



Cite this: RSC Adv., 2019, 9, 1062

## Interaction landscape of a 'C<sup>α</sup>NN' motif with arsenate and arsenite: a potential peptide-based scavenger of arsenic<sup>†</sup>

Subhankar Sahu,<sup>‡</sup> Tridip Sheet<sup>‡</sup> and Raja Banerjee  \*

Arsenic (As) is a toxic metalloid that has drawn immense attention from the scientific community recently due to its fatal effects through its unwanted occurrence in ground water around the globe. The presence of an excess amount of water soluble arsenate and/or arsenite salt (permissible limit 10  $\mu\text{g L}^{-1}$  as recommended by the WHO) in water has been correlated with several human diseases. Although arsenate ( $\text{HAsO}_4^{2-}$ ) is a molecular analogue of phosphate ( $\text{HPO}_4^{2-}$ ), phosphate is indispensable for life, while arsenic and its salts are toxic. Therefore, it is worthwhile to focus on the removal of arsenic from water. Towards this end, the design of peptide-based scaffolds for the recognition of arsenate and arsenite would add a new dimension. Utilizing the stereochemical similarity between arsenate ( $\text{HAsO}_4^{2-}$ ) and phosphate ( $\text{HPO}_4^{2-}$ ), we successfully investigated the recognition of arsenate and arsenite with a naturally occurring novel phosphate binding 'C<sup>α</sup>NN' motif and its related designed analogues. Using computational as well as biophysical approaches, for the first time, we report here that a designed peptide-based scaffold based on the 'C<sup>α</sup>NN' motif can recognize anions of arsenic in a thermodynamically favorable manner in a context-free system. This peptide-based arsenic binding agent has the potential for future development as a scavenger of arsenic anions to obtain arsenic free water.

Received 4th October 2018  
 Accepted 8th December 2018

DOI: 10.1039/c8ra08225a  
[rsc.li/rsc-advances](http://rsc.li/rsc-advances)

## 1 Introduction

Arsenic (As) contamination in ground water is a global health issue.<sup>1–5</sup> It is one of the largest worldwide health disasters affecting around 150 million people who are at high risk.<sup>6</sup> As per reported investigations, India and Bangladesh are the most victimized countries.<sup>7–11</sup> Arsenic having similar chemical properties with phosphorus, is a pnictogen element and a member of group 15 of the modern periodic table.<sup>12</sup> The most favorable oxidation states of As are +5 and +3.  $\text{AsO}_4^{3-}$ ,  $\text{HAsO}_4^{2-}$  and  $\text{H}_2\text{AsO}_4^{1-}$  are formed from arsenic acid ( $\text{H}_3\text{AsO}_4$ ; oxidation state of As +5), while  $\text{AsO}_3^{3-}$ ,  $\text{HAsO}_3^{2-}$  and  $\text{H}_2\text{AsO}_3^{1-}$  are formed from arsenious acid ( $\text{H}_3\text{AsO}_3$ ; oxidation state of As +3). From a chemistry point of view, it can be inferred that arsenate ( $\text{HAsO}_4^{2-}$ ) should behave as a molecular analogue of phosphate ( $\text{HPO}_4^{2-}$ ) and indeed, it is found that both compete for different biological processes.<sup>13</sup> Although both phosphorus and arsenic carry almost the same electronegativity, the atomic radius of arsenic is larger.<sup>12</sup> Unlike phosphorous, compounds of arsenic

are highly toxic. However, the mechanisms of arsenic toxicity in living organisms have not been extensively deciphered. Phosphate ( $\text{HPO}_4^{2-}$ ) is indispensable for life, as it regulates several biological functions<sup>14,15</sup> and is present as an important constituent of essential biological macromolecules (e.g., DNA, lipids, etc.). On the contrary, arsenic is toxic and causes hindrance in operations of processes like inhibition of ATP synthesis during oxidative phosphorylation and many others.<sup>13,16–19</sup> Most arsenic pollution goes right through ground water as excess water soluble arsenate and/or arsenite salts (permissible limit 10  $\mu\text{g L}^{-1}$  as recommended by WHO). Pentavalent and trivalent states of arsenic, along with their methylated derivatives, interfere with the metabolic pathway and normal cellular activity, but the exact mechanism is still obscure. Recently, arsenic contamination in water has been found to be linked with numerous human diseases<sup>20,21</sup> which include cardio vascular abnormalities,<sup>22–25</sup> risk of malignancy,<sup>3,26–28</sup> diabetes,<sup>29–31</sup> gastrointestinal anomalies,<sup>4,5</sup> nervous system dysfunction,<sup>32–34</sup> respiratory distress and obstructive lung diseases,<sup>35,36</sup> dermatological lesions,<sup>37,38</sup> etc. At this point, it is worthy to focus on aspects of arsenic removal from water. However, it should be noted that removal of arsenite ( $\text{As}^{+3}$ ) from water is more difficult than arsenate ( $\text{As}^{+5}$ ).<sup>39</sup> Available purification techniques range from co-precipitation with naturally occurring iron<sup>40,41</sup> to a three pitcher (3-kolshi) filtration unit.<sup>42</sup> Recombinant DNA technologies also have been employed to develop an arsenic-binding DNA aptamer for arsenic removal.<sup>43</sup>

Department of Biotechnology and Head Department of Bioinformatics, Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly Known as West Bengal University of Technology), BF-142, Salt Lake, Kolkata 700064, West Bengal, India. E-mail: [ban\\_raja@yahoo.com](mailto:ban_raja@yahoo.com); [banraja10@gmail.com](mailto:banraja10@gmail.com)

† Electronic supplementary information (ESI) available. See DOI: [10.1039/c8ra08225a](https://doi.org/10.1039/c8ra08225a)

‡ SS and TS contribute equally.



Although various approaches have been adopted globally for removing As derivatives from water, molecular recognition of As anions by proteins and peptides at the atomic scale is a new approach worth exploring.

Attempts to remove arsenic-based anions utilizing their recognition by peptide-based scaffolds would add a new dimension. However, identification of a proper peptide motif which can recognize anions of As in a context-free situation is the real challenge. Among several anion binding peptide motifs, the 'C<sup>z</sup>NN' motif,<sup>44</sup> responsible for binding phosphate in protein, would be a logical choice for recognizing arsenic salts (arsenate/arsenite), because the arsenate ion is stereochemically similar to phosphate and sulfate. The 'C<sup>z</sup>NN' motif, found mostly in the N-terminal of helices,<sup>44</sup> is a three residue motif capable of recognizing an anion (mainly phosphate and sulfate) even in a context-free non-proteinaceous environment.<sup>45-47</sup> C<sub>i-1</sub>-H, N<sub>i</sub>-H and N<sub>i+1</sub>-H of the backbone atoms in the three consecutive residues of the motif sequence participate in anion binding as mediated through hydrogen bonding (H-bond). It is established that C<sub>i-1</sub> participates in a relatively weak hydrogen bond, while N<sub>i</sub> and N<sub>i+1</sub> form rather a strong one. Further, the 'C<sup>z</sup>NN' motif bears a 'GXX' sequence pattern with a conserved geometry of  $\beta\alpha\alpha$  or  $\beta\alpha\beta$  and shows higher affinity towards sulfate (SO<sub>4</sub><sup>2-</sup>) ions than phosphate (HPO<sub>4</sub><sup>2-</sup>) ions in a context-free system.<sup>46,47</sup> Their ability to recognize anions has been characterized extensively through *in silico* approaches like molecular docking<sup>46</sup> and molecular dynamics simulation<sup>46,48</sup> along with quantum mechanical calculations<sup>48</sup> and has been further validated *via* complementary biophysical techniques like NMR and Circular Dichroism Spectroscopy.<sup>45,47</sup> All these experiments establish that efficiency of anion binding largely depends on the 'C<sup>z</sup>NN' involving residues and nature of the anion.

In this study, we explored potential peptide-based scaffolds that would recognize arsenate and arsenite compounds. Intrinsic affinity of the 'C<sup>z</sup>NN' motif for HPO<sub>4</sub><sup>2-</sup> ion and the stereochemical similarity between HPO<sub>4</sub><sup>2-</sup> and HAsO<sub>4</sub><sup>2-</sup> are the principal guiding factors of our present investigation. We used a group of arsenic compounds as anionic ligands to scrutinize the potentiality of anion recognition ability with pre-designed model scaffolds based on the 'C<sup>z</sup>NN' anion binding motif by analyzing the interaction landscape. Our results clearly indicate that the designed peptide-based scaffolds can recognize the anions of As in a thermodynamically favorable manner in a context-free system and thus would be a potential candidate to be used as a scavenger of arsenate/arsenite in a suitable form if attached to a water filtration unit for As removal.

## 2 Methods

### 2.1 Library of ligands used

In aqueous solution arsenic forms oxo-anions like arsenite (salts of H<sub>3</sub>AsO<sub>3</sub> acid) and arsenate (salts of H<sub>3</sub>AsO<sub>4</sub> acid) with oxidation states +3 and +5, respectively.<sup>12</sup> Respective pK<sub>a</sub> values of arsenous acid (H<sub>3</sub>AsO<sub>3</sub>) are 9.23 (pK<sub>a1</sub>), 12.13 (pK<sub>a2</sub>), and 13.40 (pK<sub>a3</sub>), while for arsenic acid (H<sub>3</sub>AsO<sub>4</sub>) they are 2.20 (pK<sub>a1</sub>), 6.97 (pK<sub>a2</sub>), and 11.53 (pK<sub>a3</sub>)<sup>12</sup> which are comparable to the pK<sub>a</sub> values of phosphoric acid (H<sub>3</sub>PO<sub>4</sub>). A set of arsenic compounds

were used for the molecular docking study as ligands (anions) which included AsO<sub>4</sub><sup>3-</sup>, HAsO<sub>4</sub><sup>2-</sup> and H<sub>2</sub>AsO<sub>4</sub><sup>1-</sup> from the arsenic acid group and H<sub>2</sub>AsO<sub>3</sub><sup>1-</sup>, HAsO<sub>3</sub><sup>2-</sup> and AsO<sub>3</sub><sup>3-</sup> from the arsenous acid group. Three dimensional coordinates of arsenate (HAsO<sub>4</sub><sup>2-</sup>) and arsenite (AsO<sub>3</sub><sup>3-</sup>) were retrieved from RCSB PDB<sup>49</sup> with their three letter codes, "ART" (<https://www3.rcsb.org/ligand/ART>)<sup>50</sup> and "AST" (<https://www3.rcsb.org/ligand/AST>),<sup>51</sup> respectively. Other related derivatives of As compounds were generated with the help of GaussView 5 molecular builder and visualizer<sup>52</sup> by adding or removing required hydrogen atom(s), providing suitable net charge, and finally optimizing at the semi-empirical PM6 level.<sup>53</sup> Sulfate and phosphate ions additionally were used to compare the effectiveness of binding arsenic derivatives. Because arsenic anions interfere with the metabolic pathway and normal cellular activity, priority was given to those conjugate bases of arsenic acids which exist at the physiological pH range (pH 6-8) *i.e.*, arsenate (as HAsO<sub>4</sub><sup>2-</sup>, pK<sub>a2</sub>: 6.97) and thus results related to arsenite (with pK<sub>a</sub>: 9.23) are beyond the scope of this discussion.

