NMR solution structure of the activation domain of human procarboxypeptidase A2

M. ANGELES JIMÉNEZ,[1] VIRTUDES VILLEGAS,[2] JORGE SANTORO,[1] LUIS SERRANO,[3] JOSEP VENDRELL,[2] FRANCESC X. AVILÉS,[2] AND MANUEL RICO[1]

[1]Instituto de Química-Física Rocasolano. C.S.I.C., Serrano, 119, 28006 Madrid, Spain
[2]Institut de Biotecnología i Biomedicina i Departament de Bioquímica i Biologia Molecular, Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
[3]European Molecular Biology Laboratory, Heidelberg D-69012, Germany

(RECEIVED August 9, 2002; FINAL REVISION October 21, 2002; ACCEPTED October 31, 2002)

Abstract

The activation domain of human procarboxypeptidase A2, ADA2h, is an 81-residue globular domain released during the proteolytic activation of the proenzyme. The role of this and other similar domains as assistants of the correct folding of the enzyme is not fully understood. The folding pathway of ADA2h was characterized previously, and it was also observed that under certain conditions it may convert into amyloid fibrils in vitro. To gain insight into these processes, a detailed description of its three-dimensional structure in aqueous solution is required so that eventual changes could be properly monitored. A complete assignment of the ^{1}H and ^{15}N resonances of ADA2h was performed, and the solution structure, as derived from a set of 1688 nonredundant constraints, is very well defined (pairwise backbone RMSD = 0.67 ± 0.17 Å for residues 10–80). The structure is composed of two antiparallel α-helices comprising residues 19–32 and 58–69 packed on the same side of a three-stranded β-sheet spanning residues 10–15, 50–55, and 73–75. The global fold for the isolated human A2 activation domain is very similar to that of porcine carboxypeptidase B, as well as to the structure of the domain in the crystal of the intact human proenzyme. The observed structural differences relative to the intact human proenzyme are located at the interface between the activation domain and the enzyme and can be related with the activation mechanism. The backbone amide proton exchange behavior of ADA2h was also examined. The global free energy of unfolding obtained from exchange data of the most protected amide protons at pH 7.0 and 298K is 4.9 ± 0.3 kcal.mole^{-1}, in good agreement with the values determined by thermal or denaturant unfolding studies.

Keywords: Procarboxypeptidase; activation mechanism; protein structure; protein stability; hydrogen/deuterium exchange; nuclear magnetic resonance

Carboxypeptidases are secreted from the pancreas in the form of inactive precursors known as procarboxypeptidases. Limited proteolysis on the procarboxypeptidases releases the enzymatically active carboxypeptidase and the activation segment, which inhibits enzymatic activity within the proenzyme (Vendrell et al. 1990; Villegas et al. 1995b; Reverter et al. 1998). The activation segment of procarboxypeptidases, typically spanning 94–96 residues, comprises an N-terminal globular domain of 81 residues known as the activation domain. Pancreatic carboxypeptidases have been classified as A1, A2, and B, depending on their substrate specificity (aliphatic, aromatic, or basic residues, respectively). Although the crystal three-dimensional structure of at least one member of the three types of procarboxypeptidases has been determined (Coll

Reprint requests to: Manuel Rico, Instituto de Química-Física Rocasolano, C.S.I.C., Serrano, 119, 28006 Madrid, Spain; e-mail: mrico@iqfr.csic.es; fax: 91-564-2431.

Abbreviations: ADA2h, activation domain of human procarboxypeptidase A2; ADBp, activation domain of porcine procarboxypeptidase B; CPA2h, human carboxypeptidase A2; CPBp, porcine carboxypeptidase B; PCPA2h, human procarboxypeptidase A2; PCPBp, porcine procarboxypeptidase B.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.0227303.

et al. 1991; Guasch et al. 1992; García-Sáez et al. 1997), the only activation domain for which the three-dimensional structure in solution is known is that of the porcine B form (Vendrell et al. 1991). The solution structure of the domain is well defined for residues 11–76. It consists of two α-helices packed on the same side of a four-stranded antiparallel β-sheet and is very similar to the structure found in the crystal of the intact procarboxypeptidase.

There is increasing evidence for the involvement of prosegments in the folding of several proteins. They are likely to act as intramolecular chaperones by decreasing the energetic barrier corresponding to the transition state in the folding process, which is probably the case in procarboxypeptidases (Vendrell et al. 2000). The folding of the activation domain of human procarboxypeptidase A2 (ADA2h) has been studied extensively by our group. It fits perfectly to a two-state transition, clearly showing the absence of any kinetic intermediate at neutral pH (Villegas et al. 1995a). ADA2h has also been used to develop methods to improve protein stability, and it was one of the first proteins to be stabilized by a rational redesign of its α-helices (Villegas et al. 1995c), which yielded a highly thermostable domain (Viguera et al. 1996) with an extremely fast folding rate. Moreover, it was also shown that the folding reaction of ADA2h follows a nucleation-condensation model (Villegas et al. 1998), and structural information about the transition state of the reaction was obtained by using protein engineering methods. The transition state for the folding of ADA2h is quite compact, maintains some secondary structure, and embodies a hydrophobic core in the process of being consolidated. The folding nucleus is formed by the packing of helix 2 over the two central β-strands, whereas the two strands at the edge of the final structure as well as part of helix 1 appear to be completely unfolded. The amyloid-like aggregating behavior of ADA2h at acidic pH was recently described (Villegas et al. 2000), and it was shown that the stabilization of its two α-helices impedes this behavior.

Much interest has recently focused on the capability of several proteins to form amyloid fibrils. These highly ordered aggregates share a characteristic cross-β structure, suggesting that the key elements of the fibrillation process may be common to all amyloidogenic proteins. Protein models for which the folding pathway has been well characterized could shed some light on this process. The main purpose of our present study was to determine the solution structure of the human activation domain of procarboxypeptidase A2, by using ^{1}H and ^{15}N NMR spectroscopy. The resonance assignments obtained, together with the solution structure, constitute a basic tool to monitor natural or induced conformational changes and crucial information for any further work on that line.

Results

Resonance assignments

^{1}H and ^{15}N NMR assignments were performed by the standard sequence-specific method (Wüthrich 1986) on the basis of heteronuclear 2D and 3D spectra of the uniformly ^{15}N-labeled ADA2h obtained in H_2O, and homonuclear 2D spectra of the nonlabeled protein recorded in H_2O and D_2O at 25°C, 20 mM phosphate, pH 7.0. The 2D ^{1}H-^{15}N HSQC spectrum of ADA2h (Fig. 1) shows an excellent dispersion of cross-resonances. Excluding the N-terminal residue and the four Pro residues, there are 76 potentially observable backbone NH cross-peaks in the ^{1}H-^{15}N-HSQC spectrum of ADA2h. All of them are observed except those corresponding to residues R2 and H34. All of the expected NH_2 sidechain cross-peaks, 22 in total for three Asn and eight Gln residues, were present and readily assigned to their corresponding spin systems after sequential assignment. In addition, cross-peaks corresponding to the indole $N_{\varepsilon}H$ of W40 and one Arg side chain $N_{\varepsilon}H$ (R53), out of the two expected, were identified. Spin systems were classified into different types by joint analysis of 3D ^{1}H-^{15}N-HSQC total correlated spectroscopy (TOCSY) and 2D homonuclear TOCSY. Only three residues (E5, D10, and K63) display a severe overlap at the 3D spectra due to the coincidence of the δ-values of both their amide proton and nitrogen. The presence of three spin systems at the same HSQC cross-peak was determined by the different δ-values of their $C_{\alpha}H$ protons and was facilitated by the fact that they belong to different spin system types.

