Lecture Notes for 20.320 Fall 2012

Interaction Specificity

Ernest Fraenkel

Introduction

In this section, we will examine what determines whether two macromolecules interact. We will begin with an experimental method known as alanine scanning that has been used to determine the energetics of protein-protein interfaces.

PROTEIN-PROTEIN INTERFACES

In the coming parts of the course, we will exam the structural and energetic properties of the interface macromolecules. In order to discover how to modify these interfaces we must determine the energetic contribution of each residue. One way of discovering this is through "alanine scanning" experiments. A series of mutant proteins are constructed in which a single residue is mutated to alanine. A binding assay is conducted to determine the energetic effect of the mutation.

> Why is alanine chosen?
>
> What controls are critical for interpreting the results of these experiments?

These experiments have revealed that many interfaces contain a small number of mutational "hot spots." Altering these residues has a large effect on the energetic of the interface, while most other residues have little effect.

The figure below shows one of the first examples of a hot spot to be identified. As you can see in panel "B", very few of the alanine substitutions of contact residues have a significant effect on the free energy of binding of human growth hormone (hGH) to the hGH binding protein. For example, substitution of W104 or W169 causes more than a 4.5 kcal/mole change in binding free energy. The magnitude of the free energy change does not correlate with the loss of surface area that occurs on binding (panel A).

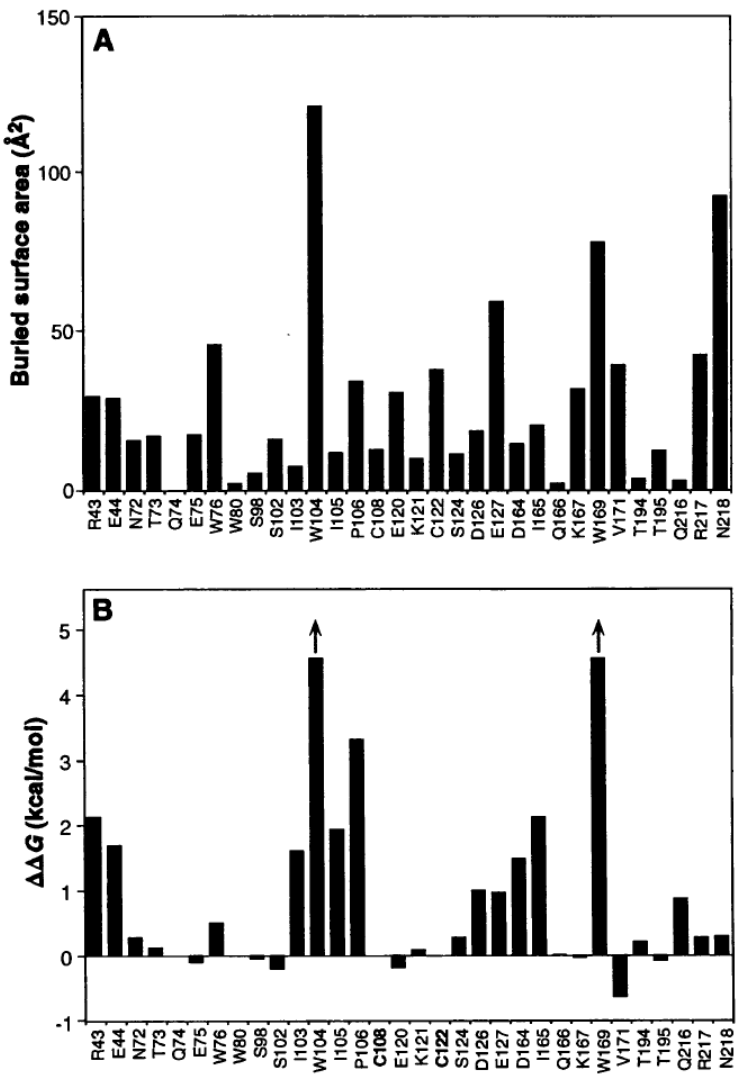


Fig. 1. Contribution of only a subset of contact residues to net binding energy. (**A**) Loss of solvent-accessible area (*7*) of the side chain portion of each residue in the hGHbp on forming a complex with hGH. (**B**) Difference in binding free energy between alanine-substituted and wild-type hGHbp $(\Delta\Delta G)_{mut\text{-}wt}$ at contact residues (*5*). Negative values indicate that affinity increased when the side chain was substituted by alanine.

Figure from Clackson & Wells (1995).

Source: Clackson, Tim, and James A. Wells. "A Hot Spot of Binding Energy in a Hormone-Receptor Interface." *Science* 267, no. 5196 (1995): 383-6.

The next figure shows that the locations of the most energetically important residues are clustered on both hGH and hGHbp and form complementary surfaces. Similar observations have now been made for many proteins (see Moreira, et al. (2007)) and some general features have emerged. Fewer than 10% of the residues at an interface contribute more than 2 kcal/mol to binding. These hot spots tend to be rich in Trp, Arg and Tyr and occur on pockets on the two proteins that have complementary shapes and distributions of charged and hydrophobic residues. The hot spots can include buried charge residues, and these tend to occur in the center of the pocket, far from solvent. In fact, most of the hot spots tend to be surrounded by residues that keep bulk solvent out of the pocket – these have been compared to O-rings.

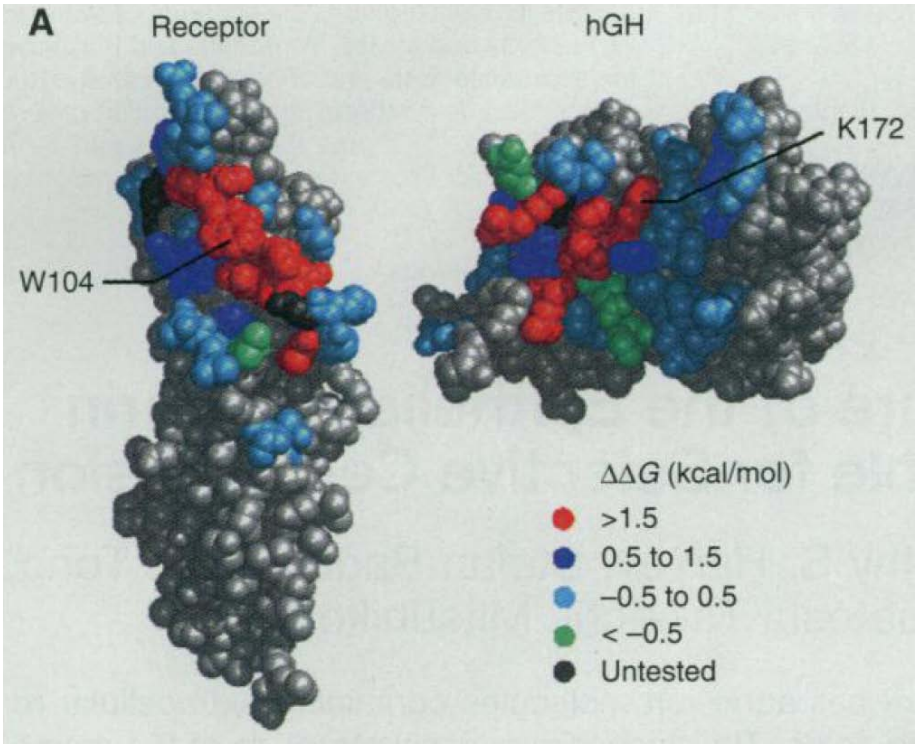


Figure from Clackson & Wells (1995).

