

Supplementary Information - Label-Free Mass and Size Characterization of Few-kDa Biomolecules by Hierarchical Vision Transformer Augmented Nanofluidic Scattering Microscopy

Henrik K. Moberg, Bohdan Yeroshenko, Joachim Fritzsche,
David Albinsson, Barbora Spackova,
Daniel Midtvedt, Giovanni Volpe & Christoph Langhammer

May 23, 2025

1 Molecular Property Characterization

Determining the MW of a biomolecule from its trajectory inside a nanofluidic channel is possible because the integrated optical contrast (iOC) of said trajectory in the kymograph is linearly dependent on the polarizability α_m of the biomolecule [1], where iOC is defined as

$$iOC = \int_{x=0}^L (I_t(x) - I_c(x))/I_c(x) dx \quad (1)$$

for a channel of length L . The polarizability α_m is in turn proportional to $a \times MW$, where $a \approx 0.46 \text{ \AA}^3/\text{Da}$ [2]. Furthermore, the iOC is inversely proportional to the cross-sectional area A of the nanochannel and proportional to $\bar{n} = (1.5n_{H_2O}^2 + 0.5n_{SiO_2}^2) / (n_{H_2O}^2 - n_{SiO_2}^2)$ [3]. The former means that nanochannels with smaller cross-section boost the iOC and is the reason why we use smaller channels for our Insulin experiments, and why BSA becomes distinctly visible in channel A_{II} (**c.f. Figure ??E**). Since both these factors are constant and easily measured before an experiment, the MW of a molecule measured by NSM can be determined as

$$MW = (iOC \cdot A)/(\bar{n}a). \quad (2)$$

Determining the R_s of a biomolecule from its trajectory is possible because its diffusivity D is correlated with its size. In particular, by approximating the molecule as a solid neutrally charged sphere, its R_s can be estimated using the Stokes-Einstein equation with a correction term K for hindrance effects arising from the diffusion of small objects in restricted volumes [4], as we have demonstrated in our seminal

work[3]. R_s is estimated as

$$R_s = Kk_BT/(6\pi\eta D), \quad (3)$$

where k_B is Boltzmann’s constant, T is temperature, and η is the viscosity of the liquid in the nanochannel.

2 Deep Learning Architecture

Neural networks, including deep learning architectures such as the Hierarchical Vision Transformer (h-ViT) employed in this work, are fundamentally piecewise linear function approximators capable of representing any function through the partitioning of hyperplanes in the input space during training. The depth of a neural network determines the complexity of these partitions, thereby enhancing the representational power of the architecture. In the case of h-ViT for NSM, subsequent layers progressively refine the latent representation of the input data, allowing for the identification of hyperplanar partitions that best approximate molecular weight and hydrodynamic radius of biomolecules inside nanofluidic channels.

A key advantage of deep learning, and specifically of the h-ViT model, is its ability to capture long-range dependencies in microscopy data through attention mechanisms and hierarchical feature extraction. Unlike conventional CNNs, which rely on local receptive fields and pooling layers to reduce dimensionality, h-ViT processes nanofluidic scattering kymographs using a transformer-based multi-scale approach. Each transformer block integrates multi-head self-attention, enabling the model to selectively focus on different regions of the input and thus improve the estimation of diffusivity and integrated optical contrast. This allows the h-ViT model to learn more physically relevant, generalizable, and interpretable features from microscopy images.

The limitations of traditional convolutional neural networks arise from their inductive biases, particularly their focus on spatial locality and translational invariance. While this is useful in conventional imaging tasks, microscopy data often requires global context awareness, which h-ViT achieves by replacing local receptive fields with attention-based mechanisms. The model encodes input kymographs using a custom PatchEncoder layer, which extracts feature-rich patches at multiple scales. These patches are then processed by transformer layers that retain spatial relationships across entire kymographs, allowing for superior feature extraction.

Traditional analytical approaches for microscopy often rely on explicit noise modeling and handcrafted filters, which limit their effectiveness in complex, non-stationary noise environments. In contrast, deep learning approaches like h-ViT leverage spectral bias to extract relevant low-frequency information while maintaining sensitivity to high-frequency molecular motion. This is particularly advantageous in denoising diffraction-limited images, where low-frequency trends hidden within high-frequency noise patterns can be effectively recovered.

In Nanofluidic Scattering Microscopy, where single molecules diffuse inside nanofluidic channels, standard CNN-based methods fail to fully capture the underlying molecular properties due to low signal-to-noise ratios (SNR) and the inherently stochastic nature of molecular motion. h-ViT addresses these challenges by employing probability maps that quantify the likelihood of molecular presence in different regions of the kymograph. These probability maps serve as intermediate feature representations that guide the final prediction of MW and Rs.

The model refines its predictions through a weighted averaging mechanism, ensuring that high-confidence regions contribute more significantly to the final MW and Rs estimation. This probabilistic approach allows h-ViT to outperform classical CNN architectures in low-SNR microscopy settings, making it an essential tool for molecular characterization at the single-molecule level.

In applications such as label-free single-molecule microscopy, the h-ViT model has significantly improved the lower detection limits (LoD) of NSM, achieving molecular weight resolutions down to 6 kDa and hydrodynamic radius resolutions below 1.5 nm. This is a major advancement over conventional analytical approaches, which previously struggled to characterize molecules below 60 kDa and 4 nm. The improved LoD is attributed to a combination of ultrasmall nanochannel fabrication (29 nm \times 60 nm cross-section) and the deep-learning-driven signal extraction capabilities of h-ViT.

Additionally, h-ViT enables real-time analysis of diffusing biomolecules without requiring prior surface immobilization, overcoming a major limitation of alternative methods such as interferometric scattering microscopy (iSCAT). The combination of hierarchical attention mechanisms and transformer-based feature learning allows h-ViT to robustly estimate MW and Rs even when traditional tracking algorithms fail due to low SNR.

The h-ViT model represents a fundamental shift in deep-learning-based microscopy, replacing conventional CNN-based approaches with hierarchical, transformer-based architectures that preserve global spatial relationships and efficiently extract biomolecular features. The integration of attention-driven, multi-scale processing within NSM enables significantly enhanced molecular characterization, setting a new benchmark for deep-learning-powered optical microscopy.

2.1 Hierarchical Vision Transformer

The deep learning architecture presented in this study, the Hierarchical Vision Transformer (h-ViT), enhances the characterization of biomolecules using NSM. It builds on recent advancements in attention mechanisms and transformer-based architectures to capture the multi-scale nature of kymograph data effectively. The input kymograph, a two-dimensional representation of scattered light intensity over time, undergoes preprocessing through temporal standard deviation normalization to ensure consistency in subsequent feature extraction.

The kymograph is first processed through a convolutional layer with a kernel size of 7

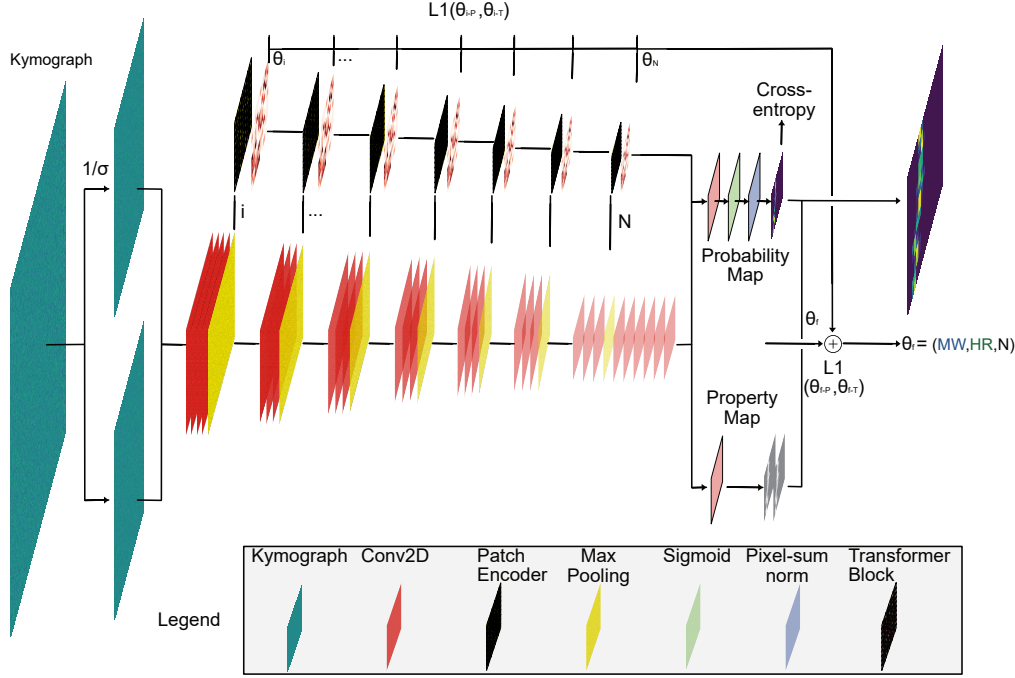


Figure 1: **Hierarchical Vision Transformer (h-ViT) architecture for single biomolecule characterization by NSM.** The h-ViT model processes NSM kymographs to predict the MW and R_s of single particles, such as biomolecules. Initially, Conv2D layers extract spatial features, followed by max pooling for downsampling. The Patch Encoder then encodes these feature maps into smaller patches, while retaining crucial spatial information. Each patch is passed through Transformer Blocks, where multi-head attention mechanisms selectively focus on different regions of the input, capturing long-range dependencies in the kymograph. The outputs of the Transformer Blocks are processed to generate two key outputs: a Probability Map (indicating the likelihood of particle trajectories at different locations) and a Property Map (predicting specific biomolecule properties, such as MW and R_s). The model utilizes cross-entropy loss for classification tasks and L1 loss to minimize the error between predicted and true molecular properties. The estimation of each parameter (θ_i) representing MW and R_s , and number of trajectory points, N , is refined through iterative back-propagation. Finally, the model weighs the Property Map with the Probability Map to output a final estimation of the biomolecule properties, weighted by the probability of said biomolecule being present in any given region.

and leaky ReLU activation to preserve subtle variations in the data. This is followed by a series of convolutional blocks that progressively downsample the data while capturing features at different spatial and temporal scales. The architecture employs a hierarchical multi-scale approach in which max-pooling operations reduce spatial dimensions at each stage. Encoded patches from feature maps are processed through transformer blocks that integrate multi-head attention mechanisms, allowing the model to selectively focus on different regions of the input. These transformer layers enhance the ability to capture long-range dependencies and refine molecular property estimations, such as diffusivity and iOC.

Instead of reconstructing explicit pixel-wise trajectories, h-ViT employs probability maps that encode the likelihood of particle presence at specific positions and times. This allows the model to aggregate information across the entire kymograph rather than being constrained by localized trajectory visibility, which is often compromised due to low signal-to-noise ratios. Multi-head attention mechanisms dynamically emphasize regions that correlate with iOC (and thus molecular weight, MW) and diffusivity (and thus hydrodynamic radius, R_s). This formulation ensures that the model extracts global patterns indicative of particle motion while mitigating noise amplification.

A key advantage of h-ViT is its mitigation of spectral bias, wherein deep learning models tend to prioritize low-frequency over high-frequency signals. In NSM kymographs, particle trajectories constitute low-frequency signals embedded within high-frequency noise. The hierarchical structure of h-ViT forces the model to extract relevant trends at different downsampling scales, improving sensitivity to weak scattering signals from small particles. Unlike conventional convolutional approaches, which often overfit to noisy pixel-wise variations, h-ViT prioritizes broader contextual features, thereby improving detection limits.

The model generates predictions for particle properties at multiple downsampling scales, which are processed through dense layers and probability maps to quantify the likelihood of particle presence. These probability maps are resized to match the original kymograph dimensions, ensuring spatial alignment with input data. The outputs from different scales are then concatenated to form a minimally dimensioned representation, condensing all relevant particle information across scales. The final predictions of MW, R_s , and the number of trajectory points (N) are computed using a weighted averaging mechanism, where higher-probability regions contribute more significantly to the output.

This hierarchical and attention-driven design marks a departure from traditional convolutional neural networks by emphasizing global feature extraction rather than local trajectory reconstruction. The ability to infer molecular properties from probability-weighted feature maps, rather than explicit trajectory tracing, significantly improves robustness in low-SNR conditions. As demonstrated in this study, the h-ViT model substantially enhances the sensitivity and accuracy of NSM-based biomolecule characterization, setting a new benchmark for deep-learning-driven optical microscopy.

3 Deep Learning Training

In this section, we delve deeper into the specific training methods which are vital to reaching the high accuracy and precision of prediction (and thus low NSM LoD) of the trained models used. In particular, we describe transfer learning; the process of transferring learnt knowledge from one network to another, and curriculum learning; the process of successively increasing the difficulty of the task during training to be learnt to facilitate learning in very challenging conditions.

3.1 Transfer Learning

Transfer learning is a technique that involves transferring knowledge gained while solving one task to another (similar but not identical) task. The underlying idea is that some of the knowledge and skills learned while solving the first problem can be applied directly to the second problem, resulting in improved performance and shorter training time. It has been shown that a neural network trained on a particular type of dataset is much better at quickly recognizing new instances of similar data, when compared to a completely untrained network [5]. For instance, though a neural network may need thousands of examples of cars to start to accurately recognize cars, such a network can learn much quicker to recognize other vehicles when introduced into its training set compared to an untrained network [6]. This is a consequence of neural networks learning increasingly complex abstractions in deeper layers of the network. Thus, a majority of the learning a network has to do to solve any given task is simply to find a good representation of the data, and transfer learning can thus be achieved by "freezing" the training of the first layers of a neural network and only training the last few layers on the new data.

There are several benefits of this approach, most evidently that it can help reduce training time and data requirements for new tasks. Another benefit is that it can improve generalization performance, meaning that models trained using transfer learning are more likely to achieve good results on unseen data than those trained from scratch specifically for the new task at hand. Additionally, because transfer learning typically relies on pre-trained models which have been already tuned for good performance on a wide range of tasks, it can help avoid overfitting problems which can plague neural networks when they are retrained from scratch for a specific purpose.

Despite its many advantages, there are also some potential drawbacks to consider with transfer learning strategies. One issue may be poor adaptation if there are significant differences between how tasks were originally learned and how they need to be applied later on, and special care should be taken when applying pre-trained models so as to not inadvertently introduce bias into results due to inaccuracies or irregularities in how the old network was transfer-learned into the new.

3.2 Curriculum Learning

Human beings often do not learn anything when approached with an insurmountable challenge, learn only the minimally required information when approached with a trivial problem, and learn optimally only when approached with a problem perfectly tuned to be slightly challenging given their particular skillset [7]. Neural networks, interestingly, have been shown to learn in analogous ways [8]. To take advantage of this, we can implement active learning [9] or curriculum learning [10] schemes, wherein we vary which portions of the datasets are fed into the neural network during training to tune the challenges of learning, for instance by automatically feeding in more examples of data-points in regimes where the network retains low accuracy. Neural networks need to train on a large amount of data only to learn optimal representations of the dataset in its initial layers, but once these representations are learnt, transfer-learning it on new data is usually only a question of tuning the specifics of the last few layers’ learnt abstractions. To take advantage of this, we might employ a curriculum learning scheme wherein we train the networks with the easiest examples or the ones we have the most data-points of first, and successively increase the difficulty of the dataset as the network’s loss decreases.

Projected onto the challenges at hand in the present work, to train the molecular weight- and hydrodynamic radius-calculating h-ViT model, a curriculum learning scheme with intermittent checkpoints to be used for later ensemble modelling prediction was employed. Specifically, the model was initially trained only on a narrow range of high MW trajectories, representing the highest SNR and in principle easiest case for the model to begin learning correlations, and then slowly curriculum-learned down to the lowest range of relevant MW values. The process is equivalent for diffusivity, with the difference being that the range of D values being trained on increases rather than decreases during curriculum learning.

To ensure optimal training progression, the molecular weight (MW) was initially sampled from the range 50-200 kDa and progressively reduced in steps of 5% down to a final range of 0-30 kDa. This allowed the model to first learn from larger, well-defined particles before adapting to the more challenging detection of smaller biomolecules with lower signal contrast. Similarly, the diffusivity (D) was initially sampled from 10-50 $\mu m^2/s$ and progressively increased in steps of 5% up to 50-300 $\mu m^2/s$. This gradual expansion of diffusivity ensured that the model first learned from slower-moving particles, where trajectory tracking is more reliable, before adapting to highly diffusive molecules, which require more robust inference techniques. The values are updated after the loss has converged, with a patience of 100 epochs. This curriculum learning approach systematically expanded the model’s training distribution, improving its ability to generalize across a broad range of biomolecular properties.

3.3 Training Configuration and Hyperparameters

The h-ViT model was trained using the Adam optimizer with an initial learning rate of 10^{-4} , employing an AMSGrad variant to improve convergence stability. A batch

size of 8 was used, ensuring sufficient generalization while maintaining computational efficiency. Weight decay was set to 5×10^{-2} to prevent overfitting, and gradient clipping was applied at 1.0 to stabilize training. The transformer layers incorporated a dropout rate of 0.1 to mitigate over-reliance on specific activation paths. Early stopping was applied with a patience of 100 epochs based on the total loss.

The loss function for training h-ViT consisted of multiple components reflecting the hierarchical nature of the architecture. At each of the seven hierarchical scales, the model predicted molecular weight (MW), hydrodynamic radius (R_s), and the number of trajectory points (N). The loss for these predictions was computed as the mean absolute error (MAE), ensuring that deviations from true values were minimized in a robust manner. The total loss across all scales was formulated as:

$$\mathcal{L}_{\text{hierarchical}} = \sum_{s=1}^7 (\text{MAE}(MW_s^{\text{pred}}, MW_s^{\text{true}}) + \text{MAE}(R_{s,s}^{\text{pred}}, R_{s,s}^{\text{true}}) + \text{MAE}(N_s^{\text{pred}}, N_s^{\text{true}})) . \quad (4)$$

In addition to hierarchical loss, the model also predicted a probability map that quantifies the likelihood of particle presence at specific spatiotemporal positions. This probability map was supervised using a binary cross-entropy loss, ensuring that the predicted spatial localization of biomolecules aligned with their true positions:

$$\mathcal{L}_{\text{mask}} = - \sum_{i,j} (y_{i,j} \log \hat{y}_{i,j} + (1 - y_{i,j}) \log(1 - \hat{y}_{i,j})) , \quad (5)$$

where $y_{i,j}$ represents the ground truth probability of a particle at position (i, j) in the kymograph, and $\hat{y}_{i,j}$ is the model’s predicted probability.