After we classified most of the spin systems, they were sequentially connected throughout $d_{\alpha N(i,i+1)}$, $d_{NN(i,i+1)}$, and/or $d_{\beta N(i,i+1)}$ nuclear Overhauser enhancements (NOEs) observed at the 3D ^{1}H-^{15}N-HSQC nuclear Overhauser enhancement spectroscopy (NOESY) so as to define stretches of residues that could match unambiguously unique segments in the ADA2h sequence. The whole stretch of sequential connectivities at the 3D spectrum was only broken at the four Pro residues and at H34 (Fig. 2). The four Pro spin systems were first identified in the 2D correlated spectroscopy (COSY) and TOCSY spectra. The residues following the prolines were then identified by the presence of a Pro $C_{\alpha}H_i$-NH_{i+1} cross-correlation in the corresponding 3D NOESY strip. The $d_{\alpha\delta(i,i+1)}$ and/or $d_{\alpha\delta'(i,i+1)}$ NOEs between the Pro preceding residues and the Pro residues themselves observed at the 2D NOESY spectra indicate that the four Pro are in a *trans* conformation. With the exception of the NH amide proton, all other resonances of the His 34 spin system were identified in the 2D spectra. Assignment of aromatic ring spin systems was based on homonuclear 2D COSY, TOCSY, and NOESY spectra and a ^{1}H-^{13}C-HSQC spectrum recorded on the nonlabeled ADA2h in D_2O at natural ^{13}C abundance. The chemical shift assignments have

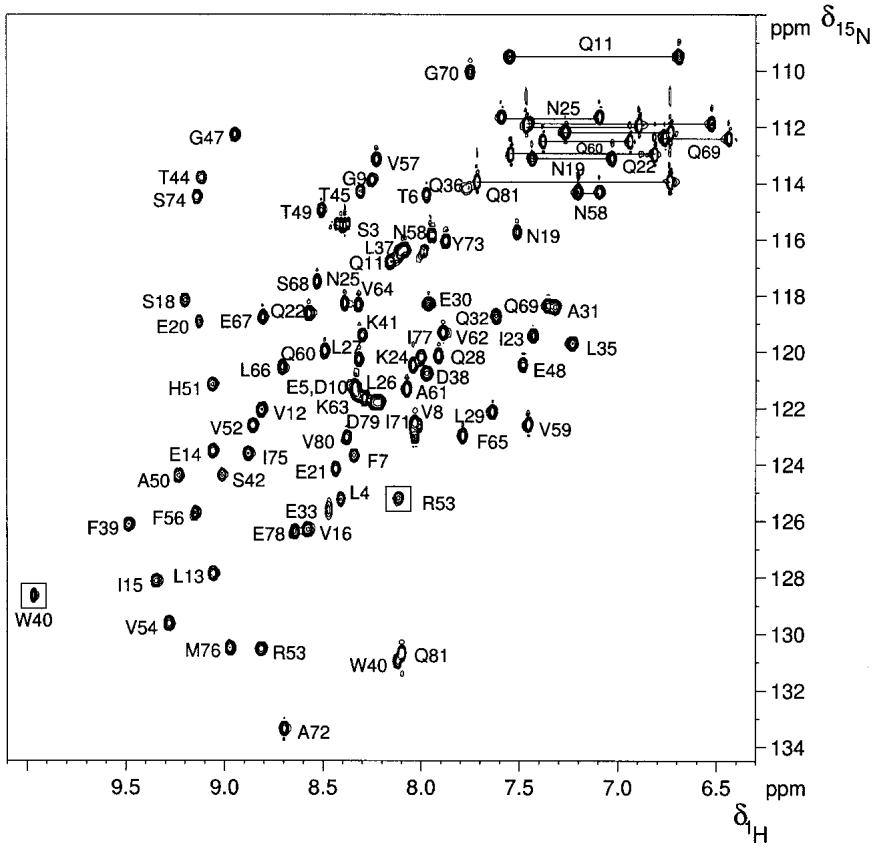

Figure 1. 2D ^{1}H-^{15}N-HSQC spectrum of uniformly ^{15}N-labeled ADA2h. Solution conditions: ADA2h at 2 mM, in 50 mM phosphate buffer pH 7.0, in H_2O/D_2O 9:1, at 25°C.

been deposited at BioMagResBank (http://www.bmrb.wisc.edu/); accession number 5561.

Stereospecific assignments of Val and Leu methyl groups, $C_\beta H$-C_β·H protons, and Pro $C_\delta H$-C_δ·H were performed on the basis of NOE intensities, $^{3}J_{\alpha\beta}$ coupling constants, and their compatibility with the preliminary structure calculations. Only those stereospecific assignments that presented no ambiguity were introduced in the final structure calculation. These were the methyl groups of V16, L29, and V64, the $C_\beta H$ and C_β·H of L29, E30, Q32, L35, H51, F56, N58, F65, and L66, and the $C_\delta H$ and C_δ·H protons of P17 and P43.

Secondary structure analysis

The elements of secondary structure along ADA2h were first delineated using information provided by NOE data and $C_\alpha H$ conformational shifts. The two stretches of strong sequential $d_{NN(i,i+1)}$ NOEs, weak or absent $d_{\alpha N(i,i+1)}$ NOEs, medium-range $d_{\alpha N(i,i+3)}$, $d_{\alpha N(i,i+4)}$, and $d_{\alpha\beta(i,i+3)}$ NOEs, and large upfield conformational shifts indicate the presence of two α-helices spanning approximately residues 19–32 and 58–69 (Fig. 2), in agreement with the helix lengths determined from the ϕ,ψ angles at the final calculated structures (see below). A four-stranded antiparallel β-sheet was identified by long-range $d_{\alpha\alpha(i,j)}$, $d_{\alpha N(i,j)}$, and $d_{NN(i,j)}$ NOEs, strong sequential $d_{\alpha N(i,i+1)}$ NOEs with weak or absent $d_{NN(i,i+1)}$ NOEs, and large positive $C_\alpha H$ conformational shifts (Figs. 2,3). Whereas the two inner strands (β1 and β3) are unambiguously determined, only one of the residues in the edge β2 strand presents a main-chain NOE which allows us to align it with the β3 strand, and only one residue (Y73) of the other edge strand (β4) shows the large positive $C_\alpha H$ conformational shift characteristic of β-strands. In the calculated structures (see below), the β2 strand is not formed.

