J Mol Model (2011) 17:3183–3193
DOI 10.1007/s00894-011-1007-6

ORIGINAL PAPER

# Exploration of the binding mode of α/β-type small acid soluble proteins (SASPs) with DNA

**Yan Ge · Jiayan Wu · Jingfa Xiao · Jun Yu**

Received: 28 October 2010 / Accepted: 1 February 2011 / Published online: 2 March 2011


**Abstract** The α/β-type small acid soluble proteins (SASPs) are a major factor in protecting the spores from being killed in *bacteria*. In this article, we perform a systematic phylogenetic analysis of the α/β-type SASP in the genus of *Geobacillus*, which indicates that the whole family can be divided into three groups. We choose one protein from each group as a representative and construct the tertiary structure of these proteins. In order to explore the mechanism of protecting DNA from damage, 15 ns molecular dynamics simulation for the four complexes of Gsy3 with DNA are performed. The sequence alignment, model structure and binding energy analysis indicate that the helix2 region of SASPs is more conserved and plays a more crucial role in protecting DNA. Pairwise decomposition of residue interaction energies calculation demonstrate that amino acids of Asn10, Lys24, Asn49, Ile52, Ile56, Thr57, Lys58, Arg59 and Val61 take major effect in the binding interaction. The differences of energy contribution of the amino acids between different complexes make us conclude that the protein structure conformation has a slight change upon more proteins binding to DNA and consequently there occur protein-protein cooperation interactions.

**Keywords** α/β-type small acid soluble proteins · Binding energy · DNA-binding protein · Molecular dynamics · Phylogenetic analysis

## Introduction

The α/β-type small acid soluble protein is one of the major factors in protecting DNA of *bacteria* from damage to desiccation, heat, toxic chemicals, enzymes and UV radiation by saturation with DNA. The α/β-type SASP belong to nonspecific DNA-binding proteins, which have one conserved helix-turn-helix DNA binding domain [1, 2]. Spores of *bacteria*, especially for *bacillus* and *clostridium* contain a number of SASPs, comprising 10–20% of the total spore proteins [3]. Experimental analyses indicate that the binding of the α/β-type SASPs variants to DNA can prevent attack on DNA bases by water and thus may be able to protect chromosomes from endogenous and exogenous reactive chemicals by excluding them from the vicinity of DNA [4]. In addition, due to its non-sequence-specific binding with DNA, it can be designed as a new antibacterial agent [5].

The crystal structure of the engineered α/β-type SASP-DNA complex reveals that three molecules of the α/β-type SASP interact cooperatively with DNA through minor groove contacts along the right-hand direction of the DNA helix. The DNA conformation changes from B type to A-B one along with the protein binding [1]. Besides, 35-40% of the protein and DNA surfaces are buried in the complex which indicates that there are many specific interactions

Yan Ge and Jiayan Wu contributed equally to this work

Y. Ge
College of Biological Sciences,
China Agricultural University,
Beijing, China

Y. Ge · J. Wu · J. Xiao (✉) · J. Yu (✉)
CAS Key Laboratory of Genome Sciences and Information,
Beijing Institute of Genomics, Chinese Academy of Sciences,
Beijing, China
e-mail: xiaojingfa@big.ac.cn

J. Yu
e-mail: junyu@big.ac.cn

specially the conserved hydrogen bond interactions. Meanwhile, it is known that the α/β-type SASPs prefer to bind to GC-rich DNA than AT-rich one, in particular to polydG–polydC [6–10]. In addition, salt concentration and concentration of pyridine-2,6-dicarboxylic acid are also crucial for the longevity of the α/β-type SASP-DNA complex [11].

In this study, we mainly use bioinformatics methods to study the binding mode of the α/β-type SASPs in *Geobacillus* with DNA. The engineered crystal protein is extracted from the genus of *bacillus subtilis*. Both *bacillus* and *geobacillus* (genus) belong to *bacillaceae* (family) and are widespread in nature. The comparison of the α/β-type SASP's binding with DNA between different organisms can help us deeply understand the mechanism of the α/β-type SASPs protecting DNA from damage. First, by systematic phylogenetic analysis, the whole family of *Geobacillus* α/β-type SASPs can be divided into three groups. Second, in order to explore the mechanism of the α/β-type SASPs protecting DNA from damage, 15 ns molecular dynamics simulation for the complexes of one kind of SASPs (Gsy3: ZP_03039201) with DNA are performed. As to Lee *et al*'s report, there are many intermolecular protein-protein interactions between the α/β-type SASPs [1]. Thus, we construct single-protein-DNA complex, two-protein-DNA complex and three-protein-DNA complex to study the detailed interaction profiles and compare the energy contribution of the residues in different complexes. Many experiments have pointed out that protein conformation will have a change upon DNA binding recently [12–14], our study will join the discussion due to the point of energy change of these residues. Above all, this research will be helpful for the further experimental study on the binding mode between DNA and the α/β-type SASPs.

## Methods

### Sequence alignment and phylogenetic analysis of the α/β-type SASPs

The α/β-type SASPs sequences of *geobacillus* are retrieved from NCBI (http://www.ncbi.nlm.nih.gov/) and the accession numbers (Table 1) are used to distinguish these sequences. The sequences alignment of 16 α/β-type SASPs is performed by ClustalW program with default settings [15]. The distance matrix is calculated based on the alignment and the phylogenetic tree of the protein family is generated by PHYLIP software [16]. To evaluate the reliability of the inferred tree, bootstrap analysis is used. The sequences are bootstrapped 1000 times by randomly choosing columns from the original alignment by using the program SEQBOOT [17]. The majority rule consensus tree is created by CONSENSUS [18] and the tree is viewed with MEGA [19].

