

DISSERTATION

TUNING INTERFACIAL BIOMOLECULE INTERACTIONS  
WITH MASSIVELY PARALLEL NANOPORE ARRAYS

Submitted by

Dafu Wang

School of Advanced Materials Discovery

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2021

Doctoral Committee:

Advisor: Matt. J. Kipper

Co-Advisor: Chris D. Snow

Travis S. Bailey

Tim J. Stasevich

Copyright by Dafu Wang 2021

All Rights Reserved

## ABSTRACT

### TUNING INTERFACIAL BIOMOLECULE INTERACTIONS WITH MASSIVELY PARALLEL NANOPORE ARRAYS

This project studied interfacial interactions of macromolecules with nanoporous materials, with an ultimate goal of exploiting these interactions in functional biomaterials. We quantified interaction forces and energies for guest molecules threaded into the pores of protein crystals via nano-mechanical atomic force microscopy (AFM) pulling experiments. We demonstrated that both double-stranded DNA and poly(ethylene glycol) are rapidly absorbed within porous protein crystals, where they presumably bind to the inner “wall” surfaces of the protein crystal nanopores. These “guest” molecules can be retrieved from the “host” crystal by chemically modified AFM tips, enabling precise measurements of the adhesion forces and interaction energies.

Based on these experiments, machine learning approaches were developed to classify hundreds of thousands of individual force-distance curves obtained in the AFM experiments. Furthermore, we showed that the interactions between protein crystal “hosts” and “guest” macromolecules can be used to modulate cell behavior, by presenting cell adhesion ligands tethered to different lengths of macromolecules that thereby modulate the maximum traction force cells can apply before rupturing bonds tethering the adhesion ligand to the porous protein crystal interior. This method affords the opportunity to create biomaterials that store an internal reservoir of cell-specific signals that can be presented to independently modulate the behavior of different cell populations in a single material.

In the first chapter, some recent advancements, and methodologies of measuring interfacial biomolecule interactions are reviewed and compared. The reviewed technics include atomic force microscopy, fluorescence recovery after photobleaching, the total internal reflection fluorescence, confocal microscopy, and optical tweezers. Furthermore, this chapter interduces the application of machine learning to assist the interfacial biomolecule interaction studies, especially the AFM measurements. This chapter further prospects of the future of interfacial biomolecule interactions studies. In the second chapter, the methodologies of probing and observing the surface of highly porous *Camphylobacter Jejuni* formed protein crystals (CJ protein crystals) by high-resolution AFM are introduced. Throughout this chapter, the morphologies of CJ protein crystals are comprehensively investigated by AFM and have been discussed in this chapter. In the third chapter, for the first time, the interactions of DNA with porous protein crystals are quantitatively measured by high-resolution AFM and chemical force microscopy. The surface structure of protein crystals with unusually large pores was observed in liquid via high-resolution AFM. Force-distance (F-D) curves were also obtained using AFM tips modified to present or capture DNA. The interactions of DNA molecules with protein crystals to be quantitatively studied while revealing the morphology of the protein crystal surface in detail, in buffer, reveals how a new protein-based biomaterial can be used to bind DNA guest molecules. In the fourth chapter, strategies of machine learning are introduced which pioneered the use of machine learning to classify and cluster the interaction patterns between DNA and protein crystals, enabling us to process thousands of F-D curves collected by AFM. Finally, in the fifth chapter, we quantitatively measure and take advantage of the interaction between poly(ethylene glycol) (PEG)-arginine-glycine-aspartic acid (RGD) complex and nanoporous protein crystals to understand how non-covalent surface presentation of peptide adhesion ligands can influence cell behavior. Through AFM, F-D curves

of interactions between PEG-RGD and host protein crystals were obtained for the first time. Furthermore, a strategy is developed that enables us to design surfaces that non-covalently present multiple different ligands to cells with tunable adhesive strength for each ligand, and with an internal reservoir to replenish the precisely defined crystalline surface.

## ACKNOWLEDGEMENTS

First of all, I would like to appreciate my Ph.D. advisor Dr. Matt J. Kipper and co-advisor Dr. Chris D. Snow at Colorado State University, with my deepest gratitude for their mentorship, who give me education and encouragement to pursue my dream. They inspire my research and help me with professional guidance. They are outstanding role models, not only as scientists with hard-working ethics, rich theoretical knowledge, strong experimental ability, and extensive experiences, but also as persons with great charm and personalities. They lead me by their examples of how I should work towards being a good researcher. They encourage me to be an active researcher and think outside the box. They not only guide me to be professional, but also encourage me to learn and engage new things in order to advance my knowledge. As a student under their guidance and education, Dr. Kipper and Dr. Snow provide me with passion, support and challenge for which make me be a better Ph.D. candidate as I continue my pathway to academia.

I would like to thank my Ph.D. committee members, Dr. Travis S. Bailey and Dr. Tim J. Stasevich for their great support and inspired ideas. My committee members are all vital to the proposals, as well as the whole research. Throughout the entire study, I fortunately have their generous help and professional guidance.

My appreciation also goes to the faculty members and colleagues who gave me guidance and helped me during the last four and half years. I would like to thank Dr. Laurance A Belfiore and Dr. Jianguo Tang for inspiring my potential at CSU. I also feel grateful for the generous help from my colleagues, Dr. Mohammadhasan Hedayati, Dr. Liszt Yeltsin, Dr. Tara Wigmosta, Dr. Jessi Vicek, Dr. Ann Kowalski, Abby Ward, Julius Stuart, Roberta Sabino, Dr. Ellen Brennan-

Pierce, Dr. Yao Wang, Dr. Yanxin Wang, and Dr. Shun Yao Wu. They passed me their professional experience with great patience and enthusiasm. I would not achieve my goals without them, who have greatly guided, encouraged, and enlightened me for being a good researcher.

Finally, I would also like to express my sincere appreciation to my parents, Dr. Liming Wang and Prof. Zhengyuan Wang, who always stand with me and support me overcome all the difficulties step by step, block by block, even in the hardest time. They made me who I am today with their unconditional love.

Pursuing a Ph.D. is supposed to reveal the truth about the nature and science. In fact, it taught me more about myself. Over the past four years, this journey taught me that my curiosity is never satisfied, and achievements themselves aren't what count, it's the fact that I worked for them. When I see myself in this bigger perspective, it makes me think differently. It allows me to realize more clearly about who I am, what my dream is, and what I am after with a much deeper sense of reverence. As T. S. Eliot put it:

We shall not cease from exploration  
And the end of all our exploring  
Will be to arrive where we started  
And know the place for the first time

DEDICATION

*To my family*

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	v
DEDICATION .....	vii
CHAPTER 1: CURRENT ADVANCEMENT OF MEASURING INTERFACIAL BIOMOLECULE INTERACTION .....	1
Chapter 1 Overview .....	1
1.1 Review of Methodology .....	2
1.1.1 Atomic Force Microscopy .....	2
1.1.1.1 Recent Development of AFM Technology .....	5
1.1.2 Measurement of Fluorescence .....	8
1.1.3 Optical Tweezers .....	9
1.1.4 Machine Learning to Assist Interfacial Biomolecule Interaction Study .....	10
1.2 Prospects and Outlook .....	11
REFERENCES .....	13
CHAPTER 2: HIGH RESOLUTION IMAGING OF BIOACTIVATED NANOPOROUS PROTEIN CRYSTALS BY ATOMIC FORCE MICROSCOPY .....	19
Chapter 2 Overview .....	19
2.1 Introduction .....	20
2.2 Experimental .....	23
2.2.1 Crystal's Growth .....	23
2.2.2 Crystal Crosslinking and Immobilization .....	24
2.2.3 AFM Probing .....	26
2.3. Results and Discussion .....	26
2.3.1 AFM Imaging on Surface Morphology .....	26
2.3.1.1 Imaging of Large Areas .....	28
2.3.1.2 Imaging of Different Orientation .....	29
2.3.1.3 Imaging of Defected Surface .....	30
Chapter 2 Summary .....	31
REFERENCES .....	32
CHAPTER 3: MEASURING INTERACTION OF DNA WITH PROTEIN CRYSTALS BY ATOMIC FORCE MICROSCOPY .....	34
Chapter 3 Overview .....	34
3.1 Introduction .....	35
3.2 Experimental .....	39
3.2.1 Materials .....	39

3.2.1.1 CJ Crystals Growth .....	39
3.2.1.2 CJ Crystal Crosslinking and Immobilization .....	40
3.2.2 AFM Tip Modification .....	41
3.2.3 DNA Loading and AFM Probing .....	43
3.3. Results and Discussion .....	47
3.3.1 AFM Imaging on Surface Morphology .....	47
3.3.2 Modification of AFM Tips .....	48
3.3.3 DNA Loading .....	49
3.3.4 Interaction of DNA and CJ Protein Crystals .....	51
Chapter 3 Summary .....	60
REFERENCES .....	61
CHAPTER 4: DISCOVERING THE INTERACTION PATTERNS OF DNA WITH PROTEIN CRYSTALS BY MACHINE LEARNING .....	63
Chapter 4 Overview .....	63
4.1 Introduction .....	64
4.2 Methodology and Experiments .....	69
4.2.1 Materials and AFM Probing .....	69
4.2.2 Automatic Classification with the Random Forest Classifier .....	69
4.2.3 Pattern Discovery with Affinity Propagation .....	70
4.3. Results and Discussion .....	74
4.3.1 Supervised Machine Learning .....	74
4.3.1.1 Binary Classification .....	74
4.3.1.2 Multi-category Classification .....	75
4.3.2 Unsupervised Machine Learning .....	77
4.3.2.1 Cluster of F-D Curves .....	77
4.3.2.2 Interpretation of Interaction Patterns .....	79
Chapter 4 Summary .....	87
REFERENCES .....	88
CHAPTER 5: NON-COVALENT PRESENTATION OF PEPTIDE ADHESION LIGANDS WITHIN POROUS PROTEIN CRYSTALS: ATOMIC FORCE MICROSCOPY MEASUREMENTS AND CELL ATTACHMENT STUDIES .....	90
Chapter 5 Overview .....	90
5.1 Introduction .....	91
5.2 Experimental .....	96
5.2.1 Materials .....	96
5.2.2 Synthesis and Characterization of PEG-RGD with <sup>1</sup> H NMR .....	96
5.2.3 PEG Loading into Porous Protein Crystals .....	97
5.2.4 Crystal Immobilization .....	97
5.2.5 AFM Tip Modification .....	97

5.2.6 AFM Imaging and Force-Distance Measurements .....	98
5.2.7 Cell Culture and Imaging .....	100
5.3 Results and Discussion .....	100
5.3.1 PEG-RGD Loading .....	100
5.3.2 Surface Morphology of the Protein Crystals by AFM .....	101
5.3.3 Modification of AFM Tips .....	103
5.3.4 Interaction of PEG <sub>10KDA</sub> -RGD and Nanoporous Protein Crystals .....	104
5.3.5 Modulation of Cell-adhesion to Protein Crystals .....	111
Chapter 5 Summary .....	114
REFERENCES .....	116
CHAPTER 6: SUMMARY .....	119
APPENDIX A: CHAPTER 3 SUPPORTING INFORMATION .....	120
APPENDIX B: CHAPTER 4 SUPPORTING INFORMATION .....	128
B.1 Experimental .....	128
B.1.1 Materials .....	128
B.1.1.1 CJ Crystal's Growth .....	128
B.1.1.2 CJ Crystal Crosslinking and Immobilization .....	129
B.1.2 AFM Tip Modification .....	130
B.1.3 DNA Loading and AFM Probing .....	131
REFERENCES OF APPENDIX B .....	135
APPENDIX C: CHAPTER 5 SUPPORTING INFORMATION .....	136
C.1 Materials .....	136
C.1.1 CJ <sup>OPT</sup> Nucleic Acid Sequence .....	136
C.1.2 CJ <sup>OPT</sup> Amino Acid Sequence .....	137
C.1.3 CJ <sup>OPT</sup> Mutations .....	137
C.2 AFM Tip Modification .....	137
C.3 Cell Culture Experiments .....	140
C.4 AFM Imaging and Force-Distance Measurements .....	141
C.5 Loading of Guest Molecules .....	143
REFERENCES OF APPENDIX C .....	145

# CHAPTER 1: CURRENT ADVANCEMENT OF MEASURING INTERFACIAL BIOMOLECULE INTERACTION

## Chapter 1 Overview

Accurately measurements and analysis of interfacial biomolecule interactions and cell-material interaction have been attracting people's attention over the recent decades.<sup>[1]</sup> By deepening the understanding of the interfacial biomolecule and cell-material interaction patterns, a series of vital scientific research problems, especially in biomedical areas, have been moving forward by giant leaps. Scientist can now sense and repair the DNA structure damages that may be causes of cancer by evaluating the interaction patterns between DNA and protein.<sup>[2, 3]</sup> Through visualizing single cell dynamics in thick tissues, people are now able to fully evaluate the dynamic nature of metastatic disease including cancer.<sup>[4-6]</sup> Taking advantage of interfacial biomolecule adhesion, scientists use adhesive surface coatings to functionalize nanomedicines with antibodies for cancer treatment.<sup>[7]</sup> In this context, the importance of such research has been reflected, and some new technologies and methodologies related to it have begun to emerge. In this chapter, some current methodologies and advancements of measuring interfacial biomolecule interactions and cell-material interactions are reviewed. By reviewing some cutting-edge technologies, we will be able to discuss, understand, and look forward to the prospects of the research of interfacial biomolecule and cell-material interactions. In the first chapter, the recent development of atomic force microscopy (AFM), fluorescence recovery after photobleaching (FRAP), the total internal reflection fluorescence (TIRF), confocal microscopy, and optical tweezers are reviewed and compared. Furthermore, this chapter introduces the application of machine learning to assist the interfacial biomolecule interaction studies, and further prospects of how people will study interfacial biomolecule interactions in the future.

## 1.1 Review of Methodology

### 1.1.1 Atomic Force Microscopy

As a widely used precision instrument, atomic force microscopy (AFM) is capable of providing the detailed nanoscale morphology of biomaterials.<sup>[8-11]</sup> AFM allows imaging of soft materials in favorable environmental conditions. It enables the samples, especially the bioactivated samples to be characterized in liquid phases.<sup>[9, 12]</sup> Moreover, AFM does not require samples to be labeled or stained.<sup>[12]</sup> AFM can also provide nanometer-level single-molecule high-resolution imaging,<sup>[13]</sup> 3-D imaging models of overall or single-molecule morphology,<sup>[14]</sup> and force–distance (F–D) curves that exhibit the characteristic point of separation of the tip from the surface and single-molecule adhesion events.<sup>[15]</sup>

AFM is not only capable of imaging nanoscale morphology, but also capable of quantitatively measuring nanoscale mechanical behaviors.<sup>[9, 12, 16]</sup> AFM can be a powerful tool for the study of the surfaces of samples, and most significantly for biological applications, in buffered solutions.<sup>[13, 17]</sup> Previous AFM studies of biological samples (Table 1.1) have successfully provided surface morphology and mechanical characterizations of soluble proteins,<sup>[18-20]</sup> lipid mono and bi-layers,<sup>[20]</sup> and membrane proteins.<sup>[18, 19, 21]</sup> The AFM technique also provides characterization of surfaces that can be used for structural studies of biological macromolecules.<sup>18</sup> However, the traditional methodology for measuring two-dimensional surface are not suitable for nanoscale characterization of some bioactivated materials with nano 3D structure but only maintain their bioactivation in the liquid phase, such as nanoporous protein crystals.<sup>[22-24]</sup> With 3D nano-structures, it is necessary to revise some strategies in order to measure the functionalized materials. These revisions of strategies include sample immobilization, AFM tip surface modification and activation, which will significantly impact the ability of AFM, as well as the

quality of characterization. The experimental methods must be extremely precise to achieve these goals. When measuring the nanoscale interfacial biomolecule interactions, the resolution and accuracy of AFM measurement can be affected by several factors, including the size of the AFM tips, the stability in buffer solution, the physical morphology of the samples, and how precisely the feedback system contours the tip over samples.<sup>[10]</sup> Due to its ability to apply very small controlled forces (pN level) between single molecules,<sup>[25]</sup> based on the intensity of approach and retraction signals, the AFM can be used as a precise indenter to determine the mechanical behaviors of materials.<sup>[26]</sup> Force curves are obtained to extract surface mechanical behaviors (both covalently and non-covalently) and interactions between biomolecules. AFM imaging modes such as Bruker's PeakForce nanomechanics enables the recording of thousands of force-distance (F-D) curves in a single experiment without interrupting the imaging process, and the registration of these forces to precise locations on a surface. The resulting force-curves can be extracted and correlated to the sample topography.<sup>[27-29]</sup>

Specifically, in the area of functional biomaterials, AFM has a brilliant capability to characterize the biomolecule samples at a molecular-level resolution under both air and liquid phases.<sup>[25, 26, 30]</sup> Thus, the characterization of biomaterials can be performed in buffer solutions, which allows biomaterials to maintain the same bioactivation that they would have in a physiological environment. These unique features distinguish AFM from conventional characterization techniques.<sup>[25, 29]</sup> So far, scientists have made effort to characterize biomolecules via AFM in many ways.<sup>[31]</sup> The published results in this field include the visualization of the mechanical behaviors process,<sup>[32]</sup> measuring volume, position, and bend angle of proteins,<sup>[29]</sup> comparison of protein-induced DNA bend angles determined by AFM tangent methods,<sup>[12, 33, 34]</sup> mapping the distribution of surface morphology on 2D bio-substrates,<sup>[32]</sup> imaging crystals in

liquid,<sup>[9, 35]</sup> high-speeding high-resolution AFM imaging on biomolecules,<sup>[36]</sup> and visualization of quantum dots on proteins.<sup>[37]</sup> From these efforts and attempts, we learn that AFM has great promise and potential in studying nanoscale morphology and biomechanics, including interfacial biomolecule interactions and cell-material interactions.

However, for the interaction measurement, AFM has shown some shortcomings in some recent studies. For example, the interaction force between some specific proteins and DNA structures is far below 1 pN.<sup>[2, 3, 38]</sup> Such small forces are significantly lower than the trustable minimum detection limit of most of AFM.<sup>[12, 27]</sup> In the meantime, AFM is capable doing single-molecule level measurements.<sup>[39, 40]</sup>

In the area of AFM measurements on cell-material interactions, the Messersmith Group at UC Berkeley has made great contributions. They applied AFM to measure the mechanical process including image the elastic properties of live cells. They used force versus displacement (F-Z) curves as an evaluation tool to approach part of the curve to extract Young's modulus, on benign and cancerous cell lines and PPAm gel samples.<sup>[14]</sup> The Messersmith Group has also published results on characterization of the microscale adhesion ability using AFM-based colloidal probe technique.<sup>[30]</sup> They have used modified Si<sub>3</sub>N<sub>4</sub> AFM cantilevers to obtain force–distance (F-D) curves to show the characteristic point of separation of the tip from the single dopa residue contacting a wet metal oxide surface and single-molecule adhesion events.<sup>[26]</sup>

To better summarize the technology of measuring interfacial biomolecule interactions and cell-material interactions using AFM, some featured publication of AFM methodologies are listed to the Table 1.1.

### 1.1.1.1 Recent Development of AFM Technology

AFM is a powerful tool to study interfacial biomolecule interactions.<sup>[12]</sup> The latest developments of AFM technology include time-lapse and high-speed AFM,<sup>[10]</sup> which helps scientists showed that RNA polymerase could transcribe DNA while adsorbed on mica.<sup>[12, 41]</sup> Fluorescence-coupled AFM is another new AFM technology, which enables the AFM to analyze complex conformations of DNA and proteins.<sup>[42-44]</sup> In addition to them, a combination of atomic force and total internal reflection fluorescence microscopy (TIRF-AFM) to visualize the location/co-location of DNA and protein reveals their differences by TIRF while imaging the surface morphology in a high-resolution by AFM.<sup>[45, 46]</sup> Dual-resonance-frequency-enhanced electrostatic force microscopy enabled high resolution visualization of the path of dsDNA around a nucleosome.<sup>[47, 48]</sup>

**Table 1.1** Selected Publications of Measuring Interfacial Biomolecule Interactions and Cell-material Interactions Using AFM

Materials	Tip and Scan Range	Methods and Conditions	Results	References
Adsorbed proteoglycan mimetic nanoparticles	5 $\mu\text{m} \times 5\mu\text{m}$ AFM topographic images acquired using the DNP tip (20 nm radius)	The surfaces are characterized by PeakForce tapping AFM, both in air and in aqueous pH = 7.4 buffer, and by PeakForce quantitative nanomechanics mode with high spatial resolution.	AFM images toward glycolyx-mimetic model surfaces	[49]
Barnacle underwater adhesive protein	Using Nanosensors PPP-NCHR probes	In a simulated seawater condition (pH = 8.0, I = 600 mM), nanoscale morphologies and its mutant at different pHs were examined by AFM imaging. After being dried in air, the samples were scanned by intermittent contact mode in air.	Surface imaging and AFM-based force spectroscopy	[30]
PEMs, PEM19 + , PCN(CS-CHI), and Vascular Endothelial Glycocalyx	The calibrated probe PF-QNM-LC-CAL with a tip radius of 65 nm, the scan size was typically 5 $\times$ 5 $\mu\text{m}$	AFM Measurements in PeakForce Quantitative Nanomechanical Property Mapping (PFQNM) Mode. The peak force QNM AFM study was done using a BioScope Resolve BIOAFM (Bruker) with a Nanoscope V controller. The measurements were performed under ambient conditions at room temperature in deionized (DI) water to characterize the morphology and mechanical properties of the sample.	Surface image and peak force quantitative nanomechanical property mapping (PF-QNM) of the PEM surfaces in fluid	[50]
M(20-10) and M(25-5) pectin-chitosan membrane	In an area of 2 $\mu\text{m} \times 2\mu\text{m}$ , AFM was using ScanAsyst Air probes	AFM images were obtained using tapping mode from a scanning image probe processor version 4.2.2.0 software. AFM was performed at room temperature in air at a rate of one-line scan per second.	3D AFM images on membrane surface	[51]
Carbonized polysaccharides from basil seed	At an average thickness value of 4.5 $\pm$ 1.6 nm, comparable to 3-8 layers of graphene flakes	N/A	AFM image on surface	[52]
Tissue culture polystyrene (TCPS) modified with heparin-terminated	A Bruker SNL probe was used with a tip radius of 2 nm, in an	An automated thermal tune was performed before each imaging session. The Peakforce setpoint, amplitude, and frequency were manually adjusted to obtain stable imaging	3D AFM images on surface	[53]

polyelectrolyte multilayers (PEMs), transforming growth factor beta (TGF $\beta$ ), and fibronectin (FN)	area of 4 $\mu\text{m} \times 4 \mu\text{m}$ and 800 nm $\times$ 800 nm	conditions and to minimize noise. The typical line scan rate was 1–1.5 Hz.	
Benign and cancerous cell lines (NIH 3T3 fibroblasts, NMuMG epithelial, MDA-MB-231 and MCF-7 breast cancer cells), and PPA $\text{m}$ gel samples	Tipless AFM cantilevers CSC38 or BL-TR400PB, modified with 5 $\mu\text{m}$ diameter silicon dioxide beads, scan at least 60–70 cells per cell line	The sample temperature was kept constant at 37 $^{\circ}\text{C}$ , while the bead was glued to the end of the cantilever using UV-curable glue under control of the inverted optical microscope. The typical spring constant of both cantilevers is 0.02–0.05 N/m. The accurate value was determined using the laser Doppler vibrometer system. The force set point was chosen individually for all samples to obtain maximal indentation depth around 500 nm. At least 60–70 cells per cell line, and 3 PPA $\text{m}$ gel samples were analyzed.	Elastic properties of live cells are evaluated from force versus displacement (F-Z) curves, to the approach part of the curve to extract Young's modulus [14]
Plant-derived proanthocyanidins	Au-coated AFM probes were coated with type I collagen solution diluted to 10 $\mu\text{g}/\text{ml}$ using 0.1 M acetic acid. Placed in collagen solution then thoroughly rinsed in Millipore water followed by drying in nitrogen stream	Force measurements of interaction were performed by placing the collagen-coated AFM tip into contact with the collagen coated substrate, allowing it to dwell and retract away from the surface in presence or absence of cross-linking solutions. All force measurements were performed with Asylum MFP-3D AFM Instrument at room temperature. The substrates were placed in a fluid cell which was then filled with about 2 mL 10 mM PBS (pH = 7.2) solution. The spring constant of each AFM tip was calibrated in the buffer by the thermal fluctuation's method. By measuring the tip deflection and its displacement, force curves (force vs tip separation) were plotted.	Force measurements of interaction: (force vs tip separation) [54]
Polymer binder from glass surface: PAA, the PPyMAA copolymer with 70 mol % pyrene and 30 mol % MAA structure	N/A	AFM force–extension curves were measuring pulling polymer binders from glass surface in 1 M LiCl, pH 6.7.	Force–extension curves on single-molecule experiment [55]
Introduce C-terminal cysteine into two cysteines present in the native Nb form a disulfide bond stabilizing the protein's tertiary structure	Types of cantilevers: MLCT (cantilever C) and Biolever mini (BL-AC40TS-C2).	The Nb–GFP bond strength was tested in a series of measurements at various pulling speeds ranging from 300 to 10 000 $\text{nm s}^{-1}$ and for different attachment geometries of GFP to the surface. A single measurement cycle consisted of approach, short dwell at the surface and retraction of the Nb-functionalized cantilever with constant velocity.	the cantilever functionalized with the nanobody is brought in contact with the GFP-decorated surface and then retracted and force extension curves are recorded [17]
Single native proteins	Multiple self-modified soft AFM cantilever (spring constant, $k = \sim 0.01\text{--}0.5 \text{ N m}^{-1}$ )	(reviewed multiple AFM imaging protocols)	high-resolution imaging of proteins in the native unperturbed state, and force-distance curves [13]
Protein films and crystals	3.8 $\mu\text{m} \times 3.8 \mu\text{m}$ and 5 $\mu\text{m} \times 5 \mu\text{m}$ , in the tapping mode using Ni-Au NSC-11 and NSC-18 cantilever	Film imaging was performed in a dry atmosphere; crystal imaging was performed in a buffer, with crystals kept stuck to the bottom of the glass slide either with grease or by gravity. The imaging of crystals was performed in order to ensure the optimal resolution. The set-point range and other acquisition parameters were frequently changed in order to achieve accurate imaging.	Image and profile of surface morphology of protein film and crystals [18]
Single dopa residue contacting a wet metal oxide surface	Chemically modified Si3N4 AFM cantilevers	A dopa-functionalized AFM tip was lowered at a constant rate onto a wet surface to a maximum load of 15–20 nN and then retracted at the same rate while force versus extension was recorded.	Force–distance (F–D) curves exhibit the characteristic point of separation of the tip from the surface and single-molecule adhesion events [26]
Protein crystals and wide variety of impurities that are incorporated into crystals of proteins,	In an area range of 470 nm $\times$ 470 nm to 80 $\mu\text{m} \times 80 \mu\text{m}$ .	(reviewed multiple AFM imaging protocols)	2D nuclei on the surfaces of protein and virus crystals visualized and dislocations on [56]

nucleic acids, and viruses.			the surfaces of a variety of protein crystals as visualized in situ	
Crystallization of ferritins from solutions containing 5% (w/w) of their inherent molecular dimers.	Standard SiN tips, and the tapping drive frequency was adjusted in the range 25–31 kHz to the resonance value for specific tip used.		AFM imaging of the crystallization processes was conducted in tapping mode. We used room temperature stabilized to $23.06 \pm 0.3^\circ\text{C}$ ; for calibration, tests, and determination of the maximum resolution of the method of 16 Å. The AFM imaging parameters were selected such that the imaging does not affect either the surface structure even at the molecular level or the processes of molecular incorporation.	High resolution AFM image on crystal surface [57]
The patterned self-assembled monolayers (SAMs)	In an area range of 500 nm × 500 nm and 5 μm × 5 μm, Olympus Si3N4 rectangular cantilevers OMCL-RC800PB-1 with a sharpened pyramidal tip with a spring constant of 0.75 N m <sup>-1</sup> , coated Cr/Au on both sides of cantilevers.		Si <sub>3</sub> N <sub>4</sub> rectangular cantilevers were with a sharpened pyramidal tip with a spring constant of 0.75 nm <sup>-1</sup> were used to take AFM topographic images, the osmium-coated cover glass plates, and the gold films sputtered on the osmium-coated plates. These topographic images were observed with a constant-force-mode AFM with a cantilever bending force of ca. 10 nN and a scan rate of 1 Hz.	AFM, FFM, and PFM-AFM topographic images on surface. [58]
Streptavidin 2-D crystals formed on biotin-containing planar lipid layers	The scanning size is around 500 nm x 500 nm, in the tapping mode, using cantilevers (BioLever mini; Olympus) with a spring constant of 0.1–0.3 N/m.		High-speed AFM imaging was carried out in the tapping mode, using small cantilevers with a spring constant of 0.1–0.2 N/m and a resonant frequency of 0.8–1.2 MHz in water.	Visualizing the biomolecular processes by AFM [32]
Damaged DNA-binding promoted dimerization of the heterodimer.	N/A		AFM was used to characterize the molecular topology of UV-DDB and to monitor substrate-induced changes in intermolecular interactions, and exam the oligomeric states of multicomponent complexes and to ascertain the nature of protein–protein interactions of globular proteins.	AFM imaging of dimerization of the heterodimer [38]
Bacteriorhodopsin (bR) molecules at the free interface of the crystalline phase that occurs naturally in purple membrane	An amorphous carbon tip was grown on the original cantilever tip by electron beam deposition. The tip length was adjusted to about 1 μm and the tip apex was sharpened by plasma etching (4 nm in radius).		The cantilever has a resonant frequency of 1 MHz in water and a spring constant of 0.1–0.2 N/m. An amorphous carbon tip was grown on the original cantilever tip by electron beam deposition.	HD 2-D imaging and dynamics of bacteriorhodopsin in 2D crystal. [35]
Actin filaments, movement of kinesin–gelsolin along a microtubule, Myosin V head et al.	(various scanning ranges, various tips)		High speed AFM imaging on protein crystals in a maximum rate of 30–60 ms/frame for a scan range of ~250 nm with a weak tip–sample interaction force being maintained.	AFM images of protein crystals [36]
Protein-DNA complex	N/A		PeakForce Tapping (Bruker) AFM, AFM was used to obtain information about interactions between full-length human XPA gene and a DNA substrate with a site-specific bulky adduct, AAF-dG, but the small size of the protein presented limitations in terms of resolution using tapping AFM.	Single molecule characterizations to gain a better understanding of XPA gene's molecular role in nucleotide excision repair [16]
DNA on protein crystals	N/A		AFM scans the samples on mica to produce topographical data. Suspension of the microscope with bungee cords provides some protection from interfering vibrations. In AFM tapping mode a cantilever (with probe tip at the end) is driven to oscillate vertically near its resonance frequency. The AFM scanner allows the probe to track a sample field in the X–Y dimensions. In tapping mode, the oscillation amplitude is kept constant.	AFM imaging of protein volume, position, and bend angle. 2D and 3D AFM images [29]
UvrB-QD, DNA repair protein conjugated to a quantum dot (QD)	Scanning range is 3-D (500 nm x 500 nm x 20 nm)	N/A		Visualization of quantum dots using AFM, measurement of complexes formed on nicked DNA fragment [37]

### 1.1.2 Measurement of Fluorescence

Fluorescence recovery after photobleaching (FRAP) is a method that measures diffusion coefficients of nanoscale and molecular fluorescent analytes. FRAP enables determination of the diffusion of fluorescent labeled target biomolecules in polymeric matrices, adsorbates on reversed-phase chromatographic surfaces, lipid membranes, and proteins.<sup>[59-64]</sup> FRAP is also capable of measuring diffusion of proteins on flat surfaces. The high-energy lasers can be used to photobleach fluorophores; the subsequent transport of labeled protein back into the bleached area, and the diffusive rate can be measured during this process.<sup>[64-68]</sup> FRAP is a power tool for studying diffusion and protein dynamics in living cells.<sup>[69]</sup> The FRAP experiments are ensemble measurements that can be fit via finite difference modeling, and interpreted via molecular simulations. Through FRAP technology, scientists studied the diffusion of nanoparticles in nanoporous materials.<sup>[70-72]</sup> Recent studies of FRAP investigated protein transport and exchange in chromatographic media, and measured the effective exchange rates in protein-loaded particles.<sup>[64]</sup> Another recent FRAP study elucidated the effect of dextran modification on translation mobilities of adsorbed proteins.<sup>[73]</sup> As a complement to the bulk measurements, total internal reflection fluorescence (TIRF) technology can quantify interfacial biomolecule interactions via single-molecule methods such as quantifying the motions of individual molecules.<sup>[74-76]</sup>

In addition to FRAP and TIRF, confocal microscopy as a quantitative measurement of fluorescent intensity, and method of monitoring live cells, is also commonly applied to measure the interfacial biomolecule interactions and cell-material interactions.<sup>[77-79]</sup> One of the most important application of confocal microscopy in measuring cell-material interaction is monitoring the cell spreading, as well as monitoring the bioactivated substance within the live cells.<sup>[80]</sup>

Confocal microscopy is capable of monitoring the dynamic process of cell-material interactions, such as dynamic interactions between cellular forces and actomyosin contractility in 3D environments.<sup>[81]</sup> Some latest developed confocal microscopy instruments are focusing on the accurate 3D reconstruction of a sample from high-resolution stacks of images.<sup>[82]</sup> Another highlighted development on confocal microscopy is “*in vivo* confocal microscopy” (IVCM). Due to its excellent performance, IVCM has been applied to the clinical diagnosis. IVCM enables the quantification of nerve parameters, and non-invasively examine immune/inflammatory cells within a live person.<sup>[79]</sup> Another trend of newly developed confocal microscopy devices is being combined with artificial intelligence to improve its resolution and accuracy of positions. Recent reports show the deep learning model has been applied to IVCM in order to automate sub-basal corneal nerve fiber segmentation and evaluation with IVCM.<sup>[83]</sup> Furthermore, 3D super-resolution imaging inside cells has become the next focus area. Combining with the confocal microscopy section within the instrument, single-molecule localization microscopy is able to visualize biological targets on a nanoscale, achieve multiplexed 3D super-resolution imaging with up to 20 nm planar.<sup>[84]</sup> These latest developments of confocal microscopy will make the methodology of measuring interfacial biomolecule and cell-material interactions more diverse. Especially, dynamic tracking directly *in vivo*, will greatly expand the research scope of cell-material interaction. As the accuracy of confocal microscopy has increased significantly, the accuracy of fluorescence measurement related to the interfacial biomolecule and cell-material interaction studies will also reach an unprecedented level in the foreseeable future.

### **1.1.3 Optical Tweezers**

As a new technology, optical tweezers can achieve a sensitivity within 0.1 pN during the measurement of biomechanics.<sup>[2, 3]</sup> Moreover, brand new methodology such like “DNA tightrope

assay” method has been developed with the help of the excellent performance of optical tweezers. Based on the ability to anchor both ends of a long DNA molecule on two nearby points via electrostatic interaction, with the rest of the DNA suspended in between the two points, forming DNA tigtropes. The DNA tigtrope assay lifts up the DNA molecules, and thereby lift up the location of interactions between DNA and other biomolecules such as proteins, allowing the optical tweezers to measure the interaction of DNA with other biomolecules in space, therefore minimizing any potential surface impact.<sup>[3, 29]</sup> The trustable limit of detection for measuring the adhesive force between DNA and proteins is able to reach as small as 0.1 pN during the force measurement.<sup>[3, 29]</sup> So far, according to the published results, some interfacial biomolecule interaction measurements using optical tweezer have already achieved a sub-pN level of sensitivity.<sup>[85]</sup>

#### **1.1.4 Machine Learning to Assist Interfacial Biomolecule Interaction Study**

Machine learning has been playing an increasingly important role to enhance image recognition, substance analysis, and autonomous operation of AFM, as a strategy to reduce the human reliance of observation and analysis.<sup>[86-88]</sup> Previously published studies reported that supervised learning has been used to enable an automated and reproducible analysis pipeline for biological samples in AFM,<sup>[87]</sup> and has also been used for the analysis of nonlinear responses to the bimodal drive at harmonics and mixing frequencies, thus to obtain quantitative material properties at high speeds and with enhanced resolution.<sup>[88]</sup> In the meantime, in addition to AFM studies, machine learning has been broadly useful in the study of interfacial biomolecules interactions. In particular, the random forest algorithm for supervised machine learning has been applied to predict the similarity of protein-protein interaction,<sup>[89]</sup> and predict the protein-protein interaction sites.<sup>[90]</sup> The affinity propagation algorithm for unsupervised machine learning has been

applied to detect protein complexes in the protein-protein interaction networks.<sup>[91]</sup> The affinity propagation has also been applied to the area in clustering the protein interaction graphs.<sup>[92]</sup> In addition to random forest and affinity propagation algorithms, other machine learning algorithms can also predict a potential energy landscape for the interatomic interactions of molecules,<sup>[93, 94]</sup> cluster the protein-protein interaction networks,<sup>[95]</sup> and predict the differentiation of a cell based on the environment.<sup>[96, 97]</sup> In sum, machine learning can be a powerful tool to automatically classify and cluster the interaction patterns between DNA-laden AFM tips and protein crystals.

## **1.2 Prospects and Outlook**

Over the past decades, the advancements of interfacial biomolecule interaction measurement has been pursued the higher resolution, and the higher accuracy. With the development of microscope technology, these two indicators are likely to reach their limit of existence in nature in the foreseeable future. I would like to take an example of building a submarine, even if the technology becomes more advanced, the depth of the Mariana Trench, the deepest place of the ocean on Earth, is and always will be only 10,984 meters deep. Microscopy technology might soon reach a similar limit in the near future. Does this mean that the development of interfacial biomolecule interaction measurement meets its ceiling? The answer is obviously no. On the one hand, the advancement of precision instrument processing technology has led to the rapid miniaturization and popularization of some originally cumbersome instruments. The development of optical tweezers can be a good example. Instruments that originally needed to be customized in the laboratory have gradually begun to be popularized and commercialized. The improvement of precision instrument processing technology also enables the instrument to adapt to more complex environments. The AFM that can work in the liquid phase, and the confocal

microscopy that is able to non-invasively examine the human body were unimaginable twenty years ago. These cutting-edge technologies will greatly promote the development of measuring interfacial biomolecule interactions.

Finally, with the popularization and application of artificial intelligence, machine learning, deep learning, and neural networks, microscopy technology will be likely to usher in a technological revolution in the near future. The combination of these technologies with microscope technology will fundamentally change the methodology of measurement of interfacial biomolecule interactions.

## REFERENCES

1. Pierschbacher, M. D.; Ruoslahti, E., Cell Attachment Activity of Fibronectin Can Be Duplicated by Small Synthetic Fragments of the Molecule. *Nature* **1984**, *309* (5963), 30-33.
2. Jang, S.; Kumar, N.; Beckwitt, E. C.; Kong, M.; Fouquerel, E.; Rapić-Otrin, V.; Prasad, R.; Watkins, S. C.; Khuu, C.; Majumdar, C., Damage sensor role of UV-DDB during base excision repair. *Nature structural & molecular biology* **2019**, *26* (8), 695-703.
3. Jang, S.; Schaich, M. A.; Khuu, C.; Schnable, B. L.; Majumdar, C.; Watkins, S. C.; David, S. S.; Van Houten, B., Single molecule analysis indicates stimulation of MUTYH by UV-DDB through enzyme turnover. *Nucleic acids research* **2021**, *49* (14), 8177-8188.
4. Paul, C. D.; Bishop, K.; Devine, A.; Paine, E. L.; Staunton, J. R.; Thomas, S. M.; Thomas, J. R.; Doyle, A. D.; Jenkins, L. M. M.; Morgan, N. Y., Tissue architectural cues drive organ targeting of tumor cells in zebrafish. *Cell systems* **2019**, *9* (2), 187-206. e16.
5. Staunton, J. R.; So, W. Y.; Paul, C. D.; Tanner, K., High-frequency microrheology in 3D reveals mismatch between cytoskeletal and extracellular matrix mechanics. *Proceedings of the National Academy of Sciences* **2019**, *116* (29), 14448-14455.
6. Jackson, S.; Meeks, C.; Vézina, A.; Robey, R. W.; Tanner, K.; Gottesman, M. M., Model systems for studying the blood-brain barrier: Applications and challenges. *Biomaterials* **2019**, *214*, 119217.
7. Black, K. C.; Yi, J.; Rivera, J. G.; Zelasko-Leon, D. C.; Messersmith, P. B., Polydopamine-enabled surface functionalization of gold nanorods for cancer cell-targeted imaging and photothermal therapy. *Nanomedicine* **2013**, *8* (1), 17-28.
8. Yang, J., AFM as a high-resolution imaging tool and a molecular bond force probe. **2004**, *41* (3), 435-449.
9. Wang, D.; Stuart, J. D.; Jones, A. A.; Snow, C. D.; Kipper, M. J., Measuring Interactions of DNA with Nanoporous Protein Crystals by Atomic Force Microscopy. *Nanoscale* **2021**, *13* (24), 10871-10881.
10. Dufrene, Y. F.; Ando, T.; Garcia, R.; Alsteens, D.; Martinez-Martin, D.; Engel, A.; Gerber, C.; Muller, D. J., Imaging modes of atomic force microscopy for application in molecular and cell biology. *Nat. Nanotechnol.* **2017**, *12* (4), 295-307.
11. Goldsbury, C. S.; Scheuring, S.; Kreplak, L., Introduction to atomic force microscopy (AFM) in biology. *Curr. Protoc. Protein Sci.* **2009**, *58* (1), 17.7. 1-17.7. 19.
12. Beckwitt, E. C.; Kong, M.; Van Houten, B. In *Studying protein-DNA interactions using atomic force microscopy*, Seminars in cell & developmental biology, Elsevier: 2017.
13. Pfreundschuh, M.; Martinez-Martin, D.; Mulvihill, E.; Wegmann, S.; Muller, D. J., Multiparametric high-resolution imaging of native proteins by force-distance curve-based AFM. *Nat. Protoc.* **2014**, *9* (5), 1113.
14. Efremov, Y. M.; Wang, W. H.; Hardy, S. D.; Geahlen, R. L.; Raman, A., Measuring nanoscale viscoelastic parameters of cells directly from AFM force-displacement curves. *Sci. Rep.* **2017**, *7* (1), 1541.
15. Dudko, O. K.; Hummer, G.; Szabo, A., Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. **2008**.

16. Beckwitt, E. C.; Simon, N.; Carnaval, I.; Kisker, C.; Carell, T.; Van Houten, B., Peakforce Tapping AFM Reveals that Human XPA Binds to DNA Damage as a Monomer Producing a 60° Bend. **2018**, *114* (3), 93a.
17. Klamecka, K.; Severin, P. M.; Milles, L. F.; Gaub, H. E.; Leonhardt, H., Energy profile of nanobody–GFP complex under force. **2015**, *12* (5), 056009.
18. Pechkova, E.; Sartore, M.; Giacomelli, L.; Nicolini, C., Atomic force microscopy of protein films and crystals. **2007**, *78* (9), 093704.
19. Butt, H.-J.; Downing, K. H.; Hansma, P. K., Imaging the membrane protein bacteriorhodopsin with the atomic force microscope. **1990**, *58* (6), 1473-1480.
20. Henderson, R.; Baldwin, J. M.; Ceska, T.; Zemlin, F.; Beckmann, E. a.; Downing, K., Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. **1990**, *213* (4), 899-929.
21. Hoh, J. H.; Lal, R.; John, S. A.; Revel, J.-P.; Arnsdorf, M. F., Atomic force microscopy and dissection of gap junctions. *Science* **1991**, *253* (5026), 1405-1408.
22. Hartje, L. F.; Munsky, B.; Ni, T. W.; Ackerson, C. J.; Snow, C. D., Adsorption-Coupled Diffusion of Gold Nanoclusters within a Large-Pore Protein Crystal Scaffold. *J. Phys. Chem. B* **2017**, *121* (32), 7652-7659.
23. Huber, T. R.; Hartje, L. F.; McPherson, E. C.; Kowalski, A. E.; Snow, C. D., Programmed Assembly of Host–Guest Protein Crystals. *Small* **2017**, *13* (7), 1602703.
24. Kowalski, A. E.; Huber, T. R.; Ni, T. W.; Hartje, L. F.; Appel, K. L.; Yost, J. W.; Ackerson, C. J.; Snow, C. D., Gold nanoparticle capture within protein crystal scaffolds. *Nanoscale* **2016**, *8* (25), 12693-12696.
25. Goldsbury, C. S.; Scheuring, S.; Kreplak, L. J. C. p. i. p. s., Introduction to atomic force microscopy (AFM) in biology. **2009**, *58* (1), 17.7. 1-17.7. 19.
26. Lee, H.; Scherer, N. F.; Messersmith, P. B., Single-molecule mechanics of mussel adhesion. *Proc. Natl. Acad. Sci.* **2006**, *103* (35), 12999-3003.
27. Monitor, P. T., *Introduction to Bruker's ScanAsyst and PeakForce Tapping AFM Technology*. Bruker application note. Bruker Nano Inc.: Bruker application note. Bruker Nano Inc.
28. Charras, G.; Lehenkari, P. P.; Horton, M., Atomic force microscopy can be used to mechanically stimulate osteoblasts and evaluate cellular strain distributions. **2001**, *86* (1-2), 85-95.
29. Kong, M.; Beckwitt, E. C.; Springall, L.; Kad, N. M.; Van Houten, B., Single-Molecule Methods for Nucleotide Excision Repair: Building a System to Watch Repair in Real Time. In *Methods in enzymology*, Elsevier: 2017; Vol. 592, pp 213-257.
30. Liang, C.; Ye, Z.; Xue, B.; Zeng, L.; Wu, W.; Zhong, C.; Cao, Y.; Hu, B.; Messersmith, P. B., Self-Assembled Nanofibers for Strong Underwater Adhesion: The Trick of Barnacles. *ACS Appl. Mater. Interfaces* **2018**, *10* (30), 25017-25025.
31. Hansma, H. G.; Hoh, J. H., Biomolecular imaging with the atomic force microscope. **1994**, *23* (1), 115-140.
32. Yamamoto, D.; Nagura, N.; Omote, S.; Taniguchi, M.; Ando, T., Streptavidin 2D crystal substrates for visualizing biomolecular processes by atomic force microscopy. *Biophys. J.* **2009**, *97* (8), 2358-67.
33. Chen, L.; Haushalter, K. A.; Lieber, C. M.; Verdine, G. L. J. C.; biology, Direct visualization of a DNA glycosylase searching for damage. **2002**, *9* (3), 345-350.