Some questions for you to consider:

- Can you think of a sequence-based approach for identifying potential hot spot residues?
- What are the consequences of the existence of hot spots for altering specificity?
- What are the consequences for designing small molecules to prevent two proteins from binding?

COILED-COIL INTERFACES.

Perhaps the simplest and most regular protein-protein interface is the coiled coil. It consists of two alpha helices, one from each protein. Two straight helices can only interact over a very small patch, which would not provide enough free energy for a stable complex. However, in the coiled coil, the helices twist around each other, creating an extensive interface. The structure repeats every seven amino acids, so the standard notation for the positions on one helix is (abcdefg) and on the second helix (a'b'c'd'e'f'g'). Positions **a** and **d** are often hydrophobic, and often leucines. In fact, the family is often called the leucine zipper because of the repeating leucines at the **d** positions. Positions **e** and **g** tend to

be polar or charged. Helices can come to together as pairs or in larger numbers, and both parallel and anti-parallel arrangements possible.

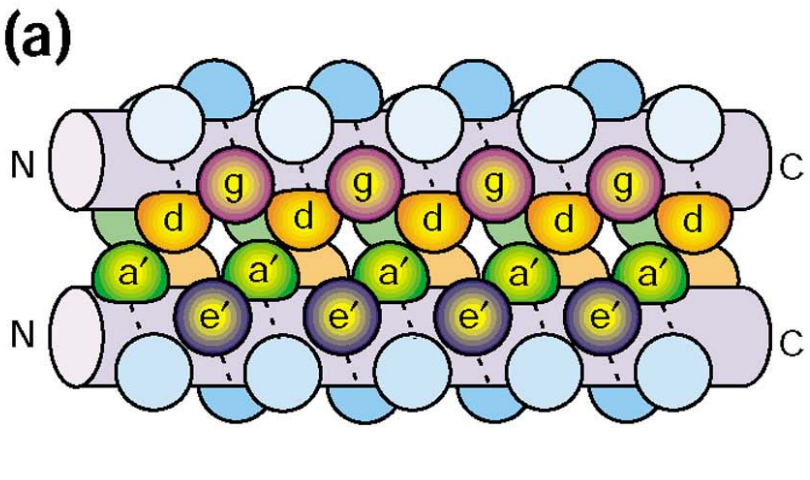


Figure from Fong et al. (2004).. "Cartoon of a parallel two-stranded coiled coil. **(a)** Side view and **(b)** top view. The interface between the α-helices in a coiled-coil structure is formed by residues at the core positions **a**, **d**, **e** and **g**. Positions in the two helices are distinguished by the prime notation; for example, **a** and **a'** are analogous positions in the two helices. N, amino terminus; C, carboxy terminus."

This simple geometry presents interesting questions. Do all coiled-coil proteins interact with each other with equal affinity? If not, what determines the specificity? What determines whether helices form dimers, trimers or tetramers. The question of how much specificity exists in the coiled coil family was addressed in a very direct way by Professor Keating and colleagues. They looked at coiled coil proteins that bind to DNA, known as bZIP proteins. Using protein arrays they examined the specificity of almost all the human bZIP proteins. As you can see in the figure below, the interactions tend to be highly specific, with most proteins interacting with relatively few partners.

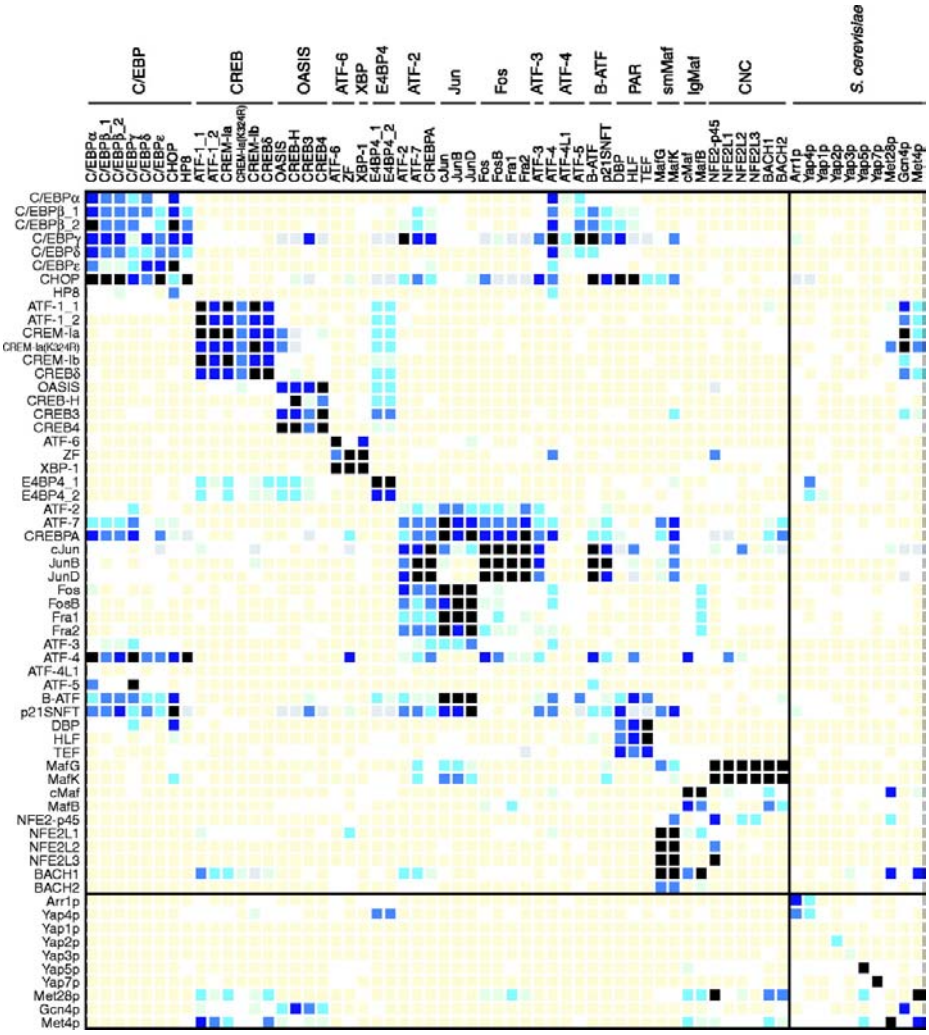


Figure from Newman and Keating (2003). The following is the original figure legend: "Consensus interaction matrix for 49 human and 10 yeast bZIP peptides. Fluorescent proteins (probes) are listed at the top, and surface-phase proteins are listed at the left. Control peptides spotted in duplicate are indicated name_1, name_2. Peptide pairs were assigned a *Z*-score (*17*) that corresponds to the highest value for which the probability of seeing the observed number of occurrences by chance is less than 10^{-4}. $Z > 20$ (black); $Z > 10$ (dark blue); $Z > 5$ (medium blue); $Z > 2.5$ (light blue); $Z > 1.5$ (light green); at least 75% observations with $Z < 1.0$ (yellow); no assignment with confidence meeting *P*-value test (white); signal observed was not reciprocal (i.e., for the heterodimer XY, $Z_{XY} > 2.5$, $Z_{YX} < 1$) (gray). Comparisons of relative intensities are most accurate within columns, which share the same probe protein. The sparseness of strong human/yeast bZIP interactions (0.5% versus ~5% for human/human or yeast/yeast) further points to the specificity of this motif."

