Modeling Reactions Catalyzed by Carbohydrate-Active Enzymes

Önder Kartal, Oliver Ebenhöh

November 12, 2014

Abstract

Carbohydrate polymers are ubiquitous in biological systems and their roles are highly diverse, ranging from energy storage over mechanical stabilisation to mediating cell-cell or cell-protein interactions. The functional diversity is mirrored by a chemical diversity that results from the high flexibility of how different sugar monomers can be arranged into linear, branched or cyclic polymeric structures. Mathematical models describing biochemical processes on polymers are faced with various difficulties. First, polymer-active enzymes are often specific to some local configuration within the polymer but are indifferent to other features. That is they are potentially active on a large variety of different chemical compounds, meaning that polymers of different size and structure simultaneously compete for enzymes. Second, especially large polymers interact with each other and form water-insoluble phases that restrict or exclude the formation of enzyme-substrate complexes. This heterogeneity of the reaction system has to be taken into account by explicitly considering processes at the, often complex, surface of the polymer matrix. We review recent approaches to theoretically describe polymer biochemical systems. All attempts address a particular challenge, which we discuss in more detail. We emphasise a recent attempt which draws novel analogies between polymer biochemistry and statistical thermodynamics and illustrate how this parallel leads to novel insights about non-uniform polymer reactant mixtures. Finally, we discuss the future challenges of the young and growing field of theoretical polymer biochemistry.

1 Introduction

Prokaryotic and eukaryotic cells synthesise a large number of chemically diverse polysaccharides (also designated as glycans) that consist of a variety of monosaccharide moieties linked by inter-sugar bonds. Chemical diversity includes both the sequence of monosaccharide residues and the type of the inter-sugar linkages. As

these linkages can be made to any hydroxyl group of the monosaccharide residues, both linear and branched glycans exist but, in terms of quantity, linear structures (i.e. glycan chains) are by far dominant.[1] Furthermore, within a given glycan the number of branching types is usually low. Polysaccharides exert many distinct biological functions, such as carbon and energy storage ([60], [26], [37]), mechanical stabilisation of cells or tissues [8], cell-cell or cell-protein interactions ([42], [14]) and organelle division [59]. In addition, glycans have attracted considerable (bio)technological interest because they are being used as starting materials or additives for many technological applications [47] and act as renewable energy source [21].

Polysaccharides constitute the most abundant polymer type present in biotic systems. As compared to the vast majority of proteins and nucleic acids, cells use, however, an entirely different mode to synthesise carbohydrate polymers. This peculiarity is due to two reasons: First, no general molecular equipment (functionally equivalent to the ribosome in protein biosynthesis) exists that is capable of forming any glycan molecule provided structural information is available. Second, (and similar to the rare cases of non-ribosomal peptide biosynthesis [45]) glycans are specified by the kinetic properties of the glycan synthesising enzymes but not encoded by any non-carbohydrate system that is comparable to those in protein biosynthesis (the sequence in base triplets in genes and in their messenger RNAs). Due to the lack of appropriate enzymes, most of the theoretically possible diversity of glycans is not real in living systems.

This mode of biosynthesis has several important implications. A large number of carbohydrate-active enzymes are required to synthesise complex glycans and all these enzymes need to be encoded in the genome. Carbohydrate-active enzymes often catalyse not a single reaction but rather perform a series of closely related reactions and repetitively act on a single glycan molecule. This implies that glycan samples of natural origin usually do not consist of a single chemical species but rather are *non-uniform*.[2] Despite sharing several chemical features, such as the building blocks (i.e. the monosaccharide moieties and/or their sequence) and the types of inter-sugar linkages, glycans in a non-uniform sample have different molar masses or degrees of polymerisation (DP). As an example, the various soluble starch synthases exert distinct yet partially overlapping functions when synthesising the various chains of the amylopectin molecule [5]. Finally, carbohydrate-active enzymes often interact with small regions of the entire polysaccharide molecule. If (as it is frequently the case) a given enzyme undergoes

[1] We ignore cyclic glycans in this review since they are less relevant for higher plants.

[2] We adopt a recent IUPAC recommendation and refer to samples as being uniform instead of monodisperse, a self-contradictory term, and non-uniform instead of polydisperse, a tautological term [44].

multiple interactions with the glycan, properties of the glycan-protein complex are largely determined by the avidity of this complex rather than by the affinity describing the interaction between a single carbohydrate-binding site and a single site of the target carbohydrate. The binding site can be closely related to the catalytically active site but, in many cases, is physically separated from the latter.

For several reasons, this mode of action of many carbohydrate-active enzymes complicates the description and characterisation of these reactions. First, any empirical determination of the usual kinetic (K_m and V_{max}) and thermodynamic (K_{eq}) parameters of the series of reactions is difficult as in most cases, any individual reaction cannot be separated from the others. From a theoretical point of view, an appropriate description of these reactions requires a large number of parameters and rate equations that often cannot be empirically determined. Furthermore, the thermodynamical equilibrium of a series of related reactions is difficult to define. Finally, the enzymatic actions at the surface of insoluble substrates (such as native starch granules) can certainly not be interpreted in terms of the classical Michaelis-Menten equation or a more advanced rate law that assumes enzymes acting in homogeneous systems. These reactions take place in an inhomogeneous system and, therefore, essential parameters such as volume-based substrate or enzyme concentrations are not well defined or insufficient. Instead, structural features of the insoluble carbohydrate substrate(s) are highly relevant for the enzymatic actions.

In the following, we summarise the current knowledge on carbohydrate-protein interactions and present a theoretical approach to appropriately describe reiterating actions of carbohydrate-active enzymes on soluble and insoluble carbohydrate substrates. We do not consider kinetic features of protein complexes consisting of several enzyme activities.

2 Challenges for modelling polymer systems

The diversity of polymeric species and of the possible chemical transitions between them requires to consider more degrees of freedom than is usually the case in kinetic models. We discuss modelling strategies to address this combinatorial complexity.

Another complication arises if intra- and intermolecular interactions of polymers leads to macroscopic boundaries like the starch granule interface. The heterogeneity that is introduced by these interactions has a profound influence on the enzymatic accessibility of parts of the substrate and thus on reaction rates.

2.1 Soluble polysaccharides

Kinetic models describe the state of a biochemical reaction system by introducing suitable state variables. These are usually copy numbers ($X = X_1, \ldots, X_N$) or concentrations ($x = x_1, \ldots, x_N$) of a fixed number N of individual chemical species. The state variables span the state space whose dimension is N. There are two potential problems with this approach when applied to glycan reaction systems. First, it is difficult to define all relevant species that make up the state space, since monomers can be combined in many different ways to form a vast number of diverse species. Second, in models of open systems at least, the choice of an upper limit on glycan size is somewhat arbitrary since, for example, cogent information on a sharp limiting DP is missing. This can introduce artificial boundary effects in computer simulations that are not observed in reality. To circumvent these effects it is possible to choose a very high maximum DP, such that the numerical error is very small compared to the measurement error under given experimental conditions. If that is not sufficient computer memory can be allocated dynamically during the simulation to extend the state space 'on the fly'.

To get an idea of the combinatorial complexity of glycans we briefly discuss the possible number of structures for a (non-cyclic) polysaccharide of a given DP.

