Journal of Biomolecular NMR, **12:** 183–188, 1998.
KLUWER/ESCOM

Local helix content and RNA-binding activity of the N-terminal leucine-repeat region of hepatitis delta antigen

Jya-Wei Cheng[a,*], I-Jin Lin[a], Yuan-Chou Lou[a], Ming-Tao Pai[a] and Huey-Nan Wu[b]
[a]*Department of Life Science, National Tsing Hua University, Hsinchu 300, Taiwan, Republic of China*
[b]*Institute of Molecular Biology, Academia Sinica, Taipei 11529, Taiwan, Republic of China*

Received 12 November 1997; Accepted 3 March 1998

Key words: CD, HDAg, HDV, RNA binding, solution conformation

Abstract

Hepatitis delta virus (HDV) is a satellite virus of the hepatitis B virus (HBV) which provides the surface antigen for the viral coat. Our results show that the N-terminal leucine-repeat region of hepatitis delta antigen (HDAg), encompassing residues 24–50, binds to the autolytic domain of HDV genomic RNA and attenuates its autolytic activity. The solution conformation of a synthetic peptide corresponding to residues 24–50 of HDAg as determined by two-dimensional ^{1}H NMR and circular dichroism techniques is found to be an α-helix. The local helix content of this peptide was analyzed by NOEs and coupling constants. Mutagenesis studies indicate that Lys^{38}, Lys^{39}, and Lys^{40} within this α-helical peptide may be directly involved in RNA binding. A structural knowledge of the N-terminal leucine-repeat region of HDAg thus provides a molecular basis for understanding its role in the interaction with RNA.

Hepatitis delta virus (HDV) is a satellite virus of the hepatitis B virus (HBV) which provides the surface antigen for the viral coat (Lai, 1995). The genome of the hepatitis delta virus consists of a single-stranded, circular RNA of 1.7 kb which is predicted to fold into a rod-like structure due to extensive intramolecular base pairing (Kos et al., 1986; Wang et al., 1986; Makino et al., 1987). HDV replicates through the synthesis of an antigenomic RNA via a rolling circle mechanism. This mechanism is governed by autocatalytic cleavage and ligation reactions (Kuo et al., 1988; Sharmeen et al., 1988, 1989; Wu and Lai, 1990; Wu et al., 1992). The genome of HDV encodes two proteins, the small delta antigen (S-HDAg, 24 kDa) and the large delta antigen (L-HDAg, 27 kDa) (Bergmann and Gerin, 1986; Bonino et al., 1986; Pohl et al., 1987; Zyzik et al., 1987). The two proteins resemble one another except for the presence in the latter species of an additional 19 amino acids at the C-terminus. Both S-HDAg and L-HDAg are nuclear phosphoproteins with RNA-binding activities, specific for HDV RNA (Chang et al., 1988; Lin et al., 1990; Chao et al., 1991; Hwang et al., 1992); however, they have different functional roles in the life cycle of HDV. While S-HDAg is required for HDV RNA replication (Kuo et al., 1989) , L-HDAg inhibits HDV RNA replication (Chao et al., 1990) and is required for HDV assembly (Chang et al., 1991; Ryu et al., 1992). Hepatitis delta antigen (HDAg) contains several functional domains (Figure 1). The N-terminal one-third, residues 69–88, contains a nuclear localization signal (Xia et al., 1992). This region also contains two stretches of basic amino acid-rich sequence for nuclear targeting of HDAg (Xia et al., 1992) and HDV RNA (Gowans et al., 1988) as well. The central one-third of HDAg contains two arginine-rich motifs (residues 97–107 and 136–146) responsible for RNA-binding activity (Lee et al., 1993). In addition, circular dichroism (CD) and radioimmunoassay studies indicated that a synthetic peptide corresponding to the N-terminal region, residues 12–60, upstream of the nuclear targeting signal, is involved in the formation of an immunoreactive α-helical multimer (Rozzelle Jr. et al., 1995). Analysis of the amino acid sequence for HDAg reveals the

*To whom correspondence should be addressed.

