# A Hybrid Machine Learning and Analytical Model of BioFETs for the Detection of Peptides

Nicholas Smoliak*, Naveen Kumar*, Pedro Miguel Parreira†, Craig Macdonald‡, Vihar Georgiev*
*Deep Nano Research Group, University of Glasgow, UK
†School of Physics and Astronomy, University of Glasgow, UK
‡School of Computing Science, University of Glasgow, UK
n.smoliak.1@research.gla.ac.uk, naveen.kumar@glasgow.ac.uk, vihar.georgiev@glasgow.ac.uk

***Abstract*—BioFETs enable the detection of a diverse range of biological analytes by transducing changes in surface potential into a measurable electrical signal. Although established analytical models, including the Gouy–Chapman–Stern and site-binding frameworks, effectively describe surface charge and protonation equilibria, they do not account for the sequence-specific chemistry, steric effects, or conformational diversity characteristic of peptides and other biomolecules. Here we present a hybrid model to address these limitations, augmenting previous BioFET simulation frameworks with molecular parameters including proton dissociation constants and topological surface areas determined through molecular modelling. While this work focuses on peptides, our framework extends to modified peptides, diverse biomolecular analytes and immobilization chemistries relevant to biosensing. We validate our model against experimental values and molecular modelling studies, confirming the physical fidelity of our hybrid model and perform simulations of representative peptides demonstrating the model's capacity to resolve fine chemical variations, providing a practical tool to guide experimental design and contribute to the development of BioFETs. We further identify notable areas for future work where we will engage with more rigorous modelling, utilizing our model outputs to guide experimental design and further validate our device modelling framework.**

***Index Terms*—Biosensor, machine learning, device modelling, analytical simulation**

## I. Introduction

BioFETs are increasingly employed for the detection of a broad spectrum of biomolecular analytes spanning peptides, proteins, nucleic acids, and other small biologically relevant molecules, with each presenting distinct charge, size, and structural properties at the sensing interface [1, 2]. As label-free sensors, BioFETs transduce their signal without fluorescent or enzymatic tags, thereby reducing sample-preparation complexity and time while preserving the native state of the analyte [3]. These advantages underpin their suitability for miniaturised, rapid diagnostic systems in clinical and point-of-care settings and have led to substantial interest in furthering their development [2, 3].

Modelling frameworks are an essential tool for the development of these BioFETs, enabling predictions that are computationally tractable yet retain sufficient physical detail to guide device analysis, optimization, and design. Without such approximate models, exhaustive atomistic simulations or purely experimental approaches are impractical, resource-intensive and too slow for timely screening of the vast chemical and structural diversity inherent to biomolecules [4, 5].

The principal innovation of this work is twofold: first, we implement a ML-driven molecular model to predict peptide-specific proton dissociation constants ($k_i$, describing functional-group specific protonation equilibria) and topological surface area ($TSA$, a proxy for steric footprint of the molecule on the sensor surface) [6]; second, we integrate these molecular parameters into a combined GCS/site-binding surrogate model for BioFETs, as shown in Fig 1. This approach enables the model to account for the unique chemistry, steric properties, and conformational properties that define the response of biomolecular analytes at the sensor interface [2]. These features which influence the BioFET signal, are not included in conventional models based on static, residue-level parameters [7]. Our framework builds on a combined site-binding and Gouy–Chapman–Stern (GCS) framework for pH-dependent titration and diffuse layer

