Modelling and Analysis of interactions between ligand D-Gluconate and receptor GntR

Mohit Kumar

MS14092

A dissertation submitted for the partial fulfilment of
BS-MS dual degree in Science

Indian Institute of Science Education and Research Mohali

December 2019

Certificate of Examination

This to certify that the dissertation titled **"Modelling and Analysis of interactions between ligand D-Gluconate and receptor GntR"** submitted by **Mr. Mohit Kumar** (Reg. No. MS14092) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Debrina Jana Dr. Rachna Chaba Dr. Monika Sharma

(Supervisor)

Dated: December 5, 2019

Declaration

The work presented in the dissertation has been carried by me under the guidance of Dr.Monika Sharma at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, or diploma, or a fellowship to any other University or Institute. Whenever the contribution of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Mohit Kumar

(Candidate)

Dated: December 5, 2019

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge

Dr. Monika Sharma

(Supervisor)

Acknowledgement

I would like to acknowledge all the people who have been around me and helped me with my project, It cannot be done without the help of my project supervisor Dr Monika Sharma who gave me this opportunity.

I cannot expect a life without friends I have made in IISER, I will always remember Jugaadis(Akash, Ankit, Ajay, Kapil, Munish, Pranshu, Ravi, Rowny, Sahil and Vishal) who always been part my life at IISER and helped me every time. My friend Rishi, Apoorv, Vodi you have been supporting too.

I really want to thank my friend Kapil for his support and help throughout the project.

List of figures

List of Tables

Notations

Deoxyribonucleic Acid- DNA

Nuclear Magnetic Resonance - NMR

Visual Molecular Dynamics - VMD

Protein Data Bank - PDB

Alanine - ala

Arginine - arg

Asparagine - asn

Aspartic acid - asp

Cysteine - cys

Glutamine - gln

Glutamic acid - glu

Glycine - gly

Histidine - his

Isoleucine - ile

Leucine - leu

Lysine - lys

Methionine - met

Phenylalanine - phe

Proline - pro

Serine - ser

Threonine - thr

Tryptophan - trp

Tyrosine - tyr

Valine - val

Abstract

GntR is the protein present in the B.Subtilis (a gram-positive, rod-shaped bacteria), Here is the study and docking of DNA with GntR and Analysis of the interactions of protein GntR docked with Gluconate.

So, for the GntR sequence, here we find a model template and align that to the sequence to make a model which is further used for the study of interactions.

Contents

3. Result and analysis

4. Bibliography

1.Introduction

1.1 Theory

1.1.1 Transcriptional regulation

Regulation is, we all know, controlling something. Transcription is a process where RNA is made from DNA template with the help of none other than an RNA polymerase. So, with this, transcriptional regulators come out to be understood as the controlling factors which regulate the rate of gene transcription either by promoting or hindering RNA polymerase binding to DNA template. The regulators involved in a huge range of mechanisms in order to increase or decrease the production of RNA or ultimately of proteins. There are different types of factors that help in the transcriptional regulation. There are promoters, sigma factors, co-activators and co-repressors.

· Promoter is that region of DNA template which initiates the process of transcription of particular gene.

· Sigma factors are bacterial co-factors that join with the RNA polymerase to encode sequence specificity.

· Co-activators and Co-repressors are nothing but proteins working with other transcription factors to increase and decrease the rate of transcription respectively.

If we look at the prokaryotes transcription is taken care of by 3 main elements viz. Promoters, Operators(recognise repressors to inhibit the process of transcription) and

Positive control elements. In prokaryotes the entire process depends on the work of promoters and presence of activators and repressors. Repressors would take on the promoter's location and inhibit further RNA polymerase binding with the template.

Transcriptional regulation in eukaryotes is much more complex than that in prokaryotes due to large number of proteins involved and presence of introns in the former. Basically, in eukaryotes, we find three types of polymerases as RNA Pol 1, Pol 2 and Pol 3, each having specific targets and activities and individual mechanisms. Moreover the DNA is highly coiled around histones and thus assistance is required from other factors in the nucleus for making the gene accessible to polymerase. Just like we have sigma factors in bacteria, there are General Transcription factors (GTFs) which help in stabilizing binding interactions and opening of DNA helix to help RNA polymerase access the DNA template.

After the binding of polymerase of DNA template , some other proteins help in elongating the nascent RNA strand. This process is called as promoter escape. Here the process can be accelerated or retarted by the regulatory elements.

1.1.2 D-Guconate

D-gluconate is a gluconate which is having D-configuration of it. Dextrorotation and levorotation are used to distinguish chiral organic compounds structure. Gluconate have a role as a human metabolite.It is a conjugate base of Gluconic acid which is an organic compound with molecular formula $C_6H_{11}O_7$ (1).

In aqueous solution of neutral pH it forms the gluconate ion. The salts of gluconic acid are known as "gluconates."

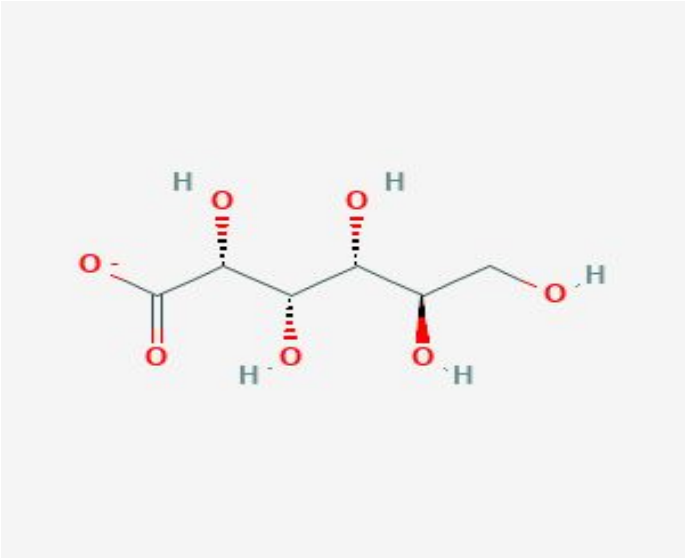


Fig. 1.1 Gluconate ion

1.1.3 GntR

GntR(Gluconate operon repressor) is a protein which is present in the B. subtilis is a gram-positive bacteria which is also present in the mammalian gut. Many bacterial transcription regulation proteins bind DNA through a helix-turn-helix (HTH) motif, which can be classified into subfamilies on the basis of sequence similarities. The HTH GntR family has many members distributed among diverse bacterial groups that regulate various biological processes. It was named GntR after the Bacillus subtilis repressor of gluconate operon(10). Gluconate operon of Bacillus subtilis includes the gntR, gntK, gntP, and gntZ genes, respectively encoding the transcriptional repressor of the operon, gluconate kinase, the gluconate permease, and an unidentified open reading frame. So here we are doing the modelling and docking analysis of GntR docked with D-Gluconate(6).

