Supporting Information: Engineering and Modeling the Electrophoretic Trapping of a Single Protein Inside a Nanopore

Kherim Willems,[†,‡,‖] Dino Ruić,[¶,‡,‖] Annemie Biesemans,[†,‖] Nicole Stéphanie Galenkamp,[§] Pol Van Dorpe,[¶,‡] and Giovanni Maglia[*,§]

†*KU Leuven, Department of Chemistry, Celestijnenlaan 200F, B-3001 Leuven, Belgium*

‡*imec, Kapeldreef 75, B-3001 Leuven, Belgium*

¶*KU Leuven, Department of Physics and Astronomy, Celestijnenlaan 200D, B-3001 Leuven, Belgium*

§*University of Groningen, Groningen Biomolecular Sciences & Biotechnology Institute, 9747 AG, Groningen, The Netherlands*

‖*Contributed equally to this work*

E-mail: g.maglia@rug.nl

Contents

1 Escape Rates Over a Potential Barrier

1.1 Dwell Times of Bound States

When measuring the dwell time of molecules in bound states such as when they are trapped in nanopores, we typically obtain a distribution of dwell times whose appearance yields direct information about the kinetics of the bound state. In this section we will discuss how to extract kinematic information from such distributions and how distributions of multi-level bound states look like.

1.1.1 Single Bound State

The simplest scenario is the decay of a single bound state. Suppose we measure the duration of the bound state many times in an experiment. This involves measuring the time trace noting the capture and release times and then plotting a histogram of events vs. dwell time in the bound state. Since the energy obtained for the release of the molecule is thermal, the process is stochastic in nature and with growing event number the histogram approaches the underlying distribution associated with the kinetics of escaping from the bound state.

For a single bound state with a rate of escape k, the probability distribution function of remaining in the bound state is given by the typical exponential distribution

$$f(t) = k\, e^{-kt}.$$

The nature of the distribution function f is sometimes confusing as it is often used interchangeably and without notational clarity as both a probability distribution and as an event distribution. In an experiment where we count events in time series data, we gather statistical data on the event distribution rather than on the probability distribution. In the simple

case of a single bound state, this event distribution is given by

$$F(t) = F_0 \; f(t), \tag{S1}$$

where F_0 is the total number of events. Note that since f is a probability distribution, it is necessarily normalized to unity

$$\langle 1 \rangle_f := \int_0^\infty dt \; f(t) = 1 \tag{S2}$$

and therefore the event distribution is normalized to the total number of events:

$$\langle 1 \rangle_F := \int_0^\infty dt \; F(t) = F_0. \tag{S3}$$

Since the event distribution and the probability distribution are so closely related in this case, the rate of escape k can simply be read off of a general exponential fit to the event histogram generated from the time-series data.

For later reference, let us note that the *expectation value* of the duration of the molecule in the bound state (*i.e.*, the average dwell time) can be computed from the distribution function as

$$\tau := \langle t \rangle_f = \int_0^\infty dt \; t \; f(t) = \int_0^\infty dt \; t \; k \; e^{-kt} = -te^{-kt}\Big|_0^\infty + \int_0^\infty dt \; e^{-kt} = \frac{1}{k}.$$

On the other hand, if we want to compute the average dwell time from the event distribution, we need to take into account that the event distribution is differently normalized:

$$\tau = \frac{\langle t \rangle_F}{\langle 1 \rangle_F} = \frac{F_0 \langle t \rangle_f}{F_0 \langle 1 \rangle_f} = \langle t \rangle_f. \tag{S4}$$

The variance is defined as the squared deviation from the expectation value

$$\mathrm{Var}(t) = \left\langle (t - \langle t \rangle_f)^2 \right\rangle_f = \langle t^2 \rangle_f - \langle t \rangle_f^2.$$

In case of an event distribution F, we also need to normalize by $\langle 1 \rangle_F$ as above. The expectation value of the squared time is given by

$$\begin{aligned} \langle t^2 \rangle_f &= \int_0^\infty dt\ t^2 f(t) = \int_0^\infty dt\ t^2\ k\ e^{-kt} = -t^2\ e^{-kt}\Big|_0^\infty + \int_0^\infty dt\ 2t\ e^{-kt} \\ &= \frac{2}{k} \langle t \rangle_f = 2 \langle t \rangle^2. \end{aligned}$$

Therefore we find for the variance of the dwell time of events

$$\mathrm{Var}(t) = \langle t \rangle^2 = \tau^2 = \frac{1}{k^2}.$$

1.1.2 Multi-Level Bound States

Now let us take a look at the general N-level system. It can be described by the set of coupled ordinary differential equations (ODE)

$$\frac{d}{dt} f_i(t) = \sum_{j=1}^{N} m_{ij} f_j(t), \tag{S5}$$

where f_i is the probability distribution of the i-th state and N is the total number of states. Here we assume that each bound state can decay into another state *via* the rate m_{ij}. If we choose one or more of the states to represent a free state where the molecule escapes, we can in principle describe trapping events with arbitrary number of transitions in between meta states before the escape of the molecule. The rates m_{ij} of Eq. (S5) are not entirely unrelated as they need to conserve probability. An obvious choice is to simply express the decrease of

probability of a state as the increase of the other states:

$$m_{ij} = \begin{cases} -\sum_j k_{ij}, & \text{if } i = j, \\ k_{ji}, & \text{else,} \end{cases} \tag{S6}$$

This means that a molecule that disappears in one state has to appear in another one. It cannot simply vanish nor can it appear out of thin air.

It is known that systems of ODEs as in Eq. (S5) describe solutions that can be expressed as the sums of exponentials of their rates:

$$f_i(t) = \sum_{\ell,m=1}^{N} a_{i\ell m} e^{-k_{\ell m} t}. \tag{S7}$$

Inserting this ansatz into Eq. (S5) tells us how the probability distributions are related:

$$\frac{d}{dt} f_i = \sum_j m_{ij} f_j = \sum_j m_{ij} \sum_{\ell,m} a_{j\ell m} e^{-k_{\ell m} t} \stackrel{!}{=} -\sum_{\ell,m} k_{\ell m} a_{i\ell m} e^{-k_{\ell m} t}.$$

Since all terms are linearly independent, we can compare the coefficients of the exponentials on each side and find

$$\sum_j m_{ij} a_{j\ell m} = -k_{\ell m} a_{i\ell m}. \tag{S8}$$

Finally using Eq. (S6), we find

$$\begin{aligned} -\sum_j k_{ij} a_{i\ell m} + \sum_{j \neq i} k_{ji} a_{j\ell m} &= -k_{\ell m} a_{i\ell m} \\ \Rightarrow \qquad a_{i\ell m} &= \frac{\sum_{j \neq i} k_{ji} a_{j\ell m}}{\sum_j k_{ji} - k_{\ell m}}. \end{aligned} \tag{S9}$$

Eq. (S9) is a manifestation of the probability conservation in terms of a constraint on the possible values of the coefficients.

This derivation tells us that we can express probability distributions of an arbitrary state in an N-level system by a simple linear combination of all the exponentials with the transition rates present in the systems. Moreover, the coefficients of the exponential terms in the distribution functions are directly related to the rates.

1.1.3 Interpretation of Experimental Data

Extracting events from time series data and plotting a histogram of the dwell time will yield the event distribution function. In a multi-state system, the observed event may be a consequence of multiple different states with transition rates between each other. If the experiment or the data analysis technique does not distinguish between these states but groups them together, what we observe may be the sum of multiple event distributions.

Event distributions are, as shown *e.g.* in Eq. (S1), mathematically equivalent to probability distributions, only their interpretation differs. When solving the ODE of Eq. (S5) for event distributions instead of probability distributions, we can proceed completely analogously. Only when we apply the initial conditions, we need to choose a normalization to events as in Eq. (S1), rather than a normalization to probabilities. This is then also reflected in the way that expectation values are calculated, as was shown in Eq. (S4).

In an N-level system with event distributions F_i, where the measurement or data analysis tools cannot distinguish between the levels $\alpha, \ldots, \beta$, we observe a compound event distribution given by

$$F_{\alpha,\ldots,\beta} = \sum_{i=\alpha}^{\beta} F_i = \sum_{i=\alpha}^{\beta} F_0 f_i.$$

Now $F_{\alpha,\ldots,\beta}$ not only contains the sum of all independent event distributions in the states $\alpha, \ldots, \beta$, it also contains all interactions between the states $\alpha, \ldots, \beta$ and all other states. Note that we only need a single overall event normalization F_0 since the probabilities of finding the molecule in one of the many states at $t = 0$ is set by the initial conditions chosen

when solving the ODE of Eq. (S5). All differences between the distributions at time t are then a consequence of the intrinsic rates in between states.

Suppose we measure such a system and now we want to extract the kinetics of the process. The compound event distribution function can be split up into exponential terms using Eq. (S7)

$$F_{\alpha,\ldots,\beta} = \sum_{i=\alpha}^{\beta} \sum_{\ell,m=1}^{N} F_0 a_{i\ell m} e^{-k_{\ell m} t} =: \sum_{\ell,m=1}^{N} F_0 A_{\ell m} k_{\ell m}\, e^{-k_{\ell m} t} \tag{S10}$$

Note that the rate $k_{\ell m}$ defined as appearing in the amplitude is to keep the analogy to the probability distribution of Sec. 1.1.1. By fitting with a sum of exponentials, we can simply ignore the results for the coefficients F_0 and $A_{\ell m}$ and extract the rates $k_{\ell m}$ that are directly related to individual transitions between states.

If we want to compute average dwell times using a compound event distribution function, we find

$$\tau = \frac{\langle t \rangle_{F_{\alpha,\ldots,\beta}}}{\langle 1 \rangle_{F_{\alpha,\ldots,\beta}}} = \frac{\sum_{\ell,m=1}^{N} \frac{A_{\ell m}}{k_{\ell m}}}{\sum_{\ell,m=1}^{N} A_{\ell m}}, \tag{S11}$$

where we used that

$$\langle 1 \rangle_{F_{\alpha,\ldots,\beta}} = \int_0^\infty dt \sum_{\ell,m=1}^{N} F_0 A_{\ell m} k_{\ell m}\, e^{-k_{\ell m} t} = F_0 \sum_{\ell,m=1}^{N} A_{\ell m}$$

and

$$\langle t \rangle_{F_{\alpha,\ldots,\beta}} = \sum_{\ell,m=1}^{N} F_0 A_{\ell m} k_{\ell m} \int_0^\infty dt\; t\; e^{-k_{\ell m} t} = F_0 \sum_{\ell,m=1}^{N} \frac{A_{\ell m}}{k_{\ell m}}.$$

Note how the event number normalization F_0 neatly cancels. This should be obvious because the kinetics of the bound states cannot depend on whether a hundred or a thousand events are collected.