### 2.2 Sequences of model peptides

Conservation of sequence in the 'C<sup>z</sup>NN' motif<sup>44</sup> was observed from analysis of different fold representative structures of proteins. A series of chimeric peptides were constructed, where the naturally occurring 'C<sup>z</sup>NN'<sup>44</sup> or designed model scaffolds based on the 'C<sup>z</sup>NN' peptide sequences<sup>48</sup> were appended at the N-terminus of a model context-free helix (ABGY).<sup>54,55</sup> All the scaffolds used here have extensive investigation portfolios with both sulfate and phosphate ions as interacting partners.<sup>46-48</sup> In the present study, arsenic related ligands were used to extract information about anion recognition by these scaffolds to explore whether they would act as effective scavengers for anion(s) of arsenic (As). Details of the peptides used are listed here where the underline segment represents the anion binding sequences having either the naturally occurring 'C<sup>z</sup>NN'<sup>44</sup> or designed model scaffolds based on the 'C<sup>z</sup>NN' peptide sequences<sup>48</sup> (constituent amino acids are designated through three letter symbols).

DS1: Ac-Ala-Gly-Lys-Ser-Ala-Aib-Ala-Lys-Ala-Aib-Lys-Ala-Lys-Ala-Aib-Gly-Gly-Tyr-NH2.

DS3: Ac-Ala-Ser-Lys-Ser-Ala-Aib-Ala-Lys-Ala-Aib-Lys-Ala-Lys-Ala-Aib-Gly-Gly-Tyr-NH2.

CPS224Ac: Ac-Leu-Gly-Lys-Gln-Ala-Aib-Ala-Lys-Ala-Aib-Lys-Ala-Lys-Ala-Aib-Gly-Gly-Tyr-NH2.

CPS226: Ac-Gly-Ser-Ala-Lys-Ala-Aib-Ala-Lys-Ala-Aib-Lys-Ala-Lys-Ala-Aib-Gly-Gly-Tyr-NH2.

SCPS224Ac: Ac-Leu-Gly-Lys-Gln-Ala-NH2.

SCPS226: Ac-Gly-Ser-Ala-Lys-Ala-NH2.

The N-terminus of all the peptides are capped with an acetyl (Ac) group while the amide (-CONH) group is used to protect the respective C-terminus. For all the following larger peptides sequences, the motif (after 4<sup>th</sup> residue) is identical and essentially fixed to a right handed helical peptide (ABGY).<sup>54,55</sup>  $\alpha$ -Aminoisobutyric acid (Aib) was included to increase helix stability because of its high helix inducing propensity. The 'C<sup>z</sup>NN' motif sequences of CPS224Ac, SCPS224Ac, and CPS226,



SCPS226 were obtained from the crystal structures 1MUG<sup>14</sup> and 1YCC,<sup>56</sup> respectively from PDB. Peptide capping was done *via* Discovery Studio Visualizer 4.1.<sup>57</sup> CPS224Ac and CPS226 were demonstrated as natural 'C<sup>2</sup>NN' motif-containing peptides whereas DS1 and DS3 represent designed scaffolds.<sup>48</sup>

### 2.3 Molecular docking

Bio-molecular interactions among the peptides and ligands were primarily investigated and analyzed *via* the AutoDock 4.2 molecular docking package.<sup>58–61</sup> MGL Tools 1.5.6 (ref. 62) was installed in Windows 7 OS and used for preparing different parameter files for docking purposes. Arsenic as a molecule is non-recognizable to the AutoDock program, so, parameterization was carried out by the standard procedure mentioned in the AutoDock official webpage (<http://autodock.scripps.edu/faqs-help/how-to/adding-new-atom-parameters-to-autodock>). Further manual additions of the required parameters (*viz* van der Waals radius, atomic solvation volume, *etc.*) were accomplished through 'AD4\_parameters.dat' and 'AD4.1\_bound.dat' files while keeping those files in the working directory where we ran the docking program. Parameters used for the arsenic atom are Rii (sum of van der Waals radius of two like atoms, in Å): 4.30, epsii (vdw well depth in kcal mol<sup>-1</sup>): 0.200, vol (atomic solvation volume, Å<sup>3</sup>): 41.6087, and solpar (atomic solvation parameter): -0.00110. Addition of the lines 'parameter\_file AD4\_parameters.dat', and 'parameter\_file AD4.1\_bound.dat' without quotes (which instructs AutoDock to read the parameter files from a default location) was done at the beginning of the .gpf (grid parameter file) and .dpf (docking parameter file) before running *autogrid4* and *autodock4* programs, respectively. All the coordinates of ligands and peptide were loaded into AutoDock Tools 1.5.6 (available under MGL Tools) to add polar hydrogens. Partial charge was calculated and saved in '.pdbqt' format. For all the arsenic compounds, partial charge was calculated using the Gaussian 2009 software package<sup>63</sup> *via* calculating the ground state energy in semi-empirical mode (as AutoDock also uses a semi empirical free energy force field) with a PM6 Hamiltonian basis set<sup>64</sup> and saved in a '.pdbqt' format. Distribution of partial charge used for docking purposes with HAsO<sub>3</sub><sup>2-</sup> and HAsO<sub>4</sub><sup>2-</sup> is mentioned in ESI Fig. S1.† Size of the grid box was set [with grid points 44(x) 44(y) 44(z)] in such a way that it included up to the Lys8 residue from the N terminal Ac-group. After generating the map files for rigid docking from the grid parameter file, final docking was carried out by a genetic algorithm approach.<sup>59</sup> The maximum number of energy evaluations was set to  $2.5 \times 10^7$  with a rate of gene mutation 0.02 and rate of crossover 0.8 for the required search parameters of the genetic algorithm and 250 conformations (number of GA runs) were generated to find all the ligand binding states. Results were validated through AutoDock Tools and also by manually written scripts by extracting information from the docking log files. Education PyMOL (version 1.3)<sup>65</sup> was used to produce the images.

### 2.4 Mass spectroscopy

Molecular association of arsenate with peptide DS3 was experimentally determined by investigating the molecular mass (*m/z*)

of the free-peptide and arsenate-bound peptide separately. Using MALDI-MS (Autoflex III TOF/TOF 200), the molecular mass of peptide DS3 (MW of DS3 1704) was obtained in the positive mode in H<sub>2</sub>O/CH<sub>3</sub>CN (1 : 1) with 0.1% HCOOH. On the other hand, the molecular mass of the arsenate added (added as Na<sub>2</sub>HAsO<sub>4</sub>) species of DS3 was determined from the corresponding *m/z* value in the negative mode of ESI-FTMS (Apex Ultra 70, Bruker Daltonics direct infusion mode), in H<sub>2</sub>O/CH<sub>3</sub>CN (1 : 1) with 0.1% NH<sub>3</sub>.

### 2.5 Circular dichroism spectroscopy

The conformational landscape of peptide DS3 upon interaction with arsenate in the fully aqueous condition was investigated using far-UV CD spectra (range of wavelength 250–190 nm). At room temperature (25 °C) before and after addition of arsenate (added as Na<sub>2</sub>HAsO<sub>4</sub>), the spectra were recorded using a Jasco J-815 CD-spectropolarimeter. Concentration of the peptide was kept at  $\sim$ 75 μM (measured from Tyr absorbance at 275 nm,  $\epsilon = 1450 \text{ cm}^{-1} \text{ M}^{-1}$ ) and arsenate was added at a ratio of 1 : 3 (peptide : arsenate). The CD spectra were recorded in a 1 mm path length cuvette and reported as mean residue ellipticity (deg cm<sup>2</sup> dmol<sup>-1</sup>).

### 2.6 Isothermal titration calorimetric

In order to measure the binding affinity of arsenate for peptide DS3, along with the associated thermodynamic parameters of the interaction, isothermal titration calorimetric (ITC) data were recorded using a Microcal ITC200 (Malvern) at 25 °C. A 400 μL aqueous solution of peptide DS3 with a concentration of  $\sim$ 50 μM was kept in the reaction cell and titrated by adding 1.5 mM arsenate solution (1 : 30 ratio of peptide and arsenate). As a control, a similar titration was carried out only with HPLC water (without the peptide) and that data was subtracted from the data obtained using the peptide.

### 2.7 Molecular dynamics simulation

To pursue a molecular dynamics simulation, we selected the ligand bound conformation generated from the molecular docking process of DS3 and HAsO<sub>4</sub><sup>2-</sup> (for reasons described in Discussion section). Simulation was done with the help of the GROMACS 5.1.2 package<sup>66</sup> using Charmm27 force field<sup>67,68</sup> within a cubic box in an explicitly water solvated system. Arsenate (HAsO<sub>4</sub><sup>2-</sup>) was properly parameterized to keep its behavior consistent with the charm27 force field. Ligand "AST" (HAsO<sub>4</sub><sup>2-</sup>) was taken from PDB [<https://www3.rcsb.org/ligand/AST>] and, with the help of GaussView 5 (ref. 52) and Gaussian09 (ref. 63) programs, its geometry optimization along with partial charge distribution were calculated with the B3LYP level and the 6-31G++ basis set.<sup>69</sup> Bonded and non-bonded parameters, adapted from the article "Structural and Functional Consequences of Phosphate–Arsenate Substitutions in Selected Nucleotides: DNA, RNA, and ATP,"<sup>70</sup> was suitably added to the charm27 force-field files within the installed GROMACS distribution. The simulation box had to be large enough, for which distance of the periphery of the peptide and the side of the box was kept approximately 2 nm. Pre-



equilibration of the system was done with simple point charge (SPC) water model molecules.<sup>71</sup> The system was neutralized by adding the requisite numbers of sodium or chloride ions and then its energy was minimized using a steepest descent algorithm. NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature)<sup>72</sup> simulations were carried out (100 ps for both NVT and NPT) as part of the standard procedure for equilibration of the system. Finally, the equilibrated system was allowed to dynamically evolve up to 50 ns and the information from it was saved at every 10 ps interval to generate an ensemble of conformational states. Constraint was applied to all bonds with the LINCS algorithm<sup>73</sup> and the non-bonded interaction was taken care of by PME method.<sup>74</sup> Upon completion of the 50 ns simulation, time dependent behavior of ligand ( $\text{HAsO}_4^{2-}$ ) with the designed scaffold (DS3) was analyzed by the default GROMACS utilities and custom written Perl scripts as a function of time.

