The data information provided in this article relate to our research article “Using patient serum to epitope map soybean glycinins reveals common epitopes shared with many legumes and tree nuts” (Saeed et al., 2016) [1]. Here we provide western blot detection of glycinin subunits by soy-sensitive human sera, ELISA screens with overlapping synthetic peptides (epitope mapping), and various database/server epitope searches.

© 2016 Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

**Specifications Table**

| Subject area | Immunology |
|--------------|------------|
| More specific subject area | Allergy |
| Type of data | Tables, Graphs, Figures |
| How data was acquired | Western blots were performed by screening total soy protein on 2D gels with soy-sensitive human sera and detecting with a secondary anti-IgE-HRP antibody. ELISAs were performed by screening a collection of synthetic peptides encompassing the glycinin sequences with soy-sensitive human sera. The IgE binding to the peptides was detected by a secondary anti-IgE-HRP antibody. |

DOI of original article: http://dx.doi.org/10.1016/j.molimm.2015.12.008

* Corresponding author.

E-mail address: steve.gleddie@agr.gc.ca (S. Gleddie).
Epitope sequence similarity searches were done using the SDAP website: [http://fermi.utmb.edu/](http://fermi.utmb.edu/)
B-cell epitope predictions were done using the following servers:
- ABCpred [http://www.imtech.res.in/raghava/abcpred/](http://www.imtech.res.in/raghava/abcpred/)
- BepiPred 1.0 [http://www.cbs.dtu.dk/services/BepiPred/](http://www.cbs.dtu.dk/services/BepiPred/)
- SVMTriP [http://sysbio.unl.edu/SVMTriP/](http://sysbio.unl.edu/SVMTriP/)

### Value of the data

- Better understanding of soy storage protein allergens may contribute to allergy management strategies.
- It may also contribute to the generation of hypoallergenic soybean cultivars.
- Provide risk assessment tools for the evaluation and characterization of the allergenicity of novel foods.

### 1. Data

The data presented here show the western blot detection of A2 or A3 subunits by soy-sensitive human sera (Fig. 1) and ELISA screens (Figs. 2 and 3) of these patient sera with overlapping synthetic peptides (Pepsets). Serum specificity is also confirmed by cross-screening the A2 Pepset with a serum that does not bind to the A2 cluster on western blot (Fig. 4). Also contained in this article is SDAP (Structural Database of Allergenic Proteins) sequence similarity search results (Tables 1 and 2) of the epitopes reported by Saeed et al. (2016) [1] and theoretical B-cell epitope prediction data on the full length sequences of A2 and A3 subunits (Table 3).

### 2. Experimental design, materials and methods

#### 2.1. Patient serum

Soy-sensitive human sera used in the western blots and epitope mapping are previously described [1].

#### 2.2. Immunoblot analysis

Western blotting of human sera was conducted as previously described [2]. Membranes were hybridized with serum dilutions ranging from 1/50 to 1/500.

#### 2.3. Epitope mapping

Two peptide sets representing the mature amino acid sequences of glycinin A2 (P04405, 90 peptides) and A3 (BAB15802, 104 peptides) were synthesized and biotinylated by Mimotopes [http://
Fig. 1. 2D western blots of patient sera on soybean seed protein gels (bottom 6 panels). Top panel is a Coomassie-stained 2D gel illustrating the A1/A2 (red) and A3 (blue) glycinins.
A2 screening

Patient 1

A2.4
#66: 5.2
#66: 5.2

Patient 2

A2.1
#34: 3.7
#34: 3.4
#42: 8.2
#89: 4.0

A2.3
#40: 4.7
#88: 6.8
#88: 8.6

A2.6
#40: 4.5
#89: 7.1

Patient 3

A2.2
#38: 4.6

A2.3
#42: 4.4

A3 screening

Patient 4

A3.1
#62: 6.5
#62: 3.9
#63: 40.6
#63: 24.6
#64: 57.7
#64: 22.6

Patient 5

A3.3
#97: 30.8
#97: 10.2

Patient 6

A3.2
#68: 3.4
#68: 4.4

www.mimotopes.com) via parallel array platform. Quality Control Assurance was provided for both peptide synthesis and biotinylation by reverse phase HPLC (RP-HPLC), and by mass spectrometry (MS) respectively. The biotinylated 12-mer peptides, frame-shifted by three residues were used as per manufacturer's instructions (Application/Method PT3013). DMSO was used to resuspend the dry peptides and streptavidin–coated high capacity plates (Pierce #15500) pre-blocked with SuperBlock™

