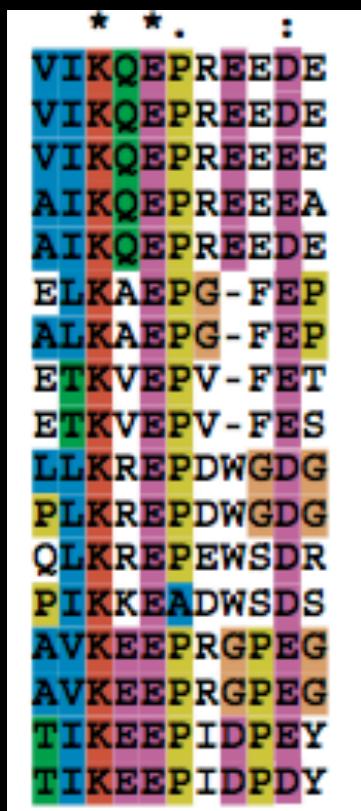




IIB-UNSAM, Buenos Aires
21-11-2017



*Warm-Up:
Thoughts and questions on
sequence-structure-function
relationships*

Toby J. Gibson
Structural & Computational
Biology Unit
EMBL, Heidelberg

Some questions

How do proteins fold?

What do you understand by intrinsically disordered protein?

Were you taught about IDP/IUP at University?

If so was it in a course text book?

Why do we talk about protein motifs?

Who is a bioinformatics predoc?

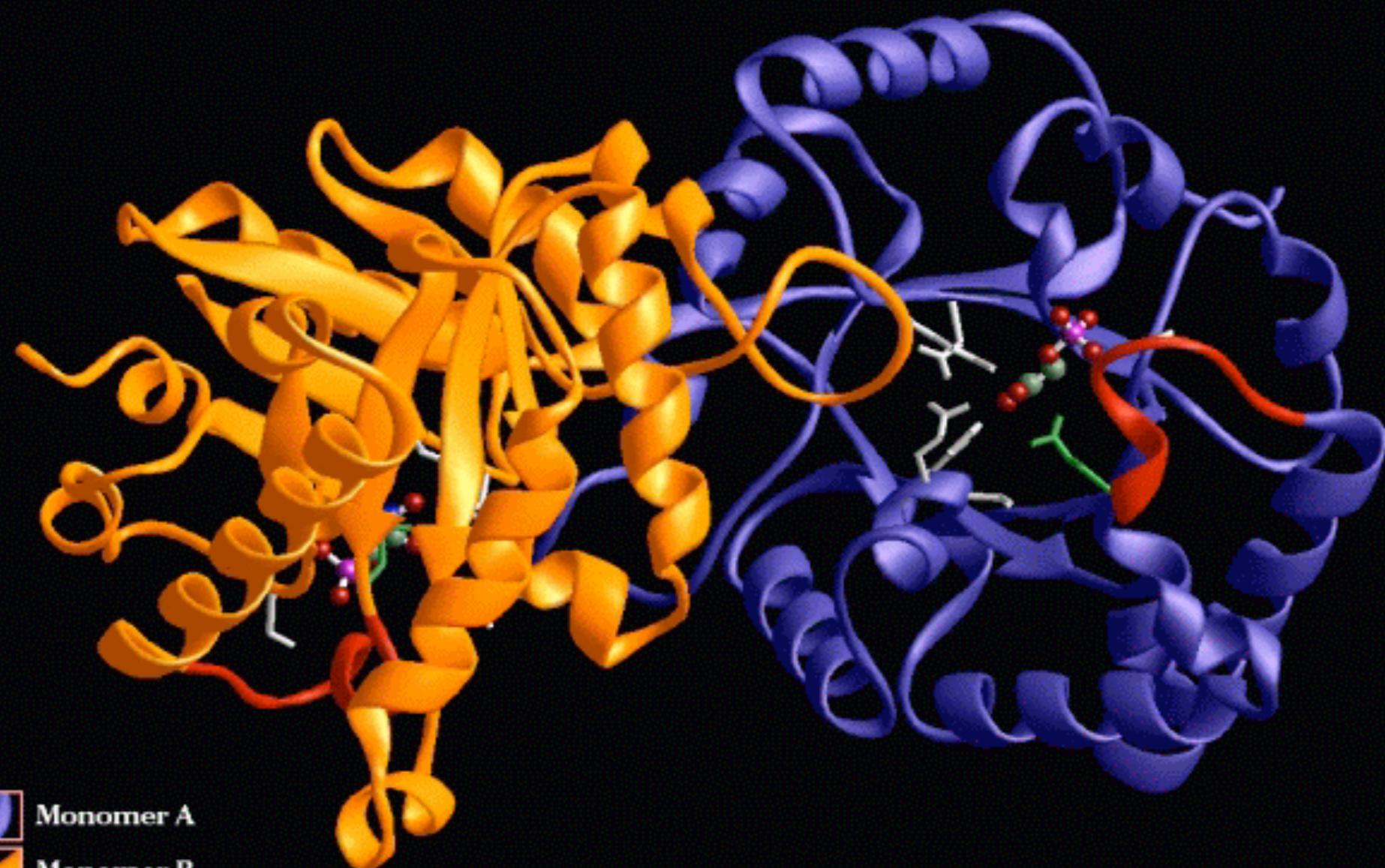
Sequence/Structure/Function/Networks

In the course, we are looking at bioinformatics tools that might give insight into these relationships

