Testing scientific models using a QR model: Application to cellulose biodegradation

[1]Kamal Kansou and [2]Bert Bredeweg

[1]INRA, Biopolymères Interactions and Assemblages, BP 71267, 44316 Nantes, France

[2]Informatics Institute, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

Kamal.Kansou@nantes.inra.fr, B.Bredeweg@uva.nl

Abstract

The rapidly growing set of scientific publications makes it difficult for researchers to keep track of the progress towards adequate mechanistic explanations of phenomena. However, high-level representations can support integrating seemingly different results and ideas presented in the literature. This paper reports on our effort to deploy the qualitative reasoning framework as an instrument towards this end.

1 Introduction

The accumulation of scientific information is enormous. Keeping up to date in some fields of natural science is getting more and more difficult for the domain specialists (Fraser and Dunstan, 2010). For example, searching for "cellulose and hydrolysis and enzyme" in the Web Of Science yields more than 3000 scientific publications since 1995. Even experts find it difficult to keep integrating new mechanistic information about ligno-cellulose hydrolysis and envision the consequences on the system dynamics.

> *An emerging question is whether the (new) pieces of knowledge found in publications about a topic provide a way forward to a better (possibly complete) understanding of the underlying mechanism.*

Higher-level representations can support literature integration by reviewing and assembling information provided in scientific papers in a computable model. Higher-level (conceptual) modelling formalisms can integrate scattered qualitative information about a mechanism and provide a valuable envisioning of the system dynamics. Our objective is to explore solutions for representing and manipulating mechanistic explanations from publications using a computational model. We focus on the analysis of cause-effect relations to identify/test putative explanations for a set of evidences.

1.1 Domain – Cellulose hydrolysis limitation

We are interested in explanations for processes limiting the cellulose hydrolysis. Cellulose is the main component of plant cell wall, and an abundant and accessible renewable source of carbon. As such, cellulose is of central interest for the many natural and industrial processes, including the production of biofuel. Hydrolysis of solid cellulosic substrate into soluble cellodextrins by a cocktail of cellulases is characterised by progress-curves determined by the amount of the carbohydrates released in a solution. The curve shows a saturation-shape that reflects the catalytic activity. It is known that the efficiency of the depolymerisation of solid cellulose chains gradually declines with time. This means that the cellulases activity gets less efficient as the reaction proceeds (Lynd et al., 2002; Zhang and Lynd, 2004).

Numerous observations pertaining cellulose hydrolysis can be found in the literature. However, establishing a mechanistic explanation of the declining rate is still an important and unsolved issue. This missing insight hampers the global conversion efficiency of cellulose into ethanol (Lynd et al., 2002; Zhang and Lynd, 2004).

1.2 Potential of QR as an instrument

Quantitative approaches (e.g. ODE) need precise data, and are very dependent on experimental conditions. Even if the model structure can be applied in a variety of experimental situations, the need to get sufficient data to perform precise parametrization is a limitation factor. Furthermore, quantitative models cannot readily represent an informal description of a mechanistic explanation in an easy manner, as for instance text or diagrams can. Finally, mathematic formulation of a physical process is not directly interpretable in terms of cause-effect.

In the work presented in this paper, we use the Qualitative Reasoning (QR) framework, which does provide representations of cause-effect and is also able to generate simulations of the system dynamics. QR modelling is complementary to quantitative approaches in the sense that it allows for formulating distinct paradigms and for providing a first assessment to a range of evidences without requiring precise measurements of parameters or specific experimental conditions.

1.3 The challenge

We present an approach to stepwise construct a mechanistic explanation from selected papers about cellulose hydrolysis rate slowing-down using the QR framework. Many studies have investigated the cause of the phenomenon; both enzyme and substrate-related factors can be held responsible for the decline of hydrolysis rate. However, an integrated or unified explanation is not available.

We have developed three QR models. Two models are derived from published mechanistic models. The third model is derived from experimental observations from the literature and analysis of the simulations of the two other models. Our paper also reflects on methodological issues relevant to creating and assessing such models exploiting observations from publications. Our primary objective is to demonstrate how the QR framework can be used for this.

2 QR for mechanism modeling

QR strives for inferring behaviour from physical system structure in a symbolic, human-like manner. We use Garp3 (Bredeweg et al., 2009), a workbench for constructing and simulating QR models. To illustrate the use of QR, consider the basic enzymatic reaction: $E+S \rightleftharpoons ES \rightarrow P$, with E (enzyme), S (substrate), ES (complex enzyme-substrate), and P (product). The ODE system representing this phenomenon computes the derivatives of the E, S, ES and P. concentrations. These simulations are well known. Fig. 1 shows the kinetic curves (coloured lines), produced with dummy values for the kinetic constants.

