*Protein Science* (1996), 5:1613–1624. Cambridge University Press. Printed in the USA.


# Preparation and properties of pure, full-length IclR protein of *Escherichia coli*. Use of time-of-flight mass spectrometry to investigate the problems encountered

LYNDA J. DONALD,[1] IGOR V. CHERNUSHEVICH,[2,3] JIE ZHOU,[2,4] ANATOLI VERENTCHIKOV,[2,5] NANCY POPPE-SCHRIEMER,[2] DAVID J. HOSFIELD,[1] JOHN B. WESTMORE,[1] WERNER ENS,[2] HARRY W. DUCKWORTH,[1] AND KENNETH G. STANDING[2]

[1] Department of Chemistry, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada
[2] Department of Physics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada
[3] Institute of Energy Problems of Chemical Physics, Chernogolovka, Russia

(RECEIVED February 5, 1996; ACCEPTED May 9, 1996)

**Abstract**

IclR protein, the repressor of the *ace*BAK operon of *Escherichia coli*, has been examined by time-of-flight mass spectrometry, with ionization by matrix assisted laser desorption or by electrospray. The purified protein was found to have a smaller mass than that predicted from the base sequence of the cloned *icl*R gene. Additional measurements were made on mixtures of peptides derived from IclR by treatment with trypsin and cyanogen bromide. They showed that the amino acid sequence is that predicted from the gene sequence, except that the protein has suffered truncation by removal of the N-terminal eight or, in some cases, nine amino acid residues. The peptide bond whose hydrolysis would remove eight residues is a typical target for the *E. coli* protease OmpT. We find that, by taking precautions to minimize OmpT proteolysis, or by eliminating it through mutation of the host strain, we can isolate full-length IclR protein (lacking only the N-terminal methionine residue). Full-length IclR is a much better DNA-binding protein than the truncated versions: it binds the *ace*BAK operator sequence 44-fold more tightly, presumably because of additional contacts that the N-terminal residues make with the DNA. Our experience thus demonstrates the advantages of using mass spectrometry to characterize newly purified proteins produced from cloned genes, especially where proteolysis or other covalent modification is a concern. This technique gives mass spectra from complex peptide mixtures that can be analyzed completely, without any fractionation of the mixtures, by reference to the amino acid sequence inferred from the base sequence of the cloned gene.

**Keywords:** DNA-binding protein; electrospray; glyoxylate cycle operon; IclR; matrix-assisted laser desorption/ionization (MALDI); time-of-flight mass spectrometry

Many microorganisms and plants can grow on fatty acid as a sole source of carbon because they possess a glyoxylate cycle, a variant of the Krebs tricarboxylic acid cycle that allows the acetyl carbons of acetyl-CoA to be converted into $C_4$ dicarboxylic acids and then to sugars, rather than to $CO_2$ (Kornberg, 1966). In the case of the Gram-negative bacterium *Escherichia coli* the enzymes of the glyoxylate cycle, isocitrate lyase, and malate synthase, are not present when the organism grows on glucose or other sugars, but are induced when it begins to grow on acetate (Kornberg, 1963). Work in several laboratories has led to a good understanding of several features of glyoxylate cycle control in *E. coli*. The structural genes for the two glyoxylate cycle enzymes, *ace*A and *ace*B, are organized into an operon located at 91 min in the *E. coli* linkage map (Brice & Kornberg, 1968; Vanderwinkel & DeVlieghere, 1968), along with a third structural gene, *ace*K, which specifies an enzyme whose role is to regulate isocitrate dehydrogenase by a well-characterized phosphorylation–dephosphorylation mechanism (LaPorte & Chung, 1985). The entire *ace*BAK region has been cloned and sequenced (Byrne et al., 1988; Cortay et al., 1988; Chung et al., 1988; Klumpp

Reprint requests to: Harry W. Duckworth, Department of Chemistry, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada; e-mail: hdckwth@cc.umanitoba.ca.

[4] Present address: Beckman Research Institute, City of Hope, Duarte, California 91010.

[5] Present address: PerSeptive Biosystems, Framingham, Massachusetts 01701.

*Abbreviations:* β-ME, β-mercaptoethanol; CNBr, cyanogen bromide; DTT, dithiothreitol; 4HCCA, α-cyano-4-hydroxycinnamic acid; ESI, electrospray ionization; IPTG, isopropyl-β-D-thiogalactoside; MALDI, matrix assisted laser desorption ionization; MS, mass spectrometry; SIMS, secondary ion mass spectrometry; TBE, tris-borate-EDTA; TOF, time-of-flight.

et al., 1988; Rieul et al., 1988), along with the nearby structural gene for an apparent repressor protein, *icl*R (Sunnarborg et al., 1990; Nègre et al., 1991). It has been demonstrated that IclR protein expressed from the cloned *icl*R gene interacts specifically with a small operator-like region of DNA in the promoter region of the *ace*BAK operon (Cortay et al., 1991; Nègre et al., 1991, 1992).

Although recent studies have provided a wealth of information about the regulation of the *ace*BAK operon, it is not known, at present, how the interaction of the IclR protein with its apparent operator DNA is related to the repression of the *ace*BAK operon during growth on glucose medium, and to its induction on acetate. In order to study further the role of the IclR protein in regulation of the glyoxylate cycle in *E. coli*, we have constructed an expression plasmid for the cloned *icl*R gene that is capable of directing the production of several tens of milligrams of IclR protein per liter of cell culture. When the purified IclR was examined by SDS-PAGE, the measured mass was about 29,000 u. A similar measurement by Cortay et al. (1991) yielded a mass of about 30,000 u. Although these numbers are consistent with the mass predicted from the IclR amino acid sequence as inferred from the sequence of the DNA, 29,739.3 u[6] (see Fig. 1), the estimated uncertainty in the measurements is about 1,000 u, so that the test they make of the integrity of the protein is not very stringent. The question therefore arises whether or not the protein has suffered any modification.

In principle mass spectrometry is capable of providing much more demanding constraints, but until recently it was very difficult to measure masses in the range of interest for intact proteins. The situation changed radically a few years ago, with the introduction of two new methods of ionization, matrix-assisted laser desorption and electrospray, and it is now reasonably straightforward to measure protein masses up to a few tens of thousands mass units with an accuracy of a few mass units (Chait & Kent, 1992).

In the following, we describe production of the IclR protein and measurements of its mass by both the above mass spectrometric methods, using time-of-flight mass spectrometers. In fact, a major modification of the protein was found. Independent mass measurements by both methods not only indicated this, but also paved the way to a method of producing the intact protein, whose DNA-binding properties are significantly different from those of the modified protein isolated originally.

## Results and discussion

### *Measurements on the original IclR preparation*

Biorex 70 chromatography (see Materials and methods) resolved two major protein peaks, designated A and B; both were largely IclR protein, as judged by SDS-PAGE, so the reason for the separation of the two peaks was unknown. Subsequent size exclusion chromatography of these fractions in Sepharose 6B or

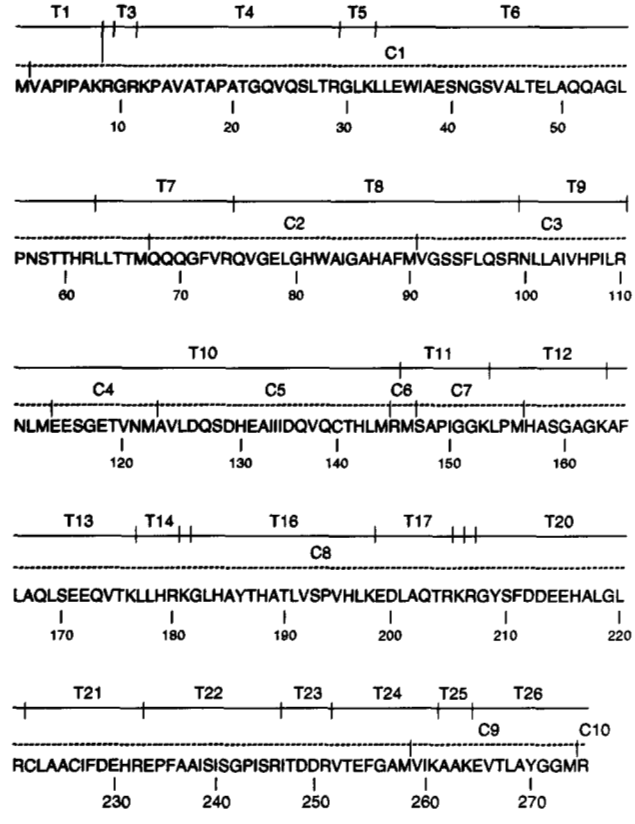


**Fig. 1.** Amino acid sequence of IclR protein, from the sequence of the cloned structural gene. The predicted products of digestion by CNBr and trypsin are indicated as C1, C2... and T1, T2..., respectively.

