An electrospray ionisation mass spectrometry (ESI-MS) study to probe the metal ion binding site in the DNA binding domain of the yeast transcriptional activator GAL4

Paresh L. Gadhavi*

Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

Received 22 September 1997

Abstract Electrospray ionisation mass spectrometry (ESI-MS) is used to detect metal ions and their stoichiometry of binding in the DNA binding domain of GAL4. In this analysis, the mass spectra of the apo- and metallo-proteins differ by both mass and charge, precluding the possibility of random adduct formation. Deuterium exchange NMR experiments of Zn(II)-GAL4(7–49) indicate that the binuclear metal ion structure, which is shown to have a net negative charge of −2, is the recipient of several hydrogen bonds, notably from the main-chain amide protons of the ligating cysteine residues, indicating the charge is stabilised in this manner.

Key words: GAL4; DNA binding; Electrospray ionisation mass spectrometry; Metal binding

1. Introduction

The zinc finger proteins are an important class of gene regulatory protein, found predominantly in the eukaryotes. This motif consists of a small protein domain whose structure is stabilised by the interaction with metal ions. The prototype of this class of protein is the transcription factor TFIIIA [1] from *Xenopus laevis*, but structural studies of several of these proteins indicate that they may be divided into at least three different classes [2–4] (see Fig. 1).

GAL4, like many other eukaryotic transcriptional activators, has functionally separable DNA binding and activation domains [5]. The DNA binding domain has been localised to the first 147 residues of GAL4, which binds sequence-specifically to the upstream activation sequence for galactose (*UASg*) to activate expression of the *GAL* genes in *Saccharomyces cerevisiae* [6]. Atomic absorption spectroscopy, EXAFS [7] and ^{113}Cd-nuclear magnetic resonance (NMR) [8] studies of the DNA binding domain of GAL4 showed that it contains two zinc ions per molecule and suggested that the metal ions were co-ordinated via the six cysteines in the sequence (residues 11–38: -Cys-X_2-Cys-X_6-Cys-X_6-Cys-X_2-Cys-X_6-Cys-). Removal of the metal ions abolishes DNA binding, implying that they are responsible for maintaining a functional structure.

To study the metal ion co-ordination and the overall structure of the DNA binding domain of GAL4, the ^{1}H-NMR spectra were first assigned for both the Zn(II) [9] and the Cd(II) forms of GAL4(7–49). Subsequently, the polypeptide-metal cluster connectivities were determined based on the pattern of ^{1}H-^{113}Cd heteronuclear couplings seen in ^{113}Cd(II)-GAL4(7–49). These results showed that cysteines 11 and 28 are bridging ligands in a C_6 two-metal ion cluster [4]. Structure determination of the Zn(II)-GAL4(7–49) using distance constraints from 2D-NOESY ^{1}H-NMR experiments showed a distinct topological fold of a two-helix-turn-strand motif packed around the Zn_2Cys_6 cluster [10]. The structure around the Zn_2Cys_6 cluster, as determined by NMR, is isomorphous with the co-crystal structure of Cd(II)-GAL4(1–65)$^{+1}$ bound to the *UASg* DNA binding site [11].

In this paper, the utility of ESI-MS is demonstrated in studies to show the presence of either the Zn(II) (native) form or Cd(II) (metal substituted) form of the DNA binding domain of GAL4, GAL4(1–49). The results show that both metal ion binding and differences in charge of the apo- and metallo-proteins are due to metal ion chelation and indicate that ESI-MS can be used to detect the net charge of the binuclear cluster in the DNA binding domain of GAL4. These studies are further correlated by deuterium exchange NMR experiments which show a pattern of hydrogen bonds from the main chain amide (NH) protons to the sulphur (Sγ) atoms of the ligating cysteine residues within the binuclear cluster.

2. Materials and methods

2.1. Protein expression and purification

GAL4 sequences encoding the following amino-terminal fragment GAL4(1–49) was expressed under the control of a *tac* promoter and over-produced in *Escherichia coli* strain JM101.

All proteins were purified by cation exchange chromatography (Fast-S and MonoS FPLC) and judged to be over 95% pure by reversed-phase HPLC, as described before [12]. Samples for mass spectrometry were concentrated by ultrafiltration using Amicon YCO5 membranes and then exchanged by dialysis into 50 μM $ZnCl_2$. Samples for cadmium substitution were prepared by initially adding a 5-fold molar excess of $CdCl_2$ to the protein, followed by additional aliquots of $CdCl_2$, monitoring the change in absorbance between 230 and 250 nm until saturation was reached. Excess cadmium was removed by dialysis against 50 μM $CdCl_2$. Protein concentrations were determined by amino acid analysis. All proteins were stored at −70°C until use.

