small structures
www.small-structures.com

Deep-Learning-Assisted Affinity Classification for Humoral Immunoprotein Complexes

*Bahar Dadfar, Safoura Vaez, Cristian Haret, Meike Koenig, Tahereh Mohammadi Hafshejani, Matthias Franzreb, and Joerg Lahann**

Immunoglobulins are important building blocks in biology and biotechnology. With the emergence of comprehensive deep-learning approaches, there are enormous opportunities for fast and accurate methods of classification of protein–protein interactions to arise. Herein, widely accessible image classification algorithms for species-specific typification of a range of different immunoglobulin G (IgG) complexes are repurposed. Droplets of various immunoglobulins mixed with a B-cell superantigen (SAg) (recombinant staphylococcal Protein A) are deposited onto hydrophobic polymer substrates and the resulting protein stains are imaged using polarized light microscopy. A comprehensive study based on 23 745 images finds that the pretrained convolutional neural network (CNN) InceptionV3 not only successfully categorizes IgGs from four different species but also predicts their binding affinity to Protein A: averaged over 36 binding pairs, the following are observed: 1) an overall accuracy of 81.4%, 2) the highest prediction accuracy for human IgG, the antibody with the highest binding affinity for Protein A, and 3) that the classification accuracy regarding the various IgG/Protein A ratios generally correlates with the binding strength of the protein–protein–complex as determined via circular dichroism spectroscopy. In addition, the CNN pretrained with IgG/Protein A stain images has been challenged with a new set of images using a different superantigen (SAg, Protein G). Despite the use of the unknown SAg, the CNN correctly classifies the images of human IgG and Protein G as indicated by a 94% accuracy over the various molar binding ratios. These findings are noteworthy because they demonstrate that appropriately pretrained CNNs can be used for the prediction of protein–protein interactions beyond the scope of the original training set. Aided by deep-learning methods, simple stains of mixed protein solutions may serve as accurate predictors of the strength of protein–protein interactions with relevance to protein engineering, self-aggregation, or protein stability in complex media.

1. Introduction

In biological systems, proteins play a vital role due to their varied functions and modes of interactions. Understanding the complexities of protein interactions is essential, because even subtle changes can have a profound effect on their biological function and stability.[1] An arsenal of methods for studying protein–protein interactions exist, such as tandem affinity purification, affinity chromatography, co-immunoprecipitation, protein arrays, fragment complementation, phage display, X-ray crystallography, and nuclear magnetic resonance spectroscopy.[2–4] However, most protein interactions are transient, characterized by small contact zones and moderate conformational changes.[5] Thus, it becomes more challenging to identify weak and nonspecific interactions between proteins. To this end, there are significant opportunities to augment conventional biological methods with machine-learning (ML)-based technologies. Further advancement of artificial intelligence (AI) in this area is expected to contribute to predictive and correlative models that will affect all fields of modern biology.[6,7]

Our previous study investigated the development of simple and accurate methods for classifying single-amino-acid mismatches in proteins using deep-learning

B. Dadfar, S. Vaez, C. Haret, M. Koenig, T. Mohammadi Hafshejani, M. Franzreb, J. Lahann
Karlsruhe Institute of Technology (KIT)
Institute of Functional Interfaces (IFG)
Hermann-von-Helmholtz-Platz1, 76344 Eggenstein-Leopoldshafen, Germany
E-mail: joerg.lahann@kit.edu

J. Lahann
Biointerfaces Institute
University of Michigan
Ann Arbor, MI 48109, USA

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/sstr.202400204.

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/sstr.202400204

approaches.[8] We demonstrated that important information about primary and secondary peptide structures can be deduced from the stains left by drying droplets.[8] Deep-learning neural networks were presented with polarized light microscopy (PLM) images obtained from the drying droplet deposits of a variety of amyloid-β peptides to assess complex stain patterns. A sessile droplet of water or another volatile solvent, which contains nonvolatile solutes or colloidal particles, leaves a nonuniform stain or deposit on a substrate when it dries. This stain pattern is highly characteristic of processes involving heat, momentum, and mass transport within a droplet during evaporation, the behavior of the three-phase contact line, and the interactions of deposited material with the interfaces between the droplet and the substrate and air during the evaporation process.[9–12] Several physical processes influence the pattern of deposited aqueous droplets on a characterized surface, referred to as coffee rings, including protein concentration, buffer, super-hydrophobicity, contact-line motion, Marangoni flow, surface-tension-driven flows, thermal-hydrodynamic instability, or liquid spreading. Stain formation also depends on more general mechanisms such as solute transport, surface tension gradients, solute diffusion, or electrostatic forces. Following the drying of complex fluid droplets, one important aspect of droplet spreading and evaporation is the creation of patterns.[13–16] Additionally, protein–protein and protein–substrate interactions play a crucial role in aggregating proteins on the substrate. The evaporation process of a drop containing two distinct proteins is primarily governed by these molecular interactions, leading to the creation of complex patterns.[17]

Conserved across most species, humoral immune responses involve polyclonal antibodies with inherent structural and functional diversity, making them useful in a variety of research and diagnostic applications.[18] Antigen binding is mediated by the variable fragment of the antibody.[19,20] IgG antibodies play pivotal roles in the immune system and their interaction with target proteins or antigens can affect body health.[21] The most important serum antibodies are immunoglobulin G (IgG) proteins and are divided into four subclasses: IgG1 (66%), IgG2 (23%), IgG3 (7%), and IgG4 (4%).[22] It is important to note that the species of origin of the IgG molecules, whether it be human, rabbit, cow, or goat, contribute to their unique characteristics, resulting in distinct binding affinities to antigens.[23–25]

An important class of B-cell superantigens (SAgs), such as Protein A, interacts with IgG molecules in an unusual way by binding to the Fab region of the IgGs outside the complementarity-determining region.[23,26] As a result of its ability to bind to antibodies via its Fc domain, Protein A has been demonstrated to inhibit opsonophagocytosis by orienting the antibodies at an incorrect angle.[27] SAgs have proven to be useful in exploring fundamental questions in immunobiology, such as mechanisms for cell activation, tolerance, and autoimmunity.

In this study, we build from our previous work with ML-based image-recognition algorithms[8] to develop and approach for species-specific typification of immunoglobulin complexes using stain patterns. To analyze the images, a trained neural network is used to stratify polyclonal immunoglobulin G from various species (human, rabbit, goat, and bovine) based on their interactions with a common binding partner, Protein A. Recombinant Protein A is a B-cell SAg with high binding affinity to human IgG and is widely used in various biological fields for the isolation and purification of human IgG.[28] For this reason, IgG was chosen to screen different protein–protein interaction levels obtained with different IgG sources and different molar ratios[29] to predict the binding affinity using the classified different protein–protein interaction levels.