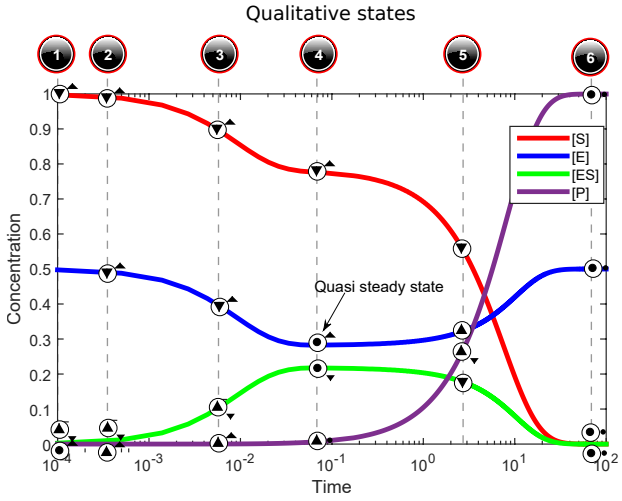

Figure 1. Simulation results for an enzymatic reaction with logarithmic time. The top row shows corresponding qualitative states, produced by simulating a QR model. Value histories of the quantities are placed on top of simulation curves. Key states are: initial state 1 (substrate starts being complexed with enzyme), state 4 (quasi-steady state), and end-state 6 (substrate conversion complete).

The Garp3 model implements a process-centric view, which emphasizes rates. Thus a Garp3 model of Equation 1 includes four entities (E, S, ES, P) each with a quantity *Concentration,* but also the rates *Ratein* and *Rateout* for respectively formation rates (for ES and P) and disappearing rate (for ES). In Garp3, quantities are characterized by: <*Magnitude, Derivative*>. The domain of allowable magnitudes associated with each quantity is called the Quantity Space (QS). *Concentration [in E]* and *Concentration [in S]* are assigned QS: {*Zero*, *Plus*, *Max*}, the other quantities have QS: {*Zero*, *Plus*}. All derivatives have QS {▼,●,▲} representing decreasing, steady, and increasing.

Garp3 provides two primitives for capturing causal dependencies between quantities, direct influence (I+ and I-) to model a *rate* influencing a concentration, and qualitative proportionality (P+; P-) to model the propagation of changes from one quantity to the next (cf. Forbus, 2008). P* is special kind of proportionality that captures the relation between the terms of a product and the result of this product.

Simulation results for the enzymatic reaction model, starting from maximum magnitudes for *Concentration [in E]* and *Concentration [in S]* includes a state-graph of 9 states. A Behaviour Path (BP) is a possible behaviour defined as a succession of qualitative states along a complete timeline. In Fig. 1 the BP [1→2→3→4→5→6] and value histories corresponding to the simulation curves are provided.

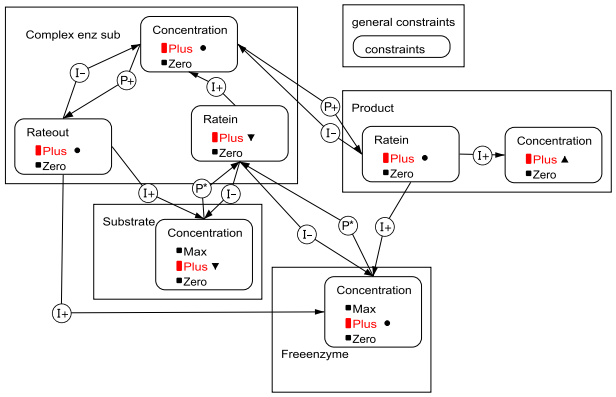

Figure 2. Causal dependencies compiled by Garp3 for state 4, providing a causal account for what is depicted by the value history graphs.

This shows that this particular BP matches the numerical simulation given Fig. 1. Key qualitative states of the process are identified this way, thus state 4 of the BP represents the quasi-steady state. The assembly of the causal chain active in state 4 is shown in Fig. 2. From this graph one can identify interacting feedbacks. For instance: two positive feedbacks, one productive including *Ratein [in P],* one unproductive including *Rateout [in ES]*, determine the reaction overall efficiency.

3 Explanatory model based on scientific publications

3.1 Behaviour path (BP)

In Garp3, a qualitative simulation of system behaviour uses a set of quantities $x_i \in X, i = \{1, \dots n\}$ linked by causal dependencies, and constrained by inequalities. A Qualitative State (QS) describes the system at time t such as: $\text{QS} = \{\langle t, < x_i, \text{magnitude} = \alpha, \text{derivative} = \beta >\rangle, \forall x_i \in X\}$ with α, some value of the QS assigned to x_i, β a value of the QS assigned for derivatives (in Garp3: {▼,●,▲}). A Behaviour Path (BP) is a finite sequence of m qualitative states that represents a possible qualitative behaviour over time. All QSs of a BP but the last one have a transition relation towards a possible and qualitatively distinct successor such as: $\text{BP} = \text{QS}_0 \rightarrow \cdots \rightarrow \text{QS}_m$

Each BP is associated to a discrete timeline, T, composed of m time periods such as $T := \langle t_0, \dots t_m \rangle$. Depending on the nature of the state, a period of time can be an instant or an interval. Note that, if two similar QSs are met at different times, as for a periodic behaviour, Garp3 refers to the same QS. The BP is then a loop.

