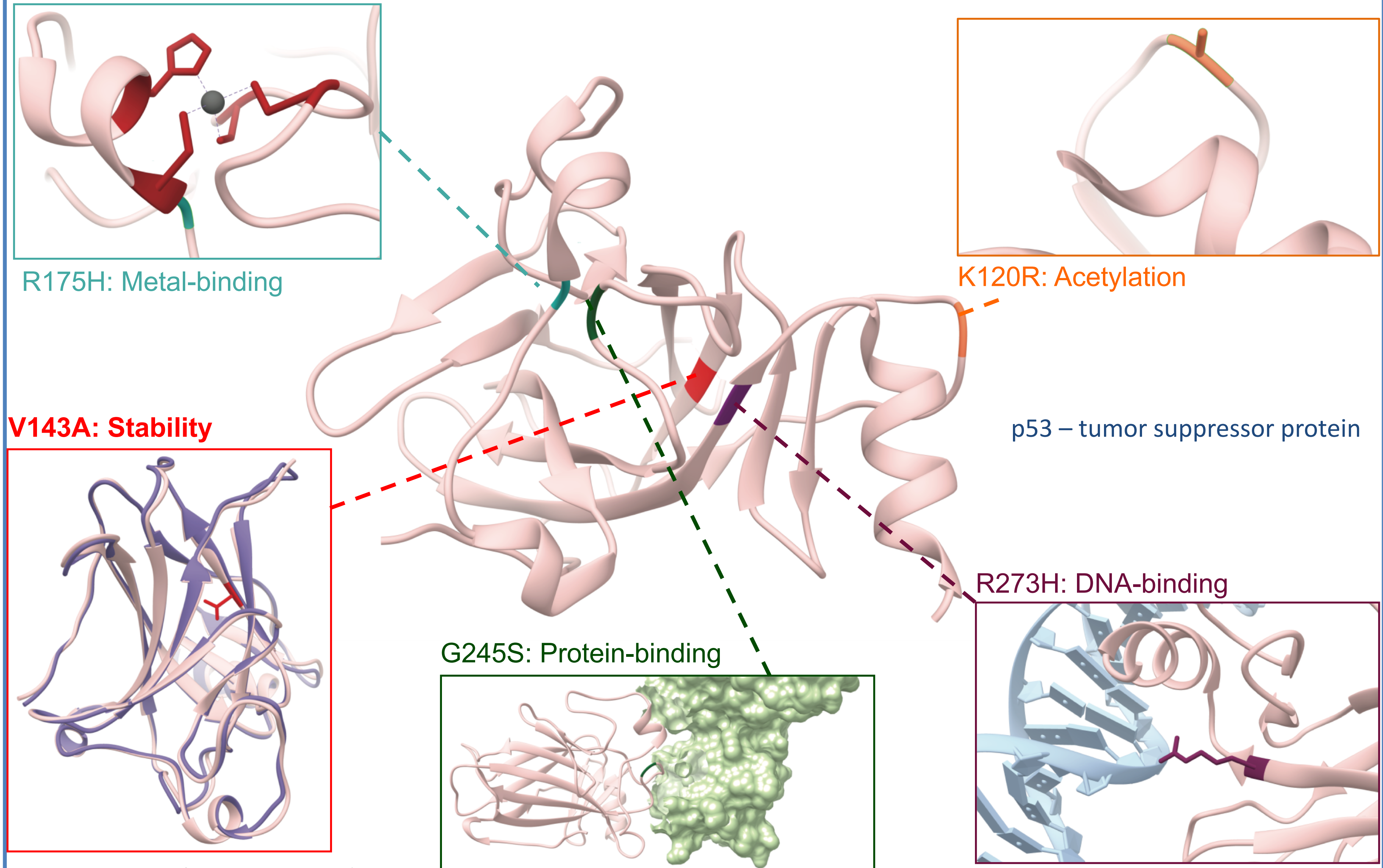
Examples of Functional Sites

A. Zinc-binding site B. Phosphorylation site C. DNA-binding site

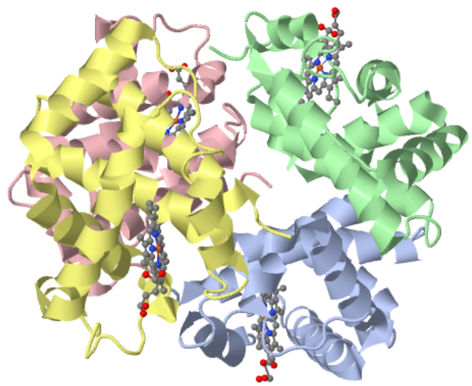
Xin & Radivojac. *Curr Prot Pept Sci* 12: 456 (2011).

Molecular Consequences (on P53)



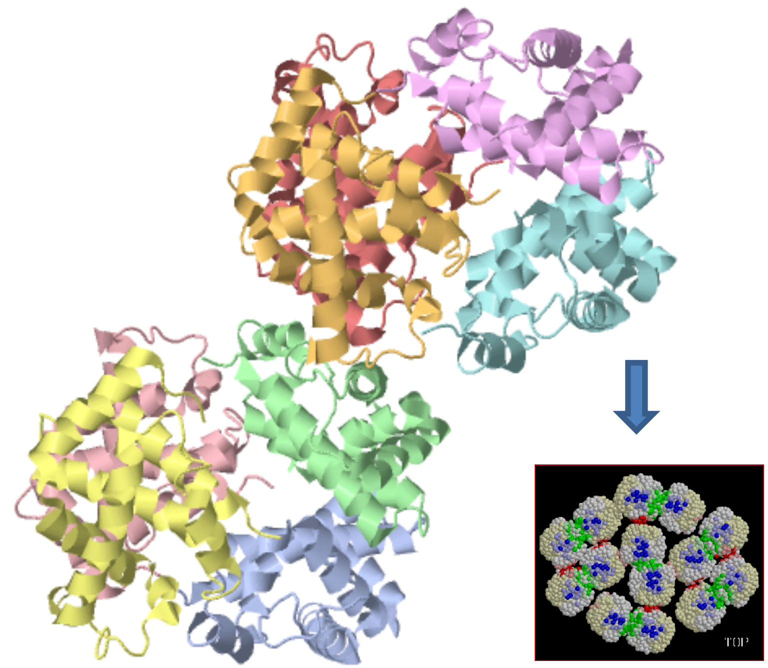
PDB structures: 2ybg, 2j1w, 1ycs and 1tup

Sickle Cell Disease: Gain of Function

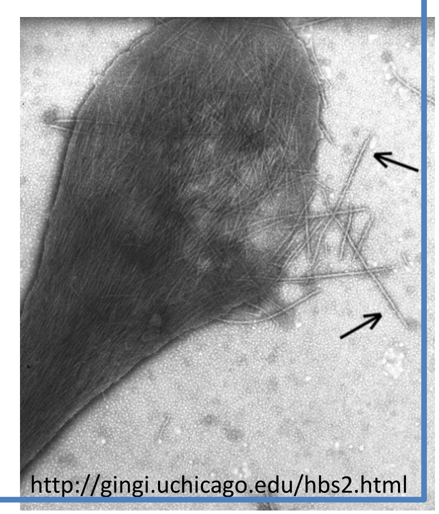


4hhb

E6V



2hbs



Sickle Cell Disease:

Autosomal recessive disorder

E6V in HBB causes interaction w/ F85 and L88

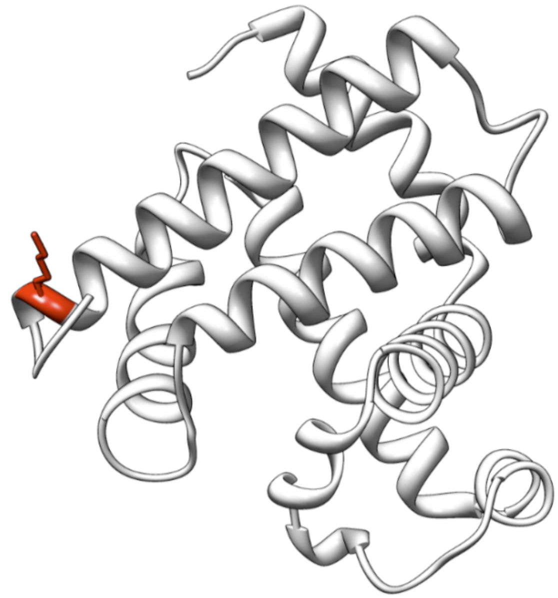
Formation of amyloid fibrils

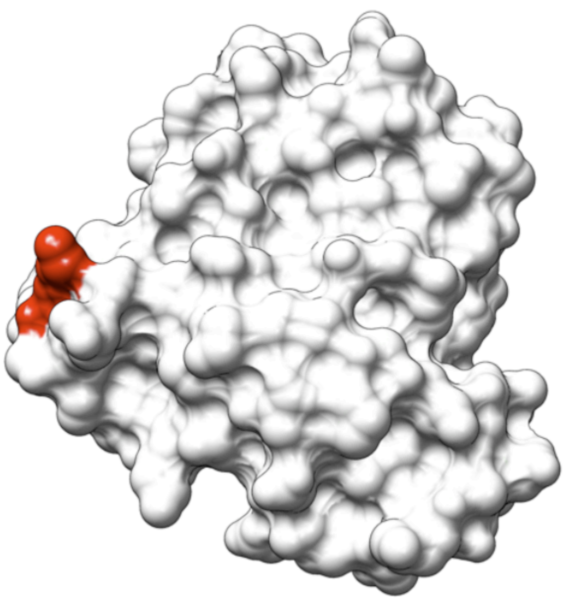
Abnormally shaped red blood cells, leads to sickle cell anemia

