, k_3^s, k_t^s and k_f^s) for the simple model and seven of them (rate constants k_1^e, k_2^e, k_3^e, k_4^e, k_t^s, k_u^e and k_f^e) for the extended model. Parameter estimations were performed in COPASI [12].

We also considered a qualitative property of the IF assembly, reported in [15]: very quickly (within approximately 10 seconds) after the initiation of the assembly, ULF is the most predominant species in the system, while tetramers are depleted. This observation only applied for the *ab initio in vitro* assembly of intermediate filaments. The dynamics could however be very different if more free tetramers were available for longer throughout the assembly (e.g., through an additional tetramer synthesis mechanism). To test it, we considered two different strategies for fitting our models: one where the tetramer-to-ULF turnover is fast, and another where it is slow. While the latter setup does not mimic the presence of a tetramer synthesis mechanism (introducing one would make it difficult to compare the models), it does allow us to analyze the system in the case where tetramers are available for a longer period for the assembly. We demonstrate in the next section that the two situations are indeed very different, in terms of which filament elongation mechanisms (with tetramers, with ULFs, or with other filaments) can explain the available experimental data.

The problem of estimating the parameters of computational models in systems biology is difficult, see e.g., [3, 20, 21]. This problem can be formulated as a minimization of a cost function which quantifies the differences between the values predicted by the model and the experimental measurements. There are numerous methods, both local and global, which can be used to tackle this problem, each with its own advantages and disadvantages. For instance, while local methods work faster to find a solution, they tend to converge to local optima. On the other

hand, global optimization methods are typically slower, but they tend to converge to a global optimum. The global optimization methods can be further divided into deterministic [6, 13] and stochastic approaches [1, 7]. Although the deterministic methods guaranty the convergence to a global optimum, they cannot ensure the termination of this process within a finite time interval [21]. On the other hand, the inherent randomness of the stochastic approaches makes it very hard to guaranty that these methods actually converge to the global optimum [21]. However, many stochastic methods are capable of locating the vicinity of global solutions with relative efficiency, i.e. they provide a very good approximation of the solution in acceptable computation time [21]. This makes the stochastic global optimization methods to be usually preferred for parameter estimation problems. We chose COPASI, [12], as a computational environment for parameter fitting since it includes a number of various optimization algorithms, searching for either local or global optimum values, see e.g., [19, 23]. This software is a widely used tool in the computational systems biology modeling community, having a documented good performance, see e.g. [3, 20, 21]. In particular, for determining the best numerical fits of our models, a suite of various global, stochastic parameter estimation procedures was used, comprising of methods such as Simulated Annealing, Genetic Algorithm, Evolution Strategy using Stochastic Ranking, and Particle Swarm. All these methods use specific strategies for sampling the parameter space looking for combinations of parameter numerical values that give better and better fits of the model predictions to the experimental data.

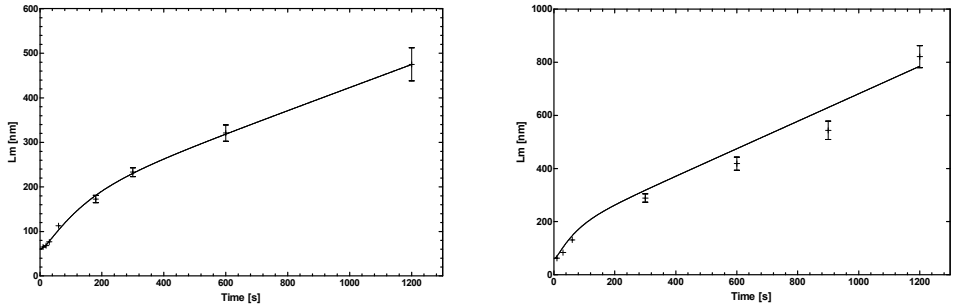
The fit of a model was performed by searching for a set of parameter values that minimizes the sum of squared deviations SS_f of the values predicted by the model from the 0.45 μM experimental data. The validation of a fitted model was performed by numerically simulating the model and by computing the sum of squared deviations SS_v of the values predicted by the model from the 0.9 μM experimental data. Moreover, the quality of the fit/validation for each model was estimated by a dimensionless number expressing the deviation of the model from the experimental data, normalized by the mean of the predicted values. This method for estimating the quality of model fit/validation was originally proposed in [16] and it allows for comparison of different models and different data sets. The formula for the quality of the fit (fq) is:

$$fq = \frac{\sqrt{SS_f/N_f}}{\text{mean of predicted values}} \cdot 100\%, \tag{23}$$

where N_f is the number of 0.45 μM experimental data points (in our case $N_f = 8$). Similarly, the formula for the quality of the validation (vq) is:

$$vq = \frac{\sqrt{SS_v/N_v}}{\text{mean of predicted values}} \cdot 100\%, \tag{24}$$

where N_v is the number of 0.9 μM experimental data points (in our case $N_v = 7$). It was argued in [16] that a low (say, lower than 15%) value of fq (vq) was con-



(a) (b)

Figure 4: Time-dependent MFL growth corresponding to the simple model without the quick filament formation requirement. (a) The model fit with respect to the EM 0.45 μM experimental data set. (b) Model validation based on the EM 0.9 μM experimental data set. The continuous line is the model prediction regarding $L_m(t)$, that is compared with the experimental data showed with crossed points. The short vertical lines represent the 0.95 confidence intervals for the experimental data.

Table 2: Kinetic rate constant values in $\mu\text{M}^{-1}s^{-1}$ for the simple model.

k_1^s	k_2^s	k_3^s	k_t^s	k_f^s
$3.39 \cdot 10^{-3}$	30	30	0.83	0.11

sidered as an indicator of a successful fit (validation). We discuss the numerical values of fq and vq for all our models in Section 3.

3 Results

3.1 Data fitting the simple model

The kinetic rate constants in Table 2 yield an excellent fit ($fq = 2.52\%$) of the simple model for the experimental data from the assay with 0.45 μM tetramers and a good validation ($vq = 12.07\%$) of the model when compared with the data from the assay with 0.9 μM initial concentration of tetramers, see Figure 4.

