Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 9900–9904, October 1995
Immunology

Direct observation of disordered regions in the major histocompatibility complex class II-associated invariant chain

ALAN JASANOFF*†, SEONG-JOON PARK*†, AND DON C. WILEY*‡§

*Department of Molecular and Cellular Biology, †Committee on Higher Degrees in Biophysics, and ‡Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138

Contributed by Don C. Wiley, June 23, 1995

ABSTRACT **Invariant chain (Ii) is a trimeric membrane protein which binds and stabilizes major histocompatibility complex class II heterodimers in the endoplasmic reticulum and lysosomal compartments of antigen-presenting cells. In concert with an intracellular class II-like molecule, HLA-DM, Ii seems to facilitate loading of conventional class II molecules with peptides before transport of the class II–peptide complex to the cell surface for recognition by T cells. The interaction of Ii with class II molecules is thought to be mediated in large part through a region of 24 amino acids (the class II-associated Ii peptide, CLIP) which binds as a cleaved moiety in the antigenic peptide-binding groove of class II molecules in HLA-DM-deficient cell lines. Here we use nuclear magnetic resonance techniques to demonstrate that a soluble recombinant Ii ectodomain contains significant disordered regions which probably include CLIP.**

The mechanism by which invariant-chain (Ii) trimers associate trivalently with most members of the polymorphic major histocompatibility complex (MHC) class II family is unknown (see ref. 1 for review). Several studies (2–6) have shown that a 24-amino acid sequence lying between the N-terminal transmembrane region and C-terminal protease-resistant trimerization domain of Ii is important for this interaction; the 24-residue sequence is designated the class II-associated Ii peptide (CLIP) (Fig. 1). One possibility for a binding mechanism is that unfolded portions of Ii including CLIP mimic *in vivo* the behavior of exogenous peptides by binding in the class II peptide-binding groove, simultaneously stabilizing and adhering with high affinity to class II molecules (7–10). A number of experiments have demonstrated that Ii inhibits peptide binding to class II molecules (11–13). Consistent with the idea that significant parts of Ii may be unfolded is the pronounced sensitivity of Ii to proteases, especially in its N-terminal third (14–18). The anomalously high apparent molecular mass of Ii observed by us and others (16, 19, 20) is also suggestive of dynamical disorder in the protein; the molecular mass estimated by light-scattering for the soluble Ii used in our studies is 100 kDa, approximately twice its actual molecular mass of 54 kDa.

Nuclear magnetic resonance (NMR) is uniquely suited to the diagnosis of disordered regions in macromolecules because of the combined dependence of NMR linewidth on local flexibility and global rotational motion. For globular proteins, linewidth is roughly proportional to the hydrodynamic volume and to the correlation time for rotational tumbling, τ_c. While peaks in the spectrum of a compact 10-kDa protein ($\tau_c \approx 7$ ns) would have characteristic linewidths around 10 Hz in water at 25°C, peaks in the spectrum of a similarly shaped 50- to 100-kDa protein under the same conditions ($\tau_c \approx 35$–70 ns) would have linewidths from 50 to 100 Hz, which are almost unobservable by NMR because of the inverse relationship between linewidth and peak height for each resonance. Flexible regions in proteins, however, enjoy short local motional correlation times and for this reason contribute narrow resonances to the spectrum of a protein, regardless of the protein's volume or overall tumbling rate (21–24). If Ii were a compact protein, its molecular weight and large hydrodynamic volume would render it too large for study by conventional solution NMR, but the observation of many narrow resonances in NMR spectra of Ii enables us to conclude that extensive portions of the protein are disordered in solution. Observation and sequential assignment of sharp resonances in the NMR spectra of the 70-kDa chaperone-associated heptamer GroES allowed Landry *et al.* (21) to conclude that a specific region of the protein, residues 17–32, was mobile in solution. Although we cannot assign most resonances in the Ii spectra to specific residues, a combination of NMR data on Ii and on fragments of the protein strongly suggest that the disordered regions of Ii span the N-terminal third of the protein, including the CLIP region implicated in binding to MHC class II proteins.