The enzyme Triose Phosphate Isomerase

A typical globular protein

Triosephosphate Isomerase (chicken)



Monomer A



Monomer B



Flexible loop involved in PGH binding



Phosphoglycolohydroxamate (PGH)



Binding pocket



Glu 165 (catalytic base)



Unstructure Determines Function

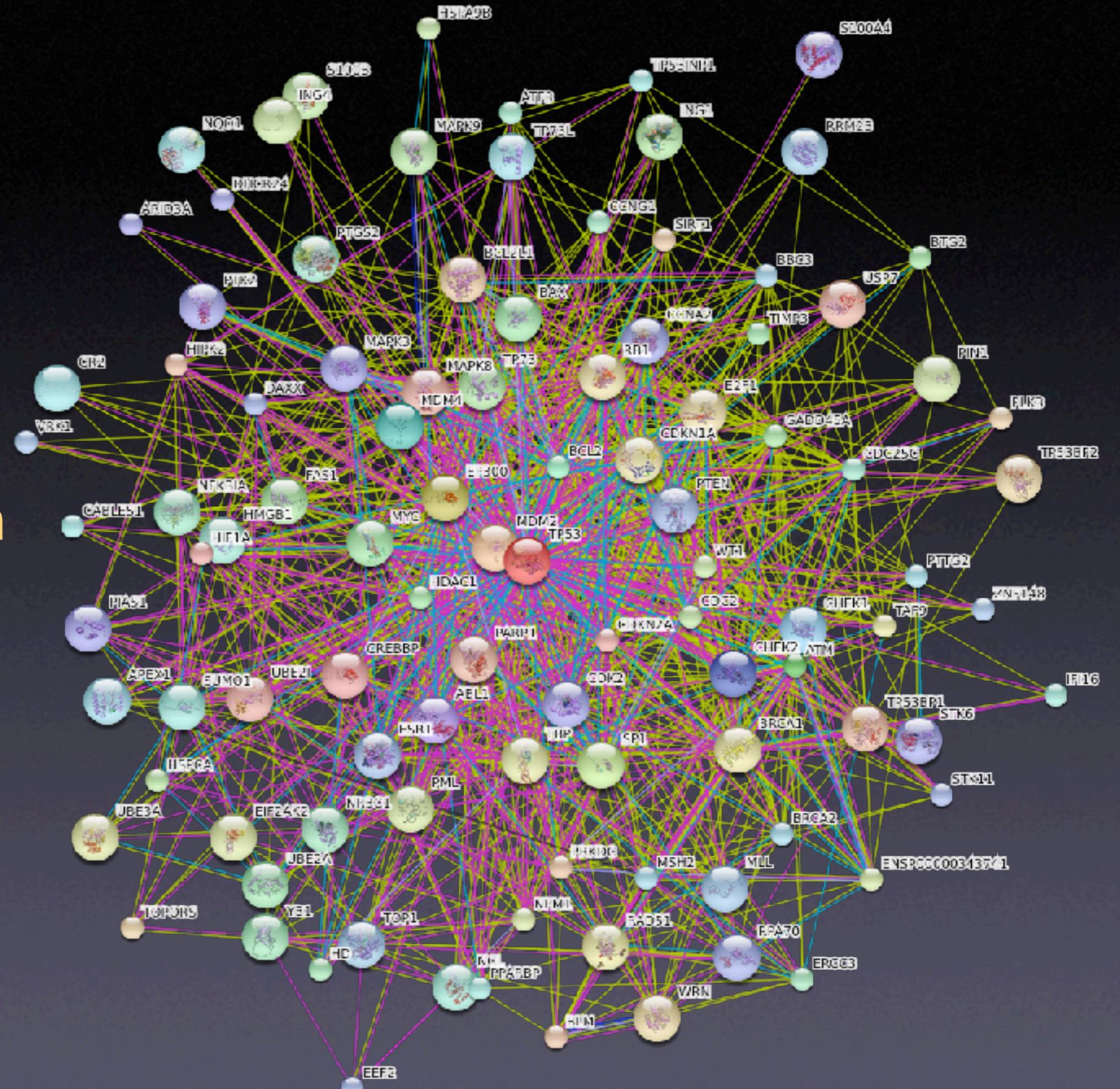
TRÈS
PASSÉ

SO LAST
MILLENNIUM

Question
How many interactions does
P53 make?

