Contract NP/EFSA/NIF/2023/02

# preDQ – Update & Upgrade

## preDQ v2 Questions & Answers

Irini Doytchinova, Ivan Dimitrov, Mariyana Atanasova
Medical University of Sofia, Bulgaria

**Ad-hoc Meeting with GMO Industry Representatives**
**27 May 2026**

# Scenario 1

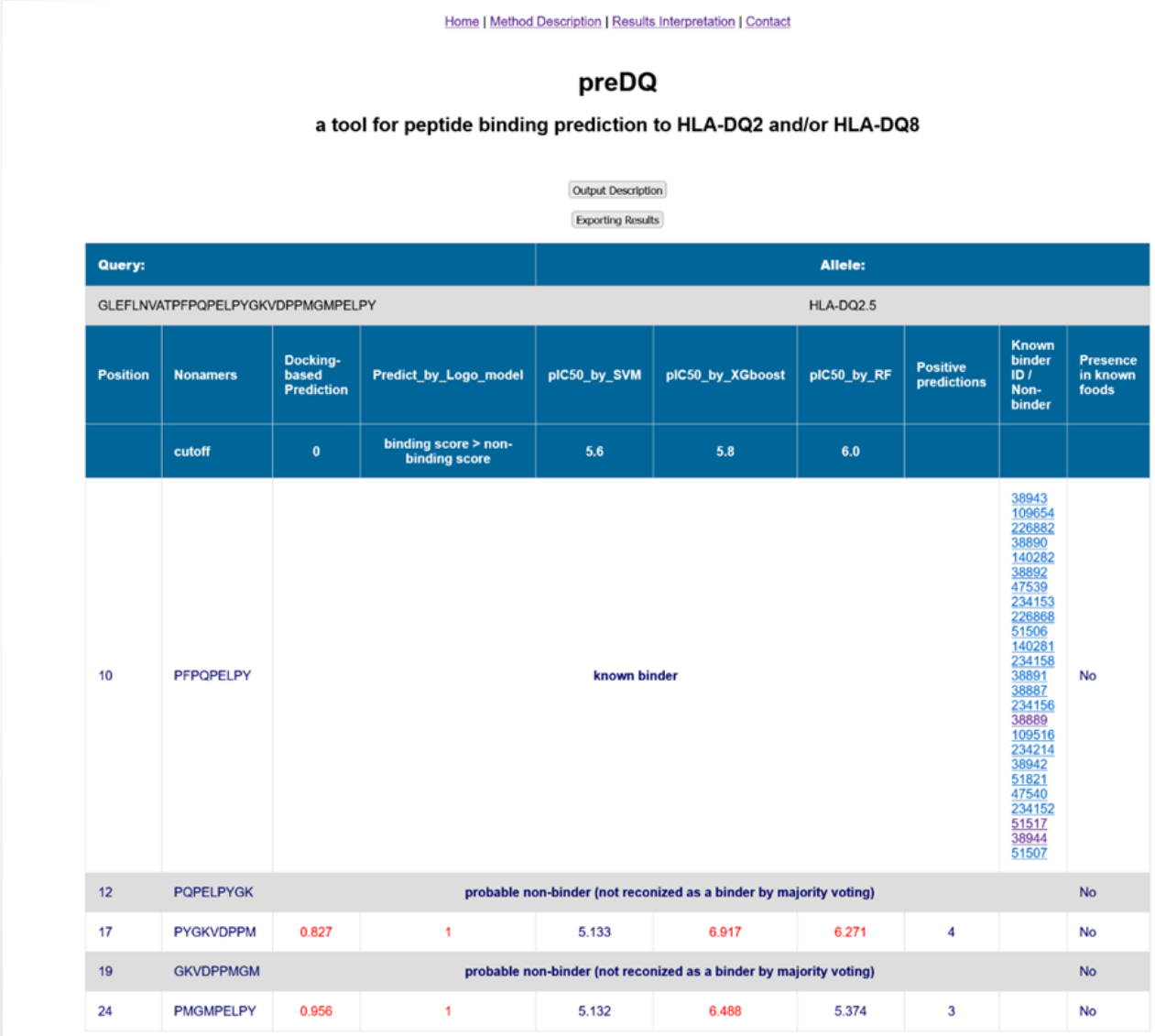
The 9mer/motif searches were performed without considering protein knowledge.

No exclusion criteria was considered.

preDQ helped eliminating some false positives from a potato protein, but still retained two false positives for both spinach and maize proteins.

| Protein | Accession | Perfect Match | Motif Match | Partial Match | preDQ result |
|---|---|---|---|---|---|
| Chicken myoglobin | NP_001161224 | 0 | 0 | 0 | N/A |
| Beef myosin heavy chain | BAB40921.2 | 0 | 0 | 0 | N/A |
| Beef myosin-2 | NP_001159699.1 | 0 | 0 | 0 | N/A |
| Beef Actin | AAI34666.1 | 0 | 0 | 0 | N/A |
| Chicken myoglobin heavy chain 1C | NP_001107181.2 | 0 | 0 | 0 | N/A |
| Potato Induced stolen tip protein TUB8 | P33191.1 | 0 | 10 | 0 | Non-binders |
| *Arabidopsis* Thioredoxin | KAG7651943.1 | 0 | 2 | 0 | 1 binder |
| Chicken actin | NP_001385203.1 | 0 | 0 | 0 | N/A |
| Spinach Rubisco | NP_054944 | 0 | 2 | 0 | 2 binders |
| Soybean 7S globin | AAA33947 | 0 | 0 | 0 | N/A |
| Soybean 11S globin | AAA33964 | 0 | 0 | 0 | N/A |
| Maize 10 kD zein | Q41881 | 0 | 4 | 0 | 2 binders |
| Tomato ripen associate membrane protein | CAA52068 | 0 | 0 | 0 | N/A |

# preDQ v2 Upgrades



Home | Method Description | Results Interpretation | Contact

**preDQ**

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

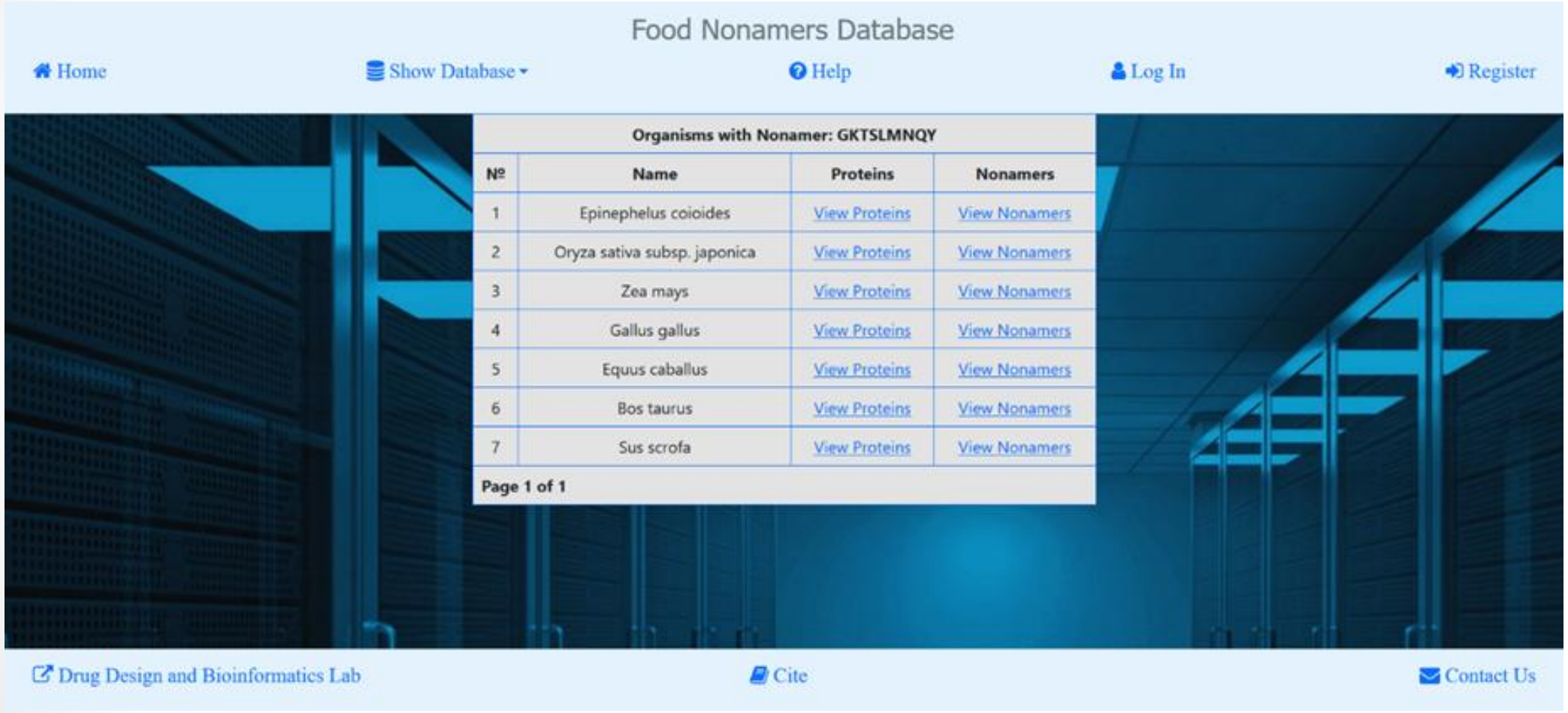
Output Description

Exporting Results

| Query: | | | | Allele: | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| GLEFLNVATPFPQPELPYGKVDPPMGMPELPY | | | | HLA-DQ2.5 | | | | | |
| Position | Nonamers | Docking-based Prediction | Predict_by_Logo_model | pIC50_by_SVM | pIC50_by_XGboost | pIC50_by_RF | Positive predictions | Known binder ID / Non-binder | Presence in known foods |
| | cutoff | 0 | binding score > non-binding score | 5.6 | 5.8 | 6.0 | | | |
| 10 | PFPQPELPY | known binder | | | | | | 38943 109654 226882 38890 140282 38892 47539 234153 226868 51506 140281 234158 38891 38887 234156 38889 109516 234214 38942 51821 47540 234152 51517 38944 51507 | No |
| 12 | PQPELPYGK | probable non-binder (not reconized as a binder by majority voting) | | | | | | | No |
| 17 | PYGKVDPPM | 0.827 | 1 | 5.133 | 6.917 | 6.271 | 4 | | No |
| 19 | GKVDPPMGM | probable non-binder (not reconized as a binder by majority voting) | | | | | | | No |
| 24 | PMGMPELPY | 0.956 | 1 | 5.132 | 6.488 | 5.374 | 3 | | No |

4. A new column, Presence in Known Foods, has been added to the Results table. If a tested nonamer is predicted as a binder, it is checked against the Food Nonamers Database. If a match is found, this column provides additional information about the associated proteins and organisms

# Food Nonamers Database




The final version of the database contains 389 organisms, 98 643 proteins, and 20 000 210 nonamers.

