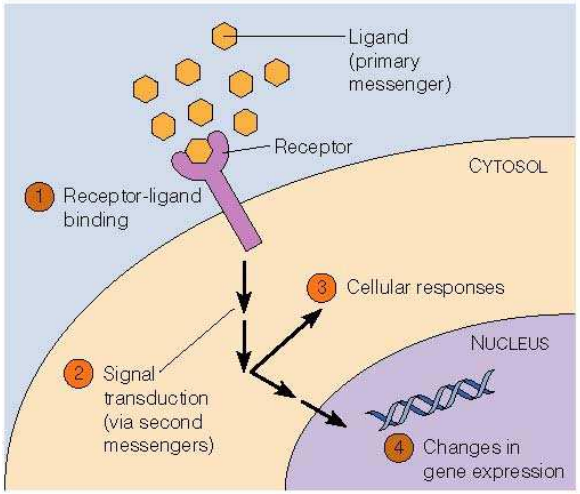
Macromolecular interactions

Biologically important macromolecular interactions are highly specific

Protein-protein interaction is essential for propagation of information from the cell surface to the nucleus and generation of appropriate response

- activation of the cell surface receptor
- regulation of the activity of signaling molecules
- nuclear localization
- expression of appropriate gene products

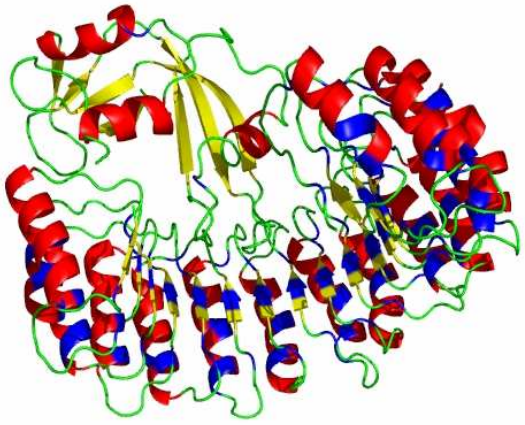


Aberrant macromolecular interaction often results in diseases—many signaling molecules are **oncogenes** (i.e. will cause cancer when mutated)

The binding affinity can be transient or extremely high

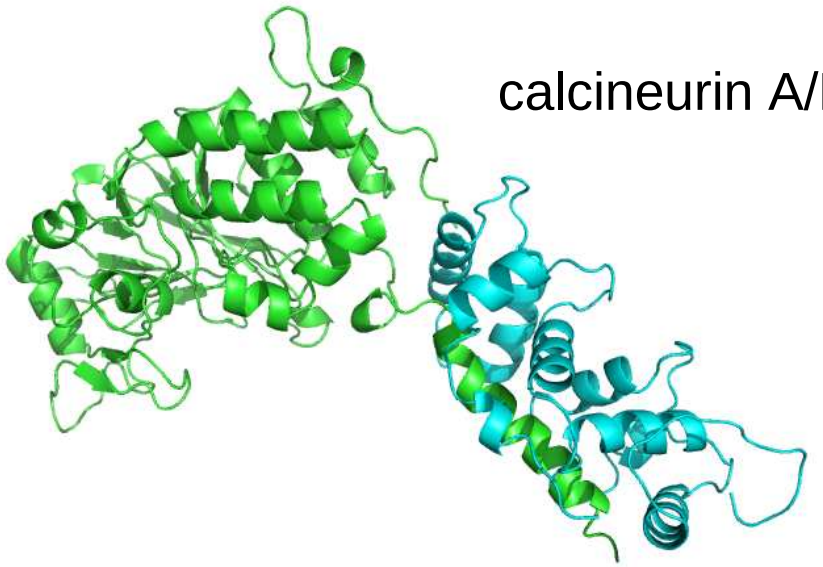
e.g. oligomerization

RNAase-inhibitor complex



RNAse-RI

(Kd ~ 4 fM)



calcineurin A/B

Interaction between signaling molecules is a potential site of regulation

Need to locate and characterize the interaction surface so that appropriate drugs can be designed

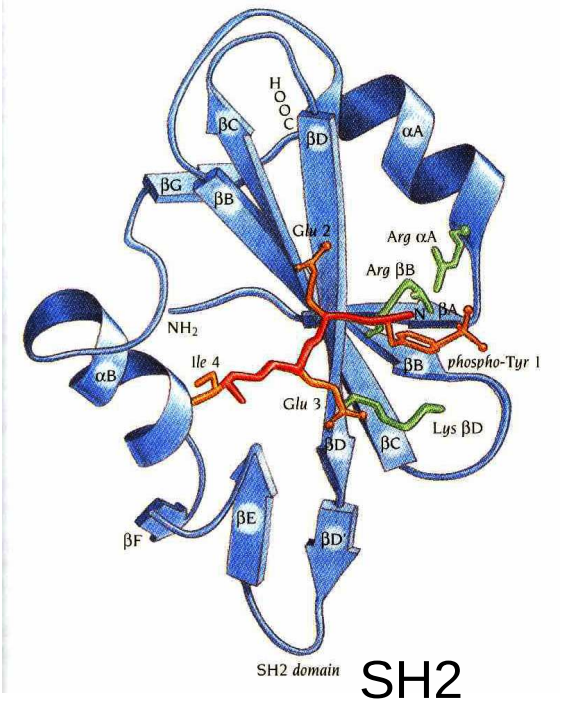
Mechanism of interaction

There are recurring structural protein modules that recognize specific sequence motifs: evidence of “co-evolution”