2.2. High performance liquid chromatography (HPLC)

Samples for mass spectral analysis of the apo-protein were purified by reversed-phase HPLC on a Hewlett-Packard 1090L liquid chromatograph using an analytical C_{18} microbore column (Brownlee Aquapore RP18 with 3.2 mm×1.5 cm guard and 2.1 mm×3 cm column cartridges). The protein was eluted with a gradient of 0–70% acetonitrile containing 0.1% (v/v) trifluoroacetic acid, whilst monitoring at either 210 or 280 nm. Purified proteins were dried in vacuo prior to further use.

*Correspondence address: Division of Membrane Biology, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK. Fax: (44) (181) 906-4477.
E-mail: p-gadhav@nimr.mrc.ac.uk

PII S0014-5793(97)01272-6

2.3. Mass spectrometry

Proteins were analysed, in both the native and denatured forms, by ESI-MS using a MicroMass (Fisons/VG) Bio-Q quadrupole mass spectrometer. The solvent flow of the system was a mixture of 50% (v/v) methanol and 1% (v/v) acetic acid in water, containing 2 μg/ml gramicidin-S as an internal standard. The flow was 2–4 μl/min with detection of positive ions. Calibrations were performed using PEG600 for the mass/charge range *m/z* 400–1400. Each sample analysed contained 50% (v/v) methanol and 20–100 pmol/μl of protein and was injected in 10 μl amounts via a loop injector into the mass spectrometer source. The cone exit voltages (B1) were set between 55 and 90 V. All spectra were recorded in multiple scan acquisition mode and processed using the manufacturer's software. Molecular weights were determined, in atomic mass units (amu), again using the manufacturer's software.

2.4. NMR spectroscopy

Samples for deuterium exchange were prepared by dialysing the protein, Zn(II)-GAL4(7–49), from high salt and buffer into 10 μM zinc chloride, 1% (w/v) ammonium bicarbonate pH 7.0. After extensive dialysis, the protein was dried in vacuo, resolubilised in 90% D_2O and dried twice and then finally resolubilised in 99.98% D_2O and adjusted to p*D* 5.40. The concentration of the protein solution was 1.5–2 mM. Two-dimensional NMR spectra were then recorded immediately at 303 K on a Bruker AMX 600 spectrometer. Data processing was carried out using either the manufacturer's software or the FTNMR package kindly provided by Dr. Dennis Hare. Two-dimensional NOESY spectra were recorded with a mixing time of 200 ms. The NOESY spectra were acquired with a data size of $4K(t_2) \times 512(t_1)$ points and 32 scans for each t_1 point, and under instrumental settings as described before [9].

3. Results

Initially the molecular weight of the apo-protein was determined by ESI-MS prior to analysis of the native form. The amino-terminal fragment of GAL4, GAL4(1–49), was studied containing the cysteine-rich sequence, known to be important for the metal and DNA binding activities. Analysis of the mass spectra showed that GAL4(1–49) contained two metal ions (see Fig. 2); the metallo-protein could be distinguished from the apo-protein not only because it had a different molecular weight but also by a shift in the distribution or envelope of its charged species (see Fig. 2). This shift by two charge units strongly suggests the actual presence of specifically bound metal ions as opposed to random adduct formation. One can deduce that in the native form of the protein, the thiolate groups from the six cysteine residues (charge −6) bind either two zinc or two cadmium ions (charge +4) to give a protein whose net charge is two units more negative than the apo-protein. In the simplest analysis, the overall charge of the metallo-protein is more negative due to the binuclear structure directly (see Fig. 3). In support of this, deuterium exchange NMR experiments for Zn(II)-GAL4(7–49) show that the main-chain amide (NH) protons of four of the ligating cysteines, Cys-11, Cys-14, Cys-28, Cys-31, exchange slowly in D_2O (see Fig. 4). An analysis of the structure shows that the amide hydrogen atoms of the zinc ligating cysteines can form hydrogen bonds to the sulphur (Sγ) atoms of other ligating cysteines (see Fig. 4). Further analysis of the structure shows that the cysteine-rich region of GAL4 contains two helix-turn-strand motifs between Cys-11 and Lys-23 and between Cys-28 and Tyr-40 and that these motifs pack around the Zn_2Cys_6 cluster and the side chains of Leu-19 and Trp-36, so that the negative charge of the zinc cluster is partially offset by the helix dipoles [10,11,13]. Another possible explanation for the overall net negative charge of the metallo-protein is that in the native conformation changes elsewhere in the tertiary fold account for the extra two negative charges, as some other ESI-MS studies have shown [14] (for review see [15]).

