

C-Terminal Sequence Determinants of T4 Bacteriophage Tail Fiber Adhesin for Specific Lipopolysaccharide Recognition

Ewa Brzozowska^{1*}, Anna Pyra², Malgorzata Miśków¹, Sabina Górńska¹ and Andrzej Gamian¹

¹Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Weigla 12, 53-114 Wrocław, Poland

²Faculty of Chemistry, Department of Crystallography, 14 F. Joliot-Curie, 50-383 Wrocław, Poland

Received: April 14, 2015; Accepted: June 01, 2015; Published: June 15, 2015

*Corresponding author: Ewa Brzozowska, Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Weigla 12, 53-114 Wrocław, Poland, Tel: +48-713-371-172; Email: ezuziak@iitd.pan.wroc.pl

Abstract

Identification of short peptides that serve as specific ligands to biological materials such as microbial cell surfaces has major implications in better understanding the molecular recognition of cell surfaces. The adsorption specificity of T4 phage is determined by an adhesive protein (adhesin) localized on the tip of the phage tail fiber. For the first time, we have shown that the specific interaction between adhesin and *Escherichia coli* B Lipopolysaccharide (LPS) occurs even if the phage protein is in denatured form. In this paper the peptide sequences, that bind to the LPS were identified and are characteristic only for T4 phage. Our results suggest that the primary structure of the phage adhesin might be responsible for the receptor recognition in the crucial role plays which the variable region of the protein the sequences are localized in its C-terminal part of the adhesin. This hypervariable area is flanked and interrupted by highly conserved stretches of histidine residues and contains amino acids which, favorably interact with terminal glucosyl- α -1,3-glucose of the bacterial LPS. Moreover, in this paper the specific LPS recognition and binding by short synthetic peptides, containing the variable regions, as well as those with histidine boxes have been studied. Of peptides of which amino acid sequences correspond to the variable region of the adhesin and contain histidine boxes have been studied.

The peptides were synthesized on pins and analyzed in the LPS binding assay. It was showed shown that these highly variable peptides imitating T4 phage variable regions, specifically interacts with the *E. coli* B LPS. The conserved regions, containing histidine residues, and which occur also in other phage adhesins, don't bind to the LPS.

Keywords: Adhesin; Bacteriophage; Determinant; Recognition; Amino-acid Sequence

Abbreviations

BSA: Bovine Serum Albumin; CBS: Carbohydrate Binding Site; DMF: Dimethylformamide; HRP: Horseradish Peroxidase; LPS: Lipopolysaccharide; SDS: Sodium Dodecyl Sulfate; TBST: Tris Buffered Saline plus 0.1% Tween 20; TMP: 3,3',5,5'-tetramethylbenzidine

Introduction

Bacteriophages (phages) are viruses specific towards

bacteria strains. Their ability to a specific recognition of bacteria is used in many fields of science and technology such as bacteria identification by phage typing or as bacteria detection sensors [1]. They are also good model systems to study viral infection and replication, gene transfer, proteins folding and assembly. Bacteriophage T4 was isolated on *Escherichia coli* B (*E. coli* B) by Demerec and Fanco [2]. T4 phage binds to the *Escherichia coli* B Lipopolysaccharide (LPS) or OmpC protein using C-terminal region of phage adhesin, called as gene product 37 (gp37) localized at the distal part of the long tail fibers [3,4]. A monomer of gp37 consists of 1026 amino acid residues with receptor-binding region encompasses C terminal amino acid residues [5]. These amino acids are involved in a specific interaction with two terminal glucoses of *E. coli* B lipopolysaccharide [6].