1.2 Software used

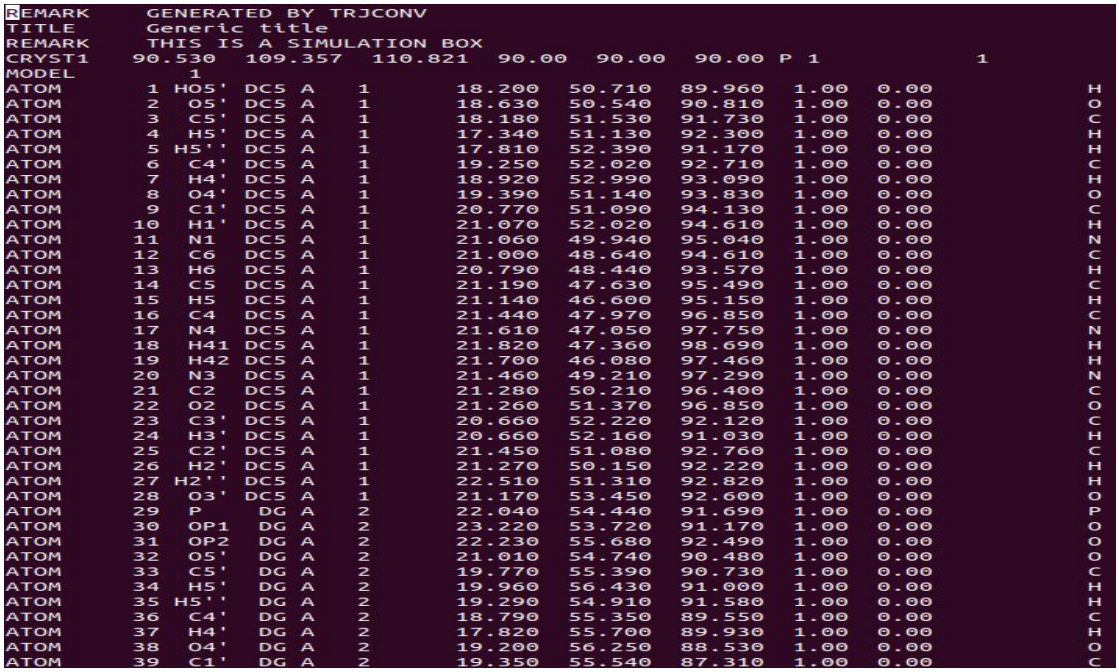
Here is the software list, which are used.

1) PyMol
2) Autodock vina(12)
3) Modeller
4) VMD(8)
5) LigPlot

1.3 PDB file

The PDB stands for Protein data bank which is provided by (www.rcsb.org)(2). PDB file gives us information about 3D structures (atomic coordinates) of protein and nucleic acids. It provides a standard representation for macromolecular structure data derived from X-ray diffraction and NMR studies.

Protein Data Bank was developed in the 1970s and today Protein Data Bank has a lot of proteins data in a PDB file format which can be freely accessible to anyone for their simulation purposes.



```
REMARK     GENERATED BY TRJCONV
TITLE      Generic title
REMARK     THIS IS A SIMULATION BOX
CRYST1   90.530  109.357  110.821  90.00  90.00  90.00 P 1           1
MODEL        1
ATOM      1 HO5' DC5 A   1      18.200  50.710  89.960  1.00  0.00           H
ATOM      2  O5' DC5 A   1      18.630  50.540  90.810  1.00  0.00           O
ATOM      3  C5' DC5 A   1      18.180  51.530  91.730  1.00  0.00           C
ATOM      4  H5' DC5 A   1      17.340  51.130  92.300  1.00  0.00           H
ATOM      5 H5'' DC5 A   1      17.810  52.390  91.170  1.00  0.00           H
ATOM      6  C4' DC5 A   1      19.250  52.020  92.710  1.00  0.00           C
ATOM      7  H4' DC5 A   1      18.920  52.990  93.090  1.00  0.00           H
ATOM      8  O4' DC5 A   1      19.390  51.140  93.830  1.00  0.00           O
ATOM      9  C1' DC5 A   1      20.770  51.090  94.130  1.00  0.00           C
ATOM     10  H1' DC5 A   1      21.070  52.020  94.610  1.00  0.00           H
ATOM     11  N1  DC5 A   1      21.060  49.940  95.040  1.00  0.00           N
ATOM     12  C6  DC5 A   1      21.000  48.640  94.610  1.00  0.00           C
ATOM     13  H6  DC5 A   1      20.790  48.440  93.570  1.00  0.00           H
ATOM     14  C5  DC5 A   1      21.190  47.630  95.490  1.00  0.00           C
ATOM     15  H5  DC5 A   1      21.140  46.600  95.150  1.00  0.00           H
ATOM     16  C4  DC5 A   1      21.440  47.970  96.850  1.00  0.00           C
ATOM     17  N4  DC5 A   1      21.610  47.050  97.750  1.00  0.00           N
ATOM     18  H41 DC5 A   1      21.820  47.360  98.690  1.00  0.00           H
ATOM     19  H42 DC5 A   1      21.700  46.080  97.460  1.00  0.00           H
ATOM     20  N3  DC5 A   1      21.460  49.210  97.290  1.00  0.00           N
ATOM     21  C2  DC5 A   1      21.280  50.210  96.400  1.00  0.00           C
ATOM     22  O2  DC5 A   1      21.260  51.370  96.850  1.00  0.00           O
ATOM     23  C3' DC5 A   1      20.660  52.220  92.120  1.00  0.00           C
ATOM     24  H3' DC5 A   1      20.660  52.160  91.030  1.00  0.00           H
ATOM     25  C2' DC5 A   1      21.450  51.080  92.760  1.00  0.00           C
ATOM     26  H2' DC5 A   1      21.270  50.150  92.220  1.00  0.00           H
ATOM     27 H2'' DC5 A   1      22.510  51.310  92.820  1.00  0.00           H
ATOM     28  O3' DC5 A   1      21.170  53.450  92.600  1.00  0.00           O
ATOM     29  P    DG A   2      22.040  54.440  91.690  1.00  0.00           P
ATOM     30  OP1  DG A   2      23.220  53.720  91.170  1.00  0.00           O
ATOM     31  OP2  DG A   2      22.230  55.680  92.490  1.00  0.00           O
ATOM     32  O5'  DG A   2      21.010  54.740  90.480  1.00  0.00           O
ATOM     33  C5'  DG A   2      19.770  55.390  90.730  1.00  0.00           C
ATOM     34  H5'  DG A   2      19.960  56.430  91.000  1.00  0.00           H
ATOM     35 H5''  DG A   2      19.290  54.910  91.580  1.00  0.00           H
ATOM     36  C4'  DG A   2      18.790  55.350  89.550  1.00  0.00           C
ATOM     37  H4'  DG A   2      17.820  55.700  89.930  1.00  0.00           H
ATOM     38  O4'  DG A   2      19.200  56.250  88.530  1.00  0.00           O
ATOM     39  C1'  DG A   2      19.350  55.540  87.310  1.00  0.00           C
```

