TCF/LEF factors earn their wings

HANS CLEVERS (h.clevers@lab.azu.nl)

MARC VAN DE WETERING (m.vandewetering@lab.azu.nl)

The founding members of the TCF/LEF family are T-cell factor 1 (TCF1)^{1,2} and lymphoid-enhancer factor 1 (LEF1)^{3,4}. *TCF1* was cloned as a gene that encodes a T-cell-specific DNA-binding protein with affinity for the AACAAAG motif in the *CD3 ϵ* enhancer². To date, there is no evidence to indicate that TCF1 directly regulates the expression of the *CD3 ϵ* gene. The gene encoding LEF1 was independently cloned by two groups: murine LEF1 was isolated in a subtractive screen for precursor B-cell-specific factors and was subsequently found to be expressed by T cells as well. Mouse LEF1 binds to the TTCAAAGG sequence element in the *TCR α* enhancer³; The gene encoding human LEF1, originally termed TCF1 α , was cloned in a classical fashion. The protein was purified from the T-cell line Jurkat by DNA-affinity chromatography using a binding site from the HIV long-terminal repeat. Peptide sequencing subsequently facilitated the cloning of its cDNA (Ref. 4).

Two additional human family members, *TCF3* and *TCF4*, were cloned by 'guessmer' PCR using primers based on the conserved DNA-binding domain of TCF1 and LEF1 (Ref. 5). Extensive cloning efforts and database searches performed in our laboratory have failed to identify additional *TCF* genes in the human or murine genome. We have cloned a chicken *TCF* gene, which is intermediate in sequence between TCF1 and LEF1, and might represent an evolutionary ancestor of the two genes⁶. The recently established link between TCFs and Wingless/WNT signaling prompted the cloning of maternally expressed TCF homologs from the two model organisms *Xenopus* and *Drosophila*. The *Drosophila* gene *Pangolin* appears to be the only TCF-encoding gene in the *Drosophila* genome; Pangolin is most closely related to TCF1, particularly in its C-terminus^{7,8}. From *Xenopus* embryos, four virtually identical genes were cloned that are most closely related to mammalian *TCF3* (Ref. 9). In *C. elegans*, a mesoderm-specifying gene called *pop-1* was found to encode a distant member of the TCF/LEF family¹⁰. No other *TCF*-like genes are currently in the *C. elegans* database.

TCFs contain a single DNA-binding high mobility group (HMG) domain. Typically, the HMG domains of TCFs have more than 90% amino acid identity. Conservation of sequence occurs throughout the remainder of the proteins, although to a lesser extent. Only the C-termini of individual members, often generated by alternative splicing, appear to be unique to one or a few members of the family (Fig. 1).

TCFs control mammalian development

In adult mammals, *TCF1* is expressed uniquely in T lymphocytes, while *LEF1* is expressed in T cells and early B cells. During murine development, however, the expression of *TCF1* and *LEF1* occurs in complex overlapping patterns in many tissues¹¹. The unique *in vivo* functions of *TCF1* and *LEF1* have been explored by gene-disruption experiments. *Tcf1*^{-/-} knockout mice are severely impaired in the generation of T cells, but are otherwise normal¹². *Lef1*^{-/-} mice lack hair, teeth, mammary glands and trigeminal nuclei and, as a consequence, die around birth¹³. As deduced from direct analyses and from transplantation experiments, the *Lef1* mutation has no major effects on the immune system, besides the

Factors of the TCF/LEF HMG domain family (TCFs) exist in vertebrates, *Drosophila melanogaster* and *Caenorhabditis elegans*. It has very recently become evident that TCFs interact with the vertebrate WNT effector β -catenin to mediate axis formation in *Xenopus*. Likewise, Armadillo (the *Drosophila* ortholog of β -catenin) is genetically upstream of a *Drosophila* TCF in the Wingless pathway. Upon Wingless/Wnt signaling, Armadillo/ β -catenin associate with nuclear TCFs and contribute a trans-activation domain to the resulting bipartite transcription factor. The cytoplasmic tumor-suppressor protein APC binds to β -catenin causing its destruction. In APC-deficient colon carcinoma cells, β -catenin accumulates and is constitutively complexed with TCF factors. In APC-positive colon carcinomas and melanomas, dominant mutations in β -catenin render it indestructible, providing an alternative mechanism to activate transcription of TCF target genes inappropriately. So, transcriptional activation of TCF target genes by β -catenin appears to be a central event in development and cellular transformation.

