

Framework for a Protein Ontology

SO Immunology Workshop
June 2007

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

[-] genomic and proteomic

[-] gene product

[+] biological process

[+] cellular component

[+] event

[x] gene product name

[+] molecular function

[+] molecule role

[+] multiple alignment

[+] pathway

[-] protein

[+] protein covalent modification

[+] protein domain

[+] protein modification

[+] protein-protein interaction

[+] proteomics data and analysis

[+] sequence types and features

[-] ⓘ GO:0006099 : tricarboxylic acid cycle (315)

• ⓘ GO:0019643 : reductive tricarboxylic acid cycle (2)

A pathway leading to the fixation of two molecules of CO2 and the production of one molecule of acetyl-CoA; essentially the oxidative TCA cycle running in reverse. Acetyl-CoA is reductively carboxylated to pyruvate, from which all other central metabolites can be formed. Most of the enzymes of reductive and oxidative TCA cycle are

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

[-] ⓘ GO:0042575 : DNA polymerase complex (117)

A multimeric DNA polymerase enzyme complex which differs

- [-] ⓘ mutation_affecting_translational_product
 - [-] ⓘ complex_change_of_translational_product
 - [+] ⓘ mutation_affecting_3D_structure_of_polypeptide
 - [+] ⓘ mutation_affecting_level_of_translational_product
 - [-] ⓘ mutation_affecting_polypeptide_amino_acid_sequence
 - [-] ⓘ amino_acid_deletion
 - [-] ⓘ amino_acid_insertion
 - [-] ⓘ amino_acid_substitution
 - [-] ⓘ conservative_amino_acid_substitution
 - [-] ⓘ nonconservative_amino_acid_substitution
 - [+] ⓘ polypeptide_elongation
 - [-] ⓘ polypeptide_fusion
 - [-] ⓘ polypeptide_truncation
 - [+] ⓘ mutation_affecting_polypeptide_function
 - [-] ⓘ no_change_of_translational_product
 - [+] ⓘ uncharacterised_change_of_translational_product
 - [-] ⓘ Sickle-cell disease, unspecified, due to structure changes
 - [-] ⓘ Other sickle cell disease with crisis

Mothers against decapentaplegic homolog 2

Smad 2

GO annotation of SMAD2_HUMAN:

Cellular Component:

