PROTEIN-PROTEIN INTERACTIONS

Catherine Royer

Introduction

The subject of protein-protein interactions represents a vast ensemble of results from biological, biochemical and biophysical studies carried out to date and cannot be treated in its entirety in any reasonable fashion. The following chapter will focus therefore on particular aspects of protein-protein interactions. It will begin with an overview of some classes of protein-protein interactions and their structural motifs, concentrating on a few important examples of homo- and hetero-oligomers. Next, biophysical methodologies for characterizing protein-protein interactions will be discussed in terms of their advantages and disadvantages for this particular application. Finally, some considerations about the energetics of protein-protein interactions and the coupling of ligand binding with protein association reactions will be presented.

Protein-protein interactions are operative at almost every level of cell function, in the structure of sub-cellular organelles, the transport machinery across the various biological membranes, packaging of chromatin, the network of sub-membrane filaments, muscle contraction, and signal transduction, regulation of gene expression, to name a few. Aberrant protein-protein interactions have implicated in a number of neurological disorders such as Creutzfeld-Jacob and Alzheimer's disease. We will be concerned mainly in this chapter with the biophysical aspects of the finely tuned specific interactions between proteins involved in the regulation of cell function, namely those proteins implicated in signal transduction and transcriptional regulation.

Because of their importance in development and disease, these systems have been the object of intense research for many years. It has emerged from these studies that nature has employed in many instances a strategy of mixing and matching of domains that specify particular classes of protein-protein interactions, modifying the amino acid sequence in order to confer specificity for particular target proteins. The regulation of cell function brought about by the interactions of these proteins is delicately balanced by the relative affinities of the various protein partners and the modulation of these affinities by the binding of ligands, other proteins, nucleic acids, ions such as Ca2+, and covalent modification, such as specific phosphorlyation or acetylation reactions. Specificity and the strength of signal transduction is encoded by the exact amino acid sequence of the domain, and it is this relationship between sequence, structure, dynamics, energetics and function that constitutes the fundamental issue for the biophysics of protein-protein interactions. This endeavor therefore requires a structural characterization of the domains and if possible their dispositions within the complete protein and their complexes with their specific protein partners. In addition to a

structural characterization, understanding the basis for specificity in these systems necessitates very careful and thorough comparative studies of similar interacting partners or mutated domains in order to bring to light the energetic properties linked to a particular sequence/structure. We must bear in mind that differences as small as 1 kcal/mol in interaction energy between pairs of protein partners can lead to profound differences in cell growth and development.

Classes of Protein-Protein Interactions

Homo-oligomerization

An enormous number of enzymes, carrier proteins, scaffolding proteins, transcriptional regulatory factors, etc. function as homo-oligomers. Incorporation of non-covalent interactions at the level of protein quaternary structure provides a number of advantages in terms of regulatory possibilities that are not afforded in cases where the functional unit is comprised of a single polypeptide chain. First, energy can be stored at the subunit interface that can serve to bind ligand, or modify the protein conformation in response to regulatory ligands. Modulation of subunit affinity in such a manner need not compromise the folded structure of the protein, yet provides a considerable energetic margin for modulation of activity. Another significant advantage to linking regulatory properties to subunit interactions lies in the simple fact that over the appropriate range, the activity of the protein becomes protein concentration dependent. Such control can function either positively or negatively, with the oligomer or the monomer exhibiting the highest activity.
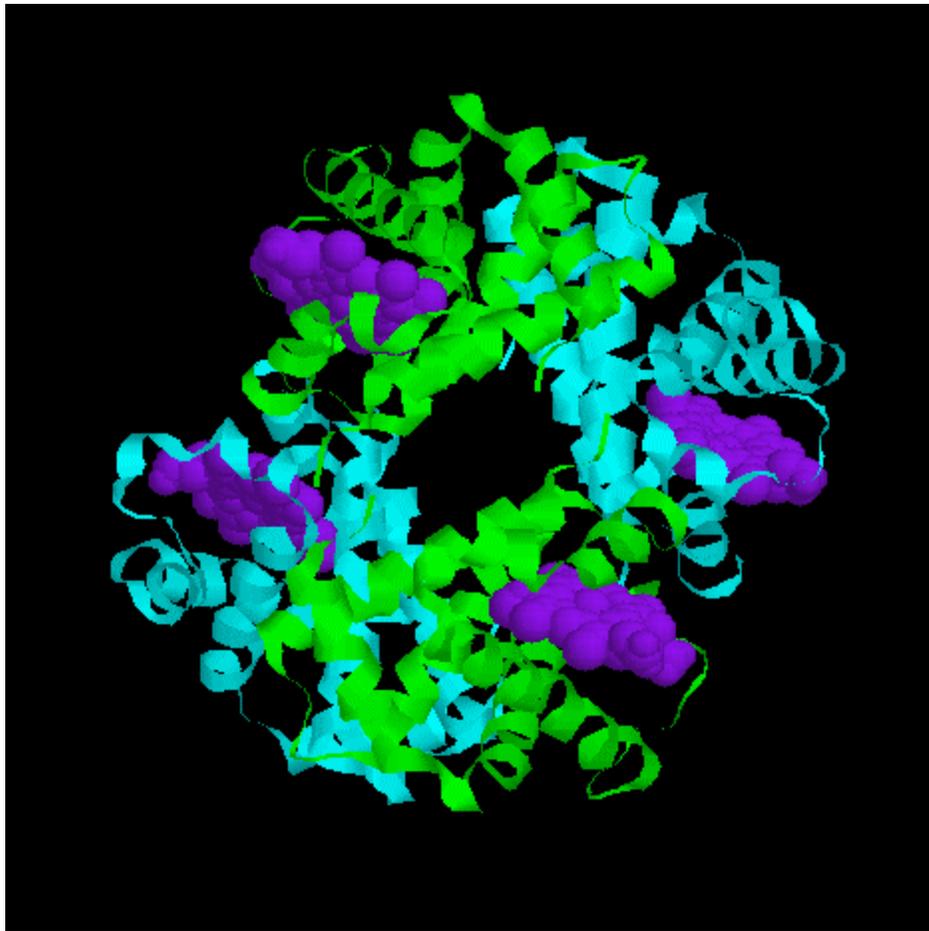

Figure 1. Schematic diagram of the structure of the hemoglobin tetramer
(Tame & Vallone, 1998)

Thus, function in oligomers can be very finely tuned by ligand concentration (including ions, substrate, allosteric ligands, protons, etc.) and by protein concentration (expression levels or degradation rates).

Although not a true homo-oligomer, the hemoglobin tetramer, (see schematic in Figure 1) is comprised of 2 _ (green) and 2 _(cyan) subunits, exhibiting very high homology, has served now for decades as a model of cooperativity and allostery in ligand binding by oligomeric proteins. The structure in Figure 1 (ref) shows the all -helical character of the hemoglobin subunits, while the heme moieties are represented in spacefilled purple. The literature on the subject of this molecule is so vast that no justice could be made here. Nonetheless, a few salient features of the relation between hemoglobin structure and function merit discussion in the present context to provide an example of the use of subunit interactions to modulate function.

In Figure 2 is a schematic diagram of the binding profile of oxygen to the monomeric myoglobin molecule and to the tetrameric hemoglobin molecule at neutral and acid pH. What is evident from these profiles, first of all, is that the affinity of myoglobin for oxygen is significantly higher than that of hemoglobin, regardless of pH. Secondly, the binding of oxygen to myoglobin presents a simple hyperbolic profile, while that for hemoglobin is sigmoidal. Moreover, it is known as well that the binding of oxygen to hemoglobin results in a decrease in the affinity between the two _ _ dimers.
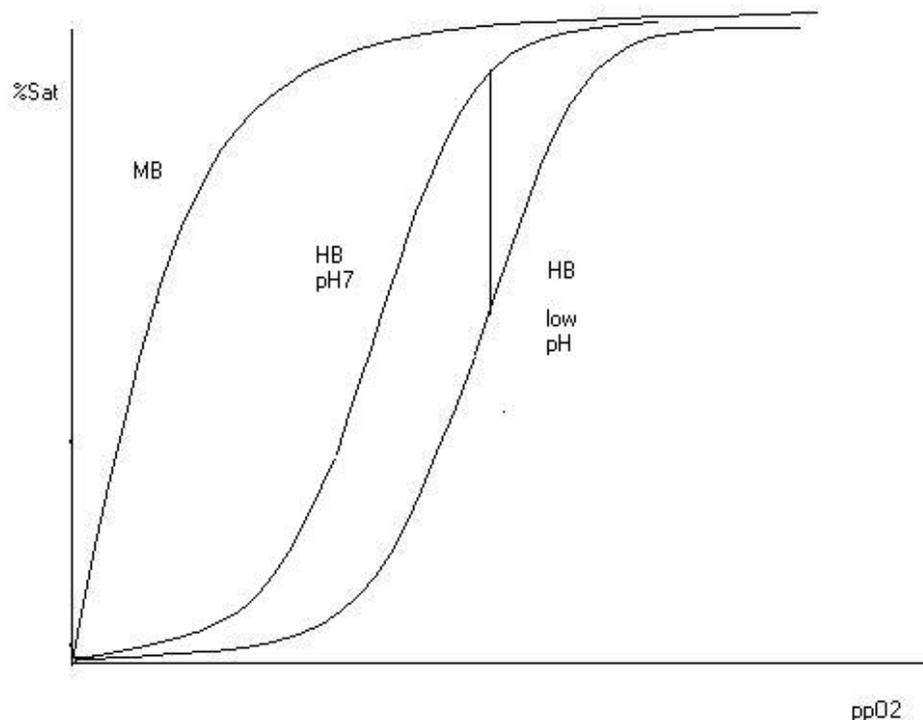

Figure 2. Schematic comparison of the oxygen binding profiles of hemoglobin and myoglobin.

In addition to exhibiting the cooperative properties described above, the binding of oxygen to hemoglobin is also under the allosteric (allo = other) control of protons and organic phosphates (eg., DPG, among others). Binding of protons to particular sites at the subunit interface leads to an increase in the constraints against oxygen binding. Physiologically this Bohr effect is important since it allows release of oxygen from hemoglobin to myoglobin in the acidic environment of muscle tissues, where the production of lactate and carbon dioxide lowers the pH.   Ascent to high altitudes is followed by an increase in production of diphosphoglycerate (DPG) which binds to a site at the interface between the two 2 _ subunits (See Figure 3) and decreases the affinity of hemoglobin for oxygen resulting in a more efficient release of oxygen to the tissues. Thus, the specific sequence of amino acids in the subunits of hemoglobin encodes not only the fold but the subunit interactions which incorporate the energetic linkages between subunit affinity and ligand binding.
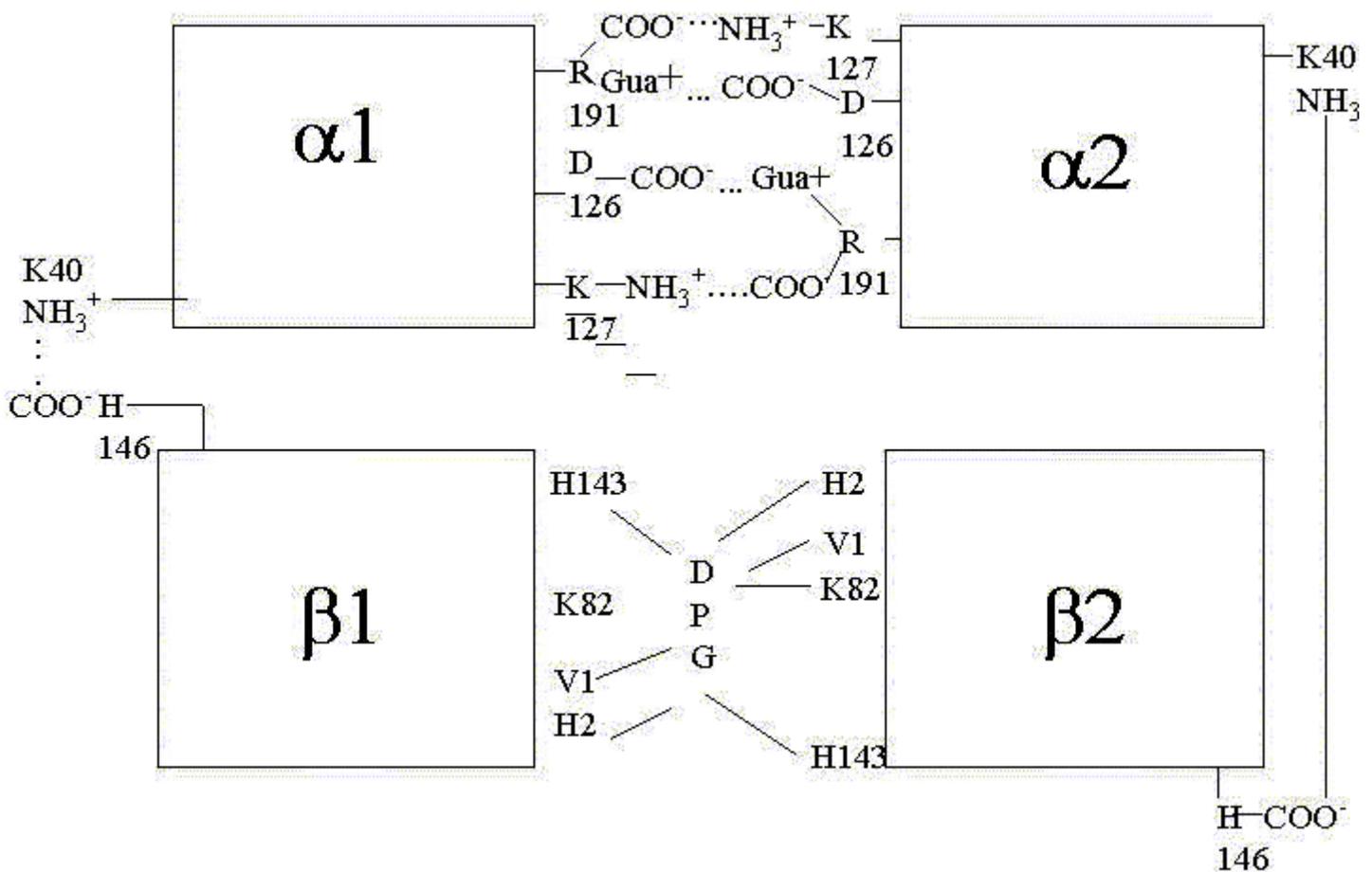