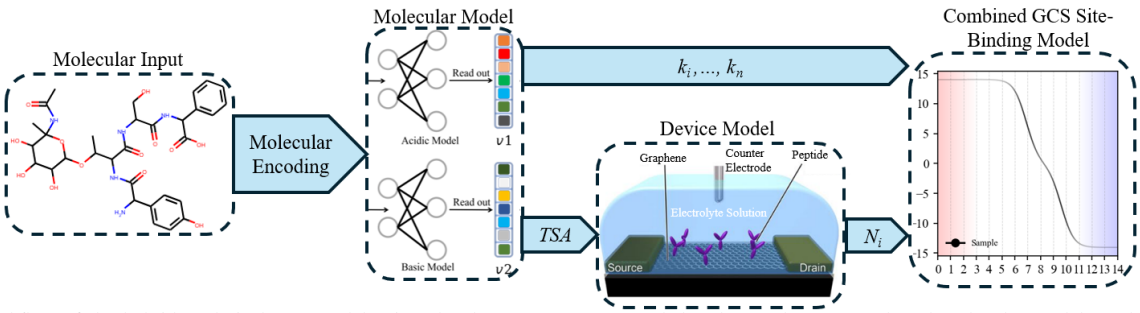


Fig. 1: Workflow of the hybrid analytical–ML model. Biomolecules are SMILES-encoded and passed to an ML-based molecular model (MolGpka) [6] to predict $k_i$ and $TSA$ values, which are then integrated into the combined site-binding and GCS analytical model to compute the transduced signal measured by the BioFET. Image components adapted from [6] and [1].

potential [7, 8], incorporating peptide-specific $k_i$ and $TSA$ directly into the analytical solution. This hybrid approach treats the biomolecular analyte as an integrated molecular system, rather than as a collection of isolated titratable groups, aligning the modelling framework with experimental evidence that immobilisation chemistry, charge distribution, and steric footprint influence BioFET signals [2, 7].

To our knowledge, this constitutes a novel union of molecular and analytical modelling, resulting in a computationally efficient, physically consistent model capable of discriminating sequence-level variation in BioFET response. The hybrid framework generalises readily to chemically modified peptides, other molecular targets, and varied surface chemistries. We validate our model through comparison of modelled $k_i$ values for alanine, glycine, and valine homopeptides with experimental values [9], and verification of $TSA$ trends against independent amino acid adsorption studies [10–12]. We then apply our model to a representative peptide of medical interest to highlight our hybrid model's capabilities.

## II. Methodology

BioFETs detect immobilised peptides by transducing changes in surface charge: charged functional groups from the immobilised peptide in the diffuse layer of the electrolyte alter the electric potential ($\Psi_0$), modulating the device's channel conductance and producing a measurable signal [1, 2, 8]. The charge states of the peptide's titratable sites are determined by their $k_i$ values, while the density of these charges is influenced by $TSA$. Together, these affect the modulation of $\Psi_0$ as a function of $pH$ [7]. Because the molecular parameters that define the BioFET response are intrinsic to each biomolecule, BioFETs can detect analytes directly without modifying the analyte from its native state - hence their label-free designation [3]. These combined effects of acid–base equilibria and analyte–surface interactions yield a BioFET response characteristic of the peptide sequence or other biomolecular analyte [1, 2].

To model the BioFET response, we implement an approximate analytical framework that combines the site-binding and GCS models, treating peptide protonation and the electrical double layer in a self-consistent manner, and linking the molecular parameters of the immobilised peptide to the measurable $\Psi_0$ [7]. In this framework, the charge contributed by each titratable group is determined by its $k_i$, while the total surface charge depends on the number of accessible sites, $N_i$, scaled according to the peptide's $TSA$. This site-binding charge acts as a boundary condition to solve the Poisson–Boltzmann equation for the diffuse layer, allowing electroneutrality to be maintained at the interface. Our method iteratively solves this coupled electrostatic and chemical system until a self-consistent solution is reached for both $\Psi_0$ and the surface protonation states at each point along the $pH$ titration. As a result, the sequence-specific molecular parameters $k_i$ and $TSA$ are directly linked to the measurable BioFET signal, converting the generic GCS/site-binding framework into the molecule-specific model illustrated in Fig. 1.

## III. Results and Discussion

### *Conformational Effects and $TSA$ Validation*

The $TSA$ parameter used in this work is a widely adopted descriptor (sometimes referred to as $TPSA$ depending on the context of the literature) that estimates molecular surface polarity by summing fragment-based contributions from polar atoms, without requiring computationally rigorous structural conformation predictions [13]. While true conformational effects reflect the three-dimensional arrangement of atoms in solution or at interfaces, our use of $TSA$ provides a tractable, two-dimensional surrogate that encapsulates the projected molecular footprint relevant for steric effects at the sensor surface. The $TSA$ parameter thus serves as a proxy for the footprint of the immobilized peptide at the sensor interface. In our work, peptides with larger $TSA$ occlude more surface sites, reducing the number of available surface sites $N_i$, and therefore the concentration of titratable functional groups and magnitude of $\Psi_0$.