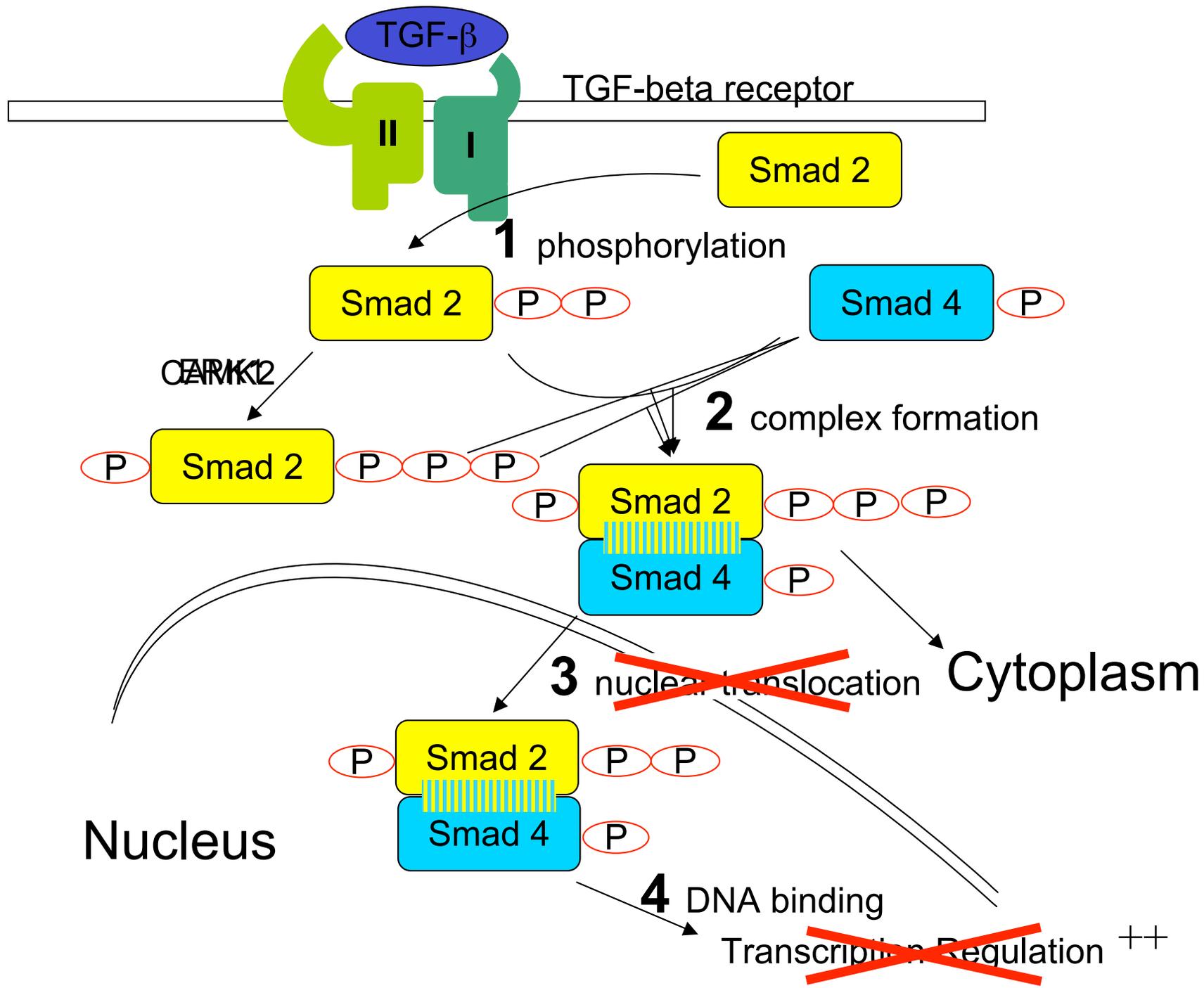
- nucleus

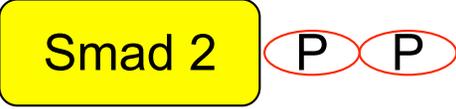
Molecular Function:

- protein binding

Biological Process:

- signal transduction
- regulation of transcription, DNA-dependent



	<p>“normal”</p>	<ul style="list-style-type: none"> •Cytoplasmic 	<p>SMAD2_HUMAN</p>
	<p>TGF-β receptor phosphorylated</p>	<ul style="list-style-type: none"> •Forms complex •Nuclear •Txn upregulation 	<p>SMAD2_HUMAN</p>
	<p>ERK1 phosphorylated</p>	<ul style="list-style-type: none"> •Forms complex •Nuclear •Txn upregulation++ 	<p>SMAD2_HUMAN</p>
	<p>CAMK2 phosphorylated</p>	<ul style="list-style-type: none"> •Forms complex •Cytoplasmic •No Txn upregulation 	<p>SMAD2_HUMAN</p>
	<p>alternatively spliced short form</p>	<ul style="list-style-type: none"> •Cytoplasmic 	<p>SMAD2_HUMAN</p>
	<p>phosphorylated short form</p>	<ul style="list-style-type: none"> •Nuclear •Txn upregulation 	<p>SMAD2_HUMAN</p>
	<p>point mutation (causative agent: large intestine carcinoma)</p>	<ul style="list-style-type: none"> •Doesn't form complex •Cytoplasmic •No Txn upregulation 	<p>SMAD2_HUMAN</p>

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

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