Manifestation of disease vastly different over patients

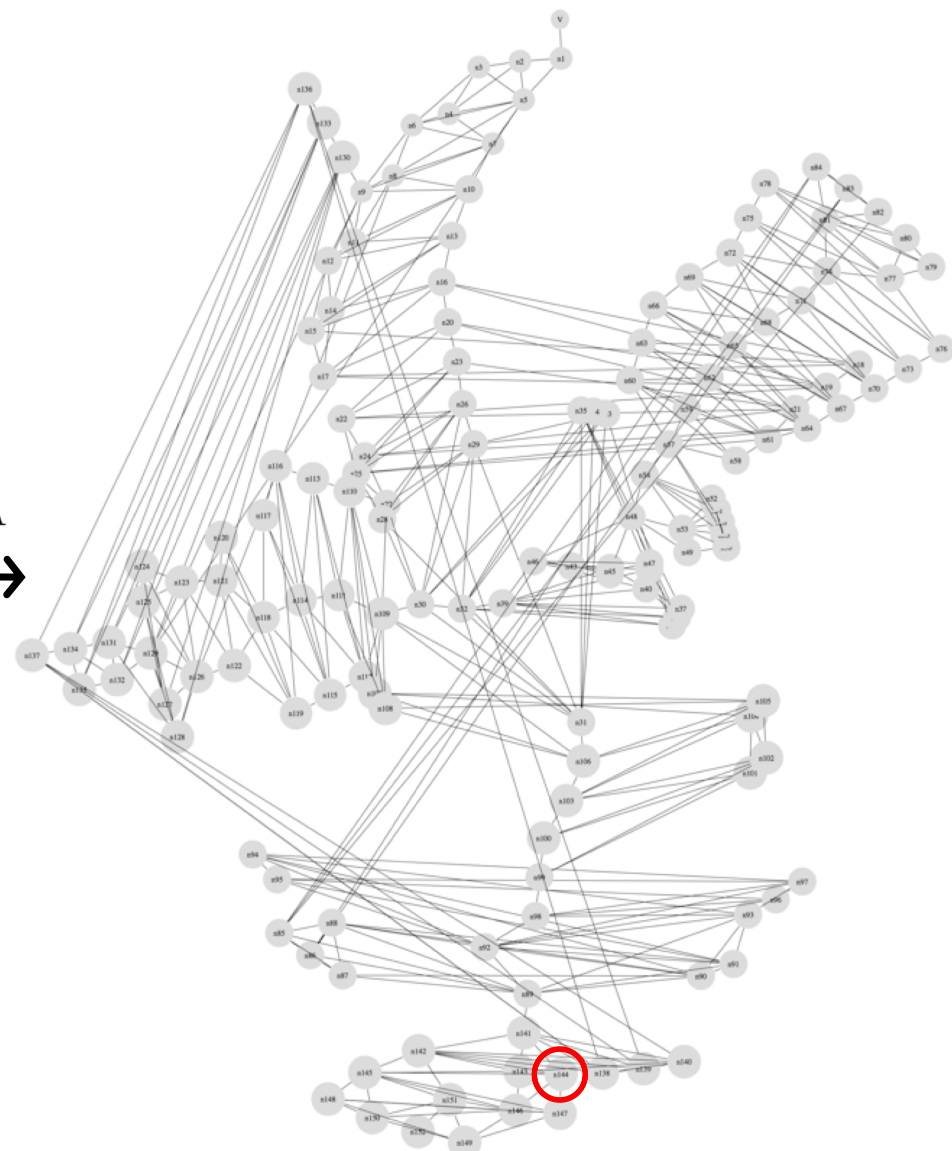
Pauling *et al*. *Science* 110: 543 (1949). Chui & Dover. *Curr Opin Pediatr*, 13: 22 (2001).

