Journal of Biomolecular NMR, 8 (1996) 417–428
ESCOM

J-Bio NMR 391

Sequential assignment of ^{1}H, ^{13}C and ^{15}N resonances of human carbonic anhydrase I by triple-resonance NMR techniques and extensive amino acid-specific ^{15}N-labeling*

Ingmar Sethson[a,**], Ulf Edlund[a], Tadeusz A. Holak[b], Alfred Ross[b] and Bengt-Harald Jonsson[c]

[a]*Department of Organic Chemistry, Umeå University, S-901 87 Umeå, Sweden*
[b]*Max-Planck Institut für Biochemie, Am Klopferspitz 18 A, D-82152 Martinsried bei München, Germany*
[c]*Department of Biochemistry, Umeå University, S-901 87 Umeå, Sweden*

Received 25 March 1996
Accepted 16 August 1996

Keywords: Triple-resonance spectroscopy; NMR resonance assignment; Secondary structure; Amino acid-specific labeling; Amide proton exchange

Summary

The backbone NMR resonances of human carbonic anhydrase I (HCA I) have been assigned. This protein is one of the largest monomeric proteins assigned so far. The assignment was enabled by a combination of 3D triple-resonance experiments and extensive use of amino acid-specific ^{15}N-labeling. The obtained resonance assignment has been used to evaluate the secondary structure elements present in solution. The solution structure appears to be very similar to the crystal structure, although some differences can be observed. Proton–deuteron exchange experiments have shown that the assignments provide probes that can be used in future folding studies of HCA I.

Introduction

The present study concerns human carbonic anhydrase I (HCA I), a globular protein that contains 260 amino acid residues and has a molecular mass of 29 kDa. HCA I is a Zn-containing enzyme, which is most prevalent in red blood cells and catalyzes the conversion of carbon dioxide and water to bicarbonate and a proton. Some of the folding properties of HCA I have been characterized (Carlsson et al., 1973,1975) and its X-ray structure has been resolved to 2.2 Å resolution (Kannan et al., 1975). The protein is composed of a single domain, with partial helical structure and a dominating open β-structure that extends throughout the entire molecule. The cDNA for HCA I can be expressed in *Escherichia coli* using the expression plasmid pHCA I (Engstrand et al., 1995).

To understand the factors that determine the stability of the native state, it is useful to characterize not only the native three-dimensional conformation, but also intermediate forms that are found at equilibrium under various denaturing conditions or when studying folding kinetics. Earlier we have characterized local structure in folding intermediates using an approach in which we introduced spectroscopic and chemical probes in human carbonic anhydrase II, employing site-directed mutagenesis in combination with chemical modification (Mårtensson et al., 1993; Carlsson and Jonsson, 1995; Lindgren et al., 1995). Despite 105 differences in the amino acid sequences of HCA I and HCA II, their secondary structures and general molecular architectures are very similar (Kannan et al., 1975; Eriksson et al., 1988), which makes them attractive targets for comparative studies of their folding mechanisms.

To obtain more detailed information about interactions at individual amino acid residues, NMR can be used to measure H/D exchange at each protonated nitrogen in the protein. A prerequisite for such studies is a complete assignment of backbone atoms.

With a molecular weight of 29 kDa, HCA I is one of the largest monomeric proteins for which a sequential

*The chemical shift data have been deposited in the BioMagResBank in Madison, WI, U.S.A.
**To whom correspondence should be addressed.

assignment has been attempted (Fogh et al., 1994; Remerowski et al., 1994). The use of isotopic labeling in combination with heteronuclear 3D experiments has made the assignment of proteins in the 30 kDa range feasible (Bax and Grzesiek, 1993; Muhandiram and Kay, 1994). However, for slowly tumbling proteins, the large line widths place a restriction on the assignment that can be obtained because of spectral overlap and reduced magnetization transfer efficiency. One way of circumventing this problem has been to use fractionated deuteration to reduce the relaxation rates (Grzesiek et al., 1995; Venters et al., 1995a). In this work we have chosen a different approach, where we have made extensive use of incorporation of selectively ^{15}N-labeled amino acids to alleviate the problems with spectral overlap. Despite the large line widths present in the spectra, it has in this way been possible to obtain a complete sequential assignment of the backbone amide resonances in HCA I along with almost all of the C^{α}, C^{β}, CO and H^{α} resonances.

The obtained assignment, in combination with studies of proton–deuteron exchange, NOEs and chemical shift deviations from random coil values, has been used to identify the secondary structure elements present in the solution structure of HCA I. A good correlation is found when comparing these results with what is observed in the crystal structure (Kannan et al., 1975), although some discrepancies are also evident. A comparison of short interproton distances in the crystal with observed NOEs confirms that the global fold in solution resembles what is observed in the crystal.

Materials and Methods

Protein production and purification

For protein production the *E. coli* strain BL21(DE3) harboring the plasmid pHCA I was grown at 23 °C. Protein production was initiated by adding IPTG to a final concentration of 0.5 mM at a cell density corresponding to $OD_{660} = 0.5$, and then allowing the cells to grow for 6–12 h.

Uniformly ^{15}N-labeled HCA I was produced by allowing the cells to grow in a defined medium consisting of 0.5 g/l (^{15}N)-NH_4Cl, 0.5 g/l NaCl, 6 g/l Na_2HPO_4, 3 g/l KH_2PO_4, 4 g/l glucose, 0.5 mM $ZnSO_4$ and 50 mg/l ampicillin. Protein labeled with both ^{15}N and ^{13}C was produced in an identical medium except that 2 g/l of uniformly ^{13}C-labeled glucose was used as a carbon source.

Incorporation of selectively ^{15}N-labeled amino acids was accomplished by growth in defined media, containing per litre: 125 mg adenine, 125 mg uracil, 125 mg cytosine, 125 mg guanosine, 50 mg thymine, 50 mg thiamine, 0.1 mg biotine, 50 mg nicotinic acid, 5 g glucose, 1 g $^{15}NH_4Cl$, 2 g sodium acetate, 2 g succinic acid, 10 g K_2HPO_4, 500 mg $MgSO_4$, 5 mg $FeSO_4$, 400 mg Ala, 400 mg Glu, 400 mg Gln, 400 mg Arg, 250 mg Asp, 100 mg Asn, 50 mg Cys, 400 mg Gly, 100 mg His, 100 mg Ile, 100 mg Lys, 250 mg Met, 100 mg Pro, 1600 mg Ser, 100 mg Tyr, 100 mg Val, 50 mg Phe, 100 mg Leu, 50 mg Trp and 100 mg carbenicillin. The specifically ^{15}N-labeled amino acid was added at a three- to fivefold higher concentration than stated above. Growth with ^{15}N-Gly in the described media resulted in extensive cross-labeling of serine residues. To suppress the cross-labeling and obtain a protein labeled only in the glycine positions, we modified the growth conditions by adding 1.5 g/l of serine at the induction with IPTG, followed by addition of 1 g/l of serine 3 h later. Amino acid-selective ^{15}N-labeling was performed for the following amino acids: Ala, Gly, His, Ile, Leu, Lys, Phe, Ser, Trp and Val.

The protein was purified to homogeneity in one step with affinity chromatography (Khalifah et al., 1977), and the purified protein was examined by SDS-PAGE. Protein concentrations were determined from the absorbance at 280 nm.

