Sequence specificity of a protein-DNA interaction

Michael Pikaart

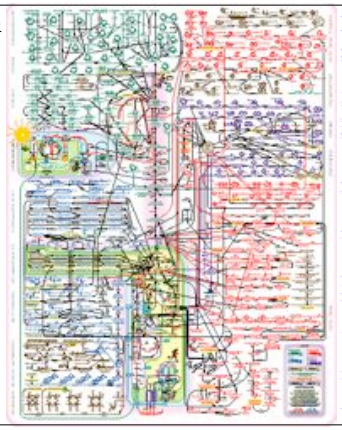
Dept of Chemistry and Biochemistry, Hope College, Holland, Michigan

Molecular Genetics and Microbiology, UNM (2006-'07)

Metabolic Pathway Map

•All these reactions are enzyme catalyzed.

•Different amounts in different cells and/or at different times.

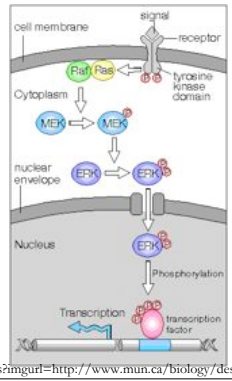


http://www.sigmaaldrich.com/Area_of_Interest/Life_Science/Metabolomics/Key_Resources/Metabolic_Pathways.html

Transcription Factors (Protein/DNA interactions)

•Allows for control of gene expression

•Often modifiable



http://images.google.com/imgres?imgurl=http://www.mun.ca/biology/desmid/brian/BIOL3530/...

Example - hematopoietic stem cells

Stem cells in the bone marrow respond to Epo.

Specific transcription factors mobilize to make cells divide and make the functional proteins of a red blood cell (like hemoglobin).

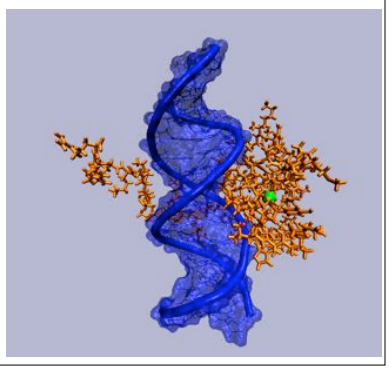
After many rounds of cell division mature red blood cells are released into circulation.

Close-up gene regulation – local binding of transcription factors

```
69901 attctatttt tagacataat ttattagcat gcatgagcaa attaagaaaa acaacaacaa
69961 atgaatgcat atatatgtat atgtatgtgt gtatatatac acacatatat atatatattt
70021 tttcttttct taccagaagg ttttaatcca aataaggaga agatatgctt agaaccgagg
70081 tagagttttc atccattctg tcctgtaagt attttgcata ttctggagac gcaggaagag
70141 atccatctac atatcccaaa gctgaattat ggtagacaaa actcttccac ttttagtgca
70201 tcaacttctt atttgtgtaa taagaaaatt gggaaaacga tcttcaatat gcttaccaag
70261 ctgtgattcc aaatattacg taaatacact tgcaaaggag gatgttttta gtagcaattt
70321 gtactgatgg tatggggcca agagatatat cttagaggga gggctgaggg tttgaagtcc
70381 aactcctaag ccagtgccag aagagccaag gacaggtacg gctgtcatca cttagacctc
70441 accctgtgga gccacaccct agggttggcc aatctactcc caggagcagg gagggcagga
70501 gccagggctg ggcataaaag tcagggcaga gccatctatt gcttacattt gcttctgaca
70561 caactgtgtt cactagcaac ctcaaacaga caccatggtg catctgactc ctgaggagaa
70621 gtctgccgtt actgccctgt ggggcaaggt gaacgtggat gaagttggtg gtgaggccct
70681 gggcaggttg [illegible]aaggt tacaagacag gtttaaggag accaatagaa actgggcatg
70741 tggagacaga gaagactctt gggtttc[illegible]ggcactga ctctctctgc ctattggtct
70801 attttcccac ccttaggctg ctggtggtct acccttggac ccagaggttc tttgagtcct
70861 ttggggatct gtccactcct gatgctgtta tgggcaaccc taaggtgaag gctcatggca
70921 agaaagtgct cggtgccttt agtgatggcc tggctcacct ggacaacctc aagggcacct
70981 ttgccacact gagtgagctg cactgtgaca agctgcacgt ggatcctgag aacttcaggg
```