Prediction of Functional Residues from Structure



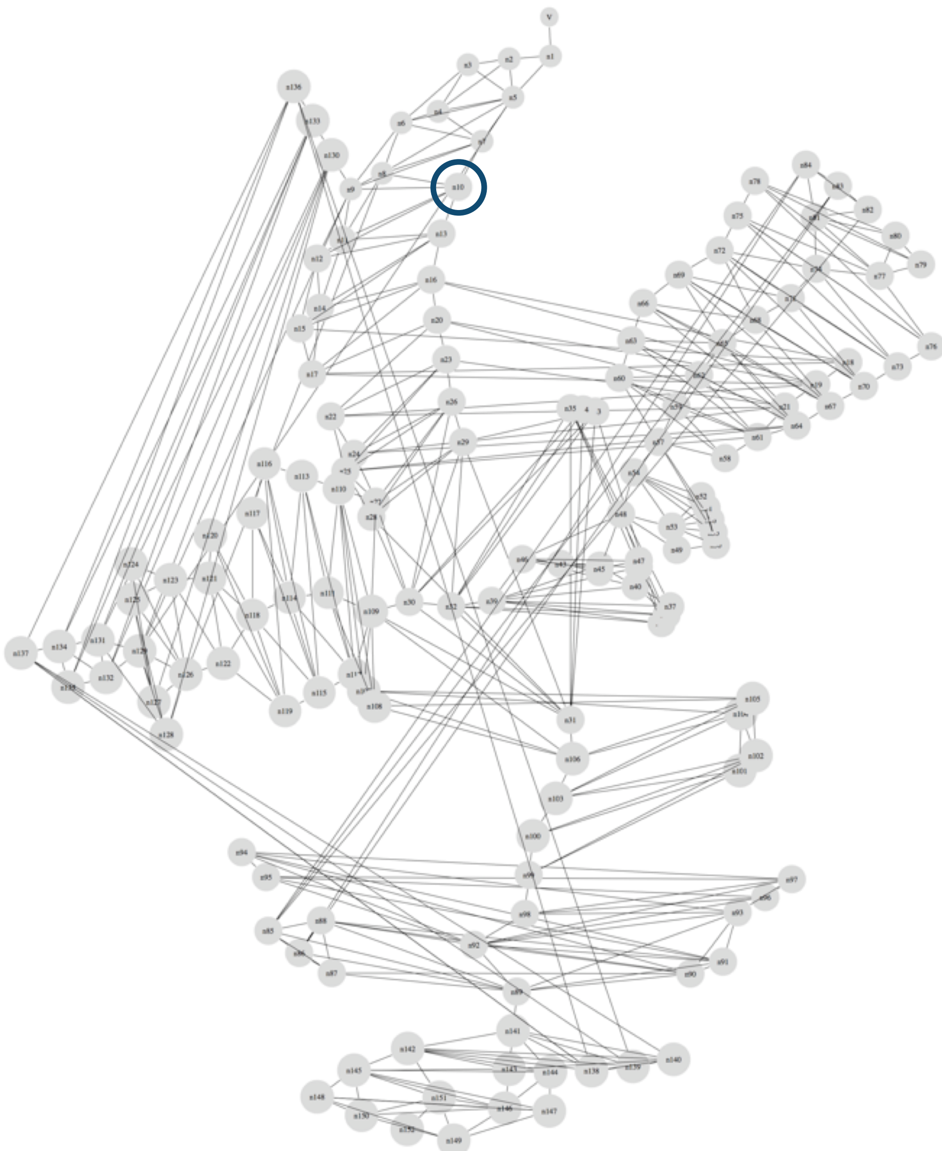


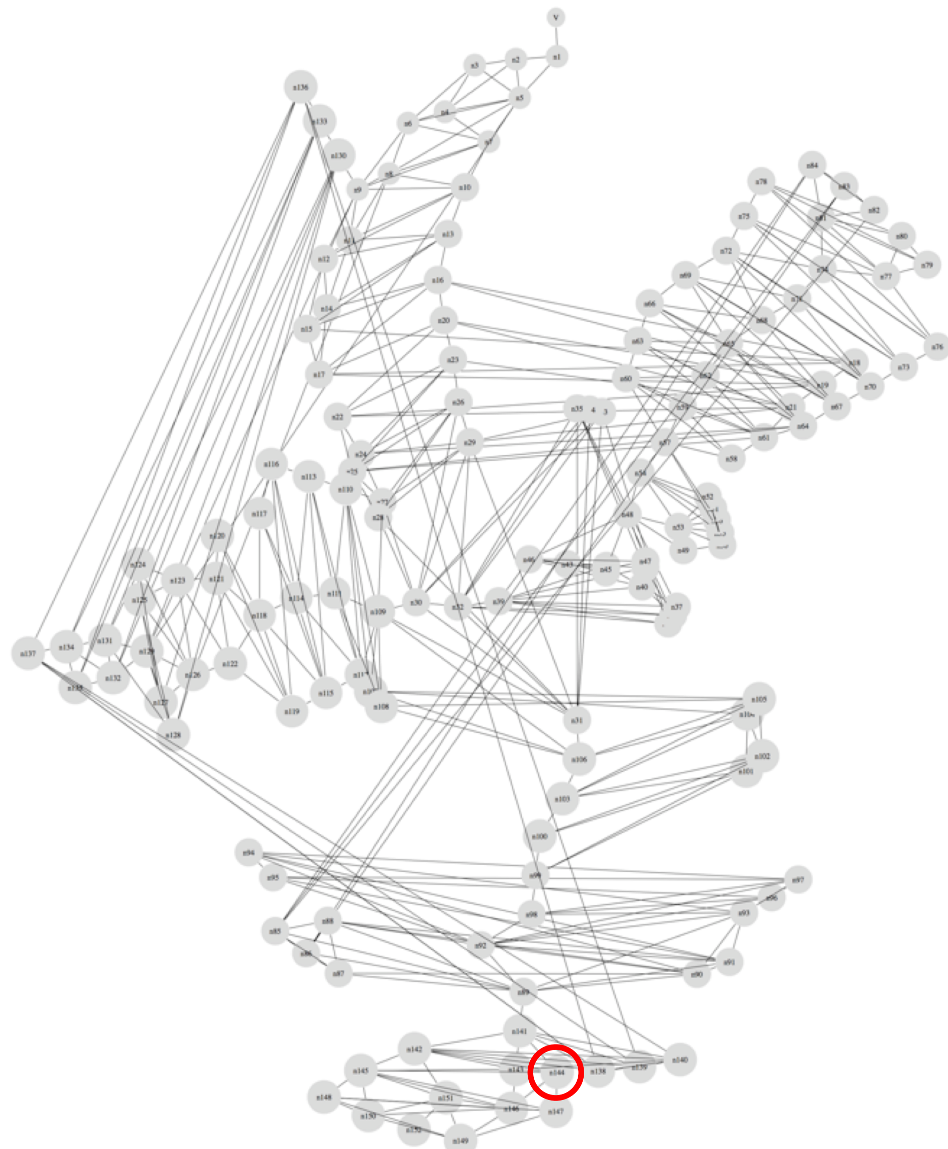
$$C_\alpha - C_\alpha \leq 6\text{Å}$$



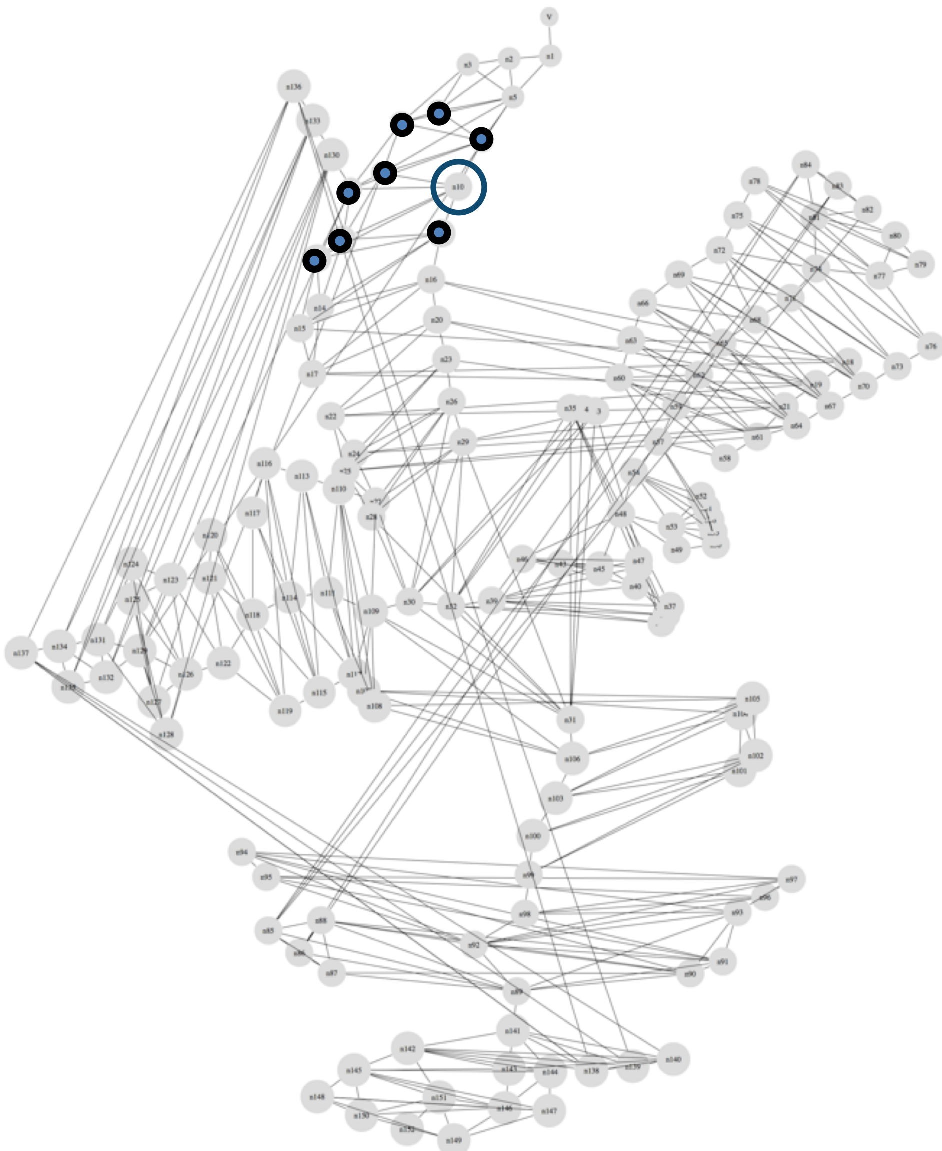
Myoglobin 1.4A X-ray PDB: 2jho 153 residues