METHODS

Protein and Peptide Purification. Soluble Ii including an N-terminal hexahistidine tag and residues 72–216 of the intact protein, Ii-(h72–216), was expressed in *Escherichia coli* and purified by elution from a nickel column followed by anion-exchange chromatography (16). A shorter version of the protein, Ii-(h94–216) was similarly expressed and purified and was digested with trypsin (≈50 μg/ml) for ≈12 hr to yield Ii-(118–193). The 24-residue CLIP was synthesized on an Applied Biosystems 431A peptide synthesizer using fluorenylmethoxycarbonyl chemistry. The peptide was cleaved and deprotected with a trifluoroacetic acid/water/ethanedithiol/thioanisole/phenol mixture and purified by high-performance liquid chromatography.

NMR Spectroscopy. NMR spectra were recorded on Bruker AM-500, Varian VXR-500, and Varian Unity-500 spectrometers at 25°C and with sample concentrations from 0.3 to 2.0 mM. Total correlated spectroscopy (TOCSY; 60-ms mixing time), nuclear Overhauser effect spectroscopy (NOESY; 150-ms mixing time), and double-quantum-filtered correlated spectroscopy (DQF-COSY) spectra were recorded for Ii-(h72–216) in 10% ^2H_2O (D_2O), and TOCSY and NOESY spectra were recorded for Ii-(h72–216) in 100% D_2O. TOCSY and NOESY spectra (same mixing times) were recorded for Ii-(118–193) and CLIP in 10% D_2O only. All NMR samples were prepared in phosphate-buffered saline at pH 5.9 (3 mM KCl/3 mM KH_2PO_4/123 mM NaCl/13 mM Na_2HPO_4), and protease inhibitors (EDTA, phenylmethanesulfonyl fluoride, iodoacetamide, aprotinin, leupeptin, and pepstatin) were added to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Ii, invariant chain; CLIP, class II-associated Ii peptide; MHC, major histocompatibility complex; TOCSY, total correlated spectroscopy; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; DQF-COSY, double quantum filtered correlated spectroscopy; D_2O, ^2H_2O.

§To whom reprint requests should be addressed.

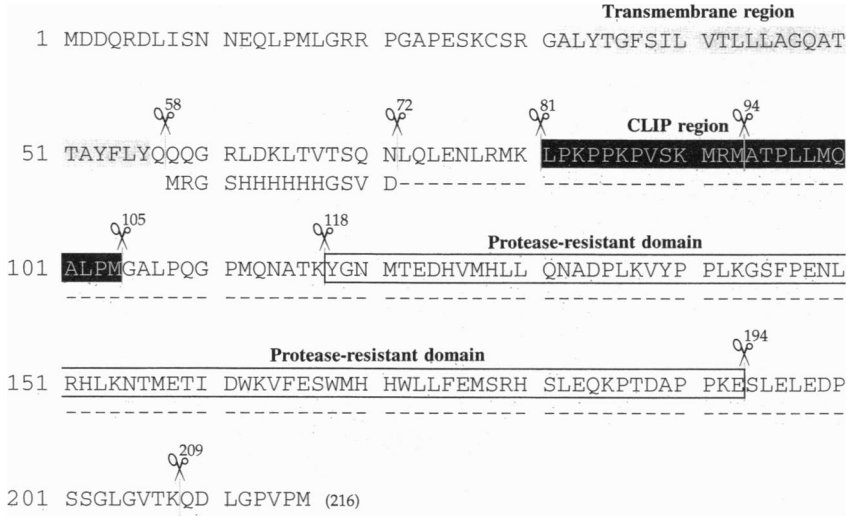

FIG. 1. Amino acid sequence of human Ii. The histidine-tagged recombinant sequence Ii-(h72–216) is shown below the cellular sequence, beginning at amino acid 58 and extending to the C terminus of the wild-type protein (dashes indicate no difference). Known proteolytic cleavage sites are marked with scissors and the number of the residue following the cleavage point. The transmembrane region of Ii is shown on a gray background and the region corresponding to CLIP is shown on a black background. The residues in Ii comprising the protease-resistant trimerization domain generated by tryptic digestion, Ii-(118–193), are boxed.

