# PreDQ Celiac Disease Modeling Tool

- Representatives of BASF, Bayer, Corteva and Syngenta have performed testing of the PreDQ tool and would like to express the following concerns.
- Testing identified numerous false positives:
  - Proteins that are not expected to trigger a celiac response return multiple potential binding sites.
- A celiac peptide sequence with a single conservative substitution was not identified by the tool.
- Minimal documentation is available regarding the tool.

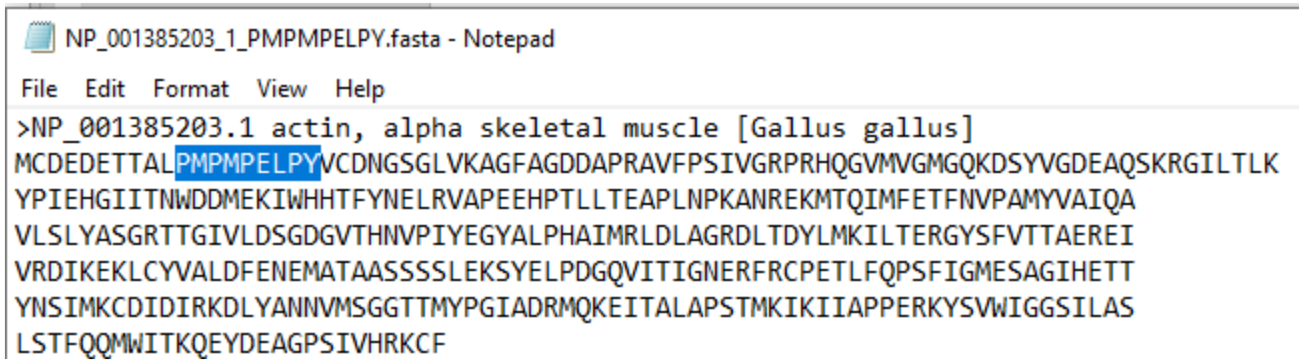
# Celiac Assessment Tool Testing Strategy

- High abundance proteins that are likely found in human diets were chosen as test sequences
- Proteins, rather than peptides, were used as this facilitates more rigorous testing of the tool
- Beef, Chicken and plant proteins were among the protein sources in the test set

| Protein description | DQ 2.5 | DQ 8.1 |
|---|---|---|
| BAB40921.2 myosin heavy chain 2x [*Bos taurus*] | 23 | 66 |
| NP_001159699.1 myosin-2 [*Bos taurus*] | 19 | 63 |
| AAI34666.1 Actin, alpha 1, skeletal muscle [*Bos taurus*] | 23 (20 known) | 8 |
| NP_001107181.2 myosin, heavy chain 1C, skeletal muscle (similar to human myosin, heavy chain 1, skeletal muscle, adult) [*Gallus gallus*] | 22 | 65 |
| NP_001385203.1 actin, alpha skeletal muscle [*Gallus gallus*] | 23 (20 known) | 8 |
| Q41881_MAIZE 10kDa zein (10kD delta zein) | 7 | |
| P33191.1\|TUB8_SOLTU RecName: Full=Induced stolen tip protein TUB8 | 7 | |
| KAG7651943.1 Thioredoxin [*Arabidopsis thaliana*] | 2 | 4 |