One useful observation for the calculation of the dwell time of a multi-exponential distribution function is that the total expectation value is simply the arithmetic mean of the expectation values from the individual exponentials:

$$\tau = \frac{\langle t \rangle_{F_{\alpha,\ldots,\beta}}}{\langle 1 \rangle_{F_{\alpha,\ldots,\beta}}} = \frac{\sum_{i=\alpha}^{\beta} \langle t \rangle_{F_i}}{\sum_{i=\alpha}^{\beta} \langle 1 \rangle_{F_i}} = \frac{\sum_{i=\alpha}^{\beta} \langle t \rangle_{f_i}}{\sum_{i=\alpha}^{\beta} \langle 1 \rangle_{f_i}} = \frac{1}{N_{\alpha,\ldots,\beta}} \sum_{i=\alpha}^{\beta} \tau_i, \tag{S12}$$

where $N_{\alpha,\ldots,\beta}$ is the number of bound states of the compound distribution function.

Another type of compounding that is often encountered is that compound distribution functions are collected from different experiments. The compound distribution function can then be expressed as

$$\mathcal{F} = \sum_{\xi} F^{\xi}_{\alpha,\ldots,\beta},$$

where ξ denotes the experiment and $F^{\xi}_{\alpha,\ldots,\beta}$ is the compound distribution function of experiment ξ. Using Eq. (S10) we find

$$\mathcal{F} = \sum_{\xi} \sum_{\ell,m=1}^{N} F^{\xi}_0 A_{\ell m} k^{\xi}_{\ell m} \, e^{-k^{\xi}_{\ell m} t}. \tag{S13}$$

Note that the rates $k^{\xi}_{\ell m}$ can vary systematically between experiments. If, for example, the temperature is not tightly controlled but measured for each experiment, the temperature dependence of the rates can in principle be taken into account to remove these kinds of systematic errors. However, for the present treatment, we will simply assume that the laboratory conditions are controlled tightly enough that we can assume the rates to be independent of the experiments, *i.e.*,

$$k^{\xi}_{\ell m} = k_{\ell m}. \tag{S14}$$

Using Eqs. (S13) and (S14) we can therefore express the average dwell time as

$$\tau = \frac{\langle t \rangle_{\mathcal{F}}}{\langle 1 \rangle_{\mathcal{F}}} = \frac{\sum_\xi F_0^\xi \sum_{\ell,m=1}^N \frac{A_{\ell m}}{k_{\ell m}}}{\sum_\xi F_0^\xi \sum_{\ell,m=1}^N A_{\ell m}} = \frac{\sum_{\ell,m=1}^N \frac{A_{\ell m}}{k_{\ell m}}}{\sum_{\ell,m=1}^N A_{\ell m}},$$

which is identical to Eq. (S11). This means that as long as the experimental conditions are sufficiently close, we can simply sum up the results of experiments. On the other hand, if we have systematic differences in rates (*i.e.*, $k_{\ell m}^\xi$ varies systematically between experiments) the above treatment can be easily extended to generalize the computation of average dwell times in the presence of systematic errors.

Lastly we want to quantify how event distributions behave when two independent processes are measured within the same event distribution. An example for this would be that the molecule is captured into a completely separate state from which it has an escape rate with kinetics that are different from the actual state we want to measure. This could potentially be a meta state with negligible transition rate to the main trapping mechanism or it could simply be a capture in a different orientation.

In that case we find that the total measured event distribution function is given by

$$\mathcal{F} = F^A_{\alpha,\ldots,\beta} + F^B_{\gamma,\ldots,\delta}. \tag{S15}$$

The main observation in that case is that F^A and F^B may contain completely different rates and that the rates of F^A and F^B are not related as in Eq. (S9). A way to distinguish these types of distribution functions would be to quantify the capture ratio F_0^A/F_0^B in terms of experimental conditions. Modifying the capture ratio would then yield predictable behavior of the ratio F^A/F^B which can be used to untangle the two independent physical processes.

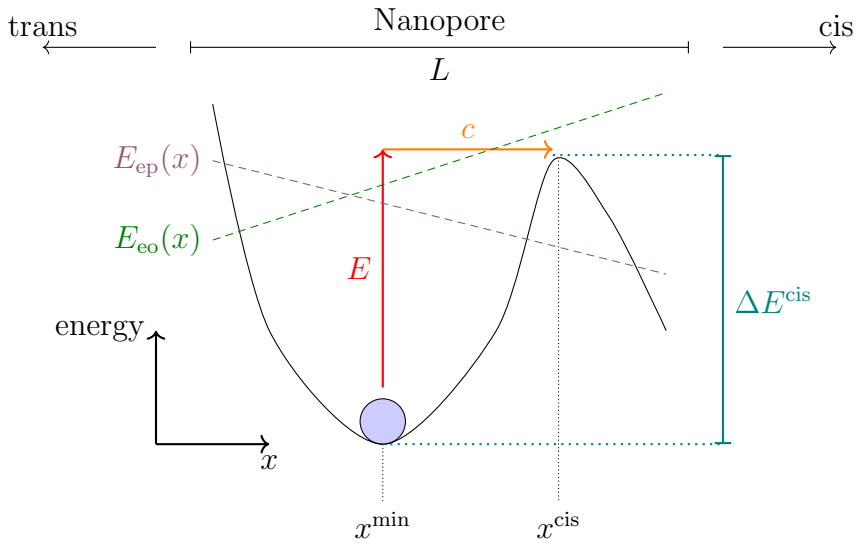

Figure S1. Molecule (blue circle) trapped in a potential well defined by a nanopore and trying to overcome the energy barrier ΔE^{cis} by picking up the kinetic energy E from its environment and then propagating over the barrier with an instantaneous transition rate c. Overlaid are the linear energy potentials from the electrophoretic force E_{ep} and from the force of the electroosmotic flow E_{eo}.

1.2 Escape From a Single Barrier System

We want to derive an expression for the dwell time of a molecule confined in a potential well with a single barrier as illustrated in Fig. S1. This treatment will be useful as a stepping stone to a double barrier model but it will also tell us why the molecule trapped in a nanopore cannot be described by a single barrier model.

In order to overcome the barrier of Fig. S1, the molecule has to pick up thermal energy randomly. The probability of finding a molecule with a kinetic energy of E in an environment at thermal equilibrium is given by the usual (normalized) Boltzmann distribution

$$f(E) = \frac{1}{k_{\mathrm{B}}T} e^{-\frac{E}{k_{\mathrm{B}}T}}.$$

Then the scattering rate k into a some final state can be expressed as

$$k = \int_0^\infty dE\ Z\ c(E)\ f(E), \tag{S16}$$

where Z is the density of states and c is the transition rate from an initial state at energy E

into a final state. The density of states is assumed to be energy independent which means that the molecule can only have a fixed number of internal states, irrespective of the kinetic energy it picks up. Note that the transition rate c could in principle depend on the internal states of the molecule but we neglected this as well. We will view the escape rate of the molecule from the potential well as an instantaneous scattering event that takes it over the barrier.

In order to derive an approximation to this rate, we will assume that the molecule cannot escape (*i.e.*, tunnel) as long as its kinetic energy is less then the required energy barrier that needs to be overcome. On the other hand, if the molecule acquires a kinetic energy equal or larger to the barrier, the transition rate c is assumed to be constant. Thus we have

$$c(E) = \begin{cases} 0, & \text{if } E < \Delta E^{\text{cis}}, \\ c_0, & \text{else.} \end{cases}$$

The escape rate over the barrier in Eq. (S16) therefore yields

$$\begin{aligned} k &= Z c_0 \int_{\Delta E^{\text{cis}}}^{\infty} dE\ f(E) \\ &= Z c_0\ e^{-\frac{\Delta E^{\text{cis}}}{k_{\text{B}} T}} \\ &=: k_0^{\text{cis}}\ e^{-\frac{\Delta E^{\text{cis}}}{k_{\text{B}} T}}. \end{aligned}$$

And the dwell time is the inverse of the escape rate:

$$\tau = \frac{1}{k}.$$

The barrier can be decomposed into steric, electrostatic, and external contributions:

$$\Delta E^{\text{cis}} = \Delta E^{\text{cis}}_{\text{st},0} + \Delta E^{\text{cis}}_{\text{es}} + \Delta E^{\text{cis}}_{\text{ex}}, \tag{S17}$$

where the steric contribution $\Delta E^{\mathrm{cis}}_{\mathrm{st},0}$ accounts for all interactions of the molecule within the potential well that are not electrostatic in nature (*e.g.* size related effects). The barrier contribution from the electrostatic interactions can be expressed as

$$\Delta E^{\mathrm{cis}}_{\mathrm{es}} = \Delta E^{\mathrm{cis}}_{\mathrm{es},0} + N_{\mathrm{tag}} e \Delta V^{\mathrm{cis}}_{\mathrm{tag}}, \tag{S18}$$

where N_{tag} is the signed net number of charges on the molecule being trapped in the electrostatic potential well. Since the molecule is large and charges may be located outside of the potential well, not all charges modify the barrier. For our present purpose we will refer to the number of charges in the well as the tag charge number in reference to the main text. Thus, $\Delta V^{\mathrm{cis}}_{\mathrm{tag}}$ is the barrier height change due to a change in the tag charge number N_{tag}. The electrostatic barrier contributions that are independent of N_{tag} are absorbed into the constant $\Delta E^{\mathrm{cis}}_{\mathrm{es},0}$.

Moreover, $\Delta E^{\mathrm{cis}}_{\mathrm{ex}}$ stems from the externally applied bias that results in an electrophoretic force and a force due to the electroosmotic flow and can therefore be split up as

$$\Delta E^{\mathrm{cis}}_{\mathrm{ex}} = \Delta E^{\mathrm{cis}}_{\mathrm{ep}} + \Delta E^{\mathrm{cis}}_{\mathrm{eo}}. \tag{S19}$$

We will assume that both of these external forces are constant throughout the nanopore and therefore the associated energy potentials are linear in space as is illustrated in Fig. S1. Thus we can express the potential energies due to the external forces as

$$\begin{aligned} E_{\mathrm{ep}}(x) &= -\mathcal{F}_{\mathrm{ep}}\, x + b_{\mathrm{ep}}, \\ E_{\mathrm{eo}}(x) &= -\mathcal{F}_{\mathrm{eo}}\, x + b_{\mathrm{eo}}, \end{aligned}$$

where $\mathcal{F}_{\mathrm{ep}}$ is the constant electrophoretic force and $\mathcal{F}_{\mathrm{eo}}$ is the constant force due to the electroosmotic force on the molecule. Note the sign in front of the force since the force points down the slope of the potential energy.

