Protein Science (1998), 7:1781–1788. Cambridge University Press. Printed in the USA.

The role of the 6 lysines and the terminal amine of *Escherichia coli* single-strand binding protein in its binding of single-stranded DNA

JIWEN CHEN, DAVID L. SMITH, AND MARK A. GRIEP
Department of Chemistry, University of Nebraska–Lincoln, Lincoln, Nebraska 68588-0304

(RECEIVED October 15, 1997; ACCEPTED May 14, 1998)

Abstract

Differential chemical modification of the lysines and amino-terminus of *Escherichia coli* single-strand binding (SSB) protein was used to determine their roles in the binding of SSB to single-stranded DNA (ssDNA). A combination of isotope labeling and mass spectrometry was used to determine the rates at which SSB was acetylated by acetic anhydride. First, SSB was labeled by deuterated acetic anhydride for given lengths of time in the presence or absence of single-stranded ssDNA. Then, the protein was denatured and completely acetylated by nondeuterated acetic anhydride. Enzymatic digests of the completely acetylated, isotopically labeled SSB were analyzed by electrospray ionization mass spectrometry. The intensities of the deuterated and nondeuterated forms of acetylated peptides provided accurate quantification of the reactivity of the amines in native SSB, either free or bound to ssDNA. Acetylation rate constants were determined from time course measurements. In the absence of ssDNA, the terminal α-amine of SSB was 10-fold more reactive than Lys residues at positions 43, 62, 73, and 87. The reactivities of Lys 7 and 49 were much lower yet, suggesting that they have very limited access to solution under any condition. In the presence of ssDNA, the reactivities of the amino-terminus and Lys residues 43, 62, 73, and 87 were reduced by factors of 3.7–25, indicating that the environments around all of these amines is substantially altered by binding of SSB to ssDNA. Three of these residues are located near putative ssDNA binding sites, whereas Lys 87 is located at the monomer–monomer interface.

Keywords: acetylation; *Escherichia coli* SSB; mapping; mass spectrometry

Escherichia coli single-stranded DNA-binding protein (SSB) is a 75-kDa homotetramer that binds to single-stranded DNA (ssDNA) and is essential for DNA replication, recombination, and repair (Meyer & Laine, 1990). How SSB participates in these different cellular processes is not clear, but the mechanism by which it binds ssDNA is fairly well understood (Lohman & Ferrari, 1994). In every cell, there are about 500 copies of SSB tetramer, and we can predict that about 50 SSB tetramers are needed for every replication fork, of which two are created every time a cell divides (McMacken et al., 1987). Thus, 20% of the cellular SSB is used every time DNA replication is initiated. In cells growing exponentially, multiple cycles of DNA replication can be initiated such that eight forks are working simultaneously. Under such conditions, most of the SSB would be bound at the replication forks. To carry out such a role, SSB must bind to ssDNA rapidly and nonspecifically. In fact, one of SSB's binding modes, the 35 mode, is predicted to have such properties. In the 35 mode, each SSB tetramer binds 35 nucleotides of ssDNA in a highly cooperative manner such that two of the four subunits are bound to the ssDNA. The other mode is the 65 mode, in which all four subunits bind a total of 65 nucleotides of ssDNA. The 65 mode may participate in DNA recombination and repair.

As a single-stranded DNA binding protein, SSB binds to the nitrogenous bases of the ssDNA and to the phosphates in the ssDNA backbone. Thermodynamic measurements have revealed large hydrophobic and electrostatic components to binding. Most current structure/function information studies have focused on four aromatic amino acids, out of the 177 total amino acids, as playing a dominant role in the hydrophobic binding (Lohman & Ferrari, 1994). At least some Lys residues in SSB likely contribute to the electrostatic component of SSB binding to ssDNA (Bandyopadhyay & Wu, 1978).

The principal goal of the present investigation was to identify and characterize Lys residues in SSB that participate in the binding of SSB to ssDNA. We have used electrospray ionization mass spectrometry (ESIMS) and deuterated acetic anhydride to determine the rates of acetylation of the amino-terminus and specific Lys residues in SSB. Performing these measurements on SSB in

Reprint requests to: Mark A. Griep, Department of Chemistry, University of Nebraska–Lincoln, Lincoln, Nebraska 68588-0304; e-mail: mgriep@unlinfo.unl.edu.

Abbreviations: ESIMS, electrospray ionization mass spectrometry; Hsmt SSB, human mitochondrial SSB; RP-HPLC, reversed-phase high-performance liquid chromatography; SSB, single-strand binding protein; ssDNA, single-stranded DNA.

the absence and presence of ssDNA revealed that the rates of acetylation of the amino-terminus and four Lys residues are highly dependent on binding, suggesting that these amines participate in the binding process. Although the general approach of differential chemical modification with acetic anhydride has been used extensively (Young & Kaplan, 1989), methods used in the present study demonstrate how stable isotopes can be used to improve the accuracy of the measurements.

Results

Sequence verification of SSB and its storage degradation product

The SSB monomer molecular mass was determined by ESIMS to be 18,845 (Fig. 1). This agreed with the calculated molecular mass of 18,844 consistent with the most recent DNA sequence determination (GenBank Accession No. P02339), but differs from that expected for the original DNA sequence-deduced value of 18,873 (Sancar et al., 1981). The molecular mass determined by ESIMS is consistent with Gly at position 133. The molecular masses of the tryptic and Endoproteinase Glu-C peptides (Fig. 2) were also in complete agreement with masses predicted from the more recent DNA sequence. In addition, the molecular mass of the degradation product of SSB, which appears after storage at 4 °C for several days (Overman et al., 1988), exhibited a loss of 262 to give a molecular mass of 18,582 (Fig. 1). This decrease in molecular mass is consistent with the C-terminal loss of Pro-Phe (calculated loss of 262), but not with the amino-terminal loss of Ala-Ser (calculated loss of 176).

SSB acetylation at pH 8.1 is sensitive to ssDNA

Because most SSB thermodynamic studies have been performed at pH 8 and 25 °C, the initial acetylation studies were carried out under those conditions. To compare the reactivity of individual amines when SSB was free or bound to ssDNA, the two forms of SSB were exposed to acetic d_6-anhydride at pH 8.1. The partially d_3-acetylated SSB was completely acetylated by h_6-acetic anhydride and digested by Endoproteinase Glu-C or trypsin. Peptides used to quantify the extent of acetylation at the amino-terminus and each of the Lys residues in SSB are indicated in Figure 2. Because all the amines were acetylated prior to the digestion, trypsin cleavage occurred on the carboxyl side of arginine, but not lysine. Natural ssDNA was used for the protection studies because preliminary experiments with poly(dT) gave very low yield after proteolytic digestion.