This model however could not confirm the quick turnover of tetramers into filaments. When this condition was taken into consideration by searching for relatively high numerical values of k_1^s, k_2^s, and k_3^s (higher than 1 $\mu M^{-1}s^{-1}$), the fit of the model to the experimental data was unsuccessful ($fq = 26.00\%$), despite numerous rounds of parameter estimation. The following mathematical argument is also indicating that this model cannot be given a reasonable fit. Based on the observation that tetramers are quickly depleted (within 10 seconds) by turning them into ULFs, the model can be split into two processes separated in time: first, the formation of filaments from tetramers, i.e. $2\,\text{T} \rightarrow \text{O}$, $2\,\text{O} \rightarrow \text{H}$, $2\,\text{H} \rightarrow \text{F}$,

and second, the elongation of filaments, i.e. $\mathrm{F} + \mathrm{F} \rightarrow \mathrm{F}$. The steady state value of F in the first process is an initial value of F in the second one. The second process is described by the differential equation $\mathrm{d}\,\mathrm{F}\,/\mathrm{d}t = -k\,\mathrm{F}^2$, which has an analytical solution of the form $\mathrm{F}(t) = \mathrm{F}_0\,/(1 + k\,t\,\mathrm{F}_0)$, where F_0 is the initial value of F. The initial concentration of tetramers in the first process is c_0, hence it follows that $\mathrm{F}_0 = c_0/8$ since all tetramers are turned into ULFs. In consequence, the mean filament length can be expressed as

$$L_m(t) = l_{ULF} + \frac{k\,c_0\,t}{8}.$$

Thus, $L_m(t)$ is a linear function. By plotting the experimental data in Table 1 for time points after 30 seconds, together with their 0.95 confidence intervals one can see that there exists no k such that the model would be fitted and validated against the data.

3.2 Data fitting the extended model

In the case of the extended model we distinguished among three modes for filament elongation: (i) with a tetramer, (ii) with a ULF, or (iii) with another filament, see Figure 2 (b). We investigated all eight possible combinations of these three mechanisms and performed parameter estimation and numerical model validation for each of them, see Figure 5. Excluding any of the three modes from the investigation was done by simply setting to 0 the corresponding rate constants, i.e. k_t^e, k_u^e, and k_f^e, respectively.

3.2.1 The extended model with fast ULF formation.

In the case of fast tetramers-to-ULF turnover, both the simple model and the extended model can be reduced. Indeed, in this case, the populations of tetramers, octamers, and hexadecamers are all quickly depleted (in a matter of seconds), leaving only the filaments as the dominant species. Consequently, the longitudinal assembly of tetramers to filaments has a negligible contribution to the overall dynamics of the model: in the first few seconds the reaction is strangled by the negligible population of filaments, whereas later on the population of tetramers is depleted. This is in agreement with [15], where it was observed that this particular elongation has insignificant role. In this case we set $k_t^e = 0$ and we searched for numerical values for the kinetic rate constants k_1^e, k_2^e, and k_3^e that are greater than $3\ \mu\mathrm{M}^{-1}s^{-1}$, to ensure a fast tetramer-to-ULF turnover. It turned out that scenario VIII, where $k_u^e = k_f^e = 0$, could be immediately excluded. Indeed, in this scenario no filament containing more than two ULFs could be assembled and so, all filaments would be at most 100 nm long, contradicting the experimental data in Table 1.

Scenarios VI and VII, where the filament elongation takes place only by ULF extension ($k_f^e = 0$), or only by filament extension ($k_u^e = 0$), respectively, could

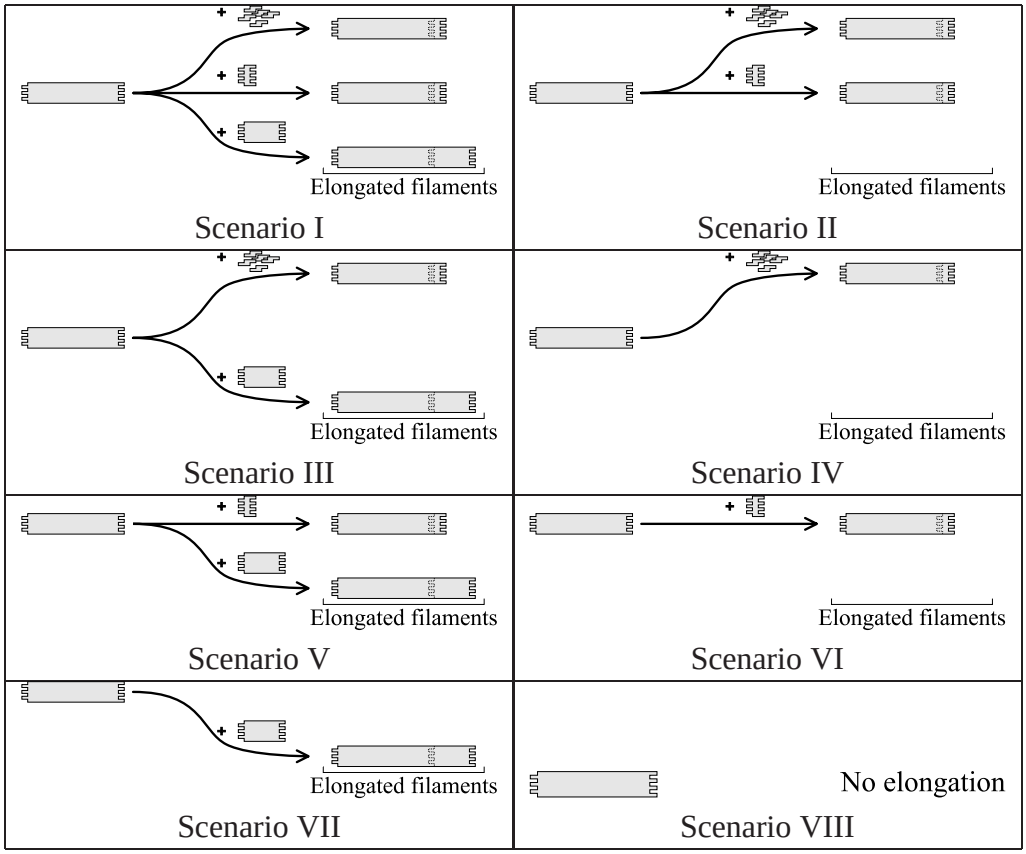


Figure 5: The eight possible scenarios for filament elongation. The tetramers/ULFs/filaments are illustrated with the same type of block as in Figure 2.

Table 3: Kinetic rate constant values in $\mu\mathrm{M}^{-1}s^{-1}$ (under fast ULF formation requirement).

k_1^e	k_2^e	k_3^e	k_4^e	k_u^e	k_f^e
3	30	30	0.25	0.95	0.11

not be fitted: for Scenario VI we obtained $fq = 22.77\%$ and for Scenario VII $fq = 14.99\%$, $vq = 16.07\%$. We concluded that these two strategies do not represent viable pathways for vimentin IFs assembly.