2. Results and Discussion

To investigate the accuracy of typification of a range of IgG complexes, their deposition patterns were analyzed. PLM imaging was applied for this purpose. Protein solutions were prepared using 0.1 M sodium phosphate buffer with a pH of 8.1 commonly used for IgG purification.[30,31] The total mass of all protein samples was kept constant at 0.3 mg mL^{-1}; however, different molar ratios were employed for protein–protein interaction screening. Several physical properties as well as environmental conditions influence the structure of the deposition patterns of biomacromolecular solutions. To ensure nearly identical droplet radii (2 mm), a hydrophobic substrate was used. Chemical vapor deposition (CVD) polymerization was used to obtain poly(*p*-xylylene) (PPX)-coated glass surfaces, a key condition for consistent and reproducible droplet deposition over broad areas. Circular droplets were deposited and dried under humidity and temperature-controlled conditions for at least 6 h (**Figure 1**A,B). After drying, images of the deposition patterns were captured using PLM, applying constant parameters like resolution, magnification, light intensity, gain, and exposure time (Figure 1C).[30,31]

Time-of-flight secondary ion mass spectrometry (ToF–SIMS) was used to elucidate the chemical composition of the dried patterns (Figure 1E). Mapping of the signal's characteristic for proteins (as CNO^-, displayed in the red channel of the red, green and blue (RGB) overlay and buffer components (as PO_2^-, displayed in the green channel of the RGB overlay) revealed complementary patterns in the droplet region. This indicates that upon drying the solution components tend to form separate agglomerates of either protein or buffer components. The presence of distinct regions was also confirmed by scanning electron microscopy (SEM) (Figure 1D). The magnified SEM image of the central region revealed the presence of crystalline structures corroborating the high salt content.

In the following, IgG from four different species and human serum albumin (HSA) as a control were investigated. Specifically, different molar ratios were used to compare different levels of protein–protein interactions.[32–34]

Figure 2A summarizes the stain patterns obtained from the four IgGs, HSA, and Protein A following the approach outlined earlier, as well as from the various protein–protein combinations. The patterns of individual IgG from various species, HSA as the primary proteins, and Protein A as an antigen are shown in Figure 2A-I,II, respectively. The patterns of the respective IgG and HSA complex with Protein A in the molar ratio 2:1 are provided with a borderline color related to their position in the relative binding strength percentage color map, explained further below and are available in Figure 2A-III. Patterns originating from the variation of the molar ratios are summarized in Figure S1, Supporting Information. The patterns of the complex of the different IgG(s) with low affinity to bind to Protein A

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH

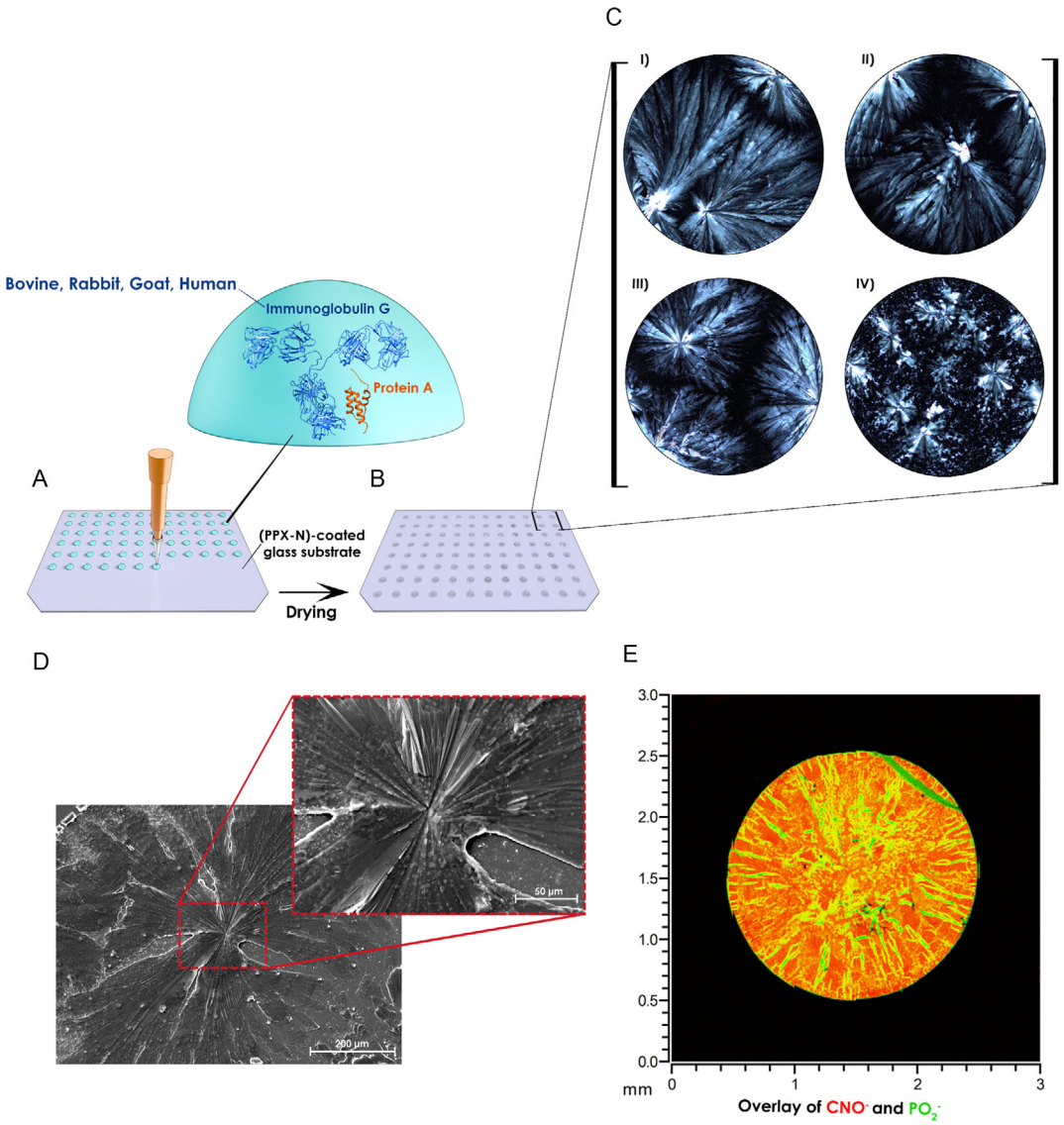


Figure 1. Formation of protein stains using controlled droplet deposition and drying process. A) Cleaned glass substrates were coated with poly(p-xylylene) via CVD polymerization to obtain reliable hydrophobic surface conditions to ensure a reproducible surface interaction. An automated pipetting system was used for dispensing several protein sample droplets (2 µL) containing different molar ratios of IgG from different species and Protein A. B) Dispensed droplets were dried under controlled environmental conditions ($T = 25$ °C, relative humidity = 40%). C) PLM imaging was used to collect all the deposited proteins' patterns under the same conditions to prepare sufficient images for each class for image classification with CNN implementation, IgG from I) bovine, II) rabbit, III) goat, and IV) human serum interacting with Protein A. D) SEM image analysis of human IgG stains. E) ToF–SIMS analysis of the deposition pattern of a complex of goat IgG with Protein A: RGB overlay image of the distribution map of PO_2^- ions (green) and CNO^- ions (red).

(e.g., goat IgG) are visually similar to the IgG patterns without the addition of Protein A. The same trend is observed for the HSA–Protein A mixtures.

The color map in Figure 2B depicts the relative binding strength of various IgGs to recombinant Protein A obtained from the circular dichroism (CD) spectroscopy measurement results. In these measurements, the CD spectra of single proteins were measured and summed up to achieve the spectra of their mixture. These sum spectra were compared to the CD spectra of the actual protein mixture solution. In the case of strong protein interaction, the CD spectra of the mixture and the sum spectra of the single proteins differ particularly as the interaction between two proteins changes their secondary structure.[35,36] Considering the CD intensity at 217 nm,[35,36] the percentage of relative change in secondary structure can be calculated. The maximum detected changes for HSA, goat IgG, bovine IgG, rabbit IgG, and human IgG in the mixture with Protein A were 6%, 9%, 16%, 16%, and 81%, respectively.