Figure 3. Diagram of the salt bridges at the subunit interfaces of the hemoglobin tetramer.

Hemoglobin represents the best studied oligomeric protein to date and the structural and energetic linkages between ligation and subunit affinity are known in exquisite detail. Literally thousands of studies have contributed to the detailed understanding of the relationships between sequence, structure, dynamics, energetics and function in this model system. Unfortunately very few other oligomeric protein systems have been characterized at such a detailed level. In many cases, this has been due to the difficulties inherent in determining the affinity between subunits, and the modulation of this affinity upon changes in ligand or solution conditions, or upon mutation of the primary sequence. Methods for measuring protein interactions will be discussed in the second section of this chapter.

Interactions between heterologous proteins

Communication at the level of the organism or the cell requires the translation of physical or chemical information signals (e.g., the presence of a particular chemical factor, a change in luminosity, the chemical modification of existing substances, a change in pH or ion concentration) from one compartment to another. In large part, this communication relies on the specific interaction between particular heterologous proteins, in response to particular chemical or physical signals. Clearly all of the mechanisms have not been elucidated that govern the specificity of the interactions between the multitude of protein partners that control cell function. Nonetheless, the resolution of the three dimensional structures of a number of the protein interaction domains by x-ray diffraction or NMR techniques has provided a great deal of insight and the basis for the development of therapeutic strategies designed to modulate some of these interactions known to play a role in various diseases. This section of this chapter is dedicated to a overview of some of the structural motifs involved in these heterologous protein-protein complexes. New structures of protein-protein complexes appear continually, and therefore, in the context of an introductory biophysics text, all known examples of important heterologous protein-protein interactions cannot be included. In the following section three classes of protein-protein interactions will be discussed. Due to their biological importance and the growing interest in the mechanism of their function, the protein-protein recognition surfaces of some of the proteins involved in signal transduction and the regulation of cell growth merit discussion. The protein-protein interactions between transcription factors, the eventual targets of the various signal transduction pathways have also become of great interest, due to their role in the control of cell growth, differentiation and development.

Protein domains involved in Signal Transduction

Src homology domains 2 and 3, commonly referred to as SH2 and SH3 domains are protein interaction domains found in proteins involved in signal transduction. The SH2 domains recognize tyrosine phosphorylated proteins, in particular autophosphorylated growth factor receptors. They are found in growth factor receptor binding proteins that intervene in signal transduction downstream of the receptors, but upsteam of Ras. They have

also been found in docking proteins. A ribbon diagram of the SH2 domain of the T-lymphocyte specific tyrosine kinase, P561ck, is shown below complexed with a carboxymethyl-phenylalanine containing peptide, designed to mimick the the natural phosphphotyrosine containing targets of this domain (Tong *et al.*, 1998). The global architecture of SH2 domains consists of a three-stranded twisted beta sheet sandwiched between two alpha-helices. The SH2 domain exhibits a deep binding pocket that presents a high affinity for phosphotyrosine, but not for either phosphoserine or phosphothreonine.
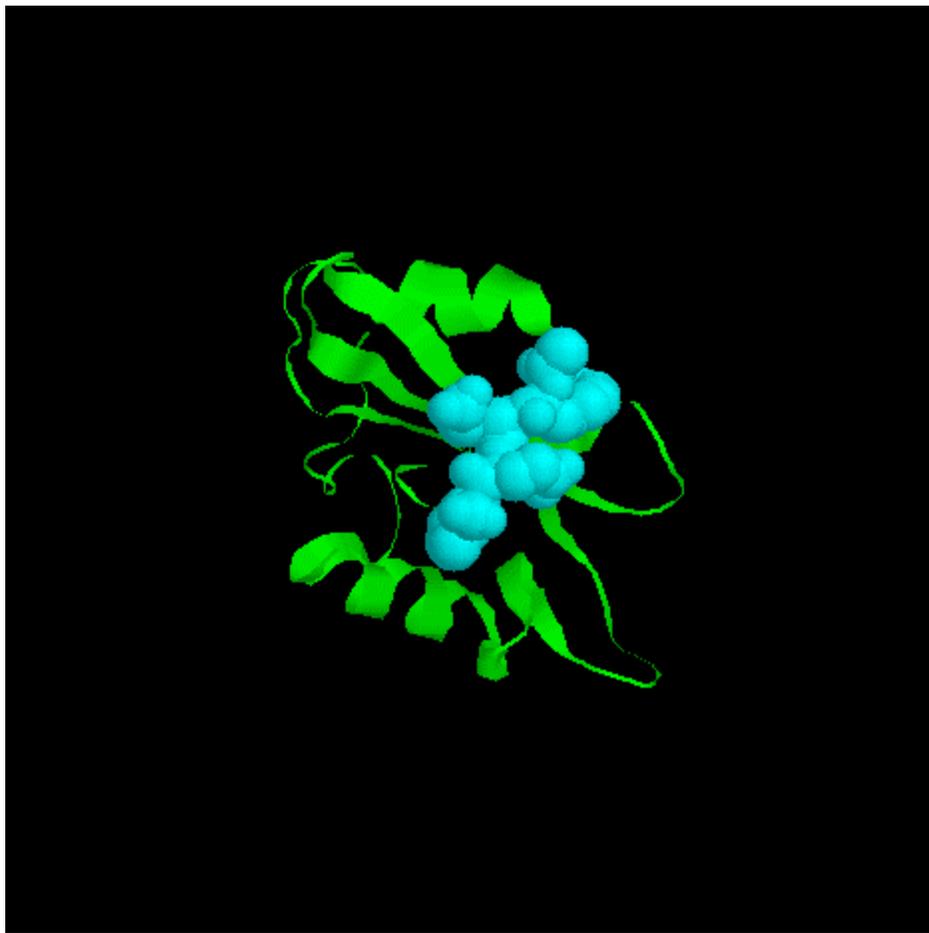

Figure 4. Ribbon diagram of the SH2 domain from p561ck (green ribbons) bound to a synthetic inhibitor (spacefilled cyan) (Tong *et al.*, 1998).

Residues carboxyterminal to the phosphotyrosine of the target protein serve to confer the specificity of the interaction. The second pocket of the SH2 domain recognizes these C-terminal specificity determinants. Different classes of SH2 domains recognize different types of C-terminal sequences. For example the class I SH2 domains recognize C-terminal tri-peptides P-P-H, where P and H stand for polar and hydrophobic. Class III SH2 domains recognize H-X-H tripeptides. The specificity determinants for the growth factor receptor binding protein 2 (Grb2) resides in the second residue C-terminal to the phosphotyrosine, which must be an asparagine. The three-dimensional structures of a number of SH2 domains

have been resolved. There is a very strong structural homology between these which is evident upon comparing the structure of the SH2 domain from p561ck to that show in Figure 5 of the complex between the C-terminal SH2 domain of the Syk tyrosine kinase and a phosphopeptide from the gamma chain of the high affinity immunoglobulin G receptor as determined by NMR (Narula *et al.*, 1995).
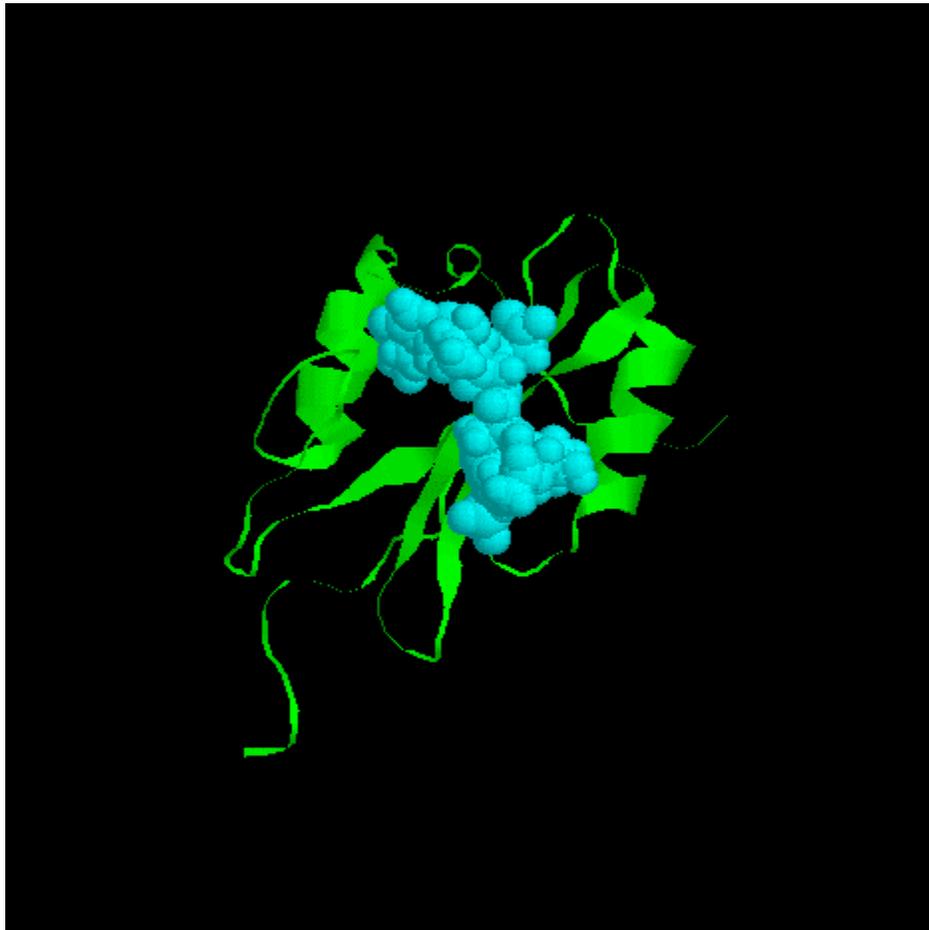

Figure 5. Complex between the C-terminal SH2 domain of the Syk tyrosine kinase and a phosphopeptide from the _ chain of the high affinity IgG receptor