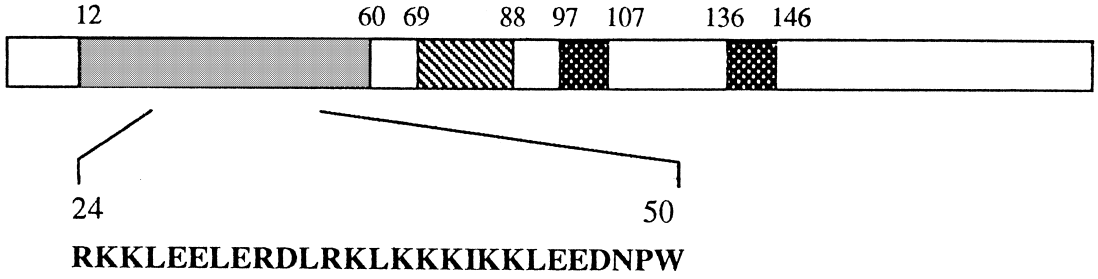

Figure 1. Domains within HDAg: oligomerization domain, residues 12–60; nuclear localization domain, residues 69–88; arginine-rich motifs for RNA-binding activity, residues 97–107 and 136–146. The amino acid sequence 24–50 of the N-terminal leucine-repeat region chosen for this study is shown.

presence of two leucine-repeat regions near the N-terminus; however, only one, residues 30–51, is conserved among five HDV isolates (Chang et al., 1993). Recently, we have found that the N-terminal leucine-repeat region of HDAg, encompassing residues 24–50, can bind to the autolytic domain of HDV genomic sense RNA (HJ12L) (Wu et al., 1997). However, the molecular basis of this RNA-binding activity is not clear. Aknowledge of the structure of the N-terminal leucine-repeat region of HDAg would help understand its RNA-binding activity. In this report, we have determined the solution conformation of a synthetic peptide corresponding to the N-terminal leucine-repeat region of HDAg (residues 24–50) using two-dimensional ^{1}H NMR techniques. Local structural variants of this peptide were analyzed based on NOEs and coupling constants. This structural knowledge provides a molecular basis for understanding the role of the N-terminal leucine-repeat region of HDAg in its interaction with RNA.

A peptide corresponding to residues 24–50 of HDAg, designated as dAg24-50 (Figure 1), was synthesized and purified as reported (Chen et al., 1996; Cheng et al., 1996). Peptides dAgAc24-50 and dAgAc24-50am, with N- and C-termini modified by acetylation and amidation, were synthesized and purified using previously reported procedures (Tan et al., 1993). All NMR experiments were performed on a Bruker DMX-600 spectrometer at 25 °C using a 7 mM sample. NMR spectra were collected in the phase-sensitive mode with time-proportional phase incrementation (TPPI) (Drobny et al., 1979; Marion and Wüthrich, 1983). The acquired data were transferred to an SGI extreme Indigo2 workstation and processed using UXNMR (Bruker). NOESY spectra were acquired at four different mixing times of 75, 150, 200, and 300 ms. Solvent suppression was done by presaturation in the relaxation and mixing period. These