The use of ESI-MS for detection of bound metal ions was sensitive to a number of conditions in these studies, three of which need particular mention. First, the sample composition of the native protein, second, the presence of acid in the sample mixture and the delivery flow of the system and third, the cone exit voltage applied during detection. Firstly, protein samples containing salts over 1 mM in the final mixture gave poorer sensitivity than those containing less (0.1 mM salt). The presence of salts in a sample mixture suppresses the electrospray process by changing the dielectric of the solution so that higher threshold voltages are required for ion production. Furthermore, a 50% (v/v) mixture of methanol, although mildly denaturing, gave better sensitivity than an aqueous sample (data not shown).

Because GAL4 binds either zinc or cadmium ions via cysteinyl thiolate ligands, changes in pH affect the stability of the protein. Under acid conditions, for example, the thiolate groups can become protonated resulting in loss of the metal. In all cases, however, some proportion of the protein was still able to withstand the acid conditions of the system allowing detection of the metallo-protein forms (note that 1% (v/v) acetic acid is 175 mM, about pH 3). Additionally, it was noted that the cadmium form of the metallo-protein was more stable than the zinc form, which is not surprising given that cadmium readily displaces zinc in the protein. Studies showing the effect of acid upon the sample have also been reported for bovine cytochrome *c*, which under different pH conditions gave markedly different electrospray ionisation mass spectra. These spectra are believed to correspond to unfolded (pH 2.6) and folded (pH 5.2) conformational states of the protein [14].

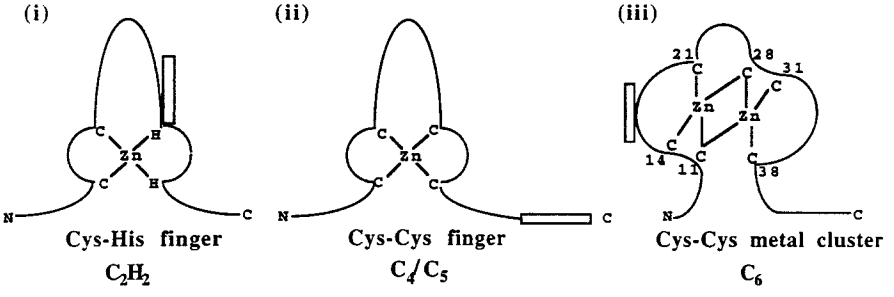

Fig. 1. Schematic representation of the three different classes of zinc finger. The regions encoding specificity in DNA binding are shown by the rectangles. (i) The C_2H_2 zinc finger typified by TFIIIA, (ii) the C_4/C_5 zinc finger found in many of the steroid hormone receptor family and (iii) the C_6 two metal ion cluster found in GAL4 and related proteins.

