

# In-Silico Structural Analysis of SoxF Protein Through Molecular Modelling and Protein-Protein Docking from *Hydrogenophilus thermoluteolus*: An Approach to Understand the Molecular Mechanism of Thiosulfate Oxidation

Sujay Ray and Angshuman Bagchi

**Abstract** Microbial redox reactions of inorganic sulphur compounds play vital role for recycling of sulphur to maintain environmental sulphur balance. These important reactions are carried out by enzyme system encoded by *sox operon*. Central player of sulphur oxidation process is SoxY–Z protein complex. Thermophilic beta-proteobacterium-*Hydrogenophilus thermoluteolus*, oxidizes sulphur compounds including thiosulfate with the help of proteins encoded by *sox operon*. Protein SoxF having sulfide dehydrogenase activity has the ability to reactivate the inactivated SoxY–Z complex. Till date no structural details are available for SoxF protein from *H. thermoluteolus*. In present work, homology modeling has been used to build 3D structures of SoxY, SoxZ and SoxF of *H. thermoluteolus*. 3D structure of SoxY–Z–F complex was obtained by ClusPro2.0. Amino acid residues responsible for protein-protein interaction were identified. Interactions in SoxY–Z–F were found to be mediated through hydrogen bonding. Probable biophysical mechanism of the interactions of SoxF with SoxY–Z complex has been identified.

**Keywords** Sulphur oxidation · Homology modelling · Sox operon · Protein-protein interaction · Molecular docking

---

S. Ray · A. Bagchi (✉)

Department of Biochemistry and Biophysics, University of Kalyani, Kalyani, Nadia, India  
e-mail: angshuman\_bagchi@yahoo.com

S. Ray

e-mail: raysujay@gmail.com

© Springer India 2015

J.K. Mandal et al. (eds.), *Information Systems Design and Intelligent Applications*,  
Advances in Intelligent Systems and Computing 340,  
DOI 10.1007/978-81-322-2247-7\_13

115

## Abbreviations

- SD Steepest descend  
CD Conjugate gradient

## 1 Introduction

Sulfide and thiosulfate are two of the most abundant forms of sulfur in the aquatic environment. Among the sulphur anions thiosulfate serves as a major sulphur anion ligand and is used by a majority of the micro-organisms to maintain environmental sulfur balance [1–3]. The sulfur oxidation process is carried out by the sulfur oxidizing operon (*sox*) in many bacteria to oxidize the environmental sulfur compounds. Thiosulfate acts as the electron donor in this sulfur oxidation process. Different molecular pathways are involved for oxidation of sulfur anions in the environment. The sulfur oxidation pathways are different in different micro-organism [4]. In *Chlorobium tepidum* (*C. tepidum*), thiosulfate anion binds to the SoxY–Z protein complex. But it has been clearly noted that the SoxY–Z protein complex often remains in inactive form in sulphur oxidizing micro-organisms. In *C. tepidum*, the SoxY–Z protein complex is often activated by SoxF protein from Sox operon [5]. SoxF protein has sulfide dehydrogenase activity and is responsible to enhance the thiosulfate oxidation activity [5]. *Hydrogenophilus thermoluteolus* is a thermophilic facultative chemoautotroph, which lives prevalently in high temperature geothermal niches. The *sox* operon of thermophilic Betaproteobacterium, *Hydrogenophilus thermoluteolus* consists of the genes *soxEFCDYZAXBH* [6, 7]. However, till date no structural or molecular details of the interactions of SoxY–Z protein complex with SoxF protein have been reported in *Hydrogenophilus thermoluteolus*.