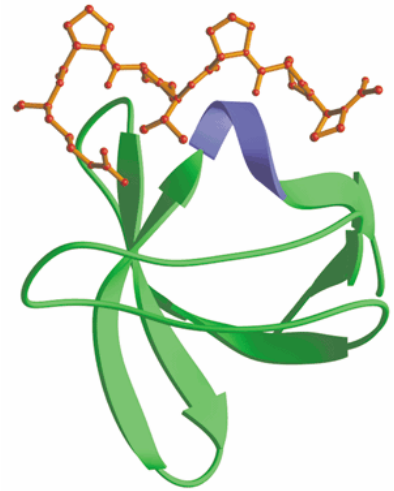
Src homology domain 2 (**SH2**) – phospho-tyrosine containing peptide

SH3, WW domain – polyproline containing peptide

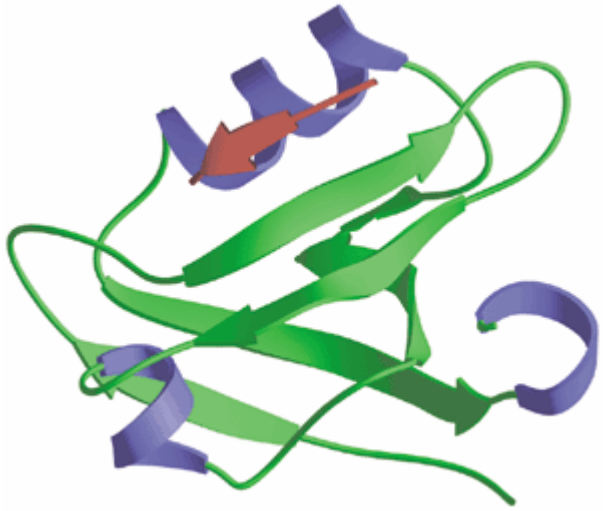
PDZ domain – Ser/Thr-Xxx-Val-COO-



SH2



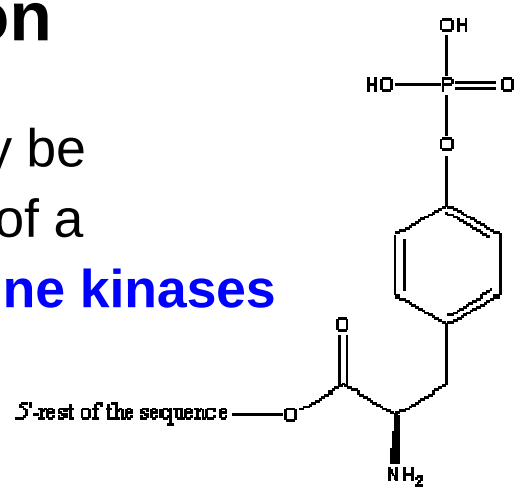
SH3



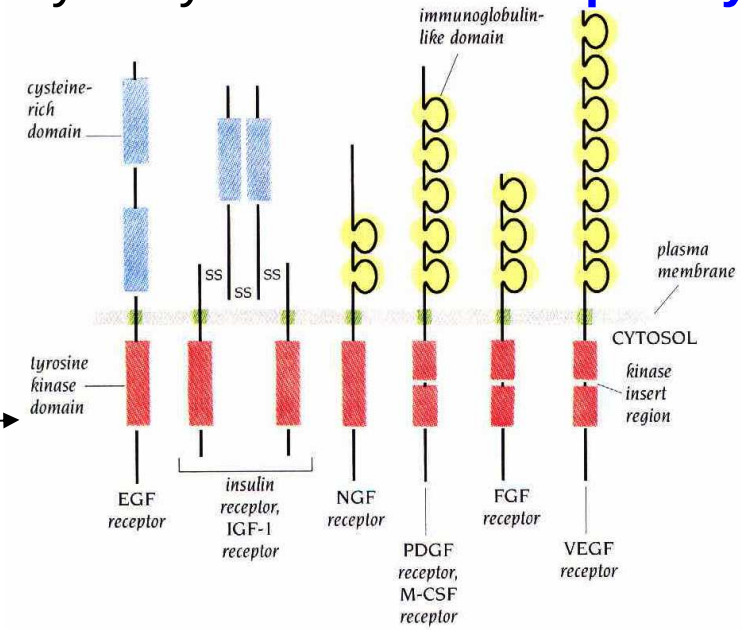
PDZ

Tyrosine phosphorylation

Some tyrosine residues in cell surface receptors may be dynamically modified ("phosphorylated") by addition of a phosphate group by enzymes called **receptor tyrosine kinases**