### Model construction of α/β-type SASPs and DNA complex

The three proteins of Gsy1, Gsy3, and Gsw1 are chosen as the representatives from the three groups according to the phylogenic tree. Two methods are employed to construct the tertiary structure of the α/β-type SASP based on the sequence identity between template and target proteins. The protein structures of Gsy1 and Gsy3 are constructed by Modeller [20] program using the corresponding protein template of the crystal structure (PDB code 2Z3X). The sequence alignment of the α/β-type SASPs and template protein is performed by means of ModellerV8.1 using the Blosum matrix. Models are generated as pdb files using spatial structure restraints on the target sequence derived

**Table 1** The α/β-type SASPs sequences of *geobacillus* (genus) collected from NCBI

| Accession number | Species | Protein abbreviation |
|---|---|---|
| YP_146482 | Geobacillus kaustophilus | Gka1 |
| YP_146787 | Geobacillus kaustophilus | Gka2 |
| YP_148647 | Geobacillus kaustophilus | Gka3 |
| ZP_03146609 | Geobacillus sp. G11MC16 | Gsg |
| ZP_02914444 | Geobacillus sp. WCH70 | Gsw1 |
| ZP_02913997 | Geobacillus sp. WCH70 | Gsw2 |
| ZP_02913229 | Geobacillus sp. WCH70 | Gsw3 |
| ZP_02914110 | Geobacillus sp. WCH70 | Gsw4 |
| ZP_03037710 | Geobacillus sp. Y412MC10 | Gsy1 |
| ZP_03038208 | Geobacillus sp. Y412MC10 | Gsy2 |
| ZP_03039201 | Geobacillus sp. Y412MC10 | Gsy3 |
| ZP_03042334 | Geobacillus sp. Y412MC10 | Gsy4 |
| ZP_03042338 | Geobacillus sp. Y412MC10 | Gsy5 |
| YP_001124668 | Geobacillus thermodenitrificans | Gth1 |
| YP_001124944 | Geobacillus thermodenitrificans | Gth1 |
| YP_001126787 | Geobacillus thermodenitrificans | Gth1 |

from its alignment with the template structure. Initially, 50 models are obtained for the α/β-type SASP conformers using ModellerV8.1. Since Gsw1 has low sequence identity with the template protein, homology modeling is not applicable for the model building of this protein. The I-TASSER [21] online web server is adopted for predicting protein structure. The quality of the candidate models is evaluated using procheck [22], Dope [23] and Ga341 [24] methods. The model that ranks the highest under these evaluation criteria is used in subsequent MD simulations. The refinement of the model is obtained by energy minimization: 5000 iterations of steepest descent (SD) calculation are carried out and then the conjugated gradient (CG) calculation is performed until the convergence on the gradient reaches 0.1kcal/(Å mol) with AMBER force field. After the above simulations, Amber software package [25] is used to perform molecular dynamics (MD) calculation. The temperature of the simulation is 300K and explicit solvent model TIP3P water is used to solve the protein with a 12 Å water cap. Finally, a conjugate gradient energy minimization of full protein is performed until the root mean-square (rms) gradient energy is lower than 0.001 kcal/(Å mol). After finishing all the above steps, the quality of the initial protein model is improved.

The next step is to construct the initial complex structure of the protein model with DNA for further molecular dynamics analysis. The initial structure is obtained by transferring the corresponding protein in the published structure of the complex (PDB code 2Z3X) [9] to the modeled protein Gsy3 (ZP_03039201) by pair-wise sequence alignment. Based on this initial structure, the single-protein-DNA complex, two-protein-DNA complex, and three-protein-DNA complex are also produced in order to make clear the difference between one protein and more proteins in DNA binding. To discriminate the position of the three proteins, they are named A, B and C, respectively. Four complexes systems including A-DNA complex, C-DNA complex, AC-DNA complex and ABC-DNA complex are constructed for further MD study.

### Molecular dynamic (MD) simulation

MD simulations are carried out at the molecular mechanics level using AMBERff99SB force field as implemented in AMBER9 programs [25]. Structures of the complex models are solvated in a truncated octahedral box of TIP3P water extending at least 12 Å in each direction from the solute as the physical state of water in bacterial spores has been made clear [26], and the cut-off distance is kept to 12 Å to compute the nonbonded interactions. Prior to MD simulations, the systems are relaxed by a series of steepest descent (SD) and conjugated gradient (CG) minimizations. MD simulations are performed based on each of the minimized systems by gradually heating over 50 ps from 0 to 300 K with the protein backbone atoms fixed using a force constant of 2 kcal/(mol $Å^2$). Then, 50 ps of density equilibration with the same restraints on the protein backbone are carried out. The following step is constant pressure equilibration at 300 K. All simulations are performed under periodic boundary conditions, and long-range electrostatics is treated by using Particle-Mesh-Ewald method. The time step is set to 2 fs and the trajectory is recorded every 2 ps with shake on hydrogen atoms and langevin dynamics for temperature control.

### Analysis

Root-mean-square deviation (RMSD) of the α/β-type SASP with DNA complex backbone atoms during MD production phase relative to the starting structure is calculated using PTRAJ module of AMBER program suite to assess stability of MD trajectories. MD simulations could make an explicit analysis of hydrogen bonds properties, such as donor-acceptor assignments and hydrogen bonds occupancies. The α/β-type SASP with DNA complex hydrogen bonds are analyzed using PTRAJ module of AMBER. Criteria for identifying hydrogen bonds are as follows: (1) the distance between proton donor (*X*) and acceptor (*Y*) atoms is less than or equal to 3.5Å (2) the angle *X*-H…*Y* is greater than or equal to 120°. Special attention is given to stable hydrogen bonds, those held this mode at least 50% of the simulation period [27].

### Binding free energy calculation

DNA binding energy is computed using molecular mechanics generalized Born solvent area (MM-GBSA) methodology [28] and molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) of Amber SANDER modules. In this approach, frames of one trajectory are stripped off counter-ions and water molecules. In order to obtain a detailed view of DNA binding, the interaction energies are further decomposed to each the α/β-type SASPs residue and DNA. These decompositions are possible for molecular mechanics and free energy of solvation, not for entropies. The details of the underlying theory are described elsewhere [29].

## Results and discussion

### Sequence alignment and phylogenetic analysis

From the phylogenetic tree, the α/β-type SASPs of *geobacillus* can be divided into three groups (Fig. 1a). By

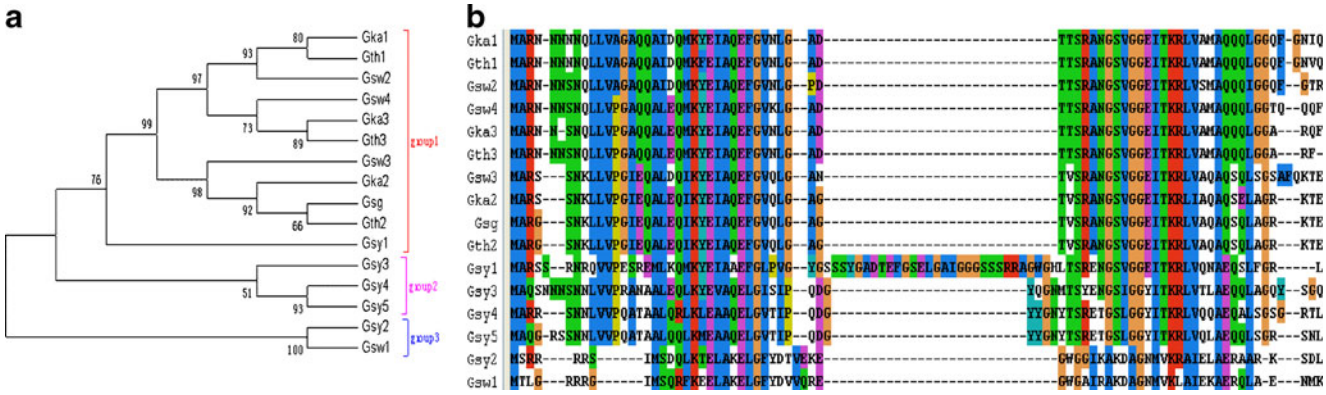


**Fig. 1** (**a**) Phylogenetic relationship among the α/β-type SASP of *geobacillus* (genus). MP bootstrap supports (>50%) are shown on the nodes. The divided three groups are characterized by different colors. (**b**) Sequences alignment of the α/β-type SASPs using ClustalW