Src homology 3 (SH3) domains bind to proline rich sequences in their target protein partners. Like SH2 domains, SH3 domains are also found in proteins involved in signal transduction, such as the protein tyrosine kinases. They recognize polyproline type II helical structures (PXXP motifs) in cell signaling proteins. For example the SH3 domain of Grb2 recognizes the PXXP motif of a guanyl nucleotide releasing factor that converts GDP-Ras to GTP-Ras, thus stimulating signal transduction. The SH3 domain of the src kinase, Fyn, can bind the PXXP motif of the HIV-1 protein Nef, and this interaction has been shown to be crucial for AIDS pathogenesis. The SH3 domain of the fyn proto-oncogene complexed with a synthetic peptide from the proline-rich binding site of the p85 subunit of PI3 kinase (Renzoni *et al.*, 1996) is shown in Figure 6 below. As can be seen, the SH3 domain

architecture consists of a beta barrel formed by two orthogonal beta sheets, made up of three anti-parallel beta strands. The polyproline residues of the target proteins pack onto aromatic residues of the SH3 domains.
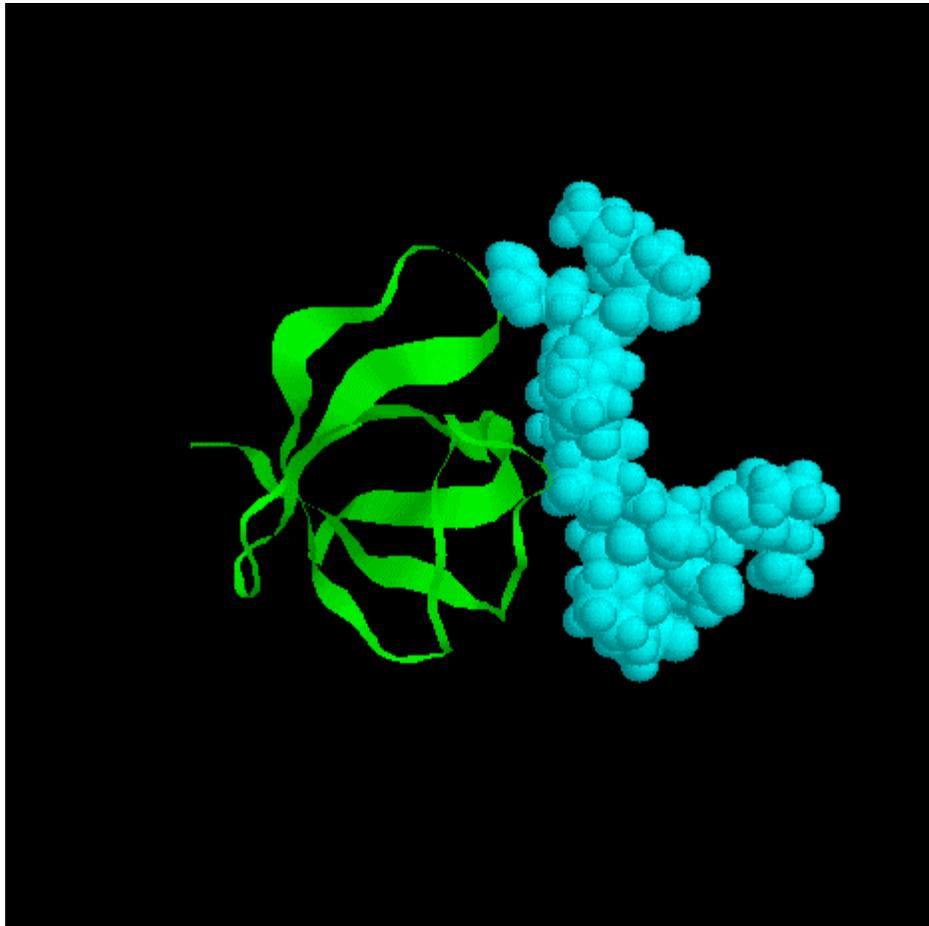

Figure 6. SH3 domain of Fyn complexed with a synthetic proline-rich peptide
(Renzoni *et al.*, 1996).

Although some auto-phosphorylated receptors can bind directly to the SH2 domains of signaling proteins such as Grb2, other receptors do not display the appropriate motif. For example, receptors such as NGF receptor must tyrosine phosphorylate adaptor protein factors such as Shc, that bind to NPXpY motifs in the receptors through a PI/PTB domain. Unlike the SH2 domains, the phosphopeptide interacting domains, or PI/PTB domains, recognize the sequence that is N-terminal, rather than C-terminal, to the phosphotyrosine of the receptor. These adapter proteins, once tyrosine phosphorlyated by the receptor, become competent to bind to the Grb2/Sos complex, thus coupling the receptor to the Ras signaling pathway. An example of the structure of a PTB domain is seen below in Figure 7. This protein is the PTB domain of the X11 protein is a neuron specific protein that has been found to bind to the internalization motif of the Alzheimer's amyloid precurseur protein (Zhang *et al.*, 1997). It contains a core -sandwich with a C-terminal -helix, termed PH-fold, so-called because it was first described as a plekstrin homology domain (Macias *et al.*, 1994) in

addition to an inserted -strand and -helix. The peptide derived from the Alzheimer's precursor protein (in spacefilled cyan) binds to the 5th -strand through hydrogen bonding interactions. It also interacts with the helix,   2, and the peptide's C-terminal 3^{10} helix makes more contacts with the X11 PTB domain.
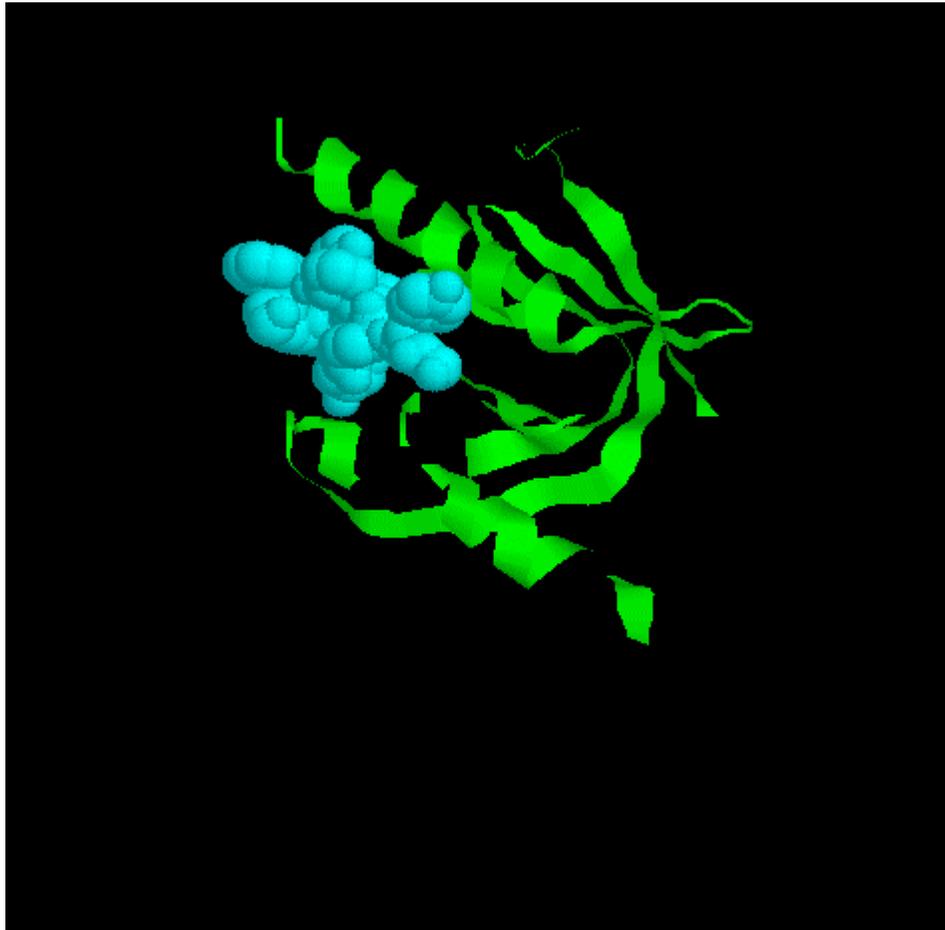

Fig. 7 Complex between the X11 PTB domain and the Alzhiemers precursor protein internalization motif.

Yet another class of protein interaction motifs involved in cell signaling pathways is represented by the PDZ domains. These domains (also termed GLFG repeats or DHR domains) were first identified as 90 amino acid segments in three proteins, **P**SD-95, **D**lgA and **Z**O-1 (hence the acronym, PDZ domain) which are all guanylate kinases. Proteins containing PDZ domains have been implicated in ion channel receptor clustering, receptor/enzyme coupling and a variety of other protein associations. The PZD domains recognize carboxy-terminal tri-peptide motifs (S/TXV), other PZD domains or even LIM domains (see below). The structure of the third PDZ domain of DglA is reminiscent of the SH2 domain, implicating a beta barrel sandwiched between two alpha-helices (refs). In this structure the PZD domain forms a dimer, which may be typical of the structure of dimerizing PZD domains such as

those between nitric oxide synthase and PSD-95. In Figure 8 is shown the structure of the third PDZ domain from PSD95 complexed with the C-terminal peptide from Cript (Doyle *et al.*, 1996). Like the PTB domain in Figure 7, this PDZ domain is made up of a six stranded β-sandwich, flanked by two α-helices. The loop between the first two β-strands contains the GLGF residues that bind the C-terminal carboxylate group of the receptors and channels. The peptide is bound in a groove on the PDZ surface with its C-terminal end protruding into a cavity in the protein.
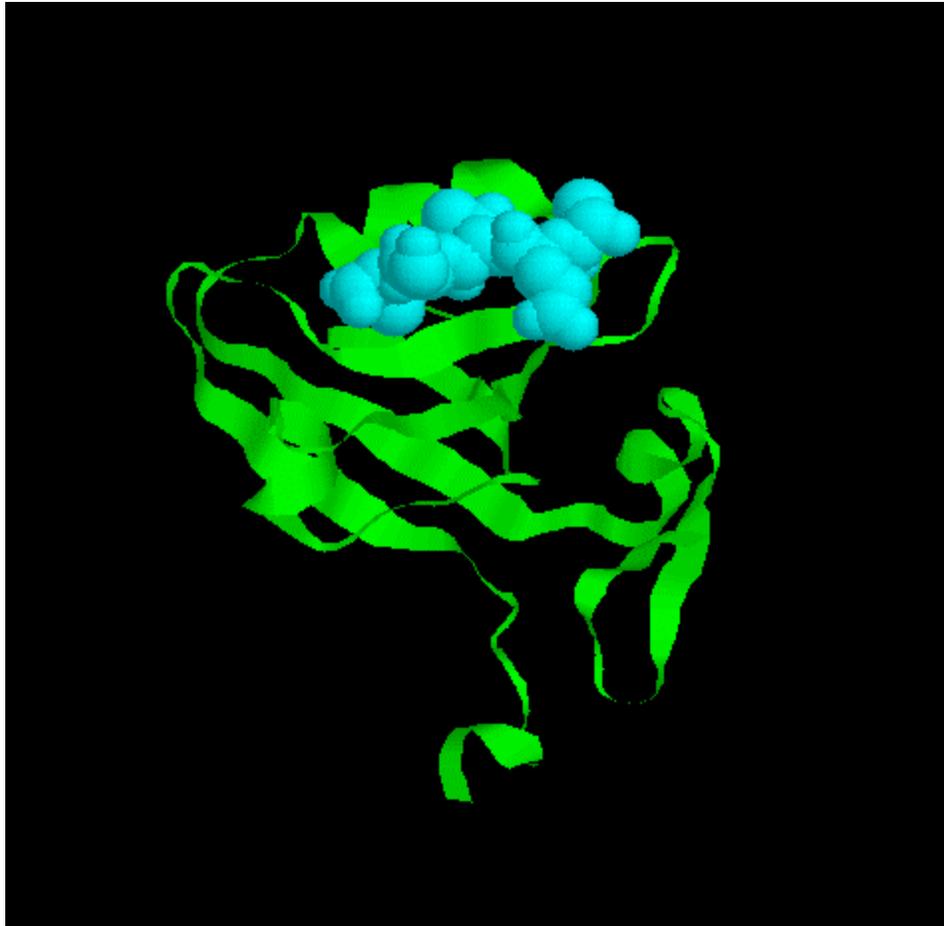

Figure 8. PDZ domain of PSD-95 complexed with the C-terminal peptide derived from the PDZ interacting protein.