Fig. 2. Epitope mapping of A2 and A3 Pepsets with patient serum 1–6. X-axis indicates Pepset peptide number and y-axis indicates ratio of colorimetric detection in the patient sample vs control sample.
buffer were used to capture the biotinylated peptides. Serum was diluted at 1/50 in TBS-BSA 2% except for Patients 4 (1/100) and 5 (1/50 or 1/100). The secondary mouse anti-human IgE-HRP (Southern Biotech, Birmingham, Alabama, #9160-05) was diluted at 1/4000 in TBS-BSA 2%. SureBlue Reserve™ TMB microwell peroxidase substrate (KPL, Gaithersburg, Maryland, #53-00-01) was added to the plate, the reaction was stopped by acidification and colorimetric detection was performed on a Tecan Sunrise microplate reader with Magellan™ data analysis software (Tecan group AG, Männedorf, Switzerland) at 450 nm. Each experiment was performed in duplicate. Negative controls were performed using the same protocol, but the addition of human sera was omitted. The data was normalized by calculating the ratio of experimental to negative control and graphed.

Fig. 3. Epitope mapping of A2 and A3 Pepsets using pooled sera. X-axis indicates Pepset peptide number and y-axis indicates ratio of colorimetric detection in the patient sample vs. control sample.

Fig. 4. Serum specificity control. Screening of A2 Pepset with Patient 4 serum which only bound to A3 subunit on western blot (see Fig. 1). X-axis indicates Pepset peptide number and y-axis indicates ratio of colorimetric detection in the patient sample vs control sample.
2.4. B-cell epitope prediction servers

Three popular B-cell epitope prediction servers were tested with the A2 and A3 sequences. ABCpred server predicts B cell epitopes using a recurrent neural network (machine based technique) using fixed length patterns [3]. Lengths of epitopes varying from 10–16 amino acids were tested. BepiPred 1.0 server uses a combination of a hidden Markov model and a propensity scale method [4]. SVMTriP uses support vector machine integrating tri-peptide similarity and propensity scores [5], where epitope lengths varying from 10–20 amino acids were tested. In all cases, a higher score reflects a higher probability that a sequence is an epitope.