Some clear answers have emerged for parallel two-helix coiled coils. Mutational analysis suggests that the **a** and **d** positions provide significant stabilization energy. Electrostatic interactions between **a** and **a'** as well as between **g** and **e'** (and, by symmetry, g' e) can be particularly important in determining specificity. Machine learning techniques have been very helpful in developing algorithms that can

predict favorable and unfavorable interactions from sequence data. However, the accuracy of these methods is still limited and probably biased by the limited experimental data. In addition, it is still hard to predict the preferred topology for a set of peptides. Even this simple domain has proven very hard to understand.

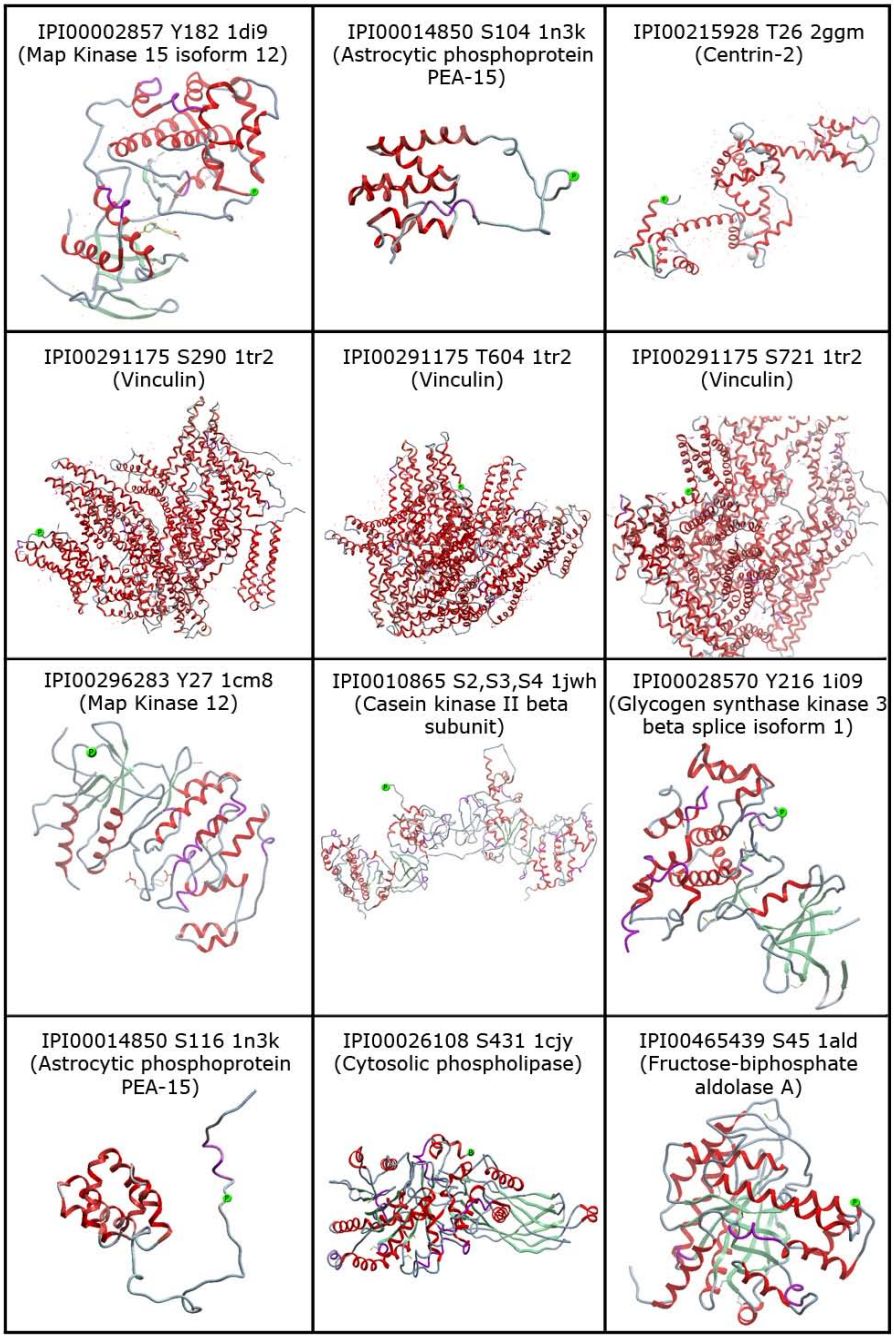
Kinase Specificity

In previous parts of the course, we looked in some detail at signaling networks, which are composed in large part of kinases and phosphatases. Almost a third of all proteins in a eukaryotic cell are phosphorylated, and many of these proteins are modified on several sites, each of which can have distinct consequences for their function. Kinases tend to be highly specific, and this specificity is crucial to the proper functioning of these networks. Phosphorylating the wrong protein or the right protein on the wrong site will lead to off-pathway effects.

Given the high specificity exhibited by the kinases it may be surprising that most kinases, including both Ser/Thr-specific and Tyr-specific kinases, share a common structure. The structures are composed of two lobes; the N-terminal one is largely composed of beta sheets, while the C-terminal lobe is largely composed of alpha helices. The ligands, ATP and the target peptide bind at the cleft between these two lobes.

Note that the peptide binds in an extended conformation. An alpha helix will not fit into this pocket. In fact, most sites of phosphorylation tend to occur in loops than in regions of regular secondary structure, as seen in the figure below from Gnad *et al.*(2007). However, as you can see from the figure, even the loops are not in an extended conformation that could fit into an active site.

How can that be?



Sites of phosphorylation for proteins in the PHOSIDA database that have known structures.
From Gnad, et al. Genome Biology 2007, 8:R250 doi:10.1186/gb-2007-8-11-r250

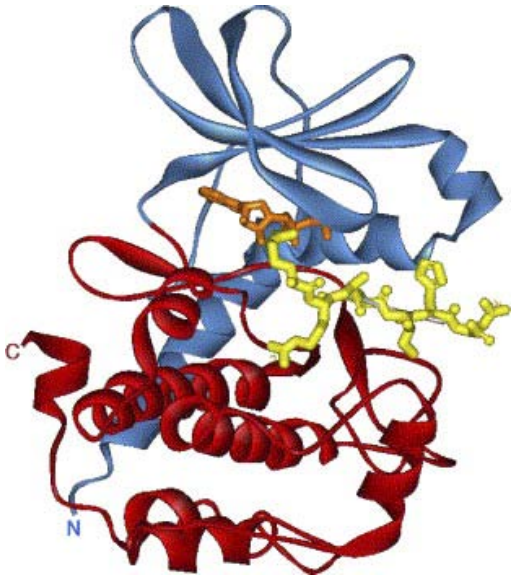


Figure from Kobe et al. (2005). Original legend: "Schematic representation of the structure of the catalytic subunit of protein kinase A (Protein Data Bank (PDB [93]) code 1JBP [94]). The small lobe is coloured light blue, the large lobe is coloured red, the peptide substrate is coloured yellow, and the ATP molecule is coloured orange. The figure was generated using the program ViewerLite (Accelrys Inc., San Diego, CA)."