### 3 Results and discussion

#### 3.1 Molecular docking experiment

A molecular docking experiment enabled us to ensure thermodynamic feasibility of recognizing the anions comprising As (oxidation states +3 and +5) with the peptide scaffolds. The experimental output corroborated affinity as well as the potential of binding different ligands with the peptide-based scaffolds. Long 18 residue peptides (e.g., DS1, DS3, CPS224Ac, and CPS226) along with the short sequences (e.g., SCPS224Ac and SCPS226) containing only the 'C<sup>α</sup>NN' fragment were allowed separately to interact with each of the ligands from the library (as mentioned in Methods). Besides the binding free energy, a few other parameters (described in following sections) were also used to infer the ligand binding quality of the peptides.

#### 3.2 Binding free energy based analysis

The respective binding free energy values obtained from the molecular docking experiment clearly emphasized that all of the peptides used have significant affinity towards the ligands of arsenic, as is similar to sulfate and phosphate ions. Appearance of a 'single cluster' of the interaction energy (binding free energy obtained from the MGL Tools [version 1.5.6]) for all the individual peptides with their respective ligands of As suggested that a single site of the peptide is responsible for recognizing the As anions. In all cases, interactions were observed between the arsenic ligands and the 'C<sup>α</sup>NN' segment of the peptides, where the interactions are mediated through typical hydrogen bonds between the  $\text{C}\alpha_{i-1}-\text{H}$ ;  $\text{N}_i-\text{H}$ ;  $\text{N}_{i+1}-\text{H}$  and oxygen atoms of the respective As ligand similar to that of sulfate and the phosphate ions.<sup>46,48</sup> It is quite appealing to note that in spite of the presence of charged side-chains with lysine residues that in all cases only the 'C<sup>α</sup>NN' segment interacts with the As-ligands as is also found for the sulfate and phosphate ions.<sup>46,48</sup> However, potentiality of the interactions varies (Table 1) depending on the type of the ligand, its geometry, and charge as well as the sequences of the 'C<sup>α</sup>NN' segment of the peptides. The designed

scaffolds [DS1: **Gly2-Lys3-Ser4** and DS3: **Ser2-Lys3-Ser4**] show comparatively higher anion recognition ability (Table 1) in terms of binding free energy for the ligands of As derivatives over the sequences containing the natural 'C<sup>α</sup>NN' motif sequence [CPS224Ac: **Gly2-Lys3-Gln4** and CSP226: **Ser2-Ala3-Lys4**], as similar to interactions observed for the  $\text{SO}_4^{2-}$  and  $\text{HPO}_4^{2-}$  ions<sup>48</sup> (ESI Fig. S2†). Out of all the participating scaffolds, DS3 showed the most favourable interaction with all the anions, having a binding free energy of  $-3.08 \text{ kcal mol}^{-1}$ ,  $-3.8 \text{ kcal mol}^{-1}$ ,  $-3.13 \text{ kcal mol}^{-1}$ , and  $-2.92 \text{ kcal mol}^{-1}$  for  $\text{HAsO}_4^{2-}$ ,  $\text{HAsO}_3^{2-}$ ,  $\text{HPO}_4^{2-}$  and  $\text{H}_2\text{AsO}_3^-$  ions respectively (Table 1). This is at par with the observation obtained with the  $\text{SO}_4^{2-}$  and  $\text{HPO}_4^{2-}$  ions.<sup>48</sup> Although both of these ligands ( $\text{HAsO}_3^{2-}$  and  $\text{HAsO}_4^{2-}$ ) are potent poisons and silent killers through pollution in groundwater,  $\text{HAsO}_3^{2-}$  is often considered to be more lethal than arsenate ( $\text{HAsO}_4^{2-}$ ). It is quite appealing to note that in terms of binding free energy,  $\text{HAsO}_3^{2-}$  was better recognized by DS3 in comparison to  $\text{HAsO}_4^{2-}$  (Table 1). One reason for a higher free energy of binding for arsenite over arsenate (having same net charge of  $-2$ ) can be speculated as follows. In comparison to  $\text{HAsO}_4^{2-}$ , distributed partial charges with oxygen atoms are somewhat higher in  $\text{HAsO}_3^{2-}$  due to the lower oxidation state of As (+3) thus enhancing the affinity of interaction (charge distribution of an arsenate and arsenite ligand can be seen in ESI Fig. S1†). Both As and P have almost comparable electronegativity, while the quickly noticeable difference is the atomic radius (As > P) due to the presence of an inner line d orbital in As. The overall observation leads one to conclude that the designed peptide scaffolds (DS1 and DS3) can be considered as potent candidates to bind  $\text{HAsO}_3^{2-}$  and  $\text{HAsO}_4^{2-}$ , the interactions of which are comparable to that of  $\text{HPO}_4^{2-}$ .

#### 3.3 Quantitative analysis of peptide and ligand interaction

The model host helix ABGY provides a native environment to the 'C<sup>α</sup>NN' motif which is appended at the N-terminus of ABGY. Interactions of anions of As with the peptide-scaffolds mediated through a hydrogen bond (H-bond) were analyzed by statistical distributions of the (X)-H $\cdots$ O bond length and X-H $\cdots$ O bond angle. Hydrogen bond distance between ( $\text{C}\alpha$ )H $\cdots$ O, (N1)H $\cdots$ O, and (N2)H $\cdots$ O atoms along with the angle formed by  $\text{C}\alpha-\text{H}\cdots\text{O}$ , N1-H $\cdots$ O, and N2-H $\cdots$ O atoms were extensively analyzed for all the 250 poses generated for each individual interaction in order to portray the binding of peptide and ligand. Detailed interaction parameters of  $\text{HAsO}_4^{2-}$  and  $\text{HAsO}_3^{2-}$  with DS3 and DS1 are shown in Fig. 1. Distribution of distance and angles between atoms of the 'C<sup>α</sup>NN' segment with the individual ligands ( $\text{HAsO}_4^{2-}$  and  $\text{HAsO}_3^{2-}$ ) are shown via a histogram plot for which the average ( $\mu$ ) and standard deviation ( $s$ ) are noted (ESI Fig. S3 and S4†). Dispersion of the distribution also was evaluated through the  $s/\mu$  ratio (relative standard deviation) (ESI Fig. S3 and S4†). Low values of  $s/\mu$  (in the range of  $10^{-2}$  to  $10^{-1}$ ) indicate that the average value ( $\mu$ ) can be approximated as the actual value of the distribution. Distribution of the distance and angle plots show average distances of 2.70 Å for  $\text{C}\alpha-\text{H}$  to O atom, 1.83 Å for N1-H to O atom, and 2.23 Å for N2-H to O atom



**Table 1** Comparison of free energy of binding of the respective ligands with the corresponding peptides expressed in kcal mol<sup>-1</sup> indicates that DS3 shows higher affinity. Values of the inhibition constant obtained from the AutoDock program are listed in parentheses

|           | HPO <sub>4</sub> <sup>2-</sup> | HAsO <sub>4</sub> <sup>2-</sup> | HAsO <sub>3</sub> <sup>2-</sup> | H <sub>2</sub> AsO <sub>3</sub> <sup>1-</sup> |
|-----------|--------------------------------|---------------------------------|---------------------------------|---|
| DS1       | -2.51 (14.19 mM)               | -2.51 (14.22 mM)                | -3.09 (5.49 mM)                 | -2.58 (11.62 mM)                              |
| DS3       | -3.13 (4.76 mM)                | -3.08 (5.21 mM)                 | -3.8 (1.53 mM)                  | -2.92 (6.33 mM)                               |
| CPS224Ac  | -1.66 (59.08 mM)               | -1.71 (54.60 mM)                | -2.37 (17.11 mM)                | -1.67 (51.28 mM)                              |
| CPS226    | -1.97 (33.81 mM)               | -2.02 (32.09 mM)                | -2.78 (8.81 mM)                 | -2.25 (39.77 mM)                              |
| SCPS224Ac | -2.39 (16.29 mM)               | -2.47 (15.03 mM)                | -3.43 (2.71 mM)                 | -2.49 (13.43 mM)                              |
| SCPS226   | -1.72 (52.18 mM)               | -1.77 (49.64 mM)                | -2.54 (13.30 mM)                | -2.14 (26.30 mM)                              |

for HAsO<sub>4</sub><sup>2-</sup> with DS3 whereas average angular values observed are 137.52° for C $\alpha$ -H···O, 164.37° for N1-H···O, and 159.34° for N2-H···O. Both the values of distance and angle demonstrate a weak H-bond interaction between the C $\alpha$  atom and O, while on the other side, H-bonds formed between N1/N2 with O of the ligand are quite strong as the distance observed is <2.5 Å and the X-H···O angle is >150°. Both the distance and angle distribution of C $\alpha$ -H have more variation with respect to those of N1-H and N2-H atoms (ESI Fig. S3 and S4†). This result is in accordance with the H-bond distance and angle as reported in the crystal structures of the 'C $\alpha$ NN' motif<sup>44</sup> and that observed for the SO<sub>4</sub><sup>2-</sup> and HPO<sub>4</sub><sup>2-</sup> ions with the scaffolds.<sup>45,46,48</sup> Respective average distance values of (C $\alpha$ )H···O, (N1)H···O and (N2)H···O for the aforesaid ligand interaction of HAsO<sub>3</sub><sup>2-</sup> with DS3 are 2.20 Å, 1.99 Å, and 2.07 Å respectively; while the corresponding angles values are 140.60°, 158.95°, and 165.92° in the similar order. This clearly infers better recognition of HAsO<sub>3</sub><sup>2-</sup> in comparison to HAsO<sub>4</sub><sup>2-</sup> as obtained through binding free energies. Comparing the distribution pattern of interactions between HAsO<sub>3</sub><sup>2-</sup> and DS3 with those of HAsO<sub>4</sub><sup>2-</sup> and DS3, one would conclude that for HAsO<sub>3</sub><sup>2-</sup> distribution of resulting distances and angles of C $\alpha$ -H and N1-H are attributed to two separate 'normal-like' distributions within a range of very small difference (ESI Fig. S3†). Although Ser2-C $\alpha$ -H of DS3 interacts with the O atom of HAsO<sub>3</sub><sup>2-</sup> with an overall average distance of 2.20 Å, upon a keen observation, the entire distribution can be seen as two separate 'normal-like' distributions with a distance range of 1.877–1.984 Å and another one of 2.44–2.54 Å (ESI Fig. S3-G†). Similarly, for Lys3-NH, although the overall average distance is 1.9 Å, actually it is a combination of two closely spaced 'normal-like' distributions (ESI Fig. S3-H†). However, for HAsO<sub>4</sub><sup>2-</sup>, a similar interaction produced a single 'normal-like' distribution (ESI Fig. S3-A–F†) that is slightly left-skewed. This observation with HAsO<sub>3</sub><sup>2-</sup> can be justified from the three-dimensional coordinates of all the generated conformations obtained from the docking algorithm. Although the built-in tool of AutoDock clusters all the 250 poses into a single cluster in terms of binding free energy, a closer look reveals that all the 250 conformations can be forced into two very closely separated clusters depending on the orientations of the ligand. Superposition of the conformations of the two small clusters (being similar in terms of binding free energy) related to HAsO<sub>3</sub><sup>2-</sup> differs in RMSD only by ~0.27 Å. This difference can be understood by the existence of two different states/orientations (S1 and S2) of the ligand (Fig. 1-a1 and a2). The interacting