**Table 1**
SDAP sequence similarity of A3 epitopes (food only).*

| Rank | Allergen       | Source | PD index | Location | Matching region                      |
|------|----------------|--------|----------|----------|--------------------------------------|
| 1    | A3.1 epitope   | Soybean| 0.00     | 214–222  | KQGQHQHQQE                           |
| 2    | A4             | Soybean| 1.15     | 214–222  | KQGQHQHQQE                           |
| 3    | Vig r 2.0201   | Mung bean| 3.82  | 218–226  | QQGQESQQQ                           |
| 4    | Tri a glutenin | Wheat  | 3.83     | 191–199  | QQGQSSQQQ                           |
| 5    | Pru du 6.01    | Almond | 3.87     | 100–108  | QQGRRQQQEQ (epitope HS5#5)            |
| 6    | Pru du 6.01    | Almond | 3.87     | 136–144  | QQGRRQQQEQ (epitope HS5#5)            |
| 7    | Pru du 6.02    | Almond | 3.87     | 209–217  | QQGRRQQQEQ (epitope)                 |
| 10   | A1a            | Soybean| 4.02     | 205–213  | QQGKHIQQQE                           |
| 12   | Pru du 6.01    | Almond | 4.06     | 122–130  | QQGQQEOQQQ (epitope HS5#5)            |
| 14   | Tri a gliadin  | Wheat  | 4.13     | 23–31    | QQQQQQEQQ                           |
| 15   | Tri a gliadin  | Wheat  | 4.34     | 142–150  | QQQQQQQQQQ                          |
| 16   | Tri a gliadin  | Wheat  | 4.48     | 260–268  | QQPQQQQQQQ                          |
| 18   | Tri a gliadin  | Wheat  | 4.84     | 208–216  | HQQQQQQQQEQ                          |
| 20   | Bra j 1        | Indian mustard | 4.93  | 88–96    | QQGQQQLOQE                           |
| 21   | Bra r 1        | Field mustard | 4.95  | 134–142  | QQGQQQQMQQ                           |
| 22   | Tri a gliadin  | Wheat  | 4.97     | 119–127  | QQQAQQQQQQ                           |
| 29   | β-conglycinin (α) | Soybean| 5.16     | 433–441  | EQQQQRQQQQ                           |
| 1    | A3.2 epitope   | Soybean| 0.00     | 226–237  | GSVLGSFSKHFL                         |
| 2    | A4             | Soybean| 0.00     | 227–238  | GSVLGSFSKHFL                         |
| 3    | Len c 1.0101   | Lentil | 3.80     | 144–155  | PSFLSGFSKNIL                         |
| 4    | A1a            | Soybean| 4.26     | 218–229  | GSILSGFTLEFL (epitope)               |
| 6    | A1b            | Soybean| 4.66     | 217–228  | GSILSGFAPEFL                         |
| 7    | Len c 1.0102   | Lentil | 4.71     | 144–155  | PSFLSGFNKSL                          |
| 8    | Ara h 3        | Peanut  | 5.39     | 237–248  | GNFSGFPTEFL (epitope)                |
| 11   | Vig r 2        | Mung bean| 5.86   | 183–194  | QSYLQGFSSNIL                         |
| 12   | β-conglycinin (α) | Soybean| 5.86     | 366–377  | QSYLQGFSSNIL                         |
| 14   | Pru du 6.01    | Almond | 6.16     | 269–280  | NNVFSGNTQQLL                         |
| 16   | A2             | Soybean| 6.17     | 215–226  | SNLGSFAPPEFL (A2.5 epitope)           |
| 18   | Vig r 2        | Mung bean| 6.30   | 181–192  | QSYLQGFSSNIL                         |
| 19   | β-conglycinin (α) | Soybean| 6.46     | 331–342  | QSYLQGFSSNIL                         |
| 21   | Fag e 1        | Buckwheat | 6.49  | 241–252  | ANILSGQDEIL                          |
| 22   | Jug r 4        | English walnut | 6.54  | 231–242  | NNVFSGFDADF (epitope)                |
| 23   | Cor a 9        | Hazelnut | 6.57   | 238–249  | NNVFSGFDEFL (epitope)                |
| 24   | Car a 14       | Pecan   | 6.57     | 232–243  | NNVFSGFDEFL (epitope)                |
| 25   | Ara h 1        | Peanut  | 6.57     | 304–315  | SYYYLGFSRNITL                        |
| 1    | A3.3 epitope   | Soybean| 0.00     | 313–324  | EEDQPPRPDPHPH                        |
| 2    | A4             | Soybean| 4.84     | 316–327  | EEDKPRPSRPS                          |
| 3    | Lup an 1.0101  | Lupin   | 7.00     | 101–112  | EQQQOPQQQRQ                          |

* Only scores up to 8 are indicated.

2.4. B-cell epitope prediction servers

Three popular B-cell epitope prediction servers were tested with the A2 and A3 sequences. ABCpred server predicts B cell epitopes using a recurrent neural network (machine based technique) using fixed length patterns [3]. Lengths of epitopes varying from 10–16 amino acids were tested. BepiPred 1.0 server uses a combination of a hidden Markov model and a propensity scale method [4]. SVMTriP uses support vector machine integrating tri-peptide similarity and propensity scores [5], where epitope lengths varying from 10–20 amino acids were tested. In all cases, a higher score reflects a higher probability that a sequence is an epitope.
### Table 2
SDAP sequence similarity of A2 epitopes (food only).

