Accurate Determination Of The Diffusion Coefficient Of Proteins By Fast Fourier Transformation With Whole Column Imaging Detection

Atefeh S. Zarabadi, Janusz Pawliszyn *

Department of Chemistry, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada

* Corresponding author: janusz@uwaterloo.ca, Tel.: +1-519-8884641; Fax: +1-519-7460435

Table of Contents

1 Time Domain Approach

According to the Einstein equation, the diffusion coefficient is related to peak area at its corresponding time:

Eq. S- 1

There are different methods to get the σ^2 value of the decaying peaks in the time domain. The spatial approaches described here are full width at half maximum (FWHM) and curve fitting. The obtained variance is then plotted against time, and the slope of the curve has a linear relationship with the diffusion coefficient.

1.1 Scanning Procedure

The following scanning procedure was employed to record the diffusion process. First, the sample is injected to the cartridge, and a reference image (I_0) of the filled column is scanned prior to the focusing or pre-concentration. Then, the absorption is acquired after applying the electric field. The camera automatically takes images almost every 30 seconds until one abort it when the protein is focused at its pI (CIEF) or the plug reaches the middle of the channel (iPF). After turning off the applied voltage, the protein band relaxes, resulting in band broadening. The CCD camera then manually scans the dynamic diffusion process. Each image (I_i) is scanned individually at desired intervals, which are timed by a stopwatch. The diffusion images are primarily in the form of light intensity. Therefore, absorption is calculated using these images by dividing all the images to the I_0, and taking a logarithm of the divided images.

1.2 Full Width At Half Maximum (FWHM)

The concentration profile is fitted to a Gaussian function and the variance is approximated from the peak width at half height ($W_{1/2}$). It is a useful approximation for systems that do not electronically integrate peaks from a baseline. The simplest method is to take the average of the two widths that surround the half height point and consider its relationship with the peak variance, which is given by:

Eq. S- 2

The flow chart is self-explanatory and describes the implemented approach, which is based on the relationship between the variation of the signal width at peak half height and time. The FWHM approach is applied using a peak-finding algorithm that locates the peak maximum (P_{max}), then locates the first positions on either side where the y-value falls to ½(P_{max}), which directly gives the $W_{1/2}$ value.

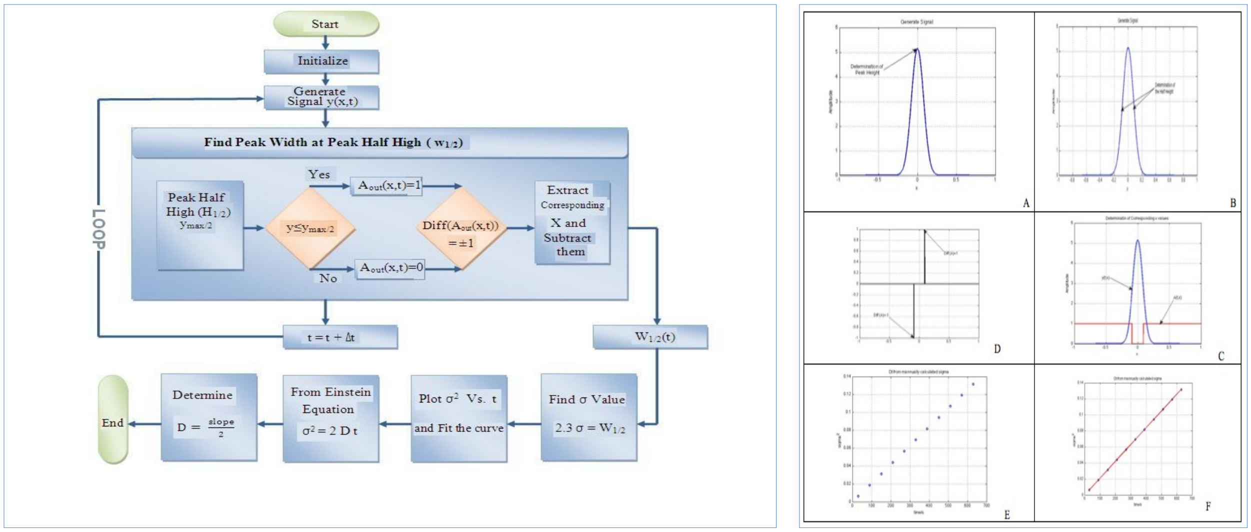

Start
Initialize
Generate Signal y(x,t)
Find Peak Width at Peak Half High (w1/2)
Peak Half High (H1/2) ymax/2
Yes
Aout(x,t)=1
y≤ymax/2
Diff(Aout(x,t)) =±1
Extract Corresponding X and Subtract them
No
Aout(x,t)=0
LOOP
t=t+Δt
W1/2(t)
End
Determine D = slope/2
From Einstein Equation σ²=2Dt
Plot σ² Vs. t and Fit the curve
Find σ Value 2.3σ = W1/2
Generate Signal
Determination of Peak Height
A
B
C
D
E
F

Figure S- 1 Algorithm for time domain calculation by full width at half maximum method

In sum, FWHM is an indicator of peak broadening and variations during the diffusion process. However, the variance can be directly extracted by curve fitting too.

1.3 Curve Fitting: Single and Multiple-term Gaussian Models

Quantitative analysis by curve fitting approach is very popular in chromatographic methods, because of the information provided by the fitted function, such as the statistical moments and the peak shape parameters. Beside its advantages, curve fitting intrinsically imposes an uncertainty to the results. Uncertainty sources of this approach include the baseline subtraction, the selected fitting boundaries, and fitting initial parameters dependence.

Electrophoresis peaks do not necessarily have a Gaussian shape. As demonstrated in the main text (Table 4), the results improved for proteins with Gaussian-like peaks while it does not work for those deviated from Gaussian shape. Examples of the non-Gaussian profiles are carbonic anhydrase and BSA, which significantly deviate from Gaussian shape (Figure S- 16 and Figure S- 17), and lead to high error by curve fitting method. A single Gaussian function did not properly fit with these experimental data, and a multiple term Gaussian equation provided a better fitting. Three major coefficients of each Gaussian term; amplitude, center, and variance are altered to get the best fit according to the root mean square error (RMSE) for peaks that deviate from normal distribution. However, myoglobin's concentration profile, Figure S- 15, as a representative of a fairly Gaussian shape signal, provides a good estimation with less than 10% error.

