



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE DELLA RICERCA

Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Food structure, function and artificial intelligence

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Mengucci, C., Ferranti, P., Romano, A., Masi, P., Picone, G., Capozzi, F. (2022). Food structure, function and artificial intelligence. *TRENDS IN FOOD SCIENCE & TECHNOLOGY*, 123(May 2022), 251-263 [10.1016/j.tifs.2022.03.015].

Availability:

This version is available at: <https://hdl.handle.net/11585/878869> since: 2022-04-04

Published:

DOI: <http://doi.org/10.1016/j.tifs.2022.03.015>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

Food structure, function and Artificial Intelligence

Carlo Mengucci ^{a,b}, Pasquale Ferranti ^c, Annalisa Romano ^c, Paolo Masi ^c, Gianfranco Picone
^a, Francesco Capozzi * ^{a,d}

^a Department of Agricultural and Food Sciences DISTAL, Alma Mater Studiorum University of Bologna, Cesena, 47521, Italy.

^b Inter-University Consortium for Magnetic Resonance of Metalloproteins CIRMMP, Sesto Fiorentino, 50019, Italy.

^c Department of Agricultural Sciences, University of Naples Federico II, Portici, 80055, Italy.

^d Interdepartmental Centre for Industrial Agrofood Research - CIRI Agrofood, Alma Mater Studiorum University of Bologna, Cesena, 47521, Italy.

*Correspondence: francesco.capozzi@unibo.it

ORCID:

C. Mengucci: 0000-0003-0602-3003

P. Ferranti: 0000-0002-6250-6512

A. Romano: 0000-0001-6628-2563

P. Masi: 0000-0003-2338-0651

G. Picone: 0000-0001-7932-6692

F. Capozzi: 0000-0002-5543-2359

1 **Abstract**

2 **Background:** The complexity of food structure is such as to hinder its inclusion in mathematical
3 models predicting food properties and transformations, although a considerable impulse is
4 being determined by using artificial intelligence. As a matter of fact, food definition currently
5 neglects the structural description, even in those fields for which structure is demonstrated to
6 have a decisive role, such as nutrition. **Scope and approach:** This review aims to analyse the
7 current knowledge about the structure of foods and its potential use to numerically define the
8 sensory and nutritional quality, as well as the stability properties. Starting from this information,
9 a possible methodology is explored to build, even in an automated way, mathematical models
10 for simulating and predicting the properties of food. A model pipeline has been proposed and
11 applied to pasta, in particular exploiting the description of the structural changes occurring upon
12 cooking. **Key findings and conclusions:** Foods may be designed *in silico*, based on automated
13 pipelines for direct extraction of information on rheological and sensory properties as derived
14 from structure images and from data on the dynamic state of the water. The ultimate goal of
15 these approaches is to make more limited use of expensive and time-consuming experiments
16 on physically prepared foods to get to use digital twins of foods designed in the laboratory.

17

18 **Key words:** food structure; bio-accessibility; bioavailability; functional food; digital twin; *in*
19 *silico* food; digestion.

20

21 **1. Introduction**

22 Although the description of food has traditionally been based on analytical chemical
23 composition, many of the important properties of food are determined by structural elements.
24 This limitation in the descriptive capacity of a food is also reflected in many mathematical
25 models that currently aim to predict the sensory, functional, and nutritional properties, including
26 for example digestibility. For this reason, the contribution of the structure of a food is often
27 overlooked including when studying the effect of diet on health. In fact, the nutritionist tends
28 to consult compositional databases when the correctness of a diet must be evaluated, having no
29 indication on how to use any structural data even when available. Nevertheless, before
30 collecting structural information, it would be necessary to establish how to use them to build
31 predictive models for nutritional functions that depend on it. Understanding how the ingredients
32 and each unit operation of food processes make up the structure of the foods and how this
33 structure changes during its life or on eating will play a main role in the development and
34 management of the food science and industry. For this reason, a tailored collection of scientific
35 work described in the literature has been examined to pave the way for a future approach using
36 matrix structural data to predict food functions, also exploiting artificial intelligence (AI).

37 **2. What is the structure of a food?**

38 Most foods are complex, heterogeneous materials composed of structural elements or domains
39 (co-) existing as solids, liquids and/or gases, where length scales span nanometres to millimetres
40 (Guo, Ye, Bellissimo, Singh, & Rousseau, 2017). Many of the important properties of foods
41 are determined by structural elements of micro-scale and above, such as bubbles, drops, strings
42 and particles (Ubbink, Burbidge, & Mezzenga, 2008). Food products consist largely of
43 carbohydrates, proteins, and lipids, forming clusters that behave as pseudo-molecules of higher
44 molecular weight than the individual constituent molecules (Ubbink, et al., 2008). These
45 interactions are primarily hydrogen-bonding interactions between the hydroxyl groups or Van
46 der Waals interactions between nonpolar molecules, but also ionic or covalent bonds, such as
47 disulphide or isopeptide, may be very important. The supramolecular organization of foods
48 gives rise to their structure. Complex food structures are formed, not because of the abundance
49 of elemental components, but because of the multiple interactions that proteins, lipids and
50 polysaccharides undergo at different conditions in an aqueous medium.

51 In natural and processed foods, the structure (or matrix) of a food is defined as the organization
52 of its constituent molecules at multiple spatial length scales (Guo, et al., 2017). At one extreme,
53 a food product is macroscopic, and at the other extreme, it is composed of molecules and atoms

54 characterized by molecular length scales (Ubbink, et al., 2008). The matrix of a food is in fact
55 scale-sensitive, i.e., interactions may take place at several scales in the same food as shown in
56 Figure 1.

57 For example, the matrix in a bakery product responsible for the textural properties of the porous
58 crumb are the protein-starch walls surrounding the air cells (Aguilera, 2019), and the relevant
59 scale is on the order of a few hundred microns (Liu & Scanlon, 2003). Starch granules
60 undergoing gelatinization may be regarded as inclusions in the continuous gluten matrix at a
61 scale of approximately 10 μm (Maeda, et al., 2013). At the nanoscale, gelatinized starch
62 granules are the matrix onto which α -amylases exert their action during digestion to release
63 glucose molecules (Li, Yu, Dhital, Gidley, & Gilbert, 2019). By and large, foods are systems
64 of dispersed phases, such as mesoscale particulate structures (colloids) derived from natural
65 food products constructed by self-assembly (e.g., granules, micelles, globules, and fibres) or
66 are created artificially via food processing (R. Van der Sman & Van der Goot, 2009). Next to
67 these mesoscale structures, food contains smaller molecular species, like salts, sugars, polyols
68 and phospholipids, which moderate the properties of the continuous or dispersed phases, or
69 their interfaces. The structure of a given food depends, however, enormously on the product,
70 its constituents and which of the many length scales are dominant in establishing the product
71 properties (Ubbink, et al., 2008). For an emulsion-based food such as mayonnaise, it is the
72 droplet size of around 1 μm which is the relevant length scale, whereas for dairy products it is
73 typically the size of a casein micelle ($\sim 50\text{-}100\text{ nm}$) (de Kruif & Huppertz, 2012) and the size
74 of the individual casein subunits ($\sim 2\text{ nm}$) that matter. The relevant length scale of food powders
75 is typically between 10 and 500 μm , and the structure of starch is described at length scales
76 between the macromolecular ($\sim 1\text{ nm}$) and the size of the starch granules ($\sim 1\text{ mm}$). Even length
77 scales substantially smaller than 1 nm matter in foods, as diffusion and the interaction of water
78 with the food matrix occur at these distances.

79 Food structure is important at all dimensional scales for texture, sensory properties, shelf life
80 and stability and can alter the kinetics and extent of food digestion (Guo, et al., 2017; H. Singh,
81 Ye, & Horne, 2009). It plays a vital role in how food interacts with the gastrointestinal tract
82 (GIT) (e.g., bodily fluids and receptors) and the resulting release and uptake of nutrients (Guo,
83 et al., 2017) and post-prandial outcomes (Turgeon & Rioux, 2011) In addition, the breakdown
84 of the food matrix is a major controlling factor for the perception of texture and flavour in the
85 mouth (Harjinder Singh, Ye, & Ferrua, 2015).

86 **3. How to quantitatively measure food structure**

87 Several techniques can be applied to measure the structure of food materials either directly
88 (optical and confocal microscopy, tomography, scanning and electron microscopy) or indirectly
89 from measurements of the mechanical response or spectroscopy (Table 1). Some challenging
90 techniques such as Differential Scanning Calorimetry (DSC) (Tester & Debon, 2000; Zhu,
91 Zhou, & Sun, 2019), Thermogravimetric analysis (TGA) (Tavares, Santos, & Noreña, 2021),
92 Nuclear Magnetic Resonance (NMR) spectroscopy and relaxometry (Kirtil & Oztop, 2016),
93 Near-Infrared Reflectance spectroscopy (NIR) (Shi, Lei, Louzada Prates, & Yu, 2019),
94 Attenuated Total Reflectance (ATR) spectroscopy (Cebi, Durak, Toker, Sagdic, & Arici, 2016)
95 and FT-Raman spectroscopy provide quantitative parameters that are related to the interactions
96 among molecules, thus making measurable physical-chemical properties that depend on the
97 supramolecular structure of the food matter. However, imaging techniques are essentially
98 dedicated to the investigation of the real 3D structure (Falcone, et al., 2006). Static Bragg-type
99 diffraction of neutrons and X-rays has been applied to either fluid or viscous food systems to
100 reveal the structure in the 10–100 nm length scale range (Ubbink, et al., 2008). Insight into lipid
101 polymorphism, liquid crystallinity, protein folding, etc. can typically be gained by using these
102 techniques. Because most common food properties are, however, directly related to the μm
103 length scale, light scattering techniques are primarily exploited. The application of the dynamic
104 light scattering (DLS) experiment to foods yields information on the diffusion coefficient of the
105 scattering objects (Ubbink, et al., 2008). Tomographic techniques such as magnetic resonance
106 imaging (MRI) and X-ray tomography are extremely powerful since they allow a full 3D
107 reconstruction of the sample structure but tend to be limited in resolution and/or slow in
108 acquisition times. Optical or Light Microscopy (LM) suffers from a similar limitation in
109 resolution, in this case due to the wavelength of visible light, even though structures of the order
110 of 1 μm can still be imaged using confocal microscopy. A further limitation of optical
111 techniques is that the food sample should be sufficiently transparent. Conversely, a major
112 advantage of optical microscopy is that dynamic processes on time scales larger than about 10
113 ms can easily be followed (Ubbink, et al., 2008). In the imaging of samples using transmission
114 electron microscopy (TEM), special staining, embedding and cutting techniques are
115 indispensable, whereas the use of scanning electron microscopy (SEM) is much more
116 straightforward (García-García, Cambero, Castejón, Escudero, & Fernández-Valle, 2019). An
117 interesting development is the progress in so called environmental scanning electron
118 microscopy (ESEM), which allows the analysis of samples at a desired relative humidity and
119 thus avoids artifacts due to the dehydration of foodstuffs (Ubbink, et al., 2008).