Sephadex G100 showed that the IclR protein of each peak behaved as a monomer with a mass of approximately 30,000 u. SDS-PAGE of these preparations showed the same size of protein in each, with a mass slightly smaller than that of carbonic anhydrase II (29,023 u). Size determination by MALDI/TOF-MS indicated that peak A, which eluted from the Biorex 70 column at about 0.55 M KCl, had a mass of 28,930 ± 10, and peak B, which eluted at 0.65 M KCl, had essentially the same mass, 28,932 ± 10 (Zhou, 1994; see Fig. 2). The open reading frame of the IclR gene (Sunnarborg et al., 1990; Nègre et al., 1991) predicts a protein with a mass of 29,739.3 u, or 29,608.1 if the initial methionine has been removed (see Fig. 1), so it was clear that the protein we had isolated was truncated in some way. In addition to the major protein, both fractions A and B contained other prominent peaks, of 9,224 ± 3 u and 9,534 ± 3 u (Zhou, 1994), suggesting the presence of two other proteins. These relatively small proteins are present on SDS gels, but they stain poorly with Coomassie blue and were not noticed initially.

Further information was obtained from automated Edman degradation of the purified protein from peak A of the same preparation (S. Kielland, University of Victoria), which showed the presence of at least two sequences. One was the amino acid sequence of IclR protein as predicted from the sequence of the gene, but commencing at amino acid Gly-10. The other was identified by protein databank searches as the N-terminal sequence of the *E. coli* DNA-binding protein HU-2. Closer examination of the sequencing data indicated that the highly

[6] Mass values quoted in this paper are calculated from chemical atomic masses; these values are appropriate for comparison with masses measured from MALDI mass spectra, for which the isotopic components of the envelope of the ionic peaks are not resolved. Note, however, that ESI/TOF mass spectrometry is capable of higher resolution, so that isotopic components can often be resolved; for these components, masses calculated from physical atomic masses are more suitable.

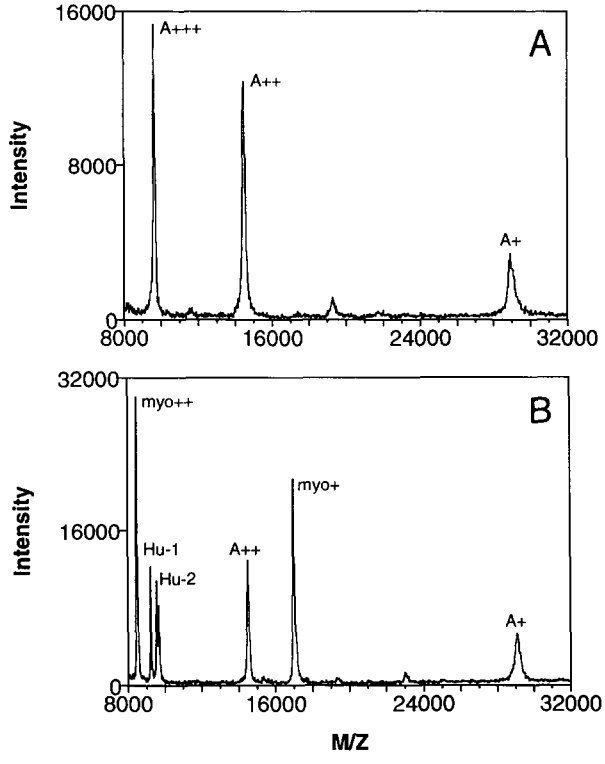


**Fig. 2. A:** MALDI mass spectrum of truncated IclR protein sample from peak B. Sample from peak A, and the full-length IclR protein examined later, gave similar results. **B:** MALDI mass spectrum of IclR protein plus myoglobin calibrant. Measurements were made with a nitrogen laser (337 nm), irradiance near threshold, and 4HCCA matrix.

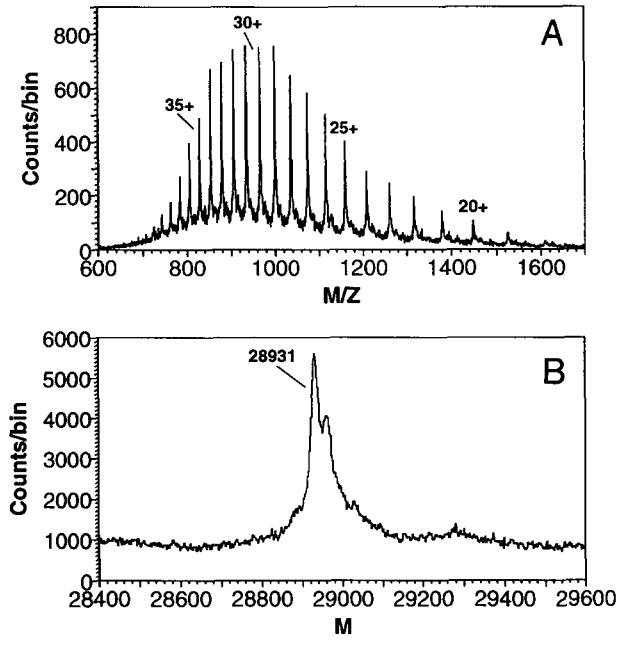


**Fig. 3. A:** ESI/TOF mass spectrum of truncated IclR protein (peak B). **B:** Deconvoluted spectrum obtained from the data of Figure 3A by a procedure similar to the one described by Mann et al. (1989).

homologous HU-1 and HU-2 sequences (Mende et al., 1978; Laine et al., 1980) were both present. The masses of HU-1 and HU-2 are 9,225.6 and 9,535.0 u, respectively, in good agreement with the masses of the extra proteins found by MALDI/TOF-MS. Thus, what we had believed to be a "pure" preparation of IclR protein was actually a truncated version, and, moreover, it was contaminated with large amounts of two nonsequence-specific DNA-binding proteins. Both difficulties could seriously compromise DNA binding studies.

The calculated mass of IclR less the nine N-terminal residues is 28,775.2 u, which disagrees with the measured value, but truncation by removal of only *eight* N-terminal residues would yield a mass of 28,931.4 u, in good agreement with observation. Further experience has shown that most IclR preparations display a mass consistent with the loss of eight amino acids from the N-terminus, but that the loss of nine is occasionally seen. At this stage, we tried to measure the IclR mass on our ESI/TOF instrument (Verentchikov et al., 1994) in order to obtain increased precision. The initial attempt was unsuccessful, owing to a very large background from the low mass impurities in the sample. Fortunately, the positive result in the MALDI measurement provided sufficient incentive to perform additional purification of the sample (see Materials and methods). Our second attempt was then successful (see Fig. 3), giving a result of 28,931 u, again in agreement with the above calculations and with the MALDI result, but now with an estimated error of only ±2 u.

Thus, the mass measurements suggest that the only difference between IclR protein as we isolated it and the predicted, full-length protein is the loss of eight amino acids from the N-terminus. To check this hypothesis, we prepared two sets of peptides, one by cleavage of IclR with CNBr, the other by digestion with trypsin. The resulting mixtures were analyzed directly, without fractionation, by MALDI for the CNBr peptides, and by both MALDI and ESI/TOF for the tryptic peptides. Tables 1 and 2 show the peptides obtained, together with the masses calculated from the inferred amino acid sequence of Figure 1. No peptide corresponding in mass to intact CNBr peptide C1 (predicted mass 6,973 u, or 6,843 if the N-terminal methionine is missing) was found, but another peptide, of mass about 6,164 u, did appear, corresponding to peptide C1 less the N-terminal eight amino acids (calculated, 6,165.1 u).

Although this result gives some support to the suggested deletion at the N-terminus, the CNBr data are somewhat unsatisfactory. In particular, the differences between calculated and experimental values of mass range up to ~6 u, no data are obtained for the large region corresponding to C8 (residues 157–258, see Fig. 1), and there are many unassigned peaks in the spectrum.

The spectra of the tryptic digest are much more definitive. The MALDI spectrum is shown in Figure 4. The ESI spectrum is not illustrated, but its quality is very similar to the one shown in Figure 5 for a different sample, discussed later. Predicted peptide T1, corresponding to the first eight amino acids of the intact IclR protein, is not found, but the remainder of the predicted protein sequence is all accounted for. The calculated and observed masses agree within 0.5 u (better than 0.02%) apart from T2+T3 in ESI (probably misassigned), T13 in MALDI (a very weak line), and T21+T22 in both techniques (see Tables 1 and 2). The last discrepancy is interesting because two of the three cysteine residues lie within this range; if these two were joined by a disulfide bond (either in the native protein, or perhaps more likely

**Table 1.** *Comparison of experimental molar masses of observed products, determined by MALDI/TOF mass spectrometry, with calculated molar masses of expected products of CNBr cleavage of IclR protein*[a]

| Observed M[b] | | | Amino acid | | |
|---|---|---|---|---|---|
| 2 h[e] | 24 h[e] | Assignment[c] | residues | Calc $M$[d] | $M_{calc} - M_{obs}$ |
| – | – | C1 | 1-67 | 6,973.2 | |
| – | – | C1(2) | 2-67 | 6,842.0 | |
| 6,161.0 m[f] | 6,164.0 s | C1(9) | 9-67 | 6,165.1 | 4.1, 1.1 |
| 2,520.0 s | 2,518.0 s | C2 | 68-90 | 2,519.9 | −0.1, 1.9 |
| 2,530.2 s | 2,529.0 s | C3 | 91-113 | 2,531.1 | 0.9, 2.1 |
| 11,276.9 s | | C1+C2+C3 | 9-113 | 11,276.1 | −0.8 |
| 11,277.6[g] s | | C1+C2+C3 | 9-113 | 11,276.1 | −1.5 |
| 3,507.0 w | 3,509.0 w | C3+C4 | 91-122 | 3,508.1 | 1.1, −0.9 |
| 3,700.8 w | | C4+C5+C6 | 114-146 | 3,696.2 | −6.4 |
| 1,209.0 m | | C6+C7 | 145-156 | 1,209.6 | 0.6 |
| 1,502.7 s | | C9 | 259-273 | 1,502.9 | 0.2 |

[a] Unassigned peaks observed: $m/z$ = 12,230, 3,522, 3,480, 3,465, 3,087, 2,980, 2,501, and 1178.

