Review

Single-stranded DNA binding protein encoded by the filamentous bacteriophage M13: structural and functional characteristics *

Alphons P.M. Stassen, Rutger H.A. Folmer, Cees W. Hilbers & Ruud N.H. Konings

Department of Molecular Biology and Department of Biophysical Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Received 29 December 1994; accepted 30 January 1995

Key words: Filamentous bacteriophage M13, single-stranded DNA binding protein, regulation DNA replication and gene expression, 3D structure, superhelical nucleoprotein complex, recurring fold

Abstract

The single-stranded DNA binding protein, or gene V protein (gVp), encoded by gene V of the filamentous bacteriophage M13 is a multifunctional protein that not only regulates viral DNA replication but also gene expression at the level of mRNA translation. It furthermore is implicated as a scaffolding and/or chaperone protein during the phage assembly process at the host cell membrane. The protein is 87 amino acids long and its biological functional entity is a homodimer. In this manuscript a short description of the life cycle of filamentous phages is presented and our current knowledge of the major functional and structural properties and characteristics of gene V protein are reviewed. In addition models of the superhelical complexes gVp forms with ssDNA are described and their (possible) biological meaning in the infection process are discussed. Finally it is described that the ‘DNA binding loop’ of gVp is a recurring motif in many ssDNA binding proteins and that the fold of gVp is shared by a large family of evolutionarily conserved gene regulatory proteins.

Abbreviations: dsDNA = double-stranded DNA; ssDNA = single-stranded DNA; dsDBP = double-stranded DNA binding protein; ssDBP = single-stranded DNA binding protein; CSD = cold-shock domain; NMR = nuclear magnetic resonance; g’X’p = gene ‘X’ protein.

Introduction

Protein-nucleic acid interactions

Interactions between proteins and nucleic acids are of vital importance for all forms of life. They play an essential role in DNA replication, DNA recombination, gene expression and mRNA translation, all of which are fundamental molecular processes underlying cell homeostasis, proliferation and differentiation. Of the nucleic acid binding proteins those that interact with DNA are characterised best. They can be divided into proteins that bind to single- or double-stranded DNA.

Single-stranded DNA binding proteins (ssDBPs) are generally multifunctional. They usually impart a regular structure to single-stranded DNA (ssDNA), which subsequently is exploited in, for instance, the assembly of higher order nucleoprotein complexes, the regulation of gene expression or processes as DNA replication, recombination and repair. A protein that is a member of the latter class is encoded by gene V (gVp) of the filamentous bacteriophage M13. In this review a survey is given of our current knowledge of the structural and functional properties of this extensively characterised ssDBP.

Protein-dsDNA interactions

Up to now the most detailed information on DNA-protein interactions is available for proteins that bind to
double-stranded DNA (dsDNA) in a sequence-specific manner. These include pro- and eukaryotic transcription factors as well as proteins involved in site-specific recombination [for reviews, see 1-4]. Structural data, obtained both from X-ray and NMR analyses, have led to the identification of several ubiquitously recurring motifs. Examples of motifs that play a major role in dsDNA binding of numerous regulatory proteins are the helix-turn-helix and the zinc-finger motifs.

**Helix-turn-helix motif.** This motif was the first to be studied in detail. It consists of two short α-helical segments connected by a tight β-turn resulting in an inter-helical angle of about 120°. When complexed with DNA, one of the two α-helices, termed the 'recognition helix' because it usually contains most of the amino acids that interact with the DNA, is positioned in the major groove. It thus provides sequence-specific recognition of its target site.

**Zinc-finger motif.** Zinc fingers constitute a second class of protein motifs for sequence specific recognition, again involving an α-helix, and comprise at least three different subclasses which share the property of stabilisation of the motif by a tetrahedrally co-ordinated zinc ion. The first subclass, denoted C_{2-H2}, consists of a β-hairpin and an α-helix held together by a zinc ion liganded by two cysteine and two histidine residues. The second subclass, designated C_{2-C2}, as found for instance in steroid and thyroid hormone receptors, contains two loop-helix elements and two zinc ions, but lacks a β-sheet structure. In each loop-helix element a zinc ion is liganded by two cysteine residues that are situated at the beginning of the loop, and by two cysteine residues located near the N-terminal end of the α-helix. The only protein which up to now has shown to contain the third type of zinc finger motif is the yeast transcription activator GAL4. 3D-structural analyses have shown that the monomeric protein contains a binuclear zinc cluster consisting of two zinc ions and six cysteine residues. Because each zinc ion is co-ordinated by four cysteine residues, two of the six cysteines are shared. Most regulatory proteins, including many transcription activators and the ssDBP that is major subject of this review, bind to DNA as dimers. As is the case for the DNA-binding motifs, the protein structural motifs that mediate dimerisation tend also to fall into one of several common and recognisable patterns. The two structural motifs that have been well characterised are the leucine zipper and the helix-loop-helix motifs, which both are connected to a basic and α-helical DNA binding domain.

**Leucine zipper motif.** This structural element consists of an amphipathic α-helix with hydrophobic amino acid residues on one side of its surface. A striking feature of this α-helix is that it contains four or five leucine residues periodically arranged every seventh residue. Because of this repeat, they lie in almost a straight line on the hydrophobic side of the α-helix. Upon dimerisation, the residues of the hydrophobic surfaces 'interdigitate', hence the name 'leucine zipper', thereby forming a parallel coiled-coil of the two α-helices. In regulatory proteins with leucine zippers, the DNA binding domain is often found in an extension of the α-helix of the leucine zipper, which contains a high concentration of the basic amino acid residues.

**Helix-loop-helix motif.** Many proteins which have been implicated in the development of multicellular organisms, share a conserved region of about 50 amino acids that include the determinants for both DNA binding and protein dimerisation. The latter sequence can form two short amphipathic α-helices connected by a 'loop' of variable length. The helix-loop-helix of one protein interacts with its counterpart in another to form (hetero)dimers. DNA binding is again mediated by a short stretch of amino acids, rich in basic residues, that is immediately adjacent to the helix-loop-helix motif. This enumeration illustrates that dsDNA-binding can be accomplished by a wide variety of structural motifs. Despite the many differences, two common features can nevertheless be distinguished. Firstly, in all the motifs positively charged amino acid residues interact with the negatively charged backbone of the DNA, thus enabling both specific and non-specific DNA binding. Secondly, sequence-specific binding is usually mediated by hydrogen bonds as well as by (hydrophobic) van der Waals interactions between side chains of the polypeptide backbone and functional groups on the exposed edges of the bases, primarily in the major groove of B-DNA.

**Protein-ssDNA interactions**

Proteins that bind to and affect the structure of ssDNA have been far less well characterised than the dsDNA binding proteins. Most of what is known is related to their non-specific, but generally co-operative ssDNA binding. They are ubiquitously present in both pro- and eukaryotic organisms [for a review, see 5] and