Volume 232, number 2, 263–268 FEB 05893 May 1988

*Hypothesis*

# A decapeptide motif for binding to the minor groove of DNA

## A proposal

William G. Turnell[+*], Sandra C. Satchwell[+] and Andrew A. Travers[+]

[+]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH* and **MRC Acute Phase Protein Research Group, Immunological Medicine Unit, Department of Medicine, Royal Postgraduate Medical School, London W12 OHS, England*

Received 23 March 1988

The pentraxin family of pentameric plasma proteins is widely distributed in both vertebrates and invertebrates [1]. In humans one of these proteins, serum amyloid P component (SAP) is the major $Ca^{2+}$-dependent DNA-binding protein of the serum and, in addition, binds highly specifically to nucleosome core particles [2]. This latter property is shared with histone H1 and it is therefore of interest to determine whether these proteins possess common structural features. In this paper we show that a decapeptide in SAP, and other pentraxins, is homologous to sequences in histones H1 and H4. On the basis of this observation and of the preferred DNA sequences at the dyad of the core particle we propose that this decapeptide forms a DNA binding motif which specifically recognises a relatively narrow minor groove of DNA. We then show that the geometry of such a groove would be consistent with recognition by an $\alpha$-helical conformation of the decapeptide.

To identify structural features common to the pentraxins and other proteins, regions of amino acid homology within the pentraxin family were first located and these regions then compared with other protein sequences using the template/sequence matching algorithm of Boswell [3] based upon the method of Bashford et al. [4]. Identical results were obtained with the computer programme ANALYSEPL written by Dr R. Staden. Of the regions conserved within the pentraxin family, two, located at residues 136–145 and 146–158, have been provisionally identified respectively as a nucleotide-binding site and a $Ca^{2+}$-binding loop (Turnell, in preparation). A third region spanning amino acids 123–132 is conserved in those pentraxins that bind to core particles but absent from those, such as *Limulus* CRP [2], which fail to interact detectably (table 1). This pattern of sequence conservation thus strongly suggests a role for this decapeptide in DNA and for histone recognition. Comparison of this sequence with the 4500 sequences in the PIR data banks revealed high homology with only two subsets of sequences: H4 histones and invertebrate gonadal H1 histones (table 1). Again this result strongly implicates the decapeptide as a DNA-binding determinant of the SAP molecule. From these data we derive a consensus sequence, P-$V^{RK^{G}}_{KR^{}_{S}}AL^{R^{N}}_{K^{}_{Q}}KG$, for the decapeptide.

Although nothing is yet known about the sites of interaction of SAP with DNA or nucleosomes those of histone H4 [5,6], and to a lesser extent,

*Correspondence address:* W.G. Turnell, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/88/$3.50 © 1988 Federation of European Biochemical Societies

Table 1

Decapeptide sequences in selected proteins

| Protein | Fraction with score ⩾ 90 | Provenance | Peptide | Score |
|---|---|---|---|---|
| (a) CRP | 2/2 | human | 123 P R V R K S L K K G 132 | 100 |
| | | rabbit | 123 P M V R K S L K K G 132 | 100 |
| SAP | 1/1 | human | 123 P L V K K G L R Q G 132 | 100 |
| Histone H1 | 1/14 | gonadal variant of sea urchin (*Parechinus angulosus*) | 84 P H V R R A L R N G 93 | 100 |
| Histone H4 | 16/16 | including yeast (*S. cerevisiae*), pea, rat | 32 P A I R R L A R R G 41 | 91 |
| Consensus from highest scores (⩾90) | | | P - V R/K K/R G/S/A L R/K N/K/Q G | |
| (b) Sperm H1 | | sea urchin (*Echinolampas crassa*) | 84 P H I R R A L K N Q 93 | 87 |
| | | annelid (*Platynereis dumerilli*) (a) | 53 P F I R K F I R K A 62 | 84 |
| | | (b) | 51 P F I R R F V R K A 60 | 84 |
| Histone H1 | | boar | 81 S R I K L G L K S L 90 | 47 |
| | | chicken | | |
| Histone H5 | | chicken | 66 L Q I K L S I R R L 75 | 59 |
| CRP | | *Limulus polyphemus* | 123 K G N A T G I A V/M G 132 | 25 |