[b] It is assumed that the species observed in the MALDI spectrum is the $[M + H]^+$ ion.

[c] As discussed in the text, the N-terminal methionine is frequently missing from *E. coli* proteins, so the calculated mass of C1 is given both with and without that residue, the latter indicated as C1(2). The corresponding fragment with eight residues deleted from the N terminus is denoted as C1(9).

[d] Mass values are calculated from chemical atomic masses; these values are appropriate for comparison with masses measured from MALDI mass spectra, for which the isotopic components of the envelopes of the ionic peaks are not resolved. In this column, calculated masses are obtained by subtracting 30.08 from the chemical atomic mass of the assigned amino acid sequence (see Fig. 1 caption).

[e] Reaction time.

[f] s, strong; m, medium; w, weak.

[g] Determined from $m/z$ 5639.8, which is assumed to be $[M + 2H]^{2+}$.

during the purification procedure or preparation of the tryptic peptides), the calculated mass of T21+T22 would be 2,702.1 u, in good agreement with observation. Thus, the tryptic digest results strongly support the protein structure suggested above (i.e., the structure predicted by the *iclR* gene sequence) but with the deletion of eight residues from the N-terminus.

Why truncated IclR protein elutes from Biorex 70 in two peaks remains unknown. The MS measurements show that the same polypeptide is present in both fractions; other measurements show that the two peaks have the same apparent molecular size, and both bind DNA equally (see below). Because truncated IclR protein is poorly soluble in low ionic strength buffer, it is possible that the first peak represents protein that remains in solution throughout chromatography, whereas the second is simply protein that actually precipitates on the Biorex 70 column, but redissolves as the salt gradient reaches a high enough level.

### *Preventing truncation of IclR protein*

The peptide bond that is cleaved in producing N-terminal removal of eight amino acids of IclR, $Lys_8$-$Arg_9$, is a typical target of the *E. coli* membrane protease OmpT, which is known to act on a number of recombinant proteins produced by expression in the bacterium (Sugimura & Nishihara, 1988). It is also known that OmpT protease is released from membranes by EDTA, and we had included 1 mM EDTA in all purification buffers. Thus, OmpT is a prime suspect as the cause of the deletion. The IclR samples with nine amino acids removed, which are less frequently observed, could arise either by a tryptic-like cleavage of the $Arg_9$-$Gly_{10}$ bond or by an aminopeptidase activity on the presumed OmpT product.

OmpT is inhibited by several divalent cations, including $Zn^{2+}$ (Sugimura & Nishihara, 1988). To test the assumption that OmpT is the culprit in our case, we decided to change the purification buffer by eliminating EDTA and including 1 mM $ZnSO_4$. The rest of the purification was as before. The IclR protein now eluted from Biorex 70 as a single peak, at about 0.35 M KCl, significantly earlier in the salt gradient than before.

When this protein sample was examined by MALDI and ESI/TOF mass spectrometry, a single peak of mass 29,608 u was found. This corresponds to the predicted mass of full-length IclR protein, less the initiation methionine that is usually missing from *E. coli* proteins. This preparation also showed the expected

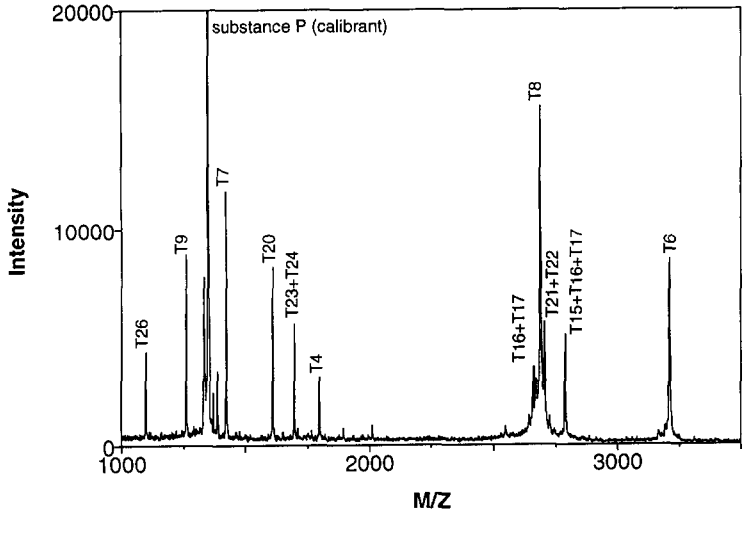


**Fig. 4.** Partial MALDI mass spectrum of a tryptic digest of the truncated IclR protein (peak A), with internal calibrants substance P and bovine insulin. Nitrogen laser, irradiance near threshold, 4HCCA matrix.

**Table 2.** *Comparison between calculated and observed molar masses of expected products of tryptic digestion of IclR protein*[a]

| Assignment | Amino acid residues | Calc $M$[b] | MALDI | | ESI | | |
|---|---|---|---|---|---|---|---|
| | | | Obs $M$[c] | $M_{calc} - M_{obs}$ | Charge | Obs $M$[d] | $M_{calc} - M_{obs}$ |
| T1(1) | 1–8 | 826.1 | | | | | |
| T1(2) | 2–8 | 694.9 | | | | | |
| T2+T3 | 9–11 | 387.6 | 387.4 s[e] | 0.2 | 1+ | 390.6[f] | −3.0 |
| T3+T4 | 10–29 | 2,009.3 | 2,009.8 w | −0.5 | | | |
| T4 | 12–29 | 1,796.1 | 1,796.1 s | 0 | 3+ | 1,796.2 | −0.1 |
| T5 | 30–32 | 316.4 | | | 1+ | 316.3 | 0.1 |
| T6 | 33–62 | 3,207.6 | 3,208.0 s | −0.4 | 3+,4+ | 3,207.9 | −0.3 |
| T7 | 63–74 | 1,421.7 | 1,421.5 s | 0.2 | 2+,3+ | 1,421.9 | −0.2 |
| T8 | 75–99 | 2,686.1 | 2,686.2 s | −0.1 | 3+,4+ | 2,686.0 | 0.1 |
| T9 | 100–110 | 1,258.6 | 1,258.5 s | 0.1 | 2+,3+ | 1,259.0 | −0.4 |
| T10 | 111–145 | 3,971.5 | 3,972.0 w | −0.5 | 3+,4+ | 3,971.9 | −0.4 |
| T11+T12 | 146–163 | 1,710.1 | | | 3+,4+ | 1,710.4 | −0.3 |
| T13 | 164–176 | 1,463.7 | 1,461.8 w | 1.9 | 2+ | 1,464.0 | −0.3 |
| T14 | 177–180 | 537.7 | 537.5 s | 0.2 | 1+ | 537.5 | 0.2 |
| T14+T15 | 177–181 | 665.9 | 665.4 w | 0.5 | 1+ | 666.1 | −0.2 |
| T15+T16 | 181–198 | 1,972.3 | 1,972.5 w | −0.2 | 3+,4+ | 1,972.6 | −0.3 |
| T15+T16+T17 | 181–205 | 2,786.2 | 2,786.5 s | −0.3 | 4+ | 2,786.3 | −0.1 |
| T16+T17 | 182–205 | 2,658.0 | 2,658.3 m | −0.3 | 4+ | 2,658.2 | −0.2 |
| T17 | 199–205 | 831.9 | 831.9 s | 0 | 2+ | 832.4 | −0.5 |
| T18+T19+T20 | 206–221 | 1,893.1 | 1,893.5 w | −0.4 | 4+ | 1,893.2 | −0.1 |
| T19+T20 | 207–221 | 1,764.9 | 1,765.0 w | −0.1 | 3+ | 1,764.8 | 0.1 |
| T20 | 208–221 | 1,608.7 | 1,608.7 s | 0 | 2+,3+ | 1,608.7 | 0 |
| T21+T22 | 222–246 | 2,704.1 | 2,702.4 s | 1.7 | 3+,4+ | 2,702.5 | 1.6 |
| T23+T24 | 247–261 | 1,695.0 | 1,694.8 s | 0.2 | 3+ | 1,695.1 | −0.1 |
| T25 | 262–264 | 288.4 | | | 1+ | 288.3 | 0.1 |
| T26 | 265–274 | 1,096.3 | 1,096.0 | 0.3 | 2+ | 1,096.5 | −0.2 |

[a] Unassigned peaks observed in MALDI spectrum: $m/z$ = 2,725, 2,670, 2,643, 2,548, and 1,332.