We use an experimental site density of $N_0 = 1\times10^{14}\ \mathrm{cm}^{-2}$, corresponding to a per-peptide footprint of $A_0 = 100\ \text{Å}^2$ from previous characterizations [7]. Peptides with $TSA > A_0$ are treated as blocking adjacent sites, reducing $N_i$ as follows: $N_i = N_0$ if $A \leq A_0$; $N_i = N_0(A_0/A)$ if $A > A_0$, where $A$ is the predicted $TSA$. This rescaling propagates through the self-consistent solution of the model, narrowing the range of $\Psi_0$ for peptides with increasingly high steric footprints which exceed $A_0$.

We confirm the use of $TSA$ as a parameter describing the steric and conformational attributes of a peptide on a sensor surface by considering its alignment with trends across three amino acid–graphene adsorption studies [10–12]. Notably, these studies investigate the adsorption behaviour of amino acids and peptides directly on graphene surfaces under aqueous conditions. This provides a relevant foundation for assessing the impact of molecular footprint and conformational features on sensor response. In our model, the effects of conformation such as molecular orientation and projected area are implicitly captured through $TSA$, which is then used to compute steric site blocking at the interface. We

TABLE I: Comparison of strongest and weakest binding residues across this work and prior amino acid-graphene adsorption studies

| Study | Metric | Strongest | Weakest |
|---|---|---|---|
| **This work** | $TSA$ (Å$^2$) | Arg > Gln > Tyr ≈ Thr ≈ Ser ≈ Phe | Gly ≈ Ala ≈ Val |
| Hughes & Walsh [10] | $\Delta G_{\text{ads}}$ (kJ mol$^{-1}$) | Arg > Gln > Tyr ≈ Gly ≈ Phe | Val < Ala ≈ Thr ≈ Ser |
| Welch *et al.* [11] | $\Delta G_{\text{bind}}$ (kcal mol$^{-1}$) | Arg > Tyr ≈ Gln | Val < Ser ≈ Thr ≈ Gly ≈ Phe |
| Dragneva *et al.* [12] | $E_{\text{ads}}$ (eV) | Tyr > Phe ≈ Arg | Ala ≈ Thr ≈ Ser ≈ Gly < Val |

*Note:* Larger $TSA$ values or more negative $\Delta G/E_{\text{ads}}$ denote stronger binding. Residues shown correspond to those ranked at the strongest and weakest binding extremes. For context, Ala is omitted in [11] as it occupies an intermediate range, and similarly, Gln is omitted in [12].

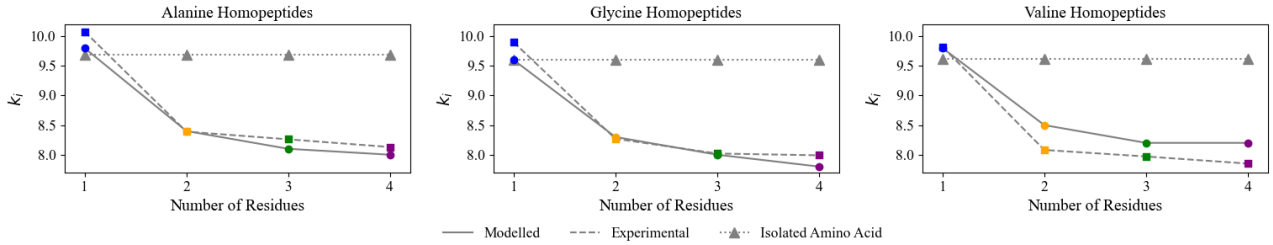


Fig. 2: Comparison of modelled $k_i$ values (solid lines) with experimental data (dashed lines) for alanine, glycine, and valine homopeptides of 1–4 residues, based on capillary zone electrophoresis measurements [9]. Dotted grey lines represent static, tabulated $k_i$ values used previously [7]. For the alanine homopeptides, we use the $k_i$ values here in the BioFET simulations of Fig. 3, where, for a monoprotic titration, these $k_i$ values correspond to the inflection points (vertical lines) in the $\Psi_0$–$pH$ plots.

summarise their findings in Tab. I, where, in all cases, residues exhibiting strong adsorption such as arginine, tyrosine, and phenylalanine show high $TSA$ values in this work, while weakly binding residues such as valine, alanine, and glycine align with a low $TSA$. This agreement across the diverse metrics of $\Delta G_{\text{ads}}$[10], $\Delta G_{\text{bind}}$[11], and $E_{\text{ads}}$ [12] provides physical justification for using TSA as an interpretable proxy for peptide conformation and site blocking, capturing both the molecular footprint and the tendency of peptides to compete for and occupy space at the BioFET interface.