| Rank | Allergen | Source | PD index | Location | Matching region |
|------|----------|--------|----------|----------|-----------------|
| 1    | A2.1 epitope | soybean | 0.00 | 121–129 | QRPQDRHQK |
| 4    | A1a | soybean | 1.36 | 124–132 | SRFQDRHQK |
| 5    | A1b | soybean | 1.36 | 121–129 | SRFQDRHQK |
| 6    | Tri a gliadin | wheat | 4.72 | 241–249 | QQPQQQQQQ |
| 8    | Pru du 6.02 | almond | 4.99 | 137–145 | EDQQDRHQK |
| 9    | Pis v 5 | pistachio | 4.99 | 126–134 | SRQFDHQK |
| 10   | Ara h 3 | peanut | 5.01 | 138–146 | QQQQDSHQK |
| 11   | Ana o 2 | ashew | 5.22 | 118–126 | GRFQDRHQK (2aa-epitope) |
| 12   | Lup an 1 | Lupine | 5.27 | 126–134 | QRPSRREE |
| 15   | Cor a 9 | Hazelnut | 5.59 | 138–146 | RSEQDRHQK (epitope) |
| 16   | Jug r 4 | English walnut | 5.70 | 210–218 | RQQQRQQQR (epitope) |
| 17   | A4 | Soybean | 5.83 | 129–137 | QQLQDSHQK |
| 18   | A3 | Soybean | 5.83 | 129–137 | QQLQDSHQK |
| 19   | Ses i 7 | Sesame | 5.83 | 141–149 | RRFMDRHQK |
| 20   | Tri a gliadin | Wheat | 5.86 | 134–142 | QQQQQQQQQ |
| 21   | Lup an 1 | Lupine | 5.90 | 51–59 | QQPRPQQ |
| 24   | Car i 4 | Pecan | 6.14 | 131–139 | EFQQDRHQK (epitope) |
| 25   | Jug r 4 | English walnut | 6.14 | 130–138 | EFQQDRHQK (epitope) |
| 26   | Ber e 2 | Brazil nut | 6.18 | 128–136 | GRFQDRHQK |
| 27   | Car i 4 | Pecan | 6.28 | 210–218 | HRRQQQHQKQ (epitope) |
| 28   | Lup an 1 | Lupine | 6.32 | 583–591 | AQPQQQQQQ |
| 31   | β-conglycinin (α’) | Soybean | 6.46 | 147–155 | PRPQHQK |
| 34   | A2 | Soybean | 6.51 | 109–117 | QEQPSQQR |
| 35   | Tri a gliadin | Wheat | 6.56 | 210–218 | QQQQQQQQK |
| 37   | Tri a glutenin | Wheat | 6.58 | 191–199 | QQQQGQQQ |
| 38   | Vlg r 2 | mung bean | 6.60 | 330–338 | QREQQQQ |
| 39   | Tri a gliadin | wheat | 6.60 | 92–100 | QQPPQQQQ |
| 40   | Ana o 1 | cashew | 6.61 | 42–50 | QKYQDEQQK |
| 42   | Ara h 6 | peanut | 6.63 | 54–62 | TRSSDDQQR |
| 43   | Tri a gliadin | wheat | 6.65 | 137–145 | QQQQQQQQ |
| 46   | Jug r 2 | English walnut | 6.78 | 111–119 | QRRQDRQDP |
| 1    | A2.2 epitope | Soybean | **0.00** | **130–141** | VHRFREGDJIAV |
| 4    | Ara h 3 | Soybean | 1.31 | 124–135 | VHRFDREGDJIAV |
| 6    | A1b | Soybean | 2.71 | 130–141 | IYHFREQDQIJAV |
| 8    | A1a | Soybean | 2.77 | 133–144 | IVNREQDFALAV |
| 9    | Ses i 6 | Sesame | 3.21 | 144–155 | VHRLQGRDIVAI |
| 10   | Pis v 5 | pistachio | 3.40 | 135–146 | IQRFFKCDIAL |
| 11   | Ana o 2 | Cashew | 4.02 | 127–138 | IRRFRQDJIAL |
| 12   | Cor a 9 | Hazelnut | 4.08 | 147–158 | IRHFREGDIJAIV (5aa-epitope) |
| 13   | Pis v2 | Pistachio | 4.25 | 151–162 | VRHIREGDIJAIV |
| 14   | Car i 4 | Pecan | 4.31 | 140–151 | IRHFREGDIJAIV |
| 15   | Jug r 4 | English walnut | 4.31 | 139–150 | IRHFREGDIJAIV (5aa-epitope) |
| 16   | Ber e 2 | Brazil nut | 4.58 | 137–148 | VHLKKGIDJIAIV |
| 17   | Pru du 6.02 | Almond | 4.70 | 146–157 | IRHFREGDIJAIV |
| 18   | Pru du 6.01 | Almond | 4.70 | 193–204 | TRRIERGDIVAI |
| 20   | Ses i 7 | Sesame | 4.73 | 150–161 | VRQFRGQIDIAIV |
| 21   | Pis v2 | Pistachio | 5.84 | 146–157 | VRQPIRQGDIVAI |
| 22   | Sin a 2 | White mustard | 6.61 | 181–192 | VEVRHRGDAIAM |
| 23   | Fag e 1 | Buckwheat | 6.94 | 161–172 | IFRQREGDQIPS |
| 24   | A3 | Soybean | 7.07 | 138–149 | IRHFNGQDLV1 |
| 25   | A4 | Soybean | 7.07 | 138–149 | IRHFNGQDLV1 |
| 1    | A2.3 epitope | Soybean | **0.00** | **136–153** | GDJIAVPTVGA/VWWMYNN |
| 2    | A1a | Soybean | 0.00 | 139–156 | GDJIAVPTVGA/VWWMYNN |
| 5    | A1b | Soybean | 1.22 | 136–153 | GDJIAVPTVGA/FAYWMYNN |
| 6    | Ara h 3 | Peanut | 2.21 | 133–150 | GDJIAVPTVGA/FWLYNDH |
| 9    | Pis v2 | Pistachio | 4.67 | 157–174 | GDIAILPGAVNIVWNNG |
| 11   | Ber e 2 | Brazil nut | 4.92 | 143–160 | GDIAILPGAVNLYCDNG |
| 12   | Pru du 6.02 | Almond | 4.92 | 152–169 | GDIAILPGAVQYWSYNQ |
| 13   | Cor a 9 | Hazelnut | 5.10 | 153–170 | GDIAILPGAVHWNCYD |