Table 2. The 9mer/motif searches were performed without considering protein knowledge. Exclusion criteria were used to screen the 9mers. preDQ helped eliminate some false positives, but still retained one false positive for the maize protein.

| protein | Motif match 9mer | Exclusion criteria Result | preDQ result |
|---|---|---|---|
| Potato Induced stolen tip protein TUB8 | EVTKVEEPA | No Hazard | No need |
| | TKVEEPAPV | No Hazard | No need |
| | AALVAEEPA | No Hazard | No need |
| | LVAEEPAAA | No Hazard | No need |
| | VAAPVEEPA | No Hazard | No need |
| | APVEEPAAA | No Hazard | No need |
| | EPAAAEEPA | No Hazard | No need |
| | AAAEEPAAA | No Hazard | No need |
| | EPAAAEEPV | No Hazard | No need |
| | AAAEEPVAA | No Hazard | No need |
| *Arabidopsis* Thioredoxin | SAVKFQFPV | No Hazard | No need |
| | VKFQFPVRR | No Hazard | No need |
| Spinach Rubisco | KEIKFEFPA | No Hazard | No need |
| | IKFEFPAMD | No Hazard | No need |
| Maize 10 kD zein | IMLQQQLPF | Step 2 needed | binder |
| | LQQQLPFMF | Step 2 needed | Non-binders |
| | PPMFLQQPF | No Hazard | No need |
| | MFLQQPFVG | Step 2 needed | Non-binders |

Given that maize is considered safe for consumption by individuals with celiac disease, the framework's flowchart indicates that a sequence identity search for Maize 10 kD zein is unnecessary. Consequently, the testing results demonstrate no false positives when the framework's flowchart is strictly adhered to.

# preDQ

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

Output Description

Download CSV

| Query: | | | | Allele: | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| IMLQQQLPF | | | | HLA-DQ2.5 | | | | | |
| Position | Nonamer | Docking-based Prediction | Predict_by_Logo_model | pIC50_by_SVM | pIC50_by_XGboost | pIC50_by_RF | Positive predictions | Known binder ID / Non-binder | Presence in known foods |
| | cutoff | 0 | binding score > non-binding score | 5.6 | 5.8 | 6.0 | | | |
| 1 | IMLQQQLPF | probable non-binder (not reconized as a binder by majority voting) | | | | | | | No |

**Page 1 of 1.**

# preDQ

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

Output Description

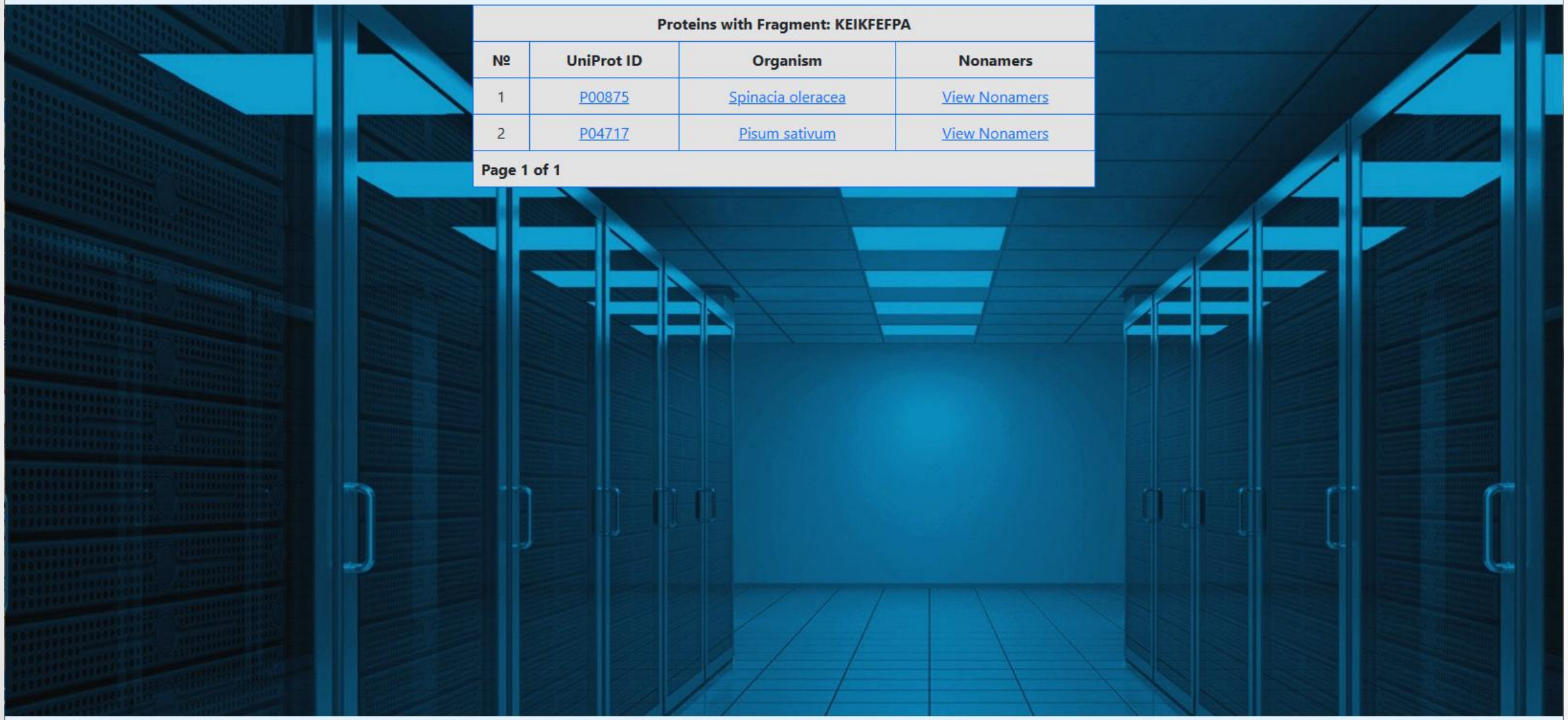
Download CSV

| Query: | | | | Allele: | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| KEIKFEFPA | | | | HLA-DQ2.5 | | | | | |
| Position | Nonamer | Docking-based Prediction | Predict_by_Logo_model | pIC50_by_SVM | pIC50_by_XGboost | pIC50_by_RF | Positive predictions | Known binder ID / Non-binder | Presence in known foods |
| | cutoff | 0 | binding score > non-binding score | 5.6 | 5.8 | 6.0 | | | |
| 1 | KEIKFEFPA | 0.481 | 1 | 6.378 | 5.552 | 5.298 | 3 | | Yes |

Page 1 of 1.

# Food Nonamers Database

Home | Show Database | Help | Log In | Register



| Proteins with Fragment: KEIKFEFPA | | | |
|---|---|---|---|
| № | UniProt ID | Organism | Nonamers |
| 1 | P00875 | Spinacia oleracea | View Nonamers |
| 2 | P04717 | Pisum sativum | View Nonamers |
| Page 1 of 1 | | | |