The final loss function combined the hierarchical loss with the probability map supervision:

$$\mathcal{L} = \mathcal{L}_{\text{hierarchical}} + \lambda_{\text{mask}} \mathcal{L}_{\text{mask}} , \quad (6)$$

where $\lambda_{\text{mask}} = 10$ was used to ensure adequate weight for the probability map supervision in relation to the hierarchical property estimation.

Following training, the h-ViT model was validated on an independent dataset of experimental kymographs. The model demonstrated high accuracy, achieving molecular weight predictions within 5% of the ground truth for $MW < 10$ kDa for higher trajectory lengths. Most notably, the model successfully reached molecular weight detection limits down to 6 kDa, approaching the theoretical lower bound for NSM-based detection as defined by the CRLB.

4 Generative Adversarial Network for Trajectory Generation

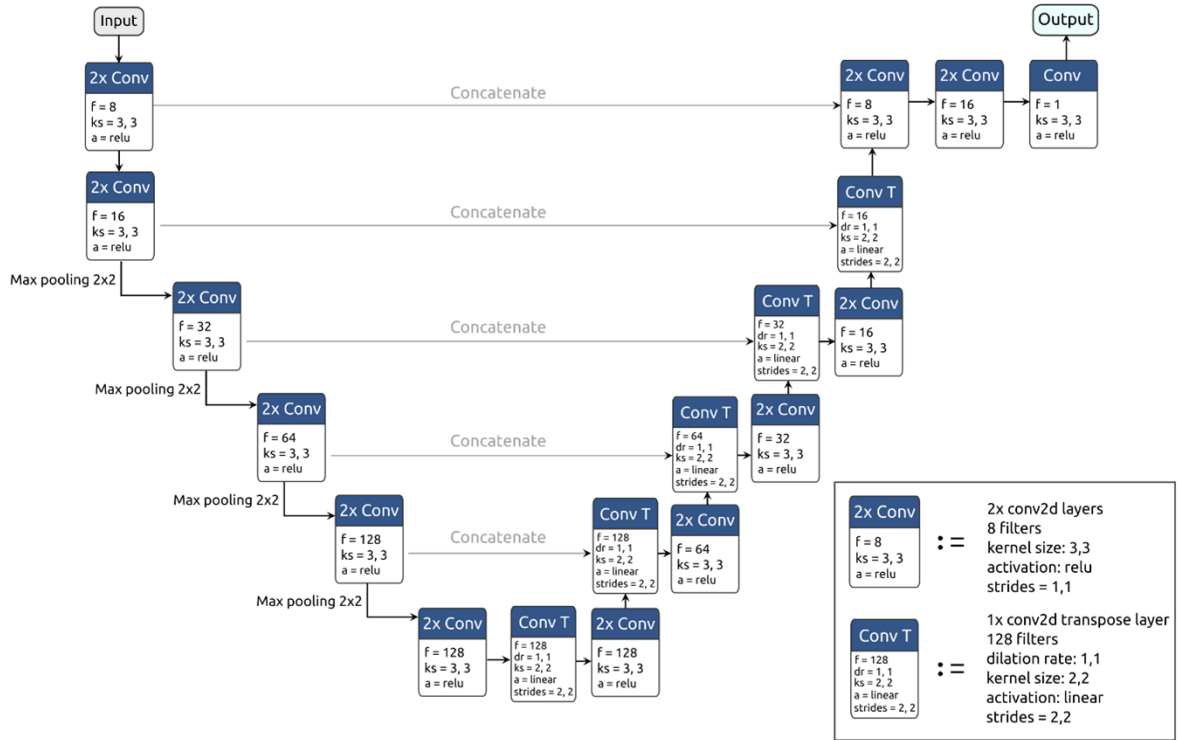
Image segmentation is used to make the value of each pixel of a given image more representative of a property of interest. In our case, we transform a kymograph where the value of each pixel represents intensity, to a segmented image where the value of each pixel represents the probability P_b of the existence of a biomolecule in that location and time. This transformation is achieved using a U-net, through the DeepTrack 2.0 Python software package [11]. The U-net’s structure is illustrated in Figure 2. The U-net’s training was integrated into a conditional Generative Adversarial Network (GAN). This GAN is composed of two neural networks (depicted in Fig. S7B): the generator network (our U-net), which produces image segmentations from input kymographs, and a discriminator network, which tries to discern if the generated segmentations truly represent the trajectories of individual biomolecules in the input kymographs. Throughout the training phase, these networks challenge each other, leading the generator to produce increasingly authentic trajectories, while the discriminator refines its ability to differentiate between real and generated trajectories. Utilizing this GAN for training was crucial to attain the necessary precision, particularly for tinier biomolecules, where the weak signal-to-noise ratio can cause their trajectories to be completely masked by noise across consecutive time frames.

The U-net’s training was based on simulated kymographs (created and pre-processed as outlined in ‘Methods’), for which the exact trajectory of a single biomolecule is established. The training employed the ADAM optimizer [12] with a learning rate that decays exponentially at 109, a decay rate set at 0.9, and spanning 50 decay steps. The U-net was trained on 300,000 simulated kymographs with particle trajectories within the bounds $1e - 9 \mu\text{m} \leq iOC \leq 3e - 3 \mu\text{m}$ $1 \leq D \leq 100 \mu\text{m}^2/\text{s}$. The training data comprised simulated images (kymographs), containing a variable number of trajectories, and corresponding segmented images of the same size. The model underwent validation every 120 epochs (equivalent to approximately 30,000 simulated kymographs) against 150 simulated kymographs that incorporated experimentally observed channel noise, following an 80-20 train-validation distribution.

5 Simulated dataset

The data used to generate figures 3 and 4 consists of $n_{\text{samples}} \cdot n_{\text{trajectories}} \cdot n_{\text{MW}} = 12 \cdot 11 \cdot 20 = 2640$ simulated kymographs, with $n_{\text{samples}} = 24$ number of samples per trajectory length and MW , for 11 trajectory lengths logarithmically equally spaced between $100 - 20480$ frames and $\text{MW} \in [1.5, 3, 5, 7, 9, 14, 28]$ kDa. The R_s of each simulated trajectory is calculated assuming the molecule is a globular protein diffusing through a (constraining) nano-channel, using the phenomenological model introduced in [4] to correct for hindrance effects related to the impact of the nanochannels’ constrictions on the diffusivity of the molecules. Specifically, the resulting R_s of the simulated particles is $R_s = [1, 1.3, 1.5, 1.8, 2.1, 2.6]$ nm.

A



B

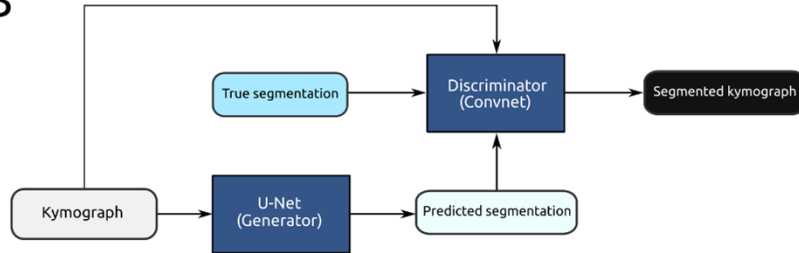


Figure 2: U-net for single biomolecule segmentation. (A) The U-net [13] consists of a series of contraction convolutional layers, a bottleneck, and a series of expansion convolutional layers, as well as a series of skip connections between corresponding contraction and expansion convolutional layers to ensure that information learnt during contraction is not lost at the bottleneck. Each 2x Conv box represents a convolutional block corresponding to 2 convolutional layers in sequence, and each Conv T box represents a single convolutional transpose layer as exemplified in the legend. Here, f is the number of filters in each block, ks is the kernel size, a is the activation function, and dr is the dilation rate. The network is visualized with Netron [14]. (B) Block diagram of the GAN training environment, where kymographs are fed both to a U-net generator network, which predicts a corresponding segmentation, and a Convnet (convolutional network) discriminator network, which takes both original kymographs and true segmentations as input to determine whether the predicted segmentation is correct. A basic Convnet consisting of 5 convolutional layers of size 4×4 and stride 1 connected to a single dense layer is used as discriminator network [15].

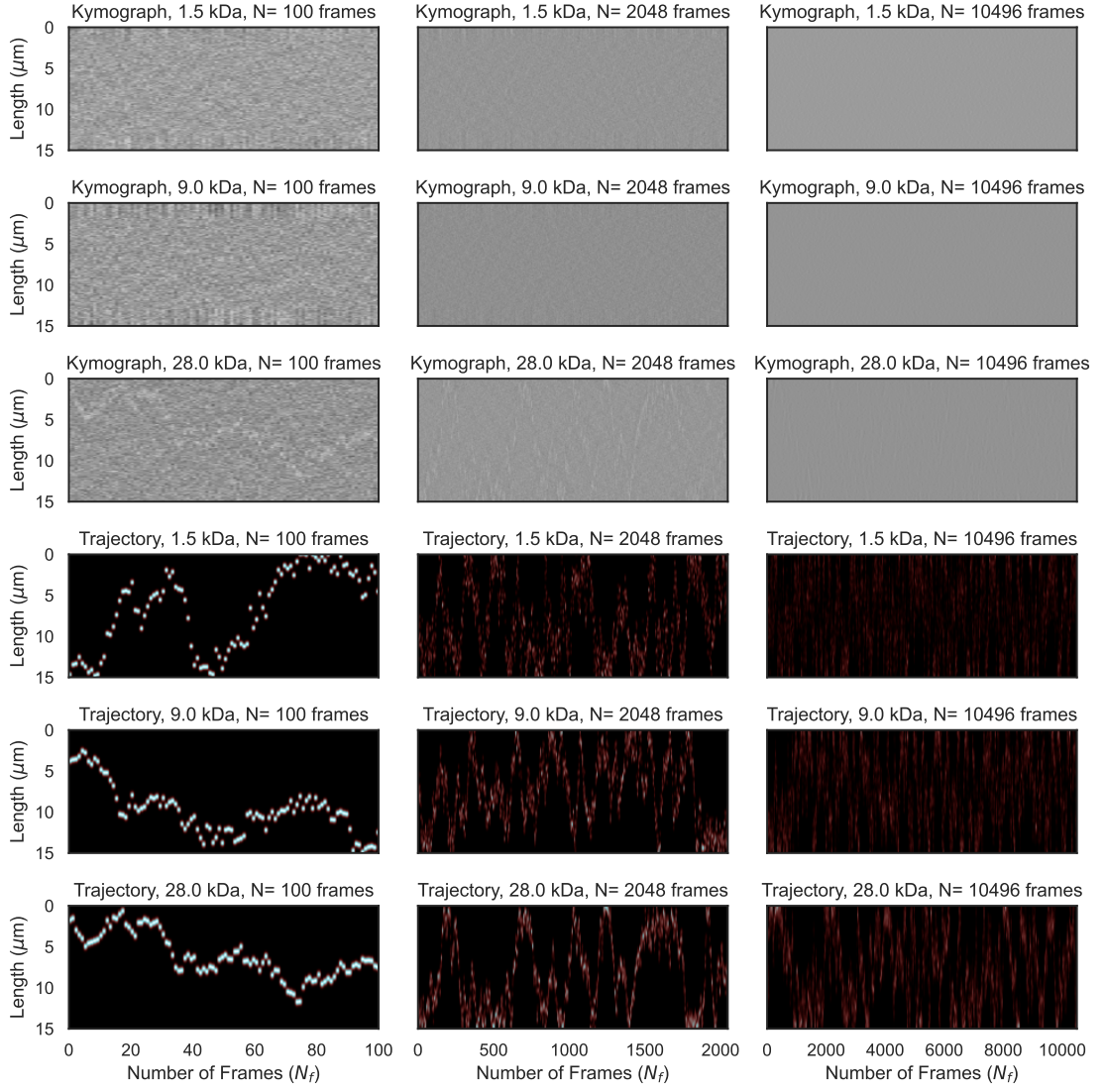


Figure 3: Representative subset of simulated dataset, consisting of simulated kymographs for $MW \in [1.5, 9, 28]$ kDa in channel A_{II} , and trajectory lengths $L \in [100, 2048, 10496]$

5.1 h-ViT as a Biased Estimator

As a key point, the h-ViT performs as a biased estimator dependant on the estimated molecular trajectories derived from its generated probability maps. Here, we show alternative visualizations of the underlying results, focusing on molecular weight and hydrodynamic radius predictions derived from our experiments, to further demonstrate this effect.

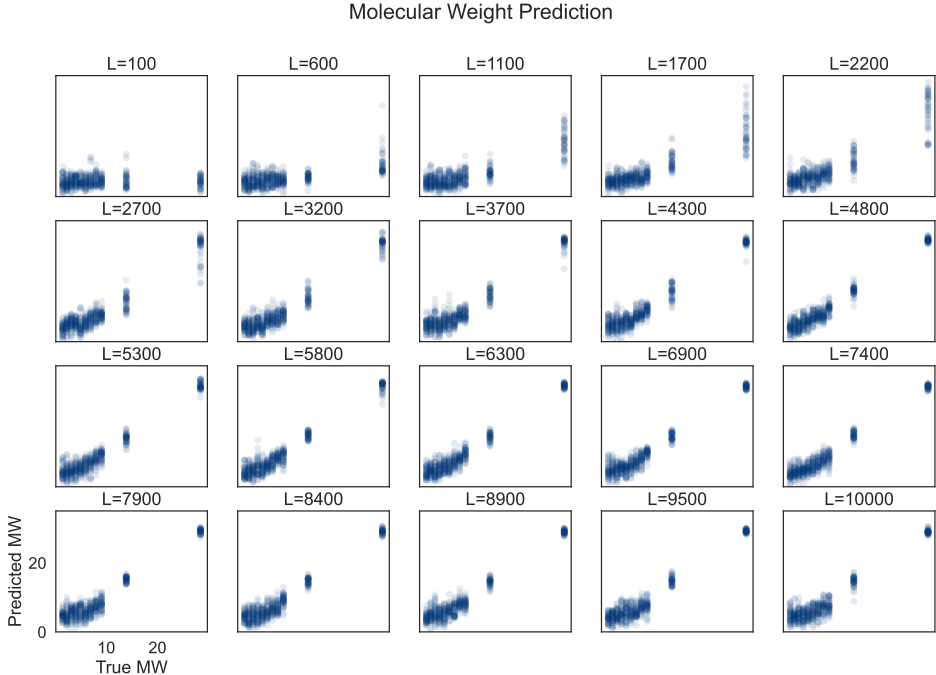


Figure 4: Predictions underlying the results in Figure 2 of the main text, consisting of predicting molecular weight of simulated molecules for $MW \in [10, 20]$ kDa for reference channel Ch_B for trajectory lengths L between 100 and 10,000 frames. True molecular weights are plotted against predicted molecular weights across a range of trajectory lengths, illustrating the convergence of the predictions with increasing L .

These visualizations highlight the robustness of our predictive framework, as longer trajectory lengths reduce uncertainties in both molecular weight and hydrodynamic radius estimations. The systematic improvement with increasing L underscores the importance of trajectory length in enhancing predictive accuracy. Specifically, we observe that for shorter trajectory lengths, the model’s predictions exhibit higher variance, reflecting limited sampling and lower statistical confidence. As trajectory lengths approach 10,000 frames, the predictions converge towards the true molecular weight and hydrodynamic radius values, with residual errors diminishing significantly. Conversely, the precision increases as the accuracy decreases for shorter trajectory lengths.

This at first somewhat surprising result can be understood as a consequence of the model being a biased estimator which tends to underfit the data at intermediate

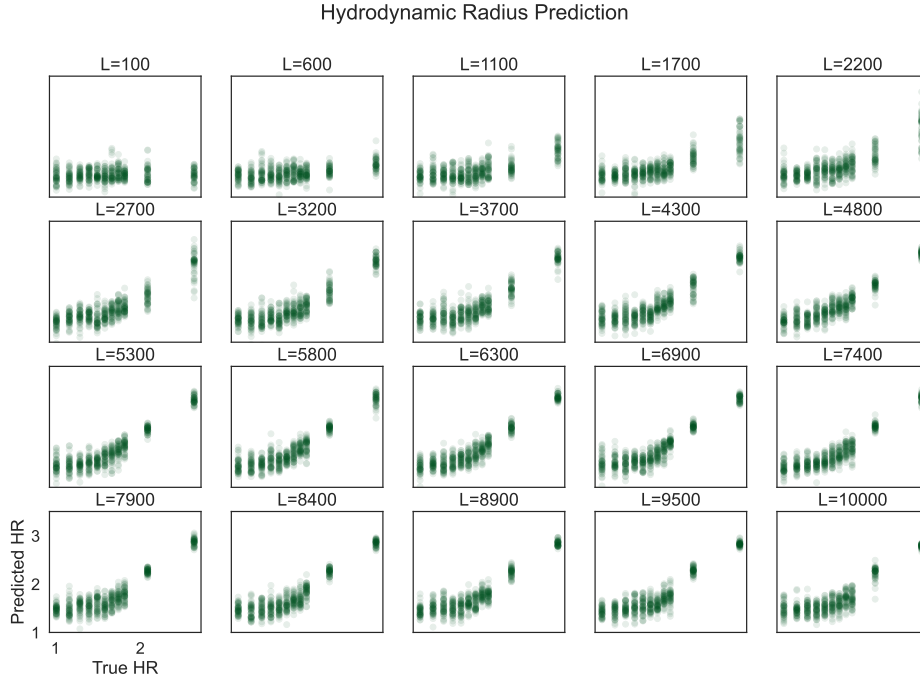


Figure 5: Predictions underlying the results in Figure 2 of the main text, consisting of predicting hydrodynamic radius of simulated molecules for $MW \in [10, 20]$ kDa for reference channel Ch_B for trajectory lengths L between 100 and 10,000 frames. True hydrodynamic radii are plotted against predicted radii, showing improvement in prediction accuracy with longer trajectory lengths.

trajectory lengths. This bias occurs because the model is designed to interpret kymographs with small N as highly likely to be noise, resulting in a systematic tendency to predict MW s close to zero and R_s close to 1 nm, in these cases, which for small true MW and R_s values indeed is a quite "good" estimate. As N increases, the network has more data at its disposal, allowing it to increasingly distinguish between actual signal due to the presence of a particle and noise. Nevertheless, initially in this regime, the precision of the model prediction decreases because the network is no longer dismissing as much data as noise. Hence, it starts to process kymographs that, for smaller N , would have been classified as noise and thus assigned MW s close to zero or R_s values close to 1 nm. Consequently, this introduces transiently larger variance into the predictions, which is reflected in the observed transiently reduced precision. However, with even further increasing N , the model continues to refine its predictions, as the additional trajectory points provide clearer particle signals, and both accuracy and precision improve. By the time the trajectory length reaches $N = 10000$, the h-ViT model is thus able to accurately and consistently predict MW and R_s of the 6 kDa particle, thereby effectively balancing the trade-off between filtering out noise and accurately capturing signals in the kymograph that stem from the presence of a particle.