further analysis to the sequences alignment result, these three groups are much different from each other (Fig. 1b). The α/β-type SASPs are usually less than 100 amino acids and have two α-helical conserved regions besides a turn region between them. In the phylogenetic tree, group 1 has the extreme similar sequence composition as the reported crystal one [1]. The separate branch, Gsy1 is a special case, which has the same conserved α-helix region as others in group1, but the turn region is very long and completely different. It has been reported that the two helix regions play an important role in the interaction with DNA and the turn region has no regular structure. We predict this long turn may be evolved to a more conserved short turn to keep the two helix regions stable, which can be verified as the majority of the α/β-type SASPs sequences have a small turn region. Group 2 represents another form of the α/β-type SASPs, the turn region of which have the same consensus sequence QDGYQGNYT. The function of these residues can be found out by further MD analysis of the DNA and proteins interactions. Group 3 has the least conserved residues compared with other groups in the whole helix–turn–helix motif region. It can be inferred that those proteins do not work as efficiently as other groups and may appear early in the α/β-type SASPs' lifestyle. With the change of the environment, the α-helix region in group 3 proteins becomes more conserved and keeps stable until so far. The division of the α/β-type SASPs provides us a new understanding to the divergence of the protein sequence compared with the report by Setlow [11]. Generally speaking, there are more conserved amino acids in helix2 than helix1, especially in group1 and 2, which probably indicate helix2 effects more in DNA binding.

## The α/β-type SASPs structure construction

Based on sequence alignment and phylogenetic analysis results (Fig. 1), three proteins including Gsy1, Gsy3 and Gsw1 are chosen as representatives for the structure construction from each group. Gsy1 has the same conserved α-helix region as the crystal one in group1 [1], but the turn region is very long and completely different. Gsy3 is more similar to the crystal protein than Gsw1 and Gsy1. It is a great candidate protein to investigate particularity and universality in their binding to DNA with the crystal one in further research. Since the phylogenetic tree and sequence alignment do not show any distinct difference in those group3 proteins, we randomly select Gsw1 as the third representative. Figure 2b shows the detail residues alignment and the identities between these proteins and the crystal protein (PDB code 2Z3X) are 46%, 37% and 14%, respectively. The protein structures of Gsy1 and Gsy3 are constructed by Modeller while Gsw1 is produced with I-TASSER [21]. From the procheck results (Fig. 2a), we can see the residues located in the most favored regions are 90.8%, 93.7% and 96.2% (all > 90%), indicating the reasonable conformations of the 3D structures. Structure comparisons confirm that all of them belong to the family of HTH motif DNA-binding proteins with the help of SSAP server [30]. The major differences between the three proteins lie in the loop regions, including the N terminal, C terminal coil and the turn region. The two helices are relatively conserved, which are important for DNA binding. Moreover, the helix1 is more changeable compared with the helix2. The RMSDs for the helix2 Ca-atoms of three α/β-type SASPs with the crystal protein's (PDB code 2Z3X) helix2 are 1.15 Å, 0.89 Å, 0.80Å, respectively.

## MD simulations of the α/β-type SASPs and DNA complexes

It has been reported that the α/β-type SASPs protect the DNA of *bacteria* from damage by deeply binding to the target DNA [31]. The comparison of the α/β-type SASP's