In the present study, the three dimensional structures of SoxY, SoxZ and SoxF proteins from *Hydrogenophilus thermoluteolus* were obtained by homology modelling. Molecular docking simulations were then performed in order to build a complex of SoxY–Z and SoxY–Z–F. The possible interaction patterns of SoxY–Z and SoxF were predicted and analysed. There have been no significant works on the analysis of the sulfur oxidation system of *Hydrogenophilus thermoluteolus* therefore an attempt has been made to analyze the plausible involvement of SoxF in the oxidation of environmental thiosulfate anion from a structural point of view. As this is the first approach regarding the structural basis of the involvements of SoxY, SoxZ and SoxF from *Hydrogenophilus thermoluteolus* in the process of biochemical oxidation of sulfur anions, this study may shed light towards the understanding of the complex thiosulfate oxidation process in the moderately thermophilic *Hydrogenophilus thermoluteolus*. Since, this is the first report regarding the molecular structural biology of SoxF protein and its interaction with SoxY–Z protein complex, the study may shed light to analyze the molecular biochemistry of environmental sulfur anion oxidation reactions.

## 2 Materials and Methods

### 2.1 Sequence Analysis and Homology Modelling of SoxY, SoxZ and SoxF Proteins

The amino acid sequences of SoxY, SoxZ proteins from *Hydrogenophilus thermoluteolus* were obtained from NCBI (Accession no: NC\_AB268546.1). The amino acid sequence of SoxF protein was retrieved from Uniprot (Uniprot id: Q08IS5). In order to identify suitable template for homology modelling, the amino acid sequences of the proteins were used separately to search Brookhaven Protein Data Bank [8] using the software tool BLAST [9]. From the BLAST results the following proteins were selected as the templates for homology modeling:

1. Crystal structure of sulphur carrier protein SoxY from *Chlorobium limicola f thiosulfatophilum* (PDB ID: 2NNC chain B) for SoxY with 38 % sequence identity.
2. Crystal structure of ComplexYZ of *Paracoccus pantrotopus* (PDB ID: 2OXG chain Z) for SoxZ with 37 % sequence identity.
3. Crystal structure of flavocytochrome C from *Chromatium vinosum* (*C.vinosum*) (PDB ID: 1FCD chain A) for SoxF with 59 % sequence identity

These crystal structures were then used separately as templates for homology modeling by Swiss Model [10–13] as well as MODELLER in Modweb [14]. The three dimensional models of the proteins were then energy minimized in two steps. In first step Steepest Decent algorithm (SD) was used and in the next step Conjugate Gradient algorithm (CG) was used to minimize the overall structures of the three models using the Discovery studio software until the structures reached the final RMS gradient of 0.0001 kcal/mole. All energy minimizations were done using CHARMM force field [12] and fixing of the backbone of proteins.

### 2.2 Validation of the Models of the Proteins

The stereo-chemical qualities of the three dimensional models were checked using PROCHECK [15], Verify3D [16] and ERRAT [17]. The software tools predicted a good model quality and no residues were found to be in the disallowed regions of the Ramachandran Plot [18].

### 2.3 Molecular Docking Simulations

In order to study the interactions between SoxY–Z protein complex with SoxF, SoxY and SoxZ proteins were first docked using the software Cluspro 2.0 [19]. Cluspro 2.0

is a fully automated web based computational program for protein-protein docking [19]. The algorithm uses the coordinate files of two protein structures through web interface for protein-protein docking. The web server provided 10 docked complexes. From them the model 0 was chosen because it had the best cluster size among all the possible docked structures. The structure of SoxY–Z protein complex was energy minimized by backbone fixation of the proteins in two steps. In first step, SD was used and in the next step, CG was used to minimize the overall structure of the model of the docked complex using the Discovery studio software until the structure reached the final RMS gradient of 0.0001 kcal/mole. All energy minimization were done using CHARMM force fields [12]. The energy minimized structure of SoxY–Z protein complex was used to find out the mode of interactions between the SoxY and SoxZ proteins. Again, to find the binding interactions among SoxY–Z protein complex with SoxF protein, the structure of SoxY–Z protein complex was used to dock with SoxF using Cluspro 2.0. The model of SoxY–Z–F protein complex so obtained was then subjected to energy minimization using the previously mentioned protocol. The possible binding modes were analysed and amino acid residues involved in protein-protein interactions were also identified.