1.3.1 Gaussian Multiple-term Model

The use of curve fitting to obtain reliable quantitative information about the electrophoresis data requires an accurate representation of the band-shape by fitted model. In case of a mixture of one-dimensional normal distributions with initial parameters; means μ_i and variances σ_i^2, the total variance is calculated by:

Eq. S- 3

Following sections, results from single Gaussian model and Gaussian Mutiple-term (GMT) model are demonstrated for carbonic anhydrase and BSA. When we are using the GMM, we need to be cautious about how far we move. By increasing the number of terms we will get a model that has smaller RMSE, which does not necessarily means a better fit. There is a possibility of over fitting that needs to come into account. We need to verify the fitted model by test set. To do so, the experimental data are divided into two; training and test sets. The former used to get the function while the latter employs for validating the fitted model.

Carbonic Anhydrase I

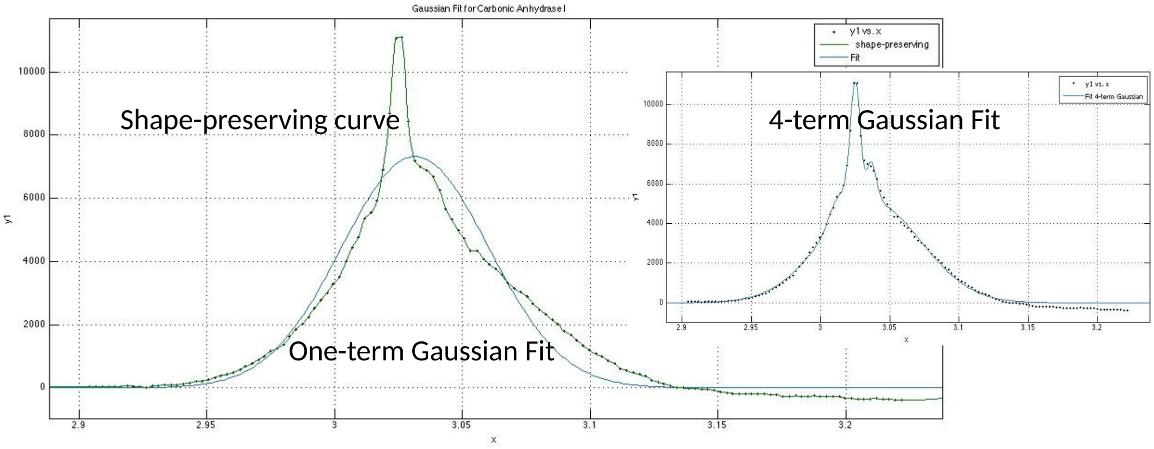


Figure S- 2 Curve fitting for carbonic anhydrase I, with one- and 4-term Gaussian models. Goodness of fit for one-term Gaussian: R-square: 0.9256, RMSE: 689.3 and Goodness of fit for GMT (4-term): R-square: 0.9953, RMSE: 179

Bovine Serum Albumin

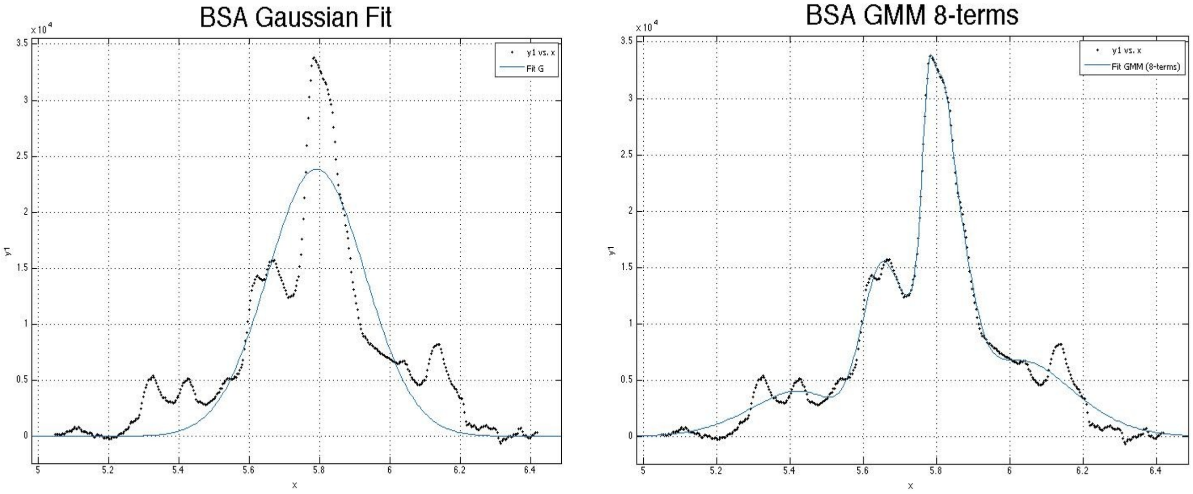


Figure S- 3 Curve fitting for bovine serum albumin (BSA), with single Gaussian and 8-term GMT Models. Goodness of fit for G: R-square: 0.8114, RMSE: 3533, and Goodness of fit for GMM (8-term): R-square: 0.9867, RMSE: 964

The curve fitting technique can be employed for more precise calculations (within 95% confidence interval). The time domain approach was performed by FWHM method for the sake of simplicity.

Protein	Gaussian Model		GMT Model		$D_{Lit.}$ [1] (cm^2/s)
	D_G (cm^2/s)	%Error	D_{GMT} (cm^2/s)	%Error	
Carbonic Anhydrase I	7.66×10^{-7}	28.11	11.62×10^{-7}	9.05	10.66×10^{-7}
Bovine Serum Albumin (BSA)	2.86×10^{-7}	51.41	4.23×10^{-7}	28.22	5.90×10^{-7}

Table S- 1 Comparison of the curve fitting results for carbonic anhydrase and BSA irregular peak shapes by Gaussian and GMT Models

Multiple-term model significantly improves the acquired results for non-Gaussian peaks in the time domain. However, the estimated diffusion coefficients from FFT method are still providing more accurate results.

2 Frequency Domain Approach

The Fourier transformation carries the signals from the time domain to the frequency domain. The visual abstract of the procedure is presented in the following flowchart. The absolute values of the Fourier transform of decaying signals at different times t_i are divided by the absolute value of the Dirac at initial time t_0. In fact, a delta Dirac function is considered a starting point, where $\sigma^2 \rightarrow 0$. The Fourier transform of a Dirac delta function is known to be a constant; hence, the Fourier transform at t=0 is set to the constant 1/p, where p is the spatial resolution, and this constant value comes from the sampling rate. The absolute value of the Fourier-transformed function is taken to remove the imaginary phase, and ‘fftshift’ command employed to centralize it to form the Gaussian-like output. Putting a negative logarithm of this ratio equal to $D\boldsymbol{\omega}^2 t$ results in a new term, $Q= -\ln(FFT(y_i)/FFT(y_0))/\omega^2$, which has a linear dependence on time and where y_0 and y_i denotes the concentration profile at the initial and given times. Plotting this term, Q versus time provides a linear curve, and its slope is the diffusion coefficient from the frequency domain.