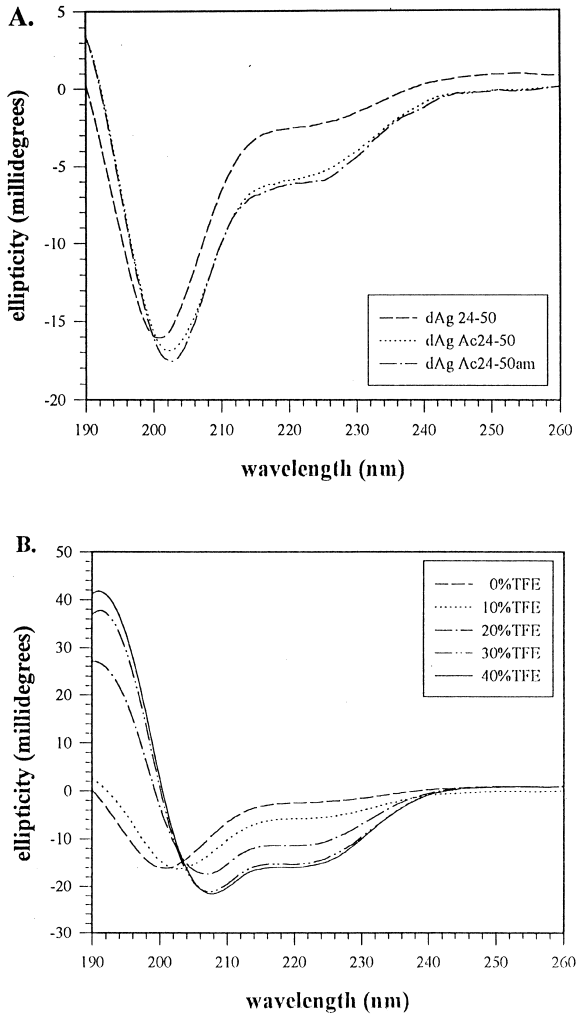

Figure 2. (A) CD spectra of dAg24-50, dAgAc24-50, and dAgAc24-50am collected in H_2O. dAg24-50 has unmodified N- and C-termini, dAgAc24-50 has an acetylated N-terminus, and dAgAc24-50am has an acetylated N-terminus and amidated C-terminus. (B) CD spectra showing the effect of TFE on the conformation of dAg24-50.