oxygen atoms belonging to these two different conformations of HAsO<sub>3</sub><sup>2-</sup> (S1 and S2) are designated as S1O1 and S2O1 (Fig. 1-a1 and a2). Distance (*D*) from the LYS3-N-H and SER2-C $\alpha$ -H atoms to both S1O1 and S2O1 varies considerably (shown in red and cyan colors, respectively). Hydrogen bonding distance of the ligand oxygen atoms of these two different states/orientations (S1 and S2) of HAsO<sub>3</sub><sup>2-</sup> from the main chain C $\alpha$ NN participating residues SER2 and LYS3 of DS3 are as follows:

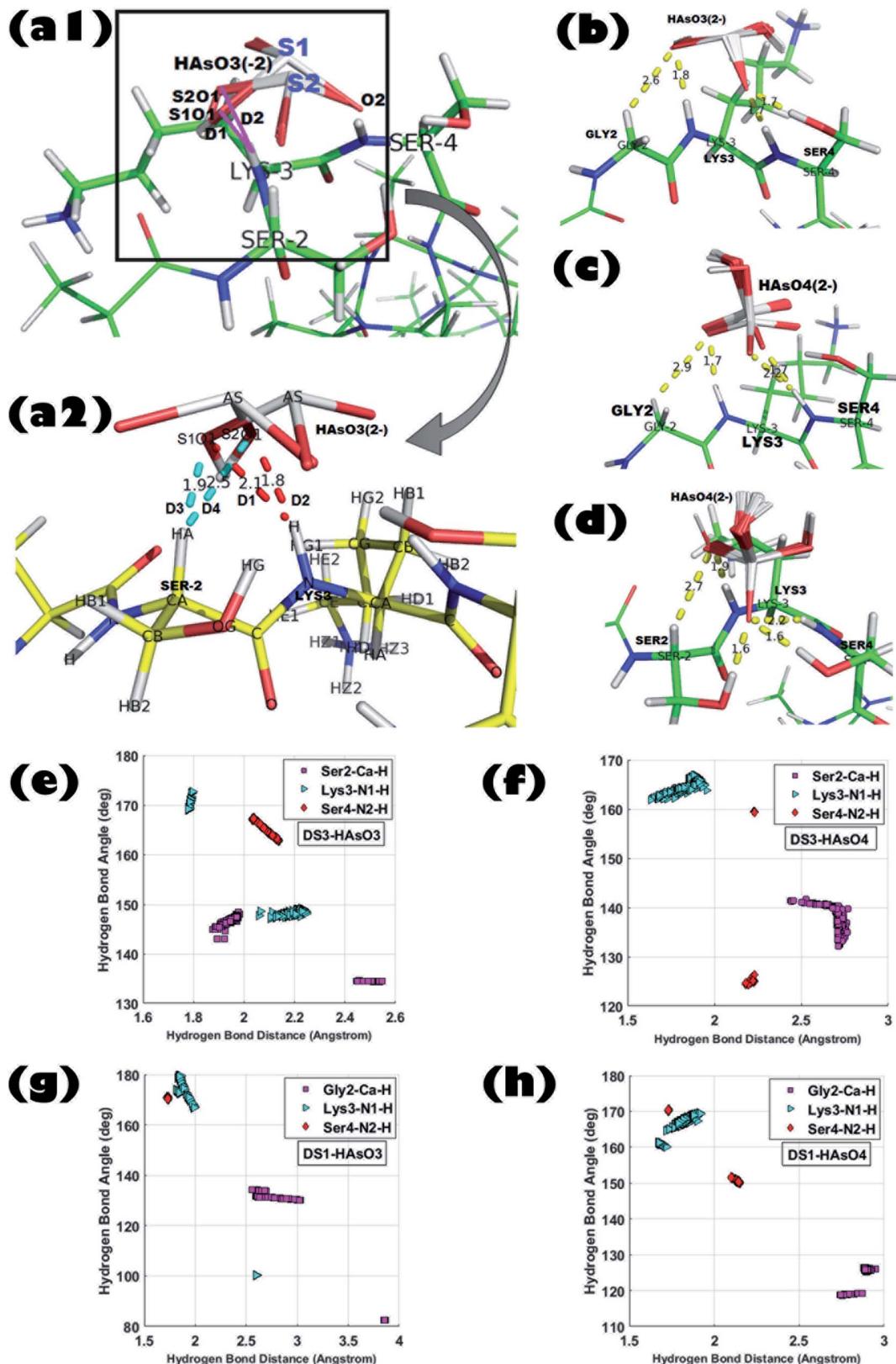
D1: LYS3-N-H to S1O1 (2.1 Å); D2: LYS3-N-H to S2O1 (1.8 Å).

D3: SER2-C $\alpha$ -H to S1O1 (1.9 Å); D4: SER2-C $\alpha$ -H to S2O1 (2.5 Å).

A similar pattern of distribution was also observed for the angle  $\angle$ X-H···O constituted by Ser2 and Lys3 with HAsO<sub>3</sub><sup>2-</sup> (ESI Fig. S3-J and K†). Such an interesting observation of having two conformations (S1 and S2) for HAsO<sub>3</sub><sup>2-</sup> (Fig. 1-a1 and a2) but not for HAsO<sub>4</sub><sup>2-</sup> can perhaps be explained by the presence of the lone pair of electrons on As in HAsO<sub>3</sub><sup>2-</sup> (oxidation state of As +3) which allows flipping of the pseudotetrahedral geometry of HAsO<sub>3</sub><sup>2-</sup> generating two different, but energetically closely related, conformations. Obviously, such structural tumbling/flipping is not feasible for HAsO<sub>4</sub><sup>2-</sup> due to its rigid tetrahedral geometry. Further, in the case of DS1, due to presence of flexible Gly in the second position, such an anion dependent distinct feature may not be clearly observed (ESI Fig. S4†). Glycine, having no side chain, can use both its enantiotropic hydrogens (connected at the C $\alpha$ -atom) to recognize the O-atom of the ligand through hydrogen bonding. Inclusion of flexibility in the anion recognition scaffold of the peptide gives less variation in the bonding parameters although producing a slight translation in the distribution range (either left or right skewed) for both anions. This observation leads to a conclusion that orientation of the ligand (anion) is not alone responsible for the perturbation in the interaction pattern; the sequence/3D-structure of the host peptide participating in the recognition would also affect the interaction parameters which are reflected in the binding free energy. It is also noted that, apart from the main chain hydrogen bond, there is also an add-on contribution from the side chain of the Ser residue(s) of DS3 in ligand binding (confirmed in the Molecular dynamics simulation section) which increases stability of the interaction.

In a nutshell, from the docking analysis one can conclude that effectiveness of the recognition of arsenic anions depends on local information embedded in the peptide sequence and conformation as observed for the sulfate and the phosphate ions. Further, the results obtained distinctly identify that in





**Fig. 1** Superposition of all 250 ligand conformations generated from interactions between peptide scaffold and the respective anion obtained from Molecular Docking experiments; (a1) DS3 (Ser2-Lys3-Ser4) : HAsO<sub>3</sub><sup>2-</sup> [ligand binding conformations are shown in zoomed view by taking one ligand conformer from each state in figure (a2). Distances of the S1O1 and S2O1 atom from the Lys3-N-H and Ser2-C $\alpha$ -H atom are shown in red and cyan colors, respectively]. (b) DS1 (Gly2-Lys3-Ser4) : HAsO<sub>3</sub><sup>2-</sup>, (c) DS1 (Gly2-Lys3-Ser4) : HAsO<sub>4</sub><sup>2-</sup>, and (d) DS3 (Ser2-Lys3-Ser4) : HAsO<sub>4</sub><sup>2-</sup>; H-bond distance and angle distribution of 'C $\alpha$ NN' motif residues of DS3 peptide with the respective anions in (e), (f), 'C $\alpha$ NN' motif residues of DS1 peptide with the respective anions in (g) and (h).

a context free system the designed peptide scaffolds (DS1 and DS3) recognize As ligands favorably over the naturally occurring 'C<sup>2</sup>NN' sequences. These designed scaffolds, DS1 and DS3, are considered as potentially active to act as scavengers of As for its removal. However, as in aqueous solution,  $\text{HAsO}_3^{2-}$  exists in a much higher than physiological pH ( $\text{p}K_2$  of  $\text{H}_3\text{AsO}_3$  is 12.13), therefore, further experiments were focused on  $\text{HAsO}_4^{2-}$  (existing at physiological pH) only.

### 3.4 Mass spectrometry

Appearance of  $m/z$  values at 1705.1, 1727.1, and 1743.1 corresponding to  $\text{M} + \text{H}^+$ ,  $\text{M} + \text{Na}^+$ , and  $\text{M} + \text{K}^+$  added species, respectively, in the positive mode of MALDI validated the existence of DS3 (MW 1704). Further, appearance of no other  $m/z$  values with appreciable intensity clearly justified the purity of the sample. Molecular association of arsenate with peptide DS3 (added as  $\text{Na}_2\text{HAsO}_4$  in 1 : 3 M ratio of peptide : arsenate) was established in the negative mode of ESI-MS [ $\text{H}_2\text{O}/\text{CH}_3\text{CN}$  (1 : 1) + 0.1%  $\text{NH}_3$ ]. Appearance of a peak corresponding to  $m/z$  value 921.9 established the binding of arsenate ( $\text{HAsO}_4^{2-}$ ) with the peptide DS3 (Fig. 2). Moreover, isotopic distribution of peaks corresponding to differences in  $m/z$  values of 0.5 (921.4, 921.9, and 922.4; data not shown) signified recognition of the divalent arsenate anion ( $\text{HAsO}_4^{2-}$ ) by the peptide DS3 and validated the observation of the molecular docking experiment. However, the  $m/z$  value at 850.9 corresponds to un-reacted DS3.