NMR sample preparation

The NMR samples were prepared by dissolving 12 mg of lyophilized HCA I in 500 µl of H_2O/D_2O (95:5) containing 40 mM phosphate buffer, to yield a protein sample of 0.8 mM concentration at a pH of 6.2. Sodium azide was added to the samples to give an NaN_3 concentration of 0.2 mM. The NMR samples for the proton–deuteron exchange were prepared in the same way, except that D_2O was used as the solvent.

NMR spectroscopy

All NMR spectra were recorded at 30 °C. The 3D NOESY-(^{15}N,^{1}H)-HMQC and TOCSY-(^{15}N,^{1}H)-HMQC spectra were recorded on a Bruker AMX-600 spectrometer. The HBHA(CO)NH (Grzesiek and Bax, 1993), HBHA(CBCA)NH (Wang et al., 1994) and all 2D NMR spectra were obtained on a Bruker AMX2-500 spectrometer. The remaining experiments were recorded on a Varian Unity 600 spectrometer. All instruments were equipped with a triple-resonance gradient probe. Quadrature detection in the triple-resonance 3D spectra was obtained by using the States-TPPI method (Marion et al., 1989a), while the TPPI method (Marion and Wüthrich, 1983) was used for the spectra recorded on the ^{15}N-labeled samples. The first data point in the indirect dimensions was delayed by 1/(2*SW) to enable the differentiation of folded and non-folded resonances and to obtain a flat baseline (Bax et al., 1991). When the sensitivity-enhanced gradient-detected echo/anti-echo method was used, the data were transformed into States complex data by appropriate processing (Palmer et al., 1992). The relaxation delay between different scans was set to 1 s in all experiments. The off-resonance ^{13}C pulses in the triple-resonance experiments were applied as phase-shifted laminar pulses (Patt, 1992) and the power and length of the pulses were

TABLE 1
PARAMETERS USED IN THE ACQUISITION OF THE 3D EXPERIMENTS

Experiment	SF (^{1}H) (MHz)	Acquisition time (days)	Nucleus			SW (kHz)			Transmitter frequency (ppm)			No. of (complex) data points		
			t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
CBCA(CO)NH/HNCACB	600	6	C^α/C^β	^{15}N	NH	10.0	4.0	12.5	49.8	126.6	4.7	64	32	512
CBCACO(CA)HA	600	4.5	C^α/C^β	CO	H^α	10	5	6	42.5	42.5[a]	4.7	64	56	512
HNCO	600	3	CO	^{15}N	NH	3	4	12.5	42.0[a]	126.6	4.7	20	40	512
HBHA(CO)NH/ HBHA(CBCA)NH	500	6	H^α/H^β	^{15}N	NH	5.2	2.5	10.0	4.7	113.9	4.7	50	25	512
NOESY-(^{13}C,^{1}H)-HSQC/ NOESY-(^{13}C/^{15}N)-HSQC	600	6	^{1}H	^{13}C (^{15}N)	NH	10.0	3.9	10.0	4.7	49.8 (126.6)	4.7	100	32	512
NOESY-(^{15}N,^{1}H)-HMQC/ TOCSY-(^{15}N,^{1}H)-HMQC	600	6	^{1}H	^{15}N	NH	7.7 (4.8[b])	2.0	4.0	4.7	115.0	7.8	120[c]	40[c]	1024[c]

[a] The CO pulses were applied as shifted laminar pulses to excite the frequencies of the CO resonances.
[b] An SW of 4.8 kHz in t_1 was used in the TOCSY-HMQC experiment.
[c] Real, qseq data points.

adjusted to yield zero excitation in the aliphatic region (Grzesiek and Bax, 1993). All the experiments were performed using the parameters stated in the cited references, apart for the parameters explicitly supplied here. The spectral parameters used in the 3D experiments are given in Table 1.

All spectra were processed using the CCNMR software package, written by Christian Cieslar at the Max-Planck Institut für Biochemie in Martinsried, Germany. The data processing included the use of Lorentzian–Gaussian windows and zero-filling in all dimensions. The residual solvent signal was suppressed without affecting nearby frequencies by means of the Kaarhunen–Loeve transformation (Mitschang et al., 1991). In the three-dimensional experiments, linear prediction was used to prolong the indirectly detected time domains by 50%.

The gradient-enhanced triple-resonance 3D experiments CBCA(CO)NH and HNCACB were collected as described previously (Muhandiram and Kay, 1994).

The CBCACO(CA)HA experiment (Kay, 1993) and the HNCO experiment (Muhandiram and Kay, 1994), where pulsed field gradients were employed to suppress the water magnetization, were recorded as described. The CO pulses were applied as shifted laminar pulses (Patt, 1992) and the CO resonances were extensively folded. A linear phase correction was applied in t_2 to achieve that all CO resonances were folded an even number of times and appeared in the same phase.

The HBHA(CO)NH experiment (Grzesiek and Bax, 1993) and the HBHA(CBCA)NH experiment (Wang et al., 1994) were performed with the following modification: the residual water signal was suppressed by means of two purging pulses (5 ms x-pulse and 3.5 ms y-pulse at a field strength $\gamma\beta_1 = 10$ kHz, followed by a spoil gradient at a time when the magnetization of interest is N_zC_z (prior to the ^{15}N evolution in t_2)).

The 3D TOCSY-(^{15}N,^{1}H)-HMQC (Marion et al., 1989b) and NOESY-(^{15}N,^{1}H)-HMQC (Kay et al., 1989) spectra on the ^{15}N-labeled samples were recorded as described previously. Solvent suppression was achieved by presaturation during the relaxation delay in the TOCSY experiment, while spin-lock pulses in the HMQC part were employed in the NOESY experiment (Messerle et al., 1989). An MLEV mixing sequence at a field strength $\gamma\beta_1 = 10$ kHz was used during the 45 ms mixing time in the TOCSY experiment. In the NOESY experiment a mixing time of 100 ms was used.

The 3D NOESY-(^{13}C,^{1}H)-HSQC experiment (Muhandiram et al., 1993) and the simultaneously ^{13}C-/^{15}N-edited NOESY-HSQC experiment (Pascal et al., 1994) on the ^{13}C-/^{15}N-labeled sample were recorded as described previously. The mixing times used were 200 and 150 ms, respectively. The total acquisition time for each experiment was 6 days.

The 2D (^{15}N,^{1}H)-HSQC spectra on the ^{15}N-labeled samples (uniformly and ^{15}N-labeled in selected amino acids) were obtained with the sensitivity-enhanced heteronuclear gradient echo method, as described previously (Kay, 1993). The spectral widths and the carrier frequency in the respective dimensions were typically set to 3.5 kHz and 113.9 ppm in the ^{15}N dimension and 10 kHz and 4.7 ppm in the acquisition dimension ($^1H^N$). The numbers of points acquired were 64 complex points in t_1 (^{15}N) and 512 complex points in t_2 (^{1}H).

A 2D constant-time (^{13}C,^{1}H)-HSQC was recorded as described previously (Vuister and Bax, 1992), modified by using the sensitivity-enhanced gradient-detected method. The constant time was set to 26.6 ms. The spectral widths and the carrier frequency in the respective dimensions were set to: 12.5 kHz and 42.1 ppm in the ^{13}C dimension and 8 kHz and 4.7 ppm in the acquisition dimension (^{1}H). The numbers of points acquired were 280 complex points in t_1 (^{13}C) and 512 complex points in the acquisition dimension (^{1}H).

