Product-specific INTERACTION annotations examples

Table of Contents

Introduction

We are planning to change the UniProt flat file format for INTERACTION comments.

Proposed format:

```
CC   -!- INTERACTION:
CC       <interactant_1>; <interactant_2>;( Xeno;)? NbExp=<value>; IntAct=<value>;
CC       <interactant_1>; <interactant_2>;( Xeno;)? NbExp=<value>; IntAct=<value>;
CC       ...
```

- The format for <interactant_1> is:

  ```
  <AC>|<IsoId>|<FTId>
  ```

- The format for <interactant_2> is the same as that for <interactant_1> if both are described in the current entry (2 interacting isoforms or chains) and otherwise it is:

  ```
  (<AC>|<IsoId>|<FTId> \[<AC>\])(: <gene_name>)?
  ```

  where <gene_name> is optional (i.e. we will no longer print a '-' when there is no gene name).

- There will no longer be a special case for self-interactions because the term "Self" is ambiguous in entries with several isoforms/chains/peptides.

Examples:

P11309.txt (isoforms)
the current format:

```
CC       Q9UNQ0:ABCG2; NbExp=5; IntAct=EBI-1018633, EBI-1569435;
CC       Q9BZS1-1:FOXP3; NbExp=3; IntAct=EBI-1018629, EBI-9695448;
```

will become:

```
CC       P11309-1; Q9BZS1-1: FOXP3; NbExp=3; IntAct=EBI-1018629, EBI-9695448;
CC       P11309-2; Q9UNQ0: ABCG2; NbExp=5; IntAct=EBI-1018633, EBI-1569435;
```

P27958 and Q9NPY3 (chains)

- P27958.txt
  will get a new line for:

  ```
  CC       PRO_0000037566; Q9NPY3: CD93; Xeno; NbExp=2; IntAct=EBI-6377335, EBI-1755002;
  ```

- Q9NPY3.txt
  will get a new line for:

```
CC       Q9NPY3; PRO_0000037566 [P27958]; Xeno; NbExp=2; IntAct=EBI-1755002, EBI-6377335;
```

More examples of product-specific interactions were identified in **/ebi/ftp/pub/databases/intact/current/psimitab/intact-micluster.txt** and are shown in the next sections. Not all of them can be found in UniProt because of the filtering rules that are applied.

Isoforms

Isoforms self interactions

```
perl -ne 'if
(/^intact:(EBI-\d+)\tintact:(EBI-\d+)\tuniprotkb:([A-Z0-9]+-\d+)\tuniprotkb:([A-Z0-9]+-\d+)\t/)' \
     -ne '{print "$3 $4 $1 $2\n" if ($1 eq $2)}' intact-micluster.txt | wc -l
402
```

Example:

```
Q04206-2 Q04206-2 EBI-289947 EBI-289947
```

Current format:

- Q04206.txt

```
ID   TF65_HUMAN              Reviewed;         551 AA.
GN   Name=RELA; Synonyms=NFKB3;
OS   Homo sapiens (Human).
...
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
...
CC   -!- INTERACTION:
CC       Self; NbExp=3; IntAct=EBI-73886, EBI-73886;
```

New format:

- Q04206.txt

```
CC   -!- INTERACTION:
CC       Q04206-2; Q04206-2; NbExp=3; IntAct=EBI-289947, EBI-289947;
```

Nicole: Is the current "Self" interaction EBI-73886/Q04206 "mapped" from EBI-289947/Q04206-2?

Example:

```
Q14790-1 Q14790-1 EBI-288309 EBI-288309
Q14790-5 Q14790-5 EBI-288326 EBI-288326
```

Current format:

- Q14790.txt

```
ID   CASP8_HUMAN             Reviewed;         479 AA.
GN   Name=CASP8; Synonyms=MCH5;
OS   Homo sapiens (Human).
...
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=9;
...
CC   -!- INTERACTION:
CC       Self; NbExp=2; IntAct=EBI-78060, EBI-78060;
```

New format:

- Q14790.txt

```
CC   -!- INTERACTION:
CC       Q14790-1; Q14790-1; NbExp=2; IntAct=EBI-288309, EBI-288309;
CC       Q14790-5; Q14790-5; NbExp=2; IntAct=EBI-288326, EBI-288326;
```

Nicole: Is the current "Self" interaction EBI-78060/Q14790 "mapped" from the 2 isoform-specific interactions?