absence of the B1 B-cell subset. In *Tcf1 Lef1* double-knockout mice, development of T cells is completely abrogated, indicating that *Lef1* can, to some degree, substitute for *Tcf1* in T-cell differentiation (R. Okamura, H. Clevers and R. Grosschedl, unpublished).

Binding and bending

Sequence-specific HMG domains (which occur in the mammalian sex-determination factor SRY, in the related SOX proteins, in some fungal mating-type proteins and in TCFs) bind to DNA in an unusual fashion. Biochemical analyses indicate that the predominant base contacts occur in the minor groove^{14,15}, and that a dramatic bend was introduced in the DNA helix¹⁵. NMR structural analyses have been performed on the non-complexed SOX4 HMG domain¹⁶, and on LEF1 and SRY HMG domains complexed to DNA (Refs 17, 18). The overall structure of a sequence-specific HMG domain resembles that of non-sequence-specific DNA-binding domains of the high mobility group proteins HMG1 and HMG-D. HMG domains comprise three α -helices arranged in an L-shape. The concave side of the L occupies the minor groove of the recognition motif and a hydrophobic amino acid, located in the first α -helix, intercalates between two AT base pairs (bp). As a result, the minor groove is widened and a dramatic bend of 70–130° is created^{17,18} (Fig. 2).

Three subfamilies of sequence-specific HMG boxes have been identified by computer comparisons: the TCF/LEF subfamily, the SRY/SOX subfamily, and the fungal mating-type subfamily¹⁹. Initial studies have reported 6–8 bp recognition motifs for HMG-domain proteins. Despite the wide divergence of sequence between subfamilies, all HMG-domain proteins bind to

FIGURE 1. Structural organization of the T-cell factor (TCF) proteins. The best-conserved regions are the β -catenin/Armadillo binding domain (blue) and the DNA-binding HMG domain (red). These are separated by a region of lower similarity that, in LEF1, contains a CAD (pink). A third domain, conserved throughout evolution, is located between the HMG box and the specific C-termini. This domain (yellow), the function of which is currently unknown, is found in TCF1E, TCF4E, PAN/dTCF and POP-1. The C-termini of TCF1B, TCF4B and chicken TCF also show a high degree of conservation (green). Prefixes: ch, chicken; d, *Drosophila*; h, human; X, *Xenopus*.

highly similar AT-rich motifs of the canonical sequence [A/T][A/T]CAA[A/T]GG. More recently, we have performed PCR-based site selection (D. Dooijes and H. Clevers, unpublished) and found conservation of a much larger 11 bp motif for several different HMG domains, for example AAGATCAAAGG for the HMG domain of Pangolin⁷.

Context-dependent *trans*-activation

Initial experiments indicated that TCF/LEF factors do not behave as classical transcription factors, in the sense that they cannot activate transcription from reporters containing multimerized binding sites. LEF1 was found to activate transcription from the *TCR α* enhancer in a fashion that is strictly dependent on adjacently bound unrelated factors. This context-dependent activity is mediated by two distinct regions in LEF1, one exerting an architectural function and the other a transcription activation function. The architectural role is mediated by the HMG domain, through the induction of a sharp bend in the DNA helix. This bend is believed to facilitate interactions between adjacently bound ETS1, AML1 and ATF/CREB transcription factors²⁰. The transcriptional activation function is mediated by a context-dependent activation domain (CAD) in the LEF1 protein^{21,22}, which binds to and is dependent on the context-dependent, ubiquitously expressed co-activator ALY (Ref. 23). Importantly, the β -catenin-binding N-terminus of LEF1 is dispensable for the context-dependent *trans*-activation of the *TCR α* enhancer²¹. TCF1 has no obvious CAD, as evidenced by its inability to activate the *TCR α* enhancer²⁴.