GATA-1

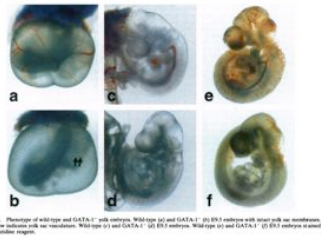
- Transcription factor
- Red blood cell formation
- Too much activity: leukemia
- Not enough activity: anemia



GATA Knockout Mice Embryos

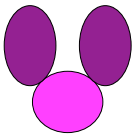
With GATA

Without GATA KO mice



Fujiwara et al., *Proc. Natl. Acad. Sci. USA*, **93**, 12355-12358 (1996)

cGata-1: An Erythroid-Specific Transcription Factor



Transcription activation domain

DNA binding domain

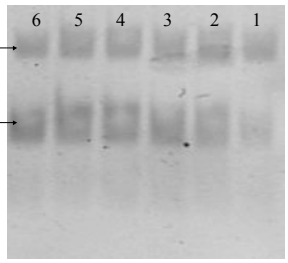
--(A/T)GATA(A/G)--
--(T/A)CTAT(T/C)--

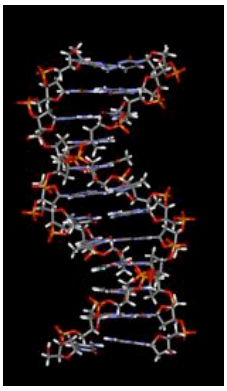
Recognition sequence in DNA

Gel Shift Assay (EMSA)

DNA/Protein Complex

Unbound DNA



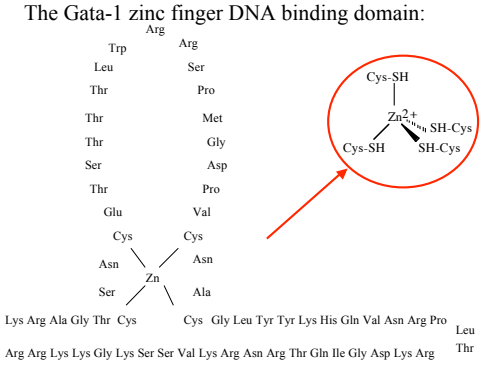


How is **sequence-specific** binding accomplished? *What makes Gata bind very tightly to GATA but not other sequence??*

Non-covalent protein-DNA interactions:

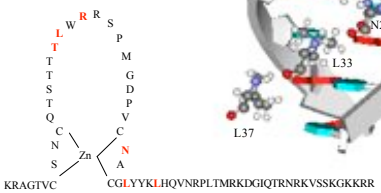
- Ionic
- Hydrogen bonding
- Hydrophobic

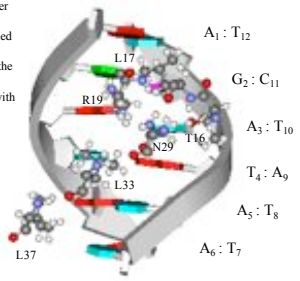
The Gata-1 zinc finger DNA binding domain:



The Gata-1-DNA Interface

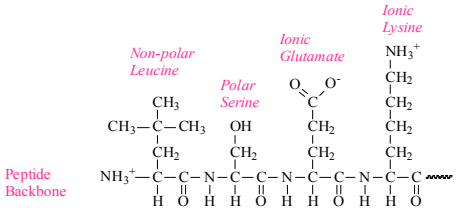
The chicken Gata-1 C-terminal zinc finger was the specific protein explored in our research. It includes four cysteines bonded to a Zinc atom. Amino acids directly involved in DNA binding are located in the C-terminal protein. Shown here is a schematic of the DNA binding domain with selected amino acids from Gata-1.





