### BI 8 LECTURE 8 EVOLUTION AND GENOME MANIPULATION; & BRIEF COMMENT ON QUANTITATING BIOLOGICAL REACTIONS

Ellen Rothenberg 28 January 2016

Reading: Ch. 3, pp. 109-144; panels 3-1, 3-2; Ch. 2, pp. 51-63 review Ch. 8 pp. 494-504; optional: Doudna/Charpentier

### Domain structures order folding even along single polypeptide chains



Figure 3-41 Molecular Biology of the Cell (© Garland Science 2008)

Related functions and structures can be predicted for proteins in distant species or among distant family members just based on sequences of genes that code for them... ... educated guessing even before starting biochemistry!

- Find possible exons predicted from continuous reading frames
- Virtual translation "in silico"
- Similarities can reveal matches to other known proteins in database

WYFGI	KIJ	RRESERL	6	GTFLVRES	E				signature sequences
WYFGI	KI I	RRESERL	LINAENPI	RGTFLVRES	ETTKGAYC	LSVS	SDFDNA	AKGL -	human
W+F	+	R+E+++L	LL ENP	GTFLVR S	е ү	LSV	D+++	+G -	• sequence matches
WFFEI	IVI	RKEADKL	LAEENPE	GTFLVRPS	EHNPNGYS	LSVE	KDWEDO	GRGY -	- Drosophila
1		10	20	30		40		50	

Human and Drosophila Src: SH2 (protein interaction) domain

Generally substantially the same folding even if only 20-25% sequence identity! (especially if conserved residues are known to be in key points of structure based on well-studied examples)

Figure 3-14 Molecular Biology of the Cell (© Garland Science 2008)

### Deep evolutionary conservation of protein structure even when amino acid sequence has drifted (homeodomain proteins)

Conservation of *particular* amino acids that are needed for structural function in particular domains is more important than *overall* conservation



# Domain based modules are the foundation of protein evolution: match, multiply, and mix



Modularity is easy to encode in genome when domains are encoded by discrete exons... introns give lots of room for copying, cutting & pasting

A fascinating story of evolution *and* alternative splicing: Down's syndrome cell adhesion molecule (Dscam)



As in many cases, exon/ domain boundaries coincide... a useful mechanism for evolution

(Crayton et al. 2006 BMC Evolutionary Biology)

### Conservation can be spotty... yet key residues are a clue to similar function in corresponding domains (Dscam)

#### A. Exon 4

human_Dscam	PVLREPYTVRVEDQKTMRGNVAVFKCIIPSSVEAYITVVSWEKDTVSLVSGRLCV
mouse Dscam	SVLREPYTVRVEDQKTMRGNVAVFKCIIPSSVEAYVTVVSWEKDTVSLVSGRLSA
rat Dscam	SVLREPYTVRVEDQKTMRGNVAVFKCIIPSSVEAYVTVVSWEKDTVSLVSGRLSA
frog 01	-VLREPYTVRVDDOKAMRGNAVVFKCIIIPSSVEAYVTVVSWEKDTVSL
zebrafish 01	VVLREPYTVRVADOTAMRGSVAVFKCIIPSSVENYITVVSWERDTVPLVSGRTLA
tetraodon_01	PVILREPYTVRVEDOKAMRGSVAVEKCI I PASVEAYI TVV SWEKDTMSTNAESKSP
human Dscaml	AVEREPYTVRVE DORS MRG NVAVEKCI I I PSSVOE YVSVV SWEKDTVSTT PGKKRP
mouse Dscaml	TVEREPYTVRVE DORS MRGNVAVEKCI I PSSVOEVVSVV SWEKDTVSI I PGKESW
rat Dscaml	
frog 02	
110g_02	- VERCET TO VOLD AND A VERCENT POWER OF VOLD AND A VERCENT A
zebrafish_02	TVIFIREPYTVRVGDQKYMRGNVAVFKCILIPSAVQENISVVSWEKDTVSIFPGKTLL
tetraodon 02	VVFREPYTVRVADQRSMRGNVAVFKCLIPAAVQEYVSVVSWERDTVSIVPGRNTW
fly_4_01	-VVPQSYTVNVMDESILRGNSAILKCHIPSFVADFIVVDSWVEDEERVIYPQEDIAESGKFTD
mosquito_4_01	-VVSQYYEVDVNKEHVILGNSAIFKCLIPSFVADFVDVVSWTSGDDEEETHVYSADAYG
honeybee_4_01	-VVAQYYDTDVNKEYAIRGNSAILKCVVPSFVADFVKVLSWHTDQGEEFVPGDDYG

