



Article

# The Effect of Oligomerization on A Solid-Binding Peptide Binding to Silica-Based Materials

Rachit Bansal <sup>1,2</sup>, Zehra Elgundi <sup>3</sup>, Sophia C. Goodchild <sup>1</sup>, Andrew Care <sup>1,2</sup>,  
Megan S. Lord <sup>3</sup>, Alison Rodger <sup>1</sup> and Anwar Sunna <sup>1,2,4,\*</sup>

<sup>1</sup> Department of Molecular Sciences, Macquarie University, Sydney, NSW 2109, Australia; rachit.bansal@hdr.mq.edu.au (R.B.); sophia.goodchild@mq.edu.au (S.C.G.); andrew.care@mq.edu.au (A.C.); alison.rodger@mq.edu.au (A.R.)

<sup>2</sup> ARC Centre of Excellence for Nanoscale Biophotonics, Macquarie University, Sydney, NSW 2109, Australia

<sup>3</sup> Graduate School of Biomedical Engineering, University of New South Wales, Sydney, NSW 2052, Australia; z.elgundi@unsw.edu.au (Z.E.); m.lord@unsw.edu.au (M.S.L.)

<sup>4</sup> Biomolecular Discovery Research Centre, Macquarie University, Sydney, NSW 2109, Australia

\* Correspondence: anwar.sunna@mq.edu.au; Tel.: +612-9850-4220

Received: 28 April 2020; Accepted: 12 May 2020; Published: 15 May 2020



**Abstract:** The bifunctional linker-protein G (LPG) fusion protein comprises a peptide (linker) sequence and a truncated form of Streptococcus strain G148 protein G (protein G). The linker represents a multimeric solid-binding peptide (SBP) comprising 4 × 21-amino acid sequence repeats that display high binding affinity towards silica-based materials. In this study, several truncated derivatives were investigated to determine the effect of the SBP oligomerization on the silica binding function of LPG (for the sake of clarity, LPG will be referred from here on as 4 × LPG). Various biophysical characterization techniques were used to quantify and compare the truncated derivatives against 4 × LPG and protein G without linker (PG). The derivative containing two sequence repeats (2 × LPG) showed minimal binding to silica, while the truncated derivative with only a single sequence (1 × LPG) displayed no binding. The derivative containing three sequence repeats (3 × LPG) was able to bind to silica with a binding affinity of  $K_D = 53.23 \pm 4.5$  nM, which is 1.5 times lower than that obtained for 4 × LPG under similar experimental conditions. Circular dichroism (CD) spectroscopy and fluorescence spectroscopy studies indicated that the SBP degree of oligomerization has only a small effect on the secondary structure (the linker unravels the beginning of the protein G sequence) and chemical stability of the parent protein G. However, based on quartz crystal microbalance with dissipation monitoring (QCM-D), oligomerization is an important parameter for a strong and stable binding to silica. The replacement of three sequence repeats by a (GGGS)<sub>12</sub> glycine-rich spacer indicated that the overall length rather than the SBP oligomerization mediated the effective binding to silica.

**Keywords:** linker-protein G; oligomerization; solid-binding peptide (SBP); silica-binding peptide; silica-based materials; quartz crystal microbalance with dissipation monitoring (QCM-D)

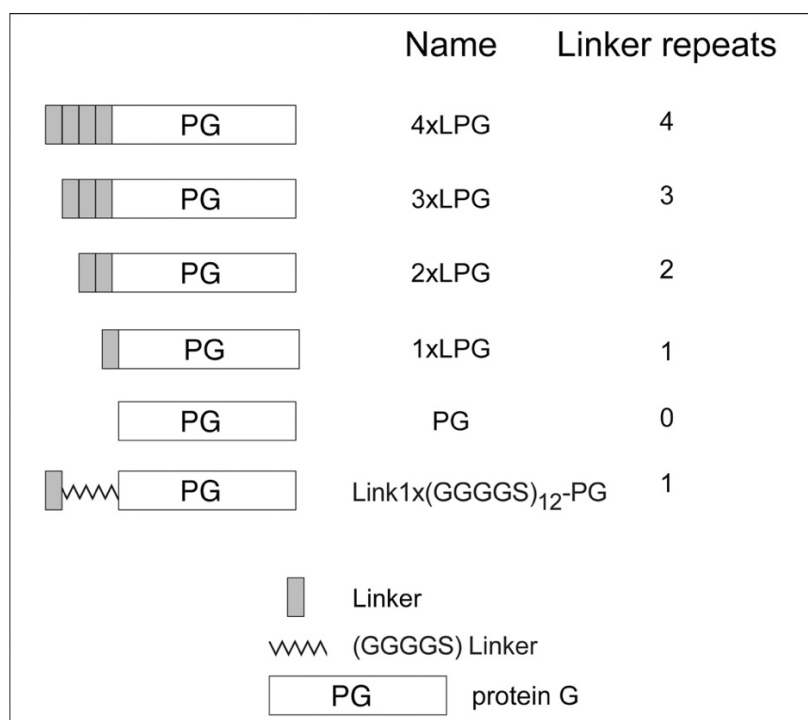
## 1. Introduction

Solid-binding peptides (SBPs) are a group of short amino acid sequences that specially bind to the surfaces of various inorganic materials including minerals, semiconductors, and polymers. In doing so, they are able to mediate the simple and controlled attachment of biomolecules onto solid surfaces, conferring biological functionality. SBPs can also act as functionalizing agents for binding and linking biomolecules to solid nanostructures [1,2]. Molecular biology protocols further allow tailoring of the selected peptides to tune their binding and material selectivity properties so that they can be used in bionanotechnological applications. Some SBPs have already been used in the development of multifunctional hybrid materials and for the oriented immobilization of biomolecules [3–5].

Proper orientation is usually crucial to preserve the bioactivity, as well as the avidity, of biomolecules. Care et al. demonstrated the oriented immobilization of antibodies onto silica-coated magnetic particles using a bifunctional linker-protein G (LPG) fusion protein, which contains a silica-binding peptide and an antibody-binding region [6]. The binding of SBPs to their corresponding materials is governed by several non-covalent interactions (i.e. electrostatic, polar, hydrophobic, hydrogen bonds, London dispersion) and these interactions typically result in equilibrium binding constants in the nM to sub-mM range [2,7,8]. However, due to the complexity [9,10] displayed at the SBP–material interface [11], many of the mechanisms involved in these interactions remain poorly understood. At present, a collection of experimental and theoretical tools are available to investigate these interactions. For example, surface plasmon resonance (SPR) [12], quartz crystal microbalance with dissipation monitoring (QCM-D) [13], and isothermal titration calorimetry (ITC) [14] can be used to investigate the selectivity, binding kinetics, thermodynamics, stoichiometry, orientation, and viscoelastic properties of these interactions. Circular dichroism (CD) spectroscopy [15], nuclear magnetic resonance (NMR) spectroscopy [16], and Fourier-transform infrared (FTIR) spectroscopy [17] provide information about the structure, conformation and dynamics of SBPs. Computational approaches such as in silico molecular modelling and simulation [18] with the development of artificial intelligence [19,20] provide in-depth knowledge of these interfaces at the atomic level [21].

The interactions between SBPs and their substrates rely on the affinity that particular chemical groups within amino acid residues have for solid surfaces. For example, SBPs that bind to metals predominantly contain hydrophobic and hydroxyl-containing polar residues [7], while those that bind to carbon-based materials are high in aromatic residues [22]. Various molecular-tailoring strategies can be applied to tune the affinity and selectivity of SBPs, the most common being site-specific mutagenesis. Another method used to tune the binding and structural features of an SBP is by increasing the number of repeats of the original SBP sequence, often referred to as multimerization; however, we use the phrase oligomerization to avoid confusion with more commonly used definitions of multimerization. For example, Seker et al. [23] reported the improved affinity and selectivity of a gold-binding peptide (GBP) as a function of polypeptide sequence repeats. Using SPR, they showed that a 3-repeat GBP (3R-GBP) was able to bind five times more strongly to gold than a 1-repeat GBP (1R-GBP). In another case, Cho et al. [24] used ITC analysis to show that a zinc oxide-binding peptide (ZBP) with triplicate tandem repeats ( $3 \times$  ZBP) displayed a nearly two-fold higher binding affinity ( $K_D = 0.69 \mu\text{M}$ ) to zinc oxide nanoparticles than the single peptide domain ( $K_D = 1.5 \mu\text{M}$ ). These findings implied that there might be a direct relationship between the binding avidity and the number of repeated sequences for some SBPs.

LPG (for the sake of clarity, LPG will be referred to from here on as  $4 \times$  LPG) was initially reported by Sunna et al. [1] and was designed as a bifunctional fusion protein for the rapid and oriented biofunctionalization of silica-based materials [6,25,26].  $4 \times$  LPG connects a recombinant form of *Streptococcus* protein G antibody-binding protein (PG) with an altered N-terminal domain to a multimeric linker made of a  $4 \times$  21-amino acid sequence repeat (Figure 1) [1] that displays high binding affinity towards silica-based materials. Here, we used various biophysical characterization techniques to examine the silica binding of truncated derivatives of  $4 \times$  LPG and to determine the effect of linker oligomerization on the stability of PG.