3.2 Target behaviour (TB)

A Target Behaviour (TB) is a qualitative abstraction of one or more observations of actual behaviours exhibited by a real (target) system, whose structure is unknown and investigated by domain scientists. A TB captures distinctive features as Target States (TS) ordered in time for which the model needs to provide an explanation. A TS describes the target system for a given time period, t, through a set V of nt quantities with known magnitudes and/or derivatives: $\text{TS} = \{\langle t, < x_i, \text{magnitude} = \alpha', \text{derivative} = \beta' >\rangle, \forall x_i \in V\}$.

A model must include the variables of the TS to have a chance to satisfy it, therefore $V \subseteq X$. Similarly α' and β' belong to QSs also included in the corresponding qualitative model. For a TS α' and β' can be subsets of quantity spaces excluding the empty set. The full QS is noted "?".

In agreement with the QR formalism, a TB represents the change of the magnitude and the derivative of some quantities at distinct time intervals. Contrary to a BP, a TB does not need to cover a complete timeline, that is, from an initial state to an end state. A TB is defined as a finite sequence of mt target states, strictly ordered in time such as: $\text{TB} = \text{TS}_0 \rightarrow \cdots \rightarrow \text{TS}_{mt}$

The successor relation indicates simply that the next TS occurs some time later. Here again two successive TSs must be distinct. A TB applies to BPs produced by a simulation model to classify the possible behaviours of the system. It imposes that the TSs of a TB are satisfied in the right order by the QSs, therefore $mt \leq m$. It is often desirable that a TB covers a continuous time-period to rule out false positives.

Suppose that the curves in Fig.1 are observations (not the result of simulation). We can select states such as (*i*) initial-state of the reaction, (*ii*) intermediate state where [ES] is at a peak, and (*iii*) end-state. Those are likely to be characteristic states of the system under investigation. Then a possible TB could describe magnitudes and derivatives for the ES and P concentrations at three moments ($t_0 < t_1 < t_2$), as shown in Table 1.

Table 1. TB capturing qualitative features of Fig. 1 curves

Time index	Concentration [ES]	Concentration [P]
t_0 (initial state)	<Zero, ? >	<Zero, ?>
t_1 (intermediate state)	<Plus, ●>	<Plus, ▲>
t_2 (end state)	<Zero, ●>	<Plus, ●>

Note, '?' can be one of {▼,●,▲}.

Trying to explain the TB using a QR model of the enzymatic reaction produces the BP: [1→2→3→4→5→6] (Fig.1) consistent with Table 1: state 1 matches the initial-state, state 4 matches the intermediate (quasi-steady) state, and state 6 the end-state. All BPs containing these 3 states in the right order are consistent with Table 1. Therefore a QR model of the enzymatic reaction would provide a sufficient explanation for the TB.

3.23 Assessing QR models versus literature information

Using QR, it is possible to capture the causal information described in publications into qualitative cause-effect models, simulate these models, and thereby envision the information in terms of system behaviour. However, capturing causal links indistinctively from a set of papers will quickly make the qualitative simulation intractable and inappropriate for conveying a meaningful explanation to domain experts. Instead, we adopt an incremental model-building approach driven by a Target Behaviour (TB).

Establishing the TB is the first step in the modelling process, as it determines the modelling goal and orients the choices of entities, quantities, and QSs relevant for simulating the observed behaviours (Kansou and Bredeweg, 2014). In the ideal case, the TB is a mapping of existing time-series data. However, in natural sciences building a TB from a dataset obtained in specific experimental conditions can be insufficient to discriminate between concurrent explanations. Qualitative abstraction smoothens the peculiarities of experimental conditions reported in papers. This enables integration of observations from different sources into a composite TB, albeit with loss of some precision. A TB is built primarily on source papers and/or experimental results. This phase involves domain experts as main beneficiaries of the work for guidance about the literature and/or conducted experiments.

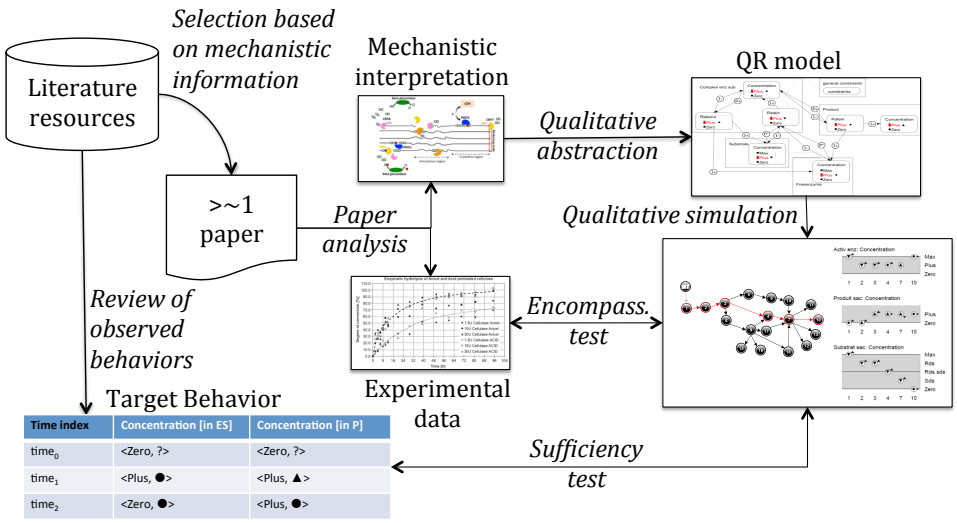

Figure 3. Using QR as an instrument to integrate scientific information from literature.