NMR samples containing Ii-(h72–216). Data were processed and plotted with the program FELIX (Biosym Technologies, San Diego).

RESULTS AND DISCUSSION

NMR Evidence for Flexibility in the Ii Ectodomain [Ii-(h72–216)]. The soluble form of Ii investigated here, Ii-(h72–216), spans residues 72–216 of the intact wild-type protein and retains an N-terminal histidine tag used in the purification. Ii-(h72–216) is trimeric, and it associates with and blocks peptide binding to MHC class II proteins as does wild-type Ii (16). The one-dimensional NMR spectrum of Ii-(h72–216) (Fig. 2*A*) contains many sharp resonances, indicating the presence of significant regions of disordered structure in the 54-kDa trimer. Also typical of randomly coiled polypeptides

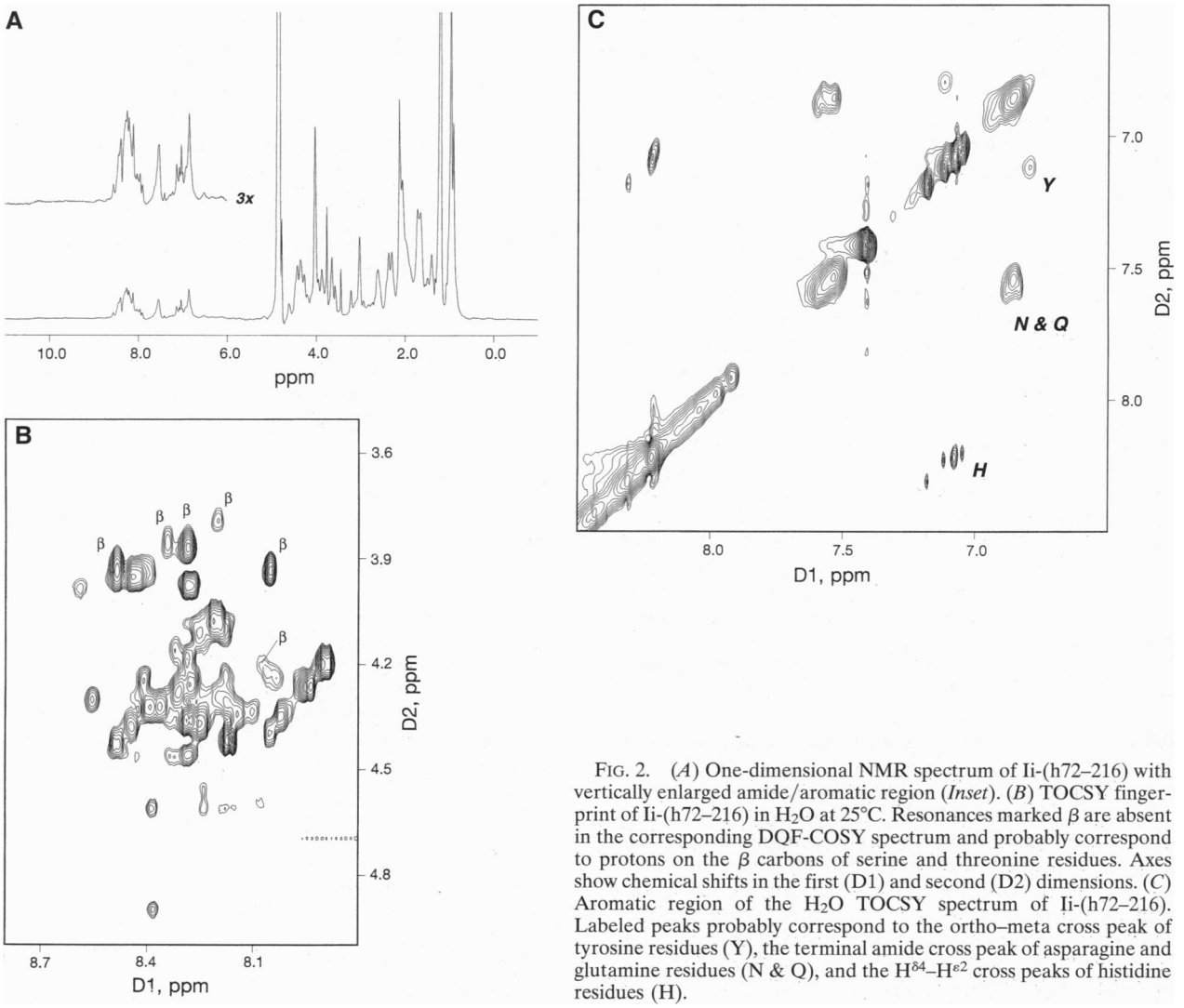

FIG. 2. (*A*) One-dimensional NMR spectrum of Ii-(h72–216) with vertically enlarged amide/aromatic region (*Inset*). (*B*) TOCSY fingerprint of Ii-(h72–216) in H_2O at 25°C. Resonances marked β are absent in the corresponding DQF-COSY spectrum and probably correspond to protons on the β carbons of serine and threonine residues. Axes show chemical shifts in the first (D1) and second (D2) dimensions. (*C*) Aromatic region of the H_2O TOCSY spectrum of Ii-(h72–216). Labeled peaks probably correspond to the ortho–meta cross peak of tyrosine residues (Y), the terminal amide cross peak of asparagine and glutamine residues (N & Q), and the $H^{\delta 4}$–$H^{\varepsilon 2}$ cross peaks of histidine residues (H).