Using the energy potentials we can compute the effect of the external fields on the energy barrier as

$$\Delta E_{\mathrm{ep}}^{\mathrm{cis}} = E_{\mathrm{ep}}(x^{\mathrm{cis}}) - E_{\mathrm{ep}}(x^{\mathrm{min}}) = -\mathcal{F}_{\mathrm{ep}}\Delta x^{\mathrm{cis}},$$

where x^{cis} is the location of the barrier and x^{min} is the location of the minimum. The distance of the barrier to the minimum is given by

$$\Delta x^{\mathrm{cis}} = \left| x^{\mathrm{cis}} - x^{\mathrm{min}} \right|.$$

The constant energy shift b_{ep} does not contribute and therefore can be chosen arbitrarily. The electrophoretic force due to an applied external bias V^{bias} can be expressed by Coulomb's law

$$\mathcal{F}_{\mathrm{ep}} = eN_{\mathrm{net}}\frac{V^{\mathrm{bias}}}{L} \tag{S20}$$

where $N_{\mathrm{net}} = N_{\mathrm{tag}} + N_{\mathrm{body}}$ is the signed net number of charges on the molecule, where we defined the body charge N_{body} in reference to the main text. Moreover, L is the length of the potential drop from cis to trans (cf. Fig. S1).

Likewise we can compute the change in barrier height due to the osmotic component as

$$\Delta E_{\mathrm{eo}}^{\mathrm{cis}} = E_{\mathrm{eo}}(x^{\mathrm{cis}}) - E_{\mathrm{eo}}(x^{\mathrm{min}}) = -\mathcal{F}_{\mathrm{eo}}\Delta x^{\mathrm{cis}},$$

where the force is linearly dependent on the potential as

$$\mathcal{F}_{\mathrm{eo}} = eN_{\mathrm{eo}}\frac{V^{\mathrm{bias}}}{L}. \tag{S21}$$

Note that for later convenience we expressed the osmotic force in complete analogy to the electrophoretic force of Eq. (S20), where we defined the parameter N_{eo} that we will refer to as

the *equivalent osmotic charge number*. This number tells us what net charge a molecule would have to have in order to experience the same force *via* electrophoresis. Thus if $N_{\rm net} = -N_{\rm eo}$, the net external force on the molecule vanishes.

Using Eqs. (S17) to (S19) we find for the total barrier height

$$\Delta E^{\rm cis} = \Delta E^{\rm cis}_{\rm st,0} + \Delta E^{\rm cis}_{\rm es,0} + N_{\rm tag} e \Delta V^{\rm cis}_{\rm tag} - (N_{\rm net} + N_{\rm eo}) \frac{\Delta x^{\rm cis}}{L} e V^{\rm bias}. \tag{S22}$$

The escape rate (*i.e.*, the inverse of the dwell time) is therefore given by

$$\begin{aligned} \frac{1}{\tau(N_{\rm tag}, V^{\rm bias})} &= k(N_{\rm tag}, V^{\rm bias}) \\ &= k_0^{\rm cis} \exp\left(-\frac{\Delta E^{\rm cis}}{k_{\rm B}T}\right) \\ &= k_{\rm eff}^{\rm cis} \ \exp\left(-\frac{N_{\rm tag} e \Delta V^{\rm cis}_{\rm tag} - (N_{\rm net} + N_{\rm eo}) \frac{\Delta x^{\rm cis}}{L} e V^{\rm bias}}{k_{\rm B}T}\right), \end{aligned} \tag{S23}$$

where the constant steric and electrostatic terms $\Delta E^{\rm cis}_{\rm st,0}$ and $\Delta E^{\rm cis}_{\rm es,0}$, respectively, have been absorbed into the effective attempt rate $k^{\rm cis}_{\rm eff}$.

We immediately see that in a single barrier model the dwell time has to monotonically increase or decrease with the applied bias. Data that exhibits a maximum as a function of $V^{\rm bias}$, such as the dwell time data of the main text, cannot be described by the single barrier model.

1.3 Escape From a Double Barrier System

Now let us assume that there is a second barrier as illustrated in Fig. S2, so that the molecule can escape towards the trans side with finite probability.

We can define the barrier to the *trans* side in complete analogy to the *cis* barrier of Eq. (S22) as

$$\Delta E^{\rm trans} = \Delta E^{\rm trans}_{\rm st,0} + \Delta E^{\rm trans}_{\rm es,0} + N_{\rm tag} e \Delta V^{\rm trans}_{\rm tag} + (N_{\rm net} + N_{\rm eo}) \frac{\Delta x^{\rm trans}}{L} e V^{\rm bias}.$$

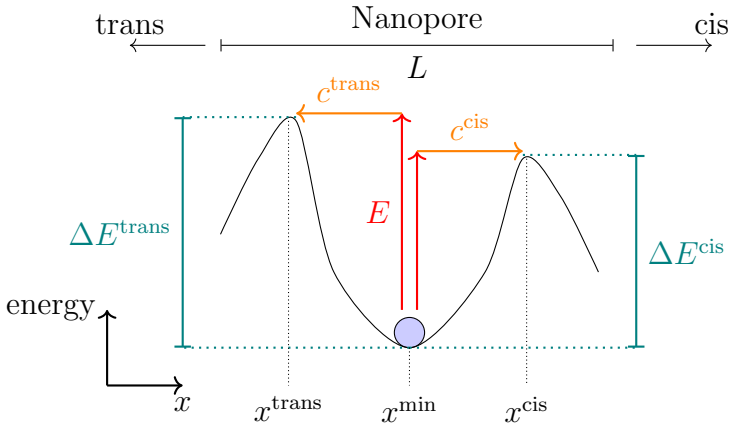

Figure S2. Molecule (blue ball) trapped in between two energy barriers ΔE^{cis} and ΔE^{trans}. By picking up kinetic energy E from its environment it can transition through either barrier towards *cis* or *trans*.

Note, however, the opposite sign in front of the external barrier contributions since the external force points in the same direction as the *trans* exit rather than opposite.

Then the dwell time is given by the inverse of the sum of rates over either barrier

$$\begin{aligned}
\frac{1}{\tau\left(N_{\text{tag}}, V^{\text{bias}}\right)} &= k(N_{\text{tag}}, V^{\text{bias}}) \\
&= k_0 \exp\left(-\frac{\Delta E^{\text{cis}}}{k_{\text{B}}T}\right) + k_0 \exp\left(-\frac{\Delta E^{\text{trans}}}{k_{\text{B}}T}\right) \\
&= k_{\text{eff}}^{\text{cis}} \ \exp\left(-\frac{N_{\text{tag}} e \Delta V_{\text{tag}}^{\text{cis}} - (N_{\text{net}} + N_{\text{eo}})\,\frac{\Delta x^{\text{cis}}}{L} e V^{\text{bias}}}{k_{\text{B}}T}\right) \\
&\quad + k_{\text{eff}}^{\text{trans}} \ \exp\left(-\frac{N_{\text{tag}} e \Delta V_{\text{tag}}^{\text{trans}} + (N_{\text{net}} + N_{\text{eo}})\,\frac{\Delta x^{\text{trans}}}{L} e V^{\text{bias}}}{k_{\text{B}}T}\right),
\end{aligned} \tag{S24}$$

Again, the parameters $k_{\text{eff}}^{\text{cis/trans}}$ are effective attemp rates related to the dwell time at zero tag charge and vanishing applied bias.

The translocation probability is given by the ratio of the *trans* escape rate to the total escape rate:

$$P = \frac{k_0 \exp\left(-\frac{\Delta E^{\text{trans}}}{k_{\text{B}}T}\right)}{k(N_{\text{tag}}, V^{\text{bias}})}.$$

The intrinsic, size-related probability for translocation without any trapping of the tag or applied fields can be directly computed as the ratio $k_{\text{eff}}^{\text{trans}}/(k_{\text{eff}}^{\text{cis}} + k_{\text{eff}}^{\text{trans}})$, which is a small number in the case of a large molecule trying to pass through a nanopore with a narrow *trans* constriction as shown in the main text.

1.4 Analytical Expression for the Threshold Voltages

An analytic expression for the threshold voltage $V_{\text{max}}^{\text{bias}}$—the bias voltage at maximum dwell time—can found as the bias voltage for which $\frac{dk}{dV^{\text{bias}}} = 0$. For the model given by Eq. (1) in the main text this becomes

$$V_{\text{max}}^{\text{bias}} = \left[\frac{\log\left(k_{\text{eff}}^{\text{trans}}/k_{\text{eff}}^{\text{cis}}\right) + \log\left(-N_{\text{eq}}^{\text{trans}}/N_{\text{eq}}^{\text{cis}}\right)}{N_{\text{eq}}^{\text{trans}} - N_{\text{eq}}^{\text{cis}}} \right] \frac{k_{\text{B}}T}{e} \tag{S25}$$

while for the more complex model given by Eq. (6) in the main text it becomes

$$V_{\text{max}}^{\text{bias}} = -\left[\frac{\log\left(k_{\text{eff}}^{\text{trans}}/k_{\text{eff}}^{\text{cis}}\right) + \log\left(\Delta x^{\text{trans}}/\Delta x^{\text{trans}}\right) + \left(\Delta V_{\text{tag}}^{\text{cis}} - \Delta V_{\text{tag}}^{\text{trans}}\right) N_{\text{tag}}}{\left(N_{\text{net}} + N_{\text{eo}}\right)\left(\Delta x^{\text{cis}} + \Delta x^{\text{trans}}\right)/L} \right] \frac{k_{\text{B}}T}{e}. \tag{S26}$$

Table S1. Summary of all estimated threshold voltages and their corresponding dwell times.

	Simple model[a]		Complex model[b]	
DHFR variant	$V_{\max}^{\text{bias}}$ [mV]	$t_{\text{d,max}}$ [s]	$V_{\max}^{\text{bias}}$ [mV]	$t_{\text{d,max}}$ [s]
$DHFR_4S$	56.1	0.31	—	—
$DHFR_4I$	65.5	0.25	—	—
$DHFR_4C$	63.7	0.44	—	—
$DHFR_4O1$	71.5	2.11	75.6	2.28
$DHFR_5O1$	70.7	3.14	69.8	4.16
$DHFR_7O1$	65.4	15.6	61.6	13.9
$DHFR_4O2$	83.0	2.69	87.3	2.28
$DHFR_5O2$	78.5	5.68	79.2	4.16
$DHFR_6O2$	70.8	11.8	73.0	7.59
$DHFR_7O2$	65.4	10.8	68.1	13.9
$DHFR_8O2$	63.6	35.1	64.1	25.3
$DHFR_9O2$	59.8	80.2	60.4	46.2

[a] Estimated using Eq. (S25) after fitting of Eq. (1) the individual mean dwell times of each mutant.

[b] Estimated using Eq. (S26) after fitting of Eq. (2) to all $DHFR_{N_{\text{tag}}}O2$ mean dwell time data.

