Transcription factors III

- Major families of transcription factors and their functions
  - zinc finger genes
    - nuclear hormone receptors
  - helix-turn-helix
    - homeobox genes
  - helix-loop-helix
    - myogenic genes
  - bZIP proteins

- Additional reading
  - Blumberg and Evans (1998) Genes and Development 12, 3149-3155

- Last year’s final exam is now posted.
  - I will post answers in a couple of weeks after you have had time to work through the questions
Regulating transcription factor activity (contd)

Figure 21.2 The activity of a regulatory transcription factor may be controlled by synthesis of protein, covalent modification of protein, ligand binding, or binding of inhibitors that sequester the protein or affect its ability to bind to DNA.

<table>
<thead>
<tr>
<th>Inactive Condition</th>
<th>Active Condition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein synthesized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ne protein</td>
<td></td>
<td>Homeoproteins</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein phosphorylated</td>
<td></td>
<td>HSTF</td>
</tr>
<tr>
<td>Inactive protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein dephosphorylated</td>
<td></td>
<td>β-catenin/</td>
</tr>
<tr>
<td>Inactive protein</td>
<td></td>
<td>armadillo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ligand binding</td>
<td></td>
<td>Steroid receptors</td>
</tr>
<tr>
<td>Inactive protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Steroid response</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleavage to release active factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Membrane-bound protein</td>
<td></td>
<td>Steroid response</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Release by inhibitor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inhibitor</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change of partner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive partner</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Regulating transcription factor activity (contd)

- How can the activity of a transcription factor be restricted to a particular cell type or time?
  - Factor is not generally present but synthesized only where it is needed
    - some developmental regulators
  - The factor is present but must be modified to be active
    - heat shock factors - phosphorylated
    - β-catenin/armadillo - dephosphorylated
  - A ligand is required for activity (or inactivity)
    - nuclear hormone receptors
  - The factor is localized to an inactive compartment (e.g. cell membrane) and required cleavage for activity
    - sterol response factors (primarily cholesterol)
  - The factor may be bound to an inhibitory factor in the cytoplasm
    - NF-κB and I-κB
  - A dimeric factor can have multiple partners. Which partner is present determines activity
    - some dimers are active
    - others are inactive
    - eg bHLH and bZip proteins
Zinc finger genes

- Zinc fingers are found in a variety of transcription factors
  - two basic types
    - Cys-His, consensus sequence is
      \[ \text{cys-X}_{2-4}-\text{cys-X}_3-\text{phe-X}_5-\text{leu-X}_2-\text{his-X}_3-\text{his} \]
      - typical gene has 3 or more fingers
      - found in factors for Pol II and Pol III
    - Cys-Cys, consensus sequence is
      \[ \text{cys-X}_2-\text{cys-X}_1-\text{cys-X}_2-\text{cys} \]
      - typical gene has only 2 fingers
      - found in steroid hormone receptor superfamily members
  - may be involved in both DNA and RNA binding, presence of finger does not indicate which
    - eg TFIIIA binds DNA and RNA product
    - eIF2β recognizes translational initiation sites
Zinc finger genes (contd)

- purpose of fingers is to arrange residues such that zn ions can be coordinated
  - fingers may form $\alpha$-helical structures that fit into the major groove of the DNA helix
  - multiple fingers may act cooperatively to bind nucleic acids
Zinc finger genes (contd)

- cys-cys fingers in nuclear receptors
  - only 1st finger binds to DNA
  - second finger is responsible for protein:protein interactions
  - spacing between fingers can vary quite a bit
- finger 1 contains a region that determines target specificity - P-box
  - CGSCKA - AGAACA
  - CEGCKA - AGTTCA
  - these can be swapped and change specificity of the receptor
- used in ecdysone-inducible system
Hormonal signaling pathways