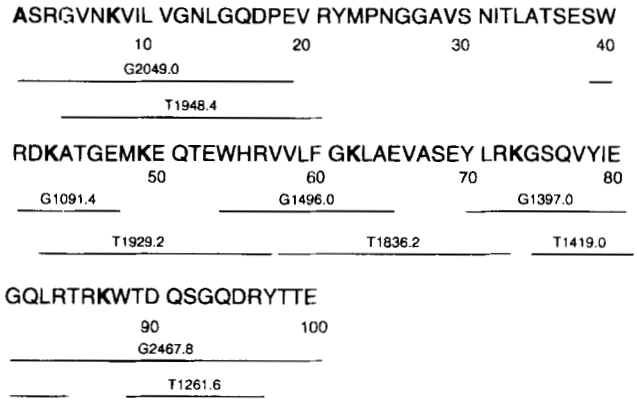

Fig. 2. Partial sequence map of SSB showing lysines and the amino-terminal residue in bold letters. The observed acetylated lysine-containing peptides are underlined below the sequence. Letter G before the monoisotopic mass indicates peptides from a Glu-C digest, and letter T indicates those from a tryptic digest.

The h_3 and d_3 acetylation of a Lys residue increases its molecular mass by 42 and 45, respectively. The extent of acetylation was determined from the relative intensities of the nondeuterated and deuterated mass envelopes, as illustrated for the segment including residues 39–47 in Figure 3. Because deuteration of the acetyl group increases the molecular weight by only 3 Da, isotope envelopes representing the h_3 and d_3 forms of the peptides overlapped. For example, the peak at m/z 1,091.4 contributed 5% of its intensity to the peak at m/z 1,094.4. Adjustments for overlapping of

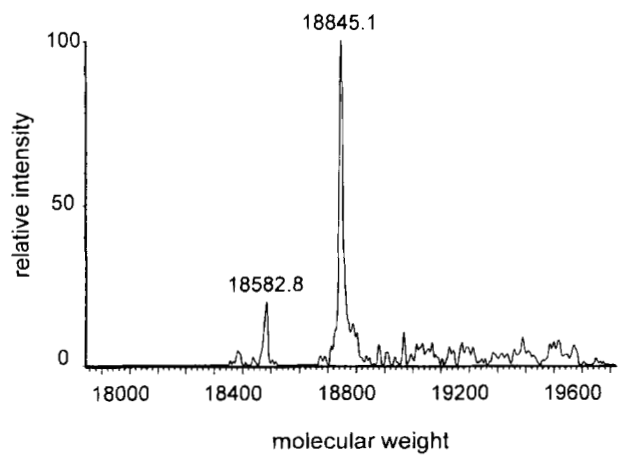

Fig. 1. Reconstructed ESIMS spectrum of SSB. The peak at 18,583 corresponds to the amino-terminal degradation product that appears after prolonged storage at 4 °C.

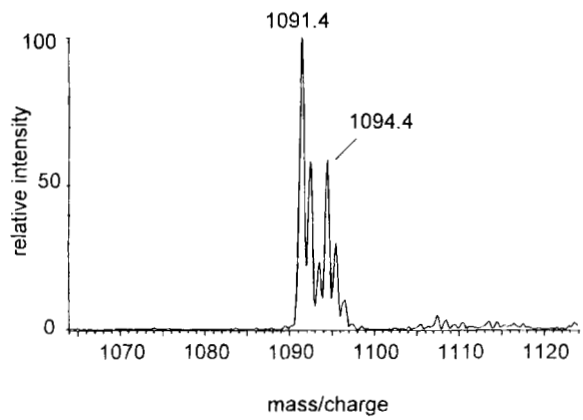

Fig. 3. ESIMS spectrum of acetylated Endoproteinase Glu-C segment 39–47. The isotope peak at mass 1,091 corresponds to the nondeuterated peptide (mass increment of 42 of the unmodified peptide), the isotope peak at 1,094 corresponds to the deuterated peptide (mass increment of 45 of the unmodified peptide).

isotope envelopes representing h_3 and d_3 acetylation were made to accurately quantify the extent of acetylation in native SSB.

Two proteases were required to give all of the peptides required to determine the reactivity of each of the six Lys residues. Because the mass spectra of tryptic fragment 4–21 showed that Lys 7 was effectively unreactive, the reactivity of the amino-terminus was determined from the Endoproteinase Glu-C fragment 1–19. The reactivity of Lys 49 could not be determined directly from the Endoproteinase Glu-C fragments, probably because the 48–50 fragment was not retained on the HPLC column. However, the reactivity of Lys 49 was determined indirectly from the tryptic fragment 42–57, which includes both Lys 43 and 49. The reactivity of Lys 43 was determined directly from the Endoproteinase Glu-C fragment 39–47. When Lys 73 acetylation was quantified from both tryptic segment 73–84 and Endoproteinase Glu-C segment 70–80 (data not shown), there was less than 5% difference between the measurements, indicating excellent reliability of the entire analytical procedure.

In general, the level of d_3 acetylation could be accurately quantified only when it was greater than 5%. The acetic anhydride concentration at which this condition is met for most of the lysines was first determined by assessing the extent of labeling of each lysine for a range of acetic anhydride concentrations. A wide range of amine reactivities was found for SSB in the absence or presence of ssDNA. Results obtained using 1.2 mM acetic anhydride are presented in Figure 4. These results show that, at pH 8.1 and 25 °C, the amino-terminal α-amine displayed the highest level of acetylation, whereas Lys 7 had the lowest level of acetylation. Although acetylation of Lys 49 was near the limit of detection, acetylation of the remaining four Lys residues, 43, 62, 73, and 87, could be quantified accurately. Addition of ssDNA substantially inhibited acetylation of the amino-terminus of SSB, as well as acetylation of the four most reactive Lys residues.