Nicole: I looked at the first 10 examples and most had no "Self" interaction. Are these currently all filtered out?

Isoforms non-self interactions

```
perl -ne 'if
(/^intact:(EBI-\d+)\tintact:(EBI-\d+)\tuniprotkb:([A-Z0-9]+-\d+)\tuniprotkb:([A-Z0-9]+-\d+)\t/)' \
     -ne '{print "$3 $4 $1 $2\n" if ($1 ne $2)}' intact-micluster.txt | wc -l
7013
```

Example:

```
Q9WVI9-1 P12023-2 EBI-288461 EBI-286828
Q9WVI9-2 Q86Y07-1 EBI-288464 EBI-1207633
Q86Y07-2 Q9WVI9-2 EBI-1207636 EBI-288464
```

Current format:

- Q9WVI9.txt

```
ID   JIP1_MOUSE              Reviewed;         707 AA.
GN   Name=Mapk8ip1; Synonyms=Ib1, Jip1, Mapk8ip, Prkm8ip;
OS   Mus musculus (Mouse).
...
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
...
CC   -!- INTERACTION:
CC       P12023:App; NbExp=3; IntAct=EBI-288461, EBI-78814;
CC       P12023-2:App; NbExp=2; IntAct=EBI-288461, EBI-286828;
CC       Q86Y07-2:VRK2 (xeno); NbExp=2; IntAct=EBI-288464, EBI-1207636;
```

  Nicole: The interaction "IntAct=EBI-288461, EBI-78814" is not in intact-micluster.txt of 21 Jul 2019. What did it represent: the displayed isoform 1, all 4 isoforms, or all but isoform 2?

- P12023.txt

```
ID   A4_MOUSE                Reviewed;         770 AA.
GN   Name=App;
OS   Mus musculus (Mouse).
...
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
...
CC   -!- INTERACTION:
CC       Q9WVI9-1:Mapk8ip1; NbExp=3; IntAct=EBI-78814, EBI-288461;
```

- Q86Y07.txt

```
ID   VRK2_HUMAN              Reviewed;         508 AA.
GN   Name=VRK2;
OS   Homo sapiens (Human).
...
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
...
CC   -!- INTERACTION:
CC       Q9WVI9-2:Mapk8ip1 (xeno); NbExp=2; IntAct=EBI-1207636, EBI-288464;
```

New format:

- Q9WVI9.txt

```
CC   -!- INTERACTION:
CC       Q9WVI9-1; P12023-2: App; NbExp=2; IntAct=EBI-288461, EBI-286828;
CC       Q9WVI9-2; Q86Y07-1: VRK2; Xeno; NbExp=?; IntAct=EBI-288464,EBI-1207633;
CC       Q9WVI9-2; Q86Y07-2: VRK2; Xeno; NbExp=2; IntAct=EBI-288464 EBI-1207636;
```

- P12023.txt

```
CC   -!- INTERACTION:
CC       P12023-2; Q9WVI9-1: Mapk8ip1; NbExp=2; IntAct=EBI-286828, EBI-288461;
```

- Q86Y07.txt

```
CC   -!- INTERACTION:
CC       Q86Y07-1; Q9WVI9-2: Mapk8ip1; Xeno; NbExp=?; IntAct=EBI-1207633, EBI-288464;
CC       Q86Y07-2; Q9WVI9-2: Mapk8ip1; Xeno; NbExp=2; IntAct=EBI-1207636, EBI-288464;
```

Chains

Chain self interactions

```
perl -ne 'if
(/^intact:(EBI-\d+)\tintact:(EBI-\d+)\tuniprotkb:([A-Z0-9]+-PRO_\d+)\tuniprotkb:([A-Z0-9]+-PRO_\d+)\t
\
     -ne '{print "$3 $4 $1 $2\n" if ($1 eq $2)}' intact-micluster.txt | wc -l
40
```

