Capping Interactions in Isolated α Helices: Position-Dependent Substitution Effects and Structure of a Serine-Capped Peptide Helix†

Pingchiang C. Lyu and David E. Wemmer

Department of Chemistry and Structural Biology Division, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720

Hongxing X. Zhou, Rachel J. Pinker, and Neville R. Kallenbach*

Department of Chemistry, New York University, New York, New York 10003

Received August 31, 1992; Revised Manuscript Received October 20, 1992

ABSTRACT: The influence of an amino acid on the stability of α-helical structure depends on the position of the residue in the helix with respect to the ends. Short α helices in proteins are stabilized both by H-bonding of the main-chain NH and CO groups and by capping interactions between side chains and unfulfilled peptide groups at the N and C termini. Peptide models based on consensus position-dependent helix sequences allow one to model capping effects in isolated helices and to establish a base line for these interactions in proteins. We report here an extended series of substitutions in the cap positions of our peptide models and the solution structure of peptide S3, with serine at the N-cap position defined as the N-terminal residue with partly helix and partly coil conformation. The resulting model, determined by 2D ^{1}H NMR, is consistent with a structure at the N-cap involving H-bonding between the serine γ oxygen and the peptide NH of the glutamic acid residue three amino acids toward the C terminus. A bifurcated H-bond of Ser O_γ with the NH of Asp5 is possible also, since this group is within interacting distance. This provides direct evidence that specific side-chain interactions with the main chain stabilize isolated α-helical structure.

α helices are among the most common secondary structures in proteins (Levitt, 1978; Fasman, 1989), for reasons that remain imperfectly understood. Side chains such as alanine and leucine tend to occur frequently in the sequences of helices, while others such as glycine do not. Differences among side chains in their helix-stabilizing propensity have been demonstrated using synthetic model peptides (O'Neil & DeGrado, 1990; Lyu et al., 1990; Padmanabhan et al., 1990). Studies with synthetic polypeptide models have also shown differences in helical propensities of natural amino acid guests but give different numerical values (Sueki et al., 1984), as discussed by Ptitsyn (1992). In contrast to the high molecular weight helical polypeptides that were used originally to analyze helix–coil transitions and determine helix propensities (Bychkova et al., 1971; Barskaya & Ptitsyn, 1971; Sueki et al., 1984), the α helices in globular proteins are short, averaging only about 12 residues in length (Presta & Rose, 1988). The effect of the ends in helices of this length is therefore significant, because the H-bonding potential of four residues at the N and C termini cannot be satisfied by the main-chain CO(i)–NH(i+4) hydrogen-bonding pattern characteristic of residues in the helix "middle" (Pauling & Corey, 1951; Presta & Rose, 1988). It follows that the mid-helix propensities now available (O'Neil & DeGrado, 1990; Lyu et al., 1990; Padmanabhan et al., 1990) apply strictly to a limited number of positions in protein helices.

The unfulfilled peptide NH and CO H-bonding capacity at the ends of α helices can be satisfied in at least two different ways (Presta & Rose, 1988). First, polar side chains can interact directly with main-chain groups to stabilize helical structure by "capping" (Presta & Rose, 1988; Richardson & Richardson, 1988). Side-chain–backbone capping of α helices appears indeed to be the predominant pattern among all side-chain–main-chain bonding interactions in proteins (Stickle et al., 1992). A second solution invokes the sterically unhindered glycine side chain, which is sufficiently flexible to allow distal peptide groups or side chains to interact with exposed NH or CO of the helix (Schellman, 1981). The frequency distribution of amino acids near the ends of helices in proteins is consistent

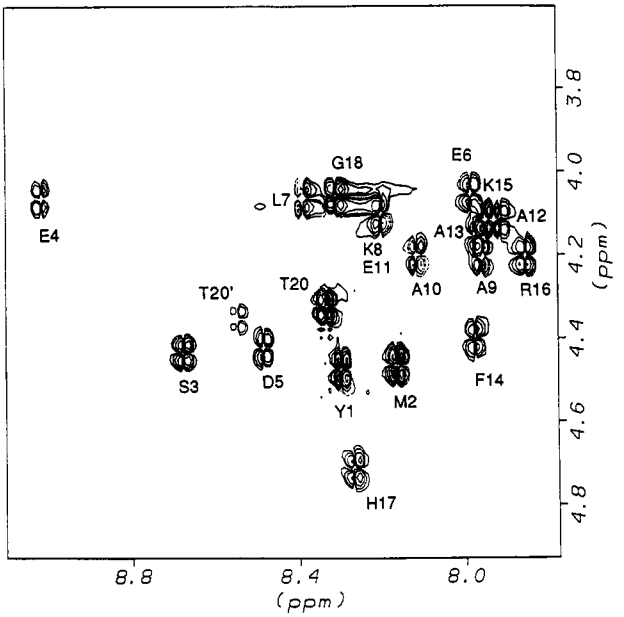

FIGURE 1: C_αH–amide region of the phase-sensitive DQF-COSY spectrum of S3 at 10 °C. The peptide sample is at ca. 4 mM concentration in H_2O/D_2O (85%/15%). Individual cross peaks are identified by standard one-letter symbols for the amino acids and the residue number in the sequence.