Drug Design and Bioinformatics Lab | Cite | Contact Us

# preDQ

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

Output Description

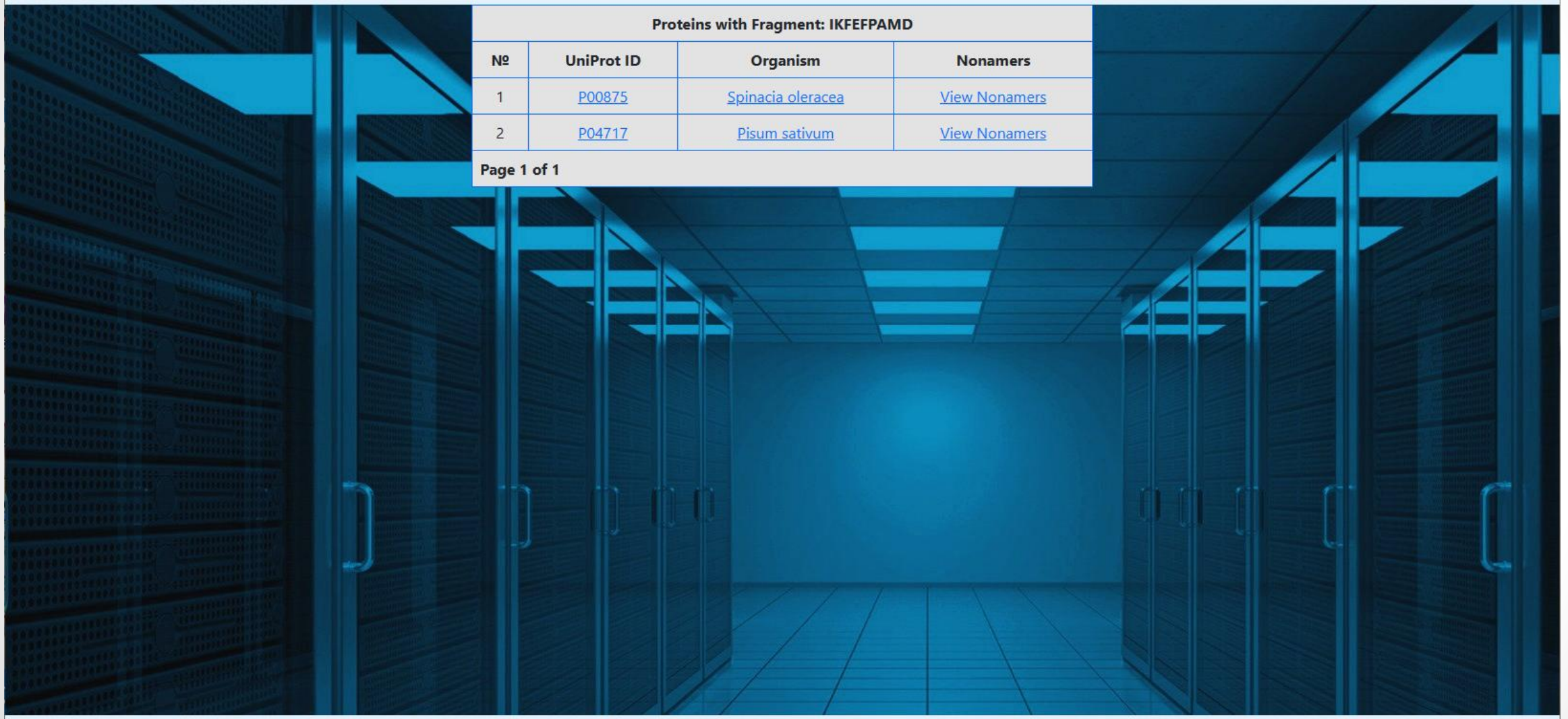
Download CSV

| Query: | | | | Allele: | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| IKFEFPAMD | | | | HLA-DQ2.5 | | | | | |
| Position | Nonamer | Docking-based Prediction | Predict_by_Logo_model | pIC50_by_SVM | pIC50_by_XGboost | pIC50_by_RF | Positive predictions | Known binder ID / Non-binder | Presence in known foods |
| | cutoff | 0 | binding score > non-binding score | 5.6 | 5.8 | 6.0 | | | |
| 1 | IKFEFPAMD | -0.029 | 1 | 6.452 | 6.526 | 6.272 | 4 | | Yes |

Page 1 of 1.

# Food Nonamers Database

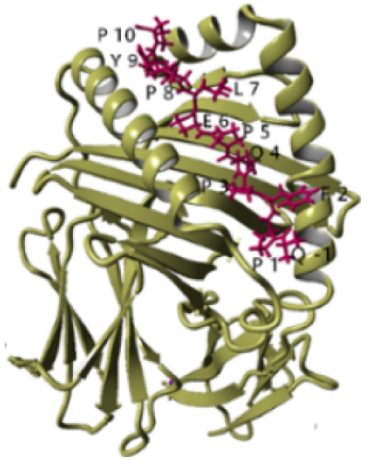




**Proteins with Fragment: IKFEFPAMD**

| № | UniProt ID | Organism | Nonamers |
|---|---|---|---|
| 1 | P00875 | Spinacia oleracea | View Nonamers |
| 2 | P04717 | Pisum sativum | View Nonamers |

**Page 1 of 1**



**1. Insert a single sequence in one CAPITAL letter code. It should contain at least 9 amino acids. The result is displayed in a table.**

**2. Choose an allele:**

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

**For prediction of multiple sequences upload a file in fasta format. The result is in a .csv file.**

Choose File No file chosen

References on deamidation at these positions for DQ2.8 epitope?

**3. Choose deamidation position**

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

**4. Search only peptides containing specific motif (recommended for celiac peptides and HLA-DQ2.5 allele)**

- [x] Q/E-X1-P-X2 motif
- [x] N/D-X1-P-X2 motif

DLPY is only presented in the epitopes from bacteria. Reference on N deamidated to D via deamidation?

**5. Choose the output format**

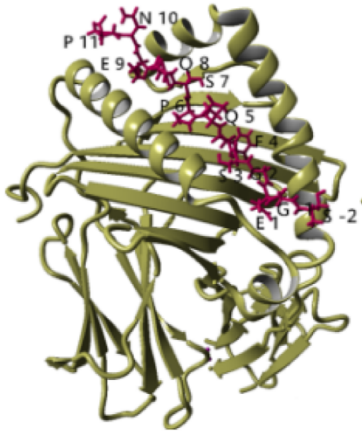
- [x] Majority voting (at least 3 models out of 5 with positive prediction)

**6. Non-binders**

- ( ) include
- (•) exclude

Predict

Reset form



Doytchinova I., Atanasova M., Fernandez A., Moreno J., Koning F., Dimitrov I. Modeling peptide-protein interactions by a logo-based method: Application in peptide-HLA binding predictions. Molecules 2024, 29, 284.

**Table 5.** Quantitative matrix (QM) for binding nonamers to HLA-DQ8.1.

| AA | p1 | p2 | p3 | p4 | p5 | p6 | p7 | p8 | p9 |
|---|---|---|---|---|---|---|---|---|---|
| Ala | 0.068 | 0.373 | 0.494 | 0.699 | 0.262 | 0.531 | 0.378 | 0.152 | 0.031 |
| Arg | -0.111 | 0.004 | -0.085 | -0.122 | -0.027 | -0.122 | -0.111 | 0.047 | -0.096 |
| Asn | -0.122 | -0.053 | -0.106 | -0.096 | -0.074 | -0.117 | -0.038 | -0.080 | -0.106 |
| Asp | 0.036 | -0.038 | 0.020 | -0.011 | -0.022 | -0.038 | 0.083 | -0.038 | 0.489 |
| Cys | -0.122 | -0.122 | -0.122 | -0.090 | -0.122 | -0.122 | -0.122 | -0.122 | -0.122 |
| Gln | -0.106 | 0.004 | -0.090 | -0.106 | -0.022 | -0.111 | 0.052 | -0.074 | 0.010 |
| Glu | 0.157 | 0.215 | -0.022 | -0.053 | 0.194 | 0.062 | 0.494 | 0.383 | 0.878 |
| Gly | -0.106 | -0.053 | 0.257 | -0.074 | -0.032 | -0.111 | -0.090 | 0.041 | 0.004 |
| His | -0.101 | -0.006 | -0.096 | -0.122 | -0.069 | -0.117 | -0.074 | -0.069 | -0.117 |
| Ile | 0.199 | -0.006 | -0.080 | 0.189 | -0.017 | 0.052 | -0.043 | -0.069 | -0.090 |
| Leu | 0.089 | -0.074 | -0.069 | 0.004 | -0.027 | -0.064 | -0.006 | 0.073 | -0.080 |
| Lys | -0.111 | 0.004 | -0.106 | -0.122 | 0.052 | -0.122 | -0.090 | 0.010 | -0.117 |
| Met | -0.106 | -0.101 | -0.101 | -0.101 | -0.090 | -0.111 | -0.117 | -0.111 | -0.096 |
| Phe | 0.099 | -0.074 | -0.048 | -0.069 | -0.074 | -0.122 | -0.111 | -0.074 | -0.111 |
| Pro | 0.099 | -0.117 | -0.038 | -0.122 | -0.032 | 0.115 | 0.020 | 0.162 | -0.122 |
| Ser | -0.048 | 0.015 | 0.226 | 0.057 | 0.062 | 0.089 | -0.006 | 0.004 | -0.011 |
| Thr | -0.074 | 0.068 | 0.147 | 0.010 | 0.052 | 0.078 | -0.053 | -0.059 | -0.080 |
| Trp | -0.038 | -0.106 | -0.111 | -0.101 | -0.106 | -0.122 | -0.106 | -0.117 | -0.106 |
| Tyr | 0.089 | -0.011 | -0.074 | -0.122 | -0.085 | -0.111 | -0.064 | -0.069 | -0.101 |
| Val | 0.210 | 0.078 | 0.004 | 0.352 | 0.178 | 0.462 | 0.004 | 0.010 | -0.059 |

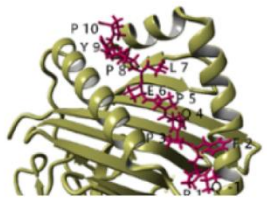
Atanasova M., Dimitrov I., Fernandez A., Moreno J., Koning F., Doytchinova I. Assessment of novel proteins triggering celiac disease via docking-based approach. Molecules 2024, 29, 138.