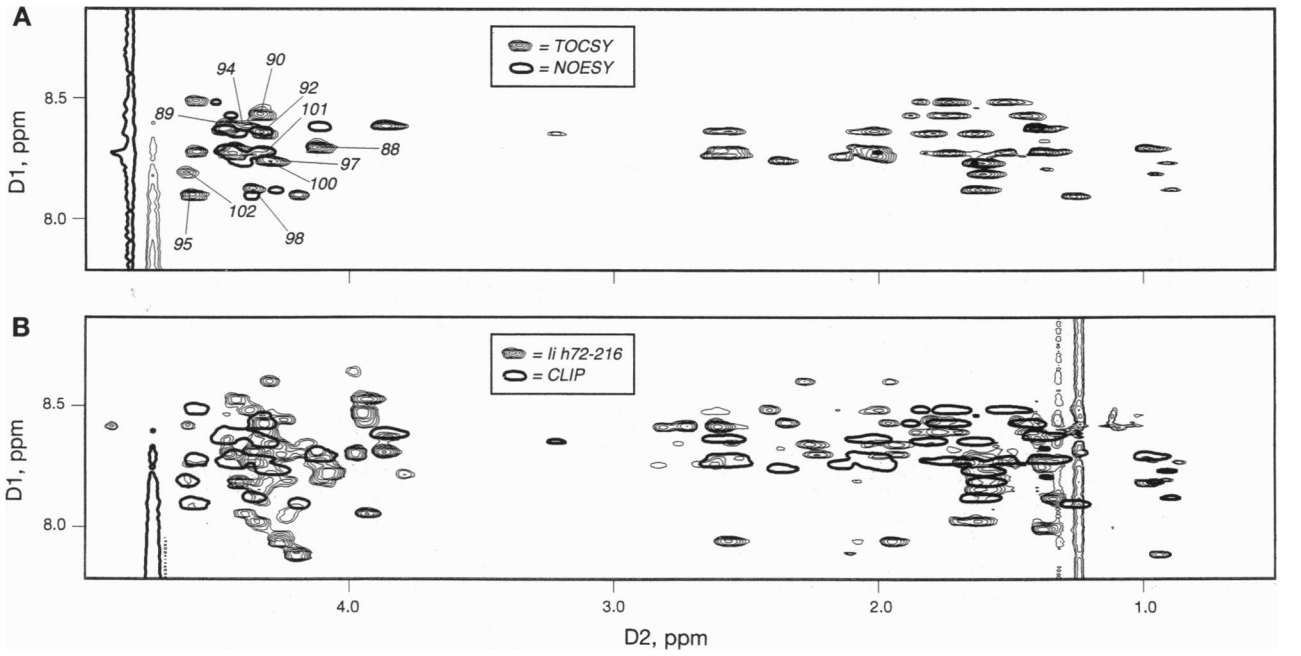

FIG. 3. (*A*) A region of the NMR spectra showing cross peaks between amide and aliphatic protons of the Ii-derived CLIP [Ii-(181–204)]. TOCSY peaks are contoured in thin lines, while NOESY peaks are shown as thick outlines. Assigned H^N–H^α peaks are labeled by the number of the corresponding residue in Ii. (*B*) Overlay of the TOCSY spectrum of Ii-(h72–216) (thin contour lines) with the TOCSY spectrum of CLIP (thick outline), showing similarity between the two.

are the absence of upfield-shifted methyl peaks among the sharp resonances and the low chemical shift dispersion in the amide region of the spectrum (7–9 ppm). A prominent high-field peak is due to the presence of protease inhibitors in the solution; addition of protease inhibitors was necessary to prevent degradation of the protein during the experiment. Two-dimensional NMR spectra (TOCSY and NOESY) of Ii-(h72–216) were obtained at 25°C and pH 5.9 in H_2O and D_2O. The TOCSY "fingerprint" region (Fig. 2*B*) typically contains one H^N–H^α cross peak for each residue except proline, and H^N–H^β cross peaks for serine and threonine; the latter can be identified due to their absence from an H_2O DQF-COSY spectrum (not shown) and are labeled β in Fig. 2*B*. Although there is considerable overlap of peaks in the fingerprint spectrum, it is highly likely, based on the number of explicitly observable peaks (40 non-β), that >40 non-proline residues contribute to the observable spectrum.