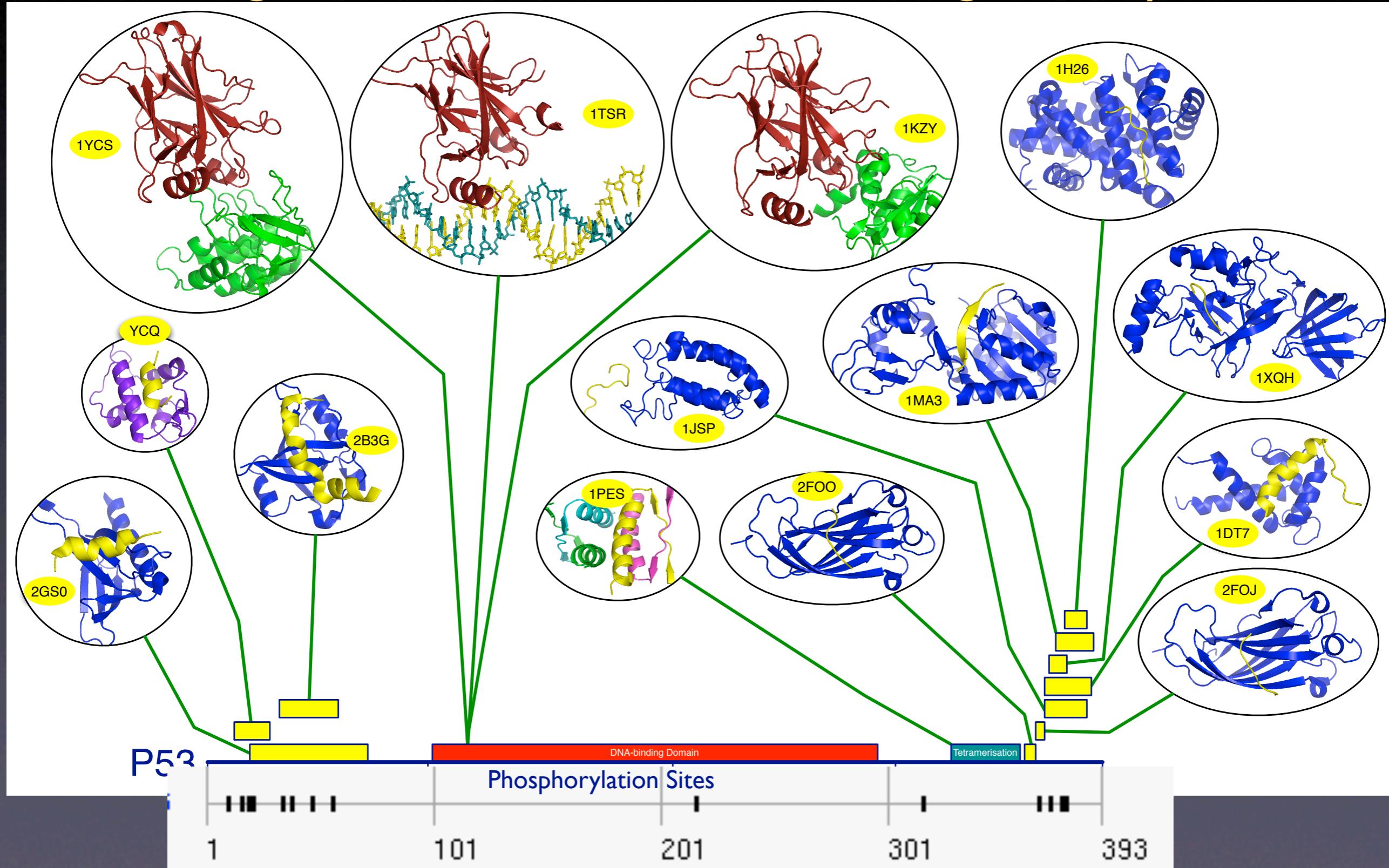
The P53 protein is a hub with many reported interactions

P53 interactions
listed in the
STRING protein-
protein interaction
resource



Molecular switching with P53

IUP makes more interactions than Globdom / Mutually exclusive binding / Alternative conformations / Regulated by PTM



Question
How many phosphorylation
sites can a protein have?

Many proteins are very heavily phosphorylated

Phosphoproteomics
champion
SRRM2_HUMAN

- 2750 AA
- 100s of Phospho-Sites
- 0 globular domains

Associated with splicing
but poorly researched
(15 PubMed citations)