In our previous paper, we have shown that the adhesin – both in native and denatured form – recognizes and binds *E. coli* B LPS with a high specificity [1]. The next goal was to establish which amino acid sequence (primary determinant) of the phage protein is responsible for specific LPS recognition and binding. We concerned with the C-terminal region of the gp37 adhesin according to prior studies of Montag et al. [7,8] or Miroshnikov et al. [9]. They suggested that the specificity is determined by repeating motifs: His-X-His-Y (His boxes) which flank very variable sequences encompassing few to tens amino acid residues (Figure 1). The authors have suggested, that the His boxes participate in bacterial receptor recognition and their relative positioning may affect host specificity [9].

Experiments with hybrid phages made by Montag et al. [8] suggest the receptor-binding region of gp37 encompasses residues 907–996, likewise structural analysis of the adhesin made by Bartual et al. [5] also indicate that LPS binding is initiated by the head domain comprising amino acid residues of 940–960. However, the LPS binding determinants have not been precisely defined yet.

To determine the probability of carbohydrate binding by single amino acid in the sequence, we made a sequence-based prediction of the Carbohydrate Binding Site (CBS) using the CBS

```
GEHSHYIEAWNGTGVGGNKMSSYAISYRAGGSNTNAAGN  
HSHTFSFGTSSAGDHSHSVGIGAHHTHTVAGSHGHTITVNS  
TGNTENTVKNI AFNYIVRLA
```

Figure 1: Amino acid sequence of the C-terminal part of adhesin gp 37 (amino acid residues from 934 to 1026) containing His-X-His-Y motifs.

Pred web-server (<http://tardis.nibio.go.jp/netasa/cbs-pred/>) the prediction methodology takes the predisposition of residues and the sequential environments – solvent accessibility and neighbor residue features as an input [10]. According to this methodology, a binary prediction was made for every residue – 1 for binding and 0 for non-binding.

In the next step, we used pin method [11] to synthesize 8 different polypeptides of which amino acid sequences correspond to the C-terminal end of T4 gp37 within 930 and 1026 amino acid residues. Two of them display a specific ability to bind LPS of *E. coli* B.

Materials and Methods

Bioinformatics analysis

The prediction of carbohydrate binding sites in peptides were performed using CBS-Pred (Carbohydrate Binding Site Prediction) web-server (<http://tardis.nibio.go.jp/netasa/cbs-pred/>) [10,11].

Synthetic peptides secondary structure predicted using PEP-FOLD 1.5 [12]. PEP-FOLD is a service for the de novo prediction of peptide conformation from amino acid sequence. It is based on a Hidden Markov Model derived Structural Alphabet that describes a conformation as a series of local overlapping canonical conformations. PEP-FOLD 3D generation is performed using a discrete set of structural prototypes and using the OPEP coarse-grained force field. PEP-FOLD will not show disulfide bonds. The output files from PEP-FOLD are .pdb files which can be opened by pymol - molecular visualization system [13].

Peptide synthesis on polyethylene pins

Peptide synthesis was performed using NCP Block of 96 Hydroxypropylmethacrylate pins (MIMOTOPES, Clayton, Victoria, Australia) according to the standard protocol described by Carter [14]. F-moc amino acids* (60 mm in Dimethylformamide DMF) were mixed with an equimolar amount of diisopropylcarbodiimide and N-hydroxybenzotriazole as coupling reagents for 4 h. Coupling reactions were tested using 10 mm bromophenol blue to check the presence of free amine groups. After synthesis, the peptides were deprotected using 20% piperidine in DMF, washed with methanol, 0.5% acetic acid in methanol and once again in methanol, drying and storage at -20°C. Pins with bonded synthetic peptides were washed using 0.1 M phosphate buffer, pH 7.2, containing 1% sodium dodecyl sulfate (SDS) and 0.1% 2-mercaptoethanol, heated to 60°C and sonicated for 10 minutes.

