PCCP

From phage display to structure: interplay of Enthalpy and Entropy in binding of LDHSLHS polypeptide to silica.

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-11-2018-007011.R2
Article Type:	Paper
Date Submitted by the Author:	22-Jan-2019
Complete List of Authors:	Oliver, Daniel; Nottingham Trent University, School of Science and Technology michaelis, monika; Nottingham Trent University, School of Science and Technology; University of Bremen, Hybrid Materials Interfaces Group Heinz, Hendrik; University of Colorado-Boulder, Chemical and Biological Engineering Volkov, Victor; Nottingham Trent University, School of Science and Technology Perry, Carole; Nottingham Trent University, School of Science and Technology

Chemical Science

ARTICLE

From phage display to structure: interplay of Enthalpy and Entropy in binding of LDHSLHS polypeptide to silica.

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Daniel Oliver,[a] Monika Michaelis,[a,b] Hendrik Heinz,[c] Victor V. Volkov[a] and Carole C. Perry[a]*

Polypeptide based biosilica composites show promise as next generation multi-functional nano-platforms for diagnostics and bio-catalytic applications. Following identification of a strong silica binder (LDHSLHS) by phage display, we conduct structural analysis of the polypeptide at the interface with amorphous silica nanoparticles in an aqueous environment. Our approach relies on modelling of Infrared and Raman spectral responses using predictions of molecular dynamics simulations and quantum studies of the normal modes for several potential structures. By simultaneously fitting both Infrared and Raman responses in the Amide spectral region, we show that the main structural conformer has a beta-like central region and helix-twisted terminals. Classical simulations, as conducted previously (Chem. Mater., 2014, 26, 5725), predict that association of the main structure with the interface is stimulated by electrostatic interactions though surface binding also requires spatially distributed sodium ions to compensate negatively charged acidic silanol groups. Accordingly, diffusion of sodium ions would contribute to a stochastic character of the peptides association with the surface. Consistent with the described dynamics at the interface, results from isothermal titration calorimetry (ITC) confirm significant enhancement of polypeptide binding to silica under higher concentrations of Na^+. The results of this study suggest that the tertiary structure of a phage capsid protein plays a significant role in regulating the conformation of peptide LDHSLHS, increasing its binding to silica during the phage display process. The results presented here support design-led engineering of polypeptide-silica nanocomposites for bio-technological applications.

Introduction

Efficient inclusion of inorganic microstructures in living organisms has occurred since the early days of evolution[1] where mineral inclusion increased the structural stability of single-celled organisms including diatoms through to higher plants.[2-4] The ability to deposit structured inorganic materials in living organisms lead to the development of internal skeletons[5,6] – the signature of all vertebrates.[1,7] Among the many possible mineral inclusions, SiO_2 is important in cellular/tissue morphogenesis.[1,8] The mineral is has varied roles depending on polymorph and crystallinity: where according to its form, silica may play a role in cancer-genesis or it may be non-hazardous.[4,9]

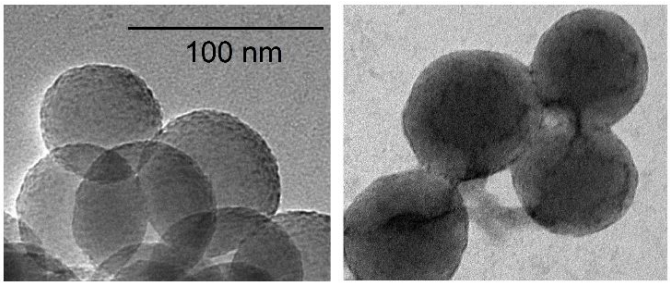

Fig. 1 TEM images of silica nanoparticles alone (left) and in presence of phages (right).

Additionally, in recent decades, silica as a wide-gap semiconductor,[10-12] has inspired numerous initiatives in engineering biosensors[13,14] and developing silica based vehicles for drug delivery.[15-17]

Borrowing from examples of hierarchical bio-mineral structures in living organisms,[18-20] we may plan next generation silica-based polypeptide functionalized sub-microscale platforms for analytical applications in living matter.[21] As the first step, we used the phage display technique[22,23] to determine silica specific polypeptides,[24] see **Fig. 1**. The technique relies on identification of generic sequences (from combinatorial libraries) that allow polypeptides expressed at the surface of a phage to bind the phage to a target site – silicon oxide interface, as is the focus of our study.[25] Furthermore, in a recent study we addressed binding mechanisms of LDHSLHS,

[a] *Interdisciplinary Biomedical Research Centre, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom. E-mail: carole.perry@ntu.ac.uk*
[b] *Hybrid Materials Interfaces Group, University of Bremen, Am Fallturm 1, Bremen D-28359, Germany.*
[c] *Chemical and Biological Engineering, College of Engineering & Applied Science, University of Colorado Boulder, Boulder, Colorado 80309, United States.*

†Electronic supplementary information (ESI) available. It provides details on: a) details on MD simulations and extracted structural properties along trajectories b) details of DFT studies on the normal modes of imidazole ring of methyl terminated histidine under different protonation conditions; c) the details on the normal modes of the structural cases including relative contributions of carbonyls and carboxyl groups into the described normal modes; d) results of isothermal titration calorimetry.

AFILPTG, and KLPGWSG hepta-peptides with amorphous SiO_2 nanoparticles.[26] Despite the fact that the outcomes of phage display provides insight on the strong binding capacity of these polypeptides,[24] the results of our studies indicate that structural and functional properties of genetically programmed segments of phage surface proteins and properties of the corresponding polypeptides in solution may not be the same. Specifically, in our studies, we observed a complex interplay of electrostatic and hydrophobic forces that modify binding efficiency.[27]

It should be expected that the conformational space and dynamics of short and polar terminated polypeptides, when in an aqueous phase, are different from the corresponding structures when present as parts of proteins at the surface of phages. To explore the situation better, we have to address the molecular mechanisms of interactions and structural behaviour at the bio-inorganic interface. This, however, requires structural analysis of polypeptides in complex environments.