Docking-based quantitative matrix for α-gliadin peptide library docked in HLA-DQ8.1

| aa | p1 | p2 | p3 | p4 | p5 | p6 | p7 | p8 | p9 |
|---|---|---|---|---|---|---|---|---|---|
| A | 0.069 | 0.105 | 0.006 | 0.006 | 0.168 | 0.033 | 0.123 | 0.186 | 0.069 |
| C | -0.012 | -0.084 | 0.024 | -0.021 | -0.021 | -0.012 | -0.084 | 0.024 | 0.006 |
| D | 0.096 | -0.166 | 0.042 | 0.006 | 0.087 | -0.021 | 0.033 | 0.123 | 0.069 |
| E | **0.078** | -0.193 | 0.078 | **0.015** | 0.078 | **-0.003** | 0.051 | 0.132 | **0.078** |
| F | 0.150 | -0.778 | 0.087 | 0.078 | 0.177 | -0.679 | 0.186 | 0.159 | 0.015 |
| G | 0.006 | 0.078 | 0.069 | 0.060 | 0.060 | 0.006 | 0.141 | 0.150 | -0.012 |
| H | 0.060 | -0.652 | 0.096 | 0.042 | 0.105 | -0.202 | 0.141 | 0.150 | 0.051 |
| I | 0.069 | -0.238 | -0.057 | 0.006 | 0.051 | 0.033 | 0.069 | 0.060 | 0.060 |
| K | 0.033 | -0.120 | -0.039 | -0.012 | -0.039 | -0.084 | 0.024 | 0.006 | -0.066 |
| L | 0.060 | -0.247 | -0.129 | 0.024 | 0.114 | -0.111 | 0.069 | 0.123 | 0.078 |
| M | 0.015 | -0.247 | -0.039 | -0.066 | 0.078 | -0.102 | -0.066 | -0.283 | -0.003 |
| N | 0.069 | -0.129 | 0.069 | 0.060 | 0.114 | -0.039 | 0.042 | -0.229 | 0.105 |
| P | -0.111 | 0.015 | 0.195 | 0.033 | 0.042 | 0.078 | 0.195 | 0.195 | -0.211 |
| Q | **-0.003** | -0.202 | 0.042 | **-0.021** | 0.078 | **-0.075** | 0.006 | 0.078 | **0.033** |
| R | 0.078 | -0.120 | -0.003 | 0.033 | 0.078 | -0.030 | 0.123 | 0.060 | 0.006 |
| S | 0.096 | -0.057 | 0.078 | -0.021 | 0.069 | -0.003 | 0.078 | 0.096 | 0.042 |
| T | 0.015 | -0.102 | 0.006 | 0.006 | 0.069 | 0.024 | 0.024 | 0.042 | 0.078 |
| V | 0.033 | -0.048 | 0.123 | -0.012 | 0.123 | -0.021 | 0.033 | 0.159 | 0.015 |
| W | 0.141 | -0.598 | 0.186 | -0.003 | 0.096 | -0.607 | 0.222 | 0.087 | -0.066 |
| Y | 0.096 | -0.643 | 0.096 | -0.120 | 0.096 | -0.697 | 0.168 | 0.123 | -0.066 |

# preDQ

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



Help

Example in plain format

**1. Insert a single sequence in one CAPITAL letter code. It should contain at least 9 amino acids. The result is displayed in a table.**

```
MAAKMLALFALLALCASATSATHIPGHLPPVMPLGTMNPCMQ
YCMMQQGLASLMACPSLMLQQLLALPLQTMPVMMPQMMTPNM
MSPLMMPSMMSPMVLPSMMSQMMMPQCHCDAVSQIMLQQQLP
FMFNPMAMTIPPMFLQQPFVGAAF
```

**2. Choose an allele:**

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

**For prediction of multiple sequences upload a file in fasta format. The result is in a .csv file.**

Choose File No file chosen

**3. Choose deamidation position**

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

**4. Search only peptides containing specific motif (recommended for celiac peptides and HLA-DQ2.5 allele)**

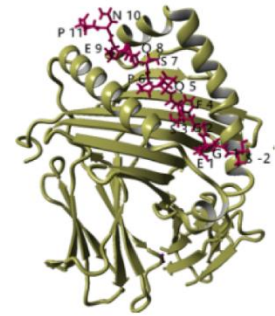
- [ ] Q/E-X1-P-X2 motif
- [ ] N/D-X1-P-X2 motif

**5. Choose the output format**

**6. Non-binders**

EFSA statistical models

**Server Error (500)**

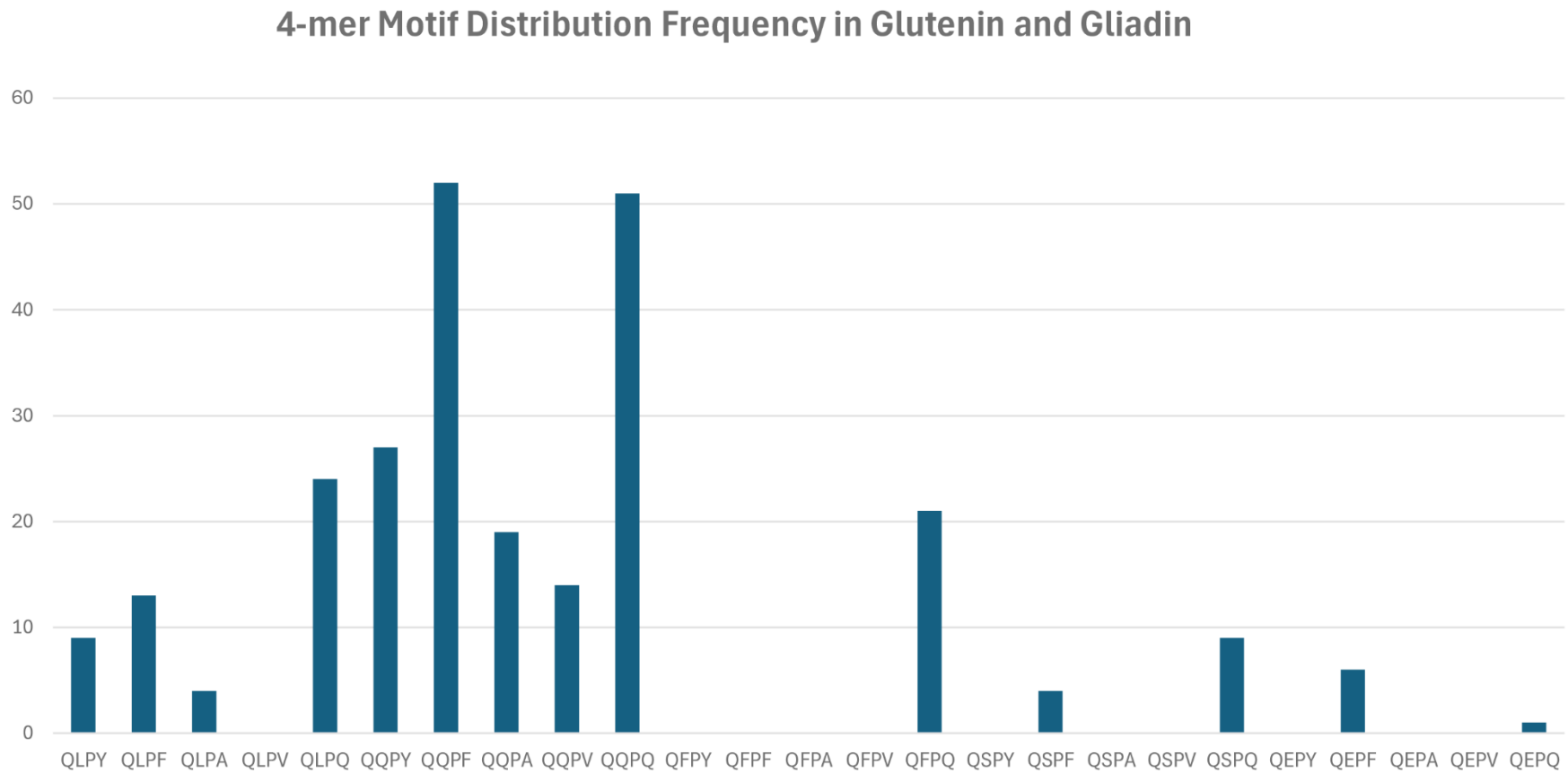


There is a server error (500) when choosing HLA-DQ2.5 allele and corresponding deamidation position, but no motifs.

# Distribution Frequency of the 25 4-aa Motifs in 72 Known Glutenin and Gliadin from Wheat, Oats, Barley and Rye (Source AOL)

| | | | |
|---|---|---|---|
| E | L | P | Y |
| Q | Q | | F |
| | F | | A |
| | S | | V |
| | E | | Q |

Only 28 possible motifs are present considering deamidation




**Suggestion**: Only using 28 validated motifs plus “DLPY” from bacteria in the sequence identity search in Step 1

# Predicted positive 9-mer peptides include positively charged amino acids and a “PP” duplex, both recognized for interfering with binding