Our modelling methodology is depicted in Fig. 3. Selected papers introduce observations or simulation results related to the TB and provide useful mechanistic interpretations. Papers describing a quantitative model, usually with ODEs, are especially interesting as they propose a formal representation of a mechanism. For each version of the QR model its legitimacy as a faithful representation of the discovered knowledge is assessed using data and observations provided in the source papers, using the *encompassment* and the *sufficiency* test. The tests are defined as follows:

Encompassment: The QR model is a consistent representation of the interpretations given in the source papers. The model generates behaviours that match the observed data, numerical simulations or qualitative observations supplied in these papers.

Sufficiency: The QR model implements a plausible explanation for the target behaviour. The model generates a behaviour from which a plausible explanation for the target behaviour can be derived.

4 Testing cellulose hydrolysis paradigms

4.1 Defining target behaviour

To compose the TB, a short review of publications pertaining to the cellulose hydrolysis rate decline over time was performed. We strived for selecting publications addressing the most basic conditions, involving common cellulosic substrates with common hydrolytic enzyme, cellulase. The most important cellulase in this system has a processive action (enzyme complexed on a cellulose strand chops it up step-by-step as small sugars of similar size). The goal was to extract observations caused by basic processes that will take place regardless of the substrate nature (cellulosic or ligno-cellulosic) or the enzymatic cocktail complexity.

Hydrolysis rate decline: decline rate is related to the absolute quantity of bound enzymes as well as the specific rate per adsorbed enzyme (Lynd et al., 2002). The phenomenon extends over different time-scales. The hydrolysis decreases exponentially, immediately after an initial burst of catalytic activity and then at a much slower pace (Praestgaard et al., 2011), up to few days (Gan et al., 2003).

Restart experiments: Amongst the experiments in the domain, typical "perturbations" of the system include the addition of fresh enzyme in the course of the reaction, so-called "restart experiment". This type of experiment provides information about the system state, in particular about the state of the enzymatic component (Lynd et al., 2002; Eriksson et al., 2002). It has been observed that the addition of fresh enzymes, shortly after the reaction initialization, causes a clear restart of the hydrolysis (Cruys-Bagger et al., 2012). After longer time, it will cause a weak restart unless the cellulose surface is cleaned up beforehand (Yang et al., 2006).

We propose three TBs, (TB1, TB2 and TB2'), to capture prominent aspects of the experimental observations reported above (Table 2-4). TB2 (Table 3) depicts the restart phenomenon as the conversion of free enzyme in solution into catalytic active enzyme so that, the recruitment of active enzyme increases as long as the free enzyme quantity is increasing. In TB2' (Table 4) there is some process(es) limiting and eventually interrupting the restart phenomenon, so that increase in free enzyme might not result in an increase of catalytic rate.

Table 2. Hydrolysis rate is declining following an initial burst of hydrolytic activity (TB1).

Time index	Free enzyme	Catalytic rate
t_0	<Max, ?>	<0, ▲>
t_1	<{Zero, Plus}, ?>	<Plus, ▲>
t_2	<{Zero, Plus}, ?>	<Plus, ●>
t_3	<{Zero, Plus}, ?>	<Plus, ▼>

Table 3. Second dose of enzyme brings about a hydrolysis restart (TB2).

Time index	Free enzyme	Catalytic rate
t_0	<Plus, ▲>	<Plus, ●>
t_1	<Plus, ▲>	<Plus, ▲>
t_2	<Plus, ●>	<Plus, {●,▲}>

Table 4. Restart, but distinct from TB2 in that it represents a limited restart due to extra processes (TB2').

Time index	Free enzyme	Catalytic rate
t_0	<Plus, ▲>	<Plus, ●>
t_1	<Plus, ▲>	<Plus, {●, ▲}>
t_2	<Plus, ▲>	<Plus, ●>
t_3	<Plus, ●>	<Plus, ?>

4.2 Establishing models

We developed three QR models to test paradigms about cellulose hydrolysis proposed in the domain literature. To present the models structure we adopted a diagrammatic representation describing the causal linkages between quantities (Figs. 4-5) where rectangular box represents *concentration* or *amount_of* something, ellipse represents rate and causal linkages are labelled "P +/-" or "I +/-". "P+" and "P-" (proportionality relations) can connect two boxes together, a box to an ellipse or two ellipses together. "I+" and "I-" are direct influences. In the graph they can relate only an ellipse to a box. Algebraic relations can be implemented in Garp3 through qualitative algebra. Operators are represented by the symbols ⊕, ⊖ and ⊗.

The first QR model (M1) implements the surface-coverage limitation explanation based on modified Langmuir-Michaelis-Menten equations (proposed by Maurer et al., 2012). The system accounts for three processes: (*i*) reversible adsorption on the surface, (*ii*) reversible formation of surface enzyme-substrate complex, and (*iii*) hydrolysis of substrate generating a product without release of the active enzyme. The principle of the model is:

Ef + AS ↔ Ea + S ↔ ESa → ESa + P.