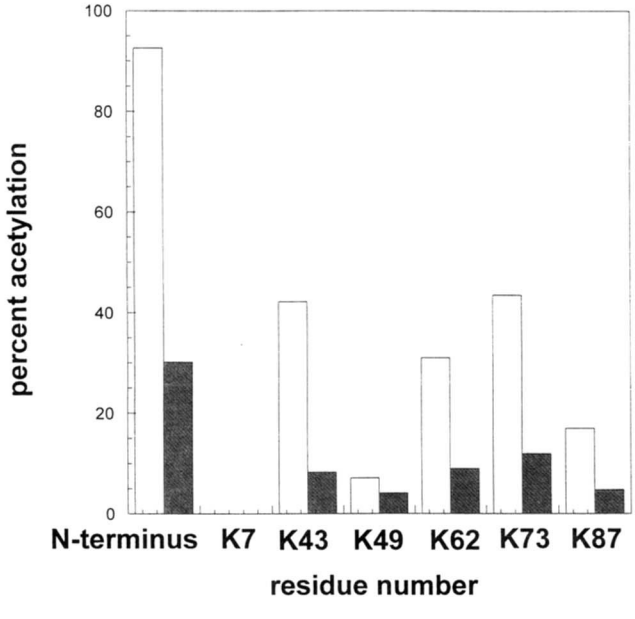

Fig. 4. Percent acetylation of SSB, with or without ssDNA, at pH 8.1, 25 °C. SSB (8 μM) was reacted with 1.2 mM deuterated acetic anhydride. The open columns represent free SSB and the shaded columns represent bound SSB.

ssDNA does not compete for the label

Because the four most reactive Lys residues were labeled less in the presence of ssDNA, a control experiment was performed to establish whether the added ssDNA caused reduced SSB acetylation by reacting with the acetic anhydride labeling agent. Substance P is a undecapeptide with the sequence Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met-NH_2 that does not bind to ssDNA but can be labeled with acetic anhydride. This peptide was acetylated under the same conditions used to acetylate SSB. In duplicate experiments, 72% of substance P was acetylated, in the absence or presence of ssDNA. This result confirmed that the decrease in SSB reactivity with d_6-acetic anhydride could only be attributed to the binding of SSB to ssDNA and not to competing ssDNA acetylation.

SSB acetylation rate constants at pH 7.0 are sensitive to ssDNA

Initial experiments performed at pH 8.1 demonstrated that the amino-terminus and the six Lys residues of SSB are characterized by a wide range of acetylation rates, and that these rates decrease in the presence of ssDNA. However, the rates of these reactions could not be determined accurately under these conditions because the reaction at pH 8.1 was complete within minutes. Because only the unprotonated form of the amines reacts with acetic anhydride, the acetylation experiments were performed at pH 7.0 where the reaction rates could be measured more accurately.

The extent of acetylation of each lysine residue was determined as a function of time following reaction of free SSB with 1.6 mM d_6-acetic anhydride (Fig. 5A). By combining the extent of acetylation of individual residues, it was determined that 0.2, 0.4 , 0.9, 1.1, and 1.3 lysines were acetylated per SSB monomer after 10, 45, 130, 270, and 600 s. Acetylation rate constants for each reactive SSB residue, calculated by fitting the data in Figure 5A to Equation 3, are presented in Table 1. Results of experiments performed at pH 7.0 are similar to those obtained at pH 8.1. The four most reactive Lys residues have acetylation rates constants spanning the range of 0.59 to 1.86 M^{-1} s^{-1}, and the amino-terminus is acetylated much faster than any of the Lys residues (23 M^{-1} s^{-1}). Acetylation of Lys 7 and 49 could not be quantified under these conditions. Thus, decreasing the pH from 8.1 to 7.0 slowed the acetylation rates substantially, but did not change the order of reactivities of amines in SSB. These reaction conditions were optimized to favor investigation of the most reactive Lys residues in SSB because they are more likely to participate in binding of SSB to ssDNA.

When SSB was bound to ssDNA, the reactivity of its Lys residues and amino-terminus decreased considerably. To facilitate measurements over the same time scale as in the absence of ssDNA (Fig. 5A), these reactions were carried out with a 5.5-fold higher concentration of acetic anhydride (Fig. 5B). By combining the extent of acetylation of individual lysines, it was determined that 0.4, 0.7, 1.2, and 1.4 lysines per SSB monomer were acetylated after 45, 130, 270, and 600 s. The reaction conditions used in this study were chosen to facilitate accurate measurement of acetylation rates while minimizing the extent of acetylation. For the longest reaction time used (600 s), 18% of the SSB monomers will have no Lys residues labeled, 38% will have one Lys labeled, and 44% will have two or more Lys labeled, if the entire population of SSB molecules were labeled randomly. The experimental data in Figures 5A and 5B fit Equation 3 very well, indicating that cooperative labeling was not significant. Rate constants for acetylation

of the amino-terminus and the four most reactive Lys residues of SSB when bound to ssDNA are presented in Table 1. These results show that amino-terminal α-amine was a factor of 25 less reactive when SSB was bound to ssDNA, while the reactivities of Lys 43, Lys 62, Lys 73, and Lys 87 were decreased four- to sevenfold.

Discussion

Advances in experimental methods

The reactivities of amino acid side chains in polypeptides has been used extensively to assess the roles played by these side chains in ligand binding. In the present study, the acetylation rates of specific Lys residues have been used to investigate the importance of these residues to the binding of *E. coli* SSB to ssDNA. Side-chain reactivities have most often been investigated using radioactive reagents to label the side chains. The extent to which specific residues were labeled was determined following proteolytic fragmentation and chromatographic isolation of the appropriate peptides. Rate constants describing the chemical reactivities of specific side chains could be determined from the radioactivity and UV absorbance of the isolated peptides. Although this approach has been successful, its application has been limited by the requirement to isolate peptides to a high level of purity before their specific activities could be determined accurately. Furthermore, increasing concerns over the availability and disposal of radioactive materials has also diminished interest in this approach.

Recent advances in mass spectrometry have made it possible to assess the reactivities of amino acid side chains without the use of radioactive materials (Qin et al., 1992; Akashi et al., 1993; Hasan et al., 1993; Glocker et al., 1994). Due to the high specificity of mass spectrometry, modified peptides can be detected and identified even when isolation is incomplete. However, quantitative assessment of the reactivities by mass spectrometry has been difficult because the mass spectrometric response to unmodified and modified peptides may differ substantially. This problem may be par-

Table 1. *Acetylation rate constants for each amine in free and bound SSB, and their ratio*[a]

Amine	$k_{acetylation}$ (M^{-1} s^{-1})		
	Free SSB	Bound SSB	Free/Bound
Amino-terminus	23.0 ± 1.6	0.91 ± 0.09	25
Lys 7	ND[b]	ND[b]	—
Lys 43	1.75 ± 0.09	0.26 ± 0.01	6.7
Lys 49	<0.25[c]	<0.10[c]	—
Lys 62	1.20 ± 0.08	0.27 ± 0.01	4.4
Lys 73	1.86 ± 0.13	0.30 ± 0.01	6.2
Lys 87	0.59 ± 0.04	0.16 ± 0.01	3.7

[a]The acetylation rate constants were determined by fitting data presented in Figures 5A and 5B to Equation 3 in which $k_{hydrolysis}$ was fixed at 0.0045 s^{-1}.