Fig. 1.2 PDB File

1.4 Modeller

Modeller is a computer program which used for homology modelling of unknown sequences to make models with template proteins. It works on the method of protein NMR, termed satisfaction of spatial restraints.

To make a model we have to find a template for our sequence through BLAST and then do alignment of sequences(position by position equivalence to the template)(3)

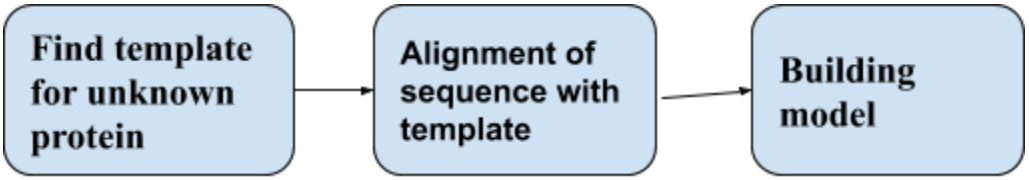


fig.1.3 Modeller's function

In Modeller, we do alignment of the sequence with a template then extract spatial restraints and then satisfy spatial restraints.

1.5 Docking

When two molecular structure a receptor and a ligand binds together to give the stable complex in which they have a good binding affinity. The receptor (Protein) is host molecule commonly bigger in size in which the ligand which is smaller in size get docked in such orientation that they have strong bonding(5).

Molecular docking has become an increasingly important tool for drug discovery due to its ability to find the good conformation and great binding affinity of the small-molecule ligand to the receptor(protein). (4)

There are software in which we can do docking by computational methods like we have used Autodock Vina plugin in PyMol.

Currently available docking tools (3)

Docking tool	License terms	URL	Reference

Autodock	Freeware	http://autodock.scripps.edu/	Forli *et al*. (2016)
PatchDock	Freeware	https://bioinfo3d.cs.tau.ac.il/PatchDock/	Schneidman-Duhovny *et al*. (2005)
GEMDOCK	Freeware	http://gemdock.life.nctu.edu.tw/dock/	Yang and Chen (2004)
Autodock Vina	Open-source	http://vina.scripps.edu/manual.html	Trott and Olson (2010)
rDOCK	Open-source	http://rdock.sourceforge.net/	Ruiz-Carmona *et al.* (2014)
PLANTS	Free for academic use	http://www.uni-tuebingen.de/	Korb *et al.* (2009)
DOCK	Free for academic use	http://dock.compbio.ucsf.edu/	Allen *et al.* (2015)
FRED	Free for academic	https://www.eyesopen.com/oedocking	McGann (2011)
HADDOCK	Free for academic	http://www.bonvinlab.org/software/haddock2.2/	Dominguez *et al.* (2003)
ICM	Commercial	https://www.molsoft.com/docking.html	Neves *et al.* (2012)
GLIDE	Commercial	https://www.schrodinger.com/glide	Repasky *et al.* (2007)
GOLD	Commercial	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/	Verdonk *et al.* (2003)

FlexX	Commercial	https://www.biosolveit.de/FlexX/	Kramer *et al.* (1999)

Table 1.1 Docking tools

2.Methodology

2.1 Modelling of protein using Modeller

After Doing PSI-BLAST search we got 1hw2 as the template for our sequence, we chose it because it has a high DOPE score.

GNTR

>sp|P10585|GNTR_BACSU Gluconate operon transcriptional repressor OS=Bacillus subtilis (strain 168) GN=gntR PE=4 SV=2 (GntR Bacillus subtilis)

MLDSKDLLYPAKWLSKASTGVRVAYELRMRIVSGLIESGTILSENTIAAEFSVSRSPV
REALKILASEKIIRLERMGAVVIGLTEKKIAEIYDVRLLLETFVFERLVKIDIEPLVKDL
SKILEMMKVSIKYEDADEFSFQDVLFHETIIRAIDHSYIQMIWNNLKPVMESFILLSMR
VRLKEKYEDFTRILDNHELYIQAIKTKDRALMIQSLHQNFDDVQDKVEDLWLSQQM
LAKGAEYNND

```
from modeller import *
from modeller.automodel import *
#from modeller import soap_protein_od

env = environ()
a = automodel(env, alnfile='GNTR.ihw2.ali',
              knowns='GNTR', sequence='GNTR',
              assess_methods=(assess.DOPE,
                              #soap_protein_od.Scorer(),
                              assess.GA341))
a.starting_model = 1
```

```
a.ending_model =100
a.make()
```