2 The Electrostatic Energy Landscape of DHFR Inside ClyA

2.1 The Poisson-Boltzmann Equation

To better understand the free energy landscape of the tagged DHFR proteins, we computed the net electrostatic energy of a coarse-grained DHFR molecule as it moves along the length of the pore. A good estimate of this energy can be derived from the electrostatic potential distribution of the nanopore system, which for an atomic system can be estimated by the Poisson-Boltzmann equation (PBE)[1,2]

$$-\nabla \cdot [\epsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})] = \rho^{\mathrm{f}}(\boldsymbol{r}) + \rho^{\mathrm{m}}(\boldsymbol{r}) \tag{S27}$$

with $\boldsymbol{r}$ the location vector, $\phi(\boldsymbol{r})$ the electrostatic potential, $\epsilon(\boldsymbol{r})$ the local permittivity, and ρ_{c} the charge density. For a molecular system in an electrolyte, the charge density can be split into a fixed part—resulting from the distribution of to the atomic partial charges—and a mobile part—resulting from the distribution of charged ions in the electrolyte. For M atomic partial charges the fixed charge density is given by

$$\rho^{\mathrm{f}}(\boldsymbol{r}) = \frac{4\pi e^2}{k_{\mathrm{B}}T}\sum_{i=1}^{M} Q_i \delta(\boldsymbol{r} - \boldsymbol{r_i}) \tag{S28}$$

with e the elementary charge, $k_{\mathrm{B}}T$ the thermal energy, δ the Dirac delta function, and Q_i the atomic partial charge and $\boldsymbol{r_i}$ the location of atom i. The charge density due to N mobile charge species can be expressed as

$$\rho^{\mathrm{m}}(\boldsymbol{r}) = \frac{4\pi e^2}{k_{\mathrm{B}}T}\sum_{j=1}^{N} c_j q_j \exp\left[-q_j\phi(\boldsymbol{r}) - V_j(\boldsymbol{r})\right] \tag{S29}$$

with c_j the bulk concentration, q_j the charge number and $V_j(\boldsymbol{r})$ the steric potential of ion species j. For a monovalent salt such as NaCl this expression can be reduced to

$$\rho^{\mathrm{m}}(\boldsymbol{r}) = -\kappa^{-2}(\boldsymbol{r}) \sinh\left(\phi(\boldsymbol{r})\right) \tag{S30}$$

with κ a coefficient that includes both the ion accessibility as the bulk ion concentration. Finally, the electrostatic energy of the system is given by[2]

$$G(\phi) = \int_{\Omega} \left[\rho^{\mathrm{f}}(\boldsymbol{r})\phi - \frac{\epsilon}{2}\left(\nabla\epsilon\right)^2 - \kappa^{-2}\left(\cosh\phi - 1\right)\right] dr \tag{S31}$$

The free electrostatic energy change caused by placing a particle inside the nanopore $\Delta G^{\mathrm{pore+part}}$ is given by

$$\Delta G^{\mathrm{pore+part}} = G^{\mathrm{pore+part}} - G^{\mathrm{pore}} - G^{\mathrm{part}} \tag{S32}$$

with $G^{\mathrm{pore+part}}$, G^{pore} and G^{part} the electrostatic free energies of systems containing respectively both the nanopore and a DHFR molecule, the empty nanopore and the DHFR molecule alone.

2.2 A Simplistic Bead Model of DHFR

We used the Adaptive Poisson-Boltzmann Solver (APBS) to compute all electrostatic energies.[1] APBS requires a list of spherical particles with a given set of coordinates (XYZ), radius and partial charge (*e.g.* a PQR file) to determine the fixed charge density distribution and to define the dielectric regions separating the protein and the electrolyte. We represented the ClyA-AS nanopore by a full atom homology model,[3] and the charges and radii of its atoms where assigned by PDB2PQR[4,5] using the CHARMM36 force-field parameters.[6] The tagged DHFR molecule was reduced to a coarse-grained 'bead' model (Fig. S3a), where the bulk of the protein (body) was defined by seven negatively charged beads ($r = 0.8\,\mathrm{nm}$, $Q_i = -1.7143\,e$) in a spherical configuration (0.8 nm spacing); and the C-terminal fusion

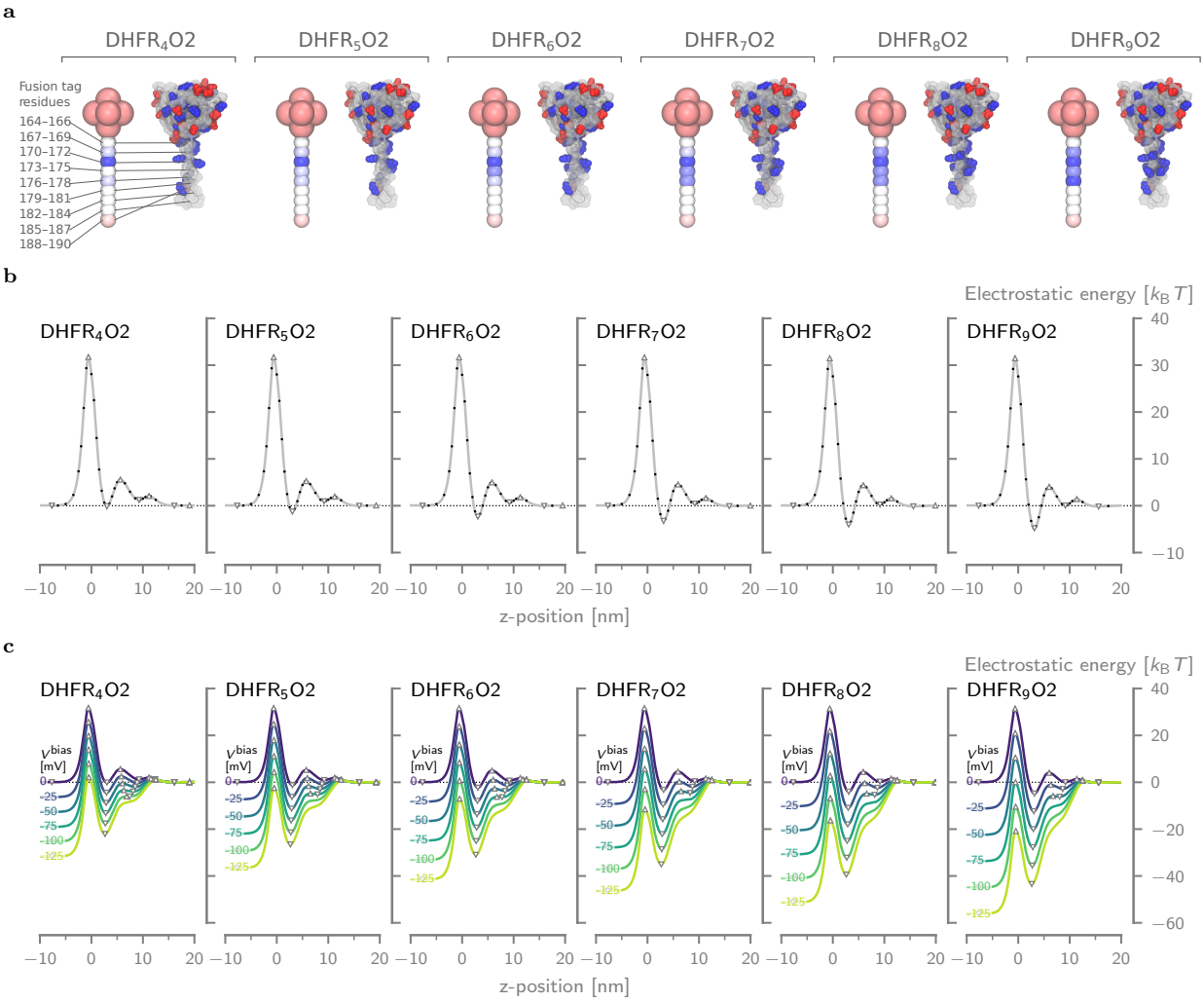

Figure S3. Electrostatic landscape of a DHFR bead model in ClyA-AS. (**a**) Simplified bead model of all DHFR$_{N_{\mathrm{tag}}}$O2 variants where the body is represented by seven evenly spaced beads (diameter of 1.6 nm, 0.8 nm center-to-center distance and a partial charge of $-1.7143\,e$) with a total net charge of $-13\,e$. The fusion tag is approximated by a linear string of nine beads (diameter of 1 nm, 0.6 nm center-to-center distance), with a net charge corresponding to that of the three amino-acids it represents (-3 to $+3\,e$). (**b**) The simulated electrostatic energy landscapes of all DHFR$_{N_{\mathrm{tag}}}$O2 variants. Increasing the number of positive charges in the tag deepens the energetic mininum at $z \approx 3$ nm. (**c**) Approximation of the tilting of the energy landscapes from **b** by an applied bias voltage. As the body is negative, increasing the number of positive charges in the tag decreases the electrophoretic force countering the electroosmotic force, resulting in a higher degree of tilting. Down- and upward faces triangles indicate the presence of a local minimum and maximum, respectively.

tag (tail) was represented by 9 smaller beads ($r = 0.5$ nm, $Q_i = -3$ to $+3\,e$ depending on the amount of charges in their corresponding amino acids) each representing 3 amino acids in an alpha-helix (0.6 nm nm spacing). Our reasons for this simplification were two-fold: (1) the high degree of axial symmetry in the bead model resulted in a free energy that was independent of the precise orientation of DHFR, significantly reducing the number of required computations; and (2) the reduced body size of the coarse-grained model compared

to the full atom model allowed for the placement of DHFR along the entire length of the pore without unphysical overlaps between the atoms of ClyA and DHFR inside the *trans* constriction, resulting in more realistic free energies.

2.3 Calculating the Energy Landscape

To obtain the energy profile of a DHFR variant as it moves along the length of the pore ($\Delta G^{\text{pore+part}}$, Eq. (S32)), the net electrostatic energy of the pore (G^{pore}), pore with DHFR ($G^{\text{pore+part}}$) and DHFR alone (G^{part}) were computed using Eq. (S31). This was achieved by placing the bead model at different locations along the central z-axis of ClyA from $z_{\text{body}} = -12.5\,\text{nm}$ to 27.5 nm relative to the center of the bilayer (0.5 nm steps inside the pore) and solving the PBE in the resulting pore–particle combinations with APBS.

All systems were solved using the non-linear PBE (`npbe`) in two steps with the automatic solver (`mg-auto`): (1) a coarse calculation in a box of $40\,\text{nm}\times40\,\text{nm}\times110\,\text{nm}$ with grid lengths of $0.138\,\text{nm}\times0.138\,\text{nm}\times0.122\,\text{nm}$ and multiple Debye-Hückel boundary conditions (`bcfl mdh`), followed by (2) a finer focussing calculation in a box of $15\,\text{nm}\times15\,\text{nm}\times70\,\text{nm}$ with grid lengths of $0.052\,\text{nm}\times0.052\,\text{nm}\times0.052\,\text{nm}$ that used the values of the coarse calculation at its boundaries. The monovalent salt concentration was set to 0.150 M with a radius of 0.2 nm for both ions. The solvent and solute relative permittivities were set to 78.15 and 10, respectively.[7] Both the charge density and the ion accessibility maps were constructed using cubic B-spline discretization (`chgm spl2` and `srfm spl2`).