In the case of scenario V we were able to obtain numerical values for the parameters, see Table 3, such that the predicted mean filament length was in good agrement with the experimental data ($fq = 3.66\%$, $vq = 11.45\%$), virtually identical to that of the simple model, showed in Figure 4. We concluded that this pathway, where the filament elongation is enabled both with ULFs and with other filaments, is the only viable strategy for vimentin IFs assembly. This is in agreement with observations of [15].

Numerically fitting this scenario, we noticed that the values of the two numerical parameter k_2^e and k_3^e can be modified arbitrarily within the $[3, 30]$ interval

Table 4: Fit and validation quality measure values for scenarios I–VII (without the fast ULF formation requirement).

	I	II	III	IV	V	VI	VII
fq	1.71%	6.50%	1.98%	6.79%	2.04%	6.54%	13.01%
vq	12.70%	29.03%	12.36%	25.83%	12.65%	29.11%	19.19%

Table 5: Kinetic rate constant values in $\mu\text{M}^{-1}s^{-1}$ of scenarios I–VII (without the fast ULF formation requirement).

	I	II	III	IV	V	VI	VII
k_1^e	0.0705	30	$4.83 \cdot 10^{-3}$	$4.58 \cdot 10^{-3}$	1.24	30	30
k_2^e	30	30	30	10^{-09}	17.78	30	30
k_3^e	11.34	$4.63 \cdot 10^{-3}$	21.25	$6.06 \cdot 10^{-5}$	$2.65 \cdot 10^{-2}$	$4.67 \cdot 10^{-3}$	30
k_4^e	0.32	10.69	30	30	11.16	10.69	2.56
k_t^e	15.48	30	0.61	0.84	0	0	0
k_u^e	0.59	30	0	0	11.57	30	0
k_f^e	0.10	0	0.10	0	0.10	0	0.15

without any significant change in the mean filament length prediction. This indicates that the extended model under the fast ULF formation exhibits almost no sensitivity of mean filament length with respect to these two parameters in the mentioned interval and, in consequence, our computational model turns to have less degrees of freedom in terms of the numerical fit.

3.2.2 The extended model with slow ULF formation.

In this case, we searched for arbitrary positive numerical values for the kinetic rate constants k_1^e, k_2^e, and k_3^e. The result of fitting and validating the extended model are very different in this case. We find that three out of the eight pathways analyzed in this paper for vimentin IFs assembly can explain the experimental data, see Figures 6 and 7.

Scenario VIII could not be fitted based on similar considerations as in the case of the fast ULF formation, see Figure 7 VIII(a) and VIII(b). In the case of the other seven pathways, the model fit with respect to the EM 0.45 μM data and the model validation with respect to the EM 0.9 μM data yielded good results, summarized in Table 5, see Figures 6 and 7V-VII. We noted that in the case of scenarios II, IV, and VI the experimental MFL measurement at 1200 seconds for the EM 0.9 μM data was an outlier. In all these three scenarios, we have $k_f^e = 0$, which indicates that the process of end-to-end filament annealing plays a crucial role in the later stages of the IFs elongation process, i.e., after the first 600 seconds. In the case of scenario VII, the model left several experimental data points as outliers, see Figure 7VII(a) and (b).

We concluded that scenarios I, III, and V are similarly good in explaining the

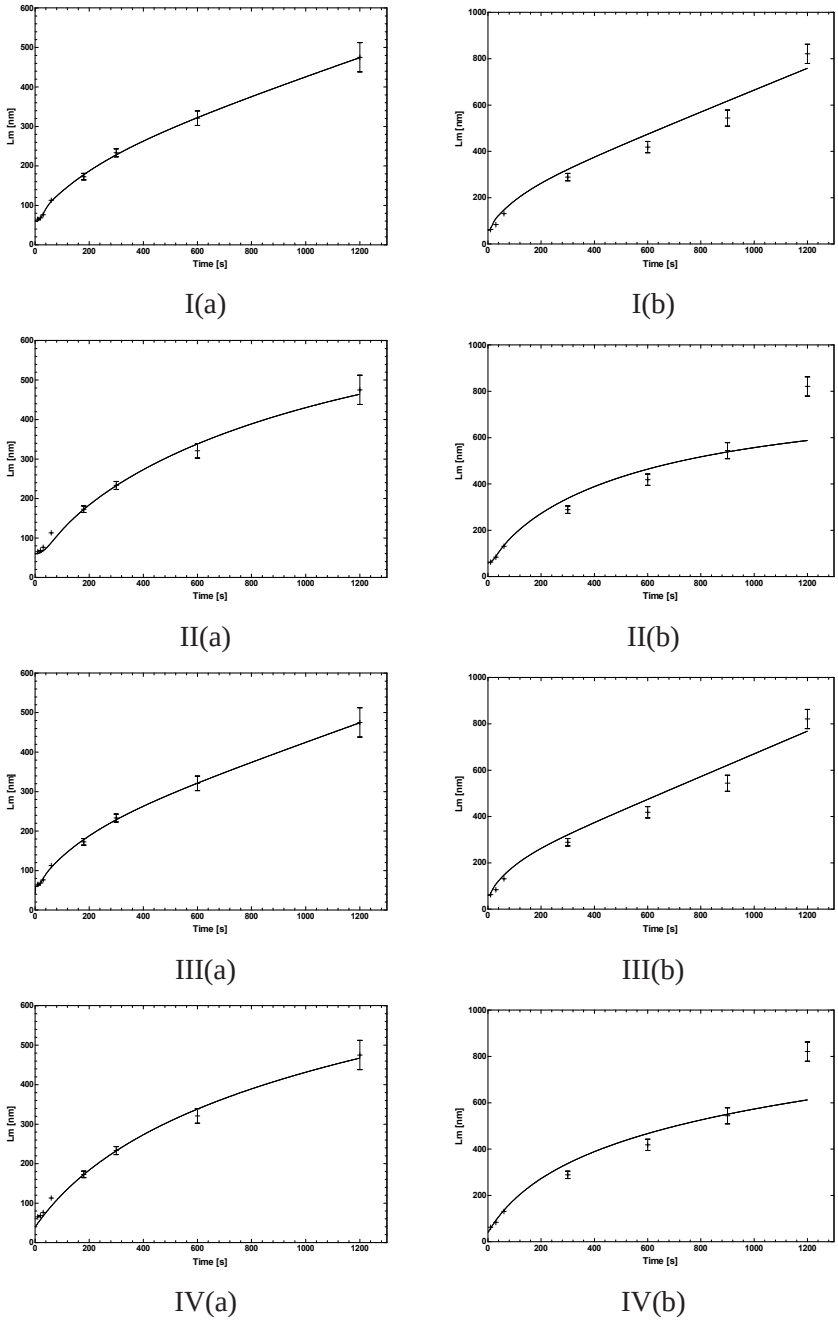


Figure 6: I(a)–IV(a) The model fit of the scenarios I to IV with respect to the EM 0.45 μM experimental data set. I(b)–IV(b) Model validation of the scenarios I to IV with respect to the EM 0.9 μM experimental data set. The continuous line is the model prediction regarding $L_m(t)$, that is compared with the experimental data showed with crossed points. The short vertical lines represent the 0.95 confidence intervals for the experimental data.