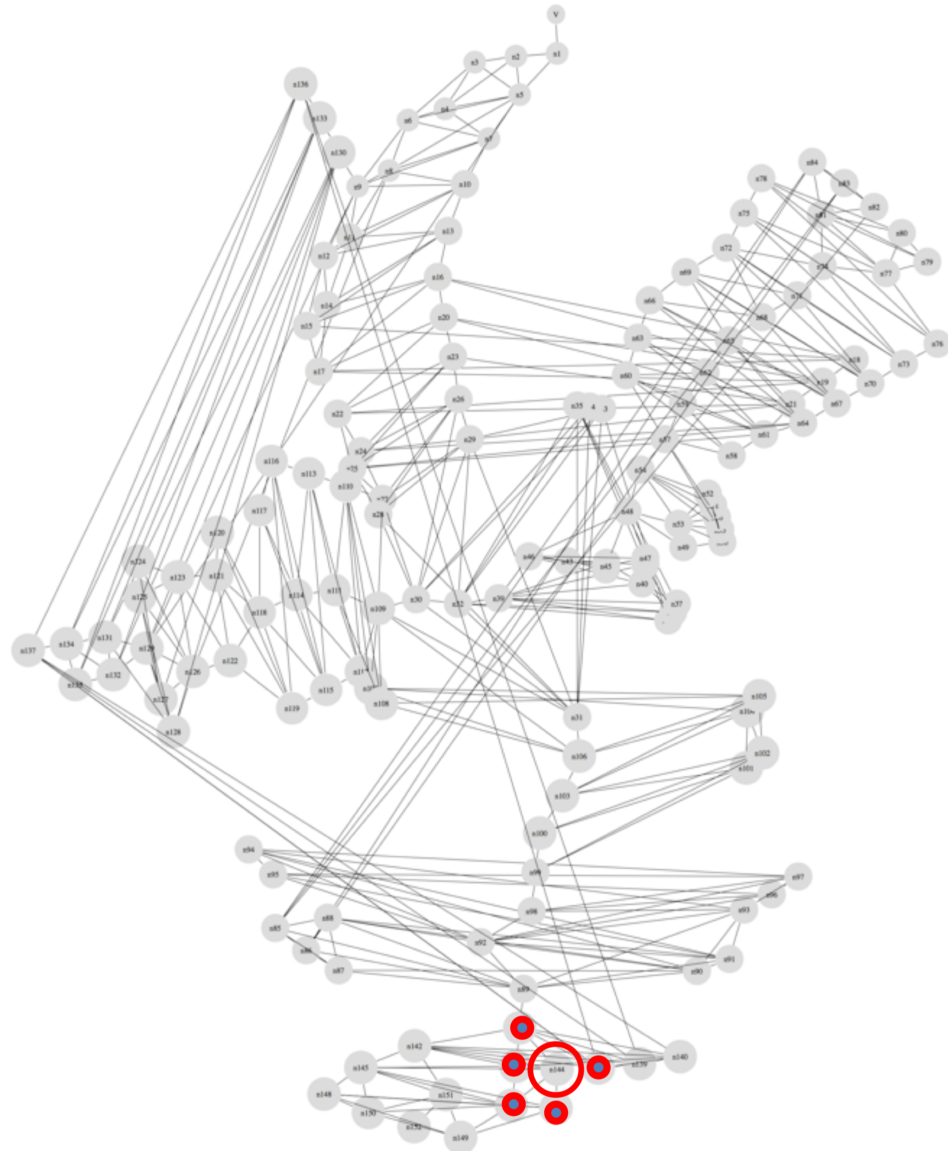
Protein Contact Graph





Protein Contact Graph





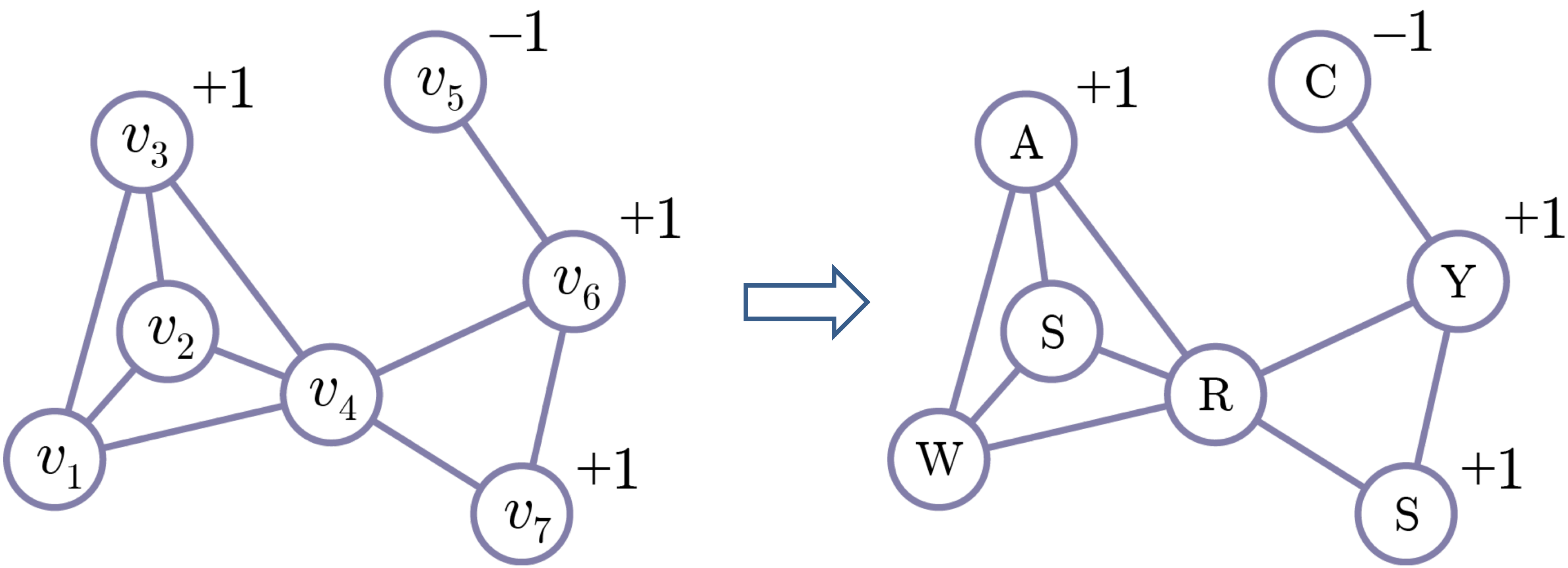
Notation and Problem Formulation

$G = (V, E)$

$t : V \rightarrow \{-1, +1\}$

$\Sigma = \{\text{A, C}, \ldots, \text{W, Y}\}$

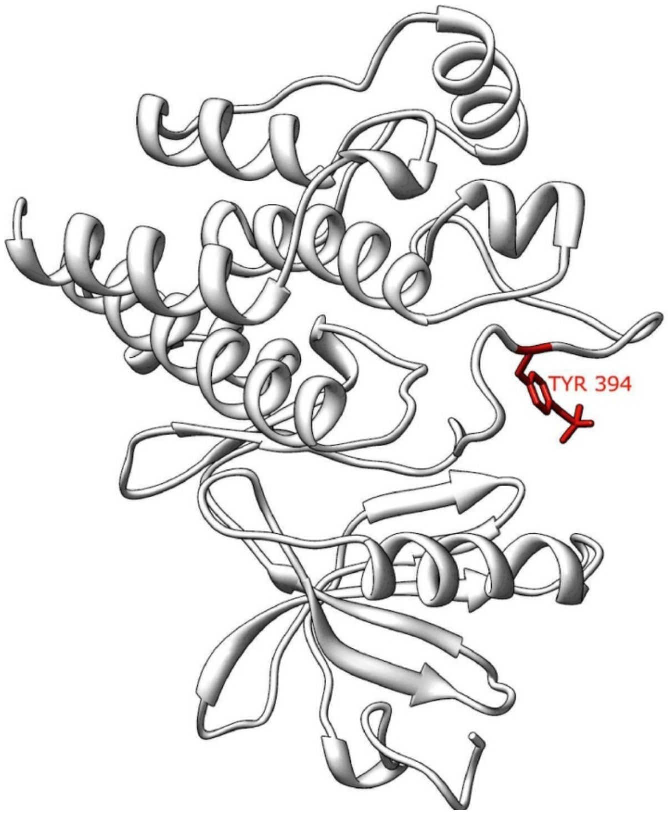
$f : V \rightarrow \Sigma$



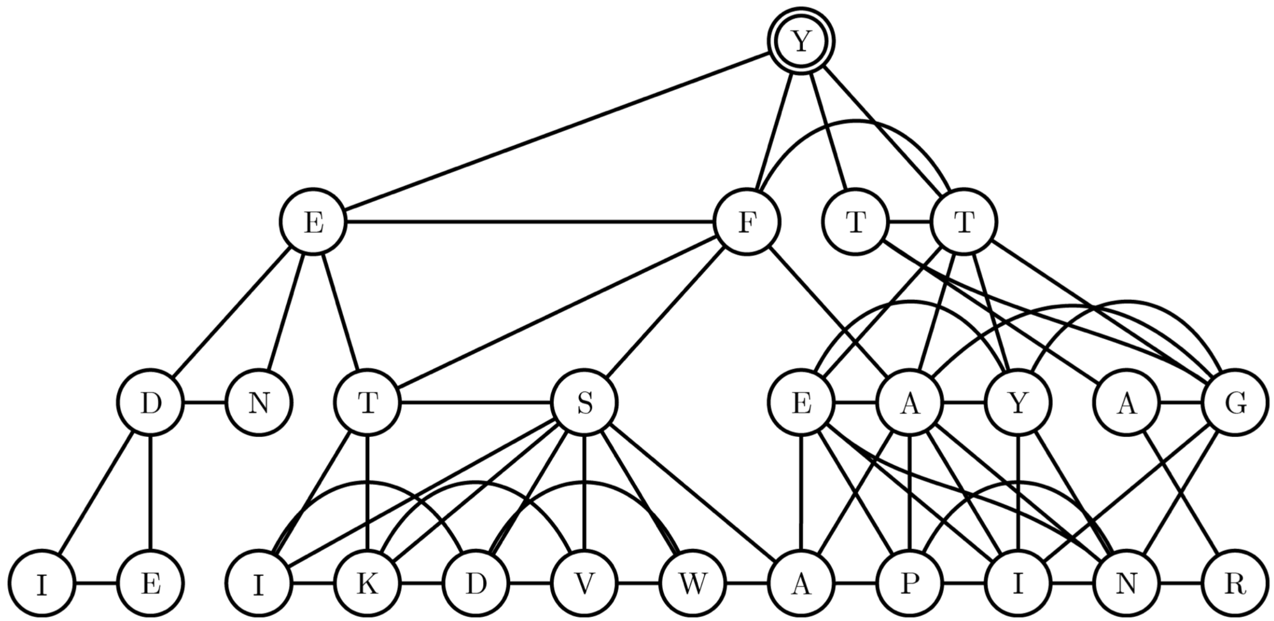
Objective: predict class label for each unlabeled vertex