**Figure 1.** Truncated derivatives used in this study.

## 2. Materials and Methods

### 2.1. Materials and Chemicals

The silica binding studies were performed using silica-coated quartz crystals (QSensor QSX 303 SiO<sub>2</sub>) purchased from ATA Scientific (Taren Point, NSW, Australia). Pefabloc, lysozyme, and Benzonase nuclease were purchased from Sigma-Aldrich (Castle Hill, NSW, Australia). Human serum was obtained from Sigma-Aldrich (Castle Hill, NSW, Australia), while the humanized anti-HER2 monoclonal antibody trastuzumab was purchased from Jomar Life Research (Melbourne, VIC, Australia). The human HER2/ErbB2 protein (His-Tag) was purchased from Sino Biological Inc. (Beijing, China). All biological assays were performed at room temperature with standard phosphate-buffered saline (1 × PBS) at pH 7.4 containing 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, 137 mM NaCl, and 2.7 mM KCl (Sigma Aldrich, Castle Hill, NSW, Australia). All other chemicals were purchased from Sigma-Aldrich (Castle Hill, NSW, Australia) unless otherwise stated.

### 2.2. Production and Purification of Truncated Derivatives

The expression plasmids (pLink1 × pET22b, pLink2 × pET22b, and pLink3 × pET22b) used to produce the recombinant proteins were previously reported by [1]. The production of each of the truncated derivatives (1 × LPG, 2 × LPG, and 3 × LPG) was as follows: 1 L Luria Bertani (LB) medium (Merck Millipore, Bayswater, VIC, Australia) supplemented with 50 µg/mL carbenicillin was inoculated with 10 mL of an overnight culture of *E. coli* Tuner (DE3) cells (Novagen) harbouring the expression plasmid. The culture was incubated at 37 °C with continuous shaking (250 rpm) until the A<sub>600</sub> was approximately 0.7–1.0. The incubation temperature was reduced to 20 °C and protein synthesis was induced by the addition of 0.2 mM isopropyl β-D-thiogalactoside (IPTG). Cells were harvested after 3–4 h induction by centrifugation for 15 min at 10,000× g and 4 °C and were stored at –20 °C.

The cells were resuspended in ice-cold lysis buffer (25 mM Tris–HCl, pH 8.0, 100 mM NaCl, 1.25 mM EDTA, and 0.05% Tween 20), supplemented with 4 mM of the Pefabloc serine protease inhibitor and 1.5 mg lysozyme. They were ruptured by three passages through a French pressure cell (Thermo Fisher Scientific Australia, Scoresby, VIC, Australia). After cell rupture, 50 units of Benzonase

nuclease and 4 mM Pefabloc were added. The sample was incubated on ice for 20 min. The debris was removed by centrifugation for 30 min at 20,000× g and 4 °C. The supernatant obtained was first filtered through a 0.45 µm and then 0.22 µm sterile filter and stored at 4 °C.

All purifications were performed using the Äkta™ start chromatography system (GE Healthcare, Parramatta, NSW, Australia). For purification, each truncated derivative soluble extract (1 × LPG, 2 × LPG, and 3 × LPG) was loaded onto a 5 mL HiTrap Q anion exchanger column (GE Healthcare) previously equilibrated with 25 mM Tris–HCl, pH 8.0, supplemented with 100 mM NaCl. The column was washed extensively with the same buffer. Under these conditions, all truncated derivatives eluted at 200 mM NaCl. The eluted protein samples were concentrated using an Amicon Ultra-15 centrifugal filter (10 kDa cut-off, Merck Millipore) followed by buffer exchange with a PD-10 desalting column (GE Healthcare). In addition, 1 × LPG was buffer exchanged into 20 mM Bis-Tris (bis-(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane) buffer pH 6.0. This fraction was applied again onto the 5 mL HiTrap Q column (GE Healthcare) previously equilibrated with 20 mM Bis-Tris buffer pH 6.0. The column was washed extensively with the same buffer and the 1 × LPG recombinant protein was eluted at 200 mM NaCl. 2 × LPG and 3 × LPG were buffer exchanged into 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer pH 8.0. The fractions were applied to a 5 mL HiTrap SP cation exchanger column previously equilibrated with 50 mM HEPES buffer pH 8.0. The column was washed extensively with the same buffer. Under these conditions, 2 × LPG eluted at 50 mM NaCl, while 3 × LPG eluted at 100 mM NaCl.

Fractions containing the truncated derivatives were identified on 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad Laboratories, Gladesville, NSW, Australia) by sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE) and stained with Coomassie Brilliant Blue. Individual fractions were concentrated using an Amicon Ultra-15 centrifugal filter and all final samples were stored in 1 × PBS buffer (after fresh addition of Pefabloc) at –80 °C.

The final protein concentration was measured using the Micro BCA protein assay kit (Thermo Fisher Scientific Australia, Scoresby, VIC, Australia) according to the manufacturer's instructions.

### 2.3. Construction, Production, and Purification of a Synthetic Linker Derivative

Three sequence repeats from the 4 × LPG peptide linker were replaced by a synthetic (GGGS)<sub>n</sub> linker sequence. This replacement was important to determine whether the oligomer or simply the distance from the PG was required for efficient binding to silica. The PG sequence was ligated into plasmid pET44a (+) via NheI/BamHI restriction sites to obtain the plasmid pET44-PG. A gene sequence containing one original SBP sequence from LPG followed by the synthetic linker sequence (GGGS)<sub>12</sub> was synthesized by Invitrogen GeneArt Gene Synthesis (ThermoFisher Scientific Australia) and ligated into plasmid pET44-PG via NdeI/NheI restriction sites to assemble the expression plasmid pLink1 × (GGGS)<sub>12</sub>-PG. *E. coli* α-Select (Bioline, Eveleigh, NSW, Australia) was used as a host for general gene cloning and vector storage, and *E. coli* Tuner (DE3) cells were used for recombinant protein expression.

Link1 × (GGGS)<sub>12</sub>-PG was produced and the soluble protein fraction was obtained as described for the truncated derivatives. For purification of Link1 × (GGGS)<sub>12</sub>-PG, the soluble extract was loaded onto a 5 mL HiTrap Q anion exchanger column previously equilibrated with 25 mM Tris–HCl, pH 8.0, supplemented with 100 mM NaCl. The column was washed extensively with the same buffer. Under these conditions, the Link1 × (GGGS)<sub>12</sub>-PG did not bind to the column and was found in the unbound fraction. The Link1 × (GGGS)<sub>12</sub>-PG was concentrated using an Amicon Ultra-15 centrifugal filter (10 kDa cut-off) followed by buffer exchange with a PD-10 desalting column and 50 mM MES (2-(N-morpholino) ethanesulfonic acid) buffer pH 6.0. This fraction was applied to a 5 mL HiTrap SP cation exchanger column previously equilibrated with 50 mM MES buffer pH 6.0. The column was washed extensively with the same buffer and the Link1 × (GGGS)<sub>12</sub>-PG was eluted at 300 mM NaCl. Fractions containing the Link1 × (GGGS)<sub>12</sub>-PG were identified by SDS-PAGE as described above. Fractions containing the purified Link1 × (GGGS)<sub>12</sub>-PG were pooled and concentrated using

an Amicon Ultra-15 centrifugal filter. Samples were stored in 50 mM MES buffer pH 6.0 at  $-20\text{ }^{\circ}\text{C}$  after sterile filtration through a  $0.22\text{ }\mu\text{m}$  filter.

#### 2.4. Silica Binding Assay

The binding of  $4\times$  LPG and the truncated derivatives used in this study ( $3\times$  LPG,  $2\times$  LPG,  $1\times$  LPG, and PG) to silica was previously reported by [1] using a silica binding assay. Similarly, the binding of purified recombinant Link1  $\times$  (GGGGS)<sub>12</sub>-PG to silica was determined as follows: 5 mg silica (BDH Ltd., Poole, Dorset, U.K.) was washed three times with washing buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 1% Triton-X100). Soluble protein in a final volume of 100  $\mu\text{L}$  was mixed with silica and incubated by rotation at room temperature for 1 h. The unbound fraction was removed after centrifugation at  $14,000\times g$  for 20 s. The silica pellet was washed three times by vortexing with 100  $\mu\text{L}$  of 100 mM Tris-HCl buffer, pH 8.0. Finally, the silica-bound protein was eluted after addition of 100  $\mu\text{L}$  of SDS PAGE-loading buffer and incubation at  $99\text{ }^{\circ}\text{C}$  for 10 min (with short mixing every 2 min). Fractions containing the Link1  $\times$  (GGGGS)<sub>12</sub>-PG protein were identified by SDS-PAGE as described above.

The percentage of protein in the different fractions was semi-quantified by analysing the intensity of bands of digital images from SDS-PAGE with ImageJ software (version 1.48, NIH, Bethesda, MD, USA) (<http://rsb.info.nih.gov/ij/index.html>).

#### 2.5. Circular Dichroism (CD) Spectroscopy

CD spectra were collected at room temperature (approximately  $25\text{ }^{\circ}\text{C}$ ) on a JASCO J-1500 spectrophotometer (JASCO Corporation, Japan). CD data were collected in water in order to avoid any interference or noise below 200 nm as PBS absorbs strongly at wavelengths below 200 nm. All protein samples were buffer exchanged to Milli-Q water using Amicon Ultra-10K 0.5 mL centrifugal filters and were then further diluted in Milli-Q water to a final concentration of 0.1 mg/mL (as determined by the Micro BCA protein assay). Wavelength scans were performed between 180 and 350 nm in a rectangular, 1 mm pathlength quartz cuvette (Starna Scientific Ltd., Ilford, UK). For each sample, 3 accumulations were recorded using a 2 nm bandwidth, a scan speed of 100 nm/min, and a digital integration time (D.I.T.) of 2 s. The data are reported in terms of molar extinction coefficient ( $\Delta\epsilon$ ), expressed in  $\text{mol}^{-1}\text{ cm}^{-1}\text{ dm}^3$ .