34. Buechner, C. N.; Maiti, A.; Drohat, A. C.; Tessmer, I. J. N. a. r., Lesion search and recognition by thymine DNA glycosylase revealed by single molecule imaging. **2015**, *43* (5), 2716-2729.
35. Yamashita, H.; Voitchovsky, K.; Uchihashi, T.; Contera, S. A.; Ryan, J. F.; Ando, T., Dynamics of bacteriorhodopsin 2D crystal observed by high-speed atomic force microscopy. *J. Struct. Biol.* **2009**, *167* (2), 153-8.
36. Ando, T.; Uchihashi, T.; Kodera, N.; Yamamoto, D.; Miyagi, A.; Taniguchi, M.; Yamashita, H., High-speed AFM and nano-visualization of biomolecular processes. **2008**, *456* (1), 211-225.
37. Wang, H.; Tessmer, I.; Croteau, D. L.; Erie, D. A.; Van Houten, B., Functional characterization and atomic force microscopy of a DNA repair protein conjugated to a quantum dot. *Nano Letters* **2008**, *8* (6), 1631-1637.
38. Yeh, J. I.; Levine, A. S.; Du, S.; Chinte, U.; Ghodke, H.; Wang, H.; Shi, H.; Hsieh, C. L.; Conway, J. F.; Van Houten, B., Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair. **2012**, *109* (41), E2737-E2746.
39. Ludwig, M.; Rief, M.; Schmidt, L.; Li, H.; Oesterhelt, F.; Gautel, M.; Gaub, H., AFM, a tool for single-molecule experiments. *Applied Physics A* **1999**, *68* (2), 173-176.
40. Janovjak, H.; Struckmeier, J.; Müller, D. J., Hydrodynamic effects in fast AFM single-molecule force measurements. *European Biophysics Journal* **2005**, *34* (1), 91-96.
41. Guthold, M.; Bezanilla, M.; Erie, D. A.; Jenkins, B.; Hansma, H. G.; Bustamante, C., Following the assembly of RNA polymerase-DNA complexes in aqueous solutions with the scanning force microscope. *Proceedings of the National Academy of Sciences* **1994**, *91* (26), 12927-12931.
42. Ebenstein, Y.; Gassman, N.; Kim, S.; Weiss, S., Combining atomic force and fluorescence microscopy for analysis of quantum-dot labeled protein-DNA complexes. *Journal of Molecular Recognition: An Interdisciplinary Journal* **2009**, *22* (5), 397-402.
43. Sanchez, H.; Kanaar, R.; Wyman, C., Molecular recognition of DNA-protein complexes: A straightforward method combining scanning force and fluorescence microscopy. *Ultramicroscopy* **2010**, *110* (7), 844-851.
44. Fronczek, D.; Quammen, C.; Wang, H.; Kisker, C.; Superfine, R.; Taylor, R.; Erie, D.; Tessmer, I., High accuracy FIONA-AFM hybrid imaging. *Ultramicroscopy* **2011**, *111* (5), 350-355.
45. Sanchez, H.; Kertokalio, A.; van Rossum-Fikkert, S.; Kanaar, R.; Wyman, C., Combined optical and topographic imaging reveals different arrangements of human RAD54 with presynaptic and postsynaptic RAD51-DNA filaments. *Proceedings of the National Academy of Sciences* **2013**, *110* (28), 11385-11390.
46. Sánchez, H.; Paul, M. W.; Grosbart, M.; van Rossum-Fikkert, S. E.; Lebbink, J. H.; Kanaar, R.; Houtsmuller, A. B.; Wyman, C., Architectural plasticity of human BRCA2-RAD51 complexes in DNA break repair. *Nucleic acids research* **2017**, *45* (8), 4507-4518.
47. Wu, D.; Kaur, P.; Li, Z. M.; Bradford, K. C.; Wang, H.; Erie, D. A., Visualizing the path of DNA through proteins using DREEM imaging. *Molecular cell* **2016**, *61* (2), 315-323.
48. Adkins, N. L.; Swygert, S. G.; Kaur, P.; Niu, H.; Grigoryev, S. A.; Sung, P.; Wang, H.; Peterson, C. L., Nucleosome-like, single-stranded DNA (ssDNA)-histone octamer complexes and the implication for DNA double strand break repair. *J Biol Chem* **2017**, *292* (13), 5271-5281.
49. Hedayati, M.; Kipper, M. J. J. C. p., Atomic force microscopy of adsorbed proteoglycan mimetic nanoparticles: Toward new glycocalyx-mimetic model surfaces. **2018**, *190*, 346-355.

50. Hedayati, M.; Reynolds, M. M.; Krapf, D.; Kipper, M. J. J. A. a. m.; interfaces, Nanostructured Surfaces That Mimic the Vascular Endothelial Glycocalyx Reduce Blood Protein Adsorption and Prevent Fibrin Network Formation. **2018**, *10* (38), 31892-31902.
51. Martins, J. G.; Camargo, S. E.; Bishop, T. T.; Popat, K. C.; Kipper, M. J.; Martins, A. F., Pectin-chitosan membrane scaffold imparts controlled stem cell adhesion and proliferation. **2018**.
52. Lee, K.; Park, E.; Lee, H. A.; Sugnaux, C.; Shin, M.; Jeong, C. J.; Lee, J.; Messersmith, P. B.; Park, S. Y.; Lee, H., Phenolic condensation and facilitation of fluorescent carbon dot formation: a mechanism study. *Nanoscale* **2017**, *9* (43), 16596-16601.
53. Lin, C.; Romero, R.; Sorokina, L. V.; Ballinger, K. R.; Place, L. W.; Kipper, M. J.; Khetani, S. R., A polyelectrolyte multilayer platform for investigating growth factor delivery modes in human liver cultures. **2018**, *106* (4), 971-984.
54. Vidal, C. M.; Zhu, W.; Manohar, S.; Aydin, B.; Keiderling, T. A.; Messersmith, P. B.; Bedran-Russo, A. K., Collagen-collagen interactions mediated by plant-derived proanthocyanidins: a spectroscopic and atomic force microscopy study. *Acta Biomater.* **2016**, *41*, 110-118.
55. Zhao, H.; Wei, Y.; Qiao, R.; Zhu, C.; Zheng, Z.; Ling, M.; Jia, Z.; Bai, Y.; Fu, Y.; Lei, J., Conductive polymer binder for high-tap-density nanosilicon material for lithium-ion battery negative electrode application. *Nano Letters* **2015**, *15* (12), 7927-7932.
56. McPherson, A.; Kuznetsov, Y. G.; Malkin, A.; Plomp, M., Macromolecular crystal growth investigations using atomic force microscopy. **2004**, *11* (1), 21-23.
57. Yau, S. T.; Thomas, B. R.; Galkin, O.; Gliko, O.; Vekilov, P. G., Molecular mechanisms of microheterogeneity-induced defect formation in ferritin crystallization. **2001**, *43* (4), 343-352.
58. Fujihira, M.; Tani, Y.; Furugori, M.; Akiba, U.; Okabe, Y., Chemical force microscopy of self-assembled monolayers on sputtered gold films patterned by phase separation. **2001**, *86* (1-2), 63-73.
59. Meyvis, T. K.; De Smedt, S. C.; Van Oostveldt, P.; Demeester, J., Fluorescence recovery after photobleaching: a versatile tool for mobility and interaction measurements in pharmaceutical research. *Pharmaceutical research* **1999**, *16* (8), 1153-1162.
60. Kovalski, J. M.; Wirth, M. J., Peer reviewed: applications of fluorescence recovery after photobleaching. *Analytical Chemistry* **1997**, *69* (19), 600A-605A.
61. Rayan, G.; Guet, J.-E.; Taulier, N.; Pincet, F.; Urbach, W., Recent applications of fluorescence recovery after photobleaching (FRAP) to membrane bio-macromolecules. *Sensors* **2010**, *10* (6), 5927-5948.
62. Geverts, B.; van Royen, M. E.; Houtsmuller, A. B., Analysis of biomolecular dynamics by FRAP and computer simulation. In *Advanced Fluorescence Microscopy*, Springer: 2015; pp 109-133.
63. Sarkar, P.; Chattopadhyay, A., Exploring membrane lipid and protein diffusion by FRAP. In *Analysis of Membrane Lipids*, Springer: 2020; pp 119-141.
64. Traylor, S. J.; Bowes, B. D.; Ammirati, A. P.; Timmick, S. M.; Lenhoff, A. M., Fluorescence recovery after photobleaching investigation of protein transport and exchange in chromatographic media. *Journal of Chromatography A* **2014**, *1340*, 33-49.
65. Axelrod, D.; Koppel, D.; Schlessinger, J.; Elson, E.; Webb, W. W., Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophysical journal* **1976**, *16* (9), 1055-1069.

66. Reits, E. A.; Neefjes, J. J., From fixed to FRAP: measuring protein mobility and activity in living cells. *Nature cell biology* **2001**, *3* (6), E145.
67. Burghardt, T.; Axelrod, D., Total internal reflection/fluorescence photobleaching recovery study of serum albumin adsorption dynamics. *Biophysical journal* **1981**, *33* (3), 455-467.
68. Yuan, Y.; Velev, O. D.; Lenhoff, A. M., Mobility of adsorbed proteins studied by fluorescence recovery after photobleaching. *Langmuir* **2003**, *19* (9), 3705-3711.
69. Hobson, C. M.; O'Brien III, E. T.; Falvo, M. R.; Superfine, R., Combined Selective Plane Illumination Microscopy and FRAP maps intranuclear diffusion of NLS-GFP. *Biophysical Journal* **2020**, *119* (3), 514-524.
70. Kärger, J.; Ruthven, D. M.; Theodorou, D. N., *Diffusion in Nanoporous Materials*, 2 Volume Set. John Wiley & Sons: 2012.
71. Kaerger, J.; Ruthven, D. M., Diffusion in nanoporous materials: fundamental principles, insights and challenges. *New Journal of Chemistry* **2016**, *40* (5), 4027-4048.
72. Beerdsen, E.; Dubbeldam, D.; Smit, B., Understanding diffusion in nanoporous materials. *Physical review letters* **2006**, *96* (4), 044501.
73. Bowes, B. D.; Traylor, S. J.; Timmick, S. M.; Czymmek, K. J.; Lenhoff, A. M., Insights into protein sorption and desorption on dextran-modified ion-exchange media. *Chemical engineering & technology* **2012**, *35* (1), 91-101.
74. Weigel, A. V.; Ragi, S.; Reid, M. L.; Chong, E. K.; Tamkun, M. M.; Krapf, D., Obstructed diffusion propagator analysis for single-particle tracking. *Physical Review E* **2012**, *85* (4), 041924.
75. Weigel, A. V.; Simon, B.; Tamkun, M. M.; Krapf, D., Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. *Proceedings of the National Academy of Sciences* **2011**, *108* (16), 6438-6443.
76. Weigel, A. V.; Tamkun, M. M.; Krapf, D., Quantifying the dynamic interactions between a clathrin-coated pit and cargo molecules. *Proceedings of the National Academy of Sciences* **2013**, *110* (48), E4591-E4600.
77. Laskowski, P. R.; Pfreundschuh, M.; Stauffer, M.; Ucurum, Z. h.; Fotiadis, D.; Müller, D., High-Resolution Imaging and Multiparametric Characterization of Native Membranes by Combining Confocal Microscopy and an Atomic Force Microscopy-Based Toolbox. *ACS nano* **2017**, *11* (8), 8292-8301.
78. Nwaneshiudu, A.; Kuschal, C.; Sakamoto, F. H.; Anderson, R. R.; Schwarzenberger, K.; Young, R. C., Introduction to confocal microscopy. *Journal of Investigative Dermatology* **2012**, *132* (12), 1-5.
79. Patel, D. V.; McGhee, C. N., Quantitative analysis of in vivo confocal microscopy images: a review. *Survey of ophthalmology* **2013**, *58* (5), 466-475.
80. de Hoog, C. L.; Foster, L. J.; Mann, M., RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers. *Cell* **2004**, *117* (5), 649-662.
81. Ghibaudo, M.; Di Meglio, J.-M.; Hersen, P.; Ladoux, B., Mechanics of cell spreading within 3D-micropatterned environments. *Lab on a Chip* **2011**, *11* (5), 805-812.
82. Elliott, A. D., Confocal microscopy: principles and modern practices. *Current protocols in cytometry* **2020**, *92* (1), e68.
83. Wei, S.; Shi, F.; Wang, Y.; Chou, Y.; Li, X., A deep learning model for automated sub-basal corneal nerve segmentation and evaluation using in vivo confocal microscopy. *Translational Vision Science & Technology* **2020**, *9* (2), 32-32.

84. Schueder, F.; Lara-Gutiérrez, J.; Beliveau, B. J.; Saka, S. K.; Sasaki, H. M.; Woehrstein, J. B.; Strauss, M. T.; Grabmayr, H.; Yin, P.; Jungmann, R., Multiplexed 3D super-resolution imaging of whole cells using spinning disk confocal microscopy and DNA-PAINT. *Nature communications* **2017**, *8* (1), 1-9.
85. Soni, G.; Hameed, F. M.; Roopa, T.; Shivashankar, G., Development of an optical tweezer combined with micromanipulation for DNA and protein nanobioscience. *Current Science* **2002**, 1464-1470.
86. Yablon, D.; Chakraborty, I., Machine learning to enhance atomic force microscopy analysis and operation. *Machine learning* **2020**.
87. Müller, P.; Abuhattum, S.; Möllmert, S.; Ulbricht, E.; Taubenberger, A. V.; Guck, J., nanite: using machine learning to assess the quality of atomic force microscopy-enabled nano-indentation data. *BMC bioinformatics* **2019**, *20* (1), 1-9.
88. Forchheimer, D.; Forchheimer, R.; Haviland, D. B., Improving image contrast and material discrimination with nonlinear response in bimodal atomic force microscopy. *Nature communications* **2015**, *6* (1), 1-5.
89. Qi, Y.; Klein-Seetharaman, J.; Bar-Joseph, Z., Random forest similarity for protein-protein interaction prediction from multiple sources. In *Biocomputing 2005*, World Scientific: 2005; pp 531-542.
90. Li, B.-Q.; Feng, K.-Y.; Chen, L.; Huang, T.; Cai, Y.-D., Prediction of protein-protein interaction sites by random forest algorithm with mRMR and IFS. *PLoS One* **2012**, *7*(8), e43927.
91. Wang, Y.; Gao, L., Detecting Protein Complexes by an Improved Affinity Propagation Algorithm in Protein-Protein Interaction Networks. *J. Comput.* **2012**, *7* (7), 1761-1768.
92. Vlasblom, J.; Wodak, S. J., Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. *BMC bioinformatics* **2009**, *10* (1), 1-14.
93. Tuckerman, M. E., Machine learning transforms how microstates are sampled. *Science* **2019**, *365* (6457), 982-983.
94. Cendagorta, J. R.; Tolpin, J.; Schneider, E.; Topper, R. Q.; Tuckerman, M. E., Comparison of the performance of machine learning models in representing high-dimensional free energy surfaces and generating observables. *The Journal of Physical Chemistry B* **2020**, *124* (18), 3647-3660.
95. Brohee, S.; Van Helden, J., Evaluation of clustering algorithms for protein-protein interaction networks. *BMC bioinformatics* **2006**, *7* (1), 1-19.
96. Chen, D.; Sarkar, S.; Candia, J.; Florczyk, S. J.; Bodhak, S.; Driscoll, M. K.; Simon Jr, C. G.; Dunkers, J. P.; Losert, W., Machine learning based methodology to identify cell shape phenotypes associated with microenvironmental cues. *Biomaterials* **2016**, *104*, 104-118.
97. Unadkat, H. V.; Hulsman, M.; Cornelissen, K.; Papenburg, B. J.; Truckenmüller, R. K.; Carpenter, A. E.; Wessling, M.; Post, G. F.; Uetz, M.; Reinders, M. J., An algorithm-based topographical biomaterials library to instruct cell fate. *Proceedings of the National Academy of Sciences* **2011**, *108* (40), 16565-16570.

## CHAPTER 2: HIGH RESOLUTION IMAGING OF BIOACTIVATED NANOPOROUS PROTEIN CRYSTALS BY ATOMIC FORCE MICROSCOPY \*

### Chapter 2 Overview

Probing and observing the surface of highly porous *Camphylobacter Jejuni* formed protein crystals by high-resolution atomic force microscope (AFM) is a technical challenge. The structure as well as the chemical and biological activation of nanoporous protein crystals is likely to change significantly if removed from aqueous solutions: crosslinked protein crystal which are 80% solvent can shrink dramatically when dried. The imaging process must be done in the liquid phase. This places great demands on the accuracy of imaging, as well as the stability of the sample. In this study, for the first time, the porous surface structure of this unusual type of protein crystal was observed via high resolution AFM in liquid. Successful observation can be attributed to the improvements in immobilization. The porous protein crystals have been systematically imaged by AFM on different substrates, namely ultraviolet glue (UV-glue) and mica. This approach has improved methods for crystal characterization by AFM. It enables the morphology of protein crystals to be studied in detail, in buffer liquid, thereby preserving the native structure and properties of protein crystals, which would typically be destroyed by drying. Experimental results for different AFM scan sizes demonstrate that a hexagonal space group of nanoporous protein crystals is manifested at the cell surface. This is consistent with the expected structure from bulking single crystal X-ray diffraction; however, AFM can probe aspects of crystals that are invisible to XRD. Optimized AFM characterization protocols provide the possibility to image and quantitatively analyze the structure, uniformity, mechanical behaviors, and interactions on the surface of protein crystals.

---

\* Portions of this chapter are reproduced from: Dafu Wang, Julius D. Stuart, Ann E. Kowalski, Abigail R. Ward, Chris D. Snow, Matt J. Kipper. "Immobilizing and Imaging Nanoscale Bioactivated Protein Crystals in a Liquid Environment"

## 2.1 Introduction

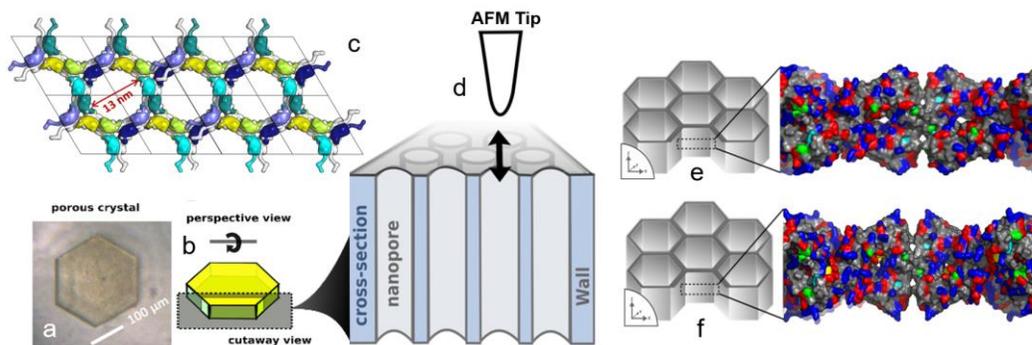
Atomic force microscopy (AFM) is a powerful tool for the study of materials' surfaces, providing molecular-scale resolution of surface features.<sup>[1-5]</sup> For biological applications and for studies of fragile biomolecules, AFM can be conducted in buffered solutions, allowing samples to remain in their natural condition.<sup>[6-8]</sup> Moreover, AFM does not require samples to be labeled or stained, which would better preserve their original chemical and biological properties.<sup>[2]</sup> AFM can provide nanometer-resolution imaging.<sup>[6, 7]</sup> Published AFM studies of biological samples have successfully provided images of soluble proteins,<sup>[9, 10]</sup> lipid mono- and bilayers,<sup>[11]</sup> and two-dimensional film and membrane of proteins.<sup>[9, 12, 13]</sup> This technique also provides images of surfaces that can be used for structural studies of biological macromolecules.<sup>[9]</sup> However, traditional AFM measurement methods for two-dimensional protein crystals are not suitable for the new class of nanomaterials based on highly crosslinked nanoporous (Fig. 2.1 (a-d)) crystals of a putative isoprenoid binding protein from *Campylobacter jejuni* (Genebank ID: CJ0420, Protein Data Bank (PDB) code: 5w17), which we refer to as CJ.<sup>[14-17]</sup> We also use a CJ protein variant, which is suitable for growing large protein crystals. We identify the modified CJ as name would be CJ-A34I-L48F-V50I-V121M-N162C-I163W-V165I. We would like to propose a new name for our large crystal variant as CJ<sup>OPT</sup>, the 'OPT' represents the optimized, large crystal-growing variant. These crystals are all highly ordered three-dimensional arrays with a surface pore size distribution of 13 nm diameter (18 nm pore-center-to-center). An important requirement in this case is that the proposed AFM imaging technique shall be able to overcome sample drift, which negatively impacts imaging. In this studies, we have found the methods to stably characterize the nanoporous crystals' surface morphology in buffer solution, and achieved the stable nanoscale AFM imaging on protein crystals' surface while maintaining their original chemical and biological activation.<sup>[18]</sup>

The objective of this paper revolves the detailed methodology of using AFM to achieve stable high resolution nanoscale characterization on nanoporous protein crystals' surface morphology, in buffer solution, under complex conditions.

In this approach, the AFM was a Bruker Bioscope Resolve, mounted on an optical microscope stage. The three-dimensional porous protein crystals have been systematically imaged by AFM on both liquid ultraviolet glue (UV-glue) and mica. A series of independent experiments have been conducted to observe the morphology while capturing clear images of the surface of protein crystals by high-resolution AFM. The core of the AFM is a force sensing device that includes a cantilever with a tip secured to one end. In this work, ScanAsyst Fluid+ cantilevers from Bruker were used. In this study, medium-large protein crystals were used with a typical diameter of 400  $\mu\text{m}$  and height of 50  $\mu\text{m}$ . The crystal habit was a hexagonal prism simplifying the alignment of the major pores towards the AFM probe. To characterize the features of the protein crystal, tips with a radius of  $\sim 12$  nm were used. Small tip diameter enabled the high-resolution observation of the protein crystal surface features, and may also be used for proposed chemical force spectroscopy experiments of biomolecules in liquid.

We hypothesized that the Bruker Bioscope Resolve AFM could reveal high-resolution details of porous crystal surfaces, including imaging the surface morphology of three-dimensional nanoporous protein crystals in the buffer liquid.<sup>[19-21]</sup> The first step was to find a practical way to obtain clear AFM images of the crystal surface, which is still a technical challenge to be overcome. Unlike electron microscopy, AFM can be conducted in solution, affording the opportunity for *in situ* imaging and molecular force probing in biological applications.<sup>[22-24]</sup> Using AFM to characterize the protein crystals enables the periodicity and morphology of crystals to be studied in their mother liquid, thereby preserving the as-grown periodic protein crystal structure, which

can be disrupted or destroyed by drying.<sup>[22]</sup> In this study, we tested several approaches for that purpose. We found that liquid ultraviolet glue (UV-glue) and mica were most suitable for immobilizing the crystals for AFM imaging in the liquid phase. So far, multiple studies and approaches related to AFM imaging have been published.<sup>[4, 9, 20, 25-31]</sup> By learning from them, here we describe a series of experiments under different imaging conditions designed to identify the optimal experimental conditions of the preparation and measurement of the protein crystals. The objective of this study is to identify the conditions under which high-resolution details of porous crystal surfaces can be reliably obtained. In this study, we also characterize surface defects, as well as different orientations of protein crystals. This work advances our ability to characterize porous protein crystals, enabling us to explore practical applications of protein crystals for DNA sensing and storage, where porous crosslinked protein crystals store functional guest macromolecules.



**Fig. 2.1** A periplasmic protein, “CJ”, from *Camphylobacter jejuni* forms (a) porous protein crystals that we stabilize via crosslinking. (b) Typical crystals are hexagonal prisms. (c) A top view of three adjacent nanopores (PDB code 5w17) (d) A hexagonal array of 13 nm diameter nanopores runs from the top to bottom of each crystal. The top face of the crystal may be probed using AFM tips.<sup>[14-17]</sup> (e) CJ and (f) CJ<sup>OPT</sup> crystal schematic with nanopores cut away, and zoomed in slice of nanopore side wall illustrating presence of ionizable amino acid. Carboxylic acids (Asp, Glu) are shown in red. Arginines are shown in cyan. Lysines are shown in dark blue. Notably, some of the Lysine sidechains have likely lost their positive charge by participating in glyoxal crosslinks. Glyoxal crosslinked crystals tend to diffract to modest resolution  $\sim 3.5\text{\AA}$  and surface lysine sidechains tend to be highly mobile. These two factors, as well as heterogeneity within the crystal, prevent us from assessing which lysines are likely to retain their positive charge. Finally, the figure shows histidine sidechains in green. Both the N- and C-terminus contain flexible regions that are not pictured in this crystal structure. The flexible C-terminal histag (not pictured) is of particular interest as a possible participant in DNA binding. Image was created by PyMOL<sup>[18]</sup>

## 2.2 Experimental

### 2.2.1 Crystal's Growth

A Millipore Synthesis water purification unit was used to obtain 18.2 MΩ cm water, used for making all aqueous solutions. First, one 24 μL tube of purified CJ protein was thawed.<sup>[17, 32-34]</sup> The samples were kept on ice at all times. In the reservoir of a plastic CRYSCHEM sitting-drop crystallization plate, 340 μL of 4 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 40 μL of 1 M bis-tris (pH = 6.5), and 20 μL of DI H<sub>2</sub>O were mixed. Then, 1 μL of the reservoir solution was pipetted into the top drop of the plastic sitting-drop crystallization plate. Finally, a 1 μL aliquot of CJ protein solution was added by pipetting the protein solution directly on top of the drop of reservoir solution. Crystals typically grew to full size within 1 to 3 days.<sup>[18]</sup>

As described in previous work for the CJ protein,<sup>[35]</sup> the CJ<sup>OPT</sup> protein was cloned into pSB3 expression vector. CJ expression was performed with BL21(DE3) *Escherichia coli* cells in Terrific broth. This needs to be modified to describe CJ<sup>OPT</sup> Induction was performed with 0.4 mM IPTG for 16 hours at 25 °C, followed by purification using immobilized metal affinity chromatography (IMAC). The target protein was then dialyzed into ammonium sulfate storage buffer (500 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 10% glycerol, pH = 7.4). CJ<sup>OPT</sup> protein was concentrated to 15 mg/mL, aliquoted and stored at -30 °C. Then, one tube (24 μL) of purified CJ<sup>OPT</sup> protein (15 mg/mL) was thawed.<sup>[17, 32-34]</sup> The samples were kept on ice at all times. In the reservoir of a plastic CrysChem sitting-drop crystallization plate, 340 μL of 4 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 40 μL of 1 M bis-tris (pH = 6.5), and 20 μL of DI H<sub>2</sub>O were mixed. Then, 1 μL of the reservoir solution was pipetted into the top drop of the plastic sitting-drop crystallization plate. Finally, a 1 μL aliquot of CJ<sup>OPT</sup> protein solution was added by

pipetting the protein solution directly on top of the drop of reservoir solution. Crystals typically grew to full size within 1 to 3 days.<sup>[18]</sup>

### **2.2.2 Crystal Crosslinking and Immobilization**

Crystals were transferred (using a nylon crystallography loop, Hampton Research) from their growth well into a drop of 4.2 M trimethylamine *N*-oxide (TMAO), 0.175 M H<sub>2</sub>SO<sub>4</sub> at pH = 7.5 to wash for a minimum of 20 min. A drop of 390 μL of 4.2 M TMAO, 0.175 M H<sub>2</sub>SO<sub>4</sub> at pH 7.5 plus 10 μL of 40% glyoxal crosslinker was then prepared. Crystals were manually transferred into the crosslinking solution and covered for 2 h. Meanwhile, a drop of 380 μL 0.1 M sodium citrate, 0.15 M NaCl at pH 5.0, 10 μL of 50% hydroxylamine, and 10 μL of 40 mg/mL dimethylamine borane complex was prepared. The crosslinked crystals were transferred to this “quenching” solution to eliminate reactive groups. After 8 hours in the quenching solution, crystals were ready to be used or stored. Crystals were stored in 4.0 M TMAO and washed briefly in water or adsorption buffer (30 mM KCL, 10 mM MES, pH = 6.0) prior to use.<sup>[14-18, 36]</sup>

The following three methods were applied for the immobilization of crystals. Mica sheets (Ted Pella, Inc.) were glued and cured overnight to glass bottom petri dishes (Ted Pella, Inc.) using super glue glass adhesive. Before crystal attachment, packaging tape was used 5-10 times to cleave a fresh sheet of mica. 100 μL of adsorption buffer was pipetted onto the mica. Crystals were looped into the solution and pressed gently to the bottom with a nylon crystallography loop (Hampton Research). Crystals were immobilized on mica by the following protocol.<sup>[14]</sup> First, the mica was cleaved 5 to 10 times.<sup>[37]</sup> The pure mica surface in H<sub>2</sub>O was imaged by AFM to ensure there was no external contamination of the surfaces that might impact the pH value.<sup>[38]</sup> Then the surfaces were cleaved two more times and soaked in 15% (3-aminopropyl) triethoxysilane (APTS) for 1 h (100 μL 100% APTS + 900 μL H<sub>2</sub>O). The surfaces were rinsed before adding the crystals. Crystals

were rinsed and settled in high-pH buffer (10 mM borate, pH = 9.5). These crystals were allowed to bind on the APTS-modified mica for 20 min. The bound crystals were gently washed with DI water three times. Finally, 10 mM sodium borate (pH = 9.5) was added to the petri dish.

The second crystal immobilization strategy uses a UV-curable glue to immobilize the protein crystals. Under this method, crystals were immobilized on glass-bottom petri dishes (Willco Wells) employing a UV-curable glue. The top of a crystal probe (Minitool HR4-217) was used to bring a drop of UV-curable glue (Bondic Inc.) on the surface of a petri dish (Ted Pella, Inc.). The glue was gently and evenly spread on the dish surface to make the layer of glue as thin as possible. Crystals were transferred to the glue with a loop. Critically, the crystal was moved inside a tiny drop of buffer, such that the crystal was not desiccated. The UV-glue was more viscous than the buffer while the glue did not mix with the buffer. The glue was then cured by exposing to the UV-light from above for 10 s. The glue completely cures after about 2 min, after which additional drops of buffer (typically ~5 mL) were added to the dish, which prevented the crystal from drying.

The third crystal immobilization strategy involved attaching crystals to a poly-lysine surface. The glass bottom of a petri dish (Willco Wells) was first washed with 100% ethanol, then rinsed 3 times with H<sub>2</sub>O. After the surface was air dried, a solution of 3 mL 0.1% wt./vol. of poly-lysine in H<sub>2</sub>O was added to the petri dish, soaked for 1 hour to keep it covered. The bottom surface of a petri dish was then rinsed 3 times with H<sub>2</sub>O and air dried again. The protein crystal was crosslinked to the surface using 5 mL 1% glutaraldehyde in 3.6 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (pH = 7.5) for 30 mins, and the crystal turned yellow after this process was finished. The crystal was then rinsed with H<sub>2</sub>O 3 times. After 4.2 M TMAO was added as a buffer solution, the crystal was immobilized on the glass bottom of the petri dish.

### **2.2.3 AFM Probing**

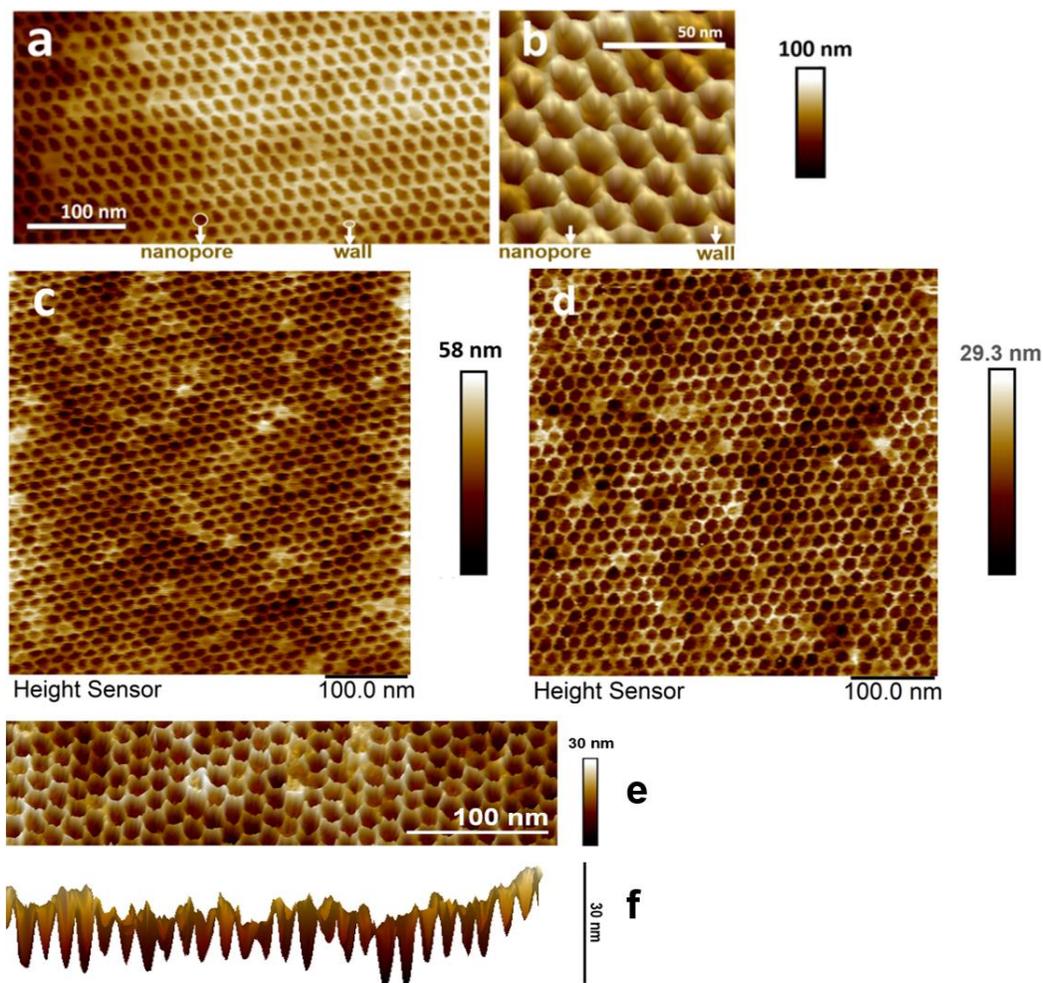
The AFM we used for imaging was a Bruker Bioscope Resolve. All the images were collected via the ScanAsyst mode, using a ScanAsyst Fluid+ probe, mounted on an optical microscope stage. The AFM scan rate was set to 1.0 Hz; the peak force tapping frequency was set to 1.0 kHz, and the peak force set point was set to below 2000 pN. This type of probe also has a scan angle of 0 or 90°. Crystal imaging was performed in the buffer, with crystals affixed to the bottom of a glass petri dish, using the method described in Section 2.2.2. The imaging of crystals was performed to ensure the optimal resolution without external influence, especially drift and vibration. To decrease the experimental noise, the laser intensity was kept between 3.5 V to 7.0 V. For each sample, the scan covers different areas on each crystal surface, while changing the position of the cantilever between successive scans. Analysis of the AFM data was performed in NanoScope (Bruker, Inc.). Immobilized (UV-curable glue) three-dimensional porous protein crystals were imaged by high-resolution AFM in a 5 mM TE / DI H<sub>2</sub>O buffer, to characterize the crystal surface morphology.

## **2.3 Results and Discussion**

### **2.3.1 AFM Imaging on Surface Morphology**

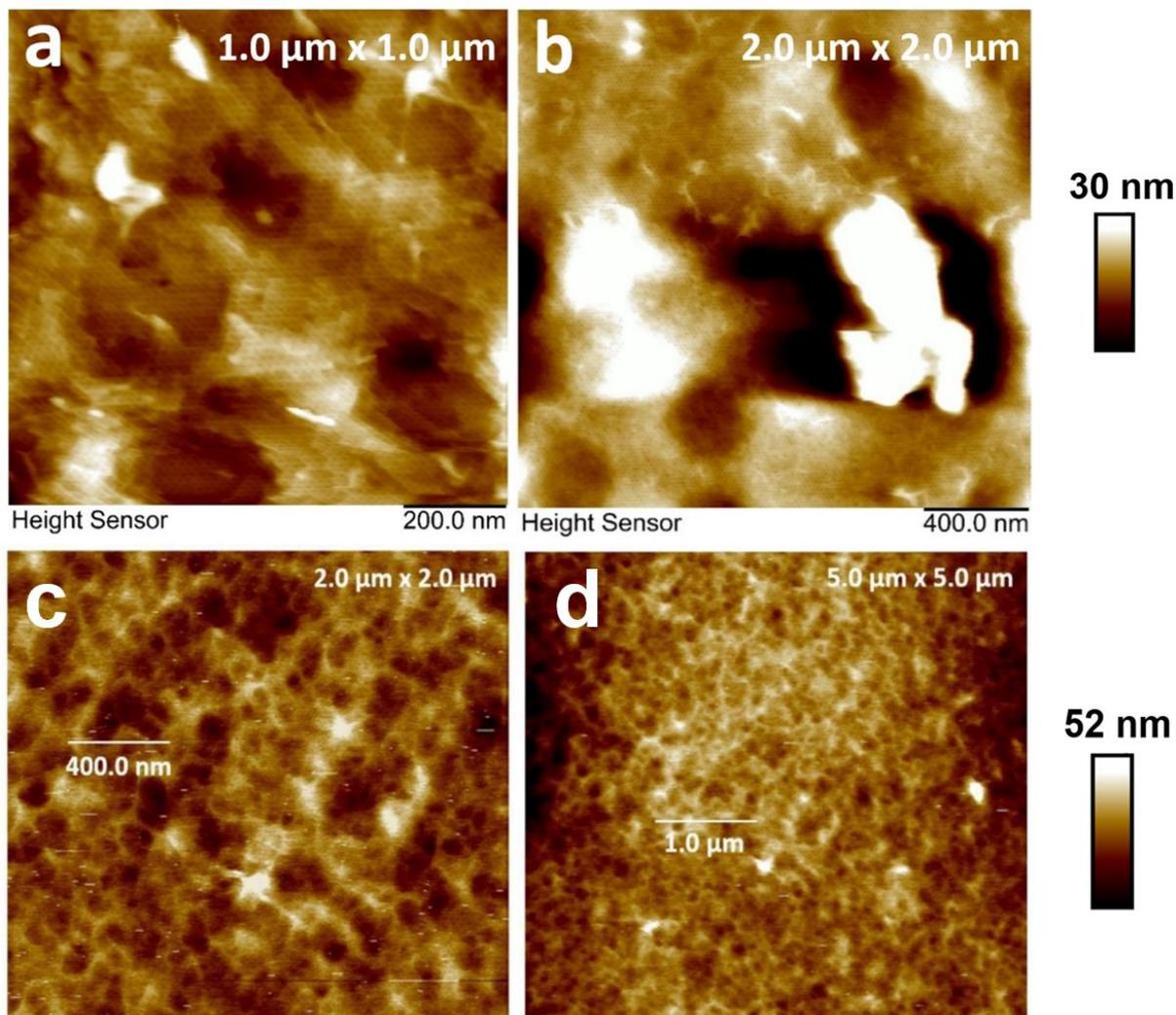
High-resolution details of porous crystal surfaces were revealed using this approach. AFM images show that the porous protein crystal surface, presents a regular honeycomb pore/hole structure consistent with bulk x-ray diffraction data.<sup>[14-17]</sup> The observed surface structure of protein crystals did not dramatically change with imaging methods. Similar surface structures were imaged by AFM on different substrates including UV-glue and mica. The diameter of the nanopores were very consistent under the same scale in different samples. AFM faults due to voltage instability of

laser detection are also observed on the images. In most of the cases the structure looks regular and uniform. Due to the very high sensitivity of AFM imaging under nanometer-level resolution, even extremely slight vibration can create stretches during imaging. It should be pointed out that the blocked pore structure in Fig. 2.2 (a, c-d) are one of the original defective morphologies of nanoporous protein crystals' surface structure. This defect structure will potentially block the guest molecules to be loaded into the nanopores of the protein crystals.<sup>[18]</sup>



**Fig. 2.2** (a) Zoomed AFM image of CJ crystal's upper surface (immobilized on UV-glue), (b) 3D rendering of a portion of the height data from Fig. 2.2 (a), (c) AFM imaging of a 100 nm × 100 nm area on CJ crystal's surface (immobilized on UV-glue), (d) AFM imaging of a 100 nm × 100 nm area on CJ<sup>OPT</sup> crystal's surface (immobilized on UV-glue), (e) 3D rendering of a portion of the height data from Fig. 2.2 (d), and (f) a perpendicular cross-section view of 3D AFM imaging of CJ<sup>OPT</sup> protein crystal surface from Fig. 2.2 (d), across the center of multiple pores in one single line.

### 2.3.1.1 Imaging of Large Areas



**Fig. 2.3** AFM height sensor image of CJ crystal surface on UV-glue substrate, in water, with a scanning size of (a)  $1.0 \mu\text{m} \times 1.0 \mu\text{m}$ , and (b)  $2.0 \mu\text{m} \times 2.0 \mu\text{m}$ . Regular distributed nanopores can be observed in the figures. The high brightness part of the picture is caused by protruding structure of surface defects, and AFM height sensor image of CJ crystal surface immobilized on mica substrate, with a scanning size of (c)  $2.0 \mu\text{m} \times 2.0 \mu\text{m}$ , and (d)  $5.0 \mu\text{m} \times 5.0 \mu\text{m}$

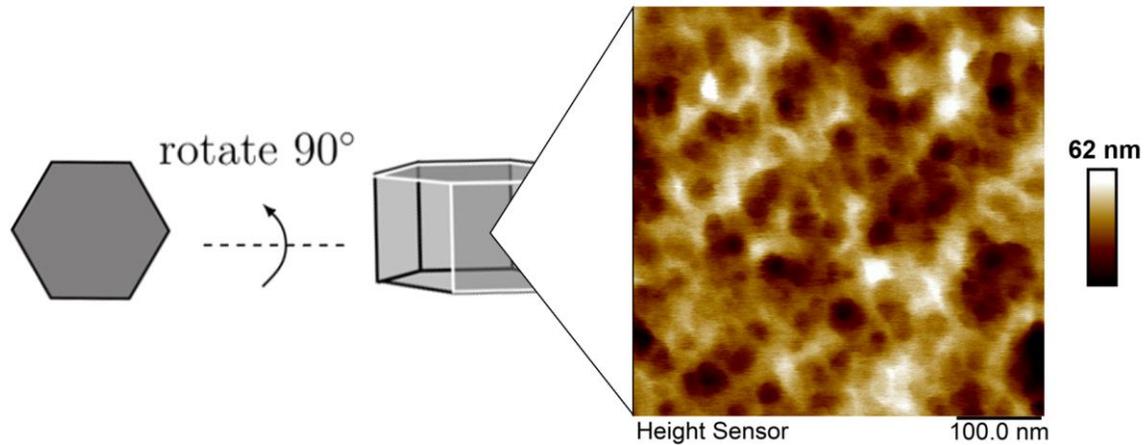
Fig. 2.3 (a-b) are the first batch of AFM images that characterize a large scanning area of CJ crystals surface over  $1.0 \mu\text{m} \times 1.0 \mu\text{m}$  and  $2.0 \mu\text{m} \times 2.0 \mu\text{m}$ , with clear observation of 13-nm diameter nanopores. The imaging process took 30 minutes for each of the individual imaging processes, requiring a higher standard of immobilization. These are the first cases where the pore structure is so cleanly visible despite large variations (multiple unit cells) in the height of the crystal.

From the microscopic image of the enlarged scan size, it can be clearly observed that the pores (with a diameter of 15~30 nm, measured via NanoScope Analysis) are uniformly distributed on the surface of the crystal over a large area. Crystals maintain a smooth morphology with very few surface defects in water. Previous experiments have shown that the surface morphology of the crystal is affected by the pH of the buffer liquid. It can be proven with these AFM images, that under the protocol of this approach, CJ crystal can be prepared with a unified surface structure, and the 15-30 nm-diameter pores are evenly and widely distributed on the surface of CJ crystals.

Fig. 2.3 (c-d) showed the overall morphology of CJ crystals' surface on mica substrate. By scanning a  $5.0\ \mu\text{m} \times 5.0\ \mu\text{m}$  area, the overall surface morphology was observed as irregularly porous. There are several potential reasons for an irregular crystal surface,<sup>[14-17]</sup> one of them can be attributed to the high pH environment (pH = 9.5) in buffer liquid. The strong alkaline pH environment is necessary for adhering crystals to the positively charged mica, however, it can damage the surface structure of protein crystals within hours. Meanwhile, during our experiments, it has also been observed that the morphology of the crystal surface was significantly altered under a strong acidic environment.

### **2.3.1.2 Imaging of Different Orientation**

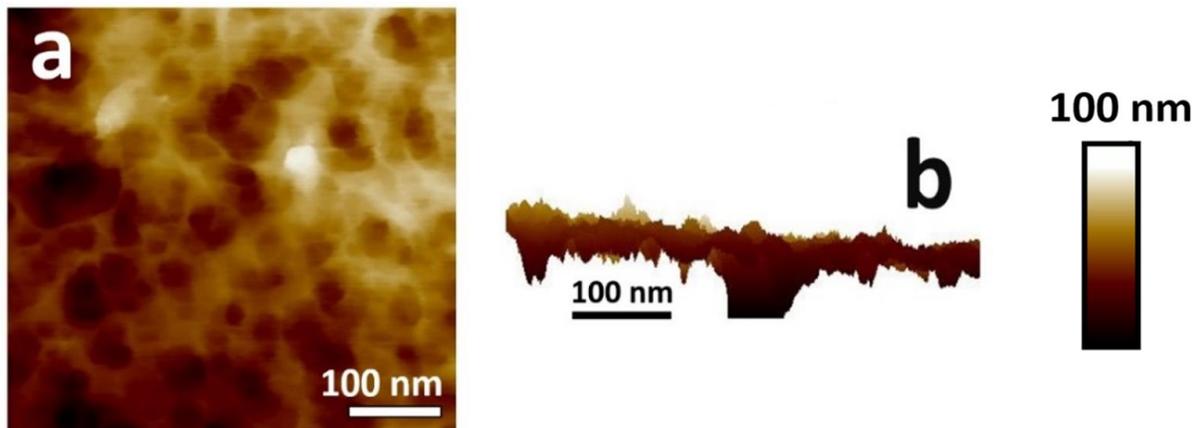
To comprehensively understand the structure and morphology of CJ protein crystals, different orientations were imaged by AFM. The AFM imaging indicated that the side surface of the crystal is also distributed with nanopores (Fig. 2.4). However, the nanopores at this orientation are not uniformly distributed. The diameter of nanopores on this side of the crystal vary between 30 to 40 nanometers, and this structure only exists at this orientation of the crystals.



**Fig. 2.4** Surface morphology of side surface of CJ crystal by AFM

### 2.3.1.3 Imaging of Defected Surface

Surface defects were commonly observed on the surface of crystals. It is not yet clear what differences in the crystal, growth crosslinking, quenching, or subsequent, handling are correlated with irregular surfaces. One common morphology (Fig. 2.5 (a)) is a sponge-like structure. The difference in height across this type of surface was significant, up to  $0.1\ \mu\text{m}$ , with hundred-fold greater value to the regular surface structure. Also, the structures are irregular and non-uniformed. In another type of structural defect (Fig. 2.5 (b)), a cavity is observed on the surface.



**Fig. 2.5** (a) AFM image of CJ crystal surface (immobilized on mica, with mess defects type (a), high sensor) (b) Side view of a 3-dementional AFM image of CJ crystal surface (immobilized on mica)

## **Chapter 2 Summary**

In this study, for the first time, the porous surface structure of an unusual type of protein crystal was observed, and imaged in the liquid phase via high resolution AFM, revealing high-resolution details of porous crystal surfaces. The methods for crystal immobilization for AFM characterization were improved. This paper also discussed the surface imperfections of protein crystals. This approach provides the possibility to image and quantitatively analyze the structure, mechanical behaviors, and interactions on the surface of nanoporous protein crystals.