120 Different methods for image acquisition (light microscopy, transmission electron microscopy
121 and scanning electron microscopy) are generally coupled to digital analysis to quantitatively
122 define, with structural parameters, food at different structural levels. This provides a
123 measurement of different aggregation descriptors. The gel network can be characterized by
124 structural parameters such as pore size, strand dimensions and how these are distributed in the
125 volume. In the case of particulate gels, the diameter size of the pore is large, up to hundreds of
126 microns, compared to the size of the particle, around microns (Langton & Hermansson, 1996).
127 At low magnifications LM is used to estimate the size of the large pores. At higher
128 magnifications TEM estimates the size of the particles forming the strands of networks. The
129 pore size is more easily measured by digital image analysis than by evaluating the difference in
130 aggregation of particles in the network. In SEM the fracture plane is visualized, and the fracture
131 will follow the weakest structure, i.e., large pores. Thus, SEM micrographs tend to show larger
132 pores. and smaller pores could be embedded in clusters or conglomerates. Stereology is a tool
133 for measuring complex biopolymer gels, where no assumptions of the shape can be made. A
134 stereological approach was used to classify the mode of aggregation by a group of experienced
135 microscopists evaluating SEM-micrographs, to quantify pore size, particle size and amount of
136 threads within the pores in volume weighted mean volumes (Langton & Hermansson, 1996).
137 Five structural descriptors were quantified, namely porosity (number of pores), clusters (many
138 particles attached to each other like bunches of grapes), conglomerates (as if the particles were
139 joined together in non-linear, irregular, inhomogeneous order), strings of beads (as if the
140 particles were attached to each other in a linear order forming strings of beads) and hairiness
141 (as if small threads were attached to the surface of the particles and their outline is indistinct).
142 The three-dimensional gel network is responsible for bulk properties such as diffusion and
143 rheological properties, sensory quality and liquid holding capacity (Langton & Hermansson,
144 1996).

145

146 **4. The properties the food structure affects: sensory, stability, digestibility and** 147 **bioaccessibility**

148 The dimensions/size and shape/form of the particles, strands and pores create the different
149 textural properties of the food products and expert panellists can detect differences between
150 very small particles $<1 \mu\text{m}^3$ in volume (Langton, Åström, & Hermansson, 1997). In fact, texture
151 is a multi-parameter attribute, that derives from the molecular, microscopic or macroscopic
152 structure of a food and is detected by several senses, the most important ones being the senses
153 of touch and pressure (Szczeniak, 2002). Food structure, food texture, nutrients digestibility

154 and consumer product preferences and choices are intrinsically linked (Figure 1). Texture
155 influences people's acceptance of food and may be more important than the flavour in some
156 products (Clark, 1998). The sensory perception during food consumption depends not only on
157 the concentrations of odour- and taste-active compounds but also on the texture of food matrix
158 (Tournier, Sulmont-Rossé, & Guichard, 2007).

159 Multivariate techniques are used to create models to describe groups of the sensory descriptors
160 by some of the microstructural parameters (Janhøj, Frøst, & Ipsen, 2008; Pereira, Singh, Munro,
161 & Luckman, 2003). Correlations between the microstructure and sensory descriptors have been
162 found: grainy appearance, gritty texture, creamy texture and tendency to fall apart have a
163 logarithmic dependence on the particle size, and size of small and large pores (Langton, et al.,
164 1997). The soft and springy textures are influenced by combinations of microstructural
165 parameters, where the formation of strands into strings of beads or in clusters and
166 conglomerates seems to play an important role. Conversely, the sticky texture is negatively
167 correlated to the proportion of threads within the pores (Langton, et al., 1997). Stability can be
168 fully grasped only if food molecular dynamics and structure are taken into consideration, i.e.,
169 an appropriate understanding of the behaviour of food products requires knowledge of its
170 composition, structure and molecular dynamics, through the three-dimensional arrangement of
171 the various structural elements and their interactions (Wu, et al., 2020). In addition to water,
172 other structural elements can be identified in foods at a supramolecular structure level, such as
173 oil droplets, gas cells, fat crystals, strands, granules, micelles and interfaces. These structural
174 elements, composed of proteins, carbohydrates and lipids (in various combinations and
175 proportions), can exist in different states (glassy/rubbery/crystalline/liquid and solubilised)
176 even at uniform temperatures and water activity. This structural heterogeneity will necessarily
177 affect the molecular dynamics in the system and consequently the macroscopic food quality
178 attributes and their behaviour along storage. Physically separating the reactants in
179 microstructural locations can control the biochemical activity by avoiding the reactants to be in
180 contact.

181 It is a matter of fact that the gastrointestinal fate of lipids depends on their level, type, and
182 structural organization in foods (McClements, 2018). Matrices could be formed by controlled
183 gelation of single or mixed biopolymer systems around lipid droplets, by dehydration of oil-in-
184 water emulsions containing biopolymers or other wall materials, or by thermal treatment or
185 extrusion of starch matrices containing lipid droplets. Several studies have recently investigated
186 the impact of the food matrix on the digestibility of lipids using either *in vitro* or *in vivo*
187 digestion models (Corstens, et al., 2017; Dias, Zhu, Thompson, Singh, & Garg, 2019; J. Singh,

188 Dartois, & Kaur, 2010). When oil droplets are dispersed in a solid-like food matrix (e.g., cheese
189 or strained-type yogurt), the structure of the surrounding food matrix becomes the dominant
190 factor controlling digestion. For instance, the size of lipid droplets dispersed of oil-in-water
191 emulsions and nano emulsions can affect, during digestion, oil-soluble vitamins (vitamins A,
192 D, E and K) bioavailability in fortified foods (Tan & McClements, 2021); increasing oil droplet
193 size reduces the bioaccessibility by inhibiting lipid digestion and reducing micelle solubilisation
194 (Tan, Zhang, Liu, Xiao, & McClements, 2020). The knowledge advances provided by these
195 studies are setting the foundation for modulating fat digestion through food structure design, as
196 exhaustively reviewed by Guo, et al. (2017). In this sense, food structure design can be a tool
197 to develop foods that enable to control the body district as well as the extent and rate of release
198 of food lipids along the digestion process.

199 During digestion, the 3D network structure within a food matrix can obstruct the diffusion of
200 enzymes towards the surface of dispersed oil droplets. That is the reason why bile salts are
201 produced by the intestinal tract and released during food digestion to create an emulsion where
202 the digestive enzymes can act onto the food lipids.

203 Compared to interfacial films, the solid like-food matrix is potentially capable of providing
204 enhanced protection against lipolysis (Guo, et al., 2017). Evidence is accumulating that a
205 structured food with a high protein content may show slower lipid digestion (Salentinig, 2019).
206 An investigation on near forty food types, based on the harmonized INFOGEST digestion
207 method (Brodkorb, et al., 2019), found that those with medium and low lipid content showed a
208 limited lipolysis extent when the content of protein or starch was high (Calvo-Lerma, Fornés-
209 Ferrer, Heredia, & Andrés, 2018). In protein-rich foods such as cheese, the disintegration of the
210 protein network occurs mainly in gastric and intestinal steps, thus facilitating the subsequent
211 release of fat aggregates from the degraded matrix (Žolnere, Arnold, Hull, & Everett, 2019).
212 These results underline the importance of microstructure and the digestive environment on the
213 release of cheese components.

214 The *in vitro* digestion rate of lipids and starch was also reduced due to the intact vegetal cell
215 walls (Dhital, Bhattarai, Gorham, & Gidley, 2016). The intact cell wall structure and protein
216 matrices are impervious to amylase and can prevent or slow down enzyme diffusion to
217 substrate. In general, the intactness of cell walls is related to particle size, which is dependent
218 on mastication habits and processing conditions, for example, milling and heating (Li, Gidley,
219 & Dhital, 2019). The hydrolysis of intracellular starch and protein in the essentially intact cells
220 was 2–3%, whereas this increased to 40–45%, when the cells were mechanically broken and
221 digested, suggesting a barrier effect of intact cell walls to digestive enzyme access to starch and

222 proteins substrate (Ogawa, et al., 2018). In support to this hypothesis, it has been shown that
223 solubilisation of pectin cell walls, induced by thermal treatment of bean, exerted higher degrees
224 of cell wall permeability so that starch hydrolysis increased proportionally to the cell damage
225 (Pallares, et al., 2018). The morphology and the particle size of starch granules from different
226 plants is also considered an important factor affecting their digestion, as smaller granules have
227 greater enzymatic susceptibility regardless of botanical origin, due to their larger specific
228 surface area (Lehmann & Robin, 2007; Romano, et al., 2018; Romano, et al., 2016). Moreover
229 starch granules vary in the level of porosity and can have openings (pores) on the surface of the
230 granule (Fannon, Hauber, & Bemiller, 1992).

231 During processing, starch granules swell and lose their crystallinity and molecular organization
232 in a process commonly known as gelatinization. *In vitro* studies have demonstrated that the rate
233 of enzyme breakdown of gelatinized starch is much higher than that of native starch; native
234 wheat starches are degraded by only 10–15%, but after partial gelatinization the rate of
235 enzymatic degradation increased three-fold (Tian, et al., 2019). Therefore, gelatinization may
236 strongly influence the rate at which starch is digested and elicits the glycaemic response.

237 Starch–protein interaction in white flours might account for a decrease in *in vivo* glycaemic
238 response as well as for a reduction in *in vitro* digestibility, so that the removal of gluten from
239 wheat flour induces a high GI value in 11 kinds of gluten-free bread. In addition to acting as an
240 enzyme barrier, proteins also affect the properties of starch (gelatinization, retrogradation, etc.)
241 which is then less digestible (de la Hera, Rosell, & Gomez, 2014). If proteins are present in a
242 structured matrix or a clot-like structure is formed in the gastric environment, gastric juice needs
243 to penetrate this structured matrix to digest the protein. A 2–10 reduction factor for the diffusion
244 coefficient of pepsin has been measured in a structured matrix as compared to water. The
245 diffusion of pepsin is one of the limiting factors in the digestion rate of a structured food matrix
246 (E. Capuano & A. E. M. Janssen, 2021). Different egg-white gel structures, with a similar
247 protein composition, induced different proteolysis kinetics and provoked the release of different
248 specific peptides (Nyemb, et al., 2016).

249 Proteins can form supramolecular assemblies also because of thermal treatment. The formation
250 of aggregates may hide peptide bonds from proteases compared to denatured but isolated
251 molecules. The effect of cooking on the digestibility of meat proteins is a good example of such
252 complex relationships. Meat digestibility of regular-cooked beef was higher (95% digested)
253 than that of ‘well-done’ cooked beef (90% digested). Meat analogues are a class of food
254 products that imitate the sensory attributes of meat products but are produced from protein from
255 more sustainable sources, e.g., plant protein isolates, that are subjected to extrusion or shear-

256 cell technology. In these products, the presence of other food ingredients or components, such
257 as lipids and polyphenols, may affect protein digestibility. These effects are still poorly
258 understood for the lack of knowledge of the matrices and by the absence of predictive models.
259 Therefore, in the design of novel foods the effects of components on protein digestibility should
260 be carefully considered in the optimization of the processing parameters (E. Capuano & A. E.
261 Janssen, 2021). The process-induced modifications, *in primis* the Maillard reaction, could also
262 play a role in modulating the food digestibility and the bioavailability of protein amino acids,
263 by altering the chemical structural of protein networks and in turn the food microstructure: this
264 is the case of bread, dairy and meat products. Not secondarily, these modifications can also
265 affect the food allergenicity, through the interactions of protein-bound advanced glycation end-
266 products (AGE) with immune cells receptors, as evidenced for egg, dairy and peanut allergens
267 (Mueller, et al., 2013; Teodorowicz, Van Neerven, & Savelkoul, 2017; Toda, Hellwig, Henle,
268 & Vieths, 2019).