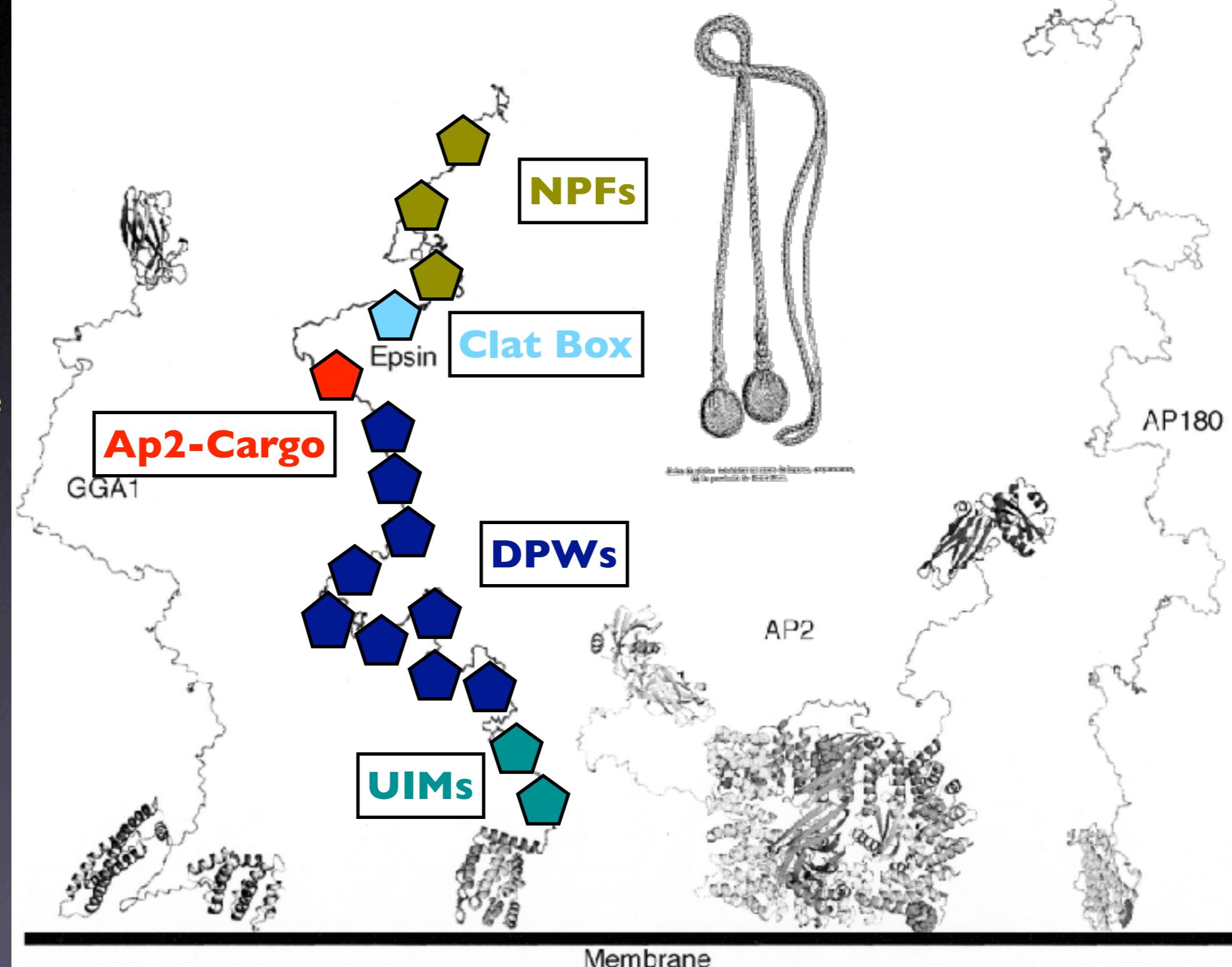
Pos/aa	Sequence	5507	SRRM2_HUMAN		104183		SRRM2_HUMAN		104183		SRRM2_HUMAN		104183		SRRM2_HUMAN		HTP			
			5507	5508	104183	104183	5507	5508	104183	104183	5507	5508	104183	104183	5507	5508	104183	104183	5507	5508
S145	RKQKQKQGIDDEYDCEK	5507	RKQKQKQGIDDEYDCEK	5508	5509	5510	5511	5512	5513	5514	5515	5516	5517	5518	5519	5520	5521	5522	5523	5524
S246	EDDKKQKQGIDDEYDCEK	5525	EDDKKQKQGIDDEYDCEK	5526	5527	5528	5529	5530	5531	5532	5533	5534	5535	5536	5537	5538	5539	5540	5541	5542
T252	LLAKAKAKAKTIPAKAKAKA	5543	LLAKAKAKAKTIPAKAKAKA	5544	5545	5546	5547	5548	5549	5550	5551	5552	5553	5554	5555	5556	5557	5558	5559	5560
T274	TKKAKAKAKAKAKAKAKAKA	5561	TKKAKAKAKAKAKAKAKAKA	5562	5563	5564	5565	5566	5567	5568	5569	5570	5571	5572	5573	5574	5575	5576	5577	5578
S295	KETTAKAKAKAKAKAKAKAKA	5579	KETTAKAKAKAKAKAKAKAKA	5580	5581	5582	5583	5584	5585	5586	5587	5588	5589	5590	5591	5592	5593	5594	5595	5596
V316	EDDKKQKQGIDDEYDCEK	5597	EDDKKQKQGIDDEYDCEK	5598	5599	5600	5601	5602	5603	5604	5605	5606	5607	5608	5609	5610	5611	5612	5613	5614
S346	EDDKKQKQGIDDEYDCEK	5615	EDDKKQKQGIDDEYDCEK	5616	5617	5618	5619	5620	5621	5622	5623	5624	5625	5626	5627	5628	5629	5630	5631	5632
T346	EDDKKQKQGIDDEYDCEK	5633	EDDKKQKQGIDDEYDCEK	5634	5635	5636	5637	5638	5639	5640	5641	5642	5643	5644	5645	5646	5647	5648	5649	5650
S361	DKEKAKAKAKAKAKAKAKA	5651	DKEKAKAKAKAKAKAKAKA	5652	5653	5654	5655	5656	5657	5658	5659	5660	5661	5662	5663	5664	5665	5666	5667	5668
S362	KKAKAKAKAKAKAKAKAKA	5669	KKAKAKAKAKAKAKAKAKA	5670	5671	5672	5673	5674	5675	5676	5677	5678	5679	5680	5681	5682	5683	5684	5685	5686
T364	EDDKKQKQGIDDEYDCEK	5687	EDDKKQKQGIDDEYDCEK	5688	5689	5690	5691	5692	5693	5694	5695	5696	5697	5698	5699	5700	5701	5702	5703	5704
S365	EDDKKQKQGIDDEYDCEK	5705	EDDKKQKQGIDDEYDCEK	5706	5707	5708	5709	5710	5711	5712	5713	5714	5715	5716	5717	5718	5719	5720	5721	5722
S373	TKKAKAKAKAKAKAKAKAKA	5723	TKKAKAKAKAKAKAKAKAKA	5724	5725	5726	5727	5728	5729	5730	5731	5732	5733	5734	5735	5736	5737	5738	5739	5740
T384	EDDKKQKQGIDDEYDCEK	5741	EDDKKQKQGIDDEYDCEK	5742	5743	5744	5745	5746	5747	5748	5749	5750	5751	5752	5753	5754	5755	5756	5757	5758
S385	EDDKKQKQGIDDEYDCEK	5759	EDDKKQKQGIDDEYDCEK	5760	5761	5762	5763	5764	5765	5766	5767	5768	5769	5770	5771	5772	5773	5774	5775	5776
T386	EDDKKQKQGIDDEYDCEK	5777	EDDKKQKQGIDDEYDCEK	5778	5779	5780	5781	5782	5783	5784	5785	5786	5787	5788	5789	5790	5791	5792	5793	5794
S387	EDDKKQKQGIDDEYDCEK	5795	EDDKKQKQGIDDEYDCEK	5796	5797	5798	5799	5800	5801	5802	5803	5804	5805	5806	5807	5808	5809	5810	5811	5812