All the energy minimizations were performed in explicit solvent with a dielectric constant of 80.

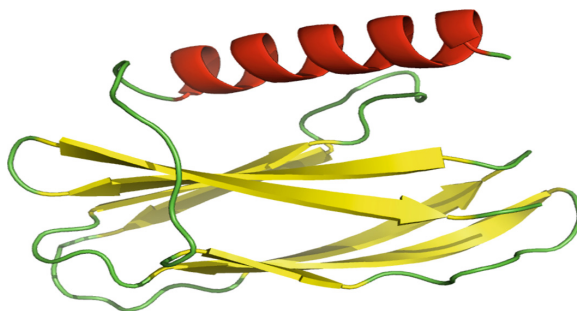
## 2.4 Calculation of Protein-Protein Interaction

Protein-protein interaction profiles from the SoxY, SoxZ and SoxF proteins were calculated using, Protein Interaction Calculator (P.I.C) web server [20]. The three dimensional structures of the protein complexes were provided to P.I.C sever as an input to calculate various kinds of interactions [20].

## 3 Results

### 3.1 Description of the Structure of SoxY

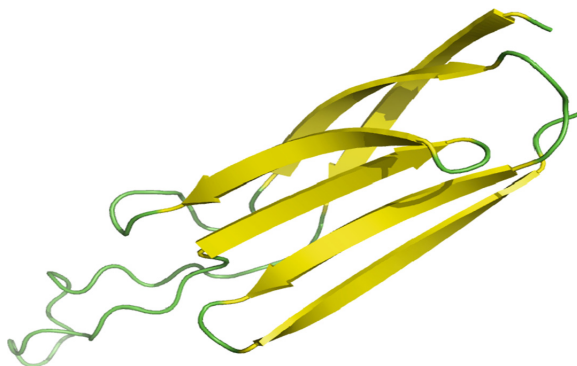
SoxY is a 114 amino acid residue long protein. The predicted structure is very similar to the structure of the Sulphur Carrier Protein SoxY from *Chlorobium limicola f thiosulfatophilum* (PDB code: 2NNC Chain A). The root mean squared deviation (RMSD) between the  $C_{\alpha}$  atoms of the modelled protein and its template is 0.703 Å. The protein comprises  $\alpha$ -helix,  $\beta$ -strands and random coil region. N-terminal position is made up of  $\alpha$ -helical region (amino acid residues 1–25). The most of the structure has  $\beta$  strands (amino acid residues 30–33, 43–48, 59–62, 70–74, 83–89, 95–100, 105–109) interspersed with coil regions. The structure is presented in Fig. 1.



**Fig. 1** Modelled structure of SoxY protein from *Hydrogenophilus thermoluteolus* with distinct secondary structure showing as  $\alpha$ -helix (red),  $\beta$ -sheet (yellow) and random coil (green)

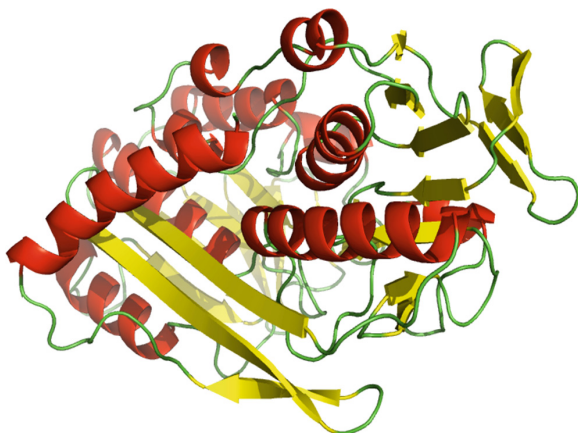
### 3.2 Description of the Structure of SoxZ