RTK →



outside the cell

inside the cell

Indicates a change in the signaling status of the molecule ("**on-off**" switch)

Removal of the phosphate group is performed by enzymes called **phosphatases**

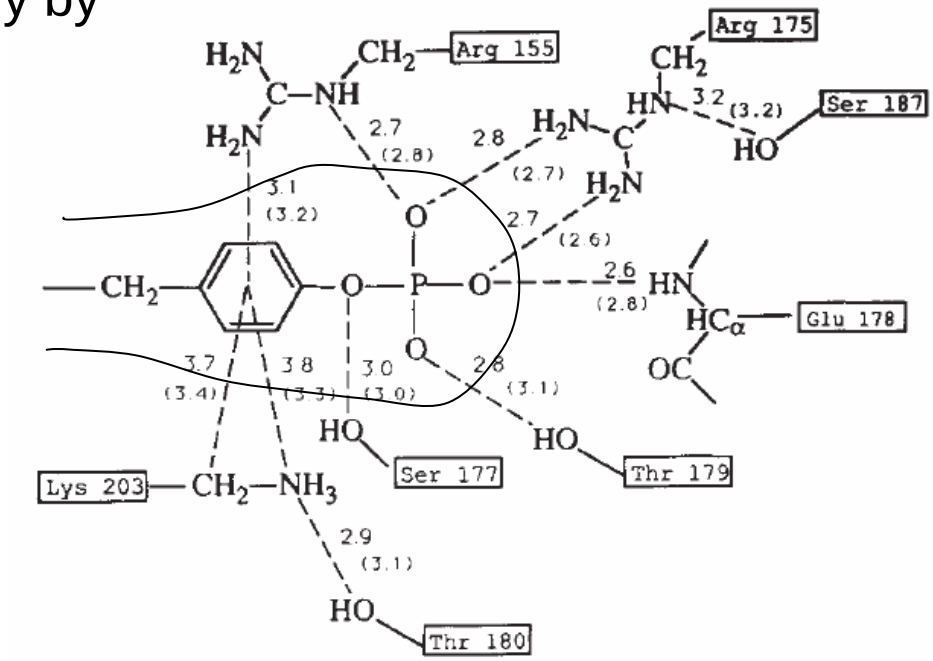
pTyr is essential for SH2 binding

Roughly 110 different human proteins contain one or more SH2 domains sharing a high level of sequence and structural homology

pTyr binding pocket contains positively charged residues (R155, R175, K20) and polar residues (S177, T179, E178), which interact with both the phosphate group and the phosphotyrosine ring

S177, T180, S187 contribute indirectly by stabilizing neighboring residues

cf. amino-aromatic interactions are also seen in hemoglobin-drug complexes



Waksman et al Nature 358, 646 (1992)

Identification of consensus motif for SH2

Different SH2 domains recognize different sequence around pTyr

Low affinity of binding (~ 1 mM) prevents identification of the determinants of sequence specificity

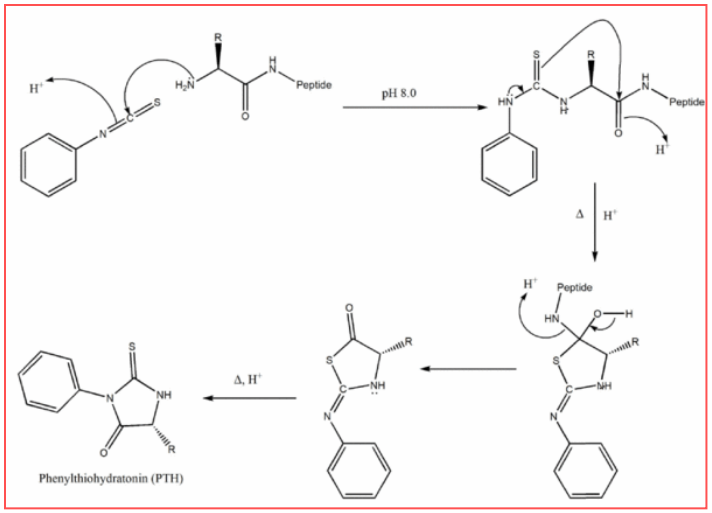
In order to improve the binding affinity, construct a randomized pTyr peptide library and screen against immobilized SH2

Composition of the library : H2N-GDG-pY-**XXX**-SPLLL-COOH

where X is any of 20 amino acids

Songyang et al, Cell 72, 767 (1993)

The peptides selected again p85 were sequenced by **Edman** degradation to identify the consensus sequence—each cycle identifies the N-terminal residue