Fmoc-Ala-OH, Fmoc-Gly-OH, Fmoc-Ile-OH, Fmoc-Leu-OH,

Fmoc-Met-OH, Fmoc-Phe-OH, Fmoc-Val-OH, Fmoc-Pro-OH, the side group protected by Boc (tboc): removed using 95% TFA, Fmoc-Trp(Boc)-OH or tboc, Fmoc-Lys(Boc)-OH or tboc, the side group protected by tbu:Fmoc-Tyr(tbu)-OH, Fmoc-Thr(tbu)-OH, Fmoc-Ser(tbu)-OH, the side group protected by otbu: removed using 95% TFA: Fmoc-Asp(otbu)-OH, Fmoc-Glu(otbu)-OH, the side group protected by Trt: Fmoc-His(Trt)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Cys(Trt)-OH

LPS binding assay

The *E. coli* B LPS binding assay was performed using biotinylated LPS and synthetic peptides tethered to pins. The LPS biotinylation was performed according to the method described by Caccavo et al. [15]. Pins with synthetic peptides were transferred into wells in 96-well plate containing 1% bovine serum albumin (BSA) in 50 mM Tris-Buffered Saline Plus 0.1% Tween 20 (TBST). The blocking reaction was performed for 1 h at room temperature (RT). After that, the biotinylated LPS was added (100 ng/well) into the wells containing peptides and incubated at 12°C overnight. The pins were washed three times using TBST and incubated for 1 h, at RT with avidin conjugated with Horseradish Peroxidase (HRP) (Thermoscientific) (avidin-HRP dilution was 1:1000). TMP (3,3',5,5'-tetramethylbenzidine) (Promega) reagent was used to measure HRP activity according to the protocol [16]. The enzymatic reaction was stopped by removal of pins from the plate and the amount of colored product was measured at $\lambda = 490$ nm in a plate reader (biotek). The positive control in this test was reactivity of 30 ng of adhesin with biotinylated LPS in the same conditions as described above.

Western blot assay

The reactivity of synthetic peptides was shown in a test in which they were first pre-incubated with a biotinylated LPS solution (1 μ g/ml) and next the same solution of LPS was used in the Western blot assay with adhesin immobilized on Immobilon-P (Millipore) membrane. Adhesin was transferred from SDS-PAGE [17]. Blocking reaction was performed using 1% BSA in Tris Base Saline buffer containing 0.05% of Tween-20 (TBS-T). The membrane was washed three times with TBS-T. Next, the solution of LPS was added for one-night incubation at 4°C. After subsequent washing with TBS-T, the avidin conjugated with Horseradish Peroxidase (HRP) (Thermo Scientific) was added for one hour incubation at RT. The membrane was washed three times using TBS-T and installed on HRP substrate - 3-amino-9-ethylcabazole - solution (0.1 M sodium acetate, pH = 5.2 supplemented with H₂O₂).

The Positive control of the test was the immobilized adhesin incubated an initial concentration of LPS (1 μ g/ml).

Results

In our studies, we took under consideration the C-terminal part of adhesin within 934 and 1026 residues because of its high content of His-boxes (five repeats). The peptides containing at least one His-box and peptides containing variable sequences localized between His-boxes, but without histidine residues were synthesized. The peptides are listed in Table 1.

The bioinformatic analysis showed that 72% of amino acid building the C-end of gp37 can interact with carbohydrates. The peptide analysis was performed using the CBS-prep software. The binary values were obtained by choosing a cutoff and marking residues with higher predicted probability as binding. The Expected accuracy scores were: the best cutoff for binary classification = 0.075793, the best binary classification, Net Pred was 60.4%, Sensitivity was 51.87% and the Specificity was 68.95% (Sensitivity $S1 = TP / (TP+FN)$, Specificity $S2 = TN / (TN+FP)$). TP - correctly identified binding sites, TN - correctly identified non-binding sites, FP - number of non-binding residues wrongly identified as binding by predictor, FN - number of binding residues predicted as non-binding)