Unique assignment of resonances on the basis of the two-dimensional spectra presented here is precluded by peak overlap, but several residues are distinguished by characteristic signals in the low-field region of the spectra (Fig. 2*C*). In particular, the $H^{\delta 4}$–$H^{\varepsilon 2}$ cross peaks of at least four histidine residues, the geminal-proton cross peaks of the side-chain amides of several glutamine and asparagine residues (at least five as judged from the fine structure of the corresponding NOESY peak, not shown), and the ortho–meta crosspeak of a tyrosine ring, one of two in Ii-(h72–216), can be distinguished and are all clustered near their expected random-coil values (24).

NMR Spectra of CLIP and Comparison with Spectra of Ii-(h72–216). We have obtained a partial assignment for the 24-residue proline-rich CLIP [Ii-(81–104)] from NOESY and TOCSY spectra recorded in H_2O at 25°C and pH 5.9 (Fig. 3*A*). The lack of medium or long-range interresidue NOEs, even with the relatively long NOE mixing time of 150 ms, indicates the absence of significant stable structure formation in CLIP. The presence of weak TOCSY signals corresponding to the side-chain amino group of one of the unassigned lysine residues argues that it may be involved in an intramolecular hydrogen bond, though no NOEs to the side chain can be found. Assigned peaks in the two-dimensional spectra of CLIP correlate well with peaks in the spectra of Ii-(h72–216) (Fig. 3*B*), notably including Val-88, Ser-89, Thr-95, Leu-97, and Leu-102, which encompass a region of CLIP known from x-ray crystallographic studies to bind in the peptide-binding site of the human class II molecule HLA-DR3 (P. Ghosh, E. Amaya, E. Mellins, and D.C.W., unpublished data). Most assigned resonances fall near published random-coil chemical-shift values.