In line with this, ALY has been reported not to interact with TCF1 (Ref. 23). Currently, it is not known whether any of the other mammalian TCFs resemble LEF1 in its capacity to activate transcription in a context-dependent fashion. Analysis of the *Ubx* enhancer in *Drosophila* indicates that a TCF/LEF binding site in this enhancer acts strictly context-dependently (see below).

β -catenin/Armadillo interacts with TCFs

Independent two-hybrid screens revealed a highly unexpected association of TCFs with the WNT/Wingless signal-transduction pathway. Genetic studies on Wingless-controlled segment polarity in *Drosophila* have led to the identification of a number of genes that constitute a linear cascade²⁵ (Fig. 3). At the top of this cascade, secreted Wingless interacts with the seven-transmembrane receptor Frizzled2 (FZ2)²⁶. This interaction activates a cytoplasmic protein of unknown biochemical function, Dishevelled²⁷, which might be bound to the C-terminus of FZ2 via a PDZ domain. Activated Dishevelled inhibits a serine kinase termed Zeste-White-3 (ZW3)²⁸. This kinase normally inhibits the action of Armadillo, probably by inducing the rapid breakdown of the latter protein. Wingless signaling inactivates ZW3, thus leading to a build-up of cytoplasmic Armadillo. It was believed that the accumulation of Armadillo, and possibly its nuclear translocation, would somehow result in genetic reprogramming²⁹. With these genetic data in hand, the *Xenopus* axis-duplication assay provided a rapid test for the involvement of vertebrate homologs of the *Drosophila* genes in WNT signaling. Thus, a striking symmetry was observed between the Wingless and WNT cascades^{25,30,31} (Fig. 3).

LEF1 was identified in a yeast two-hybrid screen for proteins interacting with β -catenin³², the most downstream-known molecule in the vertebrate WNT cascade. LEF1 interacts through its N-terminus with the Armadillo-repeat region of β -catenin. Co-injection into *Xenopus* embryos of β -catenin with LEF1 resulted in the translocation of β -catenin to the nucleus^{32,33}. Also, injection of LEF1 RNA in the ventral side of the embryo generated a secondary axis, while a dominant-negative version of LEF1 could inhibit formation of the endogenous axis. Binding of β -catenin to LEF1 altered DNA bending, providing a possible mechanism by which β -catenin could affect genetic programming³².

In a reverse two-hybrid strategy using TCF1 as a bait, we cloned the gene encoding β -catenin⁹. Having isolated four virtually identical *TCF3* orthologs from *Xenopus* embryos, we proceeded to determine a possible function of XTcf-3 in axis duplication. XTcf-3 is expressed maternally and ubiquitously in the embryo. It can bind β -catenin and, when overexpressed, XTcf-3 transports β -catenin to the nucleus. Unlike LEF1 however, overexpression of XTcf-3 has no effect on axis formation. A dominant-negative XTcf-3 inhibits β -catenin-induced axis duplication as well as formation of the endogenous axis. Like all other TCFs, XTcf-3 is incapable of activating transcription from multimerized DNA-binding motifs. However, cotransfection with β -catenin endows XTcf-3 with potent *trans*-activational properties in a context-independent fashion. This initial observation suggested a simple molecular mechanism by which β -catenin affects genetic programming, that is, through the direct activation of XTcf-3 target genes⁹.

The observation that overexpression of XTcf-3 has no effect, while LEF1 readily induces axis duplication in the *Xenopus* assay is somewhat puzzling. The dominant-positive effect of injected murine LEF1 on axis formation in *Xenopus* might reflect the heterologous nature of the test system. Alternatively, it might relate to the activity of the CAD domain that appears to be present only in LEF1.