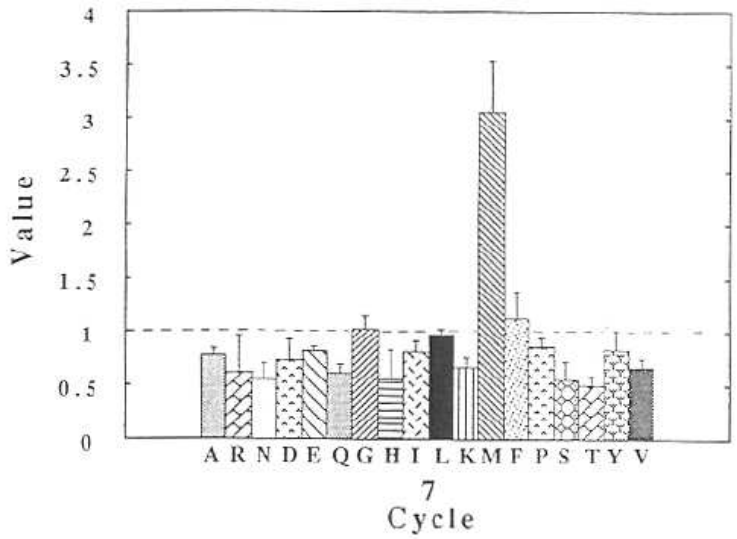


Table 3. Recognition Specificities of SH2 Domains

SH2 Domain	pY+1	pY+2	pY+3	Selectivity
Src	E (2.5) D (1.7) T (1.7)	E (2.6) N (2.4) Y (2.0)	I (3.6) M (2.5) L (2.3)	33
Fyn	**E** (3.2) T (2.0)	**E** (3.7) D (1.7) Q (1.6)	**I** (4.2) V (2.5) M (2.0)	83
Lck	**E** (3.5) T (1.7) Q (1.6)	E (2.5) D (1.5)	**I** (3.4) V (2.2) M (2.1)	45
Fgr	**E** (4.8) Y (1.6) D (1.6)	**E** (3.1) N (1.7) D (1.7)	I (2.1) V (1.7)	49
Abl	E (2.8) **T** (2.4) M (2.1)	**N** (3.5) E (2.2) D (1.8)	**P** (3.0) V (2.2) L (2.2)	44
Crk	D (2.6) K (2.3) N (1.6)	H (2.9) F (1.9) R (1.7)	**P** (7.3) L (1.7)	109
Nck	**D** (5.8)	**E** (3.6)	**P** (3.0) D (2.8) V (2.7)	117

Binding surfaces

Protein-protein interfaces tend to be large (600 – 1300 $Å^2$), involve 10 – 40 residues from each protein that interdigitate into multiple residues from the other protein, and are rich in aromatic residues (His, Phe, Tyr, Trp) and Arg

Protein interactions are determined by hydrophobic effects and shape/charge complementarity, as well as hydrogen bonding interactions

Large number of weak interactions or small number of strong interactions?

Predicting binding sites on a protein target is important for drug design

No general principles have been found

enzyme/inhibitor v. antibody/antigen

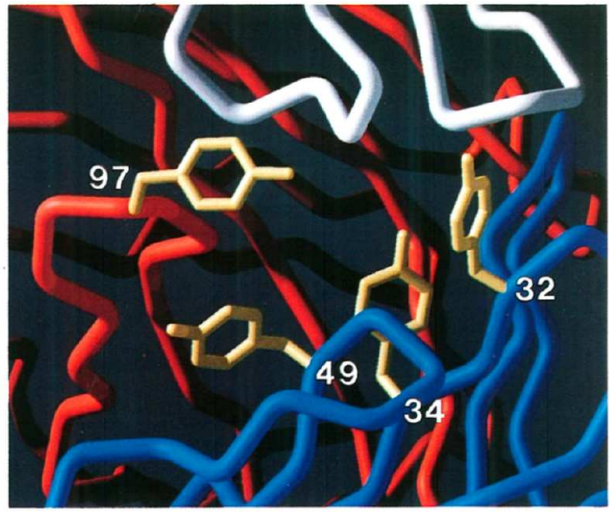
oligomers (most hydrophobic interface)

Keskin et al, JMB 345, 1281 (2005)

Types of Complex

Antibody-antigen

- variable loops of antibody recognize a wide spectrum of compounds by varying the shape and chemical properties of side chain
- polar and hydrophobic side chain-side chain or side chain–main chain interactions dominate
- over 25% of interaction energy comes from **Tyr** of antibody
- from antibody : Y > D > N > S > E > W
- from antigen : R = K > N > D > Y > S = T = G > E



antibody + lysozyme (white)

Enzyme-inhibitor

- exposed binding loop of inhibitor docks in the active site of the enzyme as a normal substrate would
- predominantly main chain–main chain
- from inhibitor : R > K = L = C > P > V > I > M
- from serine protease : S > G > D > H

Jackson, Protein Sci 8, 603 (1999)