Here, human IgG showed a strong affinity for Protein A, while rabbit IgG and bovine IgG displayed weak to medium interaction

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH

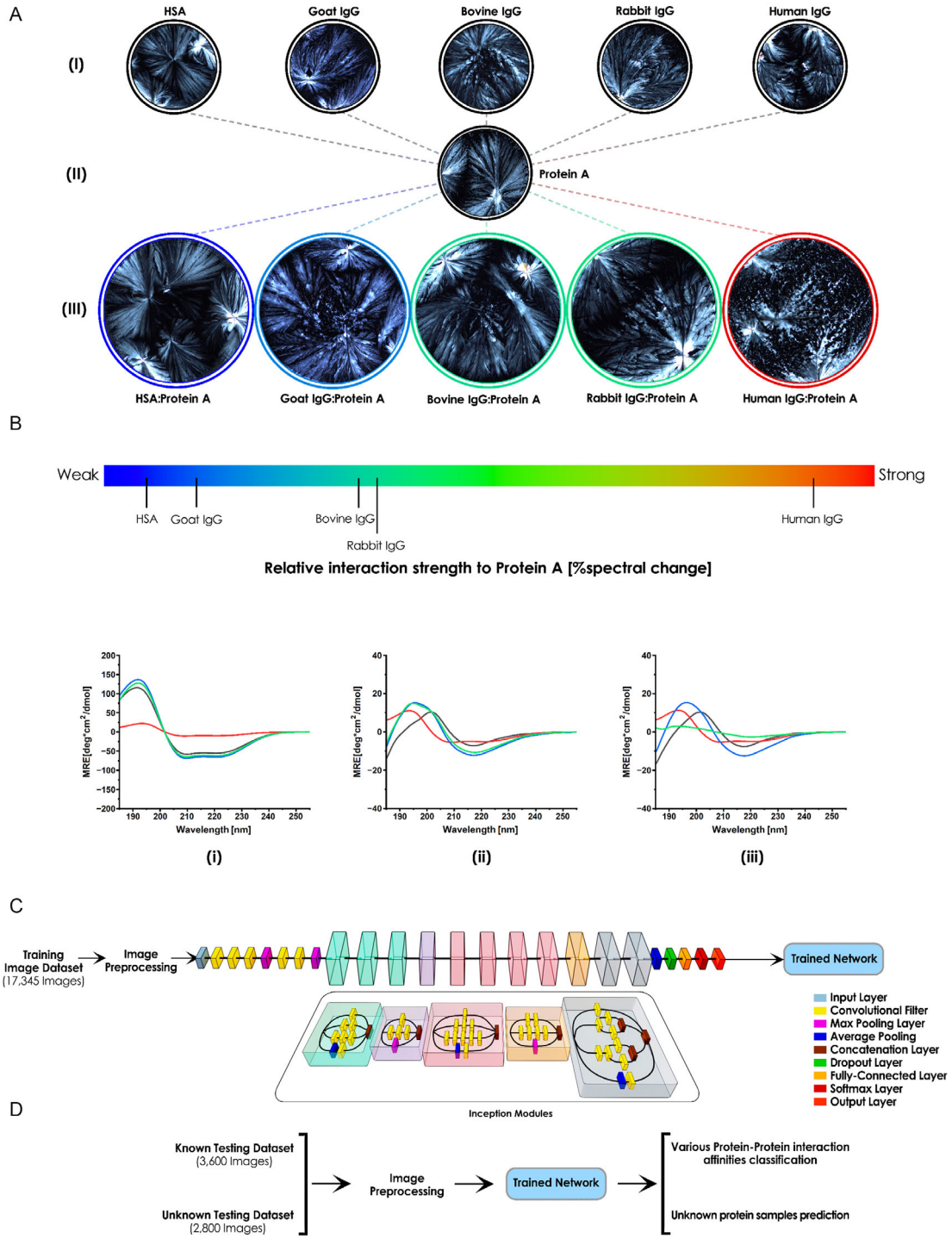


Figure 2. Analysis of protein–protein interactions based on image classification using InceptionV3. A) The patterns of single Protein A and IgG from different species (HSA was included as a negative control) and their interaction in the molar ratio of 2:1 (IgG:Protein A) are shown. Black borders specify single-protein deposition patterns. The colored borders indicate various IgG and HSA mixtures with Protein A. All images were converted to grayscale to eliminate color effects on AI classification. B) The evaluated strength of interaction is shown in the color map from weak to strong using CD spectroscopy measurement. Applying CD spectroscopy to compare the summation of two single-protein spectra with the spectra of their respective mixture to obtain the relative strength of interaction leading to changes in their secondary structure (red color is for Protein A, black for i) HSA, ii) rabbit IgG, and iii) human IgG, blue for the summation of two spectra and green for the mixture). C) The architecture of InceptionV3, a pretrained convolutional neural network (CNN) implemented for image classification. A sufficient number of images for each class was prepared for this purpose and introduced to InceptionV3 as an input image dataset. D) The trained network can classify different proteins, determine their interaction levels, and predict the similarity between the unknown testing image dataset and the training image dataset.

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH

strengths. Only negligible interactions were found for goat IgG and HSA (Figure 2B,i–iii). These results are consistent with other literature reports.[29,37–42] These similarities in secondary structure of IgG complexes were observed at different molar ratios as well and are available in Figure S2, Supporting Information. To establish an unbiased classification of binding interactions, the convolutional neural network (CNN), InceptionV3, was pretrained with at least 400 images per class.[43–47] Pretrained models, such as InceptionV3, ResNet, and NasNetLarge, are trained on large datasets like ImageNet, enabling them to learn rich, hierarchical features that can generalize across various tasks.[8] The advantage of using pretrained models in image classification algorithms stems from the fact that they leverage vast amounts of data and complex architectures. Thus, they significantly enhance accuracy and efficiency.[48–51] This network was selected because it has one of the highest accuracies among other known CNNs (with the same dataset) and requires less time for training compared to other high-performance pretrained CNNs (e.g., NasNetLarge).[8] The architecture of InceptionV3 is illustrated in Figure 2C. The images used for training and validation for 36 classes were completely distinct from the testing image datasets (3,600 and 2,800). After training the network with our image dataset, two outputs are desired: 1) screening different protein–protein interaction levels as well as 2) predicting and classifying unknown protein samples within the trained network as represented in Figure 2D.

In **Figure 3**, a confusion chart of the classification of Protein A immunocomplexes with IgGs from four different species (human, rabbit, bovine, and goat) with binding affinities ranging from weak to strong (as seen in Figure 2B) is displayed. Specifically, we evaluated a range of different molar ratios ranging from 0.5:1 to 5:1 (IgG:Protein A). This comprehensive confusion matrix developed from 36 immunoprotein complexes reveals the accurate classification of IgGs to the corresponding species regardless of the used molar ratio. The total accuracy of the 36 outlines confusion chart was 81.4%, which included not only the species classification but also the individual classification of the different molar ratios.