Residue numbers against peptide termini position the peptides in their respective protein sequences. The PIR data bank was searched for homology with a weighted template derived from the pentraxin sequences. (a) The upper part of the table shows all sequences from the data bank with template match scores ⩾90, when an average score is ~60 (for methods see [3,4] and Turnell, in preparation). These sequences represent 0.4% of the entries in the data bank. The denominator of the fraction displayed in the second column is the number of sequences of a particular protein type contained in the PIR data bank. (b) The lower part of the table shows for comparison the decapeptide in the corresponding position in other histones of the H1 class. The annelid sequences are not contained in the PIR data bank

histone H5 [7,8], the analogue of histone H1 in avian erythrocytes, are well characterised. In the nucleosome core particle each copy of histone H4 binds to the inner face of the left-handed DNA super helix, making extensive contacts with the minor groove of DNA at positions located about three and a half double helical turns on each side of the dyad [5]. By contrast histone H1 and, presumably, SAP cannot bind to the inner face of the super helix in the intact nucleosome and must therefore interact with either the upper and lower surfaces of the nucleosome or with the outer face of the super helix. However, in general, the local conformation of the outer face of the super helix is very different from that of the inner face. This is because in a tightly bent DNA, as on the nucleosome, all the grooves (both major and minor) on the inside must be compressed while those on the outside become wider to accommodate the curvature [9]. This variation in groove width and conformation is reflected in the position-dependent sequence preferences observed in nucleosomal DNA [10]. We must therefore ask whether there are any sites on the outer face of the super helical DNA which possess a structure similar to that which normally occurs on the inside.

One line of evidence points strongly to the DNA at the nucleosome dyad possessing structural features in common with the inner face of the DNA supercoil. Although at the dyad the minor groove points outwards away from the histone octamer [5,11] Satchwell et al. [10] found one major preferred trinucleotide, AAA/TTT at this position. This trinucleotide is normally located at an inward facing minor groove at other sites on the

core particle. The structural basis for this latter selection is believed to be the ability of this and related sequences to adopt a conformation with a narrow minor groove, which would then accommodate the curvature of the DNA towards the histone octamer in the regions distant from the dyad. This argument therefore suggests that at the nucleosome dyad the outward facing minor groove may be relatively narrow.

A detailed analysis of the sequence preferences at the dyad (table 2) supports this conclusion. The table shows that the most highly preferred trinucleotide sequences are those which either contain all purines in one strand and all pyrimidines in its complement (such as AAA/TTT) or which contain a purine-pyrimidine (RY) base step in addition to a purine-purine step. In contrast, with the significant exception of the trinucleotides CAA/TTG most pyrimidine-purine (YR) steps are avoided. In crystal structures of DNA oligomers RY steps generally assume a low, often negative, roll angle which closes the minor groove, while YR steps are associated with a high positive roll angle opening the minor groove [12]. Sequences such as AAA/TTT and GGG/CCC are respectively straight [13] or have a low positive roll angle [14] in the particular sequence contexts so far studied. We note that, in at least 50% of occurrences, CAA/TTG, the apparent exception to the general rule, is part of the sequence CAAA/TTTG, which is closely related to sequences such as $CA_4G$, $CA_5G$, $CA_6G$. This last sequence is known to be associated with a narrow minor groove [13]. Thus the sequence preferences at the nucleosome dyad are wholly consistent with a minor groove of average or less than average width but are not compatible with a wider than average minor groove. This conclusion is also not inconsistent with the structure (at 7 Å resolution) of the DNA located at the dyad of crystalline core particles [5].

A further piece of evidence that the decapeptide motif may be involved in the recognition of a specific DNA structure located at the nucleosome dyad is the observation that histone H5 contacts the DNA super helix in this region [8]. In this context we note that the only lysine residue in the equivalent decapeptide in histone H5, Lys 69, is protected by association with chromatin from selective radiolabelling by reductive methylation [15]. The spatial location of the decapeptide in histone H4 has yet to be experimentally established but we note that the interaction of this histone with individual phosphate chains is determined by protein sequence elements distinct from the decapeptide [6]. For this histone only the minor groove interaction ~36 bp (i.e. ~3.5 double helical turns) on either side of the dyad remains unassigned.

