Supplementary Data

Pre-steady State Kinetics for the Hydrolysis of Insoluble Cellulose by Cellobiohydrolase Cel7A

Nicolaj Cruys-Bagger[1], Jens Elmerdahl[1], Eigil Praestgaard[1], Hirosuke Tatsumi[2], Nikolaj Spodsberg[3], Kim Borch[3] and Peter Westh[1]

[1]Research Unit for Functional Biomaterials, NSM, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark [2]International Young Researchers Empowerment Center, Shinshu University, Matsumoto, Nagano 390-8621, Japan [3]Novozymes A/S, Krogshøjvej 36, 2880 Bagsvaerd, Denmark

Corresponding Author: pwesth@ruc.dk

All equations in the Supplementary Data will specified by "S" and numbers in sharp brackets (*e.g.* eq. <S1>). Equation numbers in standard brackets (as in eq. (1)) refer to the main article.

Modeling principles. The ordinary differential equations (ODE) pertaining to eq. (1) in the main manuscript are written in the usual way (where the prime mark identifies the time derivative):

$$\begin{aligned} EC_m'(t) &= k_{on}E(t)C_m(t) - EC_m(t)\left(k_{cat}+k_{off}\right) \\ EC_{m-1}'(t) &= k_{cat}EC_m(t) - EC_{m-1}(t)\left(k_{cat}+k_{off}\right) \\ EC_{m-2}'(t) &= k_{cat}EC_{m-1}(t) - EC_{m-2}(t)\left(k_{cat}+k_{off}\right) \\ &etc \end{aligned} \qquad \text{<S1>}$$

This set of ODEs can be treated numerically or solved analytically, if we assume that the substrate concentration, C_m, is constant over the (short) experimental time. As argued in the main manuscript this assumption is acceptable for the data in Figs. 1-2. The analytical solution to the cumulated concentration of "active intermediates" (*i.e.* all intermediates except the blocked EC_{m-n}) is (1)

$$\sum_{i=0}^{n-1} EC_{m-i}(t) = \frac{(1-\mathrm{e}^{-(k_{off}+k_{on}C_m)t})E_0k_{on}C_m}{(k_{off}+k_{on}C_m)} + \frac{E_0\left(\frac{k_{cat}}{k_{cat}+k_{off}}\right)^n k_{on}C_m\left(-1+\frac{\Gamma[(n-1),(k_{cat}+k_{off})t]}{\Gamma[n-1]}\right)}{(k_{off}+k_{on}C_m)} + \frac{1}{(k_{off}+k_{on}C_m)}\mathrm{e}^{-(k_{off}+k_{on}C_m)t}E_0k_{on}C_m\left(\frac{k_{cat}}{k_{cat}-k_{on}C_m}\right)^n\left(1-\frac{\Gamma[(n-1),(k_{cat}-k_{on}C_m)t]}{\Gamma[n-1]}\right) \qquad \text{<S2>}$$

where $\Gamma[n,xt] = \int_x^\infty t^{n-1}e^{-t}\,dt$ is the so-called upper incomplete gamma function. The reaction scheme in eq. (1) stipulates that all cellobiose is produced through the decay of these "active intermediates" ($\sum_{i=0}^{n-1} EC_{m-i}$), which all have the same decay constant, k_{cat}. Hence, the rate of

cellobiose production is $C'_{CB}(t) = k_{cat}\sum_{i=0}^{n-1} EC_{m-i}(t)$, and integration of this expression from t=0 will give the temporal development in the cellobiose concentration, $C_{CB}(t)$. Before performing the integration, we note that some sums and products of parameters occur repeatedly in eq. <S2>; k_{on} and C_m, for example, occur exclusively as their product. This offers a way to shorten the expression and thus facilitate the subsequent regression analysis. Specifically, we found that a useful re-parameterization was the introduction of two "composite parameters" A and B defined as $A = k_{cat} + k_{on}C_m$ and $B = k_{on}C_m$.