† This work was supported by Research Grant GM 40746 from the NIH, by NSF Grant DMB 88-15998, and through instrumentation grants from the U.S. Department of Energy, DE FG05-86ER75281, and the NSF, DMB 86-09305 and BBS 87-20134. P.C.L. holds a NRSA postdoctoral fellowship from the NIH, GM 14803.

* Author to whom correspondence should be addressed.

0006-2960/93/0432-421$04.00/0

Table I: Peptide Sequence and Helical Content[a]

peptide	sequence	$-[\theta]_{222}$	f, %
N-cap (position 3)			
parent (Ser)	YM S EDELKAAEAAFKRHGPT	16 000	50
N3 (Asn)	YM N EDELKAAEAAFKRHGPT	14 150	44
G3 (Gly)	YM G EDELKAAEAAFKRHGPT	10 580	33
A3 (Ala)	YM A EDELKAAEAAFKRHGPT	7 700	24
internal (position 6)			
parent (Glu)	YMSED E LKAAEAAFKRHGPT	16 000	50
A6 (Ala)	YMSED A LKAAEAAFKRHGPT	12 048	38
G6 (Gly)	YMSED G LKAAEAAFKRHGPT	6 941	22
internal (position 10)			
Nle10 (norleucine)	YMSEDELKA Nle EAAFKRHGPT	16 886	53
parent (Ala)	YMSEDELKA A EAAFKRHGPT	16 000	50
S10 (Ser)	YMSEDELKA S EAAFKRHGPT	9 300	29
G10 (Gly)	YMSEDELKA G EAAFKRHGPT	5 560	17
internal (position 17)			
parent (His)	YMSEDELKAAEAAFKR H GPT	16 000	50
A17 (Ala)	YMSEDELKAAEAAFKR A GPT	15 476	48
S17 (Ser)	YMSEDELKAAEAAFKR S GPT	15 449	48
C-cap (position 18)			
N18 (Asn)	YMSEDELKAAEAAFKRH N PT	20 028	63
Q18 (Gln)	YMSEDELKAAEAAFKRH Q PT	17 927	56
A18 (Ala)	YMSEDELKAAEAAFKRH A PT	17 359	54
S18 (Ser)	YMSEDELKAAEAAFKRH S PT	17 369	54
T18 (Thr)	YMSEDELKAAEAAFKRH T PT	15 951	50
parent (Gly)	YMSEDELKAAEAAFKRH G PT	16 000	50

[a] $[\theta]_{222}$ = mean residue ellipticity (deg·cm^2·dmol^{-1}) at 222 nm, 4 °C, and pH = 6. $f = ([\theta]_{obsd}/-32\,000) \times 100$ = percent helix content (Gans et al., 1991). Positions in these peptides are designated N″ N′ N_{cap} N1 N2 N3 N4 N5 (Helix Mid)$_4$ C5 C4 C3 C2 C1 C_{cap} C′ C″ in the notation of Richardson and Richardson (1988).

Table II: ^{1}H Chemical Shifts[a] of Assigned Resonances of Peptide S3

residue	NH	C_αH	C_βH	C_γH	C_δH	other ^{1}H
acetyl		1.97				
Y1	8.32	4.50	2.95			6.82, 7.12
M2	8.18	4.49	1.99/1.87	2.50		ϵCH_3 2.03
S3	8.69	4.46	4.31/4.03			
E4	9.03	4.08	2.09	2.38/2.33		
D5	8.50	4.45	2.66/2.61			
E6	8.00	4.08	2.06	2.38/2.26		
L7	8.40	4.09	1.78	1.64	0.88	
K8	8.22	4.12	1.92	1.46	1.70	ϵCH_2 2.99
A9	7.97	4.22	1.49			
A10	8.13	4.22	1.51			
E11	8.22	4.12	2.13	2.45/2.28		
A12	7.92	4.13	1.44			
A13	7.98	4.17	1.49			
F14	7.99	4.42	3.22			7.24, 7.36
K15	7.96	4.14	1.85	1.58	1.70	ϵCH_2 3.02
R16	7.87	4.23	1.80/1.66	1.57	3.19	ϵNH 7.31
H17	8.29	4.74	3.29/3.20			7.28, 8.49
G18	8.33	4.08				
P19[b]		4.54	2.34	2.06	3.68	
P19′[b]		4.52	2.34	2.02	3.62	
T20[c]	8.34	4.35	4.30	1.23		
T20′[c]	8.56	4.37	4.25	1.24		
terminal	7.58/7.24					