#### 2.6. Circular Dichroism Secondary Structure Analysis

Circular dichroism spectra were analysed using a graphical user interface (GUI) version (GUI to Implement SOMSpec, a CD Secondary Structure Fitting Approach 2018, gitlab.com) of our validated self-organizing map approach to secondary structure fitting [27,28]. In this approach, a reference set, in our case, SP175 [29] (truncated to 190 nm), augmented with a constructed 100% helix and 100% random coil peptide data [30] is organized into a map with spectra of similar shape in the same neighbourhood. Another map, which overlays the first, contains the secondary structure contents of the reference proteins and the nodes whose secondary structure is determined by interpolation from the neighbouring 5 references spectra. An unknown spectrum is placed on the map where it minimizes the distance (difference) between it and the spectral shape and magnitude of a nodal spectrum. The unknown protein's secondary structure is then assigned to be that of the node. We ran SOMSpec with a  $50\times 50$  map, 5 best matching units, and a wavelength range from 240 to 190 nm. The goodness of fit of the secondary structure estimates was assessed by considering the overlay of the experimental and predicted spectrum visually and by a normalized root-mean-square deviation (NRMSD). Visual inspection weights the maxima, minima, and where a spectrum crosses the zero-line more than other points in the spectrum, whereas the NRMSD approach equally weights all wavelengths. In general, for spectra from 240 to 190 nm, an NRMSD  $< 0.02$  indicates a good fit with only 1–2% uncertainty, whereas higher NRMSDs indicate more uncertainty in the prediction. Higher values usually require user interference to choose the best fit.

It was visually apparent (from the shift in the negative maximum from 208 nm towards 200 nm) in the spectra that addition of longer peptides increased the unfolded percentages of the proteins. However, the fitting method was not selecting the random coil member of the reference set, due to it not being one of its best matching units. We, therefore, manually subtracted different fractions of the random coil spectrum for the SufI-KK peptide MSLSKKQFIQASGIALCAGAVPLKASA [30] and reran SOMSpec to obtain the best possible fit. The derandomized molar residue  $\Delta\epsilon$  was determined from the experimental  $\Delta\epsilon$  using

$$\Delta\epsilon_{\text{derandomised}} = \frac{\Delta\epsilon_{\text{exp}} - \frac{\#_{\text{RC}}}{\#_{\text{exp}}} \Delta\epsilon_{\text{RC}}}{1 - \frac{\#_{\text{RC}}}{\#_{\text{exp}}}} \quad (1)$$

where RC denotes random coil, # denotes the number of residues, exp denotes the full protein on which data were collected, and derandomized denotes the core of the protein when the linkers and any extra unfolded residues are removed. The derivation of Equation (1) is provided in Equation (S1) (Supplementary Materials).

### 2.7. Fluorescence Spectroscopy

Protein samples were dissolved to a final concentration of 2  $\mu\text{M}$  in 1  $\times$  PBS containing various concentrations (0–7 M) of guanidinium hydrochloride (GdnHCl). All measurements were recorded at room temperature (approximately 25  $^{\circ}\text{C}$ ) in a micro-volume fluorescence cuvette with a 3 mm pathlength (Starna Scientific Ltd., Ilford, UK) using a JASCO FP-8500 Spectrofluorometer (JASCO Corporation, Japan). Fluorescence emission spectra were collected between 300 and 550 nm using a 295 nm excitation wavelength, 2.5 nm excitation bandwidth, 5 nm emission bandwidth, and scan speed of 100 nm/min.

### 2.8. Quartz Crystal Microbalance with Dissipation Monitoring (QCM-D) Interaction Analyses

The adsorption of the proteins (1  $\times$  LPG, 2  $\times$  LPG, and 3  $\times$  LPG) onto silica was monitored on a 4-channel QSense Analyzer QCM-D (Biolin Scientific AB, Gothenburg, Sweden) using  $\text{SiO}_2$ -coated sensors. Sensors were cleaned before use with 2% sodium dodecyl sulphate (SDS) for 30 min at room temperature, washed thoroughly with Milli-Q water, and dried with nitrogen gas prior to use. The temperature of the QCM-D was maintained at  $22 \pm 0.1$   $^{\circ}\text{C}$  and the flow rate was set to 150  $\mu\text{L}/\text{min}$  throughout all experiments. Sensor crystals were exposed to 1  $\times$  PBS, and stable frequency ( $f$ ) and dissipation ( $D$ ) measurements were established. Test protein solutions were prepared in 1  $\times$  PBS buffer pH 7.4 at room temperature at concentrations of 3.27–654 nM and exposed to the surfaces for 900–15,000 s and then rinsed with 1  $\times$  PBS until consistent  $f$  and  $D$  measurements were obtained. In addition, the ability of an antibody and antigen complex to bind to 3  $\times$  LPG immobilized on silica was explored. Sensor crystals were exposed to 1  $\times$  PBS, and stable  $f$  and  $D$  measurements were established. 3  $\times$  LPG was exposed to silica at a concentration of 3.27–654 nM for 900–15,000 s, and the sensors were then rinsed with 1  $\times$  PBS until consistent  $f$  and  $D$  measurements were obtained. To eliminate the risk of non-specific binding, the free binding sites on the crystal surface were blocked by injecting 1 mg/mL BSA (in 1  $\times$  PBS) followed by rinses by rinsing with 1  $\times$  PBS until stable  $f$  and  $D$  measurements were obtained. The humanized anti-HER2 monoclonal antibody, trastuzumab, was then injected into the system at a final concentration of 1  $\mu\text{g}/\text{mL}$  until a stable response was observed followed by rinses by rinsing with 1  $\times$  PBS until stable  $f$  and  $D$  measurements were obtained. The final step included the injection of HER2 antigen at a concentration of 1  $\mu\text{g}/\text{mL}$  until equilibrium was reached.

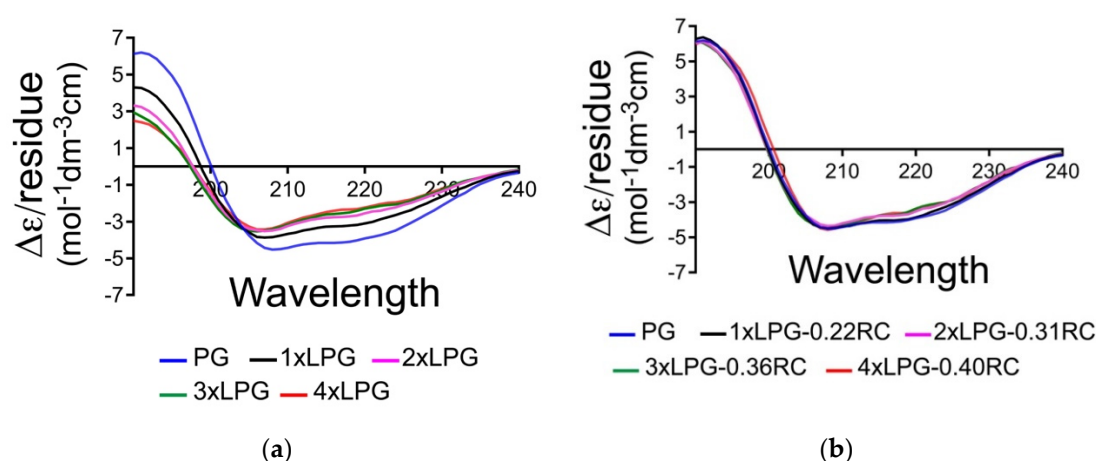
A minimum of three adsorption curves were recorded for each test protein at each concentration, which were measured at the fundamental  $f$  (5 MHz), as well as the third (15 MHz) to eleventh (55 MHz) overtones. All solutions were passed through a 0.22  $\mu\text{m}$  filter and degassed at room temperature before use. The adsorbed mass is proportional to the change in  $f$  ( $\Delta f$ ) via the Sauerbrey equation only when the adsorbed layer is rigid, uniformly distributed on the sensor surface, and the adsorbed mass is small compared to the mass of the crystal [31,32]. However, in the case of most proteins, these assumptions are not valid, as the adsorbed layer is viscoelastic as there is a dissipation change

greater than  $1 \times 10^6/10$  Hz. Under these conditions, the Voigt viscoelastic model is used to estimate the adsorbed protein layer mass and thickness from  $f$  and  $D$  measurements at multiple harmonics, assuming an adsorbed protein layer density of  $1000 \text{ g/m}^3$  [33–35]. The Voigt model was applied in the QTools software (QSense). Data are presented as  $\Delta f$  and  $\Delta D$  versus time graphs or  $Df$  plots ( $\Delta D$  versus  $\Delta f$ ) for the third overtone.