Table 2

Sequence preferred at the dyad of the nucleosome core particle

| Sequence | Occurrence at dyad/mean occurrence (nucleotide 23–123) | |
|---|---|---|
| GGT/ACC | 1.60 | |
| CAA/TTG | 1.58 | > average occurrence YR 3/16 |
| $XA_3X/ZT_3Z$ | 1.40 | RY 7/16 |
| AGG/CCT | 1.37 | RRR 5½/8 |
| GGG/CCC | 1.33 | RRY 7/8 |
| $XA_4X/ZT_4Z$ | 1.32 | YRY 0/8 |
| $XG_3X/ZC_3Z$ | 1.31 | YRR 3/8 |
| GAT/ATC | 1.29 | |
| AAC/GTT | 1.25 | |
| GAC/GTC<br>$XA_2X/ZT_2Z$ | 1.21 | |
| TGA/TCA | 1.19 | |
| AAA/TTT<br>GAA/TTC | 1.15 | |
| AGT/ACT | 1.13 | |
| AAG/CTT | 1.12 | |
| AAT/ATT | 1.05 | |
| TAG/CTA | 1.04 | |
| GGC/GCC | 1.03 | |
| GGA/TCC | 1.00 | average occurrence |
| TAT/ATA | 0.96 | |
| TGT/ACA | 0.94 | < average occurrence YR 13/16 |
| CAG/CTG | 0.91 | RY 9/16 |
| CAT/ATG | 0.91 | RRR 2½/8 |
| TGC/GCA | 0.80 | RRY 1/8 |
| CAC/GTG | 0.74 | YRY 8/8 |
| TAC/GTA | 0.73 | YRR 5/8 |
| TGG/CCA | 0.71 | |
| AGA/TCT | 0.65 | |
| AGC/GCT | 0.52 | |
| GAG/CTC | 0.51 | |
| TAA/TTA | 0.50 | |
| CGN/NCG | 0.26 | (range 0.00 → 0.40) |
| $XA_5X/ZT_5Z$ | 0.00 | |

Data shown are taken from the collection of 177 DNA sequences described by Satchwell et al. [10]. Occurrence at dyad/mean occurrence is the ratio of occurrences at nucleotide 73 (the dyad) to the mean occurrence between nucleotides 23 and 123 inclusive. The designations X and Z in the nucleotide sequences indicate homopolymer runs of defined length so far, for example, for $XA_3X/ZT_3Z$: X can be C, G or T and Z can be A, C or G. N indicates any nucleotide

Thus far we have argued that the binding of histone H4 to the inner face of the DNA supercoil and of histone H1 and SAP to the outer face can arise in part from the common recognition of a relatively narrow minor groove. We must now ask whether the structure of the decapeptide itself is compatible with this type of interaction. To answer this question we need to know the most probable conformation assumed by the decapeptide. Analysis of the conserved sequence shows that nine of the ten residues could form a strongly amphipathic $\alpha$-helix, with the polar side of the helix positively charged (fig.1). Strong amphipathicity is unusual but not unknown in globular water-soluble proteins. Values for hydrophobicity [*H*] are $-0.791$ and $-0.247$ residue$^{-1}$ for the predicted helices in human CRP and SAP, respectively, and the corresponding hydrophobic moments [*M*] are 0.736 and 0.779 residues$^{-1}$. These values are midway between those typically found in water-soluble globular proteins and the helices common in lipoproteins, thought to be aqueous/lipid surface seeking structures [16]. Independent evidence that the sequences homologous to the decapeptide can assume an $\alpha$-helical conformation is derived from two-dimensional nuclear Overhauser enhancement spectra of the globular domain of histone H5 [17]. In this protein the equivalent region (residues 66–75) is an $\alpha$-helix of two and a half turns.