H. Saeed et al. / Data in Brief 8 (2016) 123–131
### Table 2 (continued)

| Rank | Allergen | Source | PD index | Location | Matching region |
|------|----------|--------|----------|----------|-----------------|
| 14   | Car i 4  | Pecan  | 5.26     | 146–163  | GDIIFAPGVAHWYNDG |
| 15   | Ana o 2  | Cashew | 5.40     | 133–150  | GDIIFAPGVAHWYNDG |
| 16   | Pru du 6.01 | Almond | 5.43     | 179–196  | GDVIAIPGAVWYNSYD |
| 18   | Ses i 6  | Sesame | 5.75     | 150–167  | GDIIFAPGVAHWYNDG |
| 19   | Jug r 4  | English walnut | 5.80 | 145–162 | GDIIFAPGVAHWYNSYD |
| 20   | Pis v 5  | Pistachio | 6.10 | 141–158 | GDIAPGVAHWYNSYD |
| 21   | Ses i 7  | Sesame | 6.18     | 156–173  | GDIAPGVAHWYNSYD |
| 22   | A3       | Soybean | 7.83     | 144–161  | GDVLVPGPVPVWTYNTG |
| 23   | A4       | Soybean | 7.83     | 144–161  | GDVLVPGPVPVWTYNTG |
| 24   | A2.4 epitope | Soybean | 0.00     | 214–225  | GSNILSGFAPEF |
| 3    | A1b      | Soybean | 1.52     | 216–227  | GSNILSGFAPEF |
| 4    | Ara h 3  | Peanut | 2.76     | 236–247  | GNIFSFGTPFE (epitope HS#2) |
| 7    | Fag e 1  | Buckwheat | 4.61 | 240–251 | GANILSGFQE |
| 8    | A1a      | Soybean | 4.81     | 217–228  | GSNILSGFTPFE (epitope HS#2) |
| 10   | Cor a 9  | Hazelnut | 5.31     | 237–248  | GNNVGFSGDAEF (epitope HS#2) |
| 11   | Car i 4  | Pecan  | 5.31     | 231–242  | GNNVGFSGDAEF (5aa-epitope HS#2) |
| 12   | Pru du 6.02 | Almond | 5.92     | 225–236  | GNNVGFSGDAEF (epitope HS#2) |
| 13   | Sin a 2  | White mustard | 6.16 | 248–259 | QQNLSGDPDPQV |
| 14   | Jug r 4  | English walnut | 6.17 | 230–241 | GNNVGFSGDAEF (epitope HS#2) |
| 15   | A4       | Soybean | 6.17     | 226–236  | GGSVLSGFSKHF |
| 16   | A3       | Soybean | 6.17     | 225–236  | GGSVLSGFSKHF (A3.2 epitope) |
| 17   | Pru du 6.01 | Almond | 6.38     | 288–299  | GNNVGFSGDAEF (epitope HS#2) |
| 18   | Ana o 2  | Soybean | 6.54     | 167–178  | SHSLSGFEPAI |
| 20   | Gly m Bd28K | Soybean | 6.54    | 167–178  | SHSLSGFEPAI |
| 22   | Pis v 2  | Pistachio | 7.21     | 229–240  | SNILSADFDEI |
| 25   | Ses i 7  | Sesame | 7.67     | 225–236  | TKNIIFGDFDEI |
| 26   | Gal d vitellogenin (Gal d 6) | egg | 7.85     | 771–782  | ANQILNASIQW |

