Identification of N-terminal helix capping boxes by means of ^{13}C chemical shifts

Angela M. Gronenborn and G. Marius Clore

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.

SUMMARY

We have examined the ^{13}C$^{\alpha}$ and ^{13}C$^{\beta}$ chemical shifts of a number of proteins and found that their values at the N-terminal end of a helix provide a good predictor for the presence of a capping box. A capping box consists of a hydrogen-bonded cycle of four amino acids in which the side chain of the N-cap residue forms a hydrogen bond with the backbone amide of the N3 residue, whose side chain in turn may accept a hydrogen bond from the amide of the N-cap residue. The N-cap residue exhibits characteristic values for its backbone torsion angles, with ϕ and ψ clustering around $94 \pm 15°$ and $167 \pm 5°$, respectively. This is manifested by a 1–2 ppm upfield shift of the ^{13}C$^{\alpha}$ resonance and a 1–4 ppm downfield shift of the ^{13}C$^{\beta}$ resonance, relative to their random coil values, and is mainly associated with the unusually large value of ψ. The residues following the N-cap residue exhibit downfield shifts of 1–3 ppm for the ^{13}C$^{\alpha}$ resonances and small upfield shifts for the ^{13}C$^{\beta}$ ones, typical of an α-helix.

Helix start and stop signals have attracted considerable attention, both with respect to predictive value (Presta and Rose, 1988; Richardson and Richardson, 1988) and exploitation in protein design (O'Neil and DeGrado, 1990; Serrano et al., 1992; Lyu et al., 1993). Recently a unique N-terminal flanking motif with a distinct reciprocal backbone–side-chain hydrogen-bonding arrangement has been identified and dubbed the capping box (Harper and Rose, 1993). The capping box consists of a hydrogen-bonded cycle of four amino acids in which the side chain of the N-cap residue forms a hydrogen bond with the backbone amide of the N3 residue, whose side chain likewise accepts a hydrogen bond from the N-cap amide. The characteristic hydrogen-bonding geometry of these capping boxes is in turn linked to a particular set of polypeptide backbone dihedral angles for the N-cap residue, clustering around $\phi = 94° \pm 15°$ and $\psi = 167° \pm 5°$ (Harper and Rose, 1993).

Nuclear magnetic resonance chemical shift values are exquisitely sensitive to chemical environment, in particular composition and conformation; indeed, small electronic or structural changes can have pronounced effects. Recent investigations with respect to the relationship between
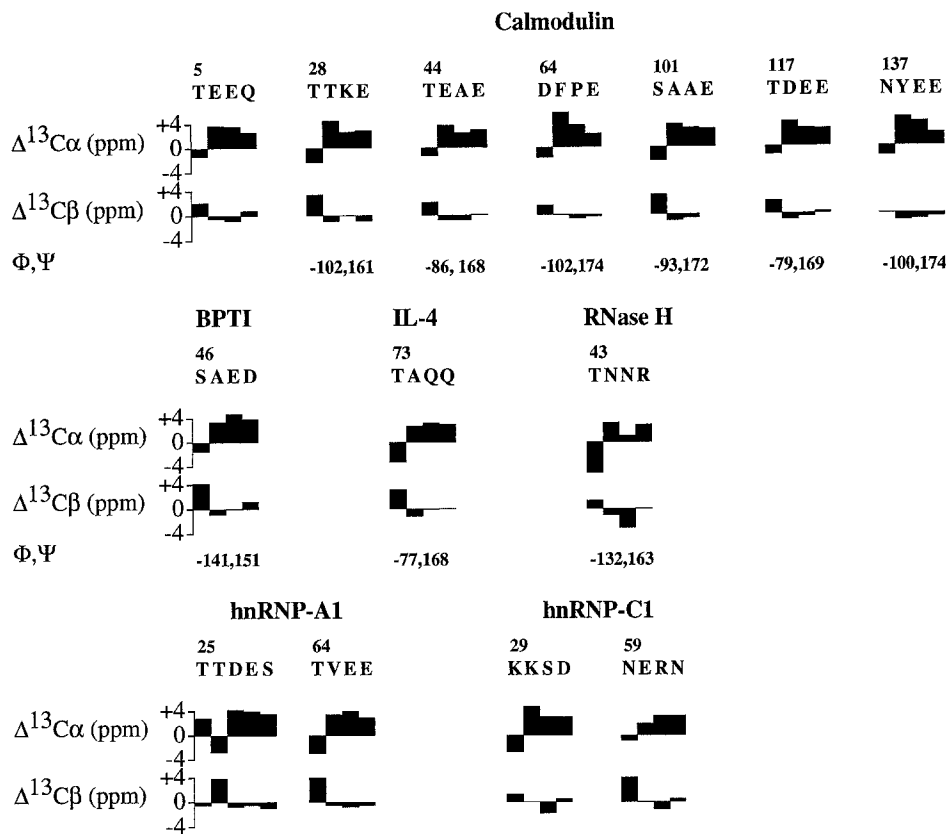
Calmodulin