[b]ND means not detectable.

[c]The acetylation rate constant was estimated from an acetylation level that was barely within the limits of detection by 600 s.

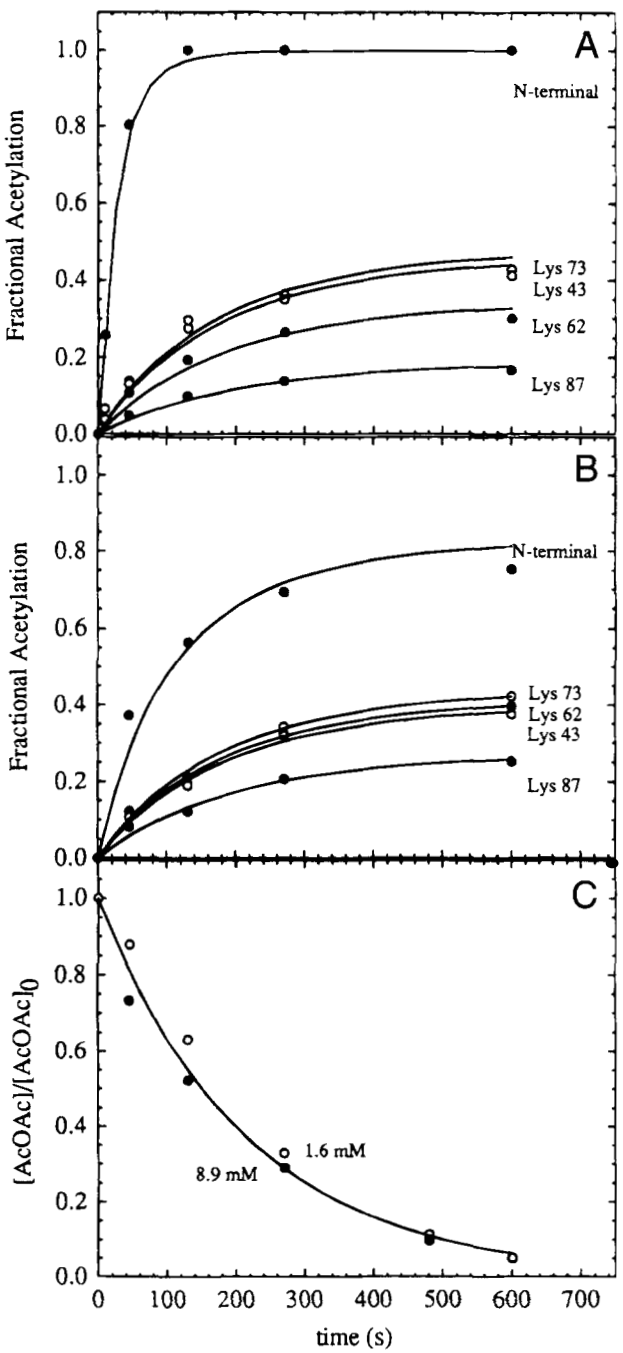

Fig. 5. Fractional acetylation as a function of time at pH 7.0 and 25 °C for the reactive lysines in free SSB (**A**) and ssDNA-bound SSB (**B**). SSB (8 μM) was reacted with 1.6 mM deuterated acetic anhydride in the absence of ssDNA and with 8.9 mM in the presence of ssDNA. The lines through each data set were generated using the rate constants presented in Table 1 and Equation 3 (the % fit error for the free SSB data sets were 3.0% for the amino-terminus, 8.2% for Lys 43, 10.6% for Lys 62, 10.6% for Lys 73, and 11.0% for Lys 87, whereas the fit error for the bound SSB data sets were 9.2% for the amino-terminus, 5.1% for Lys 43, 6.4% for Ly s 62, 4.6% for Lys 73, and 7.7% for Lys 87). (**C**) The hydrolysis of d_6-acetic acid takes place during the course of reaction. The line through these two data sets was generated using the rate constant of 0.0045 M^{-1} s^{-1} that was determined as described under Material and methods.

ticularly serious for electrospray ionization, where the distribution of charge states observed may depend on the side-chain modification. Yet electrospray ionization is emerging as the preferred ionization method for analysis of peptides because of its high sensitivity and compatibility with HPLC. Quantitative assessment of side-chain reactivities is also complicated because the quantitative release of the unmodified and modified peptides from the protein may be altered by the modified side chains. For example, trypsin does not cleave after lysine that has been acetylated. Non-quantitative release of peptides is equally important when radioactive reagents are used. Following a suggestion initially put forth by Steiner et al. (Steiner et al., 1991), we have used a two-step reaction procedure in which free and bound SSB are exposed to d_6-acetic anhydride, followed by exhaustive acetylation of lysyl residues with h_6-acetic anhydride. The reactivities of lysyl side chains in free and bound SSB were quantified from the relative levels of peptides displaying d_3 versus h_3 modification. Because all lysyl residues in SSB were acetylated, either d_3 or h_3, errors due to incomplete proteolysis or mass spectrometer response were minimized.

Correlation of lysine reactivity with SSB structure

Rate constants for acetylation at pH 7.0 of the six Lys residues in SSB varied from less than 0.25 M^{-1} s^{-1} to as high as 1.86 M^{-1} s^{-1}, a 10-fold range. The rate constant for acetylation of the amino-terminus was 23 M^{-1} s^{-1}, 10-fold larger than the rate constant for acetylation of the most reactive Lys residue. The reactivities of these primary amines are highly sensitive to their environments, facilitating their use as probes of these environments. Specific factors that may affect the reactivities of primary amines may be generally grouped as those that alter the pK_as of the protonated amines and those that alter the access the amines have to the acetylating reagent (i.e., the aqueous solvent). Although the pK_as of the amino-terminus and Lys residues in an unfolded polypeptide are predictable, their pK_as in the same protein, but folded into its native state, cannot be predicted with certainty because the protonated amines may interact with nearby residues. For example, formation of an ion-pair with aspartic or glutamic acid anions, as well as hydrogen bonding with nonbonding electrons in various heteroatoms can increase the pK_a, making the amine less reactive. The reactivities of the amines are also expected to be low if the amines are buried in folds or pockets such that they have little access to the solvent.