The chemical structure of polymers largely depends on the number and the type of functional groups of the constituting monomers [16]. In the case of glycans, functionality is determined by the OH-groups through which carbon atoms of monomers can condense. As an example, glucose has five OH-groups (at C1, C2, C3, C4 and C6) all of which can, in principle, be used for interglycosyl bonds. In starch-related glucans, however, interglucose bonds are restricted to C1, C4, and (less frequently) to C6. While amylose almost exclusively contains C1-C4 linkages, amylopectin additionally contains C1-C6 intergluose bonds. Thus, in amylose the monosaccharyl residues essentially undergo two types of linkages which allow only for chains of polymers. By contrast, C1-C6 interglucose bonds can account for non-linear, branched structures typical for amylopectin (and glycogen as well).

The notion of sequence is usually well-defined for linear polymers since both ends are chemically distinguishable by the functional group that is exposed (3'- and 5'-end in nucleic acids or C1 and C4 as reducing and non-reducing end(s), respectively, in polysaccharides). However, the sequence of monomers is relevant only in cases where different monomers are combined in which case we speak of heteropolymers, like DNA or heteroglycans. Unbranched homopolymers (consisting of a single type of monomer) are sufficiently determined by their DP.

Branched polymers are more complex than linear polymers in that they have at least two types of bonds and cannot be described anymore by a single sequence,

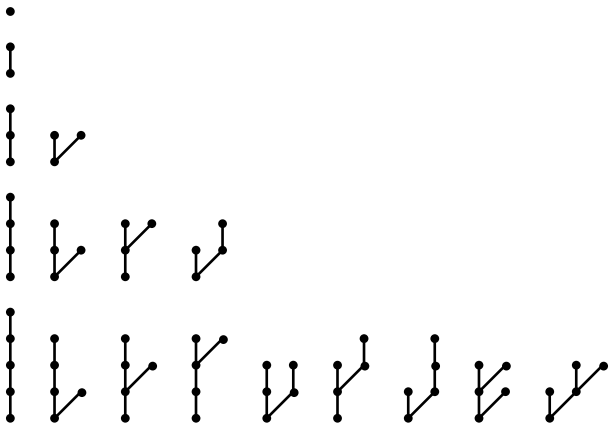
much less by a single DP. For homoglycans, an experimentally accessible observable that conveys a better description than DP of the whole polymer is the DP distribution of individual branches which can be obtained by selective enzymatic hydrolysis. Still, this distribution only partly reflects differences in the distribution of branch points within the polymer.

To illustrate the combinatorial complexity of polymers we focus on α-glucans in which a single glucosyl residue can undergo at most three interglucose bonds, i.e. at C1, C4, and C6. Typically, most internal glucosyl residues are linked to neighboring sugar residues by two α-1,4 linkages. At branching points, however, an additional α-1,6 linkage is found. Two classical examples are amylopectin and glycogen that differ, however, in the arrangement of branchings. The non-random, clustered occurrence of branch points in amylopectin leads to a different conformation and physical properties compared to glycogen, where the branch points are more randomly distributed [55]. These structural differences are correlated with their different physiological function.

To perform the counting of possible glucan structures it is useful to model the class of glucans as a mathematical object called a graph, basically a set of nodes and edges endowed with a certain relationship (connectedness). This strategy, although ignoring three-dimensional features like conformation, has a long tradition in combinatorial chemistry. A suitable model for (acyclic) glucans is a so-called *rooted, plane unary-binary tree*, in which nodes represent glucose monomers and edges the interglucose bonds.[3] Figure 1 shows all possible trees up to size DP-5, that is trees up to five nodes. The sequence in the table included in Figure 1 indicates that the number of possible glucan structures grows enormously with increasing DP. This sequence, the so-called Motzkin numbers, is well-known in combinatorics as the solution to many different combinatorial problems (A001006 in the On-line Encyclopedia of Integer Sequences) and an exact formula can be given [11, 15].

Clearly, only a tiny subset of this overwhelming number of molecular species is relevant in natural systems, and this is mainly due to enzymes that constrain possible transitions. A structure that cannot result from enzymatic activity at a reasonable time scale can be safely ignored. Thus, restriction to enzymatically possible transitions allows us to reduce the state space. Nonetheless, enzyme ac-

[3]This type of tree has a single designated root node, from which the whole tree emerges (in our case the reducing end of the glucan), at least one leaf node (the non-reducing ends) and intermediate nodes with two or three edges. If a node has a single descendant or child we assume that it is always only one type of bond (here α-1,4-glucosidic). Only in the case of two children we distinguish between them, hence we have a planar or ordered tree. If we would like to distinguish two types of bonds in general we would have to model the glucan as a *labelled* tree and this would result in a different combinatorial problem.



Motzkin Sequence	
DP	Structure Count
3	2
4	4
5	9
6	21
7	51
8	127
9	323
10	835
15	113,634
20	18,199,284
25	3,192,727,797
30	593,742,784,829

Figure 1: The combinatorial explosion of glucans. The figure shows all possible rooted and plane unary-binary trees up to DP-5. The table exemplifies the growth of the number of structures with DP as given by the Motzkin sequence (see text).

tivity can produce a highly non-uniform system even when the starting conditions are uniform. To illustrate the diversity of enzymatically catalyzed reactions, we consider state transitions that a single branched glucan can undergo in terms of changes in number of branch points k and monomers n. Table 1 summarises possible transitions and gives, for each type of transition, a carbohydrate-active enzyme (CAZyme) example from starch metabolism that can catalyse this transition under physiological conditions. Note, that the transitions referred to as grafting or cutting can be accomplished by different means, respectively. We can speak of *grafting* ($\Delta k > 0$, $\Delta n > 0$) if the glucan at hand is

- condensed through an α-1,6-bond with a branched or unbranched glucan, or
- condensed through an α-1,4-bond with a branched glucan.

Likewise, we can speak of *cutting* ($\Delta k < 0$, $\Delta n < 0$) if the glucan at hand is

- hydrolised at a α-1,6-bond (typically referred to as debranching), or
- hydrolised at an internal α-1,4-bond, such that a branched glucan is removed.

There are in principal two approaches to model large-scale polymer systems, the individual-based approach and the continuous mixture approximation.

The mechanistic, individual-based approach distinguishes each chemical species and formulates reaction rates for every reaction in the system. This leads to a set

Transition	Description	CAZyme Example
$\Delta k = 0$, $\Delta n = 0$	redistribution of branches or branch lengths	4-α-glucanotransferase
$\Delta k = 0$, $\Delta n > 0$	elongation of a branch	starch synthases, α-glucan phosphorylase
$\Delta k = 0$, $\Delta n < 0$	shortening of a branch	β-amylase, α-glucan phosphorylase
$\Delta k > 0$, $\Delta n = 0$	internal branching	branching enzymes
$\Delta k < 0$, $\Delta n = 0$	internal debranching	isoamylases, pullulanases
$\Delta k > 0$, $\Delta n > 0$	grafting with another (possibly branched) glucan	branching enzymes
$\Delta k < 0$, $\Delta n < 0$	cutting off a (possibly branched) glucan	debranching enzymes, α-amylase

Table 1: Reactions on polymers in terms of changes in number of branch points k and monomers n. In principle, the all reactions are reversible. The reactions listed in this table reflect the main direction occurring under physiological conditions.

of differential equations that describes how the polymer composition changes over time. In deterministic models, concentrations describe the composition and the time evolution is determined by ordinary differential equations. The individual-based approach to polymer dynamics goes back to Smoluchowski in the deterministic case [43, 36]. In probabilistic (or stochastic) models, each state, defined by the copy number of each component, is described by its probability. The differential equation describing the time evolution of these state probabilities is known as the chemical master equation (CME) [56, 2]. In practice, the CME is simulated by the stochastic simulation algorithm (SSA) [17] which "draws" individual trajectories for each species.