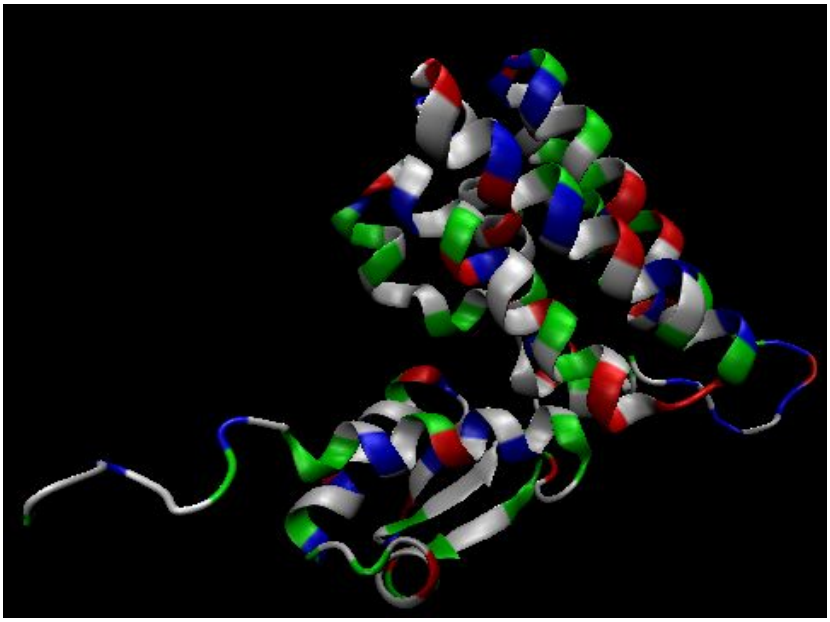


Figure 2.1 Modelled GntR

2.2 Docking of GntR with DNA and D-Gluconate

Autodock vina plugin in the PyMol software used to visualize the model of a protein, which is open-source software. its programing language is python. If the protein is capable of making dimer then PyMol adds crystallographic symmetry to that protein and make dimer by using align command of PyMol(9).

After that by Autodock Vina plugin DNA was docked in the protein. It gives different structures but we chose one with lowest dope value. To check, our structure RMSF and RMSD was done.

GNTR Docked with Gluconate

With the help of PyMol and Autodock Vina plugin Gluconate docked to GNTR.

```
ATOM      1  C5' DC5 A   1      18.190  51.400  91.870  1.00  0.00           C
ATOM      2  O5' DC5 A   1      18.600  50.160  91.310  1.00  0.00           O
ATOM      3  C4' DC5 A   1      19.270  52.100  92.720  1.00  0.00           C
ATOM      4  O4' DC5 A   1      19.440  51.410  93.950  1.00  0.00           O
ATOM      5  C3' DC5 A   1      20.660  52.210  92.050  1.00  0.00           C
ATOM      6  O3' DC5 A   1      21.230  53.460  92.400  1.00  0.00           O
ATOM      7  C2' DC5 A   1      21.440  51.100  92.740  1.00  0.00           C
ATOM      8  C1' DC5 A   1      20.830  51.190  94.150  1.00  0.00           C
ATOM      9  N1  DC5 A   1      21.000  50.000  95.020  1.00  0.00           N
ATOM     10  C2  DC5 A   1      21.290  50.210  96.380  1.00  0.00           C
ATOM     11  O2  DC5 A   1      21.370  51.340  96.860  1.00  0.00           O
ATOM     12  N3  DC5 A   1      21.420  49.160  97.230  1.00  0.00           N
ATOM     13  C4  DC5 A   1      21.260  47.940  96.740  1.00  0.00           C
ATOM     14  N4  DC5 A   1      21.410  46.980  97.600  1.00  0.00           N
ATOM     15  C5  DC5 A   1      20.890  47.670  95.400  1.00  0.00           C
ATOM     16  C6  DC5 A   1      20.760  48.730  94.560  1.00  0.00           C
ATOM     17 'HO5 DC5 A   1      18.210  50.070  90.440  1.00  0.00           H
ATOM     18  H3  DC5 A   1      21.634  49.306  98.216  1.00  0.00           H
ATOM     19  H41 DC5 A   1      21.710  47.240  98.530  1.00  0.00           H
ATOM     20  H42 DC5 A   1      21.420  46.020  97.270  1.00  0.00           H
ATOM     21  P    DG A   2      22.030  54.380  91.350  1.00  0.00           P
ATOM     22  C5'  DG A   2      19.640  55.190  90.490  1.00  0.00           C
ATOM     23  O5'  DG A   2      20.890  54.580  90.200  1.00  0.00           O
ATOM     24  C4'  DG A   2      18.590  55.110  89.370  1.00  0.00           C
```