## REFERENCES

1. Brown, B.P., et al., *Opportunities in high-speed atomic force microscopy*. *Small*, 2013. **9**(19): p. 3201-11.
2. Beckwitt, E.C., M. Kong, and B. Van Houten. *Studying protein-DNA interactions using atomic force microscopy*. in *Seminars in cell & developmental biology*. 2017. Elsevier.
3. Dufrene, Y.F., et al., *Imaging modes of atomic force microscopy for application in molecular and cell biology*. *Nat. Nanotechnol.* , 2017. **12**(4): p. 295-307.
4. Yamamoto, D., et al., *Streptavidin 2D crystal substrates for visualizing biomolecular processes by atomic force microscopy*. *Biophys. J.* , 2009. **97**(8): p. 2358-67.
5. YuG, K., et al., *Molecular resolution imaging of macromolecular crystals by atomic force microscopy*. *Biophys. J.*, 1997. **72**(5): p. 2357-2364.
6. Pfreundschuh, M., et al., *Multiparametric high-resolution imaging of native proteins by force-distance curve-based AFM*. *Nat. Protoc.*, 2014. **9**(5): p. 1113.
7. Lopez, A.E., et al., *Surface dependence of protein nanocrystal formation*. *Small*, 2010. **6**(3): p. 396-403.
8. Caballero, D., et al., *Direct patterning of anti-human serum albumin antibodies on aldehyde-terminated silicon nitride surfaces for HSA protein detection*. *Small*, 2009. **5**(13): p. 1531-4.
9. Pechkova, E., et al., *Atomic force microscopy of protein films and crystals*. 2007. **78**(9): p. 093704.
10. Edstrom, R.D., et al., *Viewing molecules with scanning tunneling microscopy and atomic force microscopy*. 1990. **4**(13): p. 3144-3151.
11. Henderson, R., et al., *Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy*. 1990. **213**(4): p. 899-929.
12. Butt, H.-J., K.H. Downing, and P.K. Hansma, *Imaging the membrane protein bacteriorhodopsin with the atomic force microscope*. 1990. **58**(6): p. 1473-1480.
13. Hoh, J.H., et al., *Atomic force microscopy and dissection of gap junctions*. *Science*, 1991. **253**(5026): p. 1405-1408.
14. Hartje, L.F., et al., *Adsorption-Coupled Diffusion of Gold Nanoclusters within a Large-Pore Protein Crystal Scaffold*. *J. Phys. Chem. B*, 2017. **121**(32): p. 7652-7659.
15. Huber, T.R., et al., *Programmed Assembly of Host-Guest Protein Crystals*. *Small*, 2017. **13**(7): p. 1602703.
16. Huber, T.R., et al., *Installing Guest Molecules at Specific Sites within Scaffold Protein Crystals*. *Bioconjugate Chem.* , 2017. **29**(1): p. 17-22.
17. Kowalski, A.E., et al., *Gold nanoparticle capture within protein crystal scaffolds*. *Nanoscale*, 2016. **8**(25): p. 12693-12696.
18. Wang, D., et al., *Measuring Interactions of DNA with Nanoporous Protein Crystals by Atomic Force Microscopy*. *Nanoscale*, 2021. **13**(24): p. 10871-10881.
19. Huang, Z., et al., *Injectable dynamic covalent hydrogels of boronic acid polymers cross-linked by bioactive plant-derived polyphenols*. 2018. **6**(9): p. 2487-2495.
20. Liang, C., et al., *Self-Assembled Nanofibers for Strong Underwater Adhesion: The Trick of Barnacles*. *ACS Appl. Mater. Interfaces*, 2018. **10**(30): p. 25017-25025.

21. Monitor, P.T., *Introduction to Bruker's ScanAsyst and PeakForce Tapping AFM Technology*. Bruker application note. Bruker Nano Inc.: Bruker application note. Bruker Nano Inc.
22. Durbin, S.D., W.E. Carlson, and M. Saros, *In situ studies of protein crystal growth by atomic force microscopy*. J. Phys. D: Appl. Phys. , 1993. **26**(8B): p. B128.
23. Keel, T., *The application of in situ AFM to the study of molecular and macromolecular crystallization*. 2004, University of Nottingham.
24. Malkin, A., Y.G. Kuznetsov, and A. McPherson, *An in situ AFM investigation of catalase crystallization*. Surf. Sci., 1997. **393**(1-3): p. 95-107.
25. McPherson, A., et al., *Macromolecular crystal growth investigations using atomic force microscopy*. 2004. **11**(1): p. 21-23.
26. Yau, S.T., et al., *Molecular mechanisms of microheterogeneity-induced defect formation in ferritin crystallization*. 2001. **43**(4): p. 343-352.
27. Yamashita, H., et al., *Dynamics of bacteriorhodopsin 2D crystal observed by high-speed atomic force microscopy*. J. Struct. Biol., 2009. **167**(2): p. 153-8.
28. Ando, T., et al., *High-speed AFM and nano-visualization of biomolecular processes*. 2008. **456**(1): p. 211-225.
29. Beckwitt, E.C., et al., *Peakforce Tapping AFM Reveals that Human XPA Binds to DNA Damage as a Monomer Producing a 60° Bend*. 2018. **114**(3): p. 93a.
30. Kong, M., et al., *Single-Molecule Methods for Nucleotide Excision Repair: Building a System to Watch Repair in Real Time*, in *Methods in enzymology*. 2017, Elsevier. p. 213-257.
31. Wang, H., et al., *Functional characterization and atomic force microscopy of a DNA repair protein conjugated to a quantum dot*. Nano Letters, 2008. **8**(6): p. 1631-1637.
32. Ilari, A. and C. Savino, *A practical approach to protein crystallography*, in *Bioinformatics*. 2017, Springer. p. 47-78.
33. Wlodawer, A., et al., *Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination*. FEBS J., 2013. **280**(22): p. 5705-5736.
34. Wlodawer, A., et al., *Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures*. FEBS J., 2008. **275**(1): p. 1-21.
35. McPherson, A., A. Malkin, and Y. G, *Atomic force microscopy in the study of macromolecular crystal growth*. Annu. Rev. Biophys., 2000. **29**(1): p. 361-410.
36. Hartje, L.F., et al., *Characterizing the Cytocompatibility of Various Cross-Linking Chemistries for the Production of Biostable Large-Pore Protein Crystal Materials*. ACS Biomaterials Science & Engineering, 2018. **4**(3): p. 826-831.
37. Amin, D.R., et al., *Size control and fluorescence labeling of polydopamine melanin-mimetic nanoparticles for intracellular imaging*. 2017. **2**(3): p. 17.
38. Hermanson, G., *Chapter 1—Introduction to bioconjugation*. 2013: p. 1-125.

## CHAPTER 3: MEASURING INTERACTION OF DNA WITH PROTEIN CRYSTALS BY ATOMIC FORCE MICROSCOPY \*

### Chapter 3 Overview

Crosslinked porous protein crystals are a new biomaterial that can be engineered to encapsulate, stabilize, and organize guest molecules, nanoparticles, and biological moieties. In this study, for the first time, the interactions of DNA with porous protein crystals are quantitatively measured by high-resolution atomic force microscopy (AFM) and chemical force microscopy. The surface structure of protein crystals with unusually large pores was observed in liquid via high-resolution AFM. Force-distance (F-D) curves were also obtained using AFM tips modified to present or capture DNA. The modification of AFM tips allowed the tips to present covalently bound DNA, or to bind DNA that was pre-loaded in the protein crystal nanopores. The modified tips enabled the interactions of DNA molecules with protein crystals to be quantitatively studied while revealing the morphology of the protein crystal surface in detail, in buffer, thereby preserving the structure and properties of protein crystals that could be disrupted or destroyed by drying. The hexagonal space group was manifest at the crystal surface, as were the strong interactions between DNA and the porous protein crystals in question. In sum, this study furthered our understanding of how a new protein-based biomaterial can be used to bind DNA guest molecules.

---

\* Portions of this chapter are reproduced from: Dafu Wang, Julius D. Stuart, Alec A. Jones, Chris D. Snow, Matt J. Kipper\*. Measuring Interaction of DNA with Nanoporous Protein Crystals by Atomic Force Microscopy, *Nanoscale*, 13 (24) 1-871-10881, 2021, © 2021 Royal Society of Chemistry, used with permission.

### 3.1 Introduction

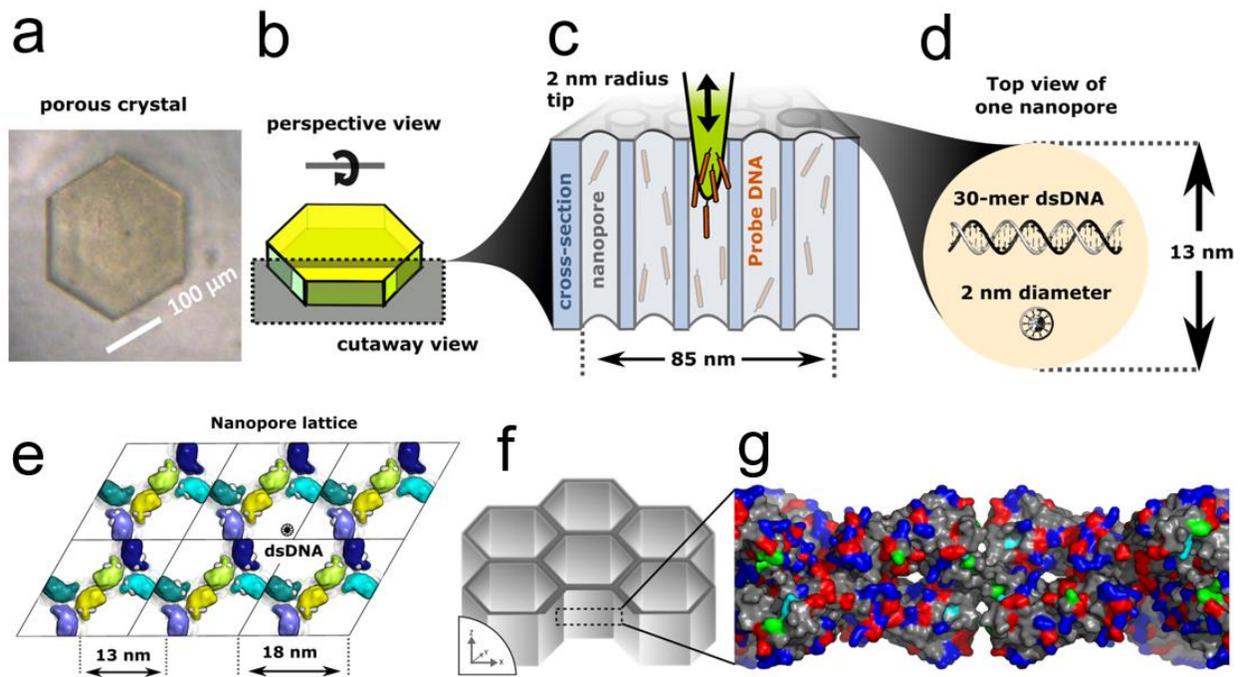
Protein crystals are a unique nanomaterial with highly ordered and well-defined three-dimensional structures. The enormous variety of crystal structures provides access to uniquely tunable and evolvable nanomaterials. We are interested in the unusual crystals of a putative isoprenoid binding protein from *Campylobacter jejuni* (Genebank ID: CJ0420, Protein Data Bank (PDB) code: 5w17, which we refer to as CJ. This protein readily and rapidly assembles into highly porous protein crystals (Fig 1).<sup>[1-4]</sup> Each unit cell within these crystals contains 12 protein monomers arranged according to the P622 space group. CJ crystals are hexagonal three-dimensional arrays with 13 nm-diameter nanopores (18 nm pore-center-to-center) that are aligned along the  $z$ -axis. The major nanopores extend from the top of the crystals to the bottom, reminiscent of a honeycomb. Once crosslinked, these materials offer an unusual combination of macroscopic stability, nanoscale precision, and a high capacity to uptake macromolecular guests. Previous studies have shown the capability of large-pore protein crystals for the capture of guest nanoparticles and proteins.<sup>[1, 2]</sup>

The uncommonly large pores of CJ crystals also provide ample space for double-stranded DNA (dsDNA, hereinafter as DNA), with a diameter of 2 nm, to be loaded and stored within each 13 nm-diameter nanopore (Fig. 3.1). We have observed strong affinity for nucleic acids to adsorb to the crystal interior. In this regard, the porous crystal shares a key attribute of viral capsids and spores. Efforts are underway to exploit this effect for applications including information storage and DNA delivery. We are also interested in determining if it is possible to use the extraction of polymers from the crystal nanopores by mechanical force as the basis for a force-sensitive signal transduction scheme.

To date, the basis of the favorable interaction has not been clear. Here, to investigate, we seek to directly quantify the attraction between DNA and the crystal nanopores. In this study, the characterization of crystals' surface morphology is also critical because the surface is where the crystal interacts with its environment. Guest molecules that are transported into or out of the crystal nanopores must traverse the interface represented by the crystal surface. Atomic force microscopy (AFM) is a powerful tool for the study of materials' surfaces, providing molecular-scale resolution of surface features.<sup>[5-11]</sup> <sup>[12, 13]</sup> Unlike electron microscopy, AFM in buffered solutions affords opportunities for *in situ* imaging and molecular force probing of fragile biomolecules in native conformations.<sup>[14-16]</sup> <sup>[12, 13, 17]</sup> Beyond revealing the surface structure, AFM can apply small controlled forces (pN level) to determine mechanical properties of materials and can characterize molecule-level adhesion events using force–distance (F–D) curves.<sup>[12, 18, 19]</sup> <sup>[17, 19-22]</sup> Accordingly, AFM studies of protein crystals have included the visualization of mechanical behaviors,<sup>[23]</sup> mapping the surface of morphology on 2-D crystal substrates,<sup>[8]</sup> imaging protein crystals in liquid phase,<sup>[14]</sup> and high-speed high-resolution imaging of crystallization dynamics.<sup>[24]</sup> Using solution AFM to characterize the protein crystals enables the periodicity and morphology of crystals to be studied in their mother liquid, thereby preserving the as-grown periodic protein crystal structure, which can be disrupted or destroyed by drying.<sup>[14]</sup> While the bulk structure of CJ protein crystals has been elucidated by X-ray crystallography, the surface morphology of CJ protein crystals with unusually large pores has not yet been characterized at the resolution of the unit cell containing the pores. Therefore, prior to this study, the degree of nanopore uniformity on the surface was unknown. This information impacts the design of protein crystals as hosts for guest molecules such as DNA. The present work was therefore undertaken to characterize the CJ protein crystal surface by AFM, and to determine the interactions of guest DNA with the protein crystal.

In our study, among other experiments, we preloaded DNA into the nanopores of CJ crystals. Modifying the AFM tip with dithionitrobenzoic acid provides a covalent attachment site for thiol-terminated DNA to be strongly connected by disulfide bonds to the AFM tip. Ideally, DNA present within the crystal can be captured by the AFM tip and pulled out of the protein crystal pores, or pulled off of the protein crystal surface. As illustrated in Fig. 3.1c, a sharp activated AFM tip is capable of penetrating deeply into the nanopores and thus capturing multiple DNA molecules. During the imaging process, the Bruker Bioscope Resolve AFM operated in PeakForce QNM mode, records a force-distance (F-D) curve at each pixel in the scanned area of the sample surface. Thus, every pixel in the AFM image contains an F-D curve. We were able to gather hundreds of thousands of (F-D) curves in a single imaging experiment without interrupting the imaging process. We were also able to exactly locate the force curves on the images, to gain a better understanding of the connection between the mechanical behaviors and the morphology of the surface.

We used AFM with modified tips to measure the interactions of DNA with the pores and surfaces of CJ protein crystals in the liquid phase. AFM tips were modified with chemistry to covalently bind thiol-terminated DNA that had been pre-loaded into the pores of the CJ crystal. For comparison, we also imaged the crystals with tips that were covalently modified with DNA prior to imaging. For the first time, we showed that the interactions of AFM-tip-conjugated DNA with a porous protein crystal can be directly and quantitatively measured. This study will enable us to tune protein crystals and solution conditions for DNA storage and release.



**Fig. 3.1** A periplasmic protein, “CJ”, from *Camphylobacter jejuni* forms (a) porous protein crystals that we stabilize via crosslinking. (b) Typical crystals are hexagonal prisms. (c) A hexagonal array of 13 nm-diameter nanopores runs from the top to the bottom of each crystal. Modified DNA molecules can be loaded into these nanopores and then pulled out using activated AFM probes capable of covalently bonding to the DNA. The top face of the crystal may then be probed using chemical force microscopy with DNA-modified AFM tips.<sup>[1-4]</sup> (d) Within each 13 nm-diameter nanopore, it is possible to fit numerous DNA double helices parallel to the host nanopore axis since DNA has a 2 nm diameter. A 30-mer DNA almost spans the nanopore diameter. (e) A top view of two adjacent nanopores (PDB code 5w17), with guest DNA to scale. (f) crystal schematic with nanopores cut away, and (g) zoomed in slice of nanopore side wall illustrating presence of ionizable amino acid. Carboxylic acids (Asp, Glu) are shown in red. Arginines are shown in cyan. Lysines are shown in dark blue. Notably, some of the Lysine sidechains have likely lost their positive charge by participating in glyoxal crosslinks. Glyoxal crosslinked crystals tend to diffract to modest resolution  $\sim 3.5\text{\AA}$  and surface lysine sidechains tend to be highly mobile. These two factors, as well as heterogeneity within the crystal, prevent us from assessing which lysines are likely to retain their positive charge. Finally, the figure shows histidine sidechains in green. Both the N- and C-terminus contain flexible regions that are not pictured in this crystal structure. The flexible C-terminal histag (not pictured) is of particular interest as a possible participant in DNA binding. Image was created by PyMOL.

## 3.2 Experimental

### 3.2.1 Materials

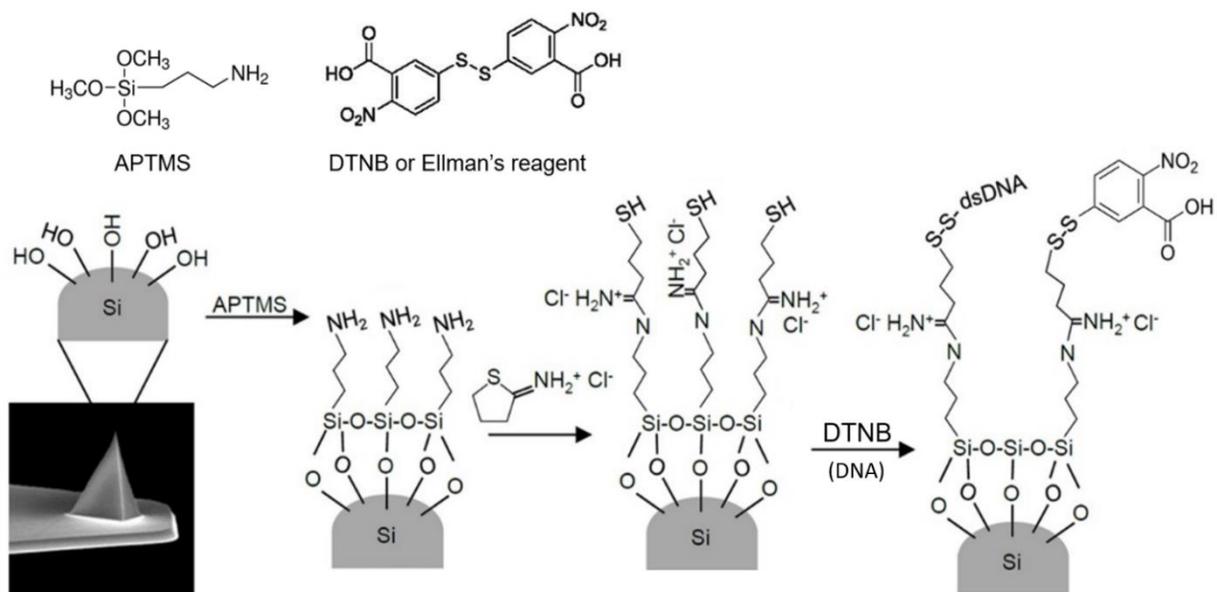
A Millipore Synthesis water purification unit was used to obtain 18.2 M $\Omega$  cm water, used for making all aqueous solutions. 3-(Trimethoxysilyl)propylamine (APTMS) for molecular vapor deposition (MVD) was purchased from EMD Millipore Corp. Traut's reagent (2-iminothiolane) used for tip modification was purchased from Chem-Impex International, Inc. 5,5'-Dithiobis-(2-nitrobenzoic acid), (DTNB, Ellman's reagent) were purchased from Thermo Fisher Scientific, Inc.

#### 3.2.1.1 CJ Crystals Growth

As described in previous work.<sup>[25]</sup> *Campylobacter jejuni* protein (CJ) was cloned into pSB3 expression vector. CJ expression was performed with BL21(DE3) *Escherichia coli* cells in Terrific broth. Induction was performed with 0.4 mM IPTG for 16 hours at 25 °C, followed by purification using immobilized metal affinity chromatography (IMAC). The target protein was then dialyzed into ammonium sulfate storage buffer (500mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 10% glycerol, pH = 7.4). CJ protein was concentrated to 15 mg/mL, aliquoted and stored at -30 °C. Then, one tube (24  $\mu$ L) of purified CJ protein (15 mg/mL) was thawed.<sup>[2, 26-28]</sup> The samples were kept on ice at all times. In the reservoir of a plastic CrysChem sitting-drop crystallization plate, 340  $\mu$ L of 4 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 40  $\mu$ L of 1 M bis-tris (pH = 6.5), and 20  $\mu$ L of DI H<sub>2</sub>O were mixed. Then, 1  $\mu$ L of the reservoir solution was pipetted into the top drop of the plastic sitting-drop crystallization plate. Finally, a 1  $\mu$ L aliquot of CJ protein solution was added by pipetting the protein solution directly on top of the drop of reservoir solution. Crystals typically grew to full size within 1 to 3 days.

### 3.2.1.2 CJ Crystal Crosslinking and Immobilization

In this study, medium-large CJ protein crystals were used, with a typical diameter of 400  $\mu\text{m}$  to 700  $\mu\text{m}$ , and typical height of 50  $\mu\text{m}$ . The crystals must be crosslinked so that we can later vary the solvent. Crystals were transferred (using a nylon crystallography loop, Hampton Research) from their growth well into a drop of 4.2 M trimethylamine *N*-oxide (TMAO), 0.175 M  $\text{H}_2\text{SO}_4$  at pH = 7.5, to wash for a minimum of 20 min. A drop of 390  $\mu\text{L}$  of 4.2 M TMAO, 0.175 M  $\text{H}_2\text{SO}_4$  at pH 7.5 plus 10  $\mu\text{L}$  of 40% glyoxal crosslinker was then prepared. Crystals were manually transferred into crosslinking solution and covered for 2 h. Meanwhile, a drop of 380  $\mu\text{L}$  0.1 M sodium citrate, 0.15 M NaCl at pH 5.0, 10  $\mu\text{L}$  of 50% hydroxylamine, and 10  $\mu\text{L}$  of 40 mg/mL dimethylamine borane complex was prepared. The crosslinked crystals were transferred to this “quenching” solution to eliminate reactive groups. After 8 hours in the quenching solution, crystals were ready to be used or stored. Crystals were stored in 4.0 M TMAO and washed briefly in water or adsorption buffer (30 mM KCL, 10 mM MES, pH = 6.0) prior to use. The surfaces of crystals were weakly negative charged after crosslinking, with a zeta potential of -16.6 mV at pH = 7.5. (Supporting Information Fig. A7). This modest negative potential suggests that simple electrostatic attraction is not the driving force for DNA adsorption. Empirically, the lack of DNA desorption in high salt washes (data not shown) further supports the idea that DNA binding is not dominated by electrostatic interactions that can be screened at high salt.



**Fig. 3.2** Schematic illustration of intended AFM tip modification process

For AFM experiments (described below), CJ crystals were immobilized on glass-bottom petri dishes (Willco Wells) employing a UV-curable glue (Bondic Inc.). The top of a crystal probe (Minitool HR4-217) was used to transfer a drop of UV-curable glue onto the surface of a petri dish (Ted Pella, Inc. 14025-20). The glue was gently and evenly spread on the dish surface to make the layer of glue as thin as possible. CJ crystals were transferred to the glue with a loop. Critically, the crystal was transported inside a tiny drop of buffer, such that the crystal was not desiccated. The UV-glue was viscous and did not noticeably mix with the buffer. The glue was then cured by exposing to UV-light LED (Bondic SK001) from above for 10 s. The glue cured after about 2 min, after which additional drops of buffer (typically ~5 mL) were added to the dish to prevent the crystal from drying

### 3.2.2 AFM Tip Modification

Bruker's ScanAsyst Fluid+ tips were modified to covalently attach DNA (Fig. 3.2). These tips have a slim shape with estimated tip radius as small as 2 nm, and a silica surface layer. To

clean the tip surface, AFM tips were placed in O<sub>2</sub> plasma chamber (Plasma Etch. Inc) with a 200–300 mTorr total pressure inside the chamber, and the power setting was adjusted to 38 W for 5 min to activate the hydroxyl groups on the silica surface of the tip.<sup>[29]</sup> Then, molecular vapor deposition (MVD) was used for amino-silane treatment of the surface of AFM tips. AFM tips were placed into a 1-L polypropylene jar. Two mL of APTMS aminosilane was added to a 10-mL scintillation vial, also placed in the polypropylene jar. The polypropylene jar was sealed using a screw cap lid, and placed in a 60 °C oven for 60 min. This allows the surface of the AFM cantilever tip to be modified with the APTMS by MVD forming an aminosilane layer anchored to the surface.<sup>[29]</sup>

Traut's reagent (2-iminothiolane) reacts spontaneously with primary amines (-NH<sub>2</sub>) at pH = 7.0 to introduce sulfhydryl (-SH) groups. We used this reaction with 1 mM Traut's reagent at room temperature in 50 mM KCl solution.<sup>[29]</sup> To activate the AFM tips for binding thiol-terminated DNA, AFM tips were modified using dithionitrobenzoic acid chemistry. Specifically, the remaining 2-iminothiolane solution was replaced with excess 5,5'-dithiobis-(2-nitrobenzoic acid), (DTNB, Ellman's reagent, 500 μM) in a 0.1 M dipotassium phosphate and sodium bicarbonate buffer (pH = 8). Activating the AFM tip with DTNB on the tip surface enables thiol-terminated DNA to be reversibly and covalently bound to the AFM tip via a thermodynamically favored disulfide exchange reaction. With the functional groups on the tip surface, the AFM tips would be capable of binding DNA.

Each step in the surface modification of the AFM tips was evaluated by X-ray photoelectron spectroscopy, using a Physical Electronics. 5800 spectrometers (Chanhassen, MN). This XPS uses a monochromatic Al K $\alpha$  X-ray beam source ( $h\nu = 1486.6$  eV), hemispherical analyzer, and multichannel detector. The binding energy scales for the samples were referenced to

the aliphatic contribution of the C1s peak at 284.8 eV. High-resolution spectra of the N1s, S2p and P2p envelopes were acquired with 0.1 eV steps, and an X-ray spot size of 800  $\mu\text{m}$ . Given this size, the XPS spectra report surface chemistry of the tip, cantilever, and probe, but the tip is composed of the same material. Analyses were performed at a photoelectron take-off angle of 45°. Peak fitting of the N1s and S2p envelope was performed in MultiPak (Ulvac-Phi, Inc.) using Gaussian/Lorentzian peaks and a Shirley background correction. The morphology of both unmodified and modified AFM tips was also imaged by field emission scanning electron microscopy (FESEM, JEOL JSM-6500F).

### **3.2.3 DNA Loading and AFM Probing**

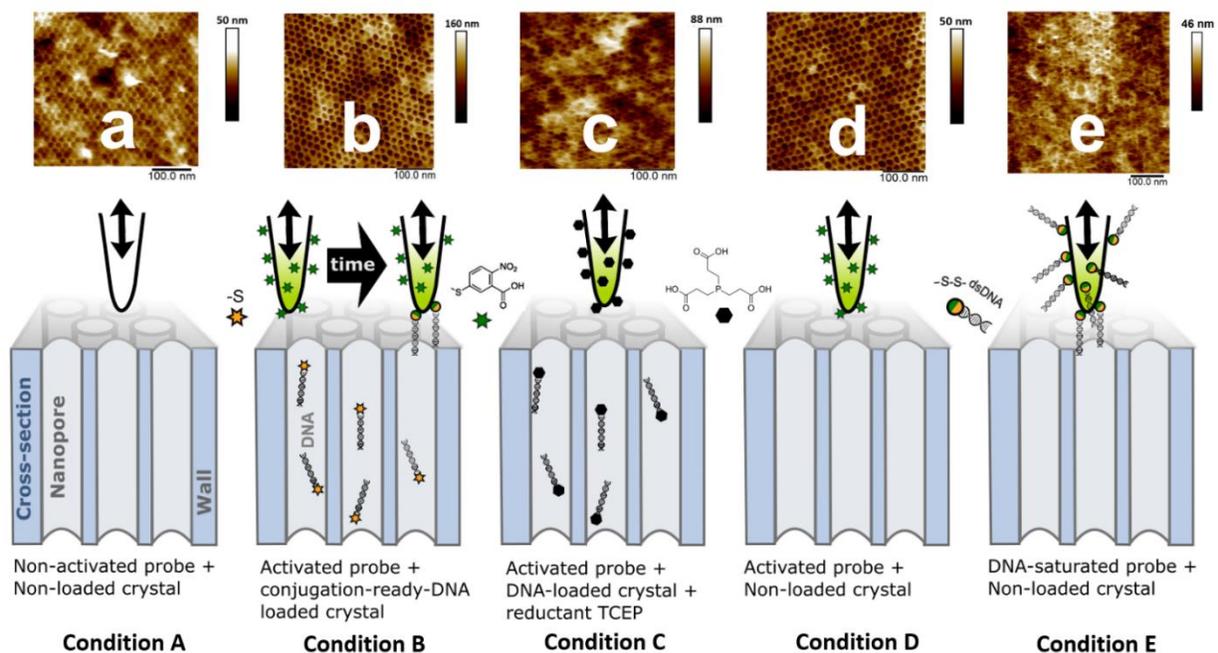
We operated the AFM (Bruker Bioscope Resolve, mounted on a spinning-disc confocal microscope built around a Nikon Eclipse TiE) in quantitative nano-mechanics (QNM) PeakForce Capture mode. All images and force curves were collected using ScanAsyst Fluid+ tips (Bruker). Crystal imaging was performed in TE (Tris-EDTA) / DI H<sub>2</sub>O buffer (5 mM Tris-HCl, 1 mM EDTA, pH = 7.5) with crystals affixed to the bottom of a glass petri dish. The AFM line scan rate was set to 1.0 Hz and the peak force tapping frequency was set to 1.0 kHz. The peak force set point was set to 2 nN. Notably, the force used here is large with respect to the forces used by investigators who use optical trap experiments to study the interactions of DNA molecules with other partners. Specifically, Dario Anselmetti et al. used peak force of ~800 pN, while Stuart Lindsay and coworkers used a peak force of 160 pN.<sup>[30, 31]</sup> The QNM PeakForce Capture mode is based on force-versus-distance measurements in which the tip oscillates sinusoidally, at a frequency far below the resonance frequency of cantilever,  $f_0$ . This mode provides a high-resolution peak force mapping as well as sensitivity to record nano-mechanical behaviors at high spatial resolution.<sup>[32]</sup> Tip-sample interactions are measured with pN-resolution by the deflection of the cantilever.

Analysis of the AFM data was performed in NanoScope (Bruker, Inc.), Origin (OriginLab, Inc.), Python (Version 2.7), and Matlab (Version 2019).

The protein crystal sample was imaged by both AFM (unmodified tip), and confocal microscopy to confirm immobilization and to verify that the crystal surface is clean. When crystals were incorrectly prepared, their surfaces could be obscured by aggregated protein. To ensure that the crystals were competent to uptake DNA, we used time-lapse confocal microscopy (z-stack imaging) to monitor and confirm the loading of fluorescently labeled DNA. First, the CJ crystals were photobleached to prevent interference from background fluorescence. Prior to DNA loading, as a control experiment, the protein crystal was imaged using an activated AFM tip (terminated with the dithionitrobenzoic acid, but without DNA) in TE buffer (Condition D in Fig. 3.3). Then, the TE buffer solution was replaced by 100  $\mu$ L 50  $\mu$ M 30mer-DNA with two terminal thiol groups (sense strand, 5'-3': /5ThioMC6-D/TAG GCG ACT CGA CGG TCT TAC GCG TTA CGT, anti-sense strand, 5'-3': ACG TAA CGC GTA AGA CCG TCG AGT CGC CTA) in TE buffer. Prior to loading, a stock of the same 30mer-DNA was fluorescently labeled with TAMRA (carboxytetramethylrhodamine) labeled DNA (Integrated DNA Tech.) for 30 minutes. During loading, 10% (90%) of the DNA was TAMRA-labeled (unlabeled). Next, after washing three times with TE buffer (30 min per wash), the samples were incubated with 100  $\mu$ L of 50  $\mu$ M tris(2-carboxyethyl)phosphine (TCEP) in TE buffer for 30 minutes to reduce disulfide bonds. After reduction of disulfide bonds with TCEP, the DNA-loaded crystal was again washed with 1 mL TE buffer for 30 min, three times, to remove the TCEP. Retention of the DNA was confirmed by confocal microscope imaging after each wash step. The DNA-loaded crystal was then imaged with an activated AFM tip (Condition B in Fig. 3.3), and F-D curves were collected at each pixel in the AFM image.

As a control experiment, the procedure for loading the protein crystal described above was repeated using DNA previously reacted with 100  $\mu$ L of 14 mM iodoacetamide (in 100 mM Tris-HCL buffer, pH = 8.3). Ideally, iodoacetamide will permanently “cap” the DNA to ensure that it cannot covalently bind to the activated AFM tip. The crystal loaded with deactivated, “capped” DNA was imaged with an activated AFM tip (Condition C in Fig. 3.3).

AFM imaging and F-D curve collection was conducted for five different experimental conditions as described above, using combinations of un-modified, activated, and DNA-modified tips and either loaded or unloaded protein crystals. The five experimental conditions are: (condition A) un-modified AFM tip on an unloaded crystal, (condition B) activated AFM tip on a crystal loaded with thiol-bearing DNA, (condition C) activated AFM tip on a DNA-loaded crystal in the presence of TCEP and iodoacetamide, (condition D) activated AFM tip on an unloaded crystal, and (condition E) 30-mer DNA-modified AFM tip on an unloaded crystal. From each AFM image, F-D curves were manually assigned to one of two classes, corresponding to protein crystal surface features: pores and walls.

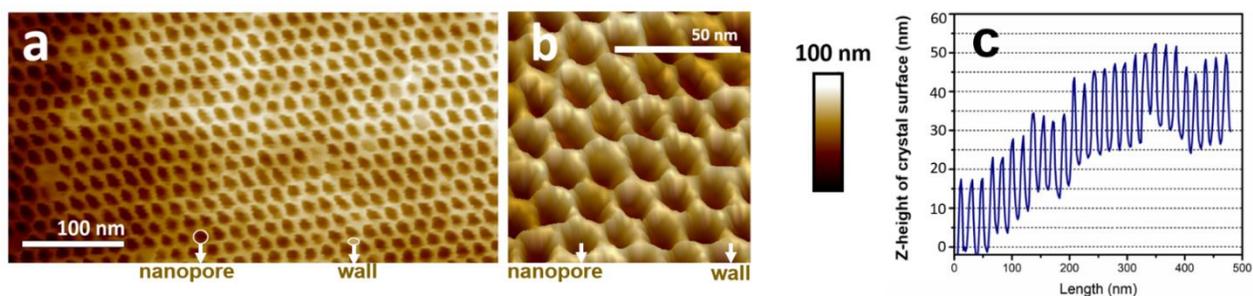


**Fig. 3.3** Force microscopy schematic illustration of AFM tips with protein crystals. The condition indices A-E in this figure correspond to the indices in Figure 3.9. All conditions accompany by the corresponding surface morphology imaging by AFM. **(a, condition A)** The non-activated AFM tip is white. **(b, condition B)** The activated AFM tip is green and covered with green stars to represent the thiol-reactive Ellman's reagent leaving group. DNA oligos preloaded into the crystal have a terminal thiol group (orange stars). In this condition, we expect the tip to conjugate and “fish” out multiple oligos at the outset of the experiment. **(c, condition C)** The reducing agent TCEP (black hexagons) should reverse any disulfide bond formation between the DNA and the activated AFM-tip, yielding detached oligos. **(d, condition D)** A no-DNA control. **(e, condition E)** In contrast, an activated AFM tip can be saturated with DNA molecules in solution prior to encountering the crystal.

### 3.3 Results and Discussion

#### 3.3.1 AFM Imaging on Surface Morphology

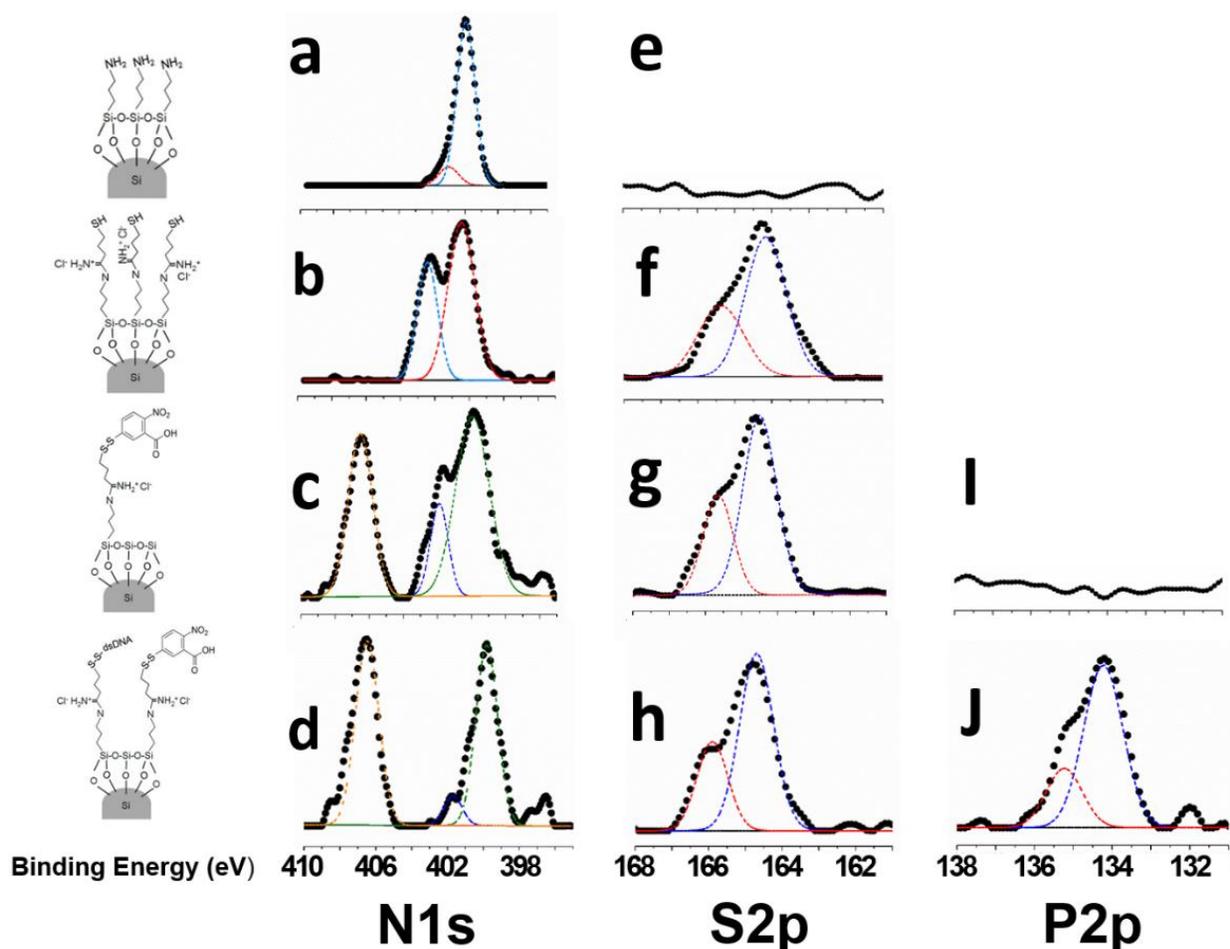
AFM imaging resolves details of the porous CJ crystal surfaces (Fig. 3.4). The CJ crystal surface presents a regular honeycomb pore/hole structure consistent with single-crystal x-ray diffraction data (PDB entry: 5w17).<sup>[1-4]</sup> The observed surface structure of CJ protein crystals did not change significantly when imaged with different tip modifications. The CJ crystal surface structure was regular and uniform, with features that are consistent among different protein crystal samples. Notably, the previously reported structure obtained from X-ray diffraction (13 nm-diameter pores in a hexagonal array) is confirmed by AFM to be manifested at the crystal surface.<sup>[1, 2, 4]</sup> The AFM z-height image is a convolution of the surface height and the geometry of the AFM tip. A one-dimensional z-height trace across the centers of multiple pores is shown in Fig. 3.4(c). When the tip is located over the center of a pore, a minimum in the z-height is recorded, and when the AFM tip is located over the center of the wall separating two pores, a maximum in the z-height is recorded.



**Fig. 3.4** (a) AFM image of CJ crystal surface (b) 3D rendering of a portion of the height data from Fig. 3.4 (a), (c) a side height section of crystal surface from Fig. 3.4 (a), across the center of multiple pores.

### 3.3.2 Modification of AFM Tips

The spring constant and tip diameter of each tip used for imaging and quantitative measurements were quantified using hardness and surface roughness standards. The spring constant of a fully modified AFM tip used to collect data for this paper was 0.94 N/m, with an estimated tip diameter of 5.09 nm (ETD, data from NanoScope). Modified and unmodified tips were also imaged by field emission scanning electron microscopy (FESEM, Supporting Information Fig. A9), confirming that the silanization and subsequent chemical modifications do not alter the tip geometry on a macroscopic scale that might be observed via SEM.<sup>[33]</sup> X-ray photoelectron spectroscopy (XPS) was used to characterize the modification of AFM tips. In this study, the high-resolution XPS spectra, as well as the changes of N1s, S2p, and P2p peaks confirm each step of modification chemistry shown in Fig. 3.5. Activation of the tip with excess 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) enables the 30mer DNA to be covalently bound to the AFM tip via disulfide bonds with the thiol-terminated DNA. This enables DNA to be bound and extracted from nanopores of CJ crystals by the AFM tip.



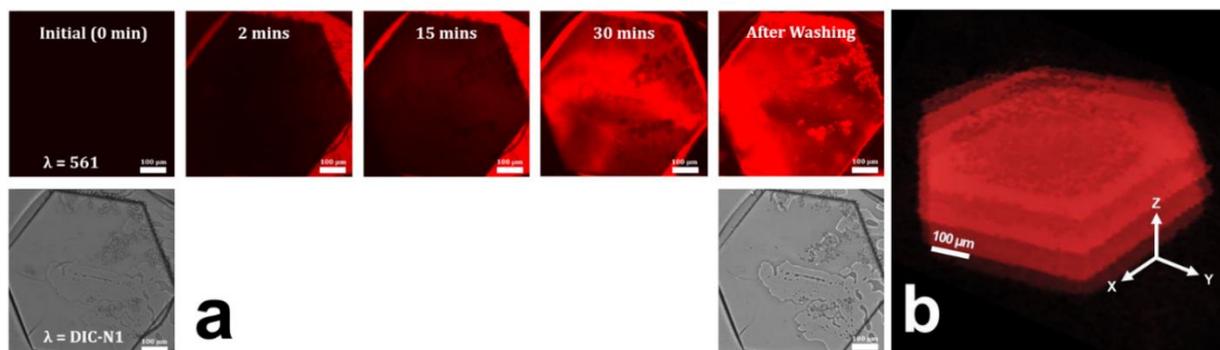
**Fig. 3.5** High-resolution XPS spectra of ScanAsyst Fluid+ tips at different stages of modification in the regions of the N1s: **(a)** +APTMS (399.82 eV N-sp<sup>2</sup> C and 400.95 eV ammonium), **(b)** +2-iminothiolane (400.40 eV N-sp<sup>2</sup> and 402.04 eV, ammonium), **(c)** +DTNB (399.84 eV N-sp<sup>2</sup>, 401.93 eV ammonium, and 406.55 eV nitrite O-N-O), **(d)** + 30mer DNA (399.84 eV pyridine N-sp<sup>2</sup>, 401.93 eV ammonium, and 406.55 eV nitrite O-N-O); and S2p: **(f)** +2-iminothiolane (thiols), **(g)** +DTNB (164.56 thiols), **(h)** + 30mer DNA (disulfide, thiols); and P2p: **(j)** +30mer DNA (O-P-O<sub>3</sub> phosphate) envelopes confirmed that all of the samples were successfully modified at each step and able to bind DNA on the surface of AFM tips, **(e)** confirms that there is no sulfur signal prior to reaction with Traut's reagent, and **(i)** confirms that there is no detectable phosphorus prior to reaction with DNA.<sup>[34, 35]</sup>

### 3.3.3 DNA Loading

Absorption of the DNA into the crystal following immersion in guest DNA solution containing 10% 30mer-DNA fluorescently labeled with TAMRA is confirmed by confocal microscopy. Throughout loading, fluorescence intensity increased in the CJ protein crystal interior.

After washing with TE buffer and 25 mM TCEP in TE buffer, confocal microscope imaging shows the retention of guest DNA (Fig. 3.6). From XRD data, each unit cell has a free volume of 1413 nm<sup>3</sup> (1.413×10<sup>-15</sup> μL).<sup>[1-4]</sup> Therefore, a medium-large crystal (400 μm diameter, 50 μm height) would contain about 3.7 trillion unit cells. According to the confocal microscopy imaging, the estimated height of this particular protein crystal was 50 μm, with a diameter of 745 μm, and the crystal's side length was 363 μm.