Figure 4 depicts a breakdown of this data by species to highlight the potential for species-specific typification of the four immunoglobulin complexes regarding the binding affinity strength of the protein complexes. An excellent correlation between the classification accuracy of the CNN and the binding affinity was observed. Although the number of free proteins in the solution varied across different molar ratios, it corresponded to the changes in secondary structure of proteins observed by CD spectroscopy. Furthermore, for each set, a t-distributed stochastic neighbor embedding (t-SNE) plot was prepared to support the cluster analysis by the trained network.[52] All t-SNE plots were applied to the "Softmax" layer of the CNN. The output of this layer is a 4D array that consists of x- and y-spatial dimensions of the images, channels of the images, and batch dimension, respectively. To illustrate the performance of the trained CNN, high-dimensional data were evaluated as shown in the t-SNE plot. Consequently, good classification was achieved by separating different image classes. Figure 4A represents the confusion matrix of human IgG and Protein A image classes. As confirmed by CD spectra (Figure S2-A, Supporting Information), human IgG has a high affinity to bind to Protein A.[25,28] In this confusion matrix, the ratio of 2:1 (IgG:Protein A) obtained the highest prediction accuracy of (97%), which matched the ratio with the highest extent of structural changes as detected by CD spectroscopy (Figure S2-A, Supporting Information).[42,53]

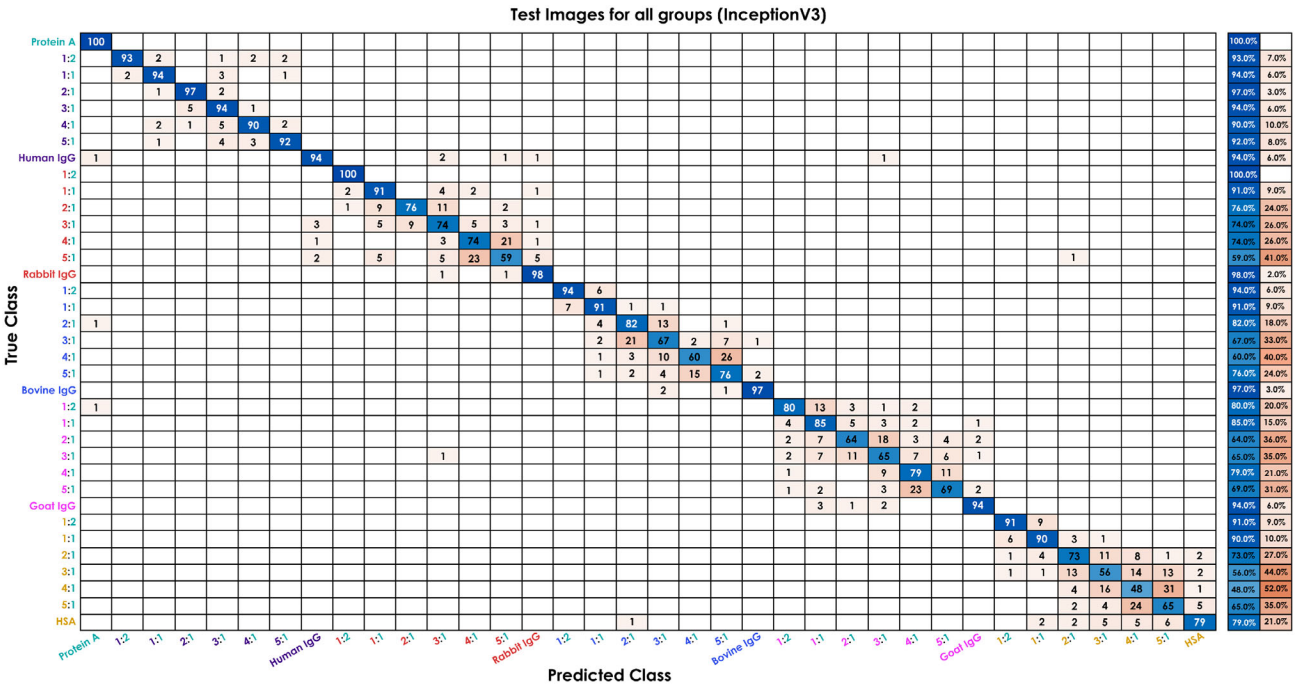


Figure 3. The confusion matrix derived from the test image dataset generated via the pre-trained InceptionV3. The total accuracy of the confusion matrix is 81.42%.

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH

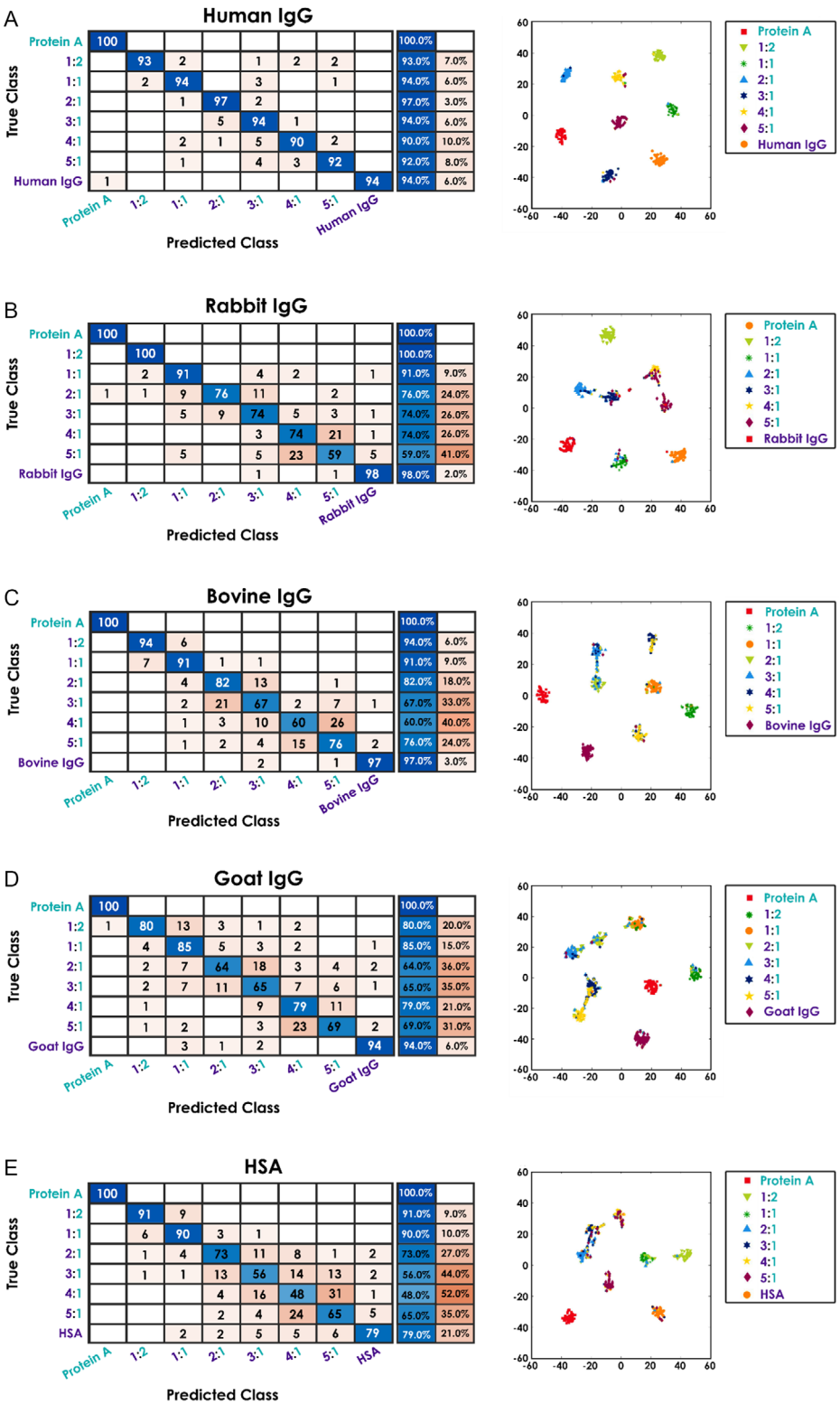


Figure 4. Confusion matrix of different protein–protein interactions: A) human IgG, B) rabbit IgG, C) bovine IgG, D) goat IgG, and E) HSA. The confusion chart was obtained for a test set of 36 categories divided into smaller charts with t-SNE plot analysis for each protein complex.