It is clear that the individual-based approach is only suitable if we know the stoichiometric coefficients and kinetic parameters of each reaction [20]. This complexity can be reduced by applying lumping techniques. Often the reactivity of a polymer depends only on the local configuration of the reactive group but not on the size and shape of the molecule more than some number of monomers away. Applying this principle of reaction shortsightedness [22] allows lumping kinetic parameters. The same principle can be applied to binding where, as a rule, the probability of forming positional isomers depends on a limited number of participating monomer units [52], that is only a small number of binding modes need to be distinguished.

The continuous mixture approximation ignores the discrete nature of the com-

ponents altogether and considers the time evolution of a concentration function $c(x)$ that varies continuously with some descriptor x of the system, like temperature or weight. Given the enormous number of species present in a mixture, it is assumed that two adjacent species differ so little that their difference can be considered infinitesimal, dx. Thus, the concentration of a polysaccharide G_i having a DP of x_i is replaced by $c(x_i)dx$, the concentration of material with DP in the interval $(x_i, x_i + dx)$. The study of the dynamics of polymer distribution functions apparently goes back to De Donder and was further developed in particular by Aris [3, 22].

2.2 Insoluble polysaccharides

In the previous section, we ignored non-covalent interactions within and between polymers. Both lead to additional complexity in describing the state of a polymer or polymer mixture. To illustrate this briefly, we focus on α-glucans but point the reader to a detailed review [9] of structural aspects of starch for further details.

In linear α-glucans one can observe conformational transitions between disordered coil and ordered helix states, depending on DP, temperature and solvent properties [6]. These pure conformations are only extremes and within long polymers helical regions may be interrupted by less ordered (melted) regions. Two helical α-glucans can form a double helix and several double helices can interact to form a crystalline phase. In starch the double helix is typically formed by adjacent amylopectin side-chains and the alignment of several double helices form the crystalline lamellae of the starch granule. These double helices can align into different configurations, which are known as allomorphs. The crystalline lamellae can melt as well – a process which is linked to starch gelatinisation [57].

From the point of view of enzyme activity, the most interesting aspect of these intermolecular interactions is the formation of interfaces. The existence and explicit incorporation of interfaces complicates modelling since the enzymes become part of a heterogeneous reaction system where the reactants are part of a different phase and not entirely accessible. Any controlled mass transfer between the aggregated and the aqueous phase (e.g. starch synthesis and breakdown) requires enzyme diffusion to and adsorption and activity at the interface. In some cases the enzyme can move on the substrate interface to act repetitively.

In heterogeneous systems, the state variables depend on spatial coordinates if diffusion or convection are significant. The resulting reaction-diffusion equations are partial differential equations (PDEs) that are more difficult to handle analytically and computationally than ODEs. The PDEs can be replaced by ODEs if the fast diffusion approximation is applied. This means that adsorption or binding at the interface is treated like the transport between two compartments. If adsorption

is reversible and is assumed much faster than substrate turnover a further simplification is possible by using adsorption isotherms. The most well-known adsorption isotherm that has also been used for modelling surface-active enzymes is the Langmuir isotherm.

3 Overview of existing models

The challenges discussed above that arise from the complexity of carbohydrate polymers such as starch have been addressed by various authors on different levels of complexity.

Rollings's review [40], apart from being a good general introduction into polymer degradation, gives an overview of deterministic modelling up to the mid 1980s. He discusses the differences between single-chain, multichain and multiple chain attack models, three action patterns that have been considered for α-amylase in pioneering studies by Robyt and French [38, 39]. He further explains how endo-acting and exo-acting enzymes lead to different product distributions and how their joint activity could lead to synergistic effects.

We will not review the models in [40] but want to point out some common features and problems. All models assume that random fission of glycosidic links basically follows Michaelis-Menten-like kinetics. Substrate multiplicity leads to inhibition terms in the denominator of the rate laws and irreversible enzyme inactivation is occasionally also considered in the models. Nearly all of the models suffer from the aforementioned combinatorial explosion, either in the form of infinite sums over substrate concentrations or many different parameters. Complicated models may, under some circumstances, give a good fit to experimental data but the general insight gained from this exercise is usually very limited. The more complicated a model, the more difficult is, as a rule, its interpretation and a parameter set that gives an exceptionally good fit for a specific experiment can miserably fail under different conditions.

For unknown reasons, Rollings omits some valuable theoretical work on depolymerizing enzymes, especially Hanson [19] who derived an analytic rate expression for short-chain cleavage, resulting in general rate laws including endo- and exohydrolases as special cases. Another early pioneer was John Thoma, who extended Hanson's work by considering endwise cleavage of mixed polymer populations and heteropolymers [53], devised mathematical models to test different attack hypotheses for α-amylase [51, 50] and who's demonstration, together with Daniel Koshland, that internal polymer segments inhibit β-amylase was important evidence for the induced-fit theory of enzyme catalysis [54].

With the similar goal to derive closed rate expressions, Chetkarov and Kolev [7]

have applied the classical approach of Michaelis and Menten [31] to enzyme-catalysed hydrolysis of homopolymers. An interesting result is their conjecture that the Michaelis constant K_M decreases with increasing DP of the substrate molecule.

The comprehensive treatment by Thoma [53] on rate laws for enzymes degrading mixed polymers by endwise cleavage demonstrates rigorously how many different phenomena may be hidden in a rate law assuming the classical Michaelis-Menten form, but that the apparent parameters of maximal rate ($V_{\max}$) and Michaelis constant (K_M) depend in general on a large variety of phenomena occurring at the molecular level. While in some cases the direction of the effect can be predicted (e.g. competitive self-inhibition, where the bonds that are not hydrolyzed act as competitive inhibitors of the reaction, will always lower the apparent K_M while the degree of repeated attack will always increase it), in many cases the effect on apparent values cannot be predicted *a priori*. Both, apparent maximal rate and Michaelis constant, depend in general on heterogeneity in monomer composition, the variation in the type of polymer distribution and on enzymatic properties such as competitive self-inhibition, the degree of repetitive attack and the occurrence of multiple intermediates.

Derivation of explicit rate laws is only possible for relatively simple systems. It is therefore not surprising that these approaches remained limited to bond hydrolysis, where the elementary reaction works on a single substrate molecule. A noteworthy approach is therefore the work by Mulders and Beeftink [32] who derived analytic expressions for enzymatic polymerisation and could show how the resulting chain length distribution depends on the Michaelis constant for the elementary elongation reaction. As a general tendency it could be shown that a Michaelis constant lower than the substrate concentration (indicating high substrate saturation) leads to narrower distributions of the degree of polymerisation.

