

○安保勲人¹⁾，○佐久間航也²⁾，嘉戸裕美子²⁾，坂本盛宇³⁾，細田和男⁴⁾，
鹿間周子²⁾，大安裕美²⁾，高木大輔³⁾，山口敦子⁴⁾，
畠中秀樹⁵⁾，小池亮太郎²⁾，廣明秀一²⁾，福地佐斗志¹⁾，太田元規²⁾

1) 前橋工科大学，2) 名古屋大学，3) (株) ホロニクス，4) 東京都市大学，
5) ライフサイエンス統合データベースセンター

<https://www.ideal-db.org/>

IDEALってなに？

- ・天然変性タンパク質のデータベース
 - 2009年活動開始（2011年公開）
 - 1,110 エントリ，13,097 領域：真核，核内
(**昨年10月に最新バージョンを公開**)
 - **世界最大規模**のアノテーションデータベース
- ・天然変性領域中の相互作用領域（ProS）を収集
 - 348 エントリにProSあり
- ・全データ，**マニュアルアノテーション**
- ・全コンテンツはXML, RDFでダウンロード可能

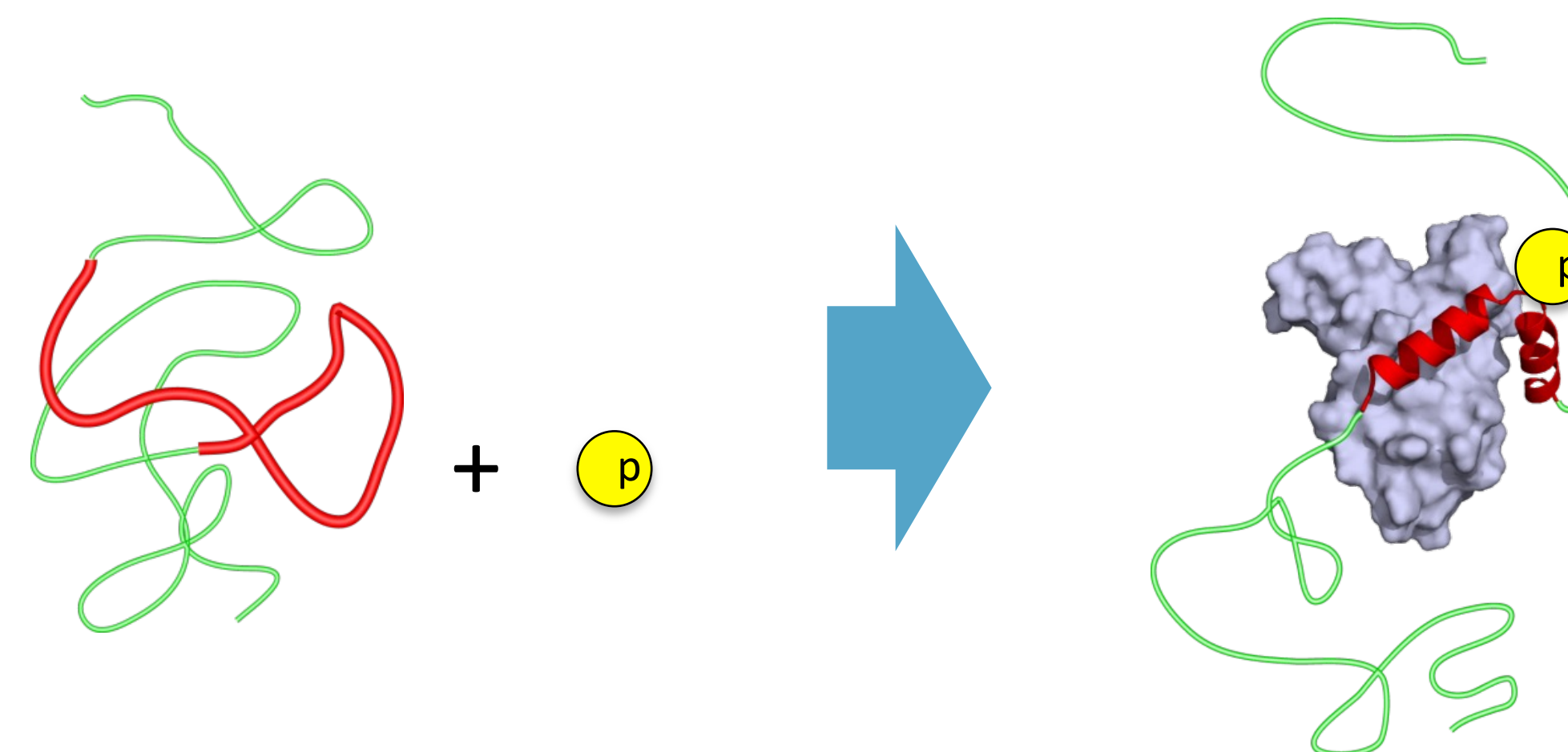


にも参画しています！

IDPcentral

天然変性タンパク質ってなに？

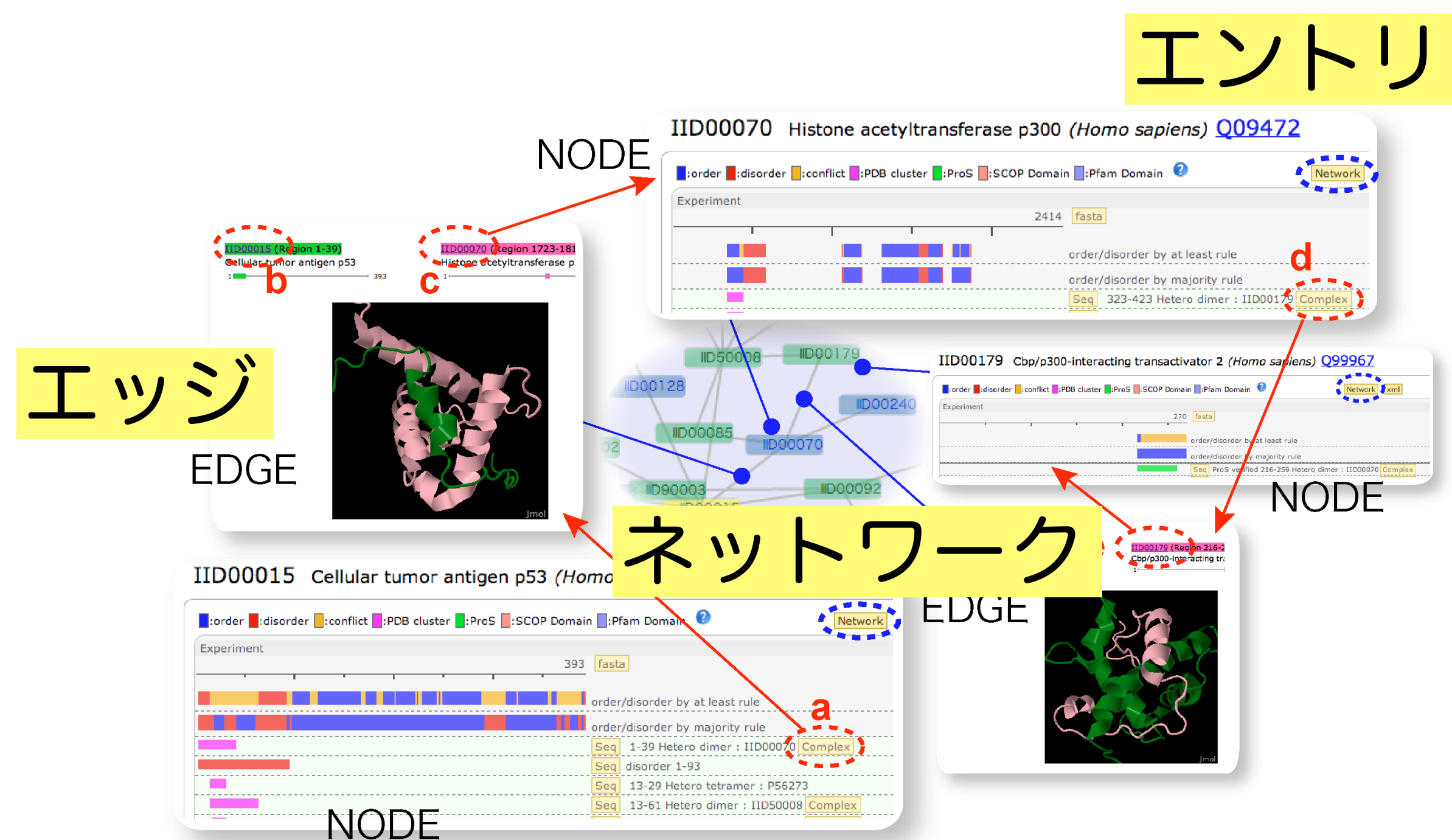
形をとらないタンパク質
機能：シグナル伝達，転写制御など
多くの翻訳後修飾部位を持つ