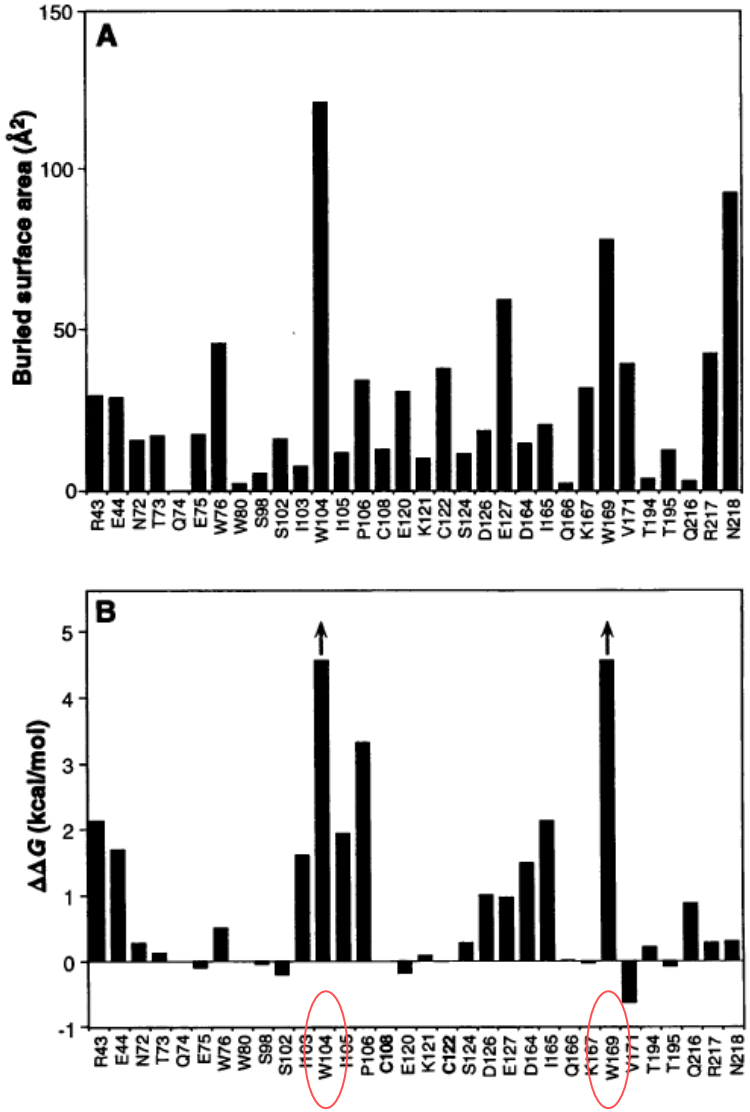
Existence of "Hot spot"

Not clear if interfacial residues are equally important—i.e. do some residues at the interface contribute more to binding than others?

Human growth hormone (hGH) and hGH receptor interact through 33 amino acids, burying 1300 Å^2 in the process, some of which are more buried than others

Ala scanning of the contact residues on hGHR shows that many of them do not contribute much (if at all) to binding affinity

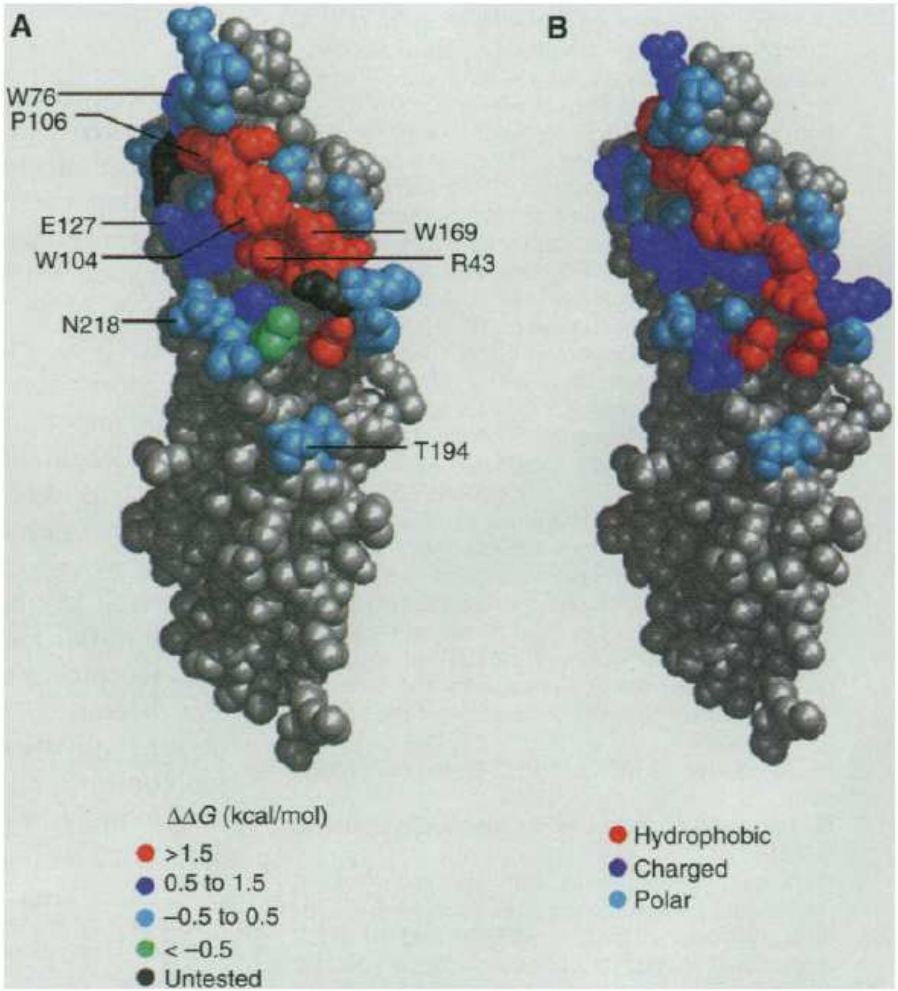
Nonfunctional region corresponds to 46% of the buried area