The range of reactivities of the amino-terminus and Lys residues of SSB can be correlated with the available folded structures. The folded structures of the homotetrameric SSB from *E. coli* (Raghunathan et al., 1997; Webster et al., 1997) and human mitochondria (Yang et al., 1997) were recently solved using X-ray crystallography. There are 36 identical and 28 highly conserved residues between residues 11 and 118 of HsmtSSB and 3 and 103 of *E. coli* SSB (Fig. 6). The sequence similarity (Fig. 6) and similar biochemical and hydrodynamic properties (Curth et al., 1994) suggested that these two proteins would fold into homologous structures. Indeed, the secondary structure of these two proteins are very homologous and the critical functional residues are clustered in homologous locations on the protein surface. However, there are significant differences between the HsmtSSB and *E. coli* SSB crystal structures with regard to the locations of some important residues. Therefore, it is instructive to consider those structural

A

```
consensus          ::: ::*  :: * :* *      %  :* :  :*::  W
Ecoli       1-ASR GVNKVILVG NLGQD PEVR YMPNGG AVANITLATSESWR
Human      10-LER SLNRVHLLG RVGQD PVLR QVEGKN PVTIFSLATNEMWR
secondary     ttt bbbbbbbbb ttttt bbbb tttttt bbbbbbbbbbbbbb
patches         B'          B        A                    CC

consensus      :      : :*:   : : W   : F       *:   :::  :*
Ecoli       42-D......KATGE MKEQTEWHRVVLF .GK LAEVASEYLR KGS
Human       51-SGDSEVY.QLGD VSQKTTWHRISVF RPG LRDVAYQYVK KGS
secondary      xxxxxxxxxxxx bbbbbbbbbbbbb ttt hhhhhhhhhh ttt
patches                       C  A A   C'C'             B

consensus      :    % *  :*  W      *:  ::  :*:
Ecoli       76-QVYIE GQL RTRK WT.DQSGQD RYTT EVVV-103
Human       91-RIYLE GKI DYGE Y.MDKNNVR RQAT TIIA-118
secondary      bbbbb ttt bbbb tttttttttt bbbb tttt
patches        B      C'       D   D   D A'
```

B

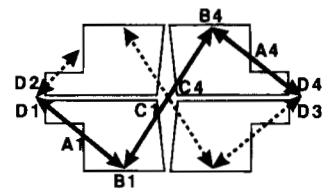

Fig. 6. A: The sequence alignment of *E. coli* and human mitochondrial SSB in their regions of greatest similarity. The "consensus" line indicates the degree of sequence similarity among bacterial SSBs for each position in the *E. coli* sequence; * indicates identical residues, : indicates highly conserved residues, and W and F indicate the four critical and identical aromatic residues. The "secondary" line indicates the type of secondary structure associated with each residue in the human mitochondrial sequence; t indicates turn or loop, b indicates beta strand, x indicates no structural resolution, and h indicates helix. The "patches" line indicates the human mitochondrial residues that belong to each electropositive patch (Yang et al., 1997) and the primes indicate the equivalent residues contributed by an adjacent subunit in the quaternary structure. **B:** A diagram showing the relationship of the electropositive patches to SSB quaternary structure and ssDNA binding. B1C1C4B4 form an electropositive channel across the front face of the tetramer and B3C3C2B2 form a channel across the back face. The bound ssDNA is shown as a double-headed arrow because its polarity, if any, with regard to the patches has not been determined. The ends are shown arbitrarily beginning and ending at patches A2 and B2. If the ssDNA were to make contact with all of the patches as shown, about 70 nucleotides of the polymeric ssDNA would be involved (Yang et al., 1997).

features that are similar and those that are different when trying to establish those features that are essential to SSB function.

The high resolution structures of the *E. coli* SSB indicate that Lys 7 is buried in the subunit binding surface, where it forms an ion pair with Glu 80. Finding Lys 7 unreactive in the present study is consistent with this location in folded *E. coli* SSB. These results are also consistent with the structure of HsmtSSB, which indicates that the equivalent residues, Arg 16 and Glu 95, form an ion pair in the subunit binding surface. Although the low reactivity of Lys 7 is consistent with the crystal structures of both human and *E. coli* SSB, the low reactivity of Lys 49 is anomalous. Lys 49 is not resolved in one *E. coli* SSB structure (Raghunathan et al., 1997), indicating that it is disordered and probably highly accessible to solvent. In contrast, Lys 49 is resolved in the other *E. coli* structure (Webster et al., 1997), points toward the ssDNA binding channel, but is still solvent accessible. Both of these structures predict that Lys 49 should be highly reactive, contrary to our observation. It may be significant that Ser 63 in HsmtSSB, the positional equivalent of Lys 49 in *E. coli* SSB, is buried and participating in a hydrogen bond. Congruence of the low reactivity found for Lys 49 with the X-ray crystallographic structure of HsmtSSB suggests a

need to reconsider the region around Lys 49 in the structures proposed for *E. coli* SSB.

The acetic anhydride labeling reactivities for the amino-terminus and Lys 43, 62, 73, and 87 in the free protein are well correlated to their surface exposure in the crystal structure. Neither the amino-terminus nor Lys 43 are resolved in either *E. coli* SSB crystal structures, but their locations suggest that both residues are located in flexible solvent-exposed regions of the protein (Raghunathan et al., 1997; Webster et al., 1997). Such well-exposed locations would yield the high reactivities that we observe for these residues. Lys 62 and 73 are well resolved and solvent exposed in both free *E. coli* SSB crystal structures. Their reactivities are as high as that for Lys 43, which is again consistent with high solvent exposure. Lys 87 is well resolved and partially buried in both *E. coli* SSB crystal structures, which correlates well with its being one-half to one-third as reactive as the other three lysines.

Relationship of the lysine residues to the other ssDNA binding residues

The high resolution structures for SSBs from all of the best-studied SSB families indicate that they carry out their functions by stacking or clustering a few aromatic residues around or near the nitrogenous bases. SSBs for which this structural similarity has been noted are Pf3 gene 5 protein, Ff gene 5 protein, the major cold shock protein *E. coli* CspA (Folmer et al., 1994), T4 gene 32 protein (Shamoo et al., 1995a, 1995b), human replication protein A (Bochkarev et al., 1997), and now human mitochondrial SSB (Yang et al., 1997), and *E. coli* SSB (Raghunathan et al., 1997; Webster et al., 1997). These SSBs share this structural feature even though there is practically no sequence conservation between, or sometimes even within, the families.