269

270 **5. The importance of structure in food design: driver for functional foods?**

271 The main objective of the food industry is to create products with specific properties and
272 characteristics which have a positive consumer impact. In recent years, the food industry, aware
273 of resource scarcity, is looking for nutritional alternatives, including functional foods, that
274 promote optimal health and help reduce the risk of disease and are “tailored”. Tailoring is a
275 process whereby the provision of information, advice and support is individualized to the user
276 (Lustria, et al., 2013). Mimic foods to be substituted, include also new functional ingredients in
277 formulation. The attempt to design new foods starting from more sustainable or more nutrient-
278 rich ingredients, with optimal characteristics for target population groups with specific needs,
279 has always clashed with the need to make these new foods at least as palatable, if not preferable,
280 to traditional ones. The limit is often in the obtainment of a desirable structure. In fact, unlike
281 some homogeneous foods, such as drinks, extracts or oils, most foods are heterogeneous
282 multiphase mixtures, having nutritional and sensory characteristics that strongly depend on the
283 placement with which the different phases are distributed in space, while forming the food
284 matrix. For this reason, the food technologists make use of structure-targeted toolboxes to
285 mimic successful matrices or invent new ones with even more performing characteristics. This
286 is usually carried out empirically in lab scale plants but, to avoid prolonged and expensive
287 physical research trials, the structure of the food could be preliminarily built *in-silico* also in
288 the design phase. This effective approach could be realized using conceptual toolboxes
289 (simulating unit operations, order of sequential steps, formulations) assisted by mathematical

290 prediction models. The purpose of designing the most suitable structures is then fulfilled,
291 through combinations of formulations and processes, to achieve the desired outcomes, like the
292 optimized durability, palatability, bioaccessibility and bioavailability of nutrients. This way,
293 food design considers not only composition, but also structure affecting chemical stability,
294 texture and dynamics of digestion and absorption of a food or its components. In this
295 perspective, tailored foods provide not only the necessary nutrients but also new functions,
296 linked to the matrix structure, targeted for specific populations groups such as the elderly,
297 babies, athletes, allergic peoples, vegans or for special diets such as low salt, sugars and fats,
298 or lactose- and gluten-free, and to increase the quantity of proteins, vitamins, dietary fibres, and
299 bioactive phytochemicals. Designer-made supramolecular food materials may form the basis
300 for personalized, health-promoting diets of the future (Norton, Espinosa, Watson, Spyropoulos,
301 & Norton, 2015). As already described in the previous section (Table 1), foods are made by
302 colloids toolboxes provided by nature, to which food technologists have added ‘artificial’
303 colloids, e.g., gas bubbles, oil droplets, ice crystals, fat crystals, and protein aggregates, created
304 by external forces (e.g., extrusion, compression, electric fields) or heating applied by food
305 processing equipment (R. Van der Sman & Van der Goot, 2009). With these ‘artificial’ colloids,
306 foods adhere to the length scales dictated by our tasting senses, which are sensitive enough to
307 detect structures of millimetre down to micrometre size (R. Van der Sman & Van der Goot,
308 2009). In this sense, a palatable food must be designed by finely modulating these structures to
309 enhance their nutritional function as well.

310 The structure of all foods can be imagined as the result of combinations of structural elements
311 provided by nature or imparted during processing and preparation. Food structure design is the
312 dedicated conception and fabrication of foods in such a way as to attain specific structures,
313 functions or properties (Guo, et al., 2017). Knowledge on how foods and beverages interact
314 with the digestive system, where they transform into supramolecular structures, can in fact have
315 a direct impact on the rational design of such advanced materials for functional food delivery
316 applications. For example, delivering a complete diet with a content of hydrophobic,
317 amphiphilic, and hydrophilic nutrients, which is personalized to the needs of the consumers,
318 could be beneficial for clinical and infant nutrition (Salentinig, 2019). Otherwise as confirmed
319 by recent studies on the use in pasta formulation of alternative flour from different sources,
320 such as potato and pigeon pea flour (Sharma, Dar, Sharma, & Singh, 2021) or flours from
321 legumes such as chickpea (El-Sohaimy, Brennan, Darwish, & Brennan, 2020; Garcia-Valle,
322 Bello-Pérez, Agama-Acevedo, & Alvarez-Ramirez, 2021) or bean (Romero & Zhang, 2019),
323 pasta nutritional profile is usually improved, leading to an increase in protein, ash, fibre

324 contents, and antioxidant compounds together with a decrease in the starch content and of *in*
325 *vitro* starch digestibility. What is missing in these approaches, solely accounting for the
326 nutritional profile, based on the composition of the ingredients, is the input related to the target
327 structural characteristics at different scale lengths. Although structure has been shown to have
328 an equally important impact on nutritional quality, a novel food is designed with great care for
329 its composition, stability and acceptability but, often, its structural optimization for nutrient
330 accessibility is omitted in the preliminary conceptualisation phase and studied only *ex post*.
331 Ultimately, the food structure design has the potential to be personalized to digestive conditions
332 and dietary nutrient requirements of the consumer or patient. From a nutritional perspective,
333 the ability to control food digestion is extremely important to design food with desired
334 characteristics: the key to control such process is to modulate the accessibility of digestive
335 enzymes to their substrate. Recently, considerable interest has also arisen in the application of
336 by-products of food processing with specific properties in food structure design, such as agar
337 or locust bean gum substitutes.

338

339 **6. Predictive models for designing the optimal structures: choice of parameters for** 340 **artificial intelligence**

341 As described in the previous section, stability, palatability, bioaccessibility and bioavailability
342 of nutrients are the target properties of food optimization. These properties must be expressed
343 using numerical descriptors, such as concentrations of degradation biomarkers, food sensory
344 scores, preferably assessed by instrumental devices (electronic nose or tongue), post-prandial
345 nutrients level in blood. Chemical and instrumental sensory analyses provide objective
346 parameters intrinsic to the food, that are independent from the individual interaction with it.
347 Conversely, parameters related to the digestive functions are strongly linked to the subjects'
348 variability. For this reason, experiments simulating different individual physiological and
349 pathological conditions are necessary, even when characterizing the target properties of a single
350 food. Whereas *in vivo* experiments give a global indication of food nutrients digestibility in its
351 full biological context, and *in vitro* experiments provide more insight into the different chemical
352 and physical mechanisms, the mathematical, or *in silico* modelling can connect these two
353 domains (E. Capuano & A. E. M. Janssen, 2021). The hydrolysis kinetics of the main
354 macronutrients (proteins, starch, and lipids) are modelled to predict the concentration and their
355 degree of hydrolysis in one or more compartments of the digestive system, or to predict the
356 transport of the food through the digestive system. The most popular approaches assume the
357 digestive tract as a series of bioreactors that can be described by mass balances, written as a set

358 of differential equations (Gim-Krumm, Donoso, Zuñiga, Estay, & Troncoso, 2018; Somaratne,
359 et al., 2020). In recent years, models that also consider the food matrix together with the reaction
360 and diffusion phenomena have been developed. Modelling of the swelling of protein gels by
361 using the Flory-Rehner theory has been combined with the Gibbs-Donnan theory to include the
362 distribution of ions between the gastric juice and the protein matrix to gain a better
363 understanding of the phenomena that are essential in the digestion of the food matrix (R. G. M.
364 van der Sman, Houlder, Cornet, & Janssen, 2020). Up to now, the role of modelling has been
365 that of linking and explaining in vivo and in vitro experiments. However, a further step is
366 required to use modelling for food properties prediction as a function of food structure. Suitable
367 numerical descriptors of structure are required as inputs for AI systems, to predict properties
368 that can define food in a functional way.

369 In the next section, available emerging approaches and those foreseen for the next future are
370 described, emphasizing how structure descriptors have been employed to predict sensory
371 properties and stability toward chemical transformations.

372

373 6.1 Describing the structure with imaging

374 The most straightforward way one can think of to parametrize food structure is through
375 descriptors extracted from imaging. Given the number of existing imaging techniques
376 (microscopy, spectral and hyperspectral imaging, nuclear magnetic resonance imaging,
377 ultrasound, microwave, etc.), many different aspects of food structure can be characterized and
378 digitalized. Furthermore, each imaging technique has its own array of analytics and descriptors,
379 capable of grasping and describing physical quantities tied to the physical nature of the specific
380 imaging technique. All these heterogeneous descriptors, together with general texture analysis
381 and computer vision descriptors, that can be obtained from images under certain conditions,
382 constitutes interesting inputs for artificial intelligence (machine and deep learning) frameworks.
383 As a matter of fact, the role of artificial intelligence in describing food structure from images,
384 is that of finding complex relationships between heterogeneous features describing different
385 aspects of the structure and the different structure-dependent properties of a food. Furthermore,
386 researcher in the field of deep learning, will rightfully argue that in the next future, a general
387 characterization of structure directly from images without a-priori features and descriptors
388 knowledge or assumptions could be possible. From an operative point of view, this means
389 feeding a neural network, as complex as needed, each pixel (or voxel in 3D) of an image as an
390 input and let the network learn how to build the best features to describe the problem (in this
391 case, predict food properties from structure description). To reach this goal, huge quantities of

392 suitable training data are however required to avoid some known problems of deep learning
393 architectures, such as overparameterization and overfitting. While some imaging techniques are
394 inherently suitable for the high-throughput standardized data production (such as magnetic
395 resonance imaging) required by deep learning architectures to achieve good prediction and
396 generalization, other imaging techniques (such as electronic microscopy) suffer from a series
397 of issues that make them less suitable for automation and high-throughput data production.
398 Overall, we are quite far from the data production required to have a huge amount of labelled
399 training data, especially regarding certain imaging techniques. In the next section, a high-
400 throughput imaging technique (MRI) and a high-resolution imaging technique (electronic
401 microscopy) are compared in terms of descriptors and suitability for automation. This is done
402 to outline possible directions to facilitate an efficient use of artificial intelligence at this stage
403 of structure description.

404

405 6.2 On the suitability of data production and imaging parameters for AI: a comparison

406 To grasp the meaning of what has been said in the previous section about data production and
407 generality of descriptors, it may be useful to focus on a comparison between electronic
408 microscopy (high-resolution, non-high-throughput) and magnetic resonance imaging (low
409 resolution, high-throughput). Table 2 sums up the main categories of descriptors that can be
410 extracted from images coming from these two different techniques, followed by a synopsis
411 highlighting the upsides and downsides of each technique as far as automation and
412 generalization are concerned. While MRI has many upsides when it comes to data production,
413 generalization, automation of analysis and feature extraction for classification, a trade off exists
414 in terms of spatial resolution. On the other hand, advocating the importance of high-resolution
415 aspects in terms of food structure description implies the necessity of high-resolution imaging
416 techniques. Electronic microscopy can fill in the role provided it becomes suitable for high-
417 throughput data production and data-driven modelling. At present, microscopic image
418 production is not optimized for automatic extraction of general features and descriptions, which
419 are at the core of frameworks using integrated data and automated workflows based on machine
420 learning. The first issue comes from image acquisitions inherently suffering from parameter
421 dependency. Lighting conditions and magnification which are obviously related to
422 experimental purposes, tend to shift microscopic imaging production toward less generalizable
423 datasets. Moreover, most canonical morphological and structural descriptors that are quantified
424 from this type of imaging, while being directly related to physical and easily interpreted
425 quantities, require specific assumptions (i.e., presence/absence of pores, spheres, shapes, fibers

426 etc.). Characterizing portions of images with ad-hoc assumptions is ill-suited for automation
427 and generalized parameter extraction. On the other hand, the power in terms of spatial resolution
428 of electronic microscopy cannot be overlooked when trying to characterize food structure. The
429 solution may lie in shifting microscopy data production toward a more pipeline-oriented way.
430 The creation of a consensus for data harmonization of microscopic images in the field, could
431 lead to parameter and feature extraction based upon low level and more general operators,
432 analogous to the ones used for MR images. This shift of paradigm in data production and
433 descriptor extraction, may contribute to boost modelling by facilitating the linking of the many
434 levels of complexity characterizing real life foods, using general parameters. A shift in data
435 production is also needed to pave the way for efficient deep learning approaches.