Example:

```
P03300-PRO_0000040090 P03300-PRO_0000040090 EBI-914162 EBI-914162
```

Current format:

- P03300.txt => Has no INTERACTION

```
FT   CHAIN      1748   2209       RNA-directed RNA polymerase.
FT                                /FTId=PRO_0000040090.
```

New format:

- P03300.txt

```
CC   -!- INTERACTION:
CC       PRO_0000040090; PRO_0000040090; NbExp=?; IntAct=EBI-914162, EBI-914162;
```

Example:

```
Q99IB8-PRO_0000045592 Q99IB8-PRO_0000045592 EBI-6858513 EBI-6858513
Q99IB8-PRO_0000045596 Q99IB8-PRO_0000045596 EBI-6901449 EBI-6901449
Q99IB8-PRO_0000045598 Q99IB8-PRO_0000045598 EBI-6901421 EBI-6901421
```

Q99IB8 also has several chain non-self interactions (details given in that section).

Chain non-self interactions

```
perl -ne 'if
(/^intact:(EBI-\d+)\tintact:(EBI-\d+)\tuniprotkb:([A-Z0-9]+-PRO_\d+)\tuniprotkb:([A-Z0-9]+-PRO_\d+)\t
\
     -ne '{print "$3 $4 $1 $2\n" if ($1 ne $2)}' intact-micluster.txt | wc -l
83
```

Example:

```
P27958-PRO_0000037570 Q99IB8-PRO_0000045602 EBI-6904269 EBI-6927873
P27958-PRO_0000037572 Q99IB8-PRO_0000045599 EBI-6919131 EBI-6858501
P27958-PRO_0000037572 Q99IB8-PRO_0000045602 EBI-6919131 EBI-6927873
Q99IB8-PRO_0000045592 P27958-PRO_0000037576 EBI-6858513 EBI-8753518
```

Q99IB8 also has interactions between different chains of the same protein:

```
Q99IB8-PRO_0000045592 Q99IB8-PRO_0000045598 EBI-6858513 EBI-6901421
Q99IB8-PRO_0000045595 Q99IB8-PRO_0000045596 EBI-6901441 EBI-6901449
Q99IB8-PRO_0000045596 Q99IB8-PRO_0000045592 EBI-6901449 EBI-6858513
Q99IB8-PRO_0000045596 Q99IB8-PRO_0000045599 EBI-6901449 EBI-6858501
Q99IB8-PRO_0000045596 Q99IB8-PRO_0000045601 EBI-6901449 EBI-6928570
Q99IB8-PRO_0000045597 Q99IB8-PRO_0000045596 EBI-8772829 EBI-6901449
Q99IB8-PRO_0000045597 Q99IB8-PRO_0000045598 EBI-8772829 EBI-6901421
Q99IB8-PRO_0000045598 Q99IB8-PRO_0000045595 EBI-6901421 EBI-6901441
Q99IB8-PRO_0000045598 Q99IB8-PRO_0000045596 EBI-6901421 EBI-6901449
Q99IB8-PRO_0000045598 Q99IB8-PRO_0000045599 EBI-6901421 EBI-6858501
Q99IB8-PRO_0000045599 Q99IB8-PRO_0000045600 EBI-6858501 EBI-9096996
Q99IB8-PRO_0000045600 Q99IB8-PRO_0000045596 EBI-9096996 EBI-6901449
Q99IB8-PRO_0000045602 Q99IB8-PRO_0000045598 EBI-6927873 EBI-6901421
Q99IB8-PRO_0000045603 Q99IB8-PRO_0000045602 EBI-6927928 EBI-6927873
```

As well as 3 chain self interactions:

```
Q99IB8-PRO_0000045592 Q99IB8-PRO_0000045592 EBI-6858513 EBI-6858513
Q99IB8-PRO_0000045596 Q99IB8-PRO_0000045596 EBI-6901449 EBI-6901449
Q99IB8-PRO_0000045598 Q99IB8-PRO_0000045598 EBI-6901421 EBI-6901421
```

