Innovative Food Science and Emerging Technologies 43 (2017) 163–172

Contents lists available at ScienceDirect

Innovative Food Science and Emerging Technologies

journal homepage: www.elsevier.com/locate/ifset

Whole soybean protein extraction processes: A review

K.E. Preece[a,*], N. Hooshyar[b], N.J. Zuidam[c]

[a] School of Chemical Engineering, University of Birmingham, Edgbaston B15 2TT, United Kingdom
[b] Unilever R & D, Colworth Science Park, Sharnbrook, Bedford MK44 1LQ, United Kingdom
[c] Unilever R & D Vlaardingen, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands

ARTICLE INFO

Keywords:
Aqueous extraction
Microstructure
Soymilk
Cavitation

ABSTRACT

Soybeans are an important raw material for those seeking vegan, lactose-free products, such as soymilk and tofu. The aim of this review article is to provide an overview of aqueous extraction of protein and other desirable components from whole soybeans. Firstly, a discussion over the microstructure of the soybean is held, including a summary of protein localisation and properties. A detailed review of common whole soybean extraction process is then given, along with extraction process parameters and process intensification steps that can improve yield. A novel extraction model is presented, based on a mass balance of the water phases. The extraction model reveals separation as the main limitation for protein recovery during aqueous soy protein extraction due to the high amount of okara waste stream and its high moisture content.

Industrial relevance: Typically, the extraction of protein from an intermediate soy-protein ingredient is studied at lab-scale. Within industry, aqueous extract from whole soybeans is commonly used for making consumer products containing both soy protein and soybean oil, and this has been the focus of this review. Key extraction process parameters are presented and challenges of each extraction step are given for the whole soybean extraction process. A novel model for determining the separation efficiency has been presented, which is useful for many other extraction systems that contain components of interest in high amounts of waste stream.

1. Introduction

Plant-based protein is more sustainable than animal-based protein when comparing fossil fuel usage, land use and water consumption (González, Frostell, & Carlsson-Kanyama, 2011). With the human population projected to increase to 9.5 billion by the year 2050 (Reynolds, Wulster-Radcliffe, Aaron, & Davis, 2015), a greater portion of the nutrients required for human nutrition will be supplied by plant-based sources. The first generation plant-based protein source for human consumption has been the soybean. Reportedly the consumption of soybeans can be dated back as early as 3rd century BC in China (Huang, DuBois, Tan, & Mintz, 2008). The consumption of soy has gained popularity in the western world over recent decades due to:

- Increased knowledge of the consumer and drive for a healthier lifestyle
- Increased prevalence of lactose intolerance
- Improved processing of soybeans with reduced off-flavour (Debruyne, 2006)

The increased consumption of soy-based products leads to the incentive for more sustainable soybean processing, by reducing the carbon footprint and/or greenhouse gas emissions compared to current processes. Not only is sustainability a motivation for industry, but financial gain is also made possible through improved utilisation of the raw material. Table 1 shows an overview of the typical oil and protein contents of soybean and the main soy-derived commercially available ingredients. The summarised production methods for these soy-ingredients can be seen in Fig. 1.

Defatted soybean flakes are common by-products from oil extraction, the most common component utilised from the oilseed. Soybeans are crushed in a roller mill and then the oil is extracted using a solvent, typically hexane. Hexane-based processing can lead to the production of greenhouse gases and concerns regarding safety due to the flammable nature of the solvent (Rosenthal, Pyle, & Niranjan, 1996). The remaining solvent within the soybean matrix is removed via heat evaporation. Mechanical extraction can also be employed; however, compared to solvent extraction, the oil yield is not as lucrative. Defatted soy flour refers to the same material as defatted soybean flakes, but with a finer particle size. It can be used as a feed, yet more value can be

Abbreviations: CLSM, Confocal laser scanning microscopy; d.b., dry basis; HPH, High pressure homogenisation; SEM, Scanning electron microscopy; SPC, Soy protein concentrate; SPI, Soy protein isolate; TEM, Transmission electron microscopy; w.b., wet basis

* Corresponding author.
E-mail address: katherine.preece@me.com (K.E. Preece).

http://dx.doi.org/10.1016/j.ifset.2017.07.024
Received 31 October 2016; Received in revised form 3 March 2017; Accepted 13 July 2017
Available online 16 July 2017

Nomenclature

S	Soybase mass
B	Soybean mass
W	Water mass
O	Okara mass
x_i	Mass fraction of component i
$x_{i,j}$	Mass fraction of component i in stream j
i	p, Protein; w, Moisture
j	s, Soybase; b, Soybean; w, Water; o, Okara

created when the proteins are extracted from it. To produce soy protein concentrate (SPC), defatted soybean flakes are added to alcohol or water to remove carbohydrates. Soy protein isolate (SPI) contains a higher protein content than SPC (see Table 1) due to the removal of insoluble carbohydrate and dietary fibres via an intermediate, acidic precipitation step.

A less commonly used extraction route is the whole soybean extraction (far left process in Fig. 1). The aqueous extract of whole soybean extraction is called soybase and it is mainly used for making consumer products containing both soy protein and soybean oil. Products like soymilk, soy-fruit beverages and tofu are produced by adding various ingredients to the soybase, such as flavours, gums, stabilisers, minerals, vitamins, sugars, fruit juices and/or coagulating agents in case of tofu. At industrial scale, the use of soybase may result in consumer products with better sensory properties and might be a commercially more attractive route than first isolating soy protein and soybean oil separately and then blending them together at a later stage. However, a large quantity of protein, oil and other components exit the process in the waste stream (30–40% of total protein, depending on the exact conditions). The waste stream, referred to as okara in the field, is typically utilised as animal feed (Li, Qiao, & Lu, 2012).