We will see that specificity is imparted as many levels, the first of which is complementarity between the shape and charge of the ligand and a "binding pocket." The proteins that phosphorylate serine and threonine tend to have a shallower catalytic cleft than those phosphorylating tyrosine. This difference in the binding pocket excludes tyrosine from the active site. The residues surrounding the active site play an important role in distinguishing among the hundreds of thousands of peptides that contain serine or threonine. The target peptide lies in an extended conformation across the cleft between the two lobes, with each amino acid capable of interacting with a distinct pocket on the kinase that can, in principle, contribute to specificity. The nomenclature for describing these interactions is derived from early work on peptidases. The site of phosphorylation (or cleavage) on the target is labeled P0, with residues N-terminal to P0 designated P-1, P-2, etc. and those C-terminal called P+1, P+2. If the size, shape and electrostatic properties of a peptide are a good match for the recognition sites, then the peptide will be more readily phosphorylated than one that has less complementarity.

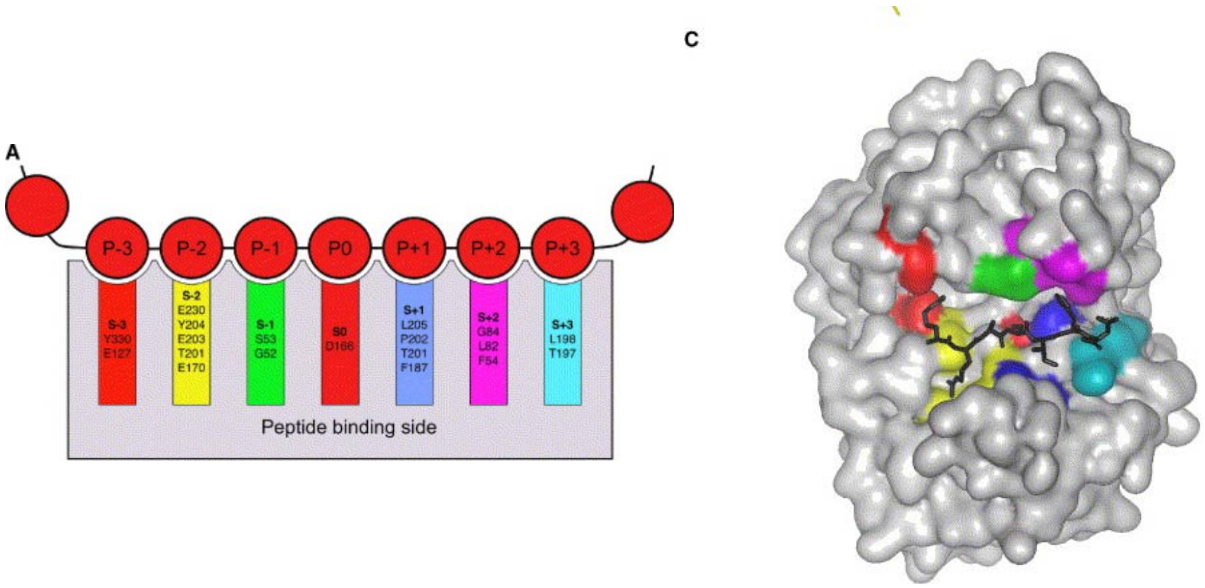


Figure from Kobe et al (2005). "Substrate binding in protein kinase A. (A) Schematic representation of the binding sites of the substrate side-chains, with the specificity-determining residues (SDRs or determinants) listed in each subsite. The sub-sites are coloured: S−3, red; S−2, yellow, S−1, green; S0, orange-red; S+1, dark blue; S+2, magenta; and S+3, light blue. The same colour scheme for the subsites is used in (B) and (C). (C) Surface representation highlighting the individual subsites, coloured as in (A), and a

The kinase CDK2 is a good example. It strongly prefers targets with the sequence S/T-P-X-K/R in the positions P0 to P+2. The basic residue at postion P+2 interacts with a Thr160 of the kinase that has been phosphorylated. (We will discuss the consequences of a required phosphorylation site on the kinase shortly.) The proline at position P+1 is preferred for two reasons that should be familiar to you from our discussion of secondary structure. First, it locks the peptide into a desirable backbone conformation that reduces the entropy of binding. Secondly, it lacks a hydrogen bond donor on the backbone. Any peptide without a Pro that fits into the cleft will have an unsatisfied hydrogen bond. (Consider the exchange reaction for desolvating the peptide).

How often do potential phosphorylation sites occur in the proteome? It appears that there is a selective pressure that drives the frequency of amino acids surrounding S, T and Y residues to reduce potential phosphorylation sites that could lead to off-pathway effects. The figure below shows these amino acid distributions for both bacterial and mammalian genomes. Note, for example, that Pro in the T+1 position is rare in mammalian genomes but relatively common in

bacterial genomes. One of the principal sequence requirements for the eukaryotic MAP kinases is the presence of Pro in position +1. For other examples, see the original publication.

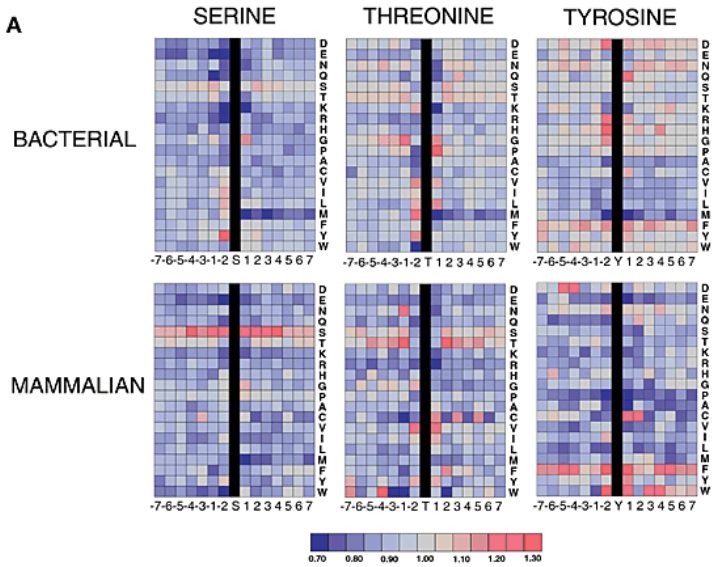


Figure from Yaffe et al. (2001). The original legend follows:

Amino acid distribution frequencies surrounding serine, threonine, and tyrosine residues in proteomes and within mapped phosphorylation sites.

(A) The relative occurrence of each amino acid in each flanking position in the bacterial and mammalian subsets of GenPept was normalized to the corresponding frequency of that amino acid within the entire database subset, and plotted topographically. Distribution values range from 0.7 (dark blue) to 1.3 (bright red). To facilitate comparisons, the color bar has been divided into two linear scales, one ranging from 0.0 (negative selection, dark blue), to 1.0 (neutral selection, gray), and one ranging from 1.0 (neutral selection) to the maximum value seen in these conserved motifs (bright red).

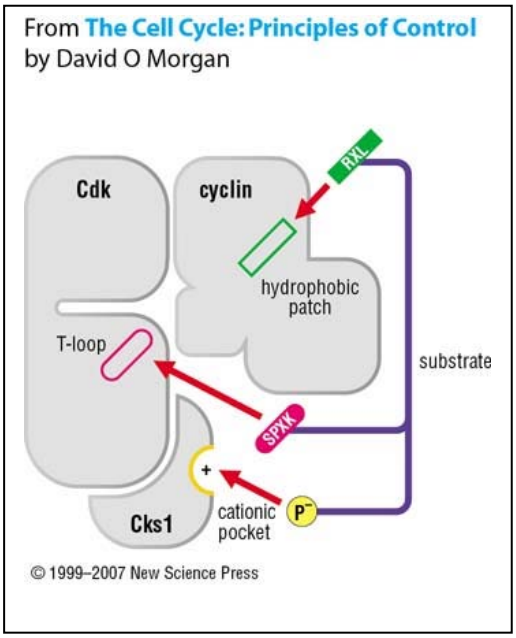
Source: Yaffe, Michael B., German G. Leparc, et al. "A Motif-based Profile Scanning Approach for Genome-wide Prediction of Signaling Pathways." *Nature biotechnology* 19, no. 4 (2001): 348-53.