### 3.5 Circular dichroism spectroscopy

In order to explore conformational changes of the peptide DS3 upon interaction with arsenate, if any, CD spectra of the peptide were recorded in absence as well as in presence of added arsenate ions at room temperature (25 °C). Peptide DS3 in a fully aqueous condition showed characteristic double minima at  $\sim 202$  ( $\pi - \pi_{\parallel}^*$ ) and  $222$  ( $\text{n}-\pi^*$ ) nm (Fig. 3) corresponding to a helical signature of the peptide,<sup>75</sup> which may arise from the contributions of admixture of  $3_{10}$ - and  $\alpha$ -helices. After addition of arsenate ions (added as  $\text{Na}_2\text{HAsO}_4$  in 1 : 3 M ratio of peptide : anion), a CD spectrum of the peptide also was recorded. The CD spectrum of the arsenate-added peptide (Fig. 3) showed

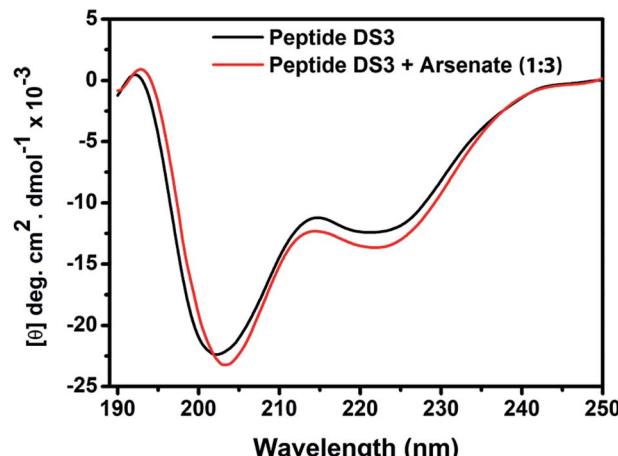


Fig. 3 CD spectrum of the peptide DS3 (represented as black line) along with its arsenate ( $\text{HAsO}_4^{2-}$ ) added (1 : 3 peptide : arsenate) species (represented as red line) measured at 298 K, showing enhancement of CD signals for both  $\text{n}-\pi^*$  ( $\sim 222$  nm) and  $\pi - \pi_{\parallel}^*$  ( $\sim 208$  nm) transitions after addition of arsenate to peptide DS3.

enhanced CD signals in both the ( $\pi - \pi_{\parallel}^*$ ) and  $\text{n}-\pi^*$  transitions (especially  $\text{n}-\pi^*$  transition at 222 nm). This signified an increment of helical population of the peptide DS3 upon interaction with the arsenate ions as observed for the  $\text{SO}_4^{2-}$  and the  $\text{HPO}_4^{2-}$  ions with the naturally occurring 'C<sup>2</sup>NN' motif sequences.<sup>45,47</sup> In order to quantify the increment in helicity upon interaction with the arsenate ion, the ratio  $R_2$  ( $[\theta]_{\text{n}-\pi^*}/[\theta]_{\pi - \pi_{\parallel}^*} \approx [\theta]_{222}/[\theta]_{208}$ )<sup>76-78</sup> was calculated. Upon addition of  $\text{HAsO}_4^{2-}$  to DS3, an enhanced magnitude of  $R_2$  (from 0.50 to 0.62) of DS3 along with a  $\sim 2$  nm bathochromic shift at ( $\pi - \pi_{\parallel}^*$ ) transition ( $\sim 204$  nm  $\rightarrow$   $\sim 206$  nm) strongly recognized the interaction of the anion with the peptide along with the enhancement of helicity upon recognition and corroborated the output of the molecular docking experiment.

### 3.6 Isothermal titration calorimetric (ITC)

The thermodynamic parameters of bio-molecular interactions can be described using isothermal titration calorimetry (ITC).<sup>79</sup>

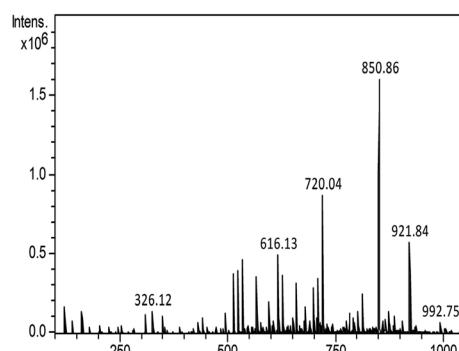
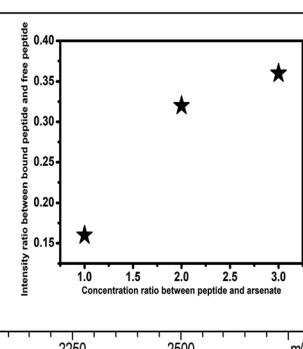


Fig. 2 ESI-MS data of arsenate ( $\text{HAsO}_4^{2-}$ ) added to peptide (DS3). Binding of arsenate with the peptide DS3 was established through appearance of the  $m/z$  value of 921.9. Moreover, un-reacted peptide can be identified by the  $m/z$  value of 850.9 (inset shows the scatter plot of intensity ratio between arsenate-bound peptide and free peptide as a function of added arsenate ion concentration with respect to the peptide. As the concentration of added arsenate was increased, the ratio was enhanced from 0.15 to 0.35 and subsequently becomes saturated when arsenate is added 3 times than that of the peptide).



These interactions involve a wide range of molecular associations, such as metal ions binding to proteins,<sup>80–82</sup> protein–protein interactions,<sup>83</sup> or ligand–protein complex formations.<sup>84,85</sup> Here, interaction of the peptide DS3 with arsenate ( $\text{HAsO}_4^{2-}$ ) was further studied through ITC. The ITC data of binding interactions between peptide DS3 and arsenate, shown in Fig. 4, established an endothermic mode of interaction. From the raw ITC thermogram of peptide DS3 titrated against arsenate (Fig. 4), the associated thermal changes were plotted against the molar ratio. To substantiate binding affinity of the peptide for arsenate, the dissociation constant ( $K_d$ ) was measured from the plotted curve. The observed  $K_d$  value of  $\sim 8$   $\mu\text{M}$  signified an appreciable amount of interaction of arsenate with the peptide DS3, which is even better than the interaction involving peptide CPS224Ac/CPS226 (containing naturally occurring ‘C<sup>z</sup>NN’ motif) with sulfate.<sup>47</sup> From the magnitude of the thermodynamic parameters like enthalpy ( $\Delta H$ : 1265 cal  $\text{mol}^{-1}$ ) and entropy ( $\Delta S$ : 27.7 cal  $\text{mol}^{-1}$  deg $^{-1}$ ), the free energy ( $\Delta G$ ) of the interaction was calculated. The observed physical parameters dictate that the interaction is an entropy driven process (positive  $\Delta S$  value) having free energy ( $\Delta G$ ) of  $-7$  kcal  $\text{mol}^{-1}$ . This amount of free energy of the system

corroborates an appreciable and spontaneous interaction between the peptide DS3 and arsenate.

### 3.7 Molecular dynamics experiment

Biophysical experiments (using mass spectrometry, CD spectroscopy, and ITC) clearly indicated the recognition of arsenate by the DS3 peptide. This interaction information of the peptide DS3 and arsenate at the atomic level can further be retrieved as a function of time from the Molecular Dynamics (MD) simulation experiment, as the previous experiments gave only time averaged results. Investigation of interactions between the arsenate ( $\text{HAsO}_4^{2-}$ ) ions and DS3 portray a major similarity in the signature of the interactions of the naturally occurring ‘C<sup>z</sup>NN’ motif and  $\text{SO}_4^{2-}$  ions<sup>46</sup> along with DS3 and  $\text{SO}_4^{2-}$  ions.<sup>48</sup> This may be due to the fact that since both the anions have similar stereochemistry along with net charge content, the inner line 3d orbital of As makes its van der Walls radius larger than that of either phosphorous or sulfur. Previous investigations using the simulation experiment of DS3 with sulfate suggested a more stable binding of the sulfate ion with DS3 ( $\sim 4.9$  ns residence time of  $\text{SO}_4^{2-}$ )<sup>48</sup> than with the natural ‘C<sup>z</sup>NN’ sequence present in the CPS224Ac peptide (residence time of  $\text{SO}_4^{2-}$  ion only 400 ps).<sup>46</sup> Here, the time dependent behavior of the  $\text{HAsO}_4^{2-}$  ion with the Ser2-Lys3-Ser4 sequence of the DS3 scaffold revealed that the anion interacts with the core of the ‘C<sup>z</sup>NN’ motif up to  $\sim 1.06$  ns (1 ns and 60 ps) mediated through a hydrogen bond (Fig. 5a). Time dependent behavior of the torsion angles and other related parameters of interactions of arsenate ( $\text{HAsO}_4^{2-}$ ) with the scaffold (DS3) is briefed in the underlying study. Average hydrogen bonding (H-bond) distance between the O-atoms of the anion and Ser2- $\text{C}\alpha$ -H is  $3.0\text{ \AA}$  ( $\pm 0.6$ ), whereas that with Lys3-N1-H and Ser4-N2-H are  $2.6\text{ \AA}$  ( $\pm 0.9$ ) and  $2.4\text{ \AA}$  ( $\pm 0.6$ ), respectively (up to the residence time of ligand with ‘C<sup>z</sup>NN’, *i.e.*, 1.06 ns) (Fig. 5b). Distance of the  $\text{C}\alpha$  atom with the ligand was more, as expected, than that of N1 and N2, indicating the comparatively weaker nature of this interaction. This is at par with the observation obtained from the molecular docking experiments. However, the standard deviation for the N1 was the largest. Along with the interactions between the main chain atoms of the ‘C<sup>z</sup>NN’ motif of the peptide DS3 with  $\text{HAsO}_4^{2-}$ , interactions from the side chain were also observed as follows: Ser4-OG (side-chain oxygen) (Fig. 6a) 1350–1450 ps, 1590–1780 ps, 2050–2210 ps; LYS3-N-H (side-chain –NH): 2380–3780 ps. Terminal presence of the Ser2 facilitated attainment of various backbone torsions relatively more freely maintaining the hydrogen bond (facilitated *via* the side-chain HG1 atom) with the ligand. This makes the N1 to point obliquely which may be one of the reasons for its greater distance fluctuation with the anion. This was not observed with the molecular docking because, due to rigid docking, the geometry of the host molecule (peptide) is fixed. Distance and torsion parameters for the Ser4 residue were mostly constant in the simulated trajectory because it has less freedom when present at the cap of the N-term of helical ABGY. An unusual behavior of torsion fluctuation was noticed in the Ser2 residue where it switches to a left handed  $\alpha$ -helical state ( $\alpha\text{L}$ ) from 200 ps to 4200 ps and