A more complex treatment is required for example for transglycosylation reactions, where a reversible bi-bi mechanism must be assumed and both substrate molecules are of an unspecified length. Such reactions are also important in the turnover of starch. To treat such enzymes, Hiroshi Nakatani employed Monte Carlo simulations and described in a series of papers various models of the action of enzymes on soluble carbohydrate polymers. In [33, 23] a model of the action of β-amylase is presented, which takes into account the possibility of repeated attack of the enzyme without dissociation of the substrate. In [34] a conceptually similar model is discussed which describes the action of α-glucanotransferases, which include central enzymes in the starch degradation pathway such as DPE1. Finally, a model of the enzyme hyaluronidase is discussed in [35]. While the carbohydrate hyarulonan is not present in plants, the action of hyaluronidase shares common principles with many starch degrading enzymes. A particularly interesting aspect of hyaluronidase is that its possible catalytic activity includes transglycosidation,

as well as condensation and hydrolysis. In so far, the presented model can serve as a prototype also for isolated multi-enzyme systems. All these models are limited to soluble carbohydrates and describe *in vitro* systems containing a single enzyme and are numerically analyzed by simple Monte Carlo simulations.

Stochastic models have also been applied to simulate the complex soluble structure of amylopectin. In [29, 28], Marchal and coworkers model amylopectin as a matrix and demonstrate how this is used in Monte Carlo simulations to compute sugar release. This model is highly illustrative and useful to investigate how molecular mechanisms determine overall characteristics of long polymers such as chain length distribution and branching patterns. In [28] the authors applied their model to evaluate various suggested subsite patterns of α-amylase and found that the inclusion of specific inhibition terms improved the predicting power of the model.

Besides its non-uniform composition, the insoluble nature of native starch poses a major challenge for any theoretical description of starch degrading or synthesising processes.

McLaren and Packer [30] summarise important earlier attempts to model enzymatic reactions in various heterogeneous systems including the action of soluble enzymes on carbohydrates like cellulose, chitin and starch. Rollings's [40] and Zhang and Lynd's [61] reviews focus on cellulose degradation but their observations regarding the influence of adsorption and surface properties (specific surface area and surface states) are valid also for starch. In particular, both insoluble cellulose and starch have ordered and less ordered interfaces that are differently susceptible towards enzymatic attack and intercompartmental mass transfer. Bansal et al. [4] is another more recent review on models of cellulases.

To our knowledge the only review dedicated to kinetic models of starch degradation exclusively is by Dona et al. [10]. It gives an impression of classical deterministic approaches geared towards biotechnological applications rather than fundamental considerations. Here we would like to emphasise models that focus on principle features or give an interesting perspective on the general problem of heterogeneous catalysis in its entirety.

An early attempt to simulate degradation of insoluble substrate was made by Suga and coworkers [46] where they describe the degradation of an insoluble cross-linked dextran by a dextranase from *Penicillium funiculosum*. Their rather elaborate model takes into account the transport of enzymes and soluble products through pores in the insoluble substrate. Tatsumi and Katano [48, 49] developed a rate expression for the enzymatic surface hydrolysis of raw starch by glucoamylase. Their results illustrated the importance of including the specific surface area into any rate equations of surface-active enzymes and they have systematically validated their rate expressions with raw starch granules from different botanical

sources having different size distributions.

Building on such results, Kartal and Ebenhöh [24] have systematically derived a generic rate law for surface-active enzymes, which can be applied to enzymatic processes at the surface-bulk interface and can easily be generalised for specific enzymatic mechanisms. The authors demonstrated how different adsorption isotherms can be used to derive the enzyme kinetics and, due to the generality of their approach, could explain how different assumptions and different adsorption models influence the kinetic parameters, in particular the apparent Michaelis and maximal rate constants. Also, in agreement with and in extension to the previous approaches mentioned above, the generic rate law provides a quantitative relation between important experimental parameters, such as particle size distribution and specific surface area, and the apparent kinetic parameters.

An insightful mechanistic model has been presented by Levine et al. [27]. This model is purely based on standard ordinary differential equations and is applied to the degradation of cellulose, but it contains a number of interesting features of general applicability: The model adapts a procedure to map a collection of arbitrarily shaped three-dimensional objects to spheres while preserving total area, volume and hydrolysis rate. This procedure, in conjunction with random sequential adsorption (RSA) simulations, allows the authors to infer an effective footprint of cellulose degrading enzymes and to calculate that the surface-active enzyme occupied nearly twice as much surface as its physical footprint would suggest.

Another interesting aspect is addressed by the model of Fenske and coworkers [13] simulating the action of a glycosidase from *Cellulomonas fimi* as an example for an enzyme which is both exo- and endo-acting in the degradation of insoluble polysaccharides. With a Monte Carlo approach simulating random endo- and exo-attacks on a two-dimensional array, which models a surface, the authors particularly investigate whether one enzyme alone can achieve synergism, so-called autosynergism. The simulations suggest that for autosynergism the enzymes should be in close vicinity to each other on the substrate surface. However, this raises the immediate question how autosynergism might be achieved while avoiding crowding and jamming of enzymes at the surface, suggesting that for autosynergism to occur further regulatory mechanisms might be required.

To confuse matters even further and to illustrate the vast amount of specific details which must be considered when theoretically describing carbohydrate polymer biochemistry, the work of Xu and Ding [58] is worth mentioning, in which they have shown that non-Fickian diffusion, resulting from small confined spaces and crowding, leads to fractal (i.e. non-integer) kinetic orders in Michaelis-Menten like rate laws. This theory was applied specifically to the catalytic action of cellobiohydrolase.

4 The entropic approach to polymer biochemistry

The review of the numerous theoretical approaches to simulate and understand the vastly complex process of biosynthesis and degradation of insoluble carbohydrate polymers demonstrates that many different peculiarities have to be considered which are not important when building mathematical models of classical pathways in the aqueous phase. Most important are the correct description of surface-active enzymes and a suitable representation of polymeric structures. Especially the latter is a novel theoretical challenge, because clearly a straight-forward enumeration of all possible structures is neither practical nor insightful due to the combinatorial explosion in theoretically possible structures.

As a consequence, various Monte Carlo based modelling approaches have been applied to simulate the action of polymer-active enzymes [33, 34, 35, 23, 29, 28, 13] (see review in the previous section). Such approaches have the advantage that not all possible configurations have to be known a priori. However, in many of these simulations, the temporal progress of the catalytic action was simulated only as a function of the reaction coordinate (see e.g. [33, 34, 35]) and it was difficult to relate the substrate formation with the actual time passed. In [25] and [41] the Monte Carlo approach was slightly modified and simulated using a Gillespie algorithm [17], allowing to explicitly include the time coordinate. As a result, the temporal progress of various glucanotransferases and a plastidial phosphorylase could be reproduced and explained by the enzymatic mechanisms and an extremely good match to experimental *in vitro* data was obtained.

However, while stochastic simulations can provide a realistic temporal representation of the action of polymer-active enzymes, they are always limited to the particular case (including the starting conditions) to which they are applied. A more general theory explaining the action of non-substrate specific enzymes acting on polymers with arbitrary length in a wider context has recently been proposed by Kartal, Ebenhöh and coworkers [25, 12].

As a hallmark of the developed theory, it accepts the fact that the number of specific reactions catalysed by polymer-active enzymes, such as glucanotransferases, is in principle infinite and so are the different chemical structures which may serve as substrate. It then draws parallels between biochemical systems with non-uniform polymer composition and canonical ensembles in statistical thermodynamics and arrives at the conclusion that polymer-active enzymes are driven by a combination of release of enthalpy and an increase in the mixing entropy of the polymer solution. In other words, polymer-active enzymes tend to maximise the disorder by creating a maximally mixed state of different chain lengths. In the interesting special case of the glucanotransferases DPE1 and DPE2, the change of enthalpy is zero [18] and consequently the increase in mixing entropy is the only

driving force of these enzymes. The theoretically predicted equilibrium distributions have been verified experimentally in [25] with high accuracy.