436

437 6.3 Structure images and sensory quality

438 Some scientific research, considered as an original reference works for these aspects, have laid
439 the foundations for the way a set of fundamental or derived parameters X , defining the food
440 structure, can be linked to a functional property Y through a mathematical function (Langton,
441 et al., 1997). For instance, the microstructural parameters may be presented as the estimated
442 model parameters A and B necessary to solve a correlating equation, e.g., $Y=A+B \log X$, where
443 Y is a sensory vector descriptor, X the model matrix for microstructural parameter. The
444 exemplary work by Langton et al. (1997), carried out on whey protein gels, defined nine
445 quantified microstructural parameters constituting the X vector feeding the model: four
446 parameters were the output of the digital image analysis (i.e., pore size at x20 magnitude; pore
447 size at x40 magnitude; particle size; amount of threads), and five parameters were mode of
448 aggregation as perceived by the test panel and already explained at the end of section 3
449 (Porosity; Clusters; Conglomerates; String of beads; Hairiness). Principal component analysis
450 (PCA) of the textural sensory data identifies two groups: (i) grainy appearance, gritty, creamy
451 and falling apart; and (ii) soft, springy, surface moisture and sticky. To find trends in groups of
452 variables (microstructural and sensory variables), PCA on the whole data set was performed.
453 The PCA had the purpose of creating, for each orthogonal component, linear combinations of
454 variables characterized by a high degree of co-variance, thus evidencing their interdependence,
455 by collecting them in different groups. One group of variables, defined by the large and small
456 star volume of pores, the star volume of particles, porosity, clusters, gritty, falling apart and
457 creamy (and acid) was found to take part in the systematic variation. Two groups of
458 microstructural parameters and sensory descriptors were found: one group depending on the

459 dimensions of the overall network and the other depending on the shape of the strands and
460 filling of the pores. This kind of data analysis made the model building a realizable approach.

461

462 6.4 Structure images, water dynamics and chemical transformations

463 Food systems behaviour is strongly dependent on water. Besides water content in a food
464 material, it is important to understand the water state and dynamics for a proper comprehension
465 of properties and stability of food structure. Understanding changes in location and mobility of
466 water represents a significant step in food stability knowledge, since water “availability” within
467 the matrix profoundly influences the chemical, physical and microbiological quality of foods.
468 Water mobility/dynamics can be described as a manifestation how “freely” water molecules
469 can participate in reactions or how easily water molecules diffuse to participate in reactions
470 occurring in different sites (Fundo, Quintas, & Silva, 2015).

471 Nuclear magnetic resonance is a powerful technique to investigate water dynamics and physical
472 structures of foods, through analysis of nuclear magnetisation relaxation times, because it
473 provides information on molecular dynamics of different components in dense complex
474 systems. The application of this technique may be very useful in predicting food systems
475 physicochemical changes, namely texture, viscosity or water migration (Fundo, et al., 2015).
476 Finding correlations amongst parameters based on time domain (TD)-NMR T2 decays,
477 describing water dynamics, and texture-derived features based on SEM images is a challenging
478 issue, when the aim is the quantitative characterization and parametrization of porous food
479 matrices and the transformation that food undergoes due to processing (such as cooking). A
480 comprehensive pipeline for parameter extraction, describing the porous food at different
481 cooking time, must be set accurately. TD-NMR raw data are preferable to classical exponential
482 fitting parameters, for building a general model accounting for the water status, as different
483 phenomena participate in the modulation of the relaxation times of the water population in the
484 compartmentalized porous matrix. For this reason, when matrix effects are investigated with
485 TD-NMR, a probabilistic PCA with Radial Base Function (RBF) kernel may constitute the
486 solution to find a latent space explaining differences in data tied to different matrices (pasta
487 type) and cooking times. The RBF kernel can take the non-linearity of decays into account,
488 projecting data into a suitable latent space, as shown in section 8.

489 The next section outlines the necessity to take another level of complexity into account when
490 trying to predict bioavailability and bioaccessibility: the physiological interaction with the
491 human organism.

492

493 **7. Digital twin of a food must include its structure**

494 Recently, a standardized food model (SFM) representing a typical US diet has been developed
495 to facilitate these investigations. This model consists of caseinate-stabilized fat droplets, free
496 casein, pectin, starch, sucrose, and sodium chloride. The SFM was stable to creaming for 2
497 days, contained small particles ($d \approx 180$ nm), and had a narrow particle size distribution (Zhang,
498 Zhang, & McClements, 2019). It would, therefore, be beneficial to have an SFM with a
499 harmonized composition and structure that could be used by researchers in different
500 laboratories to test food matrix effects. This model would allow researchers to obtain
501 reproducible results under standardized conditions, thereby leading to an improved systematic
502 understanding of the influence of the food matrix on oral bioavailability of different bioactive
503 agents. It may then be possible to establish general trends between bioactive type and the
504 magnitude of food matrix effects (Zhang, et al., 2019). However, gathering an almost infinite
505 set of model foods covering each possible category is a difficult, if not impossible, goal to
506 achieve. For this reason, having an exemplary set of model foods available, the next step could
507 be to create *in silico* models, derived from the mathematical combination of the basic models,
508 to simulate each existing real food. In other words, starting from physical model foods, virtual
509 simulator of foods can be generated.

510 As previously stressed, *in silico* simulations of food as complex particle based soft matter, are
511 strictly bound to the various length scales in the structure and occurring phenomena. As such,
512 different properties must be simultaneously investigated at different scales, from mesoscale to
513 nanoscale. While mesoscale properties (i.e. for emulsions and fat droplets) can be investigated
514 using coarse-grained particle-based simulations (Morris & Groves, 2013), at finer length scales
515 quantum-mechanical effects might occur. While hybrid multiscale models, capable of joining
516 coarse and fine level descriptions, are already available (Bolnykh, et al., 2019), making
517 predictive multiscale simulation approaches seemingly viable, the true complexity of food as a
518 system is still unaddressed. A complete review of available simulation tools, with a breakdown
519 of all the levels of complexity that must be addressed while trying to predict food properties
520 and functionalities from its structure and molecular-level interactions, is provided by Barroso
521 da Silva, et al. (2020). Amongst other issues, a predictive model relying solely upon multiscale
522 simulation, can suffer from high computational complexity. Simulating systems consisting of
523 extremely high number of particles, for which free-energy properties and kinetic properties
524 must be computed for several time-steps, can easily lead to unrealistic computational time, even
525 for specialized high-end hardware. However, machine and deep learning can prove useful in
526 decoupling multiscale descriptions from approaches based exclusively on simulation.

527 Quantitative structure-activity relationship (QSAR) based approaches are, in example, very
528 useful in predicting bio-chemical properties of compounds, including biological activity
529 (Neves, et al., 2018). These approaches are based on linking sets of molecular descriptors to a
530 given response variable; essentially the goal is to find a solution to a supervised learning
531 problem by coming up with an optimal set of user-defined molecular descriptors and a suitable
532 model to link them to the outcomes (response variable). A recent development of such a
533 framework involves the use of deep learning architectures, using recurrent and convolutional
534 neural networks (Chakravarti & Alla, 2019). The use of such neural networks allows for a
535 generalization of the learning problem, by eliminating the necessity of an a priori definition of
536 the molecular descriptors, at the cost of a very high pool of training data. Approaches of these
537 types, when the interpretability of the network-extracted descriptors is ensured, can minimize
538 the bias introduced by the users when choosing the descriptors and the difficulty of interpreting
539 descriptors that are not directly related to chemical structures. Results from these types of
540 framework, can furthermore be linked with outcomes from physiological experiments (i.e.,
541 experiments involving digestibility or involving health effects of certain compounds). In this
542 way, the molecular scale and the macroscale of physiological effects are encased in a multiscale
543 data-driven description. In a similar and more general fashion, many levels and scales of
544 complexity can be linked through machine and deep learning, by finding ways of extracting
545 general descriptors to be related to a response variable. Given the sheer complexity of food,
546 data-driven description of the various levels of complexity of food structure and food-human
547 interactions seems to be a promising way of predicting properties and health effects.

548 In the next section, an example of how to extract joint general descriptors from different scales
549 of complexity (water-matrix interaction and morphology) of a real-life food, that can be
550 ultimately related to outcomes from physiological experiments, is presented.

551 The example, set up by the authors, shows how to use SEM images in a more general way, by
552 extracting texture analysis descriptors, when the acquisition experimental design is suitable. An
553 example of how to correlate such structure descriptors to properties such as water mobility,
554 using raw data and machine learning, is also proposed.

555

556 **8. A case study: spaghetti pasta**

557 8.1 Designing food structure for food shaping

558 The structure is responsible for the sensorial, textural and nutritional properties of pasta, and its
559 formation relates to the characteristics of the raw ingredients and to several unit operations of
560 the manufacturing process (Scanlon, Edwards, & Dexter, 2005). In particular pasta structure

561 and quality depend on gluten and starch properties (Desai, Brennan, & Brennan, 2018; Witzcak
562 & Gałkowska, 2021) and on their physical-chemical modifications (protein denaturation, starch
563 gelatinization and swelling, etc.), occurring during pasta production process as well as the time
564 of cooking. Traditionally, dried spaghetti pasta is produced by mixing durum wheat (*Triticum*
565 *turgidum*, *subsp. Durum*) semolina and water (generally ~30 g /100 g), followed by a series of
566 unit operations such as extrusion, drying and packaging. The appropriate selection of
567 ingredients and technological parameters is fundamental, since it directly influences pasta
568 quality and structural features but, in turn, also affect content, digestibility and ultimately the
569 bioavailability of macro-nutrients (starch, proteins) and micro-nutrients (minerals,
570 phytochemicals). Since customers currently prefer pasta with uniform amber colour, firm
571 texture (“al dente”) and shape retention when cooked, it is of commercial importance to analyse
572 the cooking characteristics of pasta to design and develop a high-quality pasta that satisfy
573 consumer demands. Furthermore the increasing demand for innovative pasta products is
574 encouraging research on novel raw and processed materials such as dietary fibres, legume
575 flours, rice, corn, emmer, cricket flour - to meet the consumer demand in terms of nutritional,
576 sensory and technological value of pasta (Romano, Ferranti, Gallo, & Masi, 2021). In this
577 regard, cooking properties such as texture parameters (e.g. firmness and elasticity and shape
578 retention), cooking time, cooking loss, water absorption index, swelling index (Ficco, et al.,
579 2016; Susanna & Prabhasankar, 2013) are very important indicators of pasta quality. The
580 texture of pasta is the most important consumer attribute of pasta that influences consumer
581 acceptance (Susanna & Prabhasankar, 2013). In particular, firmness can be related to protein
582 content as well as the starch composition and it is a reflection of the bond strength and the
583 integrity of the protein matrix present in the pasta after the cooking process (Dexter & Matsuo,
584 1979). Microstructural changes of starch and proteins during cooking depend on water
585 availability, and the kinetics of solvation of each biopolymer have a major role on the final
586 texture of cooked pasta (Bonomi, et al., 2012). In order to control the cooking quality of pasta,
587 it is necessary to understand structural changes during the boiling process that affect textural
588 and sensorial properties of pasta. Primarily made up of carbohydrates (70 g /100 g) and proteins
589 (11.5 g / 100 g), cooked pasta is ingested as a solid food with a compact and “al dente” texture
590 and requires a low degree of mastication before swallowing, after which the pasta arrives in the
591 stomach in the form of large solid particles. It is considered to be a slowly digestible starchy
592 food with a low or medium Glycaemic Index (GI) (Gallo, Romano, & Masi, 2020; Granfeldt &
593 Björck, 1991). Generally, a compact and dense microstructure is attributed to the pasta, which:
594 i) limits water absorption and thus starch swelling and gelatinization, during cooking; ii) entraps

595 the starch granules reducing the accessibility of α -amylase (Jenkins, et al., 1983) and (iii)
596 releases α -amylase inhibitors during cooking that can immobilize the enzyme into the gluten
597 network (Zou, et al., 2019; Zou, Sissons, Gidley, Gilbert, & Warren, 2015). The major
598 challenges for pasta industry are now to increase food healthiness and customized nutrition
599 content and compositions but keeping high sensory attributes and technological performances.
600 Multiphysics simulations approaches could improve the efficiency of certain food
601 manufacturing processes and facilitate the sustainable packaging of food, for instance, by
602 creating morphing pasta that can be flat-packed, to reduce the air space in the packaging. It is
603 possible to induce temporary asynchronous swelling or deswelling that can transform flat
604 objects into designed, three-dimensional shapes (Tao, et al., 2021). How does it work with a
605 different microstructure associated to a functional pasta? Does the pasta morphing affect the
606 water-matrix interaction upon cooking?