Properties of Proteins

- Offer a wide range of chemical functionality

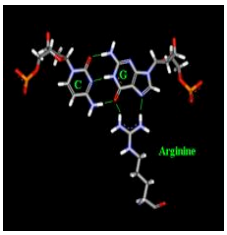


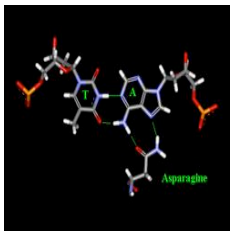
H-bonds in addition to Watson/Crick base pairs

Major groove surface presents variability

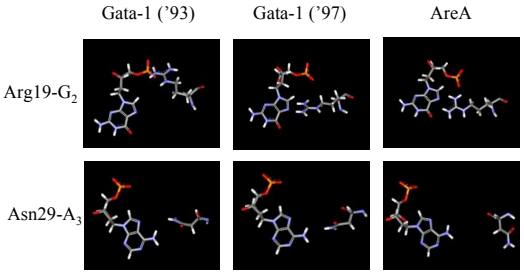
•H-bonding; for example Arg to Guanine; Asn (or Gln) to Adenine

•Methyl group on Thymine



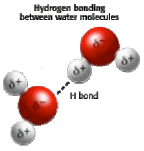


Potential amino acid-DNA hydrogen bond contacts

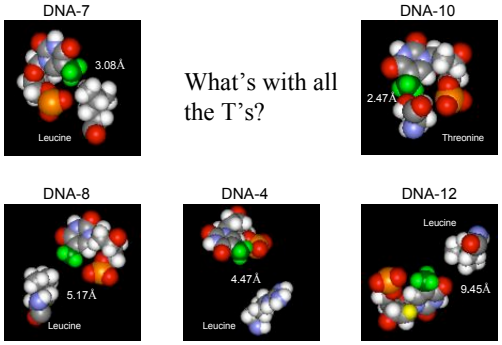


Hydrogen Bonding

- Hydrogen bonding - between a partially positively charged hydrogen atom and either a partially negatively charged Oxygen or Nitrogen atom
- This bonding is very important in DNA because it helps maintain double stranded integrity

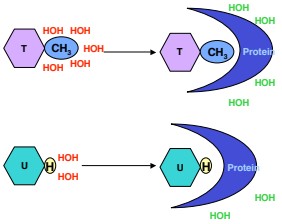


Possible van der Waals interactions:
DNA Bases to Protein Side Chains



What's with all the T's?

Hydrophobic Effect



Thymine

- Displacement of entropically unfavorable water molecules around the nonpolar thymine
- Adds stability
- ΔG of full desolvation of a thymine methyl group to be approximately –1 kCal/mole

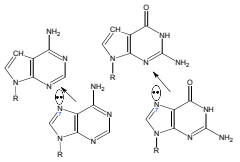
Uracil

- Uracil lacks a methyl group
- Solvation by water is less entropically disfavored.
- Protein binding results in a decreased gain in stability.

What Drives Gata-1 binding specificity?

- Hydrophobic Interactions
  - Leucine → Thymine methyl groups.
- Hydrogen Bonding
  - Arginine → N7 atoms of both G_2 and A_3
  - Asparagine → water bridge between Asn29 and A_3
- So—How can we test the importance of these interactions?

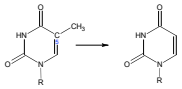
DNA Base Modifications



Adenine Guanine

Deaza Modifications

Knock out the hydrogen bonders?



Thymine Uracil

T to U

Alter the van der Waal's

Which has the greater effect?