### 3. Results and Discussion

#### 3.1. Effect of Linker Oligomerization on the Secondary Structure of PG

The bioactivity of a biomolecule is dependent on its secondary structure. CD spectroscopy was used to determine the effect of different truncations of the linker sequence on the overall secondary structure of PG. A concentration of  $0.1 \text{ mg/mL}$  (the absorbance for the CD signals measured between 180 and 350 nm was normalized to account for the small variations in concentration) was used to measure the far-UV spectra of the truncated derivatives ( $1 \times \text{LPG}$ ,  $2 \times \text{LPG}$ , and  $3 \times \text{LPG}$ ). As shown in Figure 2a, the presence of a positive peak at  $\sim 190 \text{ nm}$  and negative peaks at  $\sim 208$  and  $\sim 222 \text{ nm}$  indicates significant  $\alpha$ -helical content in all the proteins, which is evident. The SOMSpec secondary structure estimates using an augmented SP175 reference set (truncated to be from 240 to 190 nm) gave the secondary structure estimates of the first section of Table 1. The NRMSDs and the SOMSpec output (illustrated in Figure S1, Supplementary Materials) indicate that the PG analysis is of satisfactory quality with 38%  $\alpha$ -helix and 16%  $\beta$ -sheet. Goward et al.'s [36] prediction of 37% + 4%  $\alpha$ -helix and  $30\% \pm 5\%$   $\beta$ -sheet supports this helix estimate and raises doubts about the sheet. Their CD analysis (using CONTIN) [37] suggested approximately 10% less helix and 10% more sheet. CONTIN uses a ridge regression procedure, which fits the CD spectrum of the test protein as a linear combination of the CD spectra of  $N$  reference proteins by minimizing a function. CONTIN was deemed by Sreerama and Woody to seldom give the best fit [38]. The 2IGG Protein Data Bank (PDB) NMR structure [39] suggests that the 64-residue IGG-binding domain that is repeated in PG is only 23%  $\alpha$ -helix but 37%  $\beta$ -sheet. However, the N-terminal  $\beta$ -strands are curved into a 2D helix in the PDB structures, would need to be unpacked to form PG, and have high propensity helix residues on the N-terminal side of the helix. Thus, we believe our secondary structure estimate for PG is reasonably accurate. By way of contrast, both the NRMSD and visual inspection of the SOMSpec results for  $1 \times \text{LPG}$ ,  $2 \times \text{LPG}$ ,  $3 \times \text{LPG}$ , and  $4 \times \text{LPG}$  are far from satisfactory (Table 1 and Figure S1, Supplementary Materials).



**Figure 2.** (a) Far-UV  $\Delta\epsilon/\text{residue}$  spectra of protein G (PG) (blue),  $1 \times$  linker-protein G (LPG) (black),  $2 \times$  LPG (pink),  $3 \times$  LPG (green), and  $4 \times$  LPG (red) in water collected at a concentration of  $0.1 \text{ mg/mL}$ . (b) De-randomized  $\Delta\epsilon/\text{residue}$  spectra with the indicated percentage of a peptide random coil spectrum subtracted according to Equation (1).

**Table 1.** SOMSpec secondary structure estimates for all proteins. The first section is for the data sets for the experimental proteins. The second section is for the proteins where the optimized amount of random coil has been removed. The final section is for the reconstructed proteins.

	$\alpha$ -Helix	$\beta$ -Sheet	Bends	Turns	Other	NRMSD
PG	0.38	0.16	0.16	0.10	0.20	0.01
1 $\times$ LPG	0.29	0.21	0.16	0.10	0.25	0.05
2 $\times$ LPG	0.23	0.28	0.13	0.10	0.25	0.06
3 $\times$ LPG	0.21	0.31	0.13	0.10	0.25	0.07
4 $\times$ LPG	0.17	0.36	0.14	0.10	0.23	0.06
Derandomized proteins						
1 $\times$ LPG–0.22RC	0.39	0.15	0.17	0.10	0.20	0.03
2 $\times$ LPG–0.31RC	0.36	0.14	0.19	0.09	0.23	0.03
3 $\times$ LPG–0.36RC	0.36	0.14	0.19	0.09	0.23	0.03
4 $\times$ LPG–0.40	0.34	0.17	0.14	0.12	0.23	0.04
Reconstructed proteins						
1 $\times$ LPG	0.38	0.16	0.16	0.10	0.34	0.38
2 $\times$ LPG	0.30	0.12	0.13	0.08	0.38	0.30
3 $\times$ LPG	0.25	0.10	0.13	0.06	0.47	0.25
4 $\times$ LPG	0.23	0.09	0.12	0.06	0.51	0.23

We, therefore, subtracted the CD of different fractions of a random coil polypeptide, selected the fraction that had the best SOMSpec predicted spectrum, and determined the secondary structure contents. Figure 2b shows the overlay of the derandomized spectra. While not perfect, reflecting the somewhat arbitrary choice of a peptide random coil spectrum, it is sufficient for our purposes. Table 1 part 2 shows the SOMSpec output for the best fits of the derandomized proteins (fraction of random coil removed is indicated). Part 3 indicates the secondary structure content when the full random coil content is included. For each of the linker proteins, the amount of random coil required to obtain the best fit is about 14% of the residues of PG plus the appropriate linker. Although there is an error to the fitting, our data suggest that the addition of the linkers (and removal/replacement of some of the N-terminal PG residues) not only adds unstructured protein but also causes the core protein to lose some structure:  $\sim$ 10 residues of helix,  $\sim$ 6 of sheet for 1  $\times$  LPG, and more for the longer ones (Table 2). This supports the C' domain structure prediction of Goward et al. [36] with its helix near the N-terminal (and thus supports our PG structure estimates above). The loss of structured residues increases slightly for the longer linkers (Table 2).

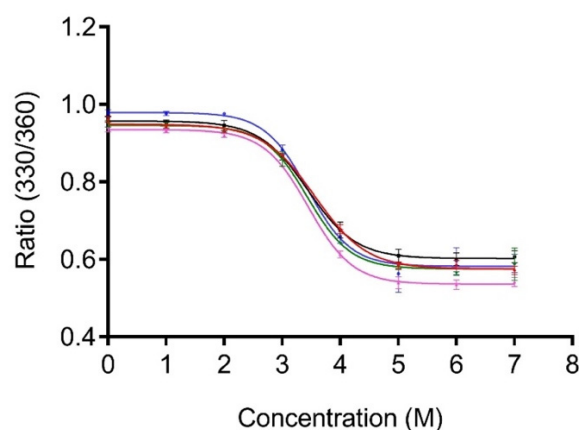
**Table 2.** Numbers of residues in the proteins (notation as in Table 1).

	$\alpha$ -Helix	$\beta$ -Sheet	Bends	Turns	Other	Total
PG	75	31	31	20	39	196
Derandomized proteins						
1 $\times$ LPG–0.22RC	65	25	28	17	33	168
2 $\times$ LPG–0.31RC	59	23	30	14	37	163
3 $\times$ LPG–0.36RC	59	23	31	14	38	164
4 $\times$ LPG–0.40	56	28	24	20	39	167
Reconstructed proteins						
1 $\times$ LPG	65	25	28	17	81	215
2 $\times$ LPG	59	23	30	14	110	236
3 $\times$ LPG	59	23	31	14	130	257
4 $\times$ LPG	56	28	24	20	150	278



### 3.2. Effect of Linker Oligomerization Chemical Stability of PG

The relative stability of all truncated derivatives was compared in the presence of the chemical denaturant GdnHCl using fluorescence spectroscopy. PG includes eight tyrosine and three tryptophan residues in the core of the protein. Accordingly, intrinsic tryptophan fluorescence (ITF) experiments were performed at an excitation of 295 nm to eliminate the contribution of tyrosine to the measured fluorescence. The linker sequence does not contain any aromatic residues, so it does not contribute to the overall fluorescence. The unfolding of 1 × LPG, 2 × LPG, 3 × LPG, and 4 × LPG was monitored by measuring the shift in the maximum tryptophan fluorescence emission with an increasing concentration of GdnHCl, and the ratio of the fluorescence intensities at 330 and 360 nm was plotted against increasing denaturant concentration (Figure 3).



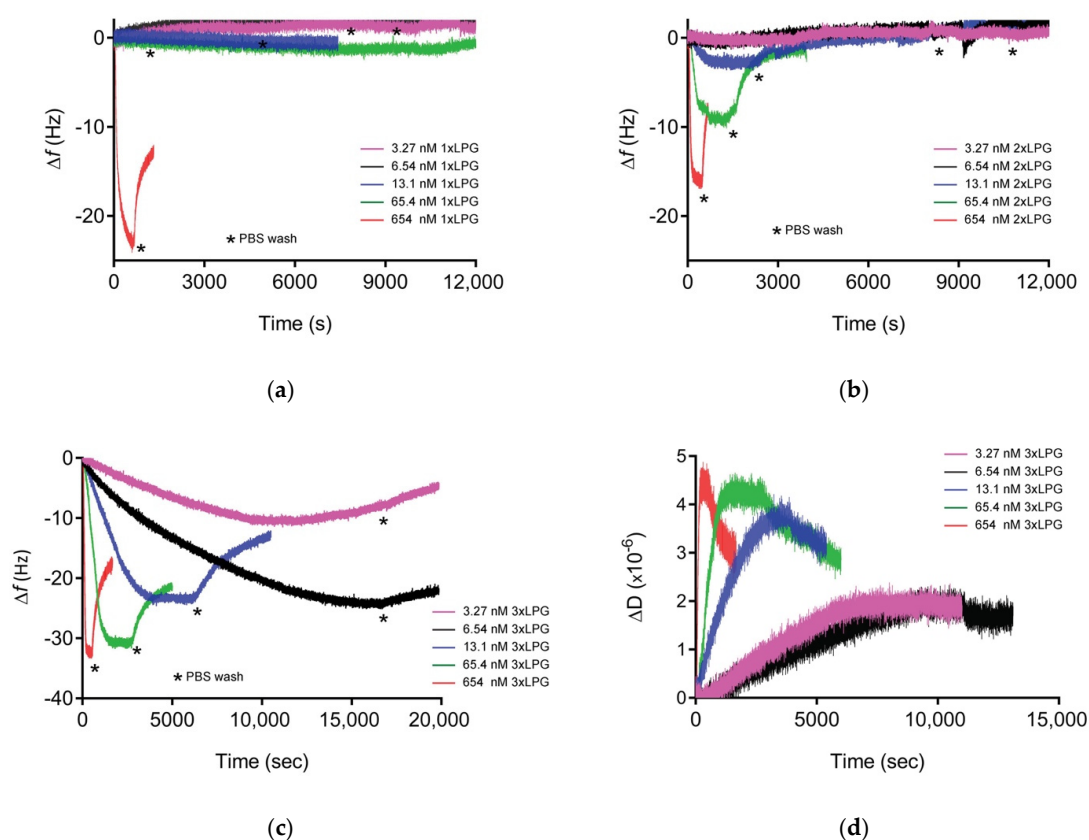
**Figure 3.** Unfolding of PG (blue), 1 × LPG (black), 2 × LPG (pink), 3 × LPG (green), and 4 × LPG (red) in the presence of guanidinium hydrochloride (GdnHCl). All protein samples were diluted to a final concentration of 2 μM in 1 × PBS, pH 7.4, containing various molar concentrations of GdnHCl. The ratio of relative fluorescence intensity at 330 and 360 nm (330/360) was monitored after excitation at 295 nm.