607

608 8.2 Cooking and water-matrix interaction

609 To date, the structure of cooked pasta has been analysed at various microscopic and mesoscopic
610 levels by means of different methods, such as MRI. In fact it can be used to evaluate water
611 distribution and mobility in dry pasta, and in pasta at various cooking time (Bernin, et al., 2014).
612 Even these studies revealed that water penetration, distribution, and mobility during cooking
613 were highly dependent on the degree of protein reticulation, which in turns is greatly affected
614 by process conditions and food formulation (Tao, et al., 2021) MRI represents a non-invasive
615 method that spatially resolves the amount and dynamics of water and macromolecules-protons.
616 For this reason, Bernin, et al. (2014) used MRI to make a real time assessment of the effect of
617 starch-gluten ratio on water distribution in dry spaghetti during cooking. Therefore,
618 investigating such properties can help to understand how pasta components (water, gluten,
619 starch, fibre, etc.) interact with each other defining its structure, quality, acceptability, and
620 stability. In this respect, Gallo, et al. (2020) investigated the impact of pasta composition
621 (semolina and durum whole-wheat semolina) on water mobility in spaghetti before and after
622 cooking by low-resolution ^1H NMR experiments. In detail T_1 and T_2 proton relaxation times
623 as indicators of the molecular water mobility, have been determined (Gonçalves & Cardarelli,
624 2019). The uncooked spaghetti had T_1 and T_2 values much lower than the cooked ones
625 suggesting a very low water mobility in the dry pasta. With increasing cooking time, it was
626 observed a significant increase of both T_1 and T_2 relaxation times, either for semolina or whole
627 wheat spaghetti, suggesting that molecular water mobility within the pasta structure increases
628 as protein coagulation and starch gelatinization proceed (Gallo, et al., 2020). According to

629 Bosmans, Lagrain, Ooms, Fierens, and Delcour (2013), this behaviour could be explained in
630 term of three phenomena: i) water uptake in pasta structure; ii) starch gelatinization with the
631 subsequent destruction of the original structure; iii) gluten polymerization accompanied by
632 water expulsion from the gluten network. By comparing the behaviour of the two samples, one
633 observes that the presence of fibre led to a reduction in water mobility, since they can keep a
634 substantial excess of water during the cooking process (Serial, et al., 2016). The intermediate
635 zone was characterized by swollen starch granules embedded in a coagulated but dense protein
636 network; the presence of fibre resulted in an irregular structure in which there were a small
637 number of still intact and therefore non-gelatinized starch granules. As reported by Manthey
638 and Schorno (2002), in whole-wheat pasta bran particles cause a dilution of the gluten proteins,
639 interfering with proper gluten development. This results in a highly porous structure in which
640 starch granules are more accessible to water molecules. Starch granules in the surface region
641 were fully gelatinized and thus completely disintegrated in amylose and amylopectin. In the
642 intermediate zone, starch granules were highly hydrated increasing in size
643 Concerning the analysis of surface roughness, laser microscopy stressed an irregular surface
644 structure for dry pasta (due to the presence of intact starch granules) which became more
645 homogeneous after 1 min of cooking, due to the starch gelatinization.

646

647 8.3 Toward the automatization of water-matrix interactions and structure characterization

648 Joining measurement of NMR T1 and T2 proton relaxation time with SEM images, seems a
649 promising way of intertwining water mobility related phenomena with morphological
650 variations, thus including structure into food characterization. Parameters extracted with these
651 techniques, can furthermore be modelled using machine and deep learning architectures.
652 However, both methodologies require a fair amount of expertise in acquisition and processing
653 of the data, making standardization and automation of modelling pipelines challenging.
654 Extracting parameters and quantities from SEM images, is especially challenging as it requires
655 the use of dedicated software (e.g., when measuring particle size) to extract the distributions of
656 nanostructures and microstructures in an image. Accurate particle size distributions can be
657 difficult to obtain, as they require images with highly detectable particles and morphologies to
658 build a suitable statistic. Furthermore, the observable size of particles and structures depends
659 drastically on the viewing angle, while measures such as porosity and surface roughness are
660 affected by lighting and zooming. A complete list of issues and standardization of measures for

661 SEM image analysis is provided by the ISO (International Organization for Standardization)¹.
662 On the other hand, NMR relaxometry, while being a high-throughput technique with relatively
663 low acquisition times and high reproducibility, requires expertise in sample preparation and
664 acquisition sequence engineering. Furthermore, studying T1 and T2 distributions with inversion
665 software such as the UPEN algorithm² requires a deep understanding of the physical and
666 mathematical nature of the inversion problem, making this kind of analyses extremely variable
667 and elaboration parameters dependent.

668

669 8.4 Is learning from raw data and general descriptors promising?

670 A possible way to bypass some of these issues and make automatization and learning easier,
671 moving toward a more general framework, is to analyse raw TD-NMR decays and study SEM
672 images by extracting general texture analysis features and learning latent components in the
673 data, instead of specific measurements and physical quantities. In this qualitative example, a
674 way to correlate water mobility phenomena and morphology related features using machine
675 learning is proposed. SEM images of different zones of semola spaghetti, acquired at different
676 cooking time points, are processed and segmented using various filtering techniques and
677 morphological operators. A set of minimum image acquisition parameter can be chosen (i.e.,
678 zoom, lighting, well defined morphological regions of the pasta to acquire), to minimize
679 variability in the final dataset related to possible acquisition biases.

680 The 13 Haralick descriptors (Haralick, Shanmugam, & Dinstein, 1973) are computed from the
681 images of the complete cooking profile of the pasta. These general descriptors, widely used in
682 texture analysis and computer vision, are moments computed from the segmented image
683 cooccurrence matrix. These moments are intended to describe the characteristics of the patterns
684 of the textures of the image, in term of the probability of occurrence of grey levels. As such,
685 they serve as general morphological descriptors, whose relationships with descriptors extracted
686 from TD-NMR can be estimated. These descriptors can be studied as a function of time-
687 dependent latent components extracted from TD-NMR raw decays, with a process summarized
688 in Figure 2, to find links with water mobility related phenomena.

689 As an example, typical raw decays of pasta at different cooking time points, are shown as
690 projection into a latent variable space using a probabilistic KPCA (Kernel Principal Component
691 Analysis). Using an RBF kernel in a self-optimizing learning pipeline, each decay curve is

¹ <https://www.iso.org/obp/ui/fr/#iso:std:iso:19749:ed-1:v1:en>

² <https://iopscience.iop.org/article/10.1088/1361-6420/33/1/015003>

692 projected into a lower dimensional space with the aim of detecting differences tied to
693 phenomena occurring during cooking (Figure 3a). Some of the Haralick descriptors appear to
694 have strong linear and non-linear correlations with the time dependent latent variables extracted
695 from TD-NMR raw decays (Figure 3b). Moreover, correlations seem to be different from zone
696 to zone, highlighting the expected behaviour of TD-NMR to discriminate information about
697 different characteristics of water populations at different cooking times in different pasta zones.
698 Some texture analysis descriptors, such as texture Sum Average (HF6, y axis of Figure 3b)
699 which is tied to “homogeneity” of the texture, describing the central zone images, show an
700 exclusively monotonous relationship with cooking time and PC scores (both PC0 and PC1)
701 after a certain cooking time (Morphologic Phase, Figure 3b). Looking at the KPC space, this
702 phenomenon corresponds to a steep variation in PC1 score and a low variation in PC0 scores.
703 On the contrary, below this time (orange to yellow points, Figure 3a), steep variations along
704 PC0 and slow variations of PC1 scores are encountered, until PC1 score variation minimum is
705 reached (red points, Figure 3a). After this, variation on PC1 scores starts to rise again
706 (Activation, Figure 3b) while PC0 scores variation starts to reach its minimum. Above this
707 threshold of cooking time, both PC1 and many HF descriptors, such as HF6 in Figure 3b, start
708 a trend with a strict monotonous dependence with time. This time point may represent the
709 threshold for which changes in the texture of the matrix start to be exclusively dependent on
710 cooking time, maybe due to the irreversible rupture of structures in the food matrix and the
711 consequent variation of the timescale of water exchanges. Looking at Figure 3, one can argue
712 that the description of the morphological changes emerging from these preliminary results, is
713 in agreement with findings from Manthey and Schorno (2002). If in the early moments of
714 cooking starch gelatinization prevails, the resulting SEM images tend to show more
715 homogeneous surfaces, with little differences from a morphological point of view. However,
716 with raising cooking time the observed increase in the inhomogeneity of pasta surface and the
717 changes in water mobility become a monotonous function of cooking time, as the partial
718 detachment of solid materials such as starch and starch-attached proteins probably becomes the
719 prevalent phenomena. Haralick descriptors for SEM images, together with self-learned latent
720 components extracted from TD-NMR raw decays, are capable of picking up this sort of
721 threshold behaviour and successfully merging description of the morphology and water-matrix
722 interaction. Learning latent features and parameters from raw NMR data and images processed
723 to a bare minimum, studying and understanding the correlation amongst the extracted
724 descriptors can help building digital twins of food with an included structural characterization
725 of the matrix. In the example, water mobility and morphology are investigated with a general

726 data-driven framework, using machine learning and canonical texture analysis to find suitable
727 features and descriptors. The main advantages of this approach are the generality and the lack
728 of assumptions needed for the description of structural elements from images. Using raw data
729 (such as T2 decays in the example) and letting AI methods learn the best way to represent them
730 is optimal when dealing with many heterogeneous datasets, in terms of automation and feature
731 discovery. Moreover, bypassing the necessity of assumptions when describing structure from
732 images, becomes an advantage when parametrizing real-life foods in which matrix structures
733 can be extremely heterogenous along the different length scales. Consequently, different types
734 of images and raw data from experiments regarding digestibility, stability and bioaccessibility
735 can be explored to shed light on their relationship with structural properties, even with complex
736 real-life food.