Another domain involved in cellular signaling is the LIM domain. LIM domains are zinc finger cysteine rich domains found in many homeodomain proteins involved in development and in non-homeodomain proteins involved in differentiation, association with the cyto-skeleton or in cellular senescence. These domains bind to PDZ motifs, bHLH transcription factors (see below), other LIM domains, and LIM binding proteins, and thus mediate protein-protein interactions implicated in important cell functions. LIM domains were first identified in three homeodomain transcription factors (**l**in11, **i**sl1 and **m**ec3, hence LIM domain), bear the consensus sequence $CX_2CX_{16-23}HX_2CX_2CX_2CX_{16-21}CX_2C/H/D$, and

have been identified in over 35 proteins to date.   The structure, solved by NMR techniques of the LIM domain from the LASP protein (Hamarstrom *et al.*, 1996) is shown in Figure 9. The zinc atom, shown in space-filled cyan, is coordinated by cysteine residues in the two loops corresponding to the double finger.
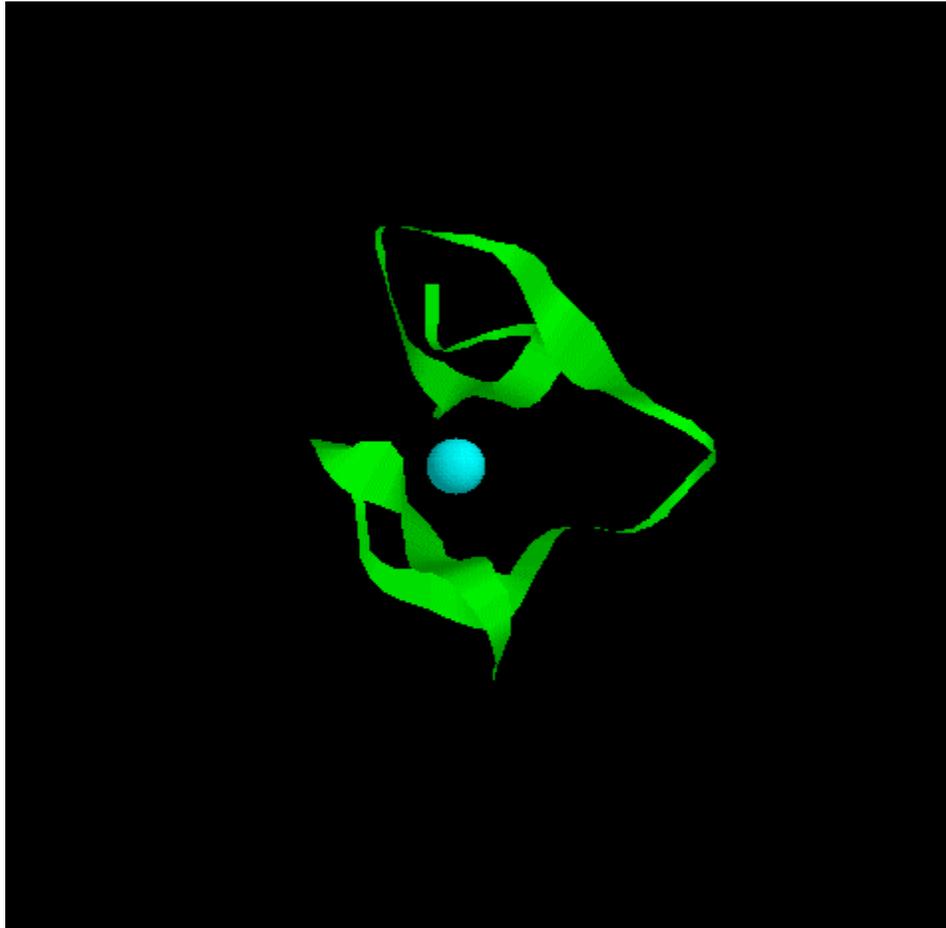

Figure 9. Structure of the LIM domain of the LASP protein.

A number of other protein interaction motifs have been identified that play a role in signal transduction and development. An incomplete list includes the pleckstrin homology domain (referred to in the section concerning the PTB domains) which binds to acidic domains in signal transduction proteins as well as to phosphoinositdes, the WW domain, a semi-conserved region of 38-40 amino acids termed WW because of the two conserved tryptophan residues spaced 20 amino acids apart and which interacts with proteins involved in cell-signaling, the WSxWS motif in cytokine receptors and the WD repeat. A shown above, structural information about these motifs is rapidly growing, and serves to guide interpretations concerning the physical basis for specific interactions in these systems. In addition to further structural characterizations, a full understanding of the function and specificity of these domains will require a great deal of work to characterize and compare the

energetic and dynamic properties of these systems. Such physical characterization of the interactions between these motifs and their various and disparate partners will surely provide fundamental information for understanding cell differentiation and growth, which can also be used in designing therapeutic strategies to combat a large number of diseases in which these biomolecules are implicated.

Protein domains involved in transcriptional regulation

A key juncture in the control of cell differentiation and growth is found at the level of transcription of DNA to mRNA. This process is mediated by a number of protein-protein and protein-DNA complexes, in addition to that between the RNA polymerase and the gene and transcript. Protein factors that serve to modulate gene transcription have been termed transcription factors. They fall into three large groups, the general transcription factors involved in the general multi-protein transcriptional machinery, the transcription associated or bridging factors, and specific factors that recognize control sequences in the non-coding regions of the DNA. Modulation of the expression of genes by these factors occurs through their ability to enhance (activation) or disfavor (repression) recruitment of the proteins involved in the transcription. This process therefore involves specific, high affinity protein-protein interactions, and understanding the structural and energetic basis for the affinity and specificity of these interactions is paramount to the understanding of transcriptional control. Biochemists have identified a very large number of procaryotic and eukaryotic transcription factors, which cannot all be presented in the context of this introduction to protein-protein interactions. Moreover, transcription factors are also discussed in a separate chapter of this volume concerning protein DNA interactions. In addition to biochemical characterizations, the three dimensional structures of a number of these factors have been determined either alone and/or complexed with DNA.   A few subclasses or categories of specific transcription factors emerge from the list of those identified to date. These categories are sometimes based on sequence or structural homology, in other instances on functional homology. An exhaustive list of these categories is not possible since the list grows continually. However, a few of these categories will be presented here.

One important family is that of the bHLH transcription factors, so named because of a structural motif that includes a basic region that binds to DNA and a helix-loop-helix region that is involved in dimerization. Members of this family include the MyoD family of transcription factors that regulate the expression of muscle specific structural genes and thus control the differentiation of muscle cells. Other members of the bHLH family of transcription factors are the ubiquitous products of the E2A gene, E12 and E47. Members of the MyoD family can form homodimers or heterodimerize with the E12 and E47 proteins, which in turn can be sequestered by inhibitor HLH proteins such as Id that lack the DNA binding basic region be retain the HLH dimerization region. Thus, differential interaction between various protein partners can modulate occupation of DNA targets. In Figure 10 is shown a representation of the bHLH domain of MyoD bound to DNA (Ma *et al.*, 1994).

Dimerization is mediated by a four helix bundle made up of two helices from each of the subunits.

Another class of transcription factors is represented by the leucine zipper transcription factors such of which c-jun and c-fos (constituents of the AP1 transcription factor) are the most well-known. AP1 is involved in regulation of gene transcription linked to cell proliferation. Like the bHLH family the interactions between bzip family members are mediated by the coiled alpha helices. In the bzip proteins the leucine residues positioned at every turn of the helix provide the hydrophobic interface for the protein-protein interactions, hence the name, leucine zipper. N-terminal to the zipper is found the basic region responsible for recognition of and binding to specific DNA sequences. These factors can homo- and hetero-dimerize, and the mixing and matching of protein partners has consequences on DNA specificity and the levels of transcriptional activation that follow.
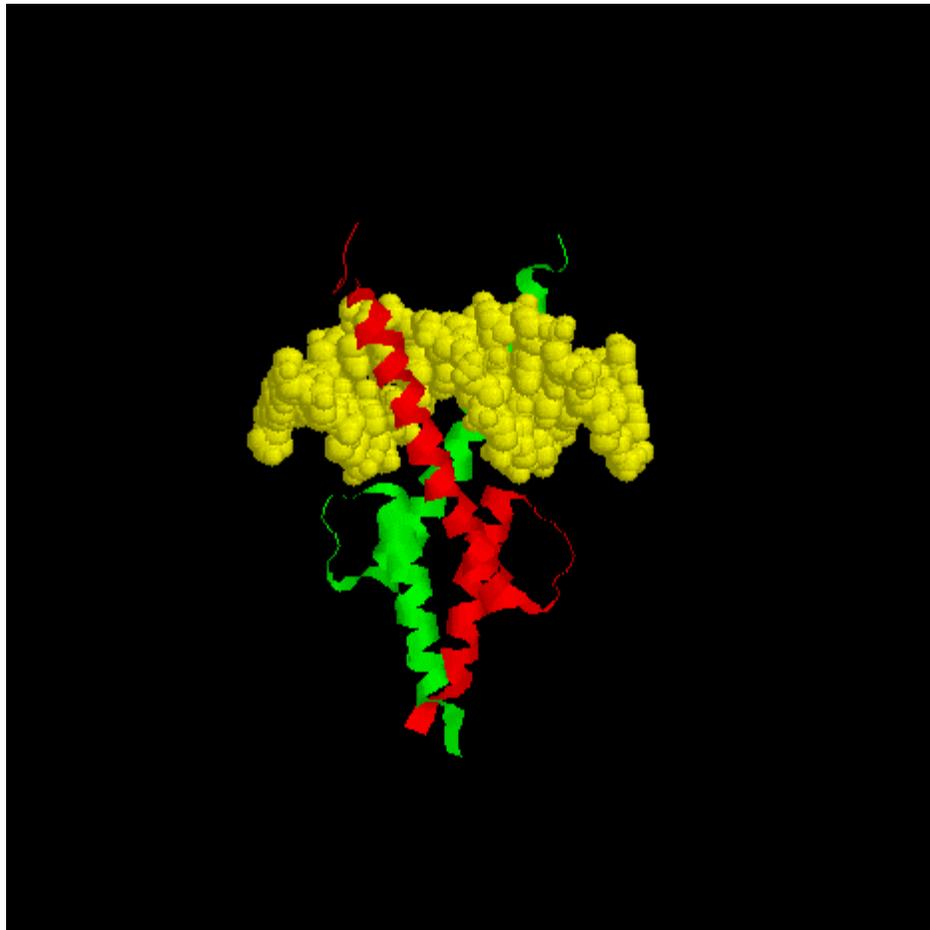
Figure 10. bHLH domain of MyoD cmplexed with DNA.

In Figure 11 is shown a schematic representation of the b-zip domains of the c-Fos/c-Jun heterodimer bound to DNA (Glover & Harrison, 1995). These proteins are eukaryotic transcription factors involved in modulating cell growth and homeostasis. Oncogenic forms of these proteins have been implicated in a variety of cancers. The salient feature of this complex is the astoundingly long -helix of each monomer, arranged in a coiled-coil manner to form the