天然変性領域にある機能部位
構造を形成し結合

Protean Segment (**ProS**) として収集

IDEALをのぞいてみよう！



クリックしていくうちに
相互作用，**機能**が実感できる

SPARQL endpoint 設置

XML → RDF

SPARQL endpoint
クエリによる検索も可能

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

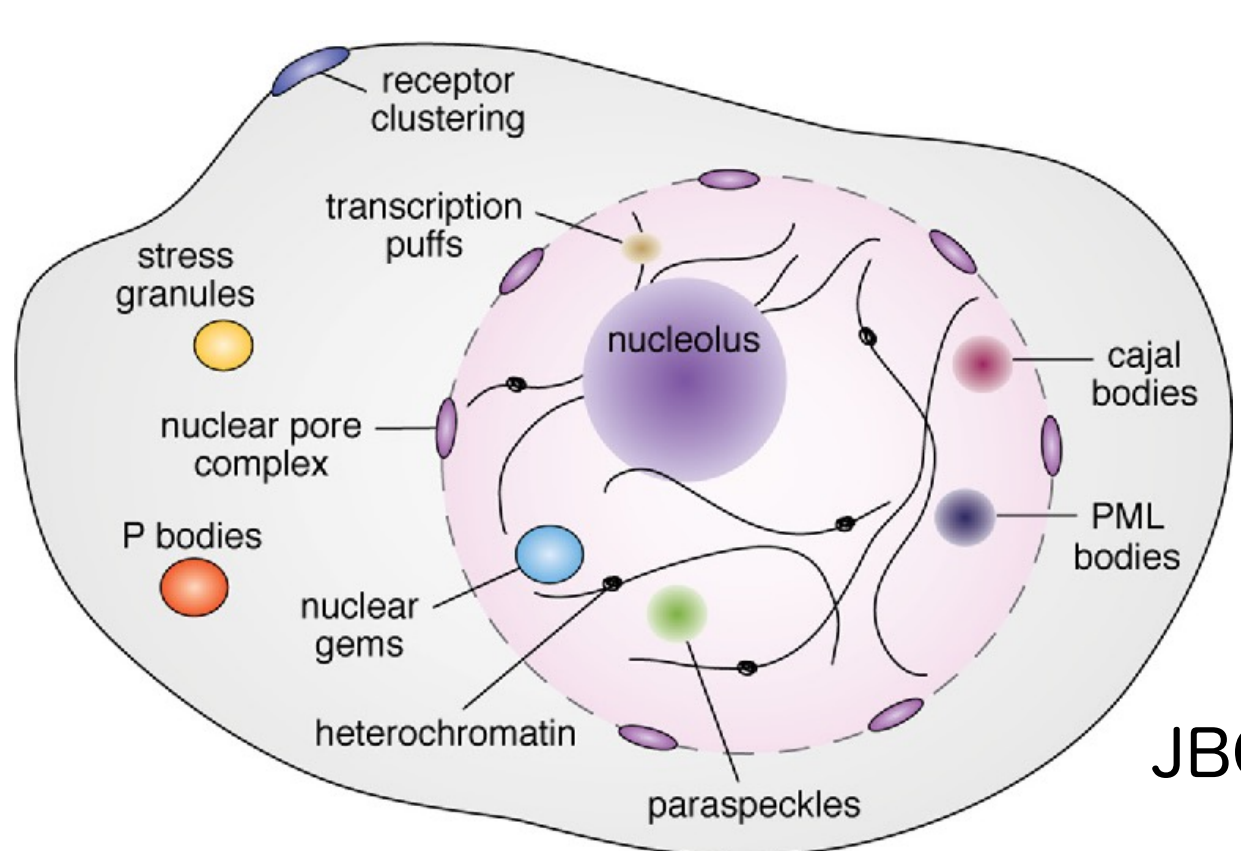
```
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX ideal:<http://www.ideal.force.cs.is.nagoya-u.ac.jp/IDEAL/>

SELECT ?protein ?idp ?region_start ?region_end ?order_disorder
WHERE
{
  SERVICE <https://integbio.jp/rdf/mirror/alt/uniprot/sparql>{
    ?protein a up:Protein .
    ?protein up:mnemonic 'MYC_HUMAN' .
  }

  ?idp ideal:uniprot ?protein .
  ?idp ideal:region ?rid.
  ?rid ideal:region_start ?region_start.
  ?rid ideal:region_end ?region_end.
  ?rid ideal:order_disorder ?order_disorder.
}
```