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH

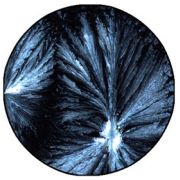
Accordingly, the CNN classified human IgG complexes with higher accuracy than the other IgGs. The CNN distinguished the different molar ratios of human IgG complexes with an overall accuracy of 93.4%. Misclassifications with rabbit IgG complexes were observed (Figure 3), which can be attributed to the similarities in molecular weight and secondary structure of human IgG and rabbit IgG. The t-SNE plot also indicates very good clustering.

Figure 4B shows the confusion chart for rabbit IgG:Protein A image classes. CD spectroscopy measurements indicated that rabbit IgG has a weak to medium affinity to bind to Protein A as the secondary structure changes were negligible in comparison to human IgG complexes.[39] As in the case for human IgG, the trained InceptionV3 classified some rabbit IgG images into human IgG classes. The accuracy is 81.7% which revealed several misclassifications between different molar ratios (Figure S2-B, Supporting Information). In Figure 4C, a confusion chart of IgG from bovine serum and Protein A including six different molar ratios is presented. The bovine IgG image classification confusion matrix had an accuracy of 81%. Misclassifications were mainly due to similarities of the patterns that only varied in their respective molar ratios. These similarities may be attributed to small changes in the secondary structure of the protein solution upon complexation, a tendency that was also observed by CD spectroscopy (Figure S2-C, Supporting Information) and several literatures.[41,54,55] Similarly, bovine IgG was found to bind with weak to medium strength to Protein A. Consequently, the obtained accuracy of classification for bovine IgG:Protein A image classes is similar to rabbit IgG:Protein A. For bovine IgG:Protein A classes, the t-SNE plot indicates no significant differences between the stain images obtained for various molar ratios. However, it still recognizes the single Protein A image class accurately (100%). Additionally, for single-bovine IgG and for the molar ratio where the Protein A content is higher than bovine IgG (0.5:1), clustering worked well with an accuracy of 97% and 94%, respectively. Furthermore, the trained InceptionV3 did not misclassify bovine IgG with IgGs from three other species (Figure 3). Figure 4D shows the confusion chart of the mixture of IgG from goat serum with Protein A. Goat IgG showed only weak interactions with Protein A[42,56] (Figure S2-D, Supporting Information). Moreover, the error rates for classifying different molar ratios are high and the local accuracy of this confusion chart is only 76.6%. Correspondingly, the t-SNE plot of this image set displays improper clustering of protein pattern classes of different molar ratios. A confusion chart of the HSA:Protein A image classification can be found in Figure 4E. For comparison, the control protein HSA has no propensity to bind to Protein A and due to this the lowest accuracy in test image classification was obtained. For all molar ratios, there was no evidence of the formation of HSA–Protein A complexes. The local accuracy of the confusion matrix for HSA:Protein A complexes is 71.7% and the t-SNE plot indicates limited clustering of these images.

To evaluate the performance of the trained network with IgG from different sources and Protein A, the patterns of Protein G and its interaction with IgG from human serum were analyzed by the trained network as unknown samples to see how the machine classifies these patterns into various categories of interaction with different molar ratios and affinities. As these two proteins have the same tendency as SAgs for binding to human IgG, their functionality and structural properties, and their deposition patterns are given in **Table 1**. Although Protein G has a lower number of binding sites to bind to human IgG in comparison to Protein A, its affinity for IgGs is higher.[57,58] CD spectroscopy analysis of human IgG:Protein G showed levels of interactions that were similar to human IgG:Protein A (Figure S3, Supporting Information). **Figure 5**A displays the obtained confusion chart of test images of human IgG:Protein G classes. Even though Protein G patterns are visually different from Protein A (Table 1), the trained network remarkably predicted 83% of them into the Protein A image class (specified with a purple bracket, last row). Also, it predicted 13% of Protein G images as HSA:Protein A with a molar ratio of 1:2, due to the aforementioned similarities between these proteins. Further classification was performed on four different image datasets of mixtures of human IgG with Protein G and similar results were obtained. The dominant prediction (94%) of the human IgG:Protein G images is the human IgG:Protein A image classes (displayed with purple brackets).

In Figure 5B, examples of patterns of human IgG:Protein G with the molar ratio of 2:1, human IgG, and human IgG:Protein

Table 1. Characteristics of recombinant protein A and protein G.

Properties ligand	MW [KDa]	Secondary structure	No. of active sites for binding IgG	Binding constant for human IgG $[M^{-1}] \times 10^{-9}$	Relative percentage bond to human IgG	Deposited pattern
Protein A	36	Three α-helices	5	44.1	81%	
Protein G	31	One α-helix packed onto a four-stranded β-sheet	2	67.4	83%	

