Nucleic Acids Research, 1995, Vol. 23, No. 21 4229–4233

The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin

Eugene V. Koonin*, Shubo Zhou[1] and John C. Lucchesi[1]

National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20894, USA and [1]Department of Biology, Emory University, Atlanta, GA 30322, USA

Received September 9, 1995; Accepted September 19, 1995

ABSTRACT

Using computer methods for detecting conserved amino acid sequence motifs, we show that the chromatin organization modifier (chromo) domain that has been previously identified in several proteins involved in transcription down-regulation is present in a much larger group of (putative) chromatin-binding proteins, some of which are positive rather than negative regulators of transcription. The most interesting new members of the chromo superfamily are *Drosophila* male-specific lethal (MSL-3) protein involved in the X chromosome gene dosage compensation in the males and human retinoblastoma-binding protein RBP-1. We show that the chromo domain is duplicated in several chromatin-binding proteins and use this observation to interpret recent results on chromatin binding obtained with chimeric chromo domain- containing proteins. We hypothesize that the chromo domain may be a vehicle that delivers both positive and negative transcription regulators to the sites of their action on chromatin.

INTRODUCTION

Recent estimates have strongly suggested that the majority of highly conserved proteins sequence motifs may be already known (1–2). Therefore identification of new occurrences of these motifs, which frequently involves detection of subtle sequence signals, is becoming an increasingly important component of protein function prediction and classification. The present paper significantly expands a previously described family of domains involved in transcription regulation, resulting in new structural and functional predictions.

Chromatin organization modifier (chromo) is a 30–50 amino acid domain that is conserved in several eukaryotic chromatin-binding proteins such as *Drosophila* heterochromatin protein 1 (HP1) and Polycomb (PC), their mammalian homologs and fission yeast SWI6 (4–8). Recently, the chromo domain has been identified in two larger, multidomain proteins, namely the murine CHD-1 (9) and *Drosophila* Su(var)3-9 (10). Heterochromatin protein 1 and PC, which may be considered 'classical' chromo-containing proteins, bind to numerous, specific, non-overlapping loci in chromatin (11–13). These proteins have been implicated, in the transcription down regulation associated with the heterochromatic position-effect variegation and in the repression of the transcription of homeotic genes by chromatin packaging, respectively (5,14). A similar function in the repression of silent mating type loci in fission yeast has been proposed for SWI6 (8). Mutations in the chromo domain abolish the binding of the PC protein to chromatin (7). In contrast, the chromo domain of HP1 is dispensable for chromatin binding, whereas the C-terminal portion that is highly conserved in HP1 homologs from different organisms is sufficient for nuclear localization and site-specific binding to the heterochromatin (15). The difference between the results obtained with HP1 and PC leaves a degree of uncertainty regarding the actual role of the chromo domain in chromatin binding and transcription regulation (15). Interestingly, the chromo domain is a specific target for autoimmune response in scleroderma patients (7).

Here we show that the chromo domain is present in a larger class of (putative) chromatin-binding proteins than previously suspected and is duplicated in some of these proteins. These findings explain the difference in the results of the experiments on the role of the chromo domain in PC and HP1 and lead to some generalizations on the structure and function of the chromo domain-containing proteins.

MATERIALS AND METHODS

Amino acid sequences were from the non-redundant (NR) protein sequence database at the National Center for Biotechnology Information (NIH). Nucleotide sequences of expressed sequence tags (EST) were from the dBEST database (16). Initial comparisons of protein sequences to the protein NR database were performed using the BLASTP program (17), and comparisons to the dBEST were performed using the TBLASTN program (18). Nucleotide sequences translated in six reading frames were compared to the protein NR database using the BLASTX program (18). The BLAST programs were used in conjunction with the SEG program in order to mask the low complexity (compositionally biased) regions in protein sequences that tend to produce artifactual alignments in database searches (18,19). Additional database searches were performed using the BLITZ