(Crayton et al. 2006 BMC Evolutionary Biology)

#### B. Exon 9

human_dscam	PPFIQPFEFPRFSIGQRVFIPCVVVSGDLPITITWQKDGRPIPGSLGVTIDNIDFTSSLRISNLSLMHN	VGNYTCIARNEAAAVEHQSQLIVRVP
mouse_dscam	PPFIQPFEFPRFSIGQRVFIPCVVVSGDLPITITWQKDGRPIPASLGVTIDNIDFTSSLRISNLSLMHN	IGNYTCIARNEAAAVEHQSQLIV
rat_dscam	PPFIQPFEFPRFSIGQRVFIPCVVVSGDLPITITWQKDGRPIPASLGVTIDNIDFTSSLRISNLSLMHN	IGNYTCIARNEAAAVEHQSQLIV
dog_01	PPF IQP FEF PRFS IGQ RVF IPC VVVS GDL PITITWQKDGRPIPAS LGVTIDNIDFTSS LR ISNLS LMHN	VGNYTCIARNEAAAVEHQSQLIV
frog_01	PPFIQPFESQRFSIGQRVFIPCVVVSGDLPITITWQKDGRPIPASLGVTIDNIDFTSSLRISNLSLMHN	VGNYTCIARNDAAAVEHQSQLIV
fugu 01	PPY IQP FEFQRFT IGQRVF IPCVVMSGDRPLDI TWQKDGRP IPVS LGVTVDN IDFTSSLR INNLTPDHN	NGNYTCIARNEAATVEHQSRLIV
fugu_02	RPIPASLGVTIDNIDFTSSLRISNLTLLHN	VGNYTCIARNQAAAVEHQSQLIV
human_dscamL	QVIISGSGVTIESKEFMSSLQISSGDMPIRITWRKDGQVIISGSGVTIESKEFMSSLQISSVSLKHN	IGNYTCIASNAAATVSRERQLIV
mouse dscamL	QVIISGSGVTIESKEFMSSLQISSGDMPIRITWRKDGQVIISGSGVTIESKEFMSSLQISSVSLKHN	IGNYTCIASNAAATVSRERQLIV
rat_dscamL	QVIISGSGVTIESKEFMSSLQISSGDMPIRITWRKDGQVIISGSGVTIESKEFMSSLQISSVSLKHN	VGNYTCIASNAAATVSRERQLIV
dog_02	QVIISGSGVTIESKEFMSSLQISSGDMPIRITWRKDGQVIISGSGVTIESKEFMSSLQISSVSLKHN	IGNYTCIASNAAATVSRERQLIV
frog_02	QVIVSGSGITIETKEFMSSLQISSGDMPIRITWRKDGQVIVSGSGITIETKEFMSSLQISSVSLKHN	VGNYTCIASNDAATVSRERQLIV
zebrafish_01	PPLIQHSDSQSASIGQRVFIPCVVISGDLPMSITWHKDGRPINASLGVTIDNIDFTSSLRISNLSEIHN	VGSYTCIARNEAAAVEHSIQLIV
tetraodon 01	QEIVPSSGITIDTKEFMSSLQISKVSLKHN	NGNYTCIASNDAATVSTERQLTV
worm_01	PTIIESPHTVRVNIERQVTLQCLAVG-IPPPEIEWQKGNVLLATLNNPRYTQLADGNLLITDAQIEDQ	QGQFTCIARNTYGQQSQSTTLMV
fly_9_9	PPQVLPFSFGESAADVGDIASANCVVPKGDLPLEIRWSLNSAPIVNGENGFTLVRLNKRTSLLNIDSLNAFHF	RGVYKCIATNPAGTSEYVAELQVN
mosquito_09	- PQIMPFEFGDEPFDSSSTVSINCVVSKGDTPIMIEWLLNGNRVTTNDGINIMKSGQKISMLSIESVQSRHA	AGNYTCVARNKAGESQHTSELKV
honeybee_10_09	09 APEIVAFDIGEGPANWGDTVTATCTVVKGDHPIQIEWALNGEPISRN-HYDISIVNTSKRVSLLTIDGVTARHA	AGEYTC SVSNAAGGTS YSATLAVN

### Yet the use of the exons is NOT the same in vertebrates as in insects...

- DSCAM
- —— 100 aa





(Brites et al. 2013 Evolution)

### In insects – not vertebrates – RAMPANT mutually exclusive alternative splicing to choose each of four exons $\rightarrow$ 38,016 potential isoforms!