2.4 Effect of Tag Charge and Bias Voltage

Our simulations showed the existence of an electrostatic energy minimum at $z = 3\,\text{nm}$, the bottom of ClyA's lumen, flanked by two maxima, located at $z = -0.6\,\text{nm}$, the *trans* constriction, and at $z = 5.7\,\text{nm}$, the middle of the *cis* lumen (Fig. S3b). Hence, when DHFR resides at the electrostatic potential minimum inside the nanopore, the largest electrostatic barrier is given by the narrower and negatively charged *trans* constriction while the barrier

at the *cis* side is much shallower.

Increasing the number of positive charges in the fusion tag gives rise to a deepening of the electrostatic minimum (Fig. S3b). This increases the barriers heights at both the *cis* and *trans* sides similarly which results in the lowering of both the *cis* and *trans* escape rates and hence longer dwell times.

As discussed in Sec. 1.2, the magnitude of external force acting on the molecule in a linear potential profile is given by the sum of the electrophoretic ($N_{\text{body}} + N_{\text{tag}}$) and electroosmotic ($N_{\text{eo}}$) contributions

$$E_{\text{ex}} = \begin{cases} (N_{\text{body}} + N_{\text{tag}} + N_{\text{eo}}) \dfrac{V^{\text{bias}}}{14\,\text{nm}} (z - 11\,\text{nm}) & \text{for } -3 < z < 11\,\text{nm}, \\ 0, & \text{for } z >= 11\,\text{nm}, \\ V^{\text{bias}}, & \text{for } z < -3\,\text{nm} \end{cases} \tag{S33}$$

In the case of tagged DHFR, the net charge of the protein is negative ($N_{\text{body}} = -13$) and the electroosmotic flow exerts an opposing force ($N_{\text{eo}} = 15.5$). This means that increasing the number of positive charges in the tag decreases the net electrophoretic force and hence a more pronounced tilting of the entire energy landscape (Fig. S3c). As the *cis* energy barrier in the lumen is relatively shallow, it disappears at moderate applied voltages ($> -50\,\text{mV}$). This suggests that, under an applied bias voltage, the *cis* barrier is located at or near the *cis* entry of the pore.

3 Experimentally Observed Behavior of Tagged DHFR

3.1 Multistate Residences of DHFR Inside ClyA

At −80 mV, in 150 mM NaCl 15 mM Tris−HCl pH 7.5, the current blockades induced by the $DHFR_4S$ variants (Fig. S4 and Tab. S2) showed a main current level (L_1) with relative residual current values (I_{res}%), expressed as a percentage of the open-pore current (I_0), of 67.4±2.1 %, 71.3±0.6 %, 72.8±1.0 %, 74.1±0.4 % and 74.9±0.7 % for $DHFR_4S$, $DHFR_4I$, $DHFR_4C$, $DHFR_4O1$, $DHFR_4O2$, respectively. As observed for other proteins,[8–11] $DHFR_4S$, $DHFR_4O1$ and $DHFR_4O2$ also displayed a second current level (L_2) with I_{res}% values of 46.4±0.2 % 57.0±1.0 % 58.0±0.9 %, respectively (Tab. S2). $DHFR_4I$ and $DHFR_4C$ also occasionally dwelled on a second current level, however, the dwell time at this level was too short to allow reliable determination of the I_{res}% (Fig. S5). $DHFR_4O1$ often visited a third current level (L_3) with I_{res}% of 38.7±1.1 %. It is likely that the multiple current levels observed for the different DHFR variants reflect the residence of the protein in different physical locations inside the nanopore.[8]

The collected time series data may contain capture events that transition to several meta states before the molecule escapes the nanopore. For the purpose of our data analysis, we count the dwell time of an event as the elapsed time from the capture up until the escape, irrespective of how many transitions to meta states have been observed. Thus, the complex kinetics of the molecule in the captured state are simply summed over in the resulting dwell time event histograms (Eq. (S10)). For the scope of this work, we make the assumptions that (1) there is only one dominant capture process involved and that (2) the molecules escape from the nanopore is dominated by a single rate, such that the dwell time distribution can then be approximated by a single exponential. Since a maximum likelihood fit of a single or multi-exponential distribution function is well represented by the arithmetic mean, we use the arithmetic mean directly as the expectation value of the entire distribution (Eq. (S12)).

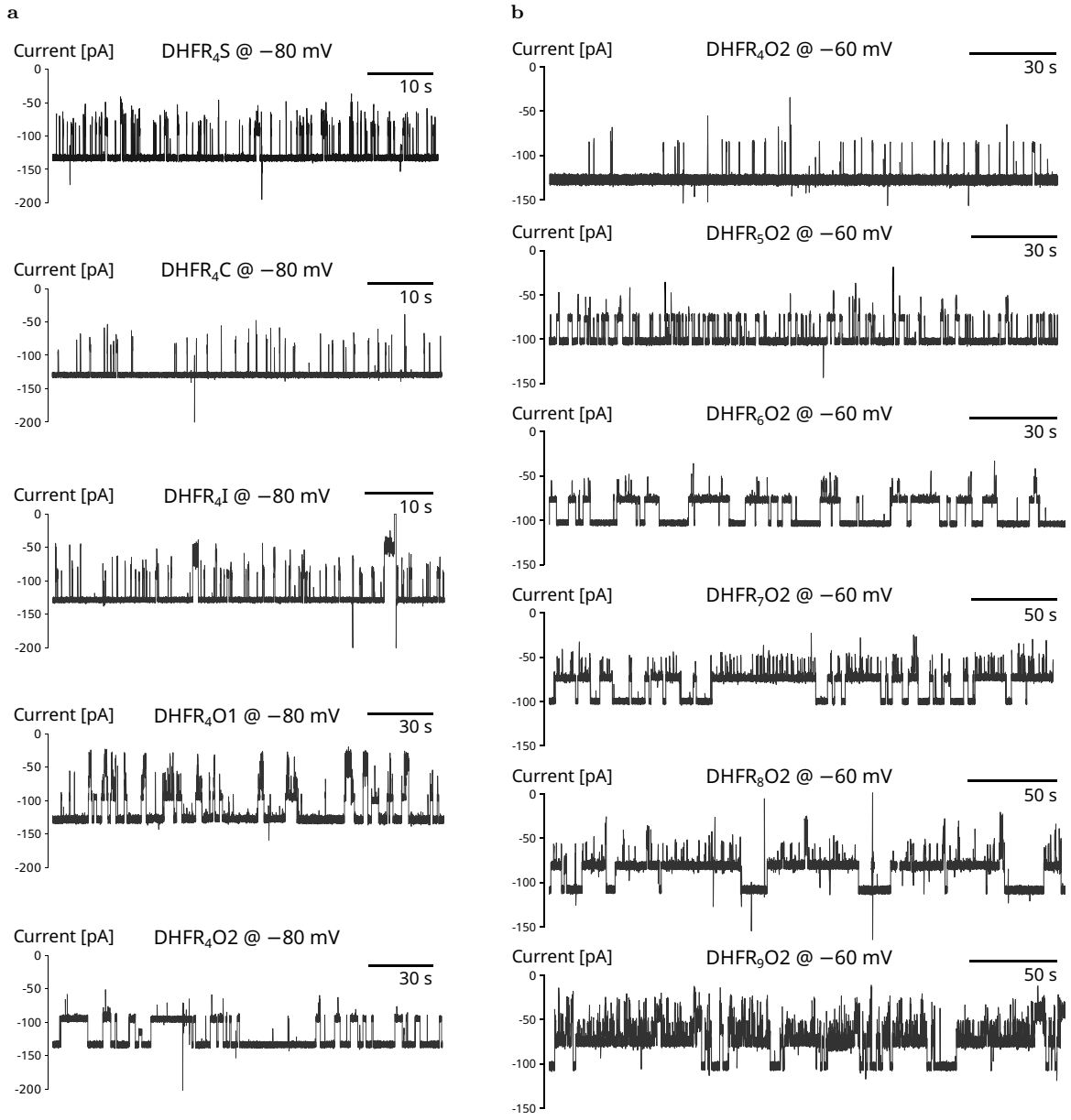

Figure S4. Current blockades of DHFR molecules with a fusion tag. (**a**) Typical current traces of the various $DHFR_4S$ body charge variants ($DHFR_4C$, $DHFR_4I$, $DHFR_4O1$ and $DHFR_4O2$) at −80 mV applied bias. The C, I and O1 mutants have the same net charge, but the location of their charges differ, which results in significantly different dwell times. (**b**) Typical current recordings for the tag charge variants of $DHFR_4O2$ ($DHFR_{N_{tag}}O2$) at −60 mV, revealing the increased dwell time with increasing positive tag charges. Note that most DHFR variants showed complex multi-level blockades. Therefore, average dwell times (t_d) were used to guarantee a fair comparison between the different mutants. All current traces were collected in 150 mM NaCl, 15 mM Tris−HCl pH 7.5 at 28 °C after adding ≈50 nM of DHFR to the *cis* side reservoir of a single ClyA-AS nanopore. Signals were sampled at 10 kHz and filtered with a 2 kHz kHz cut-off Bessel-low pass filter.

Table S2. I_{res}% **values of the different DHFR variants.**

DHFR variant	V^{bias} [mV]	I_{res}%[a] L1	L2	L3
$DHFR_4S$	−80	67.4±2.1	46.4±0.2	—
$DHFR_4I$	−80	71.3±0.6	—	—
$DHFR_4C$	−80	72.8±1.0	—	—
$DHFR_4O1$	−80	74.1±0.4	57.0±1.0	38.7±1.1
$DHFR_4O2$	−80	74.9±0.7	58.0±0.9	—
$DHFR_5O2$	−60	74.0±0.3	57.7±0.1	—
$DHFR_6O2$	−60	74.2±0.1	57.7±0.1	—
$DHFR_7O2$	−60	73.4±0.3	56.4±0.8	39.3±2.5
$DHFR_8O2$	−60	74.7±1.0	58.4±1.1	41.4±1.5
$DHFR_9O2$	−60	74.0±0.1	57.0±0.4	38.9±1.8

[a] I_{res}% values for each DHFR variant are based on at least 50 individual DHFR blockades collected from at least three different single nanopore experiments. Errors are standard deviations from the mean.