UniProt, PDBj からリンクされている

液滴を形成する天然変性タンパク質



JBC, 2019, Gomes

IID00012 Myc proto-oncogene protein (*Homo sapiens*) [P01106](#)



LLPS

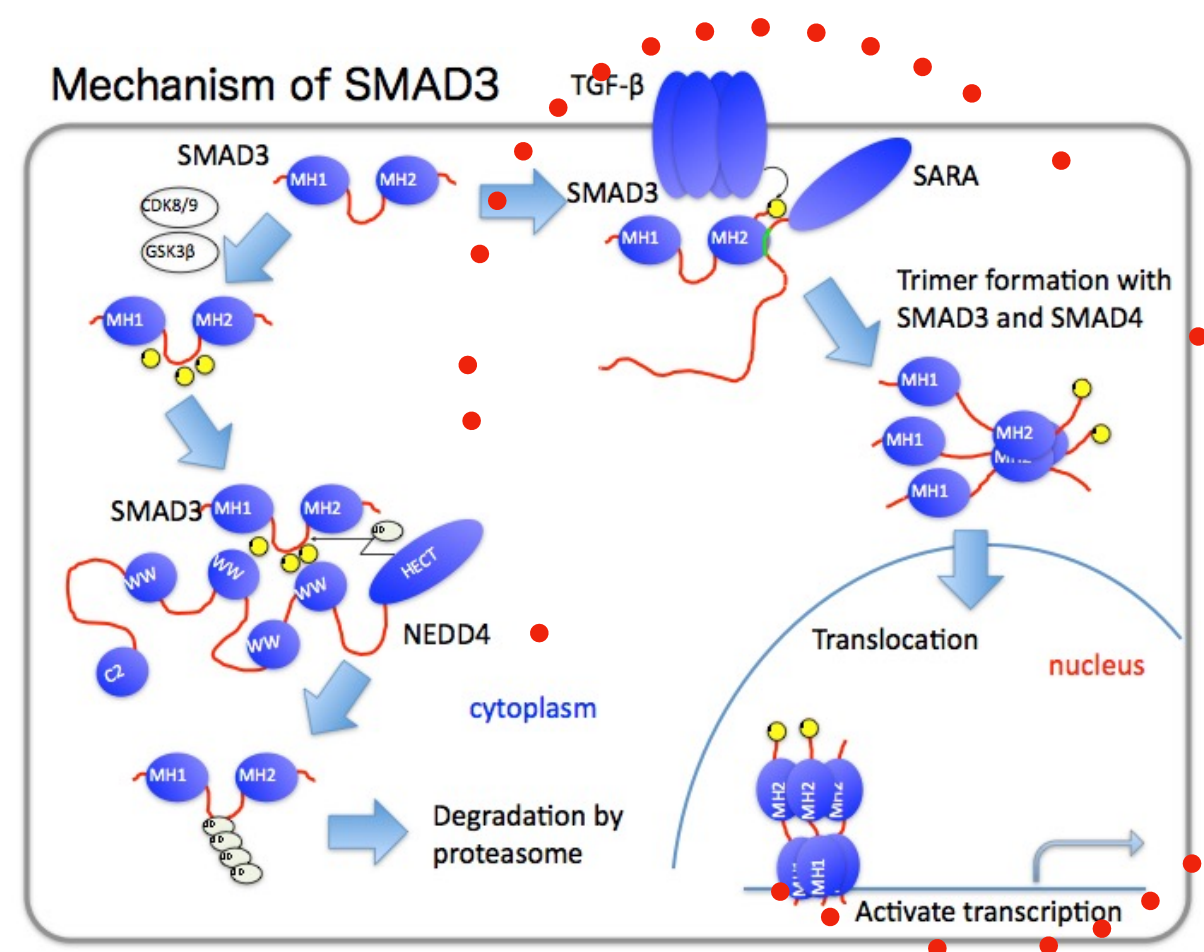
液滴の局在情報

Subcellular location

Nucleus, Nucleolus.

UniProtの局在情報から液滴をリスト化 2020年7月公開

リン酸化に伴うイベントの表現



リン酸化に起因する生命現象の記述

例) Smad3のリン酸化カスケード情報 エクセルで公開

id			Process				Additional Information		
protein_id	ideal_id	process_id	operand	operator	result	special object	location	property	
P84022	IID00113	1	SMAD3 (obj_id 1)	TGF β receptor (obj_id 2)	SMAD3-P (obj_id 3)				
		2	SMAD3-P, SMAD3-P, SMAD4 (obj_id 4)		SMAD3-P/SMAD3-P/SMAD4 (obj_id 5)	SMAD3-P/SMAD3-P/SMAD4 (obj_id 5)	Cytosol		
		3	SMAD3-P/SMAD3-P/SMAD4 (obj_id 5)		SMAD3-P/SMAD3-P/SMAD4 (obj_id 6)	SMAD3-P/SMAD3-P/SMAD4 (obj_id 6)	Nucleus	positive regulation of transcription, DNA-templated	

