Celiac disease – Using science to develop a rational approach for screening of newly expressed proteins

23 November, Parma, Italy

Andre Silvanovich
Monsanto Company
St. Louis Missouri USA

Previous Regulatory Guidance for Peptide Searches

Searches with 8 aa peptides to identify allergens have been deemed to have little value

Celiac specific peptide science is in a much better place

- Organisms that cause the celiac disease are known
- Numerous peptides that bind cellular receptors and initiate the celiac amplification cascade are known

Regulatory guidance should leverage **all available scientific knowledge** to identify proteins with putative safety concerns

Specificity that drives **consistency in the application of the guidance** is essential

Leveraging the Latest Science

- Identify a comprehensive set of wheat, barley, rye and oat peptides that trigger celiac enteropathy
- Model the comprehensive set of peptides to develop rules
- Modify the guidance to drive a consistent approach to the following:
  - Thorough informatics screening of any food protein for possible similarity with hazardous celiac proteins
  - A step-wise process that is unambiguous
  - Start the hazard identification with the knowledge of the protein
  - Conservative approach that can identify potentially hazardous proteins/peptides
  - An informatics approach that is accurate, high-throughput, and can be interpreted by all stakeholders

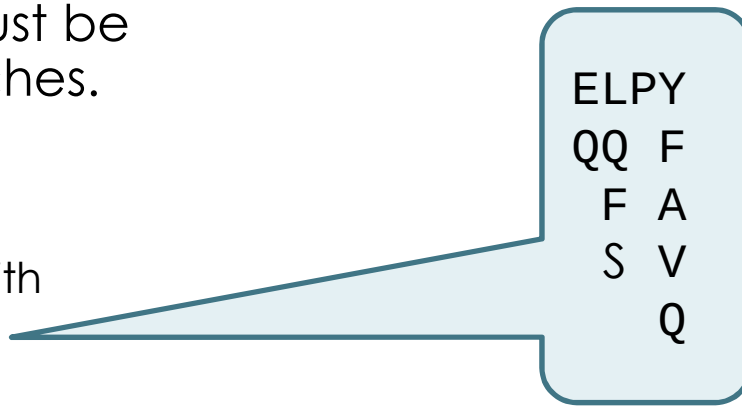
Evaluating the Feasibility of Using the Proposed Method to Identify Celiac Peptides in Novel Food Protein Sequences

For such screening to be useful it must display a reasonable level of specificity:

- ***Example***: ELISA analyses are validated to show specificity for a target protein in a matrix – if not specific, they are useless in a complex matrix.

- In bioinformatics, the large databases are our "complex matrix" and our searches must be relatively specific to find relevant matches.

The draft guidance document proposed screening with a degenerate peptide sequence[EQ][LQFS]P[YFAQV]



Evaluation of Specificity by Searching SwissProt and Genbank

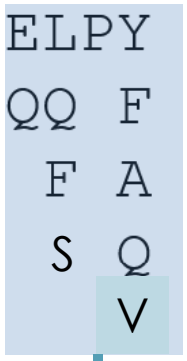
Specificity can be assessed by counting and determining the identity of aligning sequences

Imagine a bioinformatic screen that aims to identify all celiac peptides….

Specificity	Proteins identified that are known to cause celiac	The rate of matches for proteins that don't cause celiac
High	>90%	<**10%** - good specificity
Medium	~50%	~**50%** - balanced specificity
Low	<1%	>**99%** - poor balance – cannot identify hazards because they are not separated from safe proteins

The Degenerate 4 Amino Acid Peptide Has Low Specificity – Inaccurately Identifies Safe Proteins in Common Foods

Survey of UniProt
SwissProt database
552,000 entries
(Esmeralda Posada, Bayer CropScience N.V.)



