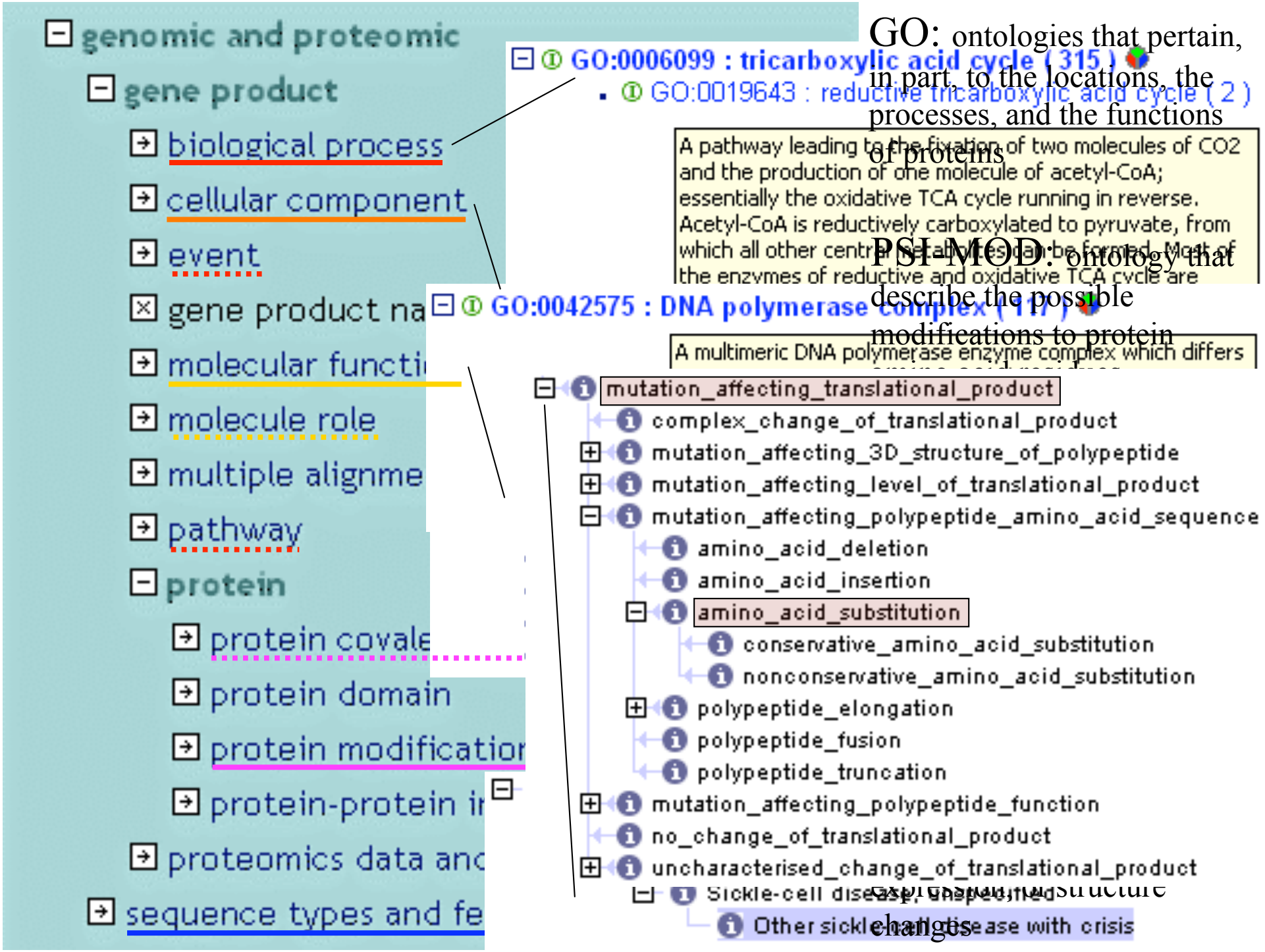
Framework for a Protein Ontology

SO Immunology Workshop
June 2007

Darren A. Natale, Ph.D.
Protein Science Team Lead, PIR
Research Assistant Professor, GUMC