Extraction of secondary structure in relatively small polypeptides in heterogeneous disordered environments and at interfaces is one of the main challenges in contemporary biophysics and physical chemistry. This is not trivial: indeed, X-ray and electron scattering experiments have limited application in the case of disordered samples. For relatively short polypeptides, which typically lack strong and specific folding, use of the nuclear Overhauser effect in NMR is rather limited. Of course, detecting homo-nuclear $^3J(HN\text{-}H_\alpha)$ coupling and using the Karplus relation,[28] one may extract Φ angles, but only if the system is at the motional line-narrowing limit. In the case of relatively large polypeptide-inorganic composites, application of the techniques listed is seriously limited. Magic angle spinning NMR experiments may be informative, but at low temperature, only. From this perspective, multidimensional infrared spectroscopy may offer a helpful insight, which could be useful for structural analysis in disordered, heterogeneous systems under physiological conditions. For example, detection of cross-peaks in two-dimensional[29,30] and three-dimensional[31,32] coherent ultrafast Infrared responses demonstrated the possibility to probe structure at time scales specific to interatomic displacements in a complex environment.[33-35]

In this work, however, we follow a different approach: taking advantage of low symmetry, we adopt the non-coincidence of linear infrared and Raman responses.[36] In particular, here, we demonstrate that assisted with theory, one may compare quantitatively predicted and experimental linear Infrared absorption and Raman spectra to extract major and minor secondary structural realizations of polypeptides at complex heterogeneous inorganic interfaces. Specifically, in this study we explore the structural properties of a polypeptide, LDHSLHS that was previously identified by phage display[24] as a strong silica binder. The outline of the article is as follows: first, we report on the experimental results of our Infrared and Raman studies on the polypeptide under different experimental conditions. Second, adopting the results of previously reported MD trajectories,[26,37] we extract the most likely molecular arrangements at the modelled interfaces of silica. Third, we review the results of quantum chemistry studies on the representative structural cases to learn how the anticipated normal modes in the extracted structural cases help us to model the experimental observations. Following this, we suggest plausible structural conformations generated by classic simulations. Using energy and entropy predictions of molecular dynamics for the extracted structural cases and the results of experimental calorimetry studies, we discuss binding mechanisms and thermodynamics of the polypeptide at the silica interface and practical implications for engineering of silica-based bio-composites.

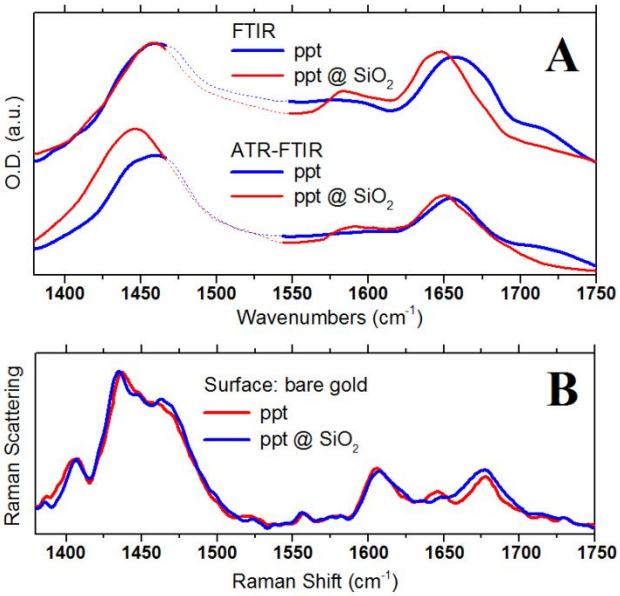

Fig. 2 Experimental optical properties of LDHSLHS polypeptide. (**A**) FTIR and ATR-FTIR spectra of the polypeptide in D_2O (blue lines) and in deuterium oxide solution of SiO_2 nanoparticles (red lines). The dashed lines indicate the ambiguous spectral region at 1467 cm^{-1}, where we subtracted background contributions of HOD bending modes. (**B**) Raman spectra of the polypeptide in deuterium oxide (blue lines) and in D_2O solution of SiO_2 nanoparticles (red lines). All samples contained 50-100 mM NaCl. Details on concentrations of the polypeptide and particles are in experimental and methods.

Results and discussion

Fig. 2 shows the experimentally detected FTIR, ATR-FTIR and Raman spectral responses of LDHSLHS polypeptide in D_2O environment when alone and in the presence of SiO_2 nanoparticles.

The FTIR responses show that upon binding to Silicon oxide: i) there is a decrease in optical density at about 1720 cm^{-1}; ii) the Amide I band at 1660 cm^{-1} experiences both a slight narrowing and a red shift; iii) the intensity of the Amide II mode, at about 1560 cm^{-1}, experiences a slight increase relative to other bands. The presence of optical absorption at 1720 cm^{-1} in FTIR spectra suggests that the aspartic acid side group is protonated. This correlates with the results of Pd measurements, see Experimental section.

The Raman responses show a dominance of the Amide I set of Raman active modes centred as a band at 1675 cm^{-1}. There is a smaller spectral signature of Amide I Raman activity at 1650

cm^{-1}. Upon introduction of SiO_2 nanoparticles, the former becomes narrower, and the latter peak gains intensity.

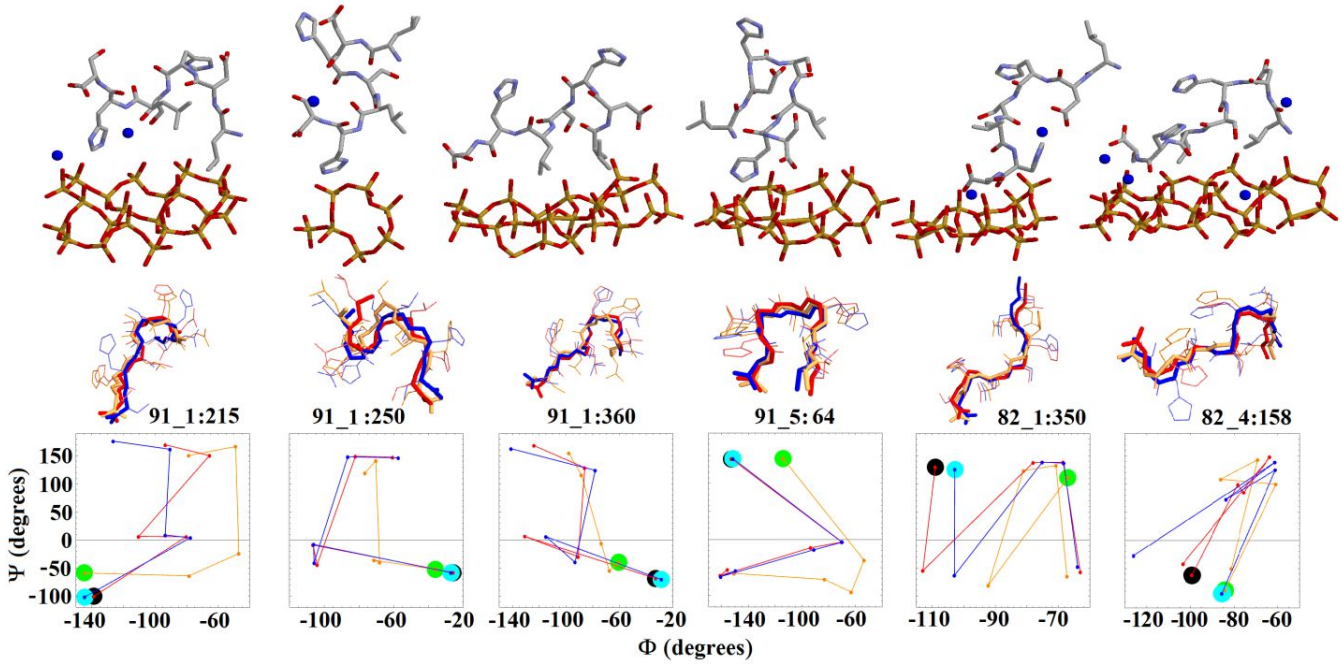

Fig. 3 Structural properties of considered composites as extracted from MD and after optimization with DFT. Upper row: graphic representations of structures of LDHSLHS polypeptide at SiO_2 interface extracted from MD simulations. Middle row: backbones of the polypeptide in the representative structures upon extractions from MD (light yellow), after first DFT optimization when at SiO_2 (red), and after second DFT optimization when alone after removal of the SiO_2 (blue). Lower row: dihedral angles of the five residues (from the N-terminal to the C-terminal) in the expanded region of the Ramachandran angular space for the extracted backbones from the MD simulations (light yellow line), after first and second DFT optimizations (red and blue lines, respectively); green, black and cyan circles indicate the dihedral angles for the first (from the N terminal) residue