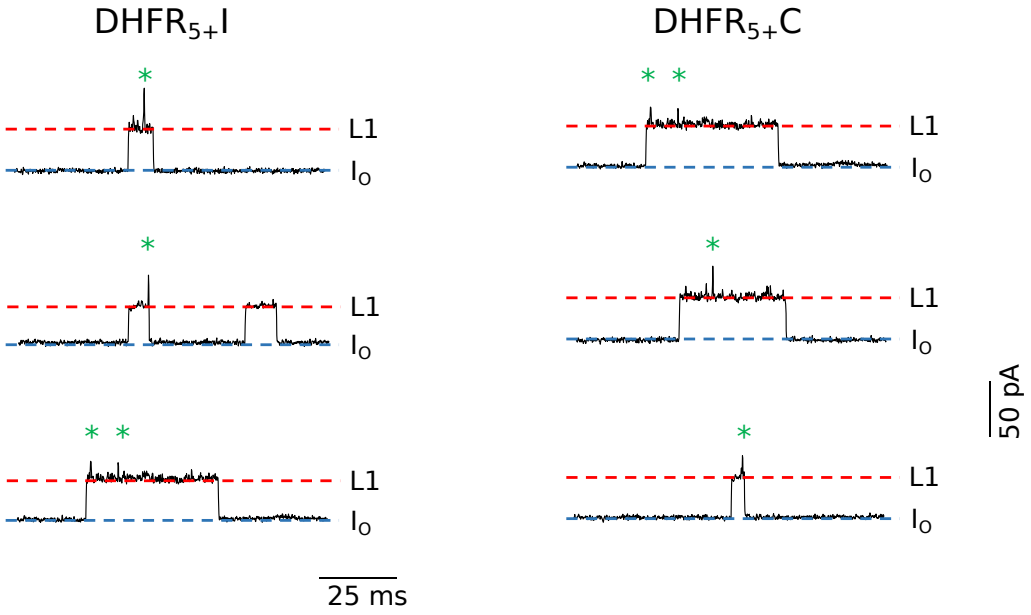

Figure S5. $DHFR_4I$ and $DHFR_4C$ blockades in ClyA-AS at −80 mV. Three individual $DHFR_4I$ (left) and $DHFR_4C$ (right) blockades to ClyA-AS at −80 mV showing the L1 current level (red dashed line) and short dwelling on a lower current level (green asterisks). The latter is too short to be properly sampled at this potential (transitions to this additional current level are observed by short, unresolved spikes). The blue dashed line represents the open-pore current I_0. The current traces were collected in 150 mM NaCl, 15 mM Tris−HCl pH 7.5 at 28 °C, by applying a Bessel-low pass filter with a 2 kHz kHz cut-off and sampled at 10 kHz.

3.2 Analysis of NADPH Binding to DHFR Variants

Typical current traces of NADPH binding to trapped DHFR molecules are shown in Fig. S6. Analysis of the on- (k_{on}) and off-rates (k_{off}), and the event amplitudes of NADPH binding to $DHFR_4O2$, $DHFR_6O2$ and $DHFR_7O2$ entrapped within the nanopore are all similar

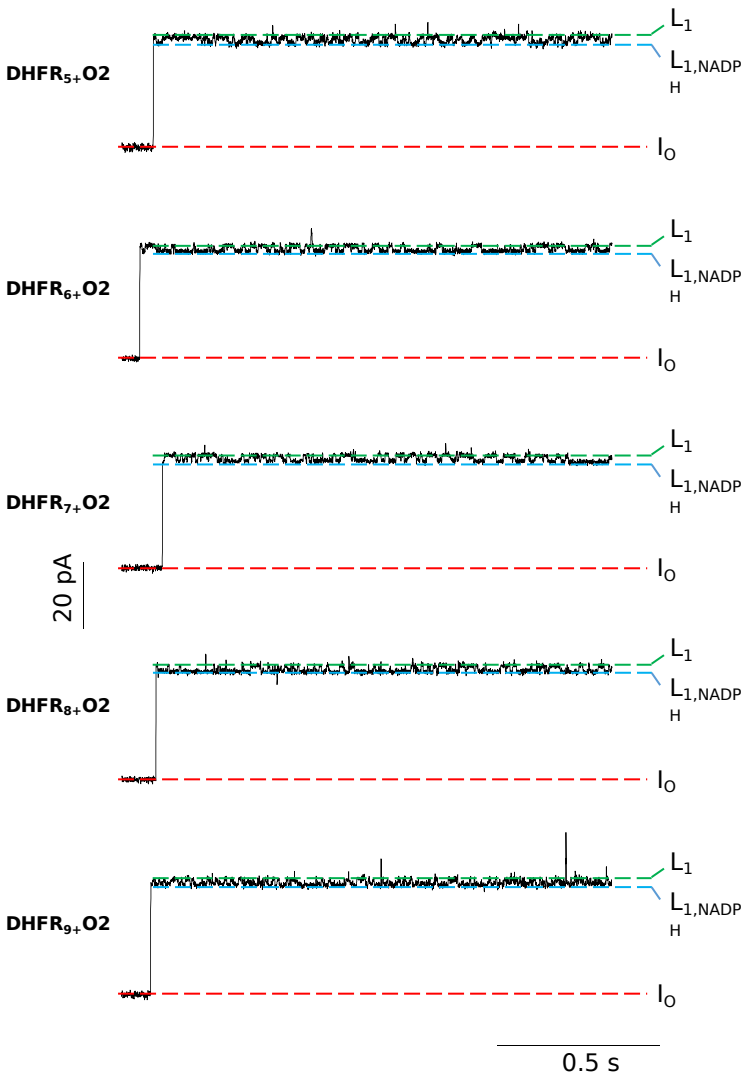

Figure S6. NADPH binding to nanopore-confined DHFR$_{N_{tag}}$O2. Typical current traces of single DHFR$_{N_{tag}}$O2 (≈50 nM, *cis*) molecules inside ClyA-AS at −80 mV applied potential after addition of 40 µM NADPH to the trans compartment. NADPH binding to confined DHFR$_{N_{tag}}$O2 is reflected by current enhancements from the unbound L1 (green dashed line) to the NADPH-bound L1-NADPH (blue dashed line) current levels. The open-pore current I_0 is represented by the red dashed line. All current traces were collected in 150 mM NaCl, 15 mM Tris−HCl pH 7.5 at 28 °C, by applying a Bessel-low pass filter with a 2 kHz kHz cut-off and sampled at 10 kHz. An additional Bessel 8-pole filter with 500 Hz cut-off was digitally applied to the current traces.

(Tab. S3), suggesting that the proteins remain folded and active inside ClyA.

Table S3. NADPH binding/unbinding kinetics to trapped DHFR variants.[a]

DHFR variant	k_{on} [$s^{-1} \cdot mM^{-1}$][b]	k_{off} [s^{-1}]	Amplitude [pA]
$DHFR_4O2$	1684±225	60.2±23.2	−1.71±0.13
$DHFR_6O2$	1390±397	55.9±5.1	−1.47±0.09
$DHFR_7O2$	2032±578	71.2±20.4	−1.46±0.07

[a] At −60 mV applied potential.

[b] k_{on}, k_{off} and amplitude values for each DHFR variant are based on at least 300 NADPH binding events on more than 15 individual DHFR blockades collected from three different single nanopore experiments.

4 Modeling of Body Charge Variations

4.1 Not All Charges on the Body Are Equivalent

The double barrier model of Eq. (6) in its current form cannot adequately account for the body charge variations of DHFR. Consider for example the body charge variants $DHFR_4I$, $DHFR_4C$, and $DHFR_4O1$ which share the same number of body charges, so that our model would predict the same dwell time (Fig. S7a). However, the body charges of these variations are at different locations and as a consequence they exhibit different dwell times as can be seen in Fig. 3b. The reason for this is that the model describes the trapping mechanism as a function of the tag charge number N_{tag} only, while body charge related barrier modifications are absorbed into the constant terms $\Delta E_{\mathrm{es},0}^{\mathrm{cis/trans}}$ of Eq. (4).

If we wanted to modify Eq. (4) such that it can account for changes in the trapping behavior we would need to account for the location of the body charges as is evident from the dwell time data sets of $DHFR_4I$, $DHFR_4C$, and $DHFR_4O1$. Such a model would drastically increase in complexity and it is not clear whether it can still be formulated analytically in a reasonable way. In that case it may in fact be more workable to use a more refined APBS simulations or even a full molecular dynamics simulation to compute parameters for the trapping.

4.2 The Distance From the Tag Matters

Despite this limitation, the double barrier model is essentially a representation of how the tag is anchored to the electrostatic minimum in the pore, and hence we can deduce the way the body charge variations will impact trapping. For example, if we modify the charge on the far end of the body with respect to the tag location (*i.e.*, far away from the tag). We hypothesize that the barriers responsible for trapping the tag will not be changed meaningfully—aside from the electrophoretic force which is included in the double barrier model (Eq. (6) of the main text) Indeed, taking the model parameters of Tab. 2 obtained from the fit to the