Integration of the re-parameterized expression in *Mathematica 8.0* (Wolfram Research, Inc. Champaign, IL) yields:

$$C_{CB}(t) = E_0 \left[\frac{k_{cat}B(-1+\mathrm{e}^{-At}+At)}{A^2} - \frac{\mathrm{e}^{(-A-k_{cat}+B)t}k_{cat}\left(\frac{k_{cat}}{A+k_{cat}-B}\right)^n Bt((A+k_{cat}-B)t)^{-1+n}}{AGamma[n]} - \right.$$
$$\frac{\mathrm{e}^{-At}k_{cat}\left(\frac{k_{cat}}{k_{cat}-B}\right)^n B\left(1-\frac{Gamma[n,(k_{cat}-B)t]}{Gamma[n]}\right)}{A^2} + \frac{k_{cat}\left(\frac{k_{cat}}{A+k_{cat}-B}\right)^n B\left(1-\frac{Gamma[n,(A+k_{cat}-B)t]}{Gamma[n]}\right)}{A^2} +$$
$$\left. \frac{k_{cat}\left(\frac{k_{cat}}{A+k_{cat}-B}\right)^n B(n-(A+k_{cat}-B)t)\left(1-\frac{Gamma[n,(A+k_{cat}-B)t]}{Gamma[n]}\right)}{A(A+k_{cat}-B)} \right]$$

<S3>

This expression for the cellobiose concentration, $C_{CB}(t)$, was used to deconvolute experimental data using different non-linear regression routines in *Mathematica*. In the simplest approach each experiment (with a certain concentration of enzyme, E_0, and substrate, C_m) was analyzed separately to obtain maximum likelihood parameters for k_{cat}, n, A and B. Values of k_{on} and k_{off} were subsequently calculated from the definitions of A and B stated above. We note that unlike k_{cat} and k_{off}, k_{on} is a second order rate constant and its value therefore depends on the choice of concentration unit for the substrate. This is readily handled in eq. <S3> because the substrate concentration does not occur explicitly in the fitting expression (it is included in the A parameter). Therefore, the substrate concentration is introduced after the regression when k_{on} and k_{off} are calculated from A and B. The value of k_{on} in Tab 1 is based on a substrate concentration in units of g/l.

Parameter dependence. As shown in Figs. 1-2 of the main manuscript, the simple analysis of separate experimental trials provided good fits to the data with consistent parameters in experiments with different enzyme- and substrate concentration. Except at the lowest substrate concentrations we found moderate parameter dependence; average correlation coefficients for the data in Fig 1 are shown in bold in Table S1.

Parameter	A	B	k_{cat}
A	-		
B	**0.60**, 0.80	-	
k_{cat}	**-0.68**, -0.86	**-0.92**,-0.91	-
n	**0.55**, 0.09	**-0.45**, -0.51	**0.31**, 0.35

Table S1: Correlation matrix for the simple regressions shown in Fig. 1 (bold font) and the global analysis described below and shown in Fig S1 (normal font). For the former, the six trials in Fig. 1 were analyzed separately and the coefficients are averages for these regressions.

To assess a possible role of parameter dependence further, we also developed a global analysis principle, where all trials in a series with different enzyme concentration were analyzed simultaneously. In other words, we used the function $C_{CB}(t, E_0)$ rather than $C_{CB}(t)$ for separate E_0-values in the analysis. This three dimensional regression provides better separation of parameters, but except for the *A*-*n* interdependence, improvements in the correlation coefficients turned out to be limited (Tab. S1). Results from global analysis of the data in Fig. 1 of the main manuscript is shown below in Fig. S1. The parameters found in the 3D analysis were equivalent (to within the experimental precision) to the average values derived from the simple approach, and we conclude that easier 2D analysis is acceptable. This is true except at low substrate concentrations (below ~0.5 g/l for the system investigated here), where the initial changes in the $C_{CB}(t)$- traces grow fainter (*c.f.* Fig 2 in the main manuscript). With this loss of a strong inflection around t=5 s, four parameters cannot be derived from a single run.