S388	EDDKKQKQGIDDEYDCEK	5813	EDDKKQKQGIDDEYDCEK	5814	5815	5816	5817	5818	5819	5820	5821	5822	5823	5824	5825	5826	5827	5828	5829	5830
T389	EDDKKQKQGIDDEYDCEK	5831	EDDKKQKQGIDDEYDCEK	5832	5833	5834	5835	5836	5837	5838	5839	5840	5841	5842	5843	5844	5845	5846	5847	5848
S390	EDDKKQKQGIDDEYDCEK	5849	EDDKKQKQGIDDEYDCEK	5850	5851	5852	5853	5854	5855	5856	5857	5858	5859	5860	5861	5862	5863	5864	5865	5866
T391	EDDKKQKQGIDDEYDCEK	5867	EDDKKQKQGIDDEYDCEK	5868	5869	5870	5871	5872	5873	5874	5875	5876	5877	5878	5879	5880	5881	5882	5883	5884
S392	EDDKKQKQGIDDEYDCEK	5885	EDDKKQKQGIDDEYDCEK	5886	5887	5888	5889	5890	5891	5892	5893	5894	5895	5896	5897	5898	5899	5900	5901	5902
T393	EDDKKQKQGIDDEYDCEK	5903	EDDKKQKQGIDDEYDCEK	5904	5905	5906	5907	5908	5909	5910	5911	5912	5913	5914	5915	5916	5917	5918	5919	5920
S394	EDDKKQKQGIDDEYDCEK	5921	EDDKKQKQGIDDEYDCEK	5922	5923	5924	5925	5926	5927	5928	5929	5930	5931	5932	5933	5934	5935	5936	5937	5938
T395	EDDKKQKQGIDDEYDCEK	5939	EDDKKQKQGIDDEYDCEK	5940	5941	5942	5943	5944	5945	5946	5947	5948	5949	5950	5951	5952	5953	5954	5955	5956
S396	EDDKKQKQGIDDEYDCEK	5957	EDDKKQKQGIDDEYDCEK	5958	5959	5960	5961	5962	5963	5964	5965	5966	5967	5968	5969	5970	5971	5972	5973	5974
T397	EDDKKQKQGIDDEYDCEK	5975	EDDKKQKQGIDDEYDCEK	5976	5977	5978	5979	5980	5981	5982	5983	5984	5985	5986	5987	5988	5989	5990	5991	5992
S398	EDDKKQKQGIDDEYDCEK	5993	EDDKKQKQGIDDEYDCEK	5994	5995	5996	5997	5998	5999	5990	5991	5992	5993	5994	5995	5996	5997	5998	5999	5990
T399	EDDKKQKQGIDDEYDCEK	5991	EDDKKQKQGIDDEYDCEK	5992	5993	5994	5995	5996	5997	5998	5999	5990	5991	5992	5993	5994	5995	5996	5997	5998
S400	EDDKKQKQGIDDEYDCEK	5999	EDDKKQKQGIDDEYDCEK	5990	5991	5992	5993	5994	5995	5996	5997	5998	5999	5990	5991	5992	5993	5994	5995	5996
T401	EDDKKQKQGIDDEYDCEK	5991	EDDKKQKQGIDDEYDCEK	5992	5993	5994	5995	5996	5997	5998	5999	5990	5991	5992	5993	5994	5995	5996	5997	5998
S402	EDDKKQKQGIDDEYDCEK	5993	EDDKKQKQGIDDEYDCEK	5994	5995	5996	5997	5998	5999	5990	5991	5992	5993	5994	5995	5996	5997	5998	5999	5990
T403	EDDKKQKQGIDDEYDCEK	5995	EDDKKQKQGIDDEYDCEK	5996	5997	5998	5999	5990	5991	5992	5993	5994	5995	5996	5997	5998	5999	5990	5991	5992
S404	EDDKKQKQGIDDEYDCEK	5997	EDDKKQKQGIDDEYDCEK	5998	5999	5990	5991	5992	5993	5994	5995	5996	5997	5998	5999	5990	5991	5992	5993	5994
T405	EDDKKQKQGIDDEYDCEK	5999	EDDKKQKQGIDDEYDCEK	5990	5991	5992	5993	5994	5995	5996	5997	5998	5999	5990	5991	5992	5993	5994	5995	5996
S4																				