How might such an amphipathic $\alpha$-helix interact with a minor groove of DNA? The side chains of the polar residues are flexible and therefore the possible distances between the basic groups can vary over a substantial range. At their closest approach the separation between the basic groups is ~4 Å (not shown). However we assume that hydrophobic [18] and charge-charge repulsive forces would favour a more extended configuration of the side chains. Owing to their flexibility the side chains can each adopt many possible conformations but at their most extended the separation between the basic groups on successive amino acids is $\leqslant 12$ Å while that between basic groups on amino acids separated by approximately one turn of the $\alpha$-helix is $\leqslant 11$ Å (fig.2b). We know also that the width of the minor groove of DNA can vary between 9 and 17 Å while that of the major groove can vary between 13 and 21 Å [19]. Clearly the configuration of the basic or dipolar residues in the extended pattern could be compatible with electrostatic interactions with the phosphate backbones across a narrow minor groove but *not* across a major groove. In addition the shorter distance between the basic or dipolar residues, that is between amino acids separated by approximately one turn of the $\alpha$-helix, is compatible with the average separation of 7.5 Å between the oxygen atoms of adjacent phosphate residues on the sugar-phosphate backbone of one DNA strand (fig.2c). We note that although the binding of counter ions to the phosphate backbone is, in general, delocalised [20], the placement of a protein or small peptide may be more strongly constrained by steric effects.

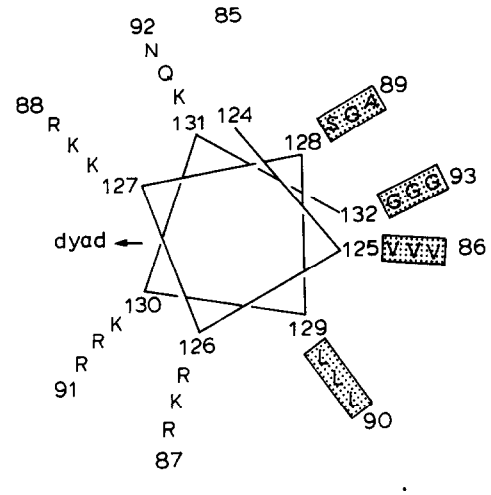


Fig.1. Projection down the axes of predicted $\alpha$-helices. Residues 124–132 in human CRP (inner) and SAP compared with 85–93 (outer) in gonadal histone H1 from the sea urchin *Parechinus angulosus*. The pentraxin sequences are shown in table 1. Residues conserved as hydrophobic or small and neutral in the five mammalian sequences as well as in all the histone H4 and invertebrate H1 sequences in the PIR data bank are boxed. The dyad axis is local and relates positively charged and/or dipolar residues (positions 126 and 130 with 131 and 127, respectively). It intersects the helix axis at right angles and intercalates the two turns of the $\alpha$-helix. We note that in chicken histones H1 and H5 a basic or dipolar residue would occupy position 85 and would potentially form an alternative or additional contact with a phosphate group bordering a minor groove.

Although the distances between the basic groups are compatible with an interaction with a minor groove of DNA it is also important to determine whether the geometry of these groups is similarly compatible. Because DNA is a double helix the four phosphate oxygens separated by a minimum distance across the minor groove are not coplanar.