*Experimental Validation and Simulation of Surface Chemistry*

Further validating our framework, we compare our modelled $k_i$ values for alanine, glycine, and valine homopeptides with experimental data obtained from capillary zone electrophoresis [9]. Fig. 2 presents this comparison for homopeptides of 1–4 residues, where the solid lines indicate modelled $k_i$ values, dashed lines show experimental measurements, and the dotted grey line denotes sequence-invariant, tabulated $k_i$ values for isolated amino acids used in previous implementations [7]. Our modelled values show strong agreement with experiment, with mean absolute errors of 0.15 (alanine), 0.14 (glycine), and 0.25 (valine).

To demonstrate how our model simulates BioFET signals, Fig.3 presents $\Psi_0$–$pH$ curves for alanine homopeptides of 1–4 residues, each immobilised on a pyrene-functionalised graphene surface. The left panel shows the simulated surface potential as a function of $pH$ for each peptide length, while the right panel displays the corresponding molecular model for the pyrene–tetrapeptide conjugate, highlighting the pyrene anchor and terminal titratable amine groups. As the number of alanine residues increases, two key effects are observed: first, the amplitude of the $\Psi_0$–$pH$ curve decreases, reflecting enhanced steric site blocking due to increased $TSA$. Second, the inflection points in each curve shift to lower $pH$ in line with the $k_i$ trends validated in Fig. 2. Each curve's colour matches the corresponding $k_i$ value from Fig. 2. For comparison, the grey plot shows the outcome from previous models using static, tabulated values, which do not distinguish between homopeptides of different lengths.

Functionalisation of the graphene surface is essential for immobilising peptides onto a BioFET [2, 7]. By explicitly incorporating surface chemistries into our model, we move beyond isolated molecular predictions and account for the surface chemistry of a BioFET. While our model remains device-

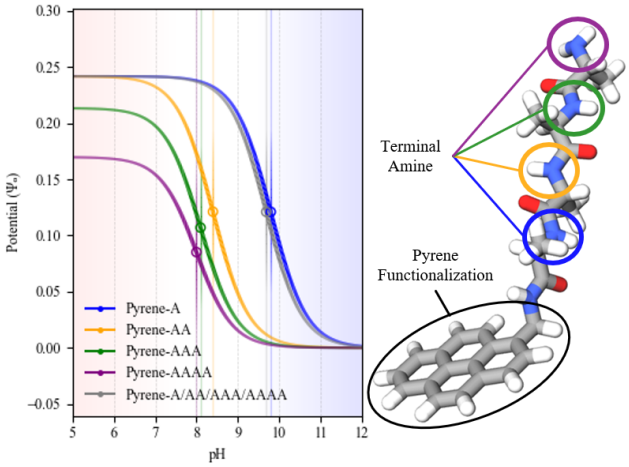


Fig. 3: Simulated BioFET plots for alanine homopeptides (1–4 residues), labelled as Pyrene-A, Pyrene-AA, Pyrene-AAA, and Pyrene-AAAA, each immobilised on pyrene-functionalised graphene (left). The plots show progressive attenuation of $\Psi_0$ amplitude and shifts in inflection points with increasing residue count. The grey curve, labelled Pyrene-A/AA/AAA/AAAA, shows the result for all peptide lengths using static, tabulated $k_i$ values as in previous models; these curves are identical and overlap because previous models did not include sequence-specific parameters. The right panel presents a molecular model of the pyrene–tetrapeptide conjugate, highlighting the pyrene anchor and terminal titratable amine sites. Colour coding links each peptide length to its corresponding $k_i$ value for the homopeptide of that residue count from Fig. 2.

independent by predicting molecular parameters directly from SMILES-encoded inputs, we demonstrate this workflow with a pyrene-functionalised system, as pyrene is widely used for graphene-based biosensing applications [14]. Importantly, our framework can also account for further chemical modifications to the peptide itself, such as post-translational modifications, synthetic modifications, which influence the resulting BioFET signal [2, 7]. By calculating the relevant $k_i$ and $TSA$ parameters for any valid input structure, our approach enables the simulation of a wide variety of analytes and surface chemistries - a notable contribution of our framework.