Three-dimensional structure

A first set of distance restraints for the structure calculation was obtained from those cross-peaks in the 2D NOESY spectra of ADA2h in H_2O and in D_2O that could be unambiguously assigned. These constraints were used to perform an initial DYANA structure calculation (Güntert et al. 1997). New NOE assignments were then made on the basis of the previous calculated structures by using the NOAH strategy and DYANA structure calculations performed it-

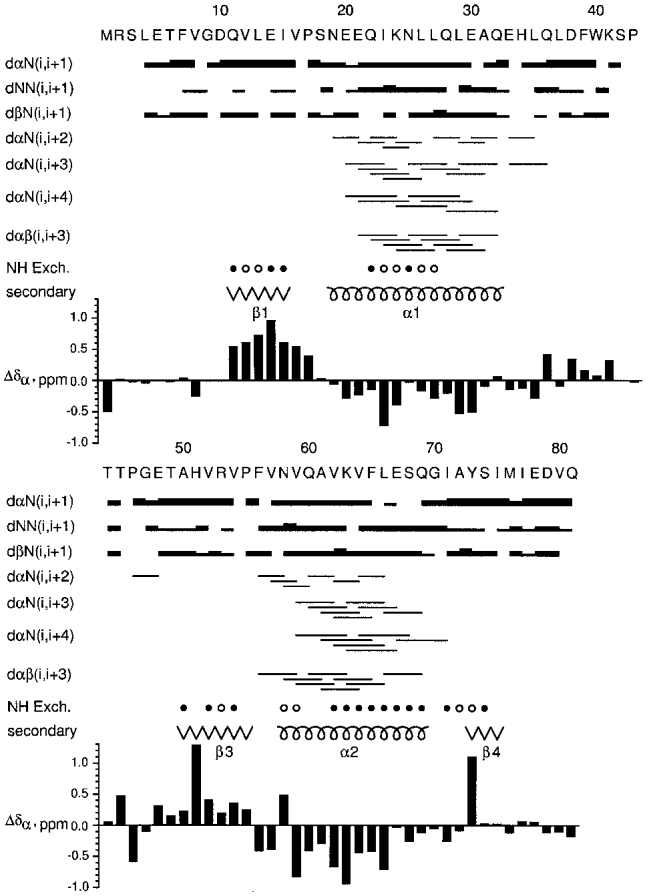

Figure 2. Summary of the sequential and medium-range NOEs, hydrogen exchange data, and $C_\alpha H$ conformational shifts ($\Delta\delta_{C\alpha H} = \delta_{observed} - \delta_{random\ coil}$, ppm) as a function of sequence for ADA2h (50 mM phosphate buffer pH 7.0, in H_2O/D_2O 9:1, 25°C). The thickness of the bars reflects the intensity of the sequential NOE connectivities, i.e., weak, medium, and strong. Amide NH protons that exchange slowly with solvent ($k_{ex} < 10^{-1}$ min^{-1}) are indicated by filled circles if log P > 3.5 and by open circles if 2.6 < log P < 3.5. The elements of secondary structure are also shown.

eratively. A total of 20 structures were finally calculated from a set of 1688 conformation relevant constraints by using the DYANA program (pdb code: 1o6x). Excluding the N-terminal nine residues and the C-terminal one, the final structure (Fig. 4) is remarkably well defined (average pairwise backbone RMSD for residues 10–80 of the 20 calculated structures is 0.67 ± 0.17 Å). It contains two α-helices, comprising residues 19–32 and 58–69, and an antiparallel β-sheet composed of three strands spanning residues 10–15, 50–55, and 73–75 (Figs. 2–4). The two α-helices have most of the characteristic hydrogen bonds between the CO of residue i and the NH of residue i+4. Hydrogen bonds linking adjacent β-strands are also found in the calculated structures (Fig. 3). Residues 38–41 are facing β-strand 3, but they are not hydrogen bonded to β-strand 3, nor do they display the ϕ and ψ dihedral backbone angles characteristic of a β-strand, as do the ones in β-strand 2 in the crystal structure of the intact procarboxypeptidase A2 (García-Sáez et al. 1997). Segment 32–38, connecting helix 1 to the β2 pseudostrand (Fig. 4), is not so well defined as the rest of the ADA2h regions. This is illustrated in Figure 5, where the RMSD of the interstructure C_α-C_α distances in the 20 converged structures is shown as a function of residue number. A second region, residues 41–48, also presents larger than average interstructure C_α-C_α distance RMSDs (see Fig. 5). It is interesting to note that both segments 32–38 and 41–48 belong to the region of the activation domain which is bound to interact with the enzyme (dotted frame in Fig. 5). Two different conformational families, one more populated than the other, can be distinguished along segment 32–38. An interconverting equilibrium between these two or more conformations is probably taking place, as deduced from the broad ^1H-^{15}N cross-peaks observed for residues E33 and Q36 in the HSQC spectrum (Fig. 1) and for the absence of the one corresponding to H34 (probably because of its large broadness). These two conformations could in principle bear some relation to the titration of H34.

A side-chain torsion angle is considered well defined when its RMSD between the values in the 20 calculated structures is less than ±30° or its order parameter is in the range 0.87–1.00 (the order parameter is zero for a totally random dihedral angle and one for a completely fixed conformation; Hyberts et al. 1992). According to these criteria, the side chains of 48 non-Ala/Gly/Pro residues are well defined. The 24 residues with a flexible side chain are those at the disordered N-terminal region (1–9, except for E5), the C-terminal residue, and mainly polar residues displaying a large accessible surface area (> 20%) in the final calculated structures (D10 and E14 in β1 strand; K24 and L29 in α1

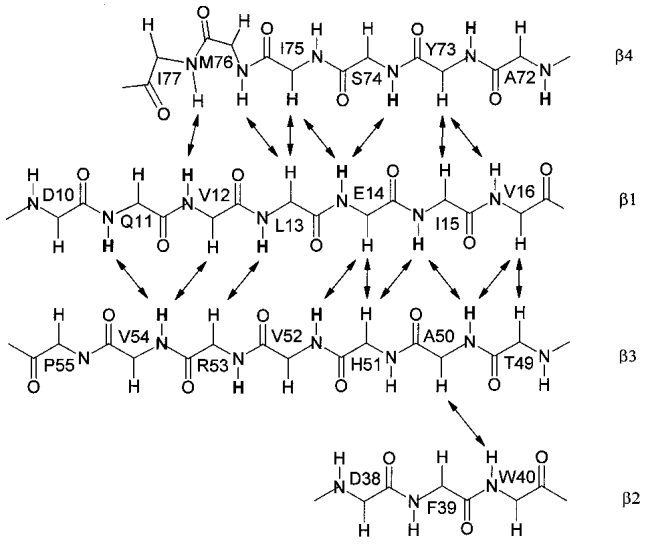

Figure 3. Schematic diagram showing the topology of the antiparallel β-sheet structure of ADA2h. An arrow between two protons indicates an observed nonsequential NOE. Slow-exchanging amide protons are in bold.