* To whom correspondence should be addressed

a

```
secondary structure     aaaaaaaaaaaalleeeeeeeellllllllllllllll
consensus                ..E.UU.....G.UEOUUKOJGU.......W.....U
MSL3/DROME          37:LNVFERRNEHGLRFYEYKIHFQGWRPS-YDRCVRATVL   S48828
MSL3/DROME         453:LNASPISNKKLEDLLPHLDAFINYLEN-HREWFDRENF   S48828
RBP1/HUMAN         592:YEASIKSTEIDDGEVLYLVHYYGWNVR-YDEWVKADRI   S66427
EST/HUMAN            ?:YEAKCVKVAIKDKQVKYFIHYSGWNKN-WDEWVPESRV   T36139
EST/RICE             ?:----VQKSENKEDEWRYHVHYLGWSKS-WDEWVTNDRL   D24751
HP1_DROVI           26:AVEKILDRRVRKGKVEYYLKWKGYAET-ENTWEPEGNL   P29227
HP1_DROVI          146:EAEKILGASDNNGRLTFLIQFKGVDQA-EMVPSTVANV   P29227
HP1/DROME           26:AVEKIIDRRVRKGKVEYYLKWKGYPET-ENTWEPENNL   M57574
HP1/DROME          146:EAEKILGASDNNGRLTFLIQFKGVDQA-EMVPSSVANE   M57574
HP1/HUMAN           22:VVEKVLDRRVVKGQVEYLLKWKGFSEE-HNTWEPEKNL   L07515
HP1/HUMAN          121:EPEKIIGATDSCGDLMFLMKWKDTDEA-DLVLAKEANV   L07515
EST/PLAFA            ?:EIGDILEIKKKKNGFIYLVKWKGYSDD-ENTWEPESNL   T02521
EST/SCHMA            ?:QVEKILKVRIRNGRKEYFLKWKGYSEE-DNTWEPEENL   T14583
MOD1_MOUSE          22 VVEKVLDRRVVKGKVEYLLKWKGFSDE-DNTWEPEENL   P23197
MOD1_MOUSE         118 EPERIIGATDSSGELMFLMKWKNSDEA-DLVPAKEANV   P23197
MOD2_MOUSE          22:VVEKVLDRRVVNGKVEYFLKWKGFTDA-DNTWEPEENL   P23198
MOD2_MOUSE         113:DPERIIGATDSSGELMFLMKWKDSDEA-DLVLAKEANM   P23198
MOD3_MOUSE          14:AAECILSKRLRKGKLEYLVKWRGWSSK-HNSWEPEENI   P30658
YO82_CAEEL          10:TVESILEHRKKKGKSEFYIKWLGYDHT-HNSWEPKENI   P34618
PC_DROME            28:AAEKIIQKRVKKGVVEYRVKWKGWNQR-YNTWEPEVNI   P26017
POL/SKIPP         1238:EAEAIRDTRKINGQREYLIKWKNYPEN-ENTWEPPKHL   L34658
ORF/CfT-I           86:ENEFEVEKILDKKGQRYLVKWKGYDES-ENTWEPRINL   S23571
CHD1_MOUSE         301:DPNAGFERNKEPGDIQYLIKWKGWSHI-HNTWETEETL   P40201
CHD1_MOUSE         392:RIIAHSNKQSAAGLPDYYCKWQGLPYS-ECSWEDGALI   P40201
SU(VAR)3-9/DROME   221:VVERIECVEMDQYQPVFFVKWLGYHDS-ENTWESLANV   S46942
T09A5.8/CAEEL       26:EVEKILAHKVTDNLLVLQVRWLGYGAD-EDTWEPEEDL   Z36753
YEZ4_YEAST         198:LEKTVPDLNNCKENYEFLIKWTDESHL-HNTWETYESI   P32657
YEZ4_YEAST         292:IDSQRASLEDGTSQLQYLVKWRRLNYD-EATWENATDI   P32657
A33_PLEWA           65:AAEIIKTRKTKENAEEFYVHYVGLNRR-QNEWVDKSRV   Q02084
XNF7/XENLA          39:AEIVKVRYNKQAGREEYYAHYVGLNRR-QNEWVDKSRL   A43906
URO2/MOLOC         160:AWYKARIIEVDNTKSEILVHFDGWRAK-FDTWYPMTSA   L19339
L8039.3/YEAST     1266:EFAFFVNNQGWRKKESHIIHWEGEPQK-FAIWYPYILA   U19103
SWI6_SCHPO          80:YVVEKVRMARKGGGYEVLLKWEGYDDPSDNTWSSEADC   P40381
SWI6_SCHPO         273:VSSIDTIERKDDGTLEIYLTWKNGAISHHPSTITNKCC   P40381
```