The LPS binding assay results showed that only peptide no 5 and no 8 bound to the biotinylated LPS of *E. coli* B (Figure 2). The Absorbance (A) at $\lambda = 490$ nm obtained in the test was 0.237 for peptide no 5 and 0.188 for peptide no 8, respectively, with regard to the negative control (A at $\lambda = 490$ nm equal 0.062) and the positive control - adhesin gp37 ($\lambda = 490$ nm equal 0.301). These two peptides - 5 and 8 - were composed of 24 amino acid residues and did not contain any His-boxes. Peptide number 5

had only 3 amino acid residues which did not bind carbohydrates according to the CBS-Pred analysis. Peptide no 8 had 50% non-binding amino acid residues. Peptide number 7 contained three His-boxes and according to bioinformatic analysis only 3 amino acid residues (12%) did not interact with carbohydrates while in the assay its activity was low. The binding levels of peptide 7 was higher compared to the negative control, but over twice lower than the binding level of peptide 8, the same peptide number 6. The peptides 6&7 bound to the LPS of *E. coli* B at a low level (the $A_{490} < 0.1$) were not analyzed further. The short peptides containing one His-box did not bind the biotinylated LPS.

The LPS binding ability of the peptides 5 and 8 were confirmed using the Western blot assay. In this test we showed, that whole pool of LPS was bound to the peptides immobilized on pins during the pre - incubation step. After pre-incubation, the solution of biotinylated LPS was used in Western blot assay in which the adhesin was transferred onto Immobilon-P membrane from SDS-PAGE gel (Figure 3).

The 24 amino acid peptides can adopt a favorable carbohydrate binding conformation in solution (in the ELISA conditions). There is such a supposition because in the test the peptides are not in denatured form. Thus, they could fold into native form and their conformation could be responsible for peptide-LPS interaction. The model structures of the peptides 5 and 8 according to Pep-fold server prediction (<http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/>) are α -helix with a disordered loop of the N-terminal end and β -sheet with loop respectively (Figure 4) [12].

Discussion

In our studies, we showed that amino acid sequences: AWNGTGVGGNKMSYAIYSRAGGS (no 5) and TVNSTGNTENTVKNI AFNYIVRLA (no 8) bind to the *E. coli* B

Table 1: Synthetic peptides immobilized on pins.

GEHSHYIE
GNHSHTFS
GDHSHVGI
GAHTHTVA
AWNGTGVGGNKMSYAIYSRAGGS
GNHSHTFSFGTSSAGDHSHSV
GDHSHSVGIGAHHTVAIGSHGHT
TVNSTGNTENTVKNI AFNYIVRLA