dimer interface, which is then prolonged across the DNA surface. The c-Fos monomer is colored in yellow, while the c-Jun is red. The coiled-coil is maintained by hydrophobic interactions between the leucine residues (shown as space-filled cyan in the Figure) placed every four residues in the sequence such that they reside at the dimer interface in three dimensions.   In addition hydrophobic residues are found adjacent to the leucines and also distributed every four residues. These residues can be leucines or valines or sometimes even a more hydrophilic residue. Thus, the conserved leucines occupy position D in the helical wheel, while the less conserved hyrophobic residues occupy position A. Residues at these two positions make up the dimer interface. One can imagine that modulation of the identity at the A position for different dimer pairs could serve to modulate the relative affinity between members of this family. Interactions at the interface are also reinforced by contacts between charged sidechains at adjacent to the hydrophobic residues. Examination of the Figure underscores the appropriateness of the name leucine zipper for this dimerization motif.
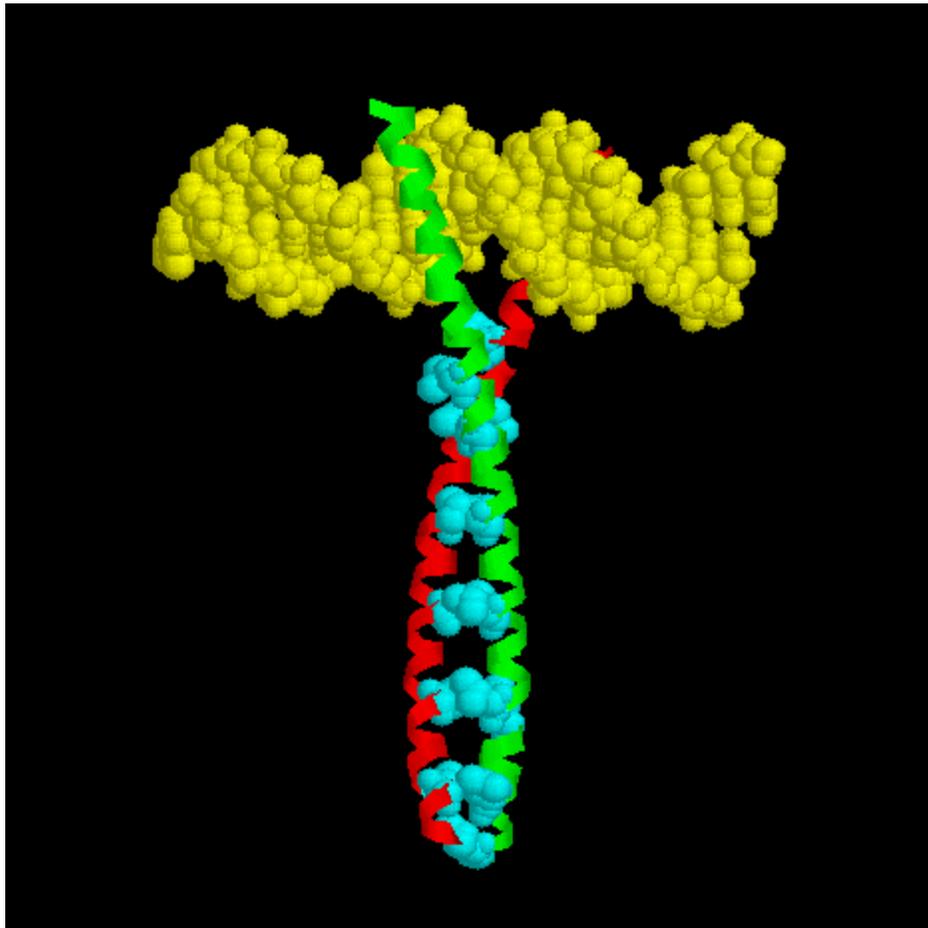

Figure 11.   c-Fos/c-Jun heterodimer complexed with DNA

The RHR or Rel Homology Region containing transcription factors represents another growing family of transcriptional modulators involved in cellular defense mechanisms or differentiation. The most well-known member of this family, NF-_B, was identified as a protein that binds to a specific DNA target site in the _ immunoglobulin light chain enhancer and that regulates expression of a variety of genes whose products are involved in the cellular response to infection. The RHR proteins bind to the _B sites as homo or heterodimers. The RHR of the two subunits of NF-_B (p50 and p65 or RelA) contains the RHR which specifies DNA binding, dimerization, nuclear localization and contains the binding site for the cytoplasmic inhibitor, I-_B. Crystallographic studies of the RHR in the p50 subunit of NF- B (Ghosh *et al.*, 1995) revealed the fascinating fact that both domains of the RHR (C-terminal and N-terminal) resemble the beta sandwich structure of the immunoglobulin folds of immunoglobulins, the expression of which they control. Both domains contact the DNA in the dimeric protein/DNA complex, and the C-terminal domain also forms the dimer interface.

Figure 12. Structure of the NF- B p50 homodimer bound to a   B site.

Another domain, termed the ankyrin repeat, implicated in protein-protein interactions, has been identified in well over 100 different proteins of very diverse function, such as

enzymes, enzyme inhibitors, transcription factors, and inhibitors of transcription factors such as I B, that inhibits the NF- B transcription factor in Figure 12 . Ankyrin repeats are units of around 32-33 amino acids first identified in the protein ankyrin and implicated in mediating protein-protein interactions. Most of the proteins that bear ankyrin contain at least 4 copies of this domain. Each ankyrin repeat is comprised of a beta strand, helix turn helix extended strand beta strand segment. The helices and beta strand are anti-parallel to each other and the plane of the strand regions is perpendicular to the helical axis (See Figure 13 below). The consensus sequence of the ankyrin repeat is:

```
- G – T P L H L A A R – G H V E V V K L L L D – G A D V N A – T K
      A     I   S Q  N N LD I A E V    K         N P D       D
            V     K      T M R       Q         S I         N
                                                E
```

The Gly residues at positions 2, 13 and 25 mediate sharp turns in between the secondary structural elements, whereas conserved small hydrophobic residues at positions 5,10,17,20 and 21 are involved in interactions between repeats. And the overall three-dimensional fold is maintained by interactions between the helices of the different repeat units.
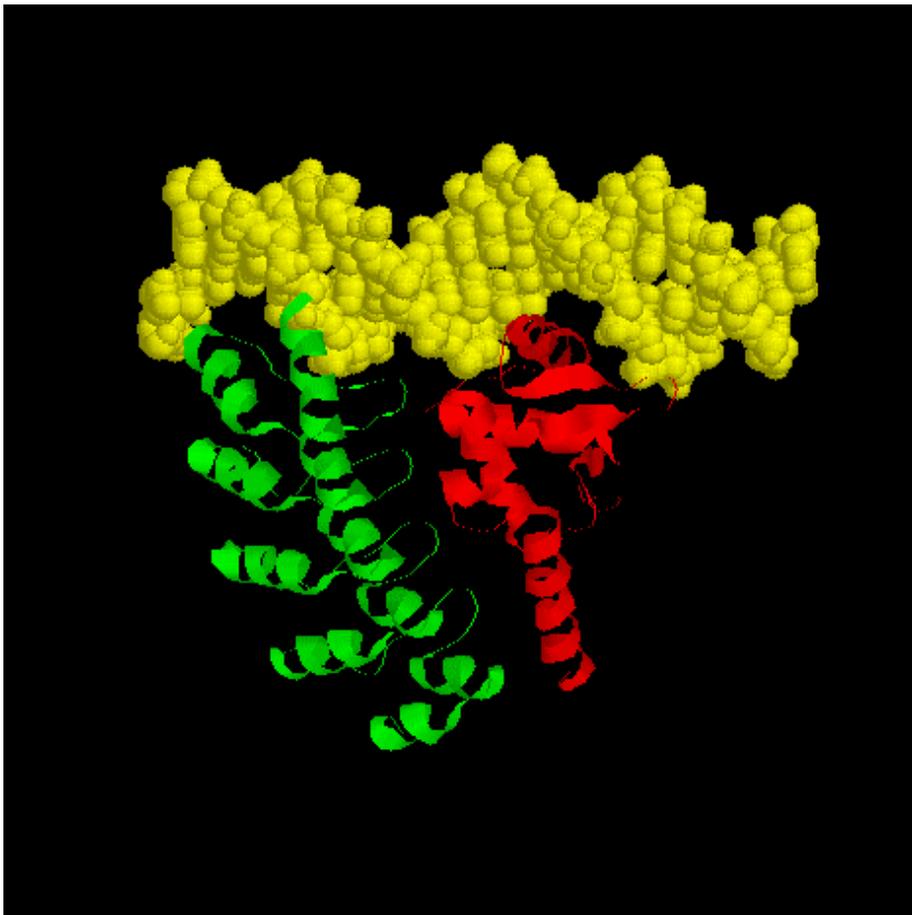

Figure 13. Structure of the GABP / heterodimer bound to DNA

In Figure 13 is shown the structure of a heterodimeric transcription factor , GABP / , bound to a 21 base pair target DNA oligonucleotide (Batchelor *et al.*, 1998). GA-binding protein (GABP) is involved in activation of nuclear genes encoding mitochondrial proteins and in viral gene expression. It is a heterodimeric transcription factor that binds to DNA sequences containing a CGA motif, and the both subunits are needed for optimal affinity. The   -subunit, colored in red, is that containing the ETS domain. ETS domain proteins have been implicated in developmental pathways and viral gene expression. The   -subunit, colored in green, contains 4 1/2 ankyrin repeats, comprised each of a pair of   -helices arranged in an anti-parallel coiled coil with an extended loop perpendicular to the coiled coil. The type I   turns at the end of each loop form a concave surface used to interact with the ETS domain subunit.

Another large family of ligand responsive transcription factors is embodied by the super-family of steroid/nuclear receptors. These proteins include the steroid hormone receptors, glucocorticoid (GR), estrogen (ER), mineralocorticoid (MR) and androgen receptors (AR), the lipophilic hormone receptors such as the vitamin D (VDR), retinoic acid (RAR and RXR) and thyroid TR hormone receptors, and a number of orphan receptors and transcription factors involved in Drosophila development. These receptors exhibit a homologous modular architecture with an N-terminal constitutive and tissue specific transcriptional activation domain (TAF1) which is quite variable in length and is responsible for interaction with tissue specific additional transcription factors, a central DNA binding and dimerization domain (DBD), which also may bear determinants for interactions with other transcriptional regulators, a linker region involved in nuclear localization and a C-terminal ligand binding domain (LBD) that also bears ligand dependent transcriptional activation activity (TAF2) and dimerization determinants. The protein-protein interactions operative in these systems are numerous and key to receptor function. Since may of the factors are implicated in human development and diseases (such as breast cancer among many others) they have become the targets for large scale drug development, many products of which are commonly used therapeutic agents. The importance of these receptors has thus spurred a strong interest in their structure/function relationships and the basis for specificity and potency in their action.

The three-dimensional structures of the DBD and LBD of a number of these receptors have been determined by crystallographic or NMR methods. They share a large degree of structural homology, even for the LBD, with lower sequence homology due to the necessity of recognizing different ligands. The steroid hormone receptors function as homodimers that bind to palindromic response elements (HRE's) in the DNA upstream of genes under their control. The spacing between the two half-sites of the palindrome is critical for cooperative binding of two monomers of the DBD's of these proteins, since the DBD fragments alone remain monomeric in solution up to millimolar concentrations. Full-length ER binds non-cooperatively, implicating pre-formed dimer in this interaction. The cooperativity of binding of the full-length versions of the other hormone receptors remains to be determined.
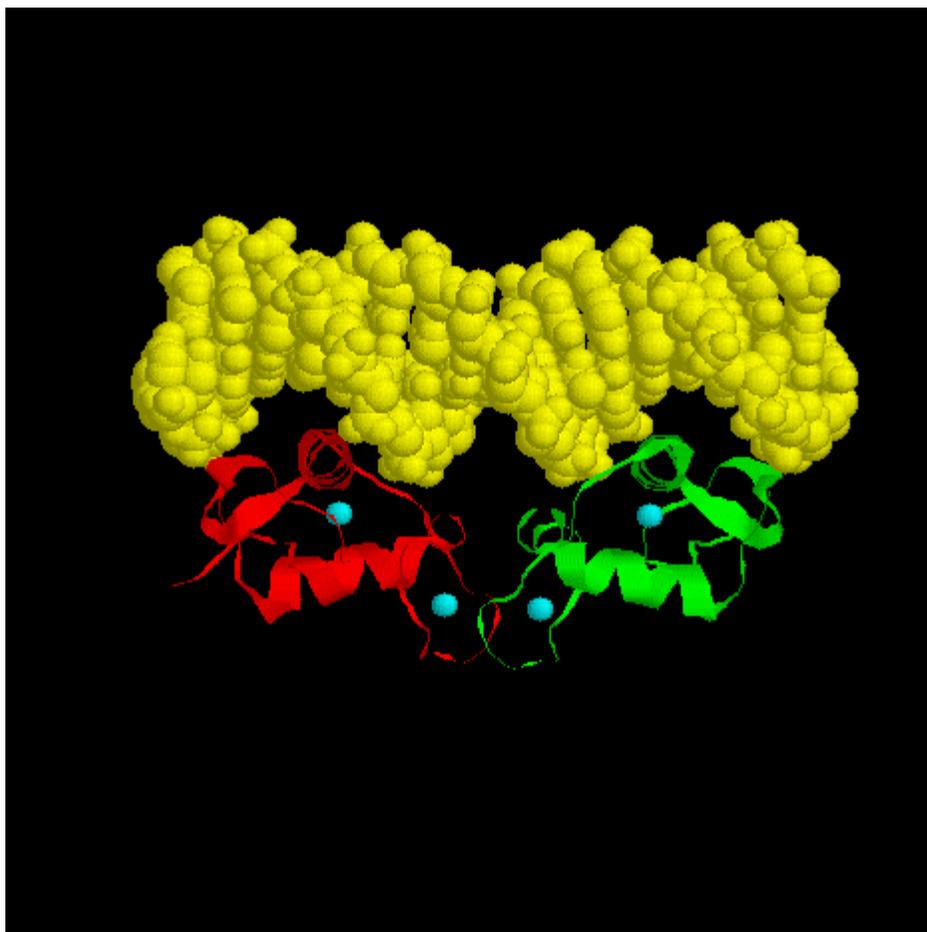