The corresponding mass balance relates the accessible surface concentration (AS) and the free enzyme concentration (Ef) to the production rate (dP/dt) via the surface concentration of adsorbed cellulase in an uncomplexed form (Ea) and in a complexed and catalytic active form (ESa). S stands for the substrate concentration surface cellulose chain, assumed constant in the model.

The model is depicted in Fig. 4. *Free enzyme* first adsorbs on *Accessible surface*, to form *Adsorbed enzyme*. *Adsorbed enzyme* can form *Active enzyme* that degrades the cellulose at *Catalytic rate*, or get back to the *Adsorbed enzyme* form. The *Covered surface*, populated by *Adsorbed enzyme* and *Active enzyme* reduces the *Accessible surface*.

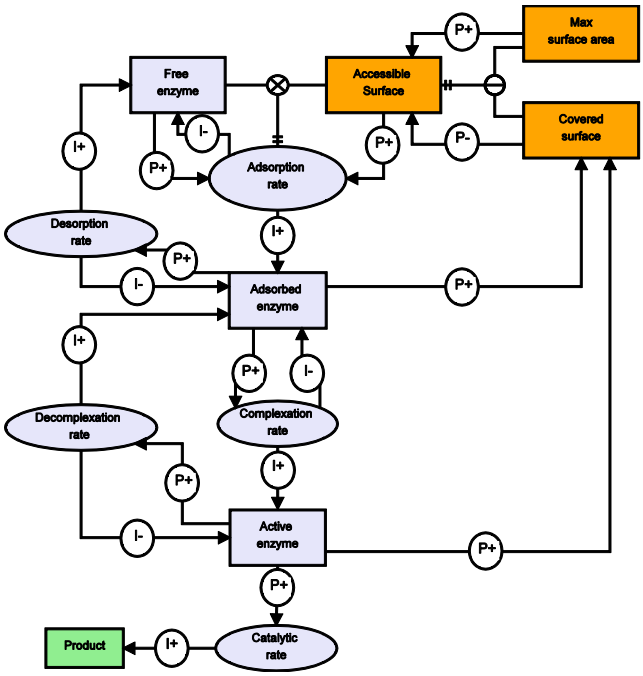

Figure 4. Model M1

The second QR model (M2, Fig. 5) implements an explanation related to putative presence of obstacles at the cellulose surface limiting the processive action of cellulase (presented in Jalak and Valjamae (2010), also implemented as kinetic model in Prastegaard et al., (2011) and in Cruys-Bagger et al., (2012)). The kinetic model implements the stalling of the processive enzyme when it reaches a surface obstacle during the catalytic process. The QR model has a global *Adsorption/complexation rate* of *Free enzyme* with cellulose, to form *Active enzyme*. *Active enzyme* degrades the cellulose strands processively at *Catalytic rate*. Next, it can either desorb (*Desorption rate1*) or get stalled if it meets an obstacle at *Stalling rate* and becomes *Stalled enzyme*. The *Desorption rates* (*Desorption rate1* + *Desorption rate2*) refill the amount of *Free enzyme* fuelling the turn-over. In our model, hydrolysis is a single step process performed by all *Active enzyme*, and not a summation of hydrolytic acts occurring along the cellulose strands as in the original model (Cruys-bagger et al., 2012). At the qualitative level, it would make the system and the ensuing explanation needlessly complicated. The relation between the *Catalytic rate* and the *Stalling rate* is modelled using a proportional dependency (P+).

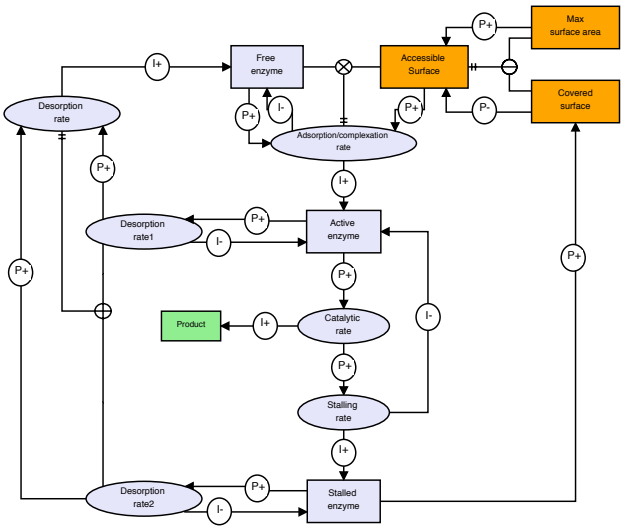

Figure 5. Model M2 and M3. M3 includes the surface limitation model fragment represented with orange boxes.

The third model (M3, Fig. 5) is an extension of M2 including the surface limitation from M1. It accounts for surface contamination by enzyme, which can hinder the hydrolytic activity. The process by which surface enzyme hinders the hydrolytic process is not clarified. Limitation of the adsorption due to *Covered surface* is similar to M1 (Fig. 4); this model also includes the case where *Covered surface* affects the complexation process. By extrapolating the impact of *Stalled enzyme* at the surface, we assume a proportional dependence (P+) between the *Stalled enzyme* and the *Covered surface*. In doing so, we test a new mechanism by which *Stalled enzyme* hinders the Adsorption and/or the complexation rate. Naturally other linkages of this kind could be tested as well. A more complete screening of the possible model structures is envisaged in future work.