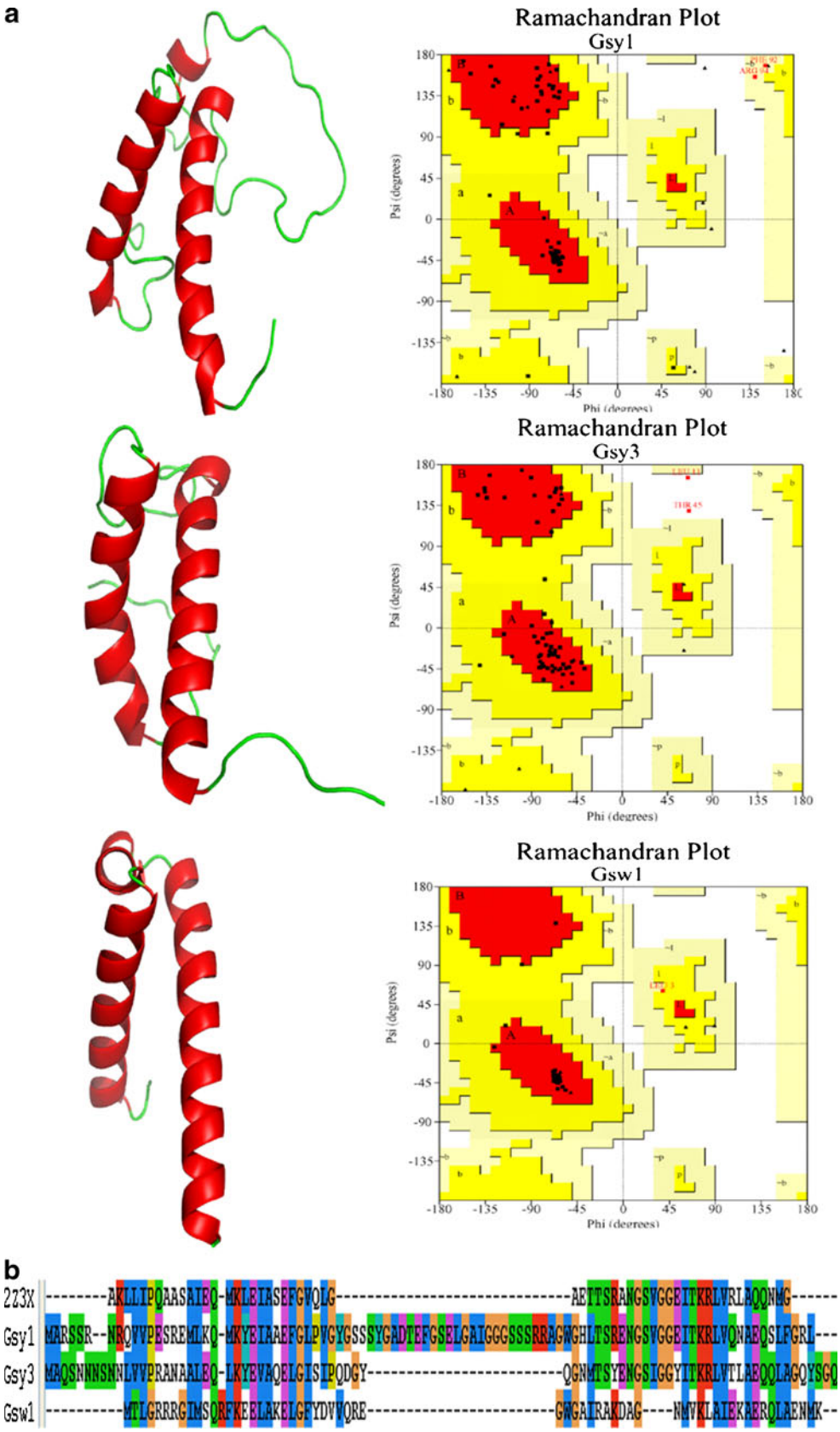


**Fig. 2** (**a**) The model structure of three representative α/β-type SASPs (left) and Ramachandran plot of the model proteins (right). The most favored regions are red, and additional allowed, generously allowed, and disallowed regions are indicated as yellow, light yellow, and white, respectively. (**b**) Sequences alignment between three representatives and the crystal protein (PDB code 2Z3X)

binding with DNA between different organisms can help us understand the mechanism of the α/β-type SASPs in DNA damage process. Gsy3 is more similar to the crystal protein, and they belong to the different groups. So we select Gsy3 as one representative protein for the further research to investigate particularity and universality in their binding to DNA with the crystal one. Consequently, 15 ns molecular dynamics for the complexes of Gsy3 with DNA, which include two single-protein-DNA complexes (A-DNA and C-DNA), one two-protein-DNA complex (AC-DNA) and one three-protein-DNA complex (ABC-DNA, Fig. 3) are performed. Classic HTH motif of DNA binding protein family is normally bound in the major groove of DNA [2], while the α/β-type SASPs show HTH motif mainly interacting with the minor groove of DNA [32]. In particular, helix1 of our constructed structure lies on the edge of the minor groove while helix2 is located in the minor groove along the right-hand direction of the DNA helix.

*Stability of MD trajectories*

In this study, we simulate as long as 15 ns to insure the equilibration phase long enough for further calculation. Root-mean-squared deviations (RMSDs) values for the backbone atoms of the complex relative to the starting structure of the heating phase are plotted to assess their stability (Fig. 4). These data show that the trajectories are stable after simulating about 5 ns. Snapshots belonging to 5 ns to 15 ns are extracted for the further hydrogen bond interactions and binding energy analysis.

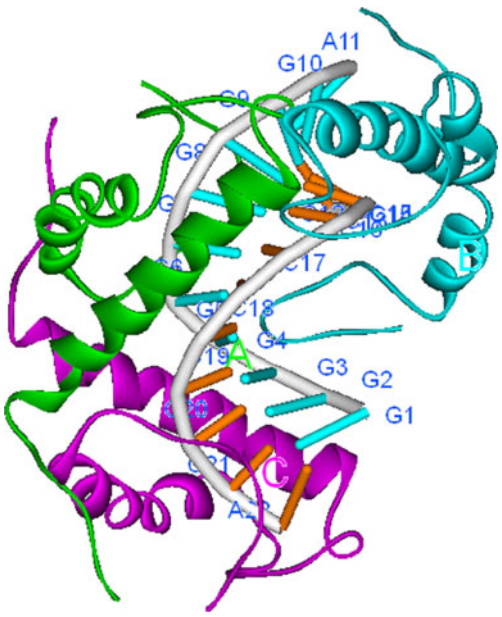


**Fig. 3** The modeled complex structure composed of three molecules of Gsy3 and 10 base pair DNA. Green color: A; Magenta: C; Cyan: B. Helix1 is located in residues 14 to 31 while helix2 is in residues 46–68

*Interactions of the α/β-type SASPs with DNA*

It has been reported that protein B assumes a dimeric arrangement by interacting with another protein molecule from a crystallographic symmetry related complex [1]. Thus, we pay more attention to the interactions between protein A, protein C and DNA. Tables 2 and 3 and Fig. 5 have shown the hydrogen bonding interactions between protein A, C and DNA in the ABC-DNA complex. The hydrogen bonds profile in our system is also compared with the crystal complex [1].

In protein A of ABC-DNA complex, the amino acids in N terminal regions, specifically Asn6, Asn7, Ser8 and Asn10, all form hydrogen bonds with the DNA, but in A-DNA and AC-DNA complexes, these hydrogen bonds do not exist (data not shown), which indicates the stable N terminal conformation in the final structure. Interestingly, the conserved Lys24 is present in A-DNA and C-DNA complexes by interacting with the base G8, the same form as the crystal complex, but it is absent in ABC-DNA complex. The reasonable explanation could be that the crystal protein belongs to group1 in the phylogenetic tree (Fig. 1), while our studied protein (Gsy3) is located in group2 of the α/β-type SASPs family.

In protein C of ABC-DNA complex, there are not as many N terminal forming hydrogen bonds amino acids as in protein A except for Ser8 which forms three hydrogen bonds with three different DNA bases. However, in C-DNA and AC-DNA complexes, these hydrogen bonds do not exist, which is the same case for Gln41. This may suggest that the conformation of protein C has a slight change in the final complex. In contrast to protein A in ABC-DNA complex, the side chain NH group of the conserved Lys24 forms the hydrogen bond with the phosphate O of base A22 in this complex, which is in accordance with the crystal structure. Besides, the conserved Asn49, Gly50 and Thr57 here also have the same behavior like Lys24 forming consistent hydrogen bonds in both the crystal complex and our constructed complex (ABC-DNA) [1].