Figure 14. Structure of the ERDBD bound to DNA

The DNA Binding Domains or DBD of the nuclear receptor proteins are made up of a highly conserved pair of Zn fingers, one of which positions the DNA recognition helix, and the other bears the determinants for dimerization. In particular, the D-loop at the base of the second finger contains residues that participate in the dimer interface, flanked by two of the cysteine residues that coordinate the Zn atom. In Figure 14 is shown the crystallographically resolved structure of the DBD from the estrogen receptor (ER) bound to an oligonucleotide bearing the ERE (estrogen response element) sequence (Schwabe *et al.*, 1993). All four Zn atoms are represented in spacefilled cyan in Figure 1, and the disposition of the two Zn fingers of each monomer, one at the DNA interface and the other at the dimer interface is

clearly distinguished. The structure of the complex of the GRDBD bound to DNA is quite similar, except that only one of the monomers is in specific interaction mode due to the incorrect spacing in the target oligonuceotide (Luisi *et al.*, 1991).
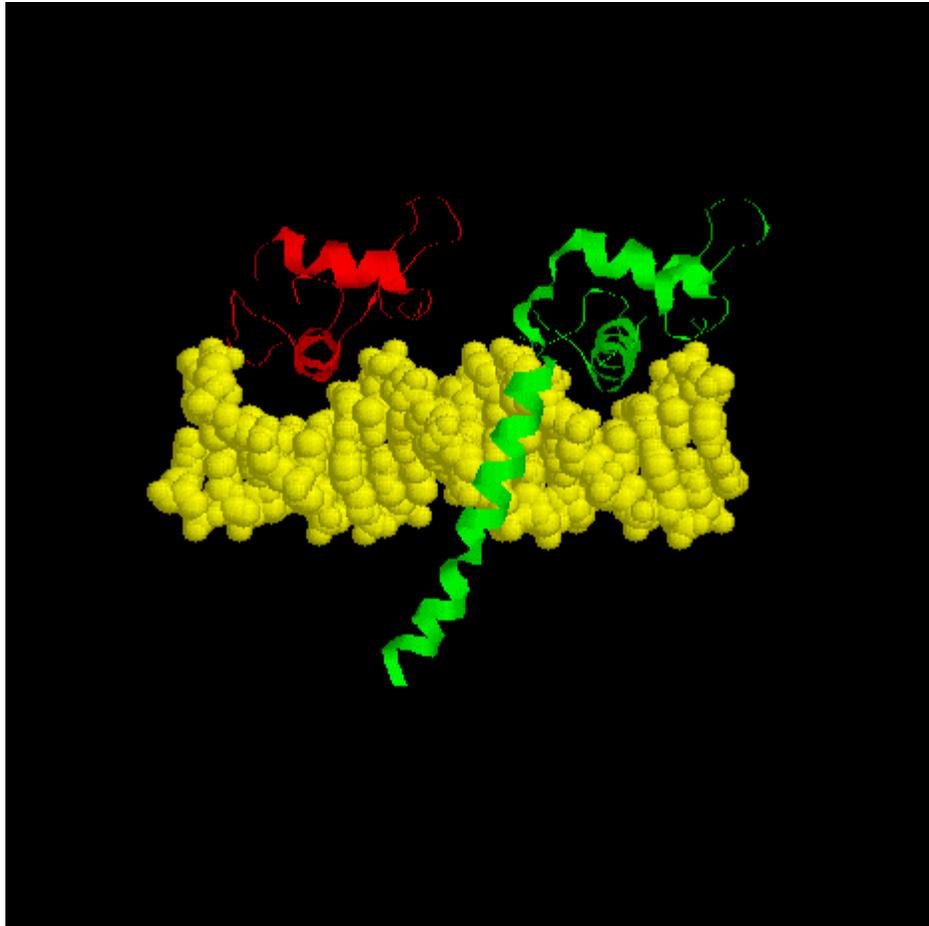

Figure 15. Structure of the RXR/TR DBD heterodimer bound to TRE bearing a 4 base pair spacer between two direct repeats.

The lipophicilic hormone receptors such as DR, RAR, and TR can form homo-dimers or heterodimers with RXR. In the case of heterodimer formation, the half sites are direct repeats and the spacing between half-sites (1-5 base pairs) determines the identity and polarity or orientation of the heterodimer. For example an RAR-RXR heterodimer will bind with the RAR subunit in the 5'-position if the spacing between the half-sites is 1 base pair, while the orientation is reversed for a spacing of 2 or 5 base pairs. Likewise the TR-VDR heterodimer will bind with the TR in the 5'-site if the spacing is 3 base pairs, and with the VDR in 5' if the spacing is 4 base pairs.   The structural basis for this orientational specificity of NR heterodimers resides in specific orientation of residues at the various dimerization interfaces that either provide key salt bridges and hydrophobic interactions for the correct orientations or, on the contrary, cause steric clashes for dimerization on in correct spacers. The structure of the heterodimer of the RXR/TR DBD bound to a tandem direct repeat TRE with a 4 base spacer (Rastinejad *et al.*, 1995) is shown in Figure 15. Contacts

between the two monomers of the heterodimer include a number of salt bridges that are found to cross the minor groove of the 4 base pair spacer element in regions that make phosphate contacts, implying direct coupling of dimerization to DNA binding, although the energetics of this coupling remain to be characterized.

The interactions between the N-terminal TAF1 of the nuclear receptors and heterologous transcription factors has not been well-characterized from a structural point of view. On the other hand, the basis for the ligand dependent function of the TAF2 is relatively well-understood with structures of LBD's free and bound by agonists and antagonists, and recently by a heterologous co-activator protein. As seen in the Figure below, the LBD's are
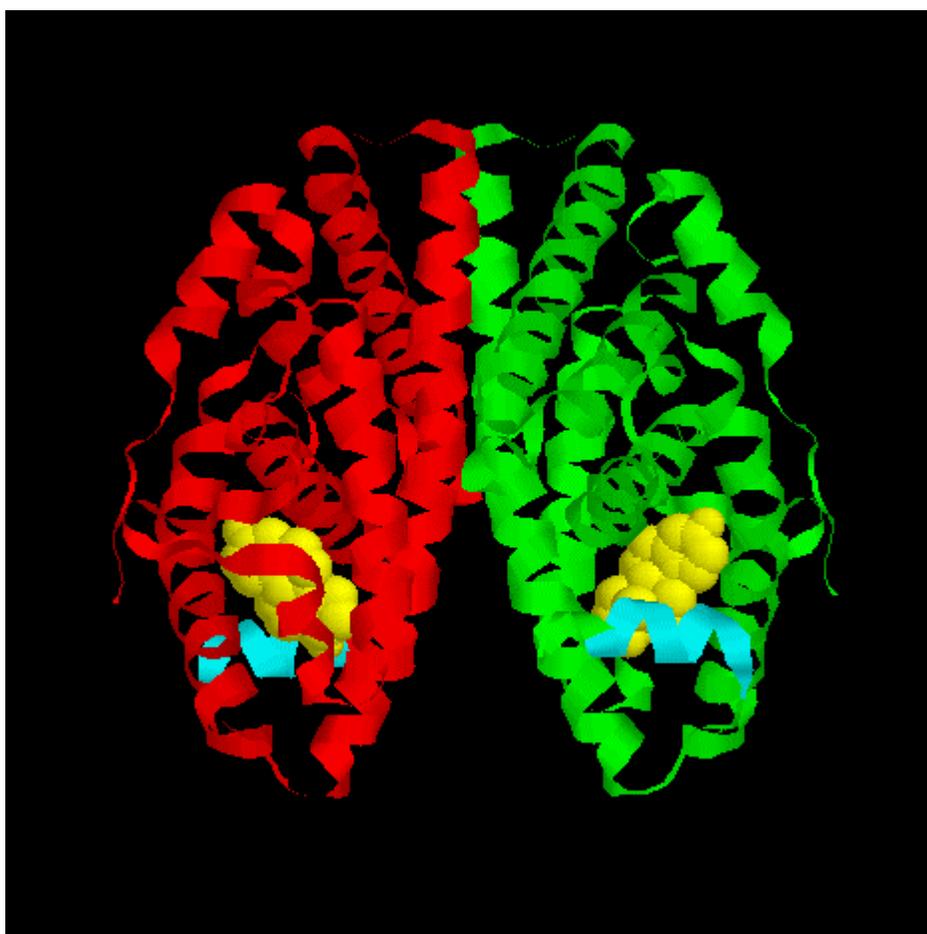
Figure 16. ERHBD dimer bound by estradiol.

dimeric mostly helical domains bearing a large pocket which presents the structural and energetic determinants that discriminate between ligands. It can be seen from the crystal structure of ERLBD bound by estradiol (Brzozowski *et al.*, 1997) that helix 12 (in cyan) folds back to cover the ligand binding pocket entrance in the core of the protein. This conformation results in shielding of the ligand and in the presentation of a protein interaction surface recognized by co-activator proteins bearing LXXLL motifs. In presence of antagonists the orientation of helix 12 is not optimal. In Figure 17, the crystal structure of the

ERHBD dimer bound by the antagonist raloxifen (Brzozowski *et al.*, 1997) is shown. The helix 12, bearing the recognition motif for co-activator binding cannot fold back over the binding site due to the bulky protrusion of the antagonist structure. In this case helix 12 is not well oriented to present the LXXLL motif and thus the co-activator proteins cannot be efficiently recruited to the complex.
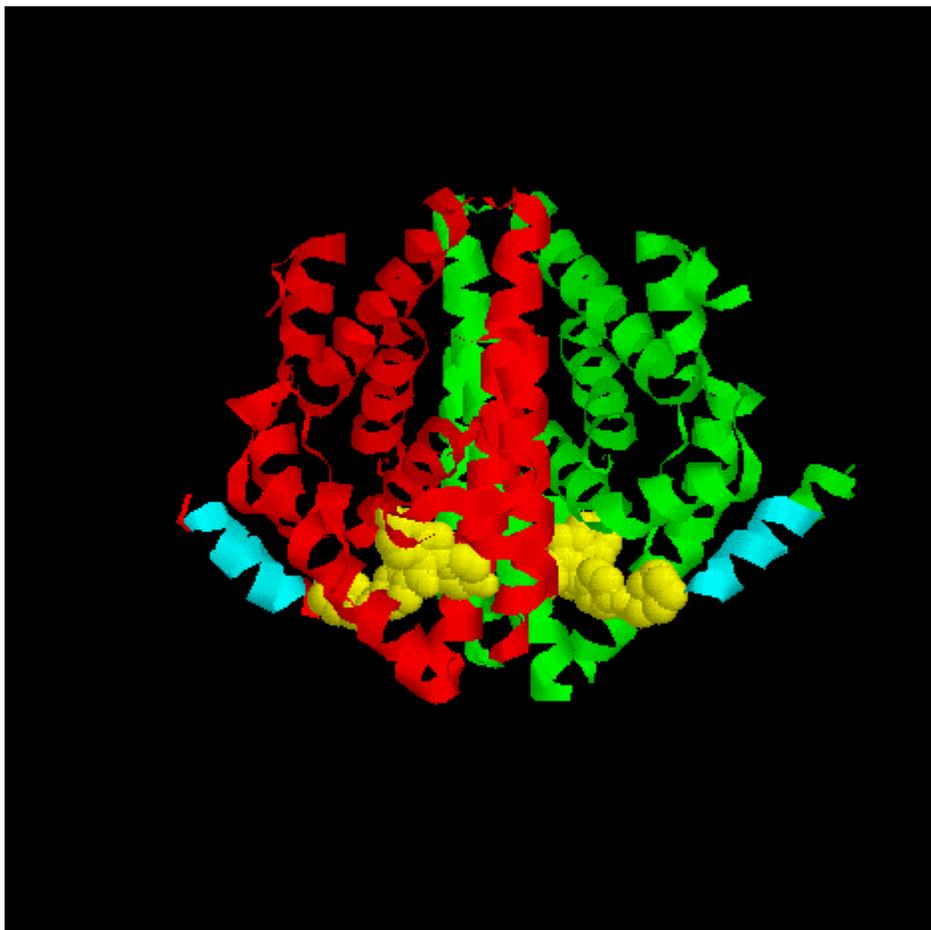
Figure 17. Structure of the ERHBD dimer bound by the antagonist raloxifen

These structural motifs represent a few examples of interesting and important protein interaction motifs. While progress is astounding in identifying by sequence homology members of various protein interaction motif families, the functional role of all the newly identified motifs is far from being determined. Moreover, structural genomics lags even farther behind functional genomics. The major question for biophysicists, that of the determinants for biological specificity within these large families will provide subjects for interrogation for many years to come, as very little solid quantitative information is available concerning the energetics of these interactions. The following sections of this chapter deal with methods for measuring the affinity between interacting protein chains, and how to analyze the data thus obtained.