From frogs to flies

Experiments in *Xenopus* are rapid, but rely on the overexpression of wild-type or mutant molecules that might interact in complex ways with the endogenous pools of WNT cascade proteins. Nevertheless, in this case, the *Xenopus* WNT pathway yielded a component not yet uncovered by *Drosophila* geneticists. The remarkable symmetry between the WNT and Wingless pathways dictated the existence of a *Drosophila* TCF homolog (dTCF). We succeeded in cloning such a gene, which mapped to the small chromosome 4, a distinctly unpopular chromosome among *Drosophila* researchers because of its lack of recombination. dTCF is contributed maternally and ubiquitously expressed in fly embryos. dTCF readily interacted with Armadillo (and β -catenin), and formation of the Armadillo-dTCF complex yielded a bipartite transcription factor, capable of activating transcription from multimerized TCF-binding motifs. The C-terminus of Armadillo, previously found to be crucial for signaling but dispensable for its adhesive functions, was found to constitute a classical *trans*-activation domain⁷.

By reverse genetics, we identified dTCF mutations in a complementation group with an *armadillo*-type segment polarity phenotype that had previously been assigned to the neighbouring gene *cubitus interruptus*. Genetically, the gene encoding dTCF was downstream from Armadillo. Furthermore, dTCF mutations affected the expression of the Wingless target genes *engrailed* and *Ultrabithorax*⁷.

The same gene was independently identified in a screen for dominant suppressors of a Wingless-induced rough eye phenotype⁸. Of four suppressors, two are recessive alleles of Armadillo, while the other two define a novel complementation group termed *Pangolin* on chromosome 4. It was correctly guessed that the gene was a member of the *Tcf* family. Essentially identical genetic observations, as described above, were made. Interestingly, one of the mutations is temperature-sensitive. Shifting flies to the non-permissive temperature at different developmental stages demonstrates that many Wingless-driven events are dependent on Pangolin.

Using a different approach, another study identified a Wingless-response element in the midgut enhancer of *Ubx* (Ref. 34). This element conspicuously resembled a TCF1/LEF1-binding site. In transgenic flies, moderate levels of murine LEF1 activated the *Ubx* enhancer in an Armadillo-dependent fashion. Overexpression of murine LEF1 bypassed the need for Wingless signaling and mimicked the *wingless* gain-of-function mutations in

FIGURE 2. Stereo view of the HMG domain of lymphoid-enhancer factor 1 (LEF1) complexed with DNA (Ref. 17). The HMG domain is represented by the red tube drawn through the backbone N, C α and C' atoms. The pronounced bend in the DNA (with its two strands depicted in pink and purple) of 120° is clearly visible (reproduced, with permission, from Ref. 17).

midgut, notum and wing. These data fit well with those of the other two studies. Interestingly, the minimal *Ubx* midgut enhancer contains another sequence element: a Decapentaplegic (DPP, a TGF- β family member) response sequence, which is a cAMP-responsive element

FIGURE 3. A model for T-cell factor (TCF)-dependent transcription, regulated by Wingless/WNT1. In the absence of Wingless, the serine kinase ZW3/GSK-3 β is active, resulting in the destruction of Armadillo/ β -catenin by the GSK-3 β -APC complex. TCF might bind to its target genes but cannot *trans*-activate transcription. After binding of Wingless/WNT, signaling through the receptor FZ2 results in the inactivation of ZW3/GSK-3 β by Dishevelled. As a consequence, Armadillo/ β -catenin is released from APC and translocates to the nucleus. It associates with TCF and the resulting bipartite transcription factor induces the transcription of TCF target genes.

(CRE). The *trans*-activation of the *Ubx* enhancer by LEF1 was strictly dependent on concomitant DPP signaling. Thus, LEF1 could help coordinate inputs from two positional signals: Wingless and DPP. The anatomy of the *Ubx* enhancer and the effect of LEF1 is strikingly reminiscent of that of the *TCR α* enhancer, which contains both a LEF-binding site and a CRE (Refs 20, 21, 34).