4.3 Results

4.3.1 Simulation scenarios

The simulations used for testing the decline of the hydrolysis rate (TB1 and TB2') start from a scenario with only substrate and free enzyme (no product nor enzyme other than free in solution). Simulations of the restart phenomenon (TB2) are produced from a perturbation scenario reproducing the addition of a second dose of enzyme. Starting from a system in a state of equilibrium, with all the rates of the model (e.g. *Catalytic rate*, *Stalling rate*) being positive and stable, the addition of new *Free enzyme* is modelled through a feeding rate, exogenous to the system. The feeding rate is imposed to decrease over time. This accounts for the enzyme diffusion in the solution and limits the perturbation in time.

4.3.2 Description of model simulations

Models M1 on one hand and M2 and M3 on the other, exhibit very distinct behaviours. The M1 state-graph has seven states ordered linearly, with one stable end-state (state 5). The simulation envisions a conversion of *Free enzyme* into *Adsorbed enzyme* and then into *Active enzyme*. Models M2 and M3 produce state-graphs of 27 and 41 states each having a characteristic water lily leaf shape with a unique end-state at the centre (state 4). After a common starting branch (states 1→2) the system either: (*i*) goes directly to state 4 via the BP [1→2→3→4], (*ii*) initiates oscillations before reaching state 4 (*e.g.* Fig. 6), (*iii*) oscillates without reaching the end state. In the present situation, the system can be interpreted as a damped oscillator moving towards a steady state. State 4 is the equilibrium state with all quantities of the system steady, except for the concentration of *Product* that increases at a constant rate. The equilibrium state is characterized by the following equalities between the rates:

Ads/Comp rate = Des rate = Des rate1 + Des rate2
$\Rightarrow \delta$(*Free enzyme*) = 0
Ads/Comp rate = Stalling rate + Des rate1
$\Rightarrow \delta$(*Active enzyme*) = 0
Stalling rate = Des rate2 $\Rightarrow \delta$(*Stalled enzyme)* = 0
$\Rightarrow \delta$(*Accessible surface*) = 0

M2 and M3 both envision the accumulation of *Stalled enzyme* governed by the balance between the *Stalling rate* and *Desorption rate2*. Inclusion of *Accessible surface* in M3 implements a negative feedback from *Stalled enzyme* concentration to *Adsorption/Complexation rate*. This leads to more complicated oscillations than for model M2. This may reflect a longer establishment of the equilibrium state.

4.3.3 Testing the Encompassment of sources

To investigate the encompassment of M1 for the interpretation by Maurer et al., (2012) simulation curves have been produced of the published ODE model (not shown here). The longest BP (seven states) produced by M1 maps exactly onto the quantitative simulation. It depicts the burst and then the decline of *Adsorbed enzyme*, while *Active enzyme* increases up to maximum level from which it stabilizes. Limitation of *Active enzyme* can be traced back to decline of *Accessible surface* and *Free enzyme*. Even if *Accessible surface* can regulate the *Adsorption rate*, (Fig. 4) deleting this model fragment does not change the system behaviour. The encompassment for M2 regarding Cruys-Bagger et al., (2012) is depicted in Fig. 6. The BP [1→2→3→5→6→4] matches the simulation curves provided in that publication. A fraction of enzyme being stalled at the cellulose surface, it is easy to infer from Fig. 5: a low *Desorption rate2* will create a bottleneck effect impacting the turnover between free and active enzymes. M2 conveys successfully the idea that obstacles at the cellulose surface would slow-down the

hydrolytic activity. Interestingly, the shortest BP of the state-graph ([1→2→3→4]) also matches one of the experimental curves of Cruys-Bagger et al., (2012) (not shown here) obtained with the lowest substrate concentration. Here, the hydrolysis rate levels out close to its maximum value so that the burst is barely noticeable. In this situation the *Adsorption/Complexation rate* is certainly limitative compared to the other rates of the system.

M3 also fulfils the encompassment test for the Cruys-Bagger et al., (2012) results. It produces the same BP [1→2→3→5→6→4] as shown in Fig. 6. *Accessible surface* inclusion in M3 is not a representation of existing theory. As such, it does not encompass specific papers.

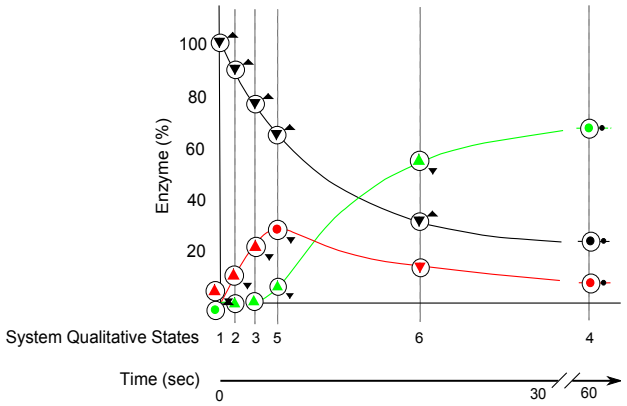

Figure 6. Enzyme evolution in M2 and M3 for the BP [1→2→3→5→6→4] placed on top on simulation curves from Cruys-Bagger et al., (2012). Red is Active, black is Free and green is Stalled enzymes.