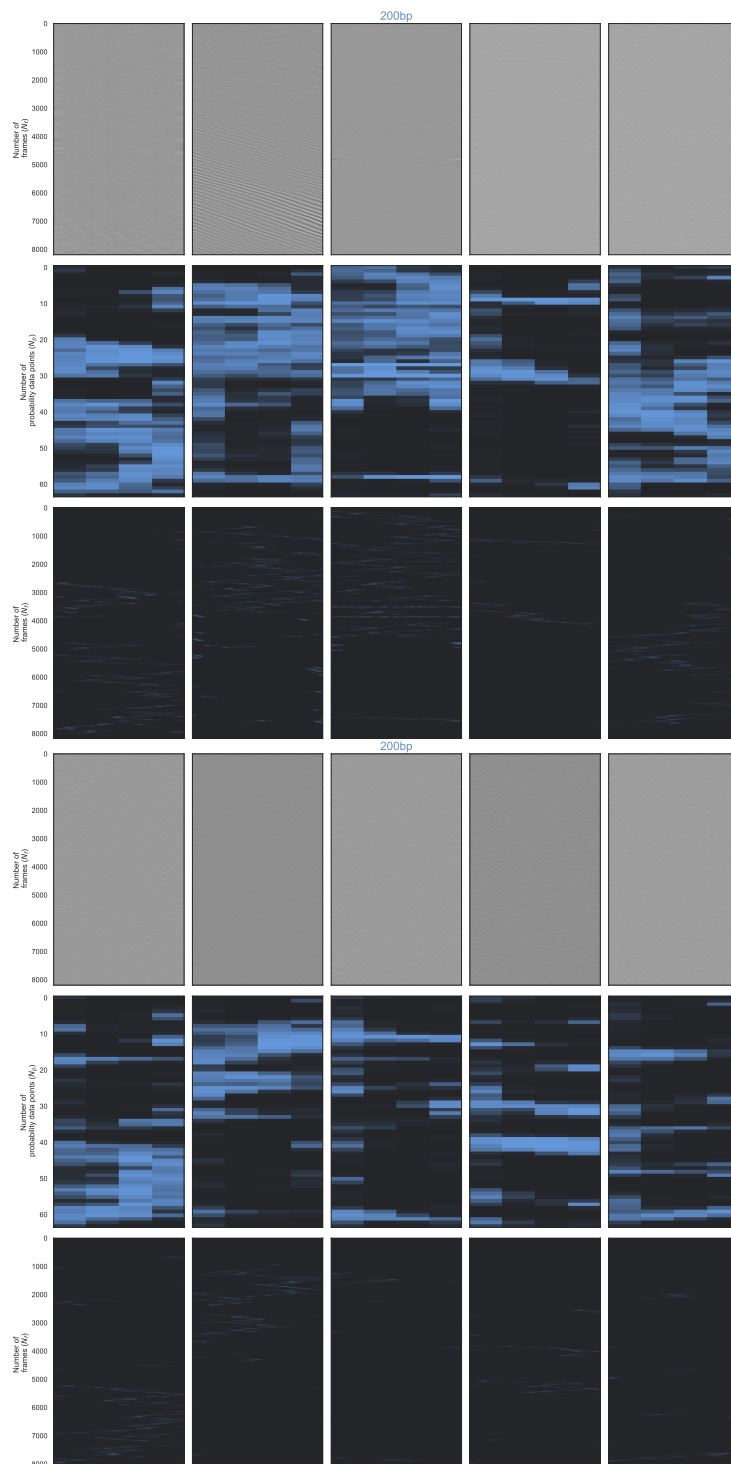
The statistical underpinnings of these trends warrant further discussion. For molecular weight predictions, the error decay aligns with a $1/\sqrt{L}$ dependence, consistent with central limit theorem principles applied to stochastic sampling. This suggests that as more frames are included, the effective sample size increases, enhancing the signal-to-noise ratio and reducing random fluctuations in the measurements. Similarly, hydrodynamic radius predictions benefit from extended trajectories through improved temporal averaging, which smooths transient deviations arising from molecular diffusion and interaction dynamics.

The convergence behavior also reflects the interplay between model fidelity and data sufficiency. For $L < 1000$, the predictions are influenced more strongly by prior assumptions and the model's intrinsic limitations, as data scarcity imposes constraints on predictive precision. Beyond this threshold, the influence of empirical data grows dominant, enabling the model to capture subtle molecular variations with high accuracy. This transition underscores the dual necessity of robust modeling frameworks and sufficient experimental data to achieve reliable characterization.

In addition, these findings have broader implications for experimental design and methodological optimization. The demonstrated improvement with longer trajectories emphasizes the need for high-throughput, long-duration measurements in nanochannel-based systems to fully exploit their analytical potential. Future studies could explore adaptive sampling strategies that dynamically adjust trajectory lengths based on real-time prediction confidence, thereby balancing measurement efficiency and accuracy. This approach could further enhance the applicability of nanochannel-based techniques across diverse fields, including polymer science, biophysics, and molecular diagnostics.

6 Full DNA Measurements

For the smallest fragment, the mean predicted MW is 26.3 ± 8.5 kDa, compared to the 30 kDa nominal value, which we consider a good agreement given the small size of the fragment. We also note that this value is very close to the mean value predicted for the empty nanochannel control with only TE-buffer, i.e. 20.5 ± 4.5 kDa. This can be understood as that the predictions for the smallest DNA fragment (50 bp/30 kDa) indeed are approaching the noise level of the system for the given channel size, indicating that the model’s detection capability is being limited by the weak scattering signal inherent to such small molecules. From the perspective of the h-ViT model, this behaviour is expected due to its probabilistic approach, which shifts the focus away from precise localization of weak signals and instead prioritizes broader property prediction based on integrated features across the kymograph. Hence, the fact that the 50 bp DNA fragment’s predicted *MW* quite closely matches that of the TE-buffer control suggests that the signal is too weak to reliably distinguish from noise at the nanochannel and molecular size at hand. However, in local regions where the scattering signal is stronger, the h-ViT model can effectively utilize its hierarchical attention mechanisms to extract meaningful patterns from the data, resulting in more accurate predictions that align closely with the known *MW*. This underscores the strength of our approach to detect and characterize molecules near the LoD, while highlighting the inherent challenge in detecting extremely small biomolecules when the scattering signal is near the noise threshold.



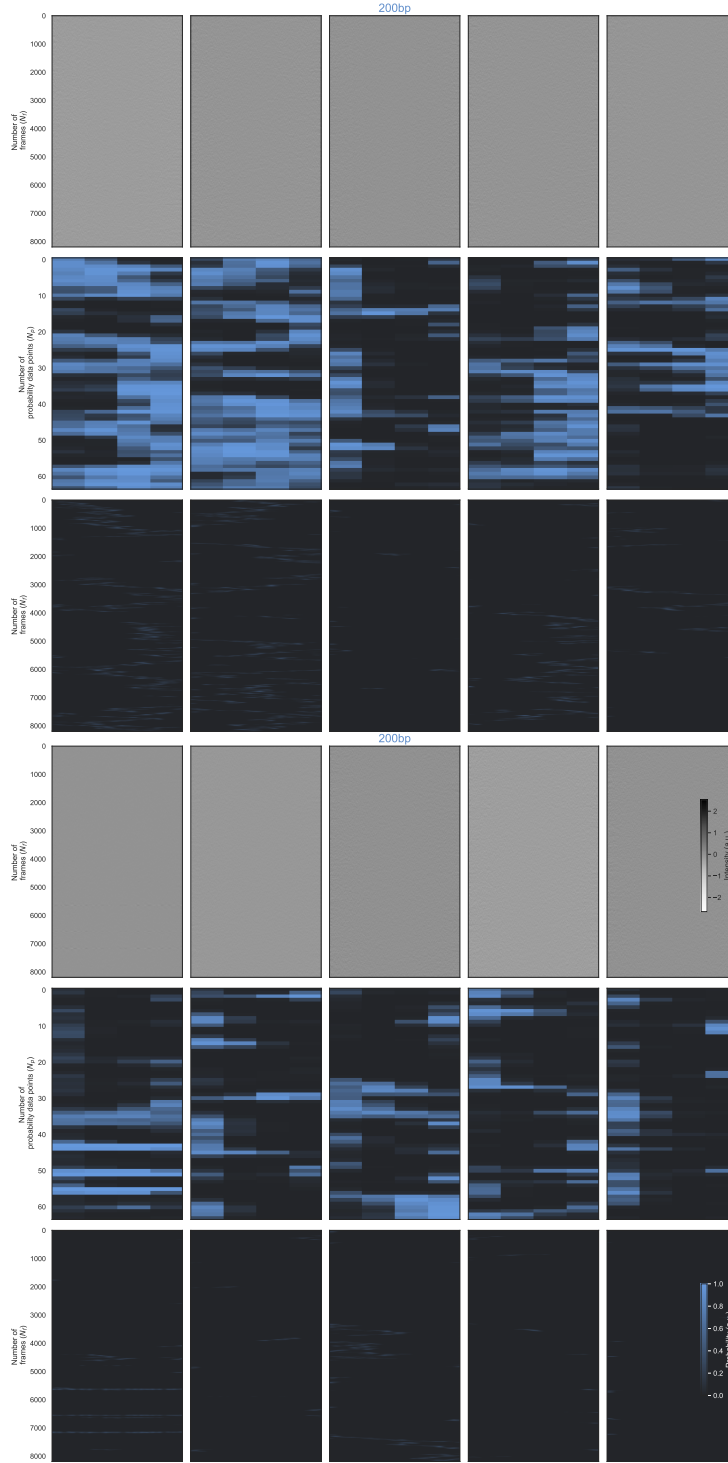
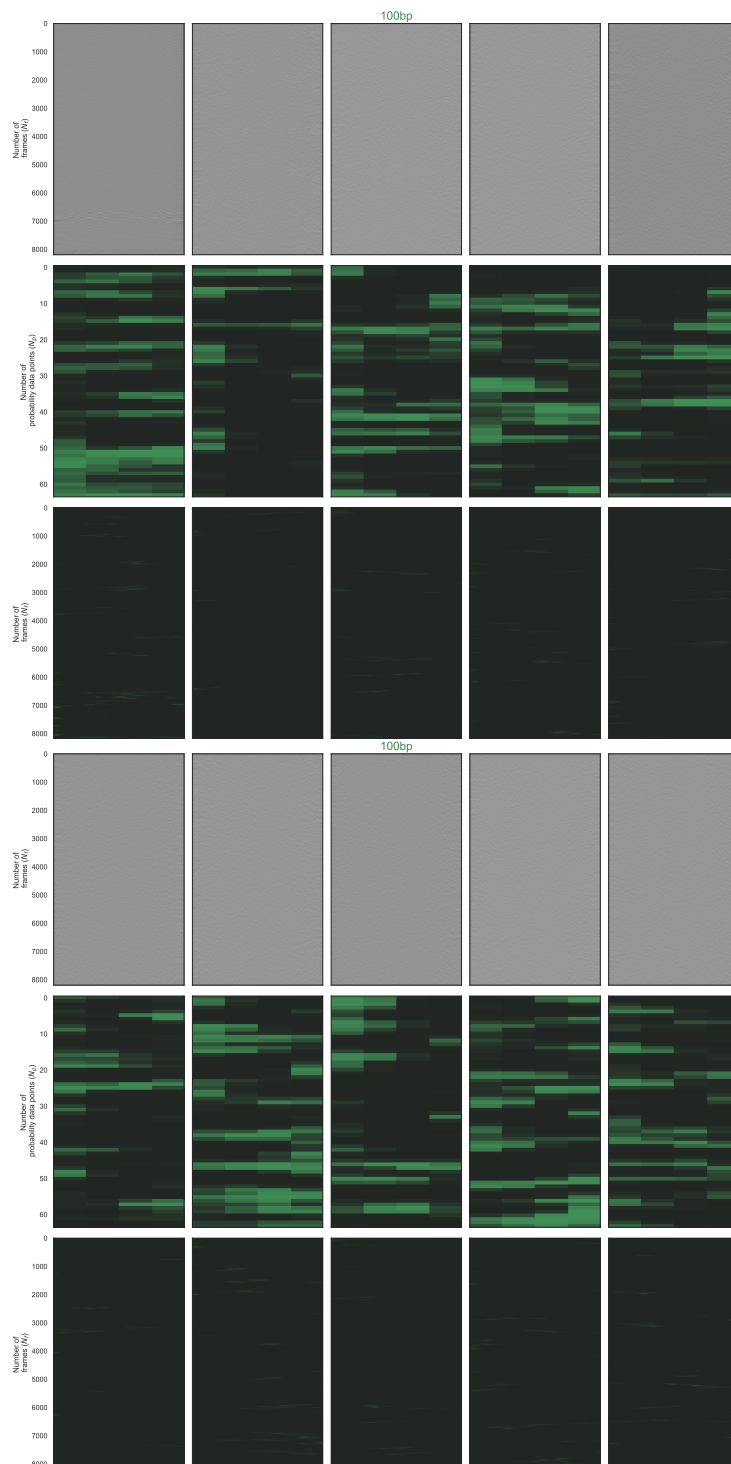


Figure 7: **DNA Measurements in a Nanofluidic Channel with 122 nm x 97 nm cross sectional dimension.** The experimental results of analyzing , and of a pure TE-buffer control are shown. (A) Experimentally measured NSM kymographs for 200 bp double-stranded DNA fragments in TE-buffer. (B) Probability maps generated by the h-ViT model using the kymographs in (A) as input. (C) Segmented kymographs generated by the cGAN model using the kymographs in (A) as input.



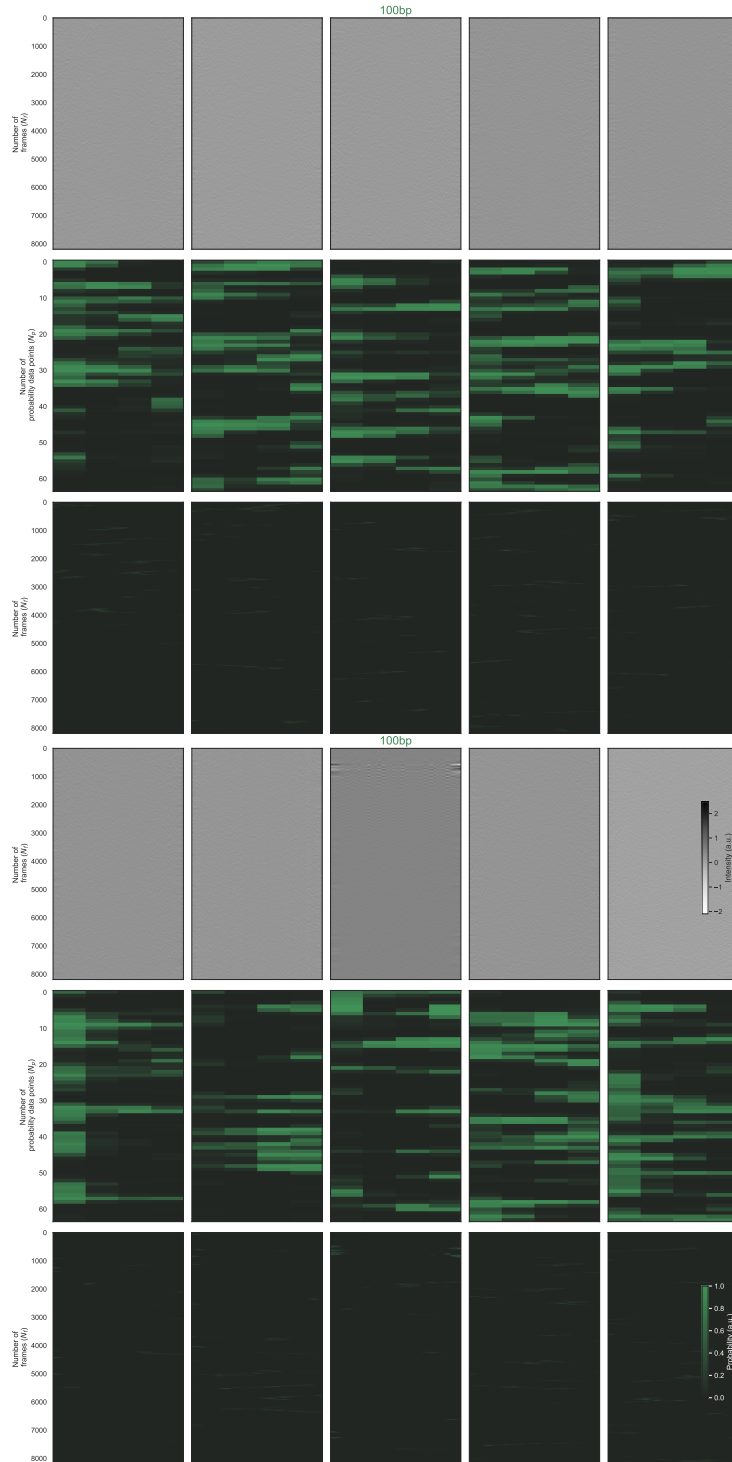
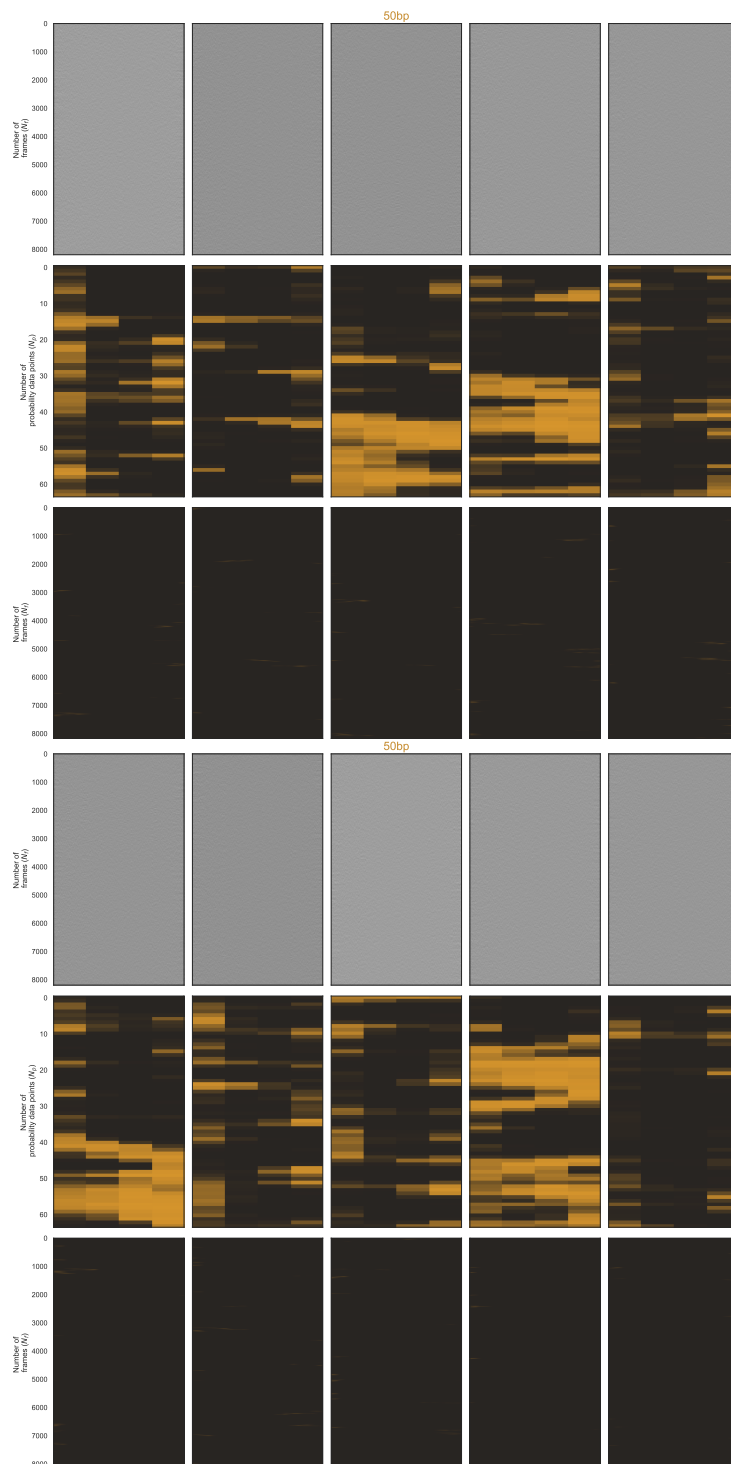


Figure 9: Equivalent to Figure 7, but for 100 bp double stranded DNA.



