**Unorganized notes about PPI type annotation decisions on BioInfer-learning**

1. Making a PPI call on each protein/gene pair identified in the XML file
2. Proteins/Genes. Ending in –in or –ins are pre-identified as structural (actin, catenin, …)  
   exceptions:  
    toxin  
    beta-catenin (could be gene regulator OR structural – it’s a dual function gene)  
    calreticulin – multifunction, mostly enzyme
3. histones and nucleosomes are not considered structural, because their “structure” is mutable and controls regulation
4. Proteins/Genes ending in –ase are pre-identified as enzymes
5. Proteins/Genes containing inhibitor, activator, transcription factor, repressor, enhancer, or regulator are pre-identified as enzymes.
6. Well known tumor suppression and/or DNA repair proteins p53, p27, DNA-pk, RAD51, are enzymes
7. if type of both proteins is structural, or if one is structural and the other is unknown, then the pair PPI is annotated as structural.  
     
   If type of both proteins is enzyme, or if one is enzyme and the other is enzyme, then the pair PPI is annotated as enzyme.  
     
   If one protein Is enzyme and the other is structural, the pair is annotated as enzyme
8. Cell movement (ie chemotaxis, chemokine) is considered structural
9. DNA binding motif is taken to be regulatory therefore enzyme type
10. Amyloid – structural?
11. Cell adhesion proteins: structural
12. Dynein/dynactin : molecular motors. Structural, which has been expanded to include mechanical mechanisms
13. snoRNAs are encoded in the genome, expressed, and are involved in regulation, therefore classified as enzymes
14. Oxytocin and vasopressin (AVP) are hormones that start out as a peptide, part of a protein. I’m counting them as proteins (they have Refseq ids). Same with melanocortin.

<<<from negative annotations in Finnish data sets>>>

1. Implied relationships & correlations between increasing one proteins level and experimental measure of another is taken to be evidence of PPI
2. Playing similar roles in some condition isn’t sufficient evidence of PPI
3. “We studied X & Y” no PPI
4. Homology between X & Y: no PPI (could be gene duplication + evolution)
5. “X has been fused to Y” no PPI (artificial construct, not natural)
6. General: look for actor -> actee relationships. If either is a physiological feature, phenotype, metabolite, or something which is not a protein, then no PPI

Problems

* Not sure how to handle receptors because they’re often membrane bound but they can be involved in signaling – for now I’m labeling them structural  
   (example: CD4 – on the outside of immune cells but its purpose is to signal other cells)
* Interleukin - enzyme?
* Extracellular signaling? Have no idea
* Chaperones – enzyme? They’re more like machinery than chemical enzymes
* Oxytocin – a modified peptide? I annotated as an enzyme
* Leptin – hormone, interacts with a receptor by not - doesn’t seem like a good idea to call this structural

Problems with Finnish annotations:

* Some examples of metabolites identified as proteins/genes, ie in I\  
  IEPA.- IEPA.d7.s13.p0. flavanone and cholesterol are not proteins  
  Similarly below that, IP3 Inosol triphosphate is not a gene/protein