737

738 **9. Conclusions**

739 Understanding how formulations of ingredients and unitary operations of food processes make
740 up the structure of food and how this structure changes during its shelf - life or eating will play
741 an important role in the development and management of food science and industry. Much of
742 the information that defines the structure of a food is currently neglected when entering the
743 domain of nutrition, as the structural dimension is too complicated to be quantitatively
744 measured and related to sensorial properties, stability, digestibility and bioaccessibility of
745 nutrients. Not even the momentum given by the considerable progress achieved in the design
746 of functional foods has so far been sufficient to assign the correct importance to the structural
747 nature of food. Certainly, the complexity of the information is such as to hinder the creation of
748 predictive-based models based on analysis of a limited amount of available data. For this reason,
749 it is certainly conceivable a considerable impulse determined using artificial intelligence
750 capable of handling certain quantities of heterogeneous data. It would be useful to be able to
751 predict the sensory quality and stability of food designed to become carriers of healthy nutrients
752 through images that shoot their supramolecular structure. It would be also desirable for these
753 same foods designed *in silico*, to predict the duration as a function of the dynamic state of the
754 water capable of modulating the chemical transformations underlying physiological or
755 anomalous phenomena, also to include the aspect of sustainability in the conception phase. A
756 model food such as pasta, widely consumed all over the world, object of studies for possible
757 functionalization as a vehicle for bioactive substances useful for health, can serve as a case
758 study to build a pipeline of an automated approach. The endpoint of such a pipeline is a direct
759 extraction of information on rheological and sensory properties starting from images of the

760 structure and from raw data of the dynamic state of the water. The main advantages of such a
761 framework are: i) an efficient automatization of parameter extraction useful for building
762 suitable inputs for AI architectures, which require high-throughput data for proper training ii)
763 a more efficient and general way of extracting parameters especially from imaging; using
764 general parameters for image analysis instead of measured technique-dependent parameters or
765 measured quantities that requires ad-hoc assumptions on structures (i.e. presence/absence of
766 pores, fibres etc.), can prove more useful given the high heterogeneity of structural elements at
767 different length scales iii) a more efficient way of linking different levels of complexity of
768 structure description and properties to be predicted, through the use of general parameters and
769 features learned directly from data with machine learning; this step is crucial to avoid
770 oversimplification generated by canonical interpretative models. However, extending this
771 framework to all the aspects of food modelling for properties prediction, poses quite a few
772 challenges. The first one is a required shift of paradigm of imaging data production. Certain
773 techniques (such as SEM) suffer from a lack of a consensus of acquisition standards, hindering
774 data harmonization which is essential for high-throughput input production. Another major
775 challenge is the complexity of modelling and parametrizing properties such as bioaccessibility
776 and bioavailability. These properties not only require a comprehensive parametrization of the
777 structure to be predicted but are also linked to the interaction with digestive functions. The
778 interaction with the human organism, especially with GIT functions, adds a whole new level of
779 complexity that must be addressed. The compartments of the GIT and their functions are
780 interlinked and impacted by food structure, while also being subjected to interindividual
781 variability. Hybrid approaches linking structure at molecular level and physiological outcomes,
782 based on deep learning architectures, are however gaining popularity (section 7) due to their
783 computational performances.

784 The ultimate goal of AI oriented frameworks is to be able to make more limited use of expensive
785 and time-consuming experiments on physically prepared foods, by using digital twins of foods
786 designed in the laboratory. This, in turn, could lead to a more efficient data production for
787 studies of physiological outcomes of functional foods.

788 Further advances for future applications of AI in food science and technology may arise, as in
789 medical sciences, from the enormous expanse of data resulting from the exploitation of different
790 types of heterogeneous information (images, chemical analysis results, physical measurements,
791 etc.) in the same system, for example a single neural network, integrating food data from
792 different scales and sources. The challenge, in this case, is to give the right importance to one
793 type of information over the others.

794

795 **Declaration of competing interests:** None of the authors has any conflict of interest relevant
796 to the content of this article.

797 **Acknowledgments:** None.

798 **Funding:** None.

10. References

- Aguilera, J. M. (2019). The food matrix: implications in processing, nutrition and health. *Critical reviews in food science and nutrition*, *59*, 3612-3629.
- Barroso da Silva, F. L., Carloni, P., Cheung, D., Cottone, G., Donnini, S., Foegeding, E. A., Gulzar, M., Jacquier, J. C., Lobaskin, V., & MacKernan, D. (2020). Understanding and controlling food protein structure and function in foods: perspectives from experiments and computer simulations. *Annual Review of Food Science and Technology*, *11*, 365-387.
- Bernin, D., Steglich, T., Röding, M., Moldin, A., Topgaard, D., & Langton, M. (2014). Multi-scale characterization of pasta during cooking using microscopy and real-time magnetic resonance imaging. *Food research international*, *66*, 132-139.
- Bolnykh, V., Olsen, J. M. H., Meloni, S., Bircher, M. P., Ippoliti, E., Carloni, P., & Rothlisberger, U. (2019). Extreme scalability of DFT-based QM/MM MD simulations using MiMiC. *Journal of chemical theory and computation*, *15*, 5601-5613.
- Bonomi, F., D'Egidio, M. G., Iametti, S., Marengo, M., Marti, A., Pagani, M. A., & Ragg, E. M. (2012). Structure–quality relationship in commercial pasta: A molecular glimpse. *Food chemistry*, *135*, 348-355.
- Bosmans, G. M., Lagrain, B., Ooms, N., Fierens, E., & Delcour, J. A. (2013). Biopolymer interactions, water dynamics, and bread crumb firming. *Journal of Agricultural and Food Chemistry*, *61*, 4646-4654.
- Brodkorb, A., Egger, L., Alminger, M., Alvito, P., Assunção, R., Ballance, S., Bohn, T., Bourlieu-Lacanal, C., Boutrou, R., & Carrière, F. (2019). INFOGEST static in vitro simulation of gastrointestinal food digestion. *Nature protocols*, *14*, 991-1014.
- Calvo-Lerma, J., Fornés-Ferrer, V., Heredia, A., & Andrés, A. (2018). In vitro digestion of lipids in real foods: influence of lipid organization within the food matrix and interactions with nonlipid components. *Journal of food science*, *83*, 2629-2637.
- Capuano, E., & Janssen, A. E. (2021). Food Matrix and Macronutrient Digestion. *Annual Review of Food Science and Technology*, *12*, 193-212.
- Capuano, E., & Janssen, A. E. M. (2021). Food Matrix and Macronutrient Digestion. *Annual Review of Food Science and Technology*, *12*, 193-212.
- Cebi, N., Durak, M. Z., Toker, O. S., Sagdic, O., & Arici, M. (2016). An evaluation of Fourier transforms infrared spectroscopy method for the classification and discrimination of bovine, porcine and fish gelatins. *Food Chem*, *190*, 1109-1115.
- Chakravarti, S. K., & Alla, S. R. M. (2019). Descriptor Free QSAR Modeling Using Deep Learning With Long Short-Term Memory Neural Networks. *Front Artif Intell*, *2*, 17.
- Clark, J. E. (1998). Taste and flavour: their importance in food choice and acceptance. *Proceedings of the nutrition society*, *57*, 639-643.
- Corstens, M. N., Berton-Carabin, C. C., Kester, A., Fokkink, R., van den Broek, J. M., de Vries, R., Troost, F. J., Masclee, A. A., & Schroën, K. (2017). Destabilization of multilayered interfaces in digestive conditions limits their ability to prevent lipolysis in emulsions. *Food Structure*, *12*, 54-63.
- de Kruif, C. G., & Huppertz, T. (2012). Casein Micelles: Size Distribution in Milks from Individual Cows. *Journal of Agricultural and Food Chemistry*, *60*, 4649-4655.
- de la Hera, E., Rosell, C. M., & Gomez, M. (2014). Effect of water content and flour particle size on gluten-free bread quality and digestibility. *Food Chem*, *151*, 526-531.
- Desai, A., Brennan, M. A., & Brennan, C. S. (2018). The effect of semolina replacement with protein powder from fish (*Pseudophycis bachus*) on the physicochemical characteristics of pasta. *LWT*, *89*, 52-57.
- Dexter, J., & Matsuo, R. (1979). Changes in spaghetti protein solubility during cooking. *Cereal Chemistry*, *56*, 394-398.

- Dhital, S., Bhattarai, R. R., Gorham, J., & Gidley, M. J. (2016). Intactness of cell wall structure controls the in vitro digestion of starch in legumes. *Food & function*, *7*, 1367-1379.
- Dias, C. B., Zhu, X., Thompson, A. K., Singh, H., & Garg, M. L. (2019). Effect of the food form and structure on lipid digestion and postprandial lipaemic response. *Food & function*, *10*, 112-124.
- El-Sohaimy, S. A., Brennan, M., Darwish, A. M., & Brennan, C. (2020). Physicochemical, texture and sensorial evaluation of pasta enriched with chickpea flour and protein isolate. *Annals of Agricultural Sciences*, *65*, 28-34.
- Falcone, P. M., Baiano, A., Conte, A., Mancini, L., Tromba, G., Zanini, F., & Del Nobile, M. A. (2006). Imaging techniques for the study of food microstructure: a review. *Advances in food and nutrition research*, *51*, 205-263.
- Fannon, J. E., Hauber, R. J., & Bemiller, J. N. (1992). Surface pores of starch granules. *Cereal Chem*, *69*, 284-288.
- Ficco, D. B. M., De Simone, V., De Leonardis, A. M., Giovanniello, V., Del Nobile, M. A., Padalino, L., Lecce, L., Borrelli, G. M., & De Vita, P. (2016). Use of purple durum wheat to produce naturally functional fresh and dry pasta. *Food chemistry*, *205*, 187-195.
- Fundo, J. F., Quintas, M. A., & Silva, C. L. (2015). Molecular dynamics and structure in physical properties and stability of food systems. *Food Engineering Reviews*, *7*, 384-392.
- Gallo, V., Romano, A., & Masi, P. (2020). Does the presence of fibres affect the microstructure and in vitro starch digestibility of commercial Italian pasta? *Food Structure*, *24*, 100139.
- García-García, A., Cambero, M., Castejón, D., Escudero, R., & Fernández-Valle, M. (2019). Dry cured-ham microstructure: A T2 NMR relaxometry, SEM and uniaxial tensile test combined study. *Food Structure*, *19*, 100104.
- García-Valle, D. E., Bello-Pérez, L. A., Agama-Acevedo, E., & Alvarez-Ramirez, J. (2021). Structural characteristics and in vitro starch digestibility of pasta made with durum wheat semolina and chickpea flour. *LWT*, *145*, 111347.
- Gim-Krumm, M., Donoso, P., Zuñiga, R. N., Estay, H., & Troncoso, E. (2018). A comprehensive study of glucose transfer in the human small intestine using an in vitro intestinal digestion system (i-IDS) based on a dialysis membrane process. *Journal of Membrane Science*, *564*, 700-711.
- Gonçalves, M. C., & Cardarelli, H. R. (2019). Changes in water mobility and protein stabilization of Mozzarella cheese made under different stretching temperatures. *LWT*, *104*, 16-23.
- Granfeldt, Y., & Björck, I. (1991). Glycemic response to starch in pasta: a study of mechanisms of limited enzyme availability. *Journal of Cereal Science*, *14*, 47-61.
- Guo, Q., Ye, A., Bellissimo, N., Singh, H., & Rousseau, D. (2017). Modulating fat digestion through food structure design. *Prog Lipid Res*, *68*, 109-118.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 610-621.
- Janhøj, T., Frøst, M. B., & Ipsen, R. (2008). Sensory and rheological characterization of acidified milk drinks. *Food Hydrocolloids*, *22*, 798-806.
- Jenkins, D. J., Wolever, T. M., Jenkins, A. L., Lee, R., Wong, G. S., & Josse, R. (1983). Glycemic response to wheat products: reduced response to pasta but no effect of fiber. *Diabetes Care*, *6*, 155-159.
- Kirtil, E., & Oztop, M. H. (2016). ¹H nuclear magnetic resonance relaxometry and magnetic resonance imaging and applications in food science and processing. *Food Engineering Reviews*, *8*, 1-22.
- Langton, M., Åström, A., & Hermansson, A.-M. (1997). Influence of the microstructure on the sensory quality of whey protein gels. *Food Hydrocolloids*, *11*, 217-230.
- Langton, M., & Hermansson, A.-M. (1996). Image analysis of particulate whey protein gels. *Food Hydrocolloids*, *10*, 179-191.
- Lehmann, U., & Robin, F. (2007). Slowly digestible starch—its structure and health implications: a review. *Trends in Food Science & Technology*, *18*, 346-355.