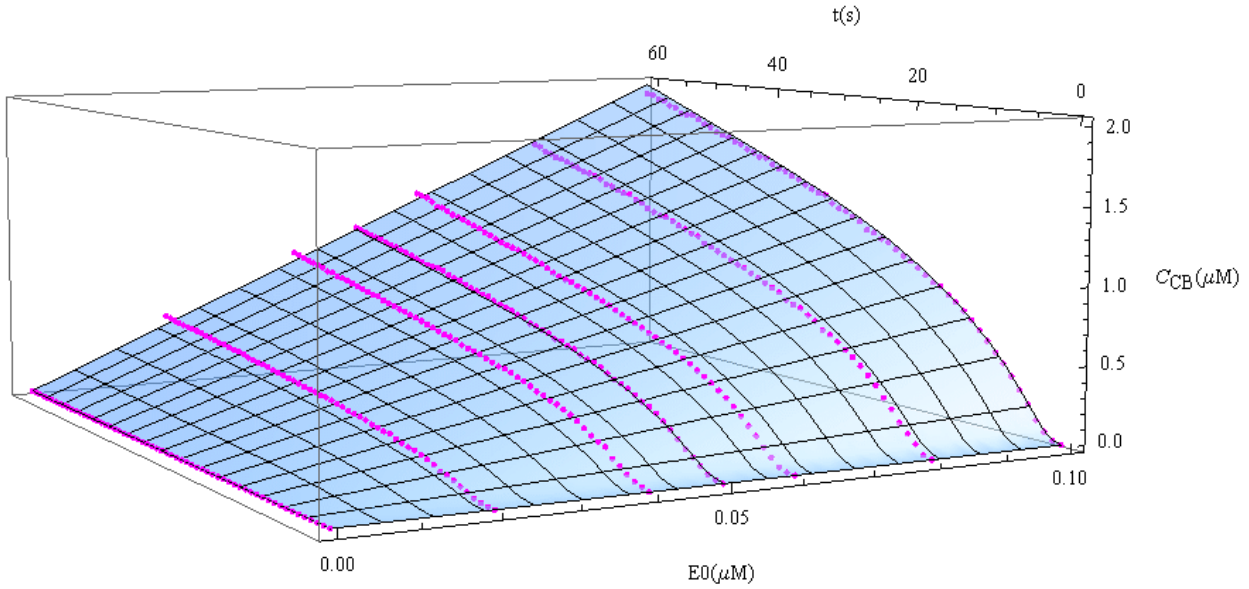

Figure S1. Global analysis of a series of measurements with variable enzyme loads. The plot shows the cellobiose concentration (C_{CB}) as a function of enzyme load (E_0) and time (t) for the first 60 sec of the reaction. Pink symbols are experimental data and the blue contour is the result of a regression analysis that considered all data in the plot simultaneously.

Enzyme distribution. The model description allows calculation of the time dependent distribution of enzyme between the states defined in eq. (1), *i.e.* active- (EC_{m-i}), inactively bound- (EC_{m-n}) or free enzyme (E), where the acronyms in brackets are defined in eq. (1). The concentration of active enzyme has already been specified in eq. <S2>. Next, we note that free enzyme, E, is removed from the bulk in the first step of eq. (1) and returned by the disintegration of enzyme –substrate complexes (governed by k_{off} for all complexes). It follows that rate equation for free enzyme may be written

$$E'(t) = -k_{on}\ C_m(t)E(t) + k_{off}\sum_{i=0}^{n} EC_{m-i}(t) = -k_{on}\ C_m(t)E(t) + k_{off}\left(E_0 - E(t)\right) \quad \text{<S4>}$$

If we again use that the substrate is in excess so that C_m can be considered constant, eq. <S4> can be solved to yield.

$$E(t) = \frac{E_0\left(k_{off} + k_{on}C_m e^{-\left(k_{off}+k_{on}C_m\right)t}\right)}{k_{off} + k_{on}C_m} \quad \text{<S5>}$$

Finally, a material balance for the enzyme can be used to express the concentration of the third form of enzyme (enzyme bound inactively in front of an obstacle).

$$EC_{m-n}(t) = E_0 - E(t) - \sum_{i=0}^{n-1} EC_{m-i} \quad \text{<S6>}$$

Where the two last terms on the right hand side are given in eqs. <S5> and <S2> respectively. The three populations of enzymes were calculated by insertion of the parameters from Tab. 1 (main article) into eqs. <S2>, <S5> and <S6> and plotted as a function of time in Fig. 3 of the main article.