Assuming the protein denaturation to be a two-state process, GraphPad Prism 7.0 (GraphPad Software, La Jolla, CA, USA) was used to non-linearly fit the data. The concentration of GdnHCl required for 50% unfolding of the core protein where the tryptophans are located was determined to be  $3.4 \pm 0.2$  M (95% CI),  $3.4 \pm 0.1$  (95% CI), and  $3.4 \pm 0.2$  (95% CI) for 1 × LPG, 2 × LPG, and 3 × LPG, respectively. The concentrations for PG and 4 × LPG were previously reported as  $3.4 \pm 0.2$  M and  $3.6 \pm 0.1$  M, respectively [31]. Thus, the chemical stability of the core PG portion of the fusion proteins is, within experimental error, unaffected by the linker sequences.

### 3.3. Minimal Number of Repeats Required for Silica Binding

The binding affinity of the various truncated derivatives to a silica surface was determined using QCM-D (Figure 4). The binding affinity of 1 × LPG was beneath the limit of quantitation as only the highest concentration showed some binding, but the adsorbed protein was removed after washing with 1 × PBS (Figure 4a  $\Delta f$  vs. time sensorgram, Supplementary Figure S2a, Supplementary Materials). The sensorgram for 2 × LPG and a silica surface (Figure 4b) showed measurable binding at the three highest protein concentrations (654, 65.4, and 13.1 nM). Although it was not possible to calculate the binding affinity for 2 × LPG, the frequency shift for the highest 2 × LPG concentration suggests that the adsorption of the protein to the silica surface is almost half (~16.5 Hz) that previously reported observed for 4 × LPG [31]. Binding of 3 × LPG to silica was observed at all concentrations analysed with both changes in  $f$  and  $D$  measurements observed (Figure 4c,d) and a binding affinity ( $K_D$ ) of  $53 \pm 5$  nM, when fitted using the Langmuir isotherm (Figure S3, Supplementary Materials). This  $K_D$  value is 1.5 times higher than that obtained for 4 × LPG under similar experimental conditions ( $K_D = 35 \pm 12$  nM),

indicating an overall lower binding affinity of the 3 × LPG truncated derivative. Based on these data, a general trend of increased binding to the silica surface was observed with linker oligomerization.



**Figure 4.** Representative  $\Delta f$  vs. time sensorgrams for (a) 1 × LPG, (b) 2 × LPG, and (c) 3 × LPG binding to silica, and  $\Delta D$  vs. time for (d) 3 × LPG binding to silica as measured at the 3<sup>rd</sup> overtone by quartz crystal microbalance with dissipation monitoring (QCM-D) crystals at various concentrations (3.27–654 nM). The measurements were performed at a flow rate of 150  $\mu\text{L}/\text{mL}$  at 22 °C in three steps: Establishment of baseline (1 × PBS buffer), adsorption of protein until saturation, and washing (1 × PBS) to remove unbound protein. Three independent measurements were performed for each concentration.

The  $D_f$  plot generated from the QCM-D data for 3 × LPG binding to silica (Figure S4, Supplementary Materials) indicated that the data were contained within the viscoelastic region consistent with the need to utilize the Voigt viscoelastic model to estimate bound thickness and mass and estimate the viscosity of the adsorbed layer (Table 3). The viscous nature of the adsorbed protein layer might be due to the entrapment of water between the peptide layers owing to the hydrophilic nature of the protein. This was also the case for the transient binding between either 1 × LPG or 2 × LPG and silica with  $\Delta D$  values greater than  $1 \times 10^6/10$  Hz (Supplementary Figure S2, Supplementary Materials), although the binding was not strong to withstand rinsing with 1 × PBS with 1 × LPG completely removed from the surface after the PBS washing step, and only half of the initially bound 2 × LPG remained on the surface after the washing step.

The maximum thickness of the peptide layer was calculated by the Voigt model at the point of maximum  $\Delta f$  prior to rinsing with 1 × PBS and was found to be 6 nm for 3 × LPG binding to silica, which equated to a layer with a mass of approximately 680  $\text{ng}/\text{cm}^2$  (Table 3). Analysis of the orientation of 3 × LPG on the silica surface was determined by taking into account its molecular weight of 28.2 kDa and globular nature, as performed previously for 4 × LPG [34], and was approximated to be 324 and 175  $\text{ng}/\text{cm}^2$  for end-on and side-on orientation, respectively. The mass of 3 × LPG on the silica surface was approximately 680  $\text{ng}/\text{cm}^2$  (Table 3), which, based on the analysis above, is most likely a hydrated

end-on configuration, as observed previously for  $4 \times$  LPG [31]. Based on these results, it appears that the three repeats of the linker ( $3 \times$  LPG) support a similar level of binding as  $4 \times$  LPG, but with comparatively weaker binding affinity.

**Table 3.** Viscoelastic parameters for  $3 \times$  LPG–silica binding interactions obtained by fitting the QCM-D data using Voigt model. Data presented as mean  $\pm$  error ( $n = 3$ ).

	Thickness (@ $R_{eq}$ ) <sup>a</sup> (nm)	Thickness (@ $k_d$ ) <sup>b</sup> (nm)	Mass Deposited (@ $R_{eq}$ ) ( $\times 10^8$ ng/cm <sup>2</sup> )	Mass Deposited (@ $k_d$ ) ( $\times 10^8$ ng/cm <sup>2</sup> )	Viscosity (@ $R_{eq}$ ) $\times$ $10^{-4}$ (kg/ms)	Viscosity (@ $R_{eq}$ ) $\times$ $10^{-4}$ (kg/ms)
$3 \times$ LPG/SiO <sub>2</sub>	5.76 $\pm$ 0.15	4.36 $\pm$ 0.05	680 $\pm$ 34	589 $\pm$ 24	19.7 $\pm$ 0.4	17.56 $\pm$ 0.15

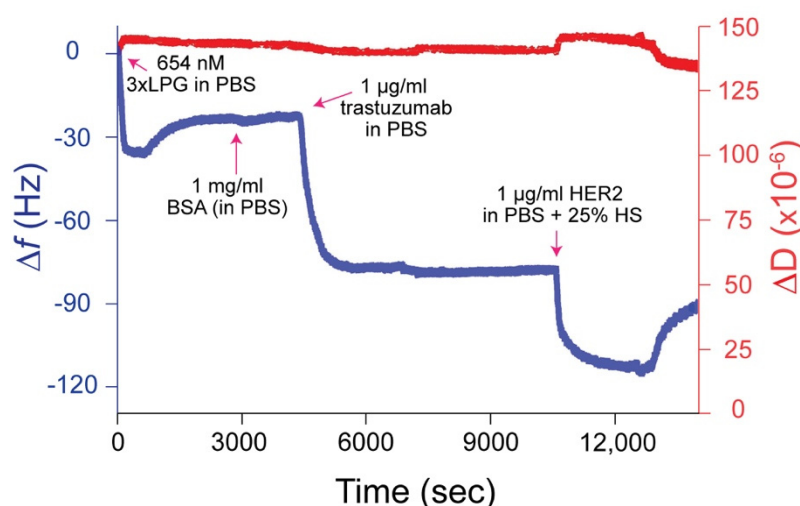
<sup>a</sup> @ $R_{eq}$ , frequency shift at equilibrium; <sup>b</sup> @ $k_d$ , saturation achieved after dissociation.

Initial work by Brown highlighted the possible dependence of a gold-binding peptide binding and avidity towards its inorganic substrate to the peptide oligomerization [40]. In this case, the binding affinity of a gold-binding peptide and alkaline phosphatase fusion protein to gold was directly dependent on the numbers of repeating gold-binding polypeptides. Seker et al. [32] studied the effect of SBP oligomerization on the binding affinity to their solid substrate. An increase in the binding affinity of a gold-binding peptide (GBP1) was observed with the increasing number of tandem repeats (3-repeats). Similarly, increased binding affinity to the surface of ZnO-coated nanoparticles was observed after the oligomerization of a zinc-binding peptide (ZBP). In this case, the trimeric ( $3 \times$  ZBP) variant displayed a binding to the ZnO-coated nanoparticles nearly twice as strong as that observed with the monomeric ( $1 \times$  ZBP) variant [24]. Interestingly, the binding stoichiometry of each ZBP per ZnO-coated nanoparticle followed a reverse pattern where it was lower for  $3 \times$  ZBP ( $\sim 2.3$  molecules) as compared to  $1 \times$  ZBP ( $\sim 3.9$  molecules). This can be attributed to the fact that the small tethered length of  $1 \times$  ZBP supported a denser packing of the peptide molecules onto the ZnO surface due to the availability of comparatively more free binding sites. The silica-binding peptides QBP1 and QBP2 displayed equilibrium binding constants of  $0.12 \times 10^6$  and  $1.2 \times 10^6$ , respectively [32]. However, only the tandem 3-repeats derivative of QBP1 (3l-QBP1) showed increased affinity to silica but not the QBP2 derivative (3l-QBP2).

Based on the current data available, there is no general trend for increased binding strength as a result of SBP oligomerization. While in some cases, oligomerization increased the binding affinity, in others, the affinity was reduced, and this was attributed to the conformational changes between the single and the multiple peptide repeats [23,32].