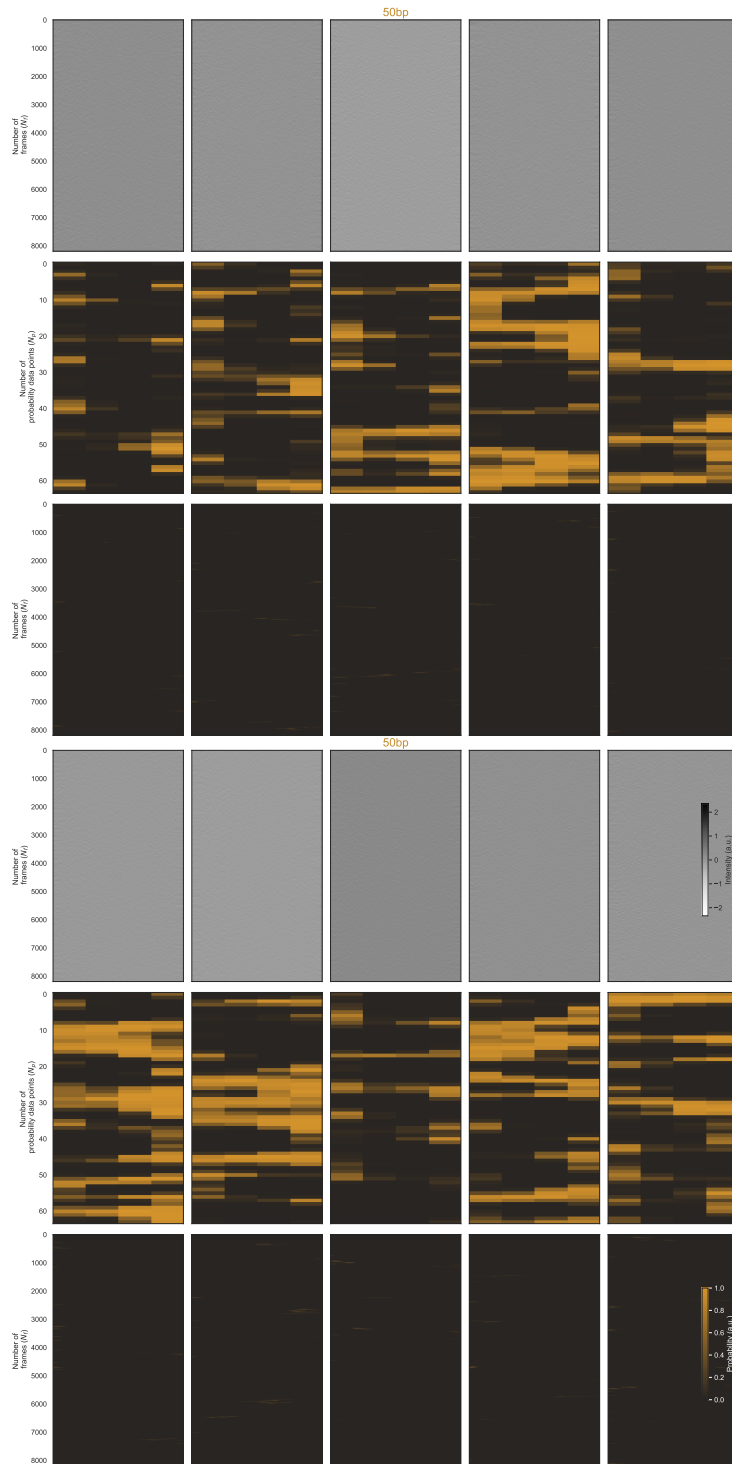
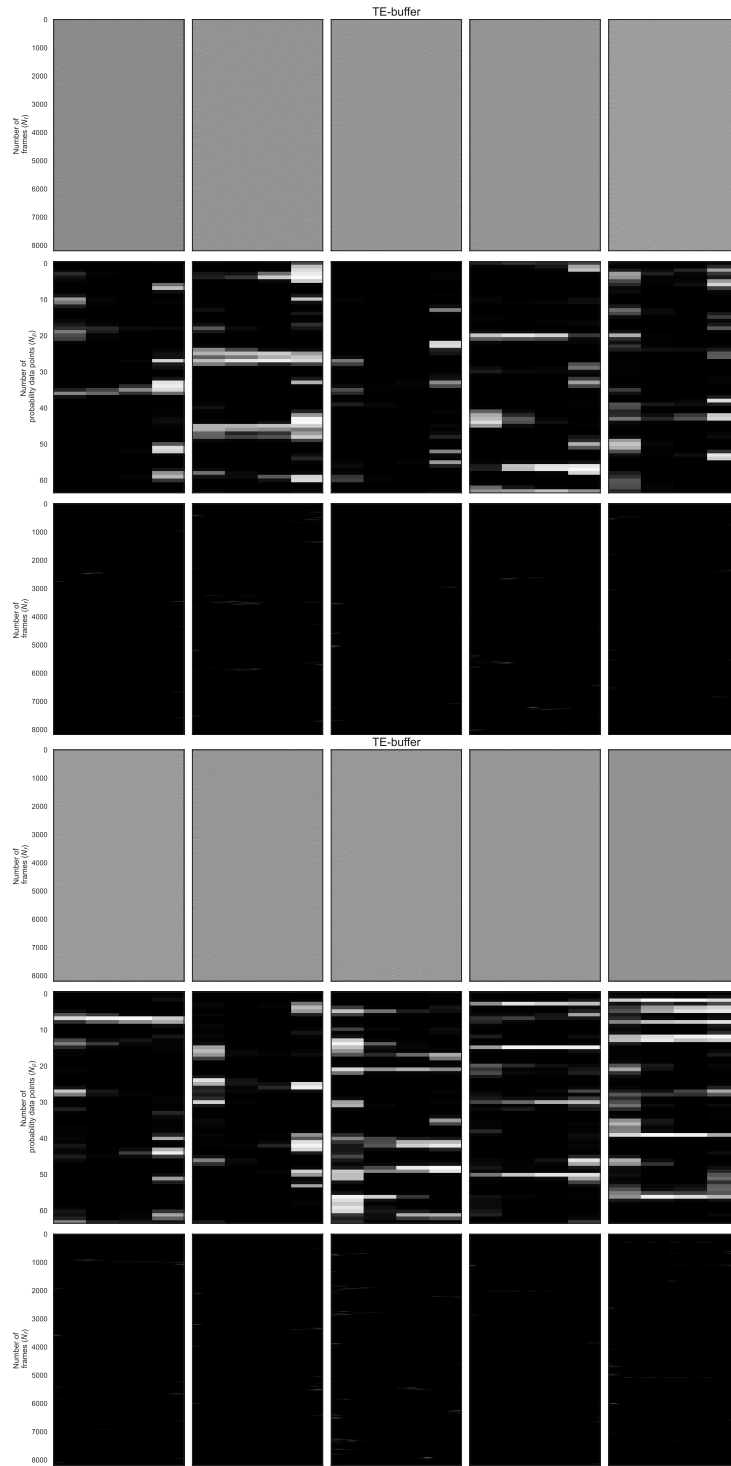


Figure 11: Equivalent to Figure 7, but for 50 bp double stranded DNA.



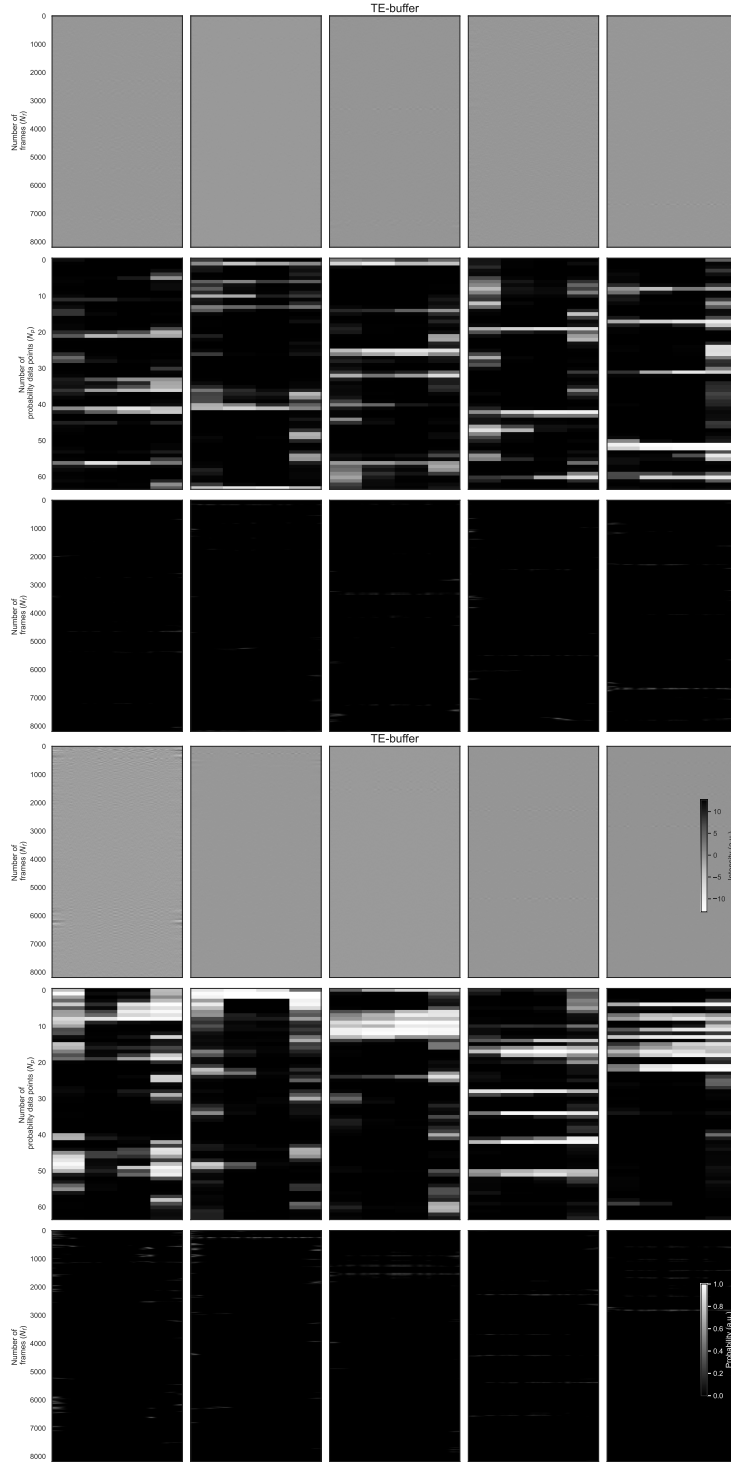


Figure 13: Equivalent to Figure 7, but for TE buffer.

7 Complete Insulin Measurement Data Set

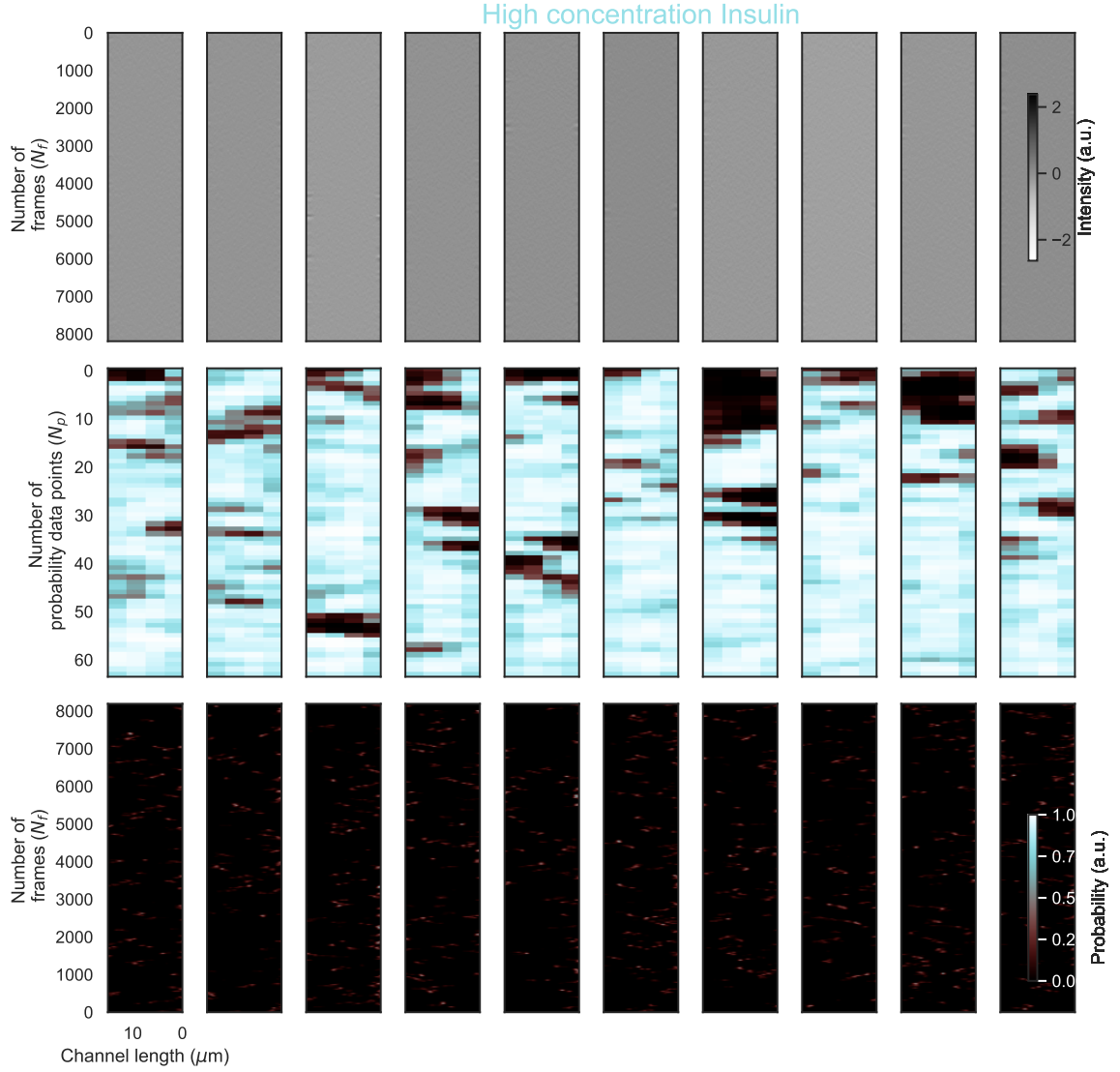


Figure 14: **Insulin Measurements in a Nanofluidic Channel with 60 nm x 29 nm cross sectional dimension.** (A) Experimentally measured NSM kymographs for Insulin in PBS-buffer. (B) Probability maps generated by the h-ViT model using the kymographs in (A) as input. (C) Segmented kymographs generated by the cGAN model using the kymographs in (A) as input.