Preparation of biosensors and electrochemical equipment. Benzoquinone-mixed carbon paste electrodes was prepared according to the protocol outlined in (2): a weighed amount of graphite powder (0.09 g) were added 8 wt% (0.01g) *p*-benzoquinone, 25 wt% (35 µl) liquid paraffin and thoroughly hand-mixed in an agate mortar until a homogenized paste was obtained. A portion of the resulting benzoquinone-carbon paste was packed into carbon paste holders from ALS (Tokyo, Japan) with a working geometric area of 0.071cm^2 and the surface was polished using waxed weighing-paper. A 10 µL droplet of a freshly prepared solution of a 1:1 mixture of cellobiose dehydrogenase enzyme stock (6.6 mg/ml) and 1% glutaraldehyde in the standard buffer was carefully cast onto the electrode surface and allowed to dry a room temperature for 25 min. The electrode was stored overnight at 4 °C in an inverted position in a closed vessel with the bottom filled with Milli-Q water. After cross-linking overnight the electrode surface was thoroughly rinsed with Milli-Q water to remove weakly adsorbed enzyme molecules and the biosensor was then stored overnight at 4 °C in an inverted position immersed in the standard buffer before measurement. In the electrochemical measurements, the working electrode potential was fixed at +0.5 volt vs. a Ag/AgCl/3M NaCl reference electrode (Bioanalytical Systems, United Kingdom) in a conventional three-electrode setup with a coiled platinum wire as counter electrode. Potential control and current detection was done by an analog potentiostat (Model 1112) from Husou Seisakusyo Co. (Kawasaki, Japan). The potentiostat was connected to a computer via a Agilent 34401A Digital Multimeter and data acquisition software was created in LabVIEW 8.6 (National Instruments, Austin, USA). The oxidation current of the reduced redox mediator, hydroquinone, produced by the cellobiose dehydrogenase-catalyzed reaction was recorded as a function of time. The oxidation current thus obtained was found to be proportional to the concentration of cellobiose up to about 25µM.

Performance of biosensors. The detection limit of the biosensors was about 25 nM and this is 1-2 orders of magnitude less than typical concentrations, when steady state is reached (Figs 1-2). Therefore, the build-up of cellobiose in the pre-steady state regime can be readily resolved, and the

main limitation of the experimental setup is the response time. The latter may be illustrated in calibration experiments, where known amounts of cellobiose are titrated into pure buffer either in steps or ramps. Figure S2 shows some examples.