Fig. 1. Deviations of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts from random coil values for selected helix termini. CaM: calmodulin; BPTI: basic pancreatic trypsin inhibitor; IL-4: interleukin-4; RNase H: ribonuclease H; hnRNP A1 and C1: heterogeneous nuclear ribonucleoprotein A1 and C1. Also listed are the ϕ,ψ angles for the N-cap residues extracted from the respective X-ray structures.

chemical shifts and protein secondary structure (Spera and Bax, 1991; Wishart et al., 1991) have revealed a clear correlation between secondary $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of an amino acid and its location in either an α-helix or a β-sheet. In particular, $^{13}C^{\alpha}$ resonances for residues located in a helix are characteristically downfield shifted by ~3 ppm relative to their random coil positions, whereas in β-sheets the $^{13}C^{\alpha}$ resonances are upfield shifted by ~2 ppm and the $^{13}C^{\beta}$ resonances are downfield shifted by ~3 ppm. Thus, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts provide easily accessible information about the backbone torsion angles ϕ and ψ. Based on this empirical correlation between ϕ, ψ and carbon chemical shifts, it should be possible to identify capping boxes at helix termini already at the resonance assignment stage, thereby permitting a more reliable delineation of the beginning of a helix. In addition, once identified, capping boxes may provide additional restraints in NMR-based structure determinations.

We have investigated the available body of ^{13}C assignments for several proteins with regard to the presence of capping box identifiers. Our database comprised the following proteins, all of which contain helices and for all of which $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts are available: basic

pancreatic trypsin inhibitor (BPTI; Wagner and Brühwiler, 1986; Hansen, 1991), calmodulin (CaM; Ikura et al., 1991; Bax, A., personal communication), Staphylococcal nuclease (SNase; Wang et al., 1992; Torchia, D.A., personal communication), interleukin-4 (IL-4; Powers et al., 1992), heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1; Garrett et al., 1994) and C1 (hnRNP C1; Wittekind et al., 1992), profilin I (Archer et al., 1993), stromelysin (Gooley et al., 1993), *E. coli* ribonuclease H (RNase H; Yamazaki et al., 1993), *L. casei* dihydrofolate reductase (DHFR; Soteriou et al., 1993) and human thioredoxin (hTRX; Quin, J., Clore, G.M. and Gronenborn, A.M., unpublished results).

The most striking example of the unique chemical shift pattern identifying capping boxes was observed for calmodulin. $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary chemical shifts for four residues at the N-terminal end of each of the seven α-helices are presented in Fig. 1. As can be easily appreciated, the N-cap residue is characterized by a 1–2 ppm upfield shift of the $^{13}C^{\alpha}$ resonance, accompanied by a 1–4 ppm downfield shift of the C^{β} resonance. The subsequent helical residues exhibit the already recognized downfield shifts of 1–3 ppm for $^{13}C^{\alpha}$ resonances and small upfield shifts for $^{13}C^{\beta}$ ones (Spera and Bax, 1991; Wishart et al., 1991). The presence of capping boxes for these helices is confirmed by the characteristic φ,ψ values for the N-cap residues, extracted from the high-resolution X-ray structure of calmodulin (Babu et al., 1988), which are also included in Fig. 1. IL-4 contains four helices, only one of which shows a clear signature of a capping box. Likewise, the helix in BPTI seems to contain a capping box, although the φ angle in the X-ray structure (Wlodawer et al., 1987) is somewhat larger (−140°) than would be expected for an idealized capping box (Harper and Rose, 1993). The latter is also true for helix α_A in *E. coli* RNase H, which shows the characteristic chemical shift pattern, but again a slightly larger φ angle in the X-ray structure (Yang et al., 1990). In this case the N3 residue is an arginine, which cannot partake in the appropriate hydrogen bonding between the side chain and the backbone amide protons of the N-cap residue. Therefore, the φ angle will not be as restricted as in the idealized cases. Both helices in hnRNP A1 and hnRNP C1 exhibit the characteristic capping box shift pattern at their N-termini, thus making it very likely that capping boxes exist. The validity of the latter prediction can be tested when high-resolution structures for these two proteins become available. The ^{13}C chemical shift pattern for helix α1 in hnRNP A1 clearly identifies Thr[25] as the N-cap residue, although based on sequence alone, Thr[24] would have been an equally likely candidate. No capping boxes were found, based on the above chemical shift criteria, for SNase, profilin, stromelysin, DHFR and human thioredoxin. Inspection of the structures of those proteins reveals no φ,ψ angles at the helix termini which lie in the −94°/167° vicinity, thus confirming the absence of capping boxes.