e.g. $\Pr(t(v_4) = +1 | D) = 0.9$

Residue Neighborhood

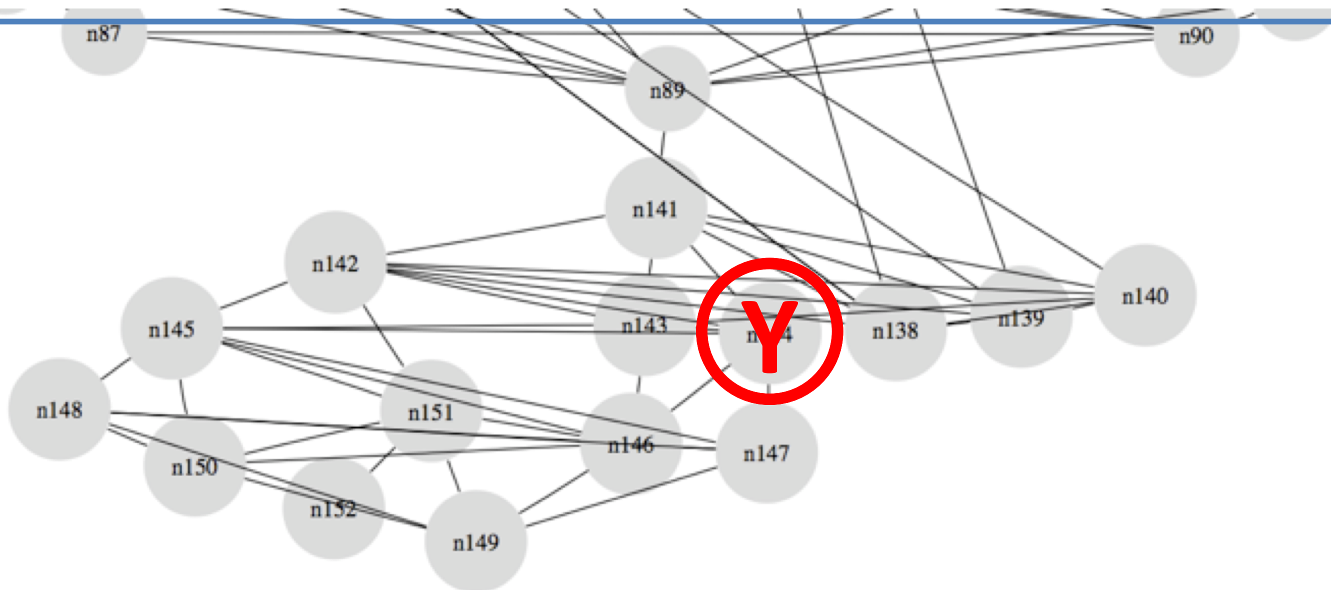


Y394 of human lymphocyte kinase



Depth-3 graph neighborhood for Y394

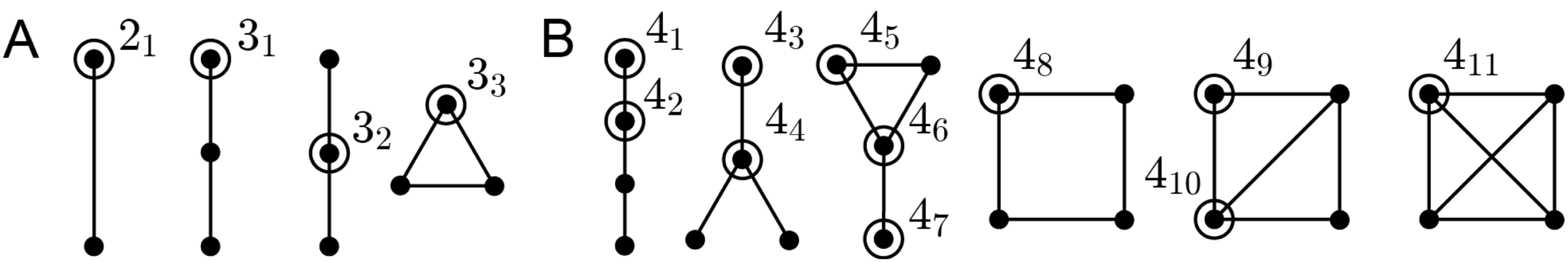
Vacic *et al. J Comput Biol* 17: 55 (2010).



Idea:

1. Count small graphs rooted at the vertex of interest
   i. *Graphlets*
2. Create a similarity measure between the counts for two vertices
   i. *Kernel function*
3. Use kernel functions for supervised classification
   i. *Support vector machines*

$\Rightarrow$ **Graphlet Kernel**



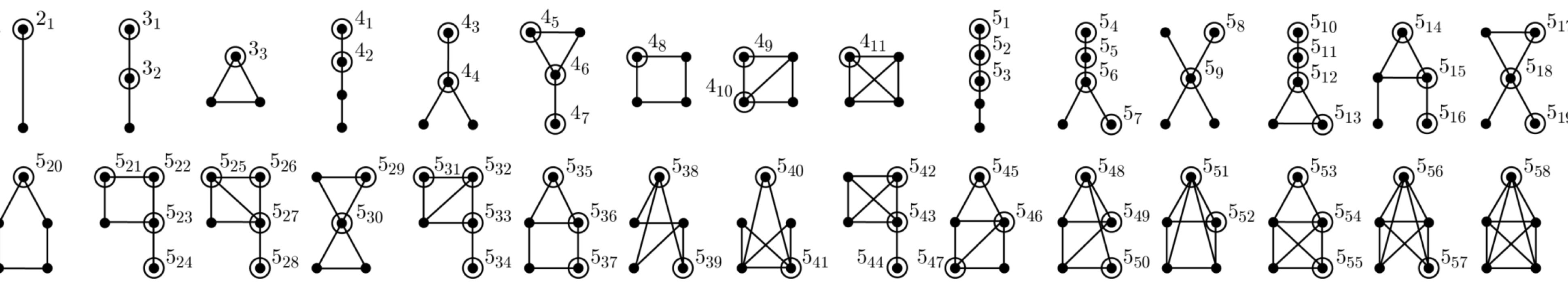
Graphlets: simple small connected rooted graphs.

Przulj *et al*. *Bioinformatics* 20: 3508 (2004).