The underlying conceptual idea which allows to apply principles from statistical thermodynamics to biochemical systems acting on non-uniform polymer mixtures is that different chain lengths (or degrees of polymerisation, DP) are identified with different energy states of a molecule. In a case of a simple unbranched chain with a single type of bond, the enthalpies of the bonds linking the monomers correspond to the energy state of the whole molecule. If, for example, the bond enthalpy is E, the energy state of a linear polymer consisting of n monomers equals $E_n = (n-1)\cdot E$. Thus, all possible configurations (chains of length $1,2,\ldots$) are represented by equidistant energy states with the ground state $E_1 = 0$ corresponding to a monomer.

As a prototype consider the reaction catalysed by disproportionating enzyme 1 (DPE1), a plastid-located glucanotransferase involved in starch metabolism. This enzyme catalyses the transfer of 1, 2 or 3 glucosyl residues from one unbranched malto-oligosaccharide to another, a reaction that can be written as

$$\mathrm{G_n} + \mathrm{G_m} \rightleftharpoons \mathrm{G_{n-q}} + \mathrm{G_{m+q}} \quad \text{with} \quad q = 1,2,3, \tag{1}$$

where $\mathrm{G_k}$ denotes an unbranched, α-1,4 linked glucan. In this special case, the catalysed reactions do not change the overall enthalpy, because in every catalytic step one bond will be opened and another one closed, and, moreover, the bond enthalpy is independent on DP and position within the polymer [18]. The action of DPE1 in the thermodynamic picture is schematically depicted in Fig. 2 for a case in which the reaction is initialised with a uniform solution of glucans of DP 8. According to Eq. (1), the catalytic action of DPE1 on a statistical ensemble of energy states corresponds to the simultaneous downward shift of one molecule from an energy state E_n to E_{n-q} (donor reaction) and an upward shift of another molecule from energy state E_m to E_{m+q} (acceptor reaction), where $q = 1,2,3$.

In statistical thermodynamics, the principle of maximum entropy provides the method to calculate the equilibrium distribution of the occupation of each energy state depending on the total energy within the system. In physical systems (such as gases or atomic ensembles), the total energy is given by the temperature of the system. In a biochemical system in which the total enthalpy is conserved (such as for DPE1), the total energy is given by the (conserved) total number of bonds between monomers in the system. This quantity can be controlled through the experimental starting conditions. If, for example, an *in vitro* assay of DPE1 is incubated with maltopentaose molecules only, there are $b = 4$ interglucose bonds per molecule and this average number will remain constant in time. Thus, in this particular scenario, the average bond number is analogous to the temperature in physical systems and

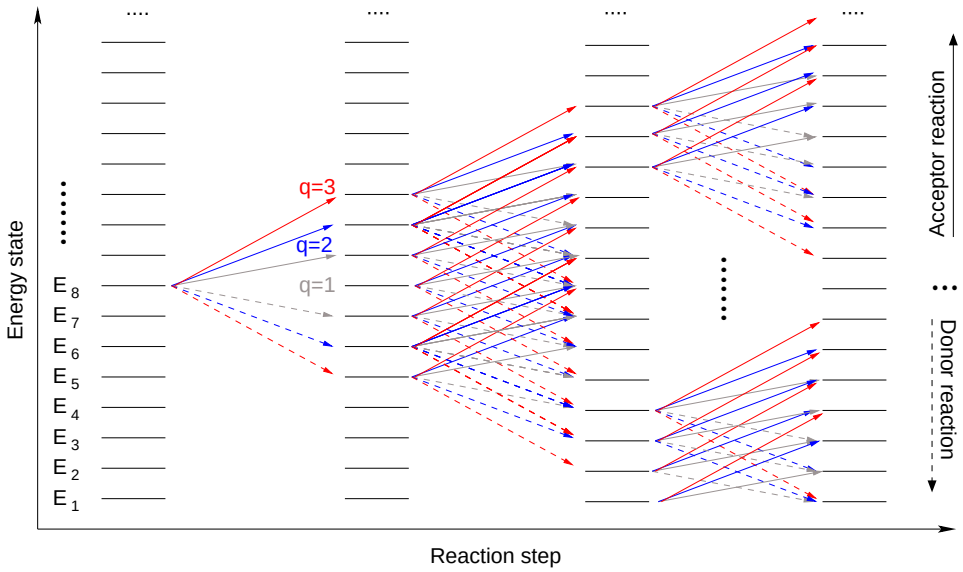


Figure 2: Scheme of the DPE1 mediated reaction system in the statistical thermodynamics picture. The energy states E_k correspond to α-1,4 linked glucans of DP k. DPE1 mediates transfers of glucose, maltose and maltotriose units, i.e. $q = 1,2,3$. In each reaction step the system follows an arbitrary dashed and solid arrow of the same colour simultaneously. This leads to a combinatorial explosion of the reaction system.

all known formulae from statistical thermodynamics can directly be applied to determine the equilibrium distribution of the enzymatic system. Thus, for polymer solutions the principle of entropy maximisation states that the action of an enzyme will continue until the distribution of DPs within the solution is maximally mixed, as is characterised by a maximal value of the mixing entropy

$$S_{\text{mix}} = -\sum x_k \ln x_k, \tag{2}$$

where x_k denotes the molar fraction of a polymer of DP k. A short calculation to determine the maximal value of S_{mix} under the constraints that the total number of molecules is conserved ($\sum x_k = 1$) and the total number of bonds is conserved ($\sum (k-1) \cdot x_k = b$) leads to the prediction that in equilibrium the molar fractions of the different DPs are distributed as

$$x_k \propto e^{-k \cdot \beta}, \tag{3}$$

where x_k is molar fraction of molecules with DP k, and $\beta = \ln((b+1)/b)$, and b the average bond per molecule.

This formula quantifies precisely how the equilibrium distribution depends on the initial conditions and, moreover, leads to a conceptual advance by introducing a novel constant, β, which is a generalisation of the classical equilibrium constant for the case of non-uniform polymeric systems.

While in the case of DPE1, the theory is presented in its simplest form and the analogies between polymer biochemistry and statistical thermodynamics become most clear, it is of general validity and thus applicable to a wide spectrum of systems, as was demonstrated in [25, 12, 41]. Besides providing experimental evidence for the soundness of the theoretical concepts, Kartal et al. [25] proved in the accompanying supplementary information that the thermodynamic formulae can be deductively derived from first principles in the case of a mixture of dilute solutions. This derivation is highly illustrative because it shows the mathematical formulation in its most general form for an arbitrary biochemical reaction system.

For the generalisation of the theory, various issues have to be taken into account. In many cases enthalpies are not conserved. For example, the reaction catalysed by phosphorylase, that reversibly transfers the non reducing glucosyl residue of an α-glucan to an orthophosphate according to the formula

$$\mathrm{G_n + P_i \rightleftharpoons G_{n-1} + G1P}, \tag{4}$$

opens an α-1,4 glucosidic linkage while forming a phosphoesther bond with quite a different bond enthalpy. In such a case, the biochemical system is analogous to a closed system in statistical physics, rather than an isolated system as was the case for DPE1. From the analogies it follows directly, that now an energetic and an entropic term have to be considered. The progress of the biochemical reaction (4) from left to right will lead to a release of enthalpy and concomitantly to a combination of reactants with a lower total Gibbs energy of formation. Kartal et al. [25] have shown that the correct way to predict the equilibrium distribution is by minimising the Gibbs free energy of the system which is related to the mixing entropy by

$$G = G^f - T \cdot S_{\mathrm{mix}}, \tag{5}$$

where G^f is the total summed Gibbs energy of formation of all reactants and T is the temperature.