GO: ontologies that pertain, in part, to the locations, the processes, and the functions of proteins

PSI-MOD: ontology that describe the possible modifications to protein

expression/structure changes

Mothers against decapentaplegic homolog 2

Smad 2

GO annotation of SMAD2_HUMAN:

Cellular Component:

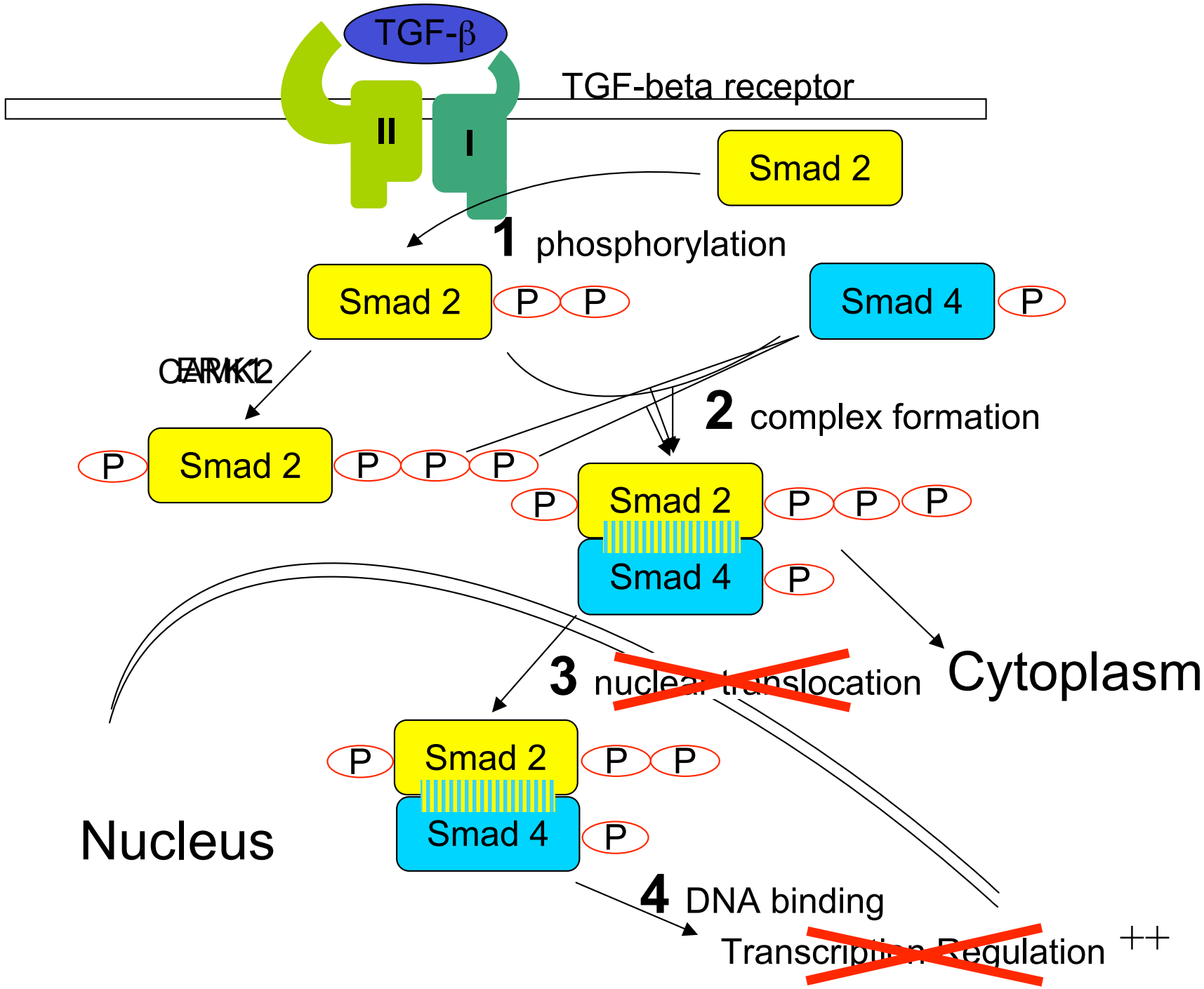
- nucleus

Molecular Function:

- protein binding

Biological Process:

- signal transduction
- regulation of transcription, DNA-dependent


TGF-β
TGF-beta receptor
II
I
Smad 2
1 phosphorylation
Smad 2
P P
Smad 4
P
Smad 2
P
2 complex formation
P
Smad 2
P P P
P
Smad 2
P P P
Smad 4
P
3 nuclear translocation
Cytoplasm
P
Smad 2
P P
Smad 4
P
Nucleus
4 DNA binding
Transcription Regulation ++

Smad 2	"normal"	•Cytoplasmic	SMAD2_HUMAN
Smad 2 P P	TGF-β receptor phosphorylated	•Forms complex •Nuclear •Txn upregulation	SMAD2_HUMAN
Smad 2 P P P	ERK1 phosphorylated	•Forms complex •Nuclear •Txn upregulation++	SMAD2_HUMAN
P Smad 2 P P	CAMK2 phosphorylated	•Forms complex •Cytoplasmic •No Txn upregulation	SMAD2_HUMAN
Smad 2	alternatively spliced short form	•Cytoplasmic	SMAD2_HUMAN
Smad 2 P P	phosphorylated short form	•Nuclear •Txn upregulation	SMAD2_HUMAN
Smad 2 X	point mutation (causative agent: large intestine carcinoma)	•Doesn't form complex •Cytoplasmic •No Txn upregulation	SMAD2_HUMAN

Important Considerations

- Need to consider the various forms a protein might take
- Need to provide connections to established ontologies
- Need to account for the possibility that a protein might not share the traits of its parent or siblings

%**PRO:00000010** Smad2
  <**PRO:00000011** Smad2 sequence 1 (long form)
    >**PRO:00000012** Smad2 sequence 1 phosphorylated form
      %**PRO:00000013** Smad2 sequence 1, TGF-β receptor I-phosphorylated
      %**PRO:00000014** Smad2 sequence 1, TGF-β receptor I and ERK1-phosphorylated
      %**PRO:00000015** Smad2 sequence 1, TGF-β receptor I and CAMK2-phosphorylated
  <**PRO:00000016** Smad2 sequence 2 (short form) - splice variant
    >**PRO:00000017** Smad2 sequence 2 phosphorylated form
      %**PRO:00000018** Smad2 sequence 2, TGF-β receptor I-phosphorylated
  <**PRO:00000019** Smad2 sequence 3 - genetic variant related to colorectal carcinoma

has_modification MOD:O-phosphorylated L-serine
has_modification MOD:O-phosphorylated L-threonine
has_function GO: TGF-β receptor, pathway-specific cytoplasmic mediator activity
has_function GO:SMAD binding
has_function GO:transcription coactivator activity
participates_in GO:signal transduction
participates_in GO:SMAD protein heteromerization
participates_in GO:regulation of transcription, DNA-dependent
located_in GO:nucleus
part_of GO:transcription factor complex

%**PRO:00000015** Smad2 sequence 1, TGF-β receptor I and CAMK2-phosphorylated
<**PRO:00000016** Smad2 sequence 2 (short form) - splice variant
>**PRO:00000017** Smad2 sequence 2 phosphorylated form
%**PRO:00000018** Smad2 sequence 2, TGF-β receptor I-phosphorylated
<**PRO:00000019** Smad2 sequence 3 - genetic variant related to colorectal carcinoma

has_agent SO: amino_acid_substitution
lacks_modification MOD: phosphorylated residue
lacks_function GO: transcription coactivator activity
agent_of DO: carcinoma of the large intestine

%	is_a
<	variant_of
>	derives_from