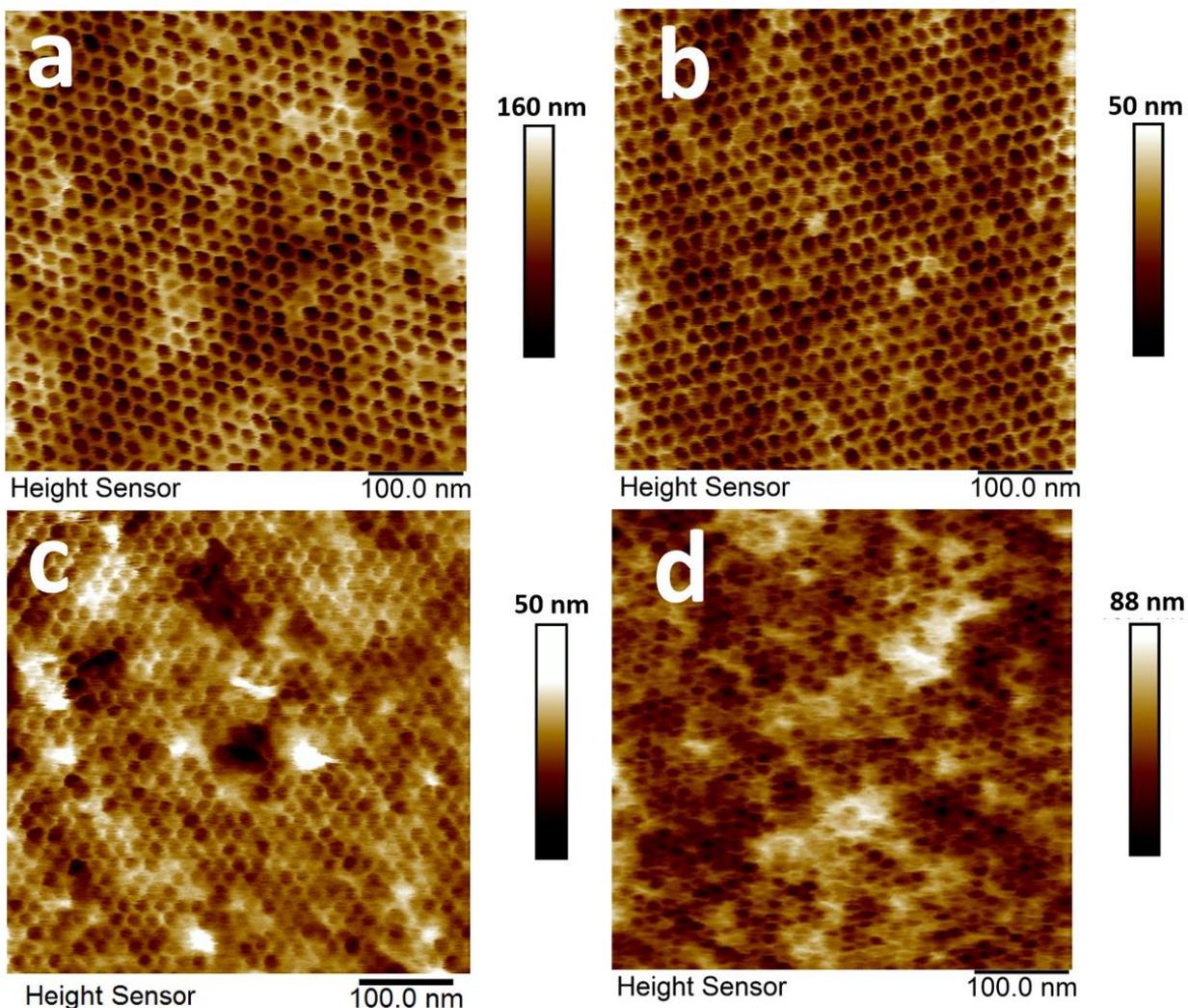
We can obtain an estimate for the 30-bp DNA concentration inside the crystal by comparing the relative volume of the crystals and the supernatant as well as the relative fluorescence intensity. Specifically, when we quantify the volume of the solution outside the crystal, the volume and concentration of DNA outside the crystal at the experiment outset (100 μL, 50 μM), the estimated volume of the crystal (0.0237 μL), and the relative fluorescence intensity of the DNA inside and outside the crystal, we can use a material balance to estimate a final average intra-crystal concentration of 138.86 μM (with a maximum of 205.77 μM) (Supporting Information Table A3). At this DNA loading density, with the known unit cell volume (1413 nm<sup>3</sup>), we can estimate that the average concentration is approximately 0.117 DNA molecules per unit cell within the crystal (maximum 0.174 DNA molecules per unit cell), if guest DNA molecules are distributed randomly throughout the crystal. Since the CJ unit cell is ~5 nm tall, and a 30-bp DNA is ~10 nm tall, we envision that DNA would not be crowded within the crystal but would be commonly encountered by chance within the nanopores.



**Fig. 3.6** (a) Confocal microscope images of fluorescent DNA loading into a CJ protein crystal (0-30 minutes time-lapse), and after washing with TE to reduce DNA outside the nanopores. (b) The 3-D distribution (z-stack) of 5 confocal microscopy images of labeled DNA after TCEP washing. The z-stack planes are separated by 10  $\mu\text{m}$  in the z direction.

### 3.3.4 Interaction of DNA and CJ Protein Crystals

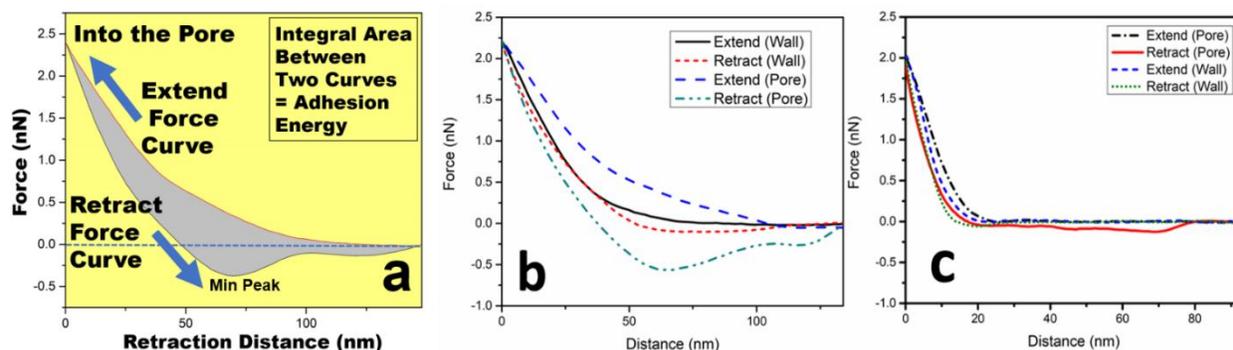
Upon interaction of the activated AFM tip with the DNA-loaded protein crystal, the AFM tip can be used to remove DNA from the protein crystal pore or surface, and to measure the force of the interaction. In PeakForce QNM mode, the AFM captures and records a force-distance (F-D) curve at each pixel of the scanned area of the CJ crystal's surface. During this study, we collected hundreds of thousands of F-D curves by AFM under different conditions. To measure the F-D curves and obtain the adhesion energy of DNA with the pores and walls of CJ protein crystals, five different combinations of tip modification (Fig. 3.2) and DNA loading were used as shown in Fig. 3.3. Respectively, each combination contains two types of areas: pores and walls. Fig. 3.7 shows that loading DNA in the nanopores of CJ protein crystals does not substantially change the surface morphology of crystals. Varying levels of surface aggregation apparent amount the images in Fig. 3.7 are likely due to crystal-to-crystal variation in growth, washing, and crosslinking processes. The high-resolution images obtained for each condition provides confidence that our AFM parameters and sample immobilization method provide data of suitable quality to further investigate the interactions between DNA and the host CJ crystals.



**Fig. 3.7** AFM images of (a) a crystal loaded with DNA imaged using an activated AFM tip, (b) an unloaded crystal imaged with an activated AFM tip, (c) an unloaded crystal imaged with a non-modified AFM tip, and (d) a crystal loaded with DNA imaged with an activated tip after deactivating thiols with iodoacetamide

The adhesion energy between the DNA and the CJ protein crystals at each tap can be calculated as the integral of the area between the extend force curve and retract force curve (corresponding to the grey area in Fig. 3.8 (a)). The measured interaction for this representative pore pixel has a larger force volume, a longer distance of force measurement, as well as a higher adhesion energy than the wall pixel. Remarkably, it can be seen in Fig. 3.8 (a) that the interaction of the AFM tip with the protein crystal in both the extend and retract portions of the curve occurs

over more than 100 nm in the z-direction. We hypothesize that the crosslinked protein crystal surface deforms elastically under the approximately 2 nN peak force used here (Fig. 3.9c). Axial and lateral deformation may contribute to the penetration distance of probing, and induce energy changes by performing extra work. Notably, the length scale of interaction is significantly reduced for an AFM tip that was DNA- modified and used to probe a crystal that is not incubated with probe DNA (Fig. 3.8 (c)).



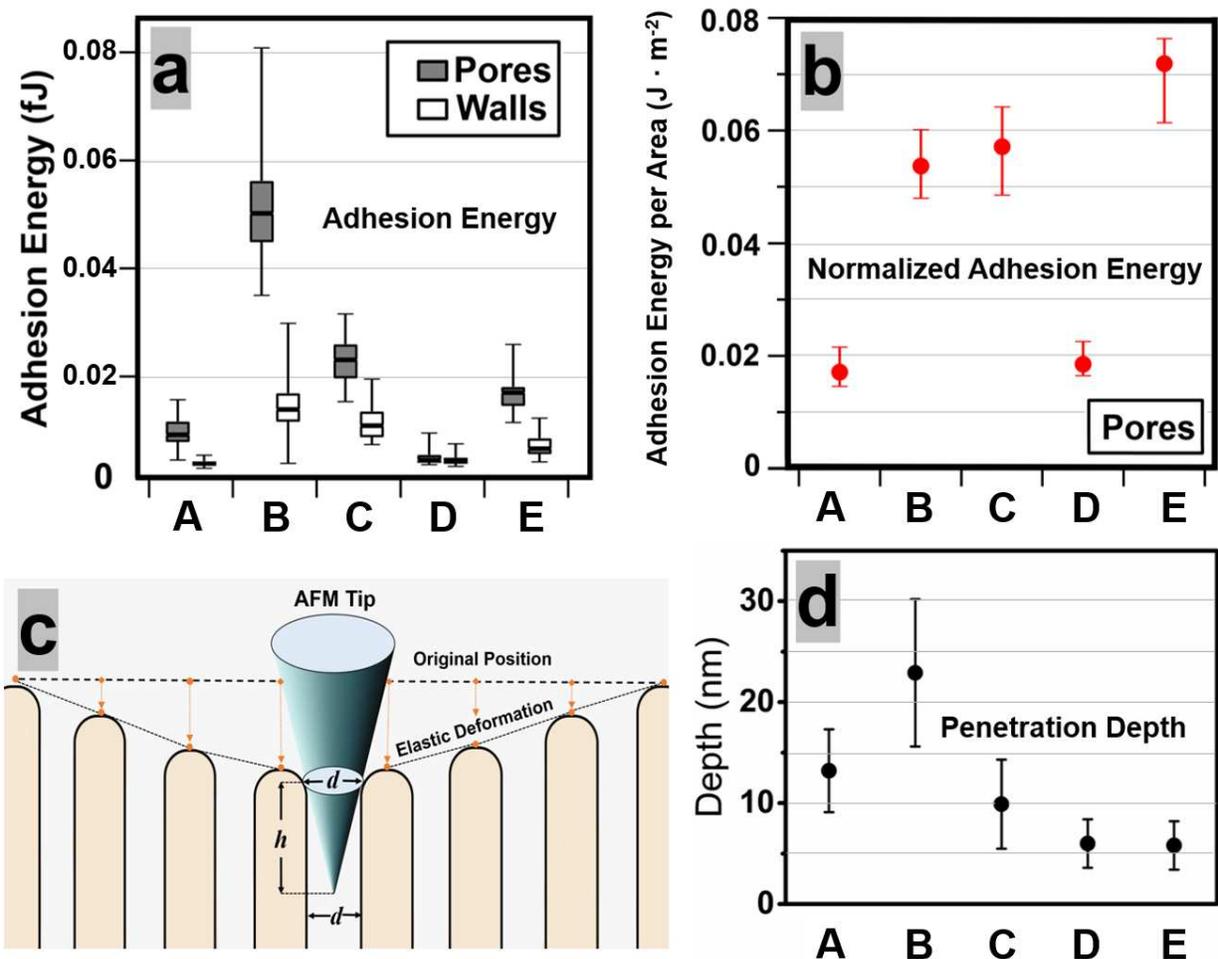
**Fig. 3.8 (a-b)** Representative examples of force-distance (F-D) curve obtained using a modified AFM tip on crystal loaded with DNA, **(a)** F-D curve of a nanopore center, **(b)** comparison of F-D curves of representative pore and wall pixels. **(c)** F-D curves obtained using a DNA-modified AFM tip on an unloaded crystal obtained from both pore and wall pixels.

We randomly selected 20 sets of F-D curves on pores and walls for each of the five experiments described in Fig. 3.3 (200 sets of F-D curves total). To account for different distances of tip penetration into the pores, the average adhesion energy in the pores was normalized by an estimated average pore area of interaction, calculated by approximating the area of a 13-nm diameter cylindrical pore ( $A_{pore\ surface} = 13nm \cdot h \cdot \pi$ ) at the local depth of penetration  $h$  of the tip (Fig. 3.9 (b)). When there was no DNA loaded in the protein crystals, there was no significant difference between the average adhesion energy of pore center pixels and wall pixels, when measuring with a modified AFM tip. When using non-activated AFM tips to measure the adhesion energy on unloaded protein crystals, both pore and wall pixels had their lowest mean adhesion

energy among tested conditions (case A in Fig. 3.3 and 3.9). The interaction measured by the modified AFM tip for pore center pixels for the CJ protein crystal loaded with DNA has the highest average adhesion energy (case B in Fig. 3.3 and 3.9). Among the sample curves, the highest adhesion energy for this class reaches  $8.08 \times 10^{-2}$  fJ, and the average adhesion energy is  $5.12 \times 10^{-2}$  fJ. This adhesion energy is much larger than the adhesion energy for the unloaded crystal and unmodified AFM probe, indicating that the adhesion of the DNA to the protein crystal is being measured. Furthermore, in many cases the total interaction persisted over a remarkably large total tip travel distance of about 150 nm. The tip-to-pore center interaction was reduced when putative disulfides were reduced to thiols with TCEP and capped with iodoacetamide (case C in Figs 3 and 10), while the average adhesion energy at the wall areas did not change much upon reduction. In addition, we also measured the interaction between an inactivated AFM tip with a DNA-loaded crystal (Supporting Information Fig. A1). The experiment showed similar results of force and adhesion energy level as condition A (of Fig. 3.3 and Fig. 3.9 (a-b)). This confirms that the inactivated AFM tip does not have a strong interaction with DNA-loaded crystals. However, when normalized by interaction area, the interaction energy per area is not different from the non-reduced experiment, because the average penetration depth for tips that were preloaded with DNA was only 5.8 nm whereas the average penetration depth for tips that acquired DNA from inside the crystal is  $22.9 \pm 7.3$  nm. This suggests that reduction with TCEP (with an average penetration depth of 9.90 nm) was insufficient to eliminate the tip-DNA interactions. For comparison, the average penetration depth for un-modified tips on unloaded crystals was 13.22 nm, while the average penetration depth for activated tips on unloaded crystals is 5.96 nm. The difference of penetration depth will directly lead to the difference of effective area of interaction, and therefore

affect the total adhesion energy. Future single-molecule studies could further elucidate the interactions of individual dsDNA with protein crystal surfaces and pores.

To further confirm that the strong affinity seen with active tips and DNA-loaded crystals was due to DNA bound to the tip surface, we conjugated DNA directly to the tips rather than allowing activated tips to capture DNA from the crystal (case E in Fig. 3.3). The DNA-conjugated tips did not penetrate as far into the pores as unmodified tips. The solution-based conjugation of DNA to the tips (Fig. 3.9 (case E)) resulted in significantly higher adhesion energy per area than tips lacking DNA, and the nanopore pixels were clearly different from the wall pixels. The normalized adhesion energy per unit area was comparable to cases B and C, where DNA was first loaded into the crystals, and retrieved with activated tips. Further evidence for the strong interaction between the DNA and the protein crystals is provided by comparing cases B and E to case D, in which an activated tip was used in the absence of DNA. In this case we found that the adhesion energy for the pore and wall areas and the normalized adhesion energy were similar to values found for the unmodified tip with no DNA.



**Fig. 3.9** Adhesion and penetration for the various AFM probe conditions (A,B,C,D,E) illustrated in Fig. 3.3. A: naked probe, B: activated DNA-capturing probe, C: after reductive DNA cleavage, D: activated probe lacking DNA, and E: probe saturated with DNA in solution. (a) Box plot (min, 25th, 50th, 75th percentiles, and max) of the adhesion energy for 20 pore center pixels and 20 wall pixels for each of the experiments depicted in Fig. 3.3, and (b) adhesion energy in the pores, normalized by the interacting pore area (adhesion energy divided by average area computed from the depth of penetration of the AFM tip into the pore) for 20 pore center pixels. The conditions labeled A-E here correspond to the conditions described in Fig. 3.3. (c) Schematic showing axial elastic deformation of the protein crystal surface (lateral deformation is not illustrated), and the AFM probe penetration into the pores, “*d*” represents the effective diameter of a nanopore, “*h*” represents the depth of probing. (d) probe penetration depth with error bar signifying the distribution and mean number of probe penetration depth into the nanopores respectively correspond to the experiment A to E. Conditions and detailed data are listed in the Supporting Information Table A2 (a-c)

Literature suggests that the mechanical force needed to rupture a disulfide bond may be 3.8 nN.<sup>[36]</sup> Therefore the peak force used in these experiments was set to about 50% of this value (2

nN). For each extension-retraction trace, in addition to extracting the adhesion energy, we can calculate the maximum force (maximum difference between the extension and retraction curves). A box plot for these maximum force values corresponding to Fig. 3.9 is shown in Supporting Information Fig. A4. The value of maximum force was in a range of 760.4 pN (min) to 2029.8 pN (max), and the mean maximum force was 1509.5 pN, well below the force needed for mechanical disulfide rupture. Furthermore, this force is likely distributed among multiple DNA molecules bound to the tip and interacting with the crystal. Similarly, the maximum force observed across DNA-modified AFM tip with unloaded protein crystal's extension-retraction traces was 2029.8 pN. Therefore, we expect that DNA molecules that successfully conjugate to the tip will typically remain conjugated for many subsequent AFM tip oscillations. To confirm this, we have verified that the AFM tip that was pre-conjugated with DNA (pores of scenario E at Fig. 3.3 and Fig. 3.9) had a mean adhesion energy of  $1.68 \times 10^{-2}$  fJ for the first 10 nanopores imaged and a mean adhesion energy of  $1.70 \times 10^{-2}$  fJ for the last 10 nanopores imaged, 16 minutes later. In accord with this quantitative test, we further noted no overall trend in the adhesion energies collected with this tip over the course of the 16-minute image collection. The inverse experiment is to look for a chronological trend in the adhesion energies for the center of nanopores scanned with the tip that was activated but not explicitly loaded with DNA prior to the AFM scan. It is conceivable that the first nanopores might have a systematically lower adhesion energy due to the AFM tip not yet having conjugated a full complement of DNA molecules. However, in fact we observed no statistically significant difference between the first 20 nanopores (mean adhesion energy was  $2.48 \times 10^{-2}$  fJ), the subsequent 20 nanopores (mean adhesion energy was  $2.47 \times 10^{-2}$  fJ), and the last 20 nanopores (mean adhesion energy was  $2.41 \times 10^{-2}$  fJ) (Supporting Information Fig. A5). We therefore conclude that the AFM tip in this experiment (pores of scenario B in Fig. 3.3 and Fig.

3.9) had ample time to conjugate to DNA during the setup time prior to the beginning of the image collection. As shown in Fig. 3.9 (a-b), there is a large difference in the adhesion energy on a per-pore basis, and these adhesion energies are quite consistent between neighboring pixels within a given pore, including pixels that are present on different horizontal scan lines, but there is no obvious correlation between nanopores that are probed consecutively. It is remarkable that individual pores demonstrate a consistent attachment strength when the tip returns to the nanopore on subsequent scan lines (~100 milliseconds later). This implies that neither the composition of the tip nor the environment inside the nanopore is significantly varying on the 100-millisecond timescale. This further supports the case that the DNA complement of the tip is not changing on a timescale that exceeds a typical expected time to move from one nanopore to the adjacent nanopore during the peak force scan.

The sinusoidal vertical oscillation profile used in the PeakForce QNM mode (as opposed to a linear ramp, used in typical force-volume mapping) ensures that the vertical tip velocity approaches a minimum of 0 as the tip reaches its fully extended position. Therefore, at the tap frequencies (1kHz) used, the AFM tip will dwell within 1 nm of the maximum extension into the nanopores for approximately 52 microseconds and will be moving at a maximum velocity of less than  $20 \cdot \pi$  nm/millisecond while the tip is within 10 nm of the maximum extension into the nanopores. At these speeds, we expect the short DNA molecule(s) that are conjugated to the AFM tip and submerged within the nanopores will have ample time to sample alternative molecular conformations and equilibrate with their local crystalline environment. This is consistent with the low intra-pore variation. To verify that the DNA had sufficient time to equilibrate (and thereby find conformations with maximal affinity for the local protein crystal matrix), we also performed peak force scans with a tap frequency of 2.0 kHz. Despite decreasing the time for DNA to

equilibrate, no change was observed in the mean adhesion energy, nor in the mean number of retraction trace minima. (Supporting Information Fig. A2 and Table A3)

In theory, the maximum interaction of DNA with protein crystals could be recorded only when the AFM tips that bind with DNA were retracted from the center of the pores. The adhesion energy map Supporting Information Fig. A8 (b) emphasized this conclusion, there was a clear correspondence between the position of the strong interaction and the position of the nanopores. Therefore, to further understand this process, we expanded the number of randomly selected F-D curves per nanopore with an activated AFM tip on the crystal loaded with DNA to 5328, and all samples were from different individual nanopores. By finding and counting the zero point(s) of the first derivative on each retract force curve, we can further obtain the number of minima on each retract force curve. One of the interesting features observed in the F-D curves was multiple minima in the retraction (Fig. 3.8 (a-c) and Fig. A8 (d)). We hypothesize that these would be related to the formation of stronger specific DNA-crystal interactions during retraction. Each of the minima on the retract curve could represent a single interaction of a DNA molecule and a binding site on the protein crystal. 58.4% of the retract curves contained between 1 and 4 minima. The distributions of adhesion energies and number of retract curve minima are similar. Meanwhile, the adhesion energy distribution of F-D retract curves with different numbers of minima in Fig. A8 (e), shows no trend or obvious correlation between the number of minima and total adhesion energy. Among all the samples, over 99% of adhesion energy was under  $6.0 \times 10^{-2}$  fJ, 78.3% of adhesion energy results were concentrated in the  $1.0 \times 10^{-2}$  to  $4.0 \times 10^{-2}$  fJ range, 50% of the adhesion energy values were between  $1.60 \times 10^{-2}$  and  $3.29 \times 10^{-2}$  fJ. F-D curves with total adhesion energy between  $2.0 \times 10^{-2}$  and  $3.0 \times 10^{-2}$  fJ counts 31.3%, which was the range where the adhesion energy was most concentrated among all the samples. The mean adhesion energy

among all samples was  $2.77 \times 10^{-2}$  fJ. The maximum adhesion energy recorded was  $8.77 \times 10^{-2}$  fJ. The comparison to case A (non-activated tip) and case D (activated tip), which were conducted in the absence of DNA, shows that the majority of this interaction energy should be attributed to the DNA-crystal interaction. Additional interactions that could also contribute include potential friction forces with the inner walls of protein crystals.

### **Chapter 3 Summary**

For the first time, the porous surface structure of this unusual type of protein crystal was observed, and imaged in the liquid phase via high-resolution AFM, revealing details of the porous crystal surface. In this study, DNA was successfully loaded into protein crystals. Loading and strong adhesion were confirmed by confocal microscope imaging. All the AFM data was likewise consistent with the confocal microscope imaging in confirming a strong interaction of DNA with CJ protein crystals. As intended, DNA-modified AFM tips had a dramatically stronger interaction with the crystal nanopores, presumably via the extra work associated with extracting DNA from the crystal during retraction. Our optimized AFM characterization quantitatively measures and analyzes the mechanical behaviors, and the nano-scale variations in the adhesion energy between DNA and protein crystals. We therefore propose that this method could be used more generally to study the interactions between guest molecules and porous crystals. In the future, we propose to use machine learning to process and classify the hundreds of thousands of F-D curves we have collected. We expect that this analysis will provide further insights into how DNA interacts with protein crystals. In turn, this knowledge will advance our understanding of CJ protein crystals as capable and reliable DNA containers, and provide a basis for engineering functional biomaterials that are responsive to small pulling forces.

## REFERENCES

1. Hartje, L.F., et al., *Adsorption-Coupled Diffusion of Gold Nanoclusters within a Large-Pore Protein Crystal Scaffold*. J. Phys. Chem. B, 2017. **121**(32): p. 7652-7659.
2. Kowalski, A.E., et al., *Gold nanoparticle capture within protein crystal scaffolds*. Nanoscale, 2016. **8**(25): p. 12693-12696.
3. Huber, T.R., et al., *Installing Guest Molecules at Specific Sites within Scaffold Protein Crystals*. Bioconjugate Chem. , 2017. **29**(1): p. 17-22.
4. Huber, T.R., et al., *Programmed Assembly of Host–Guest Protein Crystals*. Small, 2017. **13**(7): p. 1602703.
5. Brown, B.P., et al., *Opportunities in high-speed atomic force microscopy*. Small, 2013. **9**(19): p. 3201-11.
6. Beckwitt, E.C., M. Kong, and B. Van Houten. *Studying protein-DNA interactions using atomic force microscopy*. in *Seminars in cell & developmental biology*. 2017. Elsevier.
7. Dufrene, Y.F., et al., *Imaging modes of atomic force microscopy for application in molecular and cell biology*. Nat. Nanotechnol. , 2017. **12**(4): p. 295-307.
8. Yamamoto, D., et al., *Streptavidin 2D crystal substrates for visualizing biomolecular processes by atomic force microscopy*. Biophys. J. , 2009. **97**(8): p. 2358-67.
9. YuG, K., et al., *Molecular resolution imaging of macromolecular crystals by atomic force microscopy*. Biophys. J., 1997. **72**(5): p. 2357-2364.
10. Hedayati, M., et al., *Protein adsorption measurements on low fouling and ultralow fouling surfaces: A critical comparison of surface characterization techniques*. Acta Biomater. , 2020. **102**: p. 169-180.
11. Hedayati, M., et al., *Nanostructured surfaces that mimic the vascular endothelial glycocalyx reduce blood protein adsorption and prevent fibrin network formation*. ACS Appl. Mater. Interfaces, 2018. **10**(38): p. 31892-31902.
12. Pfreundschuh, M., et al., *Multiparametric high-resolution imaging of native proteins by force-distance curve-based AFM*. Nat. Protoc., 2014. **9**(5): p. 1113.
13. Lopez, A.E., et al., *Surface dependence of protein nanocrystal formation*. Small, 2010. **6**(3): p. 396-403.
14. Durbin, S.D., W.E. Carlson, and M. Saros, *In situ studies of protein crystal growth by atomic force microscopy*. J. Phys. D: Appl. Phys. , 1993. **26**(8B): p. B128.
15. Keel, T., *The application of in situ AFM to the study of molecular and macromolecular crystallization*. 2004, University of Nottingham.
16. Malkin, A., Y.G. Kuznetsov, and A. McPherson, *An in situ AFM investigation of catalase crystallization*. Surf. Sci., 1997. **393**(1-3): p. 95-107.
17. Caballero, D., et al., *Direct patterning of anti-human serum albumin antibodies on aldehyde-terminated silicon nitride surfaces for HSA protein detection*. Small, 2009. **5**(13): p. 1531-4.
18. Liang, C., et al., *Self-Assembled Nanofibers for Strong Underwater Adhesion: The Trick of Barnacles*. ACS Appl. Mater. Interfaces, 2018. **10**(30): p. 25017-25025.
19. Lee, H., N.F. Scherer, and P.B. Messersmith, *Single-molecule mechanics of mussel adhesion*. Proc. Natl. Acad. Sci. , 2006. **103**(35): p. 12999-3003.

20. Zhang, Y., et al., *Nanopore Formation via Tip-Controlled Local Breakdown Using an Atomic Force Microscope*. *Small Methods*, 2019: p. 1900147.
21. Goldsbury, C.S., S. Scheuring, and L. Kreplak, *Introduction to atomic force microscopy (AFM) in biology*. 2009. **58**(1): p. 17.7. 1-17.7. 19.
22. Pyne, A., et al., *Single-molecule reconstruction of oligonucleotide secondary structure by atomic force microscopy*. *Small*, 2014. **10**(16): p. 3257-61.
23. Li, H., et al., *Determining the molecular-packing arrangements on protein crystal faces by atomic force microscopy*. *Acta Crystallogr. D Biol. Crystallogr.* , 1999. **55**(Pt 5): p. 1023-35.
24. Yamashita, H., et al., *Dynamics of bacteriorhodopsin 2D crystal observed by high-speed atomic force microscopy*. *J. Struct. Biol.*, 2009. **167**(2): p. 153-8.
25. McPherson, A., A. Malkin, and Y. G, *Atomic force microscopy in the study of macromolecular crystal growth*. *Annu. Rev. Biophys.*, 2000. **29**(1): p. 361-410.
26. Ilari, A. and C. Savino, *A practical approach to protein crystallography*, in *Bioinformatics*. 2017, Springer. p. 47-78.
27. Wlodawer, A., et al., *Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination*. *FEBS J.*, 2013. **280**(22): p. 5705-5736.
28. Wlodawer, A., et al., *Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures*. *FEBS J.*, 2008. **275**(1): p. 1-21.
29. Bergkvist, M. and N.C. Cady, *Chemical functionalization and bioconjugation strategies for atomic force microscope cantilevers*, in *Bioconjugation Protocols*. 2011, Springer. p. 381-400.
30. Anselmetti, D., et al., *Single molecule DNA biophysics with atomic force microscopy*. *Single Mol.*, 2000. **1**(1): p. 53-58.
31. Lin, L., et al., *Recognition imaging with a DNA aptamer*. *Biophys. J.*, 2006. **90**(11): p. 4236-4238.
32. Hedayati, M. and M.J. Kipper, *Atomic force microscopy of adsorbed proteoglycan mimetic nanoparticles: Toward new glycocalyx-mimetic model surfaces*. *Carbohydr. Polym.* , 2018. **190**: p. 346-355.
33. Killgore, J.P., R.H. Geiss, and D.C. Hurley, *Continuous measurement of atomic force microscope tip wear by contact resonance force microscopy*. *Small*, 2011. **7**(8): p. 1018-1022.
34. Crist, B.V. and D.B. Crisst, *Handbook of monochromatic XPS spectra*. Vol. 1. 2000: Wiley New York.
35. Crist, B.V., *A review of XPS data-banks*. *XPS Reports*, 2007. **1**(1).
36. Iozzi, M.F., T. Helgaker, and E. Uggerud, *Influence of external force on properties and reactivity of disulfide bonds*. *J. Phys. Chem. A*, 2011. **115**(11): p. 2308-2315.

## CHAPTER 4: DISCOVERING THE INTERACTION PATTERNS OF DNA WITH PROTEIN CRYSTALS BY MACHINE LEARNING \*

### Chapter 4 Overview

During a previous study, atomic force microscopy (AFM) tips were used to covalently capture DNA molecules and probe the interaction between the DNA-laden tip and a nanopore array defined by a porous protein crystal. The resulting data set includes hundreds of thousands of detailed records of six pools of force-vs-distance interaction trace data, corresponding to 2 different nano topographical locations and 3 different scenarios for the tip and crystal preparation. The current study demonstrates that supervised learning with a random forest algorithm and cross-validation is sufficient to classify unlabeled AFM interaction data. Specifically, six binary classifiers achieve 92.2% to 98.8% accuracy, while a 6-category classifier achieves 84.8% accuracy. The most important features for classification were Euclidean distance between the data points on the force-distance curves. In one interaction class of particular interest, AFM tips that captured DNA within the crystal are centered above the 13-nm diameter crystal nanopores. Interaction traces from this class generally had a long interaction distance, a large adhesion energy, and significant variation in the force-vs-distance interaction trace curves. To further analyze this subset of the data we used unsupervised machine learning. Specifically, we used the affinity propagation algorithm to cluster these data into 12 representative interaction patterns. Extend into the center of the 13-nm diameter nanopores. These exemplar interaction types were used to generate hypotheses about the possible variation in the molecular interaction details.

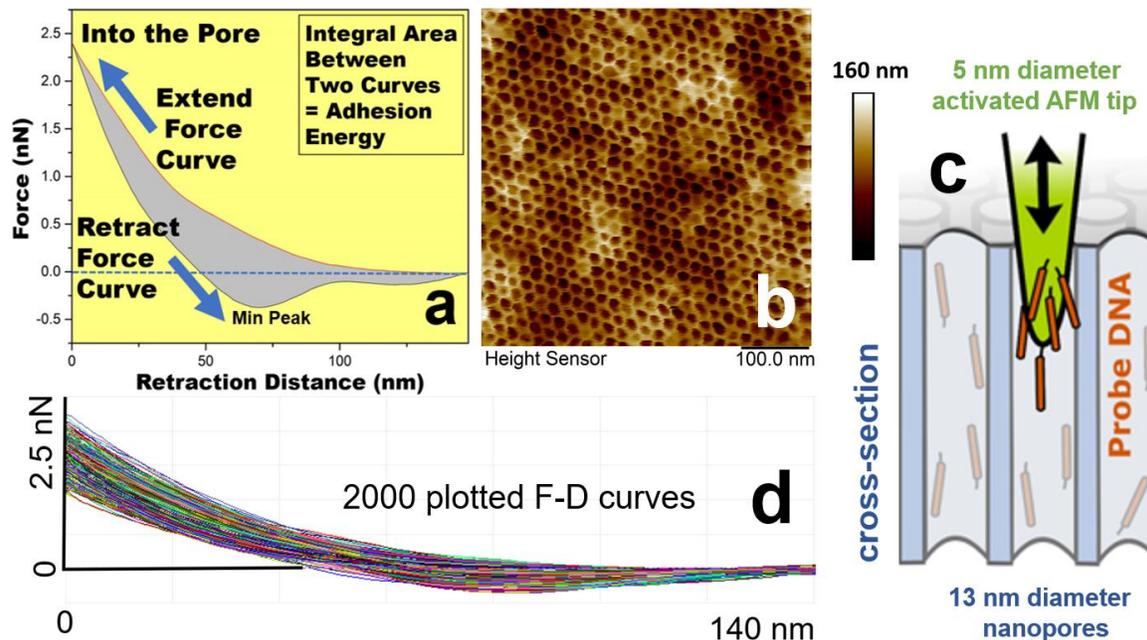
---

\* Portions of this chapter are reproduced from: Dafu Wang, Shun Yao Wu, Julius D. Stuart, Sawyer F. Halingstad, Chris D. Snow, Matt J. Kipper, “Discovering the Interaction Patterns of DNA with Nanoporous Protein Crystals by Machine Learning”

## 4.1 Introduction

Highly porous protein crystals are a biomaterial that has recently been engineered to encapsulate, stabilize, and organize guest molecules, nanoparticles, and biological moieties in highly ordered molecular arrays.<sup>[1-3]</sup> We are interested in the crystals of a putative isoprenoid binding protein from *Campylobacter jejuni* (Genebank ID: CJ0420, Protein Data Bank (PDB) code: 5w17, which we refer to as CJ). This protein readily and rapidly assembles into highly porous protein crystals.<sup>[1-4]</sup> CJ crystals are hexagonal three-dimensional arrays with 13 nm-diameter nanopores (18 nm pore-center-to-center) that are aligned along the  $z$ -axis. The uncommonly large pores of CJ crystals provide ample space for double-stranded DNA (dsDNA, hereinafter as DNA), with a diameter of 2 nm, to be loaded and stored within each 13 nm-diameter nanopore. Previous studies have shown the capability of large-pore protein crystals to capture guest nanoparticles and proteins.<sup>[1,2]</sup> During our published previous study, the interactions of DNA with this protein crystal were measured using force-distance (F-D) curves obtained using atomic force microscope (AFM) tips modified to present or capture DNA. The chemical modification of AFM tips allowed the tips to display covalently bound DNA, or to bind DNA that was pre-loaded in the protein crystal nanopores. Optimized AFM characterization quantitatively measured and analyzed the mechanical behaviors, and the nano-scale variations in the adhesion energy between DNA and protein crystals. The hexagonal space group was manifest at the crystal surface, as were the strong interactions between DNA and the porous protein crystals in question. Our previous study quantitatively measured the total and normalized adhesion energy between DNA and CJ protein crystal, furthering our understanding of how a new protein-based biomaterial can be used to bind guest molecules with multivalent DNA display.

During the previous study, we operated the AFM in quantitative nano-mechanics (QNM) PeakForce Capture mode. In this mode, one imaging scan on a 500 nm × 500 nm area of protein crystal's surface, included up to 262,144 sets of F-D curves. Each F-D curve set contained 128 extension distances, 128 force values at each extension distance, 128 retraction distances, and 128 force values at each retraction distance. As a result, we have collected approximately 30 GB of raw AFM data directly related to the interaction patterns of DNA-conjugated AFM tips (DNA-laden AFM tips, or DNA-tips) with protein crystals. Meanwhile, the interaction pattern themselves were complicated. Each F-D curve set embodies physical features including extension distance, retraction distance, force during extension, force during retraction, the derivative of the force, number and position(s) of minima on each retract force curve, and total adhesion energy (obtained as the integral area between the extend and retract force curves). Certain features were not easily observable by eye. Therefore, interaction patterns of DNA-laden tips with protein crystals were not easily classifiable by eye or any existing model. Furthermore, measuring and studying the interaction between DNA-laden AFM tips and protein crystals is an unusual measurement lacking directly comparable literature precedent. Therefore, there could be unknown features hidden within the interaction patterns. Processing the large and complicated raw data set in a timely way therefore requires the development of a suitable automated method. In this study, we demonstrate that supervised machine learning can be an effective method for automatic selection and classification of these data. Meanwhile, unsupervised machine learning allowed unbiased analysis of the raw data, extracting hidden representative clusters to facilitate analysis of subtle variations in the interaction patterns.



**Fig. 4.1** (a) Representative examples of F-D curves obtained using a modified AFM tip on a nanopore center of crystal loaded with DNA. (b) An AFM image of a crystal loaded with DNA imaged using an activated AFM tip. (c) Within a periplasmic protein, “CJ”, from *Camphylobacter jejuni* forms, a hexagonal array of 13 nm-diameter nanopores runs from the top to the bottom of each crystal. Modified DNA molecules can be loaded into these nanopores and then pulled out using activated AFM probes capable of covalently bonding to the DNA. The top face of the crystal may then be probed using chemical force microscopy with DNA-modified AFM tips.<sup>[1-4]</sup> (d) An overlay of 2000 curves plotted F-D curves from DNA loaded CJ protein crystal, collected by an activated AFM tip.

Machine learning has been playing an increasingly important role to enhance image recognition, substance analysis, and autonomous operation of AFM, as a strategy to reduce the human reliance of observation and analysis.<sup>[5-7]</sup> Previously published studies reported that supervised learning has been used to enable an automated and reproducible analysis pipeline for biological samples in AFM,<sup>[6]</sup> and has also been used for the analysis of nonlinear responses to the bimodal drive at harmonics and mixing frequencies, thus to obtain quantitative material properties at high speeds and with enhanced resolution.<sup>[7]</sup> In the meantime, in addition to AFM studies, machine learning has been broadly useful in the study of interfacial biomolecules interactions. In particular, the random forest algorithm for supervised machine learning has been applied to predict

the similarity of protein-protein interaction,<sup>[8]</sup> and predict the protein-protein interaction sites.<sup>[9]</sup> The affinity propagation algorithm for unsupervised machine learning has been applied to detect protein complexes in the protein-protein interaction networks.<sup>[10]</sup> The affinity propagation has also been applied to the area in clustering the protein interaction graphs.<sup>[11]</sup> In addition to random forest and affinity propagation algorithms, other machine learning algorithms can also predict a potential energy landscape for the interatomic interactions of molecules,<sup>[12, 13]</sup> cluster the protein-protein interaction networks,<sup>[14]</sup> predict the differentiation of a cell based on the environment.<sup>[15, 16]</sup> In sum, machine learning can be a powerful tool to achieve our goals to automatically classify and cluster the interaction patterns between DNA-laden AFM tips and protein crystals.

**Table 4.1 Scenarios of Interaction Patterns for Supervised Machine Learning**

	<b>activated Tip</b>	<b>DNA Loaded</b>	<b>pore / wall</b>
<b>a</b>	+	+	pore
<b>b</b>	+	+	wall
<b>c</b>	+	-	pore
<b>d</b>	+	-	wall
<b>e</b>	-	-	pore
<b>f</b>	-	-	wall

During our previous study, we found through quantitative AFM measurement that the status of interaction between DNA and protein crystals was closely related to its location (such as pore or wall area), activation of AFM tips, and the scenario used for DNA conjugation to the tip.<sup>[17]</sup> Here we applied the random forest classifier for supervised machine learning.<sup>[18]</sup> Due to its computational efficiency and excellent performance, so far, the random forest classifier has been widely applied into many fields, including bioinformatics.<sup>[18, 19]</sup> Using physical and geometrical features of F-D curve sets, we were able to build training models based on individually labeled sample curves for 6 different scenarios (Table 4.1). Respectively, including the pores and walls of “inactivated AFM tip with unloaded protein crystal”, “activated AFM tip with unloaded protein

crystal”, and “activated AFM tip with protein crystal loaded with DNA”. After training, we were able to sort an unlabeled test set of F-D curves into 6-categories, using an automatic classification with the random forest classifier.

Among all 6 scenarios above (Table 4.1), the interaction between DNA-laden AFM tips and protein crystals was strongest for scenario *a* when the AFM tip is probing the nanopore center, and that AFM tip was activated and captured DNA that was loaded into the crystal. To analyze the distinctions between different types of interaction patterns, we collected and labeled 5331 F-D curve sets, where all sample F-D curves were taken from the pore area of DNA loaded crystals. Specifically, 5331 pixel locations from 5331 nanopore centers were randomly selected among tens of thousands of nanopore centers. Using the affinity propagation algorithm (unsupervised machine learning), we clustered the sample curves into 12 groups, each with an exemplar F-D curve set. Each exemplar represents one type of featured interaction patterns between DNA-laden AFM tips and protein crystals.

In summary, we have built a supervised random forest classifier that can accurately predict the scenario that generated an unlabeled F-D curve set, and select the qualified samples. Analysis of the trained model revealed that the most important distinguishing features were the distribution of the total adhesion energy, and the Euclidean characteristics of the F-D curve sets. We also applied unsupervised machine learning (affinity propagation) to identify 12 sub-classes for the most interesting scenario where DNA-laden tips are deeply probing the crystal nanopores. This work provides a basis for engineering functional biomaterials that are responsive to small pulling forces. In principle, such materials could be used for signal transduction, wherein biomolecular recognition and mechanical pulling events induce downstream effects by increasing the accessibility of signaling sequences.

## **4.2 Methodology and Experiments**

### **4.2.1 Materials and AFM Probing**

Descriptions of experimental materials (**B1.1**), CJ crystal growth (**B1.1.1**), CJ crystal crosslinking and immobilization (**B1.1.2**), AFM tip modification (**B1.2**), DNA loading and AFM probing (**B1.3**) have been attached to this manuscript as the Supporting Information. Analysis of the AFM data was performed in NanoScope (Bruker, Inc.), Origin (OriginLab, Inc.), Python (Version 3.6), and Matlab (Version 2019a).

### **4.2.2 Automatic Classification with the Random Forest Classifier**

In this section, the random forest algorithm was employed to automatically classify F-D curves collected by AFM. The random forest algorithm was a parallel ensemble classifier, which constructed a multitude of decision trees through random selection of samples and features in training data and combined the predicted status of new samples by voting.<sup>[18]</sup>

We considered F-D curve set data from 6 different scenarios, respectively the pores and walls of “inactivated AFM tip with unloaded protein crystal”, “activated AFM tip with unloaded protein crystal”, and “activated AFM tip with protein crystal loaded with DNA” (Table 4.1), and adopt two strategies to automatically distinguish the six types of F-D curves. One strategy is to construct one binary classifier for each scenario. Six binary classifiers were built by random forest to identify the six kinds of F-D curves. Take the pores of “inactivated AFM tip with unloaded protein crystal” as an example, the corresponding binary classifier was constructed by taking the pores of “inactivated AFM tip with unloaded protein crystal” as positive samples and the remaining F-D curves as negative samples, which can determine whether an unknown F-D curve is a pore of “inactivated AFM tip with unloaded protein crystal”. Another strategy is to directly

construct a multiclass classifier for the six scenarios. All the F-D curves from the six scenarios were used to train a unified classifier with six categories corresponding to the six scenarios.

To construct the random forest classifiers, we extracted 17 physical features from each raw F-D interaction curve set (extension and retraction). The features of each F-D curve set included the mean, median, minimum, and maximum values of multi Euclidean distance factors for extension distance, retraction distance, extension force, and retraction force. Additionally, the total adhesion energy was estimated for each F-D curve set. One crucial parameter of random forest is “*number of trees*”, which controls the number of trees for ensembles and should be generally set as a large positive integer. During our computational experiments, “*number of trees*” was set as 100 for each random forest classifier, which means 100 decision trees was built to construct a classifier.

To evaluate the classifier performance results, accuracy, receiver operating characteristic (ROC) curve, area under the ROC curve (AUC), F1 score, precision, and true positive rate (TPR) were applied to evaluate the effect of the binary classification, while accuracy, Kappa Coefficient, Hamming Distance, and Jaccard Index were applied to evaluate the results of the 6-category classification. In this case, the 5-fold cross-validation was used to generate a “mean value  $\pm$  standard deviation” value for all the index in order to quantify the performance. In the experiments, 5-fold cross-validation was performed for the two strategies. In each fold, 80% of samples were randomly selected as the training set for model construction while the remaining 20% were the testing set for validation. The average performance results were finally recorded for each strategy.

#### **4.2.3 Pattern Discovery with Affinity Propagation**

To discover the sub-types of interaction patterns hidden within thousands of F-D curve sets, the affinity propagation (AP) algorithm was applied. Affinity propagation is a clustering algorithm

that partitions samples into several clusters on the basis of a pre-calculated matrix of dissimilarity values.<sup>[20]</sup> Clustering with any guidance, or prediction, including “pairwise constraints”, belong to supervised or semi-supervised machine learning. In this work, clustering is completely unsupervised, therefore, the clustering process are not influenced by any human impact including the known labels of F-D curves. One favorable aspect of affinity propagation is that the method does not require a pre-defined number of clusters. Instead, a single threshold parameter, combined with the input dissimilarity matrix, determines the number of output clusters. Briefly, each F-D curve set begins as a potential cluster center. Then, a message-passing procedure is enacted which allows each data point to communicate with the other points about which points are representative examples of their neighbors. When this procedure converges, a limited number of “exemplar” data points represent similar data points within a cluster. The results will then be further evaluated by the cohesion and separation between clusters.

Among all the samples, every F-D curve that we collected by AFM contained four basic elements, which included the extended distance, extend force, the retract distance, and the retract force. Furthermore, within the raw data, each sample corresponds to a file, each file contains four columns of data, and each column of data contains 128 values. Thus, each raw data file of one F-D curve set could be treated as one  $4 \times 128$  matrix. Treating this array as a  $1 \times 512$  array, each F-D curve set is considered one point in 512-dimensional space. The dissimilarity between all pairs of data points was computed as the Euclidean distance between each pair of points in 512-dimensional space. Each F-D curve is treated as a 512-dimensional vector. There are a total of 5331 sets of F-D curves, therefore the distance matrix obtained is  $5331 \times 5331$  dimensions. In this work, all calculations are based on Euclidean distance (affinity: {‘euclidean’, ‘precomputed’}, default = ‘euclidean’). The damping factor is in the range of [0.5, 1.0) (damping: float, default =

0.5), the maximum number of iterations is set as 200 (max\_iter: int, default = 200), Number of iterations with no change in the number of estimated clusters that stops the convergence is set as 15 (convergence\_iter: int, default = 15). It should be pointed out that all clustering algorithms have their limitations in theory. In most cases, we can run the algorithms directly to see whether the sample size of each cluster in the clustering result is similar, whether the compactness within the cluster and the isolation between clusters are within the acceptable ranges. However, it is more complicated to check whether the data is symbolic and spherical, especially in the case of high dimensionality ( $n > 3$ ), therefore, this paper will not discuss it. Using the AP algorithm, we obtained 12 cluster centers from 5331 F-D curves that measured within the pore area of a protein crystal that was loaded with DNA, by activated AFM tip. The AP algorithm in this work simultaneously considered all aspects of the raw data points as equally valuable and exchanged information between each pair of data points until a converged solution emerged.<sup>[20, 21]</sup>

In this work, the responsibility of a sample  $k$  to be the exemplar of sample  $i$ ,  $r(i, k)$ , is given by:

$$r(i, k) \leftarrow s(i, k) - \max[a(i, k') + s(i, k') \forall k' \neq k]$$

Where  $a$  is the availability accumulated evidence that sample  $k$  should choose sample  $i$  to be its exemplar, and considers the values for all other samples that the sample  $k$  should be an exemplar.<sup>(Open coding source 2)</sup> In the meantime, the similarity between the sample  $i$  and  $k$ ,  $s(i, k)$  is given by:

$$a(i, k) \leftarrow \min[0, r(k, k) + \sum_{i' \text{ s.t. } i' \notin \{i, k\}} r(i', k)]$$

$$r_{t+1}(i, k) = \lambda \cdot r_t(i, k) + (1 - \lambda) \cdot r_{t+1}(i, k)$$

$$a_{t+1}(i, k) = \lambda \cdot a_t(i, k) + (1 - \lambda) \cdot a_{t+1}(i, k)$$

Where  $t$  represents the iteration times. To begin with, values for  $r$  and  $a$  are set to 0, and the calculation of each iterates until convergence.  $\lambda$  represents the damping factor. <sup>(Open coding source 2)</sup>

It should be noticed that clustering uses objective indicators to find differences in patterns, and different parameters represent different angles of observation. Changing the parameters like "preference" and "damping" potentially leads to different results than the current results. Put the same types of things there, some people can see the difference in color, while some people can see the difference in shape, which is similar to clustering under different parameters. It is our purpose to find the difference from a certain angle to the interaction patterns between DNA molecules and protein crystals, instead of observing from lots of different angles. In this work, our input is a 512-dimensional vector, and the features of the interaction patterns correspond to some certain data points in the 4 columns of data. If we dig deeper into the principal component analysis (PCA) results, study PC1 and PC2 after dimensionality reduction in detail, we will therefore have to find the correspondence between PC1, PC2 and the 512-dimensional vector. However, this will lead to a very messy result. For example, PC-X can relate to multiple points from all four columns. It is not the initial purpose of this work, thus, will not be discussed in this paper.