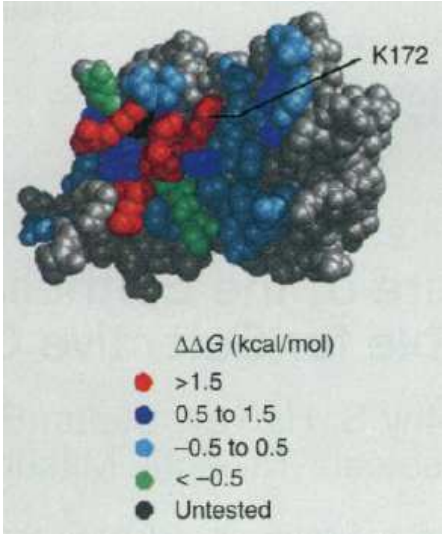
Clackson and Wells, Science 267, 383 (1995)

Functionally important residues are surrounded by nonfunctional residues

Critical residues tend to be hydrophobic



Mutations on hGH shows 8 out of 31 residues account for ~ 85% of binding energy



Most of the important residues form a hydrophobic pocket with W104 and W169

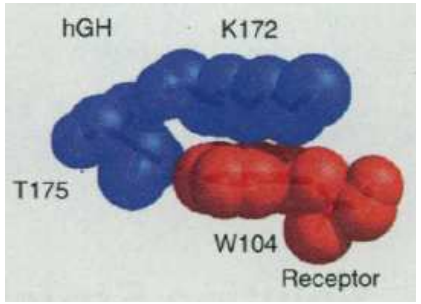
The few residues that contribute the bulk of the free energy of binding constitute the "**hot spots**" of interaction

Why are some contacts more important than others?

- not buried surface area
- not the number of van der Waals contacts
- not crystallographic temperature factors (B factors)
- not solvation parameters

But

- packing is better at the critical regions

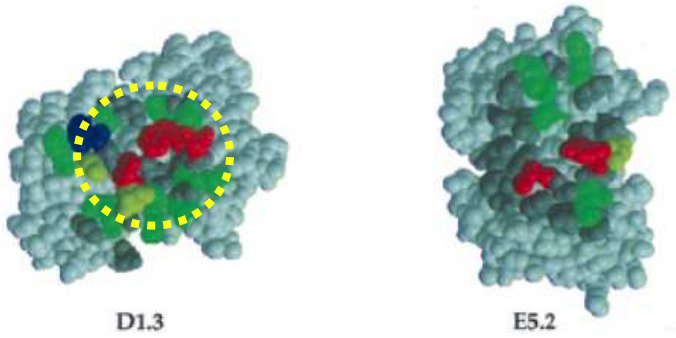


O-ring model

2325 alanine mutants with thermodynamic data show that protein-protein interfaces usually contain “hot spots” that contribute the bulk of binding energy and they are surrounded by amino acids whose primary role is to exclude solvent from the hot spot

--Bogan and Thorn, JMB 280, 1 (1998)

“O-ring” is critical because protein-protein interfaces are often flat and desolvation of the interacting surface is essential for high affinity binding



antibody-antigen

enzyme-inhibitor

Table 2. Amino acid preferences in hot spots

Residue	In database (Number)	In database (%)	Contribute ⩾ 2 kcal/mol (Number)	Contribute ⩾ 2 kcal/mol (%)	Enrichment in hot spots
Arg	218	9.38	29	13.30	2.47
Asn	99	4.26	5	5.05	0.93
Asp	177	7.61	16	9.04	1.67
Cys	3	0.13	0	0	0
Gln	160	6.88	5	3.13	0.58
Glu	220	9.46	8	3.64	0.68
Gly	28	1.20	1	3.57	0.45
His	50	2.15	4	8.00	1.49
Ile	104	4.47	10	9.62	1.79
Leu	242	10.41	2	0.83	0.01
Lys	143	6.15	9	6.29	1.17
Met	69	2.97	2	2.90	0.54
Phe	166	7.14	5	3.01	0.56
Pro	89	3.83	6	6.74	1.25
Ser	178	7.66	2	1.12	0.21
Thr	131	5.63	2	1.53	0.28
Trp	19	0.82	4	21.05	3.91
Tyr	122	5.25	15	12.30	2.29
Val	107	4.60	0	0	0

