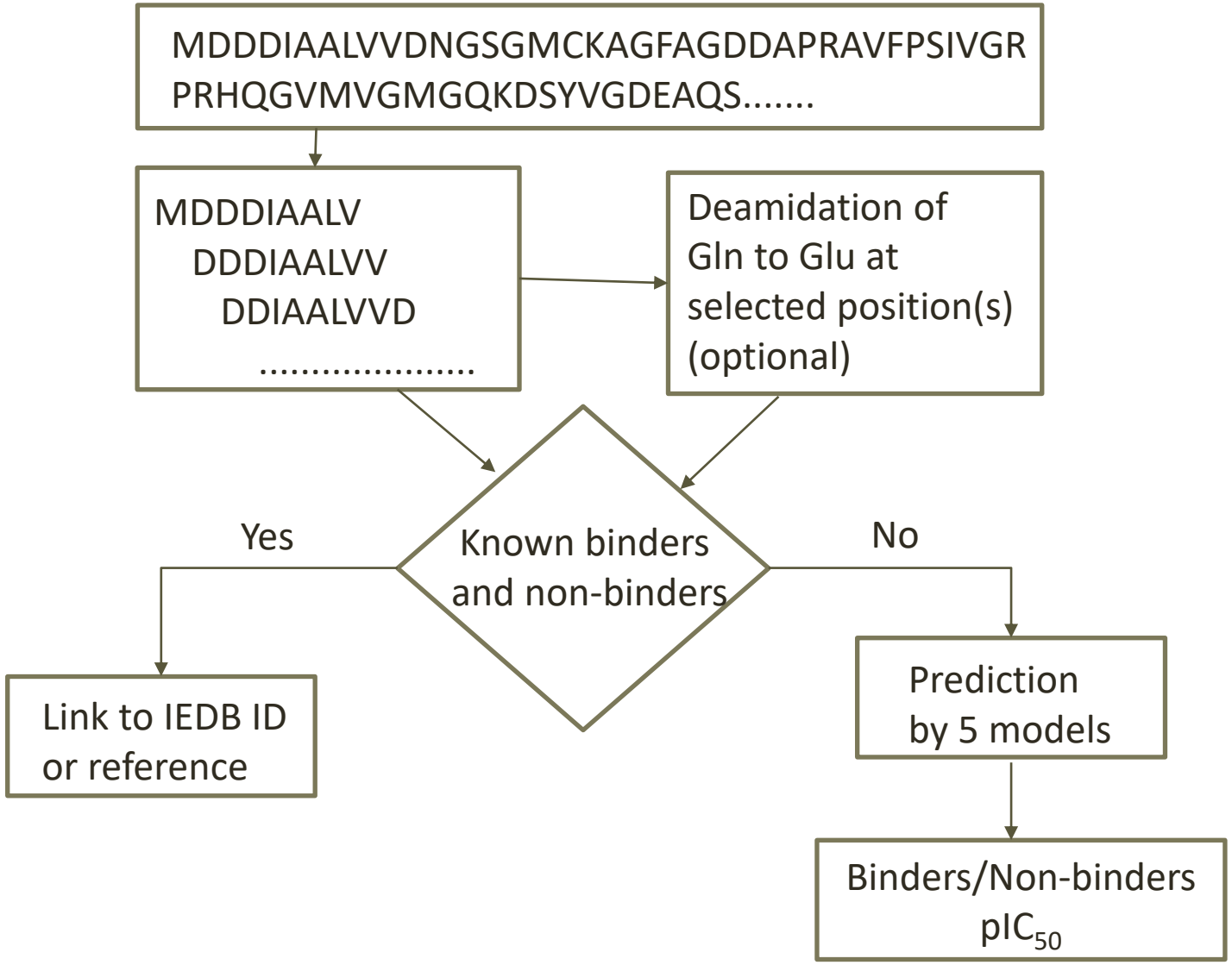
preDQ

a software tool for peptide binding prediction to HLA-DQ2 and/or HLA-DQ8

Ivan Dimitrov, Irini Doytchinova, Mariyana Atanasova, Hristo Valkov

Medical University of Sofia, Bulgaria

preDQ Algorithm



preDQ

>yneN
TTAATGCCTCTTCTCATTTCTTCTGCTGTCATCCGCACAGCAGAAGAATTCCTCATTGAC
TATTATTTCGCAATTTGCTCACATGGATTAAATTAAACTACATACTATAAGATATAAACT
TCTGCCTACAGCTGTAAGAAACTCCGCTCAGTACTGAAGCACCAGTCCTATTTCCTCTTT
TCTCCAGCCTGTTATATTAAGCATACTGATTAACGATTTTTAACGTTATCCGCTAAATAA
ACATATTTGAAATGCATGCGACCACAGTGAAAAACAAAATCACGCAAAGAGACAACTATA
A
>yegR
ACTAACGGCTGCCACCGATAAATTTCAAAAAAGAGCATATACCTAATATTCAACTAAACA
GTGGCATCTTCAATATAATATATTAAAGCCCCCATGGAGTTACCCTGAAGGGCCTCAATG
TCCGTAATTCCTACTTATGTAGGAAATGTTGTACAGAACATTTATTATAATCCTATTCAA
TTATAATAATCATGCCATTATTATATTTAAACACTAGAGAGTGTCGTTGGTATTTAATGG
GGGAAGGTGAGATGAAAAAGATAGCTGCTATATCATTAATTAGTATTTTTATTATGTCTG
G
>emrK
AAATCAGGGATTGTACCGATGATTTATAGTTTCAAGTTGGCACTATAAGTCTTCTTACTA
ATCCTACAGGCGTAAGAATTGTATTGCAAAAGCCACGGTTTAGTCCTCTGTTGTTTTTTT
TGCACCTCATTTAAATTAGGCCTCCAACGTTCCTGGGATAATGTGCAACACATGCACTGT