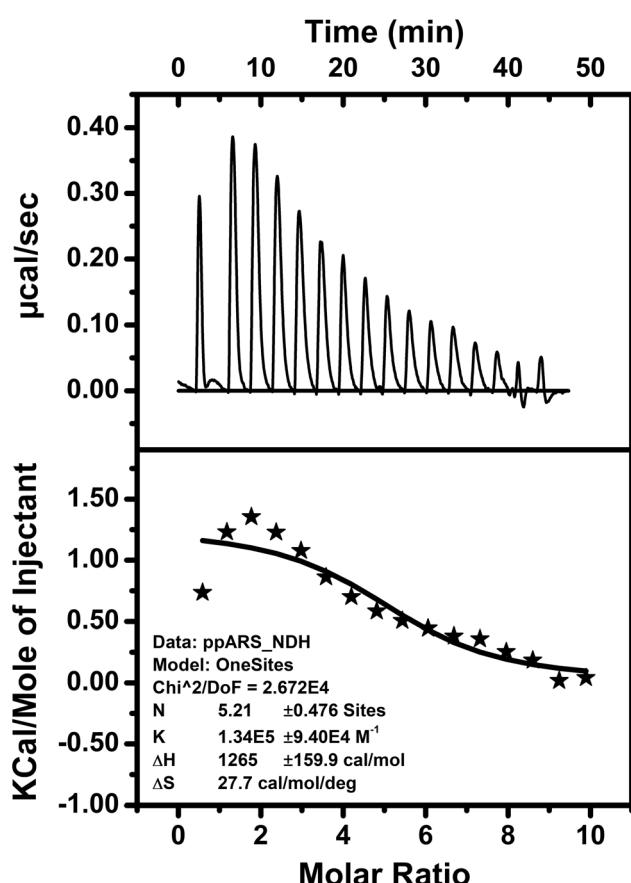


Fig. 4 ITC figure representing the peptide (DS3)–arsenate interaction measured using Microcal ITC200 (Malvern) at 25 °C. Top panel represents the raw thermogram of the interaction while the lower panel represents the best fit to a single site model along with the thermodynamic parameters.



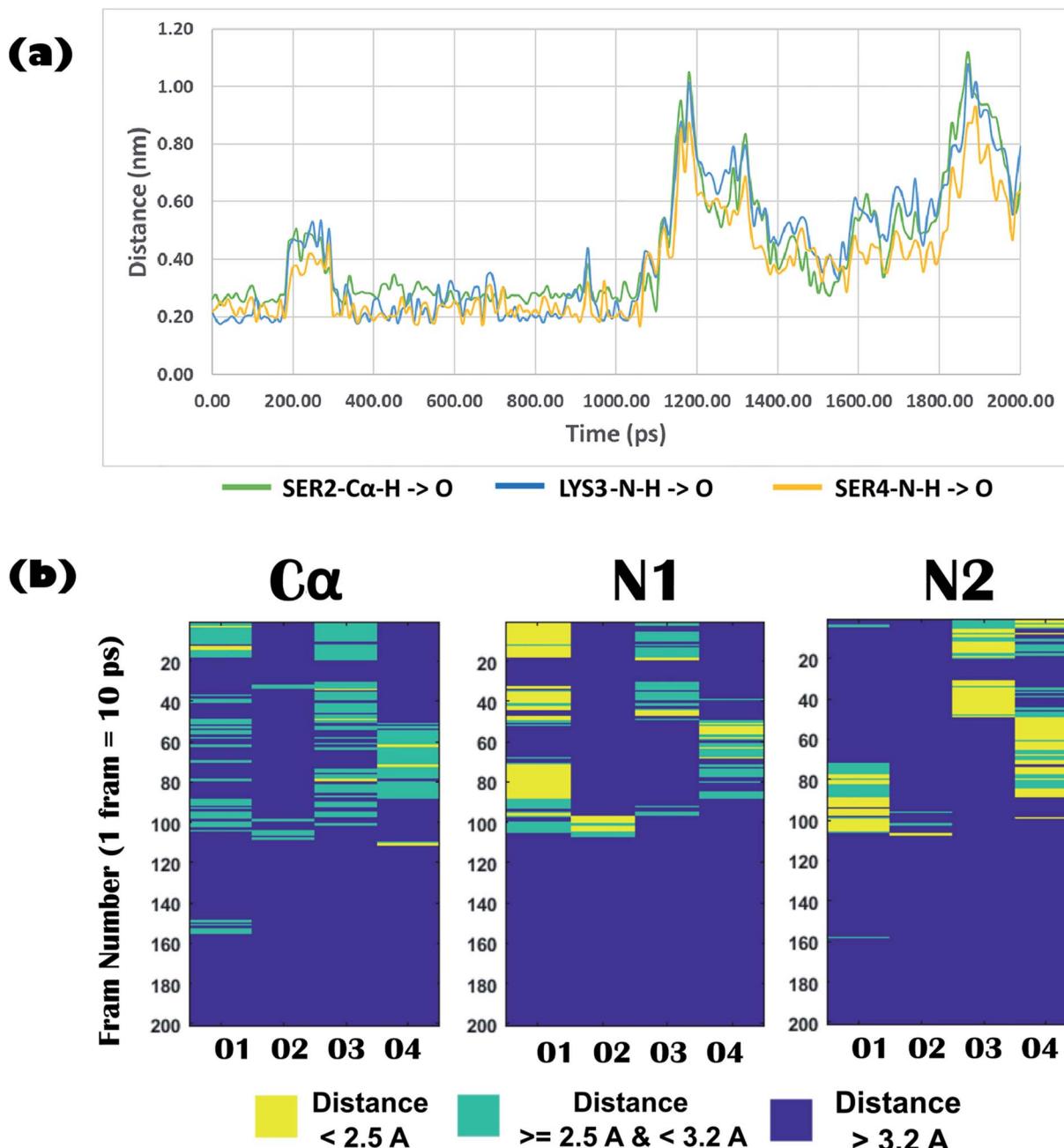


Fig. 5 (a) Time dependent variation of distance of the 'C'NN' motif residues ( $\text{C}\alpha$  of Ser2, main-chain N–H of Lys3, main-chain N–H of Ser4) of DS3 and oxygen atoms of Arsenate ligand ( $\text{HAsO}_4^{2-}$ ). Up to 2 ns of the MD trajectory is shown in this figure where it can clearly be seen that the ion  $\text{HAsO}_4^{2-}$  interacts with the C'NN region of DS3 up to 1.06 ns. (b) Distance profile of O atoms (up to 2 ns) of ligand ( $\text{HAsO}_4^{2-}$ ) from the Ser2- $\text{C}\alpha$ -H, Lys3-N1-H, and Ser4-N2-H of 'C'NN' motif residues of DS3 in time dependent manner resembling the C'NN- $\text{HPO}_4^{2-}$  interaction as reported in literature [X axis: oxygen atom number of the ligand  $\text{HAsO}_4^{2-}$ ; Y axis: frame number (1 frame = 10 ps, i.e., up to 2000 ps to 2 ns is shown)].

again from approximately  $\sim 15$  310 ps to 17 110 ps, otherwise, it remained in the  $\beta$  conformation (ESI Fig. S5†). However, effects of arsenate on this transition are not clearly understood as in the case of torsion transition for a later time interval, where the ligand was distantly located ( $\sim 60$  Å) from the scaffold. In contrast to the previous study, the Lys3 residue remained mostly in an alpha helical (right-handed) state throughout the entire simulation time (ESI Fig. S5†) which may assist with

recognizing the anion as the N1 residue in the 'C'NN', which appears with  $\alpha$ -helical geometry upon anion recognition. Within this helical adapted state, participation of the main-chain N–H atom of Lys3 and Ser4 in binding the ligand is noticeable (Fig. 6). Considering the first 5 ns of simulation trajectory, it can be seen that Ser2 mostly occupies the  $\alpha_L$  torsion state (as mentioned earlier). Up to 5 ns the average torsion angle values for the motif in DS3 are as follows: Ser2: phi

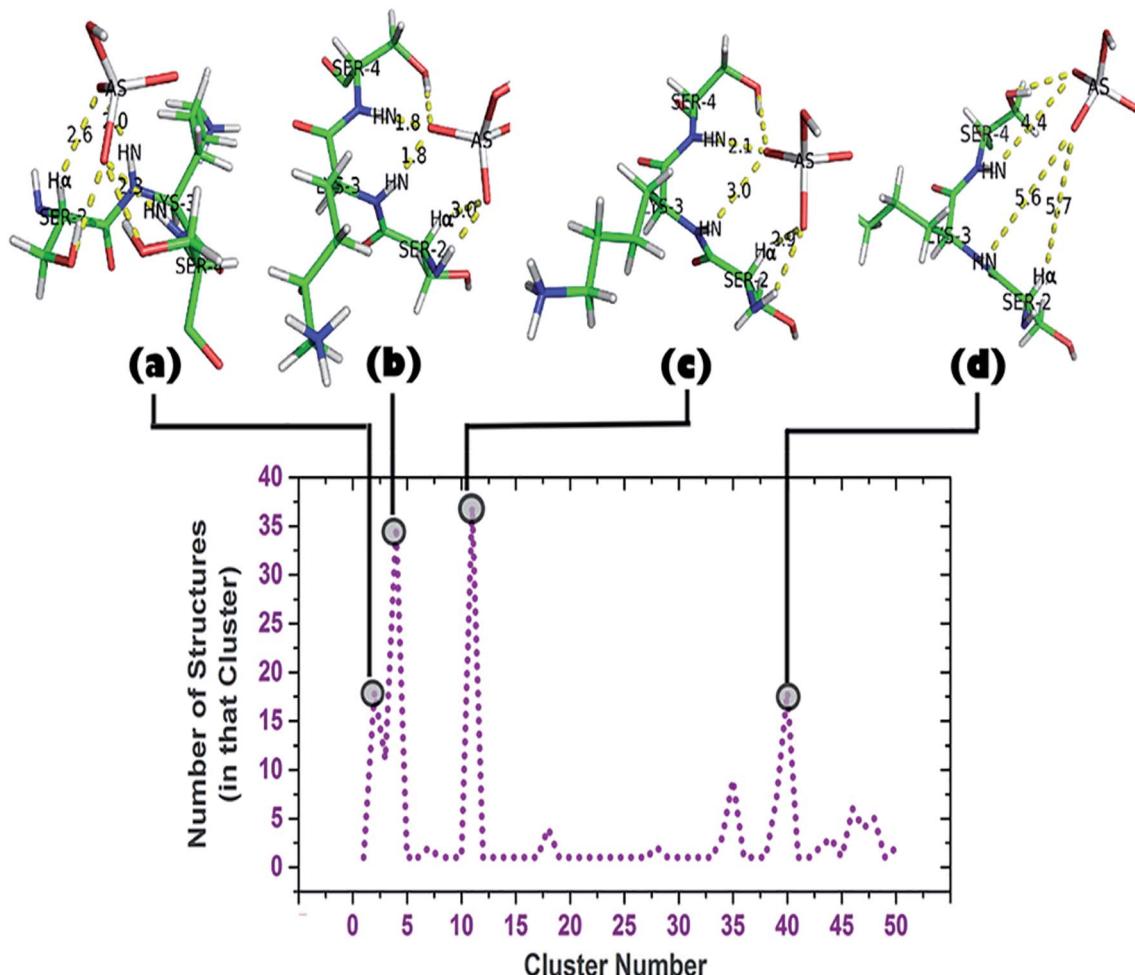


Fig. 6 Cluster analysis of the trajectory up to 2 ns ( $\text{DS3-HAsO}_4^{2-}$ ). Total of 50 clusters were found, each having a 'representative structure' and number of structures belonging to that 'representative structure' are shown beside. Top clusters (clusters containing largest numbers of structures: 2, 4, 11, and 40) and their corresponding conformations are shown.