#### 3.4. Binding Orientation of Truncated Derivative $3 \times$ LPG

In order to assess whether  $3 \times$  LPG binds to silica surfaces with its antibody-binding region available to bind an antibody, we used QCM-D to study the binding interaction between silica-bound  $3 \times$  LPG and the trastuzumab antibody followed by the ability of the  $3 \times$  LPG–trastuzumab complex to detect the HER2 antigen in human serum, a complex biological fluid, as depicted in the  $\Delta f$  and  $\Delta D$  vs. time sensorgram (Figure 5). The  $\Delta f$  and  $\Delta D$  values indicated that  $3 \times$  LPG bound to the silica, as observed in Figure 4c, and was able to support the binding of  $1000$  ng/cm<sup>2</sup> trastuzumab. This immobilized  $3 \times$  LPG–trastuzumab complex was then able to bind the HER2 antigen that was spiked into 25% human serum. The mass of trastuzumab immobilized to  $3 \times$  LPG was approximately  $1000$  ng/cm<sup>2</sup> to which approximately  $1250$  ng/cm<sup>2</sup> HER2 antigen was immobilized (Table 4). These results indicated less immobilization than reported for  $4 \times$  LPG [31]. These results are in accordance with the stoichiometric data presented above, where as many as twice the number of  $4 \times$  LPG molecules were bound on the silica surface per cm<sup>2</sup> as compared to  $3 \times$  LPG.



**Figure 5.** Representative  $\Delta f$  and  $\Delta D$  versus time plot for  $3 \times$  LPG binding to silica followed by the immobilization of trastuzumab and the HER2 antigen spiked in 25% human serum (HS) measured at the 3rd overtone by QCM-D. The measurements were performed at a flow rate of  $150 \mu\text{L}/\text{mL}$  at  $22^\circ\text{C}$ .  $1 \times$  PBS rinses were applied between protein injections.

**Table 4.** Viscoelastic parameters for the interaction between silica-bound  $3 \times$  LPG and trastuzumab followed by HER2 binding.

Immobilized Truncated Derivative	Thickness of Bound Protein (nm)	Mass of Protein Deposited ( $\text{ng}/\text{cm}^2$ )	Thickness of Bound Trastuzumab (nm)	Mass Deposited for Bound Trastuzumab ( $\text{ng}/\text{cm}^2$ )	Thickness of Bound HER2 (nm)	Mass Deposited for Bound HER2 ( $\text{ng}/\text{cm}^2$ )
$3 \times$ LPG	$5.76 \pm 0.22$	$681 \pm 32$	$15.0 \pm 0.20$	$1019 \pm 48$	$20.0 \pm 0.4$	$1251 \pm 51$

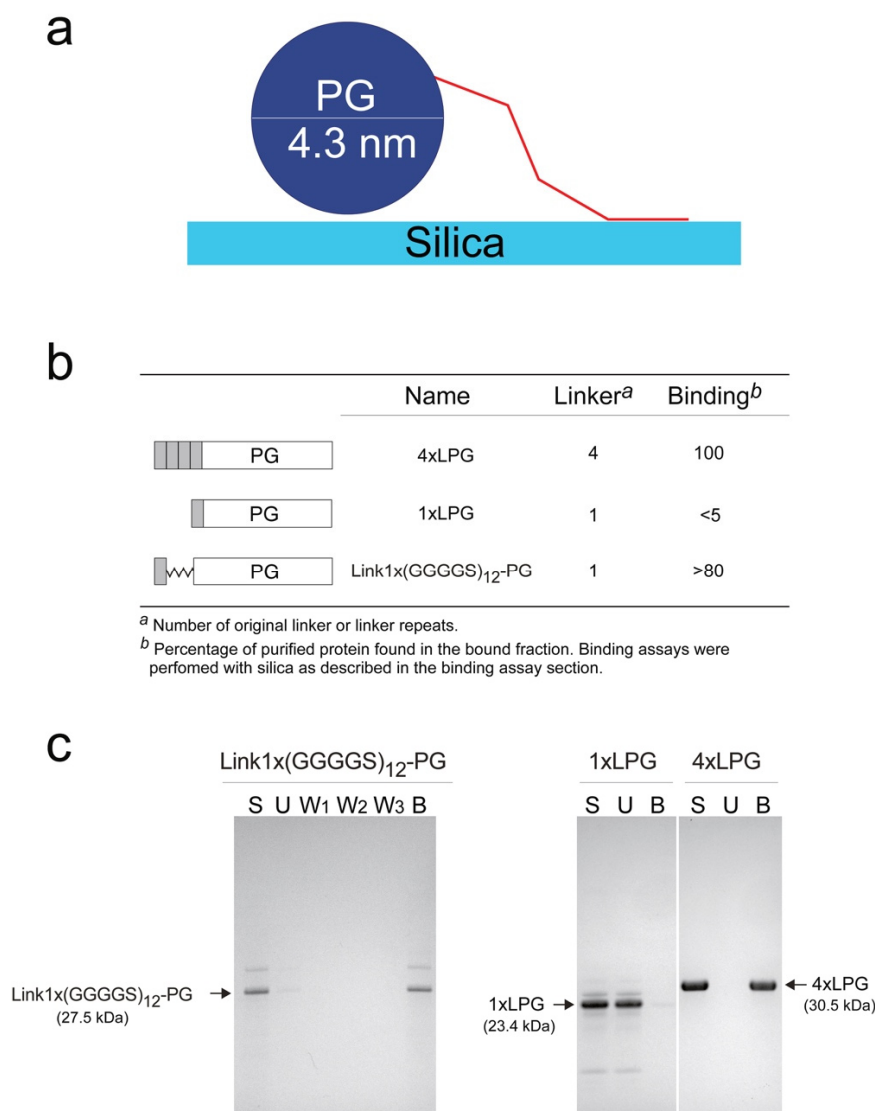
### 3.5. Physiochemical Properties of $4 \times$ LPG Derivatives and Linker Repeats

Supplementary Table S1 (Supplementary Materials) summarizes the physiochemical properties (molecular weight, charge, hydrophobicity, and isoelectric point) of the  $4 \times$  LPG derivatives and the different linker repeats. The surface of silica is negative and the adsorption of proteins to its surface is influenced by the sum of attraction and repulsion electrostatic forces. The adsorption of proteins will be driven mainly by Coulomb's electrostatic attractions at pH values below the pI (proteins will carry an overall positive charge). However, adsorption will be driven by the sum of the Coulombic repulsion and attractive forces (e.g. hydrophobic interactions) at pH values above the pI as proteins and silica will be negatively charged. At pH 7.0 (above pI), PG and all the derivatives (Table S1, Supplementary Materials) carry an overall negative charge. Considering the charge contribution by the linker region alone, the positive charge increases with oligomerization. Accordingly,  $4 \times$  L is highly positively charged at pH 7.0 (charge of 12.9) when compared to PG (charge of  $-14$ ). Silica has a negatively charged surface for the interaction of positively charged molecules, and this may indicate that the linker oligomerization and the concomitant increase in positive charge are the main drive for a strong binding to silica through electrostatic attraction forces. Taniguchi et al. [41] reported that the binding of a silica-binding tag (Si-Tag) was pH-dependent. Due to the presence of 63 positively charged amino acid residues, the peptide displayed a pI of 10.9, and a maximum binding to the silica surface was obtained at pH 8.0, when the Si-tag carried a net positive charge. Sarikaya et al. [42] found that the high affinity of polypeptides to inorganic surfaces was the result of a net high charge and a large proportion of basic amino acid residues.

### 3.6. Introduction of a (GGGGS)<sub>n</sub> Linker

Although the linker sequence is important for the binding of 3 × LPG and 4 × LPG to silica, it is not clear whether 4 × LPG binds better simply because it has a repeat of the binding peptide or simply because it is longer. Figure 6 shows a schematic diagram of the relative sizes of the components, which suggests that the sequence of the 3<sup>rd</sup> repeat unit may be less important than its length. We, therefore, replaced the first three linker units with a glycine-rich flexible soluble sequence (GGGGS)<sub>12</sub> [43] (Figure 6b) by expressing the fusion protein Link1 × (GGGGS)<sub>12</sub>-PG recombinantly in *E. coli*. The purified Link1 × (GGGGS)<sub>12</sub>-PG has the same pI (4.69), overall charge at pH 7.0 (−13.8), and very similar grand average hydropathicity (−0.64) as 1 × LPG (Table S1, Supplementary Materials). The silica binding assay is a well-established technique [1,6], which allows the semi-quantitative percentage of Link1 × (GGGGS)<sub>12</sub>-PG that remained bound to silica after several wash steps to be determined. Silica binding assays (Figure 6c) indicated that the length of the repeat sequence is an important factor for its solid binding. The insertion of the synthetic (GGGGS)<sub>12</sub> sequence into 1 × LPG dramatically increases its binding affinity from below the level of quantitation (Figure 4) to >80%. These qualitative initial results seem to indicate an important role for the length of the peptide (distance from the repeats) in an efficient binding to silica. The CD data, which showed that the linkers, as well as some of the N-terminus residues, were unfolded, support the fact that a key role of the middle linkers is simply to let the “fishing hook” of the final silica binding peptide meet the surface. However, in its current form, this experiment shows that only introducing the equivalent length of three linker repeats in the form of (GGGGS)<sub>12</sub> to 1 × LPG (Link1 × (GGGGS)<sub>12</sub>-PG) is enough to restore the binding affinity towards silica almost to the same level as 4 × LPG [1].