GTTTGATATGAAGAATGAATGCTCTTTTCATTCAATTCATAAATTTCATCTATGAGAAAT
GAGAGATAATAGTGGAACAGATTAATTCAAATAAAAAACATTCTAACAGAAGAAAATACT
T
>evgA
AATACAATTCTTACGCCTGTAGGATTAGTAAGAAGACTTATAGTGCCAACTTGAAACTAT
AAATCATCGGTACAATCCCTGATTTTATTGTTGACATTTCATTTATGCCGACTATTTATA
TGGTATACTTGTCGAATTATCTTAAAGGAAGCTCAGATTTTCTTATTTTTATTGAGAAAA
TGAGATGACGCCTTATGTCTGTATTACTACAGGGAGAAGGGAGATGCTTCATTGCAAAGG
GAATAATCTATGAACGCAATAATTATTGATGACCATCCTCTTGCTATCGCAGCAATTCGT
>yfdX
TGGCTGTATTTACATTTAATTAATCAGTATTTACATCGATATAATAAATGACATCTCTTT
GTGGTATATAAGAATAGTTCTCTGCGACAGGAAGCATATTCCTACAATTGTAAGACTAAA
ATACTTCTTGCGATAATAACTACAACTGTAAGATAACCCTTTCAAAATGACCGTTGCTCT
CTGATTTCTCATTTCATGCTCACCCAATATGATGGCGGCGTTTTCTAAAACTGTTAAAGA
ATGAGGTAAGTATGAAACGTTTAATTATGGCCACGATGGTCACAGCAATTCTGGCATCTT
C

preDQ

a tool for peptide binding prediction to HLA-DQ2 and/or HLA-DQ8

Home | Method Description | Contact

Help

1. Insert your protein sequence in one letter code

YSQPQQPISQQQQQQQQQQQKQQQQQQQQILQQILQQQL
IPCRDVVLQQHSIAYGSSQV
LQQSTYQLVQQLCCQQLWQIPEQSRCQAIHNVVHAIILHQ
QQQQQQQQQQPLSQVSFQQ
PQQQYPSQQGSFQPSQQNPQAQGSVQPQQLPQFEEIRNLA
LETLPAMCNVYIPPYCTIAP
VGIFGTN