Amide proton–deuteron exchange was measured by recording 2D (^{15}N,^{1}H)-HSQC spectra upon solvation of

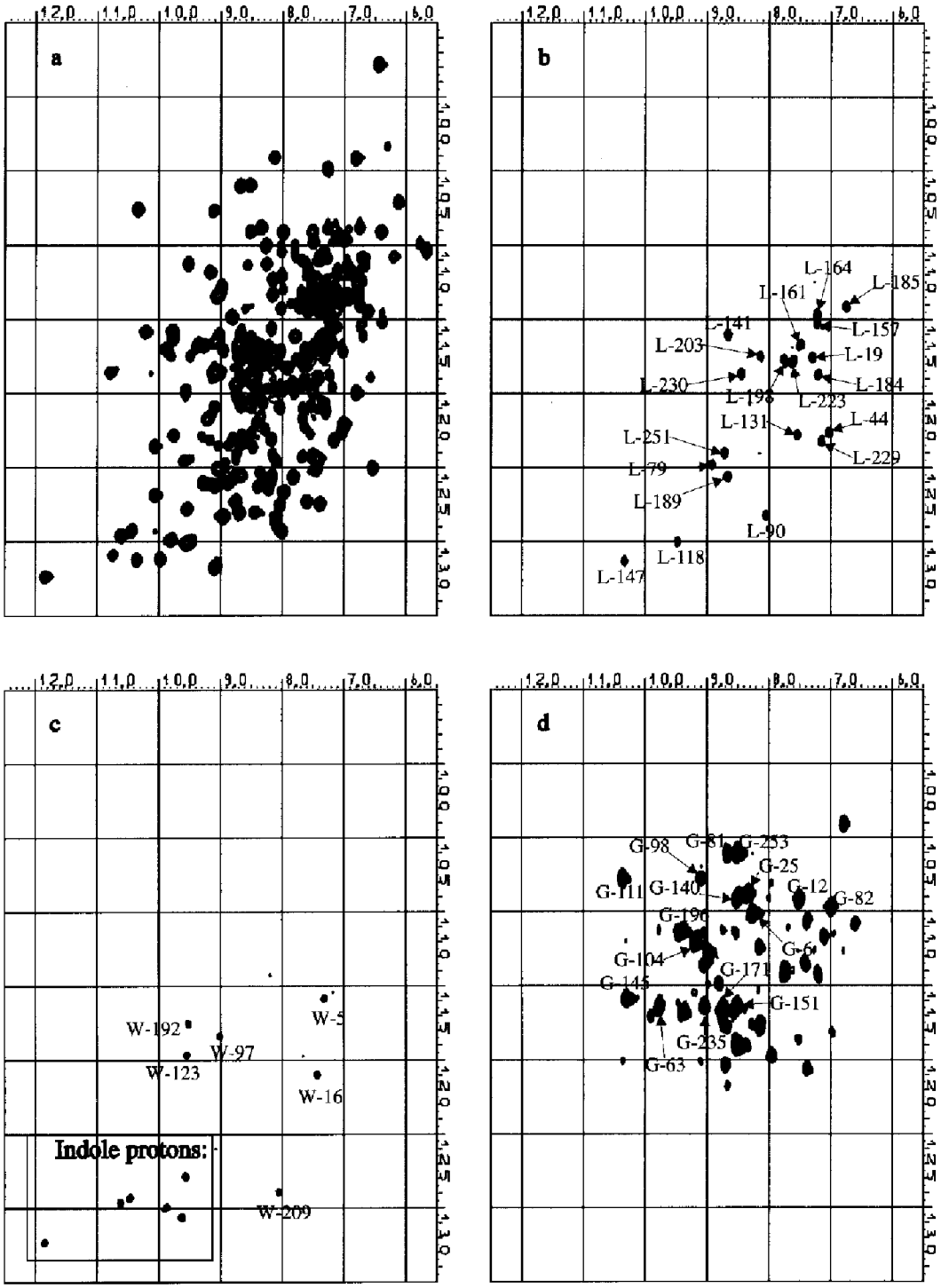

Fig. 1. (^{15}N-^{1}H)-HSQC spectra of ^{15}N-labeled HCA I. (a) Uniformly ^{15}N-labeled HCA I; (b) leucine ^{15}N-labeled HCA I; (c) tryptophan ^{15}N-labeled HCA I; (d) glycine/serine ^{15}N-labeled HCA I.

protonated ^{15}N-labeled samples of HCA I in D_2O. A total of 64 complex points, each made up of eight scans, were collected in t_1, resulting in a total acquisition time of 20 min. The first six experiments were recorded sequentially, whereafter the following time points were collected at times: 3 h, 6 h, 12 h, 24 h, 2 days, 5 days, 11 days, 25 days, 2 months, 6 months and 1 year.

Results and Discussion

Assignment strategy

The 2D (^{1}H-^{15}N)-HSQC spectrum of HCA I (Fig. 1) displayed a good signal-to-noise ratio, although the line widths in the ^{15}N dimension were found to be in the range 25–30 Hz. Unfortunately, it was not possible to reduce

the line widths by recording the NMR spectra at higher temperatures, since above 37 °C the protein gradually denatured. In the corresponding 2D ^{1}H-^{13}C-CT-HSQC experiment (Fig. 2), the sensitivity was hampered by the longer time during which the transverse relaxation was active (a constant time of 26.6 ms was used), but nevertheless a good-quality spectrum could be recorded. However, the short transverse relaxation times had a larger impact on the spectral quality in experiments where the magnetization spent more time in the transverse plane. Accordingly, the 3D (^{15}N-^{1}H)TOCSY-HMQC spectrum was of poor quality and only 40% of the $^{1}H^{N}$-H^{α} cross peaks were observable. Initial attempts to base the sequential assignment exclusively on NMR experiments with ^{15}N-labeled samples of HCA I therefore had to be abandoned. The sensitivity of the triple-resonance experiments performed on the ^{13}C-/^{15}N-labeled sample, utilizing the relatively large one-bond coupling constants for the magnetization transfer, turned out to give good S/N ratios for most of the resonances.