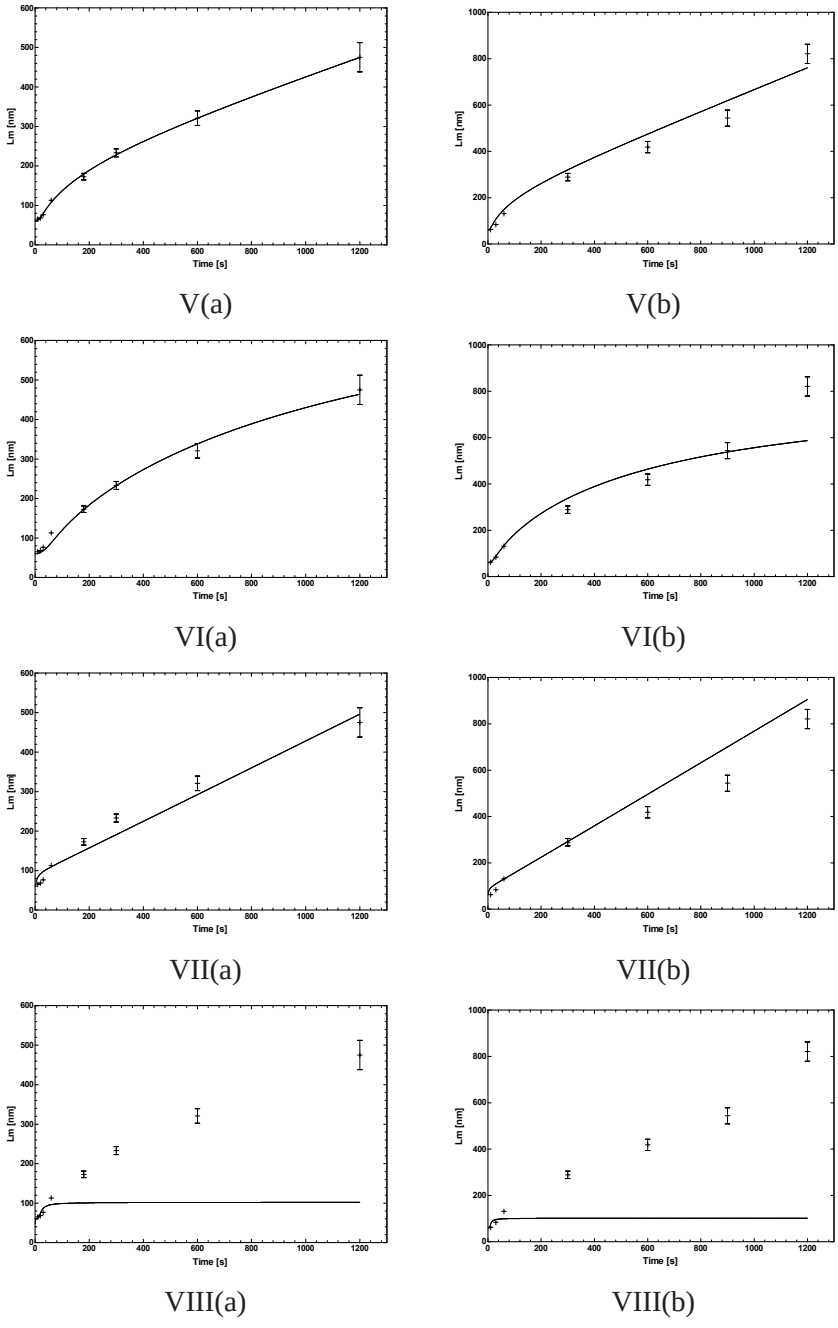


Figure 7: V(a)–VIII(a) The model fit of the scenarios V to VIII with respect to the EM 0.45 μM experimental data set. V(b)–VIII(b) Model validation of the scenarios V to VIII with respect to the EM 0.9 μM experimental data set. The continuous line is the model prediction regarding $L_m(t)$, that is compared with the experimental data showed with crossed points. The short vertical lines represent the 0.95 confidence intervals for the experimental data.

experimental data in this case. These models correspond to the following three pathways for filament elongation:

- Scenario I: by a tetramer, a ULF or another filament longitudinal elongation;
- Scenario III: by a tetramer or a filament longitudinal elongation;
- Scenario V: by a ULF or a filament longitudinal elongation.

3.3 Sensitivity analysis of the mean filament length

The effect of small variations in the model's parameters over the evolution of the entire model is estimated by the sensitivity analysis. This mathematical method consists in determining the time evolution of the partial derivatives of the solution of the system with respect to the parameters of the system. We investigated the sensitivity of the mean filament length, i.e., the $L_m(t)$ function, with respect to the parameters of the model. We compared the results of the sensitivity analysis in the case of Scenarios I-VII of the extended model in order to gain further insight into the possible pathways for IF vimentin assembly.

The concentration sensitivity coefficients are the time functions $\partial X_i/\partial \kappa_j$ for all $1 \leq i \leq 5$ and $1 \leq j \leq 7$, where $X = (X_1, \ldots, X_5)$ is the vector of the model variables ([T], [O], [H], [U], and [F], respectively) and $\kappa = (\kappa_1, \ldots, \kappa_7)$ is the vector of the model parameters ($k_1^e, k_2^e, k_3^e, k_4^e, k_t^e, k_u^e$, and k_f^e, respectively). The sensitivity of the mean filament length with respect to the parameters is obtained as follows:

$$\frac{\partial L_m(t)}{\partial \kappa_j} = \frac{\partial L_m}{\partial X}\,\frac{\partial X}{\partial \kappa_j} = \frac{\partial L_m}{\partial X_1}\,\frac{\partial X_1}{\partial \kappa_j} + \cdots + \frac{\partial L_m}{\partial X_5}\,\frac{\partial X_5}{\partial \kappa_j},$$

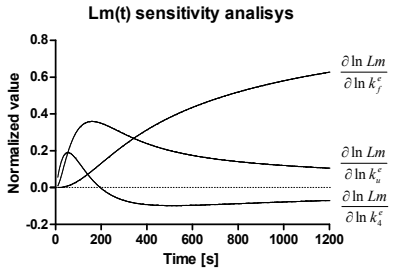
for all $1 \leq j \leq 7$.