In addition to the interactions between protein and DNA, the interactions between protein A and C also play an important role in determining the orientation of the two proteins and further affect the stability of the complex. Arg15 in protein A forms two hydrogen bonds with protein C, one of the hydrogen bonds is formed between its side chain NH group and the backbone O of Gln71 in protein C, the other is between its backbone O and the side chain NH group of Gln66. Amino acid Glu65 in helix2 of protein A forms a hydrogen bond with the side chain NH group of Gln67 in protein C, which is also located in helix2. The last

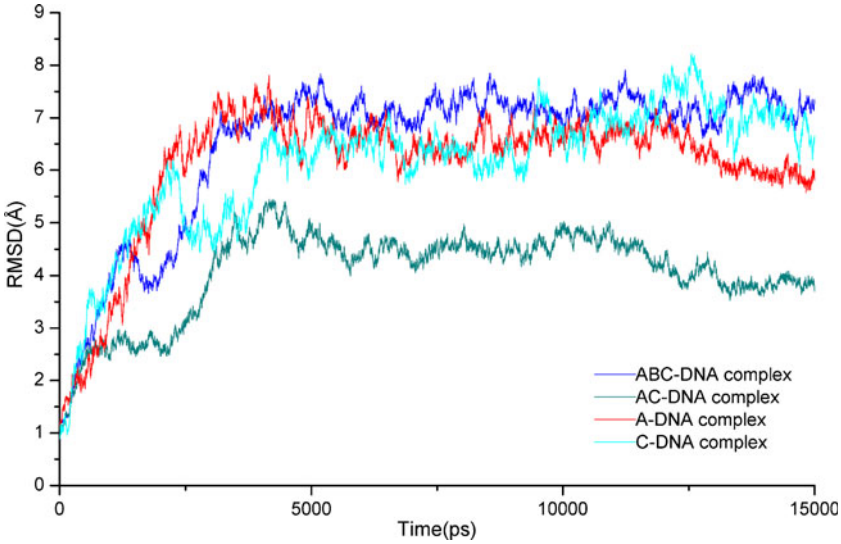


**Fig. 4** RMSD of the backbone atoms for four complexes, starting from the minimized complex structure

residue of Gln75 in protein A forms a hydrogen bond with the side chain group of Asn7 in protein C. These hydrogen bonds altogether facilitate the dimerization of the two SASPs.

To summarize briefly, the above hydrogen bonds network must have played a crucial role in keeping the whole structure stable which is similar with Lee *et al*'s report. By comparing the hydrogen bond profile of our result (Fig. 5) with theirs, it is easy to find that some conserved hydrogen bonds such as the above talking Lsy24, Gly50, Thr57 are nonspecific while others are specific in the two different organisms. In addition, after

**Table 2** The hydrogen bonds between the residues of protein A, C and the bases of DNA in ABC-DNA complex (the *s* and *b* in bracket means side chain and backbone, respectively)

| Residue | Atom | DNA | Atom | Distance(Å) | Angel(deg) | Occupancy(%) |
|---|---|---|---|---|---|---|
| Protein A | | | | | | |
| Asn6 | O(*b*) | C12 | Cytimidine N4H | 2.941 | 160.08 | 97.13 |
| Asn7 | NH(*s*) | G7 | Guanine O6 | 2.882 | 162.85 | 99.48 |
| Asn7 | O(*s*) | C14 | Cytimidine N4H | 3.029 | 146.59 | 92.62 |
| Ser8 | OH(*s*) | G7 | Guanine O6 | 2.708 | 162.40 | 98.63 |
| Asn10 | NH(*s*) | G7 | Phosphate O | 2.878 | 160.89 | 99.56 |
| Asn49 | O(*s*) | G7 | Guanine N2H | 2.849 | 159.03 | 99.64 |
| Asn49 | NH(*s*) | G7 | Guanine N3 | 3.024 | 151.03 | 98.55 |
| Asn49 | NH(*s*) | G8 | Sugar O4' | 3.157 | 140.27 | 68.85 |
| Gly50 | O(*b*) | G6 | Guanine N2H | 3.219 | 139.60 | 66.83 |
| Thr57 | OH(*s*) | G5 | Guanine N2H | 3.174 | 142.44 | 73.37 |
| Thr57 | OH(*s*) | G5 | Guanine N3 | 2.903 | 159.00 | 63.88 |
| Protein C | | | | | | |
| Ser8 | OH(*s*) | G1 | Guanine O6 | 2.717 | 159.00 | 96.97 |
| Ser8 | O(*s*) | A22 | Adenine N6H | 3.039 | 143.87 | 86.12 |
| Ser8 | O(*b*) | C21 | Cytimidine N4H | 3.041 | 147.31 | 72.84 |
| Lys24 | NH(*s*) | A22 | Phosphate O | 2.784 | 161.26 | 54.48 |
| Gln41 | NH(*s*) | A22 | Adenine N3 | 3.073 | 160.68 | 63.32 |
| Gln41 | NH(*s*) | G1 | Guanine N3 | 3.161 | 152.13 | 57.87 |
| Ser46 | O(*b*) | G1 | Guanine N2H | 3.067 | 155.79 | 62.31 |
| Asn49 | NH(*s*) | A22 | Sugar O4' | 2.912 | 154.48 | 94.11 |
| Asn49 | NH(*s*) | C21 | Cytimidine O2 | 2.958 | 141.31 | 68.68 |
| Gly50 | O(*b*) | G2 | Guanine N2H | 2.984 | 155.57 | 99.27 |
| Thr57 | OH(*s*) | C19 | Cytimidine O2 | 2.818 | 159.86 | 97.94 |

**Table 3** The hydrogen bonds between the residues of protein A and C in ABC-DNA complex

| Protein A | Atom | Protein C | Atom | Distanc(Å) | Angel(deg) | Occupancy(%) |
|---|---|---|---|---|---|---|
| Arg15 | NH(*s*) | Gln71 | O(*B*) | 2.944 | 148.30 | 52.58 |
| Arg15 | O(*b*) | Gln66 | NH(*S*) | 3.049 | 156.79 | 93.18 |
| Glu65 | OE2(*s*) | Gln67 | NH(*S*) | 2.912 | 163.07 | 94.11 |
| Gln75 | O(*s*) | Asn7 | NH(*S*) | 2.943 | 159.42 | 89.14 |

carefully examining the interacting profile, it is found that helix2's residues contribute more to hydrogen bonding interactions than helix1's.