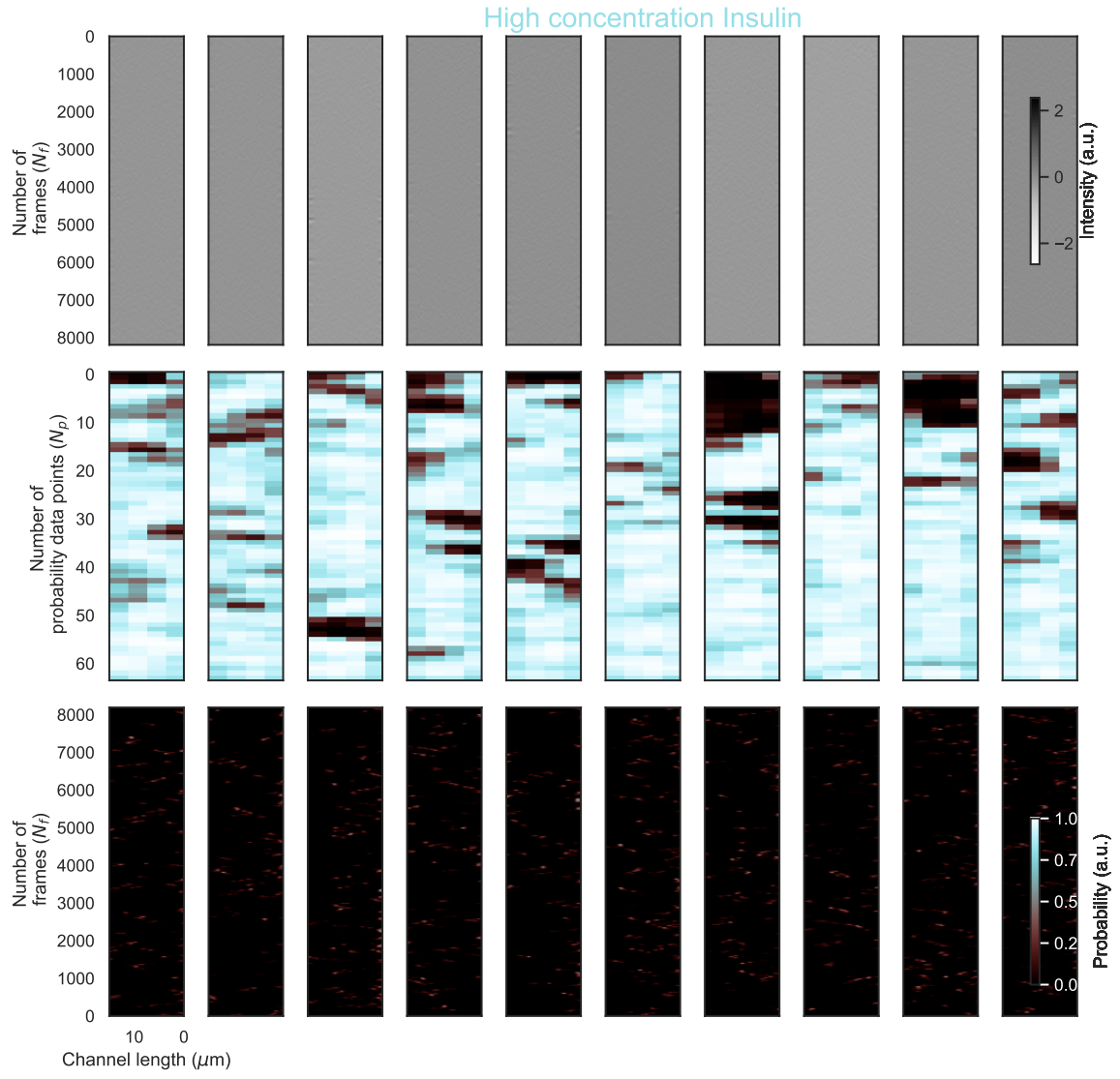


Figure 15: Equivalent to Figure 14.

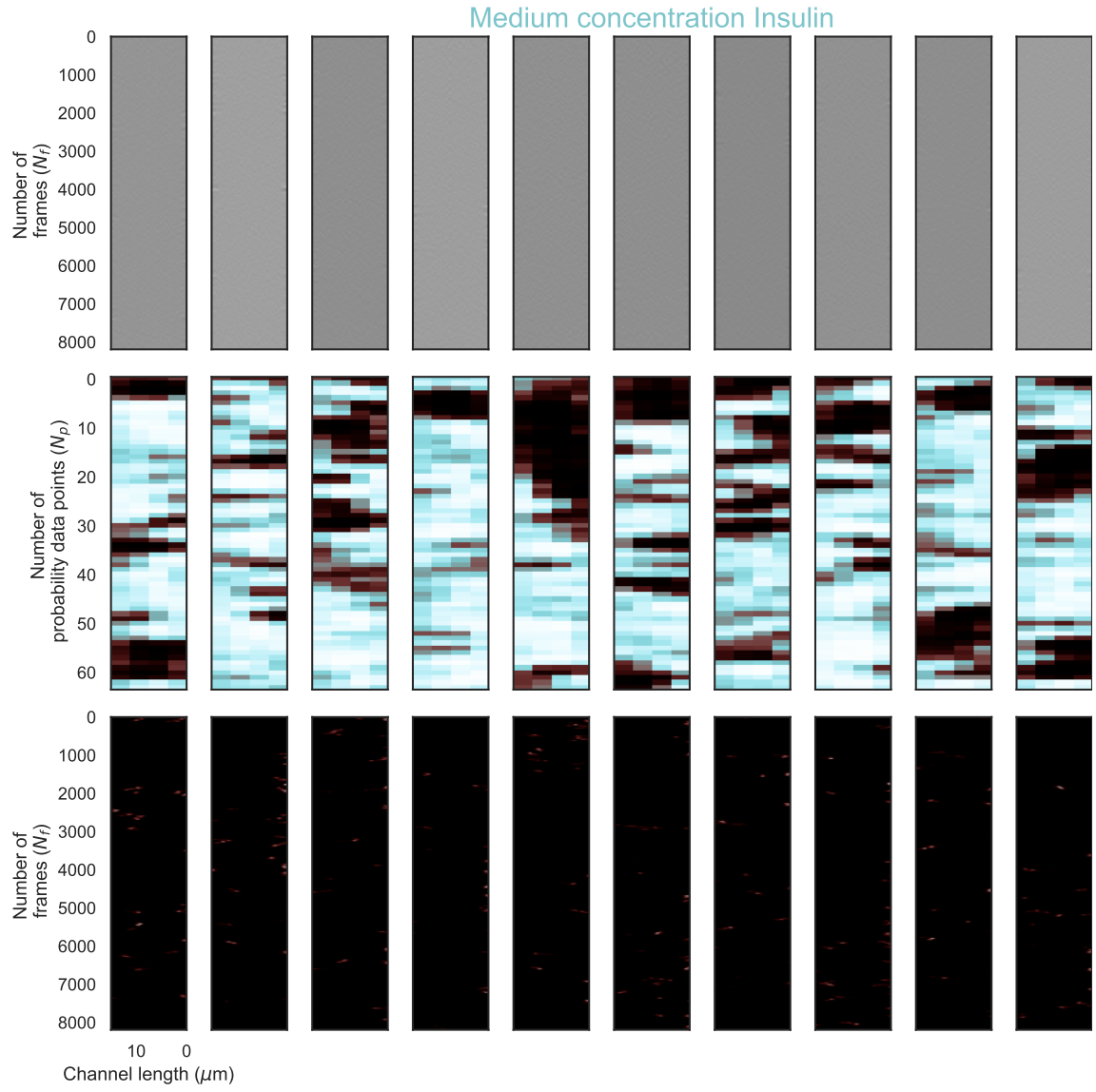


Figure 16: Equivalent to Figure 14, but for medium concentration Insulin.

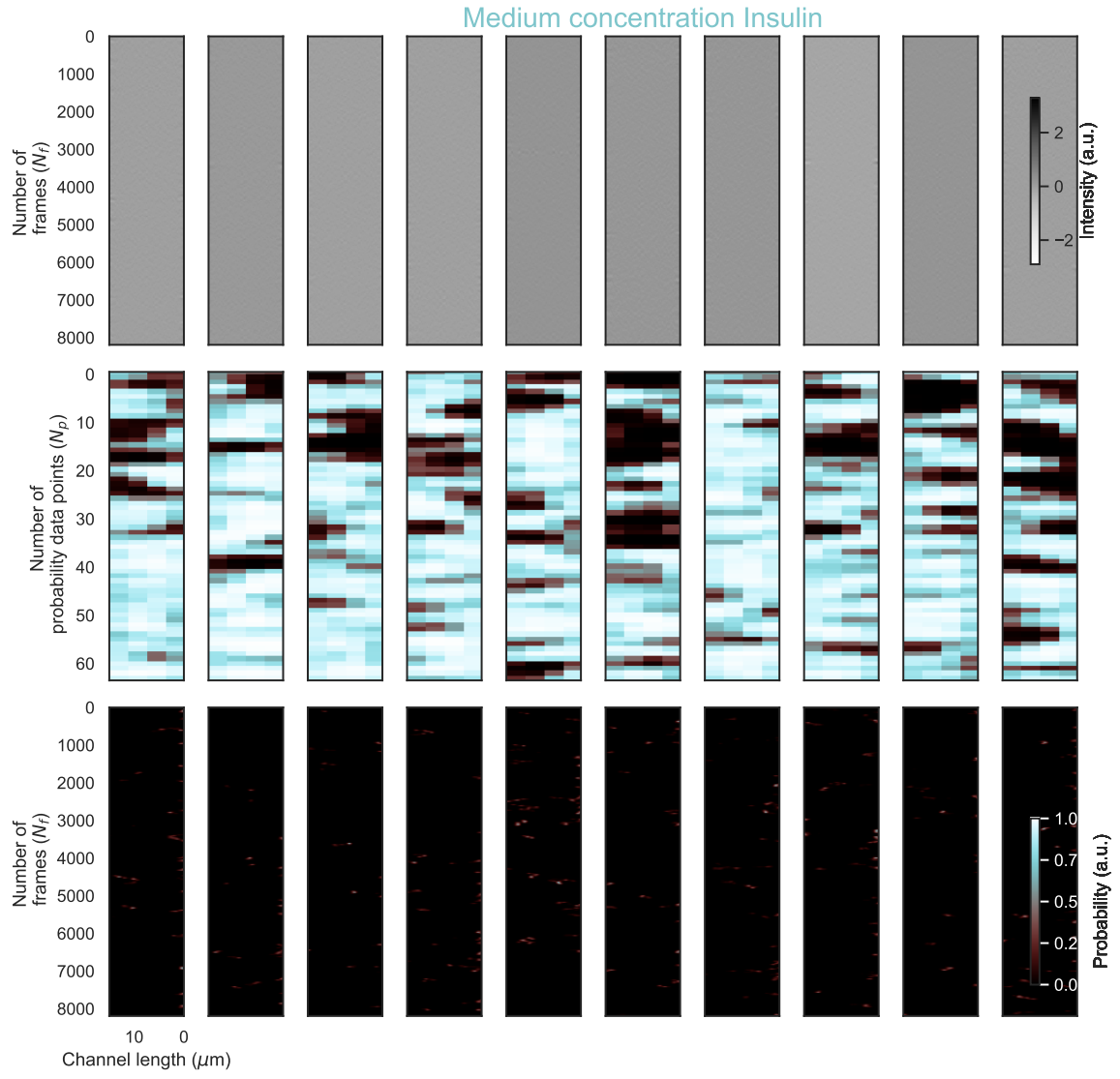


Figure 17: Equivalent to Figure 14, but for medium concentration Insulin.

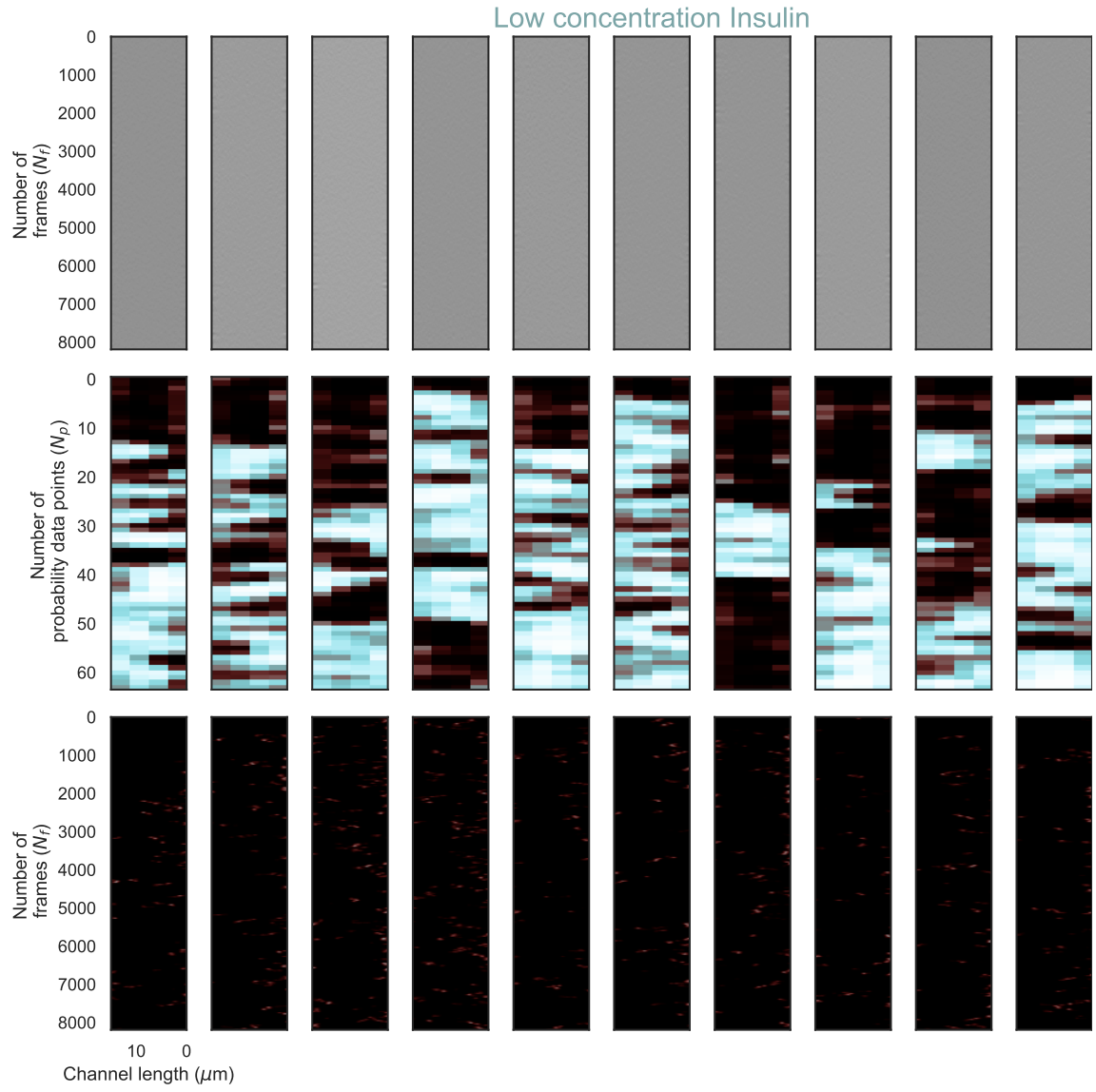


Figure 18: Equivalent to Figure 14, but for low concentration Insulin.

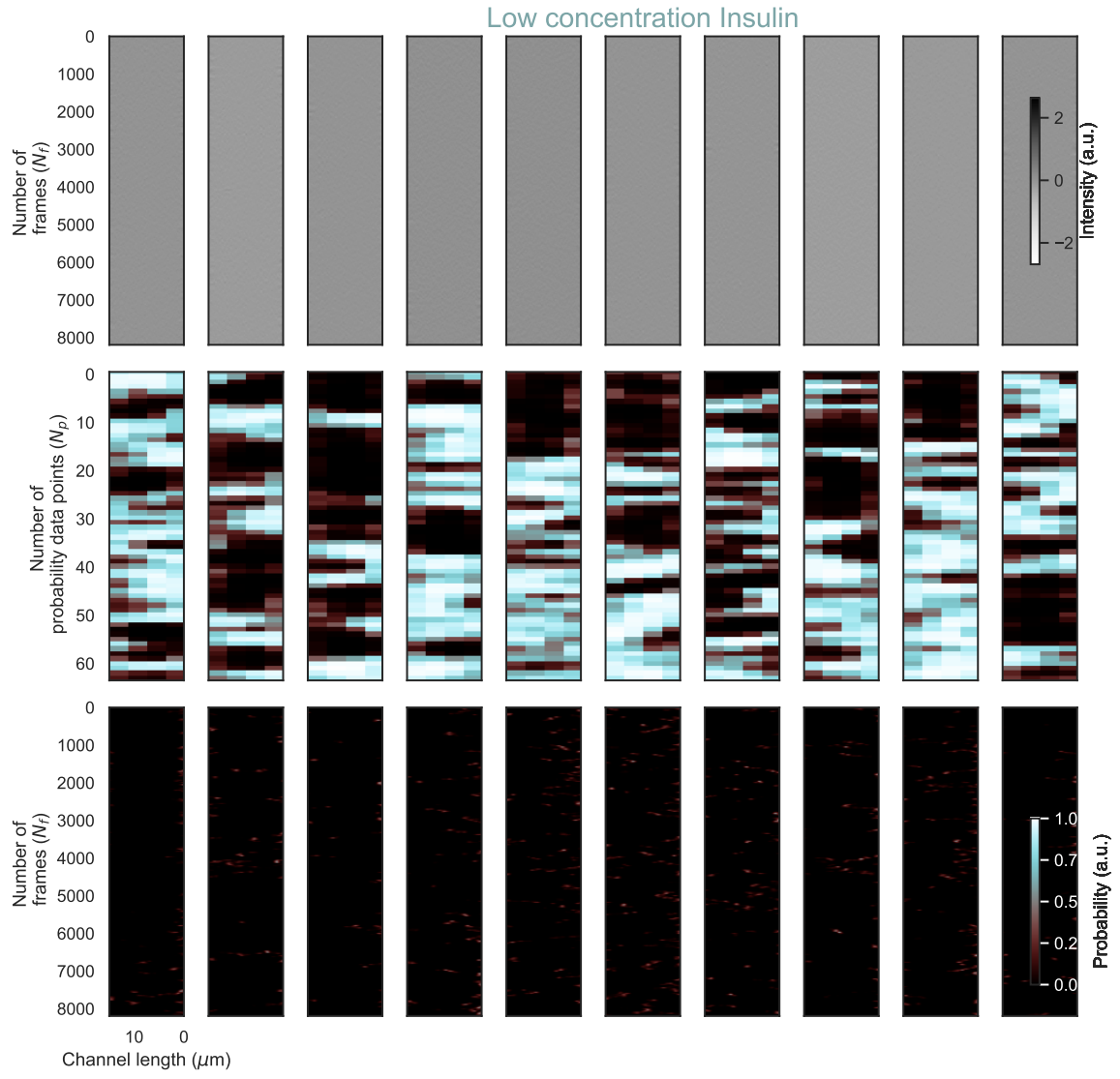


Figure 19: Equivalent to Figure 14, but for low concentration Insulin.