リン酸化に伴うイベントの表現：情報のXML・RDF・可視化計画

リン酸化情報入力エディター

Uniprot Accession : Q15797

Process List			
Operand	Operator	Result	Actions
SMAD1 obj_id1	BMP_receptor obj_id2	SMAD1(P) obj_id3	⊖
SMAD1(P), SMAD1(P), SMAD4 obj_id4		SMAD1(P)/SMAD1(P)/SMAD4 obj_id5	⊖
SMAD1(P)/SMAD1(P)/SMAD4 obj_id5		SMAD1(P)/SMAD1(P)/SMAD4 obj_id6	⊖
SMAD1 obj_id1	CDK8, CDK9 obj_id7	SMAD1(Ps) obj_id8	⊖
SMAD1(Ps), YAP obj_id9		SMAD1(Ps)/YAP obj_id10	⊖
SMAD1(Ps) obj_id8	GSK3_beta obj_id11	SMAD1(Ps) obj_id12	⊖

Object	
▼	SMAD1 obj_id1
▼	BMP_receptor obj_id2
▼	SMAD1(P) obj_id3
▼	SMAD1(P), SMAD1(P), SMAD4 obj_id4
▼	SMAD1(P)/SMAD1(P)/SMAD4 obj_id5
▼	SMAD1(P)/SMAD1(P)/SMAD4 obj_id6
▼	CDK8, CDK9 obj_id7