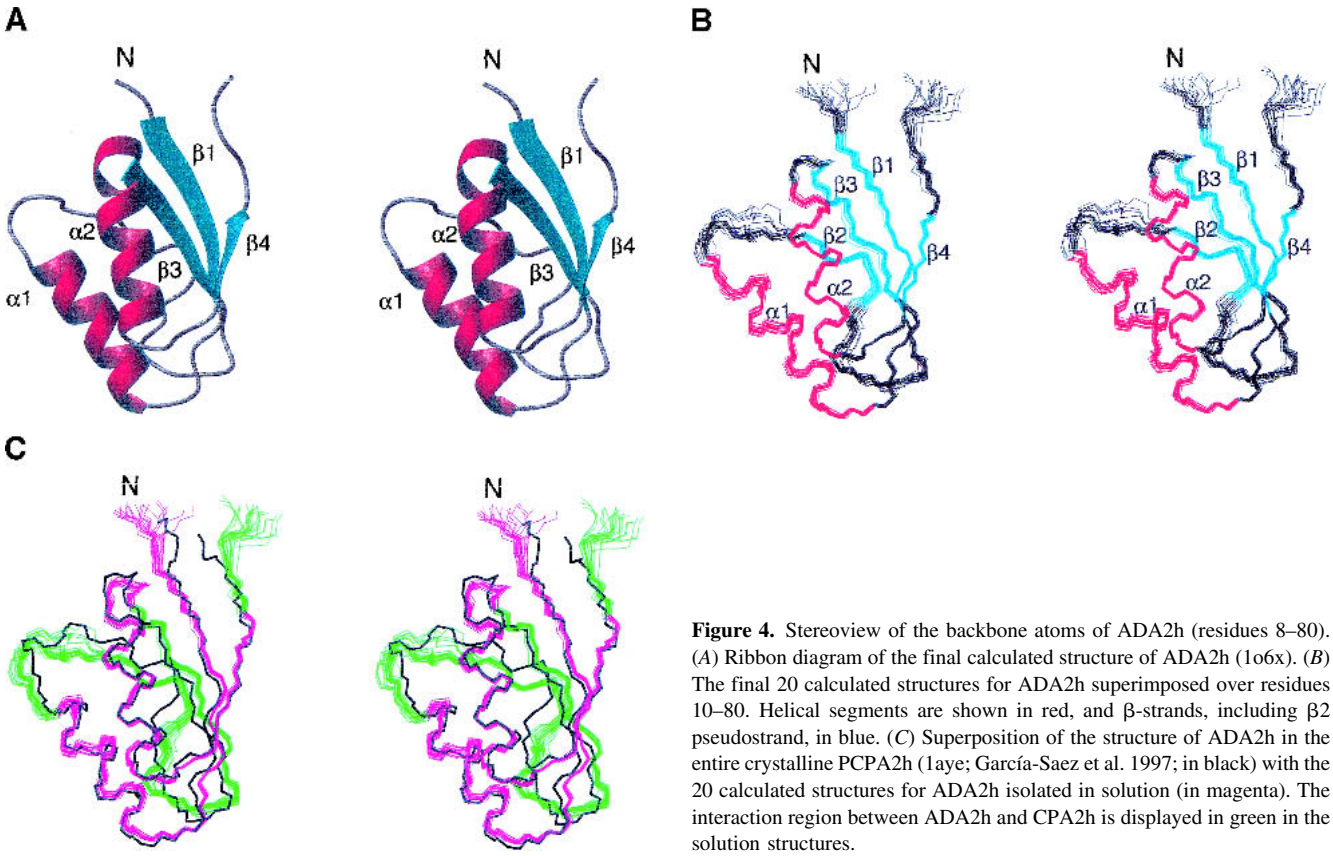

Figure 4. Stereoview of the backbone atoms of ADA2h (residues 8–80). (*A*) Ribbon diagram of the final calculated structure of ADA2h (1o6x). (*B*) The final 20 calculated structures for ADA2h superimposed over residues 10–80. Helical segments are shown in red, and β-strands, including β2 pseudostrand, in blue. (*C*) Superposition of the structure of ADA2h in the entire crystalline PCPA2h (1aye; García-Saez et al. 1997; in black) with the 20 calculated structures for ADA2h isolated in solution (in magenta). The interaction region between ADA2h and CPA2h is displayed in green in the solution structures.

helix; V52, R53, and V54 in β3 strand; S68 in α2 helix; M76 in β4 strand; and E33, H34, L35, E48, and E78 in loop regions). Residues L29, L35, V52, and V54 have a low accessible surface area (< 20%) and a flexible side chain.

pH titration

To investigate whether the signal broadening observed for residues around H34 was related to the pH titration of the imidazolium group of histidine side chain, we analyzed the pH dependence of the δ-values of appropriate protons in the pH range 5.5–9.1 by recording 1D and 2D TOCSY spectra of ADA2h in 50 mM phosphate in ^2H_2O, at 298K, at six pH values (5.5, 6.0, 6.6, 7.0, 8.1, and 9.1). NMR spectra at pH lower than 5.5 could not be obtained because of protein precipitation. Three ionizable groups are expected to titrate within the experimental pH range, the N-terminal amino group (M1) and the two histidine side chains (H34 and H51). Their pK_a values were determined by fitting to equation 1 the δ/pH curves of M1, H34, and H51 protons that, as expected, display significant titration shifts ($|\Delta\delta| > 0.08$ ppm, where $\Delta\delta = \delta_{(pH\ 9.1)} - \delta_{(pH\ 5.5)}$, ppm). Apart from them, protons sequentially or spatially close to H34 and H51 show significant titration shifts and good fits to equation 1 (Table 1). The pK_a values determined from the histidine protons themselves coincide with those obtained from protons in other residues (Table 1).

The closeness of the experimental pK_a values obtained for the N-terminal amino group and for H34 side chain to the corresponding intrinsic pK_a values (Table 1) indicates that they are solvent accessible, in agreement with their location in a disordered region and in a flexible loop of the ADA2h structure, respectively. The closeness of the H34 pK_a value (6.7) to the pH used for ADA2h spectral acquisition would suggest that the observed signal broadening might be due to intermediate exchange behavior between the protonated and unprotonated imidazolium species at the midpoint transition. However, this hypothesis is discarded due to the fact that such broadening should be larger for protons with larger protonation shifts, such as in H34 $C_{\varepsilon 1}H$ proton where it is 0.69 ppm, which is not observed. Thus, the only reasonable explanation is an equilibrium between the conformational families extant in the 32–39 region of the solution ADA2h structure that is intermediate in the NMR time scale.

The H51 pK_a value is lower than the intrinsic histidine pK_a value (Table 1), suggesting that the H51 side chain should be partially buried or have a nearby positive charge. Both effects are likely contributing to the low H51 pK_a, because H51 sidechain is close to the positively charged