The architecture of all of these SSBs is also similar in that these aromatic residues are located near the loop end of two anti-parallel beta-strands. For instance, in each protomer of the homotetrameric *E. coli* SSB, Trp 40, 54, 88, and Phe 60 are found at the ends of the anti-parallel beta-strands (Fig. 6). Site mutants of these residues had previously established the importance of these residues for ssDNA affinity (Merrill et al., 1984; Casas-Finet et al., 1987; Khamis et al., 1987a, 1987b; Bayer et al., 1989; Curth et al., 1993). The evidence accumulated from these solution studies indicated that Trp 40, Trp 54, and Phe 60 partially stack with the nitrogenous bases and that Trp 88 helped form a hydrophobic cluster for the nitrogenous bases. In fact, Phe 60 was close enough to the nitrogenous bases that it could be photochemically crosslinked to $(dT)_8$ (Merrill et al., 1984). Finally, among bacterial and human mitochondrial SSBs, these four aromatic residues are the most conserved residues with regard to their positions within the linear sequence (Fig. 6). Thus, these four aromatic residues are conserved according to sequence, function, and structure.

The structures of all of the different SSB families suggest that they bind the sugar phosphate backbone with a line of electropositive residues located a suitable distance away from the aromatic residues described above. As a rule, these cationic residues are not conserved with regard to sequence, even within a given SSB family. The homotetrameric *E. coli* and human mitochondrial SSB structures share four conserved, electropositive, surface-exposed patches named A, B, C, and D (Webster et al., 1997; Yang et al., 1997). Because these patches are adjacent to the aromatic residues Trp 40, 54, 88, and Phe 60, these patches in the free protein crystal structure were hypothesized to help form the ssDNA binding sites. Patches B and C appear to play the most prominent role in that they form an electropositive channel that extends across the surface of the tetramer to patches C and B in the subunit located furthest away (Yang et al., 1997) (Fig. 6). For the ssDNA to bind to the BCCB channel on the other face of the tetramer, the ssDNA must twist past patches A and D at the ends of the tetramer because electronegative barriers and flexible loops block other routes.

Only about half of the residues that are responsible for the cationic clusters in *E. coli* and human mitochondrial SSBs are conserved with regard to protein sequence (Fig. 6). Among the cationic cluster residues, arginines are more prevalent and more conserved in sequence than are the lysines. This result is surprising given that one early study had found that the lysines but not the arginines were critical for function in *E. coli* SSB (Bandyopadhyay & Wu, 1978). Even though the lysine residues are not conserved according to sequence, they are conserved according to their structure. They fall into one of the recently identified cationic clusters that may define the ssDNA binding sites.

Our ssDNA protection data (Fig. 5B; Table 1) indicate that the amino-terminus and the reactive lysines Lys 43, 62, 73, and 87 in *E. coli* SSB are protected by ssDNA binding. It is intriguing to note that the acetylation rates for Lys 43, 62, and 73 are all reduced to nearly the same rate by bound ssDNA. These three residues, along with the amino-terminus, belong to the patch BCCB channel along the surface of *E. coli* SSB. The reactivities of these three lysines are probably limited by the same factor, binding to ssDNA. That they are reduced to the same rate may indicate that they bind in the same way to the ssDNA because they are part of the same channel. Even though the amino-terminus is also a member of patch B, its lower pK_a would make it inherently more reactive despite being protected in the same way as these three lysines.

The labeling rate of Lys 87 is reduced by a different extent and to a different rate when ssDNA is bound. The structure of the free protein also indicates that this residue is different from the other lysines. It is more buried than the other lysines and is on the same face of a beta strand as Thr 85, which makes a critical hydrogen bond with His 55 to stabilize the SSB quaternary structure (Williams et al., 1984; Bujalowski & Lohman, 1991). When ssDNA binds to the essential patch D residue Trp 88 located on the other side of that beta strand, it must protect Lys 87 both by bonding to it and by steric hindrance to a greater extent than it does with the other Lys residues.

Materials and methods

E. coli SSB

SSB was purified as previously described (Lohman et al., 1986). Its purity was verified by RP-HPLC and ESIMS. The concentration of SSB was determined by UV spectroscopy using an extinction coefficient at 280 nm of 113,000 M^{-1} cm^{-1} per tetramer (Lohman & Overman, 1985).

ssDNA

M13Gori ssDNA, a chimeric ssDNA consisting of a 2,216-nucleotide G4 complementary strand origin inserted into a 6,407-nucleotide M13 ssDNA, was purified from M13 phage as previously described (Griep & Lokey, 1996). The concentration of nucleotides in M13Gori was determined by UV spectroscopy using an extinc-

tion coefficient at 260 nm of 7,370 M^{-1} cm^{-1} (Berkowitz & Day, 1971).

Acetylation of SSB at pH 8.1

d_6-Acetic anhydride (1.2 mM) was added to 8 μM of SSB monomer, with or without ssDNA, in 250 μL 30 mM Tris, 90 mM NaCl, pH 8.1 at 25 °C. The reaction was allowed to proceed for 15 min, at which time all the d_6-acetic anhydride had either hydrolyzed or reacted with SSB. The protein retains its ability to bind ssDNA after this partial labeling step as confirmed by fluorescence quenching. Guanidine hydrochloride was added to a final concentration of about 6 M, and the protein was allowed to denature for an hour at room temperature. SSB was then separated from ssDNA and low molecular weight compounds by size exclusion chromatography on a Superose-12 HR (Pharmacia) column pre-equilibrated with 3 M guanidine hydrochloride. About 1.5 mL of SSB containing fraction was collected and 1.5 mL of 6 M guanidine hydrochloride was added to this fraction. Then, 200 μL of 1 M acetic anhydride in acetonitrile was added to the fraction in three portions to completely acetylate the protein as suggested by Steiner et al. (1991). During this reaction, the pH was maintained between 6.5–7.5 by the addition of KOH. The pH of the reaction mixture was then raised to 11 to hydrolyze the O-acetyltyrosines (Riordan & Vallee, 1972). The fully acetylated protein was desalted on a RP-HPLC C4 column (Vydac) and vacuum dried to about 100 μL before proteolytic digestion.