[a] Chemical shifts are measured relative to internal TSP [sodium (trimethylsilyl)propionate]. [b] P19 and P19′ correspond to the trans and cis conformation. [c] T20 and T20′ correspond to the trans and cis of P19.

with both these mechanisms (Richardson & Richardson, 1988), and analyses of substitutions in proteins have confirmed the role of side chains in stabilizing helix ends (Serrano & Fersht, 1989; Bell et al., 1992; Serrano et al., 1992). We present here new data on substitutions in a system of short helical peptides that reveal capping effects in its conformation of isolated α helices (Lyu et al., 1992b) and a detailed structural model for the peptide with serine at the N-cap site, abbreviated here as S3.

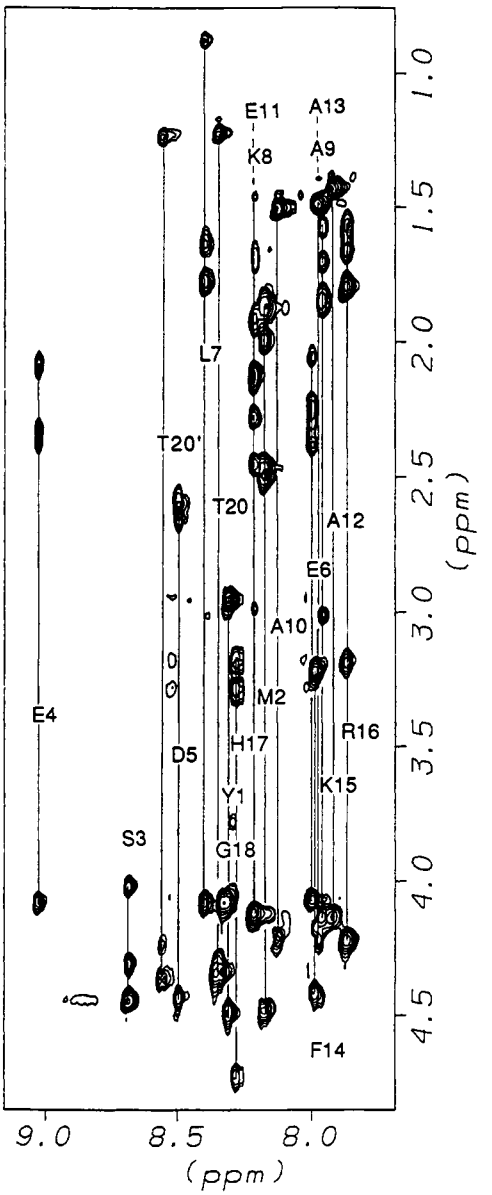

FIGURE 2: Fingerprint region of the TOCSY spectrum of S3 under the same conditions specified in Figure 1. The mixing time is 98 ms. Each intraresidue spin system is connected with solid lines and labeled as in Figure 1.

MATERIALS AND METHODS

Peptide Synthesis. Peptides were synthesized on a MilliGen/Biosearch 9600 automated synthesizer using Fmoc[1] chemistry and purified by HPLC as described previously (Lyu et al., 1990). The correct molecular weight of each product peptide was determined by electrospray ionization mass spectroscopy, using a Vestec Model 200 mass spectrometer.

CD Spectroscopy. CD spectra were recorded on an Aviv DS60 spectropolarimeter equipped with an HP 89100A temperature controller. The wavelength of the instrument was calibrated with (+)-10-camphorsulfonic acid (Chen & Yang, 1977). Measurements were performed in 10 mM KF

[1] Abbreviations: Fmoc, 9-fluorenylmethoxycarbonyl; CD, circular dichroism; 2D NMR, two-dimensional nuclear magnetic resonance; DQF-COSY, double-quantum-filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser and exchange spectroscopy; FID, Fourier induction decay; A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; K, Lys; L, Leu; M, Met; N, Asn; Nle, norleucine; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr.