2. Choose an allele:

HLA-DQ2.5 (A1*05:01/B1*02:01)
HLA-DQ2.5 (A1*05:01/B1*02:01)
HLA-DQ8.1 (A1*03:01/B1*03:02)

or upload a file with proteins in fasta format

Choose File No file chosen

3. Choose deamidation position

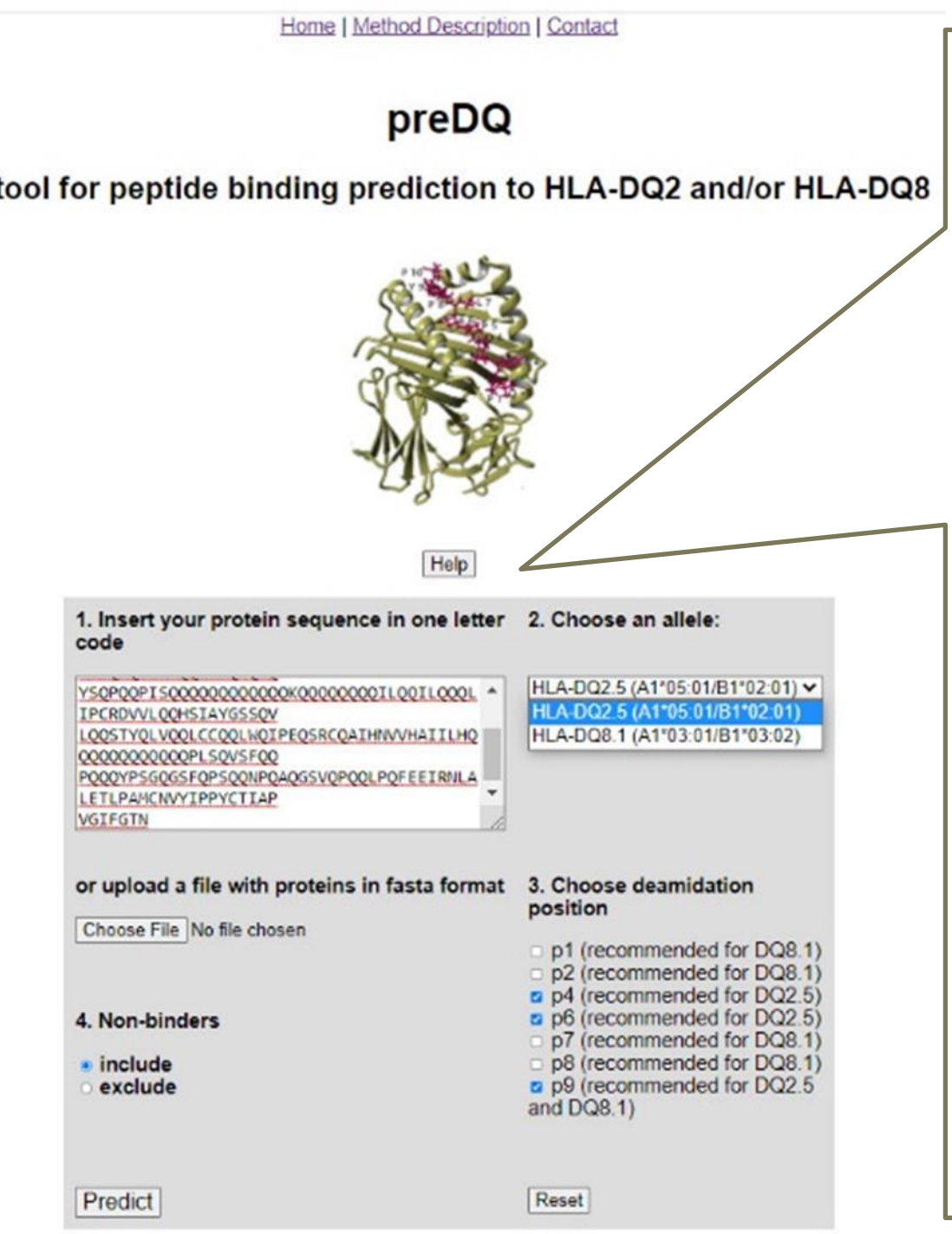
- p1 (recommended for DQ8.1)
- p2 (recommended for DQ8.1)
- p4 (recommended for DQ2.5)
- p6 (recommended for DQ2.5)
- p7 (recommended for DQ8.1)
- p8 (recommended for DQ8.1)
- p9 (recommended for DQ2.5 and DQ8.1)

4. Non-binders

- include
- exclude

Predict

Reset



PreDQ accepts queries of a single protein (one-letter code) or a batch of proteins in fasta format.