Some questions for you to consider:

- Do you think that free amino acids are good kinase substrates?
- Based on what you have just read, how might you try to re-engineer a kinase to phosphorylate non-native substrates?
- Could you convert a highly specific kinase into a non-specific kinase (one that ignores the surrounding sequence, and phosphorylates all substrates equally)?
- Since the peptide binds in an extended conformation, how will the structural stability of the substrate affect its rate of phosphorylation?
- How could you experimentally determine the substrate specificity of a kinase?

The complementarity between the substrate and the active site is only one component of kinase specificity. Other protein-protein interactions between the substrate and the kinase also contribute to specificity. Tyrosine kinases tend to have separate domains responsible for protein-protein interactions, such as the SH2 and SH3 domains. In S/T kinases the interactions tend to occur either between an adapter protein and the target or between the kinase domain itself and regions of the target known as docking domains. (These docking "domains" are not structural domains and would be better referred to as docking motifs).

Cyclin proteins are adapter proteins that are essential partners of the aptly named cyclin-dependent kinases. Cyclins recognize short sequences on the target proteins and recruit them to the kinase. In contrast, targets of the MAP kinases (mitogen-activated protein kinases) are recruited by interactions between docking motifs and the kinase domain itself. The D domain, which is often 50-100 residues away from the P-site, for example, has a consensus motif of $(R/K)_{1-2}$-X_{2-6}-hydrophobic-x-hydrophobic. As you might expect, charged residues interact with a negatively charged region on the kinase and the hydrophobic residues bind to a hydrophobic region. Variations in the D domain sequences among kinases lead them to bind different substrates with different affinities. Another docking motif in MAPK subsrates is the DEF domain, which is usually ten amino acids downstream of P0 and binds to a pocket near the active site. How do these docking domains work? Most probably function by increasing the local concentrate of the target near the active site. However, some function as either positive or negative allosteric regulators.



Other phosphorylation events also affect specificity. For example, phosphorylation of site A on a target can require previous phosphorylation of site B, a phenomenon known as priming. Thus GSK3 has a preferred target sequence of S-X_3-pS, and PLK1 phosphorylates S-pS/pT-P/X (pT= phosphothreonine; pS= phosphoserine). Such requirements can create a dependency between two signaling pathways, requiring both to be active before a substrate is fully phosphorylated. This is an example of a biological AND gate. Consider what would happen in each of N signaling pathways was inappropriately activated with a probability p_i, but the phenotypic consequences required all N phosphorylation events. What is the probability of obtaining the phenotypic consequences inappropriately?

A final and very important cause of specificity is localization. It is tempting to think of the cell as a "well mixed reaction vessel," and *in vitro* experiments measure the reactions that would occur if this was true. However, it clearly is not. Most obviously, the cell is divided into organelles, and a kinase and substrate cannot interact if they are held in separate compartments. The ERK2 kinase, for example, has been shown to have distinct targets in the nucleus and cytoplasm, and differential phosphorylation of these two sets of targets may explain the different effects of two cytokines, NGF and EGF.

Scaffold proteins also regulate localization. The best characterized scaffold protein is Ste5, which brings together three kinases that phosphorylate each other in a "kinase cascade." Ste11, a MAPKKK (MAP kinase kinase kinase), phosphorylates Ste7, a MAPKK, which phosphorylates Fus3, a MAPK. Ste11, which initiates this cascade, also initiates two other cascades that lead to very different phenotypic outcomes. (The Ste11-Ste7-Fus3 cascade

regulates mating; the other two kinases cascades regulate filamentation and response to osmotic stress). Although the role of Ste5 in controlling wild-type signaling is not completely clear, chimeric scaffold proteins have been made that alter a mating signal into an osmotic response. Scaffold proteins have other roles that we will not discuss at this point, including the ability to modulate the dynamics of signaling through allosteric control of the kinases and the recruitment of phosphatases that down-regulate the signaling pathways.

PROTEIN-DNA INTERFACES

Transcriptional regulation depends on the ability for DNA-binding proteins to identify short stretches of DNA with high-specificity from among the billions of sequences in a genome. This ability seems remarkable. However, it was appreciated early on that the regular structure of DNA may make protein-DNA interactions easier to understand than protein-protein interactions. In important early paper (Seeman, et al. 1976), proposed that DNA sequences could be readily detected by proteins if they "read out" the pattern of hydrogen bond donors and acceptors in the major groove (see the figure below). Contacts in the minor groove could distinguish AT from GC containing base pairs, but could not distinguish A-T from T-A.

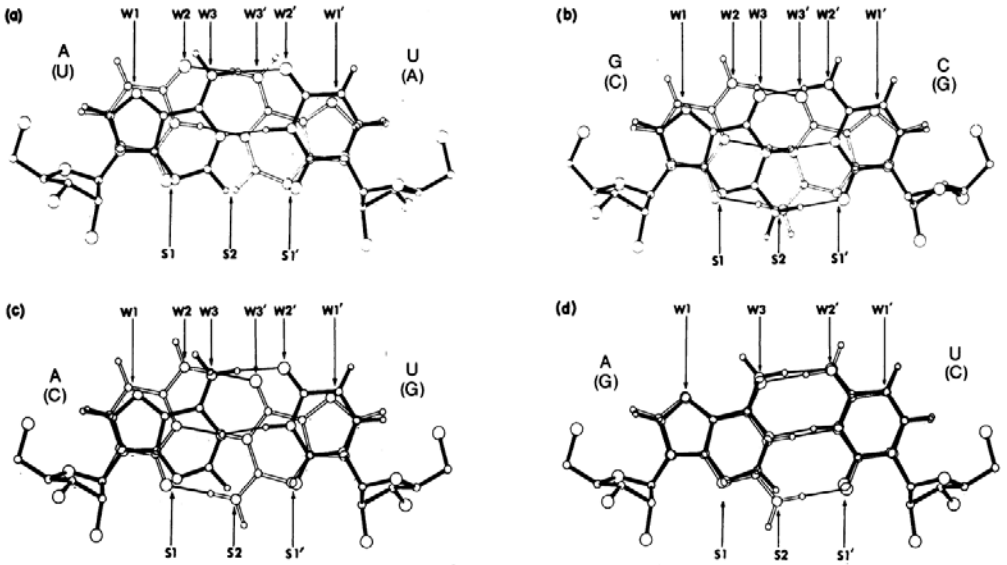


FIG. 1. Diagram showing the stereochemistry of double helical A-U and G-C base pairs. The geometry of the base pairs and the attached ribose residues were obtained from crystallographic analysis of double helical ApU (2) and GpC (3). The base pairs are superimposed upon each other with one base pair drawn with solid bonds and the other with outlined bonds. The upper letter at the side refers to the solid bases while the lower letter in parentheses refers to the outlined bases. However, both bases are drawn as attached to the same ribose residues in the antiparallel double helical conformation. W refers to a potential recognition site in the major or wide groove of the double helix; S refers to sites in the minor or small groove. The dyad axis between the two antiparallel ribose residues is vertical in the plane of the paper. (a) through (d) represent all of the possible base pair comparisons.