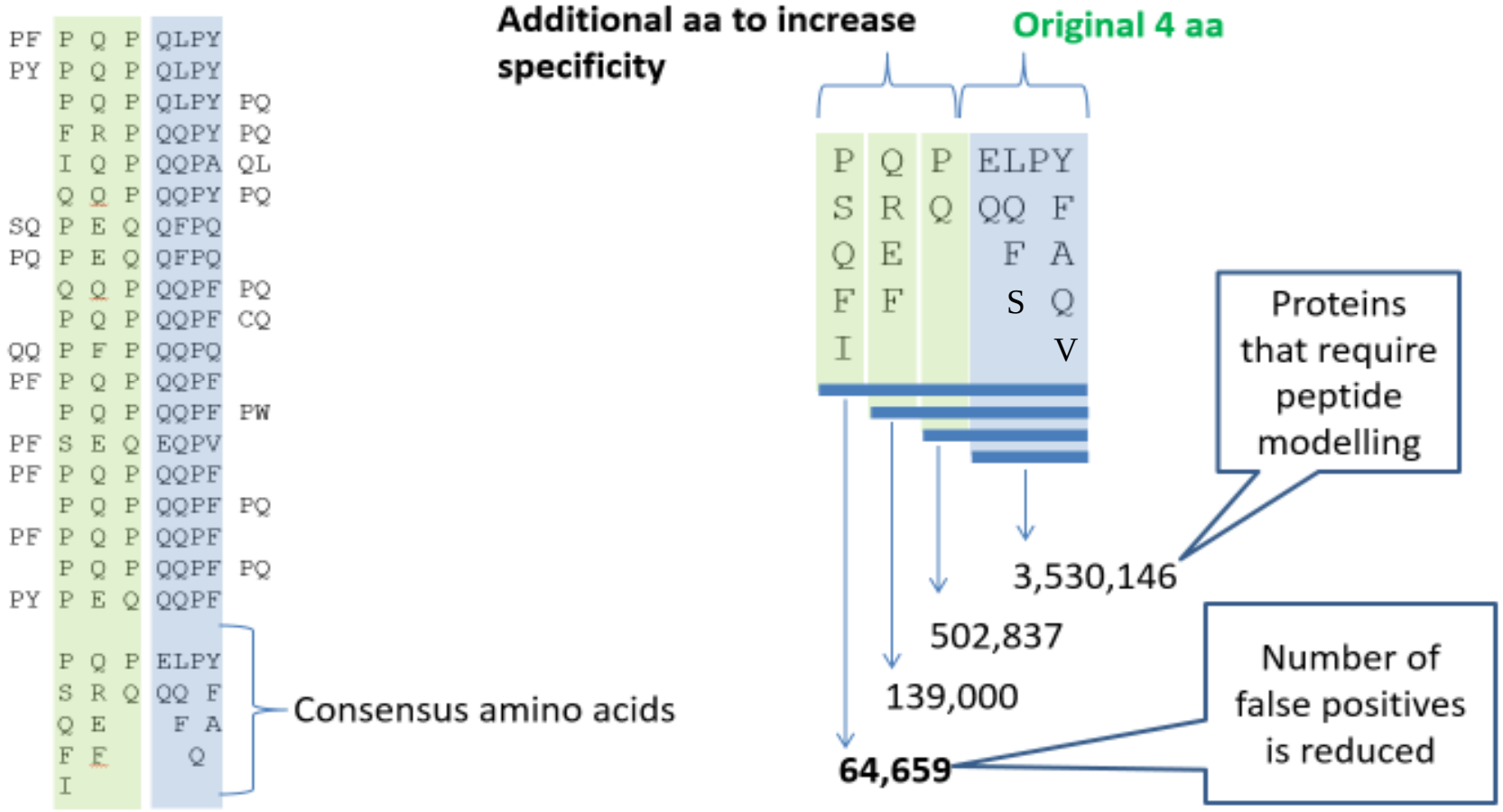
Survey of GenBank
49.5 million entries
(Andre Silvanovich, Monsanto company)

- **57,096** aligning proteins spanning **2,494** different species
- Within these species many are safely consumed by CD patients
  - Onion, garlic, cucumber, peanut, sugar beet, coffee
  - Beef, chicken, turkey,
  - Watermelon, strawberry, lemon orange and apple
  - Greater than 4,000 potato proteins

- 3.5 million produce an alignment
- Proteins from tens of thousands of species including humans, soy and corn were identified
- Among corn proteins were the high abundance zeins, a prolamin storage protein

- Not possible to distinguish celiac peptides in a meaningful way with degenerate 4 aa peptide
- Modeling of the specific celiac peptides shows us a 4 aa peptide can be improved
- Celiac peptides are very specific, specificity occurs at positions outside of the 4 aa peptide

Extending the Degenerate Peptide Length Decreases False Alignments while ***Retaining Specificity***



Evaluation of 464 Native Celiac Peptides from FARRP

Using a common denominator approach

- Shorter peptides are tested to determine if they are contained in larger peptides
- If a larger peptide contains a shorter peptide, the larger peptide is dropped
- If a protein contains a longer peptide, it will by definition also contain the shorter peptide, no need to use the long peptide as a query
- The 464 peptides are reduced to 160 by employing this filter step