SSB acetylation at pH 7.0

The reaction was carried out in 200 mM HEPES, 90 mM NaCl, pH 7.0. d_6-Acetic anhydride (1.6 mM) was added to free SSB and 8.9 mM d_6-acetic anhydride was added to SSB bound to ssDNA. The reactions were allowed to proceed for 10 s to 10 min and quenched with 80 μL of 350 mM L-Lys solution (pH 8.0). The protein was fully acetylated as described above.

Enzymatic digestion and HPLC separation

The modified protein was divided into two aliquots. One portion was digested with trypsin (Sigma) for 2 h, and the other with Endoproteinase Glu-C (Worthington Enzymes) for 3 h at 37 °C. The digestion buffer was 40 mM Tris, pH 7.8. The enzyme to substrate ratio (w/w) was 1:20 for trypsin and 1:10 for Endoproteinase Glu-C. The resulting peptides were separated on a RP-HPLC C18 Vydac column using a 2 to 90% gradient of acetonitrile in water with 0.1% trifluoroacetic acid. Peak fractions were collected and analyzed by ESI-MS.

Mass spectrometry

Mass spectrometric analysis was performed on a Micromass Platform II quadruple mass spectrometer with a Masslynx data acquisition and processing software. A Harvard Model 22 syringe pump was used to deliver a solution of water:acetonitrile (50:50) to the spray needle. The sample was introduced into the solution via a Rheodyne 8125 injector. For protein analysis, the mass range 700–2,000 was calibrated with myoglobin. The mass range (300–2,000) was calibrated with NaI for peptide analysis, and a smaller mass range (50–100) was used for each peptide to improve the signal-to-noise ratio.

Acetylation of substance P

Substance P (4 μg), with or without ssDNA, was acetylated using the same conditions as used for acetylation of SSB at pH 8.1. The fully modified substance P was separated from the reaction mixture and introduced to the mass spectrometer source by a RP-HPLC C18 microbore column using a gradient of 2 to 40% acetonitrile in water with 0.1% trifluoroacetic acid.

Derivation of the acetylation kinetic expression

Protein acetylation with d_6-acetic anhydride is a second-order reaction:

$$-\frac{d[\text{amine}]}{dt} = k_{\text{acetylation}}[\text{AcOAc}][\text{amine}] \quad (1)$$

where [amine] is unreacted amine concentration, $k_{\text{acetylation}}$ is the observed second-order rate constant, and [AcOAc] is the unreacted d_6-acetic anhydride concentration. This expression could not be used to determine the rates of lysine acetylation because acetic anhydride is rapidly hydrolyzed by water (Kemp & Kohnstam, 1956). However, an alternate rate expression, which accounts for water hydrolysis of acetic anhydride, has been developed. Because the acetic anhydride concentration (1.6 or 8.9 mM) is significantly greater than that of the amines (7×8 μM), loss of acetic anhydride due to protein acetylation can be neglected and acetic anhydride hydrolysis is described by pseudo–first-order kinetics:

$$[\text{AcOAc}] = [\text{AcOAc}]_0 \exp(-k_{\text{hydrolysis}}) \quad (2)$$

where $[\text{AcOAc}]_0$ is the initial acetic anhydride concentration.

Substituting Equation 2 into 1, integrating, and rearranging to solve for the fraction of labeled amine yields the following experimentally useful expression:

$$\frac{[\text{labled amine}]}{[\text{amine}]} = 1 - \exp\left(\frac{k_{\text{acetylation}}[\text{AcOAc}]_0\{\exp(-k_{\text{hydrolysis}}t) - 1\}}{k_{\text{hydrolysis}}}\right). \quad (3)$$

To obtain $k_{\text{acetylation}}$, the fraction of labeled amine versus time data were fit to Equation 3 using Excel 5.0 (Microsoft Corporation) and a nonlinear least-squares fit plug-in program called FIT! (WindowChem Software, Fairfield, California). The value for $k_{\text{hydrolysis}}$ was 0.0045 s^{-1}, as described in the next section.

Hydrolysis rate of d_6-acetic anhydride

The rate constant for hydrolysis of acetic anhydride under the conditions used in this study was determined to facilitate application of Equation 2. The amount of d_6-acetic anhydride remaining in solution after exposure to 200 mM HEPES, pH 7.0, for various times was determined from the amount of a dipeptide Gly-Phe that was acetylated (5.4 mM for 1.6 mM d_6-acetic anhydride and 23 mM for 8.9 mM d_6-acetic anhydride). The dipeptide reacted with the d_6-acetic anhydride for 15 min, a sufficient length of time for complete acetylation. The acetylated and nonacetylated forms of Gly-Phe were separated on a RP-HPLC C18 column (Vydac) using 11% acetonitrile in water with 0.1% trifluoroacetic acid. The

acetylated and nonacetylated forms of the dipeptide were quantified from their UV absorbance versus a standard curve. When the hydrolysis time course data set was fit to a single exponential, the value for $k_{hydrolysis}$ was determined to be 0.0045 ± 0.0002 s^{-1} (Fig. 5C).

Acknowledgments

This work was supported in part by a graduate fellowship from the University of Nebraska Center for Biotechnology to J. Chen, by NIH Grant GM40384 to D.L. Smith, and the Nebraska Center for Mass Spectrometry, and by NSF Grant MCB-9600544 to M.A. Griep. We thank Alan Linderman for isolating the SSB and Besty J. Adkins for isolating the M13Gori.

References

Akashi S, Nitsu U, Yuji R, Ide H, Hirayama K. 1993. Investigation of the interaction between enzyme and inhibitor by the combination of chemical modification, electrospray ionization mass spectrometry and frit-fast atom bombardment liquid chromatography/mass spectrometry. *Biol Mass Spectrom 22*:124–132.

Bandyopadhyay PK, Wu C-W. 1978. Fluorescence and chemical studies on the interaction of *Escherichia coli* DNA-binding protein with single-stranded DNA. *Biochemistry 17*:4078–4085.

Bayer I, Fliess A, Greipel J, Urbanke C, Maass G. 1989. Modulation of the affinity of the single-stranded DNA binding protein of *Escherichia coli* (*E. coli* SSB) to Poly(dT) by site-directed mutagenesis. *Eur J Biochem 179*:399–404.

Berkowitz SA, Day LA. 1971. Molecular weight of single-stranded fd bacteriophage DNA. High speed equilibrium sedimentation and light scattering measurements. *Biochemistry 13*:4825–4831.

Bochkarev A, Pfuetzner RA, Edwards AM, Frappier L. 1997. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature 385*:176–181.

