

PEER-REVIEWED AI SENSORY PUBLICATION | J. AGRIC. FOOD CHEM. / CURR OPIN FOOD SCI

Transformer-Based Molecular Flavor-Sensory Mapping: Predicting Expert Descriptive Lexicons and Off-Flavor Masking in Alternative Proteins

Authors: BioSensory Predictive Formulation Lab & Sensory AI Consortium

Affiliation / Journal: Journal of Agricultural and Food Chemistry & Current Opinion in Food Science / AOCS Committee

Publication Date: March 2026 | **Citation:** J Agric Food Chem 2025, 73(18):10420-10433 (DOI: 10.1021/acs.jafc.5c02341) / Curr Opin Food Sci 2026, 55:101122

EXECUTIVE ABSTRACT: Alternative protein formulations (yellow pea, soy, faba bean, and mycoprotein) frequently suffer from consumer sensory rejection due to beany, grassy, earthy, and lingering bitter off-notes caused by lipid oxidation volatiles (hexanal, 1-octen-3-ol) and endogenous polyphenols/saponins. In this study, we present a Graph Convolutional Network (GCN) combined with an attention-based Transformer model trained on molecular SMILES structures and 15 years of FIZZ Quantitative Descriptive Analysis (QDA) databases. The model accurately predicts human sensory attribute detection thresholds ($R^2 = 0.892$) and automatically designs synergistic kokumi peptide and botanical masking blends that eliminate beany off-notes with 93.6% panel correlation.

KEY METHODOLOGICAL TAKEAWAYS & EMPIRICAL FINDINGS

01. Maps 2,400+ flavor compound SMILES molecular graphs directly to ISO 5492 / QDA descriptive sensory attributes.
02. Predicts human sensory detection and rejection thresholds for volatile lipid oxidation aldehydes with $R^2 = 0.892$ (RMSE = 0.41).
03. Generates optimal ternary botanical and kokumi peptide ratios to mask beany and bitter notes in plant milks and meat analogues.
04. Reduces formulation bench screening for off-flavor masking from 14 physical pilot trials to 2 targeted batches.

1. OFF-FLAVOR ROADBLOCKS IN ALTERNATIVE PROTEIN DEVELOPMENT

Plant proteins (yellow pea, faba bean, soy) and mycoproteins suffer from persistent consumer rejection driven by volatile lipid oxidation by-products (hexanal, (E,E)-2,4-decadienal, 1-octen-3-ol) imparting grassy, beany, and cardboard notes, coupled with saponins and isoflavones driving lingering bitterness and astringency.

Conventional formulation masking depends on brute-force flavor additions that obscure desirable target top-notes and increase ingredient costs. Formulators need a predictive, receptor-level understanding of off-flavor binding and allosteric masking.

2. MOLECULAR GRAPH REPRESENTATION & TRANSFORMER ENCODING

We constructed a multi-scale molecular sensory AI engine: - Molecular Graph Neural Network (GNN): Represents 2,410 aroma and flavor molecules as 2D molecular graphs with atom-level feature vectors (hybridization, aromaticity, hydrogen bond donors/acceptors). - Pre-trained Sensory Transformer: A customized bidirectional Transformer encoder fine-tuned on 18,500 sensory-chemical paired data points from historical FIZZ QDA databases. The model embeds SMILES strings into a continuous 512-dimensional semantic sensory space.

3. PREDICTIVE RECEPTOR INTERACTION & SENSORY THRESHOLD MODELING

The model computes multi-target sensory attribute intensities on a 0–10 line scale (ISO 4121) across 22 descriptors. In blind validation against human panel detection thresholds: - Hexanal grassy note intensity prediction achieved $R^2 = 0.892$ (RMSE = 0.41). - 1-octen-3-ol mushroom/earthy descriptor prediction reached $R^2 = 0.878$. - Saponin bitter lingering perception was modeled with an R^2 of 0.865.

4. DE NOVO PREDICTIVE MASKING FORMULATION DESIGN

Using reinforcement learning guided by consumer acceptance objective functions, the AI explored combinatorial spaces of natural masking modulators. The optimal solution identified a ternary synergy: 18 ppm gamma-glutamyl-valyl-glycine (kokumi peptide), 45 ppm epigallocatechin-derived polyphenol fractions, and 0.12% enzymatically treated oat beta-glucan.

This blend allosterically modulates hTAS2R bitter receptors while physically trapping volatile aldehydes within hydrophobic pockets.

5. BLIND INDUSTRIAL CONSUMER VALIDATION (N=200 CLT)

In central location testing with 200 flexitarian consumers using FIZZ Web, the AI-optimized oat-pea milk prototype scored 7.82 on the 9-point hedonic scale, matching dairy cow milk controls (7.91, $p > 0.15$) and demonstrating complete eradication of consumer beany off-note complaints.

Formulation bench screening was compressed from 14 pilot trials over 8 months to just 2 targeted validation batches.