Modular regulatory proteins involved in endocytosis

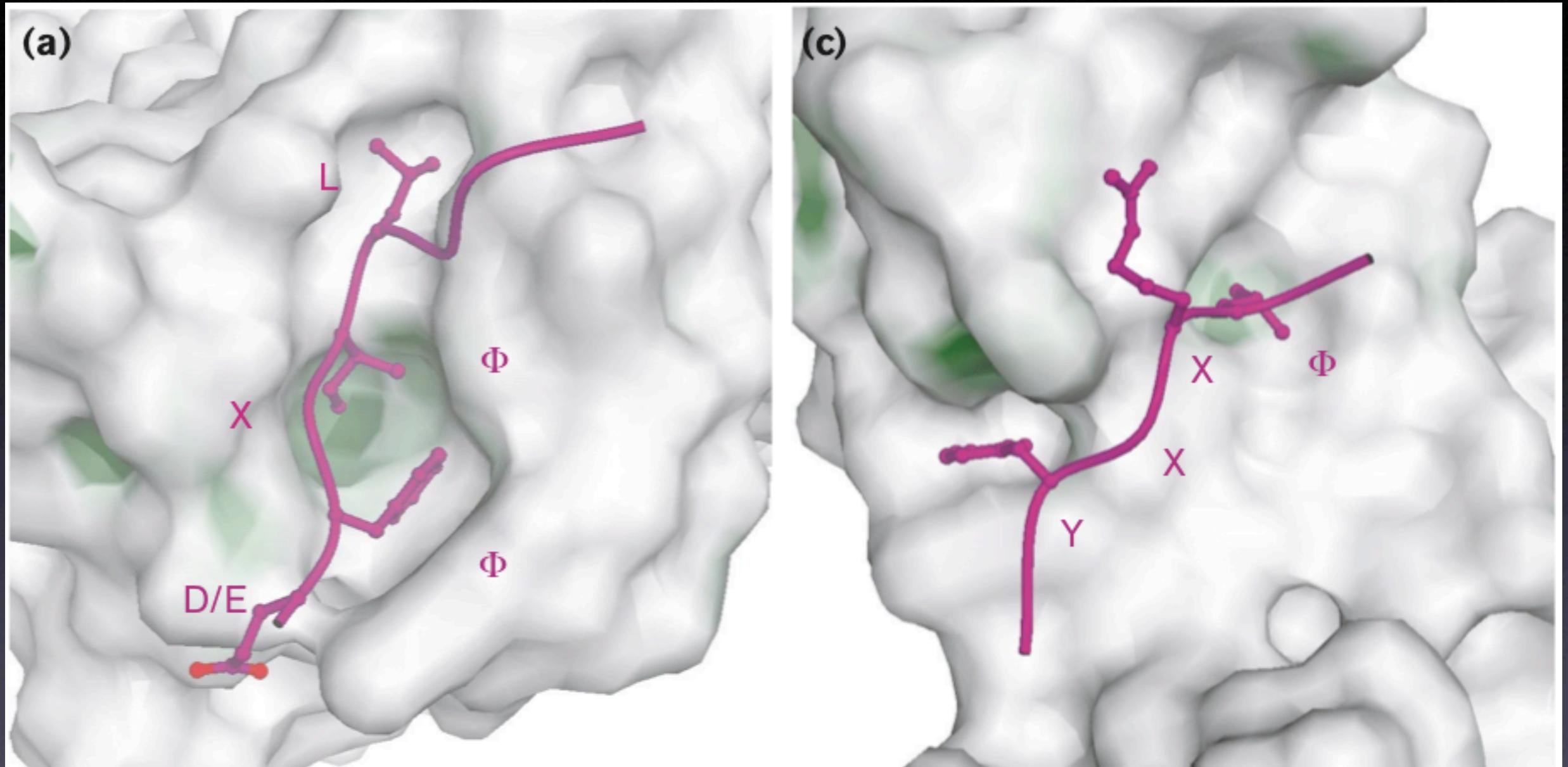
Most Endocytosis proteins have a mixture of **globular domains** and **natively disordered** regions. The disordered regions are proving to be rich in **Linear Motifs**.

Here the disordered regions are shown to scale with respect to the globular domains

Molecular Bolas - Not good for diffusion?