This review article provides an overview of the latest insights in whole soybean extraction processing and the emerging technologies employed to aid especially protein extraction. Firstly, the soybean composition and microstructure is discussed, as insights from these can be gained into extraction processes. Then an overview of these most common processes is given, as well as a more detailed discussion of the challenges and opportunities of each extraction step. Finally, a novel extraction model will highlight the location of greatest losses of protein during soybean processing. Findings in this area of research are also beneficial for the advancing generations of other plant-based protein sources, such as pea, canola and lupin, as well as many other extraction systems that contain components of interest in high amounts of waste stream.

2. Soybean composition and microstructure

The composition of soybeans can vary; for the production of soy-based products, such as soymilk and tofu, a strain of soybean containing relatively high amounts of proteins should be selected. Other criteria for soybean selection include the colour and sensory properties of its extract in the final consumer product. The composition of a typical soybean for producing soymilk on a wet and dry basis can be seen in Fig. 2 (Imram, Gomez, & Soh, 2003). Soybeans used for soybean oil production are commonly richer in oil and lower in proteins (about 40 and 20%, respectively). Soybeans are considered as the most important legume as they are one of few vegetal materials containing all of the essential amino acids required for human development. Dairy alternative products prepared from soybeans are not only selected for their high protein content, but also for their lack of cholesterol and lactose.

Table 1
Typical variations in protein and oil contents on a wet basis (w.b.) for soybean and some of its commercial ingredients. *Values derived from Riaz (2006) and own observations.

Product	Typical protein content (w.b. %)	Typical oil content (w.b. %)
Soybean	40	20
Soybase	4–5	2
Soy flour	40	20
Defatted soy flour/flakes	44–54*	0.5–1*
Soy protein concentrate (SPC)	65–70*	Trace*
Soy protein isolate (SPI)	85–90*	Trace*

For optimal extraction of components from soybeans, it is vital to understand the structure located within the soybean. There are a number of structures which make up the soybean: the hull, the hypocotyl axis and predominantly cotyledon cells (Campbell et al., 2011). Fig. 3 shows an image of soybeans and their microstructure after pre-soaking. The main constituent of the soybean, cotyledon cells, are organised within the bulk in a space-filling arrangement. Cotyledon cells are 70–80 μm in length and 15–30 μm in width (Campbell & Glatz, 2009; Rosenthal, Pyle, & Niranjan, 1998). Hydration may cause the cells volume to increase.

The cell wall of the soybean cotyledon is comprised of a series of polysaccharides, which are often cross-linked with proteins and phenolic compounds (lignin) (Ouhida, Perez, & Gasa, 2002). The primary cotyledon cell wall contains pectins, hemicelluloses and microfibrils of cellulose cross-linked with proteins (Campbell et al., 2011). There is a secondary cell wall within the primary wall containing cellulose and hemicelluloses also capable of binding to proteins. Cells are held together by adhesive substances found in the middle lamella, the extracellular space between cells, and contain pectins, glycine and hydroxyproline-rich proteins (Campbell et al., 2011; Kasai, Imashiro, & Morita, 2003).

2.1. Soybean oil

Oil consists of approximately 88.1% triglycerides, 9.8% phospholipids, 1.6% unsaponifiable components and 0.5% free fatty acids (Salunkhe, Chavan, Adsule, & Kadam, 1992). The majority of oil is located in oil bodies (oleosomes) within the cotyledon cells (Waschatko, Junghans, & Vilgis, 2012). Oil bodies are found within the cytoplasmic network of the cells and are stabilised by small molecular weight proteins termed oleosins (Rosenthal et al., 1998), which make them more hydrophilic and easy to extract aqueously. The oil bodies typically vary in size from 0.2–0.5 μm (Campbell & Glatz, 2009). Fig. 4 shows a SEM image of a dry soybean. Oil bodies are observed in this micrograph, as well as other components of interest, most notably protein bodies and phytic acid (Preece, Hooshyar, Krijgsman, Fryer, & Zuidam, 2017b).

2.2. Soy proteins

The majority of proteins are organised in protein bodies of the cotyledon cells, labelled in Fig. 4. According to Preece et al. (2017b), the protein bodies within the cotyledon cells were found to be in the size range 2.4 to 13.5 μm when examined using SEM without sample hydration. These values fell on the low side when compared to values recorded using transmission electron microscopy (TEM) of 2 to 20 μm on hydrated soybeans (Rosenthal et al., 1998). It has been reported previously (White, Welsby, & Kolar, 2013) that protein bodies swell upon hydration with water at neutral pH to double their original size, confirming these findings.

There are two major storage proteins that account for typically 60–80% of the total soybean protein: the globulins glycinin (11S) and β-conglycinin (7S) (Murphy, 2008). At neutral pH and ambient temperature, glycinin (11S) is a hexameric complex comprised of acidic and basic polypeptides linked by disulphide bridges to provide a molecular weight in the range 320–375 kDa (Lakemond, de Jongh, Hessing, Gruppen, & Voragen, 2000). β-conglycinin (7S) contains three major subunits (β, α and α’) reportedly with sizes of 50, 67 and 71 kDa,