4.3.4 Testing the sufficiency of the explanation

Results of the insufficiency test are shown in Table 5. M1 and M2 provide incomplete explanation for one of the three TBs. Particularly, M1 produces no BP with a decline of the hydrolysis rate. Indeed, following *Active enzyme* evolution, the *Catalytic rate* increases then stabilizes, which does not satisfy TB1, t_3 (Table 2). M1 produces BPs in line with TB2: addition of *Free enzyme* generates a restart of the hydrolysis process. It can also produce BPs satisfying TB2', as the reduction of *Accessible surface* due to the accumulation of *Adsorbed* and *Active enzyme*s can counteract the restart due to a second dose of *Free enzyme*. Compliance to TB2' is detailed below for M2 and M3.

Table 5. Results of the sufficiency test

Model	TB1	TB2	TB2'
M1	-	x	x
M2	x	x	-
M3	x	x	x

Fig. 6 shows that M2 provides an explanation for the decline of the *Catalytic rate* (directly proportional to the concentration of *Active enzyme*) in agreement with TB1. TB2 is assessed through a scenario that mimics the addition of *Free enzyme* in a system at the equilibrium, with a forced increase of *Free enzyme* while *Active enzyme* and *Stalled enzyme* are stable. First steps of this simulation are reported in Fig. 7a,b.

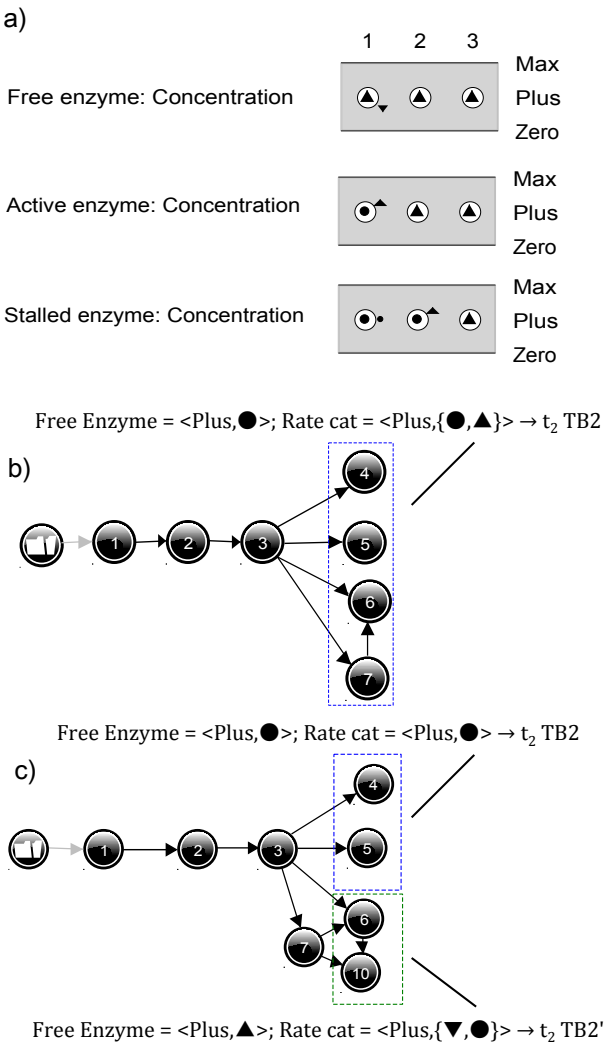

Figure 7. Partial simulation results of Restart scenario for model M2 and M3. a) Value history of the 3 first states for M2 and M3, b) first steps of M2 simulation in agreement with TB2, c) first steps of M3 simulation in agreement with TB2 and TB2'. Rate cat stands for Catalytic rate.

Addition of *Free enzyme* increases the adsorption of enzyme on the cellulose and, necessarily, brings about the increase of *Active enzyme* ([1→2→3] (Fig. 7a,b). This is consistent with TB2 (Table 3). From state 3 onwards, all possible BPs encompass the *Free enzyme* stabilization (*Free enzyme* = <Plus,●>, in states: 4, 5, 6, 7) with *Catalytic rate* = <Plus,●>, in states 4, 5 or *Catalytic rate* = <Plus, ▲> in states 6, 7. Both comply with t_2 of TB2 (Table 3) (Fig. 7b). This model implies that the second dose of

Free enzyme is completely transformed into *Active enzyme*, and causes a burst of hydrolysis anew. This behaviour was observed in concrete experiments as reported in Praestgaard et al., (2011) and in Cruys-Bagger et al., (2012). Regarding TB2' (Table 4), as shown in Fig. 7b, all the BPs produced with the restart scenario envision a stabilization of *Free enzyme* concentration prior to the stabilization of the catalytic rate, which does not match the t_2 stage. Hence, M2 does not provide an explanation for a weak restart.