Since we want to compare the MFL sensitivities of several models, we transform these coefficients into dimensionless measurements by normalizing them:

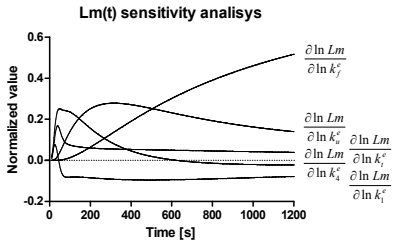
$$\frac{\kappa_j}{L_m(t)}\,\frac{\partial L_m(t)}{\partial \kappa_j} = \frac{\partial \ln L_m(t)}{\partial \ln \kappa_j}, \quad \text{for all } 1 \leq j \leq 7.$$

We can interpret these coefficients as follows: in Scenario I, an increase of 1% of the parameter k_f^e would generate at time $t = 1200\,s$ an increase of 0.5165% of the MFL, roughly as predicted by the value of $\partial \ln(L_m)/\partial \ln(k_f)$ at time $t = 1200$, see Figure 8 b).

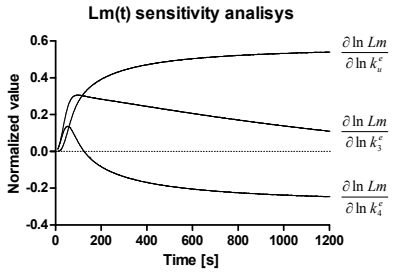
In the case of the extended model with fast ULF formation, only scenario V could be experimentally validated. The results of the sensitivity analysis in this case are presented in Figure 8 a). The most significant coefficients are with respect to the k_4^e, k_u^e, and k_f^e parameters, with the latter one being the most significant. This is consistent with the biological intuition that the mean filament length



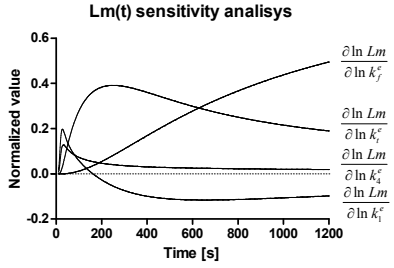
a) Scenario with fast ULF formation



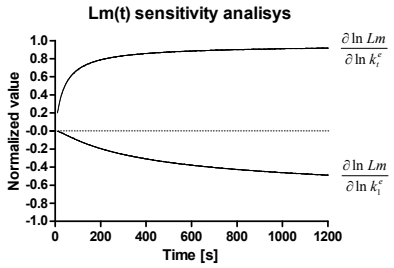
b) Scenario I



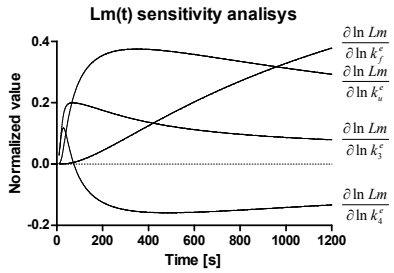
c) Scenario II



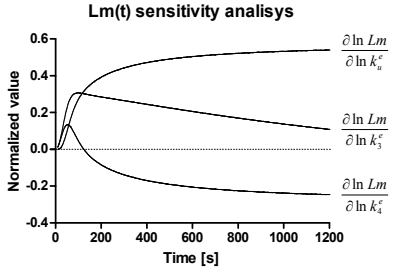
d) Scenario III



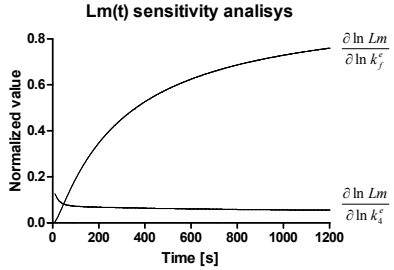
e) Scenario IV



f) Scenario V



g) Scenario VI



h) Scenario VII

Figure 8: The non-negligible sensitivity coefficients of the MFL measurement for the mathematical models corresponding to the scenario with fast ULF formation requirement and the scenarios I to VII.

is most dependent on the rate of filament formation (parameter k_4^e) and elongation (parameters k_u^e and k_f^e). Less intuitive is the fact that there is a negligible dependency of the MFL measurement on the rate constants $k_1^e, k_2^e,$ and k_3^e, which determine the fast ULF formation. The rationale for this result is that these kinetic constants play a role only in the first seconds of the assembly. Once the vast majority of tetramers are assembled into ULFs, their further contribution to the model dynamics is insignificant.

The numerical time simulation of the non-negligible normalized MFL sensitivity coefficients for scenarios I-VII without fast ULF formation requirement are presented in Figure 8 b)–h). It turned out that the mean filament length is most sensitive to k_u^e and especially to k_f^e, when these constants are non-zero. This observation helps explain why k_f^e is estimated to very similar values in most scenarios where its role is considered. Note also that while the sensitivity coefficient with respect to k_f^e increases mainly after about 200 seconds, the sensitivity coefficients for the parameters k_t^e and k_u^e have a steep increase in the first 100–200 seconds (except in scenario VII where filament elongation takes place only by longitudinal filament aggregation). The biological intuition here is that on one hand, until approximately 200 seconds the assembled filaments are relatively short and much fewer than the ULF's, while on the other hand the number of ULFs and of free tetramers becomes very low after about 200 seconds.

3.4 The length distribution of filaments in time

The models discussed so far in this paper, as well as those in [15] collect all filaments other than ULFs into one single variable denoted F, regardless of their length. This approach is indeed enough for capturing the time-dependent dynamics of the mean filament length, that could then be related to experimental data and used for parameter estimation and model validation. As pointed out also in [15], this modeling approach is however unsuitable for capturing the time-dependent distribution of the filament lengths. Indeed, the length of the assembling filaments is not directly captured in the models, which makes it impossible to reason about the time-dependant concentration of filaments of some given length. We describe in this section a refined model for the self-assembly of vimentin filaments that allows capturing the evolution of filaments of length up to n, for any given positive integer n.

