

Can we call DNA Binding proteins based on DNABIND and DNA Binder results?

During the recent workshop, two the programs [DNABIND](#) and [DNA Binder](#) were mentioned for predicting DNA Binding Proteins. We have found several genes that have been predicted by both programs to be DNA binding proteins (varying strengths), but do not necessarily have strong HHpred alignments to DNA binding. Current BLASTp hits in NCBI and phagesDb are “Hypothetical Proteins.” These genes also appear to be in either an operon or in the syntenic region with other DNA binding proteins, such as DNA methylase, translocase, resolvase, and specific-DNA-binding proteins. Can we call these genes the general name of “**DNA binding protein**” based on the two programs and the strong possibility of the operon or in syntenic region? In general, is it possible to call DNA binding proteins based on these two programs alone?

Two examples from the P1 phage Dynamo are gp 44 (start/stop: 31954-32103) and gp 51 (36150-36428).

Data for Dynamo gp 44:

DNABIND prediction

Prediction parameters:

The prediction is done from sequence. The false positive rate is set to 15.0%
Expected sensitivity: 78.3%
Expected Matthews correlation coefficient: 0.63
The score threshold is set to 0.362 (threshold probability: 0.5896)

Prediction result:

Name	Score	Probability of DNA binding	Prediction
Submitted_sequence	2.867	0.9462	DNA-binding



[ABOUT SERVER](#)



[SUBMIT SEQUENCE](#)

[HELP](#)

[CONTACT](#)

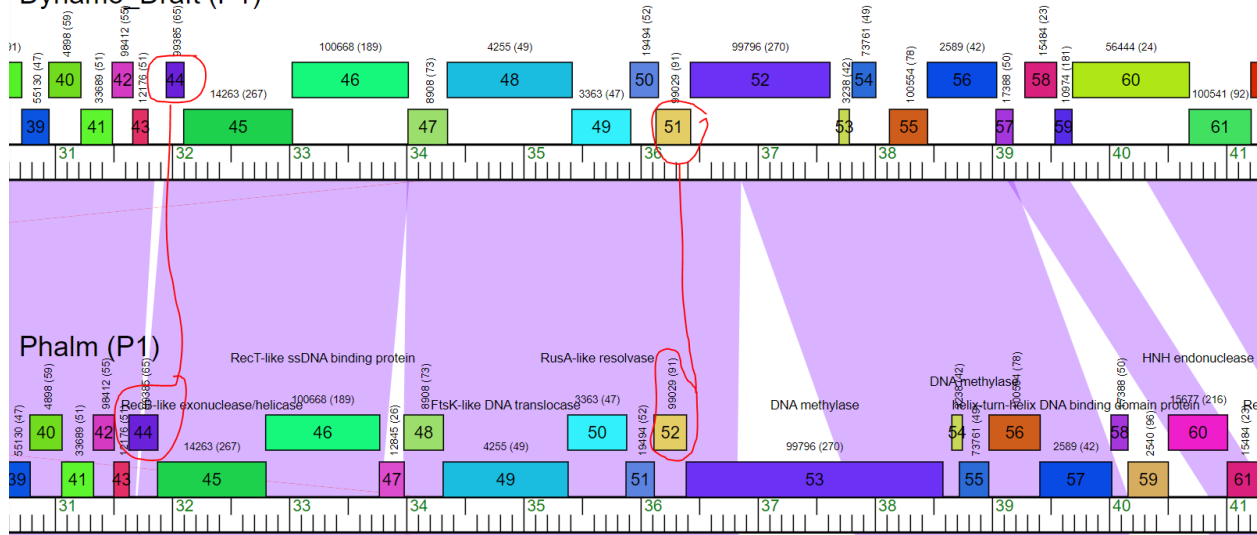


[DEVELOP](#)

Prediction parameters	
Name	Dynamo gp44
Prediction approach	Amino acid composition (Main Data)
SVM threshold	0.0

Prediction Result (Job number= 8556*) of Dynamo gp44		
Sequence name	SVM score	Prediction
seq	0.097058383	DNA-binding protein

Dynamo Draft (P1)



Dynamo gp 44_stop 32103 HHPred

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/>	1	PF14718.9 ; SLT_L ; Soluble lytic murein transglycosylase L domain	54.14	47	19.16	2.6	18	68
<input type="checkbox"/>	2	PF06323.14 ; Phage_antiter_Q ; Phage antitermination protein Q	40.08	80	22.62	2.4	20	220
<input type="checkbox"/>	3	PF07120.14 ; DUF1376 ; Protein of unknown function (DUF1376)	36.28	120	16.4	2.2	22	87
<input type="checkbox"/>	4	PF19808.2 ; DUF6291 ; Family of unknown function (DUF6291)	33.68	110	16.56	1.9	20	79

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/> 1	d3eika2	d.129.1.0 (A:114-197) automated matches {Fungus (Encephalitozoon cuniculi) [TaxId: 6035]}	93.51	1.1	29.66	5.9	43	84
<input type="checkbox"/> 2	d3eika1	d.129.1.0 (A:20-113) automated matches {Fungus (Encephalitozoon cuniculi) [TaxId: 6035]}	92.87	1.6	29.47	6	43	94
<input type="checkbox"/> 3	d1aisa1	d.129.1.1 (A:1-92) TATA-box binding protein (TBP), C-terminal domain {Pyrococcus woesei [TaxId: 2262]}	92.83	1.6	29.07	5.9	43	92
<input type="checkbox"/> 4	d1mp9a1	d.129.1.1 (A:5-96) TATA-box binding protein (TBP), C-terminal domain {Sulfolobus acidocaldarius [TaxId: 2285]}	92.46	1.9	29.03	6	43	92
<input type="checkbox"/> 5	d1mp9a2	d.129.1.1 (A:97-197) TATA-box binding protein (TBP), C-terminal domain {Sulfolobus acidocaldarius [TaxId: 2285]}	92.4	1.8	30.02	6	43	101
<input type="checkbox"/> 6	d1aisa2	d.129.1.1 (A:93-181) TATA-box binding protein (TBP), C-terminal domain {Pyrococcus woesei [TaxId: 2262]}	92.28	1.6	29	5.4	43	89
<input type="checkbox"/> 7	PF00352.24	; TBP ; Transcription factor TFIID (or TATA-	90.24	4.4	23.8	5.3	43	83



DNAbinder

A webserver for predicting DNA-binding proteins

Dynamo gp 51 (36150-36428)



[ABOUT SERVER](#)

[SUBMIT SEQUENCE](#)

[HELP](#)

[CONTACT](#)

[DEVELOPER](#)

Prediction parameters	
Name	---
Prediction approach	Amino acid composition (Main Data)
SVM threshold	0.0

Prediction Result (Job number= 7786 [*])		
Sequence name	SVM score	Prediction
seq	0.0076259802	DNA-binding protein

DNABIND prediction Dynamo gp51 (36150-36428)

Prediction parameters:

The prediction is done from sequence. The false positive rate is set to 15.0%
 Expected sensitivity: 78.3%
 Expected Matthews correlation coefficient: 0.63
 The score threshold is set to 0.362 (threshold probability: 0.5896)

Prediction result:

Name	Score	Probability of DNA binding	Prediction
Submitted_sequence	1.492	0.8163	DNA-binding