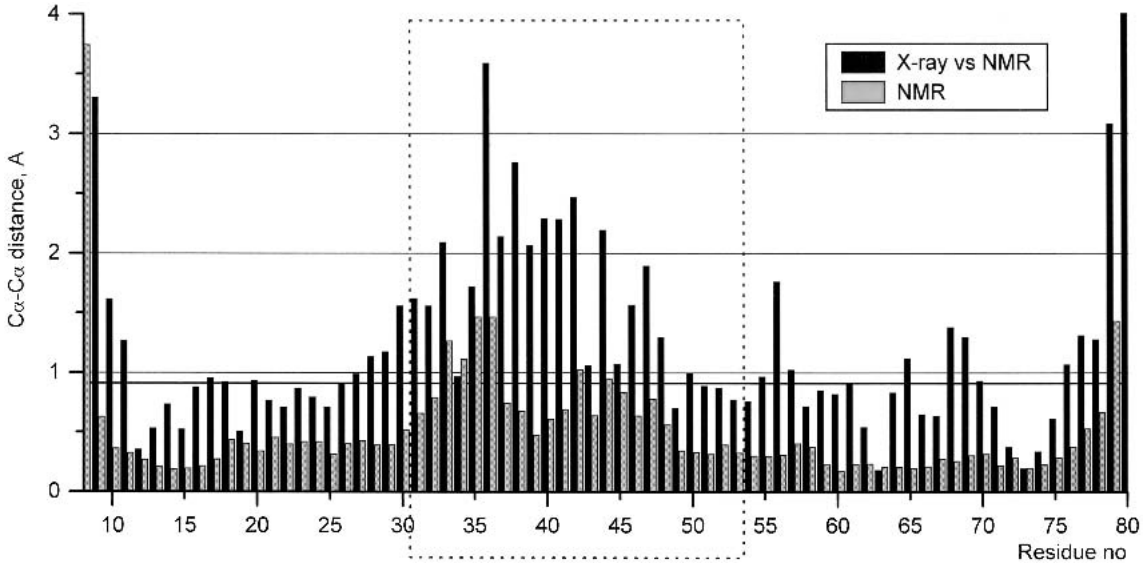

Figure 5. Pairwise averaged C_α-C_α distances between the structures calculated for isolated ADA2h are plotted in gray bars as a function of residue number for the 20 calculated structures. The black bars are the pairwise averaged C_α-C_α distances between the structure of ADA2h in the entire PCPA2h (1aye; García-Saez et al. 1997) and the structures calculated for isolated ADA2h. A horizontal line was plotted at 0.91 Å. The region of ADA2h that interacts with CPA2h in the entire PCPA2h (residues 31–53) is boxed with a dotted line.

R53 side chain in the calculated ADA2h structures, and has a very low solvent accessible surface (only 6.1% of the total H51 surface is solvent accessible, in contrast with the 50.9% exposed by H34).

Amide hydrogen exchange analysis

Amide hydrogen exchange was initiated by dissolving the freeze-dried ^{15}N-labeled ADA2h into 50 mM phosphate buffer (pH 7.0) in ^2H_2O, at 298K, and then followed throughout a series of ^{1}H-^{15}N-HSQC spectra over a 20-d period. A total of 29 cross-peaks were present at the first ^{1}H-^{15}N-HSQC spectrum that was recorded 18 min after starting the exchange. The individual hydrogen exchange rates (k_{ex}) of these amide protons were determined by fitting the time decrease of their corresponding cross-peak volumes to a single exponential decay function, and their corresponding ΔG_{op} values were calculated as described in Materials and Methods (Fig. 2). All other backbone amide protons as well as the side-chain NH protons were not measurable due to fast exchange (~$k_{ex} > 10^{-1}$ min^{-1}). An estimate of the global free energy of unfolding (ΔG_u) of 5.4 ± 0.3 kcal.mole^{-1} was obtained from the exchange rates of the most protected amide protons (those with log P > 3.5, 18 in total; Fig. 2) for ADA2h at pH 7.0 and 25°C (see Materials and Methods). To compare this value with ΔG_u obtained by other methods, the effect of slow *cis-trans* isomerization of the prolyl peptide bond must be taken into account. This effect is due to the fact that proline residues have no time to reach their isomeric equilibrium distribution in the exchange experiments, whereas they do so in the equilibrium unfolding experiments. The proline-dependent free energy increment due to the four Pro present in ADA2h, which are *trans* in the folded state, can be evalu-

Table 1. *pK_a values for the ADA2h ionizable groups that change their protonation state in the pH range of 5.5–9.1*

Ionizable group	Intrinsic pK_a	pK_a	Titrating protons
N-term	7.7	7.48 ± 0.12	3 (M1 $C_\alpha H$, $C_\beta H$ and $C_{\beta'}H$)
scH34	6.8	6.69 ± 0.06	4 (H34 $C_\beta H$, $C_{\beta'}H$, $C_{\delta 2}H$ & $C_{\varepsilon 1}H$)
		6.68 ± 0.11	8 (H34 $C_\beta H$, $C_{\beta'}H$, $C_{\delta 2}H$ & $C_{\varepsilon 1}H$, L29 $C_\gamma H$, Q32 $C_\alpha H$, L35 $C_\delta H_3$, L37 $C_\beta H$)
scH51	6.8	6.20 ± 0.16	2 (H51 $C_\alpha H$ & $C_{\varepsilon 1}H$)
		6.29 ± 0.17	7 (H51 $C_\alpha H$ & $C_{\varepsilon 1}H$, V12 $C_\gamma H_3$ & $C_{\gamma'}H_3$, A50 $C_\alpha H$, V52 $C_\beta H$ & $C_\gamma H_3$)

Reported errors are the standard deviations.

ated as 0.53 kcal.mole^{-1} at 25°C (Bai et al. 1994). Thus, the corrected global free energy of unfolding (ΔG_u) derived from exchange data is 4.9±0.3 kcal.mole^{-1}.

Discussion

The global fold of ADA2h in aqueous solution as determined here is very similar to that of the corresponding activation domain in the crystal structure of intact human procarboxypeptidase A2 (PCPA2h; pdb code: 1aye; García-Saez et al. 1997). There are however some important differences worth pointing out. As can be inferred from Figures 4 and 5, these differences are mainly located in the region of the activation domain that interacts with the enzyme in the intact procarboxypeptidase PCPA2h. The segment with maximum C_α-C_α distance RMSD values spanning residues 28–48 can be subdivided into three different subsegments: 28–33, 35–42, and 44–48. The first and the third subsegments correspond to the loops between α-helix 1 and β2 pseudostrand and between the pseudostrand and β-strand 3, respectively, which are the most flexible regions of the protein in addition to the N- and C-termini. The second subsegment includes the β2 pseudostrand itself (residues 38–41). The finding that the regions of the activation domain undergoing conformational shifts when bound correspond to the most flexible ones in the free protein is in keeping with prevailing ideas about how proteins interact.

Vendrell et al. (1991) determined the solution structure of the porcine activation domain (pdb code: 1pba;). They did not find marked differences between the solution structure of the isolated domain and the one in the crystal structure of the intact protein. This can be due to the lower definition of the solution structure of ADBp (compared to that obtained here for ADA2h). The average pairwise RMSD for backbone atoms excluding disordered N- and C-terminal regions are 0.67 ± 0.17 Å for ADA2h, residues 10–80, and 1.50 ± 0.24 Å for ADBp, residues 11–76 (calculated from deposited PDB coordinates using the MOLMOL program; pdb code: 1pba; Vendrell et al. 1991). The main differences in the solution structures of ADA2h and ADBp lie in the region following the second strand (β2 pseudostrand in ADA2h) where ADA2h (residues 41–45) has a four-residue deletion relative to ADBp (residues 39–47) and lacks the 3_{10} helix formed by residues 42–46 in ADBp. The flexible 32–38 loop in ADA2h that shows a two-residue insertion relative to ADBp corresponds to one of the three short disordered loop segments in ADBp (residues 32–34, 39–43, 56–61; Vendrell et al. 1991). Two of these disordered loops belong to the interacting interface between ADBp and CPBp. Thus, the flexible regions in the isolated activation domains of ADA2h and ADBp are located at the interacting interface between the domain and the enzyme.