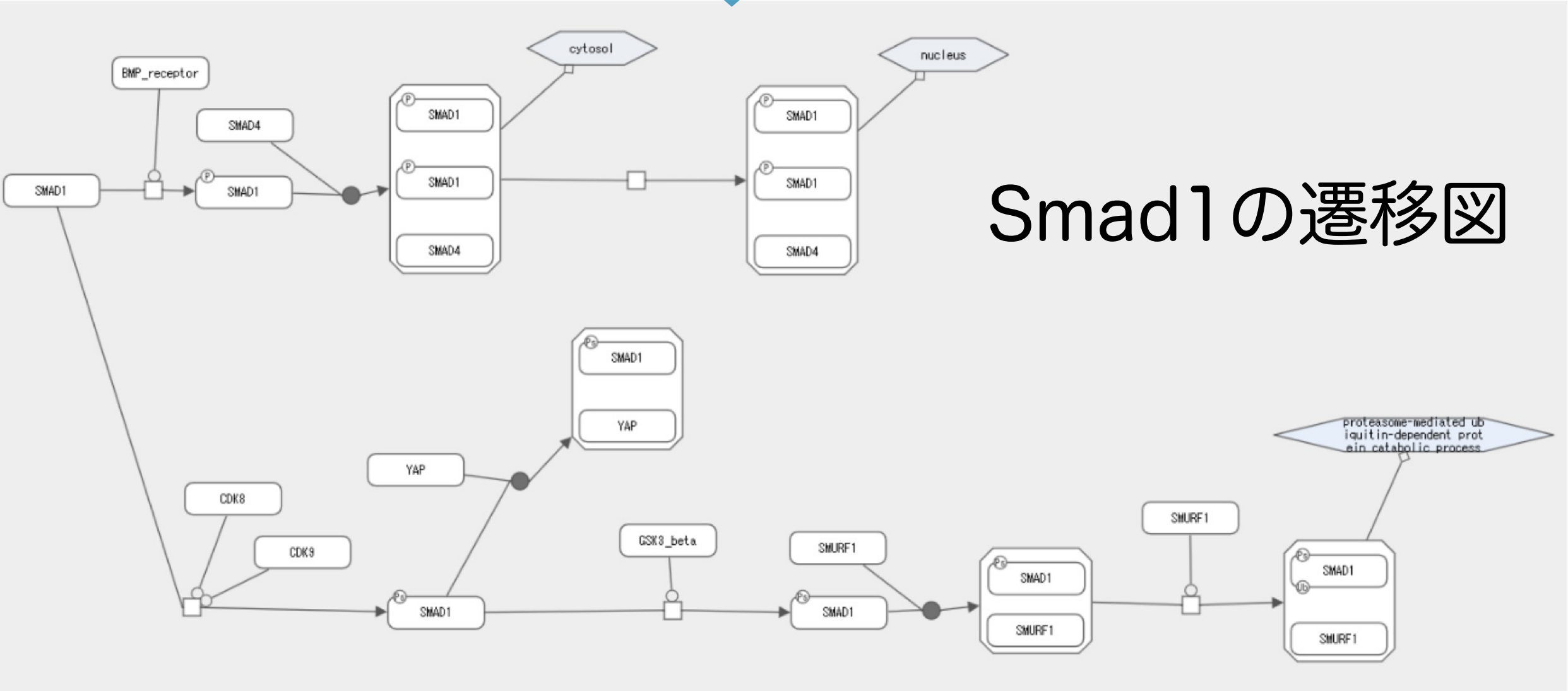


SBGN-ML

ChiSE

グラフツール

ChiSE is a library with an API based on [SBGNViz.js](#), which in turn is based on pathway models represented by process description (PD) and activity flow interaction format (SIF).



Smad1の遷移図



アノテータ

IDEAL-RDF

IDEAL-XML

リン酸化に伴うイベントのデータ化・可視化・標準化
(同じデータからは同じ図が描かれる) 作成済：60エントリ

NeProc：機械学習を用いた”天然変性領域・ProS”予測プログラム



IDEAL

*I*ntrinsically *D*isordered proteins with *E*xtensive *A*nnotations and *L*iterature

Version 05/Oct/2021 was released.