Graphlets

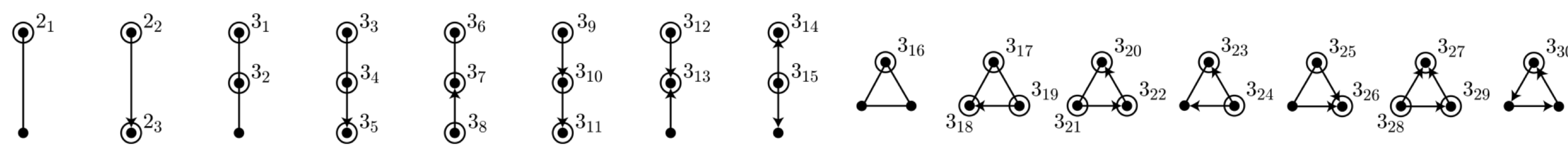
Undirected:

A



Directed:

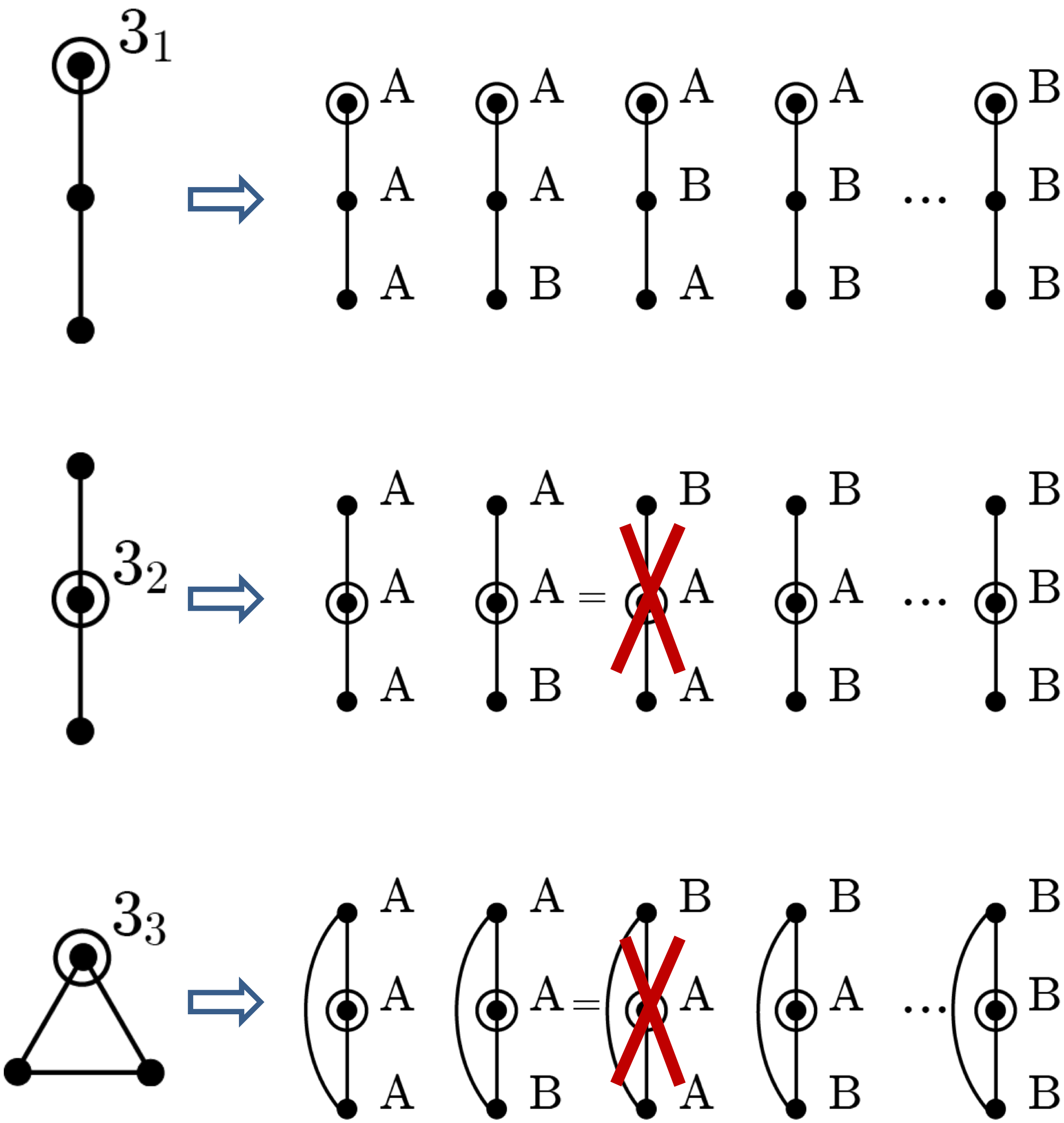
B



Graphlets: simple small (typically of order 5 or less) connected rooted graphs.

LABELED GRAPHLETS

$$\boxed{\kappa(n,\Sigma) = 20}$$



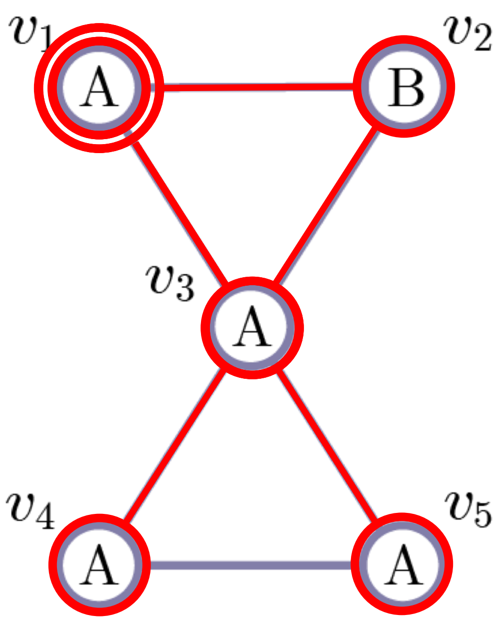
$$|\Sigma|^n = 2^3 = 8$$

$$|\Sigma| \cdot \binom{|\Sigma|+1}{|\Sigma|-1} = 6$$

same symmetry class

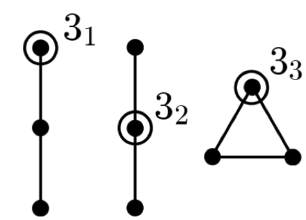
$$|\Sigma| \cdot \binom{|\Sigma|+1}{|\Sigma|-1} = 6$$

EXAMPLE



$$V = \{v_1, v_2, v_3, v_4, v_5\} \qquad \Sigma = \{\mathrm{A}, \mathrm{B}\}$$

$$f : V \to \Sigma$$
$$f(v_1) = \mathrm{A}$$
$$f(v_2) = \mathrm{B}$$
$$\vdots$$



$\phi_3(v_1)$

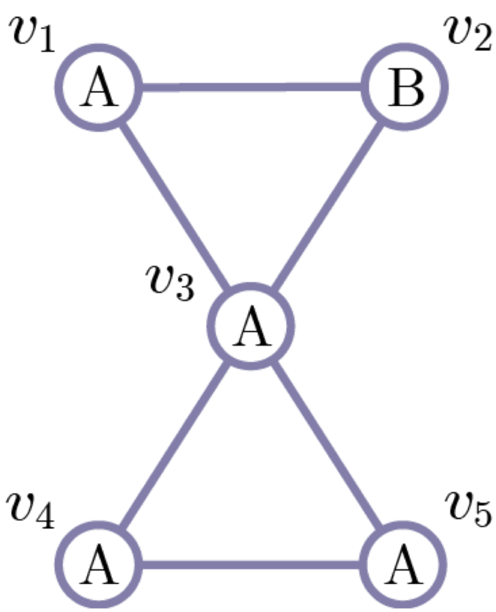
	AAA	AAB	ABA	ABB	BAA	BAB	BBA	BBB	AAA	AAB	ABB	BAA	BAB	BBB	AAA	AAB	ABB	BAA	BAB	BBB
$\phi_3(v_1)$	2															1				
	3_1								3_2						3_3					
$\phi_3(v_5)$	1	1													1					

$\phi_3(v_5)$

$$k_3(v_1, v_5) = \langle \phi_3(v_1), \phi_3(v_5) \rangle = 2$$

Vacic *et al. J Comput Biol* 17: 55 (2010).

Graphlet Kernels



$$k_n(u,v) = \langle \phi_n(u), \phi_n(v) \rangle$$

where

$$\phi_n(v) = (\varphi_{n_1}(v), \varphi_{n_2}(v), \ldots, \varphi_{n_{\kappa(n,\Sigma)}}(v))$$

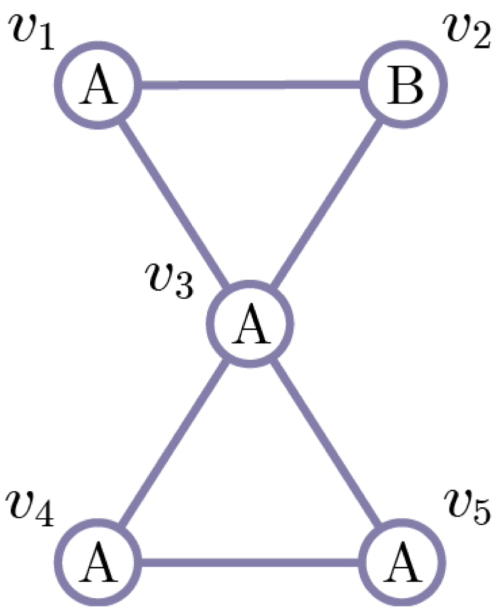
$$k(u,v) = \sum_{n=1}^{N} k_n(u,v)$$

Graphlet kernel

$$k'(u,v) = \frac{k(u,v)}{\sqrt{k(u,u)\,k(v,v)}}$$

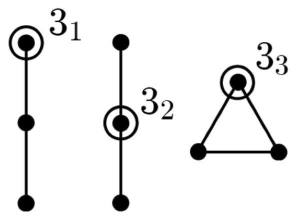
Normalized graphlet kernel

Edit Distance Graphlet Kernels



Jose Lugo-Martinez

IDEA: Allow approximate matching,
i.e. allow mismatch in labels or edges



3_1								3_2						3_3					
AAA	AAB	ABA	ABB	BAA	BAB	BBA	BBB	AAA	AAB	ABB	BAA	BAB	BBB	AAA	AAB	ABB	BAA	BAB	BBB
2	2	2		2										1	1	1		1	
2	2	1	1	1	1									1	1		1		

$$k^{l}_{(3,1)}(v_1, v_5) = \left\langle \phi^{l}_{(3,1)}(v_1), \phi^{l}_{(3,1)}(v_5) \right\rangle = 14$$

Lugo-Martinez, Radivojac. *Network Science* 2(2): 254-276 (2014).