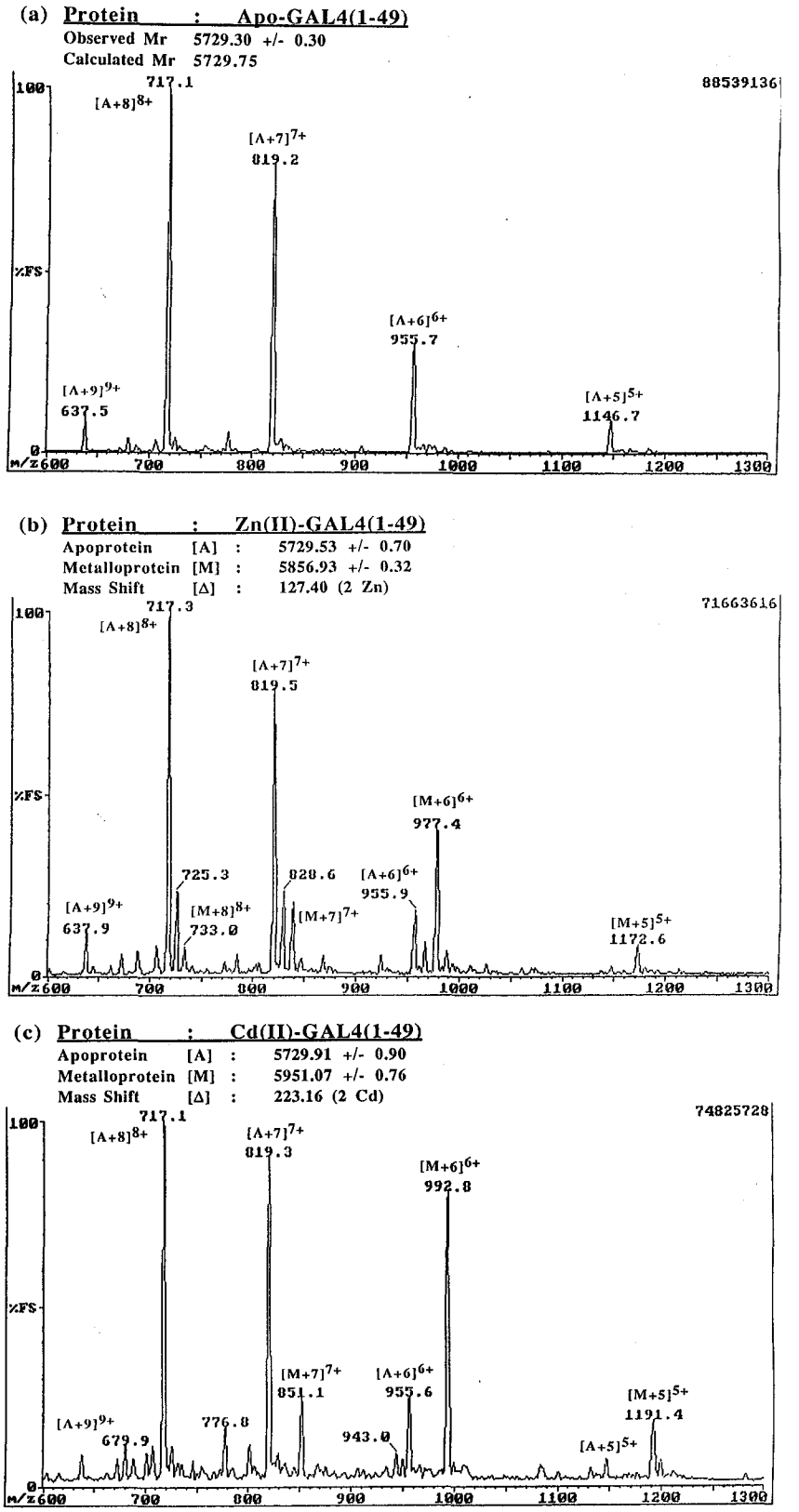

Fig. 2. Electrospray ionisation mass spectra of (a) GAL4(1–49) purified by reverse phase HPLC, (b) native Zn(II)-GAL4(1–49) and (c) cadmium substituted Cd(II)-GAL4(1–49). The peaks labeled $(A+n)^{n+}$ denote the net charge (n) of the apo (A) form and $(M+n)^{n+}$ the net charge of the metallo (M) form of the protein. Cone extraction voltages were set at 55 V for both analyses.

Interestingly, the major charge species in each envelope of the two forms differed by six negative charge units indicating that ESI-MS can detect charge differences between denatured and native conformations.

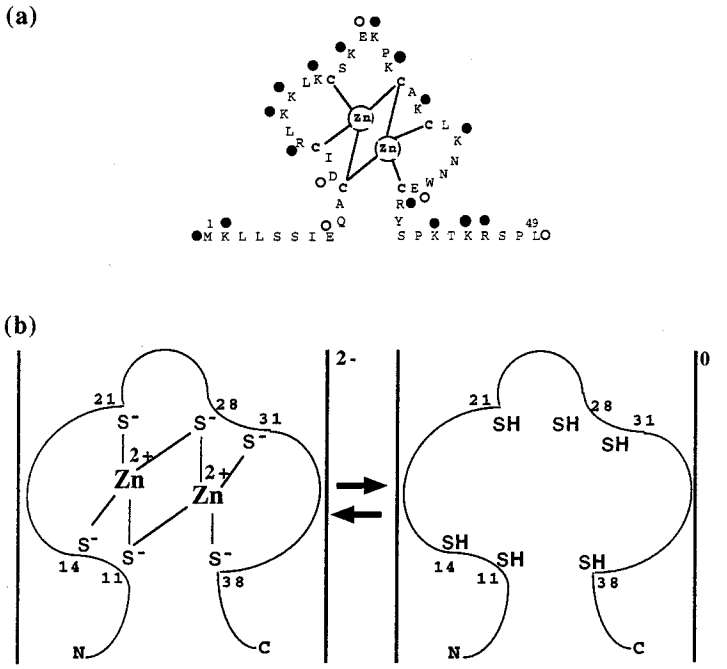

Fig. 3. Schematic representations of the binuclear metal ion cluster of GAL4 depicting: (a) the sequence of GAL4(1–49) showing the charged residues (positive groups with filled circles and negative groups with open circles); and (b) the charge effects due to metal ion binding in the metallo- (M) and apo- (A) proteins.