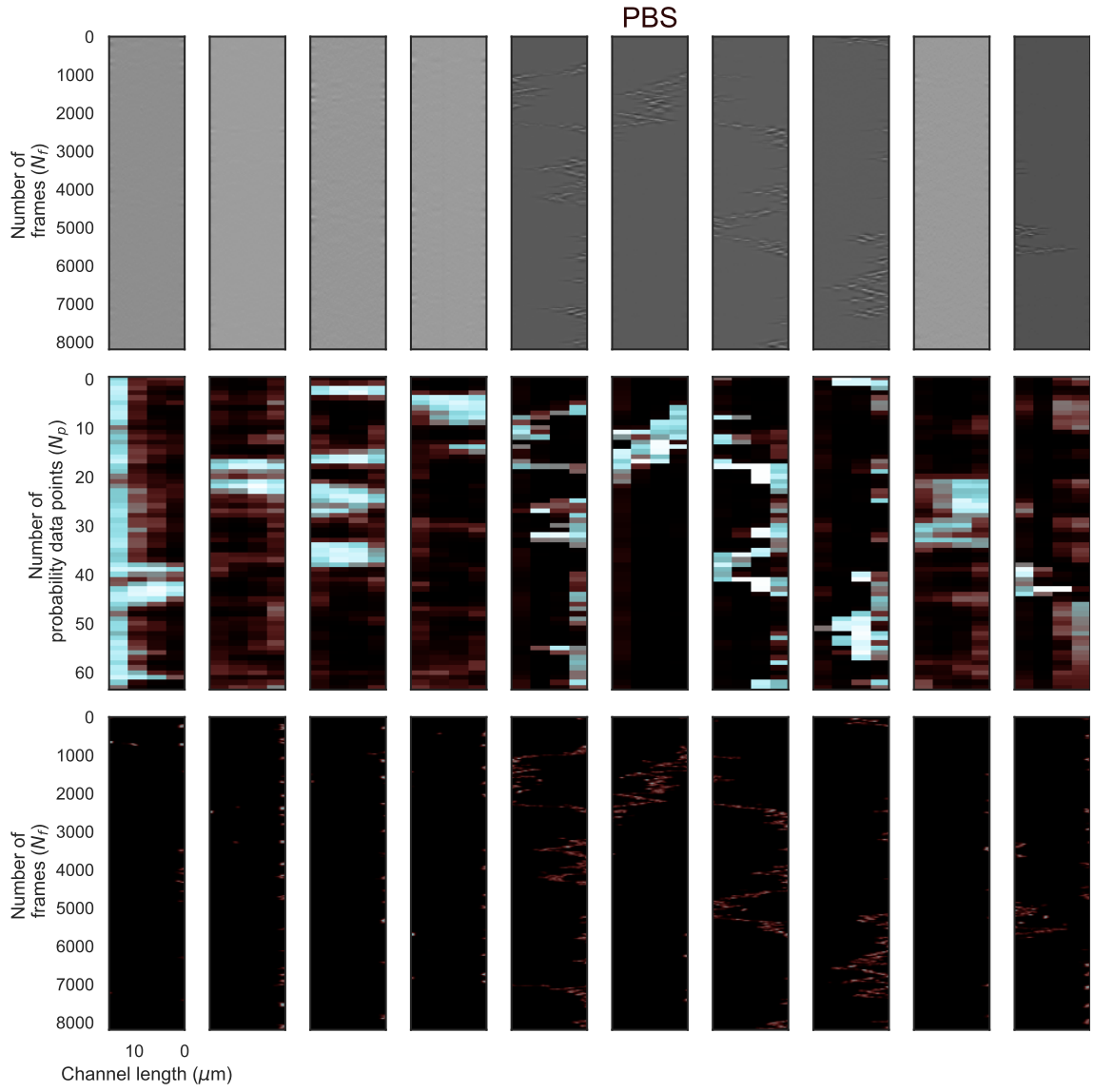


Figure 20: Equivalent to Figure 14, but for PBS buffer.

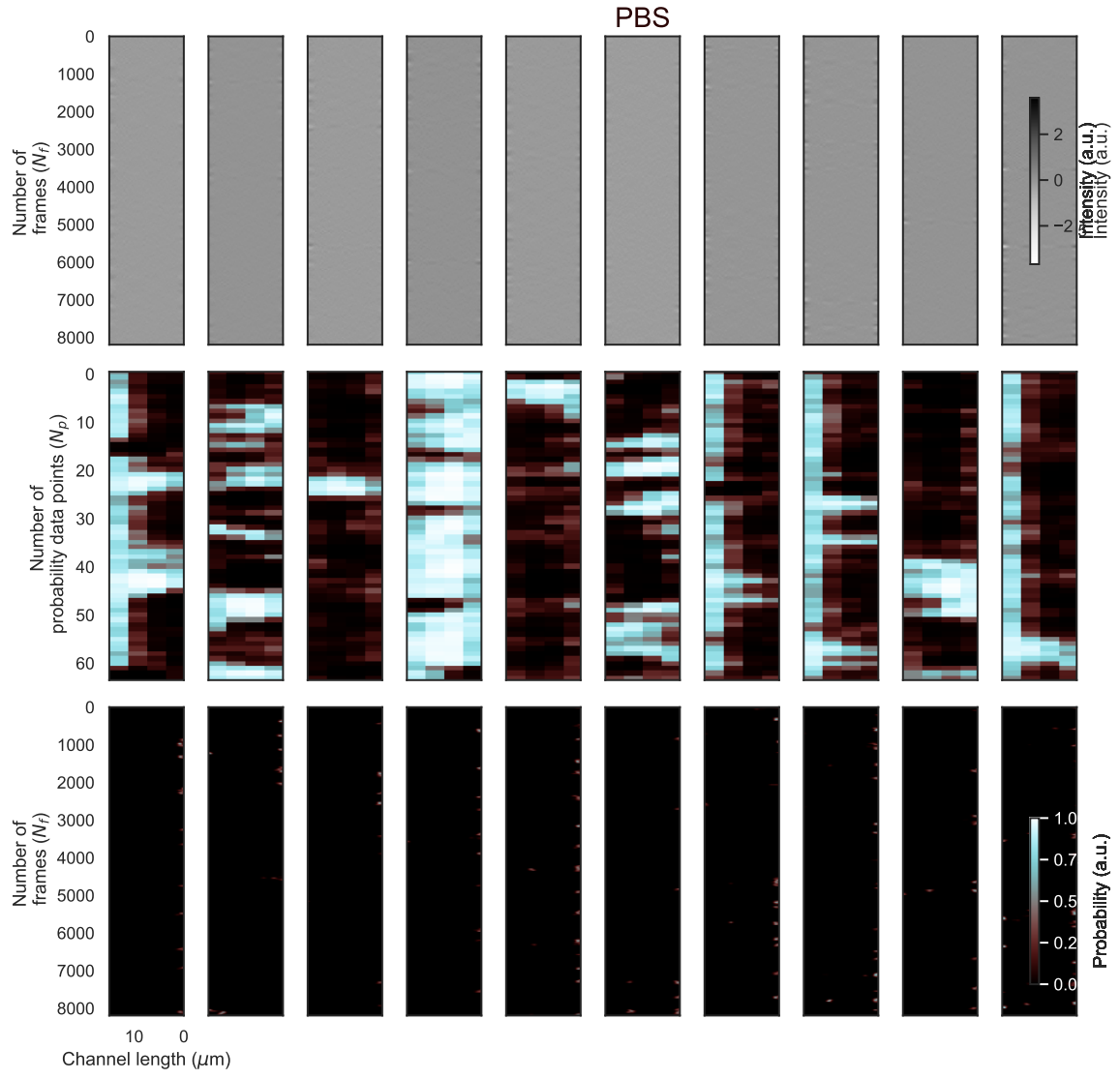


Figure 21: Equivalent to Figure 14, but for PBS buffer.

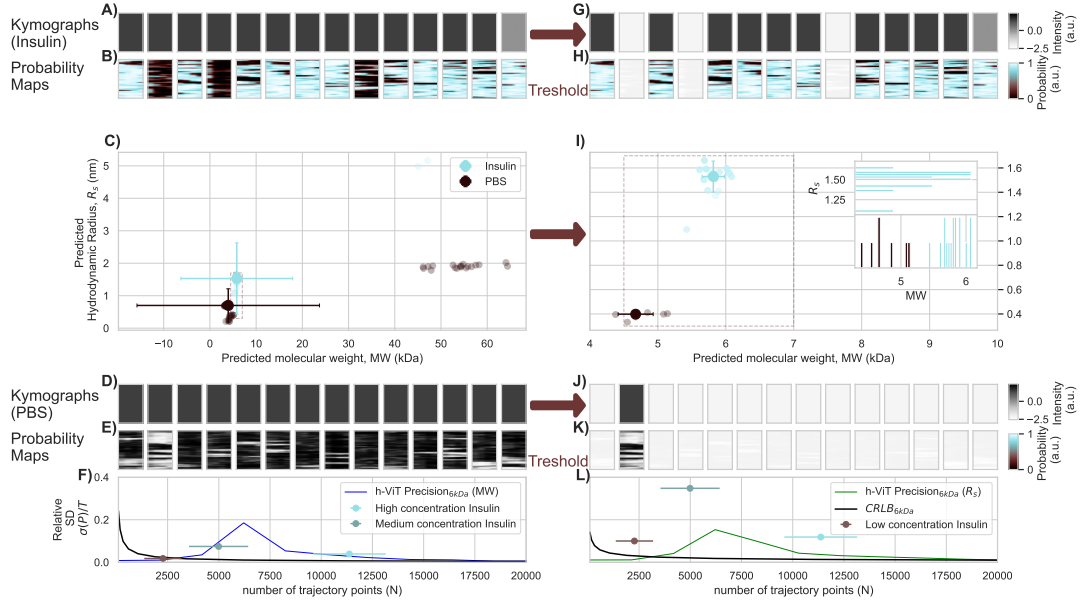


Figure 22: NSM Single Insulin Analysis in a Nanofluidic Channel with **29 nm x 60 nm cross-sectional dimensions**. (A) Representative kymographs for experimental Insulin measurements with NSM, serving as input to the h-ViT model. (B) Corresponding probability maps generated by the h-ViT model, indicating regions of high Insulin localization probability inside the nanochannel. (C) Predicted R_s versus MW for Insulin (turquoise) and pure PBS buffer control measurements (black), where each data point corresponds to a kymograph and the error bars correspond to the standard deviation across all measured kymographs. The transparency of each data point is correlated with the number of trajectory points within the respective kymograph. (D, E) Same as (A,B) but for nanochannels only filled with PBS buffer. (F) Relative standard deviation (σ) as function of number of trajectory points of predicted MW based on simulated trajectories for a 6 kDa molecule (blue line), experimental Insulin measurements at low, medium and high concentration (individual points) and the Cramér-Rao Lower Bound (CRLB) that illustrates the theoretical limit of prediction. (G-K) Same as (A-E) but for an applied probability threshold. (L) Same as (F) but for R_s .

Inputting the kymographs to the h-ViT model produces probability maps that highlight regions where the network has detected Insulin with high confidence **Figure 22B**. Corresponding probability maps obtained from kymographs measured from channels only filled with PBS buffer exhibit only very low probabilities, confirming the absence of Insulin molecules in these control samples **Figure 22E**.

Subsequently, based on these probability maps, the h-ViT model predicts the MW and R_s values for each measured kymograph for the Insulin and PBS-control samples, first without applying any probability threshold (**Figure 22C**). Clearly, identified molecules in the Insulin sample cluster around a common median value of MW

$= 5.8 \pm 13.1$ kDa and $R_s = 1.5 \pm 1.1$ nm, which is equal to the nominal $MW = 5.8$ kDa and $R_s = 1.5$ nm. Nevertheless, we also notice a significant spread in the individual predicted values, as indicated by the error bars in **Figure 22C**. This spread is the consequence of variability in the scattering signal intensity due to the inherent noise in the low SNR conditions at hand. For the pure PBS-buffer control measurements the model predicts two clusters of particles, one at lower MW and R_s than the Insulin sample and one at significantly higher values. This is the consequence of residual noise in the system being misinterpreted as ultra low-molecular-weight particles in the lower cluster when essentially no real detectable objects are in the system, and as larger objects in the higher cluster which, in principle, may correspond to real particles of a different type than Insulin. However, in particular the data points in the higher cluster are very faint, which means they were extracted from kymographs with small N , which assigns them an inherent low reliability.

As the next step, in analogy to the DNA-ladder, we applied a threshold (see 'Post-processing' section in Methods) to the probability maps to be included in subsequent MW and R_s predictions also in this case (**Figure 22G-K**). As expected, this thresholding step significantly improves the precision and accuracy, as evidenced by only probability maps with significantly concentrated high probabilities being included in the analysis (**Figure 22H**) and consequently significantly narrowed predicted MW and R_s distributions (**Figure 22I**). This is the consequence of the increased SNR in the selected probability maps, which effectively filters out low-confidence detection of molecules, and renders the identification of true insulin-related signals more consistent, and thus the MW and R_s predictions more accurate and precise. This is reflected by a much tighter distribution of the points in the scatter plot (**Figure 22I**).

As a second aspect, we also see that the thresholding effectively eliminates the low confidence points from the PBS control data ((**Figure 22J, K**) and thereby also enhances the distinction between Insulin and PBS measurements as it generates a clearer separation between the Insulin and PBS control data point clusters (**Figure 22I**). This is the consequence of the thresholding reducing the influence of noise-driven artefacts in the PBS measurements, which are "mistaken" for small particles by the model because the model is programmed to always output predicted properties, even when the probability is very low.

As the final aspect, we discuss the absolute values of predicted median Insulin MW and R_s values after thresholding, i.e. $MW = 5.85$ kDa \pm 0.18 kDa and $R_s = 1.53$ nm \pm 0.13 nm (**Figure 22I**). First, we note that the absolute values are very similar to the ones before thresholding but that the standard deviation has decreased dramatically. This advertises an impressive precision of the made predictions when a probability threshold is applied. Secondly, we note the the mean predicted Insulin MW is 0.5 kD larger than the literature value of 5.8 kDa **Jensen2014**. As the main reason for this slight discrepancy we identify the level of uncertainty in accurately determining the nanochannel cross-sectional dimensions, in particular in the very small nanochannel size regime at hand here, since such discrepancy directly affects the calibration of the *iOC* to MW conversion, and since its relative importance is

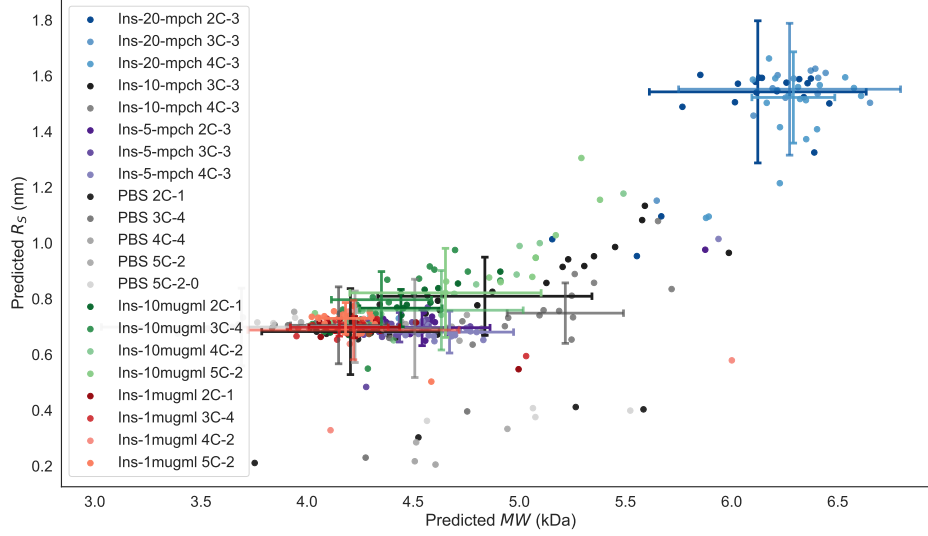


Figure 23: All results across concentrations for different channels (and therefore sizes). Is this "too much" to include? Not sure where it fits.

inversely proportional to the channel cross section [3]. When it comes to the mean predicted R_s value (is 1.53 nm), we find excellent agreement with the literature value of 1.5 nm **Jensen2014**. This is in line with the above argument for the slight discrepancy in MW prediction because the prediction of R_s , in contrast to the MW prediction, is insensitive to channel cross sectional dimension uncertainty, as long as the iOC is large enough to enable particle tracking, and thus derivation of D .

8 Nanochannel Area Estimation

The nanochannel structures analyzed in this work are characterized by their distinctive geometric features, as shown in Figure 24. To extract quantitative insights, we assessed the scattering cross sections through precise morphological evaluations using high-resolution scanning electron microscopy (SEM). These analyses were conducted to estimate the nanochannel area, which is crucial for determining molecular weights correlated with the optical contrast of molecules measured in our experiments.

In Figure 24, the leftmost panel depicts the initial SEM image of a nanochannel, with a green bounding box delineating the analyzed region. The center panel demonstrates the trapezoidal approximation of the nanochannel's cross-sectional geometry, which is the sole geometric model used for area estimation and molecular weight calculations. The rightmost panel overlays additional geometric fits, including the smallest enclosing rectangle (red), the largest inscribed rectangle (blue), and the equivalent square (dark green), which are included solely for visual demonstration and are not used in any calculations.

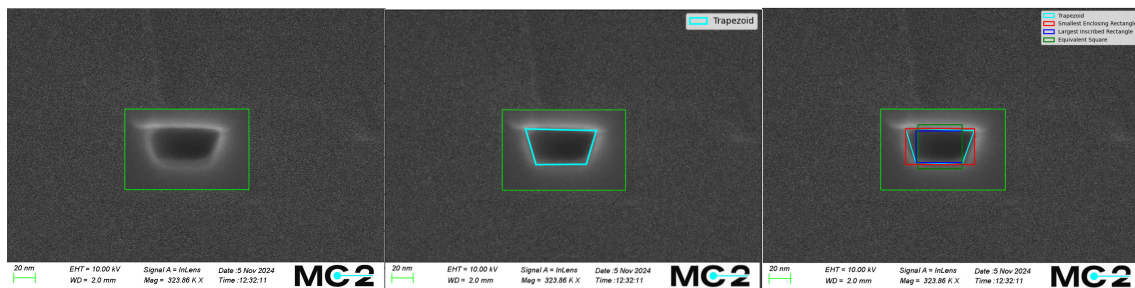


Figure 24: Visualization of the nanochannel geometries evaluated through SEM imaging. The left panel shows the raw SEM micrograph with the region of interest outlined in green. The center panel highlights the trapezoidal fit to the nanochannel’s shape, while the right panel overlays additional geometrical fits, including bounding rectangles and equivalent squares, for visual demonstration purposes. Scale bar: 20 nm.

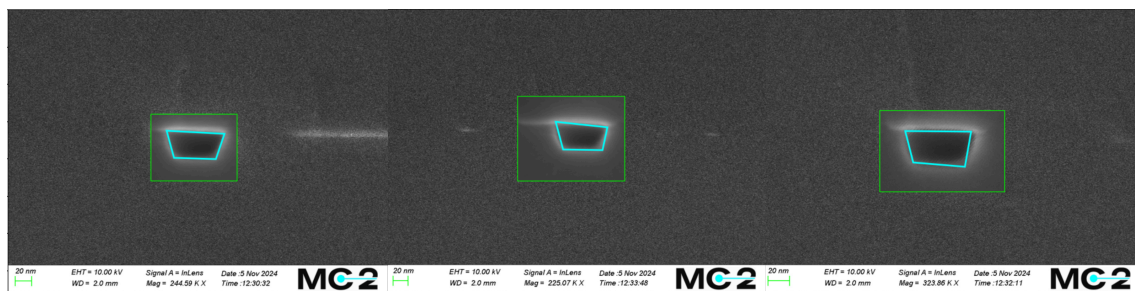


Figure 25: Scanning electron microscopy (SEM) images of a nanofluidic channel, captured three times in identical conditions to assess geometric consistency and structural uniformity. The cross-sectional shape of the channel is approximated as a trapezoid, with estimated areas of 1828, 1964, and 1880 nm² based on image analysis. The mean cross-sectional area is calculated to be 1890 nm², which provides a crucial parameter for determining optical contrast in nanofluidic experiments. These measurements and subsequent analysis serve as a basis for evaluating nanochannel fabrication precision and its impact on single-molecule analysis within the system.