For all i with $1 \leq i \leq n$, we denote by F_i the population of all filaments of length exactly i, where the length is in terms of the number of ULFs that the filament consists of. Thus, the ULFs are denoted by F_1 in the new model, the filaments formed by the longitudinal extension of a ULF with another ULF have length 2 and are denoted by F_2, etc. The population of all filaments of length higher than n is denoted by $F_{\geq n+1}$. The longitudinal extension of a filament F_i (of length $i \leq n$) with a filament F_j (of length $j \leq n$) yields a filament of length F_{i+j} if $i+j \leq n$ and a filament $F_{\geq n+1}$ if $i+j \geq n+1$. The extension of a filament

$F_{\geq n+1}$ with any other filament yields a filament $F_{\geq n+1}$.

When describing the extended model for filament self-assembly based on the populations F_i, $1 \leq i \leq n$, and $F_{\geq n+1}$, a considerable challenge is posed by the elongation of a filament with tetramers. Indeed, such a longitudinal elongation leads to a filament that ends with an incomplete ULF. Only after the lateral association of seven other tetramers would this be a complete filament of length one higher. This difficulty can be addressed by introducing a notation of the type $F_i^{j,k}$ with $1 \leq i \leq n$ and $0 \leq j, k \leq 7$ to denote filaments consisting of i complete ULFs, an incomplete ULF with j tetramers at their left end, and an incomplete ULF with k tetramers at their right end, see Figure 9. One would also denote by $F_{\geq n+1}^{j,k}$ the filaments consisting of more than n complete ULFs, an incomplete ULF with j tetramers at their left end, and an incomplete ULF with k tetramers at their right end. This approach leads however to a steep increase in the number of model variables. For example, for $n = 10$, the model would have 396 variables just to denote the different types of filaments.

To keep the size of the model manageable we can however take advantage of the kinetic observations we made on the extended model for filament assembly in Section 3.2: in the case of fast ULF formation we have demonstrated that the longitudinal elongation of filaments with tetramers has negligible kinetic influence on the dynamics of the model and that eliminating it leads to a numerically equivalent model. Consequently, we can ignore all possible filaments having incomplete ULFs at either end, since essentially all tetramers in the system assemble into ULFs within a very short period of time. In this case our model consists of the following reactions:

$$
\begin{array}{ll}
(T) & T + T \to O; \\
(O) & O + O \to H; \\
(H) & H + H \to F_1; \\
(A_{i,j}) & F_i + F_j \to F_{i+j}, \\
 & \quad \text{for all } 1 \leq i \leq j \leq n \text{ such that} \\
 & \quad i + j \leq n; \\
(B_{i,j}) & F_i + F_j \to F_{\geq n+1}, \\
 & \quad \text{for all } 1 \leq i \leq j \leq n \text{ such that} \\
 & \quad i + j \geq n + 1; \\
(C_i) & F_{\geq n+1} + F_i \to F_{\geq n+1}, \\
 & \quad \text{for all } 1 \leq i \leq n; \\
(D) & F_{\geq n+1} + F_{\geq n+1} \to F_{\geq n+1}.
\end{array}
$$

We call this a *model of resolution* n, see Figure 10 for an illustration. For example, in the case of $n = 10$, the model consists of 14 variables and 69 reactions.

The initial values of all variables except for T are set to 0, while that of T is assumed the same as in the extended model in Section 3.2. The kinetic rate constants of the new model are set in such a way that the overall number of filaments is the same as in the extended model. The kinetics of reactions (T), (O), and (H)

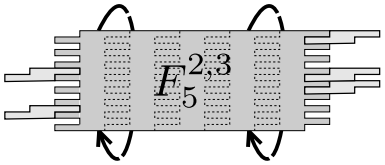


Figure 9: A filament consisting of 5 complete ULFs, an incomplete ULF with 2 tetramers at the left end, and an incomplete ULF with 3 tetramers at the right end. We denote it in our model with $F_5^{2,3}$.

are the same as in the corresponding reactions of the extended model. If $a_{i,j}$ is the kinetic rate constant of reaction ($A_{i,j}$), $b_{i,j}$ that of reaction ($B_{i,j}$), c_i that of reaction (C_i), and d that of reaction (D), then we set their values as follows:

- $a_{1,1} = k_4^e$, $a_{1,j} = k_u^e$, for all $1 < j \leq n$;
- $b_{1,j} = c_1 = k_u^e$, for all $1 \leq j \leq n$;
- $a_{i,i} = b_{i,i} = k_f^e$, for all $1 < i \leq n$, and $a_{i,j} = b_{i,j} = c_i = 2k_f^e$, for all $1 < i < j \leq n$;
- $d = k_f^e$.

Based on the corresponding ODE models, a straightforward calculation shows that with these kinetic constants, the extended model of Section 3.2 and the model of resolution n are equivalent in the following sense:

- $[F_1](t) = [U](t)$ and
- $([F_2] + \ldots + [F_n] + [F_{\geq n}])(t) = [F](t)$,

for all time points $t \geq 0$.

As an example, we have implemented in COPASI the model in the case of $n = 10$. In Figure 11 we plotted this model's prediction for the distribution in time of all filaments of length at least two. The resulting dynamics is in line with the biological expectation. For example, the number of filaments of length two, F_2, witnesses a sharp increase right after the start of the experiment, as tetramers are turned into (short) filaments. F_2 then decreases quickly as filaments start combining to each other to yield longer filaments.

4 Discussion

Related work. A recent review of the biochemistry of the intermediate filaments, including kinetic aspects of their self-assembly is in [8]. The simple and extended models for the self-assembly of vimentin proteins were originally investigated in [15]. The approach used in the fitting and the validation of the models

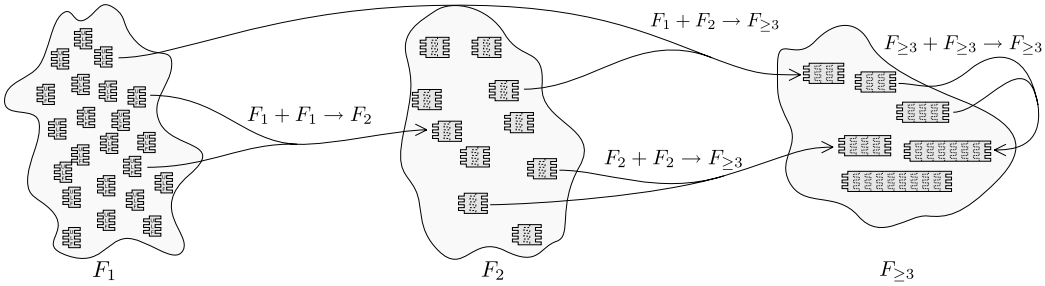


Figure 10: The scheme of a model of resolution 3 for the self-assembly of IF. We partition the population of filaments into filaments of length one (F_1), of length two (F_2), and of length at least three ($F_{\geq 3}$). The longitudinal annealing of two filaments of length one yields a filament of length two ($F_1 + F_1 \rightarrow F_2$), that of a filament of length one and another of length two yields a filament of length at least three ($F_1 + F_2 \rightarrow F_{\geq 3}$), the annealing of two filaments of length two yields a filament of length at least three ($F_2 + F_2 \rightarrow F_{\geq 3}$), and that of two filaments of length at least three results in a filament belonging to the same $F_{\geq 3}$ group ($F_{\geq 3} + F_{\geq 3} \rightarrow F_{\geq 3}$).