[b] Calculated molar masses are obtained by adding 18.0 to the chemical atomic mass of the assigned amino acid sequence (see Fig. 1 caption).

[c] It is assumed that the species observed in the MALDI spectrum is the $[M+H]^+$ ion.

[d] The isotopic peaks of the $[M + nH]^{n+}$ ions in the ESI spectra are separated so the physical atomic mass could, in principle, be determined. However, the values reported here correspond to the centroids of the isotopic distributions so as to give chemical atomic masses, as obtained from the MALDI measurements.

[e] s, strong; m, medium; w, weak.

[f] Assignment is doubtful; the error is ~0.75% compared to a typical error of ~0.01%.

N-terminal amino acid sequence, commencing at Val-2, when submitted to Edman degradation. Thus, the strategy for eliminating protease activity was successful. In addition to this, the sample no longer showed peaks corresponding to HU-1 and HU-2 proteins, or other macromolecular contaminants, presumably because the full-length IclR protein eluted from Biorex 70 at a lower salt concentration than did the HU proteins.

To verify the fact that the protein was full length, with no covalent modifications, we again prepared CNBr and tryptic peptides, which were analyzed by both MALDI/TOF and ESI/TOF-MS. The CNBr results (not shown) were again not completely satisfactory, although peptides starting at residue 2 were now clearly present. In the case of the tryptic digest, all expected peptides were located easily in both the MALDI and the ESI spectrum. Figure 5A shows the overall ESI/TOF spectrum of the tryptic digest. Features of this spectrum are illustrated in the expanded Figure 5B, C, and D, in particular, the determination of the the charge state from the isotopic spacing; note the difference in spacing between the 2+ and 4+ states. Figure 5B shows that $T1^{2+}$ is clearly separated from $(T11+T12)^{4+}$, so it is possible to establish its presence in the full-length IclR (and its absence in the previous measurement of the truncated sample). Fragments T24 and T26 differ only by 2 u in mass but they are seen to be clearly resolved in Figure 5D. In the measurement of the digest of the truncated protein, T24 was only seen in the combination (T23+T24), presumably because the digestion had not proceeded far enough to cleave the T23–T24 bond.

It is interesting to compare the MALDI and ESI spectra of the tryptic digests (Figs. 4, 5). The MALDI spectrum is much simpler and easier to interpret, thanks to the presence of only singly charged ions. Although the ESI spectrum has better resolution, the presence of multiple charge states (up to +6) makes it much more complicated. Fortunately the resolution of the ESI/TOF instrument is sufficient to determine the charge state of the ion in almost every case (see Fig. 5B,C,D), so the spectrum is effectively decomposed into a number of simple spectra, which can then be interpreted. Two conclusions can be drawn from this experience: first, there is a clear advantage in

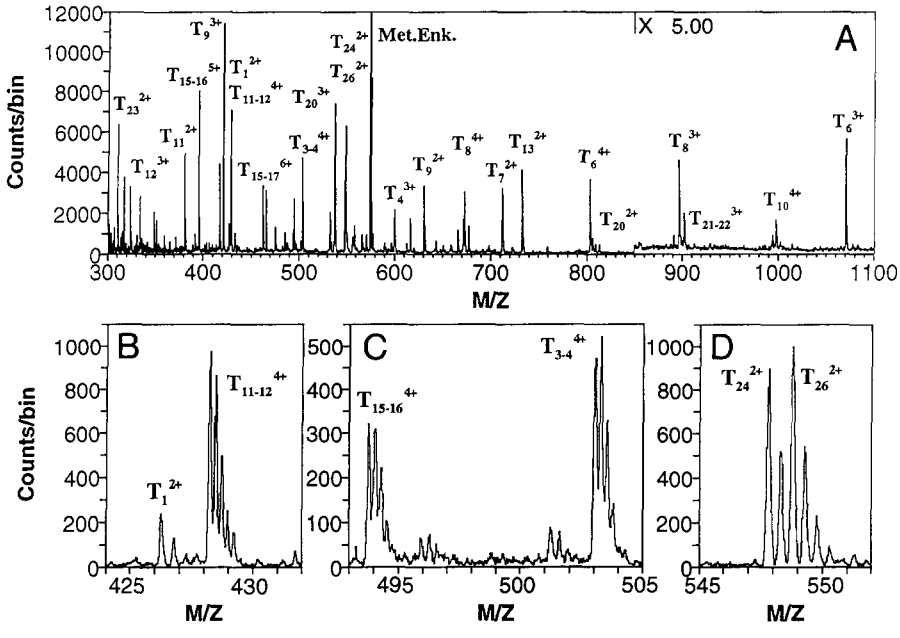


**Fig. 5. A:** ESI/TOF mass spectrum of a tryptic digest of the full-length IclR protein. **B,C,D:** Expanded sections of the spectrum of A. The truncated protein gave a similar spectrum, but with T1 absent.

having *both* MALDI and ESI available for such measurements; second, it is important to be able to resolve the isotopic distributions in order to interpret the ESI measurement.

### *Use of an ompT⁻ host to minimize proteolytic degradation*

There is a tendency, over a period of weeks, for purified, concentrated IclR to convert to a mixture of fragments of mass of approximately 18,000 u; the exact sites of proteolytic cleavage to produce these fragments have not yet been established. Because we had not observed production of these fragments in the case of our early, truncated IclR preparations, we guessed that eliminating EDTA from the purification buffers in order to prevent OmpT activation was now permitting activity of another *E. coli* protease that carries through the purification procedure in small amounts. Addition of 1 mM EDTA to the buffer used in the second purification step, the size-exclusion column, greatly reduced formation of the fragments near 18,000 u, but ESI/TOF-MS of the resulting IclR preparations showed some reappearance of truncated protein with the N-terminal eight or nine amino acids missing. We therefore decided to express our IclR clone in *E. coli* host strains carrying mutations in the *ompT* gene.

The first strain used, the *E. coli* B strain WA834 (Wood, 1966), gave IclR protein that was heavily degraded to the fragments of mass about 18,000 u (even when 1 mM EDTA had been included throughout the extraction and purification procedures), suggesting that, in WA834, one or more other proteolytic activities were more strongly expressed than in the host strain previously used, JM103. The second *ompT*⁻ strain used, the *E. coli* K strain AD202, could not be transformed with plasmid DNA by methods we used. We were able to take advantage, however, of the fact that the mutation in AD202 was made by inserting a kanamycin resistance transposon into the *ompT* gene (Akiyama & Ito, 1990). Phage P1 transduction, with selection of transductants on kanamycin, was used to transfer the *ompT*⁻ allele from AD202 into our original host, JM103. The resulting strain, LD103, was transformed readily with the IclR expression plasmid pKKICL8. IclR protein produced from this transformant, after extraction with Tris buffer containing 1 mM EDTA but no $ZnSO_4$, was full-length as demonstrated by ESI/TOF mass spectrometry, provided that the cells were harvested within 2 or 3 h of inducing IclR expression (data not shown). However, if the cells were allowed to grow overnight before harvesting, about 80% of the IclR protein was the −8 truncated form, and the rest was the −9 form. Thus, it seems that OmpT proteolysis is only part of the problem and that another unidentified host protease, perhaps an aminopeptidase, will still truncate IclR over long growth times. Our preferred method for preparing IclR, in light of all the above experience, is to use LD103 as the host strain, harvest the cells no more than 3 h after induction with IPTG, and perform the extraction and purification with Tris-EDTA buffer.

At all stages of refinement of the procedure for preparing IclR protein, as we sought to avoid the different kinds of proteolysis that the protein can suffer, the ability to monitor the state of the protein by TOF-MS has been crucial.

### *Stability of purified IclR protein*

IclR protein is not very soluble and concentrated solutions will throw down precipitate during storage. In Tris-Cl, pH 8.0, 0.5 M KCl, it is not possible to concentrate IclR solutions above about 4 mg/mL; when the salt concentration is reduced, the solubility is also much less. When stored, the stability of purified IclR protein, though not the solubility limit, is greatly improved if 50 mM β-ME is included in all buffers, including those used in the purification. This additive is removed by dialysis for long-term storage and for mass spectrometry measurements; after the addition of DTT to a final concentration of 1 mM, the dialyzed samples can be stored frozen at −20 °C in the presence of 30% glycerol.

### *Spectroscopic properties of IclR protein and native molecular weight*

The molar absorptivity of pure IclR protein (truncated form, $M$ = 28,931.4) was measured by diluting a portion of an IclR stock solution, of known $A_{280}$ and to which no β-ME had been added, into 4 volumes of 7.5 M guanidine HCl. The tyrosine concentration of this dilution was measured by the alkaline difference spectrum method of Edelhoch (1967), and this value was converted to IclR concentration by dividing by three, the number of tyrosines in one IclR polypeptide. The molar absorptivity of native IclR protein, 15,040, is consistent with the IclR content of two tryptophan and three tyrosine residues. The molar absorptivity of the full-length protein is identical, which is expected because there are no aromatic amino acids among the N-terminal nine residues (Fig. 1).

The CD spectrum of IclR protein, obtained on a sample of full-length protein that had been dialyzed thoroughly to remove β-ME, is shown in Figure 6; essentially the same result was obtained with the (−8) truncated form. Fitting of the spectrum by the program of Perczel et al. (1992) leads to an estimate of 22–25% α-helix and 15–20% β-sheet for both proteins.