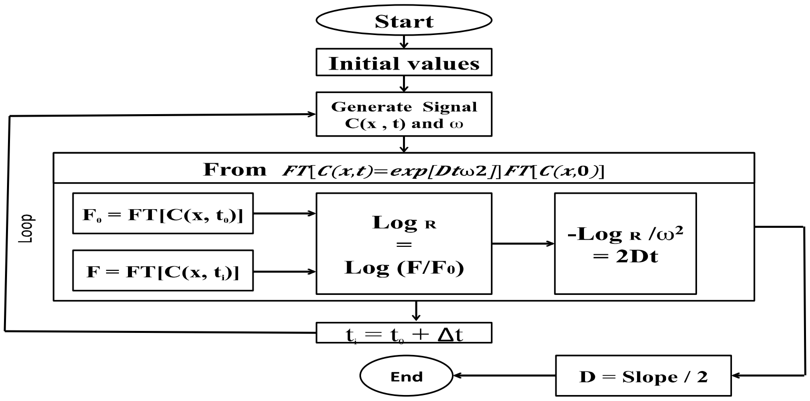


Figure S- 4 Flowchart of the FFT model procedure for determination of diffusion coefficient

To evaluate ω, a scalar vector with same size of signal is built Error: Reference source not found, representing the corresponding spatial divisions in the frequency domain:

Eq. S- 4

Where m is the half-length of the time domain signal, n_f is the number of data points in the signal, and the coefficient, $2\pi/(p\times n_f)$, is used to convert the discrete Fourier transform to a real continuous Fourier transform.

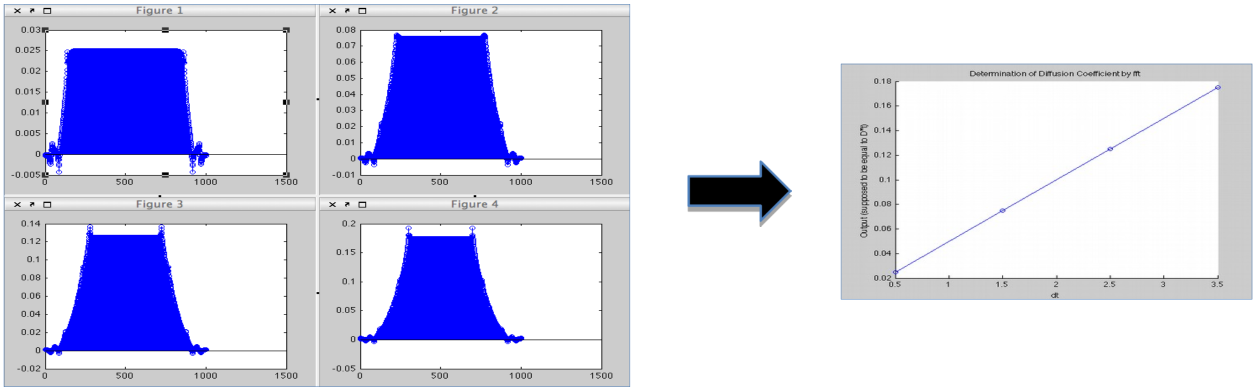


Figure S- 5 Demonstration of diffusion coefficient estimation in the frequency domain.

In theory, vector Q is expected to be constant for all Omega values. However, in practice, working within a finite interval, ±0.5cm in this example, makes vector Q vary at the endpoints. However, in the mid points it is almost constant. Thus, the value of Q is determined in the exact mid point, Q_m, to have the best approximation possible. Here, the curve is plotted for four time points. The value of D is equal to the slope of the linear curve.

2.1 Frequency component of different peak shapes

The frequency domain approach is able to extract the diffusion coefficient from any shape of the initial concentration profiles, while its direct analysis is difficult in non-regular peak shapes. However, the FFT method is not totally shape-independent, as it is affected by varying the frequency responses of different peak shapes. To investigate this characteristic of the FFT model, three mathematical functions were subjected to signal analysis through Fourier transformation.

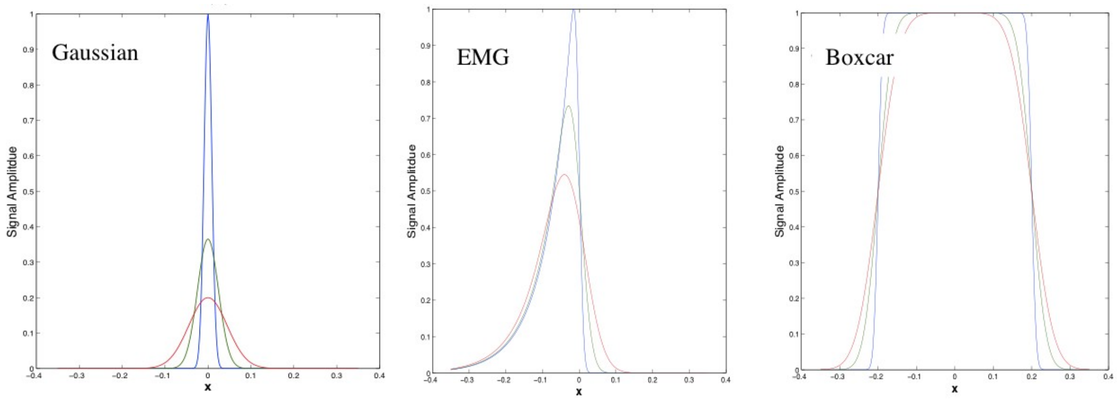


Figure S- 6 Demonstration of diffusion process in time domain via three mathematical functions: Gaussian, Exponentially Modified Gaussian, and Boxcar

In addition to the Gaussian (G) function, Boxcar (BC) and Exponentially Modified Gaussian (EMG) are studied and their corresponding Fourier analysis are explored. Gaussian is the routine assumption of the electrophoretic and chromatographic peak shapes, while the BC and EMG present a more specific description of the distorted signal. For example, EMG defines tailing or fronting peaks, while rectangular plug-like signals are simulated by BC.

2.1.1 Gaussian Model

The concentration profile C is defined as a function of space and time, with the Gaussian distribution:

Eq. S- 5

The Fourier transformation of a Gaussian gives another Gaussian, while broad peaks produce narrow and vice versa. Figure S- 7 demonstrates the diffusion pattern of Gaussian function in the time domain and its counterpart in the frequency domain.