There is also a tendency for the Amide II Raman band at 1610 cm^{-1} to increase slightly in the presence of SiO_2 particles. The observed changes of the Amide I band (more obvious in FTIR and ATR-FTIR) could be due to different processes. To predict such, it is helpful to note that the FTIR spectrum of the polypeptide in D_2O alone shows that the Amide I band has spectral densities with peaks at about 1650 and 1670 cm^{-1}. Often, in literature, these peaks are discussed as possible spectral signatures of helical, β-sheet like structures, respectively.[38-40] The ATR-FTIR spectrum does not show the higher frequency spectral component that clearly, however, the spectral heterogeneity of the Amide I band is present. It is possible that upon binding, the polypeptide may experience a redistribution of conformations that favour helical-like ones. At the same time, we cannot exclude a specific restructuring of the conformers that bind to Silicon oxide that experience a change in electronic properties. To approach structural analysis quantitatively, first, we need classical simulations to explore the diversity of the configuration space; second, we need to do quantum calculations of Transition Dipole Moments and Raman tensors for the normal modes of interest. These are the subjects of the following paragraphs.

Structural cases from MD and DFT studies

Using variances of dihedral angles and proximity of selected structural moieties to the SiO_2 interface, we extract six representative structural cases for the polypeptide attached to the Silicon oxide surface, see the upper row of images in **Fig. 3**. The details of extractions, descriptions of the structural cases are provided in the Supporting Information file. The labelling of the extracted cases details the surface ionicity, the trajectory number and the snapshot number, from left to right.

How far do the structures change upon DFT optimization from the structures extracted from the classical simulation environment? The middle row of images in **Fig. 3** demonstrates that upon either the first DFT level quantum theory optimization, when with SiO_2 clusters in water, or for the second optimization, when the polypeptide is alone in water, the backbones do not experience strong deviations from the original extracts. This is also shown in the Ramachandran plots in the lower row of images in **Fig. 3**. All the extracted structures preserve the main character of the secondary folds they possess upon their extraction from MD. The main effect of bringing the structures from the classical simulation environment into the quantum based functional is in a slight unfolding about the Φ angles. Because of the observed sensitivity specific to Φ angles, we may state that the observed small restructurings (upon the step from classic to quantum environment) is only partially due to the extent of the boundary of the extraction volumes. We consider the observed specificity to be related to the challenges in parametrization of the force fields on Φ angles, as reported recently.[41]

The Ramachandran space presentation of the secondary folds of the extracted structures indicates that the dihedral angles of the five central amino acids explore secondary structures in helical and β-like spaces. Switching from one structural space to another shows their flexible character. Results of a number of studies across structural databases indicate that the sequential interplay of α and β like structural realizations provide a rich configuration space for different turns and hairpin motifs. Finding adequate descriptions for such structural realizations is far beyond initially simplifying the sorting of a polypeptide into α-helical and β-like structures.[42-46]

Using the predictions of DFT theory for the optical responses specific to the six considered structures, we suggest two linear combinations (one for Raman and another for the Infrared spectrum) where the weightings of the Raman and Infrared responses for a particular structural case would be kept the same for fitting the two spectral responses. We fit the experimental data with such linear combinations to extract the six weightings that best suit the experimental results.

In **Fig.** 4 we show the results of simultaneous fitting of Raman and Infrared experimental data using constrained weightings for the same structural species. We anticipate that either when next to silica or away, the structure LD82_4:158, where the central region is β-like and the terminals are helix-twisted, is dominant among the considered cases: see **Fig. 4A-4D**. The probabilities (weightings) for this structure are 0.5 and 0.54 when next to silica and when away from it. It is important to notice that since we did not sample the full configurational space in an aqueous environment the results for this case reflect a best possible match, using the adopted theory, when the polypeptide is either about to associate with the surface or just after departure from the surface and would not experience a structural jump either upon association or after disengagement.

In addition to the main contributions due to the structural case LD82_4:158, we predict that, when next to silica, the structural cases where helix dominates at the N-terminal end (LD91_1:360), a random structure (LD82_1:350) and the structure, which start as an unfolded helix (LD91_1:215) may play equal roles, contributing with a probability (weighting) of 0.13. Also, we predict that the structural case LD91_1:250, where an helical motif dominates at the side of N-terminus, may contribute a little, with probability of 0.07. These minor contributions are deduced according to the following reasons (for details see Supporting Information file): infrared response