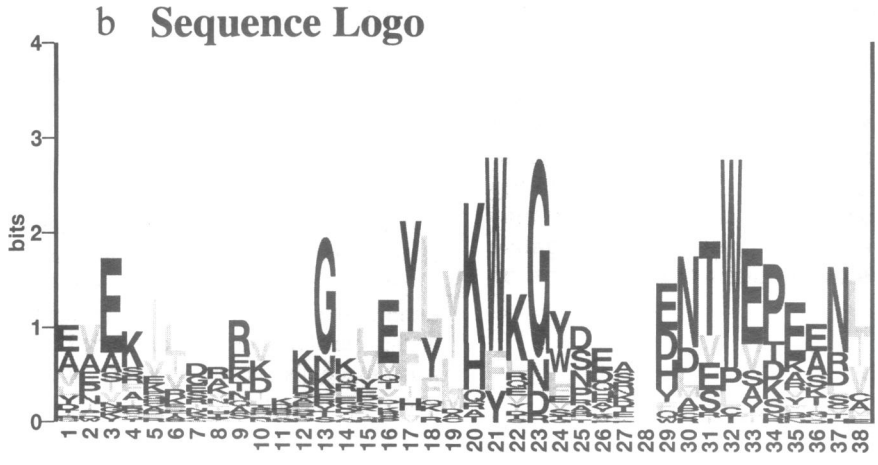

Figure 1. Sequence conservation in the chromo domain. (**a**) Multiple alignment of the chromo domain sequences. The comparisons performed using the GIBBS and MACAW programs indicated that the originally defined chromo domain size of 37 amino acid residues (5) was indeed optimal in terms of the statistical significance. The alignment was generated using the GIBBS program, with the SWI6 sequence added using the MACAW program. Duplicates of the chromo domain from the same protein sequence are grouped together. The consensus includes amino acid residues that are conserved in >1/2 of the aligned sequences. The residues conforming to the consensus are highlighted by bold type. U indicates a bulky hydrophobic residue (I, L, V, M, F, Y or W), O indicates an aromatic residue (F, Y or W) and dot indicates any residue. The numbers preceding the sequences indicate the position of the first aligned residue in each protein. The database accession numbers are given in the rightmost column. The numbers starting with P or Q are from the SWISS-PROT database, the numbers starting with A or S are from the PIR and the remaining numbers are from GenPept. The consensus secondary structure prediction derived from the prediction for all individual sequences is shown above the alignment, a indicates α-helix, e indicates extended (β) conformation and l indicates loop. The accuracy of prediction is expected to be ~72% for sequences that have several significantly similar homologs in the SWISS-PROT database and a few percentage points lower for those sequences that do not have such homologs (27). DROVI, *Drosophila viridis*; DROME, *Drosophila melanogaster*; CAEEL, *Caenorhabditis elegans*; SKIPPY, a *Fusarium oxysporum* (a phytopathogenic fungus) retrotransposon; Cft-I, a *Cladosporum fulvum* (a phytopathogenic fungus) retrotransposon; PLEWA, *Pleurodeles waltlii* (newt); XENLA, *Xenopus laevis*; PLAFA, *Plasmodium falciparum*; SCHMA, *Schistosoma mansoni*. MSL3 showed additional statistically significant similarity with an uncharacterized human ORF product (GenBank D14812), as well as a human (GenBank Z20674) and rice (GenBank D15421) ESTs. The region of similarity overlapped the distal chromo domain of MSL3 but the principal conserved motif was located upstream of the sequence shown in the figure (data not shown). This motif was not found in any other available sequence; additional sequence data are required to predict its function. Several alternatively spliced forms of RBP1 with large size differences have been identified (27–29). The indicated position of the chromo domain is for the longest form. The chromo domain has been additionally identified in a number of ESTs that are highly similar to known proteins and were not included in the alignment. Two sequences of EST-encoded putative proteins from *Schistosoma* and *Plasmodium* are shown since the chromo domain has not been previously identified in these organisms. (**b**) Information content profile: a sequence Logo. The profile was generated from the alignment shown in (a) using the ALPRO and MAKELOGO programs. The horizontal axis shows the position in the alignment and the vertical axis shows the information content in bits.