SoxZ is a 91 amino acid residue long protein. The predicted structure is very similar to the structure of the SoxZ protein from the SoxY–Z complex from *Paracoccus pantotrophus* (PDB Code: 2OXG chain Z). The root mean squared deviation (RMSD) between the  $C_{\alpha}$  atoms of the modelled protein and its template is 0.51 Å. The protein is made up of  $\beta$ -strands (amino acid residues 5–11, 17–21, 31–40, 44–51, 60–65, 72–77, 86–91) interspersed with coil regions. The structure is presented in Fig. 2.



**Fig. 2** Modelled structure of SoxZ protein from *Hydrogenophilus thermoluteolus* with distinct secondary structure showing as  $\beta$ -sheet (yellow) and random coil (green)

**Fig. 3** Modelled structure of SoxF protein from *Hydrogenophilus thermoluteolus* with distinct secondary structure showing as  $\alpha$ -helix (red),  $\beta$ -sheet (yellow) and random coil (green)



### 3.3 Description of the Structure of SoxF

SoxF is a 388 amino acid residue long protein. The predicted structure is very similar to the structure of the Crystal structure of flavocytochrome C from *Chromatium vinosum* (PDB ID: 1FCD chain A). The root mean squared deviation (RMSD) between the  $C_{\alpha}$  atoms of the modelled protein and its template is 2.94 Å. The N-terminal domain of the protein starts with a  $\beta$ -strand (amino acid residues 4–9) and C-terminal domain ends with  $\alpha$ -helical region (amino acid residues 366–383). Overall the modelled structure of SoxF has a total of 13 helices and 16 strands. The structure is presented in Fig. 3.

### 3.4 Interactions Between SoxY and SoxZ

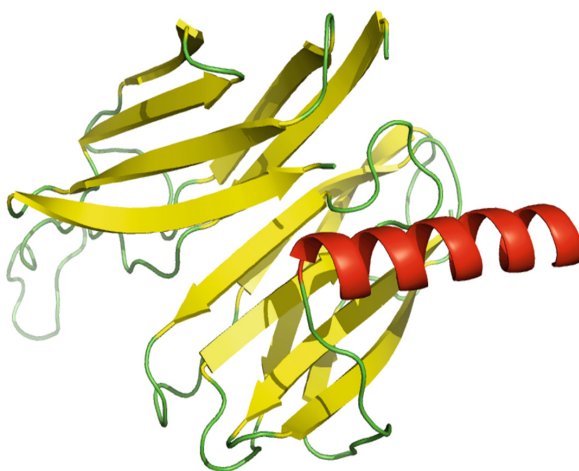
The three dimensional structures of the SoxY–Z protein complex has been generated by the protein-protein docking techniques using the software tool Cluspro2.0 [19]. The analysis using protein interactions calculator (P.I.C) shows SoxY and SoxZ interact strongly with each other by strong non-covalent forces. The predominant interactions are hydrogen bonding (H-bonding) interactions and ionic interactions. There are H-bonding interactions involving both the main chain and the side chains of the two proteins. Tables 1 and 2 represent the protein-protein hydrogen bonding interactions between the SoxY–Z protein complex. The complex structure is presented in Fig. 4.

**Table 1** List of amino acids involved in protein-protein Main Chain-Side Chain Hydrogen Bond interactions in SoxY-Z protein complex

Position	Protein name	Residue	Position	Protein name	Residue
100	SoxY	ASN	17	SoxZ	ASN
100	SoxY	ASN	17	SoxZ	ASN
107	SoxY	HIS	97	SoxZ	GLU
119	SoxY	ARG	70	SoxZ	TYR
19	SoxZ	ARG	103	SoxY	ALA
19	SoxZ	ARG	119	SoxY	ARG
19	SoxZ	ARG	119	SoxY	ARG
70	SoxZ	TYR	119	SoxY	ARG
72	SoxZ	GLU	119	SoxY	ARG
72	SoxZ	GLU	119	SoxY	ARG
97	SoxZ	GLU	105	SoxY	VAL
97	SoxZ	GLU	105	SoxY	VAL