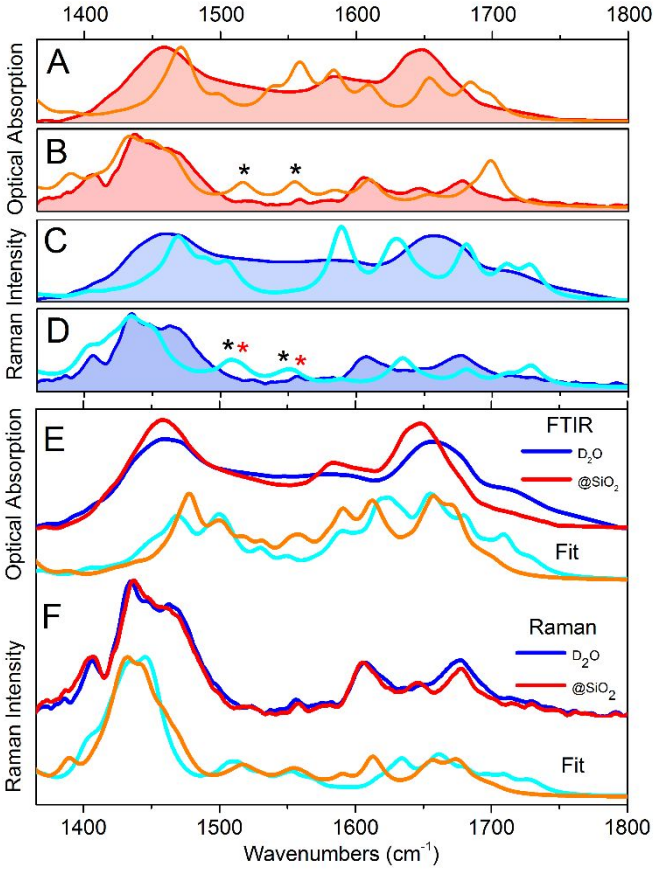

Fig. 4 Simultaneous spectral fit of FTIR and Raman spectra. (**A**) and (**B**): comparisons of calculated Infrared and Raman spectra (orange lines) for the main structural case LD82_4:158, when next to silica assuming 20 cm^{-1} line-width with the experimental results (red lines and pink backgrounds), respectively. (**C**) and (**D**): comparisons of calculated Infrared and Raman spectra (cyan lines) for the main structural case, when in water with the experimental results (blue lines and cobalt backgrounds), respectively. Stars indicate Raman responses of C=C histidine ring stretching modes, see details in Methods and in Supporting information file. (**E**): Comparisons of best calculated Infrared spectra when at silica (orange line) and in water (cyan line) using main and possible minor structural contributions with the corresponding experimental data (red and blue lines). (**F**): Comparisons of best calculated Raman spectra using the same color codes as in (**E**). For details see text and Supporting Information file.

of LD91_1:360 helps to build up optical density at 1500 and 1666 cm^{-1}; infrared absorption of LD82_1:350 contributes with the broad subset of resonances in the spectral region 1550-1630 cm^{-1} and with the resonance at 1494 cm^{-1}; the structural case LD91_1:215 improves the spectral reconstruction with the infrared optical density at 1519 cm^{-1} and with Raman scattering at 1673 cm^{-1}. The residual contribution of the structural case LD91_1:250 may be speculated as it may slightly improve reconstructed infrared absorption at the blue side of the Amide I helical response at 1673 cm^{-1}.

Spectral reconstruction of the responses when in aqueous environment using the same structural cases but properly edited, as described in Experimental and methods section, agree less well with the results of experimental studies as might be anticipated. Particularly, we fail to explain the observed

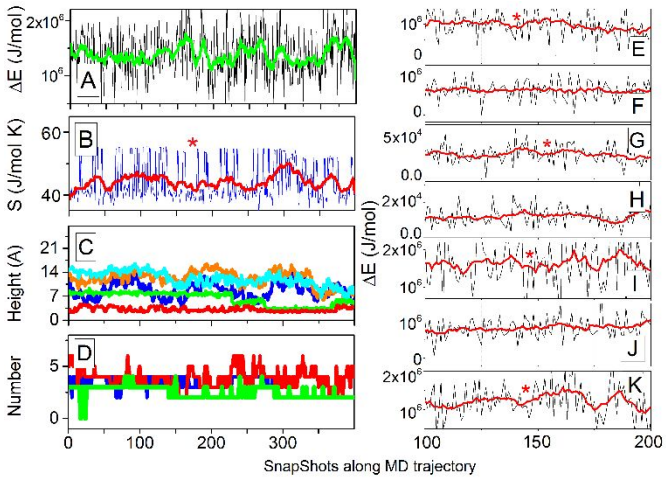

Fig. 5 Thermodynamic and structural properties sampled along a representative MD trajectory for LD82. (**A)** and (**B**): Total Energy changes (in respect to minimum reference) and configuration Entropy of the backbone, respectively. The entropies were calculate adopting the approach by Schlitter,[47] see details in the text. Green and red lines are the smoothed visual guide-lines. Star marks the snapshot of the main structure of the polypeptide at the interface. (**C**): Heights of N-terminal, COO^- of aspartic acid, centre of the side group of histidine-3, centre of the side group of histidine-6, C-terminal (by blue, orange, cyan, green and red lines, respectively) above the mean of H atom distribution at SiO_2 interface in respect to the Z axis of the box. (**D**): Number of Na^+ and SiO^- within the cut-off radius around the polypeptide: red and blue lines, respectively; green line shows number of Na^+ within the cut-off radius around the polypeptide and proximal to the interface. (**E**)-(**K**): sampled along the MD trajectory 4 for LD82 energy changes (in respect to minimum reference) specific to bond lengths (**E**), angles (**F**), dihedral angles (**G**), improper angles (**H**), electrostatic interactions (**I**), van der Waals forces (**J**) and total energy (**K**).

Raman scattering at 1610 cm^{-1}. This is likely to be due to the fact, that when in water, the polypeptide may explore a much wider structural conformational ensemble, which differs from those we extracted from MD simulations when at silica interfaces. Nonetheless, our reconstruction indicates the main contribution of LD82_4:158 (probability 0.54), the next significant helical hairpin structure LD91_5:64 with a probability 0.375, and a residual presence (probability 0.125) of the helical case LD91:360.

Here, it is important to notice that, the DFT predicted Amide 1 normal modes of the main predicted structure LD82_4:158, when next to silica, give rise to three sets of resonances. At the high frequency side, the states at 1767 (1698) and 1744 (1673), cm^{-1}, are mainly due to serine-4 and to the symmetric stretching at leucine-1 and aspartic acid-2, respectively. Here, we show both frequencies, as calculated and scaled (in brackets). In the central region of Amide 1, there is a triplet of states at 1730 (1657), 1729 (1656) and 1726 (1653) cm^{-1}. The higher frequency state is due to localization on histidine-6. The lower frequency doublet is specific to symmetric stretching of carbonyls of histidine-6 and leucine-1 and to antisymmetric activity, which involves histidine-6 stretching out of phase in respect to the synchronous stretching modes of leucine-1 and aspartic acid-2, respectively. The lower frequency Amide 1 states at 1689 (1612), 1669 (1591) and 1639 (1558) cm^{-1}. are due to localized vibrations of leucine-5 and carboxyl activities of aspartic acid side group and the C-terminal, respectively.

Here, it is clear that the often-used description of Amide 1 optical responses in terms of relative contributions of helical, β-sheet like and random coil structures,[38-40] is not applicable for our structural case. Furthermore, such an approach has no meaning in application to the other 9 structures considered in this study: see the Supporting Information file. Further, it is important to note that others predicted possible misinterpretation of spectral data using spectral signatures due to helical, β-sheet like and random coil structures.[40]

Thermodynamics at the silica interface

By further use of the structural extracts we may address the thermodynamics of binding and/or dissociation involving numerous structural realizations of the polypeptide. Here, we evaluate configuration entropies for the polypeptide according to the approach developed by Schlitter.[47] For this task we used polypeptide coordinates of consecutive time-frames along the MD trajectories. Prior to expressing covariance matrices for eigenvalue analysis, first, each structure was positioned according to its centre of mass. Next, its coordinates were projected into axes of inertia tensor specific to the structure. In **Fig. 5A** and **5B** we present the changes of energies and configurational entropies for the main structure along its trajectory. We see that the main contributing structure LD82_4:158 is, overall, in a long lasting (between snapshots 100 and 230), therefore, relatively stable Entropic valley specific to the polypeptide. Accordingly, it is important to question if we should expect any Enthalpy compensation for the predicted binding. The demonstrated total energy for the trajectory, suggests, overall, an unfavourable increase of energy. This, however, is the sum value sampled by MD. In **Fig. 5E-5K**, we specify contributions of different components in the modelled system in the expanded view for the range between snapshots 100 and 200. These indicate that upon binding, the energy specific to dihedral angles and the energy due to electrostatic interaction would contribute to binding. At the same time, at the time-frame 158, van der Waals interactions are the most unfavourable for binding. With subsequent (after time-frame 158) increase of energy due to electrostatic interactions, the polypeptide departs from the surface.