As already stated, in the absence of β-ME, freshly prepared IclR protein emerges from a Sephadex G-200 column (20 mM Tris-Cl, pH 7.8, 0.4 M KCl) at the same place as carbonic anhydrase ($M_r$ = 29,023), indicating that it behaves as a monomer; however, recovery from the column is low and older preparations fail to pass through the column at all, probably because they are greatly aggregated. When β-ME is included during preparation and storage of the protein, less equivocal results are obtained. Figure 7 shows that, under these conditions, IclR (0.2–1 mg/mL) elutes from Sephadex G-100 (20 mM Tris-Cl, pH 7.8, 0.4 M KCl, 1 mM EDTA, 50 mM β-ME) as a tetramer ($M_r$ ~ 117,000). Nègre et al. (1991) found that their IclR preparation sedimented in a sucrose density gradient with a sedimentation coefficient consistent with a dimer ($M_r$ ~ 60,000). Together these several results show that freshly prepared IclR is a tetramer at micromolar concentrations, but that it has a tendency to dissociate to dimers or even monomers when stored. It seems possible, even with fresh preparations, that at the low concentrations used for DNA binding studies ($10^{-8}$ M and lower), IclR protein could be largely in the monomeric state. We have not devised conditions to test this possibility, but it is still of interest in light of our finding that DNA binding is cooperative in the low concentration range (see below).

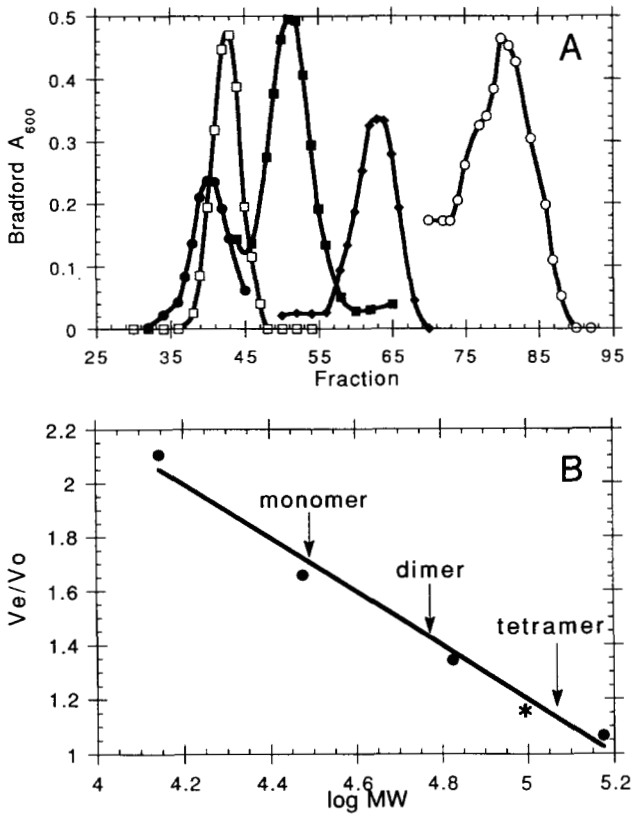


**Fig. 7.** Molecular weight determination of IclR protein by gel filtration through Sephadex G-100. **A:** Elution profiles of IclR protein and standards: ■, bovine serum albumin; ◆, carbonic anhydrase; ○, hen egg white lysozyme; ●, liver alcohol dehydrogenase; □, IclR protein. **B:** Plot showing normalized elution volumes relative to void volume, plotted against logarithm of molecular mass. Arrows show the expected positions for IclR species of molecular mass 30,000 u (monomer), 60,000 (dimer) and 120,000 (tetramer); star indicates the elution position of IclR, suggesting that the native molecular mass is that of a tetramer.

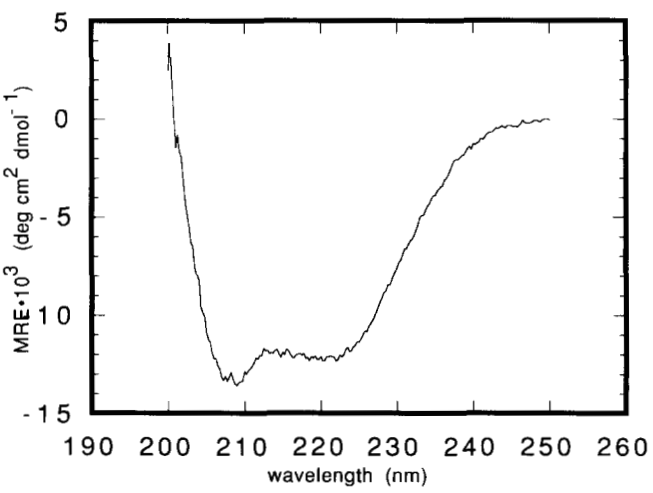


**Fig. 6.** CD spectrum of full-length IclR protein. The spectrum was collected on a 0.31 mg/mL protein sample dissolved in 20 mM Tris-Cl, pH 7.8, containing 0.4 M NaCl. An identical spectrum was obtained for (−8) truncated IclR.

### *Binding of aceBAK operator DNA by IclR protein*

We have tested DNA binding by full-length and truncated IclR protein in gel-shift experiments with the 227-bp *Eco*R I–*Hin*d III restriction fragment containing the natural operator region of the *ace*BAK operon. In Figure 8 are shown three binding curves, two obtained with truncated IclR proteins (lacking the N-terminal eight and nine amino acids, respectively) and the third with full-length IclR protein. The state of truncation of each protein sample was verified by mass spectrometry. The concentration of full-length IclR protein needed to get half-maximal binding is (0.36 ± 0.04) nM. The curves obtained with both truncated proteins lie significantly to the right, and are the same as

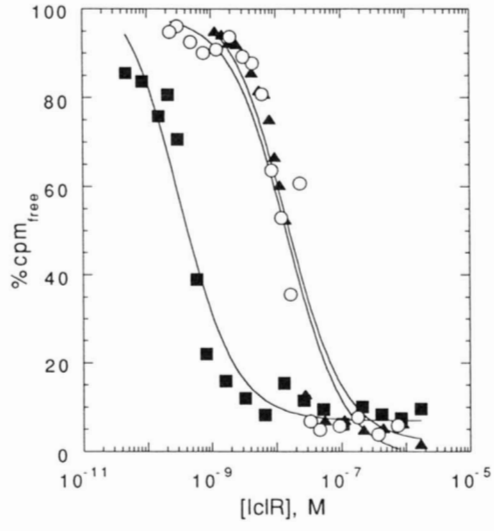


**Fig. 8.** Quantitation of Icl-DNA binding. Freshly diluted IclR protein that was either full-length (■) or truncated at the N-terminus by eight (▲) or nine residues (○), was incubated with a 227-bp *Eco*R I–*Hin*d III DNA fragment containing the *ace*BAK operator region (see Materials and methods). Samples were subjected to PAGE, and radioactive bands were excised and counted by scintillation counting. The concentration of IclR protein needed for half-maximal binding, in the case of the full-length protein, was (0.36 ± 0.04) nM, whereas that for both truncated proteins was (14 ± 0.6) nM. Curves represent nonlinear least-squares fits of the data sets to the Hill (1910) equation; Hill parameters obtained from these fits were 1.7 ± 0.2 (full length), 2.1 ± 0.2 (−8 truncated), and 2.2 ± 0.2 (−9 truncated protein).

one another within the probable errors, with a half-maximal protein concentration of (14 ± 0.6) nM. Nègre et al. (1992) reported a $K_d$ for binding by their IclR preparation of 0.7 nM, about twice our value for full-length protein.

A remarkable feature of all our DNA binding curves is the steepness of the concentration dependence (Fig. 8), indicating that IclR binding to DNA is cooperative, with polymerization of IclR subunits as they bind to the DNA, at least in the low protein concentration range in which the binding measurements were performed. When the data are fitted to the Hill (1910) equation, the Hill coefficients (an empirical measure of cooperativity) for full-length protein and for protein samples truncated by eight and nine residues, are found to be 1.7 ± 0.2, 2.1 ± 0.2, and 2.2 ± 0.2—that is, the cooperativity is greater with the truncated proteins. This cooperative shape is also seen in the curve presented by Nègre et al. (1992), though those authors did not comment on it. The finding of cooperativity is consistent with the evidence for an equilibrium between IclR subunits, dimers, and tetramers in the much higher concentration range used for molecular weight measurements on the protein alone (see above); it is explained most easily by assuming that more than one subunit binds to the DNA target sequence, so that formation of the DNA–protein complex requires assembly of a dimer if not a tetramer. Because the full-length protein shows a lower Hill coefficient than the truncated forms do, it is possible that removal of the N-terminal eight or nine residues significantly increases the degree of dissociation of the unbound protein.