Despite the high degree of chemical shift dispersion, which is typical for proteins containing a substantial fraction of β-structure, the large line widths and the large number of amino acids caused substantial resonance overlap. From the 2D HSQC spectra it is evident that, despite the high digital resolution in these spectra, a number of amide resonances are not resolved. The extent of this frequency degeneracy is such that attempts to achieve the sequential assignment based on the 3D HN(CO)CA and HNCA experiments would fail, apart from some of

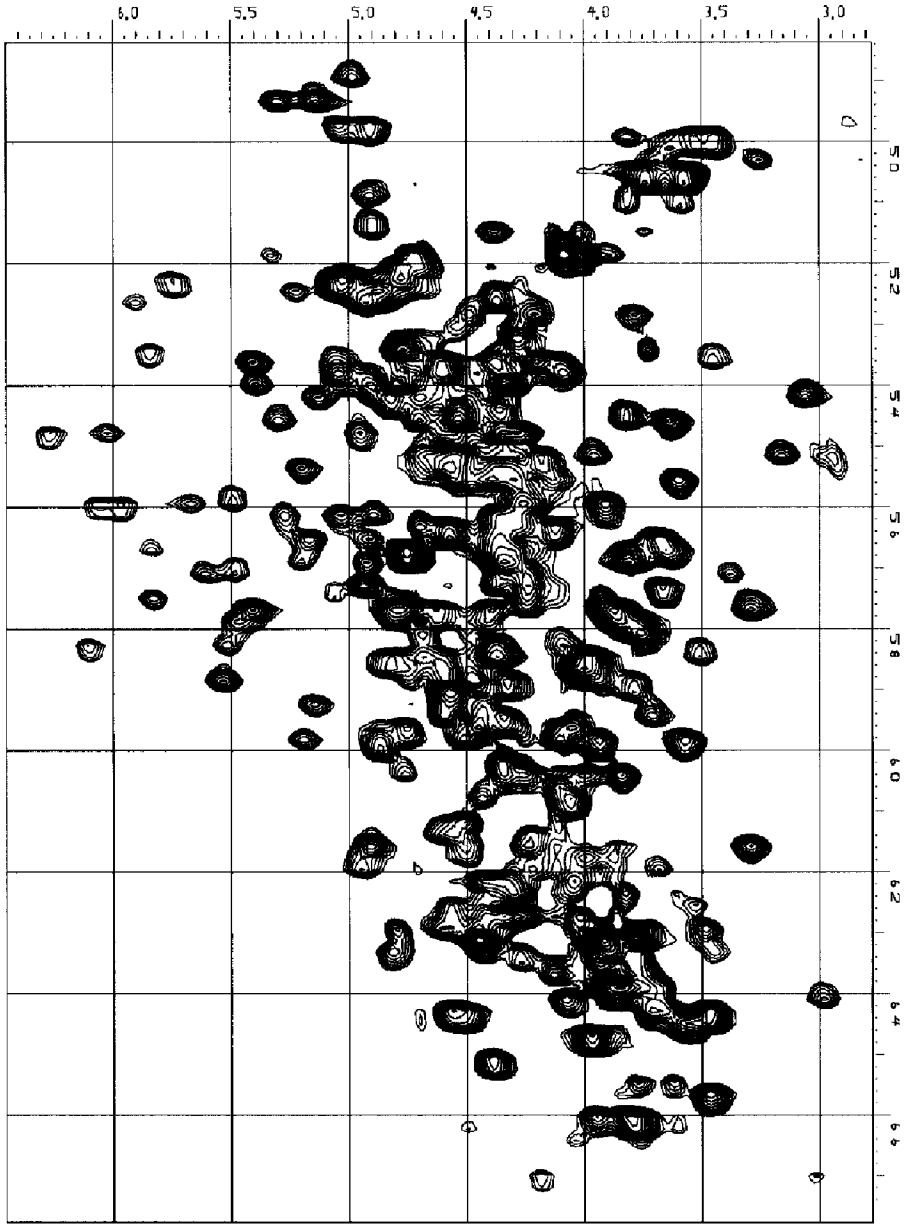

Fig. 2. Gradient-enhanced (^{13}C-^{1}H)-CT-HSQC spectrum of ^{13}C-/^{15}N-labeled HCA I, displaying the H^{α}-C^{α} region.

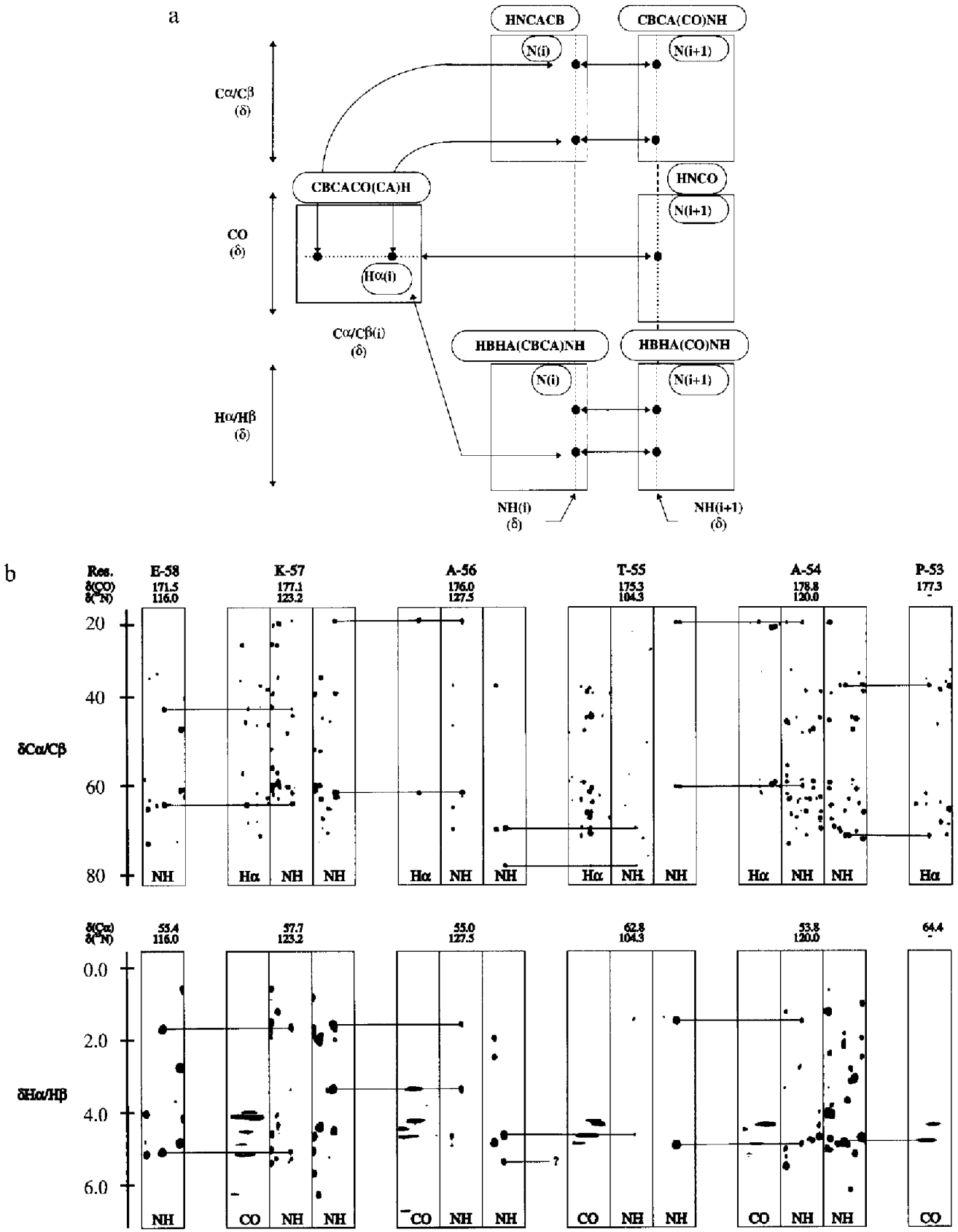

Fig. 3. (a) Schematic picture, illustrating the assignment strategy used. The spectra are 2D planes of the 3D experiments at frequencies indicated by shaded boxes (^{15}N or H^{α}). The double-headed arrows represent correlations used in the assignment procedure. (b) Sequential assignment of amino acids 53–58, using triple-resonance experiments. Upper trace: CBCACO(CA)HA, HNCACB and CBCA(CO)NH; lower trace: CBCACO-(CA)HA, HBHA(CBCA)NH and HBHA(CO)NH.

the better resolved regions of the spectra. Another complicating factor in the sequential assignment procedure is the presence of the 17 proline residues, which interrupts the sequential walk due to the lacking amide protons.