The data input consists of the following steps:

1. Insert the query protein sequence in one-letter code or upload a file with proteins in a fasta format.

2. Choose an allele.

3. Choose the position for deamidation in the peptide.

For HLA-DQ2.5, the recommended positions are p4, p6 and p9.

For HLA-DQ8.1, the recommended positions are p1, p2, p7, p8 and p9.

5. Start the prediction.

The preDQ prediction consists of several steps (See Method Description) and presents the results as a table. The table could be exported in csv format.

Home | Method Description | Contact

preDQ

a tool for peptide binding prediction to HLA-DQ2 and/or HLA-DQ8

Help

1. Insert your protein sequence in one letter code

YSQPQQPISQQQQQQQQQQQQKQQQQQQQQILQQILQQQL
IPCRDVVLQQHSIAYGSSQV
LQQSTYQLVQQLCCQQLWQIPEQSRCQAIHNVVHAIILHQ
QQQQQQQQQQPLSQVSFQQ
PQQQYPSQQGSFQPSQQNPQAQGSVQPQQLPQFEEIRNLA
LETLPAMCNVYIPPYCTIAP
VGIFGTN

2. Choose an allele:

HLA-DQ2.5 (A1*05:01/B1*02:01)
HLA-DQ2.5 (A1*05:01/B1*02:01)
HLA-DQ8.1 (A1*03:01/B1*03:02)

or upload a file with proteins in fasta format

Choose File No file chosen

3. Choose deamidation position

- p1 (recommended for DQ8.1)
- p2 (recommended for DQ8.1)
- p4 (recommended for DQ2.5)
- p6 (recommended for DQ2.5)
- p7 (recommended for DQ8.1)
- p8 (recommended for DQ8.1)
- p9 (recommended for DQ2.5 and DQ8.1)

4. Non-binders

- include
- exclude

Predict

Reset

The prediction algorithm of preDQ consists of several steps:

1. PreDQ converts the query protein into a set of overlapping nonamers.
2. The nonamers are deamidated according to the selection made by the user.
3. The nonamers are compared to the nonamers from the positive and negative sets and if any coincides, a link to a reference appears. The negative sets were generated by a total combination of the non-preferred amino acids at all 9 positions from the binding core for HLA-DQ2.5 according to Koşaloğlu-Yalçın et al., 2021 [1] and for HLA-DQ8.1 according to Tran et al., 2021 [2].
4. The nonamers are encoded binary and the binding affinity/class membership of each nonamer is predicted by four ligand-based models (Logo-based classification, Random forest regression, SVM regression and Xgboost regression) and one structure-based model (Docking-based quantitative matrix) (Figure 1).

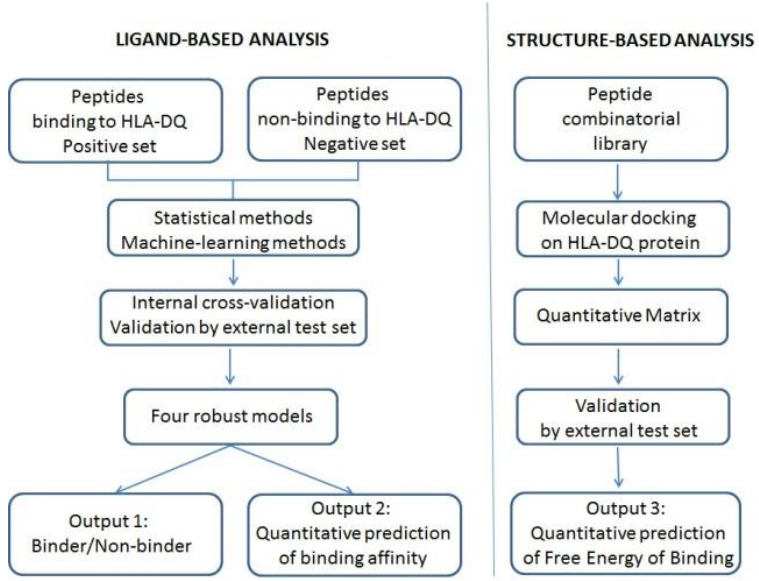


Figure 1. The ligand-based models are derived by statistical and machine-learning methods applied to training sets of known binders to HLA-DQ (positive set) and known non-binders (negative set). The models are validated by external test sets. The models give quantitative predictions of the binding affinities as pIC50 (-logIC50) and classify the given peptide as binder or non-binder on the basis of a predefined threshold. The structure-based model is a quantitative matrix derived by molecular docking of combinatorial peptide library on HLA-DQ proteins (6mfg for HLA-DQ2.5 and 2nna for HLA-DQ8.1). It classifies the given peptide as binder or non-binder on the basis of quantitative prediction.