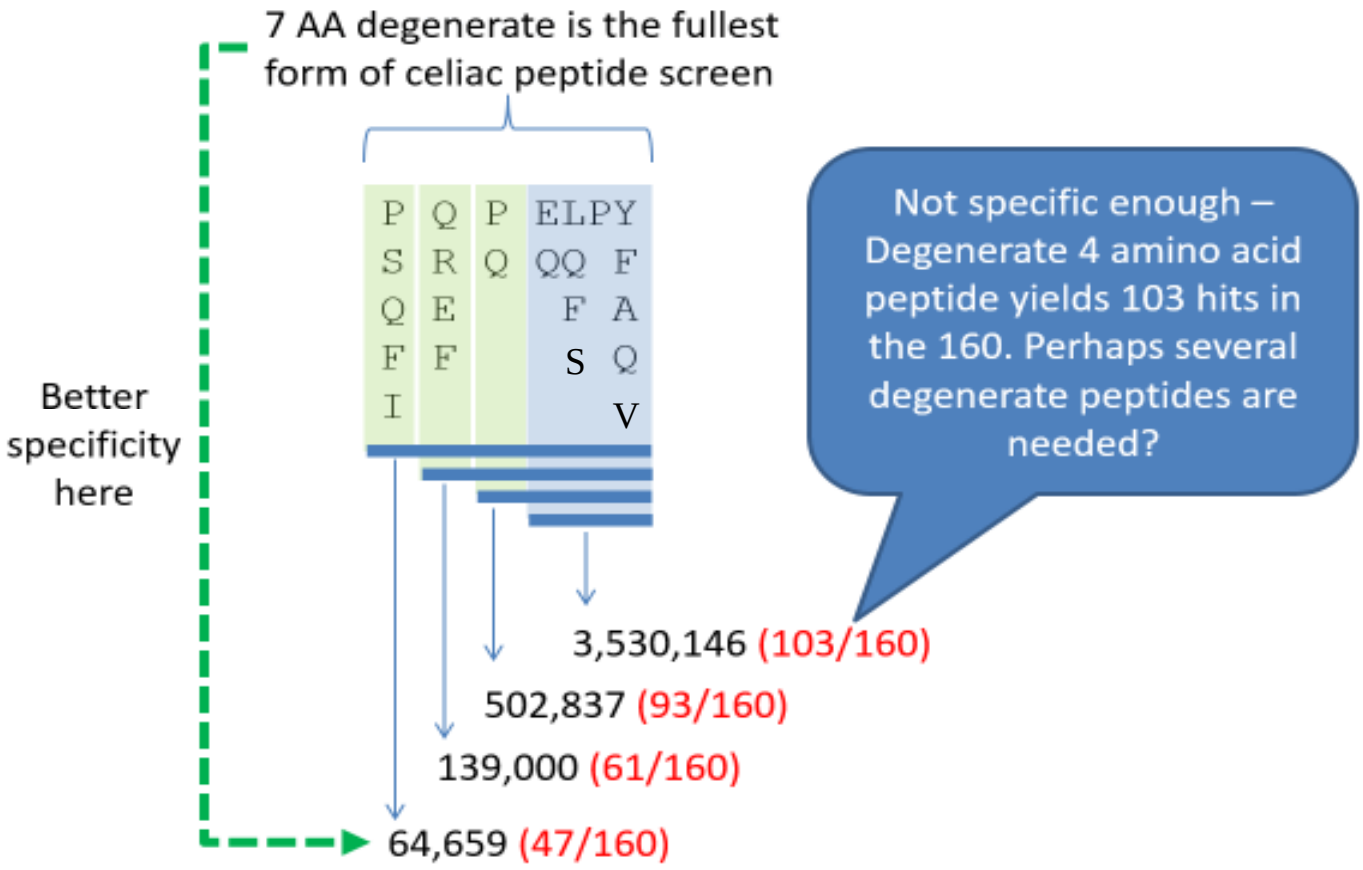
PQQPFPQQ
PFSQQQQQ
PFPQPQLPY
PQPQLPYPQ
.....
QQFLQ**PQQPFPQQ**PQQPYPQ
LQ**PQQPFPQQ**PQQPYPQQPQ
QQQFIQ**PQQPFPQQ**PQQTYP

This becomes the "screening peptide"

PQQPFPQQ

- Screening GenBank with the **160 peptides** yields 20653 exact matches
- Indicates an acceptable level of specificity from a screening perspective and is conservative enough to reliably identify celiac peptides

The Degenerate 4 Amino Acid Peptide



Use the Shortest Peptide to Screen for Potential Hazard

Goal

- To identify a screening sequence of amino acids that describes the 9-mer peptides associated with Celiac disease