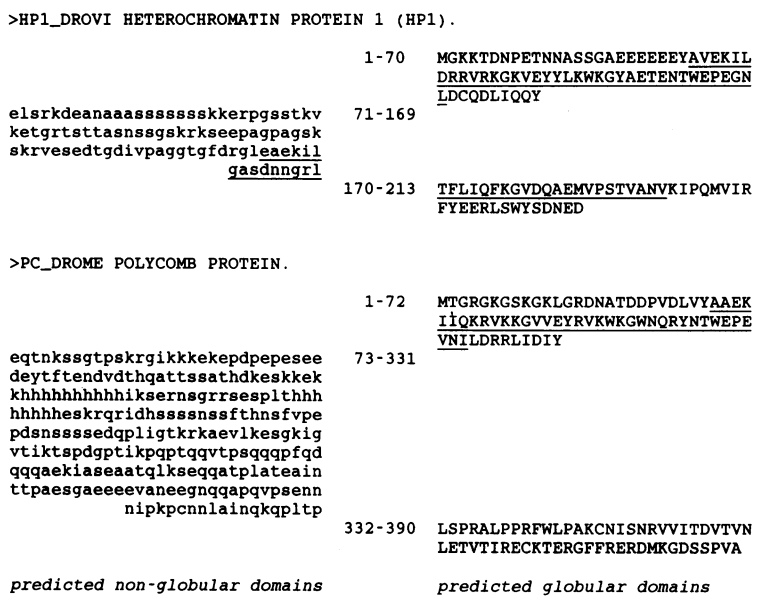

Figure 2. Partitioning of the chromo domain-containing proteins into globular and non-globular domains. The globular and non-globular domains were delineated using the SEG program with the parameters W = 45 $K_2(1) = 3.4$ $K_2(2) = 3.75$ that have been shown to predict non-globular domains in proteins with known tertiary structure with a high precision (25).

server which implements the MPSrch program (20). Multiple alignment blocks were constructed directly from BLAST outputs using the CAP program and screening of the NR database with position-dependent weight matrices generated from the derived blocks was performed using the MoST program (21). For protein superfamily delineation, alternate rounds of BLAST and MoST search were used iteratively (22). Additional protein sequence multiple alignment analysis was performed using the GIBBS (23) and MACAW (24) programs. In order to delineate probable non-globular domains in protein sequences, the SEG program was used with the parameters optimized for this purpose (25). The information content profiles (sequence Logo) for multiple amino acid sequence alignments were constructed using the ALPRO and MAKELOGO programs (26) through the WWW server supported by Steven Brenner at the University of Cambridge School of Biological Sciences. Protein secondary structure was predicted using the PHD program (27).

RESULTS AND DISCUSSION

Sequence conservation in the chromo domain: identification of new superfamily members and intramolecular duplication

The present study of the chromo domain was triggered by our analysis of the amino acid sequence of the *Drosophila* protein male-specific lethal-3 (MSL-3) which is one of the proteins required for hypertranscription of the single male X chromosome in *Drosophila* to achieve dosage compensation (28). A comparison of the MSL-3 sequence with the NR database using BLASTP showed no significant sequence similarity to any known proteins, except for an uncharacterized human open reading frame (ORF) product (28; legend to Fig. 1a). However, when the dBEST translated in six reading frames was searched using TBLASTN, a statistically significant similarity (probability of matching by chance, $P < 10^{-3}$) was detected with several human ESTs and an EST from rice. When these ESTs were in turn compared to the protein NR database using BLASTX, a highly significant similarity ($P < 10^{-9}$) was detected with the human retinoblastoma-binding protein 1 (RBP-1) and a moderate similarity was unexpectedly observed with the Zn finger protein A33 from newt ($P \sim 0.03$) and with the PC protein ($P \sim 0.08$). The detected region of PC belonged to the chromo domain, suggesting that despite the limited statistical significance, the similarity may be functionally relevant. From the BLASTX output for one of the human ESTs, a position-dependent weight matrix was constructed and the NR database was scanned iteratively using the MoST program, with the cut-off defined as the ratio of the expected number of similar segments retrieved from the database to the actual observed number of 0.01 (21). Further analysis included using BLASTP with this sequence set, in order to detect possible additional members of the chromo superfamily and using TBLASTN, in order to detect putative new chromo domains among ESTs, followed by another round of MoST search (22). The iterative database search resulted in a set of 25 chromo domain-containing proteins that included all the known chromo proteins, with the exception of SWI6, as well as several new ones. As a control for search sensitivity, additional analysis was performed using the BLITZ program. This search did not detect putative new chromo domains.