Thought to be used for immune recognition in insects Diversity helps to create a wide repertoire of possible defense structures... but neuronal pathfinding needs diversity too

(Hemani, Soller 2012 Biochem Soc Transact)

## Nucleic acid manipulation can be key to establishing function of proteins and protein domains

- DNA can be expressed into proteins of interest in vivo or in vitro
  - Transfection into cells in culture (cDNA or genomic)
  - Change in function or protein expression relative to background (hopefully low or zero!)
  - Also available: in vitro, coupled transcription/ translation systems: put in DNA (cDNA cloned in vector with promoter), get out protein
- If you have a functional assay, you can see effect of adding a protein on a system
- With a functional assay, compare effect of adding wildtype vs. protein coded by deleted or mutated cDNA
   => infer key domains, key amino acids!!

### Modifying genomes within cells

- Use sequence-specific recombination to introduce desired mutations into target DNA plasmids ...or cellular genomes
- Recombination is a key natural aspect of DNA maintenance in cells as well as an artificial result of DNA cleavage and ligation in vitro
- Starts like restriction digestion with a nick or staggered break in the DNA
- Local homology promotes rejoining
- But recombination can introduce new sequences ... or delete original sequences

Viruses in prokaryotes and eukaryotes modify host genomes "for a living": they encode their own equivalents of restriction enzymes



Integrases: high specificity for non-disruptive sites in viral DNA, varyingFigure 5-73 Molecular Biology of the Cell (© Garland Science 2008)Figure 5-73 Molecular Biology of the Cell (© Garland Science 2008)

# Manipulating recombinase (integrase) specificity for our own uses

- Recombinases made by viruses and parasites have exquisite site specificity ... though sometimes only for the "willing" recombination partner (viral genome)
- To use, learn the target site for specific cutting
- To control, clone the recombinase gene for expression on YOUR demand

Cre: A particularly useful recombination enzyme – site-specific cutting and rejoining from a single enzyme



Figure 5-42 Molecular Cell Biology, Sixth Edition © 2008 W. H. Freeman and Company



### New hottest approach around: CRISPr

- Logistical issues with Cre/LoxP system (and related systems, Flp/Frt, φ31, etc.)
  - Good: long-sequence target sites (~30 bp) are very uncommon in mammalian genomes naturally, adding specificity
  - Bad: You have to make a targeted recombinant animal first, with LoxP or Frt sites in right places, before you can use Cre to cut & rejoin.... Years of work, high cost
- CRISPr: a natural bacterial defense system that can target ANY site (almost) if a small cheap complementary RNA is introduced into system

# The heart of CRISPr... an obedient protein and a guiding RNA



### A mechanism in bacteria to use past viral infections to create "immunity" can be made into a tool



Targeting a new DNA sequence to a DNA break at will: Sequence homology, even over short distance, can enhance DNA break repair



# Key points for quantitation of biological reactions

- Synthesis / turnover
- Equilibrium vs. Kinetics
- Specific binding
- Enzyme kinetics

### Levels of all macromolecules are determined by balance between synthesis and decay rates

Asymptote=  $k_{\rm s}/k_d$ The rate of increase in any 250 molecule X is 200 # mRNA  $dX/dt = k_s - k_d X(t)$ 150 100 50 •  $X(t) = (k_s/k_d)(1 - e^{-k_d t})$ 100 200 300 400 500 t1/2 Time (minutes) Here, Initial Synthesis rate The steady state level for any •  $= k_{s}$ : molecule, when dX/dt = 0, is at

tangent to slope at t= 0

600

 $X = k_s / k_d$ 

Rates of synthesis and turnover both matter: Synthesis kinetics and decay kinetics affect both maximum level and time constants of responses



Time to half-maximal accumulation depends on decay rate  $k_{d}$ :  $t_{1/2} = (ln2)/k_{d}$ 

FIG. 3. Rate of change of two mRNAs (stable and unstable) following transcriptional up-regulation (137–139). mRNAs X and Y are transcribed at the same rate. mRNA X is more abundant, because it is 10-fold more stable. The arrow indicates the time at which transcription of both mRNAs increases 10-fold.