The introduction of (GGGGS)<sub>12</sub> to 1 × LPG results in Link1 × (GGGGS)<sub>12</sub>-PG, which seems to function as an analogue of 4 × LPG but carries only a single sequence of the silica-binding peptide rather than four repeats. This (GGGGS)<sub>12</sub> spacer addition was enough to convert 1 × LPG, which displayed negligible binding to silica in the silica binding assays [1], into a fusion protein displaying at least qualitatively comparable binding affinity to silica as 4 × LPG [1]. In terms of linker oligomerization, the additional three tandem repeats of the linker may just act as spacer regions that allow the single linker sequence to interact with the silica surface. SBPs are usually present in an unfolded state and adsorption and conformational adaptation to the substrate binding interface are fast and simultaneous events [45]. Over 76% of the amino acid residues of 4 × LPG have the tendency to promote structural disorder [46] and confer low overall hydrophobicity and high net charge, which have been reported for other polypeptide sequences exhibiting affinity to inorganic materials [44]. Thus, the sequence repeats may provide enhanced flexibility and plasticity to adopt an optimal conformation for binding substrates with differently shaped topologies or surfaces.



**Figure 6.** (a) Schematic diagram illustrating relative sizes of components assuming  $\sim 2$  nm per 21 residues of each linker sequence [44] (b) Architecture of the fusion protein with and without the glycine-rich spacer and their subsequent zeolite-binding assays. (c) sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE) of silica binding assays. Arrows indicate position of Link1  $\times$  (GGGGS)<sub>12</sub>-PG, 1  $\times$  LPG, and 4  $\times$  LPG on the gel. 1  $\times$  LPG and 4  $\times$  LPG SDS-PAGE data in (b) and (c) were from [1].

#### 4. Conclusions

There is increased evidence of the contribution of mineral-templated self-assembling systems, including biomolecular self-assemblies of peptides, on the origins of life on Earth and the complexity of early living systems [47]. Recently, SBPs have been used increasingly as molecular building blocks in several nanobiotechnological applications and have become key molecules for the engineering of biomimetic and bioinspired materials. More biophysical and quantitative analysis of the interaction between SBPs and their inorganic substrates are required to understand their fundamental binding mechanism and expand their range of practical applications. The bifunctional fusion protein 4  $\times$  LPG contains a linker made of a 4  $\times$  21-amino acid sequence repeat that displays high binding affinity towards silica. CD and fluorescence spectroscopy studies indicated no apparent negative effect to the secondary structure and chemical stability of the partner PG upon truncation of the associated linker sequence repeats. However, QCM-D analyses suggested that a linker oligomerization is an important parameter for a strong and stable binding to the silica. Removal of the linker sequence

below three repeats drastically reduced ( $2 \times$  LPG) or completely abolished ( $1 \times$  LPG) the silica-binding property of the  $4 \times$  LPG. A 12-repeat glycine rich spacer (GGGGS)<sub>12</sub> fused between the single linker sequence and PG ( $1 \times$  LPG) to create a fusion protein of the same length as  $4 \times$  LPG with  $1 \times$  L available for surface binding, Link1 $\times$ (GGGGS)<sub>12</sub>-PG, restored the silica binding property of the truncated derivative  $1 \times$  LPG. The CD data, which showed that the linkers, as well as some of the N-terminus residues, were unfolded, support the conclusion that a key role of the middle linkers is simply to let the “fishing hook” of the final silica binding peptide meet the surface. These preliminary qualitative data provide further motivation to investigate not only oligomerization but also the length of the linker as potentially important parameters for strong silica binding. The results presented here also suggest that SBP oligomerization or incorporation of flexible interdomain linkers like (GGGGS)<sub>n</sub> can be used to tailor the adsorption and affinity of a given sequence to a solid material. This provides an additional manipulation tool for future applications requiring surface-specific SBPs that act as molecular linkers and assemblers.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2079-4991/10/5/1070/s1>, Equation S1: Derivation Equation (1) to determine derandomized molar residue  $\Delta\epsilon$  of derivatives, Figure S1: SOMSpec output for PG,  $1 \times$  LPG,  $2 \times$  LPG,  $3 \times$  LPG, and  $4 \times$  LPG, Figure S2:  $\Delta D$  vs. time for the 3<sup>rd</sup> overtone of the absorbed  $1 \times$  LPG and  $2 \times$  LPG on silica-coated QCM-D crystals at various concentrations (3.27–654 nM), Figure S3: Langmuir adsorption isotherm for the adsorption of  $3 \times$  LPG to silica surface, Figure S4:  $D_f$  plot for the binding of  $3 \times$  LPG to silica, Table S1: Truncated derivatives of  $4 \times$  LPG with different linker repeats and their physicochemical properties.

**Author Contributions:** Conceptualization, R.B., A.C. and A.S.; methodology, R.B., A.C., S.C.G., Z.E., M.S.L., A.R. and A.S.; data collection, R.B., Z.E. and S.C.G.; data analysis, R.B., S.C.G., Z.E., M.S.L. and A.R.; writing—original draft preparation, R.B.; A.R.; M.S.L. and A.S.; writing—review and editing, R.B., A.C., S.C.G., Z.E., M.S.L., A.R. and A.S.; project administration, A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** R.B. is supported by an international Macquarie University Research Excellence Scholarship (iMQRES). A.C. is supported by a Cancer Institute New South Wales Early Career Fellowship (Project Number: ECF171114). This work was supported by the Australian Research Council (CE140100003).

**Acknowledgments:** In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Sunna, A.; Chi, F.; Bergquist, P.L. A linker peptide with high affinity towards silica-containing materials. *New Biotechnol.* **2013**, *30*, 485–492. [[CrossRef](#)] [[PubMed](#)]
2. Care, A.; Bergquist, P.L.; Sunna, A. Solid-binding peptides: Smart tools for nanobiotechnology. *Trends Biotechnol.* **2015**, *33*, 259–268. [[CrossRef](#)] [[PubMed](#)]
3. Rothenstein, D.; Claasen, B.; Omiecienski, B.; Lammel, P.; Bill, J. Isolation of ZnO-Binding 12-mer Peptides and Determination of Their Binding Epitopes by NMR Spectroscopy. *J. Am. Chem. Soc.* **2012**, *134*, 12547–12556. [[CrossRef](#)] [[PubMed](#)]
4. Heinz, H.; Farmer, B.L.; Pandey, R.B.; Slocik, J.M.; Patnaik, S.S.; Pachter, R.; Naik, R.R. Nature of Molecular Interactions of Peptides with Gold, Palladium, and Pd–Au Bimetal Surfaces in Aqueous Solution. *J. Am. Chem. Soc.* **2009**, *131*, 9704–9714. [[CrossRef](#)] [[PubMed](#)]
5. Kumada, Y.; Tokunaga, Y.; Imanaka, H.; Imamura, K.; Sakiyama, T.; Katoh, S.; Nakanishi, K. Screening and Characterization of Affinity Peptide Tags Specific to Polystyrene Supports for the Orientated Immobilization of Proteins. *Biotechnol. Prog.* **2006**, *22*, 401–405. [[CrossRef](#)]
6. Care, A.; Chi, F.; Bergquist, P.L.; Sunna, A. Biofunctionalization of silica-coated magnetic particles mediated by a peptide. *J. Nanopart. Res.* **2014**, *16*, 2543. [[CrossRef](#)]