NMR of a Trypsin-Generated Compact Ii Domain. A protease-resistant domain of Ii, Ii-(118–193), can be generated by trypsin treatment of a histidine-tagged recombinant Ii comprising residues 94–216, Ii-(h94–216) (16). Because of its smaller size (25 kDa), the trimeric Ii-(118–193) polypeptide produces a complete one-dimensional NMR spectrum characteristic of a large, fully folded protein, unlike Ii-(h72–216), containing upfield-shifted methyl resonances and chemical-shift dispersion in the characteristic amide and aliphatic ppm ranges (Fig. 4*A*). Examination of the two-dimensional spectra of Ii-(118–193) indicates the presence of specific resonances that are clearly absent from the spectra of Ii-(h72–216) notably corresponding to tryptophan and phenylalanine residues present only between residues 118 and 193 of Ii (Fig. 4*B*). These differences enforce the conclusion that Ii residues 118–193 are compact and well-structured both as an isolated domain and in the context of the larger Ii-(h72–216), where the slow tumbling rate of the larger protein renders them inaccessible to observation by traditional NMR methods.

Comparison with Previous Results and Possible Artifacts. Our results are fully consistent with proteolysis data (14–18) indicating cleavage sites spaced approximately every 10 residues in the N-terminal segment of the Ii ectodomain, residues 72–118 (Fig. 1). The observation of >40 H^N–H^α peaks in the TOCSY fingerprint spectrum, as well as the residue-specific resonances mentioned above, suggests that the N-terminal flexibility is continuous over the entire range 72–118, though we cannot rule out the unlikely possibility that small (≈5-residue) segments within this range are rigidly immobilized by interactions with the core domain. It is unclear to what extent residues 193–216 near the C terminus of Ii-(h72–216) are