Inspection of the BLAST outputs for the murine chromatin modifier proteins MOD1 and MOD2 suggested that these proteins may contain a second copy of the chromo domain. A similar duplication of the chromo domain was suggested by the MoST search for the uncharacterized, putative yeast protein YEZ4. This prompted a further analysis of the chromo proteins using the GIBBS program under the assumption that each protein contains two copies of the chromo domain (23). The duplication of the chromo domain was identified in eight proteins, with each

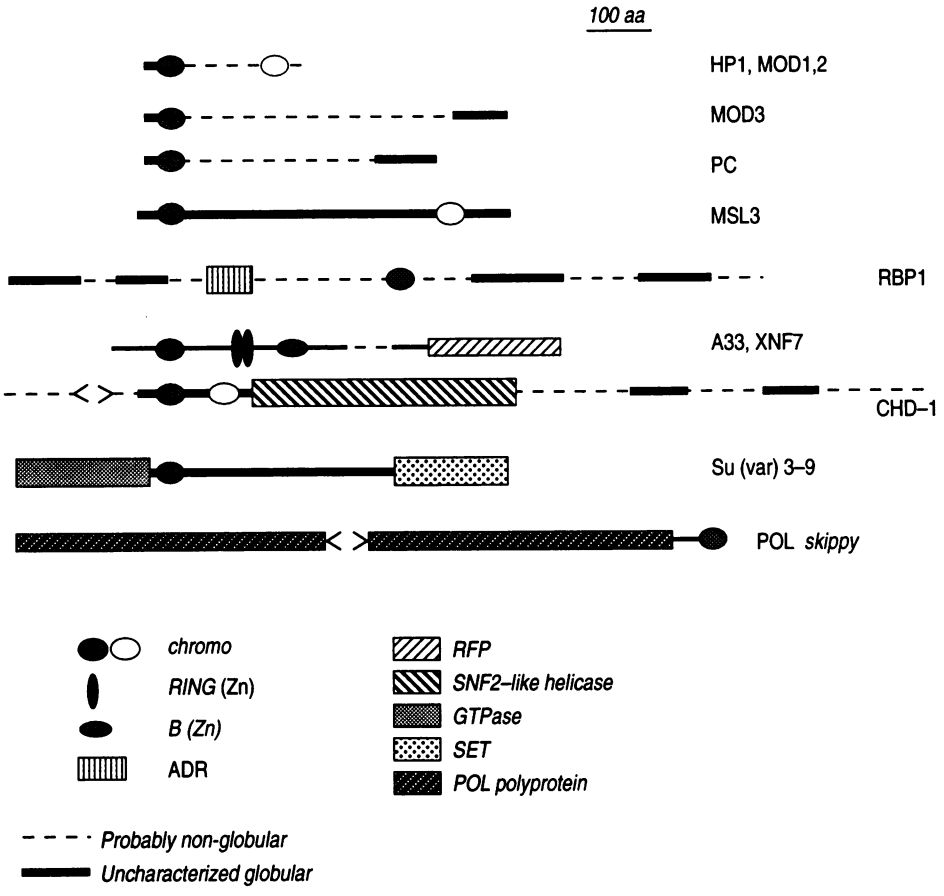

Figure 3. Scheme of the domain organization of selected chromo-containing proteins. RFP, SET and ADR are conserved domains with unknown function. RFP is conserved in A33, XNF7 and a group of other Zn finger proteins (27); SET is conserved in Su(var)3-9 and other proteins involved in transcription regulation (9); ADR is conserved in RBP1, yeast transcription factor ADR6 and human modulation recognition factors (E. V. K., unpublished observations).

of the segments scoring >6 SD above the random expectation. The statistical significance of the alignment of two copies of the chromo domains in proteins containing the duplication was also confirmed using MACAW. Specifically, the probability of the two copies of the chromo domain being detected by chance was $<10^{-5}$ for the MOD1 and MOD2 proteins and $\sim10^{-4}$ for the human HP1 protein. Further analysis using MACAW indicated that the SWI6 protein also may possess a duplicated chromo domain, with each of the copies containing a one amino acid insertion as compared to the other members of the superfamily. The presence of two chromo domain-related regions in the HP protein sequences has been noticed previously (29).