Current format:

- P27958.txt

```
ID   POLG_HCVH               Reviewed;        3011 AA.
OS   Hepatitis C virus genotype 1a (isolate H) (HCV).
...
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Ribosomal frameshifting; Named isoforms=2;
...
CC   -!- INTERACTION:
CC       Q99IB8:- (xeno); NbExp=3; IntAct=EBI-8753518, EBI-6858513;
...
FT   CHAIN         2    191       Core protein p21. {ECO:0000255}.
FT                                /FTId=PRO_0000037566.
FT   CHAIN         2    177       Core protein p19. {ECO:0000250}.
FT                                /FTId=PRO_0000037567.
FT   PROPEP      178    191       ER anchor for the core protein, removed
FT                                in mature form by host signal peptidase.
FT                                {ECO:0000250}.
FT                                /FTId=PRO_0000037568.
FT   CHAIN       192    383       Envelope glycoprotein E1. {ECO:0000255}.
FT                                /FTId=PRO_0000037569.
FT   CHAIN       384    746       Envelope glycoprotein E2. {ECO:0000255}.
FT                                /FTId=PRO_0000037570.
FT   CHAIN       747    809       Viroporin p7.
FT                                /FTId=PRO_0000037571.
FT   CHAIN       810   1026       Protease NS2-3. {ECO:0000255|PROSITE-
FT                                ProRule:PRU01030}.
FT                                /FTId=PRO_0000037572.
FT   CHAIN      1027   1657       Serine protease NS3. {ECO:0000255}.
FT                                /FTId=PRO_0000037573.
FT   CHAIN      1658   1711       Non-structural protein 4A. {ECO:0000255}.
FT                                /FTId=PRO_0000037574.
FT   CHAIN      1712   1972       Non-structural protein 4B. {ECO:0000255}.
FT                                /FTId=PRO_0000037575.
FT   CHAIN      1973   2420       Non-structural protein 5A. {ECO:0000255}.
FT                                /FTId=PRO_0000037576.
FT   CHAIN      2421   3011       RNA-directed RNA polymerase.
FT                                {ECO:0000255}.
FT                                /FTId=PRO_0000037577.
```

- Q99IB8.txt

```
ID   POLG_HCVJF              Reviewed;        3033 AA.
OS   Hepatitis C virus genotype 2a (isolate JFH-1) (HCV).
...
CC   -!- INTERACTION:
CC       Self; NbExp=4; IntAct=EBI-6674379, EBI-6674379;
CC       P27958:- (xeno); NbExp=3; IntAct=EBI-6858513, EBI-8753518;
...
FT   CHAIN         2    191       Core protein p21. {ECO:0000255}.
FT                                /FTId=PRO_0000045592.
FT   CHAIN         2    177       Core protein p19. {ECO:0000250}.
FT                                /FTId=PRO_0000045593.
FT   PROPEP      178    191       ER anchor for the core protein, removed
FT                                in mature form by host signal peptidase.
FT                                {ECO:0000250}.
FT                                /FTId=PRO_0000045594.
FT   CHAIN       192    383       Envelope glycoprotein E1. {ECO:0000255}.
FT                                /FTId=PRO_0000045595.
FT   CHAIN       384    750       Envelope glycoprotein E2. {ECO:0000255}.
FT                                /FTId=PRO_0000045596.
FT   CHAIN       751    813       p7. {ECO:0000250}.
FT                                /FTId=PRO_0000045597.
FT   CHAIN       814   1030       Protease NS2-3. {ECO:0000255|PROSITE-
FT                                ProRule:PRU01030}.
FT                                /FTId=PRO_0000045598.
FT   CHAIN      1031   1661       Serine protease NS3. {ECO:0000255}.
FT                                /FTId=PRO_0000045599.
FT   CHAIN      1662   1715       Non-structural protein 4A. {ECO:0000255}.
FT                                /FTId=PRO_0000045600.
FT   CHAIN      1716   1976       Non-structural protein 4B. {ECO:0000255}.
FT                                /FTId=PRO_0000045601.
FT   CHAIN      1977   2442       Non-structural protein 5A. {ECO:0000255}.
FT                                /FTId=PRO_0000045602.
FT   CHAIN      2443   3033       RNA-directed RNA polymerase.
FT                                {ECO:0000255}.
FT                                /FTId=PRO_0000045603.
```