Equilibrium constant from ligand titration

- Dissociation constant

$$K_D = \frac{(A_{tot} - AB)(B_{tot} - AB)}{AB}$$

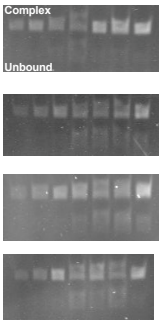
$AB_{complex}$ = bound concentration
A_{tot} = total GATA-1 protein concentration (unknown)
B_{tot} = total DNA concentration
K_D = dissociation constant (unknown)

- Total protein

$$AB_{complex} = \frac{A_{tot} + B_{tot} + K_D - \sqrt{(A_{tot} + B_{tot} + K_D)^2 - 4A_{tot}B_{tot}}}{2}$$

- $K_D \rightarrow \Delta G$ and then $\Delta G = \Delta H - T\,\Delta S$
  - $\Delta G = RT\ln(K_D)$

T to U Results

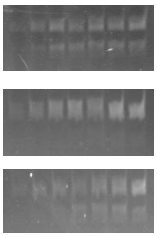


T
AGATAA
TCTATT

4
AGAUAA
TCTATT

7
AGATAA
TCTATU

8
AGATAA
TCTAUT



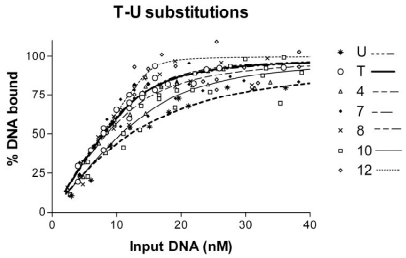
10
AGATAA
TCUATT

12
AGATAA
UCTATT

U
AGATAA
UCUAUU

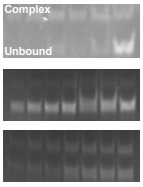
DNA-10 bound similarly to U (weak binding). In other words, modifying thymine-10 caused a loss in binding affinity similar to U. DNA-12 bound similarly to unmodified T (strong binding) so this modification did little to the protein's binding affinity.

T to U Binding Isotherms



Modified thymine to uracil DNA had the slowest approach to saturation and a lesser binding affinity than unmodified T DNA. Specific modification of thymine base 10 also showed a slow approach to saturation. This indicates a weaker binding affinity for the specific oligonucleotide to Gata-1 protein.

Deazapurine Binding Results

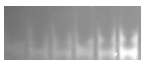


Deaza-1
AGATAA
TCTATU

Deaza-2
AGATAA
TCTATU

Deaza-3
AGATAA
TCTATU

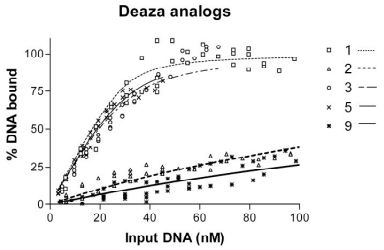
Deaza-5
AGATAA
TCTATU



Deaza-9
AGATAA
TCTATU

Deaza-2 and deaza-9 expressed the weakest binding affinity. Specifically for deaza-9, Binding was nearly invisible indicating a drastic loss in binding affinity. The elimination of hydrogen bonds at these two positions appeared to effect binding affinity the most of all deaza nucleotides.

Deazapurine Binding Isotherms



Modified deaza-2 and deaza-9 approached saturation very slowly. The binding affinity for these two analogues was much less than the other deaza analogues.

The numbers……

	K_{eq} (nM)	SE	?G (kcal/mol)	SE	P	?? G
Unmodifed	1.6	0.4	12.01	0.16		
U-4	1.6	0.4	11.98	0.15		
U-7	1.6	0.5	11.99	0.19		
U-8	1.5	0.6	12.02	0.23		
U-10	4.0	0.9	11.45	0.13	*	-0.56
U-12	0.9	0.3	12.37	0.23		
4xU	6.2	1.9	11.19	0.18	*	-0.82
Deaza-A_1	1.7	0.5	11.95	0.17		
Deaza-G_2	146	7	9.32	0.03	**	-2.69
Deaza-A_3	4.5	0.6	11.39	0.07	**	-0.62
Deaza-A_5	3.6	0.3	11.51	0.05	**	-0.51
Deaza-A_9	254	15	8.99	0.04	**	-3.02

BH Mott, J Bassman, and MJ Pikaart. "A molecular dissection of the interaction between the transcription factor Gata-1 zinc finger and DNA." *Biochemical and Biophysical Research Communications* 316: 910-917 (2004).