Another important aspect is to consider possible additional constraints imposed by the enzymatic mechanisms. Disproportionating enzyme 2 (DPE2), a cytosol-located glucanotransferase involved in maltose metabolism, catalyses a reaction according to the formula

$$\mathrm{G_n + G_m \rightleftharpoons G_{n-1} + G_{m+1}} \quad \text{with} \quad n \neq 3, m \neq 2, \tag{6}$$

with the critical limitation that maltose cannot act as an acceptor ($m \neq 2$) and maltotriose can never act as donor molecule ($n \neq 3$). This limitation results in the additional constraint that the sum of the molar fractions of glucose and maltose is constant ($x_1 + x_2 = m$). Performing the entropy maximisation as in the case of DPE1, but with this additional constraint considered, leads to the correct prediction of the equilibrium distribution, as was also experimentally demonstrated in [25]. The difference to the case of DPE1 is illustrated in Fig. 3. Similar to DPE1, the

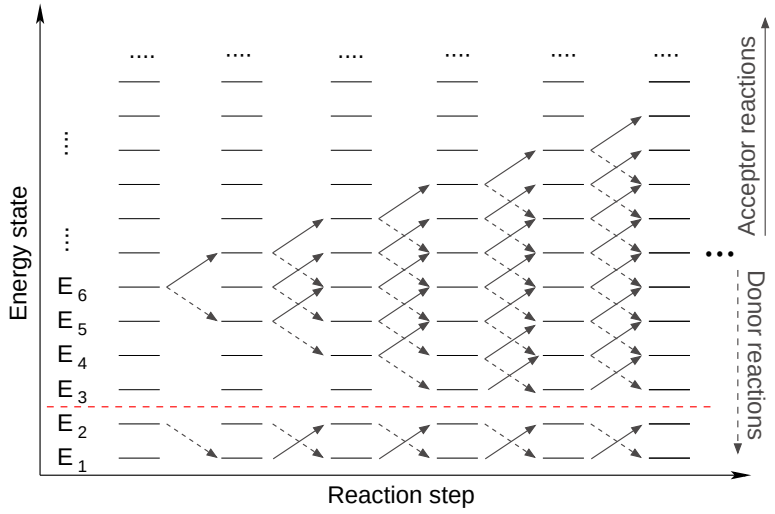


Figure 3: Scheme of the DPE2 mediated reaction system. Each DPE2 reaction step consists of one donor and one acceptor reaction depicted by a dashed and a solid arrow, respectively. Due to the restriction that maltose is never an acceptor and maltotriose is never a donor, the maltose and glucose pool is separated from the other DPs as shown by the horizontal dashed line. The scheme exhibits all possible reaction pathways starting from the two indicated initial substrates maltohexaose and maltose, where in each step one arbitrary solid and one arbitrary dashed path is taken.

catalytic action of DPE2 corresponds in the thermodynamic picture to a simultaneous occurrence of an arbitrary donor reaction (dashed arrows) and one arbitrary acceptor reaction (solid arrows). However, since maltose cannot act as an acceptor and maltotriose cannot act as donor, no arrow can cross the dashed line, resulting in a separation in two pools, one containing glucose and maltose molecules and the other all longer glucans. Fig. 3 illustrates the case in which the reaction is initialised with maltose and maltohexaose molecules only.

All the systems discussed above conserve the total number of reactants, because all elementary reactions are reversible bi-bi reactions, consuming two molecules and producing two molecules. In principle, the theory is also applicable to systems not conserving the number of molecules, but in this case the difficulty arises that the ratio between concentrations and molar fractions is no longer constant. This leads to some changes in the formulae, as has been derived and laid out in [25, 12]. However, the theory has up to now not been applied and verified experimentally for systems not conserving the total number of molecules.

Although the theory in its present form only allows to make precise predictions about the equilibrium states, and living systems are always far from equilibrium, the theoretical concepts nevertheless provide significant insight into the principles how polymer-active enzymes work. Firstly, knowledge of the equilibrium is, as for classical enzymes, prerequisite to determine in which direction a reaction will proceed and to evaluate how far from equilibrium an experimentally determined physiological state actually is. Moreover, Ebenhöh et al. [12] have shown how the theoretical knowledge can be used to indirectly determine bond enthalpies from measured equilibrium distributions.

5 Open problems and conclusions

Biosynthesis and degradation of complex and insoluble polymers, such as starch or cellulose, are multi-faceted processes which are challenging to describe in mathematical models and simulations. What makes the description of these processes more difficult than modelling classical pathways that occur in solution are mainly two facts: 1) the insoluble nature and large size make it necessary to distinguish between enzymatic processes occurring in solution and those taking place at the, probably very complicated, surface of the substrate; 2) the unlimited flexibility in combining monomers into long polymers leads to a combinatorial explosion in the numbers of theoretically possible molecular structures that appear as reactants in the biochemical pathways. Starch synthesis and degradation are excellent examples in which inhomogeneous phases and soluble polymers both play a central role. Therefore, developing a comprehensive model of starch metabolism confronts us with both of these challenges simultaneously.

The various approaches to theoretically describe and simulate processes on substrate surfaces are very promising and it appears that the difficulties to include surface-active enzymes into pathways models can soon be overcome, in particular thanks to early pioneering work [46] and the development of more and more general rate laws for surface-active enzymes [48, 49, 24]. On the theoretical and modelling side, the key issues here will be to derive simplified but sufficiently accu-

rate descriptions of the insoluble reactants and their surfaces, and to make plausible assumptions over the different adsorption models of the involved proteins, to simulate adequate available area functions, which are the key to correctly represent competition and crowding effects on the substrate surface. However, also experimental efforts are necessary to support the development of comprehensive pathway models. As was demonstrated in the various theoretical works developing surface-active rate laws, interpretation of *in vitro* data has to be performed with great care. Apparent turnover rates and Michaelis constants depend on a number of factors, which do not have to be considered for enzymes acting in solution. The specific rate, for example, decreases with increasing enzyme concentration and, moreover, depends also on the presence of other enzymes acting on the same surface. Such a dependency on enzyme concentrations does not exist in bulk solution. Further, the specific rate increases with increasing specific surface area, and the Michaelis constant increases with decreasing specific surface area and with increasing total enzyme concentration [24]. Consequently, *in vitro* experiments, in which the controllable parameters such as specific surface area and enzyme concentrations are systematically varied, are necessary to parameterise the generic rate equations.

Major challenges are posed for the simulation, and even more for the theoretical understanding, of biochemical reactions in non-uniform and complex reactants. Despite the recent progress by successfully finding analogies between polymer biochemical systems and statistical ensembles [25], it is apparent that a further development of the theoretical concepts is still necessary. Whereas it is now understood how equilibrium distributions generated by non substrate-specific polymer-active enzymes can be explained and predicted, the biologically relevant far-from-equilibrium steady states still evade a prediction from first principles. It is therefore evident that a focus of future theoretical activities in polymer biochemistry research must lie in an advancement of our fundamental understanding which factors govern the dynamics of polymer-active enzymatic processes. With the established parallels between polymer biochemistry and statistical thermodynamics it is well possible that both scientific fields will mutually benefit from each other, because experimentally observed dynamics in simple *in vitro* systems can now in principle conversely be employed to draw conclusions about non-equilibrium thermodynamic physical systems. Non-equilibrium thermodynamics is still a very active area of research and the analogies should be further elaborated to ensure that novel insights gained in physics can lead to a deeper understanding of polymer biochemistry.