Fig. 2.2 Protein data

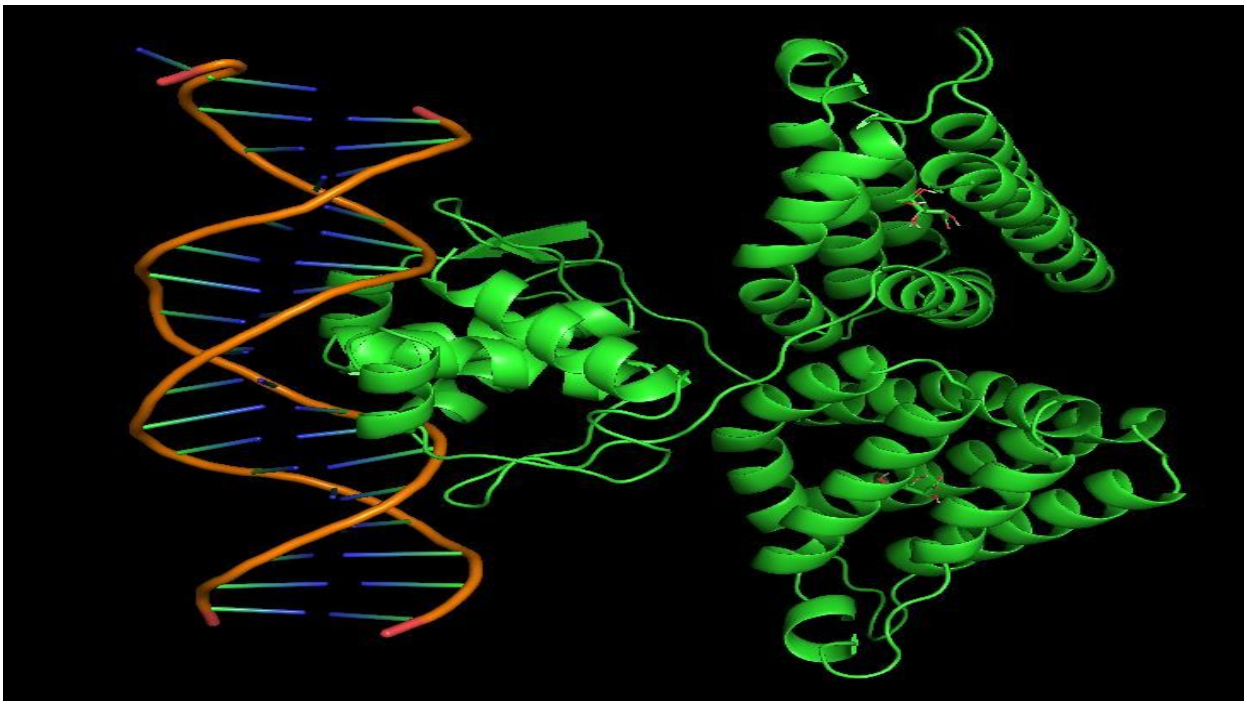


Fig 2.3 GNTR docked with Gluconate and DNA

Mode	Affinity (kcal/mol)	RMSD l.b.	RMSD u.b.
1	-3.9	0.000	0.000
2	-2.9	1.066	4.994
3	-2.7	1.436	4.791
4	-2.2	1.422	1.985
5	-1.8	1.325	4.703
6	-1.7	1.427	4.817
7	-1.6	1.669	3.178
8	-1.5	2.478	4.417
9	-1.4	1.079	4.988

Table 2.1 Conformations of docked ligand in proteins

We used first conformation of the docked ligand because it is showing more negative affinity which gibbs free energy with zero RMSD values which better conformation.

3 Analysis and Result

3.1 Ligand-protein interaction

A ligand which is a small molecule as compared to protein binds itself to the protein with hydrogen bonding ionic bonds, Van der Waals interactions, etc...Protein are dynamic molecules whose functions almost invariably depend on interactions with other molecules, and these interactions are affected in physiologically important ways by sometimes subtle, sometimes striking changes in protein conformation.A ligand binds at a site on the protein called the binding site, which is complementary to the ligand in size, shape, charge, and hydrophobic or hydrophilic character.

We have used LigPlot(11) for finding Protein-Ligand interactions for GntR and DgoR.

3.1.1GntR

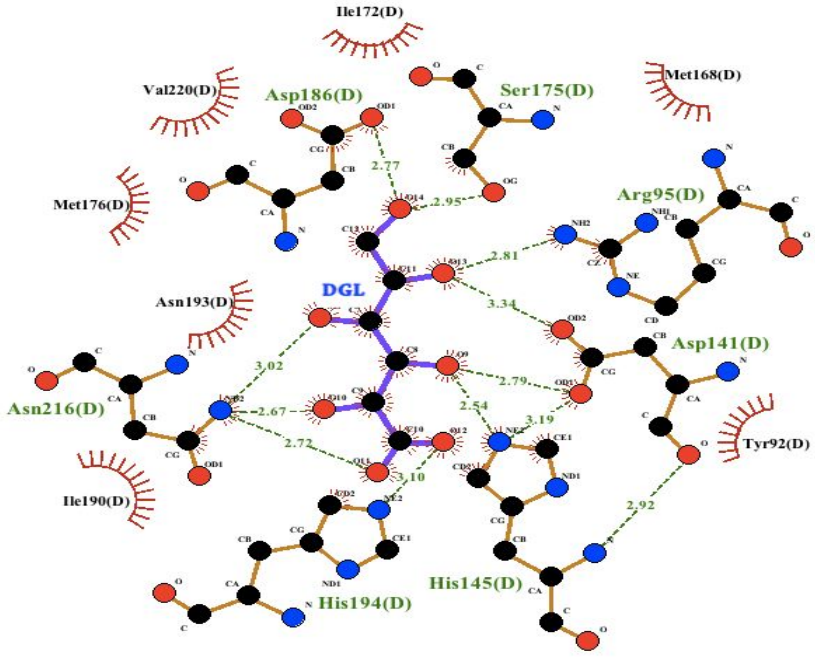


Fig. 3.1 Interactions of ligand in GntR

Following residues of protein are interacting with ligand D-Gluconate along with their bond length which is shown in figure with its nature of interaction.

Asp186, Ser175, Arg95, Asp141, His194, His145, Asn216

3.1.2 DgoR

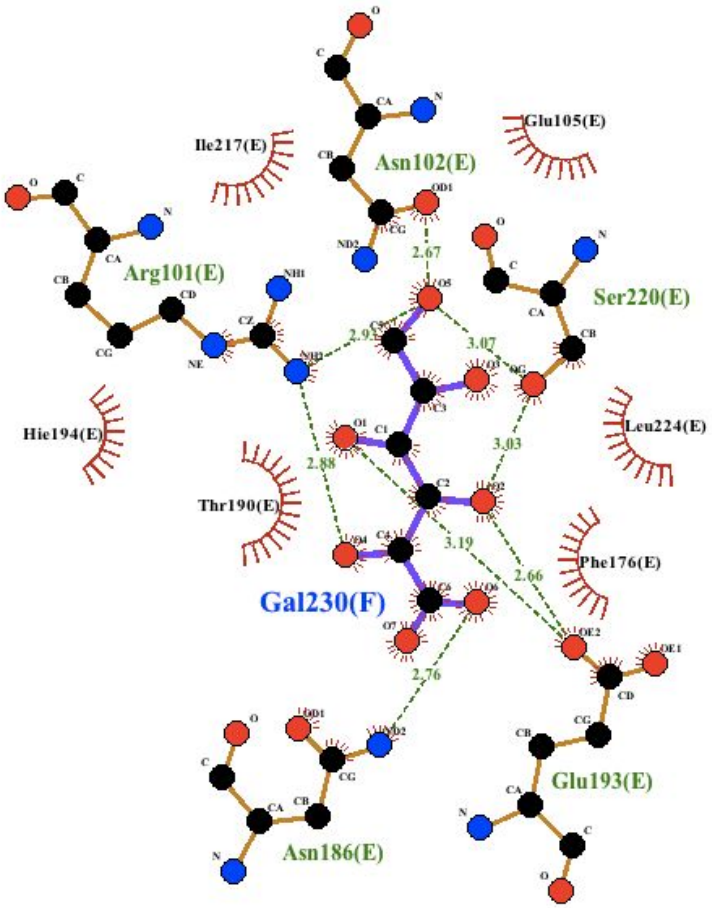


Fig. 3.2 Interactions of ligand in DgoR

Following residues of protein are interacting with ligand D-Galactonate along with their bond length which is shown in figure with its nature of interaction.