We attempted to measure the molecular size of the IclR-DNA complex by the PAGE method of Hedrick and Smith (1968) as modified for DNA–protein complexes by Orchard and May (1993). The labeled operator DNA used in these experiments was a chemically synthesized 29-bp duplex (molecular weight 17,791 u). The change in the mobility of the protein-bound DNA band on polyacrylamide gels, as the gel concentration was changed, matched the behavior of the *E. coli* citrate synthase hexamer [molecular weight 282,000 u, (Tong & Duckworth, 1975)], although the plot of mobility change versus molecular weight has a shallow slope in this range, so that the precision is low. Assuming only one DNA duplex per complex, the number of IclR subunits present in the complex would be about eight; with so many subunits present, it is quite possible that more than one DNA duplex is bound. Although this experiment failed to give a definite stoichiometry for the IclR–DNA complex, it was clear that IclR in the complex is at a high degree of polymerization.

DNase footprinting (Fig. 9) defines on the two strands a region of 27 (on the *Eco*R I–*Hin*d III strand) or 29 bp (on the *Hin*d III–*Eco*R I strand) that is protected by IclR. This region is smaller than the footprint reported by Cortay et al. (1991) (35 bases protected, on one strand only), but it is consistent with the same group's G-methylation interference and depurination/depyrimidination results (Nègre et al., 1992). The protected sequence is an approximate palindrome, very rich in AT pairs (Fig. 9). The same protection pattern has been obtained with (−8) truncated IclR as with the full-length protein. On grounds of amino acid sequence homology, it has been suggested that IclR protein is an example of the most common type of sequence-

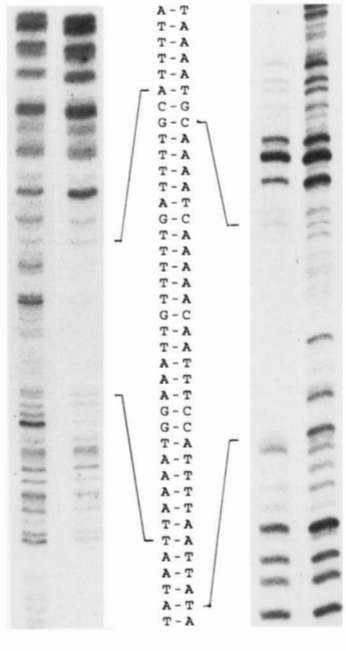


**Fig. 9.** DNase footprinting of both strands of the *Eco*R I–*Hin*d III fragment containing the *ace*BAK operator sequence (see Materials and methods). The two outside lanes are the DNase fragmentation patterns obtained in the absence of IclR protein, and the lanes immediately inside are the same patterns in the presence of saturating IclR protein. DNA sequences of the two complementary strands are given in the center of the figure; the *Eco*R I site is toward the bottom. For the left strand, the 5′ end and the shorter DNase fragments are toward the top; for the right strand, the 5′ end and the shorter fragments are toward the bottom.

specific DNA-binding protein of bacteria, those containing the helix-turn-helix motif (Sunnarborg et al., 1990; Nègre et al., 1991). If this is so, two IclR monomers would bind, one to each half of the operator, but 27–29 bp is too long a stretch of DNA for two 29,000 u monomers to protect from DNase unless the DNA were significantly bent around them, as is the case, for instance, with the dimeric helix-turn-helix cyclic AMP-binding protein (Schultz et al., 1991). Such bending, if it occurs, would be facilitated by the high AT content of the operator region.

In this paper we have shown that IclR protein is susceptible to various kinds of proteolytic degradation during purification and have established a method of preparing useful amounts of the intact, full-length molecule. The full-length preparations bind to operator DNA with only moderate affinity, showing a half-maximal binding at $3.6 \times 10^{-10}$ M protein, in comparison with other well-characterized repressors where the $K_d$ values fall in the range $10^{-10}$–$10^{-12}$ M (e.g., Vershon et al., 1987; Carey, 1988; Brenowitz et al., 1990). IclR binding of operator is 44-fold weaker when the N-terminal eight or nine amino acids are missing, suggesting that this region contributes significantly to DNA binding. There are several well-documented examples of N-terminal regions of sequence-specific DNA-binding proteins that play a role in DNA binding, such as lambda repressor (Pabo et al., 1982; Jordan & Pabo, 1988), Mnt repressor (Knight & Sauer, 1989), MAT$\alpha$2 homeodomain (Wolberger et al., 1991), and Hin recombinase (Feng et al., 1994).

A novel feature of our results is that complex formation by IclR is cooperative, occurring over a narrow concentration range of the protein, so that slight changes in IclR concentration in vivo, in the right range, could have a dramatic effect on its ability to repress the *ace*BAK operon.

The studies reported here may be of wider interest because of the extensive use we made of the powerful MALDI/TOF and ESI/TOF instruments—first in discovering that our original preparations of IclR were truncated, then in testing altered preparation methods that would avoid this, and finally in confirming that we had succeeded. It was especially gratifying that the mass spectra of the peptide mixtures could be completely interpreted without any fractionation of the mixtures before mass spectrometry. As a result of our experience, we are convinced that macromolecular mass spectrometry is a tool that should now be available to every serious protein laboratory.

## Materials and methods

### *Preparation of an expression clone for IclR protein*

Previous investigators have shown that the *ace*BAK region and the *icl*R structural gene are located within the same 10-kbp *Cla*I restriction fragment of *E. coli* genomic DNA (El-Mansi et al., 1987). We identified a clone containing this *Cla*I fragment, designated λGLY1, by probing an *E. coli* λ genomic library with two hybridization probes, generated by amplification of portions of the *ace*B and *icl*R regions. Primers were designed by reference to the known sequences of these genes (Byrne et al., 1988; Sunnarborg et al., 1990), and the sequences were amplified using Taq DNA polymerase according to Saiki et al. (1988). The smaller of the two *Cla*I–*Hin*d III fragments derived from the λGLY1 clone, which contains all of the coding region for the IclR protein, including its natural promoter, was subcloned into pBR322, giving plasmid pCHicl. This is the same as plasmid pKL5 used by Sunnarborg et al. (1990). Expression of IclR protein from this clone was too low to detect, and so we used site-directed mutagenesis (Kunkel et al., 1987) to introduce a new *Eco*R I site upstream of the initiation codon; this required changing base 224 from T to A (numbering as in Sunnarborg et al., 1990). The *Eco*R I–*Hin*d III fragment was then recloned into the vector pKK223-3 (Pharmacia), resulting in a recombinant plasmid, pKKICL8, with the entire *icl*R coding region and 30 bp of the 5′ region including a ribosome binding site and initiation codon, all under the control of the ptac promoter. All the mutations and constructions were confirmed by DNA sequencing (Sanger et al., 1977).

*E. coli* strains AD202 and WA834, both containing *omp*T$^-$ mutations, were obtained from the *E. coli* Genetics Stock Center, Yale University. In strain WA834, the *omp*T allele is believed to be a missense mutation; in AD202, the *omp*T gene is disrupted by a kanamycin resistance transposon (Akiyama & Ito, 1990). Phage P1-mediated transduction, to transfer the *omp*T$^-$ allele from strain AD202 to the JM103 background, was performed according to Miller (1972). The transductants were selected on medium containing 25 μg/mL kanamycin.

Routine molecular biology procedures followed the protocols of Maniatis et al. (1982).

### *Production of IclR protein*

Cells of the various *E. coli* strains used, containing plasmid pKKICL8, were grown in LB broth containing 100 μg/mL ampicillin to early log phase, at which time IclR production was induced by adding IPTG to 0.5 mM. Cells were allowed to grow for a further 2–3 h, or in some cases, for 18 h, with vigorous shaking, at 37 °C. Harvesting was done by centrifugation for 20 min at 5,000 × *g* and 4 °C, after which the wet pellet was weighed and suspended in three volumes of buffer. Initially the buffer used was 20 mM Tris-Cl, 1 mM EDTA, 0.2 M KCl, pH 7.8. To control OmpT protease (see Results and discussion), we used 20 mM Tris-Cl, 0.2 M KCl, 1 mM $ZnSO_4$, pH 7.8. In some experiments, we also added β-ME to 50 mM; this is now a routine precaution to stabilize IclR protein.

Cells were disrupted in a French pressure cell at 18,000 psi, and the cellular debris was removed by centrifugation at 27,000 × *g* for 1 h at 4 °C. The supernatant was loaded directly onto a Biorex 70 (Bio-Rad) column equilibrated in the same buffer as used for cell suspension. Proteins not bound to the column were eluted with the same starting buffer, and then the bound proteins were eluted with a linear gradient from 0.2 to 0.7 M KCl in the same buffer.

Fractions were monitored by their conductivity, and by their absorbance at 280 nm except when β-ME was present, in which case we used a Bradford (1976) protein assay kit (Bio-Rad) with bovine gamma globulin as standard. Fractions with a high proportion of IclR protein were identified after SDS-PAGE (Laemmli, 1970), pooled, and concentrated in an Amicon ultrafiltration cell with a YM10 membrane. At this stage, the IclR protein was about 90% pure and quite suitable for mass spectrometry studies. Further purification of this fraction by size fractionation, if desired, was done in 20 mM Tris-Cl, 1 mM EDTA, and an appropriate amount of KCl at pH 7.8. In early experiments with truncated IclR we used different concentrations of KCl for each of peaks A and B. For later experiments, we always used 0.4 M KCl and 50 mM β-ME. Preparative col-

umns were either Sephadex G75 or G100. Fractions were monitored as before. After SDS-PAGE, the fraction showing the best yield of IclR with the minimum of impurities was chosen for the experiments described in this paper. The purified samples were dialyzed at once against a storage buffer, containing 20 mM Tris-Cl, pH 7.8, 0.4 M KCl, 1 mM EDTA, 1 mM DTT, and 30% glycerol, and, if not used at once, were stored frozen at −20 °C or lower.