What is IDEAL ?

IDEAL provides a collection of knowledge on experimentally verified intrinsically disordered proteins (IDPs) [more](#)

What's new

Find the new features of IDEAL.

Browse

See the list of all proteins in IDEAL.

Search

- [Keyword](#)
- [Blast search](#)
- [SPARQL](#)

Download

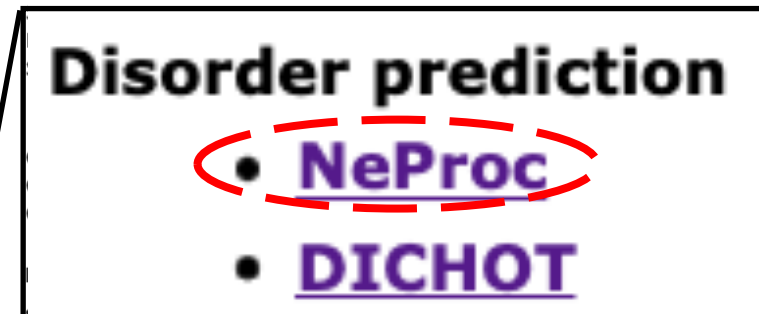
- [XML](#)
- [RDF](#)
- [Phosphorylation](#)

Help

[Statistics](#)

Disorder prediction

- [NeProc](#)
- [DICHOT](#)



Next *ProS* classifier

NeProc (Next ProS Classifier) predicts intrinsically disordered regions (IDRs) and protein binding segments in IDRs, which are called ProS (protean segment) in the [IDEAL](#) database.

Please input your e-mail address and amino acid sequence.

The result will be sent by e-mail.

Mail Address :

ideal@xxx.com

Query PROTEIN sequence (must be in the single FASTA format):

```
>sp|P04637|P53_HUMAN Cellular tumor antigen p53 OS=Homo sapiens OX=9606 GN=P53 PE=1 SV=4
MEIPQSGFVSEPPPLSGETFSQUNKLLFENNVLSPFPSQAMQMLSPDIEQWTFEGDP
DEAPRMPEAAPVAPAPAPTPAAPAPAPSWPLSSVPSQKTYQGSYGFRLGLHSGTAK
SVTCTSPALNKMFCQLAKTCVQLWYDSTPPPTQTRVRAMATYKQSQHMTFVVRCPHIE
KCSDSGLAPQHLRIVEGNLVEYLDQNTFRHSVVVPEPPEVQSDCTTHYNYMACHS
SCMGGMNRRPILITLDESSGALLGRNSFEVRCACPDQRDRTEENLRKKQEPHHELP
PGSTKRALPNNTSSSPQPKKPLDQYFTLQIRREFRERFRELNEALELKDAQAGKEPG
DSRAHSHLSKKQGSSTSRHKKLMFKTEGDSQ
```

(Don't paste DNA sequence !)

[submit](#) [clear](#)