Arg10, Asn102, Ser220, Glu193, Asn186

As from both DgoR and GntR we can see that there is a change in interacting residues of the protein DgoR and GntR with their respective ligands D-Galactonate and D-Gluconate, both of the protein are modelled using 1hw2 as template.

3.2 DNA-Protein interactions:

As we have DNA bound to our protein, So there will be interactions, which can regulate the biological function of the DNA, usually expression of genes. Among the proteins that bind to DNA are transcription factors that activate or repress gene expression by binding to DNA motifs and histones that form part of the structure of DNA and bind to it less specifically(7).

So here are DNA-Protein interactions before and after the ligand docked to the protein.

3.2.1 GntR

Before docking of D-Gluconate

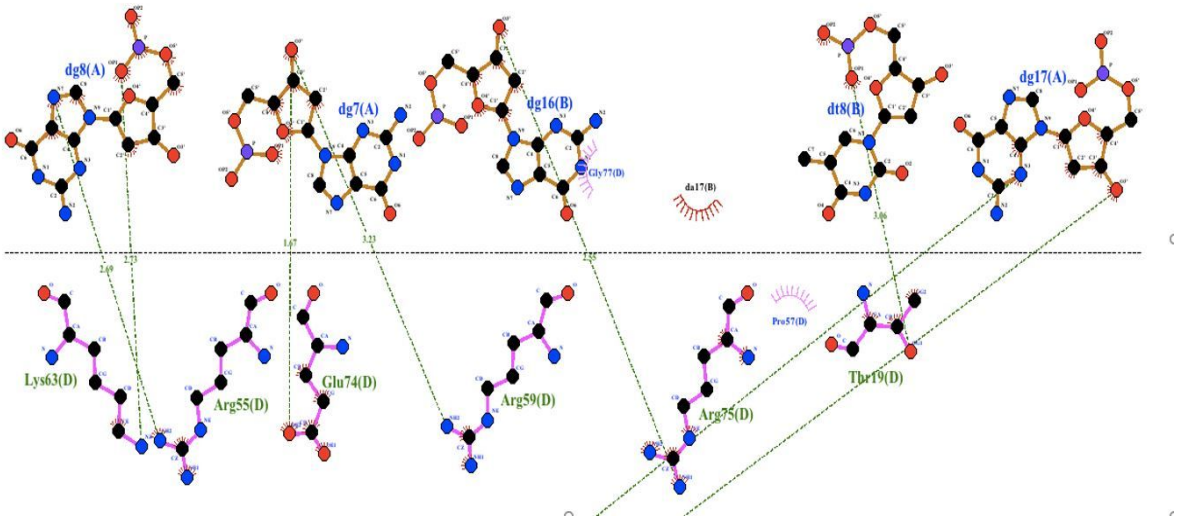


Fig. 3.3 Interactions before docking of ligand part-1

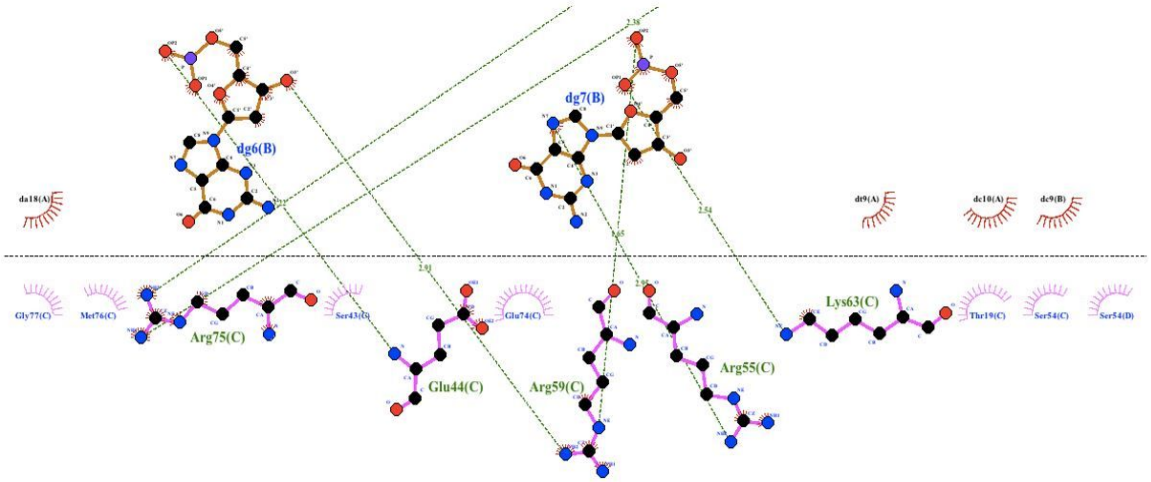


Fig. 3.4 Interactions before docking of ligand part-2

The following residues of DNA and Protein are interacting with each other before docking of D-Gluconate with detailed information about bond length etc. can be seen in the above figure.