**Table 2** Protein-protein main chain-main chain hydrogen bonds in SoxY-Z complex

Position	Protein name	Residue	Position	Protein name	Residue
19	SoxZ	ARG	103	SoxY	ALA

**Fig. 4** Protein-protein interactions between SoxY and SoxZ proteins with distinct secondary structure showing as  $\alpha$ -helix (*red*),  $\beta$ -sheet (*yellow*) and random coil (*green*)

**Table 3** Protein-protein main chain-main chain hydrogen bonds in SoxY-Z-F complex

Position	Protein name	Residue	Position	Protein name	Residue
91	SoxY	GLN	116	SoxZ	GLY
122	SoxZ	ARG	70	SoxY	ALA
205	SoxF	LYS	203	SoxZ	GLU

**Table 4** Protein-protein main chain-side chain hydrogen bonds in SoxY-Z-F complex

Position	Protein name	Residue	Position	Protein name	Residue
65	SoxY	ASN	118	SoxZ	GLU
65	SoxY	ASN	118	SoxZ	GLU
65	SoxY	ASN	118	SoxZ	GLU
65	SoxY	ASN	118	SoxZ	GLU
86	SoxY	ARG	173	SoxZ	TYR
86	SoxY	ARG	528	SoxF	PRO
86	SoxY	ARG	528	SoxF	PRO
87	SoxY	VAL	527	SoxF	ASP
88	SoxY	LYS	527	SoxF	ASP
91	SoxY	GLN	115	SoxZ	ASN
91	SoxY	GLN	115	SoxZ	ASN
115	SoxZ	ASN	90	SoxY	GLY
115	SoxZ	ASN	90	SoxY	GLY
122	SoxZ	ARG	86	SoxY	ARG
122	SoxZ	ARG	86	SoxY	ARG
122	SoxZ	ARG	86	SoxY	ARG
122	SoxZ	ARG	86	SoxY	ARG
135	SoxZ	LYS	567	SoxF	LEU
161	SoxZ	THR	228	SoxF	ARG
167	SoxZ	ALA	506	SoxF	TYR
169	SoxZ	SER	530	SoxF	LYS
173	SoxZ	TYR	86	SoxY	ARG
175	SoxZ	GLU	86	SoxY	ARG
175	SoxZ	GLU	86	SoxY	ARG
203	SoxZ	GLU	18	SoxY	ALA
203	SoxZ	GLU	18	SoxY	ALA
506	SoxF	TYR	166	SoxZ	GLY
513	SoxF	LYS	164	SoxZ	TRP
527	SoxF	ASP	86	SoxY	ARG
527	SoxF	ASP	86	SoxY	ARG
527	SoxF	ASP	175	SoxZ	GLU
534	SoxF	THR	166	SoxZ	GLY

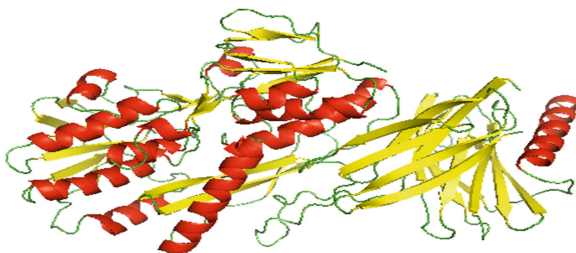


### 3.5 Interactions of SoxF with SoxY–Z Complex

The three dimensional structure of the SoxY–Z–F protein complex has been generated by the protein-protein docking techniques using the software tool Cluspro2.0 [19]. The analysis using protein interactions calculator (P.I.C) shows strong non-covalent interactions in SoxY–Z–F complex structure. There are H-bonding interactions involving both the main and the side chains of the proteins in the SoxY–Z–F protein complex. Apart from H-bonding there are hydrophobic, ionic and aromatic-aromatic interactions. Tables 3, 4 and 5 represent the protein-protein interaction schemes in the SoxY–Z–F protein complex. The complex structure is presented in Fig. 5.