| Query: | Allele: | | | | | | |
|---|---|---|---|---|---|---|---|
| NP_054944.1 | HLA-DQ2.5 | | | | | | |
| Position | Nonamer | Docking-based Prediction | Predict_by_Logo_model | pIC50_by_SVM | pIC50_by_XGboost | pIC50_by_RF | Positive_predictions |
| | cutoff | 0 | binding score > non-binding score | 5.6 | 5.8 | 6 | |
| 1 | MSPETETKA | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 2 | SPQTETKAS | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 3 | PQTETKASV | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 4 | QTETKASVG | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 5 | TETKASVGF | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 6 | ETKASVGFK | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 7 | TKASVGFKA | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 8 | KASVGFKAG | probable non-binder (not recognized as binder by majority voting) | | | | | |
| → 9 | ASVGFKAGV | 0.887 | 1 | 5.793 | 5.001 | 5.696 | 3 |
| 10 | SVGFKAGVK | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 11 | VGFKAGVKD | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 12 | GFKAGVKDY | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 13 | FKAGVKDYK | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 14 | KAGVKDYKL | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 15 | AGVKDYKLT | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 16 | GVKDYKLTY | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 17 | VKDYKLTYY | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 18 | KDYKLTYYT | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 19 | DYKLTYYTP | -0.148 | 0 | 5.774 | 6.06 | 6.123 | 3 |
| 20 | YKLTYYTPE | 0.02 | 1 | 5.745 | 5.259 | 7.026 | 4 |
| 21 | KLTYYTPEY | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 22 | LTYYTPEYE | 0.55 | 1 | 7.021 | 5.912 | 6.884 | 5 |
| 23 | TYYTPEYET | 0.05 | 1 | 6.986 | 5.946 | 5.99 | 4 |
| 24 | YYTPEYETL | 0.421 | 1 | 7.116 | 5.889 | 5.621 | 4 |
| 25 | YTPEYETLD | 0.194 | 1 | 5.91 | 5.822 | 4.882 | 4 |
| 26 | TPEYETLDT | probable non-binder (not recognized as binder by majority voting) | | | | | |
| 27 | PEYETLDTD | 0.07 | 1 | 6.428 | 5.947 | 6.487 | 5 |
| 28 | EYETLDTDI | -0.94 | 0 | 5.819 | 7.391 | 6.251 | 3 |
| 29 | YETLDTDIL | 0.565 | 1 | 6.791 | 6.242 | 6.142 | 5 |
| 30 | ETLDTDILA | probable non-binder (not recognized as binder by majority voting) | | | | | |

# Large-Scale Characterization of Natural Ligands Explains the Unique Gluten-Binding Properties of HLA-DQ2[1]

**Dariusz Stepniak,* Martina Wiesner,* Arnoud H. de Ru,*† Antonis K. Moustakas,‡ Jan Wouter Drijfhout,* George K. Papadopoulos,§ Peter A. van Veelen,*† and Frits Koning[2]***

**Celiac disease is an enteropathy caused by intolerance to dietary gluten. The disorder is strongly associated with DQA1*0501/DQB1*0201 (HLA-DQ2) as ~95% of celiac patients express this molecule. HLA-DQ2 has unique Ag-binding properties that allow it to present a diverse set of gluten peptides to gluten-reactive $CD4^+$ T cells so instigating an inflammatory reaction. Previous work has indicated that the presence of negatively charged amino acids within gluten peptides is required for specific binding. This, however, only partly explains the scale of the interaction. We have now characterized 432 natural ligands of HLA-DQ2 representing length variants of 155 distinct sequences. The sequences were aligned and the binding cores were inferred. Analysis of the amino acid distribution of these cores demonstrated that negatively charged residues in HLA-DQ2-bound peptides are favored at virtually all positions. This contrasts with a more restricted presence of such amino acids in T cell epitopes from gluten. Yet, HLA-DQ2 was also found to display a strong preference for proline at several anchor and nonanchor positions that largely match the position of proline in gluten T cell epitopes. Consequently, the bias for proline at p6 and p8 facilitates the enzymatic conversion of glutamine into glutamic acid in gluten peptides at p4 and p6, two important anchor sites. These observations provide new insights in the unique ability of HLA-DQ2 to bind a large repertoire of glutamine- and proline-rich gluten peptides. This knowledge may be an important asset in the development of future treatment strategies.** ***The Journal of Immunology,*** **2008, 180: 3268–3278.**



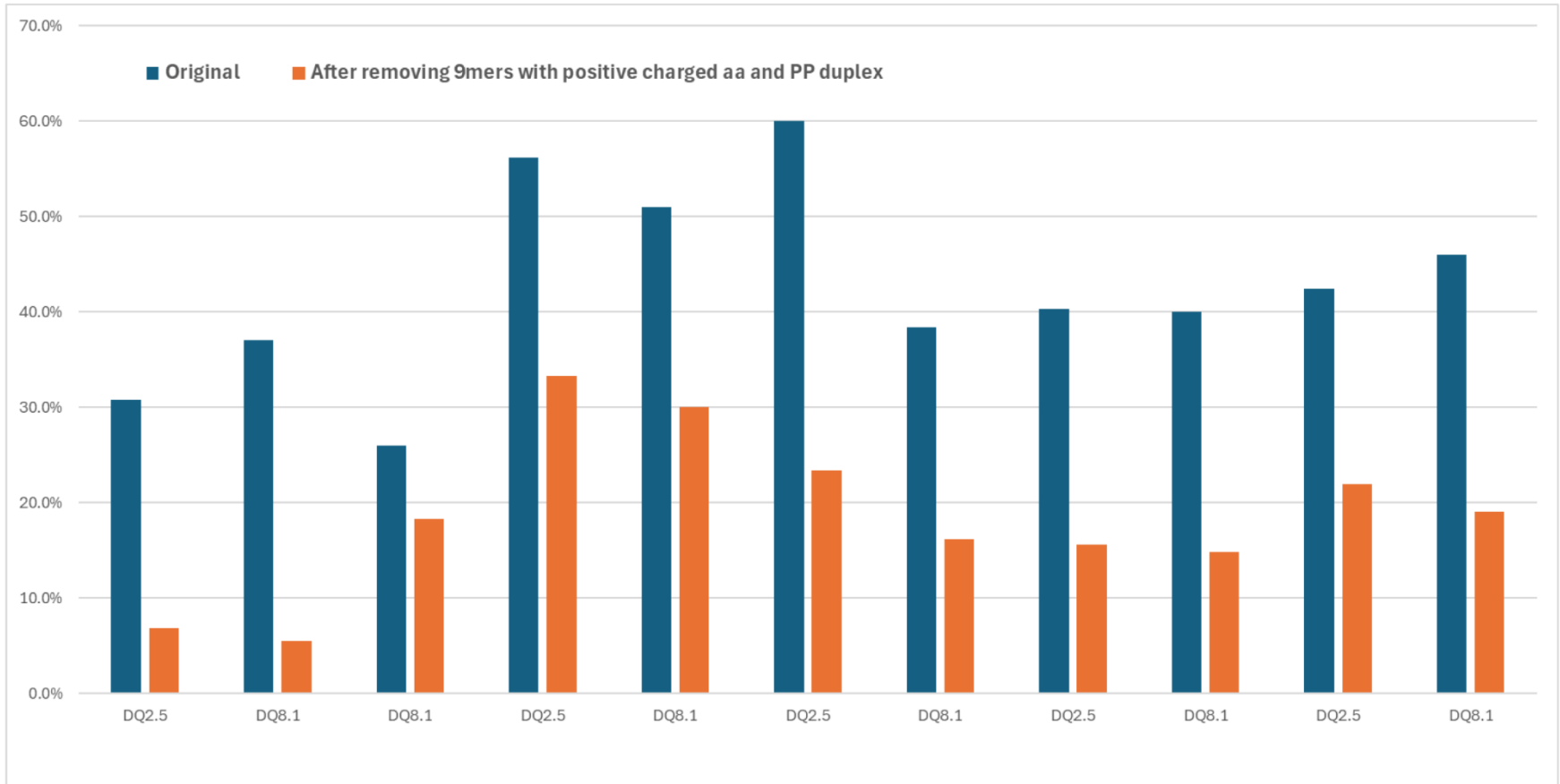
125 binders to HLA-DQ2.5
  28 binding cores contain K, R and/or H (22%)
Of them:
16 binding cores contain K (13%)
10 binding cores contain R (8%)
  5 binding cores contain H (4%)
  3 binding cores contain K and H or K and R (2%)