- Li, H., Gidley, M. J., & Dhital, S. (2019). High-amylose starches to bridge the “Fiber Gap”: development, structure, and nutritional functionality. *Comprehensive reviews in food science and food safety*, *18*, 362-379.
- Li, H., Yu, W., Dhital, S., Gidley, M. J., & Gilbert, R. G. (2019). Starch branching enzymes contributing to amylose and amylopectin fine structure in wheat. *Carbohydrate polymers*, *224*, 115185.
- Liu, Z., & Scanlon, M. (2003). Predicting mechanical properties of bread crumb. *Food and Bioprocesses processing*, *81*, 224-238.
- Lustria, M. L. A., Noar, S. M., Cortese, J., Van Stee, S. K., Glueckauf, R. L., & Lee, J. (2013). A meta-analysis of web-delivered tailored health behavior change interventions. *Journal of health communication*, *18*, 1039-1069.
- Maeda, T., Kokawa, M., Miura, M., Araki, T., Yamada, M., Takeya, K., & Sagara, Y. (2013). Development of a novel staining procedure for visualizing the gluten–starch matrix in bread dough and cereal products. *Cereal Chemistry*, *90*, 175-180.
- Manthey, F. A., & Schorno, A. L. (2002). Physical and cooking quality of spaghetti made from whole wheat durum. *Cereal Chemistry*, *79*, 504-510.
- McClements, D. J. (2018). The biophysics of digestion: lipids. *Current opinion in food science*, *21*, 1-6.
- Morris, V., & Groves, K. (2013). *Food microstructures: Microscopy, measurement and modelling*: Elsevier.
- Mueller, G. A., Maleki, S. J., Johnson, K., Hurlburt, B. K., Cheng, H., Ruan, S., Nesbit, J. B., Pomés, A., Edwards, L. L., & Schorzman, A. (2013). Identification of Maillard reaction products on peanut allergens that influence binding to the receptor for advanced glycation end products. *Allergy*, *68*, 1546-1554.
- Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., & Andrade, C. H. (2018). QSAR-based virtual screening: advances and applications in drug discovery. *Frontiers in pharmacology*, *9*, 1275.
- Norton, J., Espinosa, Y. G., Watson, R., Spyropoulos, F., & Norton, I. (2015). Functional food microstructures for macronutrient release and delivery. *Food & function*, *6*, 663-678.
- Nyemb, K., Guérin-Dubiard, C., Pézenec, S., Jardin, J., Briard-Bion, V., Cauty, C., Rutherford, S. M., Dupont, D., & Nau, F. (2016). The structural properties of egg white gels impact the extent of in vitro protein digestion and the nature of peptides generated. *Food Hydrocolloids*, *54*, 315-327.
- Ogawa, Y., Donlao, N., Thuengtung, S., Tian, J., Cai, Y., Reginio Jr, F. C., Ketnawa, S., Yamamoto, N., & Tamura, M. (2018). Impact of food structure and cell matrix on digestibility of plant-based food. *Current opinion in food science*, *19*, 36-41.
- Pallares, A. P., Miranda, B. A., Truong, N. Q. A., Kyomugasho, C., Chigwedere, C. M., Hendrickx, M., & Grauwet, T. (2018). Process-induced cell wall permeability modulates the in vitro starch digestion kinetics of common bean cotyledon cells. *Food & function*, *9*, 6544-6554.
- Pereira, R., Singh, H., Munro, P., & Luckman, M. (2003). Sensory and instrumental textural characteristics of acid milk gels. *International Dairy Journal*, *13*, 655-667.
- Romano, A., D'Amelia, V., Gallo, V., Palomba, S., Carputo, D., & Masi, P. (2018). Relationships between composition, microstructure and cooking performances of six potato varieties. *Food research international*, *114*, 10-19.
- Romano, A., Ferranti, P., Gallo, V., & Masi, P. (2021). New ingredients and alternatives to durum wheat semolina for a high quality dried pasta. *Current opinion in food science*, *41*, 249-259.
- Romano, A., Mackie, A., Farina, F., Aponte, M., Sarghini, F., & Masi, P. (2016). Characterisation, in vitro digestibility and expected glycemic index of commercial starches as uncooked ingredients. *Journal of food science and technology*, *53*, 4126-4134.
- Romero, H. M., & Zhang, Y. (2019). Physicochemical properties and rheological behavior of flours and starches from four bean varieties for gluten-free pasta formulation. *Journal of Agriculture and Food Research*, *1*, 100001.
- Salentinig, S. (2019). Supramolecular structures in lipid digestion and implications for functional food delivery. *Current Opinion in Colloid & Interface Science*, *39*, 190-201.

- Scanlon, M., Edwards, N., & Dexter, J. (2005). Pasta: Strength and structure. *New Food*, 8, 10.
- Serial, M. R., Canalis, M. B., Carpinella, M., Valentinuzzi, M. C., León, A., Ribotta, P. D., & Acosta, R. H. (2016). Influence of the incorporation of fibers in biscuit dough on proton mobility characterized by time domain NMR. *Food chemistry*, 192, 950-957.
- Sharma, R., Dar, B., Sharma, S., & Singh, B. (2021). In vitro digestibility, cooking quality, bio-functional composition, and sensory properties of pasta incorporated with potato and pigeonpea flour. *International Journal of Gastronomy and Food Science*, 23, 100300.
- Shi, H., Lei, Y., Louzada Prates, L., & Yu, P. (2019). Evaluation of near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy techniques combined with chemometrics for the determination of crude protein and intestinal protein digestibility of wheat. *Food Chem*, 272, 507-513.
- Singh, H., Ye, A., & Ferrua, M. J. (2015). Aspects of food structures in the digestive tract. *Current opinion in food science*, 3, 85-93.
- Singh, H., Ye, A., & Horne, D. (2009). Structuring food emulsions in the gastrointestinal tract to modify lipid digestion. *Prog Lipid Res*, 48, 92-100.
- Singh, J., Dartois, A., & Kaur, L. (2010). Starch digestibility in food matrix: a review. *Trends in Food Science & Technology*, 21, 168-180.
- Somarathne, G., Ferrua, M. J., Ye, A., Nau, F., Floury, J., Dupont, D., & Singh, J. (2020). Food material properties as determining factors in nutrient release during human gastric digestion: A review. *Critical reviews in food science and nutrition*, 60, 3753-3769.
- Susanna, S., & Prabhasankar, P. (2013). A study on development of Gluten free pasta and its biochemical and immunological validation. *LWT-Food Science and Technology*, 50, 613-621.
- Szczesniak, A. S. (2002). Texture is a sensory property. *Food quality and preference*, 13, 215-225.
- Tan, Y., & McClements, D. J. (2021). Improving the bioavailability of oil-soluble vitamins by optimizing food matrix effects: A review. *Food Chem*, 348, 129148.
- Tan, Y., Zhang, Z., Liu, J., Xiao, H., & McClements, D. J. (2020). Factors impacting lipid digestion and nutraceutical bioaccessibility assessed by standardized gastrointestinal model (INFOGEST): oil. *Food Funct*, 11, 9936-9946.
- Tao, Y., Lee, Y. C., Liu, H., Zhang, X., Cui, J., Mondoa, C., Babaei, M., Santillan, J., Wang, G., Luo, D., Liu, D., Yang, H., Do, Y., Sun, L., Wang, W., Zhang, T., & Yao, L. (2021). Morphing pasta and beyond. *Sci Adv*, 7, eabf4098.
- Tavares, L., Santos, L., & Noreña, C. P. Z. (2021). Bioactive compounds of garlic: A comprehensive review of encapsulation technologies, characterization of the encapsulated garlic compounds and their industrial applicability. *Trends in Food Science & Technology*, 114, 232-244.
- Teodorowicz, M., Van Neerven, J., & Savelkoul, H. (2017). Food processing: The influence of the maillard reaction on immunogenicity and allergenicity of food proteins. *Nutrients*, 9, 835.
- Tester, R. F., & Debon, S. J. (2000). Annealing of starch—a review. *International journal of biological macromolecules*, 27, 1-12.
- Tian, J., Ogawa, Y., Shi, J., Chen, S., Zhang, H., Liu, D., & Ye, X. (2019). The microstructure of starchy food modulates its digestibility. *Critical reviews in food science and nutrition*, 59, 3117-3128.
- Toda, M., Hellwig, M., Henle, T., & Vieths, S. (2019). Influence of the Maillard reaction on the allergenicity of food proteins and the development of allergic inflammation. *Current allergy and asthma reports*, 19, 4.
- Tournier, C., Sulmont-Rossé, C., & Guichard, E. (2007). Flavour perception: aroma, taste and texture interactions. *Food*, 1, 246-257.
- Turgeon, S. L., & Rioux, L.-E. (2011). Food matrix impact on macronutrients nutritional properties. *Food Hydrocolloids*, 25, 1915-1924.
- Ubbink, J., Burbidge, A., & Mezzenga, R. (2008). Food structure and functionality: a soft matter perspective. *Soft Matter*, 4, 1569-1581.
- Van der Sman, R., & Van der Goot, A. (2009). The science of food structuring. *Soft Matter*, 5, 501-510.
- van der Sman, R. G. M., Houlder, S., Cornet, S., & Janssen, A. (2020). Physical chemistry of gastric digestion of proteins gels. *Curr Res Food Sci*, 2, 45-60.

- Witczak, T., & Gałkowska, D. (2021). Sorption and thermal characteristics of ancient grain pasta of various compositions. *LWT*, *137*, 110433.
- Wu, Y., Okesola, B. O., Xu, J., Korotkin, I., Berardo, A., Corridori, I., di Brocchetti, F. L. P., Kanczler, J., Feng, J., Li, W., Shi, Y., Farafonov, V., Wang, Y., Thompson, R. F., Titirici, M. M., Nerukh, D., Karabasov, S., Oreffo, R. O. C., Carlos Rodriguez-Cabello, J., Vozzi, G., Azevedo, H. S., Pugno, N. M., Wang, W., & Mata, A. (2020). Disordered protein-graphene oxide co-assembly and supramolecular biofabrication of functional fluidic devices. *Nat Commun*, *11*, 1182.
- Zhang, Z., Zhang, R., & McClements, D. J. (2019). Establishing the impact of food matrix effects on the bioaccessibility of nutraceuticals and pesticides using a standardized food model. *Food & function*, *10*, 1375-1385.
- Zhu, Z., Zhou, Q., & Sun, D.-W. (2019). Measuring and controlling ice crystallization in frozen foods: A review of recent developments. *Trends in Food Science & Technology*, *90*, 13-25.
- Žolnere, K., Arnold, M., Hull, B., & Everett, D. W. (2019). Cheese proteolysis and matrix disintegration during in vitro digestion. *Food Structure*, *21*, 100114.
- Zou, W., Schulz, B. L., Tan, X., Sissons, M., Warren, F. J., Gidley, M. J., & Gilbert, R. G. (2019). The role of thermostable proteinaceous α -amylase inhibitors in slowing starch digestion in pasta. *Food Hydrocolloids*, *90*, 241-247.
- Zou, W., Sissons, M., Gidley, M. J., Gilbert, R. G., & Warren, F. J. (2015). Combined techniques for characterising pasta structure reveals how the gluten network slows enzymic digestion rate. *Food chemistry*, *188*, 559-568.

Caption to Figures

Figure 1. Food matrix is defined by structures at different length scales consisting of elements spanning nanometres to millimetres and above. Many of the important properties of foods are determined by structural elements at microscale. Molecules such as carbohydrates, proteins, and lipids, indeed form supramolecular clusters that behave as pseudo-molecules of higher molecular weight. Linking organised structural elements to food properties through imaging may be feasible by means of artificial intelligence applications.

Figure 2. The process behind the decomposition of T2 decays raw data into a lower dimensional space. Each time point of each decay is interpreted as variable and fed to a probabilistic PCA with an RBF kernel. Data are transformed according to coefficients which are dependent on the kernel parameters, optimized through machine learning. An example of resulting latent space is showed in the following Figure 3a, where each T2 decay, measured for each different cooking time, is represented as a point in a two-dimensional space.