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH

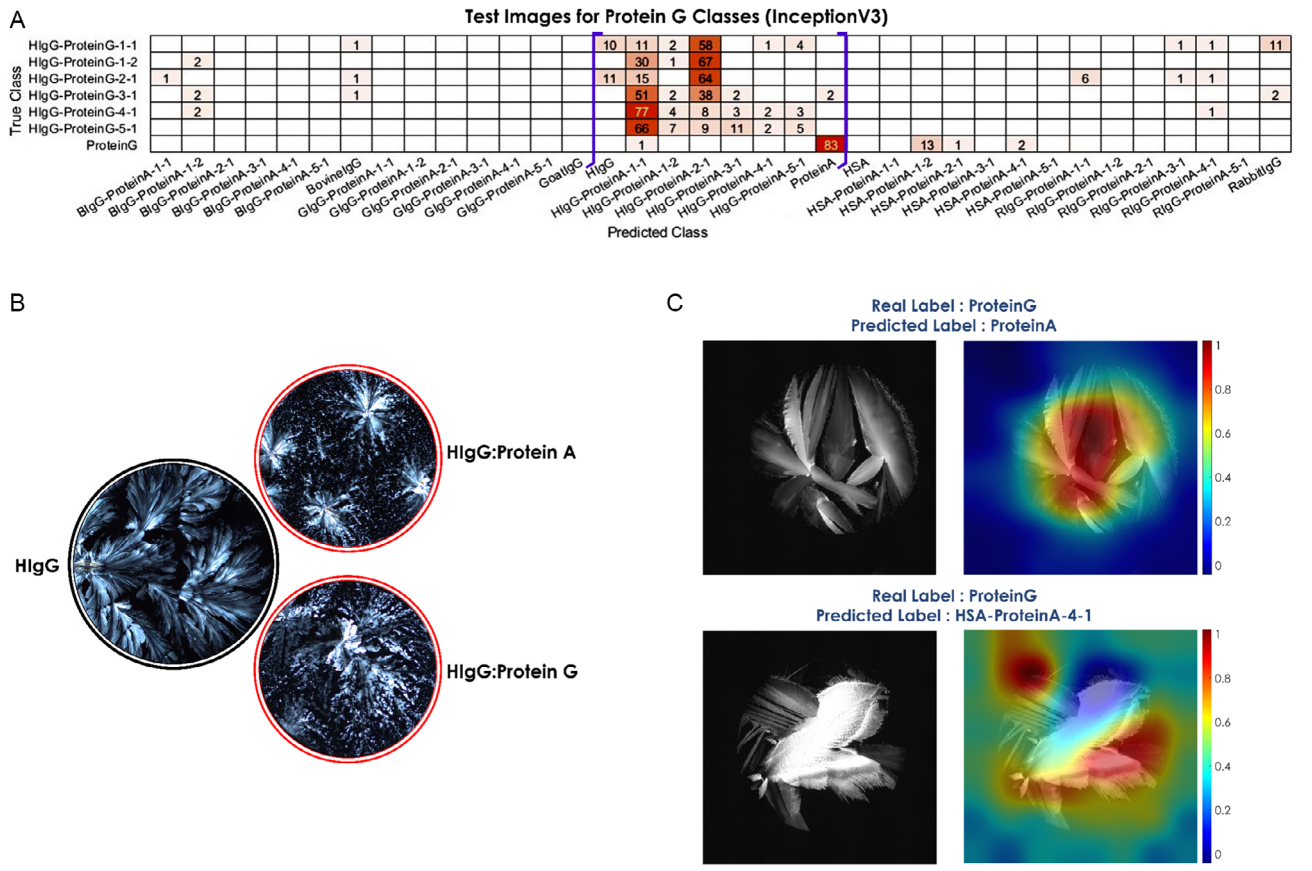


Figure 5. Performance of a pretrained network using human IgG:Protein G patterns. A) Confusion chart of human IgG:Protein G image classification as an unknown sample (not trained). B) Examples of obtained patterns of human IgG, human IgG:Protein G with a molar ratio of 2:1 in comparison to Protein A interaction with human IgG pattern with the same molar ratio. C) Single image prediction with Grad-CAM image analysis.

A are given, indicating the similarities between the patterns of the Protein G:human IgG complex and those of the Protein A:human IgG complexes compared to the single Protein G and Protein A images.

To investigate how the network makes decision based on the input images, a gradient-weighted class-activation mapping (Grad-CAM) analysis was performed. Grad-CAM is a technique applied to understand why a deep-learning network makes its classification decisions.[59] Grad-CAM calculates the gradient of the class score (that is, the score for the class of interest) based on the feature maps (also called activation maps) of the final convolutional layer in CNN. These gradients indicate the importance of each pixel in the feature map according to the partial derivatives of the class score. This analysis allows differentiation between multiple classes based on the features specific to each class in the image.[60,61] In Figure 5C, the prediction of the single images of human IgG–Protein G and their Grad-CAM image analysis is shown. The positions where this gradient is large are the places that were most heavily weighted in the analysis of the neural network algorithm. In the heat map, the regions with warmer colors are related to the areas with higher scores in prediction. This reveals that the most important information of the images for classification is drawn from the stain patterns, mostly in the center. The information from the background of the images is neglegible. On the contrary, in the case of inaccurate predictions, background interference on the network's decision making was detected in the Grad-CAM analysis as shown in Figure 5C (II). Here, the correct classificationwas "Protein G" but the trained network categorized it inaccurately as "HSA-ProteinA-4-1".

3. Conclusions

Image classification using deep-learning approaches is poised to find a wide range of applications in different fields of biology and medicine. Images of stain patterns of biomacromolecules and their mixtures contain previously untapped levels of structural and functional information. Here, we demonstrate strong correlations between structural differences in protein–protein complexes and how these can result in characteristic changes in their respective deposition patterns. Using an ML algorithm, protein–protein complexes were stratified based on the interactions of IgG from various species with the SAg recombinant Protein A. The pretrained InceptionV3 not only successfully distinguished different IgG:Protein A combinations, but also predicted binding propensities. Even unknown samples were successfully sorted into classes exhibiting similar interaction

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH

tendencies than what was observed by CD spectroscopy. This underlines the power of this approach, as it confirms effective translation to unknown protein affinities. The role of the substrate is important as it provides an inert and homogenous support for droplet deposition. In principle, any surface coating process would be appropriate as long as it results in homogeneous coatings and is scalable to the size of 96 well plates. We used CVD polymerization because it is a commercially used coating process that results in inert, homogenous, and well-defined coatings of defined thickness.[62–64] Our findings may contribute to the future development of precise, straightforward, and unbiased methods to predict protein–protein interactions. To ensure the broader applicability of this methodology, future studies may focus on performance and workflow optimization.

4. Experimental Section

Protein Solution Preparation: All IgG from various species and HSA were purchased from Sigma Aldrich (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany). Recombinant Protein A and recombinant Protein G were purchased from Abcam (Abcam plc, Cambridge, UK). To prepare the protein solutions, the lyophilized proteins were dissolved in 100 mM sodium phosphate buffer at a final concentration of 0.3 mg mL^{-1}. The pH of the buffer was 8.1, containing 94.7 mM Na_2HPO_4 (Merck Chemicals GmbH) and 5.3 mM NaH_2PO_4 (Merck Chemicals GmbH) in ultrapure water from the Mili-Q plus system (Millipore, Schwalbach, Germany), filtered twice using a 0.2 µm syringe filter (Sartorius Stedim Biotech GmbH, Gottingen, Germany). For good dissolution, each IgG was dissolved in the buffer using an tube rotator (Stuart, Stone, UK) at 10 rpm for 2 h at room temperature. Since the IgG samples had insoluble aggregates after mixing, these IgG solutions were centrifuged (Centrifuge 5804, Eppendorf, Hamburg, Germany) at 4000 rpm for 4 min. Then, the supernatant was separated, and the exact concentration was measured using a Nanodrop micro-volume spectrophotometer at 280 nm (NanoDrop One, Thermo Scientific, Darmstadt, Germany) by using the molecular weight and extinction coefficient of each. Since Protein A, Protein G, and HSA dissolved in aqueous solutions very well, after mixing for 30 min, their exact concentration was directly determined by the NanoDrop device. After the preparation of a single-protein solution (stock samples), different protein–protein solutions were prepared with various defined molar ratios. The total mass concentration was kept constant at 0.3 mg mL^{-1}. These solutions were mixed using an SB3 tube rotator at 10 rpm for an hour to maximize the interaction between each antibody and antigen pair. After mixing, the samples were stored in aliquots at −20 °C. The molar ratios were 0:1, 1:0, 1:2, 1:1, 2:1, 3:1, 4:1, and 5:1 (antibody:antigen). The ratios of 1:0 and 0:1 were related to a single-antibody/antigen solution.

Surface Preparation via CVD Polymerization: Glass plates with the exact dimension of 96-well plates (120 × 80 mm) with the specifications of extra white float, clear, and uncoated were custom-made with a thickness of 1.0 ± 0.05 mm (Optrovision, München, Germany). Before the coating step, the glass plates were cleaned using a Plasma Cleaner (PIE Scientific, San Francisco, USA) with dry air to remove surface contaminants, change the surface energy, and improve bonding strength. A power of 75 W for a 5 min emission was applied for this purpose. The cleaned glass plates were then coated with PPX via CVD polymerization following a previously described method.[65,66] The precursor, [2.2]paracyclophane (Curtiss-Wright Surface Technologies [Galway, Ireland]) was sublimed under vacuum and converted by pyrolysis into quinomethane, which spontaneously polymerized upon condensation to the glass surface. A constant argon flow of 20 sccm was used as the sweep gas. The sublimation temperature was 100–110 °C followed by pyrolysis at 660 °C. The coating pressure was set to 0.15 mbar.