- Hormones are chemical messengers that coordinate cellular activity
- Can act in different ways
  - endocrine - on distant cells
  - paracrine - on neighboring cells
  - autocrine - on cells which secrete them
- Active at very low concentrations - typically less than 1 ppb (1 ppb ~ 3 nM)
- Involved in numerous biological processes - many hundreds of hormones
  - reproduction - estrogen, testosterone, progesterone, FSH, LH, activin
  - metabolic rate - thyroid hormone, TSH, GH
  - stress - glucocorticoids, ACTH, CRF
  - blood pressure - aldosterone, renin, angiotensin, vasopressin
  - calcium homeostasis - vitamin D3, calcitonin, PH
- Some vitamins or vitamin derivatives are hormones
  - Vitamin A
    - all-trans-retinoic acid
    - 9-cis-retinoic acid
    - 14-OH-retroretinol
  - Vitamin D3
Nuclear hormone receptors

DNA  LIGAND

- Domains are assortable and transferable
- DNA-binding domain (DBD)
  - responsible for direct binding to DNA
  - discriminates half site sequence
  - determines spacing between half sites
  - contains an important dimerization motif
- Ligand binding domain (LBD)
  - responsible for ligand binding
  - has a general dimerization motif
  - contains an important transactivation domain
  - may interact with amino terminus to modulate activation
- amino terminal region (A/B domain)
  - contains an activation domain in many receptors
  - may interact with other components of the transcriptional machinery
  - many receptors have alternative splicing or promoter usage to yield different A/B domains
- linker region (D) may influence activation, repression, nuclear translocation or DNA-binding
Nuclear hormone receptors (contd)

- bind to specific target DNA sequences
- activate transcription of target genes upon ligand binding
- function at very low levels of ligand (~$10^{-9}$M or ~ ppb)
- bind to small (~300d) lipophilic molecules
  - steroids
  - retinoids
  - thyroid hormone
  - vitamin D3
Nuclear hormone receptors (contd)

- Many receptor ligands are related to cholesterol
  - steroids
  - bile acids
  - oxysterols
  - Vitamin D3
  - ecdysone
- can move freely through tissues
  - penetrate to a target
  - diffuse from a source

![Chemical structures](image)
Nuclear hormone receptors (contd)

The Nuclear Receptor Superfamily ca 1992

<table>
<thead>
<tr>
<th>Classical receptors (from biochemistry)</th>
<th>ORPHAN RECEPTORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GR</td>
<td>cortisol</td>
</tr>
<tr>
<td>MR</td>
<td>aldosterone</td>
</tr>
<tr>
<td>PR</td>
<td>progesterone</td>
</tr>
<tr>
<td>AR</td>
<td>testosterone</td>
</tr>
<tr>
<td>ERα</td>
<td>estrogen</td>
</tr>
<tr>
<td>VDR</td>
<td>1.25 (OH)₂ vit D3</td>
</tr>
<tr>
<td>TRα,β</td>
<td>Thyroid hormone</td>
</tr>
<tr>
<td>EcR</td>
<td>20-OH ecdysone</td>
</tr>
<tr>
<td>RARα,β,γ</td>
<td>all-trans RA</td>
</tr>
<tr>
<td>RXRα,β,γ</td>
<td>9-cis RA, ?</td>
</tr>
</tbody>
</table>

- more orphan than known receptors
- why study orphan receptors (not particularly easy)
  - novel Signaling Pathways
  - new Developmental Hormones
  - target gene networks
  - potential Teratogens
  - roles in adult physiology and endocrinology
  - cancer treatment
Nuclear hormone receptors (contd)

**The Nuclear Receptor Superfamily - 2000**

<table>
<thead>
<tr>
<th>DNA</th>
<th>LIGAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNOWN RECEPTORS</td>
<td>ORPHAN RECEPTORS</td>
</tr>
<tr>
<td>Classical receptors (from biochemistry)</td>
<td>Vertebrate</td>
</tr>
<tr>
<td>GR</td>
<td>cortisol</td>
</tr>
<tr>
<td>MR</td>
<td>aldosterone</td>
</tr>
<tr>
<td>PRα, β</td>
<td>progesterone</td>
</tr>
<tr>
<td>ARα, β</td>
<td>testosterone</td>
</tr>
<tr>
<td>ERα, β, γ</td>
<td>estrogen</td>
</tr>
<tr>
<td>VDR</td>
<td>1.25 (OH)₂ vit D3</td>
</tr>
<tr>
<td>TRα, β</td>
<td>Thyroid hormone</td>
</tr>
<tr>
<td>EcR</td>
<td>20-OH ecdysone</td>
</tr>
</tbody>
</table>