# Add Celiac Epitope into Test Protein

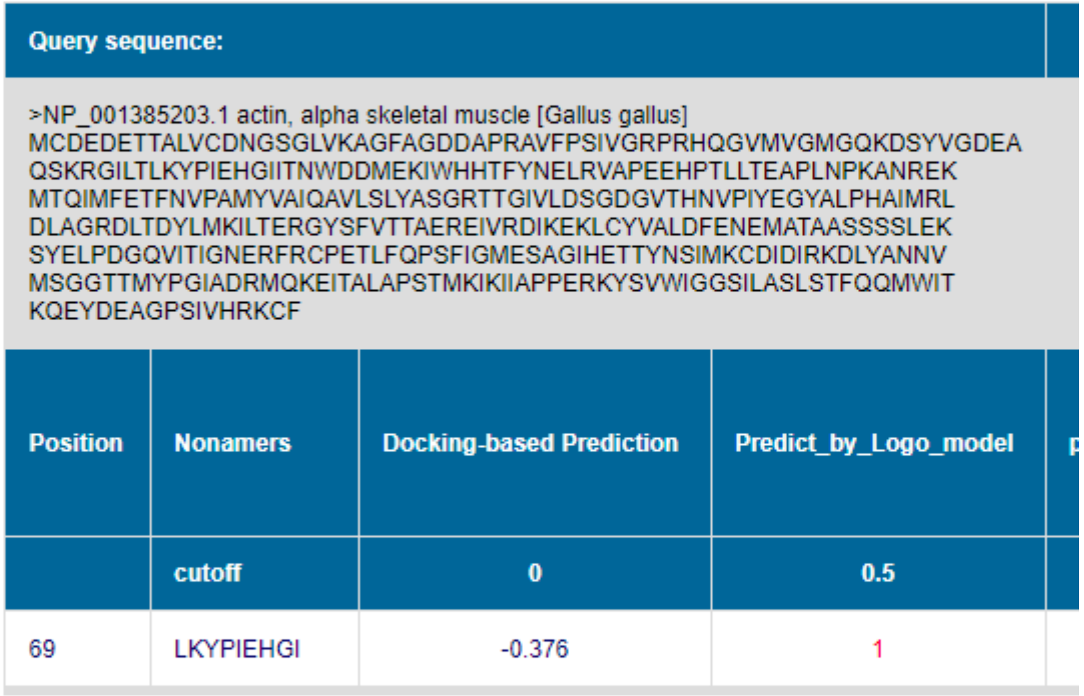
- Peptide mimic sequence PMPMPELPY was added to the chicken actin sequence



```
>NP_001385203.1 actin, alpha skeletal muscle [Gallus gallus]
MCDEDETTALPMPMPELPYVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQSKRGILTLK
YPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREKMTQIMFETFNVPAMYVAIQA
VLSLYASGRTTGIVLDSGDGVTHNVPIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYSFVTTAEREI
VRDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETT
YNSIMKCDIDIRKDLYANNVMSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILAS
LSTFQQMWITKQEYDEAGPSIVHRKCF
```

| **Protein** | **DQ 2.5** |
|---|---|
| NP_001385203.1 actin, alpha skeletal muscle [*Gallus gallus*] + PMPMPELPY | PMPMPELPY identified as celiac epitope |
| NP_001385203.1 actin, alpha skeletal muscle [*Gallus gallus*] + PMPMPDLPY | PMPMPDLPY was not identified |
| NP_001385203.1 actin, alpha skeletal muscle [*Gallus gallus*] + PMGMPELPY | PMGMPDLPY was not identified |
| NP_001385203.1 actin, alpha skeletal muscle [*Gallus gallus*] + PMPMPEVPY | PMPMPEVPY was not identified |