Notwithstanding our lack of understanding of far-from-equilibrium states, Monte-Carlo simulations [33, 34, 35] and Gillespie algorithms [25, 41] allow for a precise prediction and explanation of the temporal evolution of polymer mixtures in *in vitro* experiments towards equilibrium. Our new theoretical understanding of

polymer-active enzymes allows us now to define adequate kinetic parameters. In order to transgress from mimicking relatively simple *in vitro* systems to the biologically more interesting *in vivo* situation it will now be necessary to integrate the theoretically supported stochastic simulations into whole pathway models. Here, several difficulties can be expected. If, as is commonly the case, reactions in solution are modelled by ordinary differential equations, then the integration of these two fundamentally different types of simulation techniques is far from trivial [1]. If, on the other hand, also all reactions in bulk solution were to be modelled by stochastic simulations, a considerable increase in computation time can be expected and it is questionable whether such a pathway model could still be simulated on a standard desktop PC in reasonable time.

In conclusion, the recent advancements in theory building and model development regarding carbohydrate polymer metabolism have addressed all major problems and difficulties and solutions for most aspects have been proposed. We are therefore currently in the exciting situation that the single pieces and building blocks are at hand – at least in a prototype form – but that comprehensive mathematical models combining the various approaches, in order to simulate for example the synthesis of a starch granule, do not yet exist. One focus of theoretical research in carbohydrate metabolism should therefore lie in the development of integrative pathway models, in which processes at surfaces and in solution are combined and polymeric diversity is adequately represented. These models can then serve as valuable tools to first reproduce in simulations complex processes as starch granule synthesis and maturation, and later query in *in silico* experiments the effect of genetic and environmental perturbations in order to arrive at a comprehensive understanding how physiological regulation is accomplished with highly heterogeneous and disperse components.

References

[1] Aurélien Alfonsi, Eric Cancès, Gabriel Turinici, Barbara Di Ventura, and Wilhelm Huisinga. Exact simulation of hybrid stochastic and deterministic models for biochemical systems. Research Report RR-5435, INRIA, 2004.

[2] Steven S. Andrews, Tuan Dinh, and Adam P. Arkin. Stochastic models of biological processes. In Robert A. Meyers, editor, *Encyclopedia of Complexity and Systems Science*, pages 8730–8749. Springer New York, 2009.

[3] Rutherford Aris. Reactions in continuous mixtures. *AIChE J.*, 35(4):539–548, 1989.

[4] Prabuddha Bansal, Mélanie Hall, Matthew J Realff, Jay H Lee, and Andreas S Bommarius. Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnol Adv*, 27(6):833–848, 2009.

[5] Henrike Brust, Slawomir Orzechowski, Joerg Fettke, and Martin Steup. Starch synthesizing reactions and paths: in vitro and in vivo studies. *J. Appl. Glycosci.*, 60(1):3–20, 2013.

[6] A. Cesáro, F. Sussich, and L. Navarini. Order-disorder conformational transitions of carbohydrate polymers. In D. Lörinczy, editor, *The nature of biological systems as revealed by thermal methods*, chapter 1. Kluwer Academic Publishers, 2004.

[7] Methodi L. Chetkarov and Dimiter N. Kolev. The *Michaelis-Menten* equation in the case of enzyme-catalyzed hydrolysis of linear homopolymer substrates with different degrees of polymerization. *Monatshefte für Chemie / Chemical Monthly*, 115:1405–1412, 1984. 10.1007/BF00816339.

[8] Daniel J Cosgrove. Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.*, 6(11):850–861, Nov 2005.

[9] Iben Damager, Søren Balling Engelsen, Andreas Blennow, Birger Lindberg Møller, and Mohammed Saddik Motawia. First principles insight into the α-glucan structures of starch: their synthesis, conformation, and hydration. *Chem. Rev.*, 110(4):2049–2080, Apr 2010.

[10] A. C. Dona, G. Pages, R. G. Gilbert, and P. W. Kuchel. Digestion of starch: *In vivo* and *in vitro* kinetic models used to characterise oligosaccharide or glucose release. *Carbohydr. Polym.*, 80(3):599–617, May 2010.

[11] Robert Donaghey. Restricted plane tree representations of four Motzkin-Catalan equations. *Journal of Combinatorial Theory, Series B*, 22(2):114–121, Apr 1977.

[12] Oliver Ebenhöh, Alexander Skupin, Önder Kartal, Sebastian Mahlow, and Martin Steup. Thermodynamic characterisation of carbohydrate-active enzymes. In *Proceedings of the "5th International ESCEC Symposium on Experimental Standard Conditions of Enzyme Characterizations"*, 2013.

[13] J. J. Fenske, M. H. Penner, and J. P. Bolte. A simple individual-based model of insoluble polysaccharide hydrolysis: the potential for autosynergism with dual-activity glycosidases. *J. Theor. Biol.*, 199(1):113–118, Jul 1999.

[14] Joshua Finkelstein. Glycochemistry & glycobiology. *Nature*, 446(7139):999–999, 2007.

[15] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[16] Paul J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, 1953.

[17] D.T: Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 8:2340–2354, 1977.

[18] R. N. Goldberg, D. Bell, Y. B. Tewari, and M. A. McLaughlin. Thermodynamics of hydrolysis of oligosaccharides. *Biophys Chem*, 40(1):69–76, Apr 1991.

[19] K. R. Hanson. Enzyme kinetics of short-chain polymer cleavage. *Biochemistry*, 1:723–734, Sep 1962.

[20] R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Chapman & Hall, 1996.

[21] Michael E Himmel, Shi-You Ding, David K Johnson, William S Adney, Mark R Nimlos, John W Brady, and Thomas D Foust. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, 315(5813):804–807, Feb 2007.

[22] Teh Ho. Kinetic modeling of large-scale reaction systems. *Catalysis Reviews*, 50:287–378(92), 2008.

[23] Kazuhiko Ishikawa, Hiroshi Nakatani, Yoshio Katsuya, and Chikafusa Fukazawa. Kinetic and structural analysis of enzyme sliding on a substrate: multiple attack in β-amylase. *Biochemistry*, 46(3):792–798, Jan 2007.

[24] Önder Kartal and Oliver Ebenhöh. A generic rate law for surface-active enzymes. *FEBS Letters*, 587(17):2882 – 2890, 2013.

[25] Önder Kartal, Sebastian Mahlow, Alexander Skupin, and Oliver Ebenhöh. Carbohydrate-active enzymes exemplify entropic principles in metabolism. *Mol. Syst. Biol.*, 7:542, October 2011.

[26] Peter L Keeling and Alan M Myers. Biochemistry and genetics of starch synthesis. *Annu. Rev. Food Sci. Technol.*, 1:271–303, 2010.

[27] Seth E. Levine, Jerome M. Fox, Harvey W. Blanch, and Douglas S. Clark. A mechanistic model of the enzymatic hydrolysis of cellulose. *Biotechnol. Bioeng.*, 107:37–51, 2010.