RESULTS: PHOSPHORYLATION

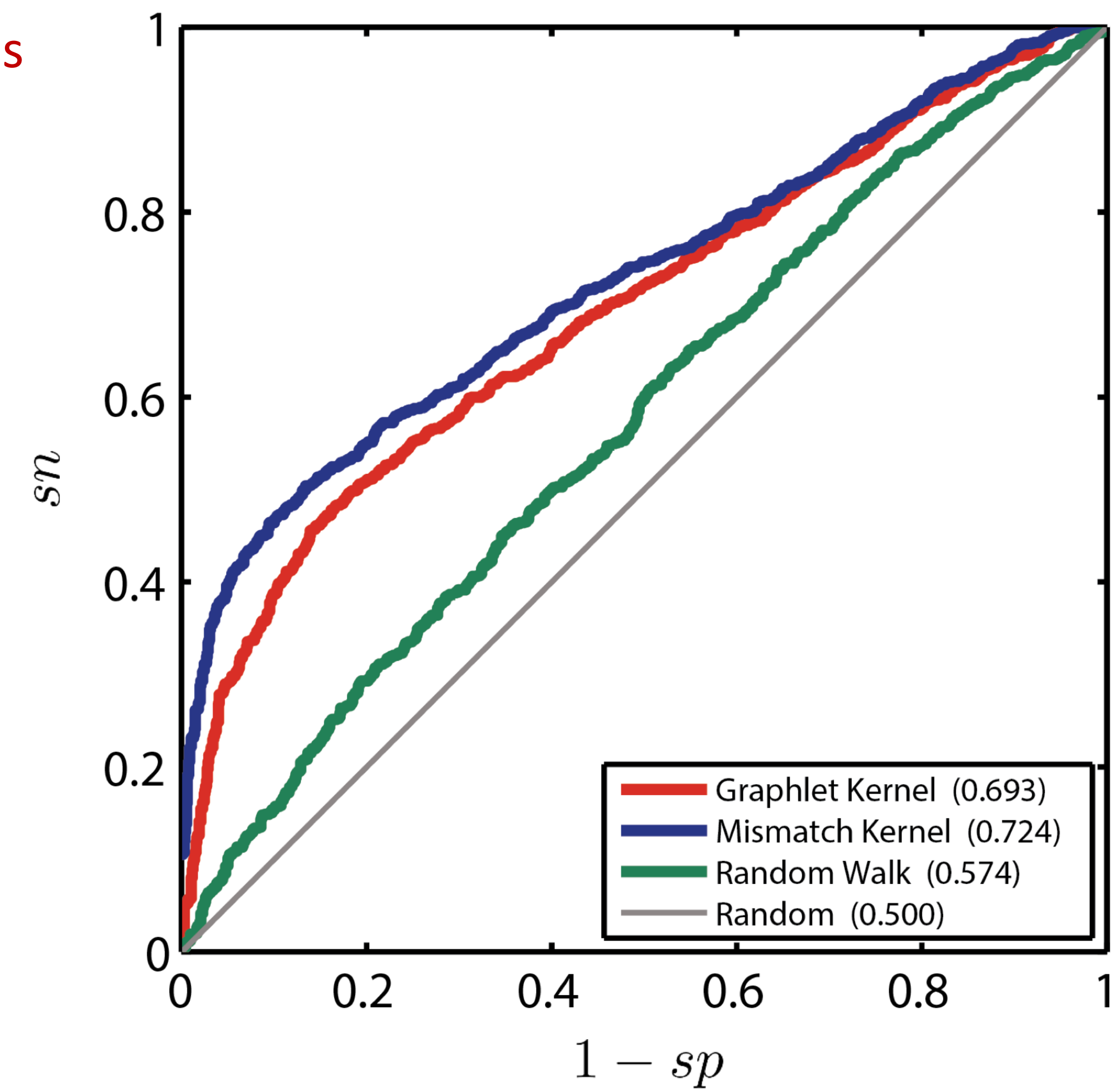
Task: predict phosphorylation sites from protein structures

Data:

$$n_+ = 1157$$

$$n_- = 1157$$

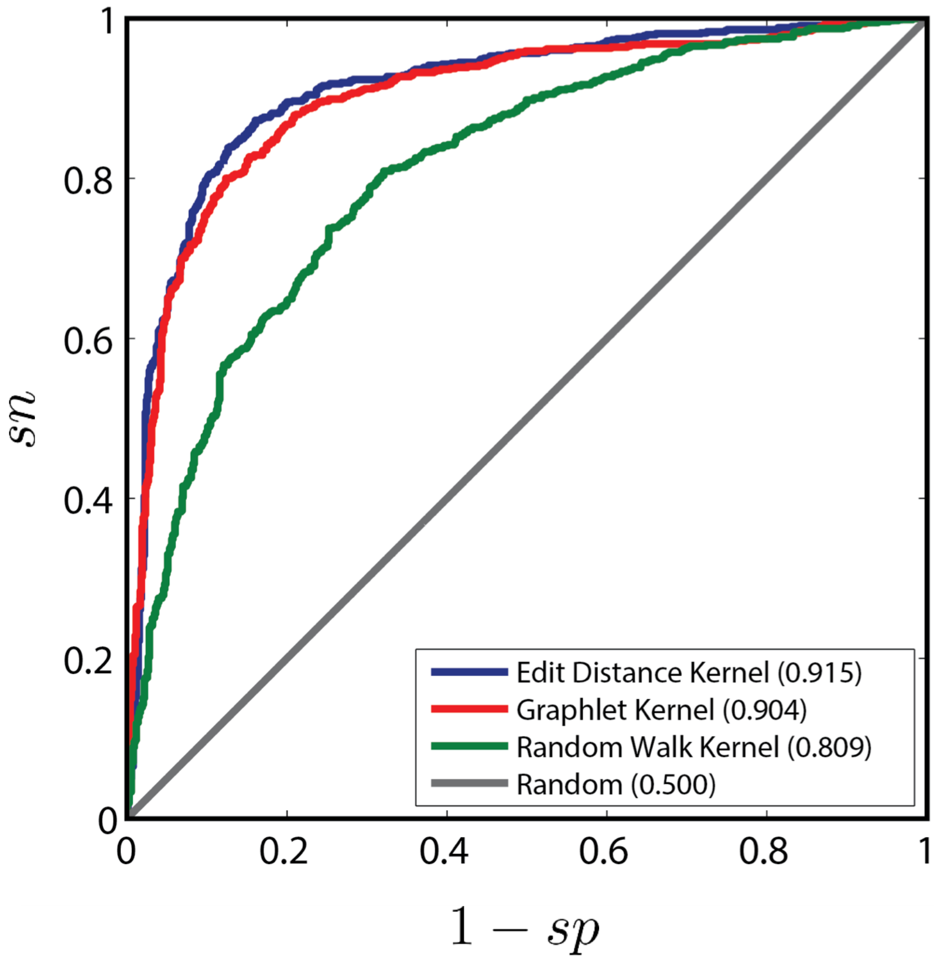
$$|\Sigma| = 40$$



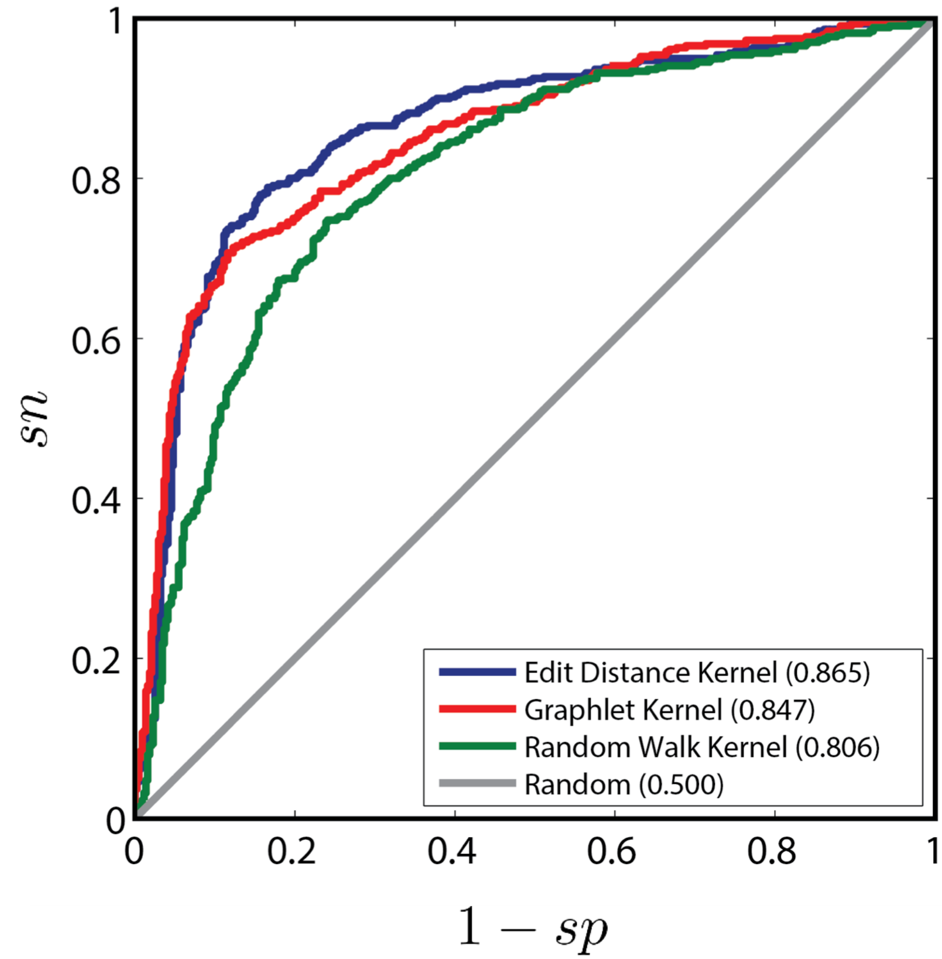
Lugo-Martinez, Radivojac. *Network Science* 2(2): 254-276 (2014).

Results: Metal Binding Residues

Iron (Fe)



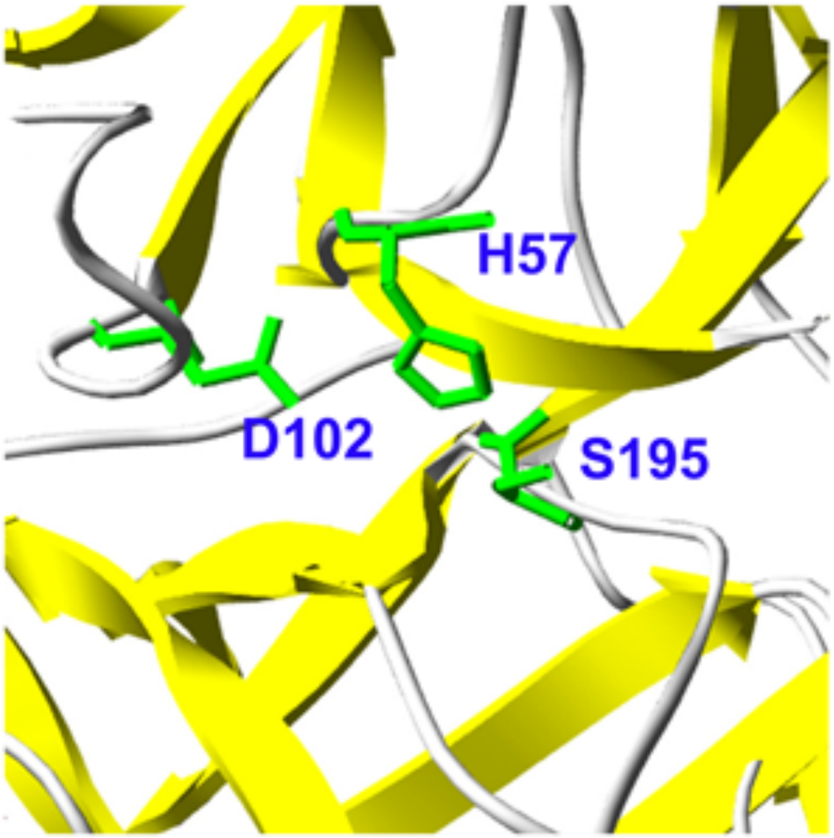
Copper (Cu)



Jose Lugo-Martinez, Indiana University

Loss of Catalytic Activity Causes Disease

Catalytic triad of F9 (1frn)



- F9 is activated in response to injury of the blood vessel and has a key role in blood clot formation;
- Mutations in F9 give rise to the X-linked recessive disorder, hemophilia B;
- H57R leads to the loss of catalytic activity.

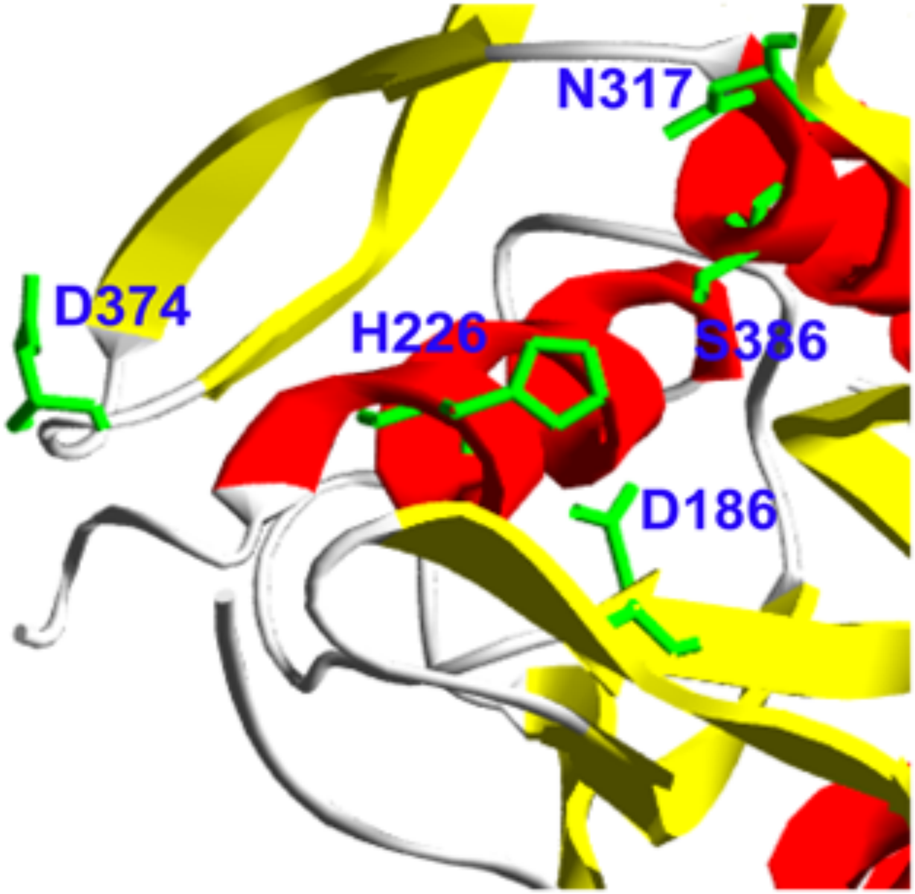
Xin *et al. Bioinformatics* 26: 1975 (2010)

Attali *et al. Thromb Haemost* 82: 1437 (1999)

Gain of Catalytic Activity Causes Disease

- a member of the proteinase K sub-family of subtilases that reduces the number of LDL receptors in liver through a posttranscriptional mechanism.
- D374Y leads to a 10-fold increase in catalytic activity that causes hypercholesterolemia.

Catalytic residues of PCSK9 (2qtw)



Lagace *et al. J Clin Invest*, 116: 2995 (2006).

Xin *et al. Bioinformatics* 26: 1975-1982 (2010).