5. The nonamers and their predicted binding affinities (as pIC50) and class membership are arranged in a table at the result page. If any of the nonamers is a known binder or non-binder, a link to the source is shown.

References

1. Koşaloğlu-Yalçın Z, Sidney J, Chronister W, Peters B, Sette A. Comparison of HLA ligand elution data and binding predictions reveals varying prediction performance for the multiple motifs recognized by HLA-DQ2.5. Immunology. 2021;162(2):235-247.
2. Tran MT, Faridi P, Lim JJ, et al. T cell receptor recognition of hybrid insulin peptides bound to HLA-DQ8. Nat Commun. 2021;12(1):5110.

Home | Method Description | Contact

preDQ

a tool for peptide binding prediction to HLA-DQ2 and/or HLA-DQ8

Help

1. Insert your protein sequence in one letter code

YSQPQQPISQQQQQQQQQQQQKQQQQQQQQILQQILQQQL
IPCRDVVLQQHSIAYGSSQV
LQQSTYQLVQQLCCQQLNQIPEQSRCQAIHNVVHAIILHQ
QQQQQQQQQQPLSQVSFQQ
PQQQYPSQQGSFQPSQQNPQAQGSVQPQQLPQFEEIRNLA
LETLPAMCNVYIPPYCTIAP
VGIFGTN

2. Choose an allele:

HLA-DQ2.5 (A1*05:01/B1*02:01)
HLA-DQ2.5 (A1*05:01/B1*02:01)
HLA-DQ8.1 (A1*03:01/B1*03:02)

or upload a file with proteins in fasta format

Choose File No file chosen

3. Choose deamidation position

- [ ] p1 (recommended for DQ8.1)
- [ ] p2 (recommended for DQ8.1)
- [x] p4 (recommended for DQ2.5)
- [x] p6 (recommended for DQ2.5)
- [ ] p7 (recommended for DQ8.1)
- [ ] p8 (recommended for DQ8.1)
- [x] p9 (recommended for DQ2.5 and DQ8.1)

4. Non-binders

- (•) include
- ( ) exclude

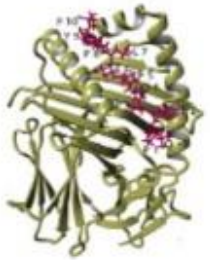
Predict

Reset

Home | Method Description | Contact

preDQ

a tool for peptide binding prediction to HLA-DQ2 and/or HLA-DQ8



Please write your name, email address and question below.

Name:

Subject:

Email address:

Message:

Submit

preDQ

preDQ

a tool for peptide binding prediction to HLA-DQ2 and/or HLA-DQ8

Output Description

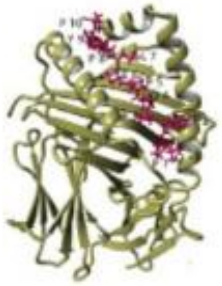
Exporting Results

Query sequence:				Allele:			
MKTFLILALLAIVATTARIAVRVPVPQLQPQNPSQQQPQEQVPLVQQQQFPGQQQPFPPQ QPYPQPQPFPSQQPYLQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQLPYPQPQPFRPQQPYPQSQPQ YSQPQQPISQQQQQQQQQQQQQKQQQQQQQQQILQQILQQQLIPCRDVVLQQHSIAYGSSQV LQQSTYQLVQQLCCQQLWQIPEQSRCQAIHNVVHAIILHQQQQQQQQQQQQPLSQVSFQQ PQQQYPSGQGSFQPSQQNPQAQGSVQPQQLPQFEEIRNLALETLPAMCNVYIPPYCTIAP VGIFGTN			DQ2.5			
Position	**Nonamers**	**Docking-based Prediction**	**Predict_by_Logo_model**	**pIC50_by_SVM**	**pIC50_by_XGboost**	**pIC50_by_RF**	**Known binder ID / Non-binder**
	cutoff	**0**	**0.5**	**5.6**	**5.8**	**6.0**	
25	VPQLQPQNP	known binder					78557
35	QQQPQEQVP	known binder					78493
37	QPQEQVPLV	known binder					78493
61	QPYPQPQPF	known binder					78495
63	YPQPQPFPS	known binder					78495
66	PQPFPSQQP	known binder					78495
77	QLQPFPQPE	known binder					38891 38864 51507 234152 109516 51517 38889 38892 234153 38887