# Amino Acids in Lower Case Not Recognized

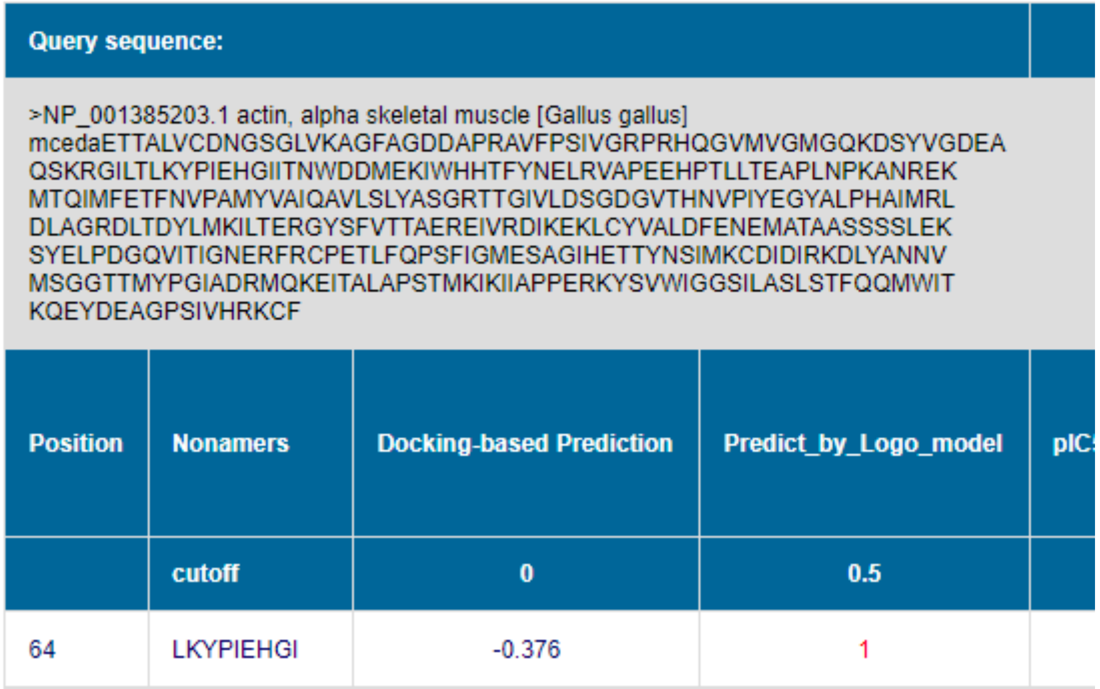


**Query sequence:**

>NP_001385203.1 actin, alpha skeletal muscle [Gallus gallus]
MCDEDETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA
QSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREK
MTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHNVPIYEGYALPHAIMRL
DLAGRDLTDYLMKILTERGYSFVTTAERЕIVRDIKEKLCYVALDFENEMATAASSSSLEK
SYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV
MSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIT
KQEYDEAGPSIVHRKCF

| Position | Nonamers | Docking-based Prediction | Predict_by_Logo_model | p |
|---|---|---|---|---|
| | cutoff | 0 | 0.5 | |
| 69 | LKYPIEHGI | -0.376 | 1 | |

Uppercase amino acids, hit at position 69



**Query sequence:**

>NP_001385203.1 actin, alpha skeletal muscle [Gallus gallus]
mcedaETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA
QSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREK
MTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHNVPIYEGYALPHAIMRL
DLAGRDLTDYLMKILTERGYSFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEK
SYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV
MSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIT
KQEYDEAGPSIVHRKCF

| Position | Nonamers | Docking-based Prediction | Predict_by_Logo_model | pIC |
|---|---|---|---|---|
| | cutoff | 0 | 0.5 | |
| 64 | LKYPIEHGI | -0.376 | 1 | |

First 5 amino acids lowercase,  hit now at position 64

# Observations and Questions

- When used to search proteins that have no known association with celiac disease, large numbers of binding sites are returned.
    - Will there be a peer-reviewed publication describing the validation of this tool when used to assess celiac disease potential?
    - Likewise, will there be a publication or documentation describing source data used in the development, training and operation of the tool?
- The tool produces 3 scores from 3 different prediction algorithms. What are the criteria to make a decision if further study is needed, one score, two scores, or all 3 scores above the threshold?
- According to the stepwise approach in EFSA guidelines, modeling is required when the peptide meets the threshold for a bioinformatic match.
    - Since this tool does not produce a structural model where interactions between amino acids in the HLA-peptide-TCR complex are reported and can be assessed, where does this tool fit into EFSA guidelines?
- Will source-code be made publicly available on GitHub for instance to support transparency and to facilitate use of the tool in a secure non-internet environment?
- What is envisioned in regard to maintenance and an update cycle for the code?