New format:

- P27958.txt

```
CC   -!- INTERACTION:
CC       PRO_0000037570; PRO_0000045602 [Q99IB8]; Xeno; NbExp=?; IntAct=EBI-6904269,
EBI-6927873;
CC       PRO_0000037572; PRO_0000045599 [Q99IB8]; Xeno; NbExp=?; IntAct=EBI-6919131,
EBI-6858501;
CC       PRO_0000037572; PRO_0000045602 [Q99IB8]; Xeno; NbExp=?; IntAct=EBI-6919131,
EBI-6927873;
CC       PRO_0000037576; PRO_0000045592 [Q99IB8]; Xeno; NbExp=?; IntAct=EBI-8753518,
EBI-6858513;
```

- Q99IB8.txt

```
CC   -!- INTERACTION:
CC       PRO_0000045602; PRO_0000037570 [P27958]; Xeno; NbExp=?; IntAct=EBI-6927873,
EBI-6904269;
CC       PRO_0000045599; PRO_0000037572 [P27958]; Xeno; NbExp=?; IntAct=EBI-6858501,
EBI-6919131;
CC       PRO_0000045602; PRO_0000037572 [P27958]; Xeno; NbExp=?; IntAct=EBI-6927873,
EBI-6919131;
CC       PRO_0000045592; PRO_0000037576 [P27958]; Xeno; NbExp=?; IntAct=EBI-6858513,
EBI-8753518;
...
CC       PRO_0000045592; PRO_0000045598; NbExp=?; IntAct=EBI-6858513, EBI-6901421;
CC       PRO_0000045595; PRO_0000045596; NbExp=?; IntAct=EBI-6901441, EBI-6901449;
CC       PRO_0000045596; PRO_0000045592; NbExp=?; IntAct=EBI-6901449, EBI-6858513;
CC       PRO_0000045596; PRO_0000045599; NbExp=?; IntAct=EBI-6901449, EBI-6858501;
CC       PRO_0000045596; PRO_0000045601; NbExp=?; IntAct=EBI-6901449, EBI-6928570;
CC       PRO_0000045597; PRO_0000045596; NbExp=?; IntAct=EBI-8772829, EBI-6901449;
CC       PRO_0000045597; PRO_0000045598; NbExp=?; IntAct=EBI-8772829, EBI-6901421;
CC       PRO_0000045598; PRO_0000045595; NbExp=?; IntAct=EBI-6901421, EBI-6901441;
CC       PRO_0000045598; PRO_0000045596; NbExp=?; IntAct=EBI-6901421, EBI-6901449;
CC       PRO_0000045598; PRO_0000045599; NbExp=?; IntAct=EBI-6901421, EBI-6858501;
CC       PRO_0000045599; PRO_0000045600; NbExp=?; IntAct=EBI-6858501, EBI-9096996;
CC       PRO_0000045600; PRO_0000045596; NbExp=?; IntAct=EBI-9096996, EBI-6901449;
CC       PRO_0000045602; PRO_0000045598; NbExp=?; IntAct=EBI-6927873, EBI-6901421;
CC       PRO_0000045603; PRO_0000045602; NbExp=?; IntAct=EBI-6927928, EBI-6927873;
...
CC       PRO_0000045592; PRO_0000045592; NbExp=?; IntAct=EBI-6858513, EBI-6858513;
CC       PRO_0000045596; PRO_0000045596; NbExp=?; IntAct=EBI-6901449, EBI-6901449;
CC       PRO_0000045598; PRO_0000045598; NbExp=?; IntAct=EBI-6901421, EBI-6901421;
```

Nicole: P27958 and Q99IB8 are 2 different HCV strains. Are the interactions between them biologically relevant?

The example Q99IB8 shows that it will still be useful to break up the virus polyprotein entries into one entry per chain, where each entry will show only the interactions relevant for that protein.