In *C. elegans*, *pop-1* was cloned by a genetic approach as a mesoderm specifier at the four-cell stage¹⁰. At this stage, the precursor cell EMS divides into the E and MS blastomeres. A maternal-effect lethal mutation in the *pop-1* gene, which like *XTcf-3* and *pan* is maternally expressed, resulted in the adoption of an E fate by the MS blastomere. The N-terminus of POP-1 has similarity to that of the other TCFs. In addition, several β -catenin homologs are present in the *C. elegans* database. This makes it likely that mesodermal fate is determined by WNT signals, relayed through POP-1.

TCFs and cancer

The interaction of TCFs with β -catenin have implications for yet another widely studied biological phenomenon: the transformation of colon epithelial cells following loss of function of the tumor suppressor protein APC (Ref. 35). APC forms complexes with GSK-3 β and β -catenin, resulting in the rapid breakdown of β -catenin. Loss of APC results in increased pools of free cytoplasmic β -catenin^{36,37}. We found that TCF4 is specifically expressed in colonic epithelium. Upon loss of APC, a constitutive β -catenin–TCF4 complex was present in the nucleus of colon carcinoma cells, concomitant with the active transcription of TCF reporter constructs. Reintroduction of APC removed β -catenin from the nucleus and abrogated the spontaneous transcriptional activity of TCF reporter genes³⁸.

So, if deregulation of TCF-target-gene expression is a pivotal event in carcinogenesis, what is the status of the pathway in APC-positive colon carcinomas? Notably, in such cells, the spontaneous transcriptional activity of TCF-reporter genes was still observed and this could be attributed to dominant mutations in the N-terminus of β -catenin. Again, these mutations resulted in the formation of nuclear β -catenin–TCF complexes³⁹. Similar mutations were found in malignant melanoma, inducing a prolonged half-life of β -catenin and the formation of constitutive complexes with LEF1 (Ref. 40). This region is critical for the normal negative regulation of Armadillo/ β -catenin by ZW3 (Refs 41, 42). Together, these data indicate that at least two different oncogenic mutations, activating mutations in β -catenin and loss-of-function mutations in APC, lead to the same transcriptional deregulation: the constitutive transcription of TCF-target genes. This deregulation is likely to be a central event in the malignant transformation of colon epithelium, melanocytes and probably other unidentified cell types.

Concluding remarks

The identification of the interaction with β -catenin/Armadillo has given TCFs an unexpected dimension. The availability of reporter-gene assays, based on the TCF-binding motif, will greatly accelerate the detailed analysis of the WNT-signaling cascade in various biological systems. Many unanswered issues spring to mind: what are the individual roles of different TCF genes within one

species? Can all TCF proteins function like LEF1 to integrate different signals at the enhancer level from more than one pathway? Is TCF-target-gene transcription deregulated in other types of cancer? What is the nature of the physiological process that is normally controlled by β -catenin/TCF4 in colon epithelium? These questions and many more will undoubtedly be answered in the near future.

Stop Press...

The Wingless/WNT field is taking off

Four very recent papers address a number of important issues. First, by X-ray crystallography, the three-dimensional structure of the Armadillo repeat region of murine β -catenin was solved¹. The 12 Armadillo repeats constitute a single structural unit. Individual repeats consist of three α -helices and are tightly packed into a superhelix, which features a long, positively charged groove. This groove probably accommodates the negatively charged interaction domains of cadherins, TCFs and APC.

Second, good evidence was provided for the identification of a direct target gene for XTcf-3. *Siamois* is the first known zygotic gene to be expressed in response to WNT signaling in the *Xenopus* embryo. Utilizing a transient reporter gene assay in *Xenopus* embryos, it was shown that the promoter of *Siamois* contains several functional XTcf-3 binding motifs², making it the first WNT-responsive TCF target gene in vertebrates.