The successful sequential assignment had to be based on a suite of triple-resonance spectra of the double-labeled sample, in combination with extensive ^{15}N-labeling of specific amino acids. It was a prerequisite for the assignment procedure that some of the triple-resonance spectra showed correlations between the NH group of one residue and resonances of the preceding amino acid (i.e., the CBCA(CO)NH, HNCO and HBHA(CO)NH experiments), while the other spectra showed correlations within the same amino acid (i.e., the HNCACB, HBHA(CBCA)-NH and CBCACO(CA)HA experiments). In these experiments (except for the HNCO), ambiguities in the resonance assignment were reduced in the ^{13}C dimension by relayed magnetization transfer between the C^{α} and C^{β} resonances (or alternatively, the H^{α} and H^{β} resonances when the ^{1}H resonances were monitored in t_1). Hence, each amino acid was represented by four frequencies in the 3D spectra, which substantially reduced the amount of spectral overlap. A cartoon displaying how the different triple-resonance spectra were used in combination to enable the resonance assignment, and an example of how this was carried out in practice for one of the better resolved regions in the spectra are shown in Fig. 3.

The spectra that turned out to be the most useful in the assignment procedure were those acquired in the CBCA(CO)NH and the HNCACB experiments (Muhandiram and Kay, 1994). In these spectra most of the expected resonances could be observed, although some resonances were missing or appeared as peaks of very low intensity. This was often observed for those amino acids that had a small B-factor in the crystal structure (Kannan et al., 1975). However, it appeared that only one amino acid was completely unobservable in these triple-resonance spectra, whereas a small number of β-resonances were missing in the HNCACB experiment (less than 5%). In addition, some regions of the spectra displayed extensive resonance overlap, which made the sequential assignment solely based on these spectra very difficult. The various 3D spectra were systematically searched for the presence of common C^{α} and C^{β} frequencies. Typically, two or more possibilities in the search for sequential candidates from one spectrum to the other were observed. These ambiguities could in most cases be resolved when suggested assignments were tested for the correct amino acid type, which was determined from the unique chemical shifts in the C^{α}/C^{β} dimension (Grzesiek and Bax, 1993) or from the 2D ^{1}H-^{15}N-HSQC spectra of selectively ^{15}N-labeled samples. Especially in crowded regions of the spectra, the identification of a resonance by amino acid type required the recording of 2D ^{1}H-^{15}N-HSQC spectra of samples that were ^{15}N-labeled in specific amino acids (Fig. 1). An example of such a region, characterized by substantial resonance overlap in the CBCA(CO)NH spectrum, is displayed in Fig. 4. The enboxed regions A, B and C contain resonances from five (Val^{109}, Lys^{168}, Val^{218}, Leu^{223} and Ala^{224}), three (Lys^{10}, Tyr^{20} and Thr^{42}) and three (Asp^{9}, His^{122} and Asn^{237}) amino acids, respectively. The chemical shift differences in the NH and ^{15}N dimensions between the amino acids in each box are substantially less than the digital resolution ($< \pm 0.02$ ppm in the NH dimension and $< \pm 0.5$ ppm in the ^{15}N dimension). Despite the extent of resonance overlap, the resonances of all these amino acids could be identified by taking advantage of the information from the 2D spectra of the specifically ^{15}N-labeled samples of HCA I. This also facilitated the identification of C^{α} and C^{β} resonances that belonged to the same amino acid. This analysis of the CBCA(CO)NH spectrum was crucial for the assignment of the resonances in the HNCACB spectrum, where the extent of the resonance overlap is even larger, since both intraresidual and sequential correlations are observed.

The 2D spectra of the selectively ^{15}N-labeled samples also turned out to be useful to detect the coexistence of the Zn-bound protein and the apo-form. When the 2D spectra of specifically labeled samples were inspected, it turned out that some amino acids existed in one major and one minor conformation (80:20). By the addition of $ZnSO_4$ to these samples, the minor peaks disappeared, which showed that this conformational complexity was caused by the dialysis towards EDTA in the purification procedure (vide supra). This observation of coexisting conformations was made during the final stages of the assignment, after all the triple-resonance spectra had been recorded, and caused additional problems during the assignment procedure. However, after including this information, all of the previously identified but unassigned resonances in the spectra could be accounted for.

Moreover, the existence of unique di- and tripeptide fragments (like Gly^{81}–G^{82}, Ala^{134}–Ala^{135}, etc.) enabled the location of stretches of sequential amino acids in the amino acid sequence of HCA I. In this way it was possible to sequentially assign about 90% of the backbone atoms in HCA I, where the consistency was checked by the correct type of amino acid in each position as judged by the specifically labeled samples and the chemical shifts. Most of the remaining resonances in the spectra could be assigned when additional triple-resonance assignments were included in the analysis, i.e., the HNCO, CBCACO-(CA)HA, HBHA(CBCA)NH and HBHA(CO)NH experiments. The inclusion of these spectra in the analysis also confirmed the obtained assignment for 90% of the hitherto assigned amino acids and provided the H^{α}, H^{β} and CO frequencies.

The only amino acid that could not be assigned on the basis of the triple-resonance experiments was Thr^{199}, since its resonances were not observable in any of these spectra.