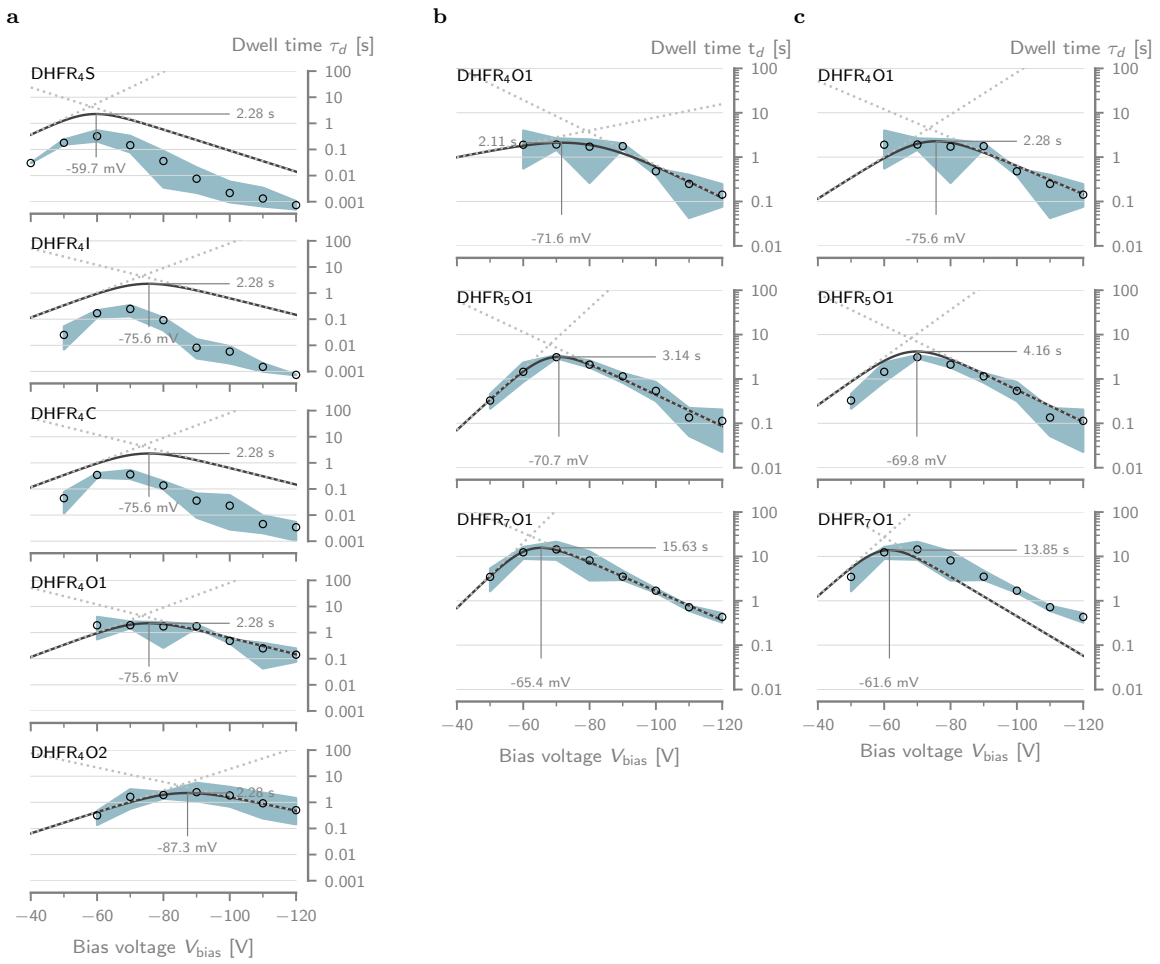

Figure S7. Effect of body and tag charge on the dwell time of DHFR. (**a**) Predicted dwell times of the body charge variations DHFR$_4$S, -I, -C, -O1 and -O2 by Eq. (2) in the main text and using the parameters in Tab. 2 of the main text. Clearly, the location of the body charge plays an important, but uncaptured, role in determining the dwell time of DHFR. (**b**) and (**c**) are the voltage dependencies of the mean dwell time (t_d), fitted with the simple barrier model of Eq. (1) and the full double barrier model of Eq. (6), respectively. The annotated threshold voltages for **b** and **c** were computed by respectively Eq. (S25) and Eq. (S26). Solid lines represent the double barrier dwell time while the dotted lines show the dwell times due the *cis* (low to high) and *trans* (high to low) barriers. Fitting parameters can be found in Tab. S4.

DHFR$_{N_{tag}}$O2 data set and *predicting* the dwell time data of DHFR$_{N_{tag}}$O1, we find good agreement as shown in Fig. S7.

On the other hand, charges on the body that are close to the tag directly impact the electrostatic energy landscape and will modify the barrier heights. Such a body charge close to the tag location can be seen as effectively modifying the net charge on the tag. In that case, we expect that our model fitted to the DHFR$_{N_{tag}}$O2 data fails to *predict* the dwell

times as can be seen in Fig. S7a.

Table S4. Fitting parameters for the simple double barrier model.[a]

	cis barrier[b]		*trans* barrier[b]	
DHFR variant	$\ln k^{\mathrm{cis}}/V_T$	$\alpha^{\mathrm{cis}}/V_T$	$\ln k^{\mathrm{trans}}/V_T$	$\alpha^{\mathrm{trans}}/V_T$
$DHFR_4S$	11.91±3.14	5.38±1.83	−5.45±0.86	2.82±0.24
$DHFR_4I$	15.15±4.88	5.88±2.34	−6.83±1.30	3.08±0.34
$DHFR_4C$	15.87±3.52	6.56±1.72	−5.73±0.79	2.50±0.21
$DHFR_4O1$	1.37±3.29	0.88±1.37	−8.03±2.64	2.16±0.60
$DHFR_4O2$	9.67±2.53	3.70±1.01	−5.54±1.43	1.31±0.34

[a] Fitting coefficients for Eq. (1) of the main text.

[b] Errors represent one sigma confidence intervals.

5 Materials and Methods

5.1 Material Suppliers

Unless otherwise specified all chemicals were bought from Sigma-Aldrich (Overijse, Belgium). DNA was purchased from Integrated DNA Technologies (IDT, Leuven, Belgium), enzymes from Fermentas (Merelbeke, Belgium) and lipids from Avanti Polar Lipids (Alabaster, USA).

5.2 Cloning of all DHFR Variants

5.2.1 Cloning of $DHFR_4S$

The $DHFR_4S$ DNA construct was built from the $DHFR_{tag}$ construct[10] by inserting an additional alanine residue at position 175 (located in the fusion tag). $DHFR_{tag}$ contains two mutations with respect to wild type *E. coli* DHFR (C85A and C152S) and has a C-terminal fusion tag which possesses five net positive charges and ends with a Strep-tag. To construct $DHFR_4S$, the $DHFR_{tag}$ circular DNA template was amplified using 175Ala frwd (forward) and T7 terminator (reverse) primers (Tab. S5) in the following PCR reaction: ≈200 ng of template plasmid and ≈16 µM of forward and reverse primers were mixed in 0.3 ml final volume of PCR mix, which contained 150 µl RED Taq Ready Mix (Sigma-Aldrich). We performed 34 PCR cycles following a pre-incubation step at 98 °C for 30 scycling protocol: denaturation at 98 °C for 10 s, annealing at 52 °C for 30 s, extension at 72 °C for 60 s; and a final elongation step at 72 °C for 10 min. The resulting PCR product was clean-upped using the QIAquick PCR purification kit (Qiagen) and further gel purified using the QIAquick Gel Extraction Kit (Qiagen) before it was cloned into a pT7 expression plasmid (pT7-SC1)[12] by the MEGAWHOP procedure:[13] ≈400 ng of the purified PCR product was mixed with ≈200 ng of the $DHFR_{tag}$ DNA template and amplification was carried out with Phire Hot Start II DNA polymerase (Finnzymes) in 50 µl final volume (pre-incubation at 98 °C for 30 s; then cycling: denaturation at 98 °C for 5 s, extension at 72 °C for 90 s, for 30 cycles; followed by a final extension for 10 min at 72 °C). The template DNA was eliminated by incubation

with DpnI (1 FDU) for 1 h at 37 °C and the enzyme was inactivated by incubation at 65 °C for 5 min. Finally, 0.5 µl of the resulting mixture was transformed into 50 µl of E. cloni® 10G electrocompetent cells (Lucigen) by electroporation. The transformed bacteria were grown overnight at 37 °C on LB agar plates supplemented with 100 $\mu g \cdot ml^{-1}$ ampicillin. The identity of the clones was confirmed by sequencing. DNA and protein sequences of $DHFR_4S$ are listed below.

```
>DHFR4S (DNA sequence)
ATGGCTTCGGCTATGATTTCTCTGATTGCGGCACTGGCTGTCGATCGTGTTATTGGTATG
GAAAACGCTATGCCGTGGAATCTGCCGGCTGATCTGGCGTGGTTTAAACGTAACACTCTG
GACAAGCCGGTCATTATGGGCCGCCATACGTGGGAAAGCATCGGTCGTCCGCTGCCGGGT
CGCAAAAATATTATCCTGAGCAGCCAGCCGGGCACCGATGACCGTGTGACGTGGGTTAAG
AGCGTCGATGAAGCAATTGCGGCGGCAGGCGACGTGCCGGAAATTATGGTTATCGGCGGT
GGCCGCGTTTATGAACAGTTCCTGCCGAAAGCCCAAAAGCTGTACCTGACCCATATCGAT
GCAGAAGTCGAAGGTGATACGCACTTTCCGGACTATGAACCGGATGACTGGGAAAGTGTG
TTCTCCGAATTTCACGACGCCGACGCTCAGAACAGCCACTCATACTCATTCGAAATCCTG
GAACGCCGTGGCAGCAGTACTCGAGCGAAAAAGAAGATTGCGgccGCCCTAAAACAGGGC
AGCGCGTGGAGCCATCCGCAGTTTGAAAAATGATAA


>DHFR4S (protein sequence)
MASAMISLIAALAVDRVIGM
ENAMPWNLPADLAWFKRNTL
DKPVIMGRHTWESIGRPLPG
RKNIILSSQPGTDDRVTWVK
SVDEAIAAAGDVPEIMVIGG
GRVYEQFLPKAQKLYLTHID
AEVEGDTHFPDYEPDDWESV
FSEFHDADAQNSHSYSFEIL
ERRGSSTRAKKKIAAALKQG
SAWSHPQFEK**
```

5.2.2 Construction of All Other Variants

The positions for introduction of negatively charged glutamate residues into $DHFR_4S$ were chosen after multiple sequence alignment of *E. coli* DHFR (PDB ID: 1RH3, BLAST, 250 results) and identification of the residues that were located on the opposite end of the molecule than the 5+tag, and which during evolution had already converted to glutamate in some sequences. In all $DHFR_4S$ variants described in this work two native residues of $DHFR_4S$ were mutated, resulting in DHFR constructs with two ($DHFR_4C$; A82E/A83E), ($DHFR_4I$; V88E/P89E), ($DHFR_4O1$; T68E/R71Q) or three ($DHFR_4O2$; T68E/R71E) extra negative charges. To construct the $DHFR_4C$, $DHFR_4I$ and $DHFR_4O2$ mutants, the $DHFR_4S$ gene was amplified using the C-frwd, I-frwd or O2-frwd primer, respectively, and the T7 terminator primer. The PCR conditions and subsequent purification and cloning steps were as described above (Sec. 5.2). The $DHFR_4O1$ mutant was constructed starting from the $DHFR_4O2$ circular DNA template as described above, using O1-frwd and T7 terminator primers in the first PCR amplification step. Following the same strategy, $DHFR_5O1/O2$ mutants were constructed using the corresponding $DHFR_4O1/O2$ circular DNA template, 6+ frwd primer and T7 terminator primer. $DHFR_6O1/O2$ and $DHFR_7O1/O2$ mutants were constructed using the corresponding $DHFR_5O1/O2$ circular DNA template, 6+ frwd or 7+ frwd primer, respectively, and T7 terminator primer. $DHFR_8O1/O2$ and $DHFR_9O1/O2$ mutants were constructed using the corresponding $DHFR_7O1/O2$ circular DNA template, 8+ frwd or 9+ frwd primer, respectively, and T7 terminator primer. All primer sequences are shown in Tab. S5.

5.3 Protein Overexpression and Purification

5.3.1 Strep-Tagged DHFR Mutants

The pT7-SC1 plasmid containing the DHFR gene and the sequence of the Strep-tag at its C-terminus was transformed into E. cloni® EXPRESS BL21(DE3) cells (Lucigen), and trans-

Table S5. Mutagenesis DNA primer sequences.