In this work, the cohesion and separation were adopted to evaluate clustering performance. Cohesion measures how closely related were objects in a cluster, while separation measures how distinct a cluster was from other clusters. The cohesion of the  $i$ -th cluster was defined as follows.

$$Cohesion(i) = \frac{1}{n_i} \sum_{a=1}^{n_i} D(x_a, \mu_i)$$

Here,  $n_i$  was the number of cluster members in the  $i$ -th cluster,  $\mu_i$  was the cluster center of the  $i$ -th cluster and  $D(x_a, \mu_i)$  was the squared Euclidean distance between  $x_a$  and  $\mu_i$ . The separation was defined as follows.

$$Separation(i, j) = \frac{1}{n_i n_j} \sum_{a=1}^{n_i} \sum_{b=1}^{n_j} D(x_a, x_b)$$

## 4.3 Results and Discussion

### 4.3.1 Supervised Machine Learning

#### 4.3.1.1 Binary Classification

The effectiveness of the binary random forest classifier to recognize the F-D curves in this work was evaluated and confirmed by multiple metrics including accuracy, receiver operating characteristic (ROC) curve (Fig. 4.1 (a-f)), area under the ROC curve (AUC), F1 score, positive predictive value (PPV), and true positive rate (TPR) (Table 4.2) respectively.

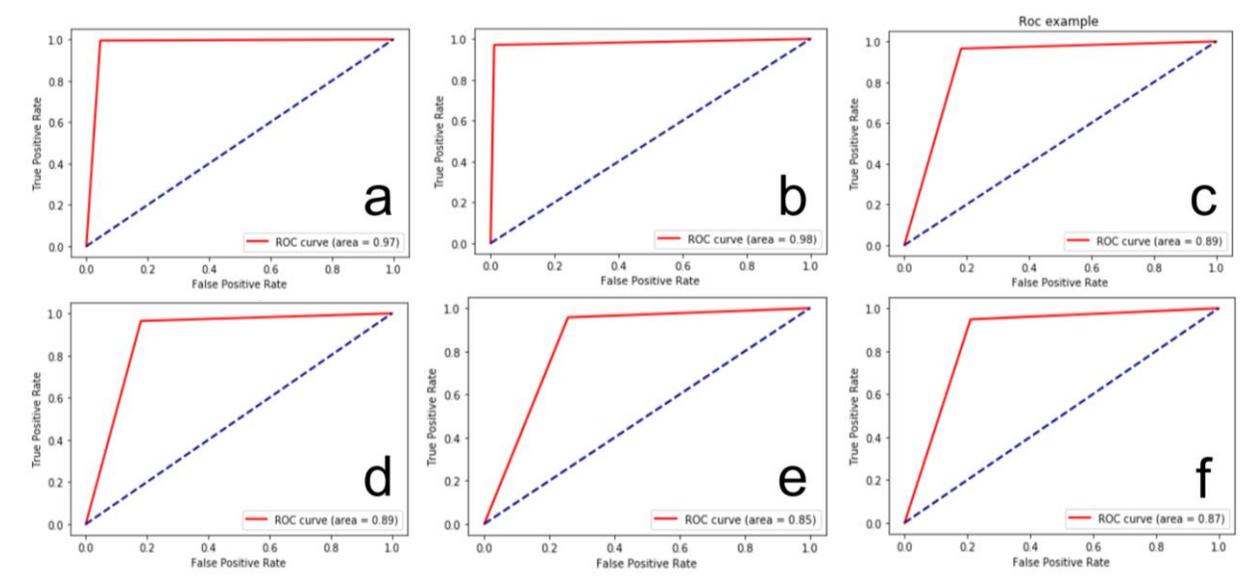
**Table 4.2 Evaluations of the binary random forest classifier**

	<b>Accuracy</b>	<b>PPV</b>	<b>TPR</b>	<b>F1 Score</b>	<b>AUC</b>
<b>a</b>	0.988	0.991	0.994	0.993	0.974
<b>b</b>	0.986	0.946	0.970	0.958	0.980
<b>c</b>	0.941	0.964	0.965	0.964	0.892
<b>d</b>	0.940	0.964	0.964	0.964	0.892
<b>e</b>	0.922	0.949	0.958	0.953	0.850
<b>f</b>	0.922	0.957	0.949	0.953	0.869

(label a – f in Table 4.2 are corresponding to the labels a – f in Table 4.1 and Fig. 4.2)

The ROC curves showed a high performance of classification at various threshold cases. The metrics indicated that the binary random forest classifier has an accuracy between 92.2% to 98.8% among all 6 cases, which showed their very closeness to the standard known value. Among them, the classifier had the highest discrimination ability for F-D curves collected by activated AFM Tip from crystals loaded with DNA, who achieved 98.8% and 98.6% for pore and wall areas respectively, they also had the largest AUC for 0.974 and 0.980 respectively, which were very close to 1, indicated that in those cases, the model was capable of well distinguishing between classes. For other cases, the model was capable as well. The average AUC was 0.909 among all 6

cases. The TPR among all 6 cases were between 94.9% to 99.4%. In this study, the PPV were between 94.6% to 99.1%. It was very close to their prediction to the true value. For the binary classification, the F1 score was a combining metric of the tests' accuracy, which was the harmonic mean of the precision and recall, where an F1 score reaches its best value at 1. Among our cases, the F1 score reached a maximum value of 0.980, and a mean value of 0.964. These metrics have confirmed that our binary random forest classifier was capable of accurately classifying and distinguishing random F-D curves from all cases.



**Fig. 4.2** The ROC curves and space of the binary random forest classifier where F-D curves collected by AFM from (a) pore area of activated AFM Tip with crystals loaded with DNA was positive while all else were negative, (b) wall area of activated AFM Tip with crystals loaded with DNA was positive while all else were negative, (c) pore area of activated AFM Tip with unloaded crystals was positive while all else were negative, (d) wall area of activated AFM Tip with unloaded crystals was positive while all else were negative, (e) pore area of inactivated AFM Tip with unloaded crystals was positive while all else were negative, and (f) wall area of inactivated AFM Tip with unloaded crystals was positive while all else were negative, scenarios a-f in Fig. 4.2 are corresponding to Table 4.1.

#### 4.3.1.2 Multi-category Classification

The metrics for the 5-fold cross-validation included the accuracy, Kappa coefficient, Hamming distance, and Jaccard index. The results indicated that the 6-category classification also

had a high accuracy of recognition, which had an average accuracy of 84.9% based on 5 test running sets. In the meantime, Kappa coefficient was used to verify and assess qualitative documents and determine agreement between different raters, including the observed agreement among the raters, and the hypothetical probability of the raters indicating a chance agreement.<sup>[22]</sup> The strength of the Kappa coefficients between 0.8 to 1.0 can be interpreted as almost perfect.<sup>[23]</sup> In our work, the Kappa coefficient in all 5 test sets were above 0.8, with an average of 0.818. The Hamming distance was for measuring the edit distance between two sequences, as well as the minimum number of substitutions required to change one string into the other. In this work, the Hamming distance with an average of 0.152 showed the similarity among the 5 test running sets. The Jaccard coefficient with an average of 0.737 also showed the similarity among all test running sets, and proved the repeatability of the test running.

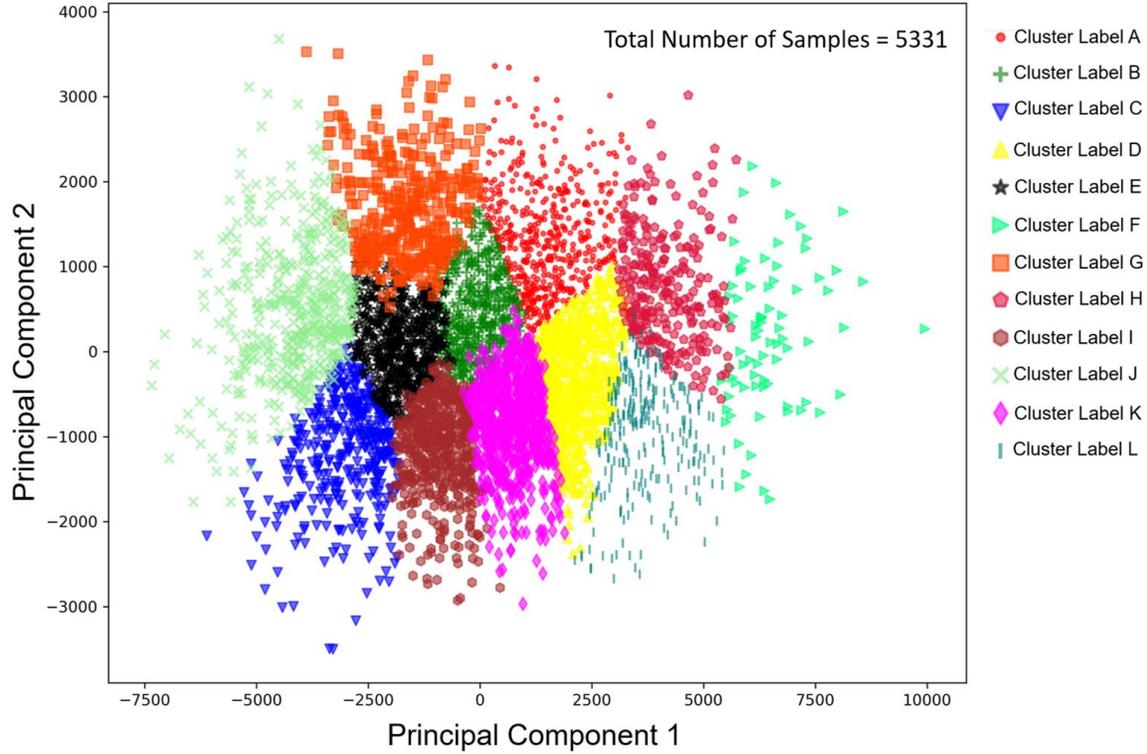
The results of supervised machine learning confirmed a capability of both the binary classifier and the 6-category classification to recognize a random unknown F-D curve, found out the location where the interaction happened, and sorted it into a known type of interaction pattern between DNA and protein crystals. More importantly, it provided us an opportunity that enabled us to further understand and illustrate the meaning behind the random unknown F-D curves.

**Table 4.3 Evaluations of 5-fold cross-validation**

	<b>Accuracy</b>	<b>Kappa Coefficient</b>	<b>Hamming Distance</b>	<b>Jaccard Index</b>
<b>Test Set 1</b>	0.860	0.832	0.140	0.754
<b>Test Set 2</b>	0.848	0.818	0.152	0.737
<b>Test Set 3</b>	0.835	0.802	0.165	0.717
<b>Test Set 4</b>	0.848	0.818	0.152	0.737
<b>Test Set 5</b>	0.850	0.820	0.150	0.739
<b>Average</b>	0.848	0.818	0.152	0.737

## 4.3.2 Unsupervised Machine Learning

### 4.3.2.1 Cluster of F-D Curves



**Fig. 4.3** Principal component analysis (PCA) for the cluster results on F-D curves from the pore area of activated AFM tip with protein crystal loaded with DNA

**Table 4.4** Cohesion and separation of clusters

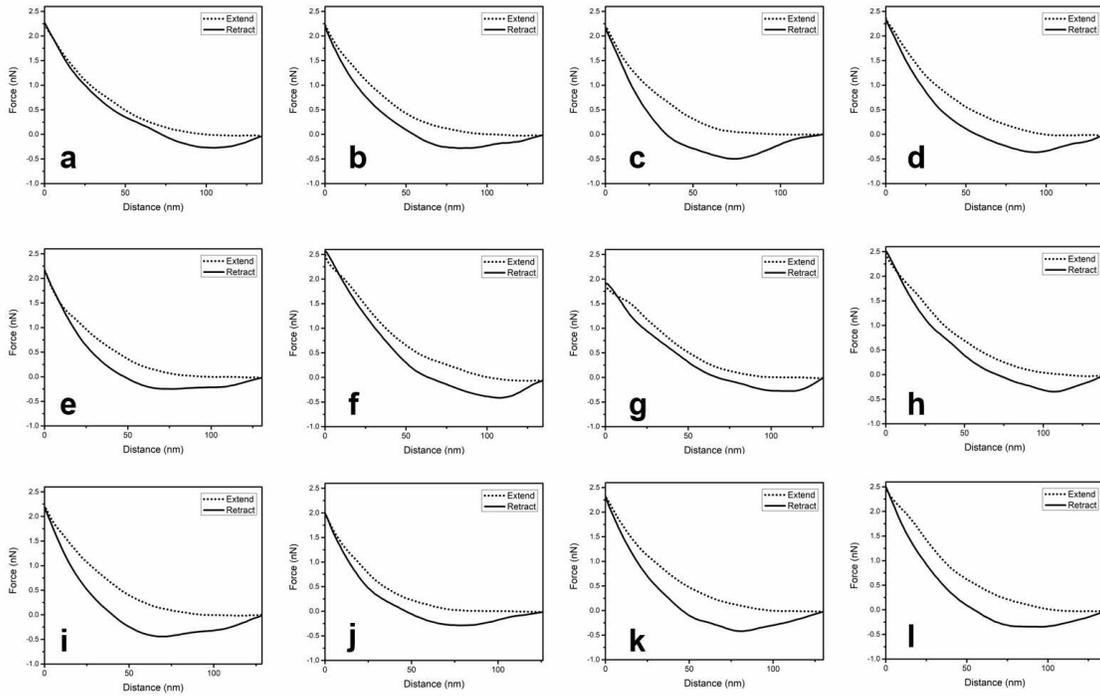
Cluster Name	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	Cluster10	cluster11	cluster12
cluster1	1777.59	2365.50	5330.44	2410.62	3766.92	5455.95	3492.93	3195.32	3743.18	5896.81	2759.18	3603.62
cluster2	2365.50	1503.81	3641.78	2777.07	2254.64	6781.28	2539.94	4501.02	2272.79	4370.61	2083.07	4294.03
cluster3	5330.44	3641.78	1699.26	5419.98	2316.89	9720.24	3575.32	7556.24	2548.80	2682.78	4001.85	6850.40
cluster4	2410.62	2777.07	5419.98	1584.20	4213.33	4694.18	4559.74	2773.61	3484.13	6552.93	2122.22	2345.38
cluster5	3766.92	2254.64	2316.89	4213.33	1479.95	8434.50	2257.42	6148.37	2082.33	2875.58	3021.70	5768.78
cluster6	5455.95	6781.28	9720.24	4694.18	8434.50	2095.30	8442.96	3023.67	7699.33	10790.29	6088.20	3540.66
cluster7	3492.93	2539.94	3575.32	4559.74	2257.42	8442.96	1795.21	6061.24	3305.72	3270.51	3779.66	6150.14
cluster8	3195.32	4501.02	7556.24	2773.61	6148.37	3023.67	6061.24	1780.75	5617.43	8461.42	4081.45	2493.22
cluster9	3743.18	2272.79	2548.80	3484.13	2082.33	7699.33	3305.72	5617.43	1461.64	3952.25	2191.24	4842.26
cluster10	5896.81	4370.61	2682.78	6552.93	2875.58	10790.29	3270.51	8461.42	3952.25	1977.76	5289.17	8129.25
cluster11	2759.18	2083.07	4001.85	2122.22	3021.70	6088.20	3779.66	4081.45	2191.24	5289.17	1485.83	3345.84
cluster12	3603.62	4294.03	6850.40	2345.38	5768.78	3540.66	6150.14	2493.22	4842.26	8129.25	3345.84	1857.01

**Table 4.5 Purity tests results of clustering for 10 times**

	Prediction 1	Prediction 2	Prediction 3	Prediction 4	Prediction 5	Prediction 6	Prediction 7	Prediction 8	Prediction 9	Prediction 10
Predictio n 1	1	1	1	1	1	1	1	1	1	1
Predictio n 2	0	1	1	1	1	1	1	1	1	1
Predictio n 3	0	0	1	1	1	1	1	1	1	1
Predictio n 4	0	0	0	1	1	1	1	1	1	1
Predictio n 5	0	0	0	0	1	1	1	1	1	1
Predictio n 6	0	0	0	0	0	1	1	1	1	1
Predictio n 7	0	0	0	0	0	0	1	1	1	1
Predictio n 8	0	0	0	0	0	0	0	1	1	1
Predictio n 9	0	0	0	0	0	0	0	0	1	1
Predictio n 10	0	0	0	0	0	0	0	0	0	1

By AP algorithm, a total number of 5331 sets of F-D curves collected from the pore area of DNA loaded crystals by activated AFM tips were clustered into 12 cluster centers (Fig. 4.3). And each cluster center was corresponding to one set of F-D curves. The cohesion of each cluster and separation between clusters were obtained as well (Table 4.4). As shown in Table 4.4, the average distance of cluster members in the same cluster was much smaller than that between different clusters. It validates the effectiveness of clustering results. This clustering model has independently run for 10 times as an examination of repeatability. According to the purity tests' results (Table 4.5), it turned out that running the algorithm 10 times came out with exactly the same cluster results for all 10 times. The results of purity tests confirmed the 100% repeatability of our cluster algorithm. All 12 clusters from the PCA plots were evenly distributed, with a proportion ranging from 1.46% to 13.07% (Table 4.6 and Fig. B1).

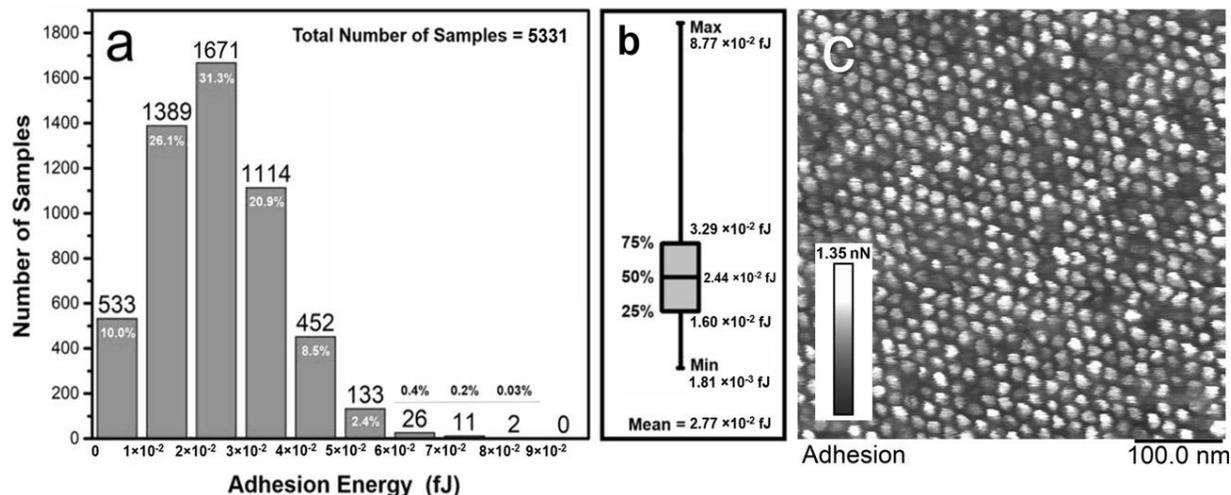
### 4.3.2.2 Interpretation of Interaction Patterns



**Fig. 4.4** (a-l) the F-D curves corresponding to the cluster centers' label "A" to "L"

**Table 4.6** Selected features of cluster centers

Cluster Center	Total Adhesion Energy (fJ)	Min Retract Force (nN)	Max Probing Force (nN)	Position of Minima (nm)	Interacting Distance (nm)	Number of Labels (sum = 5331)
A	$2.92 \times 10^{-2}$	-0.27	2.27	103.99	134.49	503 (9.44%)
B	$4.20 \times 10^{-2}$	-0.28	2.21	83.04	135.05	578 (10.84%)
C	$5.33 \times 10^{-2}$	-0.50	2.19	74.51	124.50	387 (7.26%)
D	$5.26 \times 10^{-2}$	-0.36	2.36	93.92	135.97	627 (11.76%)
E	$3.89 \times 10^{-2}$	-0.25	2.17	77.30 / 94.36 / 101.77	130.26	562 (10.54%)
F	$5.01 \times 10^{-2}$	-0.42	2.55	108.27	134.27	78 (1.46%)
G	$3.58 \times 10^{-2}$	-0.28	1.89	102.08	131.51	371 (6.96%)
H	$4.63 \times 10^{-2}$	-0.35	2.49	106.00	138.37	267 (5.01%)
I	$6.21 \times 10^{-2}$	-0.44	2.20	67.45	128.46	506 (9.49%)
J	$3.19 \times 10^{-2}$	-0.29	1.99	77.49	125.85	479 (8.99%)
K	$5.61 \times 10^{-2}$	-0.42	2.33	81.66	130.03	697 (13.07%)
L	$6.18 \times 10^{-2}$	-0.34	2.50	88.85 / 96.20	134.36	276 (5.18%)



**Fig. 4.5 (a)** Distribution of adhesion energy ranges, and **(b)** box plot (min, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles, and max) of the adhesion energy, for 5331 pore center pixels (from 5331 individual pores) with activated AFM tip, on the crystal loaded with DNA, and **(c)** a map of adhesion energy distribution of one protein crystal's surface, corresponding to the AFM height sensor imaging

12 sets of F-D curves corresponding to 12 cluster centers were shown in Fig. 4.4 (a-l). We made a hypothesis without exceeding reasonable logic and inferences, that each F-D curve at the cluster center represented a sub-type of interaction patterns between DNA and protein crystals. We then took further steps to analyze, discuss, and hypothesize the cause of each F-D curve, and what they represented, in order to further our understanding of the interaction between DNA and CJ protein crystals.

It should be noted that the F-D curves in Fig. 4.4 are exported from the Bruker's Nanoscope software which are automatically filtered before being exported separately. These F-D curves separately exported from Nanoscope software come with a background correction in order to remove the force curve baseline offset or tilt. In addition to the baseline correction, these F-D curves are filtered by the boxcar method to smooth the curves. These modifications are necessary for studying the details of the interaction patterns and would not significantly change the features from the interaction patterns.<sup>[24]</sup> In the meantime, to make sure that our results are unbiased, all the classify and cluster results by the random forest and AP algorithms are working at the raw data

of the F-D curves without any background correction or filter. For the record, the raw F-D curves of cluster centers without any background correction or filter corresponding to Fig. 4.4 have been attached to the supporting information as Fig. B2.

**Cluster Center A (Fig. 4.4a):** The interacting distance of this set of F-D curves was very long, and the location of retracting distance for the only minima was shallow compared with other cluster centers. Meanwhile, both the total adhesion energy and the minimum retract force was not very high. This presented us that there were some DNA molecules that loaded shallowly inside the nanopore of the protein crystal, and the amount of loaded DNA molecules into the nanopore was not too much either. Thus, after contacting the DNA close to the surface the nanopore, activated AFM tip pulled out the DNA molecules without costing much energy. There was another interesting phenomenon that the derivative of the retract force curve at around 65 nm had a sudden change and became much higher. This was most likely because the activated tip bound more DNA molecules around the shallow area of the nanopore, or probably different areas of the protein crystal's inner nanopore had different ability of interacting with DNA molecules. This sub-type of interaction patterns counted 9.44% among all sample F-D curves.

**Cluster Center B (Fig. 4.4b):** This sub-type of interaction pattern also contained a long interacting distance. However, the total adhesion was 40% higher than the sub-type A interaction pattern. This means the activated AFM tip might bind more DNA molecules than the sub-type A interaction pattern. The stable derivative of the smooth retract force curve between 0-105 nm showed that the interactions between DNA molecules and protein crystal were strong and stable within this distance. However, the derivative of the retract force curve became lower after 105 nm. This might indicate the loss of some short distance interactions between DNA molecules and the protein

crystal soon after the molecules are pulled out from the nanopore. About 10.48% of F-D curves among all samples belonged to this sub-type of interaction patterns.

**Cluster Center C (Fig. 4.4c):** The high adhesion energy since the very beginning of retracing movement indicated that there were lots of DNA molecules loaded deep inside the nanopore of the protein crystal. A large amount of loaded DNA molecules contacted and bound to the activated AFM tip at the very beginning, and then were pulled out all the way above the surface of the protein crystal until these interactions were gradually broken away. We have noticed this F-D curve had the largest retract force among all the cluster centers, which achieved 0.50 nN. This also confirmed that a large amount of loaded DNA molecules were bound to DNA-laden activated AFM tip as another point of evidence. In this case, a large amount of bound DNA molecules might have a premature and immediate separation of the interactions when the DNA molecules were lifted above the nanopore upon the surface of crystal; this might be the cause that this F-D curve showed the shortest interacting distance among all the cluster centers. At the end of the interaction, the intensity of interaction has become very small. The pattern also showed a decrease of slope at around 40 nm of the interacting distance, which indicated a decrease of normalized adhesion energy around that distance. This could be caused by the separation of DNA molecules away from the protein crystal. Another hypothetical reason for this can be attributed to some certain location within the nanopore of the protein crystal that didn't have a strong interaction with DNA molecules like others. This sub-type of interaction patterns counted 7.26% among all sample F-D curves.

**Cluster Center D (Fig. 4.4d):** This sub-type of interaction pattern showed us a DNA-laden activated AFM tip that had a strong and stable interaction with loaded DNA molecules from deep inside the nanopore until very high above the surface. From a very high total adhesion energy, we can learn that the DNA-laden activated AFM tips contain a large amount of DNA molecules.

However, unlike Cluster Center C, the interaction is maintained for a much longer distance and remains very stable until around 90 to 100 nm distance. According to the derivative of the retract curve, the interaction between DNA molecules bound to the activated AFM tip and protein crystal experienced two major separations after the DNA-laden tip was pulled up from the nanopore. This sub-type of interaction patterns counted 11.76% among all sample F-D curves, which was the second highest distribution among all cluster centers.

**Cluster Center E (Fig. 4.4e):** According to the derivative, the most attractive and unique feature of this interaction pattern that differed it from other cluster centers was, it contained three minima on the retract force curve. One of the hypothetical reasons that the retract force curve contained multiple minima could be attributed to the complicated process of DNA molecules' separations and continuous attachment to the protein crystal. During the process of DNA molecules being pulled out of nanopore, initially the activated AFM tip bound some DNA molecules in the deep position within the nanopore, however, some DNA molecules separated away from the interaction during the pulling movement after being lifted up above the surface. In the meantime, some surface areas of the activated AFM tip located at a higher position of height have already separated from the nanopore, while other surface areas still remained inside the nanopore. Thus, some DNA molecules that are located at a shallow position of nanopore were caught by the activated AFM tip's surface that still remain within the nanopores and eventually bind to the activated AFM tip. The repetition of the above process has resulted in the "wave" shape of this kind of F-D curves. This sub-type of interaction patterns counted 10.54% among all sample F-D curves.

**Cluster Center F (Fig. 4.4f):** This set of F-D curves showed an extreme high maximum probing force on both the extend and retract force curve. The high probing force can be caused by the liquid resistance from the buffer liquid if the movement of the AFM tip was too fast. Under certain

circumstances, the modification of AFM tip made the shape of the probe minorly changed, even without changing the radius of the probe, different hydrodynamic layout of the probe section could lead to a very different liquid resistance when the AFM tip was probing, thus higher the maximum probing force. This was a very unusual situation during AFM imaging, but sometimes it happened. This sub-type of interaction patterns only counts 1.46% among all sample F-D curves, which was the lowest percentage among all cluster centers.

**Cluster Center G (Fig. 4.4g):** Exactly opposite to the Cluster Center F, this set of F-D curves had the lowest maximum probing force among all the cluster centers. Under the same force setting, this can be attributed to the streamlined shape of activated AFM tips and a comparable slower probing speed. Other than the probing force, this set of F-D curves was entangled for about 15 nm at a deep position inside the nanopore. This phenomenon may be caused by the unsuccessfully attachment of some DNA molecules to the protein crystals in the very beginning, and then the surface areas that located at a higher height position at the activated AFM tips that later probing into the nanopores began attaching the DNA molecules. Later, with the activated AFM pulled up, more and more loaded DNA molecules were successfully caught by the activated AFM tips and maintained a strong and stable interaction with the protein crystal for a long distance, but later separated within a very short period of time. This sub-type of interaction patterns only counts 6.96% among all sample F-D curves.

**Cluster Center H (Fig. 4.4h):** This sub-type of interaction patterns combined some of the features from both Cluster Center F and Cluster Center G, which contained the large maximum probing force and the entangled F-D curve on the beginning part of retracting movement. This sub-type of interaction patterns counted only 5.01% among all sample F-D curves, which was the second smallest percentage among all the cluster centers.

**Cluster Center I (Fig. 4.4i):** The Cluster Center I came with an obvious feature that it had the highest adhesion energy among all 12 cluster centers. The total adhesion energy of Cluster Center I was within the top 1% highest among all 5331 samples, and more than 2 times higher than the adhesion energy of Cluster Center A. This could mean that the DNA-laden AFM tip bound a huge amount of loaded DNA molecules, which correspondingly caused a strong interaction between protein crystal and the DNA molecules that bond to the activated AFM tip. In addition to the huge amount of loaded guest molecules, there was another probability that the normalized interaction itself was very strong. In the meantime, it also contained the shortest minima position on the retract force curve and the third shortest interacting distance among all cluster centers, showing the interaction was strong but reached to the top of its intensity quickly and then started separating away within a short retracting distance. This sub-type of interaction patterns counted 9.49% among all sample F-D curves.

**Cluster Center J (Fig. 4.4j):** The interaction pattern of Cluster Center J had a long and narrow shape and an inconspicuous position of minima. The comparable low total adhesion energy and maximum peak force showed a potential movement that the activated AFM tip bound a small amount of loaded DNA molecules and then smoothly pulled up, elastic but without a very strong interaction. As the ramp height increased, the connection between the DNA molecules and protein crystal was gradually broken away within a short distance. Make a materialized analogy, this process was similar to picking up a small amount of brushed cheese. This sub-type of interaction patterns counted 8.99% among all sample F-D curves.

**Cluster Center K (Fig. 4.4k):** The interaction patterns represented by Cluster Center K counted 13.07% among all the sample F-D curves. This was the highest ratio among all cluster centers. The interaction pattern of Cluster Center K showed a very typical process of pulling up an activated

AFM tip bound with loaded DNA molecules. During this process, within the nanopore, the loaded DNA molecules continuously bound to the DNA-laden tip while the tip was pulling up, in the meantime, some bound DNA molecules were separated away with the interaction between protein crystals in the midway. For the Cluster Center K, similar to the hypothesis and explanation for the Cluster Center C, between the interacting distance of 50 – 80 nm, there was an obvious decrease of the slope on the retract force curve, indicating that the normalized adhesion energy was decreased. It could be potentially attributed to the separation of some DNA molecules away from the interaction. It was also possible that some certain locations within the nanopore of protein crystal might not have a strong interaction with DNA molecules. This interaction pattern, combined with Cluster Center C, F, and G, showed a hypothetical probability that within one nanopore of protein crystal, the interaction was not unified, and not all the locations have the same or similar interacting ability, as well as the normalized adhesion energy between the DNA molecules.

**Cluster Center L (Fig. 4.4I):** According to the derivative, the interaction pattern of the Cluster Center L contained two minima at the retract force curve (the potential hypothetical explanations for multiple minima has been explained at the corresponding section of Cluster Center E). However, different from Cluster Center E, between the two minima, the retract curve was relatively horizontal, a stable slope was maintained for about 8 nm. This showed the status of the pulling process was sustained during that 8 nm, with little change of normalized adhesion energy. Then after 96.20 nm, the interaction between DNA molecules and the protein crystal was broken and gradually disappeared. This sub-type of interaction patterns counted 5.18% among all sample F-D curves.

It should be noted that all the retracting force curves of cluster centers contain a higher derivative when DNA-laden AFM tips are within the nanopores than lifting up and separating from the nanopores. This indicates that the major interaction between DNA molecules and protein crystals happens during the DNA-laden AFM tips within the nanopores.

#### **Chapter 4 Summary**

Throughout, machine learning methods were able to accurately classify thousands of random unknown F-D curves obtained from AFM and discover hidden patterns of F-D curves that are unable to classify or cluster by classic statistics. This discovery has a promising biomedical application that can be potentially applied to target and select the qualified loaded guest molecules that fit the physical requirements, such as PEG-RGD-stem cell complexes that are loaded into the protein crystals. The metrics showed that our multi-category classifiers have already achieved an average accuracy of 84.8% on overall recognition, and 92.2% to 98.8% for the binary classifier recognition on individual cases. We have also partitioned thousands of DNA-protein crystal interaction traces into 12 clusters. For the first time, through the AP algorithm, we are able to find out different subtypes of interaction patterns between guest DNA molecules and protein crystals. This study will further our understanding and application on protein-based nanomaterials capable of responding to selective biomolecular recognition. Given the nature of the AFM tip raster path, the tip returns to the same nanopores multiple times. Remarkably, the interaction type for each nanopore was stable between visits (i.e., the interaction label recurs with statistics above random), which suggests that the hidden molecular features that generate force-vs-distance interaction distances are stable on the ~30 min timescale.

## REFERENCES

Some of the affinity propagation (AP) algorithms for clustering the interaction patterns in this paper draws on the structures from the following open online coding sources:

### Open coding source 1: GitHub

([https://github.com/scikit-learn/scikit-learn/blob/844b4be24/sklearn/cluster/\\_affinity\\_propagation.py#L267](https://github.com/scikit-learn/scikit-learn/blob/844b4be24/sklearn/cluster/_affinity_propagation.py#L267))

### Open coding source 2: Scikit.Cluster

(<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.cluster>)

(<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AffinityPropagation.html>)

(<https://scikit-learn.org/stable/modules/clustering.html#affinity-propagation>)

1. Hartje, L.F., et al., *Adsorption-Coupled Diffusion of Gold Nanoclusters within a Large-Pore Protein Crystal Scaffold*. J. Phys. Chem. B, 2017. **121**(32): p. 7652-7659.
2. Kowalski, A.E., et al., *Gold nanoparticle capture within protein crystal scaffolds*. Nanoscale, 2016. **8**(25): p. 12693-12696.
3. Huber, T.R., et al., *Installing Guest Molecules at Specific Sites within Scaffold Protein Crystals*. Bioconjugate Chem. , 2017. **29**(1): p. 17-22.
4. Huber, T.R., et al., *Programmed Assembly of Host–Guest Protein Crystals*. Small, 2017. **13**(7): p. 1602703.
5. Yablon, D. and I. Chakraborty, *Machine learning to enhance atomic force microscopy analysis and operation*. Machine learning, 2020.
6. Müller, P., et al., *nanite: using machine learning to assess the quality of atomic force microscopy-enabled nano-indentation data*. BMC bioinformatics, 2019. **20**(1): p. 1-9.
7. Forchheimer, D., R. Forchheimer, and D.B. Haviland, *Improving image contrast and material discrimination with nonlinear response in bimodal atomic force microscopy*. Nature communications, 2015. **6**(1): p. 1-5.
8. Qi, Y., J. Klein-Seetharaman, and Z. Bar-Joseph, *Random forest similarity for protein-protein interaction prediction from multiple sources*, in *Biocomputing 2005*. 2005, World Scientific. p. 531-542.
9. Li, B.-Q., et al., *Prediction of protein-protein interaction sites by random forest algorithm with mRMR and IFS*. PLoS One, 2012. **7**(8): p. e43927.
10. Wang, Y. and L. Gao, *Detecting Protein Complexes by an Improved Affinity Propagation Algorithm in Protein-Protein Interaction Networks*. J. Comput., 2012. **7**(7): p. 1761-1768.
11. Vlasblom, J. and S.J. Wodak, *Markov clustering versus affinity propagation for the partitioning of protein interaction graphs*. BMC bioinformatics, 2009. **10**(1): p. 1-14.
12. Tuckerman, M.E., *Machine learning transforms how microstates are sampled*. Science, 2019. **365**(6457): p. 982-983.

13. Cendagorta, J.R., et al., *Comparison of the performance of machine learning models in representing high-dimensional free energy surfaces and generating observables*. The Journal of Physical Chemistry B, 2020. **124**(18): p. 3647-3660.
14. Brohee, S. and J. Van Helden, *Evaluation of clustering algorithms for protein-protein interaction networks*. BMC bioinformatics, 2006. **7**(1): p. 1-19.
15. Chen, D., et al., *Machine learning based methodology to identify cell shape phenotypes associated with microenvironmental cues*. Biomaterials, 2016. **104**: p. 104-118.
16. Unadkat, H.V., et al., *An algorithm-based topographical biomaterials library to instruct cell fate*. Proceedings of the National Academy of Sciences, 2011. **108**(40): p. 16565-16570.
17. Wang, D., et al., *Measuring Interactions of DNA with Nanoporous Protein Crystals by Atomic Force Microscopy*. Nanoscale, 2021. **13**(24): p. 10871-10881.
18. Qi, Y., *Random forest for bioinformatics*, in *Ensemble machine learning*. 2012, Springer. p. 307-323.
19. Belgiu, M. and L. Drăguț, *Random forest in remote sensing: A review of applications and future directions*. ISPRS journal of photogrammetry and remote sensing, 2016. **114**: p. 24-31.
20. Frey, B.J. and D. Dueck, *Clustering by passing messages between data points*. Science, 2007. **315**(5814): p. 972-976.
21. Dueck, D. and B.J. Frey. *Non-metric affinity propagation for unsupervised image categorization*. in *2007 IEEE 11th International Conference on Computer Vision*. 2007. IEEE.
22. Kraemer, H.C., *Extension of the kappa coefficient*. Biometrics, 1980: p. 207-216.
23. McHugh, M.L., *Interrater reliability: the kappa statistic*. Biochemia medica, 2012. **22**(3): p. 276-282.
24. Monitor, P.T., *Introduction to Bruker's ScanAsyst and PeakForce Tapping AFM Technology*. Bruker application note. Bruker Nano Inc.: Bruker application note. Bruker Nano Inc.

## CHAPTER 5: NON-COVALENT PRESENTATION OF PEPTIDE ADHESION LIGANDS WITHIN POROUS PROTEIN CRYSTALS: ATOMIC FORCE MICROSCOPY MEASUREMENTS AND CELL ATTACHMENT STUDIES \*

### Chapter 5 Overview

Protein crystals with sufficiently large solvent pores can non-covalently adsorb polymers in the pores and be used to present ligands to cells with tunable adhesion strength. Moreover, these porous protein crystals can store an internal reservoir of additional adhesion ligands, so that the surface can be replenished. In this study, we quantitatively measure and take advantage of the interaction between poly(ethylene glycol) (PEG)-arginine-glycine-aspartic acid (RGD) complex and nanoporous protein crystals to understand how surface presentation of peptide adhesion ligands can influence cell behavior. Here, through atomic force microscopy (AFM), force-distance (F-D) curves of interactions between PEG-RGD and host protein crystals were obtained for the first time. The activation of AFM tips allowed the tips to capture PEG-RGD that was pre-loaded in the protein crystal nanopores, mimicking how a cell would attach to and pull on the ligand through integrin receptors. The activated AFM tips also enabled the interactions of PEG-RGD with nanoporous protein crystals to be quantitatively studied while simultaneously revealing the morphology of the buffer-immersed nanoporous protein crystal surface in detail. This work also demonstrates that PEG-RGD can be loaded into porous protein crystals by diffusion, and can display RGD that is available to which cells can attach. Finally, we demonstrate that non-covalently attached PEG-RGD qualitatively alters adipose-derived stem cell spreading. This strategy can be used to design surfaces that non-covalently present multiple different ligands to cells with tunable adhesive strength for each ligand, and with an internal reservoir to replenish the precisely defined crystalline surface.

---

\* Portions of this chapter are reproduced from: Hasan Hedayati, Dafu Wang (co-first author), Julius D. Stuart, Ketul C. Papat, Chris D. Snow, Matt J. Kipper. "Nanoporous Protein Crystals for Non-Covalent Presentation of Peptide Adhesion Ligands: AFM Measurements and Cell Attachment Studies"

## 5.1 Introduction

The arginine-glycine-aspartic acid (RGD) sequence was reported as a cell adhesion peptide derived from fibronectin nearly 40 years ago.<sup>[1]</sup> Over the subsequent decades, this peptide and additional adhesion ligands from fibronectin,<sup>[2]</sup> laminin,<sup>[3-6]</sup> collagens,<sup>[7, 8]</sup> and other proteins were discovered.<sup>[9]</sup> Based on these discoveries, modifying biomaterials with cell specific peptide ligands has become a ubiquitous strategy for promoting cell adhesion, spreading, and migration.

Because integrins that bind many of these peptides transduce mechanical signals through the cell cytoskeleton (mechanotransduction), these signals can also alter cell phenotype via signal transduction that modulates gene expression, thereby driving cellular functions ranging from extracellular matrix deposition to stem cell differentiation.<sup>[10, 11]</sup> The responses of cells to adhesion ligands can be further altered by adhesion ligands presentation detail. For example, cell adhesion peptide gradients on surfaces or in hydrogels have been used to study the responses of cells to varying adhesion ligand concentrations.<sup>[12-14]</sup> Researchers have also shown that clustering ligands in nanoscale domains enhances signaling and may be necessary to achieve the maximal response.<sup>[15]</sup> Because the *in vivo* cellular microenvironment is not static, techniques to dynamically control peptide presentation through external stimuli have been developed. Zhao *et al.* reviewed magnetically responsive, electrically responsive, and thermally responsive materials that can essentially switch between two states of adhesion ligand presentation.<sup>[16]</sup> In addition, irreversible release or exposure of adhesion ligands has also been triggered by photosensitive, redox, and enzyme catalyzed reactions to achieve dynamic or spatially patterned ligand presentation.<sup>[16]</sup> While most reports describe adhesion ligands covalently attached to substrates, Grewal *et al.* recently reported using non-covalent peptide coiled coil complexes to reversibly and dynamically control the presentation of adhesion ligands from hydrogels and nanofibers.<sup>[17]</sup>

The putative isoprenoid binding protein from *Campylobacter jejuni* (Genebank ID: CJ0420, Protein Data Bank (PDB) code: 5W17) has been modified to form the protein CJ,<sup>[18]</sup> which forms hexagonal crystals (with P622 space group) with unusually large, 13-nm diameter pores.<sup>[18-23]</sup> Crosslinked CJ protein crystals are capable of serving as “hosts” to “guest” nanoparticles and macromolecules, including gold nanoparticles,<sup>[18, 23]</sup> proteins,<sup>[19]</sup> and oligomeric double-stranded DNA.<sup>[20]</sup> Furthermore, these large-pore protein crystals have favorable cytocompatibility.<sup>[22]</sup> The interior surfaces of the protein crystal pores can non-covalently adsorb macromolecules, making the porous protein crystal a capacious reservoir for the presentation of biochemical signals. Macromolecules bound inside the pores very near the crystal surface can be probed by cell surface receptors.

In previous work, we showed that these large pore protein crystals can absorb oligomeric double-stranded DNA with high affinity.<sup>[20]</sup> The DNA could be subsequently removed from the protein crystal pores by attachment to the tip of an atomic force microscope probe, via a disulfide bond formation chemistry. This experimental method enables probing an individual “loaded” pore near the surface of the crystal to determine the binding energy of a cluster of macromolecules attached to the AFM tip. In the course of our previous study, we serendipitously discovered that the high affinity binding of macromolecular guests was not unique to DNA, but extends to other macromolecules, including poly(ethylene glycol) (PEG).

In the present work, we use a CJ protein variant, which is suitable for growing large protein crystals. The modified CJ is sequence variant CJ-A34I-L48F-V50I-V121M-N162C-I163W-V165I. We propose for the name CJ<sup>OPT</sup> for this large-crystal variant, the ‘OPT’ represents the optimized, large crystal-growing variant. The nucleic acid sequence (C.1.1), the amino acid sequence (C.1.2), and the mutations (C.1.3) of CJ<sup>OPT</sup> protein are outlined in the Supporting Information. The CJ<sup>OPT</sup>

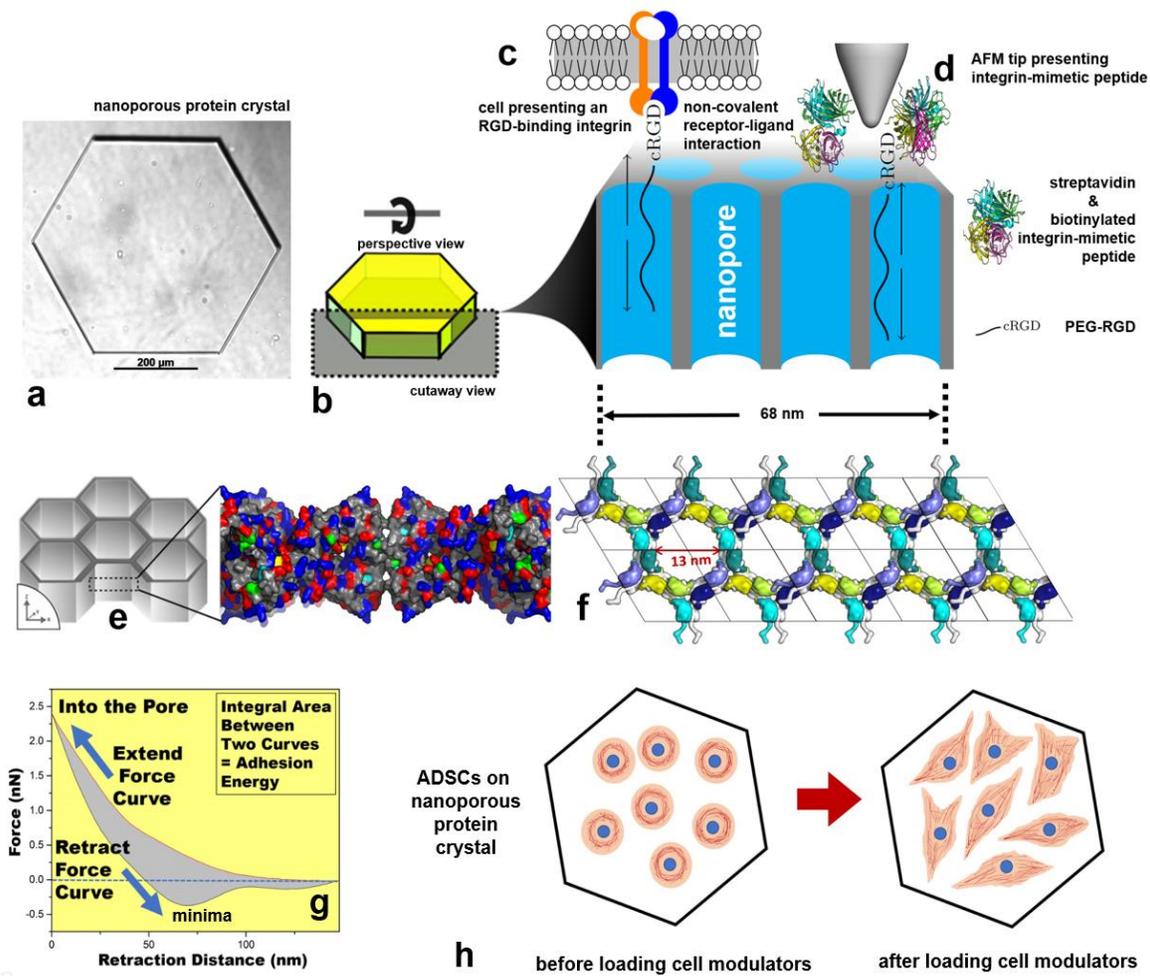
sequence is a variant of a previously reported lipid binding protein containing the following mutations designed to stabilize the inner hydrophobic cofactor-containing beta-barrel core: A34I, L48F, V50I, I163W and V165I. Additional mutations include V121M for strengthening observed hydrophobic crystal contacts, and N162C for allowing covalent installation of guest molecules using thiol chemistry as previously demonstrated.<sup>[21]</sup> Lastly, the N-terminal 20 residue signaling peptide (KKVLLSSLVAVSSLSTGLFA, UniProtKB Q79JB5)<sup>[24]</sup> was removed to enhance protein expression as previously described.<sup>[18]</sup> We hypothesized that an adhesion ligand (RGD) conjugated to PEG could be absorbed into the protein crystal pores, and that the non-covalent binding of PEG to the protein crystal pore or surface would provide an adhesive strength sufficient to enable cell adhesion and spreading. Because the attachment is non-covalent, but presumably occurs through multiple valency of the PEG chain, the attachment strength could be tuned by altering the PEG length, so that PEG molecules of different lengths could provide different strengths of receptor-surface binding. Because the PEG-RGD is absorbed into the volume of the protein crystal, the crystal provides a reservoir of ligand that can replenish the surface, if cells remove the PEG-RGD.