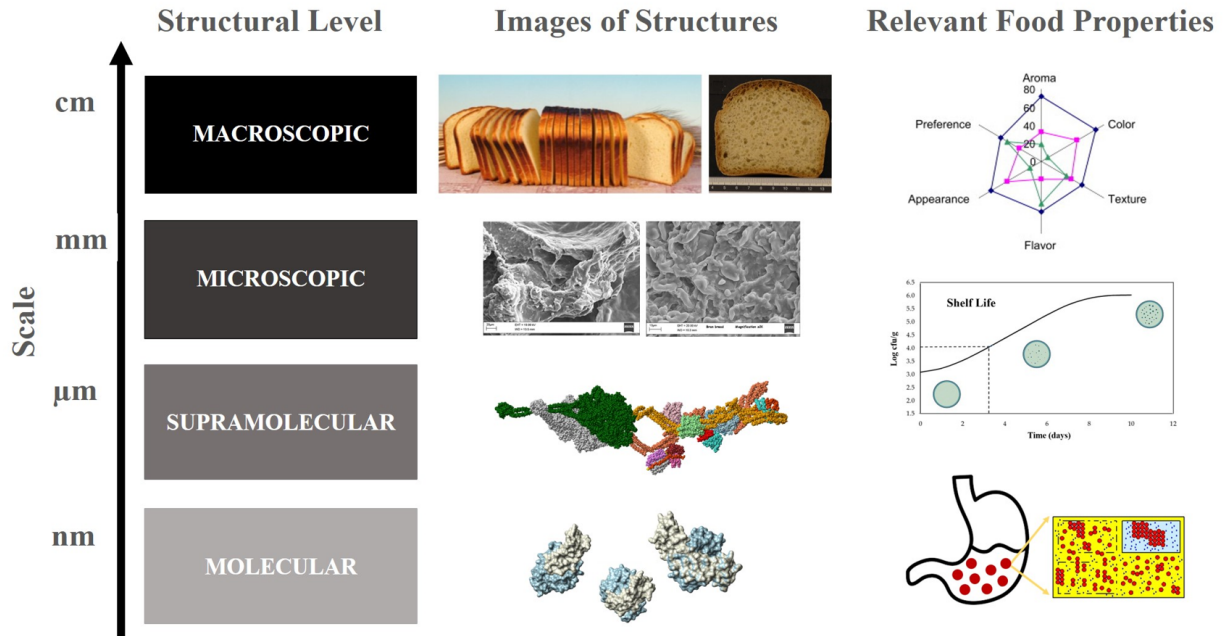
Figure 3. a) Resulting lower dimensional latent space, computed according to Figure 2. In this space, each T2 decay measured at different cooking times (indicated by the colour gradient) is represented as a point. The points are the projection in the 2-d latent features space, learned by the kernel, of each T2 decay. In this space, differences tied to effects of cooking on water mobility are the most detectable. **b)** Scatter plot of PC0 vs HF6 (Sum Average, computed from SEM images of the central zones). A qualitative interpretation of the relationship between these two variables can be given as follows: in the functionality phase, water mobility is mainly related with starch gelatinization phenomena, resulting in little morphological changes. After an activation phase, where the rupture of structures in the food matrix begin to arise, the morphological changes detected in images start a strictly monotonous trend related to cooking time (morphology phase).

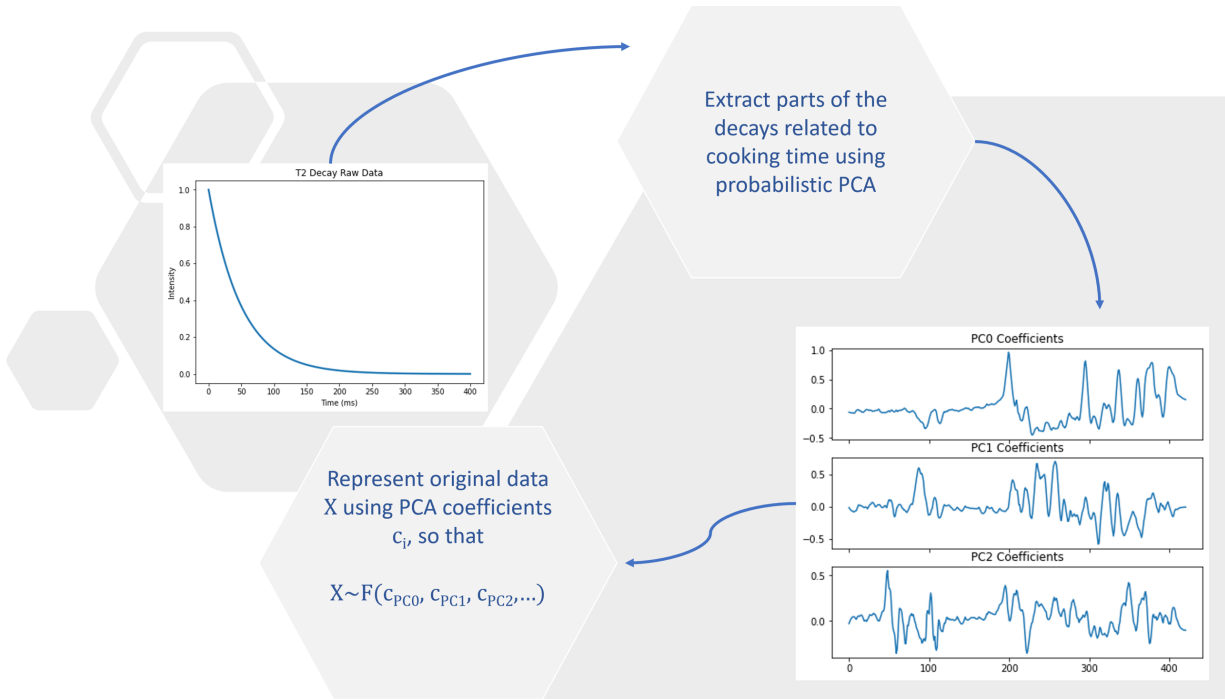
Table 1. Principal methods for structural analyses at characteristic length scales in foods, appearance of food matrix and structural elements

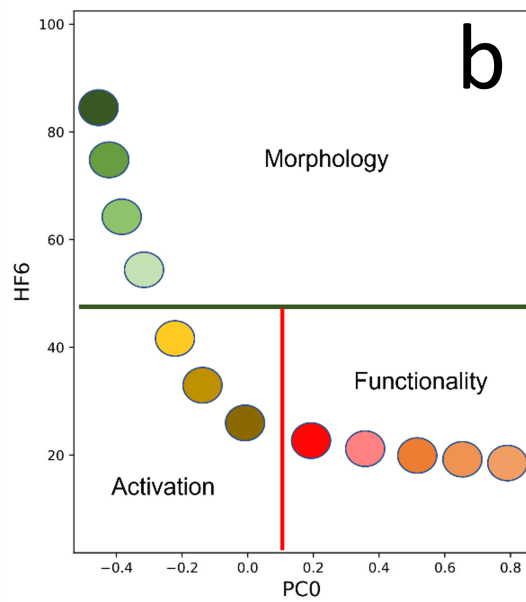
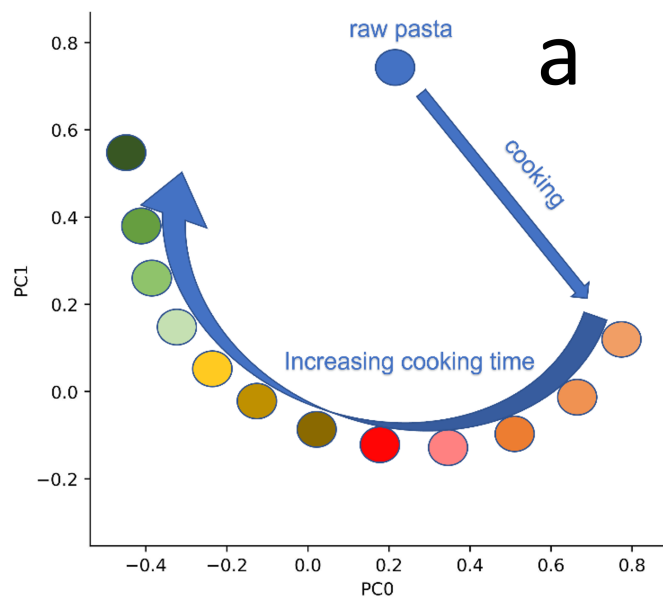
SCALE LENGTH	METHODS	PHYSICAL STATE / STRUCTURAL ELEMENTS	INFORMATION ON:
> 1 cm	<ul style="list-style-type: none"> • Texture analysis • Image analysis • Sensory panel 	liquid, gel, solid, porous solid	<ul style="list-style-type: none"> -properties of network at large deformation - size and shape macrostructural elements -sensorial attributes (e.g., appearance, colour, firmness, overall acceptability)
1 mm – 1 cm	<ul style="list-style-type: none"> • Texture analysis • Microscopy 	liquid -aqueous matrix (<i>aqueous phase in fruit juices</i>), liquid -emulsion matrix (<i>mayonnaise</i>), gels (<i>desserts, processed meats</i>), porous matrix (<i>bread, extruded snacks</i>), viscoelastic matrix (<i>dough</i>), etc.	<ul style="list-style-type: none"> -properties of network at large deformation related to eating properties -microstructure
1 - 500µm	<ul style="list-style-type: none"> • Confocal microscopy • Light microscopy • Rheology 	micelles (<i>casein micelles</i>), droplets, air cells (<i>bread bubbles</i>), crystals (<i>salt</i>), fibres, granules (<i>starch granules</i>), etc.	<ul style="list-style-type: none"> -size and shape of structures -properties of network at small deformation -ingredient interaction
10 -500 nm	<ul style="list-style-type: none"> • Light scattering • Electron microscopy 	micelles (<i>casein micelles</i>), droplets, air cells (<i>bread bubbles</i>), crystals (<i>salt</i>), fibres, granules (<i>starch granules</i>), etc.	<ul style="list-style-type: none"> -aggregation, density, arrangement -size of structures
< 10nm	<ul style="list-style-type: none"> • Raman • Chromatography • Thermal analysis • SDS Page • NIR 	carbohydrates (<i>starch</i>), proteins (<i>gluten, caseins</i>), lipids, water, etc.	<ul style="list-style-type: none"> -molecular structure -proportion of elementary parts -unfolding vs. native -denaturation /transition temperature

Table 2. Main descriptors and (dis)advantages for electronic microscopy and magnetic resonance imaging

	<u>SEM</u>	<u>MRI</u>
Descriptors	<ul style="list-style-type: none"> • Particle size and morphology • Pore size and morphology • Size distribution and morphology • Shape orientation (e.g., fibres) and diameter distribution (e.g., beads) 	<ul style="list-style-type: none"> • First order grey level statistics (e.g., Histogram of grey levels statistics, symmetry of grey levels centred about the mean, entropy of the image) • Roughness of textures • Degree of linearity • Co-occurrence matrix statistics (e.g., Haralick moments) • Structural or morphological features of ROIs (e.g., Bounding ellipsoid volume ratios) • Transform features (features extracted in frequency domains)
Pros & Cons	<ul style="list-style-type: none"> • Not immediately suitable for high-throughput production (parameter dependent acquisitions: lighting, magnification etc.) • No data harmonization standard due to heterogeneous necessities of application fields and experiments • Widely applied in many fields • Canonical descriptors immediately linkable with physical quantities • Very high resolution • Requires specific assumptions for image analysis (i.e., presence/absence of certain geometrical structures, pores, shapes etc.) 	<ul style="list-style-type: none"> • Inherently suitable for high-throughput data production • Data harmonization standards are widely supported in many biomedical fields (neuro imaging, imaging for oncology) • Descriptors comes from low-level, general texture analysis and morphological studies alike • Low resolution • Does not require specific assumptions for image analysis, due to canonical analysis based upon general first order statistics of grey levels and moments of cooccurrence matrix.







- Supramolecular structure is important for in-silico design of functional foods
- Models based on artificial intelligence may predict optimal food structures
- Water-matrix interactions and structure must be included in digital twin of food