Droplet Deposition: All protein sample solutions had a concentration of 0.3 mg ml in 100 mM sodium phosphate buffer (pH 8.1). The deposition of a predefined array of droplets onto the glass slide was performed by an automated 96-well microplate pipetting device (EpMotion 5070, Eppendorf AG, Hamburg, Germany) coupled with a 1-channel dispenser (TS10, Eppendorf AG, Hamburg, Germany). To control environmental conditions, the pipetting system was placed inside a climate chamber (ICH 750, Mommert GmbH + Co. KG, Schwabach, Germany), and the environmental conditions were controlled at a temperature of 23 ± 0.5 °C and the relative humidity at 40% ± 5%. Each droplet was deposited at a dispensing speed of 3 mm s^{-1} and a volume of 2 µL. The pipetting system was programed to dispense 96 droplets per glass plate in the form of 12 columns and 8 rows. After the droplets were allowed to dry for at least 6 h, images of the deposition patterns were captured using a polarizing optical microscope (Olympus BX-53 F, Tokyo, Japan) equipped with an automated stage. All images were acquired at a consistent light intensity using a 10x objective and stitched together using the multi-image alignment algorithm included in CellSens software (Olympus, Tokyo, Japan). The acquired images had a square dimension of 8013 × 8013 pixels in the.jpg format. The images were resized to a final dimension of 2003 × 2003 pixels for fast import into the network for training.

Training and Testing of the CNN: All raw images were imported into MATLAB (Release 2023a, Math Works Inc.) for further processing and training of the CNN. These images were resized to a size compatible with the input layer of the CNN and converted to gray scale mode. InceptionV3, a pretrained CNN, was applied for this purpose. It had an image input size of 299 × 299 pixels. InceptionV3 had 315 layers and contained five inception modules. It was one of the most accurate pretrained CNNs with respect to the number of its layers, and the time required for training was reasonable in comparison to other accurate pretrained CNNs such as NasNetLarge.

Following the transfer-learning approach, a pretrained network with a large set of image features was fine-tuned with a relatively small set of new images. During transfer learning, the final classification layer was removed from the network and retrained using the new dataset. Fine-tuning of the parameters occurred across all layers using the same global learning rate of 0.001, a minimum batch size of 64 images, and a maximum number of epochs of 80. To prevent the network from overfitting and memorizing the exact details of the training dataset, the images were augmented using a random reflection function in which each image was reflected horizontally and vertically with a 50% probability. ≈400 images per class were used for training, and 10% of these images were randomly selected for validation during training. In addition, 100 images per class were prepared for testing after the network training. There was no overlap among the training, validation, and testing datasets. This procedure was the same as that for an unknown image test set. Using the trained network with different classes, a completely different image as a test set was introduced to the network for classification and it showed similarities between the given images and the trained ones. For both types of testing datasets, the total accuracy and confusion charts were obtained.

The t-SNE algorithm, a method for visualizing high-dimensional data, was applied to the "SoftMax" layer of the trained network to demonstrate how well the network clusters different protein–protein interaction levels. A MATLAB ML package was used to perform the t-SNE.

The visualization algorithm Grad-CAM was used to understand which regions of the image had the maximum influence on the classification decision of the CNN.

ToF-SIMS: ToF–SIMS was performed using a ToF–SIMS instrument (ION-TOF GmbH, Münster, Germany) equipped with a Bi-cluster liquid-metal primary-ion source and a nonlinear ToF analyzer. For spectrometry, short primary-ion pulses (<1 ns) of the Bi source were operated in the "bunched" mode providing Bi^{1+} ion pulses at 25 keV energy and a lateral resolution of 5 µm. As the droplets were larger than the maximum deflection range of the primary-ion gun of 500 × 500 µm^2, the images were obtained using the manipulator stage scan mode. Negative polarity spectra were calibrated on the C^-, CH^-, and CH^{2-} peaks. Spectrometry was performed in static SIMS mode by limiting the primary-ion dose to $<10^{11}$ ions cm^{-2}. Charge compensation was necessary because of

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH

the glass substrate so that an electron flood gun providing electrons of 20 eV was applied and the secondary-ion reflectron tuned accordingly.

SEM: The morphology of the deposited proteins and salt in dried droplets was analyzed using an SEM (LEO 1530 Gemini, Zeiss, Germany). A thin layer of gold was sputtered onto the samples prior to SEM imaging to minimize surface charging.

CD Spectroscopy Measurements: The far-UV CD spectroscopy analysis was conducted at 20 °C using a J-1500 Spectrophotometer (JASCO, Deutschland GmbH), in quartz glass cuvettes with the path length of 1 mm (Helma GmbH & Co. KG, Müllheim, Germany) between 260 and 180 nm and with data intervals of 0.5 nm. Two repeat scans at a scan rate of 100 nm min^{-1} were carried out for each sample and its respective baseline of glass quartz. The spectrum of the protein-free buffer was also recorded. The proteins had the same concentration as in the protein stains, but in 20×10^{-3} M sodium phosphate buffer (pH 8.1). After smoothing the spectra, the final comparison was performed.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This research was supported by the Helmholtz-Initiative "Virtual Materials Design (VirtMat)" and the Helmholtz Program "Materials Systems Engineering" at the Karlsruhe Institute of Technology.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Joerg Lahann: Conceptualization (lead); Project administration (lead); Writing—review and editing (lead). **Bahar Dadfar**: Conceptualization (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Writing—original draft (lead). **Safoura Vaez**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting). **Cristian Haret**: Methodology (supporting). **Meike Koenig**: Project administration (supporting); Supervision (supporting); Writing—original draft (supporting). **Tahereh Mohammadi Hafshejani**: Methodology (supporting). **Matthias Franzreb**: Supervision (supporting); Writing—review and editing (supporting).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

machine learnings, neuronal networks, polymer coatings, protein–protein interactions

Received: May 3, 2024
Revised: August 9, 2024
Published online: September 9, 2024