The Yeast One-Hybrid Assay

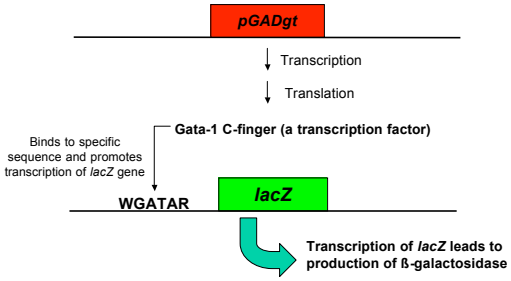
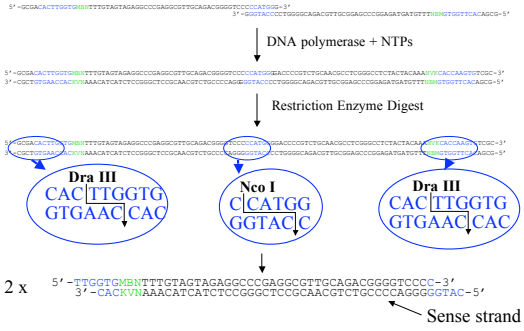


Figure 3. Yeast cells contained genes encoding the C-finger (*pGADgt*) and a Gata-1 activation construct driving *lacZ* expression (a reporter gene). The binding of Gata_1 to its sequence leads to the transcription of the *lacZ* gene and the production of __galactosidase.

Oligonucleotide Extension and Digestion



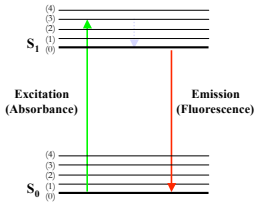
Library	Colonies screened	# of positives	Mutant amino acids found
Arg 19 KNK	100	0	None
Arg 19 NHK	100	0	None
Leu 17 RNK	150	10	Ile, Met, Val
Leu 17 NVK	100	6	Tyr
Asn 29 NNK	100	1	Trp
Leu 33 NVK	200	3	Cys, Gly
Leu 33 RNK	50	0	None
Leu 37 RNK	50	13	Ala, Asn, Ile, Met, Val
Leu 37 NVK	-	-	This sublibrary has yet to be screened

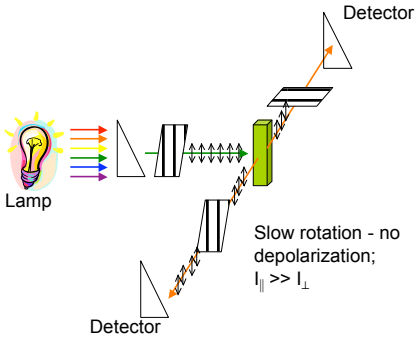
Another Approach

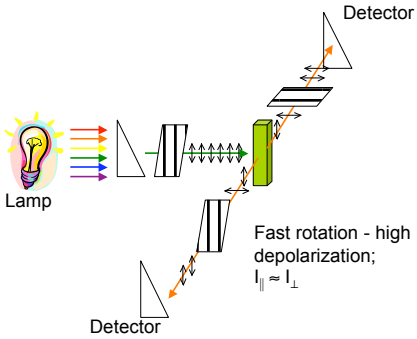
- Fluorescence technique: anisotropy
  - More accurate
  - Solution phase
  - Allows for smaller sample amount
  - Allows for temperature dependent studies

Fluorescence

- Before fluorescence can occur, a molecule must be excited by absorbing a photon, exciting an electron into its next highest energy state
- Fluorescence occurs when an excited molecule emits a photon as it returns to its original ground state







Anisotropy

Biophysical Journal Volume 84 January 2003 599–611

Constrained Analysis of Fluorescence Anisotropy Decay: Application to Experimental Protein Dynamics