To demonstrate this novel concept, we show that PEG-RGD, performs as a non-covalently bound cell adhesion promotor, and is rapidly taken up by porous protein crystals (within ~10 min). We hypothesized that protein crystals loaded with PEG-RGD could be used to present the adhesive peptide, RGD to cells, thereby enabling cell attachment. Furthermore, the force required to retract an AFM tip presenting PEG-RGD, from a single protein crystal pore can be quantitatively measured by chemical force microscopy. This experiment simulates the attachment of cell surface integrins and the application of contractile force from the cell cytoskeleton. Chemical force microscopy is performed by decorating an AFM tip with the integrin-mimetic peptide (IntP) that

has affinity for the RGD ligand and imaging the protein crystal using the peak-force quantitative nanomechanics (peak force QNM) mode, on a Bruker Bioscope Resolve microscope. Previous research has proved that the biotinylated receptor peptide used in our study (IntP, CWDDGWLC) binds RGD ligand stably and strongly, to be a mimic of the RGD-binding site of the  $\beta_3$  subunit.<sup>[25,</sup>

<sup>26]</sup> We were able to collect hundreds of thousands of F-D curves during a single imaging experiment without interrupting the imaging process. This imaging mode collects force-versus-distance data at every pixel of an image, enabling us to precisely map the force-distance curves obtained when the tip interacts with a protein crystal pore or the wall of a pore, providing details of the connection between the mechanical behaviors and the morphology of the surface. For the first time, we directly and quantitatively measured the interactions of PEG-RGD with a porous protein crystal. This study thereby provides parameters that may enable future researchers to tune protein crystals and solution conditions for cell molecule modulators storage and release. Notably, this current study is not designed to isolate the interaction of a single PEG-RGD chain with the host crystal. Instead, multiple PEG-RGD complexes are captured by the activated AFM tip, which mimics the prospective multivalent attachment of cells to protein crystals via PEG-RGD. An IntP-decorated tip may display RGD binding sites at a higher local density than a cell surface can, due to the receptor spacing. The current interaction data and analysis will also guide the design and interpretation of future single-molecule studies of this system. We show that by adding PEG-RGD to protein crystals, the spreading and cytoskeletal arrangement of ADSCs can be modulated, exhibiting more well-developed actin stress fibers. Since the force and adhesion energy generated by the interaction between PEG-RGD and protein crystals are within a specific and narrow range, PEG-RGD may serve as a cell molecule modulator, in future applications, such as guiding stem cell differentiation. The strategy in this study can also be applied to other more specific cell

adhesion peptides to prepare surfaces that selectively bind different cell phenotypes with different strengths.



**Fig. 5.1** A periplasmic protein, “CJ”, from *Camphylobacter jejuni* forms (a) porous protein crystals that we stabilize via crosslinking. (b) Typical crystals are hexagonal prisms. (c) Chemical schematic illustration of ADSCs’ attachment to the nanoporous protein crystal through biotinylated peptide and PEG-RGD complex. (d) A hexagonal array of 13 nm-diameter nanopores runs from the top to the bottom of the crystal. Modified PEG-RGD complexes can be loaded into these nanopores via diffusion. An AFM tip modified with the streptavidin-biotinylated IntP, and presenting RGD-PEG can be used to probe the interactions of PEG with the protein crystal. (e) Crystal schematic with nanopore cut away, and zoomed in on a slice of nanopore side-wall with ionizable amino acids highlighted. Cysteine residues are shown in yellow. Carboxylic acids (Asp, Glu) are shown in red. Arginines are shown in cyan. Lysines are shown in dark blue. Histidines are shown in green. Both the N- and C- terminus contain flexible regions that are not pictured in this crystal structure. Zoomed in image created using PyMOL. (f) A top view of five adjacent nanopores (PDB code: 5W17, same scale as (d)). (g) Representative illustration of the force-distance (F-D) curve of a nanopore center. (h) A schematic of ADSCs’ spreading on the upper surface of nanoporous protein crystals before and after loading cell modulators.

## 5.2 Experimental

### 5.2.1 Materials

Polyethylene glycol (5 kDa and 10 kDa) with a methyl terminus at one end and either a fluorescein or a succinimidyl group at the other end (mPEG<sub>5000</sub>-FITC, mPEG<sub>10000</sub>-FITC, mPEG<sub>5000</sub>-SC, and mPEG<sub>10000</sub>-SC) were purchased from Biochempeg (Watertown, MA). Cyclo-RGDfK peptide was purchased from MedChemExpress (Monmouth Junction, NJ). A biotinylated receptor peptide for the RGD ligand (IntP, CWDDGWLC) was purchased from Genscript (Piscataway, NJ). CJ<sup>OPT</sup> protein crystals were obtained and crosslinked according to procedures outlined in the Supporting Information as Section C.1.

### 5.2.2 Synthesis and Characterization of PEG-RGD with <sup>1</sup>H NMR

PEG<sub>10000</sub>-SC (or PEG<sub>5000</sub>-SC) (30 mg) was dissolved in DMF to make a 10 mM solution. Separately, 6 mg of cyclo-RGDfK was dissolved in an aqueous Na<sub>3</sub>PO<sub>4</sub> solution (0.1 M and pH = 7.2) to prepare 10 mM concentration. The PEG-SC and cyclo-RGDfK solutions were mixed and reacted with gentle mixing for 12 h at room temperature. The mixture (1 mL) was then transferred into a dialysis cassette (molecular weight cutoff of 7 KDa) and dialyzed against water for two days. The white powder product (PEG<sub>10000</sub>-RGD or PEG<sub>5000</sub>-RGD) was collected after freeze-drying. PEG was ~0.3 nm per repeat unit (120 repeat units in 5 kDa PEG, 240 repeat units in 10 kDa PEG, 36 to 72 nm in length).

Proton nuclear magnetic resonance (<sup>1</sup>H NMR) spectra of the reactants and products were recorded on a Bruker AV-400 spectrometer (Bruker). D<sub>2</sub>O was used as the solvent and the spectra were obtained at room temperature. The <sup>1</sup>H NMR spectrum of PEG<sub>5KDa</sub> and PEG<sub>5KDa</sub>-c(RGDfK) is shown in Fig. C2. The peak that appeared around 7.3 ppm indicates the successful conjugation of benzyl group in RGD to the PEG molecule.

### 5.2.3 PEG Loading Into Porous Protein Crystals

Confocal microscopy was used to monitor the loading of PEG-FITC. Protein crystals were immobilized on their sides in a microchannel with a cover to prevent evaporation during the experiment. PEG-FITC with 1  $\mu\text{M}$  concentration in TE (Tris-EDTA) / DI H<sub>2</sub>O buffer (pH = 7.5) was added to the solution surrounding the crystals while z-stack imaging was continued for 30 min.

### 5.2.4 Crystal Immobilization

For the AFM experiments (described below in Section 5.2.6), the CJ<sup>OPT</sup> protein crystals were immobilized on glass-bottom petri dishes (Willco Wells) employing a UV-curable glue (Bondic Inc.). The top of a crystal probe (Minitool HR4-217) was used to transfer a drop of UV-curable glue onto the surface of a petri dish (Ted Pella, Inc. 14025-20). The glue was gently and evenly spread on the dish surface to make the layer of glue as thin as possible. CJ<sup>OPT</sup> protein crystals were transferred to the glue with a loop. Critically, the crystal was transported inside a tiny drop of buffer, such that the crystal was not desiccated. The UV-glue was viscous and did not noticeably mix with the buffer. The glue was then cured by exposing to UV-light LED (Bondic SK001) from above for 10 s. The glue cured after about 2 min, after which additional drops of buffer (typically ~5 mL) were added to the dish to prevent the crystal from drying.

### 5.2.5 AFM Tip Modification

Bruker's ScanAsyst Fluid+ tips were modified to covalently attach the polyethylene glycol-arginylglycylaspartic acid (PEG-RGD) complex. The tips have a slim shape with estimated tip radius as small as 2 nm, as well as a silica surface layer. Hydroxyl groups on the tip surface were activated by oxygen plasma, enabling the tip to be modified with the (3-aminopropyl)trimethoxysilane (APTMS) by molecular vapor deposition (MVD).<sup>[27]</sup> Detailed experimental procedures are outlined in the Supporting Information, Section C.2.

1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and *N*-hydroxysuccinimide (NHS) were then dissolved in phosphate-buffered saline (PBS) buffer (pH = 7.0) to make 10 mM solution. Then the tips were immersed into the EDC-NHS mixture solution for 60 min at room temperature. Tips were later washed several times with PBS buffer (Gibco, 1×, without calcium chloride or magnesium chloride). Streptavidin was coupled to the tips by incubating with 50  $\mu$ L 1 mg/ml streptavidin at room temperature for 60 min. The streptavidin molecules are about 4-6 nm in diameter.<sup>[28]</sup> Unbound streptavidin molecules were removed by rinsing with PBS. The tips were finally modified with the biotinylated IntP (5 mg/mL) in solution at 4 °C overnight. Then tips were washed several times with PBS and DI water.

The schematic illustration of the intended AFM tip modification process is shown as Fig. C1. Each step in the surface modification of the AFM tips was evaluated by X-ray photoelectron spectroscopy (XPS). Detailed procedures of XPS characterization and data processing are outlined in the Supporting Information Section C.2.

### **5.2.6 AFM Imaging and Force-Distance Measurements**

We operated the AFM (Bruker Bioscope Resolve, mounted on a spinning-disc confocal microscope built around a Nikon Eclipse TiE) in quantitative nano-mechanics (QNM) PeakForce capture mode. All images and force curves were collected using ScanAsyst Fluid+ tips (Bruker). Crystal imaging was performed in Tris-EDTA buffer (5 mM Tris-HCl, 1 mM EDTA, pH = 7.5, dissolved in de-ionized water) with crystals affixed to the bottom of a glass petri dish. The AFM line scan rate was set to 1.0 Hz and the peak force tapping frequency was set to 1.0 kHz. The maximum peak force set point was set to 2.0 nN (activated AFM tip on loaded protein crystals) and 1.5 nN (else conditions). AFM has been used to measure the strength of bonds between biological receptor molecules and their ligands.<sup>[29-32]</sup> Notably, the retract force measured here (50

- 750 pN, under various conditions) is similar to the forces used by investigators who use similar experiments to study the receptor-ligand interactions. Specifically, Beebe *et al.* used a force of  $442 \pm 17$  pN to measure the ligand-receptor bond-rupture force between the streptavidin/biotin partners.<sup>[33, 34]</sup> Stayton *et al.* measured the force of detachment by AFM as  $433 \pm 33$  pN between biotin-functionalized tips and streptavidin-functionalized samples.<sup>[35]</sup> In the present work, there are likely multiple IntP presented on the tip, enabling multi-valent tip-PEG-RGD interactions, and in our experiments, retraction forces are for multivalent interactions. In addition, as a control experiment, to determine if the observed adhesion energy reflects equilibrated molecular interactions, we have also operated some of the experiments under the exact same condition but a much slower tapping frequency of 1 Hz (Supporting Information, Fig. C2). Tip-sample interactions are measured with pN-resolution by the deflection of the cantilever. Analysis of the AFM data was performed in NanoScope (Bruker, Inc.), Origin (OriginLab, Inc.), Python (Version 2.7), and Matlab (Version 2019).

The CJ<sup>OPT</sup> protein crystal sample was also characterized by non-modified AFM tips, confirming immobilization as well as verifying that the crystal surface is clean. It is important to independently confirm the porous morphology of the protein crystal surface prior to PEG-RGD interaction measurements, because if crystals are poorly prepared and crosslinked, the surfaces can be fouled with amorphous (non-crystallized) or aggregated protein molecules. After loading the guest molecules, the PEG-RGD-loaded crystal was then imaged with an activated AFM tip, and the corresponding force distance (F-D) curves were collected at each pixel in the AFM image. AFM imaging and F-D curve collection was conducted for three different experimental conditions, using combinations of un-modified and activated tips and either loaded or unloaded protein crystals. The three experimental regimes: (condition a) an un-modified AFM tip on an unloaded

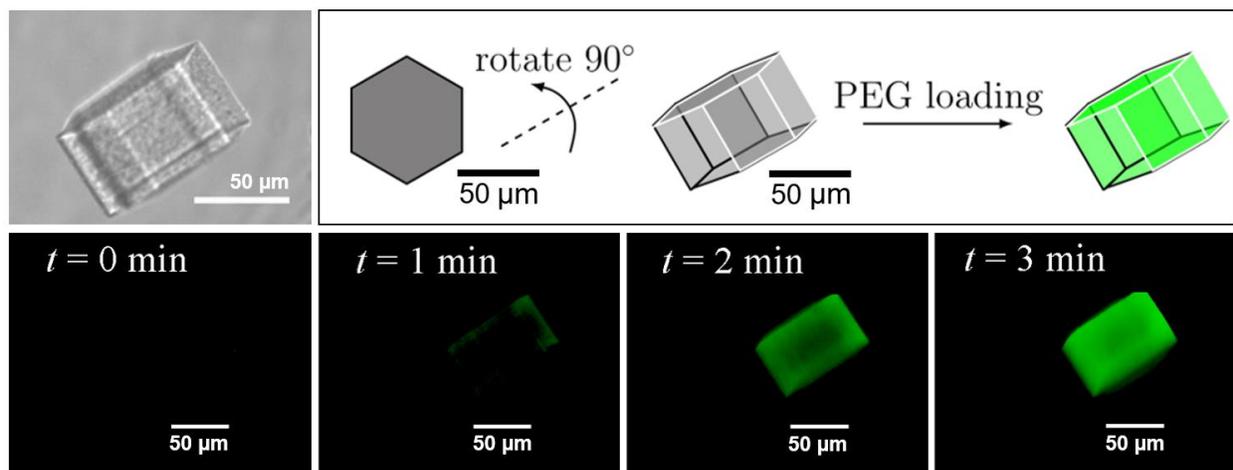
crystal, (condition b) an activated AFM tip on an unloaded crystal, (condition c) an activated AFM tip on a PEG-RGD-loaded crystal (Fig. 5.2 (d-f)). As a comparison experiment, we also modified the PEG-RGD to the AFM tip and used the modified to probe the unloaded crystals. From each AFM image, F-D curves were manually assigned to one of two classes, corresponding to protein crystal surface features: pores and walls.

### 5.2.7 Cell Culture and Imaging

The experimental procedures for culture of human adipose-derived stem cells (ADSCs), and ADSC imaging on the protein crystals are outlined in the Supporting Information, Section C.3.

## 5.3 Results and Discussion

### 5.3.1 PEG-RGD Loading

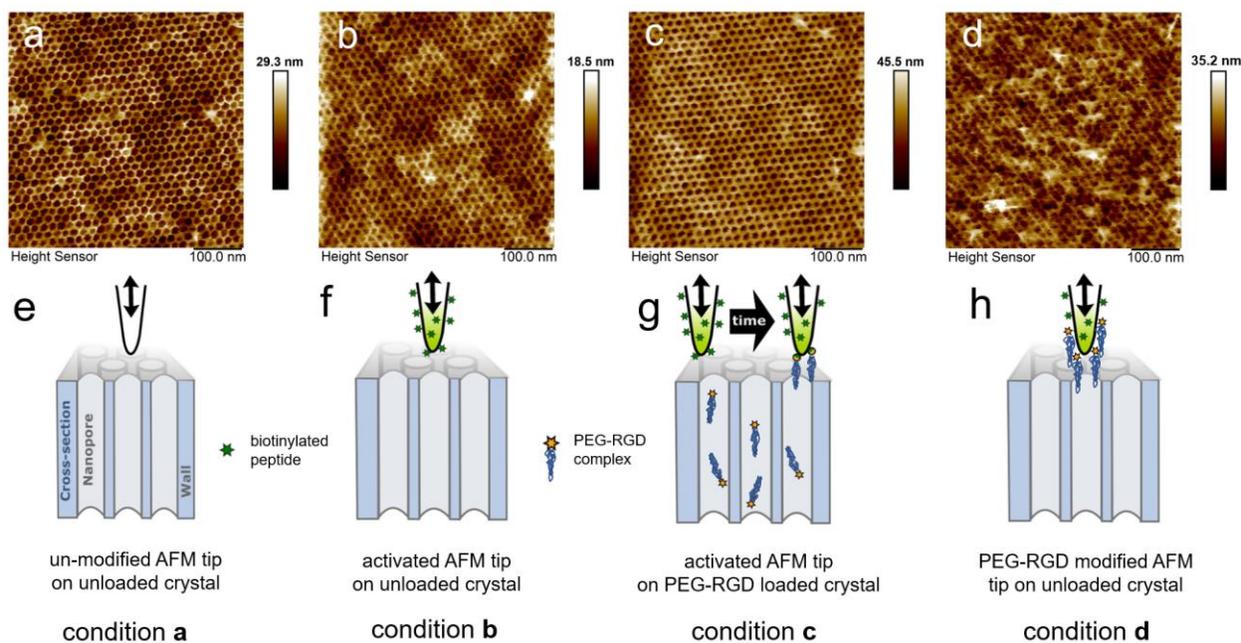


**Fig. 5.2** A  $CJ^{OPT}$  protein crystal (DIC image shown in upper left, with a 50 μm scale bar) was placed on one side and imaged during  $PEG_{10000}$ -FITC loading, by confocal microscopy. Four confocal microscopy images (bottom row) taken at 1-minute intervals are shown, from a z-plane near the center of the crystal. Diffusion of 10 kDa PEG into a crystal of this size reaches apparent equilibrium in less than 3 minutes.

PEG has an affinity for the interior of  $CJ^{OPT}$  protein crystals and is rapidly absorbed by the crystals from solution. To confirm loading, a crosslinked  $CJ^{OPT}$  protein crystal in AFM imaging

buffer was imaged on a Nikon spinning-disc confocal microscope, during loading. The crystal (approximately 50  $\mu\text{m}$  tall and 80  $\mu\text{m}$  in diameter) was placed on one side, and a fluorescein-labeled PEG (10 kDa) was added to the imaging buffer at a final concentration of 1  $\mu\text{M}$ . The crystal was imaged every 30 s, at five different  $z$  positions (separated by 5  $\mu\text{m}$ ), near the center of the crystal. The 10 kDa PEG rapidly accumulated in the protein crystal interior, reaching apparent equilibrium within 3 minutes. Four confocal images from the center of the crystal, showing the first 3 minutes of diffusion are shown in Fig. 5.2.

### 5.3.2 Surface Morphology of the Protein Crystals by AFM

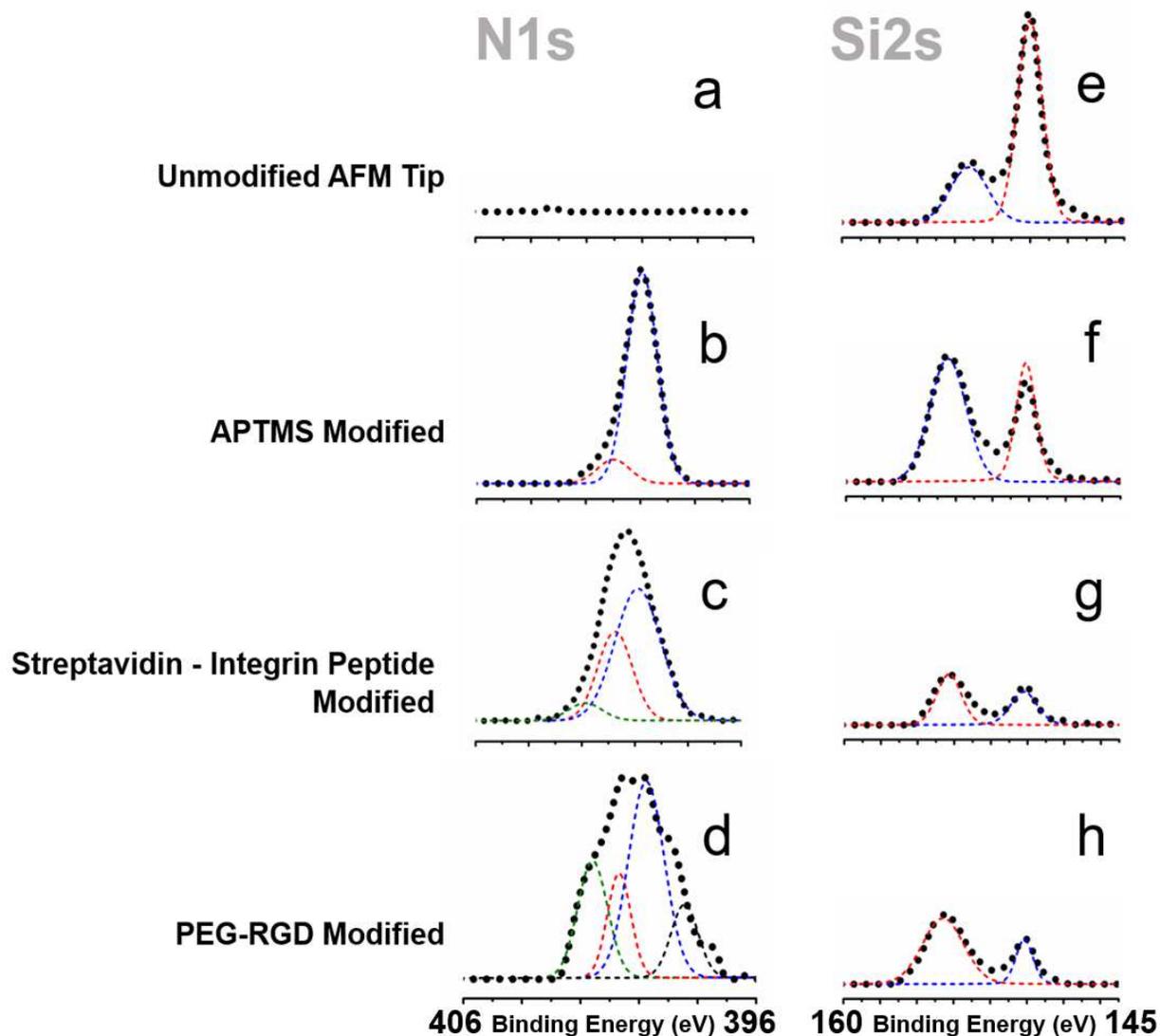


**Fig. 5.3** High-resolution AFM images of (a) an unloaded crystal imaged with a non-modified AFM tip, (b) an unloaded crystal imaged with an activated AFM tip, and (c) a crystal loaded with PEG-RGD imaged using an activated AFM tip, (d) an unloaded crystal imaged with a PEG-RGD modified AFM tip. Force microscopy schematic illustration of AFM tips with protein crystals were shown as (e-h). The condition indices a-d in this figure correspond to the indices in Section 5.2.3 and Fig. 5.2 (a-d), accompany by the corresponding surface morphology imaging by AFM.

AFM imaging shows the details of the porous  $\text{CJ}^{\text{OPT}}$  crystal surfaces (Fig. 5.3 (a-d) and Fig. 5.5 (a-b)). The  $\text{CJ}^{\text{OPT}}$  protein crystal surface has a regular honeycomb nanopore structure with features that were consistent among different protein crystal samples. We have imaged the surface

morphology under three key conditions described in Section 5.2.3 and Fig. 5.3 (d-h). For comparison, we also quantitatively measured the interaction between an inactivated AFM tip with a PEG-RGD-loaded crystal (Supporting Information Fig. C3 (b)). The observed surface structure and morphology of CJ<sup>OPT</sup> protein crystals is quite uniform and does not change significantly when imaged under different loading and tip modification conditions. It is remarkable that the tip covalently laden with streptavidin (~6 nm diameter for streptavidin tetramers<sup>[28]</sup>) as well as noncovalently laden with PEG-RGD (the lengths at maximum chain extension are 47.9 and 192 nm for 5K and 20K PEG molecules respectively<sup>[36]</sup>) can still be used to resolve the crystal surface clearly as well as to penetrate into the major nanopores. Indeed, per Fig. 5.7 (b), the fully encumbered tips penetrate *more* deeply into the nanopores. Surface modification of the AFM tip with streptavidin, biotinylated IntP, and PEG-RGD alters the both the adhesion of the tip to the crystal and the mechanical properties of the tip. Particularly, after adding PEG-RGD to the tip, a greater peak force set point (2 nN, instead of 1.5 nN) must be used to obtain image quality capable of resolving the pores and walls of the protein crystal surface.

### 5.3.3 Modification of AFM Tips



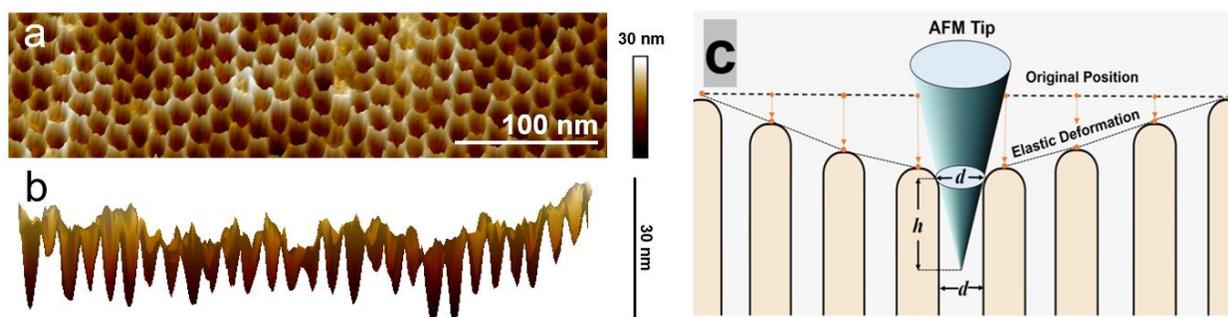
**Fig. 5.4** High-resolution XPS spectra of ScanAsyst Fluid+ tips at different stages of modification in the regions of the N1s: **(a)** confirms that there is no nitride or ammonium prior to reaction with APTMS, **(b)** +APTMS (399.69 eV N-sp<sub>2</sub> C and 400.83 eV ammonium), **(c)** +streptavidin-integrin peptide (399.72 eV NH-C=O, N-sp<sub>2</sub> C, 400.80 eV ammonium, and 401.85 eV NH-C, HO-N-C), **(d)** +PEG-RGD (398.52 eV nitride from PEG-RGD, 399.66 eV NH-C=O, N-sp<sub>2</sub> C, 400.78 eV ammonium, and 401.80 eV NH-C(=NH)-NH<sub>2</sub>, NH-C, HO-N-C) proving that PEG-RGD complex can be chemically bound to the activated AFM tip via the ligand-receptor interactions between RGD and the peptide (IntP), and **(e-h)** the Si2s from silicon on AFM tip surface during each step of the reaction. Modification attenuates the strength of Si2s signal.<sup>[37, 38]</sup>

The spring constant and tip diameter of each AFM tip used for imaging and quantitative measurements were quantified using hardness and surface roughness standards.<sup>[39]</sup> The spring

constant of a fully activated AFM tip used to collect data in this work was 0.89 N/m, with an estimated tip diameter of 5.87 nm (ETD, data from NanoScope).

Detailed XPS results are shown in Fig. 5.4 (a-h), XPS was used to characterize the modification of AFM tips. In this study, the high-resolution XPS spectra of N1s peaks confirm each step of modification chemistry. Activation of the tip with biotinylated peptide enables the PEG-RGD complexes to be bound to the AFM tip via the IntP-RGD interaction. This enables PEG-RGD complex to be bound and extracted from nanopores of CJ<sup>OPT</sup> protein crystals by the AFM tip.

### 5.3.4 Interaction of PEG<sub>10KDA</sub>-RGD and Nanoporous Protein Crystals

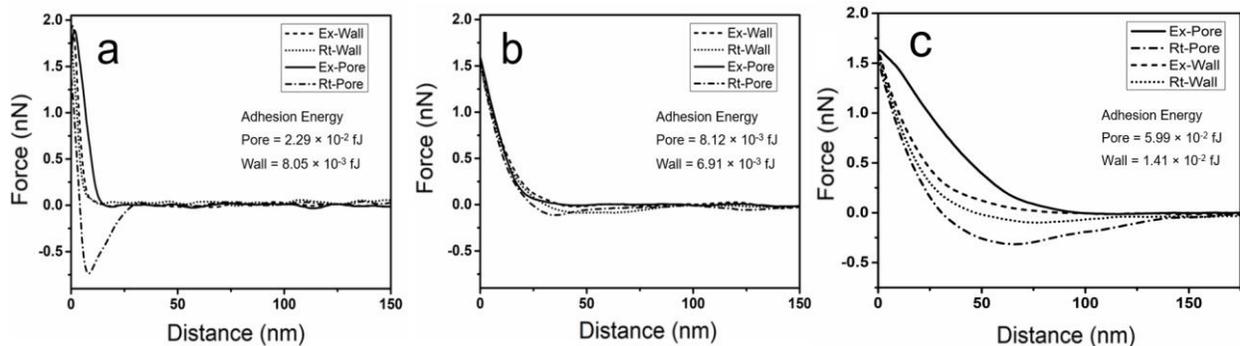


**Fig. 5.5** (a) 3D rendering of a portion of the height data from Fig. 5.3 (a), (b) a perpendicular cross-section view of 3D AFM image of the CJ<sup>OPT</sup> protein crystal surface from Fig. 5.3 (a), across the center of multiple pores in one single line, and (c) an illustration showing possible axial elastic deformation of the protein crystal surface (lateral deformation is not illustrated), and the AFM probe penetration into the pores, “*d*” represents the effective diameter of a nanopore, while “*h*” represents the actual depth of probing.

Hypothetically, activated AFM tips may bind and pull upon PEG-RGD that is adsorbed within the host crystal. Such a tip might fully extract a number of PEG-RGD molecules and carry this ligand complement in and out of the host crystal during the extension retraction cycles. As the tip visits different nanopores, such a ligand complement may undergo dynamic exchange. The plausibility of full extraction depends on the relative mechanical strength and binding kinetics for PEG-RGD with the host crystal compared to PEG-RGD with the IntP. Our main data for evaluating

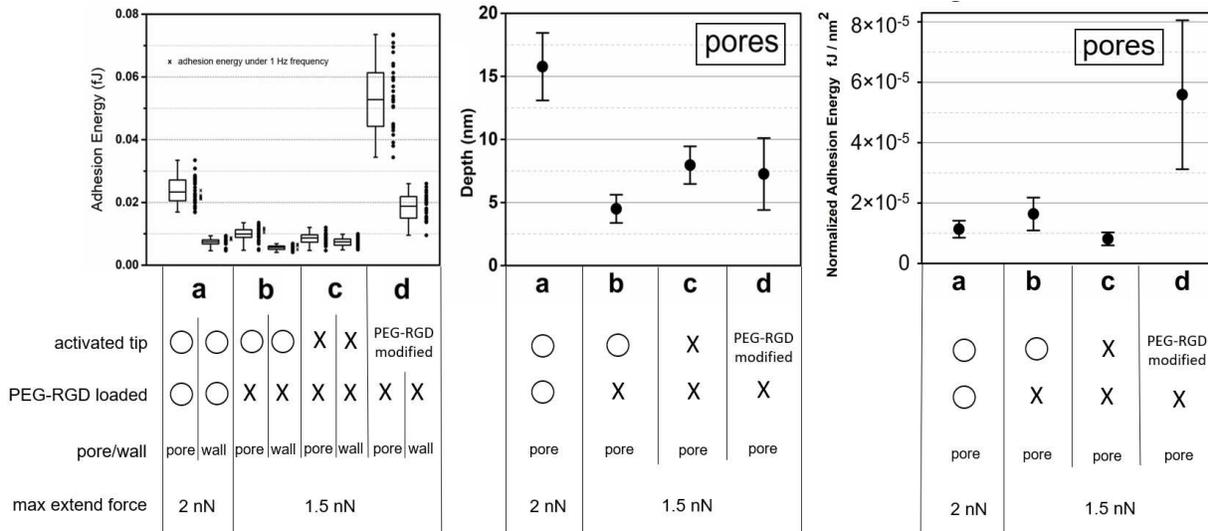
these hypotheses comes from applied force versus distance data. Bruker's PeakForce QNM imaging mode captures and records one force-distance (F-D) curve at each pixel of the scanned area of the CJ<sup>OPT</sup> crystal surface. To measure multivalent interaction effects between PEG-RGD and the CJ<sup>OPT</sup> protein crystals, four different combinations of tip activation and crystal loading were used are (Fig 3 (d-f): (condition a) an un-modified AFM tip on an unloaded crystal, (condition b) an activated AFM tip on an unloaded crystal, (condition c) an activated AFM tip on a PEG-RGD loaded crystal, and (condition d) an activated AFM tip with bound PEG-RGD on an unloaded crystal. Since the force-distance data is collected at each pixel in the image, we can classify each force-distance curve as one of two types: "pore" pixels (in which the AFM tip penetrates a 13-nm diameter crystal pore) and "wall" pixels (in which the AFM tip is interacting with a pore wall). Fig. 5.3 (a-c) shows that loading PEG-RGD into the nanopores of CJ<sup>OPT</sup> protein crystals does not change the surface morphology nor does it change the nanostructure of the protein crystals. While minor difference might be attributed to crystal-to-crystal variation in growth and crosslinking, the regular honeycomb-like lattice of 13-nm diameter pores is clearly visible at all conditions.

The dominant interaction between the PEG-RGD complex and activated AFM tip is likely via the IntP and the RGD. In this work, the forces measured during retraction were between 50 and 750 pN (Fig. 5.6 (a-b)) under various conditions. When the activated tip interacts with a protein crystal containing no PEG-RGD, there is a very little adhesion (Fig. 5.6 (b)). When a PEG-RGD modified AFM tip interacts with an unloaded crystal, the retraction force reaches a maximum magnitude of 470 pN (0.47 nN).



**Fig. 5.6** Representative examples of force-distance (F-D) curves (Ex-extend, Rt-retract) obtained using an activated AFM tip from (a) PEG-RGD loaded protein crystal, (b) unloaded protein crystal, and (c) a PEG-RGD modified AFM tip on an unloaded protein crystal.

The adhesion energy at each pixel for the tip interaction with the protein crystal can be calculated from the integral of the area between the extend force curve and the retract force curve. By precisely mapping these force curves to the “pore” and “wall” areas of the protein crystal image, we find that pixels collected from pore areas of the nanostructure have generally larger maximum retract force, and longer distance of interaction than the pixels collected from wall areas. As a result, the interaction traces collected within the nanopores have higher adhesion energy than the wall areas. Fig. 5.6 (a) shows that the interaction of the activated AFM tip with protein crystal nanopore in the retract portions of the curve occurs over 0-30 nm in the z-direction. During this process, the crosslinked protein crystal surface may deform elastically under the approximately 2 nN forces applied here Fig. 5.5 (c). Axial and lateral deformation may also contribute to the penetration distance of probing, and induce an extra energy increase by performing extra work.



**Fig. 5.7** Adhesion and penetration for the various AFM probe conditions: the min, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles, and max of the adhesion energy of (a-a) pore areas by an integrin peptide-activated tip on a PEG-RGD loaded crystal, (a-b) wall areas by an integrin-activated tip on a PEG-RGD loaded crystal, (a-c) pore areas by an activated tip on an unloaded crystal, (a-d) wall areas of an activated probe on an unloaded crystal, (a-e) pore areas by an inactivated probe on an unloaded crystal, (a-f) wall areas by an inactivated probe on an unloaded crystal, (a-g) pore areas by a PEG-RGD modified tip on an unloaded crystal, (a-h) wall areas by a PEG-RGD modified tip on an unloaded crystal, detailed data are listed in the Supporting Information Table C1 (a); and (b) probe penetration depth with error bar signifying the distribution and mean number of probe penetration depth into the nanopores respectively correspond to the experimental condition a-c to the indices in Section 5.2.3 and Fig. 5.2 (d-f), detailed data are listed in the Supporting Information Table C1 (b), (c) Adhesion energy in the pores, normalized by the interacting pore area (adhesion energy divided by average area computed from the depth of penetration of the AFM tip into the pore). The conditions labeled a–d here correspond to the conditions described in Fig. 5.7 (b).

In this work, we randomly select 30 sets of F-D curves on pores and walls respectively for each of the four experiments described in Fig. 5.2 (e-h) (240 sets of F-D curves total). As shown in Fig. 5.7 (a), measuring with an activated AFM tip, when PEG-RGD complexes are not preloaded into a nanoporous protein crystal, minimal differences were observed for average total adhesion energy between pore and wall areas. Similarly, when measuring with activated AFM tips on an unloaded protein crystal, the adhesion energy of both pore and wall areas was similar and very low. In contrast, when measuring by a PEG-RGD-modified AFM tip to nanopores of the unloaded CJ<sup>OPT</sup> protein crystal (condition **d** in Fig. 5.3 (h)), the results contain the highest average adhesion

energy within the pore areas compared with all other conditions. Within the selected sample sets, the highest normalized adhesion energy of pore areas under this condition reaches  $5.59 \times 10^{-5}$  fJ/nm<sup>2</sup>, which is more than three times higher than any other conditions. And the average total adhesion energy under condition **d** reaches  $5.39 \times 10^{-2}$  fJ, which is the highest average total adhesion energy among all conditions as well. In support of the model that the guest PEG-RGD was only accessible within the host crystal nanopore, for comparison, the average total adhesion energy for the wall areas of an unloaded crystal collected by PEG-RGD-modified tips was  $1.86 \times 10^{-2}$  fJ, which is significantly smaller than the average total adhesion energy collected in the pore areas. In addition to that, the average total adhesion energy for the wall areas of a PEG-RGD loaded crystal collected by activated tips was only  $7.37 \times 10^{-3}$  fJ, which is similar to the average adhesion energy of the wall areas from an unloaded crystal collected by inactivated tips ( $7.36 \times 10^{-3}$  fJ). The average adhesion energy for pore areas of unloaded protein crystal by activated and inactivated AFM tip are  $9.80 \times 10^{-3}$  fJ and  $8.60 \times 10^{-3}$  fJ respectively. Under this condition, the normalized adhesion energy for pore areas of unloaded protein crystal by activated and inactivated AFM tip reaches  $1.64 \times 10^{-5}$  fJ/nm<sup>2</sup> and  $8.10 \times 10^{-6}$  fJ/nm<sup>2</sup> respectively. Adhesion energy from PEG-RGD-modified nanopores is significantly larger than the adhesion energy from the pore areas for an unloaded crystal, indicating strong adhesion required both the PEG-RGD as well as the IntP-biotin-streptavidin-modified AFM tip.

In this experiment, the interaction is maintained over a distance of ~150 nm for the pore areas of an unloaded crystal collected by PEG-RGD modified tips, and ~30 nm for pore areas of a PEG-RGD loaded crystal collected by activated tips. This difference could arise due to the different total number of molecules attached to the PEG-RGD-laden AFM tip, and the different morphology of the tip surface with or without the modification of PEG-RGD. Further confirmation

for the strong interaction between the PEG-RGD and the  $CJ^{OPT}$  protein crystals is provided by comparing condition *b* to condition *c* and condition *d* (Fig. 5.3 (e-h)), in which a PEG-RGD-modified AFM top was used in the unloaded protein crystal, or an activated AFM tip was used in the absence and presence of loaded PEG-RGD. The average total adhesion energy of the pore areas significantly increased with the presence of PEG-RGD.

We also observe a large attractive force encountered in the retraction curve for the “pore” areas when PEG-RGD-laden AFM tips are interacting with the protein crystals (as shown in Fig. 5.6 (a) and (c)). The maximum magnitude of the retraction force occurs when the tip is still about 5 nm on average below the plane of the crystal surface. As the tip continues to rise out of the crystal, there is a gradual decline in the attractive force deflection until the tip is about 30 nm above the local crystal surface (where the force reaches  $\sim 0$  pN). The gradual decline in the adhesion force as the tip is retracted is consistent with a sum of many small adhesive interactions, rather than with a few strong interactions. These many small interactions are likely caused by the gradual retraction of multiple PEG chains, wherein each PEG chain can form multivalent interactions with the protein crystal pore wall. As the lengths of the PEG chains are pulled out from the protein crystal pore, their adhesive strength is gradually reduced to zero over a large (25 nm) travel distance of the tip. Moreover, the RGD-IntP bonds endure throughout the imaging experiment, wherein the tip visits many pores without losing its ability to adhere to the AFM tip. Therefore, the complement of captured PEG-RGD remains on the tip throughout multiple extension-retraction cycles of the tip, even as the tip moves between neighboring nanopores; and the measured adhesion energy represents the interaction between the PEG and the protein crystal, rather than rupture of the IntP-RGD. The adhesion maps of the  $CJ^{OPT}$  crystals for all imaging conditions, corresponding to Fig.

5.3 (a-c) are attached to the supporting information as Fig. C5 (a-c), implicated that the nature of the AFM tips are not changing during the entire imaging process.<sup>20</sup>

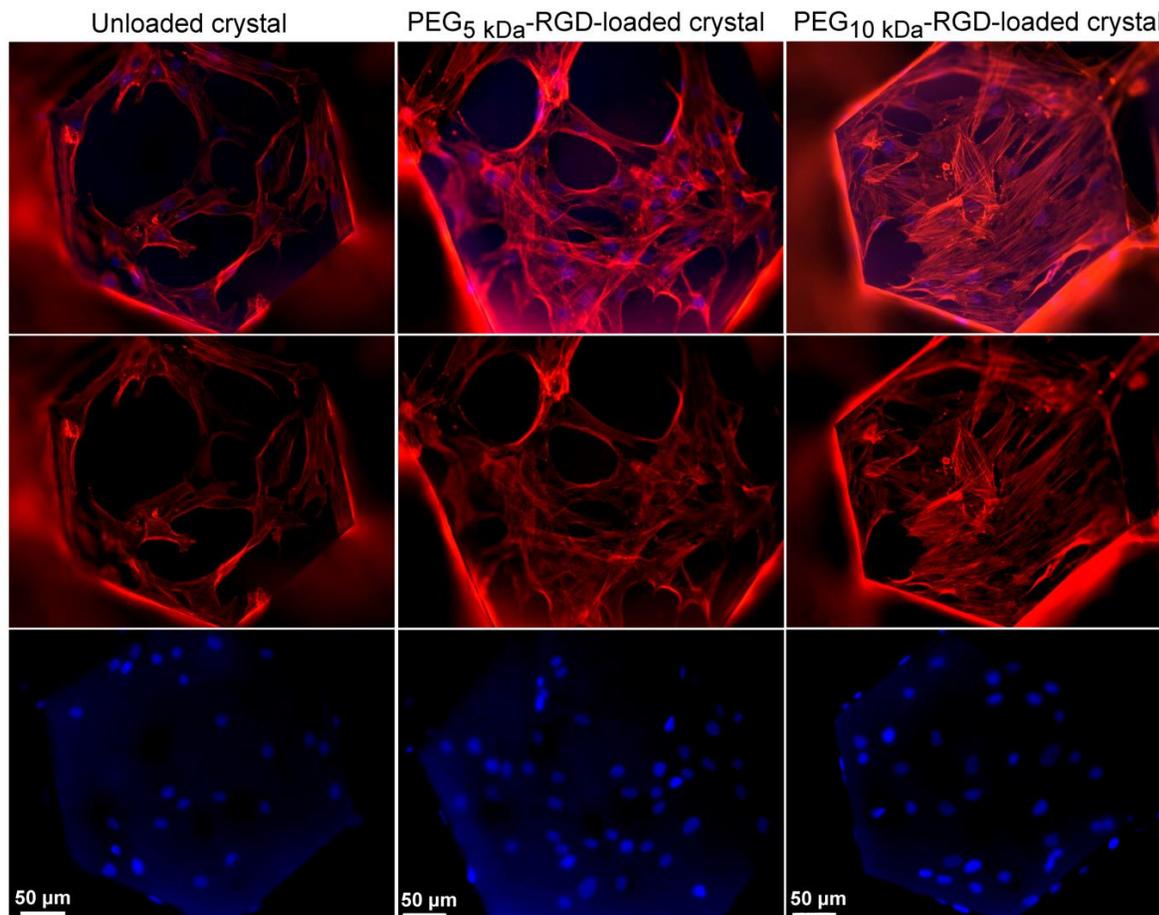
When imaging within the PeakForce QNM mode, the tip oscillates with a sinusoidal  $z$ -position profile, reaching a minimum in velocity ( $dz/dt = 0$ ) when the tip is fully extended. At 1 kHz, with a total travel of 200 nm, the tip remains within 1 nm of the fully extended position for 52 ms.<sup>20</sup> This dwell time in the pore should be sufficient for bound PEG molecules to sample the local crystalline environment to find energetically favorable bound conformations. To confirm that the PEG-modified tip can probe the adhesion of the PEG to the protein crystal pores during the 1 kHz oscillation, we also collected force-distance curves using approach-retract experiments with a linear ramp at a much slower (1 Hz frequency) using an activated AFM tip from PEG-RGD-loaded protein crystal, and with an unloaded protein crystal. Here, the frequency of 1 Hz means that the entire approach-retract cycle takes 1 second. But the speed would depend upon the retraction distance. Despite significantly increasing the time for PEG-RGD to bond the activated tip, no obvious change was observed in the adhesion energy (Supporting Information Fig. C4 (a-b)). When the probing movement is fast under the frequency of 1k Hz, due to the short period of time.