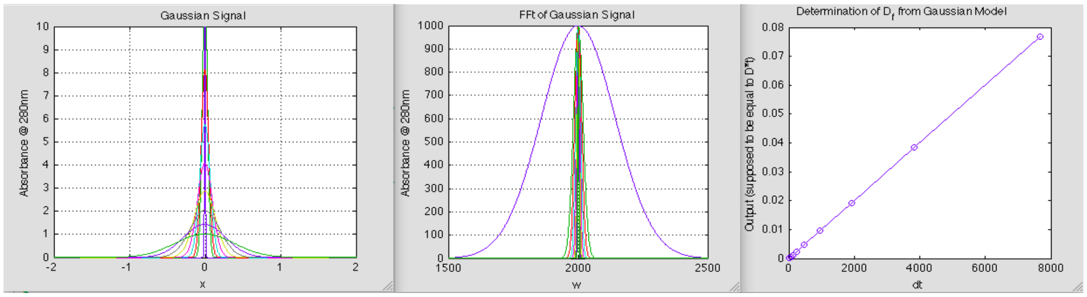


Figure S- 8 The Gaussian function, its Fourier counterpart (shifted absolute values), and determination of D

2.1.2 Exponentially Modified Gaussian (EMG)

The exponentially modified Gaussian function is the most popular model for asymmetric peaks. It is derived via the convolution of two functions; the normal Gaussian and exponential probability density.

Eq. S- 6

The truncated asymmetric signals are simulated using EMG function. Depending on the values of the exponential parameter, τ, which is inversely proportional to λ (explained in the main text), the distribution will vary from almost normal to almost exponential. Keeping all parameters constant, excluding the exponential decay value produces variation in the levels of asymmetry and truncation. In Figure S- 9, different truncated asymmetric signals are demonstrated; varying from slightly deviated from Gaussian to more exponential-like peak shape as the τ value increases.

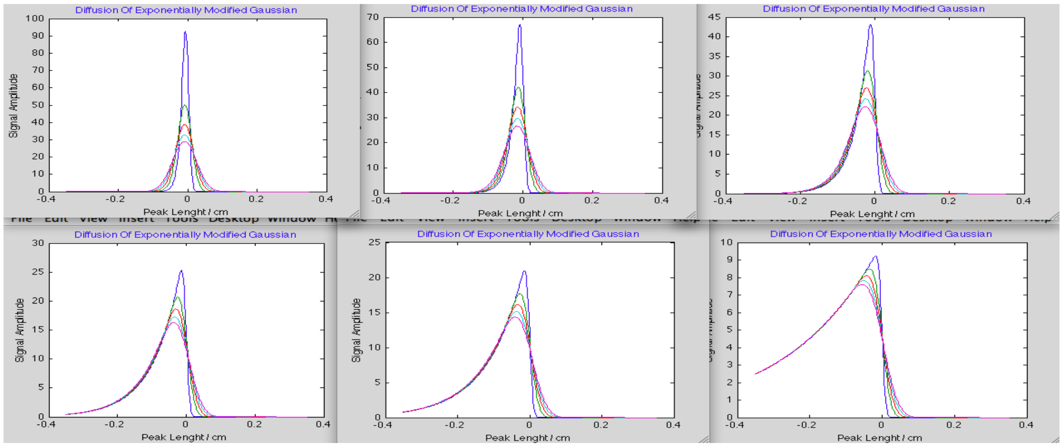


Figure S- 9 Exponentially Modified Gaussian Signals with different levels of truncation, from almost symmetric to severely asymmetric and truncated, τ=[0.01, 0.02, 0.04, 0.08, 0.1, and 0.25]

As it can be clearly understood from the graph and from data in Table S- 2, deviation from symmetric Gaussian function will affect results both in time and frequency domain, although it has been much more destructive in the time domain approach due to the full width at half maximum (FWHM) technique being severely affected by the asymmetrical shape of the EMG function for calculation of peak area.

τ value	0.01	0.02	0.04	0.08	0.1	0.25
Error FD	0.00	2×10^{-4}	1.70	3.88	6.87	24.19
Error TD	10.3	17.49	44.51	>100	Huge	Huge

Table S- 2 Estimation of diffusion coefficient from asymmetric truncated signals by frequency analysis and time domain ($w_{1/2}$)

For normal distribution, the FWHM method provides a good fit, while for BC, EMG, and non-regular shapes, other techniques, such as curve fitting and integration for calculations in the time domain, are required. However, the variance can be calculated from EMG to be $\sigma^{2}+1/\lambda^{\square}$, and this can be used in time domain estimation. On the other hand, estimation of the diffusion coefficient of asymmetric truncated signals in the frequency domain is not significantly affected by the asymmetry, although it would appear that the truncation issue drastically imposes error, an issue that has been attempted to resolve by signal processing methods.

2.1.3 Boxcar Model

The BC Function: An injection method can produce a rectangular plug, which is approximated with a BC function. The sample plug formed by imaging plug flow can be modeled as a boxcar function, described as:

Eq. S- 7

It is zero over the entire real line, except for a single interval where it is equal to the initial concentration (C_0).

where h is a constant, and it is determined experimentally.

The Boxcar function is introduced to model the non-Gaussian shapes, making a more general definition of the diffusing peaks and producing a rectangular signal. A large injection plug produced by pre-concentration injection (iPF) can be approximated as a boxcar.

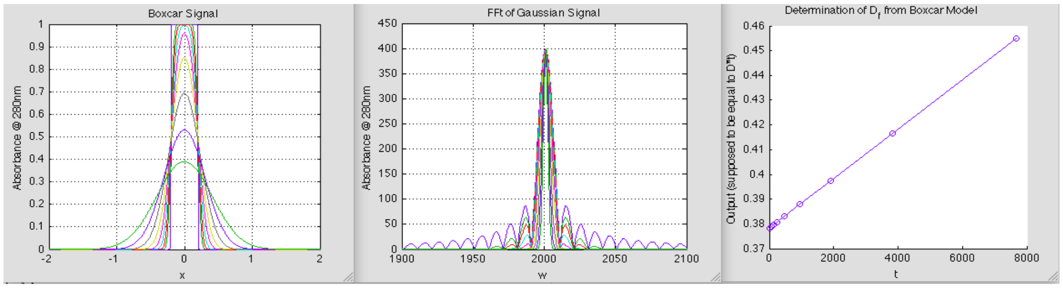


Figure S- 10 The Boxcar function, its Fourier counterpart (shifted absolute values), and determination of D

As expected, optimizing parameters such as sampling length, resolution, and initial time makes estimation more accurate in the frequency domain. For example, narrower rectangular signals (h=0.01-0.25cm) within the ±10σ provided good estimations, while for broader h values; an increase in the sampling region was needed to avoid truncation error.