**Table 5** Protein-protein side chain-side chain hydrogen bonds in Sox Y–Z–F complex

Position	Protein name	Residue	Position	Protein name	Residue
65	SoxY	ASN	117	SoxZ	ASP
65	SoxY	ASN	117	SoxZ	ASP
91	SoxY	GLN	115	SoxZ	ASN
91	SoxY	GLN	115	SoxZ	ASN
91	SoxY	GLN	115	SoxZ	ASN
91	SoxY	GLN	115	SoxZ	ASN
93	SoxY	SER	117	SoxZ	ASP
115	SoxZ	ASN	91	SoxY	GLN
115	SoxZ	ASN	91	SoxY	GLN
115	SoxZ	ASN	91	SoxY	GLN
115	SoxZ	ASN	91	SoxY	GLN
117	SoxZ	ASP	65	SoxY	ASN
117	SoxZ	ASP	65	SoxY	ASN
117	SoxZ	ASP	89	SoxY	MET
117	SoxZ	ASP	89	SoxY	MET
134	SoxZ	ARG	554	SoxF	GLU
134	SoxZ	ARG	554	SoxF	GLU
134	SoxZ	ARG	554	SoxF	GLU
134	SoxZ	ARG	554	SoxF	GLU
134	SoxZ	ARG	554	SoxF	GLU
175	SoxZ	GLU	527	SoxF	ASP
175	SoxZ	GLU	527	SoxF	ASP
205	SoxF	LYS	203	SoxZ	GLU
228	SoxF	ARG	163	SoxZ	GLU
228	SoxF	ARG	163	SoxZ	GLU
228	SoxF	ARG	163	SoxZ	GLU
228	SoxF	ARG	163	SoxZ	GLU



**Fig. 5** Protein-protein interaction between SoxY–Z protein complex and SoxF protein with distinct secondary structure showing as  $\alpha$ -helix (red),  $\beta$ -sheet (yellow) and random coil (green)

## 4 Discussion

In this study, an attempt has been made to evaluate the structural basis of the involvements of SoxY, SoxZ and SoxF in binding and their role in sulfur oxidation process. It has been documented from experimental studies that SoxF might function as an activator of the SoxY–Z protein complex [5]. In order to predict the molecular basis of interactions of SoxF with the SoxY–Z protein complex in *Hydrogenophilus thermoluteolus*, the three dimensional structures of the SoxY, SoxZ and SoxF proteins have been generated using homology modelling. After building the three dimensional structures of the proteins, protein-protein dockings have been used to first generate the SoxY–Z protein complex and then SoxY–Z–F protein complex. The interaction patterns of SoxY and SoxZ have been elucidated by protein-protein docking and P.I.C webserver. The interaction schemes revealed the presence of non-covalent interaction mainly ionic, hydrophobic and hydrogen bonding interactions in the protein complex. Mainly ionic interaction between SoxY and SoxZ protein would help in the binding of the sulfur anion to the SoxY protein. From the interaction pattern of SoxY–Z–F protein complex it has been established that SoxF interacts strongly with SoxY–Z complex with the mode of interactions being mainly non covalent in nature. Thereby, SoxF helps SoxZ in binding thiosulfate. The sulfur ligand binding domain in SoxZ protein becomes more exposed when SoxF protein binds to the SoxY–Z protein complex. In this way, SoxF protein tries to activate the SoxY–Z protein complex. Before SoxF protein binds to SoxY–Z protein complex, the sulfur ligand binding domain of SoxZ protein remains buried inside the SoxY–Z protein complex.