Characterization of protein-protein interactions

Examples of the groups of highly homologous protein interaction domain families presented in the preceding section are found in large numbers of systems involved in the complex control of cellular function. Thus although within each family, the structural and sequence homology are quite high, enough differences are present within each sequence and structure to preclude inappropriate protein pairs from forming. This implies that the relative affinities between these various protein partners have been tuned throughout evolution such that protein pairing occurs only between particular partners, and only under the appropriate conditions. Likewise in homo-oligomeric proteins, the strength of the subunit interactions has evolved to be coupled to substrate or effector binding to just the right levels, such that enzyme activity proceeds at the necessary rate. While biochemists have successfully identified these protein interaction domains and demonstrated interaction preferences and by mutational studies, certain of the necessary sequence determinants of the interactions, in most cases, thorough quantitative thermodynamic and kinetic studies of protein interactions in these systems remain to be carried out. One reason for this lack of information is that priority has been given over the past few years to the identification of these important proteins and to the determination of their three-dimensional structure, both of which are pre-requisites to a full understanding of function. Nonetheless, understanding the mechanisms underlying the function of these proteins likewise requires the characterization of their energetic and dynamic properties.   We must bear in mind that the study of the energetics and kinetics of protein subunit interactions presents certain practical experimental difficulties that will be discussed below. The good news here is that there are still plenty of interesting and important studies to be undertaken that should keep biophysicists busy for many years to come.

The first major stumbling block encountered by the experimentalist interested in characterizing protein-protein interactions is that of having access to enough purified stable protein, and if possible some interesting constructs, homologous family members or functional mutants in order to carry out a thorough comparative structure/function study. In cases where the NMR or crystal structure of the protein has been determined, this first obstacle can likely be overcome, since the production procedure has been worked out in such cases. The second issue that must be addressed is that of the observable. Given either a homo or hetero-oligomer, how will its degree of oligomerization state be assessed under different conditions? In this section a number of approaches will be discussed briefly, touching on their advantages and disadvantages. These include analytical ultracentrifugation coupled to some sort of optical detector, dynamic light scattering, fluorescence (anisotropy or intensity), surface plasmon resonance and iso-thermal titration calorimetry. Many of these techniques are presented in much more detail in specific volumes of this on-line text dedicated to particular techniques. As a general rule, the characterization will be all the more solid, the more complementary techniques are employed, since there are perturbations and uncertainties associated with all of them. The experimentalist is limited, moreover, by the concentration range of the association, relative to the sensitivity of the various techniques. Characterization of protein interactions for homo-oligomers requires a slightly different experimental approach that the study of hetero-oligomerization. It is rare that only one of

these approaches can suffice to provide all of the necessary information, such that it is a good idea to become familiar with several of these techniques. Below is given a very short overview of their usefulness and some of their limitations. The reader is encouraged to consult the appropriate accompanying chapters and volumes of this on-line text for an in-depth discussion.

Overview of biophysical techniques used to study protein-protein interactions

Analytical ultracentrifugation (AUC)

AUC is becoming an increasingly popular tool in the study of protein complexes. After having long been relegated to a very few laboratories with the good fortune of possessing the expertise (mechanical as well as theoretical), necessary for its function, AUC is making a strong comeback. This revival is based primarily on the introduction of a much more user friendly analytical ultracentrifuge, the XLA, by Beckman instruments to replace the venerable Model E. The introduction of the XLA coincided with growing interest in examining protein interactions due to the identification, purification and structural characterization of protein interaction domains involved in key aspects of a large number of human diseases. Thus, studies of protein interactions by AUC are becoming almost routine. For an in-depth discussion of this technique as applied to interactions between bio-molecules in general, the reader is referred to the corresponding chapter of this volume. Suffice it to say here that this technique can be applied in sedimentation equilibrium or sedimentation velocity mode, and that the tools for analysis of complex equilibria are much improved. Sedimentation velocity is a much more rapid experiment, whereas equilibrium sedimentation is more accurate. In sedimentation velocity, the speed of sedimentation is measured by determining the rate of movement of the sedimenting boundary. Usually , either the integral distribution of s, the sedimentation coefficient, or its differential distribution is calculated rather than relying on a single boundary point, to take into consideration boundary spreading due to diffusion, complex systems and non-idealities. In sedimentation equilibrium studies the speed of the centrifuge is low, such that the transport by sedimentation is balanced by diffusion. In this case an equilibrium concentration distribution is established along the cell, i.e., the flux of concentration at any point n the cell must be zero. Then the gradient of concentration *vs* tube distance is used to calculate the molecular weight, since the concentration distribution is exponentially dependent upon the buoyant mass of the protein. Profiles obtained using mixtures of oligomeric states at equilibrium can be analyzed for the equilibrium constant over the appropriate concentration range. Detection is based on absorption, schlieren refractive index optics or on fluorescence, with the relative sensitivity increasing in that order.   The major drawback of AUC lies in its limited sensitivity. Protein-protein interactions span from the upper micromolar range to picomolar, and AUC is really only applicable down to around 0.1 μM. With protein oligomers labeled with dyes absorbing (or fluorescing) at different wavelengths, hetero-oligomers can be distinguished from homo-oligomers for example. Readers interested in pursuing the use of analytical ultracentrifugation should consult the web site for the National Analytical Ultracentrifugation Facility in Storrs Connecticut at the following address:

http://gopher.uconn.edu/~wwwbiotc/UAF.html.

The National Facility organizes workshops and courses that would be very beneficial for those considering the use of this technique.

Light scattering

Static and dynamic light scattering represent another approach to studying protein complexes. In static light scattering, the scattering intensity is related to the molecular weight of the protein, in addition to its concentration, the scattering angle, and the wavelength which in principle are known. Dynamic light scattering is based on the auto-correlation of the time-dependent fluctuations of scattered light intensity, which in turn depends upon the diffusion constant. This auto-correlation decays more slowly for slowly diffusing particles and thus, the diffusion constant is extracted from the value of the relaxation time of this function. In the case of ideal spherical particles, this provides a measure of the molecular weight. Light scattering, like AUC, is limited principally by sensitivity, with best results around 1 mg/ml, depending upon the size of the protein or the complex. While sensitivity limits preclude the determination of affinities and association or dissociation rate constants in most cases, light scattering is quite useful in characterizing the stoichiometry of complexes at high concentration. This is very important information for the analysis of data obtained by more sensitive techniques. In fact, it should be pointed out that in many cases, biochemical methods provide the information that protein A interacts with protein B, but the stoichiometry of this interaction is often elusive. Even when crystal structures are available, the stoichiometry of the complex in the crystal may not correspond to that observed in solution under various conditions. Thus, determination of the end point stoichiometry , deduced from the molecular weight determination by light scattering is a good idea.

Fluorescence Spectroscopy

There are a number of approaches based on fluorescence that can be used in the study of protein-protein interactions. Again, this on-line text includes an entire volume on spectroscopic methods, including fluorescence, to which the reader is referred for a detailed discussion of the theory and practice. The most widely used approach in the study of protein-protein interactions is that based on the polarization or anisotropy of fluorescence. This value represents the degree to which an excited fluorophore retains the orientation of the exciting light in its emission Small molecules rotate rapidly and thus retain little of this orientation or polarization, whereas large molecules rotate more slowly in solution, and thus retain a higher polarization. Fluorescence can be carried out in static or in time-resolved modes. In the static mode, the polarization measured reflects loss of orientation by global rotation of the protein molecule to which the fluorophore is attached, in addition to loss of orientation by local motions of the fluorophore. Thus, the absolute value of the polarization measured in the static mode is not particularly informative. However, in the case of an equilibrium between associating protein subunits, this value decreases upon dilution of a homo-oligomer or increases upon titration of the labeled protein A with its protein partner B if this results in formation of a hetero-oligomer. In this manner the equilibrium binding constant, and also the kinetics of the interaction can be studied. The sensitivity of fluorescence is quite high, the limit being in the low nanomolar/high picomolar range, depending upon the probe used, such that interactions of very high affinity can be studied. The drawback is that only in cases of low affinity can the intrinsic tryptophan of the protein

be used, since it can only be detected down to about 100 nM in the best of cases. Thus, for high affinity interactions, the protein must be labeled with an extrinsic fluorescence probe. Care must be taken to control for function after labeling since the covalently bound probe can conceivably perturb the interaction under study. Time resolved fluorescence of the intrinsic tryptophan residues or of an external probe can separate the global from the local motions and provide information about the stoichiometry of the complex at micromolar concentrations that can then be used to analyze titration curves carried out at lower concentrations. An additional advantage of fluorescence techniques is that they are readily adapted to kinetics experiments (stopped flow or simple mixing).

In addition to fluorescence polarization, fluorescence resonance energy transfer between probes on different subunits can be used to study their interactions. The biochemistry involved in these experiments is somewhat more complex (1:1 labeling required for 2 probes), and the sensitivity somewhat more limited. Fluorescence energy transfer is based on the dipolar interaction of an excited donor fluorophore that will transfer its excited state energy to an acceptor fluorophore provided that the distance between donor and acceptor is not too large and also provided a strong spectral overlap between the emission of the donor and the absorption of the acceptor. Since this is a dipolar interaction, the probability of transfers goes as $1/R^6$, where R is the distance between the two fluorophores. The transfer efficiency is also related to the lifetime of the donor. The longer the donor remains in the excited state, the higher the probability that transfer will occur. The spectral overlap, the relative orientation of the two probes, the refractive index of the medium and the quantum yield of the donor in absence of the acceptor all contribute to the value of the characteristic transfer distance Ro, the distance at which 50% transfer is achieved. Typical Ro values range from 10 to 50 Å, allowing detection of transfer at inter-probe distances up to about 80 Å. Recent applications using long lived luminescing lanthanide probes as donors allow for energy transfer at significantly longer distances. Fluorescence resonance energy transfer or FRET involves therefore a large number of parameters, including the elusive orientation factor (typically, but surely not appropriately assumed to correspond to free rotation of both donor and acceptor). Despite the apparent complexity of FRET, it has proven to be a surprisingly robust technique in biophysical applications, even to the point of very accurate distance measurements in a significant number of biomolecules, later confirmed from crystallographic coordinates. Moreover, if FRET is used to detect interactions, rather than to measure accurate distances (i.e., for structural information about complexes) then much of the complexity of the technique does not come into play. The most serious limitations of the technique lie in the necessity of labeling both interacting macromolecules, and in the size of the complex that can be monitored given the distance constraints based on overlap and lifetime.

We also note here that fluorescence correlation spectroscopy (FCS) (based on principles similar to that of dynamic light scattering) is also now being used to determine the diffusion constant (and thus the size) of single protein molecules, and to study dissociation of oligomers. This technique requires relatively more sophisticated equipment (pulsed laser (preferably 2-photon) excitation, microscope), than does steady state fluorescence anisotropy

or steady-state FRET. The advantage is that such interactions may eventually be studied *in situ*.

In the United States, two national centers for fluorescence applications exist, one at the University of Maryland and the other at the University of Illinois. The web site addresses of these two centers are given below. They offer user facilities, instrument development, courses and software, for those interested in the application of fluorescence techniques.

http://cfs.umbi.umd.edu/
http://web.physics.uiuc.edu/Research/Fluorescence/index.html

The interested reader is also referred to the volume of this on-line text dedicated to fluorescence theory and practice. IT is also of note that there is a Biological Fluorescence subgroup of the Biophysical Society.