$(\phi)$  is  $41.12 (\pm 52.29)$ ,  $\psi$  ( $\psi$ ) is  $64.32 (\pm 38.36)$ , Lys3:  $\phi$  ( $\phi$ ) is  $-71.06 (\pm 17.19)$ ,  $\psi$  ( $\psi$ ) is  $-35.86 (\pm 12.44)$ , Ser4:  $\phi$  ( $\phi$ ) is  $-66.58 (\pm 7.87)$ , and  $\psi$  ( $\psi$ ) is  $-39.56 (\pm 7.85)$  (ESI Fig. S5†). Fluctuation of the terminal amino acid residues in the protein backbone are obvious. But, with up to  $\sim 1.06$  ns the torsion angles of Ser2 remain very stable having  $\phi$  ( $\phi$ ) of  $67.11 (\pm 14.19)$  and  $\psi$  ( $\psi$ ) of  $60.42 (\pm 27.14)$  and maintaining  $\alpha\text{L}$  geometry (ESI Fig. S5† inset) indicating constraints for association of anions. Alongside, torsion angular values for Lys3 and Ser4 up to 1.06 ns are as follows: Lys3,  $\phi$  ( $\phi$ ) is  $-90.87 (\pm 23.62)$  and  $\psi$  ( $\psi$ ) is  $-22.95 (\pm 14.53)$ ; Ser4 -  $\phi$  ( $\phi$ ) is  $-65.86 (\pm 6.88)$ , and  $\psi$  ( $\psi$ ) is  $-40.77 (\pm 6.44)$ .

A cluster analysis was performed on the truncated simulated trajectory (up to 2 ns) by the default 'gmx cluster' tool available within GROMACS 5.1.2 to identify the principal ligand bound with the motif under investigation. Each cluster represents a particular conformational state and also shows the number of structures belonging to that state. It helps to decide the top clusters with most favorable geometry comprising a larger number of participant states. Participation of Ser2 can be seen

in the top two representative structures of the top two clusters (Fig. 6b and c). A total of 50 clusters were considered from the cluster analysis with a RMSD range of  $0.06374\text{--}0.36899$  Å and representative structures from the most populated clusters are shown in Fig. 6. Cluster 2, having 18 structures, shows a typical anion binding pattern with the Ser2-Lys3-Ser4 residues where one oxygen interacts with  $\text{C}\alpha$ -H and N1-H, while N2-H interacts with another oxygen of arsenate ( $\text{HAsO}_4^{2-}$ ) resembling the naturally occurring ' $\text{C}^2\text{NN}$ ' motif with phosphate.<sup>44</sup> A distant location of a ligand from the scaffold can be seen in cluster 40 which captures the ejection of the arsenate. Another observation is the participation of a side chain of lysine residues in the simulation trajectory, contrary to the DS3-sulfate simulation<sup>48</sup> where, after ejection of the anion, the anion remains free in the system. Within time steps 32 600–41 000 ps (approx.) arsenate was found to interact with both the Lys8 and Lys11 side-chains. Apart from the 1.06 ns interval of  $\text{HAsO}_4^{2-}$ -DS3 interaction, the Lys3 side-chain again is very near to the ligand within a 45 000–49 000 ps (approx.) interval. A distance profile of the  $\text{HAsO}_4^{2-}$  ligand from all the residues of the DS3 peptide is also shown

(ESI Fig. S6†) to decipher the aforementioned feature of Lys residues. As mentioned previously, the bulky nature of arsenic (larger atomic radius) in comparison to sulfur may enhance the influences of a charged side-chain of lysine on the ligand in the overall conformational landscape by decreasing the mobility of  $\text{HAsO}_4^{2-}$ . Though  $\text{SO}_4^{2-}$  has a higher affinity towards the 'C<sup>z</sup>NN' segment of DS3, in a competitive binding environment the DS3-HAsO<sub>4</sub><sup>2-</sup> interaction will be favored because of retention ability by the 'C<sup>z</sup>NN' segment as well as the Ser and Lys side-chains. This motivated us to conclude that DS3 can act as a scavenger of  $\text{HAsO}_4^{2-}$ .

## 4 Conclusions

The environmental pervasiveness of arsenic is a matter of deep concern<sup>13</sup> due to its high occurrence in groundwater as arsenate or arsenite (along with its methylated derivatives), which interfere with metabolic pathways and normal cellular activities. Site specific ion recognition is the major driving force behind various biological processes<sup>86-89</sup> where loop regions of proteins with higher solvent accessible surface areas<sup>90</sup> interact with As ions utilizing various motifs. Peptide-based model scaffolds mimicking the natural protein-ligand interaction provide rich information regarding binding affinity, mode, and the contribution of motif geometry. Utilizing this concept, we initiated an effort to study peptide-arsenate/arsenite interactions that ultimately could be used for scavenging arsenic anions. In the present study, the participation of main-chain  $\text{C}\alpha_{i-1}-\text{H}$ ,  $\text{N}_i-\text{H}$ ,  $\text{N}_{i+1}-\text{H}$  residues of consecutive amino-acids, based on the 'C<sup>z</sup>NN' motif was utilized to recognize different anions of arsenic. Comparable stereochemistry between  $\text{HAsO}_4^{2-}$  and  $\text{HPO}_4^{2-}$  acted as the motivating spirit and both of the anions showed similar influence on the designed 'C<sup>z</sup>NN' motif in a context-free system that had similar interacting properties with the conserved binding pattern. Further, a kinetic model of interaction was validated through a molecular dynamics simulation of the DS3-HAsO<sub>4</sub><sup>2-</sup> docked conformation (having well-inferred the binding ability of DS3 in comparison to other peptides) where the residence time for  $\text{HAsO}_4^{2-}$  was 1.06 ns for the scaffold DS3, which is quite appreciable. Though no interaction with a side chain was found for sulfate with DS3 once the anion left the 'C<sup>z</sup>NN' (Ser2-Lys3-Ser4) segment, some extent of interaction was found for  $\text{HAsO}_4^{2-}$  through the participation of side-chains on polar residues that enhance the interactions between  $\text{HAsO}_4^{2-}$  ions and the peptide scaffold (DS3). Reduced flexibility of movement for the As ligand, due to its larger size and diffused electron density, provided an additional effect toward the retention of the anions. All these data strongly suggest that the 'C<sup>z</sup>NN'-based scaffold DS3 can well recognize anions of arsenic and be their scavenger through attachment in a water filtration unit. Ligand binding abilities by the conserved geometry presented in this work, with the help of a 'natural replica' of the 'C<sup>z</sup>NN' motif, contribute significantly towards understanding the context of free peptide-ligand recognition, which may be utilized for scavenging toxic anions through peptidomimetics.

## Conflicts of interest

There are no conflict to declare.

## Acknowledgements

Authors extend their sincere thanks to MBU-IISc, Bengaluru for providing the MS facilities; Bose Institute, Kolkata for CD facilities; and Bioinformatics Lab (DBT-BIF) WBUT for the computational facilities. Swami Vivekananda Research Grant Award, State DST, Govt. of West Bengal (Sanction No. 17.6/WBSCST/F/0530/14), and the Bioinformatics Infrastructure Facilities-Department of Biotechnology, Govt. of India (Sanction No. BT/BI/25/020/2012 (BIF)) are acknowledged for their financial support.

## References

- 1 S. Shankar, U. Shanker and Shikha, *Sci. World J.*, 2014, **2014**, 1–18.
- 2 L. Rodriguez-Lado, G. Sun, M. Berg, Q. Zhang, H. Xue, Q. Zheng and C. A. Johnson, *Science*, 2013, **341**, 866–868.
- 3 J. I. Anetor, H. Wanibuchi and S. Fukushima, *Asian Pac. J. Cancer Prev.*, 2007, **8**, 13–23.
- 4 J. P. Campbell and J. A. Alvarez, *Am. Fam. Physician*, 1989, **40**, 93–97.
- 5 A. H. Hall, *Toxicol. Lett.*, 2002, **128**, 69–72.
- 6 P. Ravenscroft, H. Brammer, and K. S. Richards, *Arsenic pollution: a global synthesis*, Wiley InterScience (Online service), Wiley-Blackwell, 2009.
- 7 C. K. Jain, *Indian J. Environ. Health*, 2002, **44**, 238–243.
- 8 A. K. Bera, T. Rana, S. Das, D. Bhattacharya, S. Bandyopadhyay, D. Pan, S. De, S. Samanta, A. N. Chowdhury, T. K. Mondal and S. K. Das, *Toxicol. Ind. Health*, 2010, **26**, 709–716.
- 9 M. M. Rahman, U. K. Chowdhury, S. C. Mukherjee, B. K. Mondal, K. Paul, D. Lodh, B. K. Biswas, C. R. Chanda, G. K. Basu, K. C. Saha, S. Roy, R. Das, S. K. Palit, Q. Quamruzzaman and D. Chakraborti, *J. Toxicol. Clin. Toxicol.*, 2001, **39**, 683–700.
- 10 M. M. Rahman, M. K. Sengupta, S. Ahamed, U. K. Chowdhury, D. Lodh, A. Hossain, B. Das, N. Roy, K. C. Saha, S. K. Palit and D. Chakraborti, *Bull. W. H. O.*, 2005, **83**, 49–57.
- 11 S. Rahaman, A. C. Sinha, R. Pati and D. Mukhopadhyay, *Environ. Geochem. Health*, 2013, **35**, 119–132.
- 12 P. A. O'Day, *Elements*, 2006, **2**, 77–83.
- 13 M. Hughes, *Toxicol. Lett.*, 2002, **133**, 1–16.
- 14 T. E. Barrett, R. Savva, G. Panayotou, T. Barlow, T. Brown, J. Jiricny and L. H. Pearl, *Cell*, 1998, **92**, 117–129.
- 15 D. Cobessi, F. Tête-Favier, S. Marchal, S. Azza, G. Branlant and A. Aubry, *J. Mol. Biol.*, 1999, **290**, 161–173.
- 16 A. O. Summers, *Curr. Opin. Microbiol.*, 2009, **12**, 138–144.
- 17 D. J. Thompson, *Chem.-Biol. Interact.*, 1993, **88**, 89–114.
- 18 R. N. Ratnaike, *Postgrad. Med. J.*, 2003, **79**, 391–396.
- 19 J. L. Webb, *Enzyme and metabolic inhibitors*, Academic Press, New York, vol. III, 1966.