The third factor, the choice of cone exit voltage, was critical for the detection of the metallo-protein species. Precisely how this affects the sample is not understood, but studies of the transition metal complex, [RuII(2,2′-bipyridyl)$_3$]Cl$_2$, showed that changes in the cone exit voltage could give rise to various levels of desolvation. Specifically, intermediary values (90 V) gave complete desolvation during detection, whereas high values (190 V) gave rise to fragmentation [16]. Another factor for the requirement of higher cone extraction voltages is the presence of anions in solution. Anion pairing to cations in proteins reduces the sensitivity of positive ion detection [17]. As with desolvation, higher cone extraction voltages counter this effect. It was discovered that the larger protein fragments of GAL4 required higher cone exit voltages for detection of the metallo-protein (data not shown). While the correlation between molecular weight and cone extraction voltage is strong, it is difficult to explain this mechanistically and these observations remain empirical. Furthermore, it is stated here that observation of metallo-proteins by ESI-MS, shown here, may be highly dependent upon the type of instrument and its electrospray and tuning capabilities.

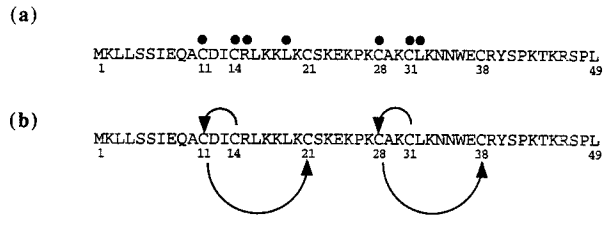

Fig. 4. Structural factors stabilizing the binuclear metal ion cluster in GAL4(1–49); (a) slow-exchanging protons of main-chain amide groups after deuterium exchange; (b) deduced hydrogen bonds between the main-chain amide (NH) and side-chain sulphurs (Sγ) of the ligating cysteinyl residues.

4. Discussion

Structural studies, particularly by NMR, have been used to probe the nature of the metal ion co-ordination of a number of fragments of the DNA binding domain of GAL4. Heteronuclear ^{1}H-^{113}Cd two-dimensional NMR experiments on ^{113}Cd(II)-GAL4(7–49) showed that all six cysteines co-ordinate cadmium with residues 11 and 28 acting as bridging ligands [4]. This two-metal ion cluster structure is more analogous to the metallothioneins [18] than to the other classes of zinc finger proteins, typified by the TFIIIA-like [2] and the glucocorticoid receptor [3] proteins. The cysteine-rich sequence responsible for metal ion binding in the DNA binding domain of GAL4 is also found in a number of other fungal gene regulatory proteins suggesting that it has been conserved as a DNA binding motif [19].

The detection of bound metal ions in proteins by ESI-MS described here demonstrates its applicability as a tool for the detection of metal-protein complexes and also shows its potential for the study of different states of a protein based on charge. This is made possible by (a) the relatively soft method of ionisation of the protein during the electrospray process and (b) the intrinsically polar nature of the proteins studied. The solvent used is a mildly acidified methanol-water mixture and the source temperature is 20°C or below, hence avoiding the denaturing conditions of acidic pH, pure organic solvents and high source temperatures. Other studies have utilised these conditions for the study of both protein-protein and ligand binding by ESI-MS [20,21] (for review see [15]). A molecular weight analysis of isopenicillin N synthetase from *Cephalosporium acromonium* CO728 by ESI-MS postulated the presence of a partially hydrated iron atom in the enzyme. This finding correlated with both spectroscopic and catalytic observations, each indicating the presence of bound iron [22]. Other studies have shown that metal ions can be detected for

stoichiometry determination in zinc finger proteins, such as in the DNA binding domain of 1,25-dihydroxyvitamin D_3 receptor [23] or for calcium ions in the EF-hands motifs of calbindin D_{28K} protein [24]. Another highly innovative use of ESI-MS has been the observation of folded states in proteins. For example, α-lactalbumin was folded in the presence of D_2O, transferred to H_2O and changes in mass via the deuteron were observed between the unfolded, transient and folded states by ESI-MS (for review see [25]).