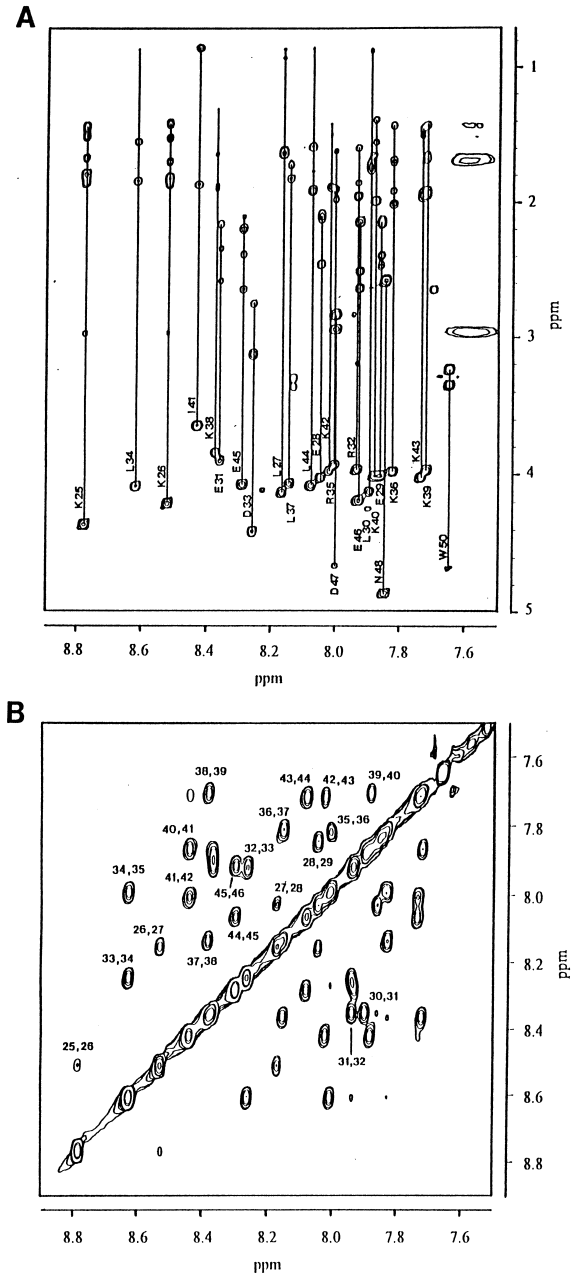

Figure 3. (A) TOCSY spectrum (600 MHz), recorded at 70 ms mixing time of dAg24-50 in solution containing 30% TFE-d_3, showing the NH(F2)-aliphatic(F1) region. Peaks are labeled at the positions of the NH-$C^{\alpha}H$ cross peaks. (B) NH(i)-NH(i+1) region of the 600 MHz NOESY spectrum recorded at 200 ms mixing time for dAg24-50 in solution, containing 30% TFE-d_3, at 25 °C.

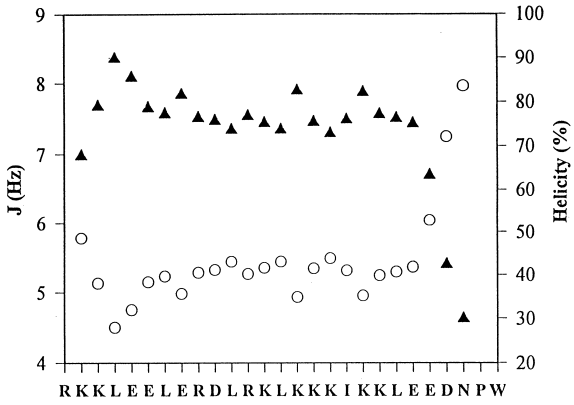

Figure 4. Plot of the $^{3}J_{HN\alpha}$ coupling constant and helix content of dAg24-50 against the peptide sequence. Empty circles represent coupling constants. Filled triangles represent helix content.

phase-sensitive NOESY spectra were collected into 2048 points in the t_2 dimension and 512 points in the t_1 dimension with 16 scans per t_1 increment and a relaxation delay of 1.5 s between each scan. The NOESY data sets were apodized with a shifted-sinebell window function and zero-filled to 4096 and 1024 points in the t_2 and t_1 dimension, respectively. DQF-COSY (Rance et al., 1983) spectra were recorded into 2048 points in t_2 and 512 increments in t_1 with 16 scans per t_1 experiment; 6 Hz of Gaussian broadening was applied in the t_2 domain and a skewed-sinebell window function was used in the t_1 domain. TOCSY spectra (Bax and Davis, 1985) were recorded using the MLEV17 pulse sequence with mixing times of 35 and 70 ms at 2048 points in t_2 and 512 points in t_1.

CD spectra of dAg24-50, dAgAc24-50, and dAgAc24-50am recorded in water are shown in Figure 2A. Evidently, dAgAc24-50 and dAgAc24-50am form stable helical conformations in water as can be seen from their ellipticity minimum at 222 nm (the helicity was calculated to be about 20%). CD spectra of dAg24-50 collected at various concentrations of 2,2,2-trifluoroethanol (TFE) showed a low helical content in the absence of TFE (Figure 2B). Upon addition of TFE, the helical content increased and reached a maximum at 30% (v/v) of TFE.

Detailed characterization of the secondary structure was undertaken by NMR. Due to severe cross-peak overlapping in the fingerprint region of spectra collected in water, it was not possible to assign most of the resonances of dAg24-50, dAgAc24-50, and dAgAc24-50am (data not shown). However, resonances in the fingerprint region of dAg24-50 in 30% TFE containing solution are well resolved, and they

could be assigned using standard procedures. The NH to aliphatic proton region of the TOCSY spectrum collected at 70 ms is shown in Figure 3A. Spin systems of leucine and isoleucine were easily characterized by their connectivities to methyl protons. Spin systems of lysine and arginine were also distinguished from other amino acid residues. Twenty-five intra-residue NH-$C^{\alpha}H$ cross peaks were detected in the DQF-COSY and TOCSY spectra and were used in conjunction with the NOESY spectrum for the sequential assignment purpose (Wüthrich, 1986).