3 Temporal resolution and sampling duration

In determination of D by capillary electrophoresis method, the time-independent factors include injection, detection, and voltage switching (high voltage "on" and "off"), while time-dependent categories include molecular diffusion, Joule heating, adsorption of species, and local electric field due to excessive sample concentration.

3.1 Random Noise Generation

The random noise generator adds white Gaussian noise to the signal, which has a uniform probability density function in the frequency domain. In the diffusion pattern, the amplitude of the decaying signals is attenuating by time, while the noise energy remains constant. Thus, the SNR value becomes poorer as time passes. The SNR can be obtained by calculating the ratio of signal-to-noise energies, and is expressed in the logarithmic decibel scale:

Eq. S- 8

Where A is the amplitude. In the present study, the A_{signal} is the undistorted average power of the signal at the initial time, which determines constant noise level according to the given SNR. By knowing the desired SNR and the signal power, appropriate noise energy can be calculated from rearranging the equation S-6, and produced with a random number generator. Keeping the noise level constant, shorter and broader peaks at longer diffusion times are subject to lower signal-to-noise ratios.

3.2 Total Time In Presence Of Different Noise Level

The signal-to-noise ratio is expected to be high for both employed methods since the sample is concentrated in a narrow zone. As time passes, the amplitude of the signal attenuates while the noise energy remains constant; as a result, the SNR decreases, and the data points obtained at a longer time become less reliable than those recorded at an earlier time.

Table S- 3 Effect of total analysis time on determination of D, at 10, 30, and 60 min total time, (D=1×10^{-6} cm^2=s, t_0 = 30 s, t_i = 30 s, p=0.001 (cm/pix), SNR=15, n=1000, PLCC=Pearson Linear Correlation Coefficient.)

Total Time	10min	30min	60min
D_{est} ($\times 10^6$)	0.973	1.02	1.11
CI_{95} ($\times 10^6$)	0.007	0.011	0.016
Error%	2.61	2.58	11.45
PLCC	0.996	0.995	0.978

This effect is demonstrated on Figure S- 11, where the diffusion coefficient is estimated varying the total time at three different noise levels. Although the line of the best fit is only plotted for the longest time period, the diffusion coefficients on right panels are estimated from the local slope of the linear curve at each individual total time, and the value is reported within a 95% confidence interval. Comparing panels A to C, it can be concluded that at a poor signal to noise ratio (panel A), the estimated D is only acceptable within short measurement times, while the lower noisy signal (panel C) shows the best estimation around half an hour.

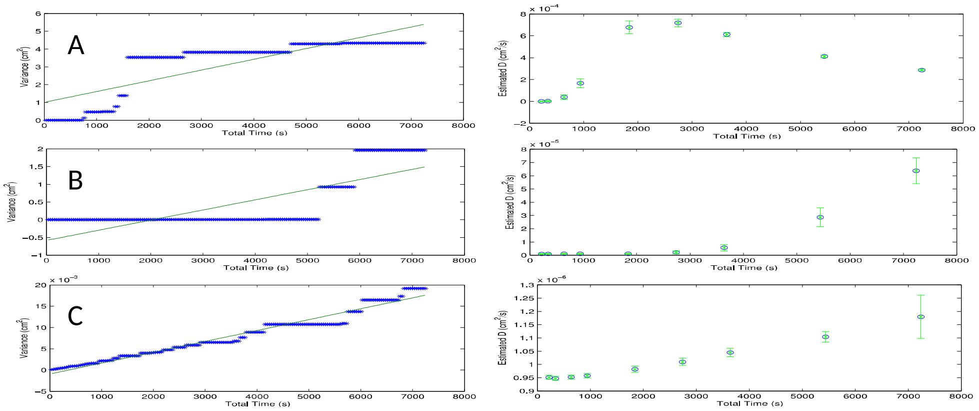


Figure S- 11 Optimization of total time according to the noise leve (D=10^{-6}(cm^2/s), t_0=30s, p=0.001 (cm/pix), n=1000, SNR values from A to C; 3, 10,and 15 respectively)

Effective analysis time is determined on the basis of the weight of the imposed error from the collected data points. The longer the analysis time is, the poorer the SNR, which leads to less reliable data. Therefore, the non-linearity increased by taking the long noisy signals into account. At higher noise levels this effect is more significant.

3.3 Optimized Data Acquisition Rate

The one hour measurements at fast scans showed more error due to a larger number of unreliable data sets, which were collected at poor SNR values, while shorter period measurements revealed better estimation at fast successive scans.

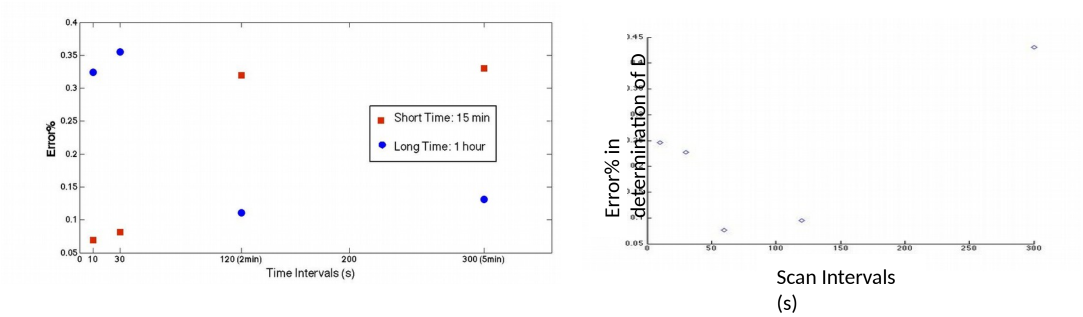


Figure S- 12 Effect of data acquisition rate on determination of diffusion coefficient (D=10^{-6}(cm^2/s), t$_0$=30s, p=0.001 (cm/pix), n=1000) at variable intervals; A) Total time 15, and 60 min, (SNR=3). B) Optimum sampling rate at 30-minute diffusing time periods (SNR=15).

The length of recording the diffusion process for the simulation condition was determined to be 30 minutes according to the optimization section in the main text. Investigation of the scan interval at this total time is demonstrated in Figure S- 12, which revealed that faster acquisition rates do not significantly improve results; however, a 1 min interval provides the lowest error for the implemented conditions.

The iCE280 instrument is capable of scanning at a 30sec rate, which fulfills the sampling rate needed for estimation of the diffusion coefficient under the employed conditions.

