© 1998 Nature Publishing Group http://www.nature.com/nsmb

Secret code

The structures of Cro[1], CAP[2] and the λ repressor DNA binding domain[3] that were determined in 1981 and 1982 had suggested defined modes of interaction with DNA through their helix-turn-helix motifs. Were these models correct? It would take several years for this question to be answered. The first glimpse of a protein bound to its DNA site was published in 1985 and was of the DNA binding domain of 434 repressor (similar in structure to λ repressor) bound to its 14 base pair operator site[4]. Although it is most unlikey that a model of a protein–DNA complex based on a 7 Å resolution map such as the one in that paper would be published today, this is only testament to the feeling in the community that DNA binding proteins were "incredibly hot stuff", as Cynthia Wolberger puts it. The structure did not allow many details to be observed, but it did confirm that the helix-turn-helix structure was used to recognize the major groove, as predicted from the early models.

Surprisingly, the first high resolution cocrystal structure of a site-specific DNA binding protein and its binding site was neither of one of the phage proteins nor of CAP, but was of EcoRI (at 3 Å resolution)[5] in 1986, and although the chain tracing had to be revised in 1990[6], the first views of this complex generated considerable excitement. Swift to follow over the next few years were the high resolution cocrystal structures of the phage proteins and other prokaryotic transcriptional regulators: the DNA binding domain of 434 repressor[7,8] (Fig. 1), Trp repressor[9], 434 Cro[10], the DNA binding domain of λ repressor[11], λ Cro[12] and CAP[13]. These structures have withstood the test of time and yielded the fine details of the interactions that everyone had been anticipating thanks to the wealth of biochemical data on these proteins. Yet, in spite of these biochemical data, with each structure came new surprises and new themes — many of which are still being addressed in papers today, such as subtle and dramatic DNA conformational changes, specific and unexpected hydrogen bond contacts, specific van der Waals contacts and water mediated interactions with the DNA.

A major theme in these early papers — now rarely discussed, however — is the lack of a one-to-one amino acid to DNA base correspondence in protein–DNA interactions, something that is not at all surprising to those of us jaded by the very great number of protein–DNA complex structures now solved. But at that time, as the first sets of detailed interactions were beginning to emerge, there was a strong wish to develop a simple 'recognition code'. As Aneel Aggarwal puts it, "A protein–DNA interface is not all that different from other interfaces, so why did we anticipate a one-to-one correspondence? It can probably be traced back to the genetic code. There was a real mindset about a code; somehow, anytime you worked with DNA, you couldn't get away from the word 'code'. " Cynthia Wolberger echoes this feeling, "A lot of people were expecting a code, just in the way there was a code that specified one DNA base pairing with another". The simplicity of the DNA genetic code had convinced the community that something equally as simple would be found for protein recognition of DNA. Given the number of times the lack of a code is discussed in the early papers, its seems that researchers were indeed unwilling to give up the idea of a very simple recognition system. But, while some generalities about base recognition by certain residues can be made, a simple one-to-one correspondence has of course not materialized.

Although there was intense interest surrounding the structures of these site-specific DNA binding proteins, it seems there was little danger of getting scooped 'out of the blue'. Cloning had revolutionized structural biology, but even so, few labs had large enough research budgets to tackle these problems. According to Cynthia Wolberger, the main obstacle was getting enough synthetic DNA. Scientists had either to develop a collaboration with a chemist or to purchase the oligos for approximately $10,000 per short duplex, a very hefty sum in the early 1980's. Although the DNA was expensive, it nevertheless soon become apparent that a useful approach that could result in a wide variety of crystal resolutions was to try many different lengths and flanking sequences of DNA. This scheme was developed by Carl Pabo and was quickly adopted by the crystallographic community as a whole.

As more structures were determined, surprises continued to emerge, such as the strong and unexpected similarities between prokaryotic and eukaryotic transcription factors. Today, there is less excitement about DNA binding proteins in general, and sadly, young graduate students often groan at the thought of seminars with 'transcription' in the title. Nevertheless, the recently published high resolution structure of the nucleosome core[14] was greeted with great enthusiasm, giving us a look at the basic architecture of chromatin. Thus, although we have learned much at a rapid pace over the last ten years, we can, no doubt, look forward to being surprised again by DNA binding proteins in the future.

TLS

1. Anderson, W.F., Ohlendorf, D.H., Takeda, Y. & Matthews, B.W. *Nature* **290**, 754–758 (1981).
2. McKay, D.B. & Steitz, T.A. *Nature* **290**, 744–749 (1981).
3. Pabo, C.O & Lewis, M. *Nature* **298**, 443–447 (1982).
4. Anderson, J.E., Ptashne, M. & Harrison, S.C. *Nature* **316**, 596–601 (1985).
5. McClarin, J.A. *et al. Science* **234**, 1526–1541 (1986).
6. Kim, Y.C., Grable, J.C., Love, R., Green, P.J., Rosenberg, J.M. *Science* **249**, 1307–1309 (1990).
7. Anderson, J.E., Ptashne, M. & Harrison, S.C. *Nature* **326**, 846–852 (1987).
8. Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. & Harrison, S.C. *Science* **242**, 899–907 (1988).
9. Otwinowski, Z. *et al. Nature* **335**, 321–329 (1988).
10. Wolberger, C., Dong, Y., Ptashne, M. & Harrison, S.C. *Nature* **335**, 789–795 (1988).
11. Jordan, S.R. & Pabo, C.O. *Science* **242**, 893–899 (1988).
12. Brennan, R.G., Roderick, S.L., Takeda, Y. & Matthews, B. *Proc. Natl. Acad. Sci. USA* **87**, 8165–8169 (1990).
13. Schultz, S.C., Shields, G.C. & Steitz, T.A. *Science* **253**, 1001–1007 (1991).
14. Luger K., Mader A.W., Richmond R.K., Sargent D.F. & Richmond, T.J. *Nature* **389** 251–260 (1997).

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Fig. 1 Electron density map of one of the high resolution structures of the DNA binding domain of 434 repressor bound to its operator, showing the N-terminal half of the recognition helix in the major groove. Reprinted with permission from ref. 7.