[redacted]

Identification of sequence-specific DNA binding proteins

DNA binding proteins play an integral role in the modulation of gene expression by binding to specific sites in the genome to enhance or repress expression of target genes. One of the earliest proteins found to bind DNA was the λ phage repressor[1-3], which was determined to bind to λ DNA by Ptashne in 1967 . This initial characterization of a new class of protein sparked the identification and isolation of a number of other DNA binding proteins, including as Sp1[4-5]. In turn, this prompted the development of new purification techniques[6] that improved our ability to isolate, identify, and characterize additional proteins[7]. As more sequence-specific DNA binding proteins were discovered, it became possible to identify recurring sequence motifs that could facilitate the binding of DNA, and to group previously identified proteins under these new families[8-12]. Improved understanding of sequence-specific DNA binding proteins and their shared motifs facilitates better prediction and mechanistic understanding of gene regulatory targets.

The λ repressor was first identified using a combination of amber mutant phages and temperature sensitive phage mutants where phage repression could be selectively toggled through environmental factors[13]. Soon after, Ptashne used radiolabeling of the λ repressor and λ DNA to show that the repressor coeluted with DNA, suggesting that the λ repressor's mode of action was through physical interaction with DNA[1]. The specific sequence targeted by the λ repressor would later be constrained to two broad regions by identifying regions where mutations reduce binding affinity[2], and these two regions would later be reduced to a series of conserved sites each measuring approximately 15 base pairs each[3]. The compilation of these efforts represented the first characterization of a sequence-specific DNA-binding protein and its associated target. Following the successful characterization of a new class of proteins, several new sequence-specific DNA binding proteins were identified, and families of sequence-specific DNA binding proteins began to emerge.

Sp1 was one of the earlier sequence-specific DNA binding proteins to be well characterized following the λ repressor. A pair of DNA-binding proteins that targeted different promoters were simultaneously characterized after isolating and separating the pair using a TGED gradient[4], which became termed Sp1 and Sp2. Purified Sp1 was determined to cause a 40-fold increase in SV40 early promoter transcription, thus demonstrating that it likely binds SV40[4]. Soon after it was found the Sp1 targets a specific hexamer motif[5], GGGCGG, which became more commonly known as a GC box motif. Up to this point, it had been relatively difficult to specifically isolate individual DNA-targeting proteins that were not significantly abundant, such as Sp1 and TFIIIA, which prompted the development of a bait-and-catch chromatography method for the affinity purification of site specific binding proteins[6] by utilizing their target sequence fused to a resin. A vital improvement added to this method was to saturate the protein fraction with competitor sequences to reduce nonspecific binding of contaminating proteins. This methodology was quickly adopted and used to identify additional sequence-specific binding proteins, notably AP-1, which targets several oncogenes that are inducible by phorbol diester (TPA)[7]. Its putative binding domain was predicted by sequence conservation across TPA-inducible genes, and this motif was successfully used in the affinity purification of

AP-1[7]. As increasing numbers of sequence-specific DNA binding proteins were identified, common elements began to emerge, and it was possible to identify sequence-specific DNA binding motifs.

TFIIIA is a transcription factor from *Xenopus laevis* that was an appealing target for early characterization efforts due to its high copy-number and easy purification[8]. Upon digestion it was found to have multiple similarly sized domains which could be isolated by aligning the protein against itself to find a repeating motif hypothesized to specifically bind grooves in DNA known as a zinc finger domain[8]. Following the characterization of this new motif, truncation mutants of Sp1 were used to localize its DNA binding activity to a small region at its C-terminus[9]. This region was found to contain three zinc finger binding domains[9], suggesting that this class of sequence-specific DNA binding structure is likely to be a recurring motif. At the same time, it was known that not all sequence-specific DNA binding proteins shared this motif, as a group of six proteins, including Fos and Jun, previously identified were known not to have a zinc-finger like domain from primary sequence analysis[10]. Instead, they had a repeating leucine based motif which was predicted to facilitate the formation of long and particularly stable α-helices. Given this predicted conserved helix motifs across multiple DNA binding proteins from multiple organisms, a model in which these helices dimerize and form a complex capable of binding specific DNA grooves known as a leucine zipper[10] was formulated. After the leucine zipper motif was hypothesized, AP-1 was revisited, and found to be a dimer of a pair of leucine zippers monomers, Fos and Jun, two of the original leucine zippers identified[11-12]. Zinc finger proteins and leucine zippers were two of the major classes of sequence-specific DNA binding proteins discovered by comparisons across recently identified DNA binding proteins.