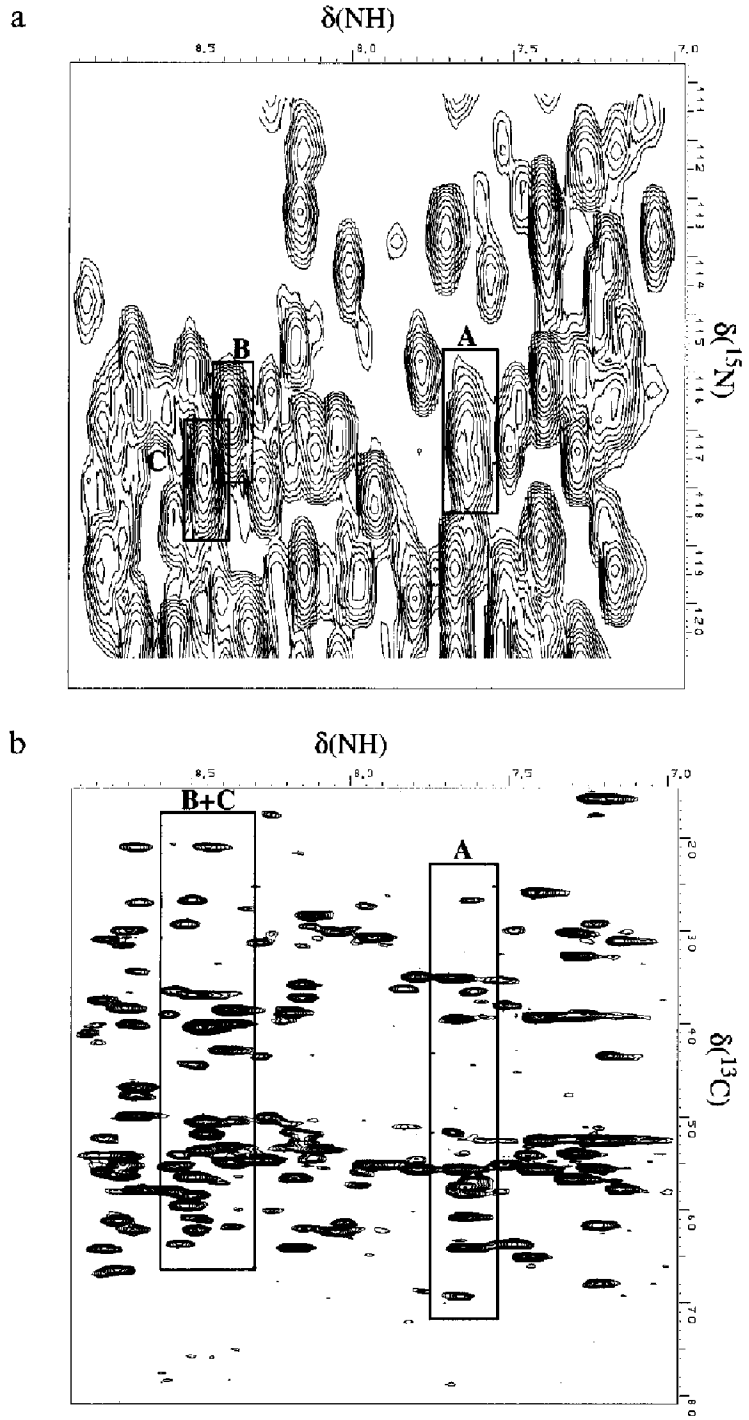

Fig. 4. 2D projections of the CBCA(CO)NH spectrum. Top: projection of the complete ^{13}C dimension onto the NH-^{15}N plane. Bottom: projection of the ^{15}N dimension between 115.5–119 ppm onto the NH-^{13}C plane. The boxes represent regions of resonance degeneracy as described in the text.

However, since both His^{200} and Leu^{198} had been assigned it was possible to search the (^{15}N-^{1}H)-NOESY-HSQC spectrum for candidates showing sequential NOEs to these residues. One of the amide resonances in this spectrum, which was not observed in the triple-resonance spectra, showed $d_{\alpha N}(i+1)$ NOEs to His^{200} and Leu^{198} and was accordingly assigned to Thr^{199}. The observation that the neighboring amino acid Leu^{198} also appeared as a very weak peak in the triple-resonance spectra suggests that the mobility of this part of the protein is extensively restricted.

Another amino acid that was difficult to assign was Ser^{29}. In the crystal structure, its side chain is pointing towards the interior of the protein with the hydroxyl group involved in hydrogen bonding. This gives rise to a very small crystallographic B-factor for this amino acid.

The (i+1) neighbor to Ser^{29} is Pro^{30}, and any correlation between these amino acids is unobservable in the triple-resonance spectra. The only sequential (i+1) candidate to Gln^{28} found in the CBCA(CO)NH spectrum that matched with a serine appeared as weak peaks and was partially overlapping with other resonances. Interestingly, peaks at the corresponding C^{α} and C^{β} resonances were found in the CBCA(CO)HA experiment at two H^{α} frequencies, while having the same CO frequency, and were impossible to assign to any other serine residue. Possible explanations for the appearance of serine in one major and one minor conformation (the peak intensities are 80:20) are that Pro^{30} exists in both a cis and a trans conformation, or that it is the apo-form of the protein that gives rise to the minor conformation (vide supra).

Finally, three amino acids, located between residues 244–246, were assigned only at the final stage of the assignment procedure, since their resonances were weak and showed substantial resonance overlap. Furthermore, some of the C^{β} peaks were missing in the spectra. Nevertheless, their chemical shifts and sequential correlations of the observable peaks support the suggested assignment. On the other hand, the CSI analysis (vide infra) suggests that these three amino acids should constitue a β-structure, which was not observed in the crystal structure. Hence, the confirmation of the assignment of these three residues has to await a more rigorous NOE analysis.

Sequential assignment data have been submitted to the BioMagResBank in Madison, WI, U.S.A., where all the amide-^{15}N and $^{1}H^{N}$ resonances and almost all of the other backbone nuclei have been assigned. We were unable to assign 18 CO, 1 C^{α}, 3 C^{β} and 5 H^{α} resonances. Additionally, about 70% of the H^{β} protons were assigned, but without any attempt to obtain a stereospecific assignment. The unassigned resonances were either caused by missing resonances or by ambiguities due to resonance overlap. The reason why so many CO resonances are left unassigned is that the CBCACO(CA)HA experiment was set up with delays optimized for the detection of CH groups, which makes all the glycines unobservable. The guanidino protons of the arginine side chains, the indole protons of the tryptophans and the imidazole protons of the histidines (only four observable) have also been identified, but their assignment has to await the recording of triple-resonance experiments designed to observe correlations where these resonances are involved.

Secondary structure identification

In the ^{15}N- or ^{13}C-edited 3D NOESY-HSQC spectra the number of unique correlations was not sufficient to allow a comprehensive evaluation of secondary structures based on NOEs. However, it has been shown that the deviations of the chemical shifts from the 'random coil' values of H^{α}, C^{α}, C^{β} and CO resonances are indicative of what secondary structure they are located in (Wishart and Sykes, 1994). Hence, the secondary structure elements in the solution structure of HCA I were evaluated by the chemical shift index method (CSI) of Wishart and Sykes

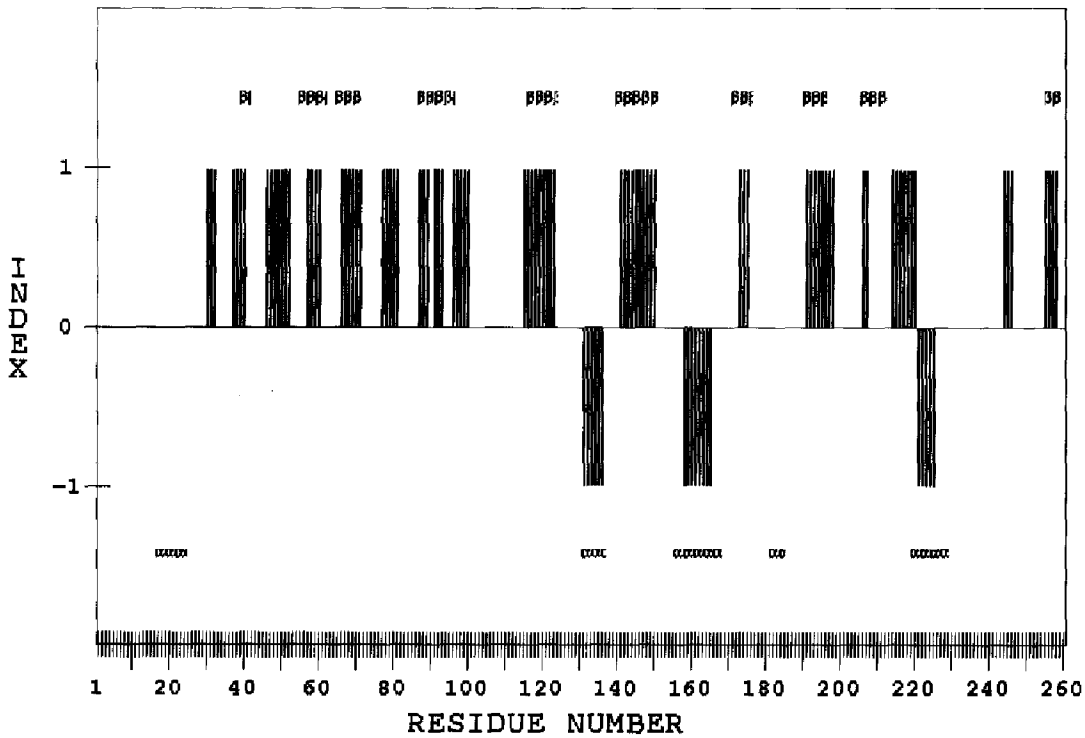

Fig. 5. Secondary structure elements in HCA I, as suggested by the CSI analysis. −1 = α-helix ; +1 = β-sheet; ααα and βββ indicate secondary structure elements found in the crystal structure.