DG8-LYS63, ARG55
DG7-GLU74, ARG59
DG16-ARG75
DT8-THR19
DT17-ARG75
DG6-GLU44
DG6-ARG59
DG7-ARG59
DG7-LYS63

After docking of D-Gluconate

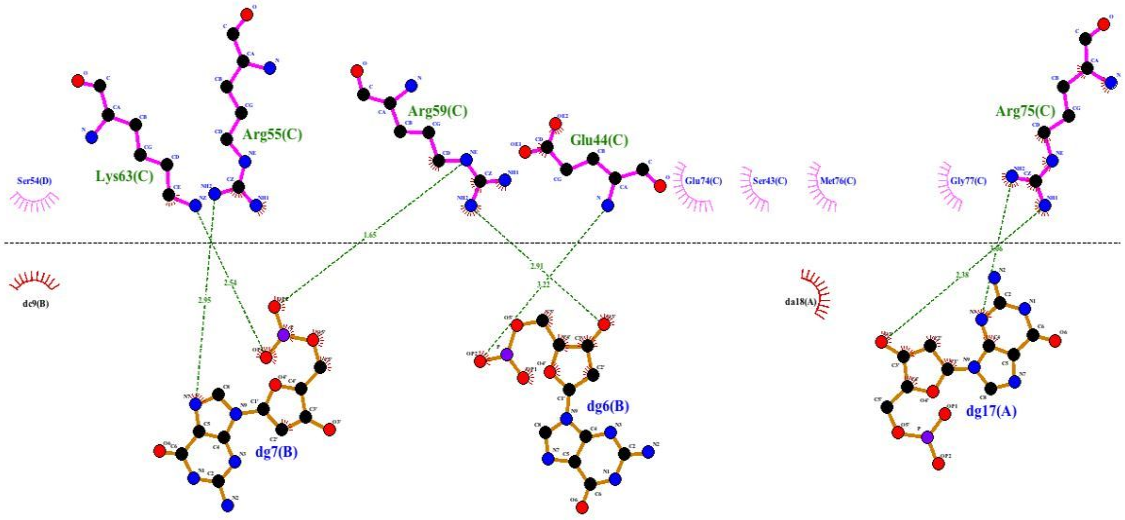


Fig. 3.5 Interactions after docking of ligand part-1

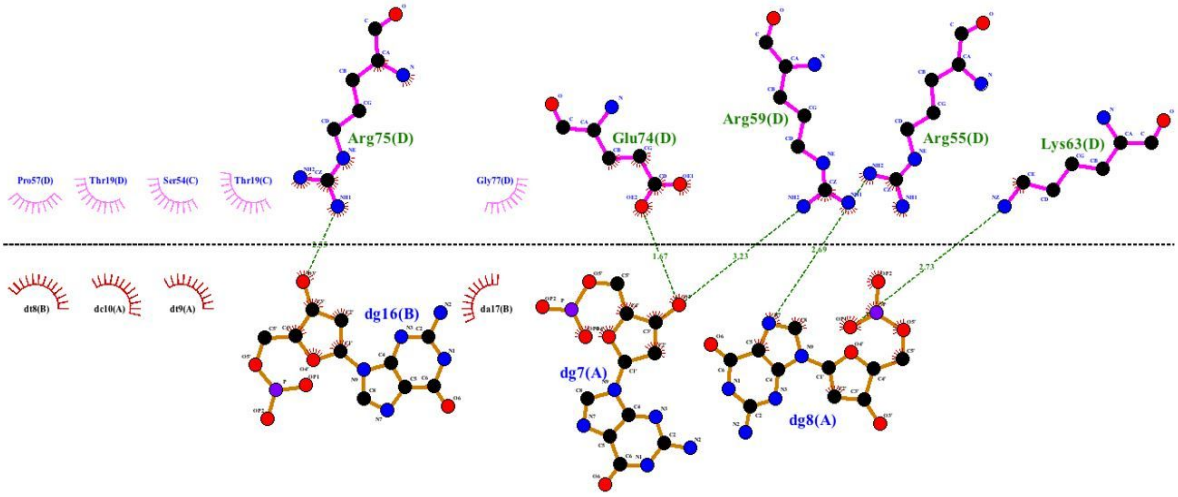


Fig. 3.6 Interactions after docking of ligand part-2

The following residues of DNA and Protein are interacting with each other after docking of D-Gluconate with detailed information about bond length etc. can be seen in the above figure.

DG7-LYS63, ARG55, ARG59
DG6-ARG59, GLU44

DG17-ARG75
DG16-ARG75
DG7-GLU74, ARG59
DG8- ARG55, LYS63

3.2.2 DgoR

Before docking of D-Galactonate

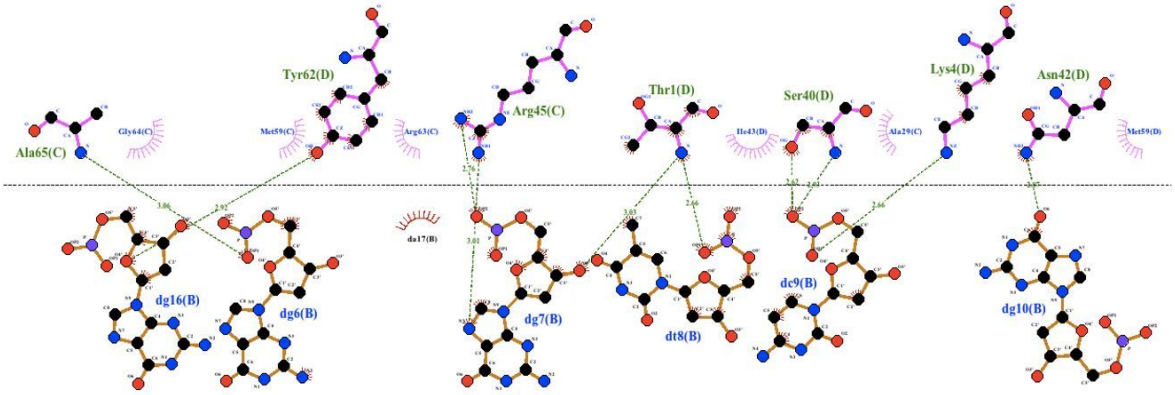


Fig. 3.7 Interactions before docking of ligand part-1

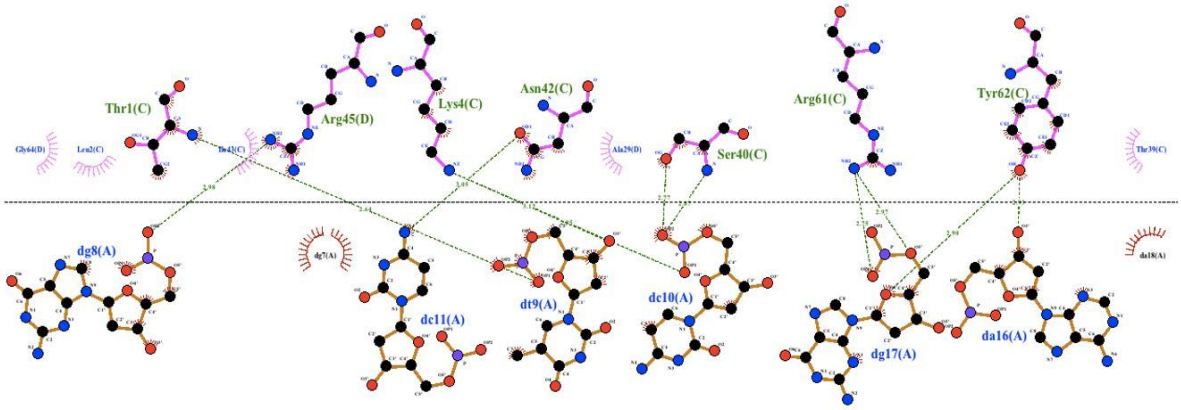


Fig. 3.8 Interactions before docking of ligand part-2

The following residues of DNA and Protein are interacting with each other before docking of D-Galactonate with detailed information about bond length etc. can be seen in the above figure.