3.4 Protein Mixture Analysis in Short Time

Experiments on β-lactoglobulin A and B and also myoglobin isoforms represent the capability of CIEF-WCID for protein mixture analysis. In case of β-lactoglobulin mixture, two peaks are resolved by a small pI difference and combine shortly after diffusion commenced, providing a short period to record the diffusion process for each individual isoform. The two peaks were so close that they convoluted after 10 minutes, so keeping a record of the combined peaks for longer periods of time underestimated the D values.

Table S- 4 Diffusion coefficients of β-lactoglobulin A/B at different analysis periods.

$D_{Freq.}$(×10^7 s/cm^2)	5 min	10 min	25 min	$D_{Lit.}$	Ref.
β-lactoglobulin A	6.87	7.18	6.23	7.38	[3]
β-lactoglobulin B	4.82	3.52	2.13	3.14	[4]

The Diffusion path was monitored for 25 minutes, and total analysis time required for estimation of the D of these two proteins was determined to be 10 minutes, with accuracy of 2.71% and 12.10% error in comparison to the literature values. Precision in the determination of β-lactoglobulin A and B variants' diffusion coefficients at a 10 minute sampling time length was defined by relative standard deviations of 4.88% and 11.47% respectively.

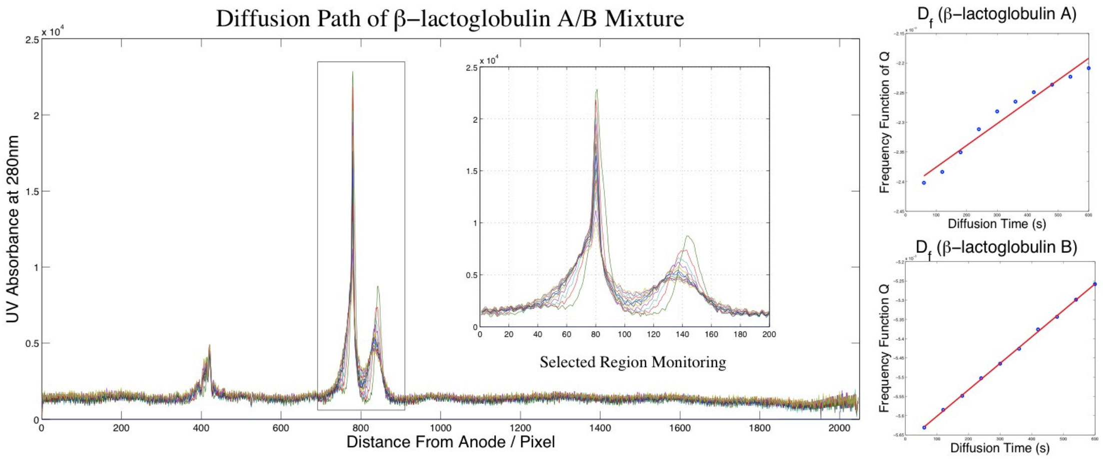


Figure S- 13 Determination of diffusion coefficients of β-lactoglobulin A/B variants from bovine milk; using FFT approach, experimental conditions the same as CIEF method explained in the methodology section of the main text.

The quantitative results are given in Table S- 4 and in the above figure the electropherogram of the diffusion pattern of β-lactoglobulin mixture is demonstrated.

4 Experimental Section Supplementary Figures

In this section, the iPF approach is graphically illustrated and the electropherograms of some CIEF-analyzed compounds are demonstrated with an emphasis on variety of the peak shapes.

4.1 Imaging Plug Flow Procedure

The sample is eletrokinetically injected and accumulated under the membrane. After the stacking step, the generated sample plug is pushed to the channel by applying the electric field between two reservoirs of the commercial cartridge. The whole-column imaging detector monitors the sample position; when it reaches midway through the channel, the voltage is turned off and the diffusing plug is scanned frequently.

The sample must dissolve in a buffer with lower concentration and ionic strength compared to the run buffer, to make stacking works. In addition the pH of the buffer is set lower than the pI of the protein to make the analytes positively charged to move toward cathodic end. In this study the phosphate buffer with pH= 2.4 is used for myoglobin (pI=6.88, 7.33), and BSA (pI= 4.29).

The applied voltage amplitude and duration depends on the analyte and the buffer system. According to the velocity of the analyte and the ionic strength of the media.

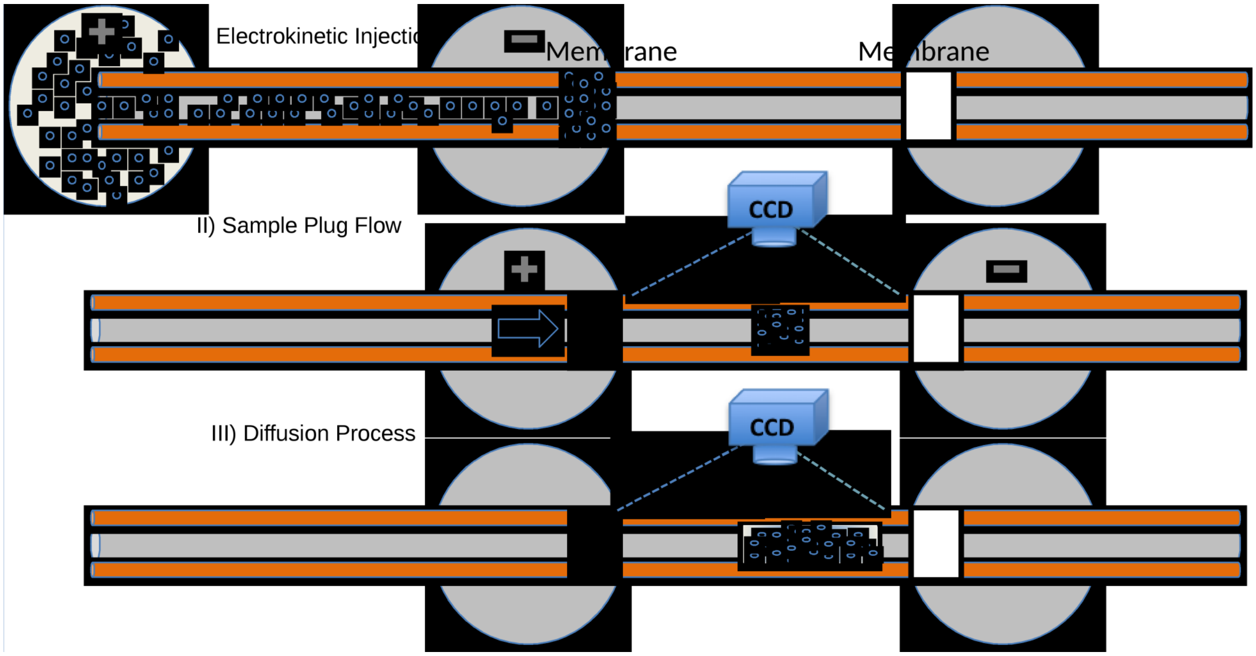
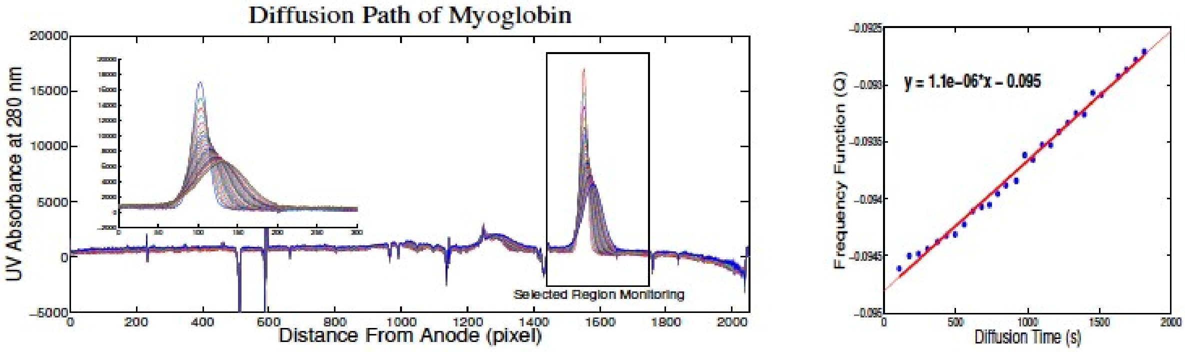