[28] L. M. Marchal, R. V. Ulijn, C. D. De Gooijer, G. T. Franke, and J. Tramper. Monte Carlo simulation of the α-amylolysis of amylopectin potato starch. 2. α-amylolysis of amylopectin. *Bioprocess Biosyst. Eng.*, 26(2):123–132, Dec 2003.

[29] L. M. Marchal, J. Zondervan, J. Bergsma, H. H. Beeftink, and J. Tramper. Monte Carlo simulation of the α-amylolysis of amylopectin potato starch. Part I: modeling of the structure of amylopectin. *Bioprocess Biosyst. Eng.*, 24:163–170, 2001. 10.1007/s004490100247.

[30] A. D. McLaren and L. Packer. Some aspects of enzyme reactions in heterogeneous systems. *Adv. Enzymol. Relat. Areas Mol. Biol.*, 33:245–308, 1970.

[31] L. Michaelis and M.L Menten. Kinetik der Invertinwirkung. *Biochem. Z.*, 49:333–369, 1913.

[32] K. J. M. Mulders and H. H. Beeftink. Chain length distribution and kinetic characteristics of an enzymatically produced polymer. *e Polymers*, 13(1):261–272, 2013.

[33] H. Nakatani. Monte Carlo simulation of multiple attack mechanism of β-amylase-catalyzed reaction. *Biopolymers*, 42(7):831–836, Dec 1997.

[34] H. Nakatani. Monte Carlo simulation of 4-α-glucanotransferase reaction. *Biopolymers*, 50(2):145–151, Aug 1999.

[35] Hiroshi Nakatani. Monte Carlo simulation of hyaluronidase reaction involving hydrolysis, transglycosylation and condensation. *Biochem. J.*, 365(Pt 3):701–705, Aug 2002.

[36] Douglas Poland. Time evolution of polymer distribution functions from moment equations and maximum-entropy methods. *J. Chem. Phys.*, 111(17):8214–8224, 1999.

[37] Peter J. Roach, Anna A. Depaoli-Roach, Thomas D. Hurley, and Vincent S. Tagliabracci. Glycogen and its metabolism: some new developments and old themes. *Biochem. J.*, 441(3):763–787, Feb 2012.

[38] John F Robyt and Dexter French. Multiple attack hypothesis of α-amylase action: Action of porcine pancreatic, human salivary, and aspergillus oryzae α-amylases. *Arch. Biochem. Biophys.*, 122(1):8–16, 1967.

[39] John F Robyt and Dexter French. The action pattern of porcine pancreatic α-amylase in relationship to the substrate binding site of the enzyme. *J. Biol. Chem.*, 245(15):3917–3927, 1970.

[40] James Rollings. Enzymatic depolymerization of polysaccharides. *Carbohydr. Polym.*, 5(1):37 – 82, 1985.

[41] Christian Ruzanski, Julia Smirnova, Martin Rejzek, Darrell Cockburn, Henriette L Pedersen, Marilyn Pike, William G T Willats, Birte Svensson, Martin Steup, Oliver Ebenhöh, Alison M Smith, and Robert A Field. A bacterial glucanotransferase can replace the complex maltose metabolism required for starch to sucrose conversion in leaves at night. *J. Biol. Chem.*, 288(40):28581–28598, Oct 2013.

[42] Peter H Seeberger. Exploring life's sweet spot. *Nature*, 437(7063):1239, Oct 2005.

[43] Marian von Smoluchowski. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Zeitschrift f. phys. Chemie*, 92:129–168, 1917.

[44] Robert F. T. Stepto. Dispersity in polymer science (IUPAC Recommendations 2009). *Pure Appl. Chem.*, 81(2):351–353, 2009.

[45] Matthias Strieker, Alan Tanović, and Mohamed A Marahiel. Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol*, 20(2):234–240, Apr 2010.

[46] K. Suga, G. Van Dedem, and M. Moo-Young. Enzymatic breakdown of water insoluble substrates. *Biotechnol. Bioeng.*, 17:185–201, 1975.

[47] T. Takaha and S. M. Smith. The functions of 4-α-glucanotransferases and their use for the production of cyclic glucans. *Biotechnol. Genet. Eng. Rev.*, 16:257–280, 1999.

[48] H. Tatsumi and H. Katano. Kinetics of the surface hydrolysis of raw starch by glucoamylase. *J. Agric. Food Chem.*, 53:8123–8127, 2005.

[49] Hirosuke Tatsumi, Hajime Katano, and Tokuji Ikeda. Kinetic analysis of glucoamylase-catalyzed hydrolysis of starch granules from various botanical sources. *Biosci., Biotechnol., Biochem.*, 71(4):946–950, Apr 2007.

[50] J. A. Thoma. Models for depolymerizing enzymes. Application to α-amylases. *Biopolymers*, 15:729–746, 1976.

[51] J. A. Thoma. Models for depolymerizing enzymes: criteria for discrimination of models. *Carbohydr. Res.*, 48(1):85–103, May 1976.

[52] J. A. Thoma, G. V. Rao, C. Brothers, J. Spradlin, and L. H. Li. Subsite mapping of enzymes. Correlation of product patterns with Michaelis parameters and substrate-induced strain. *J. Biol. Chem.*, 246(18):5621–5635, Sep 1971.

[53] John A. Thoma. The kinetics of enzyme degradation of mixed polymer systems. endwise cleavage*. *Biochemistry (Mosc.)*, 5(4):1365–1374, 1966.

[54] John A. Thoma and D. E. Koshland. Competitive inhibition by substrate during enzyme action. evidence for the induced-fit theory. *J. Am. Chem. Soc.*, 82(13):3329–3333, 1960.

[55] D. B. Thompson. On the non-random nature of amylopectin branching. *Carbohydr. Polym.*, 43(3):223–239, November 2000.

[56] T. E. Turner, S. Schnell, and K. Burrage. Stochastic approaches for modelling *in vivo* reactions. *Comput. Biol. Chem.*, 28(3):165–178, Jul 2004.

[57] T. A. Waigh, M. J. Gidley, B. U. Komanshek, and A. M. Donald. The phase transformations in starch during gelatinisation: a liquid crystalline approach. *Carbohydr. Res.*, 328:165–176, 2000.

[58] Feng Xu and Hanshu Ding. A new kinetic model for heterogeneous (or spatially confined) enzymatic catalysis: Contributions from the fractal and jamming (overcrowding) effects. *Applied Catalysis A: General*, 317(1):70–81, January 2007.

[59] Yamato Yoshida, Haruko Kuroiwa, Osami Misumi, Masaki Yoshida, Mio Ohnuma, Takayuki Fujiwara, Fumi Yagisawa, Shunsuke Hirooka, Yuuta Imoto, Kazunobu Matsushita, Shigeyuki Kawano, and Tsuneyoshi Kuroiwa. Chloroplasts divide by contraction of a bundle of nanofilaments consisting of polyglucan. *Science*, 329(5994):949–953, Aug 2010.

[60] Samuel C Zeeman, Jens Kossmann, and Alison M Smith. Starch: its metabolism, evolution, and biotechnological modification in plants. *Annu. Rev. Plant Biol.*, 61:209–234, Jun 2010.

[61] Y.-H. P. Zhang and L. R. Lynd. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol. Bioeng.*, 88:797–824, 2004.