### Table 3

#### B-cell epitope prediction.

| Name | Amino acid | Sequence | ABCpred | Bepipred | SVMTriP |
|------|------------|----------|---------|----------|---------|
| A3.1 | 214–222    | KQGQHQQQE  | 0.72 (Rank 8) | 1.63     | –       |
| A3.2 | 226–237    | GSVLSGFSKHF | – | – | 0.737 |
| A3.3 | 313–324    | EEEDQQPRPDHPP | 0.81 (Rank 7) | 2.72     | –       |
| A2.1 | 256–261    | KGGLRV  | 0.80 (Rank 9) | 1.33     | 0.293   |
| A2.2 | 130–141    | EGRGLRV  | 0.80 (Rank 9) | 1.33     | 0.293   |
| A2.3 | 136–153    | GDIIFAPGVAHWYNSYD | 0.87 (Rank 7) | 0.428   |
| A2.4 | 214–225    | GSNILSGFAPEF | 0.82 (Rank 7) | 0.319 |
| A2.5 | 256–261    | KGLLRV  | 0.80 (Rank 9) | –        | 0.302   |
| A2.6 | 283–291    | QCVETDKGC | 0.00     | 283–291  | QCVETDKGC |

*Amino acid number corresponds to position of full length sequence. A dash (–) indicates that no epitope was found. Scores are listed for the 3 different methods tested. Only the highest score is listed if the epitope was found in multiple lengths tested (10–20 mer).*
Acknowledgment

The authors would like to acknowledge the late Dr. Susan Hefle and Dr. Steve Taylor (University of Nebraska) for their generous gift of patient sera. Supported by Agriculture and Agri-Food Canada, and the Genomics Research and Development Initiative of the Government of Canada.

Appendix A. Transparency document

Transparency document associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.05.027.

References

[1] H. Saeed, C. Gagnon, E.R. Cober, S. Gleddie, Using patient serum to epitope map soybean glycinins reveals common epitopes shared with many legumes and tree nuts, Mol. Immunol. 70 (2016) 125–133.
[2] C. Gagnon, V. Poyya, E.R. Cober, S. Gleddie, Soybean allergens affecting North American patients identified by 2D gels and mass spectrometry, Food Anal. Methods 3 (2010) 363–374. http://dx.doi.org/10.1007/s12161-009-9090-3.
[3] S. Saha, G.P.S. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, Proteins 65 (2006) 40–48.
[4] J.E.P. Larsen, O. Lund, M. Nielsen, Improved method for predicting linear B-cell epitopes, Immunome Res. 2 (2006) 2.
[5] B. Yao, L. Zhang, S. Liang, C. Zhang, SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity, PLoS ONE 7 (2012) e45152.