*Sequence Discrimination and Medical Applications*

We further demonstrate the sequence-level discrimination capabilities of our model using variants of the KYD motif, a peptide of medical interest recognised by specific antibodies for affinity tagging and detection [15]. Fig. 4 presents simulated BioFET responses for KYD, alongside AKYD, DKYD, NYD, and DNYD - derivatives of the KYD motif - where each curve reflects unique combinations of protonation and steric

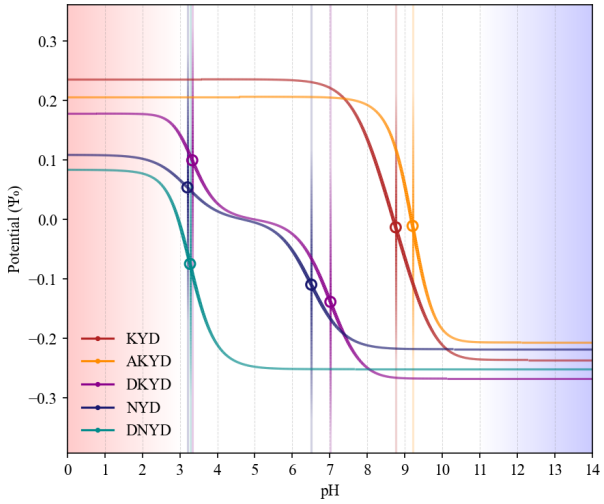


Fig. 4: Simulated BioFET responses for five peptides based on the KYD motif: KYD, AKYD, DKYD, NYD, and DNYD showing $\Psi_0$ as a function of $pH$. Vertical lines indicate $pH$ values at which major charge transitions occur, derived from ML-predicted $k_i$ values. The hybrid model captures both sequence-specific protonation behaviour and conformational differences between peptides (via $TSA$), yielding separable $\Psi_0$ curves.

parameters related to the specific amino acid sequence. The model yields clearly distinguishable $\Psi_0$-$pH$ profiles and can additionally accommodate for analytes with numerous titratable groups, highlighting its ability to resolve single-residue changes relevant to medical and diagnostic applications.

## IV. Future Work

Building on the present work, future efforts will refine our model of the peptide–electrolyte interface by explicitly accounting for the spatial distribution of molecular charges and their screening by the surrounding electrolyte. This will enhance the physical fidelity of device simulations, especially for peptides with complex conformations or charge arrangements. In parallel, we aim to systematically validate model predictions through close collaboration with experimental partners, comparing simulated responses with sensor data from ongoing device fabrication and measurement efforts within our consortium. Such validation will provide iterative feedback to refine both the modelling assumptions and device protocols. In addition, we aim to develop inverse modelling approaches to infer properties of the biomolecular analyte from BioFET sensor outputs, with this being a key contribution which will encourage the application of BioFETs for medical applications.

## V. Conclusion

This work demonstrates an hybrid ML-analytical model for simulating peptide-specific BioFET responses. By incorporating ML-predicted $k_i$ values and $TSA$-based scaling of accessible surface sites into a GCS/site-binding framework, we achieve increased physical fidelity and sequence-level resolution compared to prior analytical models based on static, tabulated molecular parameters.

Validation against experimentally measured $k_i$ values for alanine, glycine, and valine homopeptides (Fig. 2) confirmed strong agreement between predicted and measured values. Similarly, $TSA$-based scaling of $N_i$ reproduced trends observed in three independent adsorption studies (Tab. I), supporting the physical interpretability of the model's molecular descriptors. Ongoing work will focus on enhancing the physical fidelity of the peptide–electrolyte model, experimental validation, and the development of inverse methods for extracting analyte properties from experimental BioFET signals.