*Binding energies*

In order to evaluate the binding activity between protein A, C and DNA, MM-PBSA analysis is performed for the four systems including A-DNA complex, C-DNA complex, AC-DNA complex and ABC-DNA complex. The interaction energies correspond only to the enthalpy contributions of free energy of binding. Since these model structures are homology modeling structures, the absolute values of these energies may vary. However, even with these qualifications, the relative importance of each protein or residue can be inferred by its rank order of interaction energy [33, 34]. From Table 4, we can find that the binding energy of protein A become more negative when more proteins bind to DNA, especially when protein C binds forming AC-DNA complex, the absolute value of the total binding free energy calculated by the MM-PBSA method (PBTOT)

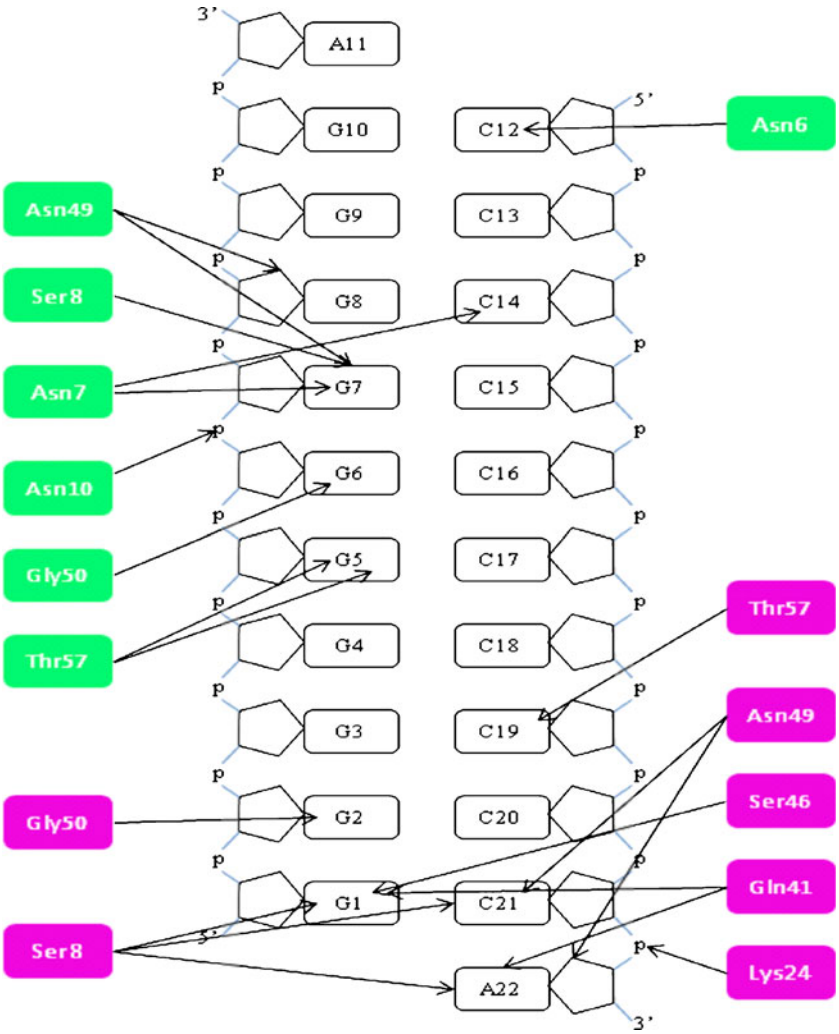


**Fig. 5** Summary of the hydrogen bonding interactions between the α/β-type SASPs and DNA (indicated by arrows). Magenta: protein A; Cyan: protein C

**Table 4** Binding free energies (kcal $mol^{-1}$) between the α/β-type SASPs and DNA in the complexes

| Method | Contribution | Complex | | | | | |
|---|---|---|---|---|---|---|---|
| | | A-DNA A binding | AC-DNA | ABC-DNA | C-DNA C Binding | AC-DNA | ABC-DNA |
| MM | ELE | -10.42 | -214.17 | -188.05 | -85.51 | -8.03 | -36.93 |
| | VDW | -119.43 | -217.37 | -239.03 | -105.17 | -216.11 | -212.00 |
| | GAS | -129.84 | -431.54 | -427.08 | -190.67 | -224.14 | -248.93 |
| PBSA | PBSUR | -13.52 | -28.44 | -32.55 | -13.40 | -29.20 | -28.31 |
| | PBCAL | 59.07 | 315.30 | 298.50 | 139.71 | 114.75 | 118.28 |
| | PBSOL | 45.55 | 286.86 | 265.94 | 126.31 | 85.55 | 89.98 |
| | PBELE | 48.66 | 101.13 | 110.44 | 54.21 | 106.72 | 81.36 |
| | PBTOT | -84.29 | -144.67 | -161.14 | -64.36 | -138.59 | -158.95 |
| GBSA | GBSUR | -13.52 | -28.44 | -32.55 | -13.40 | -29.20 | -28.31 |
| | GB | 58.22 | 311.72 | 288.75 | 136.06 | 114.03 | 127.36 |
| | GBSOL | 44.70 | 283.28 | 256.20 | 122.66 | 84.84 | 99.05 |
| | GBELE | 47.81 | 97.55 | 100.70 | 50.55 | 106.00 | 90.43 |
| | GBTOT | -85.14 | -148.26 | -170.88 | -68.01 | -139.30 | -149.88 |

ELE, electrostatic interactions; VDW, van der Waals interactions between the fragments; GAS, addition ELE+VDW+INT being the binding enthalpic contributions in vacuum; PBSUR, nonpolar contribution to solvation; PBCAL, polar contribution of solvation; PBSOL, the PBSUR+PBCAL; PBELE, PBCAL+ELE addition; PBTOT, total binding free energy calculated by MM-PBSA method

has a great increase. It can be deduced that there are cooperative interactions between protein C and A to make the whole binding more effective, which can also be proved by the change of the binding energy of protein C from C-DNA complex to AC-DNA complex. Above all, the major contribution to the binding free energy is from van der Waals contribution (VDW) in these complexes.

*Decomposition of energy on each amino acid residues*

To determine the key residues in DNA binding, the interaction energies of each individual amino acid of protein A and C with DNA in single-protein-DNA complex, two-protein-DNA complex and three-protein-DNA complex are calculated. Figure 6 shows the relative decomposed energies with the absolute relative energy larger than 1 kcal $mol^{-1}$ versus amino acids of protein A and C in their corresponding three complexes. Here, the different contributions of every residue between these complexes are also compared. Figure 6a shows that Asn10, Lys24, Asn49, Ile52, Ile56, Thr57, Lys58, Arg59 and Val61 in protein A have lower decomposition energies compared with other residues, which affect more in DNA binding in three A-protein containing complexes. There are clear differences for the amino acids of Asn5, Asn7, Leu20, Gln41, Ser46, Tyr47, Asn49, Tyr62, Gln65 and Gln66 between these three complexes, which definitely mean that the conformations of these residues have a change in transformation to reach the final stable conformations. This is in accordance with the previous report that minor groove-binding architectural proteins change the conformation of the DNA [35]. Figure 6b shows a similar result for protein C in all three C-protein containing complexes. Amino acids of Met1, Asn10, Arg15, Lys24, Tyr47, Asn49, Ile52, Ile56, Thr57, Lys58, Arg59 and Val61 contribute much for protein C binding in three complexes. It also can be concluded that Ala2, Asn5, Leu11, Val12, Arg15 and Gln41 have different conformations between these three complexes.