The resulting alignment of the chromo domains produced using the GIBBS program, with the SWI6 sequence added based on the results of the BLASTP search and additional analysis using MACAW, is shown in Figure 1a. There are no positions containing the same amino acid residue in all chromo domains. Nevertheless, the information content profile (sequence Logo) for the chromo domain alignment clearly reveals two conserved motifs centering at the tryptophan residues in positions 21 and 32, respectively (Fig. 1b). The first of these motifs, located in the middle of the chromo domain, is the most highly conserved part, with the lysine in position 20, tryptophan in position 21 and glycine in position 23 showing the highest information content. However, positions with a relatively high information content span the entire 37 residue range of the chromo domain (Fig. 1a and b).

Most of the chromo domain-containing proteins are enriched in compositionally biased regions that probably comprise non-globular domains (25). Specifically, HP1 and its mammalian homologs may not contain globular domains other than the two chromo domains (Fig. 2). This leads to an intriguing speculation that in order to form a stable globule, the two chromo domains may have to interact with one another, with the non-globular domains looping out. Alternatively, each chromo domain of HP1 may interact with the chromo domains of other HP1 molecules to form dimeric or multimeric complexes; the formation of HP1 homodimers *in vitro* has been recently detected (30).

The chromo domain is found in association with a striking variety of other domains (Fig. 3). These include Zn finger domains, a helicase domain, coiled coil domains, and several conserved domains whose functions have not yet been determined. Furthermore, the chromo domain is encoded by two fungal retrotransposons, either as a stand-alone ORF or as the C-terminal domain of the POL polyprotein (Fig. 3).

Functional implications: the chromo domain may be a general vehicle for the delivery of transcription regulators to their action sites on chromatin

The duplication of the chromo domain in HP1 is in accord with the recent results showing that HP1 contains two distinct chromatin-binding domains in its N-terminal and C-terminal halves (30). It has been shown that a chimeric protein containing the chromo domain from PC and the C-terminal chromo domain from HP1 binds both to the HP1-binding sites in heterochromatin and to the PC-binding sites in polytene chromosomes. Furthermore, in flies expressing the HP1/PC chimera, endogenous HP1 and PC are reciprocally misdirected to the binding sites of the other protein (30). These results imply that the chromo domain may be involved in a network of multiple protein–protein contacts, both with the targets in chromatin and with chromo domains in other molecules.

The identification of the chromo domain in RBP1 and in MSL3 is of particular interest. RBP1 is a nuclear phosphoprotein that has been identified based on its binding to the retinoblastoma (RB) protein, a wide-spread tumor suppressor (31–33). Retinobalstoma is a global transcription repressor that is thought to function, at least in some cases, by binding to and sequestering the transactivation domains of various transcription factors (34,35). The association of RB with chromatin and, specifically, its accumulation at the heterochromatin/euchromatin boundary have been demonstrated (36). In a RBP1-RB complex, the chromo domain of RBP1 may be involved in the delivery of the bound RB to the sites of its action on chromatin.

In contrast to the initially described chromo-containing proteins, MSL3 is involved in transcription activation rather than repression (37 and refs therein). Similarly, XNF7 has been shown to function as a transcription transactivator (38), and in CHD-1, the chromo domains are associated with an SNF2-like helicase domain implicated in transcription activation (9,39). Taken together, these observations indicate that the chromo domain may be a vehicle that delivers both positive and negative transcription regulators to the sites of their action on chromatin.

ACKNOWLEDGEMENT

We are grateful to Joel Eissenberg for a critical review of this manuscript and for communicating experimental results from his laboratory prior to their publication.