Table 1. Discrimination of Watson–Crick base pairs by single interactions

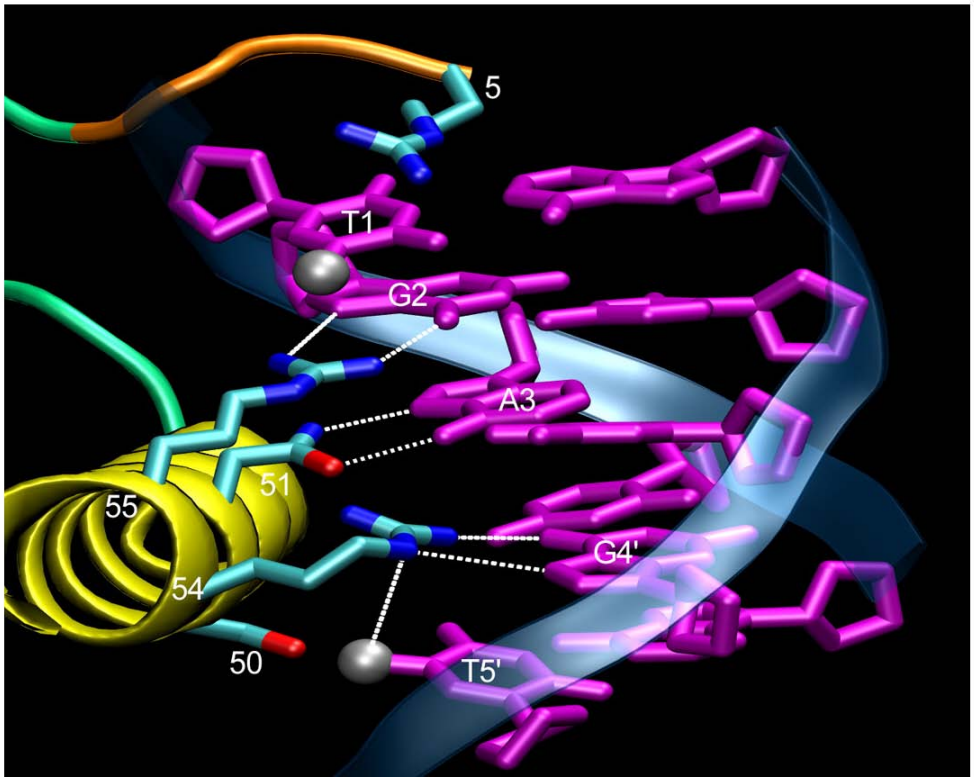
Sites	A·U / U·A	G·C / C·G	A·U / C·G	G·C / U·A	A·U / G·C	U·A / C·G
Outer major groove (W1/W1′)	+	+	+	+	0	0
Central major groove (W2, W3/W2′, W3′)	+	+	(0)	(0)	+	+
Outer minor groove (S1/S1′)	*	*	*	*	0	0
Central minor groove (S2)	0	(0)	+	+	+	+

The columns in this table refer to Figs. 1a, 1b, 1c, 1c(rev.), 1d, and 1d(rev.), respectively. This table applies to A·T as well as A·U pairs, except for the case of the outer major groove. In that case the two pyrimidines, cytosine and thymine, could be distinguished because of the thymine methyl group. However, the purines would still be degenerate. The symbols are defined as follows: +, indicates sites which could give strong discrimination between the alternatives listed. 0, indicates virtually identical sites resulting in potential ambiguities. (0), indicates only small steric differences, which might result in ambiguities if the interacting atom from the protein is free to move slightly. *, indicates that the hydrogen bonding properties of the site appear identical, but that discrimination could possibly occur through preferential ion binding.

From Seeman et al. (1976).

Source: Seeman, Nadrian C., John M. Rosenberg, et al. "Sequence-Specific Recognition of Double Helical Nucleic Acids by Proteins." *Proceedings of the National Academy of Sciences* 73, no. 3 (1976): 804-8.

The subsequent determination of many crystal structures of protein-DNA complexes has show that this hypothesis is largely correct. Almost all proteins that recognize specific DNA sequences use contacts in the major groove to directly "read" the DNA sequence. The figure below shows contacts between a homeodomain and DNA.



From Noyes et al. (2008), showing contacts in both the major and minor grooves.

Source: Noyes, Marcus B., Ryan G. Christensen, et al. "Analysis of Homeodomain Specificities Allows the Family-Wide Prediction of Preferred Recognition Sites." *Cell* 133, no. 7 (2008): 1277-89.

Proteins use several other recognition methods in addition to base contacts in the major groove, which are summarized in the figure above. "Indirect readout" is an important mechanism that is not obvious from static structures: optimal interactions between the protein and both the bases and the DNA backbone require the DNA to be deformed. Since the flexibility of a DNA sequence is influenced by its sequence, the free energy obtained from the contacts has a sequence dependence that goes beyond the base-specific contacts.