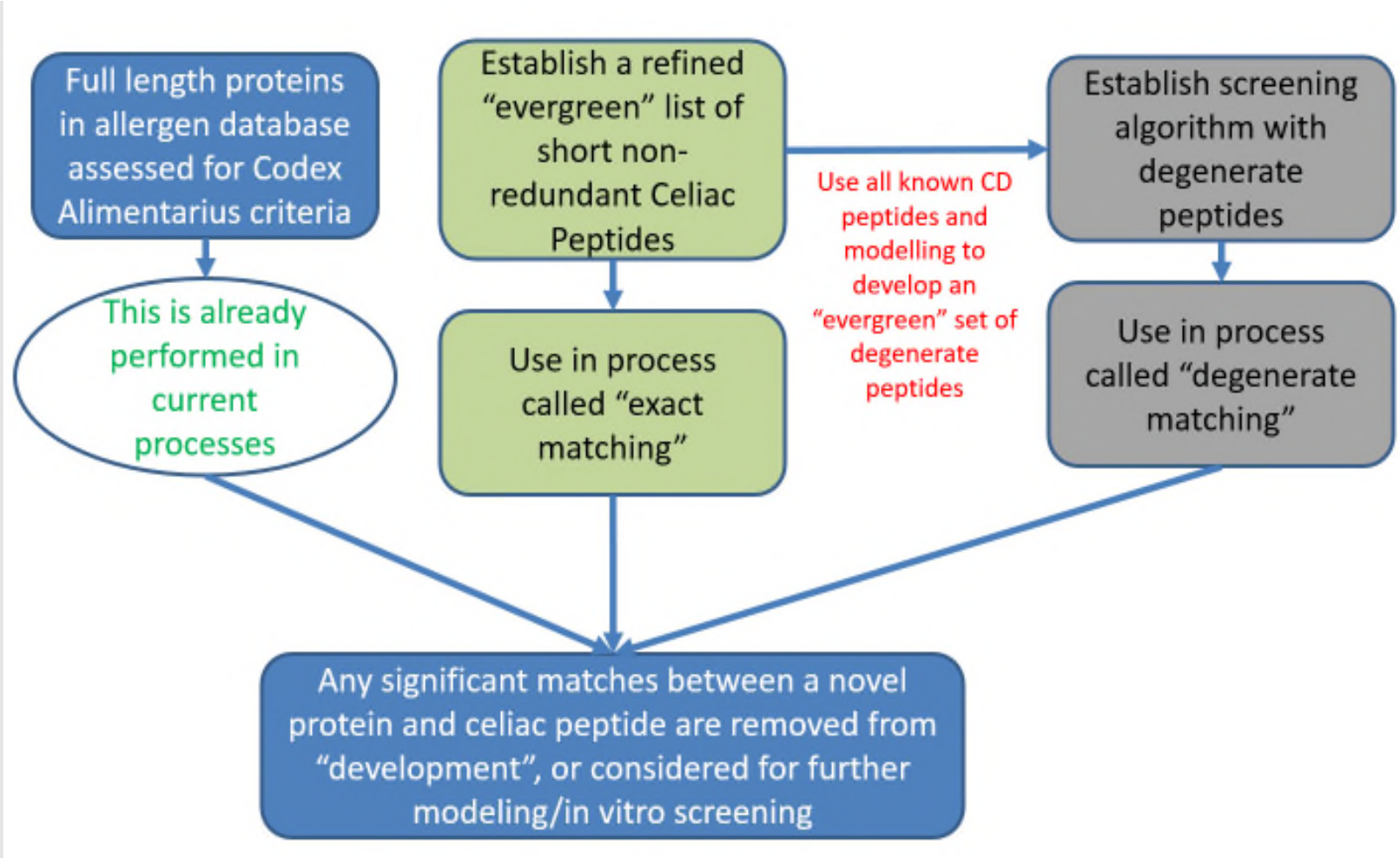
Premise

- There are 9-mers AND there are longer peptides that have been listed and associated with wheat, barely and rye that also contains these smaller (9 AA) peptides

Proposed Screening Objective

- Determine the smallest screening sequence or set of sequences that identify both the 9-mers and the longer sequences.

Proposed Informatics Workflow



Summary

Guidance should take into account origin and knowledge on newly expressed proteins

Current draft guidance has a high false positive rate and is likely to miss CD proteins

- False positives and risk of missing important matches limit the goal of conservative safety screening

A comprehensive list of celiac peptides should be evergreen and reside outside of the guidance document

- Keeps the safety screening up to date

Any bioinformatics that includes degeneracy should be based upon modelling and the comprehensive evergreen list

- Modelling supports the most accurate way to build a screening process – degeneracy alone is broad, but not accurate, so limits the goal of finding hazards

As we have seen with the 8-mer allergen search, old guidance does not go away

- Limits the ability to deliver to the newest science

EuropaBio

Thank You