Table I. *Overview of peptides eluted from HLA-DQ2[a]*

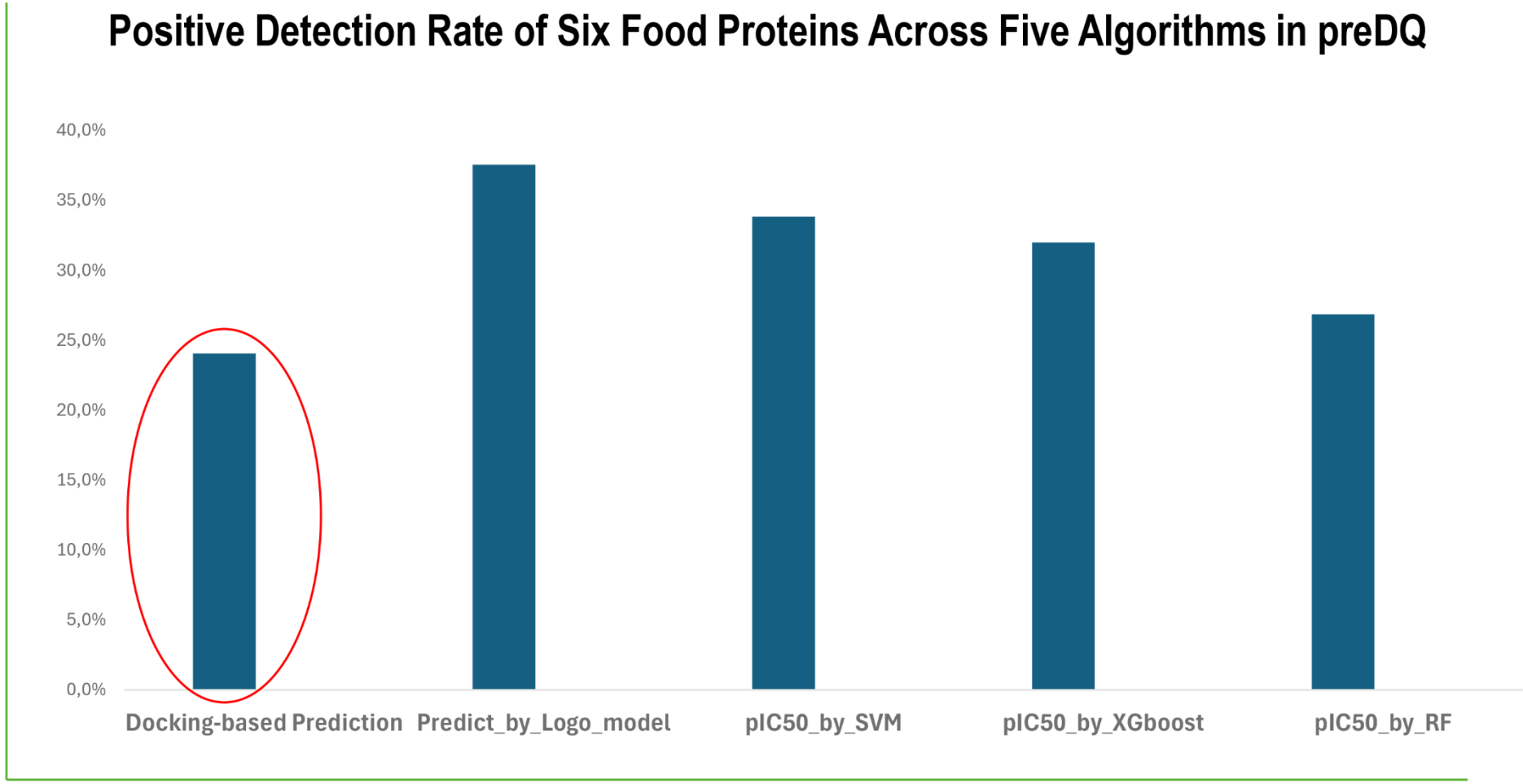
| Peptide Source | Sequence | Length Variants |
|---|---|---|
| Actin | KEITALAPSTMKIK | 3 |
| α tubulin | APVISAEKAYHEQ | 2 |
| α tubulin | DYEEVGADSADGEDEG | 1 |
| $\beta_2$-microglobulin | HPSDIEVDLLKN | 2 |
| β tubulin | DATAEEEEDFGEEAEEE | 4 |
| β tubulin | DPTGTYHGDSDLQLDR | 2 |
| β-actin | EEEIAALVVDN | 4 |
| β-actin | YVALDFEQEMA | 3 |
| β-actin | VPIYEGYALPHAI | 1 |
| Calcium-binding protein (stromal cell-derived factor 4) | DGIVTAEELESYMDP | 3 |
| Calmodulin | LQDMINEVDADGN | 4 |
| CD20 | IEIIPIQEEEEEE | 15 |
| CD20 | IKEEVVGLTETSSQPK | 1 |
| CD74 | MRMATPLLMQALPM | 22 |
| Desmocollin type 4 | ADGYSADLPLPLPIR | 2 |
| Endoplasmin precursor | ESDDEAAVEEEEEEK | 1 |
| Eukaryotic translation initiation factor 4A | DIETFYNTTVEEMP | 1 |
| Glyceraldehyde 3-phosphate dehydrogenase | GKVDIVAINDPFID | 2 |
| Glyceraldehyde 3-phosphate dehydrogenase | DIQVVAINDPFID | 1 |
| Golgi phosphoprotein | EFEEAEQVREENLPDE | 1 |
| Golgi-localized phosphoprotein 130 kDa | VDEQYQEEAEEEVQED | 1 |
| GP73 Golgi phosphoprotein 2 | AALSVSQENPEMEGPE | 2 |
| Heat shock cognate 71-kDa protein | FDVSILTIEDGIFE | 1 |
| Heparan sulphate 2 sulphotransferase | GPRQDATLDEEEDM | 5 |
| HLA DM α-chain | ADWAQEQGDAPAILFDK | 2 |
| HSP-90 | DPTADDTSAAVTEEMPP | 2 |
| HSP-90 | IDEDEVAAEEPN | 3 |
| HSP-70 | KSINPDEAVAYG | 1 |
| HSP-70 | TPSYVAFTDTERLIG | 2 |
| Hyaluronan-binding protein | YPEDEVGQEDEAESD | 1 |
| Hypoxanthine-guanine phosphoribosyltransferase | SPGVVISDDEPGYD | 2 |
| Ig κ-chain V-III region POM | TLTISSLQSEDFAV | 1 |
| IgE FcR 2 | NLNGLQADLSSFK | 7 |
| IgG | YPSDIAVEWESNGQPE | 1 |
| IgG | FTLTISSLQSEDF | 1 |
| Importin β-1 subunit | EAADVADDQEEPAT | 1 |
| Initiation factor 4A-I | GPDGMEPEGVIESNWNEIV | 1 |
| Integral membrane protein 2A | DSDPAAIIHDFE | 1 |
| Integrin $\beta_1$ | DDLENVKSLGTDLMNEM | 1 |
| IFN-γ receptor α | APTSFGYDKPHVLVDL | 1 |
| IL-21 receptor | GSPLAGLDMDTFDSG | 1 |
| IL-2 receptor β-chain precursor | SPFPSSSFSPGGLAPEISPLE | 1 |
| IL-4 receptor α chain precursor | APVECEEEEEVEEE | 1 |
| ITM2A | REDLVAVEEIRD | 1 |
| ITM2C protein | LPQTYLIQEEMV | 1 |

# Positive Hit Rate Reduction in Six Food Proteins after Removing 9-mer Peptides Containing Positively Charged Amino Acid and "PP" Duplex

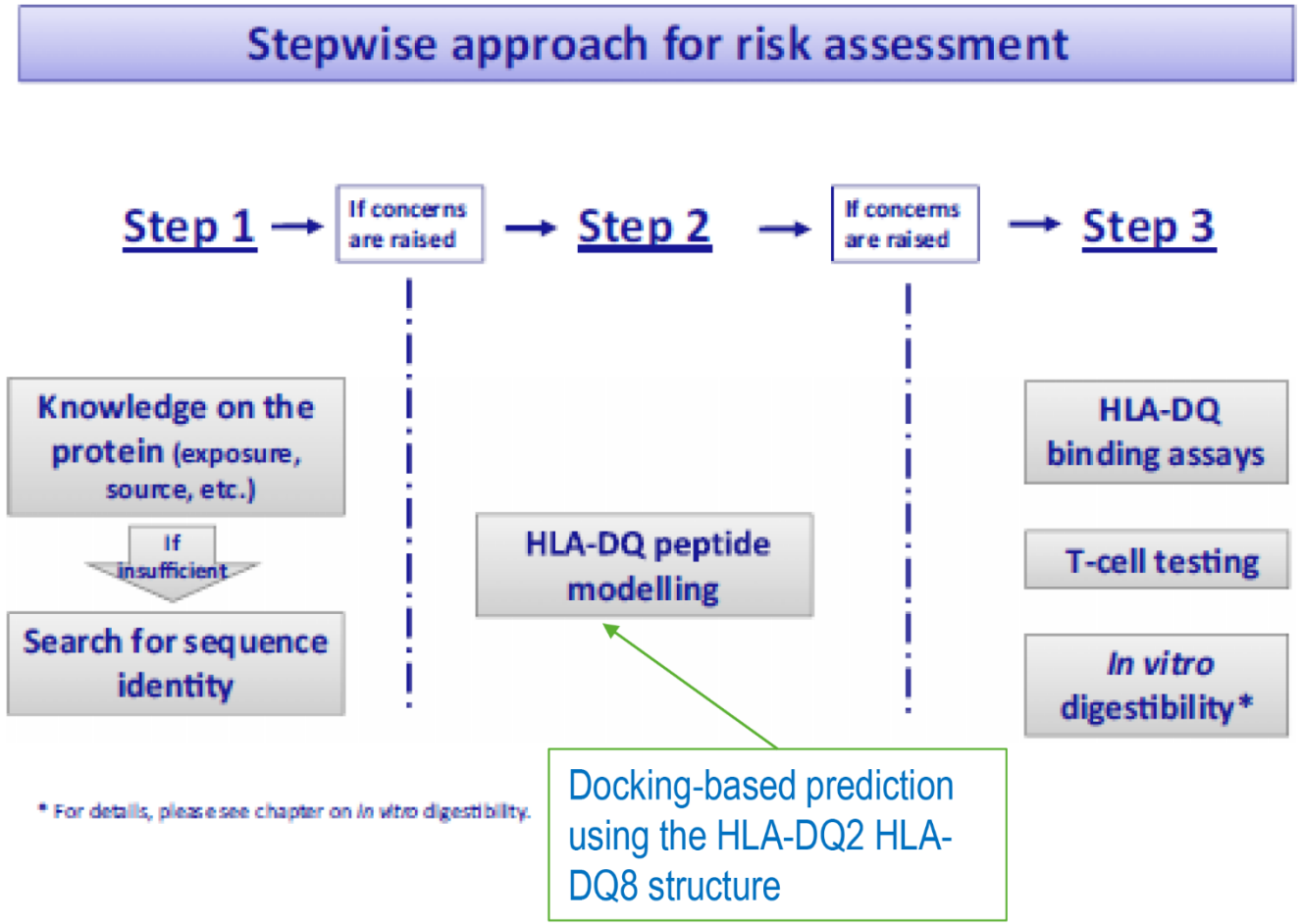



**Recommendation**: Exclude 9-mer peptides with positively charged amino acids at positions P1, P4, P6, P7, P9, sequences containing "PP," and serine at position 8 prior to model analysis.

Ensemble method increase positive rate




**Recommendation**: Concentrate on docking-based predictions utilizing the structures of HLA-DQ2 and HLA-DQ8, consistent with Step 2.