REFERENCES

1 Chothia, C. (1992) *Nature*, **357**, 543–542.
2 Green, P., Lipman, D. J., Hillier, L., Waterston, R., States, D. J. and J. M. Claverie (1993) *Science*, **259**, 1711–1716.
3 Green, P. (1994) *Curr. Opin. Struct. Biol.*, **4**, 404–412.
4 Shaffer, C. D., Wallrath, L. L. and Elgin, S. C. R. (1993) *Trends Genet.*, **9**, 35–37.
5 Paro, R. and Hogness, D. S. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 263–267.
6 Singh, P. B., Miller, J. R., Pearce, J., Kothary, R., Burton, R. D., Paro, R., James, T. C. and Gaunt, S. J. (1991) *Nucleic Acids Res.*, **19**, 789–794.
7 Saunders, W. S., Chue, C., Goebl, M., Craig, C., Clark, R. F., Powers, J. A., Eissenberg, J. A., Elgin, S. C. R., Rothfield, N. F. and Earnshaw, W. C. (1993) *J. Cell. Sci.*, **104**, 573–582.
8 Lorentz, A., Ostermann, K., Fleck, O. and Schmidt, H. (1994) *Gene*, **143**, 139–143.
9 Delmas, V., Stokes, D. G. and Perry, R. P. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 2414–2418.
10 Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. and Reuter, G. (1994) *EMBO J.*, **13**, 3822–3831.
11 Messmer, S., Franke, A. and Paro, R. (1992) *Genes Dev.*, **6**, 1241–1254.
12 James, T. C. and Elgin, S. C. R. (1986) *Mol. Cell Biol.*, **6**, 3862–3872.
13 Zink, B. and Paro, R. (1989) *Nature*, **337**, 468–471.
14 Eissenberg, J. C., Morris, G. D., Reuter, G. and Hartnett, T. (1992) *Genetics*, **131**, 345–352.
15 Powers, J. and Eissenberg, J. C. (1993) *J. Cell. Biol.*, **120**, 291–299.
16 Boguski, M. S., Lowe, T. M. and Tolstoshev, C. M. (1993) *Nature Genet.*, **4**, 332–333.
17 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) *J. Mol. Biol.*, **215**, 403- 410.
18 Altschul, S. F., Boguski, M. S., Gish, W. and Wootton, J. C. (1994) *Nature Genet.*, **6**, 119–129.
19 Wootton, J. C. and Federhen, S. (1993) *Comput. Chem.*, **17**, 149–163.
20 Rice, C. M., Fuchs, R., Higgins, D. G., Stoehr, P. J. and Cameron, G. N. (1993) *Nucleic Acids Res.*, **21**, 2967–2971.
21 Tatusov, R. L., Altschul, S. F. and Koonin, E. V. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 12 091–12 095.
22 Koonin, E. V. and Tatusov, R. L. (1994) *J. Mol. Biol.*, **245**, 125–132.
23 Lawrence, C. E., Altschul, S. F., Boguski, M. S., Neuwald, A. J. and Wootton, J. C. (1993) *Science*, **262**, 208–214.
24 Schuler, G. D., Altschul, S. F. and Lipman, D. J. (1991) *Prot. Struct. Funct. Genet.*, **9**, 180–190.
25 Wootton J. C. (1994) *Comput. Chem.*, **18**, 269–285.
26 Schneider, T. D. and Stephens, R. M. (1990) *Nucleic Acids Res.*, **18**, 6097–6100.
27 Rost, B. and Sander, C. (1993) *J. Mol. Biol.*, **232**, 584–99.
28 Gorman, M., Franke, A. and Baker, B. S. (1995) *Development*, **121**, 463–475.
29 Epstein, H., James, T. C. and Singh, P. B. (1992) *J. Cell. Sci.*, **101**, 463–474.
30 Suso Platero, J., Hartnett, T. and Eisenberg, J. C. (1995) *EMBO J.*, **14**, 3977–3986.
31 Kaelin, W. G. Jr., Krek, W., Sellers, W. R., DeCaprio, J. A., Ajchenbaum, F., Fuchs, C. S., Chittenden, T., Li, Y., Farnham, P. J., Blanar, M. A., Livingston, D. M. and Flemington, E. K. (1992) *Cell*, **70**, 351–364.
32 Otterson, G. A., Kratzke, R. A., Lin, A. Y., Johnston, P. G. and Kaye, F. J. (1993) *Oncogene*, **8**, 949–957.
33 Fattaey, A. R., Helin, C., Dembski, M. S., Dyson, N., Harlow, E., Vuocolo, G. A., Hanobik, M. G., Haskell, K. M., Oliff, A., Defeo-Jones, D. and Jones, R. E. (1993) *Oncogene*, **8**, 3149–3156.
34 Dynlacht, B. D. (1995) *Nature*, **374**, 114.
35 Weinberg, R. A. (1995) *Cell*, **81**, 323–330.
36 Szekely, L., Uzvolgyi, E., Jiang, W. Q., Durko, M., Wiman, K. G., Klein, G. and Sumegi, J. (1991) *Cell Growth Differ.*, **2**, 287–295.
37 Baker, B. S., Gorman, M. and Marin, I. (1994) *Annu. Rev. Genet.*, **28**, 491–521.
38 Li, X., Shou, W., Kloc, M., Reddy, B. A. and Etkin, L. D. (1994) *Exp. Cell. Res.*, **213**, 473–481.
39 Stokes, D. G. and Perry, R. P. (1995) *Mol. Cell. Biol.*, **15**, 2745–2753.