preDQ

preDQ

a tool for peptide binding prediction to HLA-DQ2 and/or HLA-DQ8

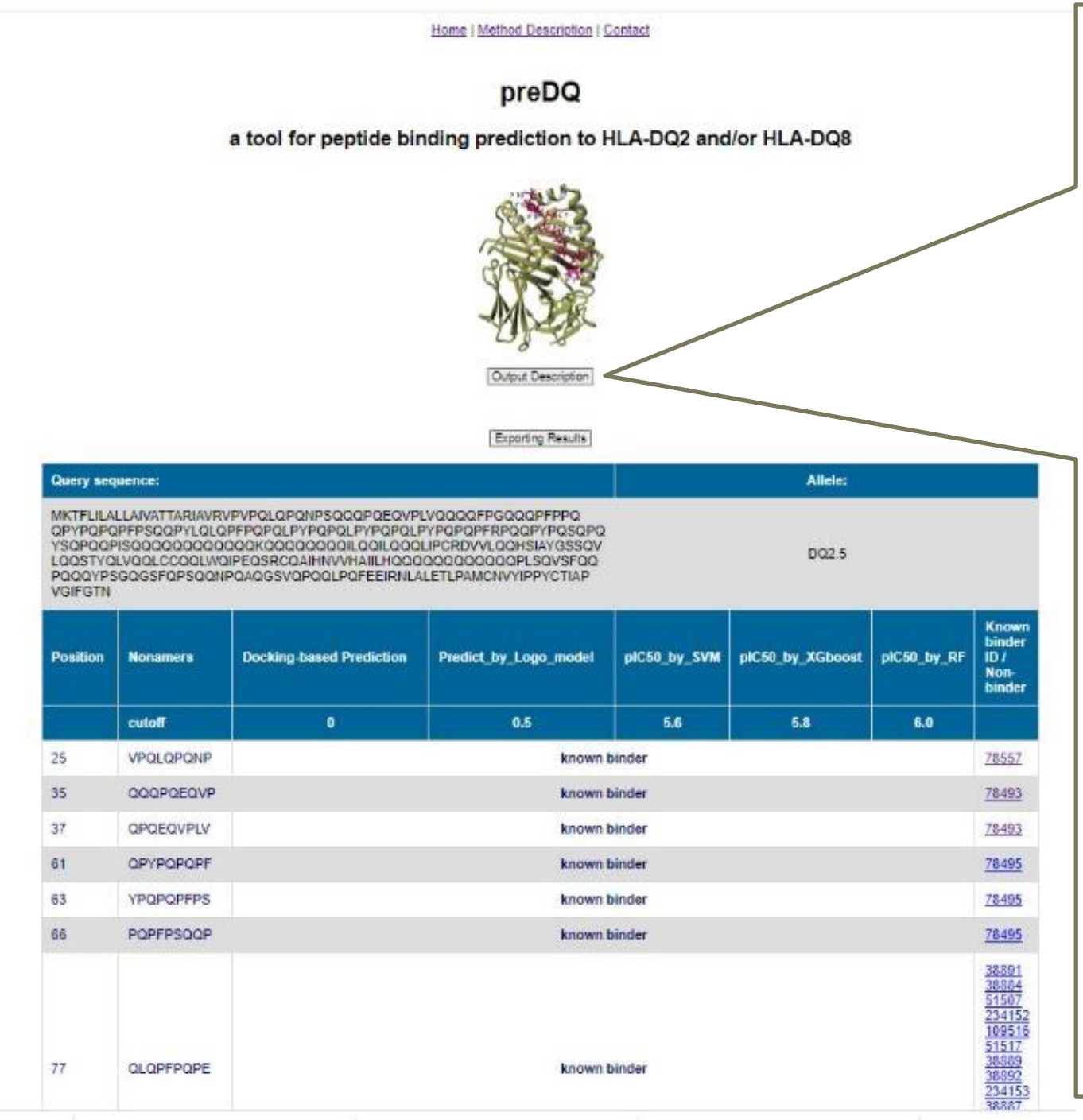


Output Description

Exporting Results

Query sequence:	Allele:
MKTFLILALLAIVATTARIAVRVPVPQLQPQNPSQQQPQEQVPLVQQQQFPGQQQPFPPQ QPYPQPQPFPSQQPYLQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPFRPQQPYPQSQPQ YSQPQQPISQQQQQQQQQQQQKQQQQQQQQILQQILQQQLIPCRDVVLQQHSIAYGSSQV LQQSTYQLVQQLCCQQLWQIPEQSRCQAIHNVVHAIILHQQQQQQQQQQQQPLSQVSFQQ PQQQYPSGQGSFQPSQQNPQAQGSVQPQQLPQFEEIRNLALETLPAMCNVYIPPYCTIAP VGIFGTN	DQ8.1

Position	Nonamers	Docking-based Prediction	Predict_by_Logo_model	pIC50_by_SVM	pIC50_by_XGboost	pIC50_by_RF	Known binder ID / Non-binder
	cutoff	0	0.5	5.6	5.5	5.6	
1	MKTFLILAL	0.157	0	5.546	5.088	6.135	
2	KTFLILALL	0.0	0	5.674	5.961	6.46	
3	TFLILALLA	-0.719	0	6.25	5.562	5.808	
4	FLILALLAI	0.0	0	6.172	6.076	5.208	



Home | Method Description | Contact

preDQ

a tool for peptide binding prediction to HLA-DQ2 and/or HLA-DQ8

Output Description

Exporting Results

Query sequence:				Allele:			
MKTFLILALLAIVATTARIAVRVPVPQLQPQNPSQQQPQEQVPLVQQQQFPGQQQPFPPQ QPYPQPQPFPSQQPYLQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPFRPQQPYPQSQPQ YSQPQQPISQQQQQQQQQQQQKQQQQQQQQILQQILQQQLIPCRDVVLQQHSIAYGSSQV LQQSTYQLVQQLCCQQLWQIPEQSRCQAIHNVVHAIILHQQQQQQQQQQQQQPLSQVSFQQ PQQQYPSGQGSFQPSQQNPQAQGSVQPQQLPQFEEIRNLALETLPAMCNVYIPPYCTIAP VGIFGTN				DQ2.5			