**EFSA journal. Naegeli, et al (2017)**

## Detection of Recognized Epitopes in the Food 9-mer Database (partial results)

| | Epitope Name | Sequence | Food 9-mer Database | Organism |
|---|---|---|---|---|
| 1 | DQ2.5-glia-α1a | PFPQP**E**LPY | PFPQP**Q**LPY | Wheat |
| 2 | DQ2.5-glia-α1b | PYPQP**E**LPY | PYPQP**Q**LPY | Wheat |
| 3 | DQ2-P.fluor-α1a | PMPMPELPY | no | |
| 4 | DQ2-P.fluor-α1b | PMPLPDLPY | no | |
| 5 | DQ2.5-glia-α2 | PQP**E**LPYPQ | PQP**Q**LPYPQ | Wheat |
| 6 | DQ2-P.fluor-α2a | PMPELPYPA | no | |
| 7 | DQ2-P.fluor-α2b | PLPELPYPA | no | |
| 8 | DQ2-B.copro-α2a | PLPDLPYPV | no | |
| 9 | DQ2-P.aerug-α2a | VQSELPYPE | no | |
| 10 | DQ2.5-glia-α3var | FLPELPYPQ | no | |
| 11 | DQ2.5-glia-α3 | FRP**E**QPYPQ | FRP**Q**QPYPQ | Wheat |
| 12 | DQ2.5-glia-α4 | PQPEQPYPQ | no | |
| 13 | DQ2.5-glia-α5 | PQPEYPQPQ | no | |
| 14 | DQ2.5-glia-γ1 | PQQSFP**E**QQ | PQQSFP**Q**QQ | Wheat |

**Recommendation**: Add deamidated peptides to the Food 9-mer Database.

# Questions on preDQ v.2

How to make a final decision based on the analysis results. Specifically:
1) How to balance the weight of the five model results;
2) We are unsure how to further take into account the other analysis results involving positive predictions, known binders and presence in known foods.

When we tried to analyse the Vip3Aa19 protein under the HLA-DQ8.1 model, the software stopped working when we selected p4.

EQNQVLNDV

# Methodology (3)

For the 9-mers selected with motif, the following conditions were set:

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

**3. Choose deamidation position**

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

**4. Search only peptides containing specific motif (recommended for celiac peptides and HLA-DQ2.5 allele)**

- [x] Q/E-X1-P-X2 motif
- [x] N/D-X1-P-X2 motif

**5. Choose the output format**

- [x] Majority voting (at least 3 models out of 5 with positive prediction)

**6. Non-binders**

- ( ) include
- (•) exclude

For the 9-mers selected with partial match, the following conditions were set:

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

**3. Choose deamidation position**

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

**4. Search only peptides containing specific motif (recommended for celiac peptides and HLA-DQ2.5 allele)**

- [ ] Q/E-X1-P-X2 motif
- [ ] N/D-X1-P-X2 motif

**5. Choose the output format**

- [x] Majority voting (at least 3 models out of 5 with positive prediction)

**6. Non-binders**

- ( ) include
- (•) exclude

# Testing the preDQ Tool (2)

| Query: | | | | Allele: | | | | |
|---|---|---|---|---|---|---|---|---|
| ETSDIQEPY | | | | HLA-DQ2.5 | | | | |
| **Position** | **Nonamer** | **Docking-based Prediction** | **Predict_by_Logo_model** | **pIC50_by_SVM** | **pIC50_by_XGboost** | **pIC50_by_RF** | **Positive predictions** | **Known binder ID / Non-binder** |
| | **cutoff** | **0** | **binding score > non-binding score** | **5.6** | **5.8** | **6.0** | | |
| **Page 1 of 1.** | | | | | | | | |

- In this instance, no result was shown.
- This finding was considered an uncertain result.
- This problem was found only when looking for HLA-DQ2.5. The search for DQ8.1 using did not give uncertain results.

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

## 6. Non-binders

- (•) include
- ( ) exclude

Home | Method Description | Results Interpretation | Contact

# preDQ

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

Output Description

Download CSV

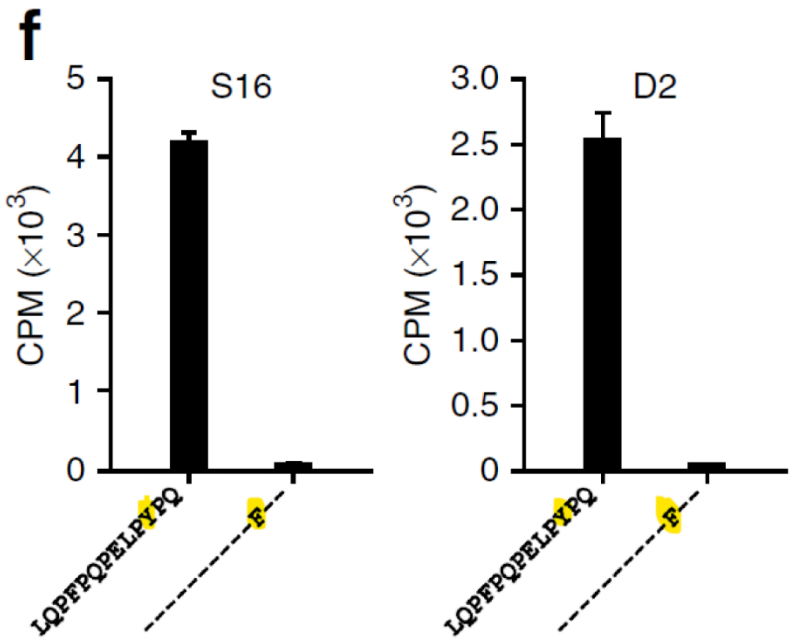
| Query: | | | | Allele: | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ETSDIQEPY | | | | HLA-DQ2.5 | | | | | |
| Position | Nonamer | Docking-based Prediction | Predict_by_Logo_model | pIC50_by_SVM | pIC50_by_XGboost | pIC50_by_RF | Positive predictions | Known binder ID / Non-binder | Presence in known foods |
| | cutoff | 0 | binding score > non-binding score | 5.6 | 5.8 | 6.0 | | | |
| 1 | ETSDIEEPY | non-binder | | | | | | *** | |

Page 1 of 1.

# Testing the preDQ Tool (3)

- Attempts to send 9-mers in multi-Fasta format using the [Choose File] button, always returned results for about 4 sequences, independent of the number of sequences sent.
- When checked independently, those 9-mer sequences with no result, which could be thught not to be binders, pasting them in the web tool, many of those 9-mers had a positive result.
- Therefore, this option cannot  be used.
- If this multi-Fasta function is used, this issue may inadvertently hide many results.

# Observations (1)




Petersen et al. 2014 showed that the 4 amino acid combination ELPF motif abolishes the reactivity of the celiac epitope.

PreDQ tool nevertheless uses the ELPF motif to rank a peptide as positive, suggesting this peptide as a concern for celiac patients.

| | D | E | F |
|---|---|---|---|
| | **Sequence** | **result** | **Docking** |
| | FLVQLPFHN | FLVELPFHD | 0.209 |

*Petersen et al 2014. T-cell receptor recognition of HLA-DQ2–gliadin complexes associated with celiac disease. Nature structural & Molecular biology VOLUME 21 NUMBER 5. p. 480-488*

nature
structural &
molecular biology

# T-cell receptor recognition of HLA-DQ2–gliadin complexes associated with celiac disease

Jan Petersen[1,8], Veronica Montserrat[2,8], Jorge R Mujico[2], Khai Lee Loh[1], Dennis X Beringer[1], Menno van Lummel[2], Allan Thompson[2], M Luisa Mearin[3], Joachim Schweizer[3], Yvonne Kooy-Wi Jeroen van Bergen[2], Jan W Drijfhout[2], Wan-Ting Kan[4], Nicole L La Gruta[4], Robert P Anderson[5], Frits Koning[2,9] & Jamie Rossjohn[1,6,7,9]

Celiac disease is a T cell–mediated disease induced by dietary gluten, a component of which is gliadin. 95% of i celiac disease carry the *HLA* (human leukocyte antigen)-*DQ2* locus. Here we determined the T-cell receptor (TC fine specificity of patient-derived T-cell clones specific for two epitopes from wheat gliadin, DQ2.5-glia-α1a and We determined the ternary structures of four distinct biased TCRs specific for those epitopes. All three TCRs spe glia-α2 docked centrally above HLA-DQ2, which together with mutagenesis and affinity measurements provide biased TCR usage. A non–germline encoded arginine residue within the CDR3β loop acted as the lynchpin withi docking footprint. Although the TCRs specific for DQ2.5-glia-α1a and DQ2.5-glia-α2 docked similarly, their int respective gliadin determinants differed markedly, thereby providing a basis for epitope specificity.

preDQ is not able to recognize T-cell epitopes because it is not trained to recognized T-cell epitopes.