20 C. O. Abernathy, Y. P. Liu, D. Longfellow, H. V Aposhian, B. Beck, B. Fowler, R. Goyer, R. Menzer, T. Rossman, C. Thompson and M. Waalkes, *Environ. Health Perspect.*, 1999, **107**, 593–597.

21 K. Jomova, Z. Jenisova, M. Feszterova, S. Baros, J. Liska, D. Hudecova, C. J. Rhodes and M. Valko, *J. Appl. Toxicol.*, 2011, **31**, 95–107.

22 H. Y. Chiou, W. I. Huang, C. L. Su, S. F. Chang, Y. H. Hsu and C. J. Chen, *Stroke*, 1997, **28**, 1717–1723.

23 C.-H. Tseng, C.-K. Chong, C.-P. Tseng, Y.-M. Hsueh, H.-Y. Chiou, C.-C. Tseng and C.-J. Chen, *Toxicol. Lett.*, 2003, **137**, 15–21.

24 Y. M. Hsueh, W. L. Wu, Y. L. Huang, H. Y. Chiou, C. H. Tseng and C. J. Chen, *Atherosclerosis*, 1998, **141**, 249–257.

25 J. C. States, S. Srivastava, Y. Chen and A. Barchowsky, *Toxicol. Sci.*, 2009, **107**, 312–323.

26 A. H. Smith, C. Hopenhayn-Rich, M. N. Bates, H. M. Goeden, I. Hertz-Pannier, H. M. Duggan, R. Wood, M. J. Kosnett and M. T. Smith, *Environ. Health Perspect.*, 1992, **97**, 259–267.

27 R. Philipp, *Rev. Environ. Health*, 1985, **5**, 27–57.

28 P. B. Tchounwou, J. A. Centeno and A. K. Patlolla, *Mol. Cell. Biochem.*, 2004, **255**, 47–55.

29 M. L. Kile and D. C. Christiani, *JAMA*, 2008, **300**, 845–846.

30 C. F. Huang, Y. W. Chen, C. Y. Yang, K. S. Tsai, R. Sen Yang and S. H. Liu, *Kaohsiung J. Med. Sci.*, 2011, **27**, 402–410.

31 C.-H. Tseng, *Toxicol. Appl. Pharmacol.*, 2004, **197**, 67–83.

32 H. Sińczuk-Walczak, *Med. Pr.*, 2009, **60**, 519–522.

33 S. Kim, A. Takeuchi, Y. Kawasumi, Y. Endo, H. Lee and Y. Kim, *J. Occup. Health*, 2012, **54**, 344–347.

34 Y. Chen, F. Parvez, M. Gamble, T. Islam, A. Ahmed, M. Argos, J. H. Graziano and H. Ahsan, *Toxicol. Appl. Pharmacol.*, 2009, **239**, 184–192.

35 B. P. Chattopadhyay, A. K. Mukherjee, P. K. Gangopadhyay, J. Alam and A. Roychowdhury, *J. Environ. Sci. Eng.*, 2010, **52**, 147–154.

36 F. Parvez, Y. Chen, M. Yunus, C. Olopade, S. Segers, V. Slavkovich, M. Argos, R. Hasan, A. Ahmed, T. Islam, M. M. Akter, J. H. Graziano and H. Ahsan, *Am. J. Respir. Crit. Care Med.*, 2013, **188**, 813–819.

37 P. Bhattacharjee, N. Das, D. Chatterjee, A. Banerjee, J. K. Das, S. Basu, S. Banerjee, P. Majumder, P. Goswami and A. K. Giri, *Mutat. Res.*, 2013, **755**, 1–5.

38 J. Nriagu, T.-S. Lin, D. G. Mazumder and D. Chatterjee, *Sci. Total Environ.*, 2012, **420**, 65–72.

39 X. Zhou, X. Sun, C. Mobarak, A. J. Gandolfi, S. W. Burchiel, L. G. Hudson and K. J. Liu, *Chem. Res. Toxicol.*, 2014, **27**, 690–698.

40 A. Hussam and A. K. M. Munir, *J. Environ. Sci. Health, Part A*, 2007, **42**, 1869–1878.

41 L. C. Roberts, S. J. Hug, T. Ruettimann, M. M. Billah, A. W. Khan and M. T. Rahman, *Environ. Sci. Technol.*, 2004, **38**, 307–315.

42 J. Ahmad, B. Goldar and S. Misra, *J. Water Health*, 2006, **4**, 463–477.

43 M. Kim, H.-J. Um, S. Bang, S.-H. Lee, S.-J. Oh, J.-H. Han, K.-W. Kim, J. Min and Y.-H. Kim, *Environ. Sci. Technol.*, 2009, **43**, 9335–9340.

44 K. A. Denessiouk, M. S. Johnson and A. I. Denesuk, *J. Mol. Biol.*, 2005, **345**, 611–629.

45 T. Sheet and R. Banerjee, *J. Struct. Biol.*, 2010, **171**, 345–352.

46 T. Sheet, S. Supakar and R. Banerjee, *PLoS One*, 2013, **8**, e57366.

47 T. Sheet and R. Banerjee, *RSC Adv.*, 2016, **6**, 54129–54141.

48 T. Sheet, S. Ghosh, D. Pal and R. Banerjee, *Pept. Sci.*, 2017, **108**, e22921.

49 H. Berman, K. Henrick, H. Nakamura and J. L. Markley, *Nucleic Acids Res.*, 2007, **35**, D301–D303.

50 S.-Y. Kim, T. Hakoshima and K. Kitano, *Sci. Rep.*, 2013, **3**, 1–10.

51 H. Cao, J. Hall and R. Hille, *J. Am. Chem. Soc.*, 2011, **133**, 12414–12417.

52 R. Dennington, T. Keith and J. Millam, *GaussView* 5, 2009.

53 I. Tubert-Brohman, C. R. W. Guimaraes and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2005, **1**, 817–823.

54 R. Banerjee and G. Basu, *ChemBioChem*, 2002, **3**, 1263–1266.

55 R. Banerjee, S. Chattopadhyay and G. Basu, *Proteins*, 2009, **76**, 184–200.

56 G. V. Louie and G. D. Brayer, *J. Mol. Biol.*, 1990, **214**, 527–555.

57 Accelrys Software Inc., 2013.

58 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785–2791.

59 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J. Comput. Chem.*, 1998, **19**, 1639–1662.

60 R. Huey, G. M. Morris, A. J. Olson and D. S. Goodsell, *J. Comput. Chem.*, 2007, **28**, 1145–1152.

61 D. S. Goodsell, G. M. Morris and A. J. Olson, *J. Mol. Recognit.*, 1996, **9**, 1–5.

62 P. A. Holt, J. B. Chaires and J. O. Trent, *J. Chem. Inf. Model.*, 2008, **48**, 1602–1615.

63 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, 2009.

64 I. Tubert-Brohman, A. Cristiano Ruch Werneck Guimaraes and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2005, **1**, 817–823.

65 L. L. C. Schrödinger, *The {PyMOL} Molecular Graphics System, Version~1.8*, 2015.



66 M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS and D. Team, 2016.

67 A. D. MacKerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

68 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

69 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728.

70 Y. Xu, B. Ma and R. Nussinov, *J. Phys. Chem. B*, 2012, **116**, 4801–4811.

71 H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.

72 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

73 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

74 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

75 J. T. Yang, C.-S. C. Wu and H. M. Martinez, *Methods Enzymol.*, 1986, **130**, 208–269.

76 C. Toniolo, A. Polese, F. Formaggio, M. Crisma and J. Kamphuis, *J. Am. Chem. Soc.*, 1996, **118**, 2744–2745.

77 M. C. Manning and R. W. Woody, *Biopolymers*, 1991, **31**, 569–586.

78 R. Banerjee and T. Sheet, *Proteins: Struct., Funct., Bioinf.*, 2017, **85**, 1975–1982.

79 N. E. Grossohme, A. M. Spuches and D. E. Wilcox, *JBIC, J. Biol. Inorg. Chem.*, 2010, **15**, 1183–1191.

80 C. F. Quinn, M. C. Carpenter, M. L. Croteau and D. E. Wilcox, in *Methods in enzymology*, 2016, vol. 567, pp. 3–21.

81 Y. Zhang, S. Akilesh and D. E. Wilcox, *Inorg. Chem.*, 2000, **39**, 3057–3064.

82 F. Zidane, A. Matéos, C. Cakir-Kiefer, L. Miclo, S. Rahuel-Clermont, J.-M. Girardet and C. Corbier, *Food Chem.*, 2012, **132**, 391–398.

83 M. M. Pierce, C. S. Raman and B. T. Nall, *Methods*, 1999, **19**, 213–221.

84 R. Perozzo, G. Folkers and L. Scapozza, *J. Recept. Signal Transduction Res.*, 2004, **24**, 1–52.

85 L. Damian, in *Methods in molecular biology* (Clifton, N.J.), 2013, vol. 1008, pp. 103–118.

86 M. Marquart, J. Walter, J. Deisenhofer, W. Bode, R. Huber and IUCr, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1983, **39**, 480–490.

87 A. Via, F. Ferrè, B. Brannetti, A. Valencia and M. Helmer-Citterich, *J. Mol. Biol.*, 2000, **303**, 455–465.

88 P. Chakrabarti, *J. Mol. Biol.*, 1993, **234**, 463–482.

89 D. Markovich, *Physiol. Rev.*, 2001, **81**, 1499–1533.

90 M. V. S. Kumar and R. Swaminathan, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 900–916.