| EX-orphans | | ORPHAN RECEPTORS |
| RARα, β, γ | all-trans RA | No known homologs |
| RXRα, β, γ | 9-cis RA, ? | ERRα, β, γ |
| PPARα, β, γ, δ | fatty acids, eicosanoids | knirps |
| LXRα, β | oxy-sterols | knirps-related |
| FXR | bile acids | DAX-1 |
| BXR | benzoates | SHP |

| Nearly EX-orphans | | egon |
| CARα, β | androstans | GCNF |
| SXR | steroids, xenobiotics | DHR96 |
| PXR, 1,2 | steroids, xenobiotics | |

- Payoff from orphan receptor research so far (3 biotechs)
  - LXR and FXR regulate cholesterol metabolism
    - LXR diverts cholesterol into bile acid pathway
    - FXR negatively regulates uptake of bile acids
  - PPARs regulate fat metabolism
    - PPARγ is insulin sensitizer
  - SXR and PXR regulate metabolism of steroids, xenobiotics and environmental compounds
  - CAR also mediates drug breakdown
Nuclear hormone receptors (contd)

- Nuclear receptors interact with each other and DNA
  - can form homodimers, heterodimers and monomers
  - four possible modes of DNA binding
    - IR, DR, ER, monomer
    - steroid receptors very closely related, others not so much
      - bind to HSPs and stay in cytoplasm until ligand bound.
      - GR, MR, PR and AR all bind each others response elements. Significant crossover between pathways at pharmacological levels of ligands (eg. anabolic steroid use)
        - RXR heterodimers is the largest and most diverse family
        - monomeric orphan receptors may also dimerize and/or interact with RXR depending on the response element.
Nuclear hormone receptors (contd)

Known ligands
- RAR\(\alpha,\beta,\gamma\) all-trans RA
- TR \(\alpha,\beta\) thyroid hormone
- VDR \(1,25-(OH)_{2}-VD3\)
- EcR ecdysone

Recent EX-orphans
- PPAR \(\alpha,\beta,\gamma,\delta\) fatty acids, eicosanoids
- FXR bile acids
- BXR benzoates
- LXR \(\alpha,\beta\) oxysterols

Activatable orphans
- SXR/PXR steroids, xenobiotics
- CAR \(\alpha,\beta\) androstans, xenobiotics

- RXR is a common partner in >10 different pathways
  - can be silent (non-permissive)
  - can be active (permissive)
  - offers another means to regulate dimeric receptors
    - rexinoids (RXR selective compounds) are being used clinically in treatment of diabetes, breast cancer and other diseases.
  - at least part of the reason that vitamin A levels are tightly regulated in vivo
    - too much or too little both very harmful, particularly during development.
Nuclear hormone receptors (contd)

- Retinoids affect many processes in the body
  - majority of developmental effects are mediated through RAR:RXR heterodimers
  - some effects in CNS by presumed RXR homodimers
  - orphan receptors can potentially regulate many features of homestasis
    - rexinoids totally inhibit cholesterol uptake
  - RXR:Nurr1 heterodimer may be important for differentiation of dopaminergic neurons (Parkinson’s)
Nuclear hormone receptors (contd)

- P-box determines half-site specificity
  - CEGCKGFF in many receptors
  - therefore they all bind to the same or similar half sites
  - where does specificity come from?
- spacing between half-sites encodes specificity
  - 3-4-5 rule of Kaz Umesono
  - DR-3 VDRE SXRE
  - DR-4 TRE SXRE, LXRE, FXRE, BXRE
  - DR-5 RARE SXRE, CARRE
  - DR-1 RXRE PPRE
  - DR-2 RARE
- selectivity is not absolute but these provide a good model for determining response elements
Nuclear hormone receptors (contd)

- Transcriptional regulation by RXR heterodimers
  - most, or all of these bind DNA in the absence of ligand
    - unliganded receptor is a repressor
      - effect of activators and repressors together?
    - ligand causes a conformational change that kicks off corepressor
    - liganded receptor can now recruit coactivators and activate transcription
      - coactivators and corepressors alter chromatin conformation by modulating histone modification
      - offers possible ways to specifically disrupt complex
Nuclear hormone receptors (contd)