[1] O. Keskin, A. Gursoy, B. Ma, R. Nussinov, *Chem. Rev.* **2008**, *108*, 1225.
[2] J. A. Miernyk, J. J. Thelen, *Plant J.* **2008**, *53*, 597.
[3] T. Berggård, S. Linse, P. James, *Proteomics* **2007**, *7*, 2833.
[4] V. S. Rao, K. Srinivas, G. N. Sujini, G. N. Kumar, *Int. J. Proteomics* **2014**, *2014*, 147648.
[5] F. A. Aprile, P. Sormanni, M. Podpolny, S. Chhangur, L.-M. Needham, F. S. Ruggeri, M. Perni, R. Limbocker, G. T. Heller, T. Sneideris, *Proc. Natl. Acad. Sci.* **2020**, *117*, 13509.
[6] S. Hassoun, F. Jefferson, X. Shi, B. Stucky, J. Wang, E. Rosa, *Integr. Comp. Biol.* **2021**, *61*, 2267.
[7] J. M. Vaz, S. Balaji, *Mol. Divers* **2021**, *25*, 1569.
[8] A. Jeihanipour, J. Lahann, *Adv. Mater.* **2022**, *34*, 2110404.
[9] D. Brutin, V. Starov, *Chem. Soc. Rev.* **2018**, *47*, 558.
[10] R. D. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S. R. Nagel, T. A. Witten, *Nature* **1997**, *389*, 827.
[11] R. G. Larson, *AIChE J.* **2014**, *60*, 1538.
[12] R. G. Larson, *Angew. Chem., Int. Ed.* **2012**, *51*, 2546.
[13] A. Pal, A. Gope, G. Iannacchione, *Biomolecules* **2021**, *11*, 231.
[14] R. G. G. Larson, M. A. López, D. W. Lim, J. Lahann, *MRS Online Proc. Libr.* **2010**, *1273*, 1273.
[15] D. Brutin, *Colloids Surf., A* **2013**, *429*, 112.
[16] H. Hu, R. G. Larson, *J. Phys. Chem. B* **2006**, *110*, 7090.
[17] Y. J. P. Carreón, J. González-Gutiérrez, M. I. Pérez-Camacho, H. Mercado-Uribe, *Colloids Surf., B* **2018**, *161*, 103.
[18] M. Caffrey, *Biochim. Biophys. Acta, Mol. Basis Dis.* **2001**, *1536*, 116.
[19] C. Haußner, J. Lach, J. Eichler, *Curr. Opin. Chem. Biol.* **2017**, *40*, 72.
[20] A. C. A. Roque, C. S. O. Silva, M. Â. Taipa, *J. Chromatogr. A* **2007**, *1160*, 44.
[21] S. W. de Taeye, T. Rispens, G. Vidarsson, *Antibodies* **2019**, *8*, 30.
[22] X. Chen, O. Schneewind, D. Missiakas, *Proc. Natl. Acad. Sci.* **2022**, *119*, e2114478119.
[23] A. M. Deacy, S. K.-E. Gan, J. P. Derrick, *Front Immunol.* **2021**, *12*, 731845.
[24] C. Aybay, *Immunol. Lett.* **2003**, *85*, 231.
[25] M. Eliasson, R. Andersson, A. Olsson, H. Wigzell, M. Uhlén, *J. Immunol.* **1989**, *142*, 575.
[26] A. L. Anderson, R. Sporici, J. Lambris, D. LaRosa, A. I. Levinson, *Infect. Immun.* **2006**, *74*, 1196.
[27] P. G. Fox, F. Schiavetti, R. Rappuoli, R. M. McLoughlin, F. Bagnoli, *mBio* **2021**, *12*, 10.
[28] S. Hober, K. Nord, M. Linhult, *J. Chromatogr. B* **2007**, *848*, 40.
[29] E. V. Sidorin, T. F. Solov'Eva, *Biochemistry* **2011**, *76*, 295.
[30] P. L. Ey, S. J. Prowse, C. R. Jenkin, *Immunochemistry* **1978**, *15*, 429.
[31] S. Ghose, M. Allen, B. Hubbard, C. Brooks, S. M. Cramer, *Biotechnol. Bioeng.* **2005**, *92*, 665.
[32] R. Lindmark, K. Thorén-Tolling, J. Sjöquist, *J. Immunol. Methods* **1983**, *62*, 1.
[33] E. B. Myhre, G. Kronvall, *Comp. Immunol. Microbiol. Infect. Dis.* **1981**, *4*, 317.
[34] J. J. Langone, *Biochem. Biophys. Res. Commun.* **1980**, *94*, 473.
[35] S. Sechi, P. P. Roller, J. Willette-Brown, J. P. Kinet, *J. Biol. Chem.* **1996**, *271*, 19256.
[36] N. J. Greenfield, *Protein-Protein Interactions: Methods and Applications* Humana Press, Totowa, NJ, **2004**, pp. 55.
[37] R. C. Duhamel, E. Meezan, K. Brendel, *Mol. Immunol.* **1980**, *17*, 29.
[38] N. L. Brown, S. P. Bottomley, M. D. Scawen, M. G. Gore, *Mol. Biotechnol.* **1998**, *10*, 9.
[39] C. Endresen, *Acta Pathol. Microbiol. Scand. Sect. C* **1978**, *86*, 211.
[40] B. Akerström, E. Nielsen, L. Björck, *J. Biol. Chem.* **1987**, *262*, 13388.
[41] M. J. P. Lawman, S. Joiner, D. R. Gauntlett, M. D. P. Boyle, *Comp. Immunol. Microbiol. Infect. Dis.* **1985**, *8*, 1.
[42] M. D. P. Boyle, K. J. Reis, *Bio/Technology* **1987**, *5*, 697.

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH

[43] L. Hamadeh, S. Imran, M. Bencsik, G. R. Sharpe, M. A. Johnson, D. J. Fairhurst, *Sci. Rep.* **2020**, *10*, 3313.
[44] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436.
[45] C. Angermueller, T. Pärnamaa, L. Parts, O. Stegle, *Mol. Syst. Biol.* **2016**, *12*, 878.
[46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, Las Vegas, NV, June **2016**, pp. 2818–2826.
[47] C. Wang, D. Chen, L. Hao, X. Liu, Y. Zeng, J. Chen, G. Zhang, *IEEE Access* **2019**, *7*, 146533.
[48] N. Abou Baker, N. Zengeler, U. Handmann, *Mach. Learn Knowl. Extr.* **2022**, *4*, 22.
[49] J. Plested, T. Gedeon, *arXiv preprint arXiv:2205.09904* **2022**.
[50] E. D. S. Puls, M. V. Todescato, J. L. Carbonera, *arXiv preprint arXiv:2310.02037* **2023**.
[51] M. S. Hasan, in *2017 20th Inter. Conf. Computer and Information Technology (ICCIT)*, IEEE, Piscataway, NJ **2017**, pp. 1–6.
[52] L. Van der Maaten, G. Hinton, *J. Mach. Learn. Res.* **2008**, *9*, 2579.
[53] D. Lancet, D. Isenman, J. Sjödahl, J. Sjöquist, I. Pecht, *Biochem. Biophys. Res. Commun.* **1978**, *85*, 608.
[54] W. A. Wallner, M. J. P. Lawman, M. D. P. Boyle, *Appl. Microbiol. Biotechnol.* **1987**, *27*, 168.
[55] Y. Kanamaru, S. Nagaoka, Y. Kuzuya, *Anim. Sci. Technol.* **1992**, *63*, 385.
[56] M. D. P. Boyle, W. A. Wallner, G. O. Von Mering, K. J. Reis, M. J. P. Lawman, *Mol. Immunol.* **1985**, *22*, 1115.
[57] K. Saha, F. Bender, E. Gizeli, *Anal. Chem.* **2003**, *75*, 835.
[58] B. Nilsson, T. Moks, B. Jansson, L. Abrahmsén, A. Elmblad, E. Holmgren, C. Henrichson, T. A. Jones, M. Uhlen, *Protein Eng., Des. Sel.* **1987**, *1*, 107.
[59] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, D. Batra, *arXiv preprint arXiv:1611.07450* **2016**.
[60] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* IEEE, Las Vegas, NV, June **2016**, pp. 2921–2929.
[61] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, in *Proc. IEEE Inter. Conf. on Computer Vision,* IEEE, Venice, October **2017**, pp. 618–626.
[62] P. Katira, A. Agarwal, T. Fischer, H. Chen, X. Jiang, J. Lahann, H. Hess, *Adv. Mater.* **2007**, *19*, 3171.
[63] X. Jiang, H.-Y. Chen, G. Galvan, M. Yoshida, J. Lahann, *Adv. Funct. Mater.* **2008**, *18*, 27.
[64] H.-Y. Chen, J. Lahann, *Langmuir* **2011**, *27*, 34.
[65] X. Deng, C. Friedmann, J. Lahann, *Angewe Chem. Int. Ed Engl.*, **2011**, *50*, 6522.
[66] H.-Y. Chen, Y. Elkasabi, J. Lahann, *J. Am. Chem. Soc.* **2006**, *128*, 374.

© 2024 The Author(s). Small Structures published by Wiley-VCH GmbH