Surface Plasmon Resonance
A technique that been the object of a great deal of interest and activity for studying biomolecular interactions is that of surface plasmon resonance (SPR). This technique is based on the excitation of surface plasmons of a gold layer by the light waves of a laser beam impinging on the gold surface. In fact, the reflectivity of the laser beam drops to near zero at a particular angle, termed the surface plasmon resonance angle, due to this phenomenon. The intensity of the optical fields in these surface waves decays exponentially with distance from the surface over about 100 nm, and the value of the SPR angle is therefore very sensitive to the refractive index of the medium adjacent to the surface. SPR gold chips are first derivatized and then one partner of a protein pair is covalently linked to the surface. Buffer solution bearing the other partner flows over the chip and the changes in the SPR angle are monitored. If the protein in the flow solution binds to the immobilized partner protein, the refractive index near the surface changes, leading to a change in the SPR angle. This technique has proven very useful in demonstrating interactions between particular partners. One must bear in mind however, that the situation deviates slightly from solution thermodynamics, and is more akin to the interaction of soluble molecules with membrane components, due to the surface set-up. Secondly, since the measurements are in fact made under flow conditions, the on and off rates derived from the profiles upon addition or deletion of the binding factor from the flow buffer are used to calculate the affinities. These calculations are usually valid for simple systems, provided there are no non-idealities due to transport effects. The situation is a bit more complicated when the interactions, themselves, are complex. Nonetheless, the ease of use and adaptability to large scale screening situations has contributed to the growing popularity of SPR in the study of biomolecular interactions.

Calorimetry
Iso-thermal titration calorimetry represents a very powerful tool for the characterization of protein-protein interactions. This technique is based on the relatively straightforward measurement of the heat absorbed or released upon titration of a solution of

molecule A with that of molecule B. If the interactions are in the appropriate concentration range, then ITC experiments yield both the enthalpy and the free energy changes of the reaction. Moreover, using the Gibbs fuction, one can calculate the entropy change. Since protein-protein interaction free energies are usually the result of strong enthalpy-entropy compensations, a thorough understanding of the energetics truly requires the knowledge of all of these thermodynamic parameters. The experiment is done by sequential additions of the titrant to a cell containing a solution of reactant. The heat change associated with each successive addition of titrant is measured using thermopiles whose signals are proportional to the rate of heat transfer from the titration cell to the heat sink, which integrated over time yields the total heat of reaction. The ITC is carried out at constant temperature. However, if the experiment is repeated at multiple temperatures one can obtain the change in the enthalpy change of reaction with respect to the temperature. This value corresponds to the change in heat capacity that occurs upon complexation. Upon complex formation between large biomolecules, a significant amount of hydrophobic surface area is removed from contact with water, which in turn usually results in a large decrease in the heat capacity. Thus, determination of the heat capacity change can provide information about the change in exposed surface area upon complex formation. The only drawback to ITC is its relatively low sensitivity. The lower limit of dissociation constants that can reasonably be determined is about 10 nM, and this under the best of conditions (large heats of reaction, etc.). Nonetheless, one can always obtain values for the enthalpy change of the reaction, as well as the heat capacity change at concentrations much higher than the dissociation constant. If this latter can be obtained under equilibrium binding conditions (i.e. [reaction] << K]) using another more sensitive technique such as fluorescence anisotropy, then all of the thermodynamic parameters can be obtained. One must take care to ensure that the stoïchiometry of the interaction under the concentration conditions of the ITC experiments is identical to that observed under equilibrium binding conditions. The interested reader is referred to the volume of this on-line text specifically dedicated to calorimetric methods for further information. A national center for biocalorimetry exists at the Johns Hopkins University with the following web site address:

http://biocal2.bio.jhu.edu/

Like the national centers for analytical ultracentrifugation and fluorescence, this center is involved in user resources and instrumentation development, offering in addition courses and software.

Energetics of Protein-Protein Interactions

The techniques outlined above will provide the experimentalist with binding or dissociation profiles that must be analyzed using some sort of model in order to extract the thermodynamic information of interest. For a simple dissociating dimer, A2 $\rightleftharpoons$ A, or a simple interaction between two heterologous proteins, A + B $\rightleftharpoons$ AB, the situation is quite straightforward and little need be discussed along these lines. However, the interacting protein systems of wide-spread interest are rarely limited to such straightforward

considerations. Typically the subunit affinity is linked to the degree of ligation or the competition between two different proteins for binding to a third. Alternatively several different proteins can interact with multiple DNA sites in response to various ligands. These sorts of systems are more typical of the issues currently of interest to biophysicists, and imply a much different level of complexity.

For example, returning to the well-known example of the hemoglobin tetramer discussed briefly in the first section of this chapter, ligation is coupled to subunit affinity in very subtle manners, in order to bind oxygen in the lungs and deliver it to muscle. An in-depth treatment of the linkages between ligation and subunit affinity is provided in another chapter in this text. Below is given a simple example of how the cooperative binding of ligand to a dimer can drive dimerization, or inversely how increasing protein concentration can drive ligation. It can be seen from the free energy diagram in Figure 18 that the binding of the ligand to the monomer, $G11$, is less favorable than that of binding of the first ligand to the dimer, $G21$, and that binding of the second ligand to the dimer, $G12$, is more favorable than binding the first (i.e., binding is cooperative to the dimer). In this case, ligand binding also favors dimer formation ($Gd0 <$ $Gd1 <$ $Gd2$, in absolute value), whereas the opposite situation arises in the case of hemoglobin. In fact many different linkage scenarios are possible. The situation is even more complex in cases where more than one ligand binds or where there are multiple protein partners in the complex. Ideally in order to obtain detailed structure-function relations one wants to be able to resolve all of the different binding free energies and hence cooperativity values in such systems and to compare the values obtained under different conditions and as a function of changes in the sequence by comparing the behavior observed for homologous systems or site-specific mutants.

In order to resolve the parameters of interest for systems involving many different components that can form complexes of various stoichiometric combinations of these components, the observables obtained from the experiment must provide data a sufficient quality, but also of sufficient complexity, i.e., the data must first of all contain the information about which species are populated under which conditions. This means that the experiments must be designed in order to resolve complex binding interactions before complex analysis approaches are used to fit the data. This warning may seem a bit obvious, but it is not always easy to devise approaches for observing as many different species as possible. As mentioned in the previous section, multiple techniques are usually required.

In-depth derivations of analytical expressions that can be used to fit titration cures, (fractional saturation of A by B) derived from some experimental observable for various cases of energetic linkages between ligation and protein-protein associations can be found in an accompanying chapter. These expressions are used in the context of non-linear least squares fitting routines in order to obtain best estimates of the various binding affinities involved in the model. A general word of caution is warranted. These analytical approaches are only so valid as the model used. If in reality this model (various species populated) does not correspond to what actually occurs in the test tube, cuvette, centrifuge cell, etc. then the numbers obtained from such fits have little meaning. Thus it is paramount that the validity

G11   Gd0

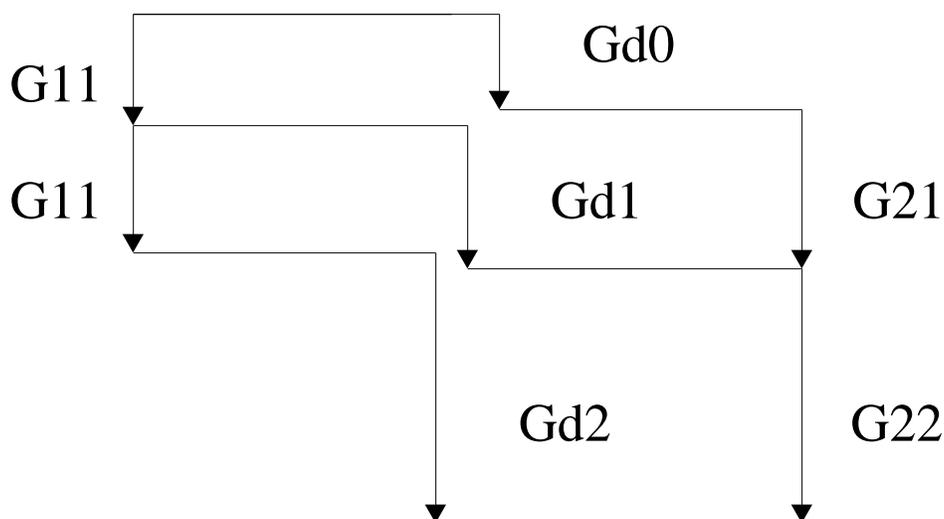G11   Gd1   G21

Gd2   G22

Figure 18. Free energy diagram of a system in which the binding of ligand is cooperative and more favorable to the dimer than to the monomer. In this case, ligand binding promotes dimerization.

of the model be demonstrated (i.e., protein A and protein B do actually form a hetero-tetramer in which only protein A is bound by ligand X, etc.) This means that to the extent possible, the populated species must be clearly identified. This is not always an easy task, as mentioned above. Moreover, in very complex cases involving multiple oligomeric states, it is no longer possible to derive an analytical expression which contains only one independent parameter (i.e., free ligand concentration). In such cases, it is simplest to move to an analysis methodology based on numerical, rather than analytical solutions to the binding equations. Such approaches have been discussed in depth, elsewhere (Royer & Beechem, 1992). One more word of caution. In fitting macromolecular association data, regardless of whether the solver used is analytical or numerical, one should always try to directly fit the raw data, rather than transforming the values (radioactive counts, fluorescence anisotropy, etc.) into some percentage bound. This allows for the asymptotic values of the observable to be floating parameters in the fit, leading to fewer artefacts in the estimation of subtle quantities such as linkage free energies.

Overall, this chapter has tried to stress the idea that a thorough characterization of protein-protein interactions is necessary for a thorough understanding of cell function. This thorough characterization includes, but is by no means limited to, the determination of the three-dimensional structures of these molecules. The examples of the structures of a few protein interaction motifs given in the first section of this chapter provides an idea of the beauty and diversity of protein structure. In addition to structural considerations, the dynamic and energetic properties of these systems reveal the exquisite subtlety involved in biological specificity and control.

LITERATURE CITED

Batchelor, A.H., Piper, D. E., Charles de la Brousse, F., McKnight, S.L. 7 Wollberger, C. (1998) *Science* **279**, 1037-1041.

Brzozowski, A.M., Pike, A.C.W., Dauter, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustaffson, J.-A. 7 Carlquist, M. (1997) *Nature* **389**, 753-758.

Doyle A. D., Lee, A., Lewis, J., Kim, E. Sheng, M., MacKinnon, R. (1996) *Cell* **85**, 1067-1076.

Ghosh, G., Van Duyne, G., Ghosh, S. 7 Sigler, P.B. (1995) *Nature* **373**, 303-310.

Glover, J.M.N. & Harrison, S. C. (1005) *Nature* **373**, 257-261.

Hammarstrom, A., Berndt, K.D., Sillard, R., Adermann, K. & Otting, G. (1996) *Biochemistry* **35**, 12723-12732.

Ma, P. C. M., Rould, M.A., Weintraub, H. & Pabo, C. O. (1994) *Cell* **77**, 451-459.

Narula, SS., Yuan, R.W., Adams, S.E., Green, O.M., Philips, T.B., Zydowsky, L.D., Botfield, M.C., Hatada, M., Laird, E.R., Zoller, M.J., Karas, J.L. 7 Dalgarno, D.C. (1995), *Structure* **15** 1061-1073.

Rastinejad, F., Perlmann, T., Evans, R.M. & Sigler, P. (1995) *Nature* **375**, 203-211.

Renzoni, D.A., Pugh, D.J.R., Siligardi, G., Das, P., Morton, C.J., Rossi, C., Waterfield, M.D., Campbell, I.D. & Ladbury, J.E., (1996) *Biochemistry* **36**, 15646-15653.

Royer, C. A. and Beechem, J.M. (1992). *Methods in Enzymology,* **210**, 481-497.

Schwabe, J.W.R., Chapman, L., Finch, J.T. & Rhodes, D. (1993) *Cell* **75**, 567-578.

Tame, J. and Vallone, B. entry 1a3n of the Brookhaven Protein Data Base, Deoxyhuman hemoglobin, deposited April 29, 1998.

Tong, L., Warren, T.C., Lukas, S., Schembri-King, J., Betageri, R., Proudfoot, J.R. & Jakes, S. (1998) *J. Biol. Chem.* **273**, 20238-20242.

Zhang, Z., Lee, C.-H., Mandiyan, V., Borg, J.-P., Margolis, B., Schlessinger, J., Kuriyan, J. (1997) *EMBO J.* **16**, 6141-6150.