The nanochannel area estimation directly impacts the accuracy of molecular weight determination. Variations in the trapezoidal fit introduce uncertainties in the area measurement, which propagate into the calculated molecular weights. These uncertainties, combined with intrinsic fluctuations in optical contrast, necessitate rigorous error analysis to ensure the reliability of our results. By focusing on the trapezoidal approximation and minimizing its associated uncertainties, we aim to improve the correlation between molecular weight and optical contrast, enabling more precise characterization of the molecules studied.

Following Figure 25, the scanning electron microscopy (SEM) images depict the cross-sectional geometry of a nanofluidic channel, captured three times in identical conditions to assess its structural consistency and fabrication precision. The cross-sectional shape is approximated as a trapezoid, with measured areas of 1818, 1974,

and 1880 nm², yielding an average of 1891 nm². These measurements provide critical insights into the uniformity of nanochannel fabrication, which directly impacts nanofluidic experiments.

The precise determination of the nanochannel cross-sectional area is essential for accurately estimating molecular weight in the NSM experiments. The integrated optical contrast (iOC), a key parameter in molecular weight calculations, is again inversely proportional to the nanochannel area. Consequently, variations in the channel dimensions affect iOC normalization and can introduce systematic errors in molecular weight determination. For insulin, with a molecular weight of 5.8 kDa, the accuracy of iOC measurements is particularly important due to its relatively weak scattering signal. A larger-than-expected nanochannel cross-section would reduce iOC and lead to an overestimated molecular weight, while a smaller cross-section would result in an underestimated molecular weight. Given that the measured nanochannel areas exhibit a variance of approximately 8% around the mean of 1890 nm², this variability must be accounted for in the NSM analysis pipeline to ensure reliable molecular weight estimations.

By incorporating these nanochannel measurements into data processing and calibration procedures, systematic biases can be minimized, thereby improving the reproducibility and accuracy of molecular weight determination for small biomolecules such as insulin. Furthermore, the results highlight the necessity of precise nanochannel fabrication and post-experimental corrections to maintain the fidelity of single-molecule analysis at the lower detection limits of NSM.

8.1 Dependence on channel cross-sectional area

To explore how the model’s performance depends on channel cross-sectional area, we use the same simulated dataset as above and go through a few extra steps. Firstly, we consider only predictions of simulated trajectories of length $N = 8192$ frames and (linearly) interpolate values of RME (RSD) as function of optical contrast to generate a continuous function of prediction RME (RSD) as function of optical contrast. The interpolation is achieved using SciPy [16]. Secondly, we consider nanochannel areas linearly dispersed between 50x20 and 50x120 nm, and re-calibrate the effective conversion between iOC and MW and (assuming hindered globular proteins) HR.

The cross-sectional area of the nanochannel was determined using high-resolution scanning electron microscopy (SEM) images, with the boundaries of the trapezoid identified by analyzing pixel intensity differences. To segment the nanochannel from the surrounding material, the *Otsu thresholding algorithm* was applied to the intensity histogram of the image. This algorithm calculates an optimal intensity threshold by maximizing the variance between two pixel intensity classes: the bright boundary of the channel and the darker interior material. Specifically, Otsu’s method computes the intensity threshold T that minimizes the intra-class variance, defined as:

$$\sigma_w^2(T) = q_1(T)\sigma_1^2(T) + q_2(T)\sigma_2^2(T),$$

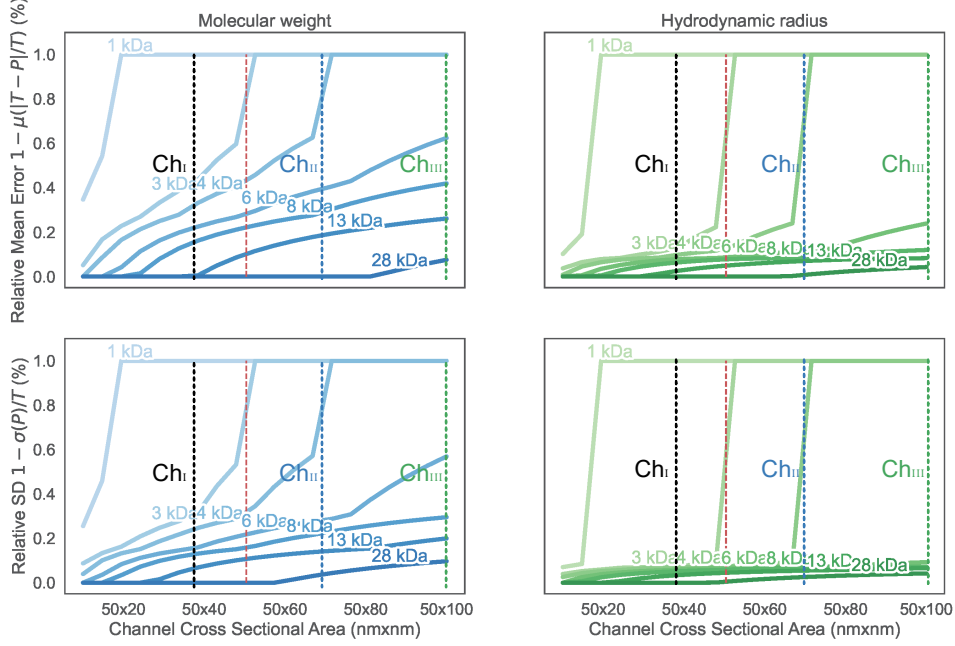


Figure 26: Accuracy and precision of the model as a function of channel cross-sectional area. The figure illustrates the relative mean error and relative standard deviation for both molecular weight (left panels) and hydrodynamic radius (right panels) predictions across different channel sizes. The plots show how the model's performance varies with the cross-sectional area of the nanochannels, highlighting the impact of channel geometry on the estimation accuracy and precision. The dashed vertical lines in black blue and green indicate specific channel sizes (Ch_I , Ch_{II} , Ch_{III}), respectively, used in the study. The red dotted line corresponds to the size of Channel 1 in [3], used to characterize the properties of different proteins.

where $q_1(T)$ and $q_2(T)$ are the probabilities of the pixel intensities being below and above T , respectively, and $\sigma_1^2(T)$ and $\sigma_2^2(T)$ are the variances of those classes. The resulting threshold T was used to binarize the SEM image, separating pixels into either the channel region or the background. The resulting binary image was overlaid onto the original SEM image for manual inspection to verify that the detected boundaries corresponded accurately to the trapezoidal geometry of the nanochannel.

To calculate the trapezoid area, the boundary points were extracted from the binary image using and the trapezoid's area was then calculated using the formula:

$$A = \frac{1}{2}(b_1 + b_2)h,$$

where b_1 and b_2 are the lengths of the parallel sides, and h is the perpendicular height between them. These dimensions were derived by converting the pixel coordinates of the vertices into physical distances using the SEM's resolution calibration factor (e.g., nanometers per pixel).

The variation in cross-sectional area, ranging from 1652 nm² to 1823 nm², arises from several interconnected factors that influence the segmentation and measurement process. While the Otsu thresholding algorithm is deterministic, its output depends on the intensity histogram of the SEM image, which is affected by noise and slight fluctuations in illumination. Noise in the image, stemming from electron beam instability or detector artifacts, can alter the intensity distribution, leading to subtle shifts in the threshold value T and, consequently, the delineation of the channel boundaries.

The edges of the nanochannel, particularly at sloped regions, introduce ambiguity in defining the exact boundary, as pixel intensities in these regions transition gradually rather than sharply. This ambiguity results in minor variations in the placement of the detected edges, especially when the contour-detection algorithm interpolates between pixels to trace continuous boundaries. Such interpolations, while necessary to refine the geometry, can introduce additional uncertainty in the measured dimensions of b_1 , b_2 , and h .

Moreover, the calibration factor used to convert pixel dimensions to physical distances introduces a systematic source of variation. Even minor inaccuracies in the SEM’s magnification settings or calibration standards can propagate through the area calculations, compounding the observed range. Manual inspection of the binary image, used to validate the detected boundaries, also contributes to variability, as human judgment in confirming or adjusting edges can be influenced by noise and resolution limits.

Taken together, these factors—histogram sensitivity, edge ambiguity, resolution limits, and calibration variability—interact to produce the range of 1652 nm² to 1823 nm². This range reflects the inherent uncertainty of the measurement process, providing a robust estimate of the nanochannel’s cross-sectional area. This variability in area estimation directly affects the predicted molecular weight of insulin, as reflected in the red error bars in Figure 27, underscoring the impact of this measurement uncertainty in the estimation of molecular properties in this regime.

9 Impact of Nanochannel Roughness on Scattering Cross Section

Nanochannel surface roughness introduces critical variability in scattering cross sections, influencing molecular weight and hydrodynamic radius predictions. The roughness directly impacts the effective optical path and interaction area, leading to potential deviations from the idealized geometrical models.

9.1 Quantitative Analysis of Roughness Effects

** I Do not have the figures ready here, maybe include in review? **

Experimental observations indicate that nanochannel roughness affects scattering

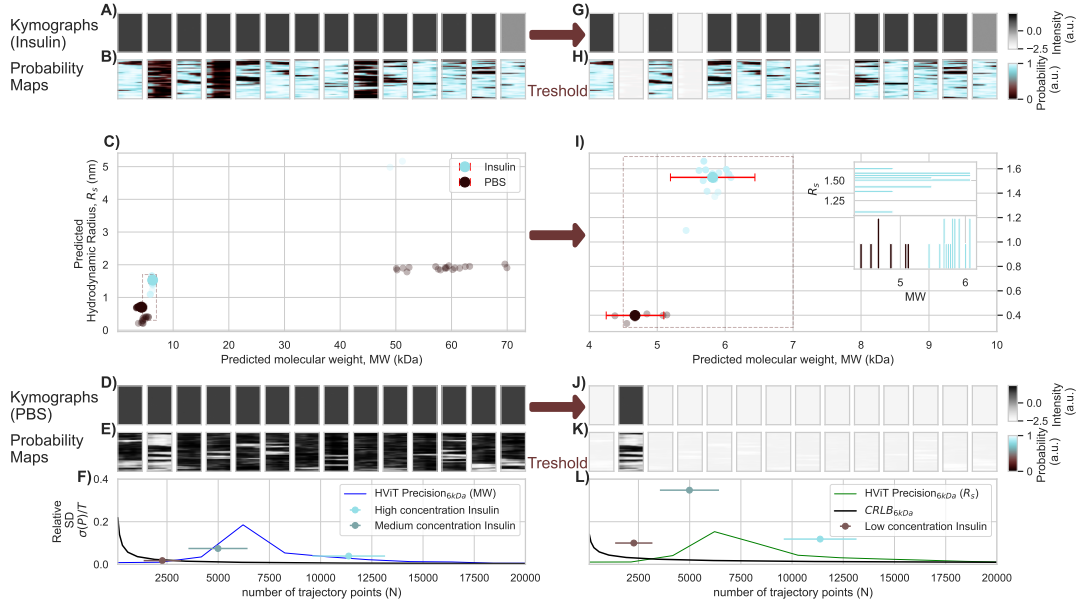


Figure 27: Effect of nanochannel area estimation uncertainty on the inferred properties of Insulin.

intensities non-uniformly. Surface irregularities create localized variations in refractive index, altering the interaction of incident beams with analytes. This effect is captured in the roughness-corrected scattering cross section (σ_r):

$$\sigma_r = \sigma_0(1 + \Delta_r), \quad (7)$$


where σ_0 is the ideal scattering cross section and Δ_r represents the roughness-induced deviation factor, which depends on roughness parameters such as root-mean-square (RMS) height and correlation length.

Figure ?? illustrates examples of nanochannel roughness effects from the provided experimental data. These include variations in scattering cross sections and the corresponding impacts on molecular weight predictions.

9.2 Impact on Molecular Weight and Hydrodynamic Radius Predictions

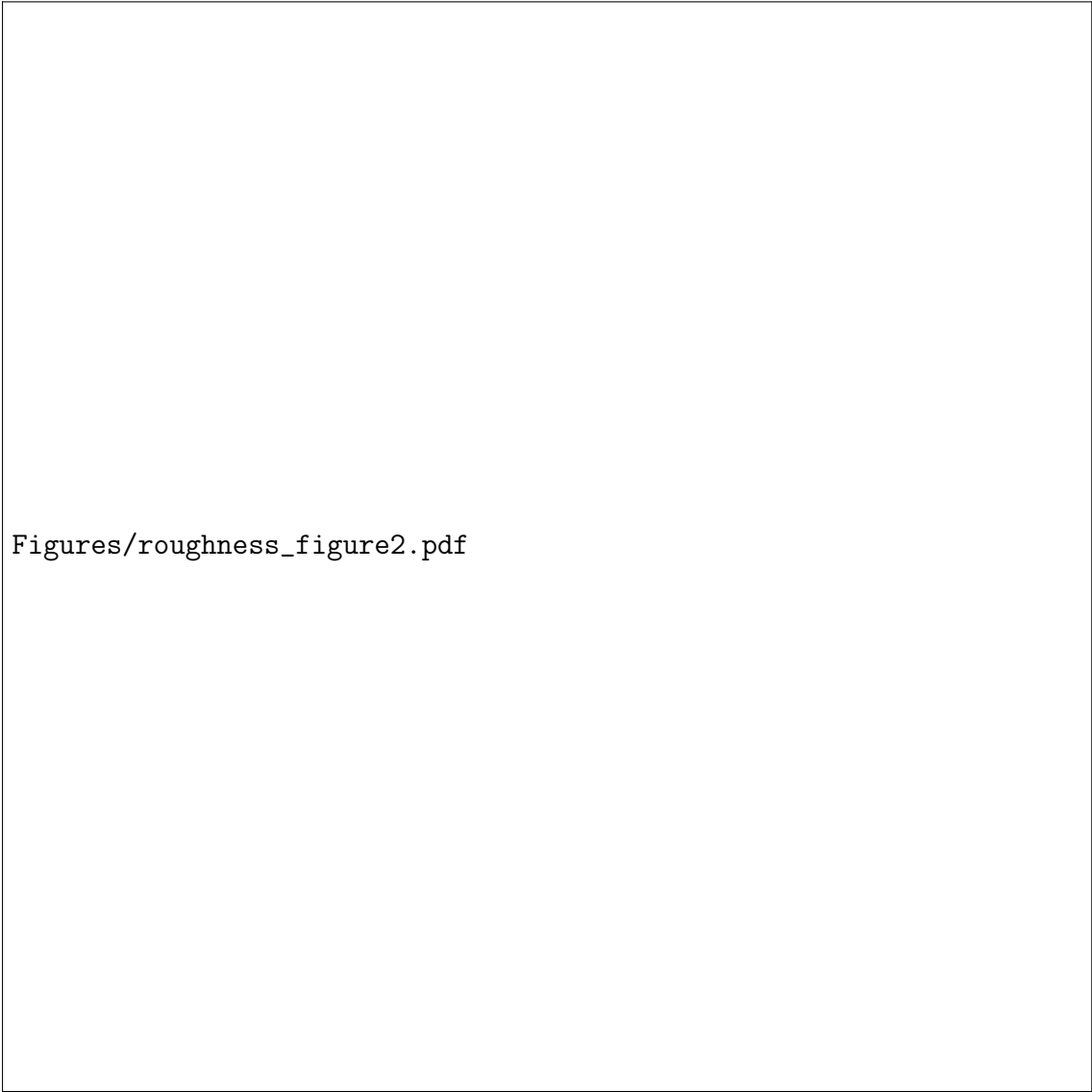
Roughness-induced variations in nanochannel geometry propagate into uncertainties in molecular weight and hydrodynamic radius estimations. Figure ?? demonstrates these effects by comparing prediction errors across channels with varying RMS roughness values, based on the experimental datasets.

The results show a marked increase in prediction errors with higher roughness levels,



Figures/roughness_figure1.pdf

Figure 28: Representative data showcasing the influence of nanochannel roughness. (a) Scattering intensity deviations for channels with increasing roughness. (b) Predicted molecular weight variations as a function of roughness parameters.



Figures/roughness_figure2.pdf

Figure 29: Impact of nanochannel roughness on prediction accuracy from experimental data. (a) Molecular weight prediction errors for channels with RMS roughness values ranging from 1 nm to 10 nm. (b) Corresponding hydrodynamic radius prediction errors. Error bars indicate standard deviations from three independent measurements.

particularly for RMS heights exceeding 5 nm. This trend highlights the importance of minimizing surface irregularities during nanochannel fabrication to ensure reliable analytical outcomes. Moreover, incorporating roughness corrections into predictive models significantly reduces discrepancies, as evidenced by the convergence of corrected predictions towards experimental measurements.

The influence of nanochannel roughness necessitates careful design and manufacturing controls. Techniques such as chemical mechanical polishing and plasma-based surface smoothing have proven effective in reducing roughness. Future work should focus on quantifying the interplay between roughness parameters and scattering effects across different analyte sizes, developing adaptive correction algorithms that dynamically adjust for roughness variations and exploring the impact of nanoscale heterogeneities on multi-channel systems.

By addressing these challenges, the field can advance towards more robust and precise nanochannel-based analytical platforms, further extending their applicability in molecular diagnostics and nanomaterial characterization.

10 Cramér-Rao Lower Bound (CRLB)

The Cramér-Rao Lower Bound (CRLB) provides a theoretical lower bound on the variance of any unbiased estimator of a parameter, indicating the best precision achievable under a given statistical model [17], [18]. It is a fundamental concept in estimation theory and is widely used in statistical signal processing [19]. The CRLB is directly tied to the Fisher information, denoted as $I(\theta)$, which measures the amount of information that an observable random variable conveys about an unknown parameter θ . The mathematical expression for the CRLB is given by:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

where $\text{Var}(\hat{\theta})$ represents the variance of the estimator $\hat{\theta}$. This inequality suggests that the variance of any unbiased estimator cannot be smaller than the inverse of the Fisher information.

It serves as a benchmark for assessing the efficiency of estimators. An estimator that achieves this lower bound is considered efficient, as it has the smallest possible variance among all unbiased estimators for that parameter. The amount of data influences the Fisher information—generally, an increase in data leads to higher Fisher information, which implies a tighter bound and hence, a reduced potential estimation error.