M3 extends M2. M3 also meets TB1 and TB2. Regarding TB2', first steps of the simulation are given in Fig. 7c. It shows a restart of the hydrolysis in the path [1→2→3], Fig. 7a. For the next steps, some BPs satisfy TB2'. One of them starts with [1→2→3→6]. For this path *Free enzyme* = <Plus, ▲> so the addition of new enzyme is still ongoing. However, in state 6 the *Adsorption/Complexation rate* and the *Catalytic rate* stabilize (<Plus,●>), in agreement with the stage (t_2) of TB2'. The t_3 of TB2' is met in the following steps (not shown here). Given the model structure (Fig. 5), it can be inferred from *Adsorption/Complexation rate* = <Plus,●> and *Free enzyme* =<Plus, ▲> that *Accessible surface* = <Plus, ▼>. Therefore, the reduction of the *Accessible surface* limits the adsorption of *Active enzyme*, canceling out the restart phenomenon. Including *Accessible surface* in M3 does provide an explanation for a weaker restart effect.

5 Discussion

The presented work prepares the ground for a structured approach of literature integration using QR, using TB as a cornerstone. In addition, M3, shows that it is relatively simple to move from known paradigms to new ones.

Despite the fact that several mechanistic models of cellulose hydrolysis have been proposed in the literature and match experimental data, a scientist of the domain must still feel unsure about which one explains the observed the kinetics best, not to mention the parametrization or data collection techniques. This illustrates the difficult problem of verification and validation of numerical models in natural sciences (Oreskes et al., 1994). QR techniques can help overcome some of these difficulties as they focus on reproducing more abstract and generic, and accounting for diverse observations of the phenomenon.

Mechanistic interpretations, as well as observations, are available in the literature. Extraction qualitative information from selected papers has been performed manually for the work presented in this paper. Automatic composition of QR model structure is expected from treatment of natural language in the future (McFate et al., 2014).

A key question to be addressed concerns the assessment of the models, especially the genericity of the explanation they convey. Our approach used TB as reference for assessing explanations. Hence, it is the properties of the TB that determines the property of an explanation model.

References

Bredeweg, B., Linnebank, F., Bouwer, A. & Liem, J. 2009. Garp3 – Workbench for qualitative modelling and simulation. Ecological informatics, 4(5-6), 263–281.

Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K. & Westh, P. 2012. Pre-steady-state Kinetics for Hydrolysis of Insoluble Cellulose by Cellobiohydrolase Cel7A. *J. Biol. Chem*., 287(22), 18451-18458.

Eriksson, T., Karlsson, J. & Tjerneld, F. 2002. A model explaining declining rate in hydrolysis of lignocellulose substrates with cellobiohydrolase I (Cel7A) and endoglucanase I (Cel7B) of Trichoderma reesei. *Applied Biochemistry and Biotechnology*, 101(1), 41-60.

Forbus, K. 2008. Qualitative modeling. In F. Harmelen, V. Lifschitz, & B. Porter eds. *Handbook of knowledge representation (pp.* 361–394*)*. New York: Elsevier

Fraser, A.G. & Dunstan, F.D. 2010. On the impossibility of being expert. *BMJ*, 341.

Gan, Q., Allen, S.J. & Taylor, G. 2003. Kinetic dynamics in heterogeneous enzymatic hydrolysis of cellulose: An overview, an experimental study and mathematical modeling. *Proc Biochem* 38, 1003–1018.

Jalak, J. & Väljamäe, P. 2010. Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. *Biotechnology and Bioengineering*, 106(6), 871-883.

Kansou, K. & Bredeweg, B. 2014. Hypothesis assessment with qualitative reasoning: Modelling the Fontestorbes fountain. *Ecological Informatics*, 19, 71-89.

Lynd, L.R., Weimer, P.J., van Zyl, W.H. & Pretorius, I.S. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiol Mol Biol Rev.* 66, 506–577.

Maurer, S.A., Bedbrook, C.N. & Radke, C. J. 2012. Cellulase Adsorption and Reactivity on a Cellulose Surface from Flow Ellipsometry. *Ind. Eng. Chem*. Res., 51, 1389−11400.

McFate, C.J., Forbus, K. & Hinrichs, T. 2014. Using Narra-tive Function to Extract Qualitative Information from Natural Language Texts. *Proceedings of AAAI 2014*, Québec, Canada

Oreskes, N., Shrader-Frechette, K. & Belitz, K. 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, 263(5147), 641-646.

Praestgaard, E., Elmerdahl, J., Murphy, L., Nymand, S., McFarland, K.C., Borch, K. & Westh, P. 2011. A kinetic model for the burst phase of processive cellulases. *FEBS J.*, 278(278), 1547–1560.

Yang, B., Willies, D.M., & Wyman, C.E. 2006. Changes in the enzymatic hydrolysis rate of Avicel cellulose with conversion. *Biotechnology and Bioengineering*, 94(6), 1122-1128.

Zhang, Y.-H. P. & Lynd, L. R. 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering*, 88(7), 797-824.