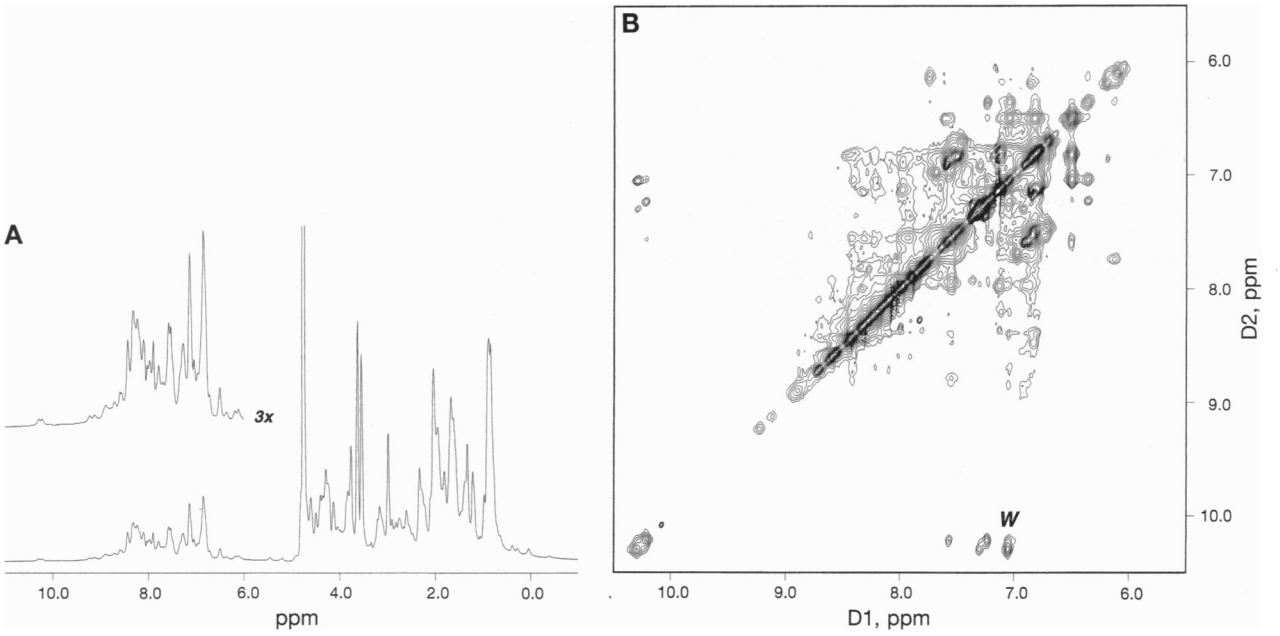

FIG. 4. (*A*) One-dimensional spectrum of trimeric Ii-(118–193) with vertically enlarged amide/aromatic region (*Inset*). (*B*) NOESY spectrum of Ii-(118–193) showing the region containing cross peaks among amide and aromatic protons. Intraresidue NOEs involving the $H^{\varepsilon 1}$ indole protons of tryptophan residues are labeled (W).

flexible and contribute to the NMR spectra in Fig. 2. There is also a possibility that Ii-(h72–216) contains regions of flexibility, particularly near its truncated N terminus, that are not present in the full-length native Ii, which is normally anchored to the membrane via residues 32–56. Previous structural studies on solubilized membrane proteins such as influenza hemagglutinin (26, 27) and human growth hormone receptor (28) have indicated that domains appear to be well folded within very few residues of their transmembrane regions, so that structural perturbations due to cleavage from the membrane are minimal. Though the possibility of more drastic structural changes due to expression of Ii without its transmembrane region cannot be ruled out, it seems unlikely in the context of previous research that perturbations introduced by solubilization would unfold the protein as extensively as we have observed. The possible structural significance of such perturbations specific to our construct is further restricted because Ii-(h72–216) retains the two main activities attributed to the ectodomain of wild-type Ii, association with class II molecules and inhibition of peptide binding to class II molecules (16).

Conclusion. We have shown that the NMR spectra of Ii-(h72–216), a soluble ectodomain of MHC class II-associated Ii, contain resonances corresponding to at least 40 non-proline residues in locally flexible regions of the protein. Among the sharp resonances of the Ii-(h72–216) spectra, histidine, glutamine, asparagine, and tyrosine side-chain spin systems with random-coil chemical shifts occurred in numbers consistent with flexibility of the N-terminal third of the protein. This finding, the correlation of Ii-(h72–216) resonances with assigned resonances in the spectra of synthetic CLIP [Ii-(81–104)], and the observation of specific differences between spectra of Ii-(h72–216) and its protease-resistant core domain, Ii-(118–193), provide strong evidence that the N-terminal third of Ii-(h72–216) extending up to Tyr-118 and containing CLIP is largely structurally disordered in solution whereas residues 118–193 are compact. Previous studies have identified the N-terminal (membrane-proximal) region of Ii as essential for the interaction of Ii with MHC class II molecules (2, 3). X-ray crystallography of a DR3–CLIP complex has shown that as an isolated peptide, CLIP binds in the MHC binding groove (P. Ghosh, E. Amaya, E. Mellins, and D.C.W., unpublished data). It therefore seems likely that flexibility in the N-terminal region of Ii may have important functional consequences for the interaction of Ii and class II molecules *in vivo* (9, 25).

We are grateful to Gregory Heffron (Harvard Medical School) for help with NMR data acquisition and to Lila Gierasch for reading and commenting on the manuscript. We thank Anastasia Haykov for technical assistance and Don Caspar, James Hogle, and Stuart Schreiber for helpful discussions. National Science Foundation Grant CHE88-14019 and National Institutes of Health Grant 1-S10-RR04870-01 provided support for the Bruker AM-500 spectrometer (Harvard Chemistry Department). A.J. and S.-J.P. are supported by Howard Hughes predoctoral fellowships. D.C.W. is an investigator of the Howard Hughes Medical Institute.