There are no significant differences between the accessible surface areas in the solution structures of the isolated ADA2h domain and the intact PCPA2h, except for two residues (D10 and R53) that are solvent-exposed (ASA > 30%) in isolated ADA2h and buried (ASA < 10%) in the crystal structure. Considering these results, the effect of mutations belonging to helices α1 and α2 and to strands β1, β3, and β4 on the stability and folding kinetics of ADA2h that was interpreted on the basis of the structure of the activation domain in the intact PCPA2h is validated. Regarding mutations of F39 and K41, they belong to strand β2 in the crystal structure that is not present in the isolated ADA2h in solution, but they produce a significant decrease in protein stability (Villegas et al. 1998). The fact that the side chains of F39, W40, and K41 are well defined (see Results) accounts for their important contribution to the stability of ADA2h, even though they do not belong to a regular element of secondary structure. This means that the previous interpretation of the effect of different mutations on the folding kinetics of ADA2h (Villegas et al. 1998) is correct. However, our results indicate that some caution must be taken when assuming that the structure of an isolated domain is identical to that adopted in its corresponding multidomain protein.

Considering the exchange behavior, the most stable regions of ADA2h, which show slow-exchanging amide protons, coincide with the identified regular elements of secondary structure (Fig. 2). The C-terminal helix (α2) and the strands β1, β3, and β4 (Figs. 2–4) are the more stable. Two of these, β1 and β3, are inner strands, whereas the remaining strand (β4), although it is an edge strand, packs on the C-terminal helix (Fig. 4). The degree of protection of the N-terminal helix α1 is lower. This distribution of slow-exchanging amide protons along the secondary structure elements is analogous to that reported in a qualitative way for ADBp (Vendrell et al. 1991).

The stability obtained from the H-exchange data of the most protected protons of ADA2h, once corrected for the proline effect (see above), is 4.9 ± 0.3 kcal.mole^{-1}. This value is slightly larger than those obtained for the unfolding-refolding processes under the same experimental conditions by calorimetry (4.6 kcal.mole^{-1}, Fernández et al. 2000), chemical denaturation at the equilibrium (4.2 kcal.mole^{-1}, Villegas et al. 1995a), and kinetic studies (4.4 kcal.mole^{-1}, Villegas et al. 1995a). This discrepancy could simply be due to the H/D isotopic effect, because the exchange data are obtained in D_2O, whereas the protein is in H_2O in the other methods. This agrees with reports on some other proteins for which a higher stability in D_2O was observed (Huyghues-Despointes et al. 1999). An alternative explanation would be the occurrence of some local structures in the denatured state of the protein, although the fact that the higher stability values are obtained in different parts of the protein (Fig. 2) disproves this possibility. Hydrogen

exchange rates at equilibrium have been proposed as useful to probe intermediates on the protein folding pathway (Woodward 1993), although that proposal has been seriously contested (Clarke and Fersht 1996). However, in many cases a simultaneity between thermodynamic and kinetic aspects of protein folding has been observed, and it has been suggested that, in particular, slow exchange protein cores could be the refolding cores (Lacroix et al. 1997). In the case of ADA2h, this empirical rule is also roughly respected. Thus, region 38–41 that has phi (Φ) values close to zero exchanges very rapidly (Villegas et al. 1998). Similarly, α-helix 2 is more protected than α-helix 1, and it also exchanges more slowly. Moreover, recent studies of deuterium exchange followed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) also point out the coincidence between protection and early folding (Villanueva et al. 2000).

Materials and methods

Protein expression and purification

To obtain nonlabeled ADA2h, the XL-1-BLUE/pTZU18:*ADA2h E. coli* strain was expressed and purified with improved procedures relative to those described (Villegas et al. 1995a,c; Viguera et al. 1996). *E. coli* cells were grown overnight in LB-ampicillin at 37°C from a single colony. Inoculation, at 1:1000, was done in fresh medium and the culture grown until OD_{600nm} of 0.4 was reached. Isopropyl β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.17 mM, and the cultures were left for expression for 10 h. Cells were harvested by centrifugation and resuspended in 15 mL of 50 mM sodium phosphate buffer (pH 7.0) per L of initial culture. After sonication and ultracentrifugation at 100,000 *g*, the soluble fraction was submitted to ammonium-sulphate precipitation (70% saturation), and the pellet recovered after centrifugation was dialyzed and loaded into a QMA plus cartridge (Waters) in 20 mM Tris buffer (pH 8.0). The same buffer containing 0.1 M ammonium sulphate was used to eliminate contaminant proteins, and the elution step was performed with 0.4 M ammonium sulphate. After concentration up to 4 mL in centripreps (Amicon), the sample was loaded in a HiLoad 26/60 Superdex 75 (Pharmacia) column in 1 mM ammoniumtrifluoroacetate (pH 8.0). The pool containing the recombinant protein was lyophilized.

To prepare ^{15}N-labeled protein, the pTZU18:*ADA2h* construct was used to transform TG1 *E. coli* strain. *E. coli* cells were grown in LB-ampicillin at 37°C from a single M9 (glucose 0.4%) ampicillin-selected colony until OD_{600nm} of 0.4 was reached. A 1:200 inoculation was then done in fresh M9 (glucose 0.4%) ampiciline medium and grown until OD_{600nm} of 0.4 was reached. IPTG was then added to a final concentration of 0.17 mM, and the cultures were left for expression for 16 h. The cells were harvested and treated as described above for the nonlabeled protein.

Both the molecular weight of the recombinant protein and the incorporation of the labeling were checked by MALDI-TOF-MS. More than 99.5% of the N content of the ^{15}N-labeled protein incorporated the appropriate isotope.

The final yield of the expression and purification was about 15 mg/L of initial culture for the nonlabeled ADA2h and about 5 mg/L of initial culture for the ^{15}N-labeled ADA2h.

NMR spectroscopy

Protein samples for NMR experiments were prepared by dissolving the lyophilized protein in 0.5 mL of 50 mM phosphate buffer (pH 7.0) in either H_2O/D_2O (9:1 ratio by volume) or in D_2O, at a concentration of ~2 mM. The pH of the samples was checked with a glass microelectrode and was not corrected for isotope effects. The temperature of the NMR probe was calibrated using a methanol sample. Sodium [3-trimethylsilyl 2,2,3,3–$^2H_4$] propionate (TSP) was used as an internal reference for ^1H δ. The ^{13}C and ^{15}N δ-values were indirectly referenced by multiplying the spectrometer frequency that corresponds to 0 ppm in the ^1H spectrum, assigned to internal TSP, by 0.25144954 (Bax and Subramanian 1986; Spera and Bax 1991) and 0.101329118 (Wishart et al. 1995), respectively. NMR spectra were acquired on a Bruker AMX-600 pulse spectrometer operating at 600.13 MHz for the proton. Two-dimensional homonuclear COSY (Aue et al. 1976), TOCSY (Rance 1987), nuclear Overhauser enhancement spectroscopy (NOESY; Jeener et al. 1979; Kumar et al. 1980) spectra were recorded for the nonlabeled protein and acquired in the phase-sensitive mode using the time-proportional phase incrementation mode (Redfield and Kuntz 1975). A mixing time of 150 msec was used for NOESY spectra. TOCSY spectra were recorded by using the clean-CITY sequence (Briand and Ernst 1991) and an 80-msec mixing time. Two-dimensional heteronuclear ^1H-^{15}N HSQC (Bodenhausen and Ruben 1980) and 3D heteronuclear HSQC-TOCSY and HSQC-NOESY were recorded for the ^{15}N-labeled sample and acquired by using the States-TPPI method (Marion et al. 1989). Two-dimensional ^1H-^{13}C HSQC spectra (Bodenhausen and Ruben 1980) at natural ^{13}C abundance were recorded in nonlabeled protein samples in D_2O. Water suppression was achieved either by selective presaturation or by including a WATERGATE module (Piotto et al. 1992) in the original pulse sequences prior to acquisition. Two-dimensional acquisition data matrices were defined by 2048×512 points in t_2 and t_1, respectively. Data were processed using the standard XWIN-NMR Bruker program on a Silicon Graphics computer. The 2D data matrix was multiplied by a square-sine-bell window function with the corresponding shift optimized for every spectrum and zero-filled to a $4K \times 2K$ complex matrix prior to Fourier transformation. Baseline correction was applied in both dimensions.