Two Structures of Linear Motif Complexes involved in Endocytosis: Clathrin box and YxxΦ LMs

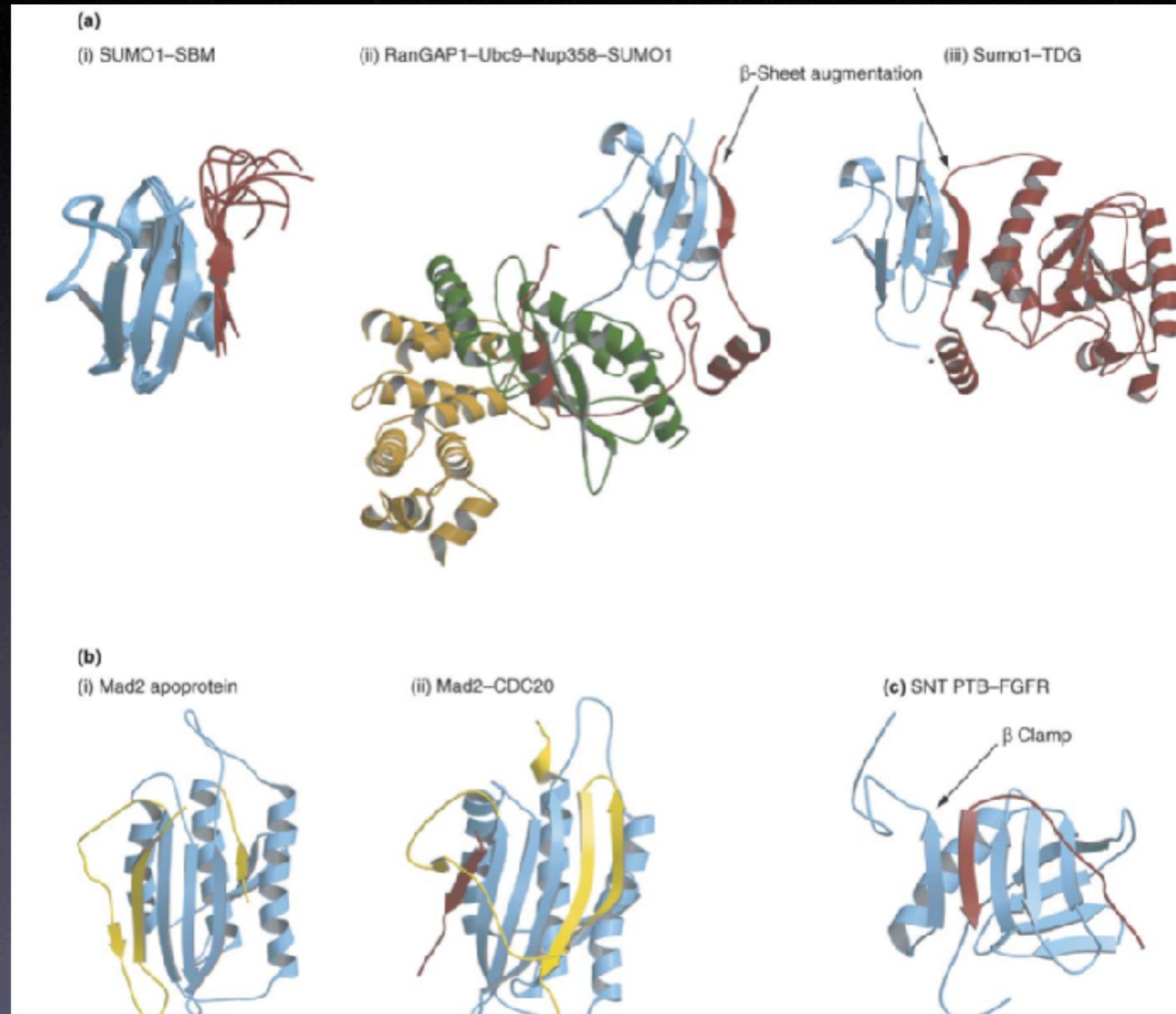


Clathrin box LΦ.Φ[D/E]
bound to Clathrin

Y..Φ bound to $\mu 2$
subunit of AP2

Up to ~30% of Linear Motifs bind by
“ β -sheet Augmentation”
First seen for PDZ and PTB binding motifs

β -sheet
Augmentation



Protein Architecture Modules in Cell Regulation

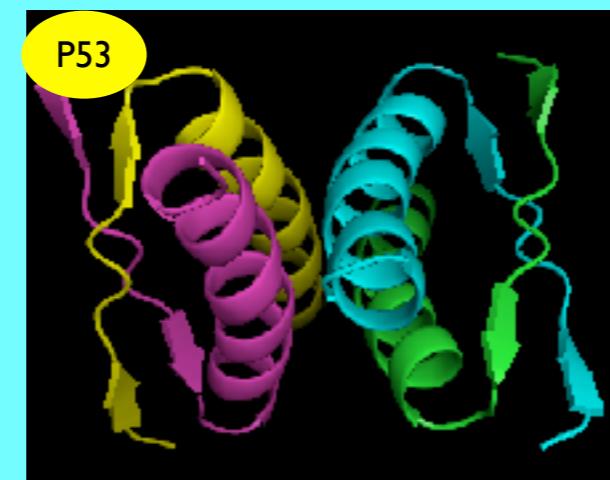
Globular ~70%



Kinase
Phosphatase
Acetylase
Deacetylase
SH3
SH2
PH
PDZ
Bromo ...