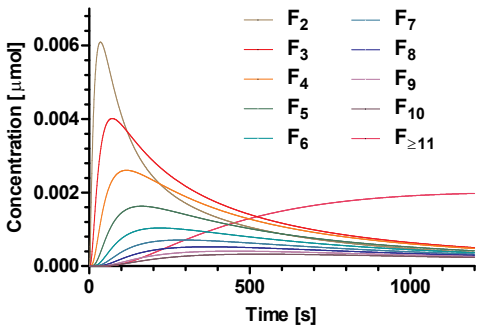


Figure 11: Model prediction for the distribution in time of all the filaments containing from two to ten ULFs.

was somewhat ad-hoc in [15], as discussed below. We made in our paper a systematic investigation of the kinetics of the two models for intermediate filament self-assembly, based on well-established techniques of model fit and model validation. Some of our results confirm those of [15], while others bring a new insight into the nature of filament assembly. We discuss in the following the main points of divergence between our approach and that of [15].

A main difference concerns the mathematical modeling of the simple and the extended models. The models in [15] assume that the lateral association of two tetramers, of two octamers, and of two hexadecamers have the same kinetic rate constants. This strong model assumption is however unsubstantiated by experimental evidence and leads to limiting the range of possible model behaviors. We assign different kinetic constants for each different reaction to allow maximum flexibility in the predictive power of the models.

Our mathematical expression for the mean filament length differs from the one presented in [15]. In there, the authors use a so-called *linear density variable* d_l, set at 43.5 nm, representing the length of a ULF inside a filament, regardless of whether the ULF is the first of the filament, or a subsequent one. This distinction is however crucial for estimating the mean filament length. Indeed, ignoring this distinction introduces an approximation error which is proportional to the length of each filament. For example, according to the formula from [15], the length of a filament consisting of only two ULFs is 2×43.5 nm $= 87$ nm, while according to the current knowledge regarding filaments measurements, see [2], its length is 63 nm $+ 42$ nm $= 105$ nm. Consequently, [15] introduces a so-called correction factor that only partially addresses the problem. Our approach for computing the MFL value is not influenced by this approximation error and leads to a correct interpretation of the experimental data.

For the experimental data fit of the models, [15] performs a so-called preassessment of the eight variants of the extended model. Based on some *fixed parameter values*, the eight variants are classified into four classes of dynamics. Three of the classes are then quickly dismissed from the analysis and only one representative of the remaining class is chosen for further assessment. This approach is however assuming that the classification of the dynamics of the eight model variants is independent of the parameter values, which is most likely not true for mathematical models with 5 or more parameters, such as those in [15]. In our case the approach was different. During parameter estimation we fitted all the variants of the extended model with respect to the EM 0.45 μM experimental data set. We then took advantage of the available data from the EM 0.9 μM experiment and performed model validation by comparing the predictions of the models with the experimental data. On the contrary, the second set of data was used in [15] in a second round of model fit, yielding different numerical values for the model parameters.

For the sake of having models of small size, in the first part of the paper we do not distinguish between filaments of different sizes and we use for the filament-

filament extensions a "generic" kinetic constant. However, in the second part of the paper we explicitly address the problem of extending the molecular model to distinguish between filaments of different sizes, recognizing that different constants may/should be used depending on the size of the filaments. We approach the problem from a numerical point of view, aiming to build the extended model in such a way that the numerical fit of the original model is preserved. On the other hand, in [18] a physical approach to estimate how the size of the complexes influences the binding rates is taken. However, this approach is based on the hypotheses that: i) reactants are shaped like balls and, especially, ii) the diameter of the balls representing larger complexes is the same as the diameter of the balls representing small complexes. Unfortunately, these assumptions make the approach of [18] unsuitable for filament-filament interactions. The approach might be developed further to suit our models by modifying the reactants-as-balls assumption and/or the assumption regarding the size of the larger complexes. This would require the recalculation of the collision probabilities in the stochastic approach to chemical kinetics. This however is a project in itself, distinct from the aim and scope of this paper.

Conclusions and further work. Our mathematical models show that if tetramers are very quickly (in just a few seconds) assembled into ULFs, then the elongation of filaments with ULFs and with other filaments both play a crucial role in the formation of long intermediate filaments. The elongation with tetramers on the other hand, has negligible quantitative contribution to the filament assembly. One reason for this is that in the case of fast ULF formation, the population of tetramers is very quickly depleted. However, this leaves open the question of the filament assembly dynamics in the case when tetramers would be continuously added to the system, i.e. by an additional synthesis mechanism. To address this problem, we investigated our mathematical models in the case when the turnover of tetramers into ULFs is slower. It turned out that in case the tetramers persist in the system for a longer time, the dynamics of the filament assembly is much richer and several different mechanisms can equally well explain the available experimental data. In fact, even the simple model discussed in [15] and in our paper could be fitted to the experimental data. An *in vitro* experiment where tetramers were added either continuously or at well-chosen time points could offer more insight into the role of tetramer longitudinal aggregation for the process of filament elongation. Choosing the time points when the additional amount of tetramers should be added to the solution could be done based on the analysis of our mathematical models. For example, one could choose the time points where the number of filaments in the solution is close to its maximum, so that the possible interplay between tetramers and filaments has maximum flux.

It is visible already from the experimental data that the system does not reach a steady state within 20 minutes, our time interval of choice. Similarly as in the study in [15], we have focused on the early dynamics of the vimentin filament

assembly, where the kinetics of the system is fast, with tetramers and ULFs being quickly replaced by emerging filaments of various lengths. During this phase, the presence of a large amount of tetramers and, a little later, of short filaments in the solution make far more likely assembly/elongation events rather than disassembly events. For this reason our models turn out to be able to explain the experimental data during the early phase of the assembly, even though they do not include any disassembly or filament breaking mechanisms. The applicability of the models is however tied to the early part of the assembly. Over longer time intervals (e.g., long enough so that the experimental data may potentially show a steady state), the lack of a disassembly mechanism in the models makes them limited in their predictive power. For example, a model with no disassembly or filament breaking mechanism would predict that the system will reach (albeit in a huge interval of time) a steady state where all initial tetramers are integrated into one single filament (of huge length).