The secondary structure of peptides can be characterized based on a qualitative interpretation of inter-residue NOE distances (Wüthrich, 1986). In the present NMR studies, inter-residue $d_{\alpha N}(i, i+3)$ NOE cross peaks were observed in the NOESY fingerprint region (data not shown). Cross peaks representing the $d_{\alpha N}(i, i+4)$ interactions were found at a lower contour level. The other NOE cross-peaks such as $d_{\alpha\beta}(i, i+3)$ were also assigned in the α to aliphatic proton region. These $d_{\alpha N}(i, i+3)$, $d_{\alpha N}(i, i+4)$ and $d_{\alpha\beta}(i, i+3)$ NOE cross peaks illustrate an α-helical conformation for dAg24-50. Further evidence for this contention is obtained from strong $d_{NN}(i, i+1)$ NOE cross peaks observed for residues 25–47 (Figure 3B). Strong $d_{NN}(i, i+1)$ cross peaks are only found in peptides with an α-helical conformation (Wüthrich, 1986).

A detailed analysis of the local helix content of dAg24-50 based solely on the CD and ^{1}H NOE data was, however, not possible due to conformational averaging and spectral overlapping. The $^3J_{HN\alpha}$ coupling constant can provide complementary information to NOE for the determination of local structural variants in polypeptide backbones (Wüthrich, 1986). However, partial cancellation of the antiphase multiplet components in COSY cross peaks often results in a systematic overestimation of these coupling constants in proteins (Neuhaus et al., 1985). Several methods have been proposed to overcome this problem (Kessler et al., 1985; Kim and Prestegard, 1989; Ludvigsen et al., 1991) , although they are limited by the loss of intensity in the COSY spectra due to mutual cancellation and spectral overlapping in the fingerprint region. Recently, Szyperski et al. (1992) have developed a novel method for extracting $^3J_{HN\alpha}$ coupling constants of proteins using an inverse Fourier transform and iterative fitting in the time domain of individual NOE cross-peak lineshape. This method greatly increases the possibility of obtaining the $^3J_{HN\alpha}$ coupling constant for a particular residue. Well-separated $d_{NN}(i, i+1)$ NOE cross peaks (Figure 3B) and/or $d_{\beta N}(i, i)$ cross peaks in a NOESY spectrum recorded at 200 ms mixing time were used to determine the $^3J_{HN\alpha}$ coupling constant for a given residue according to the INFIT procedures of Szyperski et al. (Szyperski et al., 1992; Millhauser et al., 1996). Some of the resulting coupling constants were confirmed using the method of Kim and Prestegard (1989) from the fingerprint region of the DQF-COSY spectrum (data not shown). Wüthrich (1986) has suggested that the presence of three or more consecutive $^3J_{HN\alpha}$ coupling constants lower than 6 Hz is a good indication of helical structure. Figure 4 shows the $^3J_{HN\alpha}$ coupling constant plotted against the peptide sequence. Helical coupling constants are found for residues 25–46. Extended and end-fraying conformations are found for the C-terminal residues. These results are consistent with the secondary structure derived from NOE data. The $^3J_{HN\alpha}$ coupling constant is a weighted population average and depends on the distribution angles over the population (Kessler et al., 1988; Bradley et al., 1990). Using a two-state model, the local helix content (f_{hlx}) was estimated by the equation

$$f_{hlx} = \frac{J_{mean} - J_{ext}}{J_{hlx} - J_{ext}}$$

where J_{mean} is the experimentally measured coupling constant and J_{hlx} (3.9 Hz) and J_{ext} (9.7 Hz) are ideal α-helix and extended (β) coupling constants, respectively (Wüthrich, 1986; Bradley et al., 1990). Figure 4 also shows the helix population for each residue of dAg24-50. It is obvious that the helix population of the N-terminal residues is higher than that of the C-terminal ones. This may be due to the higher possibility of (i, i+3) and (i, i+4) salt bridge formation between basic and acidic residues in the N-terminus than in the C-terminus (Figure 5).