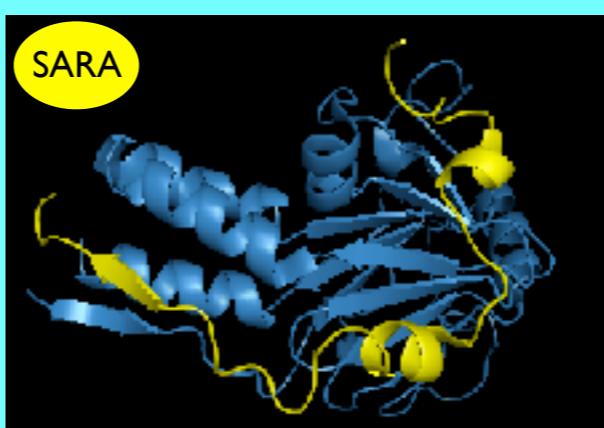
Natively Disordered ~30%

Mutual Fit



Coiled Coil
Collagen Helix
P53 tetramerisation
T4 Endonuclease VII
HLH DBD
ACTR/NCBD ...

Induced Fit



SARA > Smad2
Tcf > beta-Catenin
Hif1-alpha > CBP-TAZ
Cited2 > CBP-TAZ
P27kip1 > CDK
ERM C-tail ...

Linear Motif



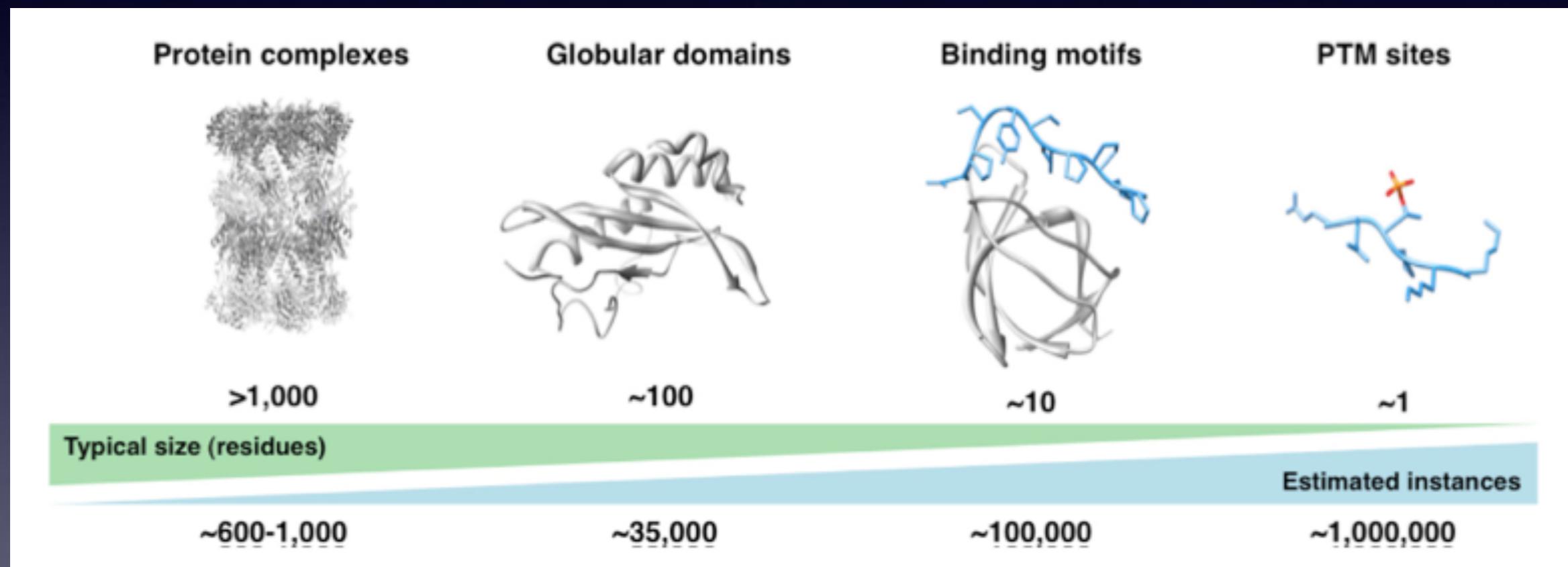
NLS / NES / PTSI /
KDEL / YPWM / EHI /
WRPW / LXXLL /
NPF / DPW / RGD ...

Effectors of regulation

Passive components involved in building regulated, often highly dynamic, complexes

A Million Peptide Motifs for the Molecular Biologist

Peter Tompa,^{1,2,*} Norman E. Davey,³ Toby J. Gibson,⁴ and M. Madan Babu^{5,*}



We focus a lot on
short linear motifs because we
think they provide the most
interactions in the cell.

But the course covers much
more