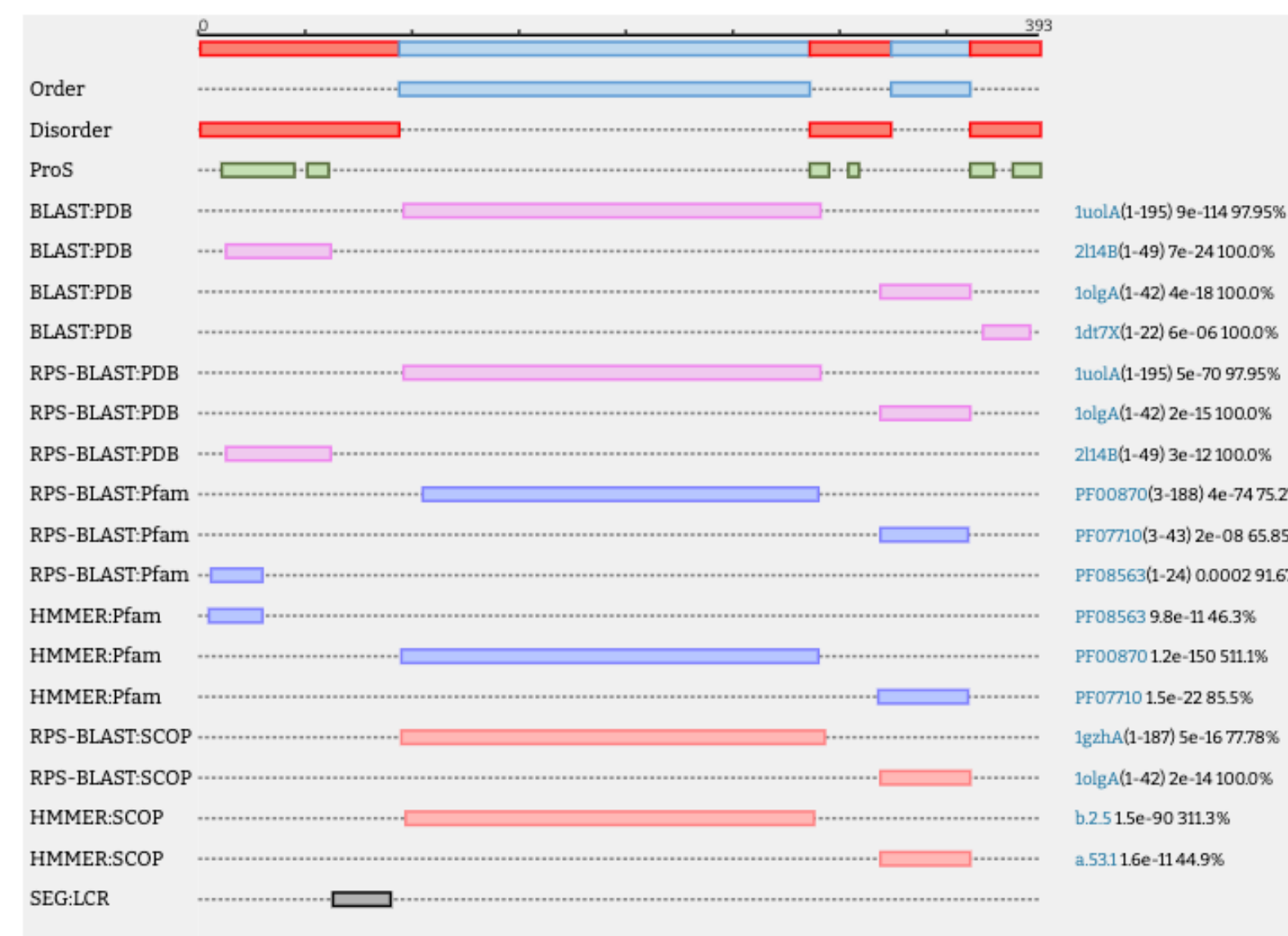
アミノ酸配列を入力



Input sequence

Sequence name	P04637	Length	393
Disorder content	41.7%	ProS content	20.9%
IDEAL	IID00015	Job ID	N20220829145141

Prediction



天然変性領域

構造領域

ProS

立体構造やPfamドメインに相当する領域

低複雑性領域, coiled coil領域, 膜貫通領域

IDEALに登録がないタンパク質はNeProcで予測ができる

IDEAL

おわりに

*Intrinsically **D**isordered proteins with **E**xtensive **A**nnotations and **L**iterature*

PDBが整備されてタンパク質の立体構造研究が飛躍的に進展したように、まだ未知のことだらけの**天然変性タンパク質研究の進展には、質の高いデータベースが必要不可欠**です。

IDEALは欧州のDisProtとともに、研究のインフラストラクチャとして**継続的、安定的な情報提供**を行ってきたいと考えています。

IDEAL開発チームの活動をご理解いただけると幸いです。