Recently, we found that dAg24-50 binds to HJ12L, the autolytic domain of HDV genomic sense RNA (Figure 6A) (Wu et al., 1997). Moreover, we found that preincubation of HJ12L with dAg24-50 attenuated the autolytic activity of HJ12L (Figure 6B) (Wu et al., 1997). Many nucleic acid binding proteins use basic amino acid residues to interact with the backbone phosphate groups of nucleic acids. It is conceivable that arginine and lysine residues in the N-terminal leucine-repeat region of HDAg are involved in such interaction. By inspecting the sequence of dAg24-50, we found that the RNA-binding region may locate within residues 35–43 consisting of RKLKKKIKK. Mutant HDAg peptides, where Lys^{38}, Lys^{39}, and Lys^{40} are changed to glutamic acid, bound weakly to HJ12L

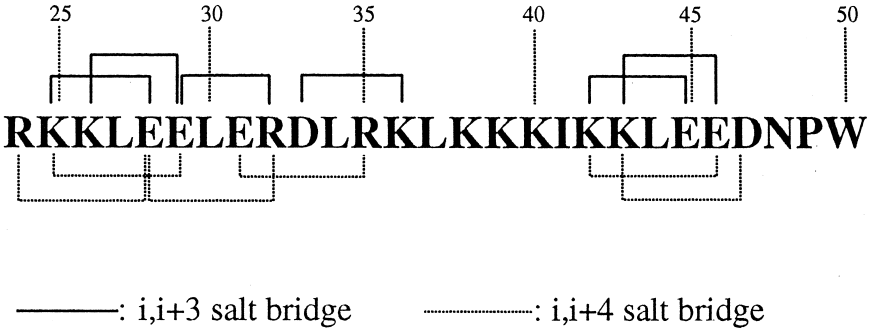

Figure 5. Schematic representation of potential salt bridges in dAg24-50.

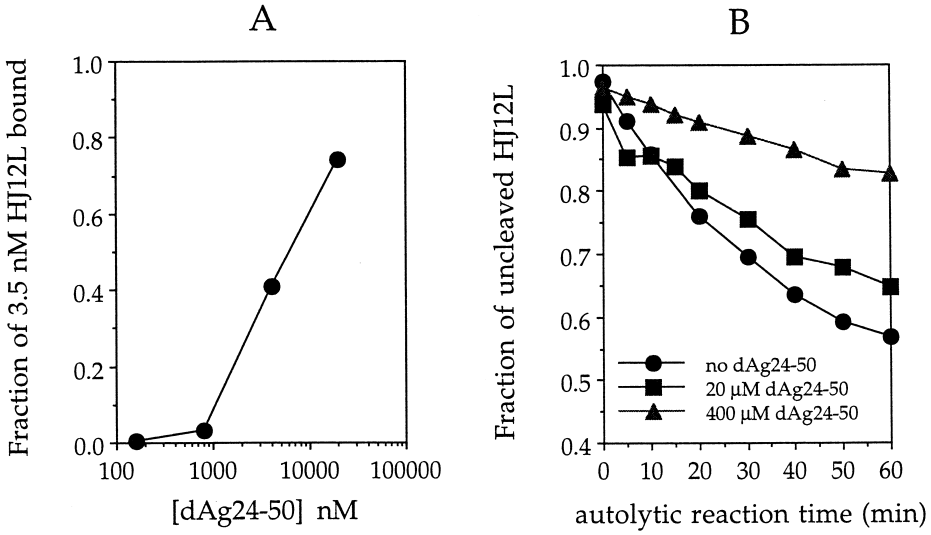

Figure 6. (A) Binding of dAg24-50 to the autolytic domain of HDV genomic sense RNA (HJ12L). The reactions were carried out at 37 °C using a nitrocellulose filter binding assay (Wu et al., 1997). The fraction of HJ12L bound was plotted versus the concentration of delta peptide. (B) Inhibition of the autolytic activity of HJ12L by dAg24-50. HJ12L was incubated with various concentrations of dAg24-50 before the initiation of the autolytic reaction. The reaction was carried out with 12 mM $MgCl_2$ at 37 °C.