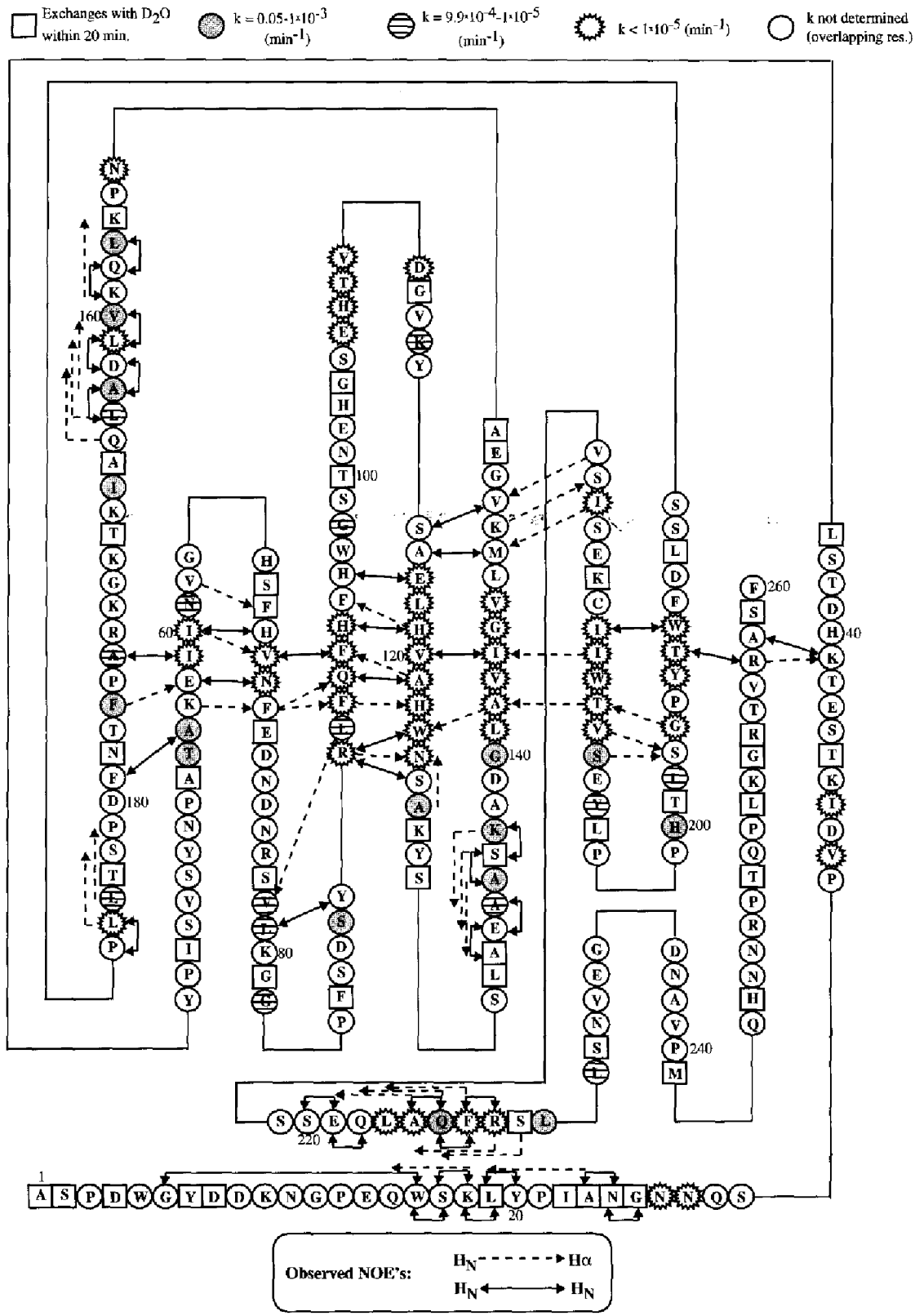

Fig. 6. Proton–deuteron exchange rates and NOEs observed in HCA I.

as incorporated in their program CSI (Wishart and Sykes, 1994). The result of this evaluation is shown in Fig. 5, along with the secondary structure elements observed in the crystal (Kannan et al., 1975). The CSI method predicts all the 10 β-sheets observed in the crystal, although the beginning and end of the β-sheets display some dis-

crepancies. Additionally, the CSI also predicts four β-structures (between residues 30–32, 46–52, 214–220 and 244–246), which in the crystal structure do not appear as secondary structure elements. Three of the α-helices are correctly predicted, whereas the helices between 16–20, 20–25 and 181–185 are suggested to be absent. Whether these observed differences should be attributed to the robustness of the CSI or if they represent true differences between the structure of HCA I in solution and in the solid state was further evaluated by qualitative analysis of NOEs and amide proton–deuteron exchange rates.

The ^{15}N- and ^{13}C-edited 3D NOESY-HSQC spectra were searched for NOEs corresponding to short interproton distances within α-helices and between different β-sheets. The criterion used for accepting a NOE peak as representing a unique interproton distance was that it appeared as a symmetry-related cross peak in the NOESY spectra. That is, a $^1H^N$ / ^{15}N-H^α peak in the ^{15}N-edited 3D NOESY-HSQC spectrum should have a related H^α / ^{13}C-$^1H^\chi$ peak of similar intensity in the ^{13}C-edited 3D NOESY-HSQC spectrum, and likewise for amide–amide NOEs in the ^{15}N-edited 3D NOESY-HSQC spectrum. Moreover, only those peaks resolved from other NOEs, representing sequential and intraresidual correlations, were accepted in this analysis. The NOEs identified in this way are shown in Fig. 6. The results show that all the β-structures are oriented in the same way relative to each other, both in the crystal and in solution. The same qualitative result has also been obtained for the isozyme HCA II (Venters et al., 1995b). The α-helices suggested from the CSI are confirmed, but the NOEs supply additional evidence for the existence of the helix between residues 181–185. Interestingly, the NOE analysis supports the CSI suggestion that the helices between residues 16–25 are absent, because the NOEs representing the $d_{\alpha N}(i + 3)$ distances are missing and, if existing, these peaks would be resolved and observable. This does not necessarily imply that the N-terminus is unstructured. On the contrary, NOEs between residues 17/23 and 6/16 are observable, which represent short interproton distances also present in the crystal structure. Hence, it appears that the solution structure has a defined structure in this region, but exhibits some distinct differences in the local structure.