From Ben-Tabou de Leon & Davidson, Devel. Biol. 325: 317-328 (2009)

#### Decay rates have enormous impact on biological response *kinetics* as well as molecular levels

For molecules differing only in decay rates:

Molecules with same steadystate levels but different turnover rates can respond to differences in synthesis rate very differently ... Key point is what fraction of molecules at each timepoint are new!



Figure 15-11 Molecular Biology of the Cell (© Garland Science 2008)

Equilibria between alternative states are the "constraints" but also the tools of biology

- For aggregates of individual molecules, equilibria between synthesis and degradation
- For chemical recognition, between bound and free
- For reactions, between forward and reverse reaction



Time  $\longrightarrow$ 

Figure 2-22 Molecular Cell Biology, Sixth Edition © 2008 W.H. Freeman and Company Gibbs free energy



The free energy of Y is greater than the free energy of X. Therefore  $\Delta G < 0$ , and the disorder of the universe increases during the reaction  $Y \rightarrow X$ .

this reaction can occur spontaneously

Gibbs free energy



If the reaction X → Y occurred, ∆G would be > 0, and the universe would become more ordered.

this reaction can occur only if it is coupled to a second, energetically favorable reaction

Figure 2-50 Molecular Biology of the Cell (© Garland Science 2008)

# Gibbs free energy changes predict the direction of chemical and biochemical reactions

$$\Delta \mathbf{G} = \Delta \mathbf{H} - \mathbf{T} \Delta \mathbf{S}$$

Free energy is negative if reaction is strongly exothermic ( $\Delta H < 0$ ), or if entropy ( $\Delta S$ ) increases enough to offset endothermic  $\Delta H > 0$ 



# Equilibrium constant ${\rm K}_{\rm eq}$ for a reaction depends directly on $\Delta {\rm G}^{\rm 0'}$

 $\Delta G = \Delta G^{0'} + RT \ln K_{eq}$ 

So at equilibrium, when  $\Delta G = 0$ ,

 $\Delta G^{0'} = -RT \ln K_{eq}$ or  $K_{eq} = e^{(-\Delta G^{0'/RT)}}$ Or for a binding reaction,  $[AB]/([A][B]) = e^{(-\Delta G^{0'/RT)}}$ If  $\Delta G^{0'}$ is negative to the second second

If  $\Delta G^{0^{\circ}}$  itself is negative, this makes exponent positive and [AB] > [A][B]

#### A major type of reaction is binding vs. unbinding of two \_\_\_\_\_\_molecules



Figure 3-43a Molecular Biology of the Cell (© Garland Science 2008)

The power of a few noncovalent bonds in setting chemical equilibria

Recall: the strength of *ONE* Hbond in water is ~ 1 kcal/mole ... a difference of 2-3 H-bonds in a complex can shift equilibrium constant by a factor of 30-100

The relationship between free-energy differences and equilibrium constants (37°C)						
equilibrium	free-energy	free-energy				
constant	difference	difference				
$\frac{[AB]}{[A][B]} = K$ (liters/mole)	of AB minus free energy of A + B (kcal/mole)	of AB minus free energy of A + B (kJ/mole)				
1	0	0				
10	-1.4	-5.9				
10 <sup>2</sup>	-2.8	-11.9				
10 <sup>3</sup>	-4.3	-17.8				
10 <sup>4</sup>	-5.7	-23.7				
10 <sup>5</sup>	-7.1	-29.7				
10 <sup>6</sup>	-8.5	-35.6				
10 <sup>7</sup>	-9.9	-41.5				
10 <sup>8</sup>	-11.3	-47.4				
10 <sup>9</sup>	-12.8	-53.4				
10 <sup>10</sup>	-14.2	-59.4				
10 <sup>11</sup>	-15.6	-65.3				

Enzymes speed up reactions (sometimes immensely) by lowering the free energy cost of the transition states: Direction that reaction will run depends on overall free energy change in current conditions



Catalysis can work both ways in principle... not necessarily a net increase in one direction

Figure 2-53 Molecular Biology of the Cell (© Garland Science 2008)

But enzymes can also add to reaction specificity and directionality by coupling them with other reactions

- Siphon off accumulated products to slow reverse reaction
- Couple a reaction physically with an energetically favorable one

(e.g. ATP +  $H_2O \rightarrow AMP + PP_i$  might be reversible... but after  $PP_i + H_2O \rightarrow 2P_i$ , overall  $ATP + 2H_2O \rightarrow AMP + 2P_i$  gives  $\Delta G \sim -8$  kcal/mol)