It is interesting that over the considered time-frame (during the trajectory), the polypeptide orients with its negative C-terminal to be oriented towards the surface in a region, where the number of $Si\text{-}O^-$ defects and Na^+ are equally present in proximity to the polypeptide, see **Fig. 5C** and **5D**. However, beginning from snap shot 148, the number of Na^+ around the polypeptide and, at the same time, close to the interface (green line, **Fig. 5D**) start to decrease. The decrease is due to the diffusion of sodium ions from the contact region of the polypeptide and the surface. This changes the charge balance at the contact point, leaving negatively charged acidic silanol groups uncompensated. The resultant unfavourable electrostatic interactions would stimulate dissociation of the peptide from the surface. It is apparent that fluctuating engage–

Table 1. Association constant, ΔH, ΔS and ΔG for the polypeptide binding to silica nanoparticles under low and relatively high salinity.

NaCl (mol)	K_a (mol^{-1}) 10^{-4}	ΔH (kJ mol^{-1})	ΔS(kJ K^{-1} mol^{-1})	ΔG (kJ mol^{-1})
0	1.1 ± 0.1	-3.4 ± 0.1	0.066 ± 0.001	-33.1 ± 0.3
0.2	34 ± 12	-161 ± 6	-0.43 ± 0.02	-41.5 ± 0.7

disengage character of the polypeptide association with the surface is according to electrostatic interactions that are modulated by slow diffusion processes and entropy losses due to structural constraints of the backbone when next to the surface. In our previous study,[27] using fluorescence titration, we proposed that van-der-Waals interactions could play a role to help association of the polypeptide with a negatively charged surface. In our current study we extend the picture to include the possible complexity of dynamics next to the surface due to diffusion of sodium ions.

Extractions of the eigenvalues for the force field components (see **Fig. 5E-K**) are characteristic for the whole simulation box and only indicative in respect to dynamics of binding, which is local. To probe the predicted sensitivity of binding to the relative proximity of Na^+ ions, we experimentally measure the changes of thermodynamic functions under low and relatively high concentration of NaCl, Table 1. Considering the particles size, the surface density of defects (20% of ≡Si-O- at the surface), and the concentrations of particles used; the concentration of 0.2M of NaCl would provide a sodium ion within 2 nm next to a surface defect at any time. The results in Table 1, indeed, indicate that upon salinity increase, the Enthalpic component favourable for binding increases fifty times and this compensates the observed slight Entropic decrease. This data agrees with our FTIR and Raman studies where we observed spectral changes in both responses only in the presence of NaCl as we describe in the Experimental and methods section. The ITC experimental data has some support in weak contributing modulations in energies specific to bond lengths, dihedral angles, and electrostatic interactions, as shown with red stars in **Fig 5E**, **5I** and **5G**, respectively. As we already noticed, the anticipated binding has some correlation with a relative decrease of Entropy associated with the backbone fluctuation dynamics (the backbone fluctuates less) during the time of association with the surface, see **Fig. 5B**. Here, it is necessary to notice that due to averaging over the simulation box, one cannot expect strong agreement between ITC experiment and MD simulation, particularly, in the case of weak binding. However, the simulations provide several indications in modelled energetics that are in agreement with the results of ITC. Accordingly, here, we suggest that the LDHSLHS polypeptide represents a case of a non-systematic silica surface grazer, where binding propensity is dynamically modulated by the local ionic environment.

The outcome does not correspond well with the initial analysis, which suggested that the polypeptide is an effective silica binder.[24] This indicates strongly that when the peptide is present as a part of the phage capsid decoration, the secondary structural fold induced by the complete viral protein may alter the structure and affect thermodynamics of binding comparing to the case we explore in this study where the sequence as a short polypeptide has both C and N terminals. It is likely that when part of phage protein, the hepta-peptide segment may occupy a relatively narrow and more constrained region(s) in the configurational space, than is the case for the corresponding hepta-peptide in an aqueous environment. In this respect, our research suggests that a systematic review of the outcomes of phage display assays is necessary if bioinorganic engineering with polypeptides detected in this matter is to be pursued further. Further, given that the main pragmatic motivation of such studies is to search for novel effective bio-composites, should it be possible to suggest an edited primary sequence to provide a more stable association with silica? This may be possible upon replacement of two hydrophobic leucines with amino acids of polar nature. This, however, is beyond the scope of the current study.

Experimental and methods

Silica nanoparticles were synthesized by a modified Stöber method.[48,49] In our studies we use 200 ± 12 nm diameter particles, the sizes of which we determined employing dynamic light scattering Zetasizer Nano-S, Malvern Instruments, Malvern, UK. For infrared (IR) optical absorption studies in mid-IR spectral range we prepared samples where water, as solvent, was exchanged with deuterium oxide. To prepare such samples we used a rotary evaporator IKA RV10/HB10 basic, Sigma-Aldrich Co., St. Louis, MO, USA, operated at 60-80 °C under reduced pressure to remove the solvent after each hydration step.

The polypeptide was prepared using a Liberty Blue automated microwave peptide synthesizer, CEM Corporation, Buckingham, UK. The general procedure was as described previously.[27] The concentrations of the polypeptide for FTIR and ATR-FTIR measurements were about 20 mM and 0.5 mM, respectively. Prior to optical studies we added NaCl to the aqueous solutions to give a sodium ion concentration of 50-100 mM. Using standard pH litmus paper, Fisherbrand, we identify Pd (since in D_2O) of the samples (typically 2-3 microliters of sample). In particular, we observe Pd = 6 and Pd = 3-4 for the samples with SiO_2 particles present and for the aqueous samples with polypeptide alone, respectively.