### *DNA binding by IclR protein*

The larger of the two *Cla*I-*Hin*d III restriction fragments from clone λGLY1, which contains the entire *ace*BAK operon, was cloned into pBR322, and from this was subcloned a 170-bp *Rsa*I-*Bam*H I fragment, which contains the *ace* operator sequence. This subclone, in the vector pTZ19U, was designated pTZop and was the standard source of ace operator DNA. Labeled DNA for gel shift and DNase footprinting experiments was prepared by digesting pTZop with one of *Eco*R I or *Hin*d III and filling in the cohesive ends with Klenow fragment of DNA polymerase I using a dNTP mix that included $\alpha$-$^{32}$P-dATP. After ethanol precipitation, the plasmid was digested with the second restriction enzyme (*Hin*d III or *Eco*R I, respectively). The insert band, 227 bp, was purified on a 5% polyacrylamide gel (Maxam & Gilbert, 1980).

Operator DNA binding was monitored by electrophoresis mobility shift assays (Carey, 1991) in which IclR protein was used to titrate a constant amount of labeled operator. Reaction solutions (12 μL) containing approximately 1,500 cpm (0.07 fmol) of $^{32}$P-labeled operator and varying amounts of freshly diluted IclR protein, were incubated 20 min at room temperature in a buffer containing 20 mM Tris-Cl, pH 7.8, 50 mM KCl, 8.3 mM $MgCl_2$, 0.6 mM EDTA, 0.4 mM DTT, 42 μg/mL bovine serum albumin, and 5% glycerol. After incubation, 10 μL of each mixture was loaded on a pre-run 5% polyacrylamide gel (19:1, acrylamide:methylene-bis-acrylamide) in 1× TBE buffer (Maniatis et al., 1982), and electrophoresed at 100 volts until the bromphenol blue dye marker had run about 8 cm. Gels were then fixed in 10% acetic acid/10% methanol, dried under vacuum, and subjected to autoradiography. Using the exposed film as a template, gel regions corresponding to the bound and free DNA were excised, placed in vials, and the radioactivity determined by scintillation counting. The percentage of free DNA was plotted against IclR concentration (Fig. 8). Data were fitted to the Hill (1910) equation using the nonlinear least squares fitting program Kaleidagraph Version 2.1.1 (Abelbeck Software).

DNase I footprint titrations were conducted as described by Brenowitz et al. (1989). Samples were run on 8% polyacrylamide gels (5% crosslinker) containing 8 M urea and 0.5× TBE buffer. Sequencing ladders were prepared by the dideoxy chain termination method (Sanger et al., 1977) using as template the same *Eco*R I-*Hin*d III fragment cloned into M13mp18 or M13mp19 and the primers described by DiRusso et al. (1992). The sequence was also confirmed by using a ladder generated with the A+G reaction mix of Maxam and Gilbert (1980).

### *CNBr cleavage and trypsin digestion*

CNBr cleavage of IclR protein was performed according to the original conditions of Gross and Witkop (1961), with 0.1 M HCl as solvent. After incubation overnight in the dark, at room temperature, the reaction mixture was freeze-dried. Tryptic digestions were performed on 0.2 mg samples of IclR protein in 0.44 mL of 0.1 M ammonium bicarbonate buffer, adding 20 μg of TPCK-trypsin (Sigma), and incubating 2.0 h at 37 °C; the reactions were freeze-dried. These conditions were chosen deliberately to be mild, and the result was that some trypsin-sensitive bonds were not cleaved, so that some of the expected peptides were obtained only in combination with a neighbor (see Results and discussion).

### *Mass spectrometry*

Our first experiments were performed using a MALDI/TOF mass spectrometer built in our laboratory for SIMS measurements and later modified for MALDI (Tang et al., 1988). Protein samples were dialyzed against nanopure water, which caused the protein to precipitate, and lyophilized in aliquots of about 100 μg. The sample was dissolved in 0.1% trifluoroacetic acid in water, to give a concentration in the range 0.03–5 mg/mL, together with Substance P ($M$ = 1,347.6 u), insulin ($M$ = 5,733.5 u), or myoglobin ($M$ = 16,951.5 u) to act as $m/z$ calibrants. A few microliters of this solution was mixed with 20 μL of a 1% solution of 4HCCA in water and acetonitrile (1:4 v/v) as a matrix. This mixture was then dropped onto a 1-cm diameter etched silver foil and allowed to dry in air. TOF mass spectra of ions produced by irradiation of the sample with pulses from a nitrogen laser operating at 337 nm were recorded.

Later experiments employed a novel ESI/TOF instrument, also constructed in our laboratory (Verentchikov et al., 1994). Instead of lyophilization, the samples were dialyzed as before, and then solubilized with 5% acetic acid and transferred to a Centricon 10 membrane cartridge (Amicon) and concentrated to about $10^{-5}$ M. This sample was usually diluted to $10^{-6}$ M in 5% acetic acid containing 40% methanol before electrospray. Substance P, $[M+H]^+$ = 1,347.7 u, was used in a separate measurement to calibrate the $m/z$ scale.

In this spectrometer, ions are counted one at a time. A particular advantage of this technique is the ability to reject background from ions having small mass, but the same $m/z$ ratio as the ions of interest, by accepting only large pulses from the detector (Axelsson et al., 1994; Standing et al., 1994). An example of interest in the present context is shown in Figure 10.

It proved to be a considerable advantage to have both MALDI and ESI ionization methods available constantly. Although the precision of mass measurements obtained by MALDI/TOF was lower than those from ESI/TOF experiments, MALDI is much more tolerant of impurities in protein preparations, so that it was possible to obtain successful preliminary measurements by MALDI when ESI experiments gave nothing. This success then encouraged us to perform further purification steps, especially removing buffer cations such as Tris and potassium, which interfere with ESI, in order to use the higher precision of the ESI/TOF instrument. In addition, the availability of both ionization techniques was of considerable help in interpreting the results of the digests, as remarked above. The recent application of the delayed extraction technique (Wiley & McLaren, 1955) to MALDI measurements (Brown & Lennon, 1995; Ens et al., 1995; Vestal et al., 1995) promises to increase the precision of such measurements, but we believe that the addition of ESI capability is still a significant advantage.

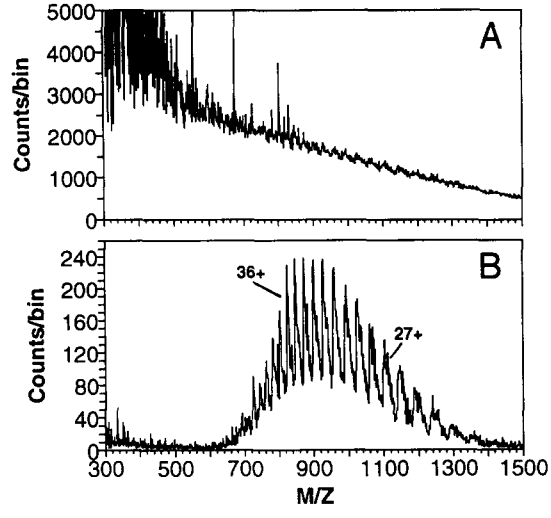


**Fig. 10. A:** ESI/TOF spectrum of an old sample of IclR protein (after ~2 months in frozen solution), including all pulses greater than 10 mV. **B:** The same spectrum including only pulses greater than 200 mV. Spectra A and B were taken simultaneously by using two discriminators set at the corresponding thresholds. Note the change in counting rate of the background.

## Acknowledgments

We thank Glen Klassen and Georg Hauser for instruction in PCR; Isamu Suzuki for unrestricted access to the French press; Mike Mulvey for a gift of his *E. coli* genomic library; Peter Loewen for the phage P1 lysate; Ayeda Ayed for assistance with the CD measurements and analysis and for reading this manuscript; and Mark Soliven and Tanya Boyaniwsky for technical assistance. This work was supported by NSERC (Canada) research grants to H.W.D. and K.G.S., by an NSERC Collaborative Project Grant to H.W.D., K.G.S., W.E., and J.B.W., and by a grant to K.G.S. from the U.S. National Institutes of Health (GM 30605). D.J.H. was the holder of an NSERC PGSA Graduate Research Award.

## References

Akiyama Y, Ito K. 1990. SecY protein, a membrane-embedded secretion factor of *E. coli*, is cleaved by the ompT protease in vitro. *Biochem Biophys Res Commun 167*:711-715.

Axelsson J, Reimann CT, Sundqvist BUR. 1994. Secondary electron resolved mass spectrometry of electrosprayed ions. *Int J Mass Spectrom Ion Processes 133*:141.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein using the principle of protein-dye binding. *Anal Biochem 72*:248-254.