Primer name	Primer sequence
175Ala frwd	GGCAGCAGTACTCGAGCGAAAAAGAAGATTG CGgccGCCCTAAAACAGGGCAGCGCGTGG
T7 terminator	GCTAGTTATTGCTCAGCGG
O2-frwd	CCTGAGCAGCCAGCCGGGCGAAGATGACGA AGTGACGTGGGTTAAGAGCGTCG
I-frwd	GAGCGTCGATGAAGCAATTGAAGAAGCAGG CGACGTGCCGGAAATTATGGTTATCGGCGG
C-frwd	GCAATTGCGGCGGCAGGCGACGAAGAGGAA ATTATGGTTATCGGCGGTGGCCGCG
O1-frwd	CCTGAGCAGCCAGCCGGGCGAAGATGACCA GGTGACGTGGGTTAAGAGCGTCG
5+ frwd	CTCGAGCGAAAAAGAAGATTGCGAAAGCCC TAAAACAGGGCAGCGCGTGGAGCCATCCGC
6+ frwd	CGTGGCAGCAGTACTCGAGCGAAAAAGAAG ATTAAGAAAGCCCTAAAACAGGGCAGCGCG
7+ frwd	CGTGGCAGCAGTACTCGAGCGAAAAAGAAG ATTAAGAAAAAGCTAAAACAGGGCAGCGCG
8+ frwd	GGCAGCAGTACTCGAAAGAAAAAGAAGATT AAGAAAAAGCTAAAACAGGGCAGCGCGTGG
9+ frwd	GGCAGCAGTACTCGAAAGAAAAAGAAGATT AAGAAAAAGAAGAAACAGGGCAGCGCGTGG

formants were selected on LB agar plates supplemented with $100\,\mu g \cdot ml^{-1}$ ampicillin after overnight growth at 37 °C. The resulting colonies were grown at 37 °C in 2xYT medium supplemented with $100\,\mu g \cdot ml^{-1}$ ampicillin until the O.D. at 600 nm was ≈0.8 (200 rpm shaking). The DHFR expression was subsequently induced by addition of 0.5 mM IPTG (isopropyl β-D-1-thiogalactopyranoside), and the temperature was switched to 25 °C for overnight growth (200 rpm shaking). The next day the bacteria were harvested by centrifugation at 6000 g at 4 °C for 25 min and the resulting pellets were frozen at −80 °C until further use.

Bacterial pellets originating from 50 ml culture were resuspended in 30 ml lysis buffer (150 mM NaCl, 15 mM Tris–HCl pH 7.5, 1 mM $MgCl_2$, 0.2 units/ml DNase, $10\,\mu g \cdot ml^{-1}$ lysozyme) and incubated at 37 °C for 20 min. After further disruption of the bacteria by probe sonication the crude lysate was clarified by centrifugation at 6000 g at 4 °C for 30 min.

The supernatant was allowed to bind to ≈150 µl (bead volume) of Strep-Tactin® Sepharose® (IBA) pre-equilibrated with the wash buffer (150 mM NaCl, 15 mM Tris – HCl pH 7.5)—'end over end' mixing. The resin was then loaded onto a column (Micro Bio Spin, Bio-Rad) and washed with ≈20 column volumes of the wash buffer. Elution of DHFR from the column was achieved by addition of ≈100 µl of elution buffer (150 mM NaCl, 15 mM Tris – HCl pH 7.5, ≈15 mM D-Desthiobiotin (IBA)). Proteins were aliquoted and frozen at −20 °C until further use. New aliquots of DHFR were thawed prior to every experiment.

5.3.2 His-Tagged Type I ClyA-AS

E. cloni® EXPRESS BL21 (DE3) cells were transformed with the pT7-SC1 plasmid containing the ClyA-AS gene. ClyA-AS contains eight mutations relative to the *S. Typhi* ClyA-WT: C87A, L99Q, E103G, F166Y, I203V, C285S, K294R and H307Y (the H307Y mutation is in the C-terminal hexahistidine-tag added for purification).[9] Transformants were selected after overnight growth at 37 °C on LB agar plates supplemented with 100 µg · ml^{-1} ampicillin. The resulting colonies were grown at 37 °C (200 rpm shaking) in 2xYT medium supplemented with 100 µg · ml^{-1} ampicillin until the O.D. at 600 nm was ≈0.8. ClyA-AS expression was then induced by addition of 0.5 mM IPTG, and the temperature was switched to 25 °C for overnight growth (200 rpm shaking). The next day the bacteria were harvested by centrifugation at 6000 g for 25 min at 4 °C and the pellets were stored at −80 °C.

Pellets containing monomeric ClyA-AS arising from 50 ml culture were thawed and resuspended in 20 ml of wash buffer (10 mMimidazole, 150 mM NaCl, 15 mM Tris – HCl pH 7.5), supplemented with 1 mM $MgCl_2$ and 0.05 units/ml of DNaseI. After lysis of the bacteria by probe sonication, the crude lysates were clarified by centrifugation at 6000 g for 20 min at 4 °C and the supernatant was mixed with 200 µl of Ni – NTA resin (Qiagen) equilibrated in wash buffer. After 60 min, the resin was loaded into a column (Micro Bio Spin, Bio-Rad) and washed with ≈5 ml of the wash buffer. ClyA-AS was eluted with approximately ≈0.5 ml of wash buffer containing 300 mM imidazole. ClyA-AS monomers were stored at 4 °C until

further use.

ClyA-AS monomers were oligomerized by addition of 0.5 % β-dodecylmaltoside (DDM, GLYCON Biochemicals, GmbH) and incubation at 37 °C for 30 min. ClyA-AS oligomers were separated from monomers by blue native polyacrylamide gel electrophoresis (BN-PAGE, Bio-rad) using 4 to 20 % polyacrylamide gels. The bands corresponding to Type I ClyA-AS were excised from the gel and placed in 150 mM NaCl, 15 mM Tris – HCl pH 7.5, supplemented with 0.2 % DDM and 10 mM EDTA to allow diffusion of the proteins out of the gel.

5.4 Single Nanopore Experiments

5.4.1 Electrical Recordings in Planar Lipid Bilayers

By convention, the applied potential refers to the potential of the trans electrode in the planar lipid bilayer set up. ClyA-AS nanopores were inserted into lipid bilayers from the *cis* compartment, which is connected to the ground electrode. The *cis* and *trans* compartments are separated by a 25 µm thick polytetrafluoroethylene film (Goodfellow Cambridge Limited) containing an orifice of ≈100 µm in diameter. After pre-treatment of the aperture with ≈5 µl of 10 % hexadecane in pentane, a bilayer was formed by the addition of ≈10 µl of 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) in pentane (10 mg · ml^{-1}) to both electrophysiology chambers. Typically, the addition of 0.01 to 0.1 ng of pre-oligomerised ClyA-AS to the *cis* compartment (0.5 ml) was sufficient to obtain a single channel. Since ClyA-AS nanopores displayed a higher open-pore current at positive than at negative applied potentials, the orientation of the pore could be easily assessed. All electrical recordings were carried out in 150 mM NaCl, 15 mM Tris – HCl pH 7.5. The temperature of the recording chamber was maintained at 28 °C by water circulating through a metal case in direct contact with the bottom and sides of the chamber.

5.4.2 Data Recording and Event Analysis

Electrical signals from planar lipid bilayer recordings were amplified using an Axopatch 200B patch clamp amplifier (Axon Instruments, San Jose, USA) and digitized with a Digidata 1440 A/D converter (Axon Instruments, San Jose, USA). Data were recorded using the Clampex 10.5 software (Molecular Devices, San Jose, USA) and the subsequent event analysis was carried out with the Clampfit software (Molecular Devices). Ionic currents were sampled at 10 kHz and filtered with a 2 kHz low-pass Bessel filter.

Residual current values ($I_{res}\%$) of the different DHFR variants were calculated by $I_{res}\% = I_b/I_o$, in which I_b and I_o represent the blocked an open-pore current values, respectively. I_b and I_o values were calculated from Gaussian fits to all point current histograms (0.1 pA bin size) from at least 3 individual single channels each displaying at least 50 current blockades. The average residence time of the DHFR mutants was determined using the 'single channel search' feature in Clampfit. The detection threshold was set to ≈75 % of the open-pore current and events shorter than 1 ms were ignored for all mutants except for $DHFR_4S$, $DHFR_4I$ and $DHFR_4C$, as they exhibited very short dwell times inside ClyA-AS. The process of event collection was monitored manually.

The average of the mean dwell times obtained from at least three single channels each displaying at least 100 blockades was used to describe the average dwell time (t_d) at every potential. For analysis of NADPH binding events to $DHFR_{N_{tag}}O2$, traces were filtered digitally with a 8-pole low-pass Bessel filter with a 500 Hz cut-off. Current transitions from L1 to L1-NADPH were analysed with the 'single channel search' option in Clampfit. The detection threshold to collect the NADPH-induced events was set to 2 pA and events shorter than 0.1 ms were ignored. The process of event collection was monitored manually. The resulting event dwell times (τ_{off}) and the times between events (τ_{on}) were binned together as cumulative distributions and fitted to a single exponential to retrieve the NADPH-induced lifetimes (τ_{off}) and the NADPH-induced inter-event times (τ_{on}). The average amplitude of the events was derived from Gaussian fits to the conventional distributions of the events'

amplitudes. Values for off and on rates were determined as $k_{\text{off}} = 1/\tau_{\text{off}}$ and $k_{\text{off}} = (\tau_{\text{on}}c)^{-1}$ with c the concentration of NADPH added to the *trans* solution. Final values for τ_{on}, τ_{off}, k_{on}, k_{off} and event amplitudes are the averages derived from three single channel experiments, each analysing at least 100 binding events on more than five different $\text{DHFR}_{N_{\text{tag}}}\text{O2}$ blockades.

References

(1) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037–10041.

(2) Baker, N. A. *Rev. Comput. Chem.*; John Wiley & Sons, Inc., 2005; pp 349–379.

(3) Franceschini, L.; Brouns, T.; Willems, K.; Carlon, E.; Maglia, G. DNA Translocation Through Nanopores at Physiological Ionic Strengths Requires Precise Nanoscale Engineering. *ACS Nano* **2016**, *10*, 8394–8402.

(4) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.

(5) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* **2007**, *35*, W522–W525.

(6) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.

(7) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.

(8) Soskine, M.; Biesemans, A.; Moeyaert, B.; Cheley, S.; Bayley, H.; Maglia, G. An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry. *Nano Lett.* **2012**, *12*, 4895–4900.

(9) Soskine, M.; Biesemans, A.; Maeyer, M. D.; Maglia, G. Tuning the Size and Properties of ClyA Nanopores Assisted by Directed Evolution. *J. Am. Chem. Soc.* **2013**, *135*, 13456–13463.

(10) Soskine, M.; Biesemans, A.; Maglia, G. Single-Molecule Analyte Recognition with ClyA Nanopores Equipped with Internal Protein Adaptors. *J. Am. Chem. Soc.* **2015**, *137*, 5793–5797.

(11) Biesemans, A.; Soskine, M.; Maglia, G. A Protein Rotaxane Controls the Translocation of Proteins Across a ClyA Nanopore. *Nano Lett.* **2015**, *15*, 6076–6081.

(12) Miles, G.; Cheley, S.; Braha, O.; Bayley, H. The Staphylococcal Leukocidin Bicomponent Toxin Forms Large Ionic Channels. *Biochemistry* **2001**, *40*, 8514–8522.

(13) Miyazaki, K. *Methods Enzymol.*; Elsevier, 2011; pp 399–406.