Fourier transform infrared (FTIR) and polarization sensitive attenuated total reflection Fourier transform infrared (ATR-FTIR) measurements were conducted with a Nicolet 6700 FTIR spectrometer and an Attenuated Total Reflection (ATR) unit Model: 0055290, from Thermo Electron (now, Thermo Fisher Scientific) equipped with BaF_2 holographic wire grid polarizers, WP25H-B, from Thorlabs Ltd., Ely UK. The spectral resolution used was 1 cm^{-1}. The polarization sensitive ATR-FTIR spectra confirmed the isotropic character of the prepared samples. We measured Raman spectra with a DXR microscope station from Thermo Fisher Scientific, Madison, WI 53711, using 530 nm excitation radiation at 10 mW. The spectral resolution in Raman

experiments was 2 cm^{-1} according to the instrumental limit by choice of a 25 micron confocal slit. We deposited solutions on an unprotected gold mirror, PFSQ05-03-M03 (Thorlabs), and measured from the liquid droplets without any cover. Spectral properties were recorded using 3 µl solution of 50 mM for polypeptides and of 1 mM for nanoparticles (about $2.5 \cdot 10^{15}$ particles in 3 µl). As a single polypeptide may occupy a surface area of ca. $2.9 \cdot 10^{-18}$ m^2, we anticipate about 36-40 polypeptides per particle. Considering the surface area of a silica particle, according to the provided average diameter, this implies that less than 0.1% of the available silica surface was involved in binding in samples used for optical studies.

Isothermal titration calorimetry (ITC) studies were conducted using a Malvern MicroCal VP instrument, Malvern, UK. In particular, peptide and silica nanoparticles were dispersed in either type 1 water (H_2O) or in 200 mM NaCl saline solution at a pH of 7.4 to prepare 1 mM and 0.1 mM stock solutions, respectively. The cell temperature and the reference power were 25 ˚C and 15 µcal·s^{-1}, respectively. A stirring speed of 307 rpm was used to maintain nanoparticles in suspension. The time delay between injections was half an hour to allow equilibrium to be reached after each injection. Each experiment had a total duration of around 8.5 hours and was repeated at least in duplicate to ensure reproducibility of the results. To obtain the values for the association constant a linear fitting was used, while enthalpy was taken from the amplitudes of the data.

In our analysis we rely on the results of MD studies reported earlier.[26,37] Briefly, the simulation box of 7094 atoms included the polypeptide of 109 atoms, 1600 water molecules, and a six layered silica slab, which contained 672 silicon atoms. Models of unit cells of α-cristobalite obtained from published X-ray crystal structures.[50,51] Models of deprotonated surface areas are obtained by deletion of hydrogen atoms of randomly selected ≡Si–OH groups and addition of sodium counter ions to create sodium siloxide groups (≡SiO^- ⋯ Na^+) to represent the influence of pH. We use two types of simulations, where 9 % and 18 % of ≡SiOH groups were ionized to ≡SiO^-. These imply that for every 56 silicon and oxygen atoms at each interface, there is a lack of 5 or 10 hydrogen atoms to determine the corresponding number of ≡SiO^- defects, respectively. Accordingly, in this article we use the notations LD91 and LD82 for the structural extracts from these systems with surfaces of different ionizations, respectively. Accordingly, in application to the extracted structure of the polypeptide next to the two types of the surface, in this article, we employ names LD91 and LD82, respectively.

Introduction of the polypeptide into simulation box is achieved using Hyperchem, Materials Studio, and the program psfgen.[52,53] The peptides were ionized at N termini, C termini, and ionic groups along with necessary chloride or sodium ions to balance net positive or negative charges depending on pH value. For each of the two cases, where surfaces of silica were under different ionization conditions, we prepared and ran five independent simulations of the polypeptide approaching silica surfaces. For each of the five simulations, the initial secondary structure for the polypeptide was determined independently as a lowest energy structural state gained in result of thermalisation and consequent MD simulation in water.

For the MD simulations, the CHARMM-SILICA force field (with 12-6 Lennard-Jones potential) is employed along with the program NAMD.[54,55] All atoms were flexible, the time step was 1 fs, the cut-off for van-der-Waals interactions 12 Å, and the accuracy of the summation of long-range Coulomb interactions high at 10^{-6} using Particle Mesh Ewald. The simulation production runs were performed as NVT ensembles.

Extraction of representative structural cases was conducted reviewing 4000 selected molecular arrangements, for details see Supporting Information file. Extractions of representative cases were to fulfil several criteria: that at least one of the atoms of the polypeptide would be no further than 3 angstroms from the average height of the H atoms of the Si-OH groups at the silica surface, and that the (Φ;Ψ) angular pairs for the five residues of the polypeptide would be unique. The structural extractions are conducted to sample from MD i) the coordinates of the atoms of the polypeptide, ii) the location of water molecules and Na^+ ions within 3 angstroms from any atom of the polypeptide, and iii) the segment of the first layer of the silicon oxide under a perimeter determined by the outer area of the polypeptide at the considered contact. In the case where several contacts are considered, the extracted silicon oxide segment would include all the perimeters at the considered contacts. After extraction of the representative structures from the MD simulation boxes, we edit the atomic composition. In particular, we remove Si atoms which are not involved in rings, and we remove water molecules that are not involved in hydrogen bonding. The structural extractions are done preserving first neighbour water molecules. Inclusion of the first neighbour water shell allows density functional theory (DFT) optimizations without any structural constraints. This secures frequencies to be all positive upon DFT based normal modes analysis, and enables the role of first neighbour water (as predicted by classic simulations) in the DFT based definition of the normal modes to be accounted for.

We conduct structural optimizations with DFT implemented in the Gaussian 09 package.[56] When compared to local spin density approximation functionals, Hartree-Fock, and Møller-Plesset, the generalized gradient approximation approach predicts structural properties, hydrogen bond interaction energies, ionization potentials and heats of formation better.[57] Among hybrid generalized gradient approximation functionals, B3LYP has been shown to be fast and most accurate in normal modes analysis.[57,58] Accordingly, we adopted the restricted hybrid B3LYP[59,60] functional with 6-31g* basis set. This basis set was chosen to optimize the systems with very large number of atoms. At the same time, this type of basis set is still used in current state of the art studies.[61] For each DFT computation, we used 16 cores, set to 128GB per processor.

For structural optimizations, we employed the default settings of Gaussian 09 for the convergence criteria. Specifically, the convergence threshold for self-consistent field integral accuracy was reached at 4.0×10-9 Hartree, the threshold for maximum force was 4.5×10-4 Hartree $Bohr^{-1}$, for force root mean square was 3×10-4 Hartree $Bohr^{-1}$, for maximum

displacement was 2×10-3 Bohr, and for displacement root mean square was 0.1×10-3 Bohr. Here, we specifically confirm that upon DFT optimizations of the extracted silica segments, structural properties, such as Si-O bonds lengths, Si-O-Si angles, O-Si-O angles and tetrahedral angles converge with the reported values and variances, specific to amorphous silica.[62-64] Since contacts of the polypeptide with the surface are local, we consider DFT anticipations of electronic and structural properties of moieties involved in local coordination to be adequate in representing the physics of coordination at the silica surface, including an amorphous one. This is sufficient, since the surface of the prepared particles is not porous to support partitioning and diffusion through silica.