Figure S- 14 Schematic of iPF Method (I) Injection: protein is eletrokinetically loaded and accumulated under the membrane. (II) Plug Flow: conventional capillary electrophoresis operation under electric field. (III) Imaging Diffusion Process: CCD camera scans the diffusing sample plug in absence of the electric field.

It is of great importance to optimize the stacking and flow stages, since the way samples are introduced may affect obtained results. The stacking time and voltage were optimized according to the generated current in order to acquire the sample plug within a short time and without precipitation. A pre-concentration time of 1 min was sufficient to stack the sample efficiently. Moreover, 0.5kV was adjusted for the moving step, which was disconnected after the plug reached the middle of the channel. In the iPF model, total analysis time is considerably shorter by skipping the focusing step and replacing it by the stacking and plug flow steps.

4.2 Fairly Gaussian and Distorted Gaussian Peaks

Although concentration profiles in CIEF are mostly well defined by the Gaussian function due to the narrow focused band, employing the FFT analysis improves the results considerably. The diffusion coefficient of myoglobin and carbonic anhydrase I has been greatly improved in the frequency domain. The diffusion path of the second isoform of myoglobin is recorded for 3 minutes, and it nearly follows the Gaussian assumption.

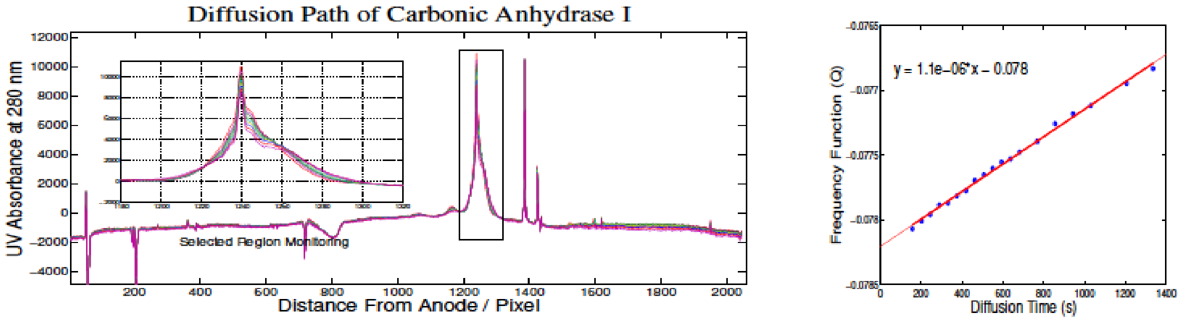


(a) Myoglobin diffusion path

(b) Mb D_f

Figure S- 15 myoglobin diffusion pattern represents fairly Gaussian signal; experimental conditions the same as CIEF method explained in the methodology section of the main text.

Theoretically, the broadening peaks on each trial should be centralized, but experimentally, there are slight influences from the hydrodynamic (open capillaries), electro osmotic flow, and sample wall adsorption to be considered. However, this shift in peak positions does not affect the calculations in either the time or the frequency domains. In the frequency domain calculations, the real part is only considered by taking the absolute value; hence, the phase variation of the frequency component is ignored.



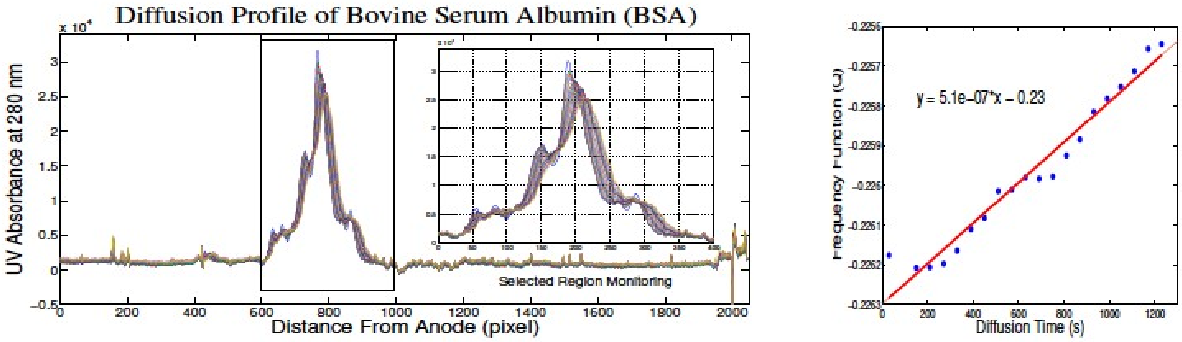
(a) Carbonic Anhydrase I diffusion path (b) CA I D_f

Figure S- 16 Diffusion pattern of carbonic anhydrase I represents asymmetric Gaussian; experimental conditions the same as CIEF method explained in the methodology section of the main text.

Carbonic anhydrase I is an example of an asymmetric Gaussian distribution that imposed error in the full width at half maximum method, but that can still be estimated with time domain calculations.