Third, two recent papers in *Cell* establish the existence of a full set of WNT cascade genes in *C. elegans* that cooperate with the TCF homolog POP-1 to specify the fates of the E and MS daughter cells of the four-cell stage EMS blastomere^{3,4}. These *C. elegans* genes encode proteins that are related to WNT, frizzled, β -catenin and APC, respectively. Somewhat puzzling are the apparent opposite effects of the APC, β -catenin and TCF homologs, when compared with the prototypic pathways as described in the frog and fruitfly model systems. Nevertheless, the power of *C. elegans* genetics can now be added to already existing tools for the study of Wingless signaling.

- 1 Huber, A.H., Nelson, W.J. and Weis, W.I. (1997) *Cell* 90, 871–882
- 2 Brannon, M. et al. (1997) *Genes Dev.* 11, 2359–2370
- 3 Thorpe, C.J. et al. (1997) *Cell* 90, 695–705
- 4 Rocheleau, C.E. et al. (1997) *Cell* 90, 707–716

Acknowledgements

We thank M. Peifer, J. Bos, K. Georgopoulos and M. Schilham for their comments on the manuscript, and P. Wright for making available Fig. 2.

References

- 1 Oosterwegel, M. et al. (1991) *J. Exp. Med.* 173, 1133–1142
- 2 van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) *EMBO J.* 10, 123–132
- 3 Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991) *Genes Dev.* 5, 880–894
- 4 Waterman, M.L., Fischer, W.H. and Jones, K.A. (1991) *Genes Dev.* 5, 656–669
- 5 Castrop, J., van Norren, K. and Clevers, H.C. (1992) *Nucleic Acids Res.* 20, 611

- 6 Castrop, J., Hoevenagel, R., Young, J.R. and Clevers, H.C. (1992) *Eur. J. Immunol.* 22, 1327–1330
- 7 van de Wetering, M. *et al.* (1997) *Cell* 88, 789–799
- 8 Brunner, E., Peter, O., Schweizer, L. and Basler, K. (1997) *Nature* 385, 829–833
- 9 Molenaar, M. *et al.* (1996) *Cell* 86, 391–399
- 10 Lin, R., Thompson, S. and Priess, J.R. (1995) *Cell* 83, 599–609
- 11 Oosterwegel, M. *et al.* (1993) *Development* 118, 439–448
- 12 Verbeek, S. *et al.* (1995) *Nature* 374, 70–74
- 13 Van Genderen, C. *et al.* (1994) *Genes Dev.* 8, 2691–2704
- 14 van de Wetering, M. and Clevers, H. (1992) *EMBO J.* 11, 3039–3044
- 15 Giese, K. and Grosschedl, R. (1992) *Cell* 69, 185–195
- 16 van Houte, L. *et al.* (1995) *J. Biol. Chem.* 270, 30516–30524
- 17 Love, J.J. *et al.* (1995) *Nature* 376, 791–795
- 18 Werner, H.M., Huth, J.R., Gronenborn, A.M. and Clore, G.M. (1995) *Cell* 81, 705–714
- 19 Laudet, V., Stehelin, D. and Clevers, H. (1993) *Nucleic Acids Res.* 21, 2493–2501
- 20 Giese, K., Kingsley, C., Kirshner, J.R. and Grosschedl, R. (1995) *Genes Dev.* 9, 995–1008
- 21 Carlson, P., Waterman, M. and Jones, K. (1993) *Genes Dev.* 7, 2418–2430
- 22 Giese, K. and Grosschedl, R. (1993) *EMBO J.* 12, 4667–4676
- 23 Bruhn, L., Munnerlyn, A. and Grosschedl, R. (1997) *Genes Dev.* 11, 640–653
- 24 van de Wetering, M., Castrop, J., Korinek, V. and Clevers, H. (1996) *Mol. Cell. Biol.* 16, 745–752
- 25 Orsulic, S. and Peifer, M. (1996) *Curr. Biol.* 6, 1363–1367
- 26 Bhanot, P. *et al.* (1996) *Nature* 382, 225–230
- 27 Yanagawa, S. *et al.* (1995) *Genes Dev.* 9, 1087–1097
- 28 Cook, D. *et al.* (1996) *EMBO J.* 15, 4526–4536
- 29 Orsulic, S. and Peifer, M. (1996) *J. Cell Biol.* 134, 1283–1300
- 30 Miller, J.R. and Moon, R.T. (1996) *Genes Dev.* 10, 2527–2539
- 31 Moon, R.T., Brown, J.D., Yang-Snyder, J.A. and Miller, J.R. (1997) *Cell* 88, 725–728
- 32 Behrens, J. *et al.* (1996) *Nature* 382, 638–642
- 33 Huber, O. *et al.* (1996) *Mech. Dev.* 59, 3–10
- 34 Riese, J. *et al.* (1997) *Cell* 88, 777–787
- 35 Kinzler, K.W. and Vogelstein, B. (1996) *Cell* 87, 159–170
- 36 Rubinfeld, B. *et al.* (1996) *Science* 272, 1023–1026
- 37 Moon, R.T. and Miller, J.R. (1997) *Trends Genet.* 13, 256–258
- 38 Korinek, V. *et al.* (1997) *Science* 275, 1784–1787
- 39 Morin, P.J. *et al.* (1997) *Science* 275, 1787–1790
- 40 Rubinfeld, B. *et al.* (1997) *Science* 275, 1790–1792
- 41 Yost, C. *et al.* (1996) *Genes Dev.* 10, 1443–1454
- 42 Pai, L.M., Orsulic, S., Bejsovec, A. and Peifer, M. (1997) *Development* 124, 2255–2266