Bujalowski W, Lohman TM. 1991. Monomers of the *Escherichia coli* SSB-1 mutant protein bind single-stranded DNA. *J Mol Biol 217*:63–74.

Casas-Finet JR, Khamis MI, Maki AH, Ruvolo PP, Chase JW. 1987. Optically detected magnetic resonance of tryptophan residues in *Escherichia coli* ssb gene product and *Escherichia coli* plasmid-encoded single-stranded DNA-binding proteins and their complexes with poly(deoxythymidylic) acid. *J Biol Chem 262*:8574–8583.

Curth U, Greipel J, Urbanke C, Maass G. 1993. Multiple binding modes of the single-stranded DNA binding protein from *Escherichia coli* as detected by tryptophan fluorescence and site-directed mutagenesis. *Biochemistry 32*:2585–2591.

Curth U, Urbanke C, Greipel J, Gerberding H, Tiranti V, Zeviani M. 1994. Single-stranded-DNA-binding proteins from human mitochondria and *Escherichia coli* have analogous physicochemical properties. *Eur J Biochem 221*:435–443.

Folmer RHA, Folkers PJM, Kaan A, Jonker AJ, Aelen JMA, Konings RNH, Hilbers CW. 1994. Secondary structure of the single-stranded DNA binding protein encoded by filamentous phage Pf3 as determined by NMR. *Eur J Biochem 224*:663–676.

Glocker MO, Borchers C, Fiedler W, Suckau D, Przybylski M. 1994. Molecular characterization of surface topology in protein tertiary structures by amino-acylation and mass spectrometric peptide mapping. *Bioconjugate Chem 5*:583–590.

Griep MA, Lokey ER. 1996. The role of zinc and the reactivity of cysteines in *Escherichia coli* primase. *Biochemistry 35*:8260–8267.

Hasan A, Smith JB, Qin W, Smith DL. 1993. The reaction of bovine lens αA-crystallin with aspirin. *Exp Eye Res 57*:29–35.

Kemp IA, Kohnstam G. 1956. Decomposition of inorganic cyanates in water. *J Chem Soc 191*:900–911.

Khamis MI, Casas-Finet JR, Maki AH, Murphy JB, Chase JW. 1987a. Investigations of the role of individual tryptophan residues in the binding of *Escherichia coli* single-stranded DNA binding protein to single-stranded polynucleotides. *J Biol Chem 262*:10938–10945.

Khamis MI, Casas-Finet JR, Maki AH, Murphy JB, Chase JW. 1987b. Role of tryptophan 54 in the binding of *Escherichia coli* single-stranded DNA-binding protein to single-stranded polynucleotides. *FEBS Lett 211*:155–159.

Lohman TM, Ferrari ME. 1994. *Escherichia coli* single-stranded DNA-binding protein: Multiple DNA-binding modes and cooperativities. *Annu Rev Biochem 63*:527–570.

Lohman TM, Green JM, Beyer RS. 1986. Large-scale overproduction and rapid purification of the *Escherichia coli ssb* gene product: Expression of the *ssb* gene under λ P_L control. *Biochemistry 25*:21–25.

Lohman TM, Overman LB. 1985. Two binding modes in *Escherichia coli* single strand binding protein–single stranded DNA complexes: Modulation by NaCl concentration. *J Biol Chem 260*:3594–3603.

McMacken R, Silver L, Georgopoulos C. 1987. DNA replication. In: Neidhardt FC., ed. *Escherichia coli and Salmonella typhimurium: Cellular and molecular biology*. Washington, DC: American Society for Microbiology. pp 564–612.

Merrill BM, Williams KR, Chase JW, Konigsberg WH. 1984. Photochemical cross-linking of the *Escherichia coli* single-stranded DNA-binding protein to oligodeoxynucleotides: Identification of phenylalanine 60 as the site of cross-linking. *J Biol Chem 259*:10850–10856.

Meyer RR, Laine PS. 1990. The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol Rev 54*:342–380.

Overman LB, Bujalowski W, Lohman TM. 1988. Equilibrium binding of *Escherichia coli* single-strand binding protein to single-stranded nucleic acids in the $(SSB)_{65}$ binding mode: Cation and anion effects and polynucleotide specificity. *Biochemistry 27*:456–471.

Qin W, Smith JB, Smith DL. 1992. Rates of carbamylation of specific lysyl residues in bovine α-crystallins. *J Biol Chem 267*:26128–26133.

Raghunathan S, Ricard CS, Lohman TM, Waksman G. 1997. Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength X-ray diffraction on the selenomethionyl protein at 2.9 Å resolution. *Proc Natl Acad Sci USA 94*:6652–6657.

Riordan JF, Vallee BL. 1972. Acetylation. *Methods Enzymol 25*:494-499.

Sancar A, Williams KR, Chase JW, Rupp WD. 1981. Sequences of the *ssb* gene and protein. *Proc Natl Acad Sci USA 78*:4274–4278.

Shamoo Y, Friedman AM, Parsons MR, Konigsberg WH, Steitz TA. 1995a. Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. *Nature 376*:362–366.

Shamoo Y, Friedman AM, Parsons MR, Konigsberg WH, Steitz TA. 1995b. Erratum. *Nature 376*:616.

Steiner RF, Albaugh S, Fenselau C, Murphy C, Vestling M. 1991. A mass spectrometry method for mapping the interface topography of interacting proteins, illustrated by the mellitin–calmodulin system. *Anal Biochem 196*:120–125.

Webster G, Genschel J, Curth U, Urbanke C, Kang C-H, Hilgenfeld R. 1997. A common core for binding single-stranded DNA: Structural comparison of the single-stranded DNA-binding proteins (SSB) from *E. coli* and human mitochondria. *FEBS Lett 411*:313–316.

Williams KR, Murphy JB, Chase JW. 1984. Characterization of the structural and functional defect in the *Escherichia coli* single-stranded DNA-binding protein encoded by the *ssb-1* mutant gene: Expression of the *ssb-1* gene under λ_{pL} regulation. *J Biol Chem 259*:11804–11811.

Yang C, Curth U, Urbanke C, Kang C-H. 1997. Crystal structure of human mitochondrial single-stranded DNA binding protein at 2.4 Å. *Nat Struct Biol 4*:153–157.

Young NM, Kaplan H. 1989. Chemical characterization of functional groups in proteins by competitive labelling. In: Creighton TE, ed. *Protein function: A practical approach*. Oxford: IRL Press. pp 225–246.