Extraction of normal modes, transition dipole moments and Raman tensors were conducted on optimized structures. After we conduct DFT based optimization and normal mode analysis for each structure next to silica, we remove atoms specific to silica and fill the voids with water molecules to conduct DFT studies on the same structure but in an aqueous environment. Of course, this does not help to explore the role of the full possible configurational space when in water as detected by optical responses, but does help to discuss properties of the polypeptide when next to silica versus when it is fully hydrated in an aqueous environment before or after binding to the surface.

We develop structural analysis comparing spectral dispersions of predicted Infrared and Raman activities with those detected in experiments in the frequency range from 1350 to 1750 cm^{-1}. This spectral range is rather wide to use one scaling factor for the calculated normal modes. To address this, we explored and developed a linear scaling function 1.091 ω_{nm} - 230 to map the frequencies of calculated normal modes, ω_{nm}, in the chosen spectral range. Consistently, in this article, when we provide a calculated frequency, we present its value as calculated and scaled (in brackets). It is important to notice that according to the results of our recent study,[65] we can rule out systematic biasing effects of sharp edges of truncated silicon oxide on the frequencies of normal modes of polypeptides.

Our classical mechanics studies[26,37] did not account for proton transfer processes. In the reported MD trajectories, the polypeptide is charged: with positively charged N-terminal, and negatively charged deprotonated C-terminal and the functional group of aspartic acid. The polypeptide in classic simulation contains two single protonated (therefore neutral) histidines, where the nitrogen atoms next to the aliphatic linkers are deprotonated. However, the measured pD in sample micro-droplets and the experimentally measured spectra suggest that when next to silica and when in an aqueous environment the functional group of aspartic acid should be deprotonated (as modelled in MD) and the histidines protonated (not modelled in MD), respectively. Also, in all our samples we should expect the presence of double protonated ionic rings of histidine. To identify the correct nature of the charge state of aspartic acid when in an aqueous environment, we protonated its functional group when conducting DFT calculations for selected structures in the aqueous environment only. To address the possible contribution of histidine rings when single and double protonated to the optical response, we perform a series of DFT studies on the normal modes of the imidazole ring of methyl terminated histidine under different protonation conditions (see Supporting Information file). As a result, we determine that addition of a second protonation proximal to the linker nitrogen atom leads to about a 38 cm^{-1} shift to higher frequency for the imidazole mode specific to the ring C=C stretching. Considering that half of the histidines in our cases may be completely protonated, we calculate spectra for the polypeptide (either next to silica or alone in aqueous environment) by averaging spectral responses of polypeptide where histidine side groups are single and double protonated. The spectral response for the latter case is anticipated by giving a 38 cm^{-1} shift to higher frequency for the imidazole mode specific to the ring C=C stretching.

Conclusions

Combining Infrared and Raman spectroscopies with methods of classical and quantum computations, we present a structural analysis of the LDHSLHS polypeptide when at the silica interface and when in an aqueous environment. Molecular dynamics trajectories, development of which we reported previously,[26,37] partially confirm the results of bio-panning studies, that predict that the polypeptide should have the capacity to associate with silica. Using classical simulations, we extracted six representative structures: one case has extended and nearly random conformation; three structures demonstrate helical propensity at the N-terminal side but become more random towards the C-terminal; one shows a hairpin structure with β-like N-terminal but mainly helical for the rest of the residues; and one more structure where the central three residues are β-like, while terminals “twist out” in a helical fashion. Supported by quantum studies we compare calculated Infrared and Raman responses for the extracted structures with experimental measurements. Simultaneous fitting of the two optical responses in the Amide spectral region, we propose that either when next to silica or away the structure with β-like central region and helix-twisted terminals is dominant among other considered cases. Furthermore, we also predict that the three cases, where helix dominates at the side of N-terminal and the random structure should also be present when the polypeptide is next to the inorganic interface. For the peptide when in an aqueous environment, the simultaneous spectral fitting suggests that a hairpin structure and a structure including a helical component should be also present to match the experimental results. Using structural extracts and assisted with isothermal titration calorimetric data, we explore the thermo-dynamics of the binding process to confirm the predictions of both optical experiment and molecular dynamics simulations that under lower salinity LDHSLHS polypeptide represents a case of a non-systematic silica surface grazer, where binding propensity is dynamically modulated by local ionic environment. Accordingly, we discuss the implications of our study for phage display analysis. The results should help engineering of polypeptide-silica nanocomposites as next

generation multi-functional platforms for diagnostics and bio-catalytic applications.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Funding from AFOSR FA9550-1-16-2013 is gratefully acknowledged. The authors thank Dr. Joanna Aizenberg, Harvard University, for continued access to the Odyssey cluster at Harvard University. The computations in this paper were run on the Odyssey cluster sup-ported by the Faculty of Arts and Sciences division, Research Computing Group at Harvard University.