Considering the different distances of tip penetration into the pores, the average adhesion energy within the nanopores can be normalized by the average individual pore area of interaction, more specifically, calculated by approximating the area of a 13-nm diameter cylindrical pore<sup>18-21</sup>,<sup>23</sup> ( $A_{pore\ surface} = 13nm \cdot h \cdot \pi$ ) at the local depth of penetration  $h$  of the AFM tip (as shown in Fig. 5.5 (c)). In this work, an average of normalized adhesion energy for the pore areas of the PEG-RGD loaded LCJ protein crystals, collected by activated AFM tip, was  $1.14 \times 10^{-5}$  fJ/nm<sup>2</sup>, with an average probe penetration depth of  $15.8 \pm 2.7$  nm. For comparison, the average penetration depth

on unloaded crystals was respectively 4.5 nm and 8.0 nm for AFM tips before and after activation. This somewhat counterintuitive result suggests that favorable interactions are a more important driver for probe depth than steric, since a steric model would predict a lower penetration depth for AFM tips that are encumbered with bulky streptavidin tetramers as well as bound PEG-RGD. Chemically, penetration depth will make a difference on the effective interacting area when the tip penetrates the pore, and thereby affect the adhesion energy of interaction. Single-molecule studies in the future could further elucidate the interactions of individual PEG-RGD with CJ<sup>OPT</sup> nanoporous protein crystal nanostructures.

### **5.3.5 Modulation of Cell-adhesion to Protein Crystals**

We hypothesized that protein crystals loaded with PEG-RGD could be used as an unusual platform for the display of the adhesive RGD peptide to cells, thereby enabling cell attachment and spreading. Although the RGD is not covalently bound to the protein crystals, the force required to remove the guest molecules from the protein crystals is sufficiently high that the cells can attach and form adhesive contacts. This is illustrated by the images of adipose-derived stem cells on protein crystals. Fig. 5.8 shows confocal microscopy images of ADSCs cultured on protein crystals.



**Fig. 5.8** Confocal microscopy images of ADSCs cultured on large protein crystals (original magnification is 60 ×). Cells are cultured on either an unloaded protein crystal (left column), a protein crystal loaded with PEG(5 kDa)-RGD (middle column), or PEG(10 kDa)-RGD (right column). The top row shows merged images of the red (rhodamine phalloidin, for F-actin) and blue (DAPI for cell nuclei) channels, in the middle and bottom rows, respectively. The protein crystal exhibits some auto fluorescence, appearing also in the blue channel. Between 2 and 5 protein crystals were imaged for each condition, and representative images are shown, with between 30 and 50 cell nuclei on each crystal surface. Scale bars on all images represent 50 μm. The thickness of the cell membrane was between 7.5 to 10 nm.

The unloaded protein crystal also permits cell attachment to the surface (Fig. 5.8, left column). Although a similar number of cell nuclei are observed on this crystal as on the loaded crystals, there is less cell spreading on the unloaded crystal. Next, we sought additional evidence to support the hypothesis that cell adhesion and spreading was dependent on the RGD domain (as

opposed to the presence of PEG). Passivation of the protein crystal surface with covalent attachment of PEG results in no cell adhesion.

Cells on the PEG-RGD-loaded crystals, loaded with either PEG(5 kDa)-RGD or PEG(10 kDa)-RGD exhibit more well-developed actin stress fibers. The formation of actin stress fibers is important for multiple cellular processes, including cell migration and cellular morphogenesis. Furthermore, stress fiber formation is evidence that the adhesive contacts provided by the PEG-RGD are sufficiently strong to enable the cells to form cell-ECM-like contacts with the modified protein crystal surface that can endure the tension necessary for cell spreading. We therefore propose that the PEG-RGD-loaded porous protein crystals may be used to tune cell adhesion, by tuning the strength of the PEG-RGD binding to the protein crystal. This could be accomplished by varying the PEG length, rather than by varying the surface density of adhesion ligands.

RGD is an adhesion ligand for multiple cell-surface receptors, including integrins that permit cell attachment and mechanotransduction. This strategy could be expanded to other more specific cell adhesion peptides to prepare surfaces that selectively bind different cell phenotypes with different strengths. Adhesion ligands conjugated to shorter PEG chains would provide weaker adhesive contacts, or would prohibit mechanotransduction of some cell types, by breaking free from the surface when pulled. In contrast other adhesive ligands conjugated to longer PEG chains would provide stronger adhesive contacts. In this way the same surface could be tuned to selectively present different adhesiveness to multiple cell types.

Cells use adhesive contacts with their surrounding matrix to probe the local mechanical properties and thereby make cell fate decisions. For example, substrate or matrix mechanical properties are a key driver of mesenchymal stem cells toward adipogenic or osteogenic differentiation,<sup>[40]</sup> the transdifferentiation of vascular smooth muscle cells in arterial medial

calcification,<sup>[41]</sup> prometastatic signaling in tumors,<sup>[42]</sup> and macrophage polarization.<sup>[43]</sup> To achieve precise control over the downstream effects of adhesion ligand presentation, it is therefore essential to control the restoring force with which the surrounding matrix responds to the cell-imposed contractile forces. We propose that materials with tunable nanostructures containing non-covalently attached adhesion ligands can provide signals to adhered cells; rather than modifying the stiffness of the substrate itself, non-covalent attachments can be tuned to rupture if cells pull with too much force, thereby obviating undesirable stiffness-induced responses, such as metastatic transition, calcification, or inflammation. In principle, if the contractile force imparted by the cell on a specific ligand is greater than the force required to rupture the non-covalent bond between the ligand and the material, then the cell would be unable to attach. By decoupling the mechanotransduction from the material mechanical properties, different apparent stiffness values could be presented to different cell types through specific receptor-ligand interactions<sup>[43]</sup> using non-covalent attachments of different strengths. The presentation of ligands with tunable and cell-specific elasticity or rupture energy will offer a new dimension through which peptide ligands can be used to control cell behavior, or to simultaneously control multiple behaviors of cells that express different receptors.

## **Chapter 5 Summary**

In this study, the nanoporous surface morphology of CJ<sup>OPT</sup> protein crystal was observed, and imaged in the liquid phase via high-resolution AFM, revealing details of the nanoporous crystal surface. PEG-RGD as a cell adhesion ligand, was successfully loaded into the nanoporous protein crystals through diffusion, confirmed by the confocal microscopy results. Furthermore, ASDCs were successfully attached to the surface of nanoporous protein crystals. Confocal imaging

results also suggested that non-covalently attached PEG-RGD can alter the spreading of adipose-derived stem cells. As hypothesized, activated AFM tips that display a peptide that mimics integrin receptors (IntP) had a dramatically stronger interaction with crystal nanopores that were pre-loaded with PEG-RGD. We attributed the extra work associated with AFM tip extraction in that scenario with the energy required to extract PEG-RGD from the crystal. The optimized AFM characterization in this work quantitatively measures and analyzes the mechanical behaviors, as well as the nanoscale variations in the adhesion energy, between the interactions of PEG-RGD and nanoporous protein crystals.

## REFERENCES

1. Pierschbacher, M. D.; Ruoslahti, E., Cell Attachment Activity of Fibronectin Can Be Duplicated by Small Synthetic Fragments of the Molecule. *Nature* **1984**, *309* (5963), 30-33.
2. Massia, S. P.; Hubbell, J. A., Vascular Endothelial-Cell Adhesion and Spreading Promoted by the Peptide Redv of the Iiics Region of Plasma Fibronectin Is Mediated by Integrin Alpha-4-Beta-1. *J Biol Chem* **1992**, *267* (20), 14019-14026.
3. Graf, J.; Iwamoto, Y.; Sasaki, M.; Martin, G. R.; Kleinman, H. K.; Robey, F. A.; Yamada, Y., Identification of an amino acid sequence in laminin mediating cell attachment, chemotaxis, and receptor binding. *Cell* **1987**, *48* (6), 989-96.
4. Tashiro, K.; Sephel, G. C.; Weeks, B.; Sasaki, M.; Martin, G. R.; Kleinman, H. K.; Yamada, Y., A synthetic peptide containing the IKVAV sequence from the A chain of laminin mediates cell attachment, migration, and neurite outgrowth. *J Biol Chem* **1989**, *264* (27), 16174-82.
5. Nomizu, M.; Kim, W. H.; Yamamura, K.; Utani, A.; Song, S. Y.; Otaka, A.; Roller, P. P.; Kleinman, H. K.; Yamada, Y., Identification of cell binding sites in the laminin alpha 1 chain carboxyl-terminal globular domain by systematic screening of synthetic peptides. *J Biol Chem* **1995**, *270* (35), 20583-90.
6. Malinada, K. M.; Nomizu, M.; Chung, M.; Delgado, M.; Kuratomi, Y.; Yamada, Y.; Kleinman, H. K.; Ponce, M. L., Identification of laminin  $\alpha$ 1 and  $\beta$ 1 chain peptides active for endothelial cell adhesion, tube formation, and aortic sprouting. *The FASEB Journal* **1999**, *13* (1), 53-62.
7. Chelberg, M. K.; Mccarthy, J. B.; Skubitz, A. P. N.; Furcht, L. T.; Tsilibary, E. C., Characterization of a Synthetic Peptide from Type-Iv Collagen That Promotes Melanoma Cell-Adhesion, Spreading, and Motility. *J Cell Biol* **1990**, *111* (1), 261-270.
8. Staatz, W. D.; Fok, K. F.; Zutter, M. M.; Adams, S. P.; Rodriguez, B. A.; Santoro, S. A., Identification of a tetrapeptide recognition sequence for the alpha 2 beta 1 integrin in collagen. *J Biol Chem* **1991**, *266* (12), 7363-7.
9. Huettner, N.; Dargaville, T. R.; Forget, A., Discovering Cell-Adhesion Peptides in Tissue Engineering: Beyond RGD. *Trends Biotechnol* **2018**, *36* (4), 372-383.
10. Finney, A. C.; Stokes, K. Y.; Pattillo, C. B.; Orr, A. W., Integrin signaling in atherosclerosis. *Cell Mol Life Sci* **2017**, *74* (12), 2263-2282.
11. Wagh, K.; Ishikawa, M.; Garcia, D. A.; Stavreva, D. A.; Upadhyaya, A.; Hager, G. L., Mechanical Regulation of Transcription: Recent Advances. *Trends Cell Biol* **2021**, *31* (6), 457-472.
12. Kipper, M. J.; Kleinman, H. K.; Wang, F. W., Covalent surface chemistry gradients for presenting bioactive peptides. *Anal Biochem* **2007**, *363* (2), 175-84.
13. Kipper, M. J.; Kleinman, H. K.; Wang, F. W., New method for modeling connective-tissue cell migration: improved accuracy on motility parameters. *Biophys J* **2007**, *93* (5), 1797-808.
14. Vega, S. L.; Kwon, M. Y.; Song, K. H.; Wang, C.; Mauck, R. L.; Han, L.; Burdick, J. A., Combinatorial hydrogels with biochemical gradients for screening 3D cellular microenvironments. *Nat Commun* **2018**, *9* (1), 614.
15. Karimi, F.; O'Connor, A. J.; Qiao, G. G.; Heath, D. E., Integrin Clustering Matters: A Review of Biomaterials Functionalized with Multivalent Integrin-Binding Ligands to Improve

Cell Adhesion, Migration, Differentiation, Angiogenesis, and Biomedical Device Integration. *Adv Healthc Mater* **2018**, *7* (12), e1701324.

16. Zhao, X. L.; Jin, L. L.; Shi, H. F.; Tong, W. J.; Gorin, D.; Kotelevtsev, Y.; Mao, Z. W., Recent advances of designing dynamic surfaces to regulate cell adhesion. *Colloid Interfac Sci* **2020**, *35*, 100249.

17. Grewal, M. G.; Gray, V. P.; Letteri, R. A.; Highley, C. B., User-defined, temporal presentation of bioactive molecules on hydrogel substrates using supramolecular coiled coil complexes. *Biomater Sci-Uk* **2021**.

18. Kowalski, A. E.; Huber, T. R.; Ni, T. W.; Hartje, L. F.; Appel, K. L.; Yost, J. W.; Ackerson, C. J.; Snow, C. D., Gold nanoparticle capture within protein crystal scaffolds. *Nanoscale* **2016**, *8* (25), 12693-12696.

19. Huber, T. R.; Hartje, L. F.; McPherson, E. C.; Kowalski, A. E.; Snow, C. D., Programmed Assembly of Host–Guest Protein Crystals. *Small* **2017**, *13* (7), 1602703.

20. Wang, D.; Stuart, J. D.; Jones, A. A.; Snow, C. D.; Kipper, M. J., Measuring Interactions of DNA with Nanoporous Protein Crystals by Atomic Force Microscopy. *Nanoscale* **2021**, *13* (24), 10871-10881.

21. Huber, T. R.; McPherson, E. C.; Keating, C. E.; Snow, C. D., Installing Guest Molecules at Specific Sites within Scaffold Protein Crystals. *Bioconjugate Chem.* **2017**, *29* (1), 17-22.

22. Hartje, L. F.; Bui, H. T.; Andales, D. A.; James, S. P.; Huber, T. R.; Snow, C. D., Characterizing the Cytocompatibility of Various Cross-Linking Chemistries for the Production of Biostable Large-Pore Protein Crystal Materials. *Acs Biomater Sci Eng* **2018**, *4* (3), 826-831.

23. Hartje, L. F.; Munsky, B.; Ni, T. W.; Ackerson, C. J.; Snow, C. D., Adsorption-Coupled Diffusion of Gold Nanoclusters within a Large-Pore Protein Crystal Scaffold. *J. Phys. Chem. B* **2017**, *121* (32), 7652-7659.

24. Consortium, T. U., UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **2021**, *49* (D1), D480-D489.

25. Pasqualini, R.; Koivunen, E.; Ruoslahti, E., A peptide isolated from phage display libraries is a structural and functional mimic of an RGD-binding site on integrins. *J. Cell Biol.* **1995**, *130* (5), 1189-1196.

26. Ruoslahti, E., RGD and other recognition sequences for integrins. *Annu. Rev. Cell Dev. Biol.* **1996**, *12* (1), 697-715.

27. Bergkvist, M.; Cady, N. C., Chemical functionalization and bioconjugation strategies for atomic force microscope cantilevers. In *Bioconjugation Protocols*, Springer: 2011; pp 381-400.

28. Williams, E. H.; Davydov, A. V.; Motayed, A.; Sundaresan, S. G.; Bocchini, P.; Richter, L. J.; Stan, G.; Steffens, K.; Zangmeister, R.; Schreifels, J. A., Immobilization of streptavidin on 4H–SiC for biosensor development. *Appl. Surf. Sci.* **2012**, *258* (16), 6056-6063.

29. Merkel, R.; Nassoy, P.; Leung, A.; Ritchie, K.; Evans, E., Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature* **1999**, *397* (6714), 50-53.

30. Florin, E.-L.; Moy, V. T.; Gaub, H. E., Adhesion forces between individual ligand-receptor pairs. *Science* **1994**, *264* (5157), 415-417.

31. Moy, V. T.; Florin, E.-L.; Gaub, H. E., Intermolecular forces and energies between ligands and receptors. *Science* **1994**, *266* (5183), 257-259.

32. Hinterdorfer, P.; Baumgartner, W.; Gruber, H. J.; Schilcher, K.; Schindler, H., Detection and localization of individual antibody-antigen recognition events by atomic force microscopy. *Proceedings of the National Academy of Sciences* **1996**, *93* (8), 3477-3481.

33. Lo, Y.-s.; Zhu, Y.-J.; Beebe, T. P., Loading-rate dependence of individual ligand– receptor bond-rupture forces studied by atomic force microscopy. *Langmuir* **2001**, *17* (12), 3741-3748.
34. Lee, C.-K.; Wang, Y.-M.; Huang, L.-S.; Lin, S., Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein–ligand interaction. *Micron* **2007**, *38* (5), 446-461.
35. Chilkoti, A.; Boland, T.; Ratner, B. D.; Stayton, P. S., The relationship between ligand-binding thermodynamics and protein-ligand interaction forces measured by atomic force microscopy. *Biophysical journal* **1995**, *69* (5), 2125-2130.
36. Stan, G.; DelRio, F. W.; MacCuspie, R. I.; Cook, R. F., Nanomechanical properties of polyethylene glycol brushes on gold substrates. *The J. Phy. Chem. B* **2012**, *116* (10), 3138-3147.
37. Crist, B. V.; Crisst, D. B., *Handbook of monochromatic XPS spectra*. Wiley New York: 2000; Vol. 1.
38. Crist, B. V., A review of XPS data-banks. *XPS Reports* **2007**, *1* (1).
39. Monitor, P. T., *Introduction to Bruker's ScanAsyst and PeakForce Tapping AFM Technology*. Bruker application note. Bruker Nano Inc.: Bruker application note. Bruker Nano Inc.
40. Han, S. B.; Kim, J. K.; Lee, G.; Kim, D. H., Mechanical Properties of Materials for Stem Cell Differentiation. *Adv Biosyst* **2020**, *4* (11), e2000247.
41. Van den Bergh, G.; Opdebeeck, B.; D'Haese, P. C.; Verhulst, A., The Vicious Cycle of Arterial Stiffness and Arterial Media Calcification. *Trends Mol Med* **2019**, *25* (12), 1133-1146.
42. Bauer, J.; Emon, M. A. B.; Staudacher, J. J.; Thomas, A. L.; Zessner-Spitzenberg, J.; Mancinelli, G.; Krett, N.; Saif, M. T.; Jung, B., Increased stiffness of the tumor microenvironment in colon cancer stimulates cancer associated fibroblast-mediated prometastatic activin A signaling. *Scientific Reports* **2020**, *10* (1), 1-11.
43. Liu, X. Q.; Chen, X. T.; Liu, Z. Z.; Gu, S. S.; He, L. J.; Wang, K. P.; Tang, R. Z., Biomimetic Matrix Stiffness Modulates Hepatocellular Carcinoma Malignant Phenotypes and Macrophage Polarization through Multiple Modes of Mechanical Feedbacks. *Acs Biomater Sci Eng* **2020**, *6* (7), 3994-4004.

## CHAPTER 6: SUMMARY

Through this dissertation, the interaction forces and energies for guest molecules threaded into the pores of protein crystals are quantified via nano-mechanical AFM pulling experiments, as well as the dynamics of molecules-nanopore interactions via fluorescence tracking for guest molecules confined within host nanopores. This dissertation demonstrates the applications for guest molecules-entrapping porous crystals in three areas. First, it quantifies the stability of guest molecules in protein crystal pores. And it modulates cell traction forces by coupling adhesion ligands to guest molecules threaded into protein crystal pores. Finally, the research demonstrates the adsorption-enhanced and confinement-enhanced fluorogenic detection of target guest molecules.

In sum, for the first time, the porous surface structure of an unusual type of protein crystal was observed, and imaged in the liquid phase via high resolution AFM, revealing high-resolution details of porous crystal surfaces. The optimized AFM characterization then quantitatively measures and analyzes the mechanical behaviors, and the nano-scale variations in the adhesion energy between protein crystals and guest biomolecules including DNA, PEG-RGD complex, advance our understanding of porous protein crystals as capable and reliable containers of guest biomolecules, and provide a basis for engineering functional biomaterials that are responsive to small pulling forces. This project applies and confirms that machine learning methods are able to accurately classify thousands of random unknown F-D curves obtained from AFM, and discover hidden patterns of F-D curves by clustering. This project also suggests that non-covalently attached guest molecules to the porous protein crystals can alter the spreading of adipose-derived stem cells, enable us to develop instructive materials of driving cell fate decisions.

APPENDIX A: CHAPTER 3 SUPPORTING INFORMATION

**Table A1** Time scaled fluorescence intensity of the DNA inside and outside the crystal

Integrated Density/Area	Solution Fluorescence	Internal Crystal Fluorescence
Mean. (AU, time = 0)	980.896 / 979.894 / 974.104 / 966.443 / 972.423 ( <i>avg. mean = 974.752</i> )	989.490 / 989.684 / 997.717 / 983.990 / 977.960 ( <i>avg. mean = 987.769</i> )
Min. (AU, time = 0)	931 / 903 / 918 / <b>753</b> / 922	957 / 940 / 945 / 955 / <b>839</b>
Max. (AU, time = 0)	<b>1083</b> / 1055 / 1029 / 1039 / 1030	1015 / 1011 / <b>1083</b> / 1017 / 1029
Mean. (AU, time = end)	268.603 / 239.252 / 251.488 / 238.022 / 264.273 ( <i>avg. mean = 252.328</i> )	2852.836 / 2783.137 / 2944.677 / 2337.633 / 2797.874 ( <i>avg. mean = 2743.232</i> )
Min. (AU, time = end)	167 / 168 / <b>151</b> / 166 / 164	1979 / 2052 / 2059 / <b>1602</b> / 1755
Max. (AU, time = end)	<b>415</b> / 370 / 414 / 389 / 393	3805 / 3558 / 3791 / 3400 / <b>4065</b>

known:  $V_{\text{crystal}} = 0.0237 \mu\text{L}$   
 $V_{\text{unit cell}} = 1.413 \times 10^{-15} \mu\text{L}$   
 $V_{\text{solution}} = 100 \mu\text{L}$   
 $[\text{DNA}]_{\text{solution, time 0}} = 50.0 \mu\text{M}$

For dilute solutions,  $I = k\epsilon lc$  ( $I$  = intensity of fluorescence,  $k$  = constant number,  $\epsilon$  = the molar attenuation coefficient,  $l$  = the optical path length,  $c$  = concentration), since  $\epsilon$ , and  $l$  are known numbers, and  $I \neq 0$  in this case, therefore as a result,  $\frac{I_1}{I_2} = \frac{c_1}{c_2}$ .

we can learn:  $[\text{DNA}]_{\text{avg. solution, time end}} = 12.93 \mu\text{M}$   
 $[\text{DNA}]_{\text{avg. intra-crystal, time end}} = 138.86 \mu\text{M}$

This calculation of concentration ratio relies on the same linear relationship between concentration and fluorescence intensity that underlies a traditional fluorescence standards curve. Notably, it is likely not quantitatively accurate to attempt to use a traditional fluorescence standard curve to directly convert fluorescence intensity from the confocal microscope images into concentration due to inconstant optical effects (e.g. out of plane excitation)

**Table. A2 (a)** The min, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles, and max of the adhesion energy for 20 pore center pixels and 20 wall pixels for each of the experiments detected in Fig. 3.3, **(b)** adhesion energy in the pores, normalized by the interacting pore area (adhesion energy divided by average area computed from the depth of penetration of the AFM tip into the pore) for 20 pore center pixels, and **(c)** detailed penetration depth data correspond to the Fig. 3.10 (d). The conditions labeled A-E here correspond to the conditions described in the Fig. 3.3.

**(a) Adhesion energy (fJ)**

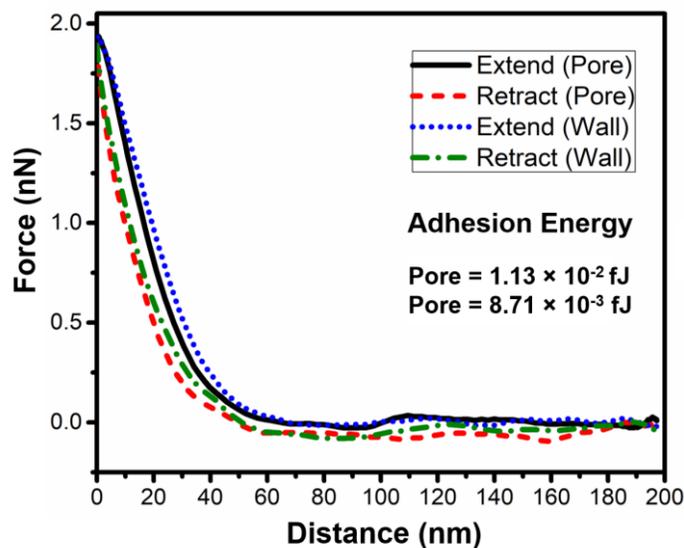
Condition	Pores					Walls				
	min	25 %-ile	50 %-ile	75 %-ile	max	min	25 %-ile	50 %-ile	75 %-ile	max
<b>A</b>	$4.55 \times 10^{-3}$	$7.88 \times 10^{-3}$	$9.23 \times 10^{-3}$	$1.17 \times 10^{-2}$	$1.57 \times 10^{-2}$	$3.02 \times 10^{-3}$	$3.52 \times 10^{-3}$	$3.93 \times 10^{-3}$	$4.29 \times 10^{-3}$	$5.43 \times 10^{-3}$
<b>B</b>	$3.50 \times 10^{-2}$	$4.49 \times 10^{-2}$	$5.02 \times 10^{-2}$	$5.62 \times 10^{-2}$	$8.08 \times 10^{-2}$	$3.91 \times 10^{-3}$	$1.16 \times 10^{-2}$	$1.39 \times 10^{-2}$	$1.69 \times 10^{-2}$	$2.99 \times 10^{-2}$
<b>C</b>	$1.54 \times 10^{-2}$	$1.96 \times 10^{-2}$	$2.31 \times 10^{-2}$	$2.60 \times 10^{-2}$	$3.16 \times 10^{-2}$	$7.38 \times 10^{-3}$	$8.78 \times 10^{-3}$	$1.09 \times 10^{-2}$	$1.35 \times 10^{-2}$	$1.95 \times 10^{-2}$
<b>D</b>	$3.69 \times 10^{-3}$	$4.02 \times 10^{-3}$	$4.50 \times 10^{-3}$	$5.49 \times 10^{-3}$	$9.55 \times 10^{-3}$	$3.37 \times 10^{-3}$	$3.84 \times 10^{-3}$	$4.51 \times 10^{-3}$	$5.00 \times 10^{-3}$	$7.54 \times 10^{-3}$
<b>E</b>	$1.15 \times 10^{-2}$	$1.46 \times 10^{-2}$	$1.70 \times 10^{-2}$	$1.81 \times 10^{-2}$	$2.59 \times 10^{-2}$	$4.20 \times 10^{-3}$	$5.63 \times 10^{-3}$	$6.64 \times 10^{-3}$	$8.56 \times 10^{-3}$	$1.23 \times 10^{-2}$

**(b) Normalized Adhesion energy ( $J \cdot m^{-2}$ )**

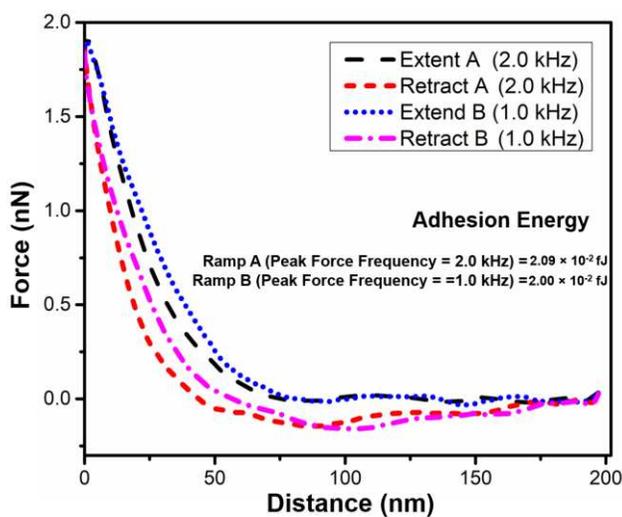
Condition	25 %-ile	50 %-ile	75 %-ile
<b>A</b>	$1.46 \times 10^{-2}$	$1.71 \times 10^{-2}$	$2.16 \times 10^{-2}$
<b>B</b>	$4.80 \times 10^{-2}$	$5.37 \times 10^{-2}$	$6.02 \times 10^{-2}$
<b>C</b>	$4.86 \times 10^{-2}$	$5.72 \times 10^{-2}$	$6.42 \times 10^{-2}$
<b>D</b>	$1.65 \times 10^{-2}$	$1.85 \times 10^{-2}$	$2.26 \times 10^{-2}$
<b>E</b>	$6.14 \times 10^{-2}$	$7.19 \times 10^{-2}$	$7.63 \times 10^{-2}$

**(c) Penetration Depth of Probing**

Condition	Penetration Depth (nm)
<b>A</b>	$13.2 \pm 4.1$ (n = 579)
<b>B</b>	$22.9 \pm 7.3$ (n = 335)
<b>C</b>	$9.9 \pm 4.4$ (n = 164)
<b>D</b>	$6.0 \pm 2.4$ (n = 324)
<b>E</b>	$5.8 \pm 2.4$ (n = 173)



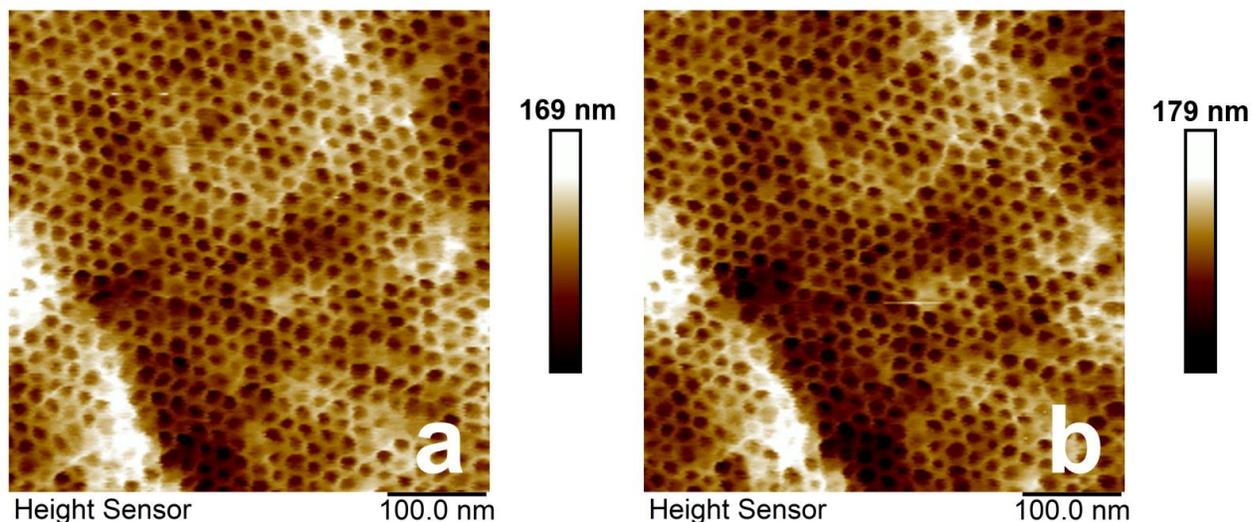
**Fig. A1** One randomly selected set of F-D curves, using an inactivated AFM tip on a DNA loaded crystal, obtained from both pore and wall pixels.



**Fig. A2** Two sets of ramping F-D curves in different peak force frequency, respectively 1.0 kHz and 2.0 kHz, using an activated AFM tip on a DNA loaded crystal. The curves were from the same randomly selected pixel, in addition to the peak force frequency, all other imaging parameters were the same.

**Table A3** Selected imaging parameters related to the Fig. A2

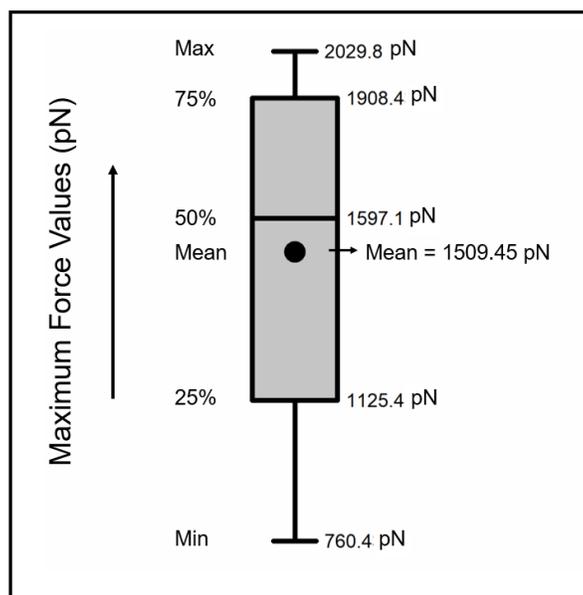
Feedback		Peak Force Tapping Control	
Peak Force Setpoint	2.0 nN	Peak Force Amplitude	100 nm
		Peak Force Frequency	(A) 2.0 kHz / (B) 1.0 kHz



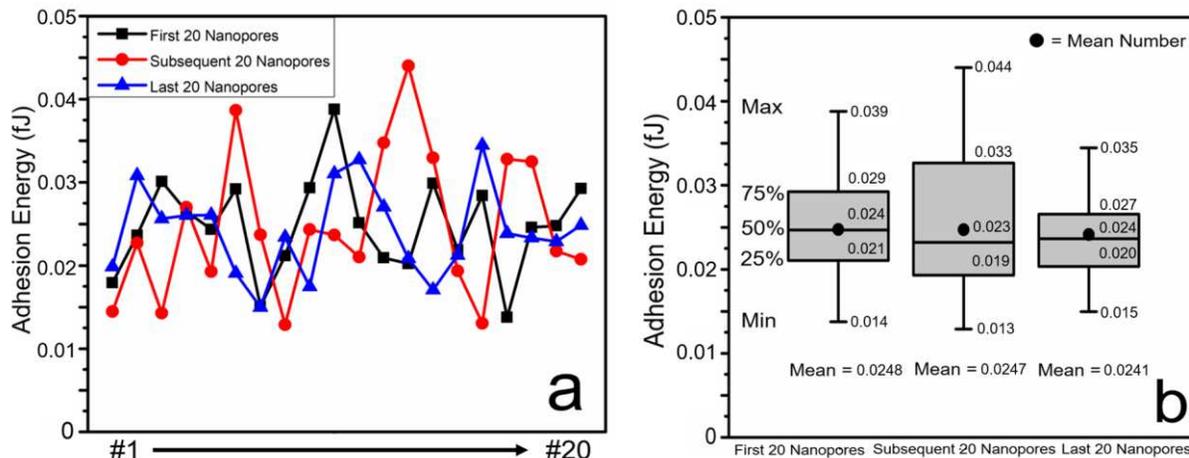
**Fig. A3** AFM images of a crystal loaded with DNA imaged using an activated AFM tip in a scan rate of (a) 1.0 Hz and (b) 2.0 Hz, respectively. In addition to the scan rate, all other imaging parameters were the same.

**Table A4** Selected imaging parameters related to the Fig. A3 (a-b)

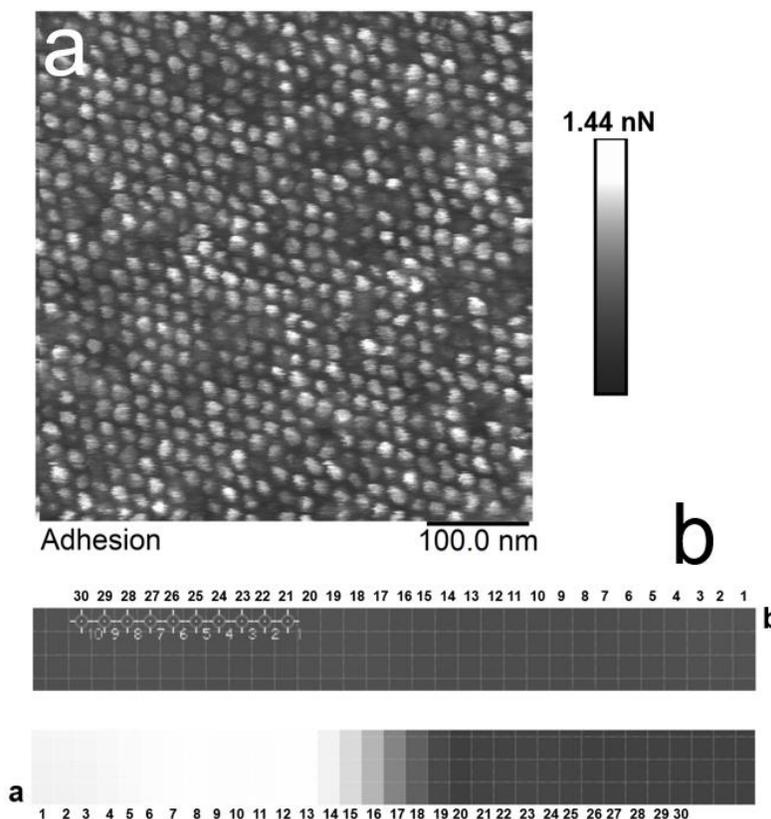
Scan		Feedback		Peak Force Tapping Control	
Scan Size	500 × 500 nm	Peak Force Setpoint	2.0 nN	Peak Force Amplitude	100 nm
Scan Rate	(a) 1.0 / (b) 2.0 Hz			Peak Force Frequency	1.0 kHz
Scanning Lines	512				



**Fig. A4** Box plot (min, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles, mean, and max) for the distribution of maximum force values between DNA and protein crystals, in the pore areas, under scenario B and E of Fig. 3.3 and Fig. A8.



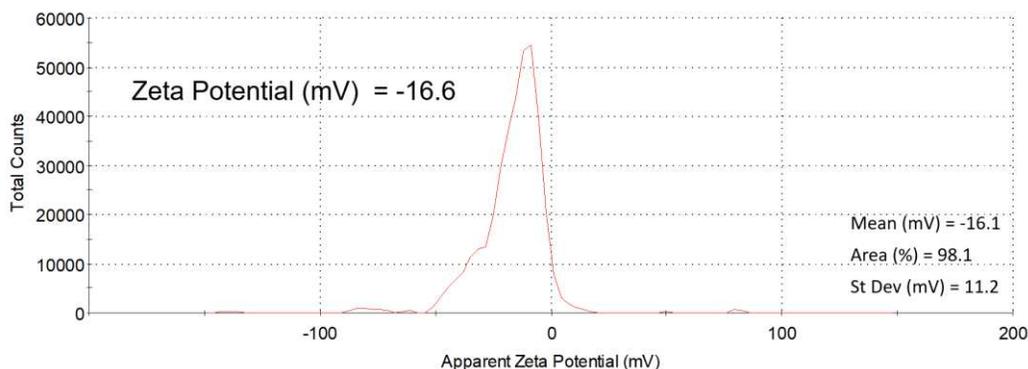
**Fig. A5 (a)** Line charts and **(b)** box plot (min, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles, mean, and max) for the pore area adhesion energy of the first, the subsequent, and the last 20 nanopores, during one AFM imaging, from activated AFM tips with DNA loaded crystal.



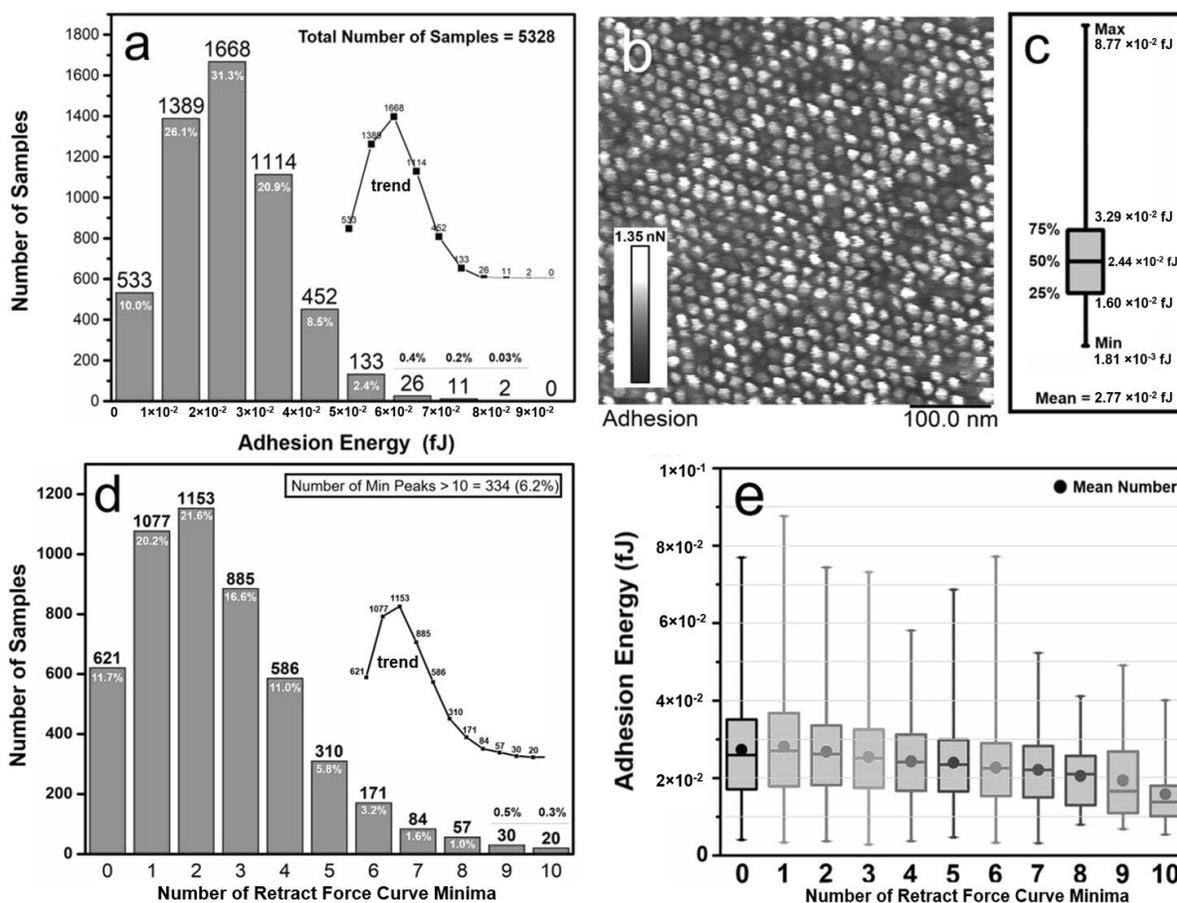
**Fig. A6 (a)** an adhesion energy map for activated AFM tip with DNA loaded protein crystal, **(b)** the very first 30 scanned pixels (from a1 to a30) of imaging, and after 16 minutes, the very last 30 scanned pixels (from b30 to b1) of imaging, from the adhesion energy map Fig. A6 (a). The dark pixels locate at the wall areas while the bright pixels locate at the pore areas. Along with the corresponding grayscale data, the results indicate that during that 16 mins, the ability of attachment for activated tips was stable and consistent.

**Table A5** The grayscale of very first 30 scanned pixels (from a1 to a30) of imaging, and after 16 minutes, the very last 30 scanned pixels (from b30 to b1) of corresponding to the Fig. A6 (b), from the adhesion energy map Fig. A6 (a).

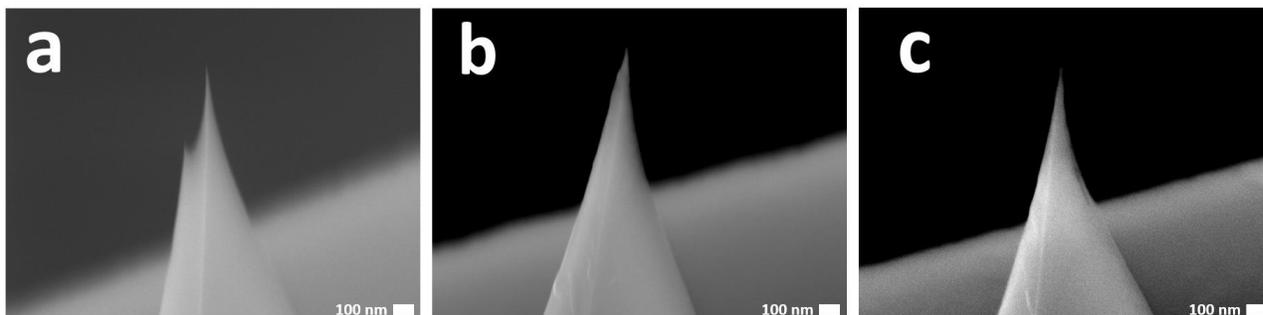
Grayscale (K)		Grayscale (K)	
Pixel	16-bit (0-32768)	Pixel	16-bit (0-32768)
<b>a1</b>	30435	<b>b1</b>	8120
<b>a2</b>	30600	<b>b2</b>	8120
<b>a3</b>	30765	<b>b3</b>	8007
<b>a4</b>	31262	<b>b4</b>	7780
<b>a5</b>	31762	<b>b5</b>	7555
<b>a6</b>	32263	<b>b6</b>	7442
<b>a7</b>	32432	<b>b7</b>	7332
<b>a8</b>	32600	<b>b8</b>	7220
<b>a9</b>	32432	<b>b9</b>	7332
<b>a10</b>	32432	<b>b10</b>	7332
<b>a11</b>	32432	<b>b11</b>	7332
<b>a12</b>	32600	<b>b12</b>	7332
<b>a13</b>	32600	<b>b13</b>	7220
<b>a14</b>	30600	<b>b14</b>	7332
<b>a15</b>	26708	<b>b15</b>	7442
<b>a16</b>	21010	<b>b16</b>	7555
<b>a17</b>	14086	<b>b17</b>	7893
<b>a18</b>	9526	<b>b18</b>	8007
<b>a19</b>	6890	<b>b19</b>	8007
<b>a20</b>	5820	<b>b20</b>	7893
<b>a21</b>	6242	<b>b21</b>	7780
<b>a22</b>	6455	<b>b22</b>	7666
<b>a23</b>	6564	<b>b23</b>	7666
<b>a24</b>	6564	<b>b24</b>	7666
<b>a25</b>	6349	<b>b25</b>	7893
<b>a26</b>	6135	<b>b26</b>	8120
<b>a27</b>	6135	<b>b27</b>	8236
<b>a28</b>	6135	<b>b28</b>	8236
<b>a29</b>	6135	<b>b29</b>	8236
<b>a30</b>	6135	<b>b30</b>	8120



**Fig. A7** Zeta potential distribution on the surface of crosslinked CJ protein crystals



**Fig. A8** (a) Distribution of adhesion energy ranges, (b) a map of adhesion energy corresponding to Fig 3.8 (a), along with the corresponding grayscale data for the very initial and very last scanned pixels (Supporting Information Fig. A6 and Table A5), the results can also indicate that during that 16 mins of imaging, the ability of attachment for activated AFM tips was stable and consistent, and (c) box plot (min, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles, and max) of the adhesion energy, for 5328 pore center pixels (from 5328 individual pores) with activated AFM tip, on the crystal loaded with DNA, and (d) distribution of numbers of minimum peak(s) (as labeled on Fig. 3.9 (a)) on each individual retract force curve, (e) box plot (min, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles, mean, and max) of the adhesion energy distribution, corresponding to the number of minimum peak(s) among retract force curves in Fig. A8 (d).



**Fig. A9** The field emission scanning electron microscope imaging (FESEM,  $\times 50000$ ) of (a) unmodified ScanAsyst Fluid+ tip, (b) activated AFM tip without DNA, (c) modified AFM tip with 30mer DNA.

## APPENDIX B: CHAPTER 4 SUPPORTING INFORMATION

### B.1. Experimental

#### B.1.1 Materials

A Millipore Synthesis water purification unit was used to obtain 18.2 M $\Omega$  cm water, used for making all aqueous solutions. 3-(Trimethoxysilyl)propylamine (APTMS) for molecular vapor deposition (MVD) was purchased from EMD Millipore Corp. Traut's reagent (2-iminothiolane) used for tip modification was purchased from Chem-Impex International, Inc. 5,5'-Dithiobis-(2-nitrobenzoic acid), (DTNB, Ellman's reagent) were purchased from Thermo Fisher Scientific, Inc.