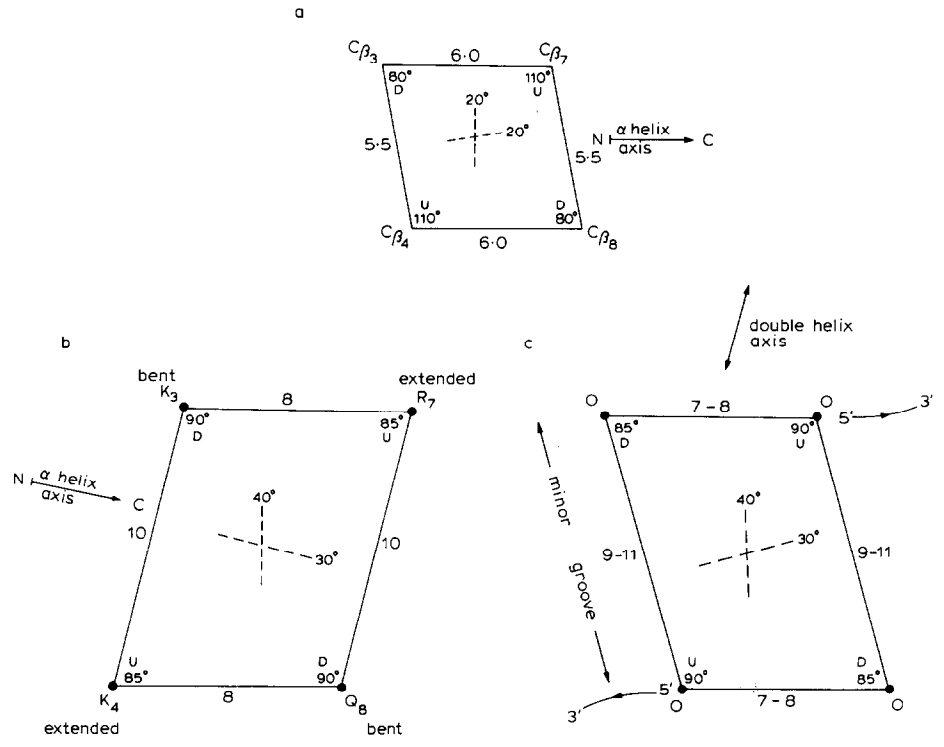


Fig.2. (a) Distances and geometry between the $C_\beta$ atoms of residues 3,4,7,8 in the $\alpha$-helical conformation of a decapeptide. (b) Distances and geometry between basic groups (the $\epsilon$-N of lysine, the imino-N of the $\delta$-guanidinium group of arginine and the $\epsilon$-2-N of glutamine) in the $\alpha$-helical conformation of the decapeptide (numbered 0–9) from human serum amyloid P component: with partially extended configuration of the side chains at 3,4,7,8. (c) Distances and geometry between neighbouring phosphate oxygen atoms across a moderately narrow DNA minor groove. All diagrams are presented looking perpendicular to a helical axis. D and U indicate vertices directed respectively away (Down) or towards (Up) the observer. Virtual bond-lengths are in ångstroms, angles in degrees. b and c represent complementary saddle-shaped surfaces which can be fitted together by folding one over the other about a line drawn vertically on the page between the two. The fit can also be achieved after *either* of the surfaces shown in b and c have been rotated by 180° about an axis passing through the centre of the surface perpendicular to the page. Axes of twist are represented by dashed lines with associated angles in degrees. Angles at vertices of these parallelograms do not sum to 180° because of the non-planarity of the surfaces. The skewness of the parallelograms has been exaggerated for clarity.

Instead the line connecting two oxygen atoms of successive phosphate groups on one strand is twisted by ~30–40° in a left-handed sense relative to the line connecting the equivalent two oxygen atoms on the opposing strand. This then imposes a similar relative twist on the lines connecting oxygen atoms on opposite strands (fig.2c). Optimal interaction between the DNA and the decapeptide thus requires the complementary spatial relationship between the basic or dipolar groups of the peptide.

A related consideration is that the handedness of the B-DNA double helix and of the $\alpha$-helix is the same. This means the constrained $C_\beta$ atoms of the four basic or dipolar amino acids are in the opposite configuration to that required for interaction with the DNA minor groove (fig.2a). To compensate for this constraint we have to assume that two of the side chains, $K_3$ and $Q_8$ are bent slightly towards each other, while the remaining two, $K_4$ and $R_7$ are fully extended. This arrangement, which can be built without violation of Van der Waals contacts, generates a spatial orientation between the basic groups that is precisely complementary to that of the phosphate oxygen atoms on DNA, thus allowing the superimposition of the basic groups on the negatively charged oxygens (fig.2b and c). We note that in this model the decapeptide motif interacts with both DNA strands bordering a minor groove. However the

most conserved residues, 3 and 7 (table 1), bind to only one of the two DNA strands. This means that, for example, in chicken histone H5 in which residue 4 is leucine the binding to the opposite DNA strand would be substantially less. Consequently in this situation the steric constraints for recognition of a narrow minor groove of DNA could, in principle, be relaxed. We note, in this context, that those histones of the H1 class which contain the best fit to the consensus decapeptide are those that are, in general, associated with highly condensed chromatin e.g. in the sperm of invertebrates.