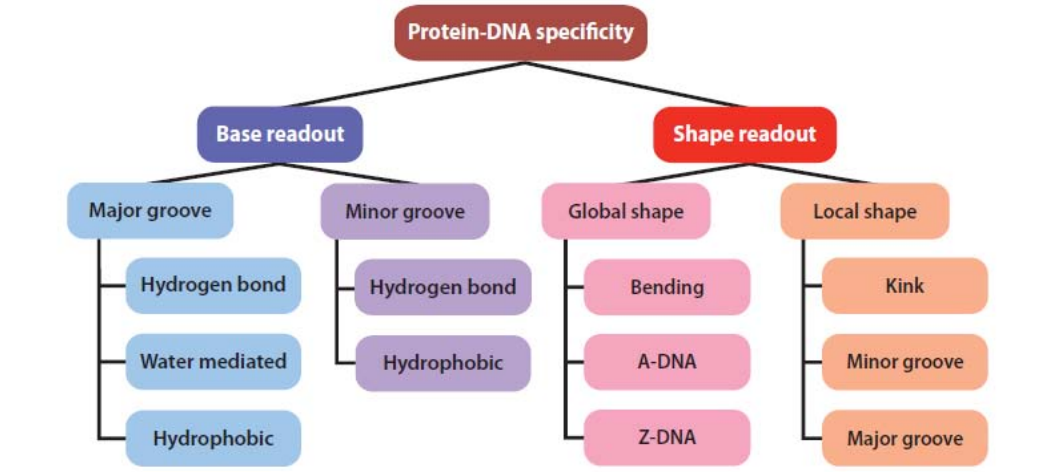


Figure 4

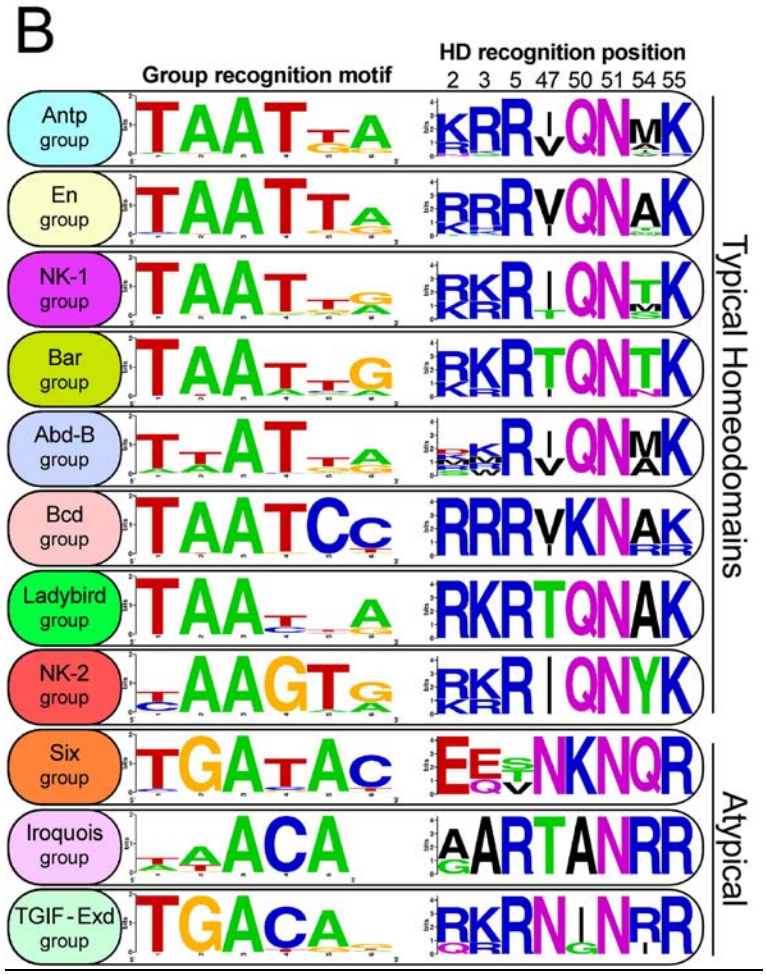
Types of protein-DNA recognition mechanisms used for specificity. We distinguish between two main classes of recognition: base readout and shape readout, which are further subdivided as illustrated.

From Rohs (2010)

Rohs *et al.*(2009) have shown how many DNA-binding proteins indirectly detect stretches of "A" bases. Even very short stretches of A bases tend to cause the minor groove of the DNA to narrow, producing a very distinct electrostatic potential surface. Insertion of an arginine residue into the groove is very electrostatically favorable, even more so than a lysine. (Can you guess why Lys is less favorable that Arg?) When the sequence contains a "T" followed by an "A" (denoted TpA, where the "p" represents the phosphate) or a GC base pair, the groove tends to be significantly wider and the interaction less favorable. See Rohs *et al.* (2010) for a review of the mechanisms of protein-DNA recognition.

Domain Families

The domain structure of a DNA-binding protein imposes constraints on the range of sequences to which it can bind. Two studies extensively characterized the sequence specificities of homeodomains from two species (Noyes et al (2008) and Berger et al. (2008)). The figure below shows the resulting classifications from one of these papers. Although there is clearly variation, most of the families, some of which are quite large, recognize sequences with the core sequence TAAT. The differences between the Antp and En group, for example are very unlikely to explain the phenotypic differences that can be traced back to mutations in proteins from each class. The full specificity of the protein is determined by the protein-DNA interactions of multiple DNA-binding proteins that interact with each other through protein-protein contacts.



From Noyes et al. (2008). "The typical and atypical homeodomains are distributed into separate groups. The average specificity of each group is indicated under the group recognition motif, and to the right is the sequence logo of the key recognition positions." (The designation of "typical" and "atypical" is based on the sequence of the protein.

Source: Noyes, Marcus B., Ryan G. Christensen, et al. "Analysis of Homeodomain Specificities Allows the Family-Wide Prediction of Preferred Recognition Sites." *Cell* 133, no. 7 (2008): 1277-89.

Versatile Families

Two families of DNA-binding proteins are unusual in that they are capable of recognizing a wide range of sequences: the TFIIIA-style zinc fingers and the TAL effectors. In both cases, the proteins are composed of small repeating domains that bind to adjacent regions of the DNA. This makes them very useful for biotechnology applications, as we saw in the introduction to this section. By modifying the contact residues, it is possible to create proteins that can recognize arbitrary regions of the genome.

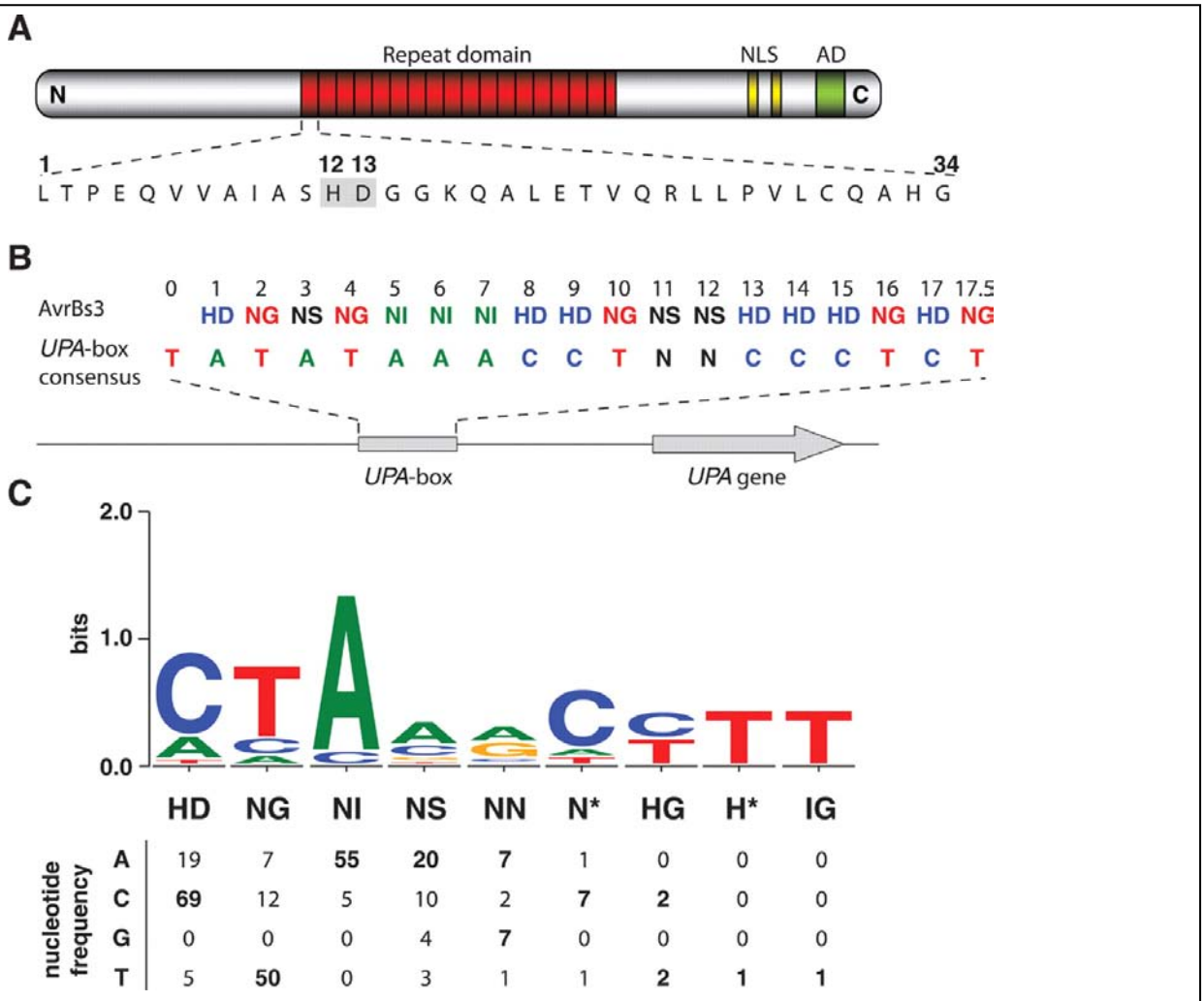


Figure and Legend from Boch et al 2009.