The spectra were assigned within the X-EASY program (Bartels et al. 1995) after converting the XWIN-NMR processed spectra to the X-EASY format.

Structure calculation

Distance constraints for structure calculations were derived from the 2D 120-msec mixing time NOESY spectra recorded in H_2O and in D_2O. The NOE cross-peaks were integrated using the automatic integration subroutine of the X-EASY program (Bartels et al. 1995) and then calibrated and converted to upper-limit distance constraints within the DYANA program (Güntert et al. 1997). Structures were calculated on a Silicon Graphics Indigo Computer using the DYANA program (Güntert et al. 1997) and an annealing strategy.

Limits for the elements of secondary structure in the final structures calculated for ADA2h as well as in those previously reported for ADBp and for the intact PCPA2h were determined by using the Promotif v 2.0 program (Hutchinson and Thornton 1996).

pH titration

The effect of pH on the chemical shifts of protons was determined by analyzing a series of TOCSY spectra recorded at different pH

values ranging from 5.5 to 9.1. ADA2h precipitates at pH values lower than 5.5. The pH dependence of protons in the TSP reference was corrected as described previously (De Marco 1977; Bundi and Wüthrich 1979). Measurements of the pH sample were taken at room temperature before and immediately after the NMR experiment. For most resonances, the chemical shifts between consecutive spectra were small and allowed the facile assignment by comparison with the spectra assigned at pH 7.0.

Calculation of pK_a values

pK_a values were determined by nonlinear least-squares fit of the experimental pH titration curves of the ^1H chemical shifts to the following equation:

$$\delta = [\delta_1 + \delta_2 \times 10^{(pH - pKa)}]/[1 + 10^{(pH - pKa)}] \quad (1)$$

where δ_1 and δ_2 are the chemical shift values at the low and high extremes of pH, respectively. This equation was derived from the Henderson-Hasselbach equation by considering the titration of one group and assuming a rapid equilibrium between protonated and unprotonated forms (Forman-Kay et al. 1992).

All reported standard errors reflect the data fitting precision. The uncertainty of the experimental determination of the sample pH value, which is estimated to be about ±0.1 pH units, was not included.

Amide proton exchange

The NH exchange was followed by a tandem method. The exchange reaction was started by dissolving the lyophilized ^{15}N-labeled ADA2h in D_2O. After transfer to a 5-mm NMR tube and shimming, a series of consecutive 2D heteronuclear ^1H-^{15}N HSQC (Bodenhausen and Ruben 1980) experiments were run. The first 2D spectrum was recorded 18 min after dissolving the protein. The acquisition time for each 2D ^1H-^{15}N HSQC spectrum that was acquired with 2048 data points and 256 t_1 increments with eight scans per increment was approximately 45 min.

Hydrogen exchange rates were determined by fitting cross-peak volumes that were measured using the XWIN-NMR program (Bruker) to a first-order exponential decay:

$$I(t) = I(0) \exp(- k_{ex} t)$$

where I represents the volume of the cross-peak, I(0) is the cross-peak volume at $t = 0$, k_{ex} is the experimental rate of hydrogen exchange, and t is the time in minutes. Data were fitted with the program Microcal Origin 5.0.

Hydrogen exchange analyses can be described in terms of a structure unfolding model, in which exchange only takes place from an 'open', or O, form of the amide hydrogen atom, but not from the 'closed', or C, form (Linderström-Lang 1955; Hvidt and Nielsen 1966; Englander and Kallenbach 1984):

$$C^{N\text{-}H} \underset{kf}{\overset{ku}{\leftrightarrow}} O^{N\text{-}H} \underset{D_2O}{\overset{kc}{\leftrightarrow}} O^{N\text{-}D} \underset{ku}{\overset{kf}{\leftrightarrow}} C^{N\text{-}D}$$

In this scheme, k_u and k_f are the local unfolding and folding rates, respectively, and k_{rc} is the intrinsic rate constant for the exchange reaction, which is dependent on the primary sequence. The intrinsic exchange rate constants for each amide proton, k_{rc}, were calculated as described by Bai et al. (1993). The dominant mechanism of exchange for most proteins at moderate pH and temperature tends towards the limiting condition $k_f >> k_u$ and $k_f >> k_{rc}$, known as the EX2 limit. Hence, the measured exchange rate, k_{ex}, reduces to:

$$k_{ex} = k_u * k_{rc}/k_f = K_{op} * k_{rc},$$

Where K_{op} is the equilibrium constant for local transient opening of a hydrogen bonded site. This equilibrium constant relates to ΔG_{op}, the structural free energy difference between the closed and open states:

$$\Delta G_{op} = -RT \ln K_{op} = -RT \ln (k_{ex}/k_{rc}),$$

with R the gas constant and T the absolute temperature. One of the advantages of the hydrogen exchange method over equilibrium denaturation techniques lies in the capacity to pinpoint local differences in structural free energy. Furthermore, exchange of the most protected amide protons has been shown to often occur via global unfolding, in which case ΔG_{op} approximates ΔG_u, the free energy of unfolding of a protein (Wagner and Wüthrich 1979; Woodward 1994; Jeng and Dyson 1995; Englander et al. 1996).

Acknowledgments

V.V., J.V., and F.X.A. gratefully acknowledge financial support from the CICYT (Spain; grant BIO01-2046) and from the Centre de Referència en Biotecnologia (Generalitat de Catalunya). L.S. acknowledges financial support from an EU TMR grant (CT96-0013). M.A.J., J.S., and M.R. acknowledge financial support from DGICYT (PB98-0677).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Aue, W.P., Bertholdi, E., and Ernst, R.R. 1976. Two-dimensional spectroscopy. Application to NMR. *J. Chem. Phys.* **64:** 2229–2246.

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17:** 75–86.

———. 1994. Protein stability parameters measured by hydrogen exchange. *Proteins* **20:** 4–14.

Bartels, C., Xia, T.-H., Billeter, M., Güntert, P., and Wüthrich, K. 1995. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **5:** 1–10.

Bax, A. and Subramanian, J. 1986. Sensitivity-enhanced two-dimensional heteronuclear shift correlation NMR spectroscopy. *J. Magn. Reson.* **67:** 565–570.

Bodenhausen, G. and Ruben, D.J. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69:** 185–189.

Briand, J. and Ernst, R.R. 1991. Computer-optimised homonuclear TOCSY experiments with suppression of cross relaxation. *Chem. Phys. Lett.* **185:** 276–285.

Bundi, A. and Wüthrich, K. 1979. Use of amide ^1H NMR titration shifts for studies of polypeptide conformation. *Biopolymers* **18:** 299–311.