Efraim Feinstein,* Gintaras Deikus,† Elena Rusinova,‡ Edward L. Rachofsky,* J. B. Alexander Ross,‡ and William R. Laws§

$$I_M(t) = \frac{1}{3}\sum_{i=1}^{n} \alpha_i e^{-t/\tau_i}$$

Fluorescence decay (unpolarized)

$$r_i(t) = \sum_{j=1}^{m} \beta_{ij} e^{-t/\phi_j}$$

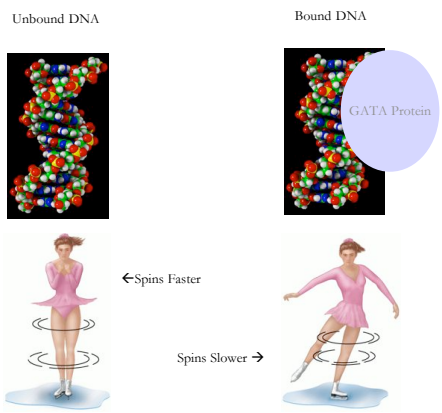
Rotationally oriented decay

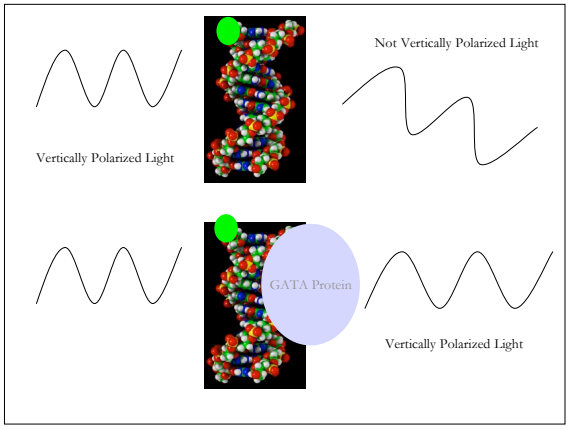
$$I_V(t) = I_M(t)\{1 + 2r(t)\}$$

$$I_H(t) = I_M(t)\{1 - r(t)\}.$$

$$r = \frac{I_{VV} - I_{VH}}{I_{VV} + 2I_{VH}}$$

- Faster/free-moving sample = low anisotropy
- Slower/restricted sample = high anisotropy

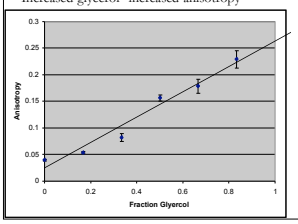




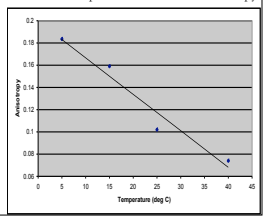
Steady-state measurements

- Fluorophors are continually being excited with polarized light, then rotate (or not) and emit light that is depolarized to the extent of the rotation speed.
- Validation: Eosin dissolved in water with different concentrations of glycerol and at different temps.

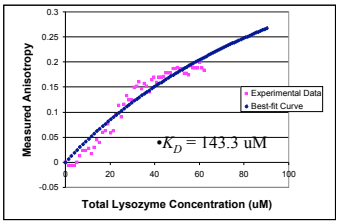
Increased glycerol=increased anisotropy



Increased temperature=decreased anisotropy



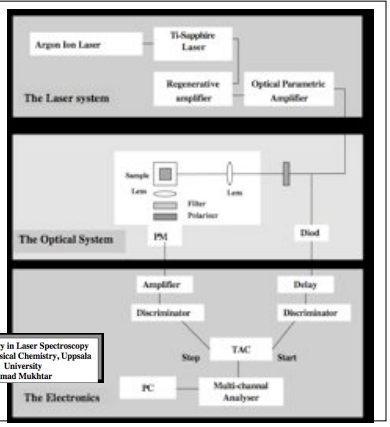
Titration of eosin with lysozyme:

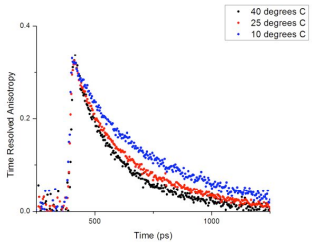


$$K_D = \frac{([M_{Total(T)}] - [ML]) \times ([L_T] - [ML])}{[ML]}$$

TCSPC:

Instrumental setup for time-resolved anisotropy measurements.





Time-resolved anisotropy decays of Pacific Blue in methanol as a function of temperature.