H. Clevers and **M. van de Wetering** are in the Department of Immunology, University Hospital Utrecht, PO Box 85500, 3508GA, Utrecht, The Netherlands.

'I confess that I cannot recall any case within my experience which looked at the first glance so simple, and yet which presented such difficulties.'
The Man with the Twisted Lip¹.

'Chance has put in our way a most singular and whimsical problem, and its solution is its own reward.'
The Adventure of the Blue Carbuncle¹.

The centromere is cytologically visible as the primary constriction of the chromosome, and is associated with the trilaminar kinetochore, which serves as the key site of attachment between the chromosome and the spindle during mitosis and meiosis². Centromere function is essential for mediating chromosome movement along microtubules, and also for normal sister-chromatid cohesion and separation³. In this review we attempt to bring together information about the structure and function of the centromere in different eukaryotic systems; we emphasize the emerging, surprising view that epigenetic mechanisms might play a key role in determining centromere identity and regulation. The term 'epigenetic' has acquired different meanings in different contexts. Here, we use epigenetic to mean heritable changes in gene or centromere activity without a corresponding change in primary DNA sequence.

Centromeric DNA sequence is not conserved in different eukaryotes

A mere decade ago it seemed so elementary. The *cis*-acting DNA sequences required to confer complete centromere function had been dissected in great detail in

The case for epigenetic effects on centromere identity and function

GARY H. KARPEN (karpen@salk.edu)

ROBIN C. ALLSHIRE (robin.allshire@hgu.mrc.ac.uk)

The centromere is required to ensure the equal distribution of replicated chromosomes to daughter nuclei. Centromeres are frequently associated with heterochromatin, an enigmatic nuclear component that causes the epigenetic transcriptional repression of nearby marker genes (position-effect variegation or silencing). The process of chromosome segregation by movement along microtubules to spindle poles is highly conserved, yet the putative cis-acting centromeric DNA sequences bear little or no similarity across species. Recently, studies in several systems have revealed that the centromere itself might be epigenetically regulated and that the higher-order structure of the underlying heterochromatin contributes to centromere function and kinetochore assembly.

the unicellular budding yeast *Saccharomyces cerevisiae*. The *S. cerevisiae* minimal centromere (CEN) is only 125 bp in length, and all 17 centromeres contain the same three functionally distinct elements (CDEs I, II and III) (Fig. 1). Thus, centromere function in this yeast is known to be mediated by specific DNA sequences, some of which