The amide proton–deuteron exchange experiments reveal that the amide protons are extremely well protected from exchange, since roughly 25% of the amides are still fully protonated after having spent one year in D_2O at pD 8.3. Another 25% of the amide protons are exchanging slowly and exhibit mean lifetimes ranging from 15 min to 100 days, while the remaining 50% are exchanged within the dead time of the experiments, i.e., 20 min. Unfortunately, it has not been possible to unambiguously assign all of the slowly exchanging amide protons. This uncertainty was caused by spectral overlap, in combination with small chemical shift differences observed for some resolved resonances when comparing spectra recorded in H_2O and D_2O (the cause for these discrepancies is currently under investigation). Of the instantly (within the experimental dead time) exchanging amide protons, 49 have been identified. In addition, we have identified 79 slowly exchanging resonances and their exchange rates are shown in Fig. 6. These exchange rates further supported the secondary structure analysis (vide supra), since these amino acids are mainly located in structures where slow exchange is expected. The fact that slowly exchanging amide protons are observed for all secondary structures will allow future studies of folding kinetics and accordingly supply information about the folding pathway of HCA I. Another interesting observation concerns the amino acids which, according to the crystal structure, should be involved in hydrogen bonds in helices but for which the CSI does not indicate any secondary structure. Those protons are observed as slowly exchanging in the α-helix located between residues 181–185, whereas they are found to be rapidly exchanging in the α-helix located between amino acids 20–25. Accordingly, the observed proton–deuteron exchange rates also indicate that the N-terminal helices are absent in the solution structure.

Studies of the exchange as a function of pH and denaturing agent (guanidinium hydrochloride) are currently in progress and will yield information about relative stabilities in different regions of the protein (A. Kjellsson, B.-H. Jonsson and I. Sethson, to be published).

Conclusions

In this work it has been demonstrated that it is possible to obtain resonance assignments for a large, monomeric protein with unfavorable line widths. Prerequisites for such an assignment are the use of a suite of 3D triple-resonance experiments in combination with ^{15}N-labeling of specific amino acids. CSI, NOE and amide proton exchange studies confirm that the solution structure has great similarities with the crystal structure. Some observed distinct differences will make a comparison between a future solution structure and the crystal structure interesting. The identification of slowly exchanging amide protons in almost all secondary structure elements in HCA I will permit further kinetic studies to be done in order to characterize the folding pathway.

Acknowledgements

We thank the Swedish NMR centre for the use of their Varian Unity-600 spectrometer and Dr. Johan Kördel and Dr. Lotta Johansson for their kind assistance. Prof. Lewis Kay is acknowledged for supplying the code for the gradient-selected sensitivity-enhanced triple-resonance pulse programs. Katarina Wallgren is acknowl-

edged for technical assistance in the protein preparation and purification. Grants from the Swedish Natural Science Research Council to Ingmar Sethson, Ulf Edlund and Bengt-Harald Jonsson and from the Deutsche Forschungsgemeinschaft to Tadeusz. A. Holak are acknowledged. Dr. Michael Sharp is acknowledged for linguistic revision.

References

Bax, A., Ikura, M., Kay, L.E. and Zhu, G. (1991) *J. Magn. Reson.*, **91**, 174–178.

Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.*, **26**, 131–138.

Carlsson, U., Henderson, L.E. and Lindskog, S. (1973) *Biochim. Biophys. Acta*, **310**, 376–387.

Carlsson, U., Aasa, R., Henderson, L.E., Jonsson, B.-H. and Lindskog, S. (1975) *Eur. J. Biochem.*, **52**, 25–36.

Carlsson, U. and Jonsson, B.-H. (1995) *Curr. Opin. Struct. Biol.*, **5**, 482–487.

Engstrand, C., Jonsson, B.-H. and Lindskog, S. (1995) *Eur. J. Biochem.*, **229**, 696–702.

Eriksson, A.E., Jones, T.A. and Liljas, A. (1988) *Proteins Struct. Funct. Genet.*, **4**, 274–282.

Fogh, R.H., Schipper, D., Boelens, R. and Kaptein, R. (1994) *J. Biomol. NMR*, **4**, 123–128.

Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.

Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J.D. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 9594–9595.

Kannan, K.K., Notstrand, B., Fridborg, K., Lövgren, S., Ohlsson, A. and Petef, M. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 51–55.

Kay, L.E., Marion, D. and Bax, A. (1989) *J. Magn. Reson.*, **84**, 72–84.

Kay, L.E. (1993) *J. Am. Chem. Soc.*, **115**, 2055–2057.

Khalifah, R.G., Strader, D.J., Bryant, S.H. and Gibbons, S.M. (1977) *Biochemistry*, **28**, 4914–4922.

Lindgren, M., Svensson, M., Freskgård, P.-O., Carlsson, U., Jonasson, P., Mårtensson, L.-G. and Jonsson, B.-H. (1995) *Biophys. J.*, **69**, 202–213.

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

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989a) *J. Magn. Reson.*, **85**, 393–399.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989b) *Biochemistry*, **28**, 6150–6156.

Mårtensson, L.-G., Jonsson, B.-H., Freskgård, P.-O., Kilgren, A., Svensson, M. and Carlsson, U. (1993) *Biochemistry*, **32**, 224–231.

Messerle, B.A., Wider, G., Otting, G., Weber, C. and Wüthrich, K. (1989) *J. Magn. Reson.*, **85**, 608–613.

Mitschang, L., Cieslar, C., Holak, T.A. and Oschkinat, H. (1991) *J. Magn. Reson.*, **92**, 208–217.

Muhandiram, D.R., Farrow, N.A., Xu, G.Y., Smallcombe, S.H. and Kay, L.E. (1993) *J. Magn. Reson.*, **B102**, 317–321.

Muhandiram, D.R. and Kay, L.E. (1994) *J. Magn. Reson.*, **B103**, 203–216.

Palmer III, A.G., Cavanagh, J., Byrd, R.A. and Rance, M. (1992) *J. Magn. Reson.*, **96**, 416–424.

Pascal, S.M., Muhandiram, D.R., Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1994) *J. Magn. Reson.*, **B103**, 197–201.

Patt, S.L. (1992) *J. Magn. Reson.*, **96**, 94–102.

Remerowski, M.L., Domke, T., Groenewegen, A., Pepermans, H.A.M., Hilbers, C.W. and Van de Ven, F.J.M. (1994) *J. Biomol. NMR*, **4**, 257–278.

Venters, R.A., Huang, C.-C., Farmer II, B.T., Trolard, R., Spicer, L.D. and Fierke, C.A. (1995a) *J. Biomol. NMR*, **5**, 339–344.

Venters, R.A., Metzler, W.J., Spicer, L.D., Mueller, L. and Farmer II, B.T. (1995b) *J. Am. Chem. Soc.*, **117**, 9592–9593.

Vuister, G.W. and Bax, A. (1992) *J. Magn. Reson.*, **98**, 428–435.

Wang, A.C., Lodi, P.J., Qin, J., Vuister, G.W., Gronenborn, A.M. and Clore, M. (1994) *J. Magn. Reson.*, **B105**, 196–198.

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.