The methodology introduced in this paper for increasing the resolution of the filament assembly model helps provide a deep insight into the dynamics of filament self-assembly. Details on the assembly of filaments of various lengths will help in designing finer grained experimental assays that would focus on filaments of different lengths at different time points. In terms of model complexity, increasing the resolution of the model implies a considerable increase in the size of the model, linear in the number of variables and quadratic in the number of reactions. We showed however that the kinetic rate constants can be set from a model of low resolution to one of higher resolution in such a way that the model predictions on the dynamics of the total amount of filaments, regardless of their length, are preserved. In particular, this implies that given generic data on, for example, the mean filament length, the model fit and the model validation problems can be solved on the (smaller) model of low resolution and then extrapolated to the models of higher resolution.

Acknowledgements. We are grateful to Robert Kirmse for the EM data on measurements of the mean filament length of vimentin intermediate filaments. The work of Eugen Czeizler, Elena Czeizler, Andrzej Mizera and Ion Petre was supported by Academy of Finland, grants 129863, 108421, and 122426. Andrzej Mizera is on leave of absence from the Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland.

References

[1] M. M. Ali, C. Storey, and A. Törn. Application of stochastic global optimization algorithms to practical problems. *J. Optim. Theory Appl.*, 95(3):545–563, 1997.

[2] S. Ando, K. ichiro Nakao, R. Gohara, Y. Takasaki, K. Suehiro, and Y. Oishi. Morphological analysis of glutaraldehyde-fixed vimentin intermediate filaments and assembly-intermediates by atomic force microscopy. *Biochimica et Biophysica Acta*, 1702(1):53–65, 2004.

[3] S. M. Baker, K. Schallau, and B. H. Junker. Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. *J Integr Bioinform*, 7(3):133, 2010.

[4] G. Bonne, M. R. D. Barletta, S. Varnous, H.-M. Bécane, E.-H. Hammouda, L. Merlini, F. Muntoni, C. R. Greenberg, F. Gary, J.-A. Urtizberea, D. Duboc, M. Fardeau, D. Toniolo, and K. Schwartz. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nat. Genet.*, 21:285–288, 1999.

[5] D. Fatkin, C. MacRae, T. Sasaki, M. R. Wolff, M. Porcu, M. Frenneaux, J. Atherton, H. J. Vidaillet, S. Spudich, U. D. Girolami, J. G. Seidman, C. Seidman, F. Muntoni, G. Müehle, W. Johnson, and B. McDonough. Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction-system disease. *N. Engl. J. Med.*, 341(23):1715–1724, 1999.

[6] I. E. Grossmann. *Global optimization in engineering design*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[7] C. Guus, E. Boender, and H. E. Romeijn. Stochastic methods. In R. Horst and P. M. Pardalos, editors, *Handbook of Global Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[8] H. Herrmann and U. Aebi. Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular scaffolds. *Ann Rev Biochem*, 73:749–789, 2004.

[9] H. Herrmann, H. Bär, L. Kreplak, S. V. Strelkov, and U. Aebi. Intermediate filaments: from cell architecture to nanomechanics. *Nature Reviews Mol Cell Biol*, 8:562–573, 2007.

[10] H. Herrmann, M. Häner, M. Brettel, N.-O. Ku, and U. Aebi. Characterization of distinct early assembly units of different intermediate filament proteins. *J. Mol. Biol.*, 286(5):1403–1420, 1999.

[11] K. C. Holmes, D. Popp, W. Gebhard, and W. Kabsch. Atomic model of the actin filament. *Nature*, 347:44–49, 1990.

[12] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI – a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[13] R. Horst and H. Tuy. *Global optimization: Deterministic approaches*. Springer-Verlag, Berlin, 1990.

[14] R. Kirmse. Personal communication, 2008.

[15] R. Kirmse, S. Portet, N. Mücke, U. Aebi, H. Herrmann, and J. Langowski. A quantitative kinetic model for the in vitro assembly of intermediate filaments from tetrameric vimentin. *J. Biol. Chem.*, 282(52):18563–18572, 2007.

[16] M. Kühnel, L. S. Mayorga, T. Dandekar, J. Thakar, R. Schwarz, E. Anes, G. Griffiths, and J. Reich. Modelling phagosomal lipid networks that regulate actin assembly. *BMC Systems Biology*, 2:107, 2008.

[17] E. Lazarides. Intermediate filaments as mechanical integrators of cellular space. *Nature*, 283(5744):249–256, 1980.

[18] L. Lok and R. Brent. Automatic generation of cellular reaction networks with moleculizer 1.0. *Nat. Biotechnol.*, 23:131–136, 2005.

[19] P. Mendes, S. Hoops, S. Sahle, R. Gauges, J. O. Dada, and U. Kummer. Computational modeling of biochemical networks using COPASI. *Methods Mol Biol*, 500:17–59, 2009.

[20] P. Mendes and D. Kell. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10):869–883, 1998.

[21] C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res*, 13:2467–2474, 2003.

[22] E. Nogales and K. H. Downing. Tubulin and microtubule structures. In T. Fojo, editor, *The Role of Microtubules in Cell Biology, Neurobiology, and Oncology*. Humana Press, 2008.

[23] S. Sahle, R. Gauges, J. Pahle, N. Simus, U. Kummer, S. Hoops, C. Lee, M. Singhal, L. Xu, and P. Mendes. Simulation of biochemical networks using COPASI: a complex pathway simulator. In *WSC '06 Proceedings of the 38th conference on Winter simulation*, pages 1698–1706, 2006.

[24] M. Schliwa. *The Cytoskeleton*, volume 13 of *Cell Biology Monographs*. Springer-Verlag, Vienna, Austria, 1986.

[25] M. O. Steinmetz, D. Stoffler, A. Hoenger, A. Bremer, and U. Aebi. Actin: from cell biology to atomic detail. *J. Struct. Biol.*, 119:295–320, 1997.

[26] H. J. Worman and J. C. Courvalin. The nuclear lamina and inherited disease. *Trends Cell. Biol.*, 12:591–598, 2002.

Turku
Centre *for*
Computer
Science

Joukahaisenkatu 3-5 B, FI-20520 Turku, Finland | www.tucs.fi

University of Turku
- Department of Information Technology
- Department of Mathematics

Åbo Akademi University
- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration
- Institute of Information Systems Sciences

ISBN 978-952-12-2387-7
ISSN 1239-1891