In summary, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift values at N-terminal helix ends seem to have good predictive value as to the presence of capping boxes, defined in terms of a characteristic set of φ,ψ backbone angles for the N-cap residue, with the N-cap residue exhibiting a 1–2 ppm upfield shift of the $^{13}C^{\alpha}$ resonances and a 1–4 ppm downfield shift of the $^{13}C^{\beta}$ resonances, relative to their random coil values. This chemical shift effect seems to be mainly due to the unusually large value of the backbone torsion angle ψ, clustering around 160–170°. It is interesting that substantially larger variations occur in φ, possibly due to the fact that the side-chain–backbone hydrogen bond between the N-cap and the N3 residue is the more frequent and important one, being associated with ψ = 167 ± 5°, whereas the reciprocal interaction between the side chain of the N3 residue and the backbone amide of the N-cap residue will not always be present. Only if the latter is

present will ϕ be restricted to values around $-90°$, otherwise a much broader range of angles can occur.

ACKNOWLEDGEMENTS

REFERENCES

Archer, S.J., Vinson, V.K., Pollard, T.D. and Torchia, D.A. (1993) *Biochemistry*, **32**, 6680–6687.

Babu, Y.S., Bugg, C.E. and Cook, W.J. (1988) *J. Mol. Biol.*, **204**, 191–204.

Garrett, D.S., Lodi, P., Shamoo, Y., Williams, K.R., Clore, G.M. and Gronenborn, A.M. (1994) *Biochemistry*, **33**, 2852–2858.

Gooley, P.R., Johnson, B.A., Marcy, A.I., Cuca, G.C., Salowe, S.P., Hagmann, W.K., Esser, C.K. and Springer, J.P. (1993) *Biochemistry*, **32**, 13098–13108.

Hansen, P.E. (1991) *Biochemistry*, **30**, 10457–10466.

Harper, E.T. and Rose, G.D. (1993) *Biochemistry*, **32**, 7605–7609.

Ikura, M., Spera, S., Barbato, N., Kay, L.E., Krinks, M. and Bax, A. (1991) *Biochemistry*, **30**, 9216–9228.

Lyu, P.C., Wemmer, D.E., Zhou, H.X., Pinker, R.J. and Kallenbach, N.R. (1993) *Biochemistry*, **32**, 421–425.

O'Neil, K.T., and DeGrado, W.F. (1990) *Science*, **250**, 646–651.

Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. and Clore, G.M. (1992) *Biochemistry*, **31**, 4334–4346.

Presta, L.G. and Rose, G.D. (1988) *Science*, **240**, 1632–1641.

Richardson, J.S. and Richardson, D.C. (1988) *Science*, **240**, 1648–1652.

Serrano, L., Sancho, J., Hirshberg, M. and Fersht, A.R. (1992) *J. Mol. Biol.*, **227**, 544–559.

Soteriou, A., Carr, M.D., Frenkiel, T.A., McCormick, J.E., Bauer, C.J., Sali, D., Birdsall, B. and Feeney, J. (1993) *J. Biomol. NMR*, **3**, 535–546.

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.

Wagner, G. and Brühwiler, D. (1986) *Biochemistry*, **25**, 5839–5843.

Wang, J., Hinck, A.P., Loh, S.N., LeMaster, D.M. and Markley, J.L. (1992) *Biochemistry*, **31**, 921–936.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.

Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G. and Mueller, L. (1992) *Biochemistry*, **31**, 6254–6265.

Wlodawer, A., Deisenhofer, J. and Huber, R. (1987) *J. Mol. Biol.*, **193**, 145–156.

Yamazaki, T., Yoshida, M. and Nagayama, K. (1993) *Biochemistry*, **32**, 5656–5669.

Yang, W., Hendrickson, W.A., Crouch, R.J. and Sato, Y. (1990) *Science*, **249**, 1398–1405.