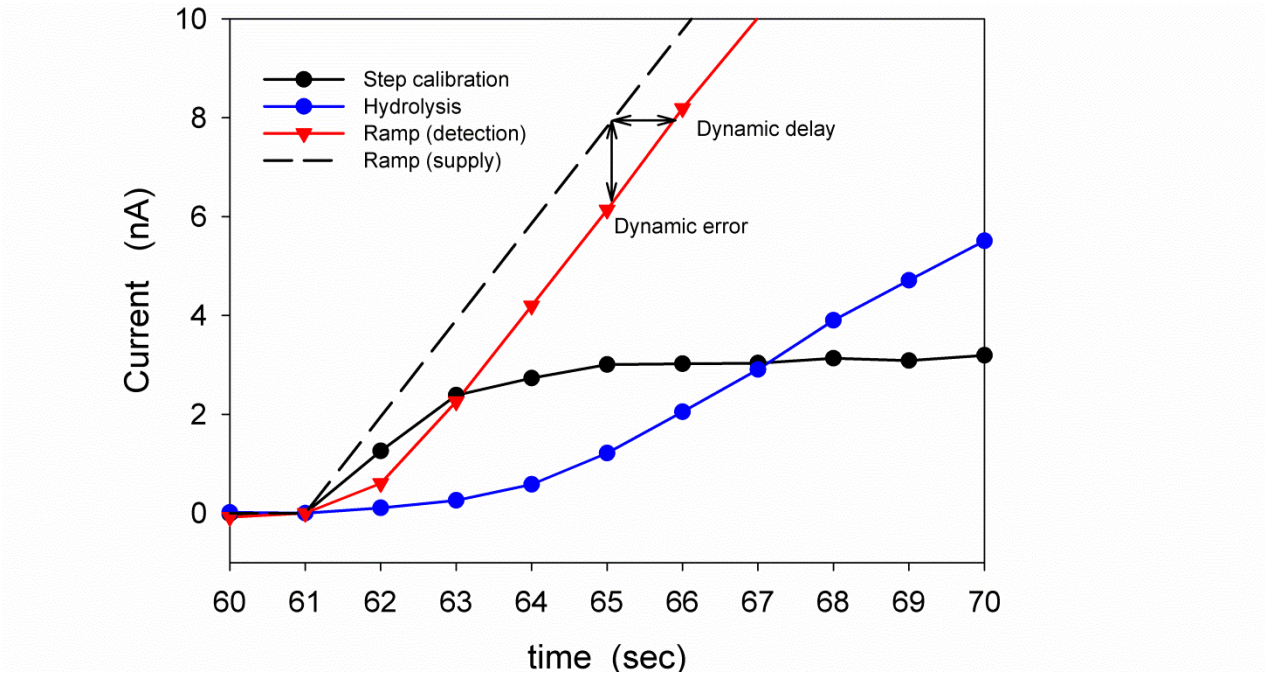

Figure S2. Illustration of the response time for the biosensors. Black symbols show the detector response for a 0.5 μM step-increase in the cellobiose concentration. The injection commenced at t=61 s and lasted 0.1 sec. The red curve shows the sensor response when an injection ramp (252 nM cellobiose/sec, indicated by the dashed line) is started at t=61 s. The dynamic delay (~0.8 sec) and -error (~1.3 nA) are specified by the two arrows. Also shown (blue) is an example of a hydrolysis measurement (from Fig. 1 in main article) where 100nM TrCel7A is injected to 2 g/l RAC.

The results in Fig. S2 intuitively show that the sensor is able to respond more rapidly than the cellobiose concentration builds up in a typical hydrolysis experiment. To assess this quantitatively, we first treat the step calibration (black symbols) as a first order approach towards the steady state current (*i.e* an exponential rise towards the nominal value). The half-time for this rise was about 0.7 s (*i.e.* a time constant, τ~1 s), and simple correction with respect to this delay(3) showed that the biosensor signal is slightly smeared for the first 10-20 s where the cellobiose concentration changes most rapidly (Fig S3). More rigorous treatments of a biosensor's response can be made on the basis of control experiments where cellobiose is added at a constant rate for times that are much longer than τ (4). This type of control defines so-called dynamic error and -delay, as the difference between the supply (dashed line) and response (red curve) as illustrated in Fig. S2. In this case the values were respectively 150 nA and 0.8 s and this again implies a minor smearing of the initial

signal. To assess the influence of this smearing on the kinetic parameters we conducted non-linear regression analyses (eq. <S2>) on data-sets, which had been corrected by a single exponential version of the so-called Tian equation (3) with a time constant of τ=1 s (Fig. S3). These regressions showed that $k_{on,}$ k_{off} and n were only slightly affected by the time correction (changes were well below the experimental error-margins given in Tab. 1 of the main manuscript). For k_{cat}, the time corrected data gave values of 5 - 6 s^{-1}, and this is slightly larger than for the raw data (Tab. 1). Thus, limitations in the time resolution of the current set-up does not seem to effect $k_{on,}$ k_{off} or n but it may lead to an underestimation of k_{cat}. However, the shift is quite small and not critical with respect to the current discussion. In the light of this and the fact that the single exponential correction is only a coarse approximation (particularly so as the data sampling interval and duration of the injection (both 1 s) is comparable to τ) we conclude that general correction of the experimental data in Figs. 1 and 2 is not warranted.