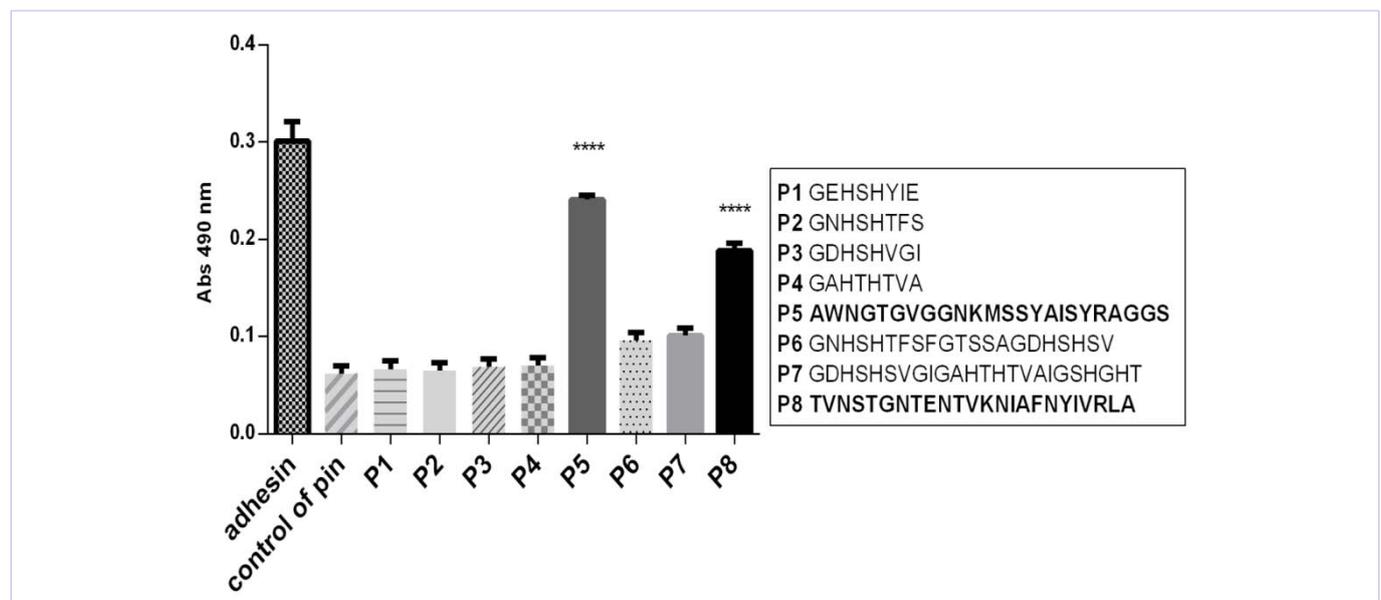


Figure 2: The results of *E. coli* B LPS binding by the synthetic peptides (1 - 8) immobilized on pins in the LPS binding assay - ELISA - compared to the positive control- adhesin.

LPS in a specific manner. Synthetic peptides containing His-box motifs, the long and the short ones, are not capable to bind to the lipopolysaccharide molecule. The studies included the part of adhesin gp37, which was previously indicated by Montag et al. [8]; as a part responsible for phage-bacteria interaction. The sequence no 5, that determines the specificity of LPS binding belongs to the variable region of T4 adhesin and our studies showed that the region of the adhesin is responsible for LPS specific recognition and binding. As was previously noted, the His-X-His-Y boxes (which X and Y are predominantly Ser or Thr) flank the variable regions and are present at the sites almost exactly corresponding to the sites in adhesive proteins of phages: T4, tuia, tuib, Stf and λ . However, the only T4 phage can infect *E. coli* B using LPS as receptor [8,14,17]. The other phages use bacterial outer membrane proteins (porins) to infect their host bacteria (but not the *E. coli* B). It is known, that the His-box motifs are metal-binding sites which interact with divalent ions such as Fe^{2+} or Zn^{2+} [5]. The divalent cations as Zn^{2+} enhance phage T4 infectivity and contribute to cell lysis in environmental systems [19]. The amino acid sequence of peptide no 5, which showed the highest ability of the *E. coli* B LPS binding is present only in T4 phage adhesin, while the sequence of peptide no 8 is highly homologous to those present in all adhesive proteins of phages listed above. The amino acid sequence of the peptide no 5 is

identical to the sequence belonging to the head domain of the adhesin located at the extreme distal end of the long tail fiber. Previously, it was suspected to play a primary role in receptor binding [5]. The peptide no 8 can also bind the LPS of *E. coli* B, so the presence of the homological sequence should also allow to bind the LPS by the other phages (tuia, tuib, Stf and λ). The adhesive protein of tuib phage contains exactly the same amino acid sequence like peptide no5 and the T4 adhesin - gp37. The experiments presented by Yu et al. [20] confirmed that tuib phage also uses LPS as receptor initiating the infection process. It is very likely, that the other phages can also bind to the same lipopolysaccharide receptor. The LPS binding is an intermediate stage of bacterial infection, but it is not enough to infect the bacteria. It should be emphasized, that if the infection does not occur, it does not mean the lack of LPS binding. It is known, that the adhesin gp37 interacts with the terminal glucosyl- α -1,3-glucose of lipopolysaccharides [21] and protein-saccharide interactions almost always involve stacking of sugar residues onto aromatic amino acid side chains [22]. So the side chain of aromatic amino acids of the peptide no 5 (Trp-936, Tyr-949, and Tyr-953) are the attractive candidates for receptor binding. Moreover, Lys-945 (near Trp-936) and Arg-954 present in the peptide 5 may also interact with the lipopolysaccharides phosphate groups. Peptide no 8 also contains aromatic amino