10.0.1 Define the Statistical Model

We begin by specifying the probability density function (pdf) or probability mass function (pmf) of our data, including the scale parameter σ that we aim to estimate.

The choice of model depends on the nature of our data (e.g., normal distribution, exponential distribution).

Example: Consider a normal distribution with mean zero and unknown scale (standard deviation) σ :

$$f(x; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

This model is appropriate for many natural phenomena due to the Central Limit Theorem [20].

10.0.2 Find the Likelihood Function

For a sample of independent and identically distributed observations $X = (X_1, X_2, \dots, X_n)$, we construct the likelihood function by taking the product of individual pdfs:

$$L(\sigma; X) = \prod_{i=1}^n f(x_i; \sigma)$$

Example:

$$L(\sigma; X) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}\right)$$

10.0.3 Compute the Log-Likelihood Function

Next, we take the natural logarithm of the likelihood function to simplify differentiation:

$$\ell(\sigma; X) = \ln L(\sigma; X)$$

Example:

$$\ell(\sigma; X) = -n \ln \sigma - \frac{\sum_{i=1}^n x_i^2}{2\sigma^2} + \text{constant terms}$$

10.0.4 Calculate the First Derivative (Score Function)

We differentiate the log-likelihood function with respect to σ to obtain the score function:

$$\frac{\partial \ell(\sigma; X)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^n x_i^2}{\sigma^3}$$

10.0.5 Compute the Second Derivative

We now differentiate the score function with respect to σ to get the observed information:

$$\frac{\partial^2 \ell(\sigma; X)}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3 \sum_{i=1}^n x_i^2}{\sigma^4}$$

10.0.6 Calculate the Fisher Information $I(\sigma)$

The Fisher Information quantifies the amount of information that our observable data carries about the unknown parameter σ :

$$I(\sigma) = -\mathbb{E} \left[\frac{\partial^2 \ell(\sigma; X)}{\partial \sigma^2} \right]$$

Given that $X_i \sim N(0, \sigma^2)$ and $\mathbb{E}[X_i^2] = \sigma^2$, we have:

$$I(\sigma) = - \left(\frac{n}{\sigma^2} - \frac{3n\sigma^2}{\sigma^4} \right) = \frac{2n}{\sigma^2}$$

This result is consistent with standard statistical texts [21].

10.0.7 Compute the Cramér-Rao Lower Bound

Finally, the CRLB provides the minimum variance bound for any unbiased estimator $\hat{\sigma}$ of σ :

$$\text{Var}(\hat{\sigma}) \geq \frac{1}{I(\sigma)} = \frac{\sigma^2}{2n}$$

This implies that no unbiased estimator of σ can have a variance lower than $\frac{\sigma^2}{2n}$, as established in estimation theory [19].

10.1 Cramér-Rao Lower Bound for NSM

In NSM, the intensity of scattered light from biomolecules diffusing in a nanofluidic channel provides insights into molecular weight and hydrodynamic radius. For NSM, the observed intensity I_t of the scattered light from a biomolecule within a nanochannel can be modeled as:

$$I_t = cI_0L|\alpha_t|^2 \frac{k^3}{4}$$

where:

- $\alpha_t = \alpha_c + \frac{\alpha_m}{L^2}$ represents the total polarizability, composed of the polarizability of the nanochannel α_c and the biomolecule α_m .
- α_m, α_c : Polarizabilities of the biomolecule and the nanochannel.
- I_0 : Incident light intensity.
- k : Wavenumber of the light.
- L : Length of the illuminated part of the nanochannel.
- c : Collection efficiency.

10.1.1 Assumptions for the Likelihood Function

Assume the measured intensity I_t is subject to Gaussian noise. Thus, the probability density function of observing I_t given the parameters α_m is modeled as:

$$f(I_t; \alpha_m) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(I_t - \mu(\alpha_m))^2}{2\sigma^2}\right)$$

where:

$$\mu(\alpha_m) = cI_0L \left(\alpha_c + \frac{\alpha_m}{L^2}\right)^2 \frac{k^3}{4}$$

and σ is the standard deviation of the measurement noise.

The log-likelihood function $\log f(I_t; \alpha_m)$ is:

$$\log f(I_t; \alpha_m) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(I_t - \mu(\alpha_m))^2}{2\sigma^2}$$

10.1.2 Derivation of the Fisher Information

The Fisher Information $I(\alpha_m)$ is defined as the negative expected value of the second derivative of the log-likelihood with respect to the parameter α_m . In the Gaussian noise model with known variance, the Fisher Information simplifies to:

$$I(\alpha_m) = \frac{1}{\sigma^2} \left(\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} \right)^2$$

Compute the first derivative of $\mu(\alpha_m)$ with respect to α_m :

$$\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} = cI_0L \cdot 2 \left(\alpha_c + \frac{\alpha_m}{L^2}\right) \cdot \left(\frac{1}{L^2}\right) \cdot \frac{k^3}{4}$$

Simplify:

$$\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} = \frac{cI_0k^3}{2L} \left(\alpha_c + \frac{\alpha_m}{L^2}\right)$$

Note that $\alpha_t = \alpha_c + \frac{\alpha_m}{L^2}$, so:

$$\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} = \frac{cI_0k^3}{2L} \alpha_t$$

Therefore, the Fisher Information is:

$$I(\alpha_m) = \frac{1}{\sigma^2} \left(\frac{cI_0 k^3}{2L} \alpha_t \right)^2$$

10.1.3 Cramér-Rao Lower Bound (CRLB)

The CRLB provides a lower bound on the variance of any unbiased estimator of α_m :

$$\text{Var}(\hat{\alpha}_m) \geq \frac{1}{I(\alpha_m)} = \frac{\sigma^2}{\left(\frac{cI_0 k^3}{2L} \alpha_t \right)^2}$$

Simplify:

$$\text{Var}(\hat{\alpha}_m) \geq \sigma^2 \left(\frac{2L}{cI_0 k^3 \alpha_t} \right)^2$$

This expression shows that the variance of the estimator depends on the true value of α_t , which includes the parameter α_m we aim to estimate.

10.1.4 Model for Multiple Measurements

For N independent measurements, the total Fisher Information is:

$$I_{\text{total}}(\alpha_m) = N \times I(\alpha_m)$$

Thus, the CRLB for estimating α_m from N measurements is:

$$\text{Var}(\hat{\alpha}_m) \geq \frac{1}{I_{\text{total}}(\alpha_m)} = \frac{\sigma^2}{N} \left(\frac{2L}{cI_0 k^3 \alpha_t} \right)^2$$

10.1.5 CRLB for Diffusivity in NSM

The relationship between the diffusivity D and the hydrodynamic radius R_s is given by the Stokes-Einstein equation:

$$D = \frac{k_B T}{6\pi\eta R_s}$$

Considering the displacement x of a diffusing particle over time t , the probability density function is:

$$f(x; D) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right)$$

The log-likelihood function is:

$$\log f(x; D) = -\frac{1}{2} \log(4\pi Dt) - \frac{x^2}{4Dt}$$

Compute the first derivative with respect to D :

$$\frac{\partial}{\partial D} \log f = -\frac{1}{2D} + \frac{x^2}{4D^2t}$$

Compute the second derivative:

$$\frac{\partial^2}{\partial D^2} \log f = \frac{1}{2D^2} - \frac{x^2}{2D^3t}$$

Calculate the Fisher Information by taking the negative expected value of the second derivative. Since $E[x^2] = 2Dt$, we have:

$$I(D) = -E \left[\frac{\partial^2}{\partial D^2} \log f \right] = \frac{1}{2D^2}$$

Thus, the CRLB for a single measurement is:

$$\text{Var}(\hat{D}) \geq \frac{1}{I(D)} = 2D^2$$

For N independent measurements:

$$I_{\text{total}}(D) = N \times I(D) = \frac{N}{2D^2}$$

So the CRLB becomes:

$$\text{Var}(\hat{D}) \geq \frac{1}{I_{\text{total}}(D)} = \frac{2D^2}{N}$$

10.1.6 Final CRLB Expressions

$$\boxed{\text{Var}(\hat{\alpha}_m) \geq \frac{\sigma^2}{N} \left(\frac{2L}{cI_0 k^3 \alpha_t} \right)^2}$$

$$\boxed{\text{Var}(\hat{D}) \geq \frac{2D^2}{N}}$$

10.2 Including Localization Error in CRLB

So far, we have assumed a theoretically optimal but practically impossible localization error of 0. If we include the theoretical limits of localization error, for n independent displacement measurements over a total observation time $T = n\Delta t$, the CRLB is derived as follows.

Let σ_L be the standard deviation of the localization error. The measured position $x_m(t)$ at time t is:

$$x_m(t) = x(t) + \epsilon$$

where $x(t)$ is the true position, and ϵ is a zero-mean Gaussian random variable with variance σ_L^2 .

The observed displacement Δx_m between two time points separated by Δt is:

$$\Delta x_m = x_m(t + \Delta t) - x_m(t) = \Delta x + \epsilon_2 - \epsilon_1$$

where $\Delta x = x(t + \Delta t) - x(t)$ is the true displacement due to diffusion.

The variance of the observed displacement Δx_m is:

$$\text{Var}(\Delta x_m) = 2D\Delta t + 2\sigma_L^2$$

10.2.1 Likelihood Function and Fisher Information

Assuming that the observed displacements Δx_m are independent and normally distributed, the likelihood function for observing a displacement Δx_m given diffusivity D is:

$$f(\Delta x_m; D) = \frac{1}{\sqrt{4\pi(D\Delta t + \sigma_L^2)}} \exp\left(-\frac{(\Delta x_m)^2}{4(D\Delta t + \sigma_L^2)}\right)$$

The log-likelihood function is:

$$\log f(\Delta x_m; D) = -\frac{1}{2} \log(4\pi(D\Delta t + \sigma_L^2)) - \frac{(\Delta x_m)^2}{4(D\Delta t + \sigma_L^2)}$$

The first derivative with respect to D is:

$$\frac{\partial}{\partial D} \log f = -\frac{\Delta t}{2(D\Delta t + \sigma_L^2)} + \frac{(\Delta x_m)^2 \Delta t}{4(D\Delta t + \sigma_L^2)^2}$$

The second derivative is:

$$\frac{\partial^2}{\partial D^2} \log f = \frac{\Delta t^2}{2(D\Delta t + \sigma_L^2)^2} - \frac{(\Delta x_m)^2 \Delta t^2}{2(D\Delta t + \sigma_L^2)^3}$$

The Fisher information $I(D)$ from a single displacement measurement is the negative expected value of the second derivative:

$$I(D) = -E\left[\frac{\partial^2}{\partial D^2} \log f\right]$$

Since $E[(\Delta x_m)^2] = 2(D\Delta t + \sigma_L^2)$, we have:

$$I(D) = -\left(\frac{\Delta t^2}{2(D\Delta t + \sigma_L^2)^2} - \frac{2(D\Delta t + \sigma_L^2)\Delta t^2}{2(D\Delta t + \sigma_L^2)^3}\right) = \frac{\Delta t^2}{2(D\Delta t + \sigma_L^2)^2}$$

For n independent measurements:

$$I_{\text{total}}(D) = n \times I(D) = \frac{n\Delta t^2}{2(D\Delta t + \sigma_L^2)^2}$$

10.2.2 Cramér-Rao Lower Bound

The CRLB for diffusivity D , considering the localization error, is:

$$\text{Var}(\hat{D}) \geq \frac{1}{I_{\text{total}}(D)} = \frac{2(D\Delta t + \sigma_L^2)^2}{n\Delta t^2}$$

Simplifying:

$$\text{Var}(\hat{D}) \geq \frac{2}{n} \left(\frac{D\Delta t + \sigma_L^2}{\Delta t} \right)^2$$

The localization precision σ_L is related to the measurement noise σ and the experimental parameters:

$$\sigma_L = \frac{\sigma}{\sqrt{cI_0L|\alpha_t|^2 \frac{k^3}{4}}}$$

Substituting σ_L into the CRLB expression:

$$\sigma_L^2 = \left(\frac{\sigma}{\sqrt{cI_0L|\alpha_t|^2 \frac{k^3}{4}}} \right)^2 = \frac{4\sigma^2}{cI_0L|\alpha_t|^2 k^3}$$

Thus, the CRLB becomes:

$$\text{Var}(\hat{D}) \geq \frac{2}{n} \left(\frac{D\Delta t + \frac{4\sigma^2}{cI_0L|\alpha_t|^2 k^3}}{\Delta t} \right)^2$$

Simplify the numerator inside the parentheses:

$$D\Delta t + \frac{4\sigma^2}{cI_0L|\alpha_t|^2 k^3} = \Delta t \left(D + \frac{4\sigma^2}{cI_0L|\alpha_t|^2 k^3 \Delta t} \right)$$

Therefore, the CRLB simplifies to:

$$\text{Var}(\hat{D}) \geq \frac{2}{n} \left(D + \frac{4\sigma^2}{cI_0L|\alpha_t|^2 k^3 \Delta t} \right)^2$$

10.2.3 Final CRLB Expression for Diffusivity in NSM

The modified CRLB for diffusivity D in NSM, incorporating the localization precision, is:

$$\boxed{\text{Var}(\hat{D}) \geq \frac{2}{n} \left(D + \frac{4\sigma^2}{cI_0L|\alpha_t|^2 k^3 \Delta t} \right)^2}$$

The term $\frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3\Delta t}$ represents the influence of localization error on the variance of \hat{D} . Reducing measurement noise σ or increasing the detected signal improves localization precision, thus reducing this term. Increasing Δt reduces the impact of localization error relative to diffusion, enhancing estimation precision. The variance decreases inversely with the number of measurements, emphasizing the benefit of collecting more data.

In the absence of localization error ($\sigma_L^2 = 0$), the CRLB reduces to:

$$\text{Var}(\hat{D}) \geq \frac{2D^2}{n}$$

which matches the standard result derived without considering localization error.

References

- [1] M. Piliarik and V. Sandoghdar, “Direct optical sensing of single unlabelled proteins and super-resolution imaging of their binding sites,” *Nature Communications*, vol. 5, no. 1, p. 4495, Jul. 2014, ISSN: 2041-1723. DOI: [10.1038/ncomms5495](https://doi.org/10.1038/ncomms5495). [Online]. Available: <https://doi.org/10.1038/ncomms5495> (cit. on p. 1).
- [2] M. Piliarik and V. Sandoghdar, “Direct optical sensing of single unlabelled proteins and super-resolution imaging of their binding sites,” *Nature communications*, vol. 5, no. 1, pp. 1–8, 2014 (cit. on p. 1).
- [3] B. Špačková, H. Klein Moberg, J. Fritzsche, *et al.*, “Label-free nanofluidic scattering microscopy of size and mass of single diffusing molecules and nanoparticles,” *Nature Methods*, vol. 19, no. 6, pp. 751–758, Jun. 2022, ISSN: 1548-7105. DOI: [10.1038/s41592-022-01491-6](https://doi.org/10.1038/s41592-022-01491-6). [Online]. Available: <https://doi.org/10.1038/s41592-022-01491-6> (cit. on pp. 1, 2, 35, 38).
- [4] P. Dechadilok and W. M. Deen, “Hindrance factors for diffusion and convection in pores,” *Industrial & engineering chemistry research*, vol. 45, no. 21, pp. 6953–6959, 2006 (cit. on pp. 1, 9).
- [5] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *Journal of Big Data*, vol. 3, no. 1, p. 9, May 2016, ISSN: 2196-1115. DOI: [10.1186/s40537-016-0043-6](https://doi.org/10.1186/s40537-016-0043-6). [Online]. Available: <https://doi.org/10.1186/s40537-016-0043-6> (cit. on p. 6).
- [6] M. D. Zeiler and R. Fergus, *Visualizing and understanding convolutional networks*, 2013. DOI: [10.48550/ARXIV.1311.2901](https://arxiv.org/abs/1311.2901). [Online]. Available: <https://arxiv.org/abs/1311.2901> (cit. on p. 6).
- [7] M. Csikszentmihalyi, “Flow: The psychology of optimal experience,” in Jan. 1990 (cit. on p. 7).
- [8] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham: Springer International Publishing, 2014, pp. 818–833, ISBN: 978-3-319-10590-1 (cit. on p. 7).

- [9] B. Settles, “Active learning literature survey,” 2009 (cit. on p. 7).
- [10] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML ’09, Montreal, Quebec, Canada: Association for Computing Machinery, 2009, pp. 41–48, ISBN: 9781605585161. DOI: [10.1145/1553374.1553380](https://doi.org/10.1145/1553374.1553380). [Online]. Available: <https://doi.org/10.1145/1553374.1553380> (cit. on p. 7).
- [11] B. Midtvedt, S. Helgadottir, A. Argun, J. Pineda, D. Midtvedt, and G. Volpe, “Quantitative digital microscopy with deep learning,” *Applied Physics Reviews*, vol. 8, no. 1, p. 011310, Feb. 2021, ISSN: 1931-9401. DOI: [10.1063/5.0034891](https://doi.org/10.1063/5.0034891). eprint: https://pubs.aip.org/aip/apr/article-pdf/doi/10.1063/5.0034891/14577703/011310_1_online.pdf. [Online]. Available: <https://doi.org/10.1063/5.0034891> (cit. on p. 9).
- [12] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014. DOI: [10.48550/ARXIV.1412.6980](https://arxiv.org/abs/1412.6980). [Online]. Available: <https://arxiv.org/abs/1412.6980> (cit. on p. 9).
- [13] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. LNCS, (available on arXiv:1505.04597 [cs.CV]), vol. 9351, Springer, 2015, pp. 234–241. [Online]. Available: <http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a> (cit. on p. 10).
- [14] L. Roeder, *Netron*, <https://github.com/lutzroeder/netron>, 2020 (cit. on p. 10).
- [15] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, “Phasestain: The digital staining of label-free quantitative phase microscopy images using deep learning,” *Light: Science & Applications*, vol. 8, no. 1, p. 23, Feb. 2019, ISSN: 2047-7538. DOI: [10.1038/s41377-019-0129-y](https://doi.org/10.1038/s41377-019-0129-y). [Online]. Available: <https://doi.org/10.1038/s41377-019-0129-y> (cit. on p. 10).
- [16] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2) (cit. on p. 37).
- [17] H. Cramér, *Mathematical Methods of Statistics*. Princeton University Press, 1946 (cit. on p. 43).
- [18] C. R. Rao, “Information and the accuracy attainable in the estimation of statistical parameters,” *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945 (cit. on p. 43).
- [19] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993 (cit. on pp. 43, 45).
- [20] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd. John Wiley & Sons, 1968, vol. 1 (cit. on p. 44).
- [21] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd. Springer, 1998 (cit. on p. 45).