- How does one go about identifying orphan receptor ligands?
  - requirements
    - receptor expression construct
    - response element to make reporter
    - cofactors (use tissue where receptor is active)
  - good news!
    - assay is very sensitive sub parts per billion
  - analytical bad news
    - chemistry requires parts per thousand-ppm

TYPICAL ORPHAN RECEPTOR LIGAND SCREENING ASSAY

DBD

\[ \text{GAL4} \]

Ligand Binding Domain

\[ \text{RAR, OR, etc.} \]

RE

\[ \text{GAL4}_{\text{UAS}} \]

Reporter Gene

\[ \text{Firefly luciferase} \]

Lipofect cultured cells in 96-well format

Add candidate ligand cocktails or HPLC fractions

Assay for reporter gene activity

Demonstrate specific binding of agonist to receptor
Nuclear hormone receptors (contd)

- What are some effects of mutations in nuclear receptors?
  - steroid receptors
    - knockout of SF-1 removes adrenal/gonad axis
    - human mutations in DAX-1 similar
    - overproduction of adrenal steroids - Cushing’s syndrome
    - underproduction - Addison’s disease
  - nonfunctional AR - testicular feminization
    - genotypic males develop as females externally but male internally (rumor about well-known actress)
  - nonfunctional ER
    - male, osteoporosis, coronary artery disease, continuous growth
    - female lethal in utero, osteoporosis later
  - nonfunctional MR - hypotension
  - nonfunctional GR - hypertension and low renin
  - other receptors
    - thyroid hormone receptor - mutations can lead to alterations in metabolism and ADHD
    - vitamin D receptor - vitamin D resistant rickets
    - retinoic acid receptor
      - several types of leukemia result from fusion of RAR to other transcription factors
      - some are treatable with RA, others not
Homeobox genes

- helix-turn-helix motif is widely used in transcription factors
  - phage repressors
  - 2 helices come to lie at almost right angles to each other
  - the recognition helix fits in the major groove of DNA while other helices make minor groove contacts
  - target sequence discrimination is by only a few residues in the recognition helix
Homeobox genes (contd)

- Homeobox is a highly-conserved, 180 bp DNA sequence
  - only DNA has a homeobox
- homeodomain is the protein product of the homeobox and
  - proteins encoded by homeobox genes are homeodomain proteins
- very large family of genes
  - first discovered in Drosophila homeotic selector genes, Antennapedia and Ultrabithorax
  - low stringency hybridization showed that there were many such genes in the Drosophila genome
  - in a “secret” experiment, Bill McGinnis (Walter Gehring’s lab) and Andrés Carrasco (Eddy De Robertis’s lab) decided to test whether such sequences occurred in vertebrates
    - low stringency Southern blot was performed, the first “zoo blot”
    - several genes were identified, Cell papers published and feelings were hurt
- Drosophila (Hom-C) and vertebrate Hox genes control the identity of body segments during development. Loss of function mutations cause changes in segment identity
  - incredibly, corresponding vertebrate genes can completely rescue fly mutations
Homeobox genes (contd)

- this residue is very important for determining specificity
Homeobox genes (contd)

• We will talk more about homeobox genes in the last two lectures
  – for now, it is sufficient to note that homeobox genes are critical for normal development
  – more than 400 different types already known
  – homeodomain proteins can act as transcriptional activators or repressors
    • many people spent years trying to demonstrate activation of reporter genes by homeodomain proteins with little success
    • it later turned out that the ones that were being tested were repressors

• mutations in homeobox genes cause developmental defects in humans
  – mutations in emx2 homeobox gene (related to Drosophila empty spiracles) causes schizencephaly
    • cortical malformation that manifests developmental delay, blindness, seizures, and other neurological disabilities
  – mutations in MSX-2 lead to Boston-type craniosynostosis
    • cranial bones fuse inappropriately
Helix-loop-helix proteins