Figure 3: Western blot assay of biotinylated LPS binding by the adhesin using LPS solutions which were first pre-incubated with the synthetic peptides on pins. Numbers 5 and 8 show the absence (no pink band) of the LPS that was previously bound to the peptides. (1-8 peptide numbers, C- positive control).

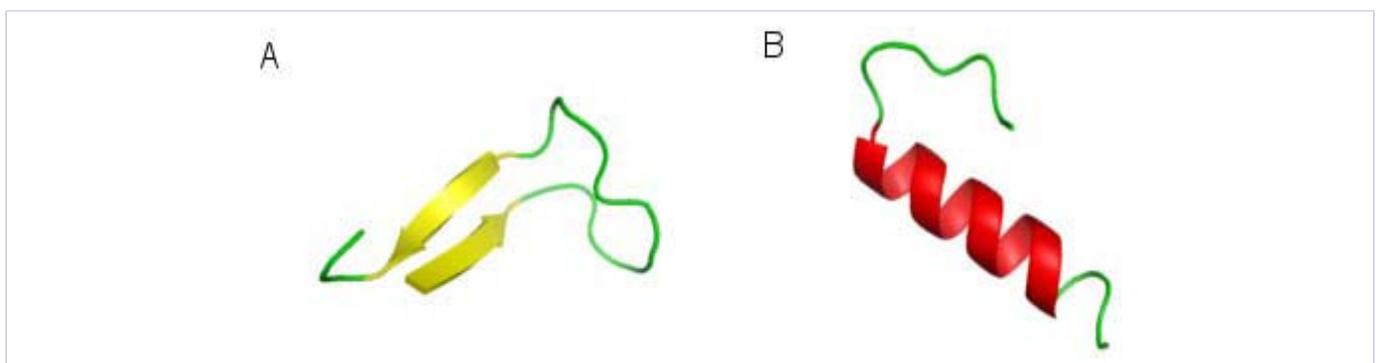


Figure 4: Model structures of peptide 5 (A) and 8 (B) according to the PEP-FOLD 1.5. [12].

acid side chains (Phe-1018 and Tyr-1020) capable to bind the carbohydrate receptor as well as charged amino acid (Lys-1014 and Arg-1024) involved in the LPS phosphate group interaction.

E. coli B infection by T4 phage has been studied for a long time and it is considered as a model system of bacteria – phage interaction. In this paper, the amino acid sequence responsible for the interaction has been described. However, identification of short peptides those serve as specific ligands to biological materials such as microbial cell surfaces become more and more popular. Short peptides recognizing bacterial cells in specific manner have major implications in better understanding the molecular recognition of cell surfaces and for bacteria detection.

Acknowledgements

The authors gratefully acknowledge support for this work from the Polish National Science Centre under grant No. 2011/03/D/NZ7/02054.