4.3 Non-Gaussian and Irregular Peaks

The CIEF is also capable of monoclonal antibody charge heterogeneity determination, which provides a jagged concentration profile, making the estimation of D more challenging via the conventional time domain approach. In cases such as the one of albumin, the peak shapes cannot be mathematically defined, and the calculations from curve fitting or the full width at half maximum are not applicable any more. In the following figure, the significant deviation of the diffusing peaks of albumin from the Gaussian peak assumption is graphically demonstrated.



(a) Albumin diffusion path (b) Albumin D_f

Figure S- 17 Diffusion pattern of albumin as a representative for ill-shaped signals; experimental conditions the same as CIEF method explained in the methodology section of the main text.

As albumin analysis consists of multiple jagged peaks, limitations due to Gaussian-shape assumptions for calculations in the time domain leads to errors in the determination of its diffusion coefficient. The proposed method of Fourier transformation gives acceptable estimation in good agreement with the literature value, while the full width at half maximum approach does not converge.

5 Stability Study On The Basis Of Diffusion Coefficient

Proteins selected for research purposes are usually handled in powder form, and as long as they are kept in the fridge under recommended storage conditions, they have a long lifespan. However, when it comes to the liquid state, they need to be prepared and implemented fresh.

The stability assessment of β-lactoglobulin is presented as an example of the aggregation and shelf life of the protein. It has been already studied vastly with different techniques and Majhi et al. proved that measurement at isoelectric point does not necessarily lead to precipitation [2]. Hence the capillary isoelectric focusing is an appropriate method for this purpose.

The changes to the electropherogram of β-lactoglobulin are graphically shown in Figure S- 18, and the diffusion coefficient of fresh and one-month old samples were measured and compared to the literature values, presented in Table S- 5. There is a shift in the pI values of both variants, and extra peaks show up in the electropherogram. The protein is fairly stable at room temperature within the first four days, however, it deforms to smelly gel after three weeks. The observed charge heterogeneity and shift in the peak position can provide information on the aggregate mechanism, in addition to estimation of degradation or aggregation from changes in diffusion coefficient values.

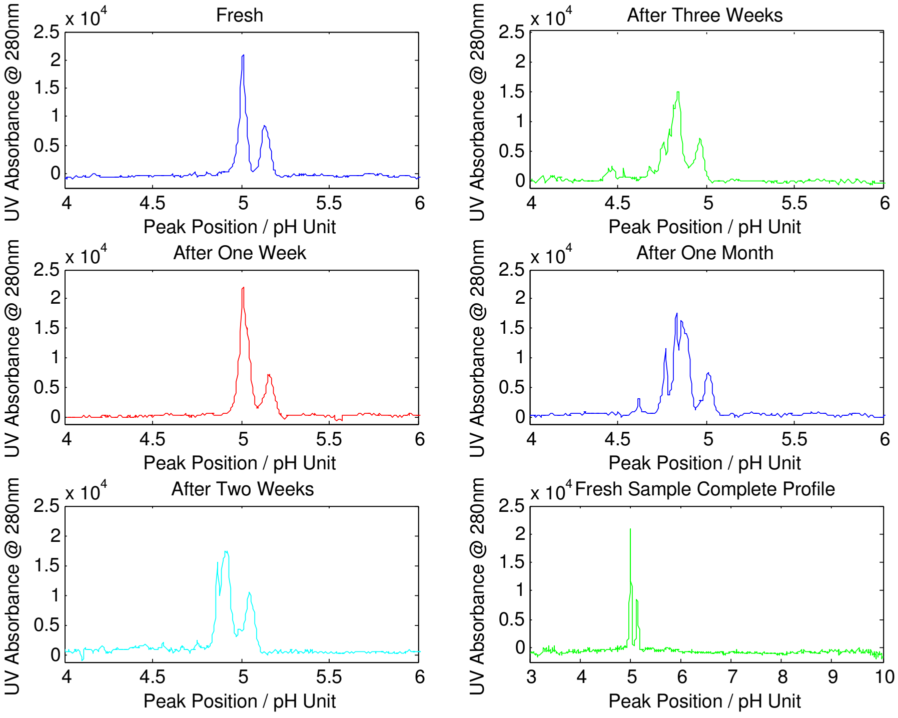


Figure S- 18 Stability assessment of β-lactoglobulin variants by CIEF-WCID during one month; experimental conditions the same as CIEF method explained in the methodology section of the main text.

The fresh sample's diffusion coefficients measured by the Fourier domain approach are in good agreement with reference values in comparison to the old samples Ds, revealing that both variants has changed during time, and isomer A is more significantly altered. Further study of the aggregation can provide valuable information on the association/dissociation mechanisms.

β-lactoglobulin (Fresh)		β-lactoglobulin (Old)		β-lactoglobulin (Literature)	
A	B	A	B	A[3]	B[4]
7.18 (±0.35)	3.52 (±0.40)	0.31 (±0.08)	1.23 (±0.13)	3.14	7.38

Table S- 5 Comparison of the diffusion coefficient (× 10^7) of the fresh and one month old β-lactoglobulin variants (n=3) with literature values; experimental conditions the same as CIEF method explained in the methodology section of the main text.

Generally due to D alteration one can decide on either aggregation or dissociation. For example the increase in D value can be interpreted as a result of the dissociation that caused the protein falls apart. In this study the old sample D value decreased that proves formation of the aggregate or particulate.

Estimation of the impurity level by resolving the diffusion profile of oligomer mixture is a useful tool from pharmaceutical viewpoint. However, the diffusive behavior of the mixture components needs to differ significantly, typically by factor two, in order to discriminate efficiently, e.g. study of monomer-dimer or higher oligomeric states. The original pure sample solution protein is found in the monomeric form; if it partially aggregates to dimer, trimer, and higher oligomeric states, then a change in the diffusion coefficient value of the particulates happens due to the molecular weight increase. The final diffusion coefficient profile of the mixture is overlapped by profiles of the oligomers. The impurity level from the diffusion profile of a mixture of different oligomeric states of the same protein can be estimated, and the contribution coefficient can be determined accordingly.

6 Reference

[1] Smith, M. H. In Handbook of Biochemistry, 2nd ed.; Sober, H. A., Ed.; Chemical Rubber Company: Cleveland, OH, 1970.

[2] Majhi, P.R.; Ganta, R.R.; Vanam, R.P.; Seyrek, E.; Giger, K.; Dubin, P.L. Langmuir **2006**, 22, 9150–59.

[3] Chu, B.; Yeh, A.; Chen, F. C.; Weiner, B. Biopolymers **1975**, 14, 93–109.

[4] Chen, B.; Andreas, C.; David, R. Anal. Biochem. **1979**, 130, 120–130.