The identification of family of sequence-specific DNA binding proteins and the characteristic motifs that define them were made possible through the identification and characterization of multiple sequence-specific DNA binding-proteins, including Sp1, AP-1, TFIIA, and others[4,5,7-12]. The techniques and methodology underlying how to identify a sequence-specific DNA binding protein and its target were pioneered during the characterization of the λ repressor[1-3,13]. The continued identification of both specific sequence-specific DNA binding proteins and new classes of sequence-specific DNA binding proteins are integral to improving our understanding of recurring motifs vital for gene regulation.

[redacted]

References:

1. Ptashne, Mark. "Specific Binding of the λ Phage Repressor to λ DNA." *Nature* 214.5085 (1967): 232–234.
2. Ptashne, M, and N Hopkins. "The Operators Controlled by the Lambda Phage Repressor." *Proceedings of the National Academy of Sciences of the United States of America* 60.4 (1968): 1282–7.
3. Maniatis, T, and M Ptashne. "Multiple Repressor Binding at the Operators in Bacteriophage Lambda." *Proceedings of the National Academy of Sciences of the United States of America* 70.5 (1973): 1531–5.
4. Dynan, William S., and Robert Tjian. "Isolation of Transcription Factors That Discriminate between Different Promoters Recognized by RNA Polymerase II." *Cell* 32.3 (1983): 669–680.
5. Dynan, William S., and Robert Tjian. "The Promoter-Specific Transcription Factor Sp1 Binds to Upstream Sequences in the SV40 Early Promoter." *Cell* 35.1 (1983): 79–87.
6. Kadonaga, James T, Robert Tjian, and Bruce M Alberts. "Affinity Purification of Sequence-Specific DNA Binding Proteins (DNA-Agarose Affinity Chromatography/competitor DNA/transcription Factor Spl/synthetic Oligodeoxynucleotides)." *Biochemistry* 83. (1986): 5889–5893.
7. Angel, P et al. "Phorbol Ester Inducible Genes Contain a Common Cis Element Recognized By a Tpa-Modulated Trans-Acting Factor." *Cell* 49.6 (1987): 729–739.
8. Miller, J., A. D. McLachlan, and A. Klug. "Repetitive Zinc-Binding Domains in the Protein Transcription Factor IIIA from Xenopus Oocytes." *Journal of Trace Elements in Experimental Medicine* 14.2 (2001): 157–169. Web.
9. Kadonaga, J T et al. "Isolation of cDNA Encoding Transcription Factor Sp1 and Functional Analysis of the DNA Binding Domain." *Cell* 51.6 (1987): 1079–1090. Web.
10. Landschulz, William H, Peter F Johnson, and Steven L Mcknight. "The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins *Science*. **240** (1988): 1759–1764. Print.
11. Chiu, Robert et al. "The c-Fos Protein Interacts with c-Jun AP-1 to Stimulate Transcription of AP-1 Responsive Genes." *Cell* 54.4 (1988): 541–552. Web.
12. Halazonetis, Thanos D. et al. "C-Jun Dimerizes with Itself and with c-Fos, Forming Complexes of Different DNA Binding Affinities." *Cell* 55.5 (1988): 917–924. Web.
13. Ptashne, Mark. "Isolation of the Λ Phage Repressor." *Proceedings of the National Academy of Sciences of the United States of America* 57.2 (1967): 306.