In conclusion, the ESI-MS technique offers a simple method for the detection of bound metal ions in a protein, either suspected or known to require metal binding for its biological activity. Significantly, the presence of bound metal ions can also give rise to a change in the charge of the protein, a feature which the electrospray spectrum is able to reflect uniquely by changes in the envelope of the ionised species.

Acknowledgements: Thanks to Paul Skelton and Dick Barton (University Chemical Laboratory, University of Cambridge) for their technical help in using the mass spectrometer, Juliette Jacoby for amino acid analysis, Peter Alefounder for help with construction of the GAL4 subclone, and Professor Mark Ptashne (Harvard University) for the generous gift of the plasmid expressing GAL4(1–147). Thanks also to Brian Green at MicroMass Ltd (previously Fisons/VG MassLabs) Altrincham, UK, for help during the initial work and Ernest Laue in the Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge for help with the NMR experiments.

References

[1] Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609–1614.
[2] Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. and Wright, P.E. (1989) Science 245, 635–637.
[3] Härd, T., Kellenbach, E., Boelen, R., Maler, B.A., Dahlman, K., Freedman, L.P., Carlstedt-Duke, J., Yamamoto, K., Gustafsson, J.-A. and Kaptein, R. (1990) Science 249, 157–160.
[4] Gadhavi, P.L., Davis, A.D., Povey, J.F., Keeler, J. and Laue, E.D. (1991) FEBS Lett. 281, 223–226.
[5] Ptashne, M. (1988) Nature 335, 683–689.
[6] Keegan, L., Gill, G. and Ptashne, M. (1986) Science 231, 699–704.
[7] Povey, J.F., Diakun, G.P., Garner, C.D., Wilson, S.P. and Laue, E.D. (1990) FEBS Lett. 266, 142–146.
[8] Pan, T. and Coleman, J.E. (1989) Proc. Natl. Acad. Sci. USA 86, 3145–3149.
[9] Gadhavi, P.L., Raine, A.R., Alefounder, P.R. and Laue, E.D. (1990) FEBS Lett. 276, 49–53.
[10] Kraulis, P.J., Raine, A.R.C., Gadhavi, P.L. and Laue, E.D. (1992) Nature 356, 448–450.
[11] Marmorstein, R., Carey, M., Ptashne, M. and Harrison, S.C. (1992) Nature 356, 408–414.
[12] Gadhavi, P., Morgan, P.J., Alefounder, P. and Harding, S.E. (1996) Eur. Biophys. J. 24, 405–412.
[13] Baleja, J.D., Marmorstein, R., Harrison, S.C. and Wagner, G. (1992) Nature 356, 450–453.
[14] Chowdhury, S.K., Katta, V. and Chait, B.T. (1990) J. Am. Chem. Soc.
[15] Loo, J.A. (1997) Mass Spectrom. Rev. 16, 1–23.
[16] Katta, V., Chowdhury, S.K. and Chait, B.T. (1991) J. Am. Chem. Soc. 112, 5348–5349.
[17] Mirza, U.A. and Chait, B.T. (1994) Anal. Chem. 66, 2898–2904.
[18] Frey, M.H., Wagner, G., Vasák, M., Sørensen, O.W., Neuhaus, D., Kägi, H.R., Ernst, R.R. and Wüthrich, K. (1985) J. Am. Chem. Soc. 107, 6847–6851.
[19] Evans, R. and Hollenberg, S. (1988) Cell 52, 1–3.
[20] Smith, R.D. and Light-Wahl, K.J. (1993) Biol. Mass Spectrom. 22, 493–501.
[21] Light-Wahl, K.J., Schwartz, B.L. and Smith, R.D. (1994) J. Am. Chem. Soc. 116, 5271–5278.
[22] Aplin, R.T., Baldwin, J.E., Fujishima, Y., Schofield, C.J., Green, B.N. and Jarvis, S.A. (1990) FEBS Lett. 264, 215–217.
[23] Craig, T.A., Veenstra, T.D., Naylor, S., Tomlinson, A.J., Johnson, K.L., Macura, S., Juranic, N. and Kumar, R. (1997) Biochemistry 36, 10482–10491.
[24] Veenstra, T.D., Johnson, K.L., Tomlinson, A.J., Naylor, S. and Kumar, R. (1997) Biochemistry 36, 3535–3542.
[25] Robinson, C.V. and Radford, S.E. (1995) Structure 3, 861–865.