So far we have only considered the interaction of the four basic or dipolar amino acids with the sugar-phosphate backbone of DNA. It is clearly important that these basic residues be sufficiently exposed at the protein surface to permit this interaction. However, when embedded in a continuous $\alpha$-helix the side chains protruding from the same face of the flanking $\alpha$-helical turns could interfere substantially with DNA binding of the type we have postulated. We therefore suggest that the delimitation of the conserved $\alpha$-helical region to two-and-one-half turns permits full exposure of the basic groups.

In summary, we have proposed that the decapeptide motif found in SAP and histones H1 and H4 binds to a minor groove of average to less than average width. This structure in unconstrained DNA could be associated with runs of oligo$(dA)_n \cdot (dT)_n$ where $n \geqslant 4$ [13,19]. The minor groove in such regions is sufficiently narrow to exclude the minor groove interactions with the enzyme DNase I [21]. We would therefore predict, first, that the decapeptide motif in histones H4 and H1 (H5) be directly involved in minor groove contacts and second that SAP and the globular domain of histone H1 (H5) containing the decapeptide motif should bind preferentially to DNA sequences associated with a narrow minor groove width (e.g. [10]). We note finally that a protein with precisely this sequence selectivity has been purified from African green monkey cells [22]. It will be interesting to ascertain whether this protein also contains the decapeptide motif.

*Acknowledgements:* We thank Dr J.T. Finch for suggesting this line of enquiry, Drs R. Boswell and R. Staden for computer programs, Drs P.J.G. Butler, R. Henderson, A. Klug and J.O. Thomas for constructive suggestions and Professor M.B. Pepys for his interest and support.

## REFERENCES

[1] Pepys, M.B. and Baltz, M.L. (1983) Adv. Immunol. 34, 141–212.
[2] Pepys, M.B. and Butler, P.J.G. (1987) Biochem. Biophys. Res. Commun. 148, 308–313.
[3] Boswell, J.R. (1988) Nucleic Acids Res., in press.
[4] Bashford, D., Chothia, C. and Lesk, A.M. (1987) J. Mol. Biol. 196, 199–216.
[5] Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984) Nature 311, 532–537.
[6] Ebralidse, K.K. and Mirzabekov, A.D. (1986) FEBS Lett. 194, 69–72.
[7] Allan, J., Hartman, P.G., Crane-Robinson, C. and Aviles, F.X. (1980) Nature 288, 675–679.
[8] Drew, H.R. and McCall, M.J. (1987) J. Mol. Biol. 197, 485–511.
[9] Drew, H.R. and Travers, A.A. (1985) J. Mol. Biol. 186, 773–780.
[10] Satchwell, S.C., Drew, H.R. and Travers, A.A. (1986) J. Mol. Biol. 191, 659–675.
[11] Lutter, L. (1978) J. Mol. Biol. 124, 391–420.
[12] Calladine, C.R. and Drew, H.R. (1984) J. Mol. Biol. 178, 773–782.
[13] Nelson, H.C.M., Finch, J.T., Luisi, B.A. and Klug, A. (1987) Nature 330, 221–226.
[14] McCall, M., Brown, T. and Kennard, O. (1985) J. Mol. Biol. 183, 385–396.
[15] Thomas, J.O. and Wilson, C.M. (1986) EMBO J. 5, 3531–3537.
[16] Eisenberg, D. (1984) Annu. Rev. Biochem. 53, 595–623.
[17] Clore, G.M., Gronenborn, A.M., Nilges, M., Sukumaran, D.K. and Zarbock, J. (1987) EMBO J. 6, 1833–1842.
[18] Eisenberg, D. and McLachlan, A.D. (1986) Nature 319, 199–203.
[19] Drew, H.R. and Travers, A.A. (1984) Cell 37, 491–502.
[20] Manning, G.S. (1978) Q. Rev. Biophys. 11, 179–246.
[21] Suck, D. and Oefner, C. (1986) Nature 321, 620–625.
[22] Solomon, M.J., Strauss, F. and Varshavsky, A. (1986) Proc. Natl. Acad. Sci. USA 83, 1276–1286.