7. Tang, Z.; Palafox-Hernandez, J.P.; Law, W.-C.; Hughes, Z.E.; Swihart, M.T.; Prasad, P.; Knecht, M.R.; Walsh, T.R. Biomolecular Recognition Principles for Bionanocombinatorics: An Integrated Approach to Elucidate Enthalpic and Entropic Factors. *ACS Nano* **2013**, *7*, 9632–9646. [[CrossRef](#)]
8. Rimola, A.; Sodupe, M.; Ugliengo, P. Affinity Scale for the Interaction of Amino Acids with Silica Surfaces. *J. Phys. Chem. C* **2009**, *113*, 5741–5750. [[CrossRef](#)]
9. Sapsford, K.E.; Algar, W.S.; Berti, L.; Gemmill, K.B.; Casey, B.J.; Oh, E.; Stewart, M.H.; Medintz, I.L. Functionalizing nanoparticles with biological molecules: Developing chemistries that facilitate nanotechnology. *Chem. Rev.* **2013**, *113*, 1904–2074. [[CrossRef](#)]
10. Nel, A.E.; Mädler, L.; Velegol, D.; Xia, T.; Hoek, E.M.V.; Somasundaran, P.; Klaessig, F.; Castranova, V.; Thompson, M. Understanding biophysicochemical interactions at the nano–bio interface. *Nat. Mater.* **2009**, *8*, 543–557. [[CrossRef](#)]
11. Naik, R.R.; Stringer, S.J.; Agarwal, G.; Jones, S.E.; Stone, M.O. Biomimetic synthesis and patterning of silver nanoparticles. *Nat. Mater.* **2002**, *1*, 169–172. [[CrossRef](#)]
12. Yamaguchi, A.; Isozaki, K.; Nakamura, M.; Takaya, H.; Watanabe, T. Discovery of 12-mer peptides that bind to wood lignin. *Sci. Rep.* **2016**, *6*, 21833. [[CrossRef](#)] [[PubMed](#)]
13. Sultan, A.M.; Westcott, Z.; Hughes, Z.E.; Palafox-Hernandez, J.P.; Giesa, T.; Puddu, V.; Buehler, M.J.; Perry, C.C.; Walsh, T.R. Aqueous Peptide–TiO<sub>2</sub> Interfaces: Isoenergetic Binding via Either Entropically or Enthalpically Driven Mechanisms. *ACS Appl. Mater. Interfaces* **2016**, *8*, 18620–18630. [[CrossRef](#)] [[PubMed](#)]
14. Jaworski, J.W.; Raorane, D.; Huh, J.H.; Majumdar, A.; Lee, S.-W. Evolutionary Screening of Biomimetic Coatings for Selective Detection of Explosives. *Langmuir* **2008**, *24*, 4938–4943. [[CrossRef](#)] [[PubMed](#)]
15. Sawada, T.; Okeya, Y.; Hashizume, M.; Serizawa, T. Screening of peptides recognizing simple polycyclic aromatic hydrocarbons. *Chem. Commun.* **2013**, *49*, 5088. [[CrossRef](#)] [[PubMed](#)]
16. Suzuki, Y.; Shindo, H. Binding sites and structure of peptides bound to SiO<sub>2</sub> nanoparticles studied by solution NMR spectroscopy. *Polym. J.* **2018**, *50*, 989. [[CrossRef](#)]
17. Gwak, Y.; Park, J.-I.; Kim, M.; Kim, H.S.; Kwon, M.J.; Oh, S.J.; Kim, Y.-P.; Jin, E. Creating Anti-icing Surfaces via the Direct Immobilization of Antifreeze Proteins on Aluminum. *Sci. Rep.* **2015**, *5*, 12019. [[CrossRef](#)]
18. Yu, J.; Becker, M.L.; Carri, G.A. The Influence of Amino Acid Sequence and Functionality on the Binding Process of Peptides onto Gold Surfaces. *Langmuir* **2011**, *28*, 1408–1417. [[CrossRef](#)]
19. Heffernan, R.; Paliwal, K.; Lyons, J.; Dehzangi, A.; Sharma, A.; Wang, J.; Sattar, A.; Yang, Y.; Zhou, Y. Improving prediction of secondary structure, local backbone angles and solvent accessible surface area of proteins by iterative deep learning. *Sci. Rep.* **2015**, *5*, 11476. [[CrossRef](#)]
20. Sun, T.; Zhou, B.; Lai, L.; Pei, J. Sequence-based prediction of protein protein interaction using a deep-learning algorithm. *BMC Bioinform.* **2017**, *18*, 277. [[CrossRef](#)]
21. Walsh, T.R.; Knecht, M.R. Biointerface Structural Effects on the Properties and Applications of Bioinspired Peptide-Based Nanomaterials. *Chem. Rev.* **2017**, *117*, 12641–12704. [[CrossRef](#)] [[PubMed](#)]
22. Kuang, Z.; Kim, S.N.; Crookes-Goodson, W.J.; Farmer, B.L.; Naik, R.R. Biomimetic chemosensor: Designing peptide recognition elements for surface functionalisation of carbon nanotube field effect transistors. *ACS Nano* **2010**, *1*, 452–458. [[CrossRef](#)] [[PubMed](#)]
23. Seker, U.O.S.; Wilson, B.; Kulp, J.; Evans, J.S.; Tamerler, C.; Sarikaya, M. Thermodynamics of Engineered Gold Binding Peptides: Establishing the Structure–Activity Relationships. *Biomacromolecules* **2014**, *15*, 2369–2377. [[CrossRef](#)] [[PubMed](#)]
24. Cho, N.-H.; Cheong, T.-C.; Min, J.H.; Wu, J.H.; Lee, S.J.; Kim, D.; Yang, J.-S.; Kim, S.; Kim, Y.K.; Seong, S.-Y. A multifunctional core–shell nanoparticle for dendritic cell-based cancer immunotherapy. *Nat. Nanotechnol.* **2011**, *6*, 675–682. [[CrossRef](#)] [[PubMed](#)]
25. Lu, Y.; Zhao, J.; Zhang, R.; Liu, Y.; Liu, D.; Goldys, E.M.; Yang, X.; Xi, P.; Sunna, A.; Lu, J.; et al. Tunable lifetime multiplexing using luminescent nanocrystals. *Nat. Photonics* **2013**, *8*, 32–36. [[CrossRef](#)]
26. Liang, L.; Care, A.; Zhang, R.; Lu, Y.; Packer, N.H.; Sunna, A.; Qian, Y.; Zvyagin, A.V. Facile Assembly of Functional Upconversion Nanoparticles for Targeted Cancer Imaging and Photodynamic Therapy. *ACS Appl. Mater. Interfaces* **2016**, *8*, 11945–11953. [[CrossRef](#)]
27. Hall, V.; Sklepari, M.; Rodger, A. Protein Secondary Structure Prediction from Circular Dichroism Spectra Using a Self-Organizing Map with Concentration Correction. *Chirality* **2014**, *26*, 471–482. [[CrossRef](#)]



28. Sklepari, M.; Rodger, A.; Reason, A.; Jamshidi, S.; Prokes, I.; Blindauer, C.A. Biophysical characterization of a protein for structure comparison: Methods for identifying insulin structural changes. *Anal. Methods* **2016**, *8*, 7460–7471. [[CrossRef](#)]
29. Lees, J.G.; Miles, A.J.; Wien, F.; Wallace, B. A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* **2006**, *22*, 1955–1962. [[CrossRef](#)]
30. San-Miguel, M.A.; Marrington, R.; Rodger, P.M.; Rodger, A.; Robinson, C. An Escherichia coli twin-arginine signal peptide switches between helical and unstructured conformations depending on the hydrophobicity of the environment. *JBC J. Boil. Inorg. Chem.* **2003**, *270*, 3345–3352. [[CrossRef](#)]
31. Bansal, R.; Elgundi, Z.; Care, A.; Goodchild, S.C.; Lord, M.; Rodger, A.; Sunna, A. Elucidating the Binding Mechanism of a Novel Silica-Binding Peptide. *Biomolecules* **2019**, *10*, 4. [[CrossRef](#)] [[PubMed](#)]
32. Seker, U.O.S.; Wilson, B.; Sahin, D.; Tamerler, C.; Sarikaya, M. Quantitative Affinity of Genetically Engineered Repeating Polypeptides to Inorganic Surfaces. *Biomacromolecules* **2009**, *10*, 250–257. [[CrossRef](#)] [[PubMed](#)]
33. Lord, M.; Whitelock, J.; Simmons, A.; Williams, R.L.; Milthorpe, B.K. Fibrinogen adsorption and platelet adhesion to silica surfaces with stochastic nanotopography. *Biointerphases* **2014**, *9*, 41002. [[CrossRef](#)] [[PubMed](#)]
34. Liu, S.X.; Kim, J.-T. Application of Kevin–Voigt Model in Quantifying Whey Protein Adsorption on Polyethersulfone Using QCM-D. *J. Lab. Autom.* **2009**, *14*, 213–220. [[CrossRef](#)]
35. Hovgaard, M.B.; Dong, M.; Otzen, D.E.; Besenbacher, F. Quartz Crystal Microbalance Studies of Multilayer Glucagon Fibrillation at the Solid-Liquid Interface. *Biophys. J.* **2007**, *93*, 2162–2169. [[CrossRef](#)]
36. Goward, C.R.; Irons, L.I.; Murphy, J.P.; Atkinson, T. The secondary structure of protein G', a robust molecule. *Biochem. J.* **1991**, *274*, 503–507. [[CrossRef](#)]
37. Provencher, S.W.; Gloeckner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **1981**, *20*, 33–37. [[CrossRef](#)] [[PubMed](#)]
38. Sreerama, N.; Woody, R.W. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON and CDSSTR Methods with an Expanded Reference Set. *Anal. Biochem.* **2000**, *287*, 252–260. [[CrossRef](#)]
39. Lian, L.-Y.; Derrick, J.P.; Sutcliffe, M.; Yang, J.; Roberts, G. Determination of the solution structures of domains II and III of protein G from Streptococcus by 1H nuclear magnetic resonance. *J. Mol. Boil.* **1992**, *228*, 1219–1234. [[CrossRef](#)]
40. Brown, S. Metal-recognition by repeating polypeptides. *Nat. Biotechnol.* **1997**, *15*, 269–272. [[CrossRef](#)]
41. Taniguchi, K.; Nomura, K.; Hata, Y.; Nishimura, T.; Asami, Y.; Kuroda, A. The Si-tag for immobilizing proteins on a silica surface. *Biotechnol. Bioeng.* **2007**, *96*, 1023–1029. [[CrossRef](#)] [[PubMed](#)]
42. Sarikaya, M.; Tamerler, C.; Jen, A.-Y.; Schulten, K.; Baneyx, F. Molecular biomimetics: Nanotechnology through biology. *Nat. Mater.* **2003**, *2*, 577–585. [[CrossRef](#)] [[PubMed](#)]
43. Chen, X.; Zaro, J.; Shen, W.C. Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev.* **2013**, *65*, 1357–1369. [[CrossRef](#)] [[PubMed](#)]
44. Kohn, J.E.; Millett, I.S.; Jacob, J.; Zagrovic, B.; Dillon, T.M.; Cingel, N.; Dothager, R.S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T.R.; et al. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12491–12496. [[CrossRef](#)]
45. Ikeda, T.; Kuroda, A. Why does the silica-binding protein “Si-tag” bind strongly to silica surfaces? Implications of conformational adaptation of the intrinsically disordered polypeptide to solid surfaces. *Colloids Surf. B Biointerfaces* **2011**, *86*, 359–363. [[CrossRef](#)]
46. Campen, A.; Williams, R.M.; Brown, C.J.; Meng, J.; Uversky, V.N.; Dunker, A.K. TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder. *Protein Pept. Lett.* **2008**, *15*, 956–963. [[CrossRef](#)]
47. Gillams, R.J.; Jia, T.Z. Mineral Surface-Templated Self-Assembling Systems: Case Studies from Nanoscience and Surface Science towards Origins of Life Research. *Life* **2018**, *8*, 10. [[CrossRef](#)]