Model for DNA-target specificity of TAL effectors. (**A**) TAL effectors contain central tandem repeats, NLSs, and an AD. Shown is the amino acid sequence of the first repeat of AvrBs3. Hypervariable amino acids 12 and 13 are shaded in gray. (**B**) Hypervariable amino acids at position 12 and 13 of the 17.5 AvrBs3 repeats are aligned to the*UPA* box consensus (*14*). (**C**) Repeats of TAL effectors and predicted target sequences in promoters of induced genes were aligned manually. Nucleotides in the upper DNA strand that correspond to the hypervariable amino acids in each repeat were counted on the basis of the following combinations of eight effectors and experimentally identified target genes:
AvrBs3/*Bs3*,*UPA10*, *UPA12*, *UPA14*, *UPA19*, *UPA20*, *UPA21*, *UPA23*, *UPA25*, AvrBs3Δrep16/*Bs3-E*, AvrBs3Δrep109/*Bs3*, AvrHah1/*Bs3*, AvrXa27/*Xa27*, PthXo1/*Xa13*, PthXo6/*OsTFX1*, and PthXo7/*OsTFIIAγ1*(fig. S1). An asterisk indicates that amino acid 13 is missing in this repeat type. Highest nucleotide frequencies are in bold. Nucleotide frequencies are displayed in a logo (http://weblogo.threeplusone.com).

Source: Boch, Jens, Heidi Scholze, et al. "Breaking the Code of DNA Binding Specificity of TAL-type III Effectors." *Science* 326, no. 5959 (2009): 1509-12.

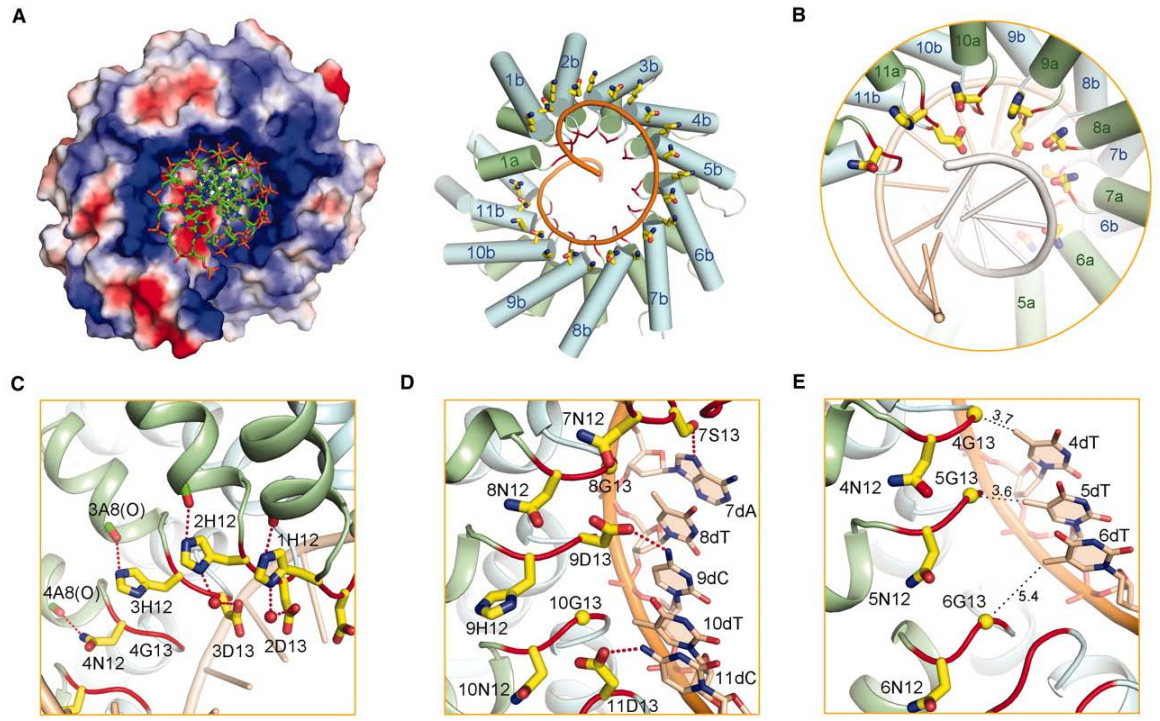


Figure and legend from Deng et al 2012.

DNA recognition by TAL repeats. (**A**) The phosphate groups of the DNA sense strand is embraced by the positively charged ridge of the dHax3 TAL repeats. The surface electrostatic potential was calculated with PyMOL (*30*) (left). The invariant residues Lys^{16} and Gln^{17} (yellow sticks), located at the beginning of helix b in each TAL repeat, contribute to the positive electrostatic potential (right). The RVD loops are highlighted in red. (**B**) The two hypervariable residues in each TAL repeat are placed in the major groove of DNA. The sense and antisense strands of DNA are colored gold and gray, respectively. (**C**) The hypervariable residues at position 12 do not contact DNA bases. These residues, either His or Asn in dHax3 repeats, form hydrogen bonds with the carbonyl oxygen of Ala^{8} in the same repeat, which may help stabilize the conformation of the RVD loop. When consecutive repeats containing HD are present, His^{12} forms a water-mediated H bond with Asp^{13} from the previous repeat. (**D**) The hypervariable residues at position 13 are direct determinants of DNA base specificity. Shown here are repeats 7 to 11 and the corresponding nucleotides from the DNA sense strand. (**E**) Recognition of base T by NG. A close-up view on the RVD loops in TAL repeats 4 to 6 in molecule A is shown. Note that the RVD loop of repeat 6 adopts a conformation different from all other RVD loops. All distances are shown in the unit of Å.

References:

Alon, U. (2007) Nat Review Genetics 8(6):450-61.

Berger et al. (2008). Cell 133: 1266-76.

Boch et al. (2009). Science 326:1509-1512.

Clackson and Wells (1995). Science 267:383-386.

Deng et al (2012) Science 335:720-723.

Deremble and Lavery. (2005) Current Opinion in Structural Biology 15: 171-175.

Fong, et al. (2004). Genome Biology 5:R11.

Gnad, et al. Genome Biology 2007, 8:R250.

Grigoryan and Keating. Current Opinion in Structural Biology 18:1-7.

Grigoryan *et al.* (2009) Nature 458, 859-864

Kaplan, *et al.* (2005). PLoS Comput Biol. 1: e1.

Keating and Newman (2003). Science 300:2097-2101.

Kobe et al. (2005). Bioch. Biophys. Acta 1754:200-9.

Linding et al. (2007). Cell 129:1415-26.

Moreira, et al. (2007). Proteins 68: 803-812.

Noyes, et al. (2008). Cell 133:1277-89.

Rohs, et al. (2009). Nature. 461:1248-53.

Rohs, et al. (2010). Annu. Rev. Biochem. 79:233-69.

Seeman, et al. (1976). PNAS 73:804-8.

Ubersax and Ferrell (2007). Nature Reviews Molecular Cell Biology 8:530-541.

Woods and Schier (2008). Nature Biotech. 26:650-1.

Yaffe et al. (2001). Nature Biotech. 19:348-353.

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012