Taking these two figures together, we can see that Asn10, Lys24 and Lys58 in protein A and C all have significant negative value in any complex (except for Asn10 in C-complex) reflecting their essential effect in DNA binding of these residues in two proteins. The tight binding according with the extreme lower decomposed energy is partly derived from the intensive hydrogen bond interactions, such as Asn10 in protein A and Lys24 in protein C. These polar residues of Asn10, Lys24 and Lys58 are also mainly through electrostatic interactions with DNA.

From the sequence alignment and the structure produced by homology modeling, the two conserved helices in our model protein correspond to two regions, one is Pro14 to Leu31, and the other is Ser46 to Leu68. The energy contribution and H-bond analysis indicate that helix2 plays a significant role in promoting the α/β-type SASPs binding to DNA, which agrees with sequence alignment and phylogenetic analysis results.

*The mechanism of protection for DNA*

The α/β-type SASPs belong to nonspecific DNA-binding proteins. It is one of the major factors in protecting DNA from damage to desiccation, heat, toxic chemicals, enzymes and UV radiation by saturation with DNA. Previous study points out a set of recognition rules existing in DNA and protein binding interface no matter the protein is specific or nonspecific, and the interface conservation follows trends

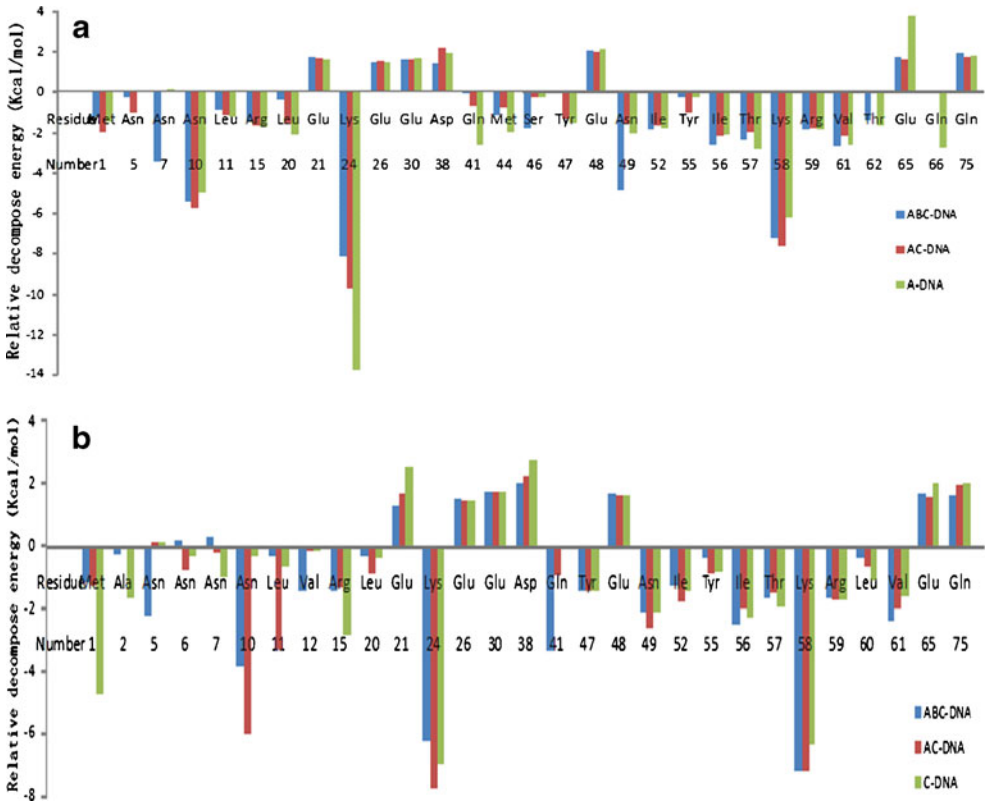


**Fig. 6** List of the decomposed energies of the amino acids contributing absolutely larger than 1 kcal $mol^{-1}$ in protein A and C. (**a**) The energy distribution of protein A residues in three containing A complexes. (**b**) The energy distribution of protein C residues in three containing C complexes

that are superfamily-specific [36]. In our system, the change of accessible surface areas (ASA) of DNA in monomer is from 4296Å$^2$ to 915 Å$^2$ (Fig. 7) after interacting with those three α/β-type SASPs, which indicates the efficient protection by the α/β-type SASPs to keep the DNA away from been attacked by any agents approaching the target DNA. According to systematic phylogenetic analysis, the whole family of *geobacillus* can be divided into three groups. MD results of one α/β-type SASPs complexes with DNA show that helix2 is more conserved compared with helix1 and takes more charge of the interaction with DNA.

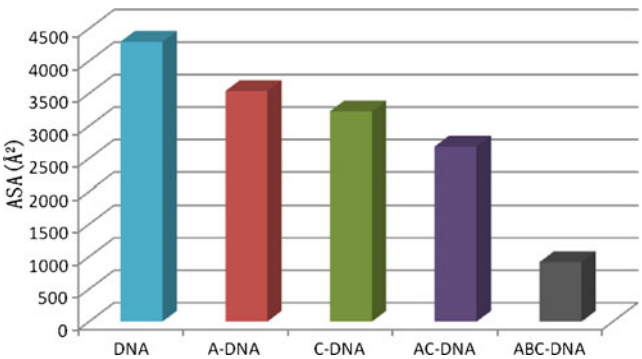


**Fig. 7** The accessible surface areas (ASA) of DNA in four complexes systems. The left column DNA means the amount of area without proteins

## Conclusions

The systematic phylogenetic analysis of the *geobacillus* α/β-type SASPs indicates these proteins can be divided into three groups. The α/β-type SASPs contain a helix–turn–helix motif, the sequence alignment and model structure show that the second helix is more conserved in the protein evolution, which suggests this helix may play a more essential role in DNA binding. In order to explore the mechanism of protection of DNA from damage, 15 ns molecular dynamics simulation for the complexes of one α/β-type SASPs (Gsy3:ZP_03039201) and DNA are performed. MM-PBSA and pairwise decomposition energies are calculated from post analysis of molecular dynamics structures to determine the key residues involved in the interactions of DNA and the α/β-type SASPs. Several amino acids, Asn10, Lys24, Asn49, Ile52, Ile56, Thr57, Lys58, Arg59 and Val61 are defined as their major contributions to the binding efficiency. The differences between one protein-, two protein-, and three protein-binding to DNA are investigated and the same residues taking different effects are found, which indicates the conformations of the proteins have a slight change in the final complex. Integrating the results from sequences alignment, constructed proteins' 3D structure and energy contributions, we find that helix2 of the α/β-type SASPs is more conserved than helix1 and generally plays a more important role in protection of DNA.