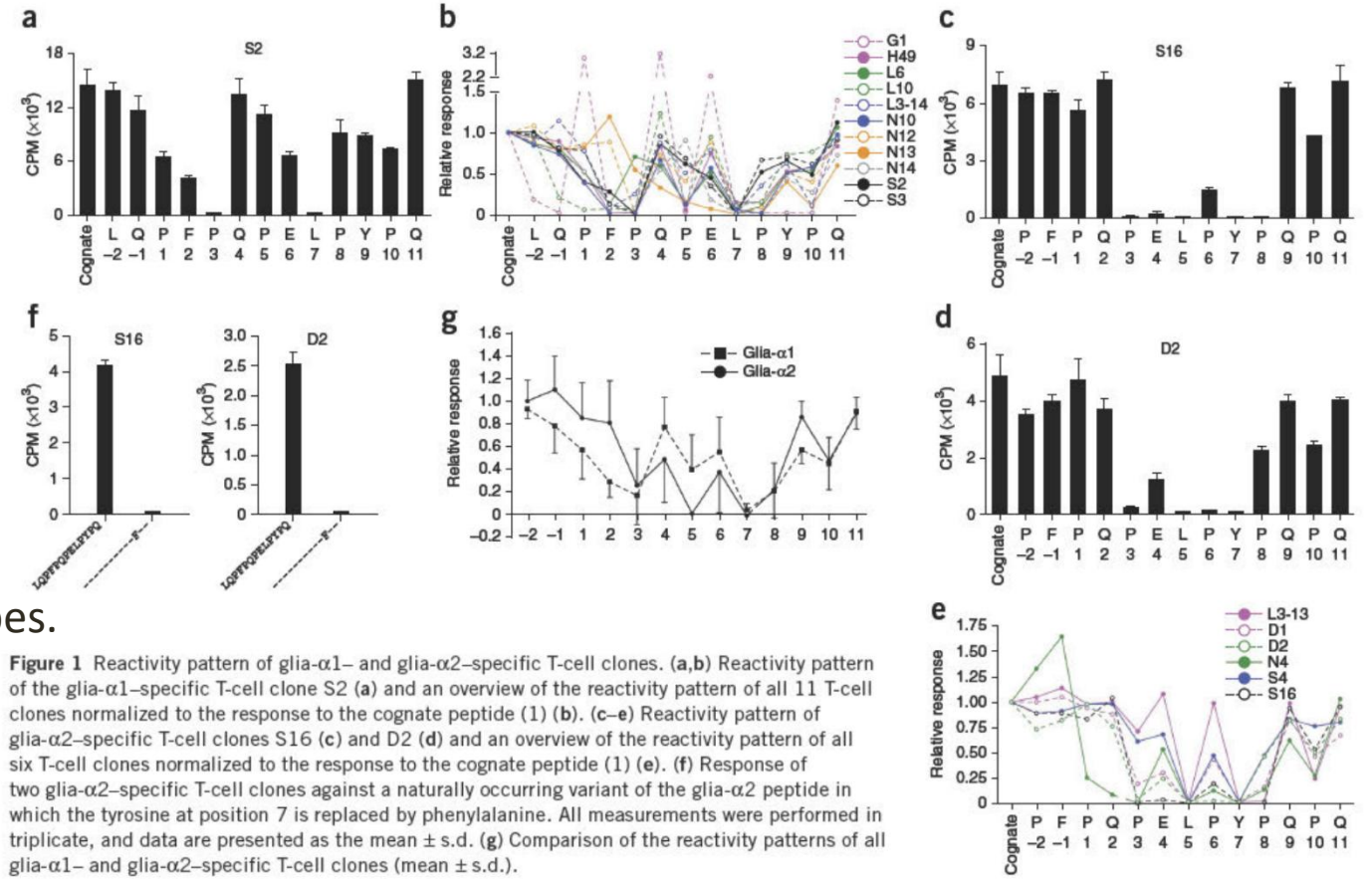



**Figure 1** Reactivity pattern of glia-α1– and glia-α2–specific T-cell clones. (**a**,**b**) Reactivity pattern of the glia-α1–specific T-cell clone S2 (**a**) and an overview of the reactivity pattern of all 11 T-cell clones normalized to the response to the cognate peptide (1) (**b**). (**c**–**e**) Reactivity pattern of glia-α2–specific T-cell clones S16 (**c**) and D2 (**d**) and an overview of the reactivity pattern of all six T-cell clones normalized to the response to the cognate peptide (1) (**e**). (**f**) Response of two glia-α2–specific T-cell clones against a naturally occurring variant of the glia-α2 peptide in which the tyrosine at position 7 is replaced by phenylalanine. All measurements were performed in triplicate, and data are presented as the mean ± s.d. (**g**) Comparison of the reactivity patterns of all glia-α1– and glia-α2–specific T-cell clones (mean ± s.d.).

# Observations (2)

| Test (clone 6) | CPM | SI | IFN-γ | IL-4 |
|---|---|---|---|---|
| T+APC only | 1053 (911) | — | 31 | 0 |
| PQPELPYPQPQLPY | 14764 (3870) | 14 | 232 | 0 |
| PQ**A**ELPYPQPQLPY | 343 (110) | <1 | 0 | 0 |
| PQPE**A**PYPQPQLPY | 244 (59) | <1 | 0 | 0 |
| PQPEL**A**YPQPQLPY | 182 (42) | <1 | 0 | 0 |
| PQPELP**A**PQPQLPY | 236 (25) | <1 | 0 | 0 |
| PQPELPY**A**QPQLPY | 211 (33) | <1 | 0 | 0 |
| PQPELPYP**A**PQLPY | 798 (123) | <1 | 31 | 0 |
| PQPELPYPQ**A**QLPY | 341 (70) | <1 | 0 | 0 |
| PQPELPYPQP**A**LPY | 14454 (197) | 14 | 205 | 0 |
| PQPELPYPQPQ**A**PY | 13681 (1209) | 13 | 225 | 0 |

Ellis et al. 2003 showed that alanine (A) substitutions on several central positions of the epitope abolish its reactivity.

preDQ is not able to recognize T-cell epitopes because it is not trained to recognized T-cell epitopes.

The preDQ tool ranks A in one or more of those positions as a highly relevant positive binder to HLA-DQ2 finding up to 5 positive predictions.

| | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|
| | **Sequence** | **result** | **Docking** | **logo** | **by_SVM** | **XGboost** | **by_RF** | **predictions** |
| | GSSQLPAAE | GSSELPAAE | 0.288 | 1 | 5.96 | 6.563 | 6.856 | 5 |

*Ellis HJ et al 2003. Investigation of the putative immunodominant T cell epitopes in coeliac disease. Gut. 52:2012-2017*

# Observations (3)

The tool identified a known binder (EGSTIPIQE) within human sequences. HLA-DQA1*03:01/DQB1*03:02, which is not related to celiac disease.

Does this mean that non-celiac related sequences were used in the training set of this tool?

Would adding known binders of non-celiac related HLAs to the negative training set improve prediction?

HLA-DQ8.1 (DQA1*03:01\DQB1*03:02) is associated with susceptibility to celiac disease.
8-11% of celiac patients carry this allele.

The models in preDQ were trained on peptides binding to HLA-DQ2.5 and HLA-DQ8.1, regardless of whether the peptides were associated with celiac disease.

The negative test sets include peptides - combinations of non-preferred amino acids at all 9 positions of the binding core. The pools contain > 20 mln non-binding nonamers.

# HLA-DQ8.1 (DQA1*03:01/DQB1*03:02)

Home | Method Description | Results Interpretation | Contact

## preDQ

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

Output Description

Download CSV

| Query: | | | | Allele: | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| EGSTIPIQE | | | | HLA-DQ8.1 | | | | | |
| Position | Nonamer | Docking-based Prediction | Predict_by_Logo_model | pIC50_by_SVM | pIC50_by_XGboost | pIC50_by_RF | Positive predictions | Known binder ID / Non-binder | Presence in known foods |
| | cutoff | 0 | binding score > non-binding score | 5.6 | 5.5 | 5.6 | | | |
| 1 | EGSTIPIQE | known binder | | | | | | 422699 | No |
| Page 1 of 1. | | | | | | | | | |

Pending Filters

Reset | Search

Filter Options

MHC

Epitope

- Any
- Linear peptide
  - Length
  - Sequence
- Discontinuous
- Non-peptidic

3D structure assays

Amino acid modification

Epitope Source

Organism

Ex: influenza, peanut

Antigen

Ex: core, capsid, myosin

Include related structure

Select multiple options

MHC Assay

Outcome: Positive

Current Filters: Epitope: 422699 | Include Positive Assays | No T cell assays | No B cell assays | MHC Restriction Type: MHC Assay Type

Epitopes (1) | Antigens (1) | Assays (1) | Receptors (0) | References (1)

T Cell Assays (0) | B Cell Assays (0) | MHC Ligand Assays (1)

IQ-API | Go To Records Starting At A,b GO | Export Results

1 Records Found | Page 1 of 1 | 25 Per Page

| IEDB ID | Reference | Epitope | Antigen Processing | MHC Restriction | Assay Description | Quantitative Measure |
|---|---|---|---|---|---|---|
| 2466854 | Jonathan A Trujillo; J Biol Chem 2014 | SCPREEEGSTIPIQE DYRKP + OX(C2) CD27 antigen isoform X1 (189-208) Homo sapiens (human) | Host:Homo sapiens (human). No immunization was performed | HLA-DQA1*03:01/DQB1*03:02 | cellular MHC/mass spectrometry ligand presentation **Positive** | |

1 Records Found | Page 1 of 1 | 25 Per Page

IQ-API | Go To Records Starting At A,b GO | Export Results

| MHC Ligand Assay | | |
|---|---|---|
| **Qualitative Measurement** | Positive | |
| **Method/Technique** | cellular MHC/mass spectrometry | OBI:0001489 |
| **Measurement of** | ligand presentation | |

| Antigen Presenting Cells | | |
|---|---|---|
| **Cell Tissue Type** | blood | UBERON:0000178 |
| **Cell Type** | B cell | CL:0000236 |
| **Cell Culture Conditions** | Cell Line / Clone (EBV transformed, B-LCL) | |

| MHC Allele | | |
|---|---|---|
| **MHC Allele Name** | HLA-DQA1*03:01/DQB1*03:02 | MRO:0001220 |
| **MHC Evidence Code** | Elution with MHC specific antibody | ECO:0008039 |