Notes and references

1 H. A. Lowenstam, *Science*, 1981, **211,** 1126-1131.
2 H. C. Schroder, X. H. Wang, W. Tremel, H. Ushijima and W. E. G. Muller, *Nat. Prod. Rep.*, 2008, **25**, 455-474.
3 C. C. Perry, *Biomineralization: chemical and biochemical perspectives*, VCH: Cambridge, New York, 1989, 223-256.
4 R. H. Newman and A. L. Mackay, *Ann. Bot.*, 1983, **52**, 927-929.
5 J. Arivalagan, T. Yarra, B. Marie, V. A. Sleight, E. Duvernois-Berthet, M. S. Clark, A. Marie and S. Berland, *Mol. Biol. Evol.*, 2017, **34**, 66–77.
6 P. Barden, H. W. Herhold and D. A. Grimaldi, *Systematic Entom.*, 2017, **42**, 837–846.
7 D. O. Wagner and P. Aspenberg, *Acta Orthop*., 2011, **82**, 393–398.
8 C. C. Perry, *Morphogenesis, and Nanobiotechnology, Progress in Molecular and Subcellular Biology, Marine Molecular Biotechnology,* Springer-Verlag: Berlin, Heidelberg, 2009, **47**, 295-313.
9 C. C. Leung, T. S. Yu and W. Chen, *The Lancet*, 2012, **379**, 2008-2018.
10 R. P. Diez and J. I. Amalvy, *J. Mol. Struct. (Theochem.)*, 2003, **634**, 187-193.
11 A. Abbasi, E. Nadimi, P. Plänitz and C. Radehaus, *Surf. Sci.,* 2009, **603**, 2502-2506.
12 D. Costa, A. Tougerti, F. Tielens, C. Gervais, L. Stievano and J. F. Lambert, *Phys. Chem. Chem. Phys.* 2008, **10**, 6360-6368.
13 M. C. McAlpine, H. D. Agnew, R. D. Rohde, M. Blanco, H. Ahmad, A. D. Stuparu, W. A. Goddard and J. R. Heath, *J. Am. Chem. Soc.*, 2008, **130**, 9583–9589.
14 M. Saadaoui, I. Fernandez, A. Sanchez, P. Diez, S. Campuzano, N. Raouafi, J. M. Pingarron and R. Villalonga, *Electrochem. Comm*. 2015, **58**, 57-61.
15 M. Pagliaro, *Silica-Based Materials for Advanced Chemical Applications*, RSC Publishing: Cambridge, U.K., 2009.
16 K. K. Qian and R. H. Bogner, *J. Pharm. Sci.*, 2012, **101**, 444–463.
17 A. Pedone, G. Prampolini, S. Monti and V. Barone, *Phys. Chem. Chem. Phys.* 2001, **13**, 16689-16697.
18 J. N. Cha, G. D. Stucky, D. E. Morse and T. J. Deming, *Nature,* 2000, **403**, 289-292.
19 M. Sumper and N. Kröger, *J. Mater. Chem.,* 2004, **14**, 2059-2065.
20 A. Rai and C. C. Perry, *Langmuir*, 2010, **26**, 4152-4159.
21 E. E. Oren, R. Notman, I. W. Kim, J. S. Evans, T. R. Walsh, R. Samudrala, C. Tamerler and M. Sarikaya, *Langmuir*, 2010, **26**, 11003-11009.
22 G. P. Smith, *Science*, 1985, **228**, 1315–1317.
23 J. McCafferty, A. Griffiths, G. Winter and D. Chiswell, *Nature*, 1990, **348**, 552–554.
24 S. V. Patwardhan, F. S. Emami, R. J. Berry, S. E. Jones, R. R. Naik, O. Deschaume, H. Heinz and C. C. Perry, *J. Am. Chem. Soc.,* 2012, **134**, 6244−6256.
25 R. R. Naik, L. L. Brott, S. J. Clarson and M. O. Stone, *J Nanosci. Nanotechnol.,* 2002, **2**, 95-100.
26 F. S. Emami, V. Puddu, R. J. Berry, V. Varshney, S. V. Patwardhan, C. C. Perry and H. Heinz, *Chem. Mater.*, 2014, **26**, 5725−5734.
27 V. Puddu and C. C. Perry, *ACS Nano*, 2012, **6**, 6356–6363.
28 M. Karplus, *J. Chem. Phys.*, 1959, **30**, 11-15.
29 P. Hamm, M. Lim, W. F. DeGrado and R. M. Hochstrasser, *Proc. Natl. Acad. Sci.,* 1999, **96**, 2036-2041.
30 O. Golonzka, M. Khalil, N. Demirdoven and A. Tokmakoff, *Phys. Rev. Lett.*, 2001, **86**, 2154-2157.
31 F. Ding and M. T. Zanni, *Chem. Phys.*, 2007, **34**1, 95-105.
32 J. A. Borek, F. Perakis and P. Hamm, *Proc. Natl. Acad. Sci.*, 2014, **111**, 10462-10467.
33 V. Volkov, D. J. Palmer and R. Righini, *Phys. Rev. Lett.*, 2007, **99**, 078302.
34 V. Volkov, R. Chelli, W. Zhuang, F. Nuti, Y. Takaoka, A. M. Papini, S. Mukamel and R. Righini, *Proc. Natl. Acad. Sci.*, 2007, **104**, 15323-15327.
35 V. Volkov, R. Chelli, F. Muniz-Miranda and R. Righini, *J. Phys. Chem. B*, 2011, **115**, 5294-1303.
36 H. Torii, *J. Phys. Chem. A*, 2006, **110**, 4822-4832.
37 F. S. Emami, V. Puddu, R. J. Berry, V. Varshney, S. V. Patwardhan, C. C. Perry and H. Heinz, *Chem. Mater.*, 2014, **26**, 2647-2658.
38 S. Krimm and Y. Abe, *Proc. Natl. Acad. Sci.*, 1972, **69**, 2788-2792
39 Y. N. Chirgadze, B. V. Shestopalov and S. Y. Venyaminov, *Biopolymers*, 1973, **12**, 1337-1351
40 G. Martinez and G. Millhauser, *J. Struct. Biol.*, 1995, **114**, 23-27
41 F. Jiang, C.-Y. Zhou and Y.-D. Wu, *J. Phys. Chem. B*, 2014, **118**, 6983-6998.
42 K.-C. Chou, *Anal. Bioch.*, 2000, **286**, 1-16.
43 S. J. Lahr, D. E. Engel, S. E. Stayrook, O. Maglio, B. North, S. Geremia, A. Lombardi and W. F. DeGrado, *J. Mol Biol.*, 2005, **346**, 1441-1454.
44 X. Shi, X. Hu, S. Li and X. Liu, *J. Theor. Biol.*, 2011, **286**, 24-30.
45 S. Aravinda, U. S. Raghavender, R. Rai, V. V. Harini, N. Shamala and P. Balaram, *Org. Biomol. Chem.*, 2013, **11**, 4220-4231.
46 A. G. de Brevern, *Sci. Rep.*, 2016, **6**, 33191:1-15.
47 J. Schlitter, *Chem. Phys. Lett.*, 1993, **215**, 617-621.
48 P. Roach, D. Ferrar and C. C. Perry, *J. Am. Chem. Soc.*, 2006, **128**, 3939-3945.
49 W. Stober, A. Fink and E. Bohn, *J. Colloid Interface Sci.*, 1968, **26**, 62-69.
50 P. Dera, J. D. Lazarz, V. B. Prakapenka, M. Barkley and R. T. Downs, *Phys. Chem. Miner*. 2011, **38**, 517-529.
51 K. Kihara, *Eur. J. Mineral*. 1990, **2**, 63-77.
52 HyperChem(TM) Professional 7.51, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.
53 Material Studio 4.1 and Discover Program, Accelrys Inc: San Diego, CA, 2006.
54 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781-1802.
55 H. Heinz, T.-J. Lin, R. K. Mishra and F. S. Emami, *Langmuir*, 2013, **29**, 1754-1765.
56 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, et al., *Gaussian 09*, revision B.01, 2010.
57 K. E. Riley, B. T. Op't Holt and K. M. Merz Jr., *J. Chem. Theory Comp.* 2007, **3**, 407-433.

58 A. A. El-Azhary and H. U. Suter, *J. Phys. Chem.* 1996, **100**, 15056-15063.
59 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098.
60 C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.
61 A. Muzdalo, P. Saalfrank, J. Vreede and M. Santer, *J. Chem. Theory Comp.* 2018, **14**, 2042−2051.
62 Y. P. Li and W. Y. Ching, *Phys. Rev. B*, 1985, **31**, 2172-2179.
63 V. A. Gritsenko, J. B. Xu, I. H. Wilson and R. M. Kwok, *Phys. Rev. Lett.*, 1998, **81**, 1054-1057.
64 S. S. Nekrashevich and V. A. Gritsenko, *Phys. Sol. St.*, 2014, **56**, 207-222.
65 V. V. Volkov, D. Belton and C. C. Perry, *J. Phys. Chem. A*, 2018, **122**, 4997-5003.