Brenowitz M, Jamison E, Majumdar A, Adhya S. 1990. Interaction of the *Escherichia coli* Gal repressor protein with its DNA operators in vitro. *Biochemistry 29*:3374-3383.

Brenowitz M, Senear DF, Kingston RE. 1989. DNase I footprint analysis of protein-DNA binding. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, eds. *Current protocols in molecular biology*. New York: Wiley.

Brice CD, Kornberg HL. 1968. Genetic control of isocitrate lyase activity in *Escherichia coli*. *J Bacteriol 96*:2185-2186.

Brown RS, Lennon JJ. 1995. Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer. *Anal Chem 67*:1998-2003.

Byrne C, Stokes HW, Ward KA. 1988. Nucleotide sequence of the *ace*B gene encoding malate synthase A in *Escherichia coli*. *Nucleic Acids Res 16*: 9342.

Carey J. 1988. Gel retardation at low pH resolves trp-repressor-DNA complexes for quantitative study. *Proc Natl Acad Sci USA 85*:975-979.

Carey J. 1991. Gel retardation. *Methods Enzymol 208*:103-117.

Chait BT, Kent SBH. 1992. Weighing naked proteins; practical, high accuracy mass measurement of peptides and proteins. *Science 257*:1885-1894.

Chung T, Klumpp DJ, Laporte DC. 1988. Glyoxylate bypass operon of *Escherichia coli*: Cloning and determination of the functional map. *J Bacteriol 170*:386-392.

Cortay JC, Bleicher F, Rieul C, Reeves HC, Cozzone AJ. 1988. Nucleotide sequence and expression of the *ace*K gene coding for isocitrate dehydrogenase kinase/phosphatase in *Escherichia coli*. *J Bacteriol 170*:89-97.

Cortay JC, Nègre D, Galinier A, Duclos B, Perrière G, Cozzone AJ. 1991. Regulation of the acetate operon in *Escherichia coli*: Purification and functional characterization of the IclR repressor. *EMBO J 10*:675-679.

DiRusso CC, Heimert TL, Metzger AK. 1992. Characterization of FadR, a global transcription regulator of fatty acid metabolism in *Escherichia coli*. *J Biol Chem 267*:8685-8691.

Edelhoch H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry 6*:1948-1954.

El-Mansi EM, MacKintosh C, Duncan K, Holms WH, Nimmo HG. 1987. Molecular cloning and over-expression of the glyoxylate bypass operon from *Escherichia coli* ML308. *Biochem J 242*:661-665.

Ens W, Dworschak RG, Spicer V, Standing KG. 1995. Delayed extraction in matrix-assisted laser desorption. Atlanta, Georgia: *Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics*. p 684.

Feng JA, Johnson RC, Dickerson RE. 1994. Hin recombinase bound to DNA. The origin of specificity in major and minor groove interactions. *Science 263*:348-354.

Gross E, Witkop B. 1961. Selective cleavage of the methionyl peptide bonds in ribonuclease with cyanogen bromide. *J Am Chem Soc 83*:1510-1511.

Hedrick JL, Smith AJ. 1968. Size and charge isomer separation and estimation of molecular weights of proteins by disc gel electrophoresis. *Arch Biochem Biophys 126*:155-164.

Hill AV. 1910. A new mathematical treatment of changes of ionic concentration in muscle and nerve under the action of electric currents with a theory as to their mode of excitation. *J Physiol (Lond) 40*:190-224.

Jordan SR, Pabo CO. 1988. Structure of the lambda complex at 2.5 Å resolution: Details of the repressor-operator interactions. *Science 242*: 893-899.

Klumpp DJ, Plank DW, Bowdin LJ, Stueland CS, Chung T, LaPorte DC. 1988. Nucleotide sequence of *ace*K, the gene encoding isocitrate dehydrogenase kinase/phosphatase. *J Bacteriol 170*:2763-2769.

Knight KL, Sauer RT. 1989. Identification of functionally important residues in the DNA binding region of the Mnt repressor. *J Biol Chem 264*:13706-13710.

Kornberg HL. 1963. The role of acetate in isocitrate lyase induction. *Biochim Biophys Acta 73*:517-519.

Kornberg HL. 1966. The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochem J 99*:1-11.

Kunkel TA, Roberts JD, Zakour RA. 1987. Rapid and efficient site-directed mutagenesis without phenotypic selection. *Methods Enzymol 154*:367-382.

Laemmli UK. 1970. Cleavage of structural proteins during assembly of bacteriophage T4. *Nature 227*:280-285.

Laine B, Kmiecik D, Sautière P, Biserte G, Cohen-Solal M. 1980. Complete amino acid sequence of DNA-binding proteins HU-1 and HU-2 from *Escherichia coli*. *Eur J Biochem 103*:447-461.

LaPorte DC, Chung T. 1985. A single gene codes for the kinase and phosphatase which regulate isocitrate dehydrogenase. *J Biol Chem 260*:15291-15297.

Maniatis T, Fritsch EF, Sambrook J. 1982. *Molecular cloning. A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Mann M, Meng CK, Fenn JB. 1989. Interpreting mass spectra of multiply charged ions. *Anal Chem 61*:1702-1708.

Maxam A, Gilbert W. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol 65*:499-560.

Mende L, Timm B, Subramaniam AR. 1978. Primary structure of two homologous ribosome-associated DNA-binding proteins of *Escherichia coli*. *FEBS Lett 96*:395-398.

Miller JH. 1972. *Experiments in molecular genetics*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Nègre D, Cortay JC, Galinier A, Sauve P, Cozzone AJ. 1992. Specific interactions between the IclR repressor of the acetate operon of *Escherichia coli* and its operator. *J Mol Biol 228*:23-29.

Nègre D, Cortay JC, Old IG, Galinier A, Richaud C, Saint Girons I, Cozzone AJ. 1991. Overproduction and characterization of the *iclR* gene product of *Escherichia coli* K-12 and comparison with that of *Salmonella typhimurium* LT2. *Gene 97*:29-37.

Orchard K, May GE. 1993. An EMSA-based method for determining the molecular weight of a protein-DNA complex. *Nucleic Acids Res 21*: 3335-3336.

Pabo CO, Krovatin W, Jeffrey A, Sauer RT. 1982. The N-terminal arms of λ repressor wrap around the operator DNA. *Nature 298*:441–443.

Perczel A, Park K, Fasman GD. 1992. Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: A practical guide. *Anal Biochem 203*:83–93.

Rieul C, Bleicher F, Duclos B, Cortay JC, Cozzone AJ. 1988. Nucleotide sequence of the *aceA* gen coding for isocitrate lyase in *Escherichia coli*. *Nucleic Acids Res 16*:5689.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science 239*:487–491.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74*:5463–5467.

Schultz SC, Shields GC, Steitz TA. 1991. Crystal structure of a CAP–DNA complex: The DNA is bent by 90 degrees. *Science 253*:1001–1007.

Standing KG, Chernushevich IV, Ens W, Verentchikov AN. 1994. Time-of-flight methods for analysing large molecules. Chicago, Illinois: *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics*. p 1148.

Sugimura K, Nishihara T. 1988. Purification, characterization and primary structure of *Escherichia coli* protease VII with specificity for paired basic residues: Identity of protease VII with ompT. *J Bacteriol 170*:5625–5632.

Sunnarborg A, Klumpp D, Chung T, Laporte DC. 1990. Regulation of the glyoxylate bypass operon: Cloning and characterization of *iclR*. *J Bacteriol 172*:2642–2649.

Tang X, Beavis R, Ens W, Lafortune F, Schueler B, Standing KG. 1988. A secondary ion time-of-flight mass spectrometer with an ion mirror. *Int J Mass Spectrom Ion Processes 85*:43–67.

Tong EK, Duckworth HW. 1975. The quaternary structure of citrate synthase from *Escherichia coli* K12. *Biochemistry 14*:235–241.

Vanderwinkel E, DeVlieghere M. 1968. Physiologie et génétique de l'isocitratase et des malate synthases chez *Escherichia coli*. *Eur J Biochem 5*:81–90.

Verentchikov A, Ens W, Standing KG. 1994. A reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal injection. *Anal Chem 66*:126–133.

Vershon KA, Liao SA, McClure WR, Sauer RT. 1987. Bacteriophage P22 Mnt repressor. DNA binding and effects on transcription in vitro. *J Mol Biol 195*:311–322.

Vestal ML, Juhasz P, Martin SA. 1995. Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom 9*:1044–1050.

Wiley WC, McLaren IH. 1955. Time-of-flight mass spectrometer with improved resolution. *Rev Sci Instrum 26*:1150–1157.

Wolberger C, Vershon AK, Liu B, Johnson AD, Pabo CO. 1991. Crystal structure of a MAT$\alpha$2 homeodomain-operator complex suggests a general model for homeodomain–DNA interactions. *Cell 67*:517–528.

Wood WB. 1966. Host specificity of DNA produced by *Escherichia coli*: Bacterial mutations affecting the restriction and modification of DNA. *J Mol Biol 16*:118–133.

Zhou J. 1994. Investigations of matrix-assisted laser desorption with a reflecting time-of-flight mass spectrometer [thesis, part 2]. Winnipeg, Manitoba, Canada: University of Manitoba.