1. Cresswell, P. (1994) *Annu. Rev. Immunol.* **12,** 259–293.
2. Bijlmakers, M. E., Benaroch, P. & Pleogh, H. L. (1994) *J. Exp. Med.* **180,** 623–629.
3. Freisewinkel, I. M., Schenk, K. & Koch, N. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9703–9706.
4. Mellins, E., Cameron, P., Amaya, M., Goodman, S., Pious, D., Smith, L. & Arp, B. (1994) *J. Exp. Med.* **179,** 541–549.
5. Riberty, J. M., Newcomb, J. R., Surman, M. J., Barbosa, J. A. & Cresswell, P. (1992) *Nature (London)* **360,** 474–477.
6. Sette, A., Ceman, S., Kubo, R. T., Sakaguchi, K., Appela, E., Hunt, D. F., Davis, T. A., Michel, H., Shabanowitz, J., Rudersdorf, R., Grey, H. M. & Demars, R. (1993) *Science* **258,** 1801–1804.
7. Germain, R. N., Rinker, A. G., Jr. (1993) *Nature (London)* **363,** 725–728.
8. Malcherek, G., Gnau, V., Jung, G., Rammensee, H.-G. & Melms, A. (1995) *J. Exp. Med.* **181,** 527–536.
9. Romagnoli, P. & Germain, R. N. (1994) *J. Exp. Med.* **180,** 1107–1113.
10. Sette, A., Southwood, S., Miller, J. & Appella, E. (1995) *J. Exp. Med.* **181,** 677–683.
11. Roche, P. A. & Cresswell, P. (1990) *Nature (London)* **345,** 615–618.
12. Newcomb, J. R. & Cresswell, P. (1993) *J. Immunol.* **150,** 499–507.
13. Teyton, L., O'Sullivan, D., Dickson, P. W., Lotteau, V., Sette, A., Fink, P. & Peterson, P. A. (1990) *Nature (London)* **348,** 39–44.
14. Avva, R. R. & Cresswell, P. (1994) *Immunity* **1,** 763–774.
15. Marks, M. S. & Cresswell, P. (1991) *J. Immunol.* **136,** 2519–2525.
16. Park, S.-J., Sadegh-Nasseri, S. & Wiley, D. C. (1995) *Proc. Natl. Acad. Sci. USA* **92,** in press.

17. Roche, P. A. & Cresswell, P. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 3150–3154.
18. Xu, M., Capraro, G. A., Daibata, M., Reyes, V. E. & Humphreys, R. E. (1994) *Mol. Immunol.* **31,** 723–731.
19. Marks, M. S., Blum, J. S. & Cresswell, P. (1990) *J. Cell Biol.* **111,** 839–855.
20. Roche, P. A., Marks, M. S. & Cresswell, P. (1991) *Nature (London)* **354,** 392–394.
21. Landry, S. J., Zellstra-Ryalls, J., Fayet, O., Georgopoulos, C. & Gierasch, L. M. (1993) *Nature (London)* **364,** 255–258.
22. Oswald, R. E., Bogusky, M. J., Bamberger, M., Smith, R. A. G. & Dobson, C. M. (1989) *Nature (London)* **337,** 579–582.
23. Perham, R. N., Duckworth, H. W. & Roberts, G. C. K. (1981) *Nature (London)* **292,** 474–477.
24. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
25. Sant, A. J. & Miller, J. (1994) *Curr. Opin. Immunol.* **6,** 57–63.
26. Wiley, D. C., Wilson, I. A. & Skehel, J. J. (1981) *Nature (London)* **289,** 373–378.
27. Wilson, I. A., Wiley, D. C. & Skehel, J. J. (1981) *Nature (London)* **289,** 368–373.
28. de Vos, A. M., Ultsch, M. & Kossiakoff, A. A. (1992) *Science* **255,** 306–312.