##### B.1.1.1 CJ Crystal's Growth

As described in previous work.<sup>[B1]</sup> *Campylobacter jejuni* protein (CJ) was cloned into pSB3 expression vector. CJ expression was performed with BL21(DE3) *Escherichia coli* cells in Terrific broth. Induction was performed with 0.4 mM IPTG for 16 hours at 25 °C, followed by purification using immobilized metal affinity chromatography (IMAC). The target protein was then dialyzed into ammonium sulfate storage buffer (500mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 10% glycerol, pH = 7.4). CJ protein was concentrated to 15 mg/mL, aliquoted and stored at -30 °C. Then, one 24  $\mu$ L tube of purified CJ protein (15 mg/mL) was thawed.<sup>[B2-B5]</sup> The samples were kept on ice at all times. In the reservoir of a plastic CrysChem sitting-drop crystallization plate, 340  $\mu$ L of 4 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 40  $\mu$ L of 1 M bis-tris (pH = 6.5), and 20  $\mu$ L of DI H<sub>2</sub>O were mixed. Then, 1  $\mu$ L of the reservoir solution was pipetted into the top drop of the plastic sitting-drop crystallization plate. Finally, a 1  $\mu$ L aliquot of CJ protein solution was added by pipetting the protein solution directly on top of the drop of reservoir solution. Crystals typically grew to full size within 1 to 3 days.

### **B.1.1.2 CJ Crystal Crosslinking and Immobilization**

In this study, medium-large CJ protein crystals were used, with a typical diameter of 400~700  $\mu\text{m}$ , and typical height of 50  $\mu\text{m}$ . The crystals must be crosslinked so that we can later vary the solvent. Crystals were transferred (using a nylon crystallography loop, Hampton Research) from their growth well into a drop of 4.2 M trimethylamine *N*-oxide (TMAO), 0.175 M  $\text{H}_2\text{SO}_4$  at pH = 7.5, to wash for a minimum of 20 min. A drop of 390  $\mu\text{L}$  of 4.2 M TMAO, 0.175 M  $\text{H}_2\text{SO}_4$  at pH 7.5 plus 10  $\mu\text{L}$  of 40% glyoxal crosslinker was then prepared. Crystals were manually transferred into crosslinking solution and covered for 2 h. Meanwhile, a drop of 380  $\mu\text{L}$  0.1 M sodium citrate, 0.15 M NaCl at pH 5.0, 10  $\mu\text{L}$  of 50% hydroxylamine, and 10  $\mu\text{L}$  of 40 mg/mL dimethylamine borane complex was prepared. The crosslinked crystals were transferred to this “quenching” solution to eliminate reactive groups. After 8 hours in the quenching solution, crystals were ready to be used or stored. Crystals were stored in 4.0 M TMAO and washed briefly in water or adsorption buffer (30 mM KCL, 10 mM MES, pH = 6.0) prior to use.

For AFM experiments (described below), CJ crystals were immobilized on glass-bottom petri dishes (Willco Wells) employing a UV-curable glue. The top of a crystal probe (Minitool HR4-217) was used to transfer a drop of UV-curable glue (Bondic Inc.) onto the surface of a petri dish (Ted Pella, Inc. 14025-20). The glue was gently and evenly spread on the dish surface to make the layer of glue as thin as possible. CJ crystals were transferred to the glue with a loop. Critically, the crystal was transported inside a tiny drop of buffer, such that the crystal was not desiccated. The UV-glue was viscous and did not noticeably mix with the buffer. The glue was then cured by exposing to UV-light LED (Bondic SK001) from above for 10 s. The glue cured after about 2 min, after which additional drops of buffer (typically ~5 mL) were added to the dish to prevent the crystal from drying.

The surface of crystals was weakly negatively charged after crosslinking, with a zeta potential of -16.6 mV at pH = 7.5. This modest negative potential suggests that simple electrostatic attraction is not the driving force for DNA adsorption. Empirically, the lack of DNA desorption in high salt washes (data not shown) further supports the idea that DNA binding is not dominated by electrostatic interactions that can be screened at high salt.

### **B.1.2 AFM Tip Modification**

Bruker's ScanAsyst Fluid+ tips were modified to covalently attach DNA (Fig 2). These tips have a slim shape with estimated tip radius as small as 2 nm, and a silica surface layer. To clean the tip surface, AFM tips were placed in O<sub>2</sub> plasma chamber (Plasma Etch. Inc) with a 200–300 mTorr total pressure inside the chamber, and the power setting was adjusted to 38 W for 5 min to activate the hydroxyl groups on the silica surface of the tip.<sup>[B6]</sup> Then, molecular vapor deposition (MVD) was used for amino-silane treatment of the surface of AFM tips. AFM tips were placed into a 1-L polypropylene jar. Two mL of APTMS aminosilane was added to a 10-mL scintillation vial, also placed in the polypropylene jar. The polypropylene jar was sealed using a screw cap lid, and placed in a 60 °C oven for 60 min. This allows the surface of the AFM cantilever tip to be modified with the APTMS by MVD forming a very thick aminosilane layer anchored to the surface.<sup>[B6]</sup>

Traut's reagent (2-iminothiolane) reacts spontaneously with primary amines (-NH<sub>2</sub>) at pH = 7.0 to introduce sulfhydryl (-SH) groups. We used this reaction with 1 mM Traut's reagent at room temperature in 50 mM KCl solution.<sup>[B6]</sup> To activate the AFM tips for binding thiol-terminated DNA, AFM tips were modified using dithionitrobenzoic acid chemistry. Specifically, the remaining 2-iminothiolane solution was replaced with excess 5,5'-dithiobis-(2-nitrobenzoic acid), (DTNB, Ellman's reagent, 500 μM) in a 0.1 M dipotassium phosphate and sodium

bicarbonate buffer condition (pH = 8). Activating the AFM tip with DTNB on the tip surface enables thiol-terminated DNA to be reversibly and covalently bound to the AFM tip by a disulfide bond. With the functional groups on the tip surface, the AFM tips would be capable of binding DNA in the nanopores of the CJ protein crystal surface.

Each step in the surface modification of the AFM tips was evaluated by X-ray photoelectron spectroscopy, using a Physical Electronics 5800 spectrometer (Chanhassen, MN). This XPS uses a monochromatic Al K $\alpha$  X-ray beam source ( $h\nu = 1486.6$  eV), hemispherical analyzer, and multichannel detector. The binding energy scales for the samples were referenced to the aliphatic contribution of the C1s peak at 284.8 eV. High-resolution spectra of the N1s, S2p and P2p envelopes were acquired with 0.1 eV steps, and an X-ray spot size of 800  $\mu\text{m}$ . Given this size, the XPS test is reflecting surface chemistry beyond the tip, which is of the same material. Analyses were performed at a photoelectron take-off angle of 45°. Peak fitting of the N1s and S2p envelope was performed in MultiPak (Ulvac-Phi, Inc.) using Gaussian/Lorentzian peaks and a Shirley background correction. The morphology of both unmodified and modified AFM tips was also imaged by field emission scanning electron microscopy (FESEM, JEOL JSM-6500F).

### **B.1.3 DNA Loading and AFM Probing**

We operated the AFM (Bruker Bioscope Resolve, mounted on a spinning-disc confocal microscope built around a Nikon Eclipse TiE) in quantitative nano-mechanics (QNM) PeakForce Capture mode. All images and force curves were collected using ScanAsyst Fluid+ tips (Bruker). Crystal imaging was performed in the TE (Tris-EDTA) / DI H<sub>2</sub>O buffer (5 mM Tris-HCl, 1 mM EDTA, pH = 7.5) with crystals affixed to the bottom of a glass petri dish. The AFM line scan rate was set to 1.0 Hz and the peak force tapping frequency was set to 1.0 kHz. The peak force set point was set to 2 nN. Notably, the force used here is quite large with respect to the forces used by

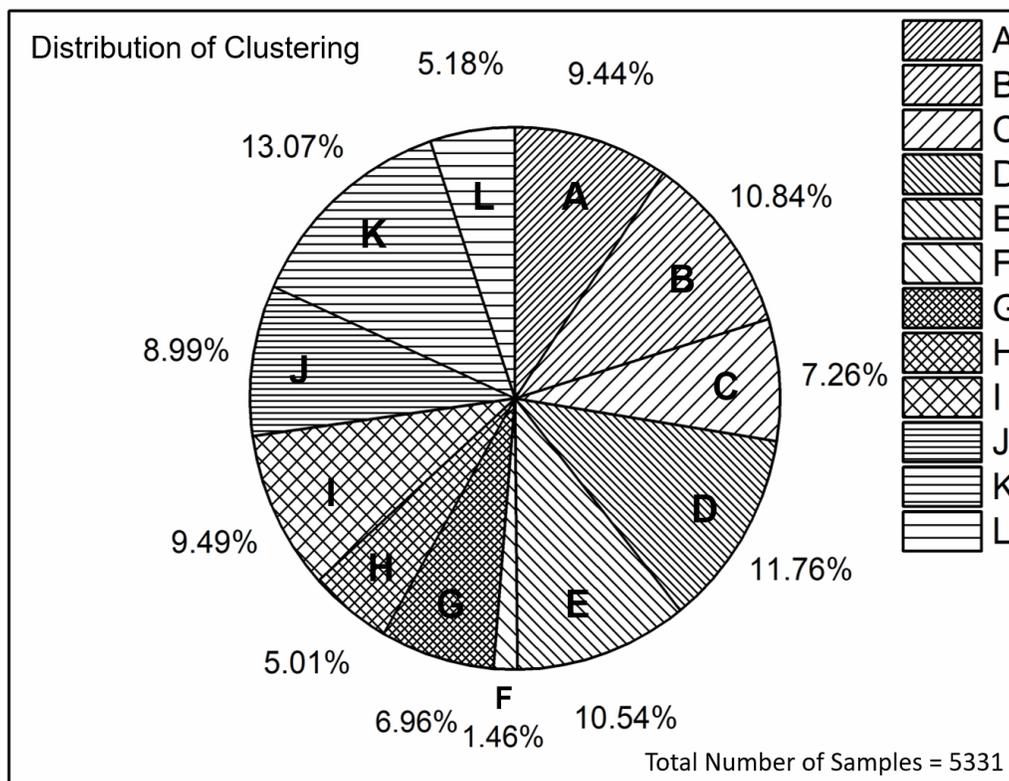
investigators who use optical trap experiments to study the interactions of DNA molecules with other partners. Specifically, Dario Anselmetti et al. used peak force of  $\sim 800$  pN, ) while Stuart Lindsay and coworkers used a peak force of 160 pN.<sup>[B7, B8]</sup> The QNM PeakForce Capture mode is based on force-versus-distance measurements in which the tip oscillates at a frequency below the resonance frequency of cantilever,  $f_0$ . This mode provides a high-resolution peak force mapping as well as sensitivity to record nano-mechanical behaviors at high spatial resolution.<sup>[B9]</sup> Tip-sample interactions are measured with pN-resolution by the deflection of the cantilever. Analysis of the AFM data was performed in NanoScope (Bruker, Inc.), Origin (OriginLab, Inc.), Python (Version 2.7), and Matlab (Version 2019).

The protein crystal sample was imaged by both AFM (unmodified tip), and confocal microscopy to confirm immobilization and to verify that the crystal surface is clean. When crystals were incorrectly prepared, their surfaces could be obscured by aggregated protein. To ensure that the crystals were competent to uptake DNA, we used time-lapse confocal microscopy (z-stack imaging) to monitor and confirm the loading of fluorescently labeled DNA. First, the CJ crystals were photobleached to prevent interference from background fluorescence. Prior to DNA loading, as a control experiment, the protein crystal was imaged using an activated AFM tip (terminated with the dithionitrobenzoic acid, but without DNA) in the TE buffer. Then, the TE buffer solution was replaced by 100  $\mu$ L 50  $\mu$ M 30mer-DNA with two terminal thiol groups (sense strand, 5'-3': /5ThioMC6-D/TAG GCG ACT CGA CGG TCT TAC GCG TTA CGT, antisense strand, 5'-3': ACG TAA CGC GTA AGA CCG TCG AGT CGC CTA) in TE buffer. Prior to loading, a stock of the same 30mer-DNA was fluorescently labeled with TAMRA (Carboxytetramethylrhodamine labeled DNA, Integrated DNA Tech.) for 30 minutes. During loading, 10% (90%) of the DNA was TAMRA-labeled (unlabeled). Next, after washing three times with TE buffer (30 min per wash),

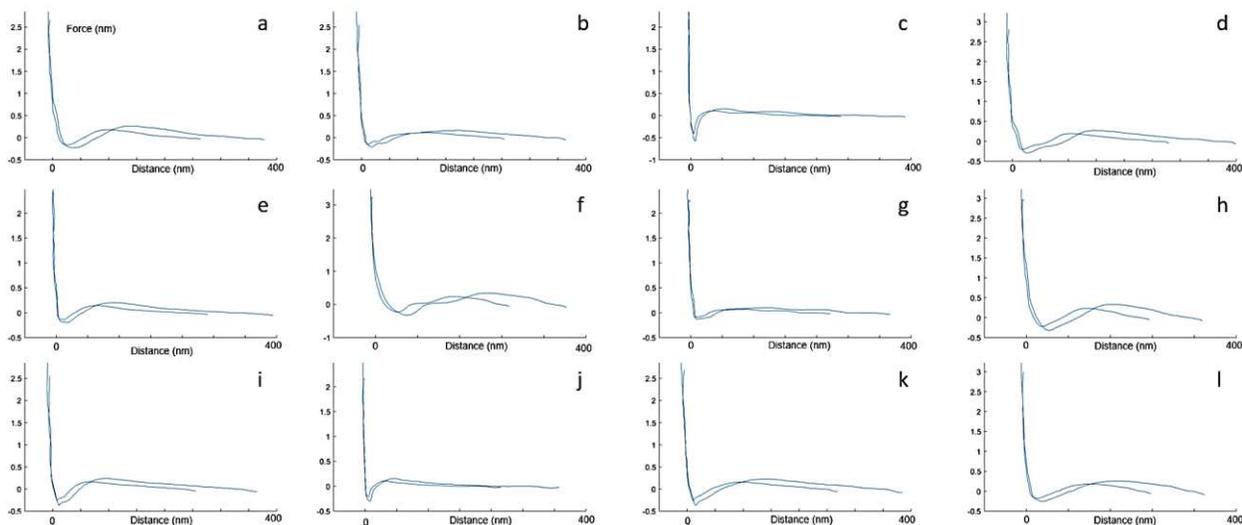
the samples were incubated with 100  $\mu$ L of 50  $\mu$ M tris(2-carboxyethyl)phosphine (TCEP) in TE buffer for 30 minutes to reduce disulfide bonds. After reduction of disulfide bonds with TCEP, the DNA-loaded crystal was again washed with 1 mL TE buffer for 30 min, three times, to remove the TCEP. Retention of the DNA was confirmed by the confocal microscope imaging after each wash step. The DNA-loaded crystal was then imaged with an activated AFM tip, and F-D curves were collected at each pixel in the AFM image.

As a control experiment, the procedure for loading the protein crystal described above was repeated using DNA previously reacted with 100  $\mu$ L of 14 mM iodoacetamide (in 100 mM Tris-HCL buffer, pH = 8.3). Ideally, iodoacetamide will permanently “cap” the DNA to ensure that it cannot covalently bind to the activated AFM tip. The crystal loaded with deactivated, “capped” DNA was imaged with an activated AFM tip.

AFM imaging and F-D curve collection was conducted for five different experimental conditions as described above, using combinations of un-modified, activated, and DNA-modified tips and either loaded or unloaded protein crystals. The five experimental conditions are: (A) un-modified AFM tip on an unloaded crystal, (B) activated AFM tip on a crystal loaded with thiol-bearing DNA, (C) activated AFM tip on a DNA-loaded crystal in the presence of TCEP and iodoacetamide (D) activated AFM tip on an unloaded crystal, and (E) 30-mer DNA-modified AFM tip on an unloaded crystal. From each AFM image, F-D curves were manually assigned to one of two classes, corresponding to protein crystal surface features: pores and walls.



**Fig. B1** Distribution of labels corresponding to the cluster centers



**Fig. B2** The raw F-D curves of cluster centers from AFM without background correction or filtering, corresponding to Fig. 4.4, label a-l in Fig. B2 are corresponding to Fig. 4.4.

## REFERENCES OF APPENDIX B

- B1. McPherson, A., A. Malkin, and Y. G., *Atomic force microscopy in the study of macromolecular crystal growth*. *Annu. Rev. Biophys.*, 2000. **29**(1): p. 361-410.
- B2. Ilari, A. and C. Savino, *A practical approach to protein crystallography*, in *Bioinformatics*. 2017, Springer. p. 47-78.
- B3. Wlodawer, A., et al., *Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination*. *FEBS J.*, 2013. **280**(22): p. 5705-5736.
- B4. Wlodawer, A., et al., *Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures*. *FEBS J.*, 2008. **275**(1): p. 1-21.
- B5. Kowalski, A.E., et al., *Gold nanoparticle capture within protein crystal scaffolds*. *Nanoscale*, 2016. **8**(25): p. 12693-12696.
- B6. Bergkvist, M. and N.C. Cady, *Chemical functionalization and bioconjugation strategies for atomic force microscope cantilevers*, in *Bioconjugation Protocols*. 2011, Springer. p. 381-400.
- B7. Anselmetti, D., et al., *Single molecule DNA biophysics with atomic force microscopy*. *Single Mol.*, 2000. **1**(1): p. 53-58.
- B8. Lin, L., et al., *Recognition imaging with a DNA aptamer*. *Biophys. J.*, 2006. **90**(11): p. 4236-4238.
- B9. Hedayati, M. and M.J. Kipper, *Atomic force microscopy of adsorbed proteoglycan mimetic nanoparticles: Toward new glycocalyx-mimetic model surfaces*. *Carbohydr. Polym.* , 2018. **190**: p. 346-355.

## APPENDIX C: CHAPTER 5 SUPPORTING INFORMATION

### C.1 Materials

As described in previous work for the CJ protein,<sup>[C1]</sup> the CJ<sup>OPT</sup> protein was cloned into pSB3 expression vector. CJ expression was performed with BL21(DE3) *Escherichia coli* cells in Terrific broth. This needs to be modified to describe CJ<sup>OPT</sup> Induction was performed with 0.4 mM IPTG for 16 hours at 25 °C, followed by purification using immobilized metal affinity chromatography (IMAC). The target protein was then dialyzed into ammonium sulfate storage buffer (500 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 10% glycerol, pH = 7.4). CJ<sup>OPT</sup> protein was concentrated to 15 mg/mL, aliquoted and stored at -30 °C. Then, one tube (24 µL) of purified CJ<sup>OPT</sup> protein (15 mg/mL) was thawed.<sup>[c2-c5]</sup> The samples were kept on ice at all times. In the reservoir of a plastic CrysChem sitting-drop crystallization plate, 340 µL of 4 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 40 µL of 1 M bis-tris (pH = 6.5), and 20 µL of DI H<sub>2</sub>O were mixed. Then, 1 µL of the reservoir solution was pipetted into the top drop of the plastic sitting-drop crystallization plate. Finally, a 1 µL aliquot of CJ<sup>OPT</sup> protein solution was added by pipetting the protein solution directly on top of the drop of reservoir solution. Crystals typically grew to full size within 1 to 3 days.

#### C.1.1 CJ<sup>OPT</sup> Nucleic Acid Sequence

```
ATGAAAGAATATACCCTGGATAAAGCCCATACCGATGTTGGCTTTAAAATCAAACAT
CTGCAGATTAGCAATGTGAAAGGCAACTTTAAAGATTATAGCATAGTGATCGATTTT
GATCCGGCAAGTGCAGAATTCAAAAAATTCGATATAACCATTTAAAATCGCCAGCGT
GAATACCGAAAATCAGACCCGTGATAATCATCTGCAGCAGGATGACTTCTTCAAAG
CCAAAAAATACCCGGATATGACCTTTACCATGAAAAAATACGAGAAAATCGATAAC
```

GAAAAAGGCAAAATGACCGGCACCCTGACCATTGCCGGTGTAGCAAAGATATTGT  
TCTGGATGCAGAAATTGGTGGTATGGCCAAAGGTAAAGATGGCAAAGAAAAAATTG  
GCTTTAGCCTGAACGGCAAAATCAAACGTAGCGATTTCAAATTTGCAACCAGCACCA  
GCACCATTACCCTGAGTGATGACATTAATCTGTGTTGGGAAATAAAAAGCCAACGAG  
AAAGAAGGTGGTAGTCATCACCACCACCATCACTAATAA

### **C.1.2 CJ<sup>OPT</sup> Amino Acid Sequence**

MKEYTLDKAHTDVGFKIKHLQISNVKGNFKDYSIVIDFDPASAEFKKFDITIKIASVNTEN  
QTRDNHLQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTGTLTIAGVSKDIVLDAEIG  
GMAKKGKD GKEKIGFSLNGKIKRSDFKFATSTSTITLSDDINLCWEIKANEKEGGSHHHH  
HH\*\*

\* denotes stop codon

### **C.1.3 CJ<sup>OPT</sup> Mutations**

A34I

L48F

V50I

V121M

N162C

I163W

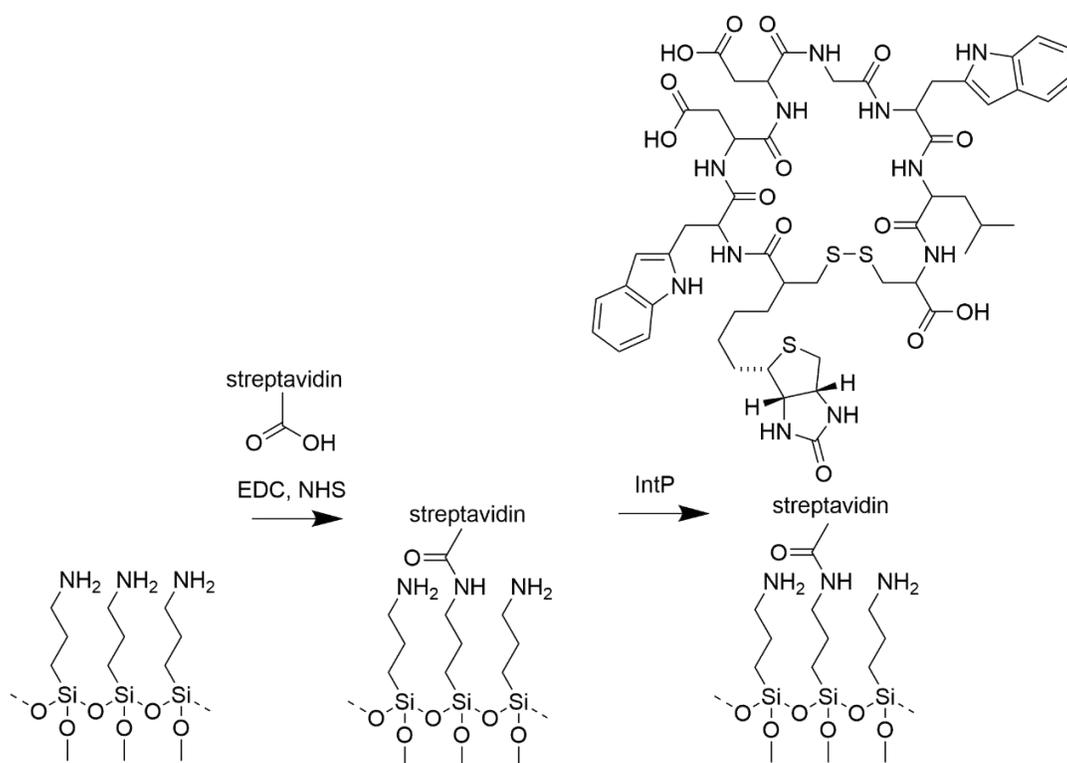
V165I

### **C.2 AFM Tip Modification**

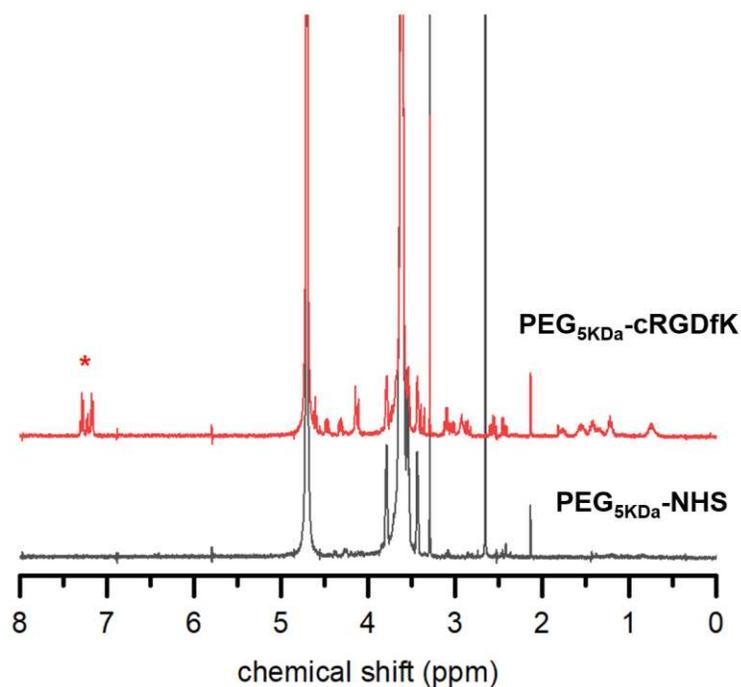
Bruker's ScanAsyst Fluid+ tips were modified to covalently attach the polyethylene glycol-arginylglycylaspartic acid (PEG-RGD) complex. These tips have a slim shape with estimated tip radius as small as 2 nm, and a silica surface layer. To clean the tip surface and to

activate the hydroxyl groups on the silica surface of the tip, AFM tips were placed in O<sub>2</sub> plasma chamber (Plasma Etch. Inc) with a 200–300 mTorr total pressure inside the chamber, and the power setting was adjusted to 38 W for 5 min.<sup>[c6]</sup> Then, molecular vapor deposition (MVD) was used for amino-silane treatment of the surface of AFM tips. AFM tips were placed into a 1-L polypropylene jar. Two mL of (3-aminopropyl)trimethoxysilane (APTMS) aminosilane was added to a 10 mL scintillation vial, also placed in the polypropylene jar. The polypropylene jar was sealed using a screw cap lid, and placed in a 60 °C oven for 15 min. This allows the surface of the AFM cantilever tip to be modified with the APTMS by MVD forming an aminosilane layer anchored to the surface.<sup>6</sup> The spring constant of a separate fully PEG-RGD modified AFM tip used to collect data was 0.96 N/m, with an estimated tip diameter of 5.94 nm (ETD, data from NanoScope).

Each step in the surface modification of the AFM tips was evaluated by X-ray photoelectron spectroscopy (XPS), using a Physical Electronics 5800 spectrometer (Chanhassen, MN). This XPS uses a monochromatic Al K $\alpha$  X-ray beam source ( $h\nu = 1486.6$  eV), hemispherical analyzer, and multichannel detector. The binding energy scales for the samples were referenced to the aliphatic contribution of the C1s peak at 284.8 eV. High-resolution spectra of the N1s and Si2s envelopes were acquired with 0.1 eV steps, and an X-ray spot size of 800  $\mu$ m. Given this size, the XPS spectra report surface chemistry of the tip, cantilever, and probe, but the tip is composed of the same material. Analyses were performed at a photoelectron take-off angle of 45°. Peak fitting of the N1s and Si2s envelope was performed in MultiPak (Ulvac-Phi, Inc.) using Gaussian/Lorentzian peaks and a Shirley background correction.



**Fig. C1** Schematic illustration of intended AFM tip modification process



**Fig. C2** A comparison of proton nuclear magnetic resonance (<sup>1</sup>H NMR) spectra between PEG<sub>5kDa</sub>-cRGD and PEG<sub>5kDa</sub>-NHS. The peak appeared around 7.3 ppm indicates the successful conjugation of benzyl group in RGD to the PEG molecule.

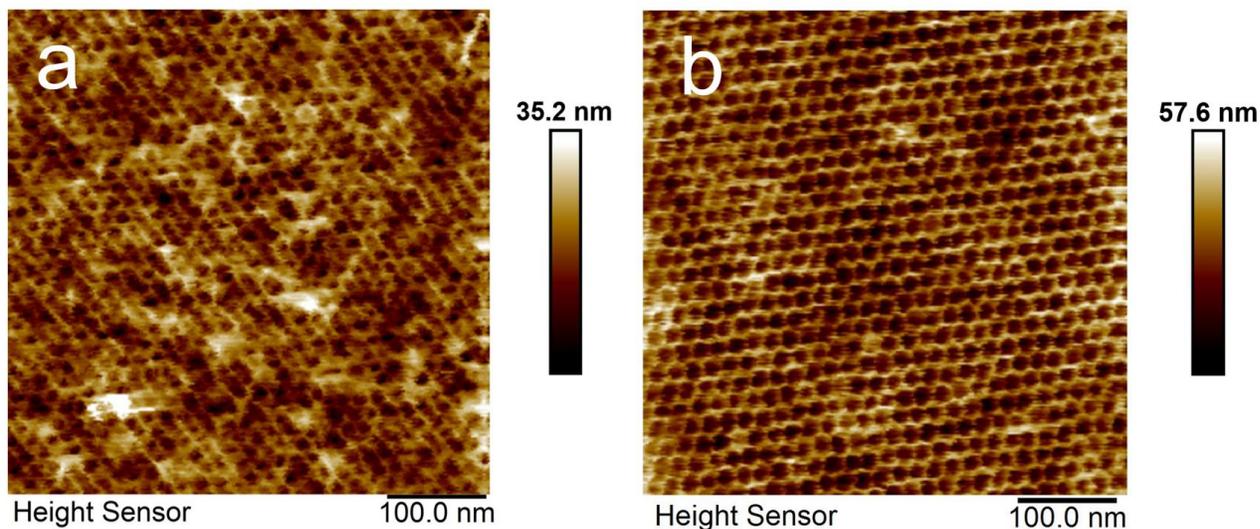
### C.3 Cell Culture Experiments

For the cell study, protein crystals were immobilized on 12 mm siliconized glass cover slides employing a UV-curable glue as described previously.<sup>[C1]</sup> Briefly, the glue was gently spread on the glass surface to make a very thin layer of glue. Protein crystals were transferred to the glass surface and pushed towards the glue with a loop and then the glue was cured by exposing to UV-light LED for 30 s. During this process, careful attention was given to make sure the surface of the crystals does not get contaminated by the glue. Protein crystals glued on the glass slides were then transferred to 24-well plate to sterilize by incubating with 70 % ethanol for 30 min, followed by three rinses and incubation with sterile PBS for another 30 min. Human adipose-derived stem cells (ADSCs) isolated from adipose tissue and at passage three were obtained from Dr. Kimberly Cox-York's laboratory at Colorado State University. The protocol for ADSC isolation from healthy individuals was approved by Colorado State University Institutional Review Board. The cells were cultured at 37 °C and 5% CO<sub>2</sub> in growth media composed of  $\alpha$ -MEM Media (HyClone™) with 10 % (v/v) Fetal Bovine Serum (FBS) and 1% (v/v) penicillin/streptomycin. The cells were seeded on the protein crystals at a final concentration of  $2.0 \times 10^4$  cells mL<sup>-1</sup>. 1 ml of cells in media was added to each well where the glued protein crystals on the glass is located. Cell culture media was removed and replaced with fresh media every other day.

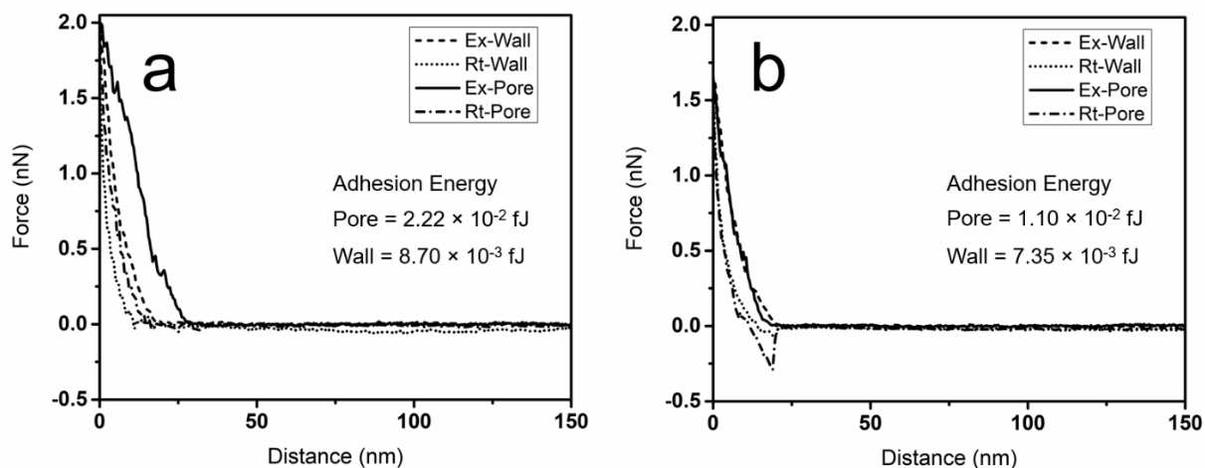
The cell adhesion and proliferation on the surfaces were characterized using fluorescence microscopy. After 4 days of culture, the media was removed and the cells adhered on the surfaces were fixed in 4% paraformaldehyde in PBS for 15 min, followed by two rinses (5 min each) with PBS. Adhered ADSCs were permeabilized by incubation with 0.1% Triton X-100 in PBS for 4 min and rinsed twice with PBS and applied blocking solution (1 % BSA in PBS) for 30 min. Actin fibers were stained using 1:200 TRITC-conjugated phalloidin in PBS for 60 min, followed by three

rinses. The nuclear stain DAPI (1:1000 in PBS) was added and after 5 min the surfaces were rinsed with PBS three times and then imaged using a fluorescence microscope (Zeiss).

#### C.4 AFM Imaging and Force-Distance Measurements



**Fig. C3** High-resolution AFM images of (a) an unloaded crystal imaged by a PEG-RGD modified AFM tip, and (b) a PEG-RGD loaded crystal imaged by an inactivated AFM tip



**Fig. C4** F-D curves collected in a slow tapping frequency of 1 Hz by the activated AFM tip from (a) PEG-RGD loaded protein crystal, and (b) an unloaded protein crystal.

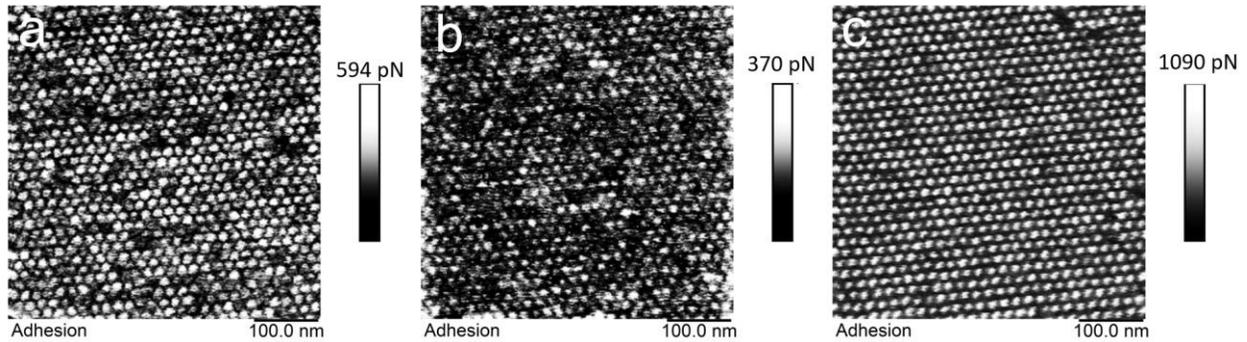
**Table C1.** The max, min, and median of the adhesion energy for 30 pore center pixels and 30 wall pixels for each of the experiments detected corresponding to Fig. 5.5 in the main text (a), (L = loaded, U = unloaded, I = inactivated, M = PEG-RGD modified, A = activated, P = pore, and W = wall), and (b) detailed penetration depth data corresponding to the images in Fig. 5.5 in the main text (b).

**(a) Adhesion energy (fJ)**

	LA-P	LA-W	UA-P	UA-W	UI-P	UI-W	UM-P	UM-W
<b>Max</b>	$3.34 \times 10^{-2}$	$9.38 \times 10^{-3}$	$1.35 \times 10^{-2}$	$6.83 \times 10^{-3}$	$1.20 \times 10^{-2}$	$9.86 \times 10^{-3}$	$7.35 \times 10^{-2}$	$2.60 \times 10^{-3}$
<b>Min</b>	$1.69 \times 10^{-2}$	$4.58 \times 10^{-3}$	$4.75 \times 10^{-3}$	$4.06 \times 10^{-3}$	$4.70 \times 10^{-3}$	$4.88 \times 10^{-3}$	$3.44 \times 10^{-2}$	$9.54 \times 10^{-3}$
<b>Median</b>	$2.33 \times 10^{-2}$	$7.46 \times 10^{-3}$	$9.91 \times 10^{-3}$	$5.64 \times 10^{-3}$	$8.67 \times 10^{-3}$	$7.41 \times 10^{-3}$	$5.28 \times 10^{-2}$	$1.88 \times 10^{-2}$
<b>Mean</b>	$2.38 \times 10^{-2}$	$7.37 \times 10^{-3}$	$9.80 \times 10^{-3}$	$5.50 \times 10^{-3}$	$8.60 \times 10^{-3}$	$7.36 \times 10^{-3}$	$5.39 \times 10^{-2}$	$1.86 \times 10^{-2}$

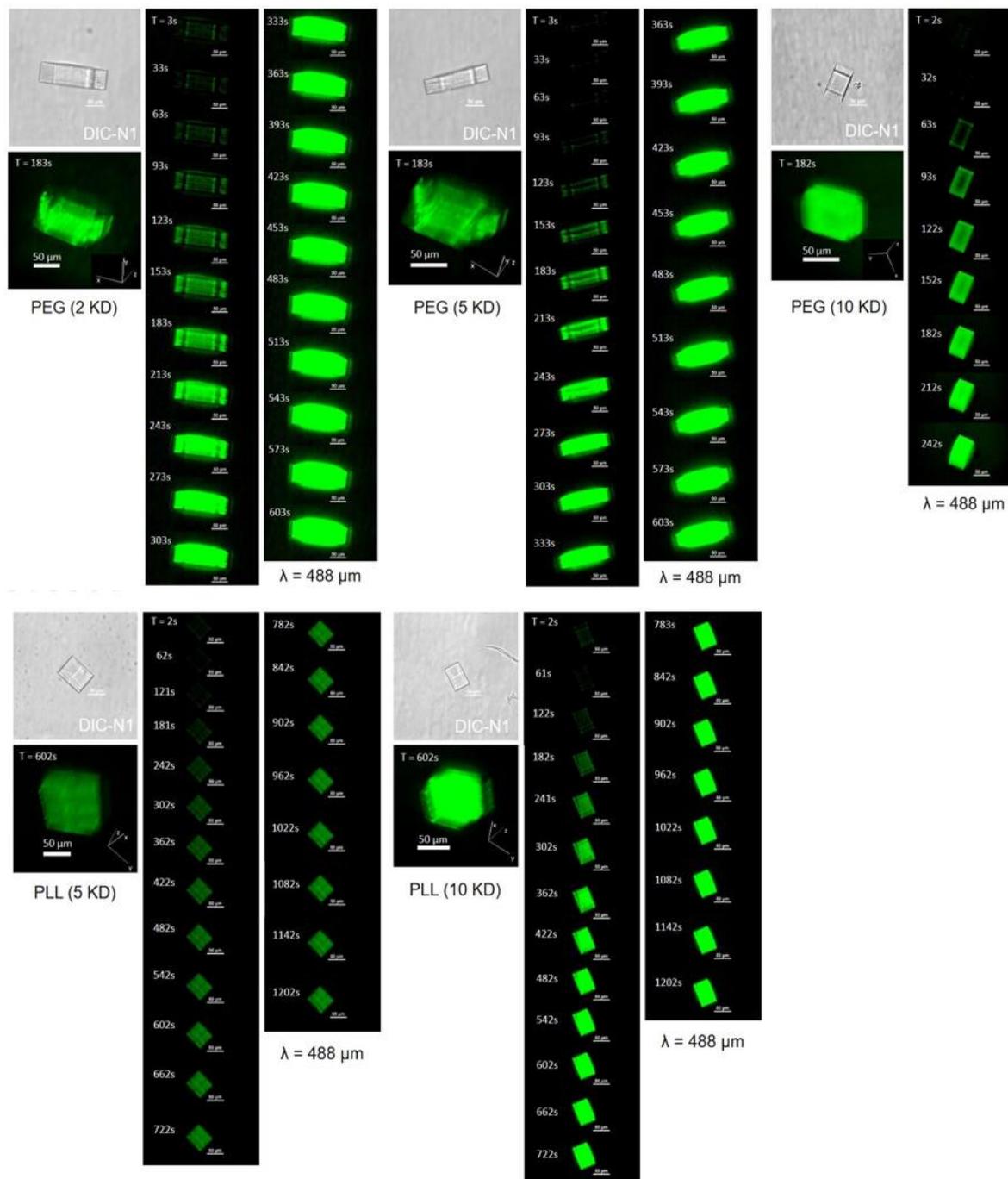
**(b) Penetration depth of tip into pores**

Condition	Penetration Depth (nm)
PEG-RGD loaded crystal by activated tip	$15.8 \pm 2.7$ (n = 251)
Unloaded crystal by activated tip	$4.5 \pm 1.1$ (n = 250)
Unloaded crystal by inactivated tip	$8.0 \pm 1.5$ (n = 234)
Unloaded crystal PEG-RGD modified tip	$7.3 \pm 2.9$ (n = 177)

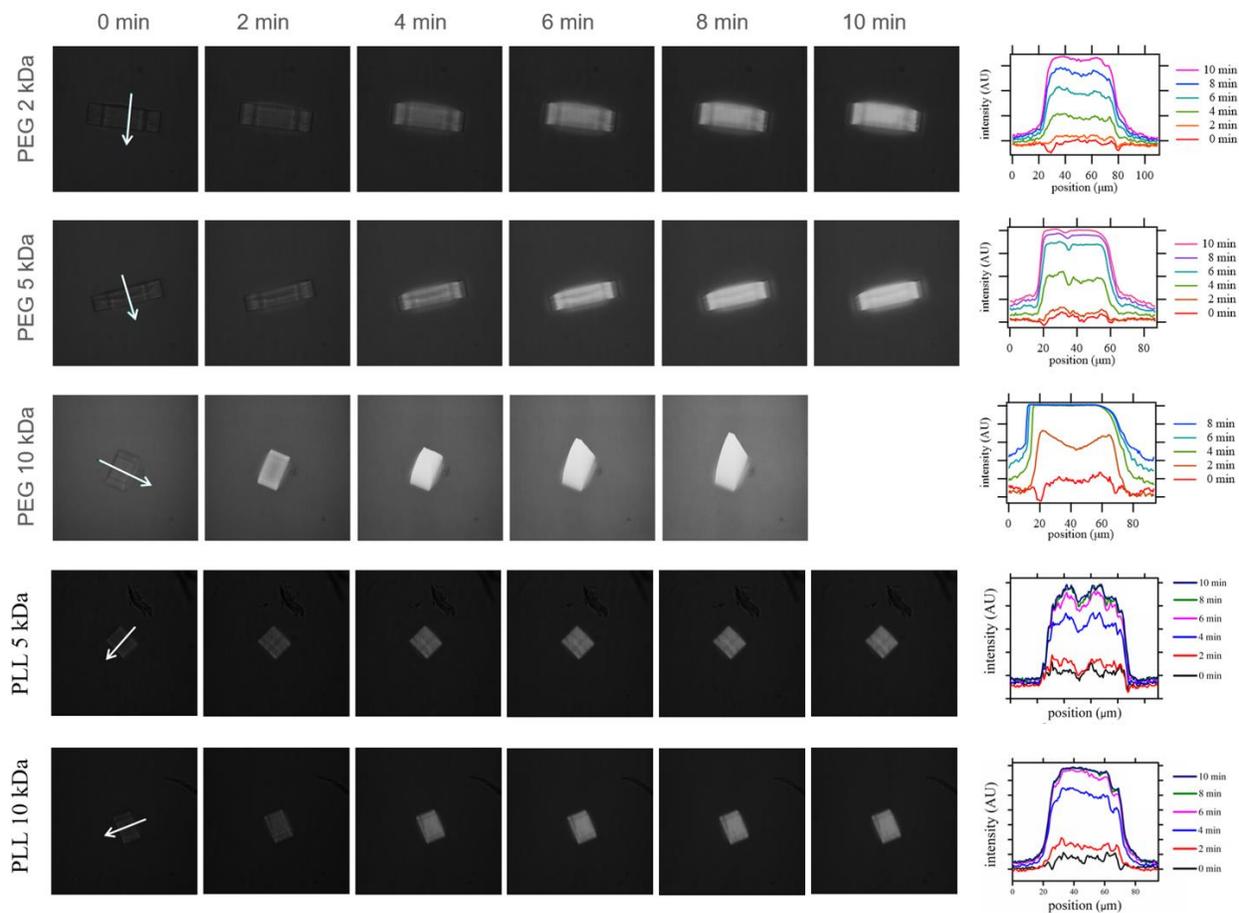


**Fig. C5** The adhesion maps of the  $CJ^{OPT}$  crystals for all imaging conditions, corresponding to the high resolution AFM images in Fig. 5.3 (a-c): (a) an unloaded crystal imaged with a non-modified AFM tip, (b) an unloaded crystal imaged with an activated AFM tip, and (c) a crystal loaded with PEG-RGD imaged using an activated AFM tip.

## C.5 Loading of Guest Molecules



**Fig. C6**  $CJ^{OPT}$  protein crystals (DIC image shown in upper left, all the scale bars in the figures are 50  $\mu\text{m}$  scale bar) were placed on one side and imaged during PEG-FITC (2 KDa, 5 KDa, and 10 KDa) and PLL-FITC (5 KDa and 10 KDa) loading, by confocal microscopy under the wavelength of DIC-N1 and  $\lambda = 488 \mu\text{m}$ . Confocal microscopy images taken at 0 to 20 minutes intervals are shown, from a  $z$ -plane near the center of the crystal (3 out of 5 plain by  $z$ -stack).



**Fig. C7** Spinning disk confocal microscopy images of fluorescently labeled PEG diffusing into protein crystals for three different molecular weights of PEG and two different molecular weight of PLL. Intensity profiles in the direction parallel to the protein crystal pores. The direction of the pores and intensity profiles is indicated in the  $t = 0$  min image for each sample. Diffusion of 10 kDa PEG into a crystal of this size reaches apparent equilibrium in less than 3 minutes, which is the fastest among all five types of biomolecules. We assume that the fluorescence intensity of each pixel has a linear relationship with the concentration of molecules at that point.

## REFERENCES OF APPENDIX C

- C1. McPherson, A.; Malkin, A.; G, Y., Atomic force microscopy in the study of macromolecular crystal growth. *Annu. Rev. Biophys.* **2000**, *29* (1), 361-410.
- C2. Ilari, A.; Savino, C., A practical approach to protein crystallography. In *Bioinformatics*, Springer: 2017; pp 47-78.
- C3. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M., Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.* **2013**, *280* (22), 5705-5736.
- C4. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M., Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.* **2008**, *275* (1), 1-21.
- C5. Kowalski, A. E.; Huber, T. R.; Ni, T. W.; Hartje, L. F.; Appel, K. L.; Yost, J. W.; Ackerson, C. J.; Snow, C. D., Gold nanoparticle capture within protein crystal scaffolds. *Nanoscale* **2016**, *8* (25), 12693-12696.
- C6. Bergkvist, M.; Cady, N. C., Chemical functionalization and bioconjugation strategies for atomic force microscope cantilevers. In *Bioconjugation Protocols*, Springer: 2011; pp 381-400.