- HLH proteins are a large group of dimeric proteins
  - signature motifs are two stretches of amphipathic $\alpha$-helices flanking a central loop (linker) of variable size
    - protein protein interaction is mediated via hydrophobic interactions
    - regions of strong sequence conservation within the helices among related proteins
  - not all have the ability to bind DNA, these are typically negative regulators
  - those that can bind DNA tend to have a basic region adjacent to the HLH motif
    - these are called bHLH proteins
- dimerization regulates function
  - two basic regions are required for DNA binding
  - two groups of bHLH proteins exist
    - Class A are ubiquitously expressed (eg E12/E47)
    - Class B are tissue-specific (MyoD, myogenin)
  - a common strategy among tissue-specific proteins is to heterodimerize with ubiquitous partners
  - homodimers are not very stable and do not bind DNA with high affinity
  - heterodimers between bHLH and HLH proteins are typically nonfunctional, an important regulatory mechanism
Helix-loop-helix proteins (contd)

Figure 21.13 All HLH proteins have regions corresponding to helix 1 and helix 2, separated by a loop of 10–24 residues. Basic HLH proteins have a region with conserved positive charges immediately adjacent to helix 1.

<table>
<thead>
<tr>
<th>MyoD</th>
<th>Ala Asp Arg Arg Lys Ala Ala Thr Met Arg Gin Arg Arg Arg</th>
<th>Basic region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Id</td>
<td>Arg Leu Pro Ala Leu Leu Asp Gin Glu Glu Val Asn Val Leu</td>
<td>6 conserved residues are absent from Id</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MyoD</th>
<th>Leu Ser Lys Val Asn Gin Ala Phe Gln Thr Lys Arg Cys Thr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Id</td>
<td>Leu Tyr Asp Met Asn Gly Cys Tyr Ser Arg Leu Lys Gin Leu Val</td>
</tr>
</tbody>
</table>

Conserved residues are found in both MyoD and Id

<table>
<thead>
<tr>
<th>MyoD</th>
<th>Lys Val Gin Ile Lau Arg Asn Ala Ile Arg Tyr Ile Gin Gly Lle Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Id</td>
<td>Lys Val Gin Ile Lau Glu His Val Ile Asp Tyr Ile Arg Asp Leu Glu</td>
</tr>
</tbody>
</table>

Helix 1

Helix 2

bHLH proteins dimerize and bind DNA

Nonbasic HLH proteins prevent DNA-binding

Insufficient affinity to bind to DNA
Helix-loop-helix proteins (contd)

- bHLH proteins and muscle development
  - MyoD was the first discovered. Identified in an expression screen as a single protein that could transform cultured fibroblasts (3T3) into muscle (myotubes)
  - MyoD acts first to kick ID off of E12 and/or E47 and initiates the muscle program. Later bHLH genes such as myogenin and myf5 are also important

- this family of genes illustrates the general principle that combinatorial associations of transcription factors can yield complexes with different functions
  - DNA binding
  - transcriptional regulation
Leucine zipper (b-ZIP) proteins

- Leucine zipper is a protein:protein interaction domain characterized by coiled-coil $\alpha$-helical structure
  - coiled-coil is a common structural motif in proteins (e.g. myosin)
  - coiled-coil is formed from two helices wound around each other and typified by large hydrophobic amino acids (leu, ile) repeated every 7 residues
  - the helices are usually amphipathic
  - leucine zipper flanked by a basic region is common in transcription factors, so-called b-ZIP motif.

![Diagram of leucine zipper interaction](image-url)
Leucine zipper (b-ZIP) proteins (contd)

- bZIP is a common motif in viral transcriptional activators and some enhancer binding proteins
  - eg C/EBP (CAAT box enhancer binding protein)
- like bHLH proteins, bZIP proteins are regulated by heterodimerization.
  - dimers have distinct functions
  - not all proteins can homodimerize
    - eg c-jun can homodimerize to bind DNA
    - c-fos can not
    - c-jun and c-fos can heterodimerize to produce the transcription factor AP-1
  - the jun/fos heterodimer binds DNA ~10x better than the jun homodimer although both prefer the same DNA target sequence

- target sequences for all dimeric proteins have two half-sites.
  - dyad symmetry
    - AGGTCATGACCT
    - AGGTCAAGGAGGTCA
  - this is a signature feature and should always cause you to suspect a dimeric transcription factor