Clarke, J. and Fersht, A.R. 1996. An evaluation of the use of hydrogen exchange at equilibrium to probe intermediates on the protein folding pathway. *Fold. Des.* **1:** 243–254.

Coll, M., Guasch, A., Aviles, F.X., and Huber, R. 1991. Three-dimensional structure of porcine procarboxypeptidase B: A structural basis of its inactivity. *EMBO J.* **10:** 1–9.

De Marco, A. 1977. pH dependence of internal references. *J. Magn. Reson.* **26:** 527–528.

Englander, S.W. and Kallenbach, N.R. 1984. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **16:** 521–655.

Englander, S.W., Sosnick, T.R., Englander, J.J., and Mayne, L. 1996. Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **6:** 18–23.

Fernández, A.M., Villegas, V., Martínez, J.C., Van Nuland, N.A., Conejero-

Lara, F., Aviles, F.X., Serrano, L., Filimonov, V.V., and Mateo, P.L. 2000. Thermodynamic analysis of helix-engineered forms of the activation domain of human procarboxypeptidase A2. *Eur. J. Biochem.* **267:** 5891–5899.

Forman-Kay, J.D., Clore, G.M., and Gronenborn, A.M. 1992. Relationship between electrostatics and redox function in human thioredoxin: Characterization of pH titration shifts using two-dimensional homo- and heteronuclear NMR. *Biochemistry* **31:** 3442–3452.

García-Sáez, I., Reverter, D., Vendrell, J., Aviles, F.X., and Coll, M. 1997. The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of inhibition, activation and intrinsic activity of the zymogen. *EMBO J.* **16:** 6906–6913.

Guasch, A., Coll, M., Aviles, F.X., and Huber, R. 1992. Three-dimensional structure of porcine pancreatic procarboxypeptidase A. *J. Mol. Biol.* **224:** 141–157.

Güntert, P., Mumenthaler, C., and Wüthrich, K. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273:** 283–298.

Hutchinson, E.G. and Thornton, J.M. 1996. PROMOTIF—A program to identify and analyse structural motifs in proteins. *Protein Sci.* **5:** 212–220.

Huyghues-Despointes, B.M.P., Scholtz, J.M., and Pace, C.N. 1999. Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat. Struct. Biol.* **6:** 910–912.

Hvidt, A. and Nielsen, S.O. 1966. Hydrogen exchange in proteins. *Adv. Protein Chem.* **21:** 287–386.

Hyberts, G.S., Golberg, M.S., Havel, T.F., and Wagner, G. 1992. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1:** 736–751.

Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. 1979, Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71:** 4546–4553.

Jeng, M.F. and Dyson, H.J. 1995. Comparison of the hydrogen-exchange behavior of reduced and oxidized *Escherichia coli* hioredoxin. *Biochemistry* **34:** 611–619.

Kumar, A., Ernst, R.R., and Wüthrich, K. 1980. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95:** 1–6.

Lacroix, E., Bruix, M., López-Hernández, E., Serrano, L., and Rico, M. 1997. Amide hydrogen exchange and internal dynamics in the chemotactic protein CheY from *Escherichia coli*. *J. Mol. Biol.* **271:** 472–487.

Linderström-Lang, K. 1955. Deuterium exchange between peptides and water. *Chem. Soc. (London), Spec. Publ.* **2:** 1–20.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., and Clore, G.M. 1989. Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1 β. *Biochemistry* **28:** 6150–6156.

Piotto, M., Saudek, V., and Sklenar, V. 1992. Gradient-tailored excitation for single quantum NMR spectroscopy in aqueous solutions. *J. Biomol. NMR* **6:** 661–665.

Rance, M. 1987. Improved techniques for homonuclear rotating frame and isotropic mixing experiments. *J. Magn. Reson.* **74:** 557–564.

Redfield, A.G. and Kuntz, S.D. 1975. Quadrature Fourier NMR detection: Simple multiplex for dual detection. *J. Magn. Reson.* **19:** 250–254.

Reverter, D., Ventura, S., Villegas, V., Vendrell, J., and Aviles, F.X. 1998. Human procarboxypeptidase A2: Overexpression in *Pichia pastoris* and detailed characterization of its activation pathway. *J. Biol. Chem.* **273:** 3535–3541.

Spera, S. and Bax, A. 1991. Empirical correlation between protein backbone conformation and C_α and C_β ^{13}C NMR chemical shifts. *J. Am. Chem. Soc.* **113:** 5490–5492.

Vendrell, J., Cuchillo, C.M., and Aviles, F.X. 1990. The tryptic activation pathway of monomeric procarboxypeptidase A. *J. Biol. Chem.* **256:** 6949–6953.

Vendrell, J., Billeter, M., Wider, G., Aviles, F.X., and Wüthrich, K. 1991. The NMR structure of the activation domain isolated from porcine procarboxypeptidase B. *EMBO J.* **10:** 11–15.

Vendrell, J., Querol, E., and Aviles, F.X. 2000. Metallocarboxypeptidases and their potent inhibitors. Structure, funtion and biomedical properties. *Biochim. Biophys. Acta* **1477:** 284–298.

Viguera, A.R., Villegas, V., Aviles, F.X., and Serrano, L. 1996. Favourable native-like helical local interactions can accelerate protein folding. *Fold. Des.* **2:** 23–33.

Villanueva, J., Canals, F., Villegas, V., Querol, E., and Aviles, F.X. 2000. Hydrogen exchange monitored by MALDI-TOF mass spectrometry for rapid characterization of the stability and conformation of proteins. *FEBS Lett.* **472:** 27–33.

Villegas, V., Azuaga, A., Catasús, L., Reverter, D., Mateo, P.L., Aviles, F.X., and Serrano, L. 1995a. Evidence for a two-state folding transition in the folding process of the activation domain of human procarboxypeptidase A2. *Biochemistry* **34:** 15105–15110.

Villegas, V., Vendrell, J., and Aviles, F.X. 1995b. The activation pathway of Procarboxypeptidase B from porcine pancreas: Participation of the active enzyme in the proteolytic process. *Protein Sci.* **4:** 1792–1800.

Villegas, V., Viguera, A.R., Aviles, F.X., and Serrano, L. 1995c. Stabilisation of proteins by rational design of α-helix stability using helix/coil transition theory. *Fold. Des.* **1:** 29–34.

Villegas, V., Martínez, J.C., Aviles, F.X., and Serrano, L. 1998. Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.* **283:** 1027–1036.

Villegas, V., Zurdo, J., Filimonov, V., Aviles, F.X., Dobson, C.M., and Serrano, L. 2000. Protein engineering as an strategy to avoid formation of amyloid fibrils. *Protein Sci.* **9:** 1700–1708.

Wagner, G. and Wüthrich, K. 1979. Correlation between the amide proton exchange rates and the denaturation temperatures in globular proteins related to the basic pancreatic trypsin inhibitor. *J. Mol. Biol.* **130:** 31–37.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. 1995. ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest neighbor effects. *J. Biomol. NMR* **5:** 67–81.

Woodward, C.K. 1994. Hydrogen exchanges rates and protein folding. *Curr. Opin. Struct. Biol.* **4:** 112–116.

Wüthrich, K. 1986. *NMR of proteins and nucleic acids.* Chapter 8. J. Wiley, New York.