(Wu et al., 1997). Since Lys^{38}, Lys^{39}, and Lys^{40} are not involved in the formation of salt bridges with other acidic residues (Figure 5), it is likely that these Lysine residues are involved in the RNA-binding activity.

In conclusion, the solution conformation of the N-terminal leucine-repeat region of HDAg, encompassing residues 24–50, is determined to be an α-helix. The local helix content of this α-helix was analyzed using NOEs and coupling constants. Based on this structural knowledge, Lys^{38}, Lys^{39}, and Lys^{40} are proposed to be the RNA-binding region of HDAg. These results contain significant potential for the development of diagnostic and therapeutic methods for HDV.

Acknowledgements

We would like to thank Dr. Ming F. Tam for mass spectral analysis, Prof. Ping-Chiang Lyu for helpful discussions, Prof. Glenn Millhauser and Dr. Peter Güntert for technical help with the INFIT program. Dr. A. Vyas and P. Rajan are acknowledged for their critical reviewing of this manuscript. The software package XEASY was kindly provided by Prof. K. Wüthrich. This study was carried out on the 600 MHz NMR spectrometer at the Regional Instrument Center at Hsinchu, National Science Council, Taiwan, Republic of China. We thank the National Science Council of the Republic of China for research grants.

References

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **65**, 355–360.

Bergmann, K.F. and Gerin, J.L. (1986) *J. Infect. Dis.*, **154**, 702–705.

Bonino, F., Heermann, K.H., Rizzetto, M. and Gerlich, W.H. (1986) *J. Virol.*, **58**, 945–950.

Bradley, E.K., Thomason, J.F., Cohen, F.E., Kosen, P.A. and Kuntz, I.D. (1990) *J. Mol. Biol.*, **215**, 607–622.

Chang, F.L., Chen, P.J., Tu, S.J., Wang, C.J. and Chen, D.S. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8490–8494.

Chang, M.F., Baker, S.C., Soe, L.H., Kamahora, T., Keck, J.G., Makino, S., Govindarajan, S. and Lai, M.M.C. (1988) *J. Virol.*, **62**, 2403–2410.

Chang, M.F., Sun, C.Y., Chen, C.J. and Chang, S.C. (1993) *J. Virol.*, **67**, 2529–2536.

Chao, M., Hsieh, S.Y. and Taylor, J. (1990) *J. Virol.*, **64**, 5066–5069.

Chao, M., Hsieh, S.Y. and Taylor, J. (1991) *J. Virol.*, **65**, 4057–4062.

Chen, Y.J., Lin, S.J., Tzeng, S.Y., Patel, H., Lyu, P.C. and Cheng, J.W. (1996) *Proteins*, **26**, 465–471.

Cheng, J.W., Cheng, C.C., Lyu, P.C., Chen, S.T. and Lin, T.H. (1996) *Intl. J. Pept. Protein Res.*, **47**, 117–122.

Drobny, G., Pines, A., Sinton, S., Weitkamp, D.P. and Wemmer, D.E. (1979) *Faraday Symp. Chem. Soc.*, **13**, 49–55.

Gowans, E.J., Baroudy, B.M., Negro, F., Ponzetto, A., Purcell, R.H. and Gerin, J.L. (1988) *Virology*, **167**, 274–278.

Hwang, S., Lee, C.Z. and Lai, M.M.C. (1992) *Virology*, **190**, 413–422.

Kessler, H., Müller, A. and Oschkinat, H. (1985) *Magn. Reson. Chem.*, **23**, 844–852.

Kessler, H., Greisinger, C., Lautz, J., Müller, A., van Gunsteren, W.F. and Berendsen, H.J.C. (1988) *J. Am. Chem. Soc.*, **110**, 3393–3396.