This hybrid model is not intended to replace atomistic simulations or empirical experiments but offers a computationally efficient, physically grounded approach that bridges molecular-level properties and device-scale response. By resolving biochemical variations, including chemical modifications and specific surface chemistries, the model provides a practical tool to contribute to the development of BioFETs for biomolecular sensing.

## References


[1] D. Sung and J. Koo, "A review of BioFET's basic principles and materials for biomedical applications," *Biomed. Eng. Lett.*, vol. 11, no. 2, pp. 85–96, Apr. 2021.

[2] S. Szunerits, T. Rodrigues, R. Bagale, H. Happy, R. Boukherroub, and W. Knoll, "Graphene-based field-effect transistors for biosensing: Where is the field heading to?," *Anal. Bioanal. Chem.*, vol. 416, no. 9, pp. 2137–2150, Apr. 2024.

[3] V. R. Samuel and K. J. Rao, "A review on label free biosensors," *Biosens. Bioelectron. X*, vol. 11, p. 100216, Sep. 2022.

[4] D. Passeri, A. Morozzi, K. Kanxheri, and A. Scorzoni, "Numerical simulation of ISFET structures for biosensing devices with TCAD tools," *Biomed. Eng. Online*, vol. 14, suppl. 2, p. S3, Aug. 2015.

[5] F. Pasadas, T. El Grour, E. G. Marin, A. Medina-Rull, A. Toral-Lopez, J. Cuesta-Lopez, F. G. Ruiz, L. El Mir, and A. Godoy, "Compact modeling of two-dimensional field-effect biosensors," *Sensors*, vol. 23, no. 4, p. 1840, Jan. 2023.

[6] X. Pan, H. Wang, C. Li, J. Z. H. Zhang, and C. Ji, "MolGpka: A Web Server for Small Molecule pKa Prediction Using a Graph-Convolutional Neural Network," *J. Chem. Inf. Model.*, vol. 61, no. 7, pp. 3159–3165, Jul. 2021.

[7] N. Kumar, R. P. S. Dhar, C. P. García, and V. Georgiev, "Discovery of Amphoteric Fingerprints of Amino Acids with Field-Effect Transistors," *IEEE Access*, vol. 12, p. 3411168, Jul. 2024.

[8] R. E. G. van Hal, J. C. T. Eijkel, and P. Bergveld, "A general model to describe the electrostatic potential at electrolyte–oxide interfaces," *Adv. Colloid Interface Sci.*, vol. 69, no. 1, pp. 31–62, Dec. 1996.

[9] R. Plasson and H. Cottet, "Determination and Modeling of Peptide pKa by Capillary Zone Electrophoresis," *Anal. Chem.*, vol. 78, no. 15, pp. 5394–5402, Aug. 2006.

[10] Z. E. Hughes and T. R. Walsh, "What makes a good graphene-binding peptide? Adsorption of amino acids and peptides at aqueous graphene interfaces," *J. Mater. Chem. B*, vol. 3, no. 16, pp. 3211–3221, Apr. 2015.

[11] C. M. Welch, A. N. Camden, S. A. Barr, G. M. Leuty, G. S. Kedziora, and R. J. Berry, "Computation of the binding free energy of peptides to graphene in explicit water," *J. Chem. Phys.*, vol. 143, no. 4, p. 045104, Jul. 2015.

[12] N. Dragneva, W. B. Floriano, D. Stauffer, R. C. Mawhinney, G. Fanchini, and O. Rubel, "Favorable adsorption of capped amino acids on graphene substrate driven by desolvation effect," *J. Chem. Phys.*, vol. 139, no. 17, p. 174711, Nov. 2013.

[13] S. Prasanna and R. J. Doerksen, "Topological polar surface area: a useful descriptor in 2D-QSAR," *Curr. Med. Chem.*, vol. 16, no. 1, pp. 21–41, 2009.

[14] V. Mishyn *et al.*, "The holy grail of pyrene-based surface ligands on the sensitivity of graphene-based field effect transistors," *Sensors Diagn.*, vol. 1, no. 2, pp. 235–244, Mar. 2022.

[15] J. W. Slootstra, D. Kuperus, A. Plückthun, and R. H. Meloen, "Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5," *Mol. Divers.*, vol. 2, no. 3, pp. 156–164, Mar. 1997.