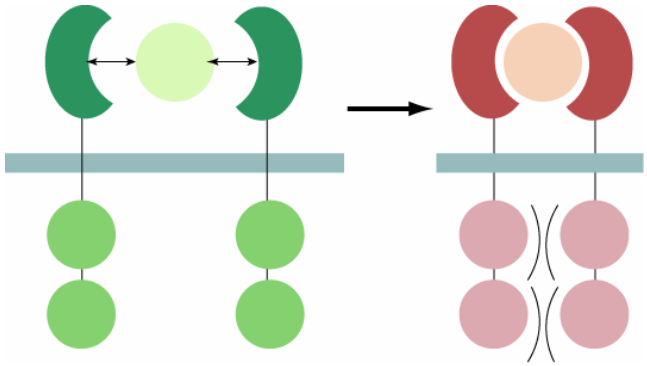
The number and percentage of amino acids in the database of 2325 alanine mutations studied are shown. The composition of the database is in rough agreement with the distribution of amino acids on protein surfaces (Janin *et al.*, 1988). The number and percentage of each residue type with $\Delta\Delta G \geqslant 2$ kcal/mol is also given. Percentage of amino acid type with $\Delta\Delta G \geqslant 2$ kcal/mol is calculated by dividing the number of a given residue type with $\Delta\Delta G \geqslant 2$ kcal/mol by the number of that type of residue in the database. Enrichment in hot spots gives the fold enrichment of that residue type in hot spots ($\Delta\Delta G \geqslant 2$ kcal/mol) over the database as a whole (e.g. a value of 2 indicates that the residue is twice as frequent in hot spots as in the database as a whole). The distribution of these energetically important residues is very non-random, with certain amino acid types (Trp, Tyr and Arg) much more likely to be in hot spots than others (Val, Leu and Ser).

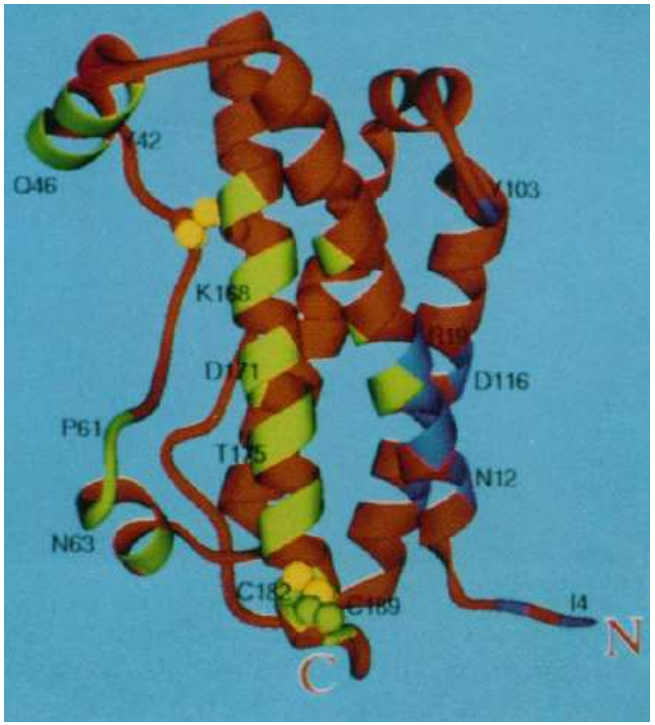
Bogan and Thorn, JMB 280, 1 (1998)

Plasticity of the interface

Protein may bind multiple unrelated targets using a common binding surface

Human growth hormone (hGH) is a four helix bundle that binds to two molecules of hGH receptor and results in dimerization of hGHR required to initiate signals essential to muscle, bone growth





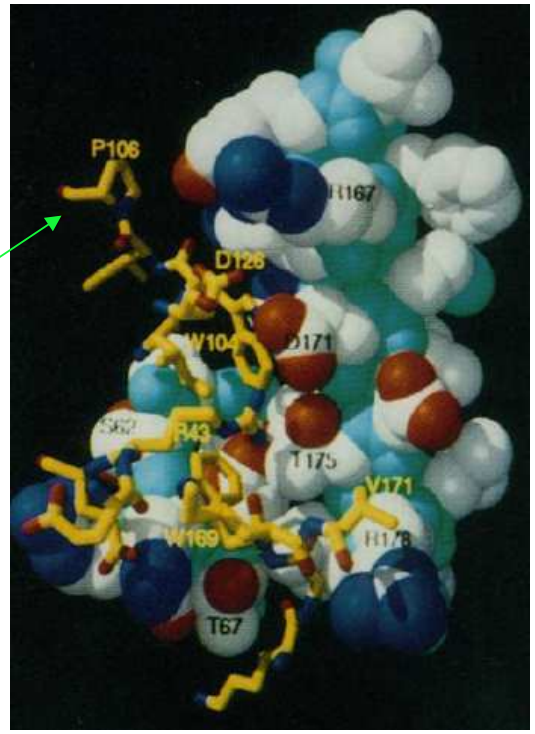
Same binding surface of hGHR is used to interact with two unrelated surfaces of hGH