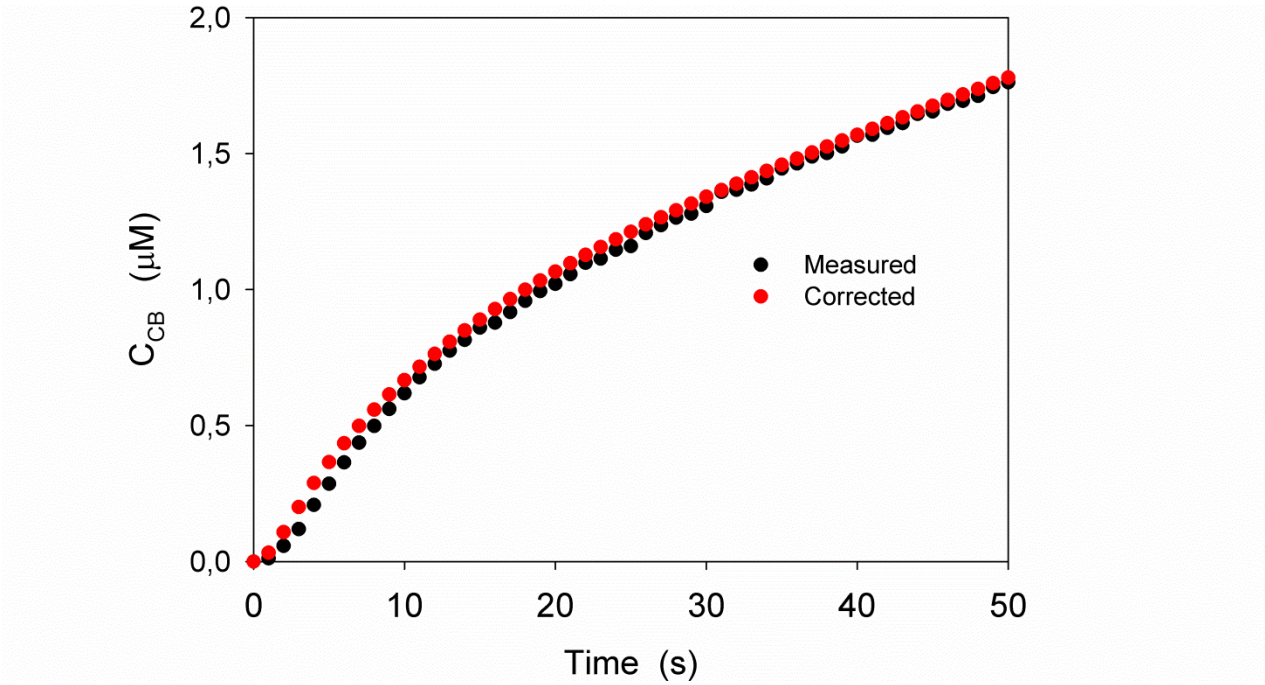

Figure S3. Illustration of a simplified time correction (red symbols) of an experiment (black symbols) from Fig. 1 (100nM *Tr*Cel7A and 2 g/L RAC). The correction relies on the assumption that the recorded signal rises exponentially towards the correct value with the time constant τ =1 s (*c.f.* Fig. S2). This is undoubtedly and oversimplification but it provides an estimate of the dead-time in the current measurements.

Double injections. The rapid slow-down in enzymatic activity could rely on the depletion of particularly reactive forms of cellulose (*e.g.* small soluble cello oligosaccharides or frayed ends). To test this, we conducted double-injection trials as illustrated in Fig S4. In accordance with earlier studies of the same enzyme-substrate system (1), we found that a second dose of enzyme also generated a burst in activity, and this strongly speaks against depletion as the main origin of the slow-down.