References

1. Brzozowska E, Smietana M, Koba M, Górka S, Pawlik K, Gamian A, et al. Recognition of bacterial lipopolysaccharide using bacteriophage-adhesin-coated long-period gratings. *Biosensors & Bioelectronics*. 2015; 67: 93-99. doi: 10.1016/j.bios.2014.07.027.
2. Demwrec M, Fano U. Bacteriophage resistant mutants in *Escherichia coli*. *Genetics*. 1945; 30(2): 119–136.
3. Hashemolhosseini S, Montag D, Kramer L, Henning U. Determinants of receptor specificity of coliphages of the T4 family. A chaperone alters the host range. *J Mol Biol*. 1994; 241(4): 524-33.
4. Riede I, Degan M, Henning U. The receptor specificity of bacteriophages can be determined by a tail fiber modifying protein. *EMBO J*. 1985; 4(9):2343-6.
5. Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC, et al. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proc Natl Acad Sci USA*. 2010; 107(47): 20287-92. doi: 10.1073/pnas.1011218107.
6. Prehm P, Jann B, Schmidt G, Strim S. On a bacteriophage T3 and T4 receptor region within the cell wall Lipopolysaccharide of *Escherichia coli* B. *J Mol Biol*. 1976;101(2):277-81.
7. Montag D, Riede I, Eschbach M.L, Degen M, Hening U. Receptor-recognizing proteins of T-even type bacteriophages. Constant and hyper variable regions and an unusual case of evolution. *J Mol Biol*. 1987; 196(1): 165-74.
8. Montag D, Hashemolhosseini S, Henning U. Receptor-recognizing proteins of T-even type bacteriophages. The receptor-recognizing area of proteins 37 of phages T4, Tula and Tulb. *J Mol Biol*. 1990; 216(2): 327-34.
9. Miroshnikov KA, Marusich EI, Cerritelli ME, Cheng N, Hyde CC, Steven AC, et al. Engineering trimeric fibrous protein based on bacteriophage T4 adhesins. *Protein Eng*. 1998; 11(4): 329-32.
10. Malik A, Firoz A, Jha V, Ahmad S. PROCARB: A Database of Known and Modelled Carbohydrate-Binding Protein Structures with Sequence-Based Prediction Tools. *Adv Bioinformatics*. 2010; 436036. doi: 10.1155/2010/436036.
11. Malik A, Ahmad S. Sequence and structural features of carbohydrate binding in proteins and assessment of predictability using a neural network. *BMC Struct Biol*. 2007; 7: 1.
12. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Research*. 2012; 40(Web Server issue): W288-W293. doi:10.1093/nar/gks419.
13. The PyMOL Molecular Graphics System, Version 1.2r3pre. Schrödinger LLC. Available from <https://www.pymol.org/pymol>.
14. Michael CJ, Jacq B, Argues G, Bickle TA. A remarkable amino acid sequence homology between a phage T4 tail fibre protein and ORF314 of phage lambda located in the tail operon. *Gene*. 1986; 44(1): 147-50.
15. Caccavo D, Afeltra A, Pece S, Giuliani G, Freudenberg M, Galanos C, et al. Lactoferrin-lipid A-lipopolysaccharide interaction: inhibition by anti-human lactoferrin monoclonal antibody AGM 10.14. *Infect Immun*. 1999; 67(9): 4668-72.
16. Gosling JP. *Immunoassays: A Practical Approach*. Oxford: Oxford University Press. 2000.
17. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970; 227(5259):680-5.
18. Montag D, Schwarz H, Hennin U. A component of the side tail fiber of *Escherichia coli* bacteriophage lambda can functionally replace the receptor-recognizing part of a long tail fiber protein of the unrelated bacteriophage T4. *J Bacteriol*. 1989; 171(8): 4378-84.
19. Adams MH. *Bacteriophages*. New York: Interscience publishers. 1959.
20. Yu F, Yamada H, Mizushima S. Role of lipopolysaccharide in the receptor function for bacteriophage Tulb in *Escherichia coli*. *J Bacteriol*. 1981; 148(2): 712-5.
21. Yu F, Mizushima S. Roles of lipopolysaccharide and outer membrane protein OmpC of *Escherichia coli* K-12 in the receptor function for bacteriophage T4. *J Bacteriol*. 1982; 151: 718–722.
22. Vyas NK. Atomic features of protein-carbohydrate interactions. *Curr Opin Struct Biol*. 1991; 1(5): 732–740.