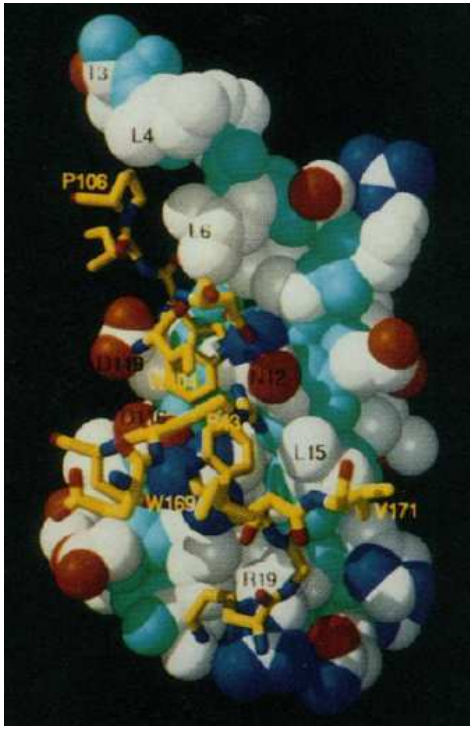
Residues at the interface are arranged to accommodate structural differences

de Vos et al, Science 255 (1992)

hGH

hGHR





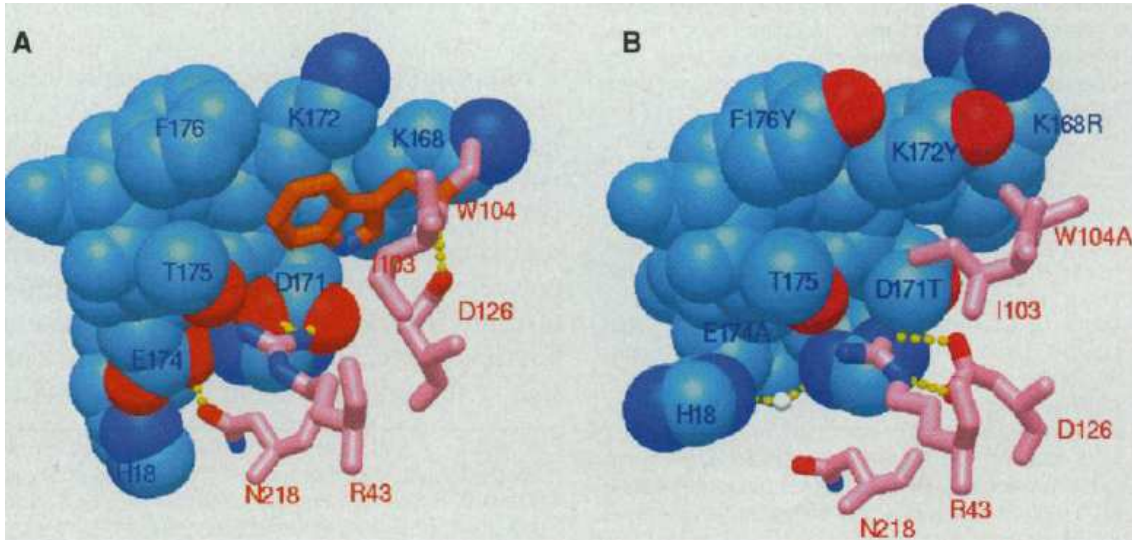
binding site I

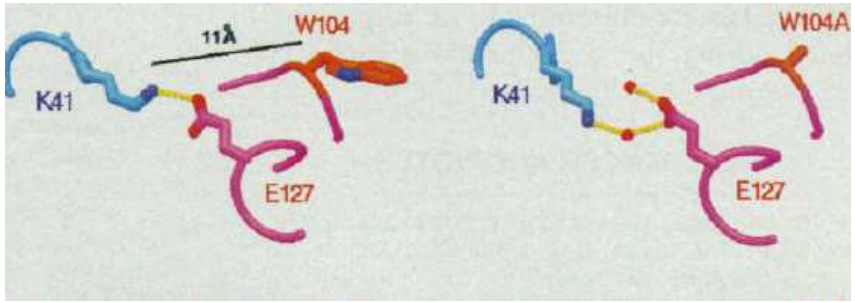
binding site II

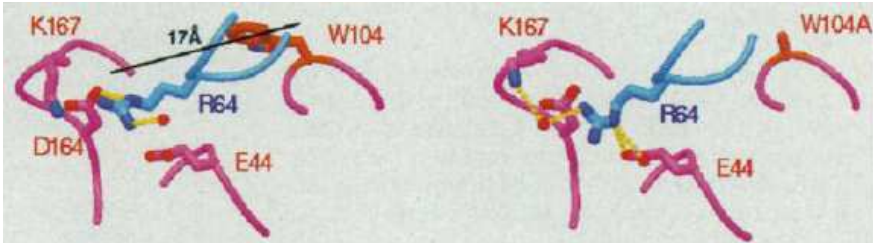
W104A receptor mutant has a ~2500 lower affinity for hGH compared to wt

Complementary mutations can be introduced in hGH to restore binding by screening a random library of ~10^7 diversity

Residues that are up to 17 Å away from W104 make structural adjustments







Atwell et al, Science 278, 1125 (1997)