Kim, Y. and Prestegard, J.H. (1989) *J. Magn. Reson.*, **84**, 9–13.

Kos, A., Dijkema, R., Arnberg, A.C., van der Merde, P.H. and Schellekens, H. (1986) *Nature*, **323**, 558–560.

Kuo, M.Y.P., Sharmeen, L., Dinter-Gottlieb, G. and Taylor, J. (1988) *J. Virol.*, **62**, 4439–4444.

Kuo, M.Y.P., Chao, M. and Taylor, J. (1989) *J. Virol.*, **63**, 1945–1950.

Lai, M.M.C. (1995) *Annu. Rev. Biochem.*, **64**, 259–286.

Lee, C.Z., Lin, J.H., Chao, M., McKnight, K. and Lai, M.M. (1993) *J. Virol.*, **67**, 2221–2227.

Lin, J.H., Chang, M.F., Baker, S.C., Govindarajan, S., and Lai, M.M.C. (1990) *J. Virol.*, **64**, 4051–4058.

Ludvigsen, S., Andersen, K.V. and Poulsen, F.M. (1991) *J. Mol. Biol.*, **217**, 731–736.

Makino, S., Chang, M.F., Shieh, C.K., Kamahora, T., Vannier, D.M., Govindarajan, S. and Lai, M.M.C. (1987) *Nature*, **329**, 343–346.

Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.

Millhauser, G.L., Stenland, C.J., Bolin, K.A. and van de Ven, F.J.M. (1996) *J. Biomol. NMR*, **7**, 331–334.

Neuhaus, D., Wagner, G., Vasak, M., Kägi, J.H.R., & Wüthrich, K. (1985) *Eur. J. Biochem.*, **151**, 257–273.

Pohl, C., Baroudy, M., Bergmann, K.F., Cote, P.J., Purcell, R.H., Hoofnagle, J. and Gerin, J.L. (1987) *J. Infect. Dis.*, **156**, 622–629.

Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Rozzelle Jr., J.E., Wang, J.G., Wagner, D.S., Erickson, B.W. and Lemon, S.M. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 382–386.

Ryu, W.S., Bayer, M. and Taylor, J. (1992) *J. Virol.*, **66**, 2310–2315.

Sharmeen, L., Kuo, M.Y.P., Dinter-Gottlieb, G. and Taylor, J. (1988) *J. Virol.*, **62**, 2674–2679.

Sharmeen, L., Kuo, M.Y.P. and Taylor, J. (1989) *J. Virol.*, **63**, 1428–1430.

Szyperski, T., Güntert, P., Otting, G. and Wüthrich, K. (1992) *J. Magn. Reson.*, **99**, 552–560.

Tan, R., Chen, L., Buettner, J.A., Hudson, D. and Frankel, A.D. (1993) *Cell*, **73**, 1031–1040.

Wang, K.S., Choo, Q.L., Weiner, A.J., Ou, J.H., Najarian, R.C., Thayer, R.M., Mullenbach, G.T., Denniston, K.J., Gerin, J.L. and Houghton, M. (1986) *Nature*, **323**, 508–514.

Wu, H.N. and Lai, M.M.C. (1990) *Mol. Cell. Biol.*, **10**, 5575–5579.

Wu, H.N., Wang, Y.J., Hung, C.F., Lee, H.J. and Lai, M.M.C. (1992) *J. Mol. Biol.*, **223**, 233–245.

Wu, H.N., Lin, I.J., Lou, Y.C., Pai, M.T. and Cheng, J.W. (1997) in preparation.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley New York, NY.

Xia, Y.P., Yeh, C.T., Ou, J.H. and Lai, M.M.C. (1992) *J. Virol.*, **66**, 914–921.

Zyzik, E., Ponzetto, A., Frozani, B., Hele, C., Heermann, K. and Gerlich, W.H. (1987) In *Hepadnaviruses* (eds., Robinson, W., Kolike, K. and Will, H.), Alan R. Liss, New York, NY, pp. 565–577.