**Acknowledgments** This study was supported by a grant from the National Science Foundation of China (31071163), grant from the National Basic Research Program (973 Program) (2010CB126604) and the Special Foundation Work Program (2009FY120100) of the Ministry of Science and Technology of the People's Republic of China. We also thank Supercomputing Center of Computational Network Information Center in Chinese Academy of Sciences for computer time and support.

## References

1. Lee KS, Bumbaca D, Kosman J, Setlow P, Jedrzejas MJ (2008) Structure of a protein-DNA complex essential for DNA protection in spores of Bacillus species. Proc Natl Acad Sci USA 105:2806–2811
2. Brennan RG, Matthews BW (1989) The helix-turn-helix DNA binding motif. J Biol Chem 264:1903–1906
3. Fairhead H, Setlow B, Setlow P (1993) Prevention of DNA damage in spores and in vitro by small, acid-soluble proteins from Bacillus species. J Bacteriol 175:1367–1374
4. Sohail A, Hayes CS, Divvela P, Setlow P, Bhagwat AS (2002) Protection of DNA by alpha/beta-type small, acid-soluble proteins from Bacillus subtilis spores against cytosine deamination. Biochemistry 41:11325–11330
5. Fairhead H (2009) SASP gene delivery: a novel antibacterial approach. Drug News Perspect 22:197–203
6. Setlow P (1994) Mechanisms which contribute to the long-term survival of spores of Bacillus species. J Appl Bacteriol 76 (Symposium Supplement):49S–60S
7. Setlow P (1995) Mechanisms for the prevention of damage to the DNA in spores of Bacillus species. Annu Rev Microbiol 49:29–54
8. Hayes CS et al (2000) Equilibrium and kinetic binding interactions between DNA and a group of novel, non-specific DNA binding proteins from spores of bacillus and clostridium species. J Biol Chem 275:35040–35050
9. Hayes CS et al (2001) N-terminal amino acid residues mediate protein-protein interactions between DNA-bound a/b-type small, acid-soluble spore proteins from Bacillus spores. J Biol Chem 276:2267–2275
10. Kosman J, Setlow P (2003) Effects of carboxyl-terminal modifications and pH on the binding of a Bacillus subtilis small, acid-soluble spore protein to DNA. J Bacteriol 185:6095–6103
11. Setlow P (2007) I will survive: DNA protection in bacterial spores. Trends Microbiol 15:172–180
12. Neira J, Roman-Trufero M, Contreras L, Prieto J, Singh G et al (2009) The transcriptional repressor RYBP Is a natively unfolded protein which folds upon binding to DNA. Biochemistry 48:1348–1360
13. Yamane T, Okamura H, Nishimura Y, Kidera A, Ikeguchi M (2010) Side-chain conformational changes of transcription factor PhoB upon DNA binding: a population-shift mechanism. J Am Chem Soc 132:12653–12659
14. Aguado-Llera D, Goormaghtigh E, de Geest N, Quan X, Prieto A et al (2010) The basic helix- loop- helix region of human neurogenin 1 is a monomeric natively unfolded protein which forms a "fuzzy" complex upon DNA binding. Biochemistry 49:1577–1589
15. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace I, Wilm A, Lopez R (2007) Clustal W and clustal X version 2.0. Bioinformatics 23:2947–2948
16. Felsenstein J (1989) PHYLIP-phylogeny inference package (version 3.2). Cladistics 5:164–166
17. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
18. Margush T, McMorris F (1981) Consensusn-trees. Bull Math Biol 43:239–244
19. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9:299–306
20. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2006) Comparative protein structure modeling using Modeller. Curr Protoc Bioinf 50:2.9.1-2.9.31
21. Zhang Y (2008) I-TASSER server for protein 3 D structure prediction. BMC Bioinform 9:40–40
22. Laskowski R, MacArthur M, Moss D, Thornton J (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
23. Shen M, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15:2507–2524
24. Melo F, Sánchez R, Sali A (2002) Statistical potentials for fold assessment. Protein Sci 11:430–448
25. Case D, Cheatham T III, Darden T, Gohlke H, Luo R, Merz K Jr, Onufriev A, Simmerling C, Wang B, Woods R (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688
26. Sunde EP, Setlow P, Hederstedt L, Halle B (2009) The physical state of water in bacterial spores. Proc Natl Acad Sci USA 106:19334–19339
27. Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 44:97–179
28. Kollman P, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33:889–897
29. Gohlke H, Kiel C, Case D (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. J Mol Biol 330:891–913
30. Orengo C, Michie A, Jones S, Jones D, Swindells M, Thornton J (1997) CATH-a hierarchic classification of protein domain structures. Structure 5:1093–1109
31. Moeller R, Setlow P, Horneck G, Berger T, Reitz G, Rettberg P, Doherty AJ, Okayasu R, Nicholson WL (2008) Roles of the major, small, acid-soluble spore proteins and spore-specific and universal DNA repair mechanisms in resistance of Bacillus subtilis spores to ionizing radiation from X rays and high-energy charged-particle bombardment. J Bacteriol 190:1134–1140
32. Tubbs JL, Pegg AE, Tainer JA (2007) DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. DNA Repair (Amst) 6:1100–1115
33. Payne V, Chang Y, Loew G (1999) Homology modeling and substrate binding study of human CYP2C9 enzyme. Proteins Struct Funct Bioinf 37:176–190
34. Chang Y, Loew G (1999) Homology modeling and substrate binding study of human CYP4A11 enzyme. Proteins Struct Funct Bioinf 34:403–415
35. Bewley C, Gronenborn A, Clore G (1998) Minor groove-binding architectural proteins: structure, function, and DNA recognition 1. Annu Rev Bioph Biom 27:105–131
36. Contreras-Moreira B, Sancho J, Angarica V (2010) Comparison of DNA binding across protein superfamilies. Proteins Struct Funct Bioinf 78:52–62