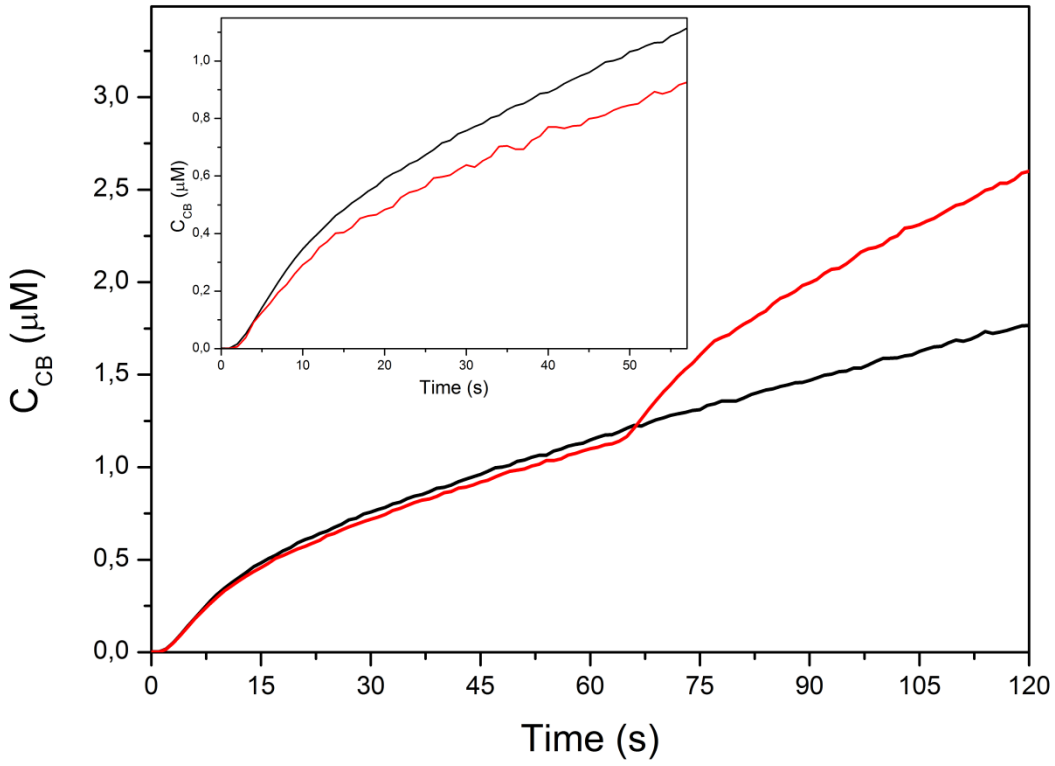

Figure S4. Biosensor recordings over 2 min in which 50nM doses *Tr*Cel7A is added to 2 g/L RAC at respectively t=0 sec. (black curve) and both t=0 sec. and t=60 sec. (red curve). The red curve in the inset is the increase in activity associated with the second dosage (*i.e.* the difference between the red and black curves from 60 to 120 sec in the main panel). This measure of the second burst is compared to the first burst (black curve in the inset).

Enzyme purity. The purity of *Tr*Cel7A was assessed by SDS-PAGE.

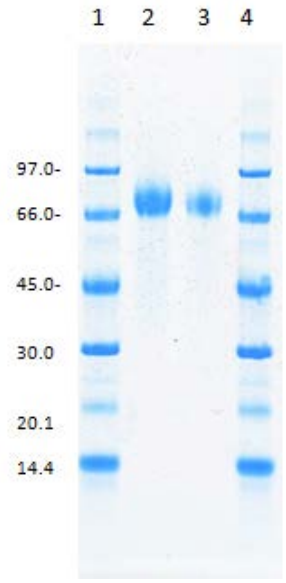

Figure S5: SDS-PAGE of *Tr*Cel7A at different loads (lane 2-3) and MW-markers (lane 1 and 4)

References to the Supplementary Data

1. Praestgaard, E., Elmerdahl, J., Murphy, L., Nymand, S., McFarland, K. C., Borch, K., and Westh, P. (2011) *FEBS J.* **278**, 1547-1560
2. Tatsumi, H., Katano, H., and Ikeda, T. (2006) *Anal. Biochem.* **357**, 257-261
3. Loblich, K. R. (1994) *Thermochim. Acta* **231**, 7-20
4. Baker, D. A., and Gough, D. A. (1996) *Anal. Chem.* **68**, 1292-1297