Position	**Nonamers**	**Docking-based Prediction**	**Predict_by_Logo_model**	**pIC50_by_SVM**	**pIC50_by_XGboost**	**pIC50_by_RF**	**Known binder ID / Non-binder**
	cutoff	**0**	**0.5**	**5.6**	**5.8**	**6.0**	
25	VPQLQPQNP	known binder					78557
35	QQQPQEQVP	known binder					78493
37	QPQEQVPLV	known binder					78493
61	QPYPQPQPF	known binder					78495
63	YPQPQPFPS	known binder					78495
66	PQPFPSQQP	known binder					78495
77	QLQPFPQPE	known binder					38891 38884 51507 234152 109516 51517 38889 38892 234153 38887

The preDQ results are presented in a table with 8 columns.

The column 'Position' contains the number of the first peptide residue in the query protein sequence.

The column 'Nonamers' contains the sequence of the nonamer binding core.

PreDQ uses five models to predict binding to DQ alleles: docking-based QM, logo-based classification, Random Forest regression, SVM regression and xgboost regression. The results of the predictive models are presented in 5 columns.

PreDQ distinguishes binders from non-binders by cutoffs. The cutoff for each of the models is presented in the second row of the table. A nonamer with a predicted value lower than the cutoff is predicted as a non-binder. The predicted values for binders are given in red.

If a nonamer is a known binder or non-binder to HLA-DQ, a link to the source appears in column 'Known binder/non-binder'.

The table could be exported as a csv table.