DG16-TYR62
DG6-ALA65
DG7-ARG45,THR1
DT8-TRH1
DC9-SER40,LYS4
DG10-ASN42
DG8-ARG45
DT9-TRH1
DC11-ASN42
DC10-LYS4,SER40
DC17-TYR62,ARG61
DA16-TYR62

After docking of D-Galactonate

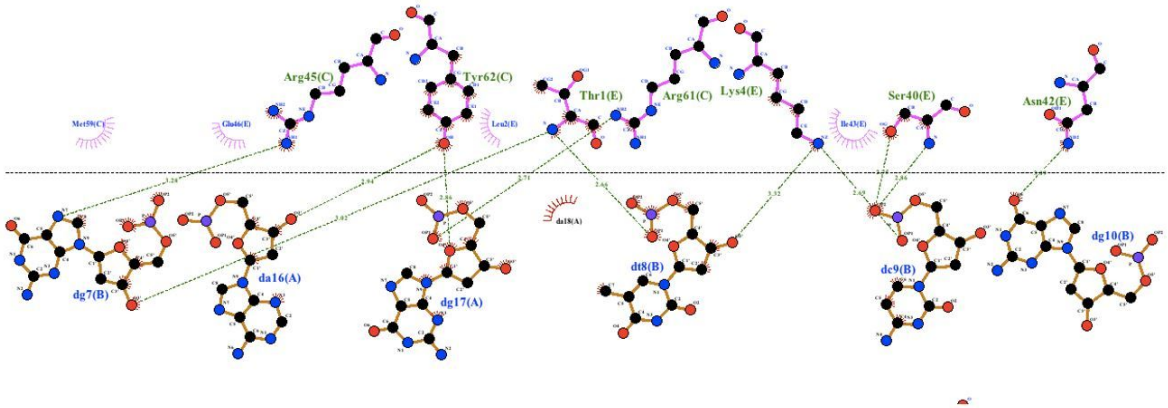


Fig. 3.9 Interactions after docking of ligand part-1

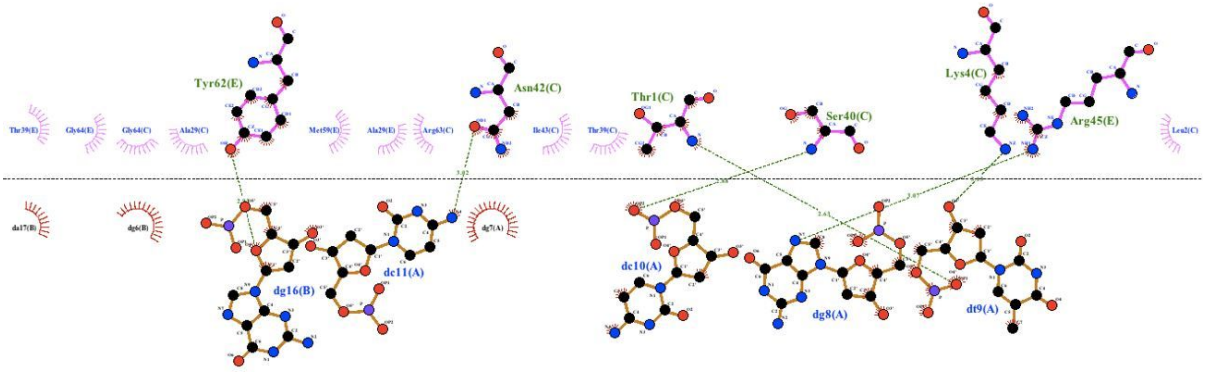


Fig. 3.10 Interactions after docking of ligand part-2

The following residues of DNA and Protein are interacting with each other after docking of D-Galactonate with detailed information about bond length etc. can be seen in the above figure.

DG7-ARG45,THR1
DA16-TYR62
DG17-TYR62,ARG61
DT8-THR1,LYS4
DC9-SER40,LYS4
DG10-ASN42
DG16-TYR62
DC11-ASN42
DC10-SER40
DG8-ARG45
DT9-THR1

As here comparing GntR and DgoR interacting residues before and after docking, we can see that both protein's residues are different which are bound with DNA.

Bibliography

1) National Center for Biotechnology Information. PubChem Database. D-Gluconate, CID=6419706, https://pubchem.ncbi.nlm.nih.gov/compound/D-Gluconate (accessed on Nov. 30, 2019)

2) www.rcsb.org Protein Data bank.

3) B. Webb, A. Sali. Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics 54, John Wiley & Sons, Inc., 5.6.1-5.6.37, 2016.

4) Molecular Docking: A powerful approach for structure-based drug discovery Xuan-Yu Meng,1,2 Hong-Xing Zhang,1, Mihaly Mezei,3 and Meng Cui2, Curr Comput Aided Drug Des. 2011 Jun 1; 7(2): 146–157.

5) Rational Structure-Based Drug Design,
Varun Khanna, Shoba Ranganathan, Nikolai Petrovsky, Sep 2018.

6) GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. Inna A. Suvorova ,Yuri D. Korostelev,Mikhail S. Gelfand, Published: July 7, 2015

7) DNA-protein interactions: methods for detection and analysis. Dey B1, Thukral S, Krishnan S, Chakrobarty M, Gupta S, Manghani C, Rani V. Mol Cell Biochem. 2012 Jun;365(1-2):279-99. Epub 2012 Mar.

8) Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", J. Molec. Graphics, 1996, vol. 14, pp. 33-38.

9) PyMOL, The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

10) Haydon DJ, Guest JR (April 1991). "A new family of bacterial regulatory proteins". FEMS Microbiol. Lett. **63** (2-3): 291–5.

11) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng., Wallace A C, Laskowski R A, Thornton J M (1996). **8**, 127-134.

12) O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461