The results from this study, may lead to an elucidation of the structural basis behind the molecular functions of these proteins. This model provides a rational platform to design experiments for the determination of the contribution of the various amino acid residues in SoxY, SoxZ and SoxF proteins to predict the molecular basis of their interactions in sulfur oxidation mechanism.

**Acknowledgment** The authors would like to thank the DST-PURSE programme 2012–2015 going on in the Department of Biochemistry and Biophysics, University of Kalyani for providing

different instrumental and infrastructural support. The author is also thankful to the DBT sponsored Bioinformatics Infrastructure Facility in the Department of Biochemistry and Biophysics, University of Kalyani for the necessary support.

## References

1. Freidrich, C.G., Bardischewsky, F., Rother, D., Quentmeier, A., Fischer, J.: Prokaryotic sulfur oxidation. *Curr. Opin. Microbiol.* **8**, 253–259 (2005)
2. Freidrich, C.G., Rother, D., Bardischewsky, F., Quentmeier, A., Fischer, J.: Oxidation of reduced inorganic sulphur compounds by bacteria: emergence of a common mechanism? *Appl. Environ. Microbiol.* **67**, 2873–2882 (2001)
3. Bagchi, A.: Structural insight into the mode of interactions of SoxL from *Allochromatium vinosum* in the global sulfur oxidation cycle. *Mol. Biol. Rep.* **39**, 10243–10248 (2012)
4. Freidrich, C.G.: Physiology and genetics of sulfur-oxidizing bacteria. *Adv. Microb. Physiol.* **39**, 235–289 (1998)
5. Ogawa, T., Furusawa, T., Shiga, M., Seo, D., Sakurai, H., Inoue, K.: Biochemical studies of a soxF-encoded monomeric flavoprotein purified from green sulfur bacterium *Chlorobium tepidum* that stimulates in Vitro thiosulfate oxidation. *Biosci. Biotechnol. Biochem.* **74**, 771–780 (2010)
6. Sano, R., et al.: Thiosulphate oxidation by a thermo-neutrophilic hydrogen-oxidizing bacterium *Hydrogenobacter thermophilus*. *Biosci. Biotechnol. Biochem.* **74**, 892–894 (2010)
7. Miyake, D., et al.: Thiosulfate oxidation by a moderately thermophilic hydrogen oxidizing bacterium, *Hydrogenophilus thermoluteolus*. *Arch. Microbiol.* **188**, 199–204 (2007)
8. Berman, M.H., et al.: The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000)
9. Altschul, S.F., et al.: Basic local alignment search tool. *J. Mol. Biol.* **25**, 403–410 (1990)
10. Arnold, K., Bordoli, L., Kopp, J., Schwede, T.: The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006)
11. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., Schwede, T.: The SWISS-MODEL repository and associated resources. *Nucleic Acids Res.* **37**, D387–D392 (2009)
12. Peitsch, M.C.: Protein modeling by E-mail. *Bio/Technology* **13**, 658–660 (1995)
13. Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M.: CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* **4**, 187–217 (1983)
14. Eswar, N., Marti-Renom, M.A., Webb, B., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., Sali, A.: Comparative Protein Structure Modeling With MODELLER. *Current Protocols in Bioinformatics* **15**, 5.6.1–5.6.30 (2006)
15. Laskowski, R.A., et al.: PROCHECK: a program to check the stereochemistry of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993)
16. Eisenberg, D., et al.: VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **277**, 396–404 (1997)
17. Colovos, C., Yeates, T.O.: Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519 (1993)
18. Ramachandran, G.N., Sashisekharan, V.: Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–438 (1968)
19. Comeau, S.R., et al.: ClusPro: An automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **20**, 45–50 (2004)
20. Tina, K.G., Bhadra, R., Srinivasan, N.: PIC: protein interactions calculator. *Nucleic Acids Res.* **35**, W473–W476 (2007)