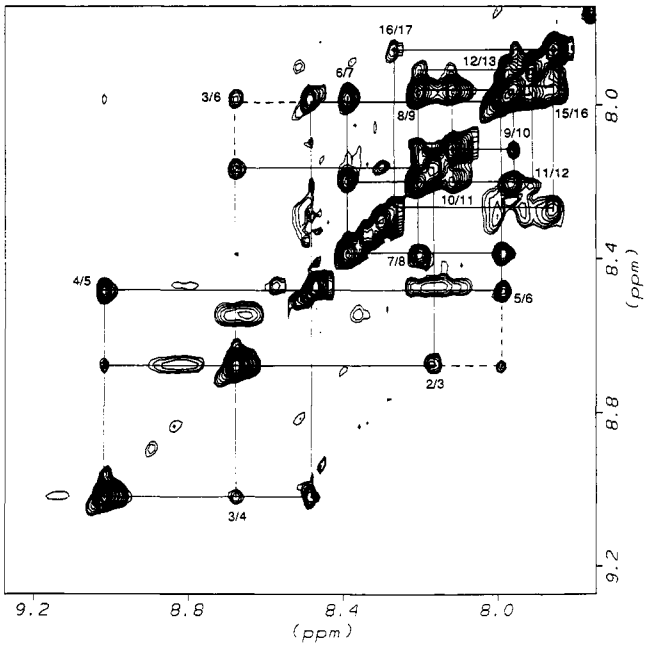

FIGURE 3: Amide–amide region of the NOESY spectrum of S3; τ_m = 350 ms. A series of cross peaks between adjacent peptide NH gruops are indicated by the sequence numbers of the pair of adjacent residues.

solutions at a peptide concentration of ca. 30 μM, averaging three scans with a 0.5-nm step size in each case.

^1H NMR Spectroscopy. Spectra were recorded on a General Electric GN500 spectrometer at 10 °C. Time-proportional phase incrementation (TPPI; Bodenhausen et al., 1984) was used to obtain phase-sensitive DQF-COSY (Rance et al., 1983), TOCSY (Bax & Davis, 1985), and NOESY (Macura & Ernst, 1981) spectra. Typical 2D data sets contained 512 FIDs with 2K data points each, requiring 32–96 scans/FID, a spectral width of 5208 Hz, and a recycle delay of at least 1 s. Spectra were Fourier transformed in both t_1 and t_2 dimensions after apodization with a skewed sine-bell function, typically with a 60° phase shift and skew of 0.6. Data in the second dimension were zero filled to obtain a final matrix of 1024 × 1024 real points. The first data point in t_1 was multiplied by 0.5 to suppress t_1 ridges (Otting et al., 1986). All data processing was carried out on a personal IRIS 4D/25GT workstation using the FELIX program (Hare Research, Inc.). The peptide sample was ca. 4 mM in H_2O/D_2O (85%/15%) at pH 5.7.

RESULTS

Position-Dependent Substitution Effects in Our Peptide Models. A system of model peptides containing each of the side chains that occur with highest frequency in the capping positions of helices in proteins according to the survey of Richardson and Richardson (1988) has been shown to form stable intramolecular α helices in water (Lyu et al., 1992b). These allow detailed investigation of the stability and mechanism of capping interactions in isolated helices. Table I summarizes the sequences and relative stabilities of 18 peptides in this series. The "parent" model peptide sequence contains serine at the third position, which has been assigned as the N-cap site in a preliminary 2D NMR study (Lyu et al., 1992b). S3 has a mean helix content of 50% α helix, calculated from the CD spectrum of the peptide at 4 °C and pH = 6 (Gans et al., 1991). For 20mers, 100% helix corresponds to $[\theta]_{222}$ = −32 000 deg·cm^2·dmol^{-1} (Gans et al., 1991). As in the case of other intramolecular peptide helices (Lyu et al., 1990, 1992a), the mean helix content is that of a population of chains in which there is rapid exchange at each residue between helix and coil conformations (Liff et al., 1991). The helix content of other chains in the series is summarized in Table I, revealing position-dependent effects of each substitution on the helicity of the molecule. At position 3, for example, the N-cap site, the relative order of stabilization is Ser > Asn > Gly > Ala. This differs from the order of stabilization seen at a mid-helix site in the same peptides, where at position 10, for example, we find Nle > Ala > Ser > Gly. This order is consistent with results from other models for mid-helical positions (O'Neil & DeGrado, 1990; Lyu et al., 1990, 1991; Ptitsyn, 1992). At the presumptive C-cap site, position 18, we observe instead Asn > Gln > Ser ~ Ala > Gly ~ Thr. This order differs from that at N-cap as well as the mid-helix one, reflecting the distinctive donor vs acceptor potentials and stereochemistry of these amino acid side chains (Presta & Rose, 1988). The behavior at intermediate end sites is also of interest: at position 6, for example, Glu ≫ Ala > Gly. Substitution of Ala or Ser

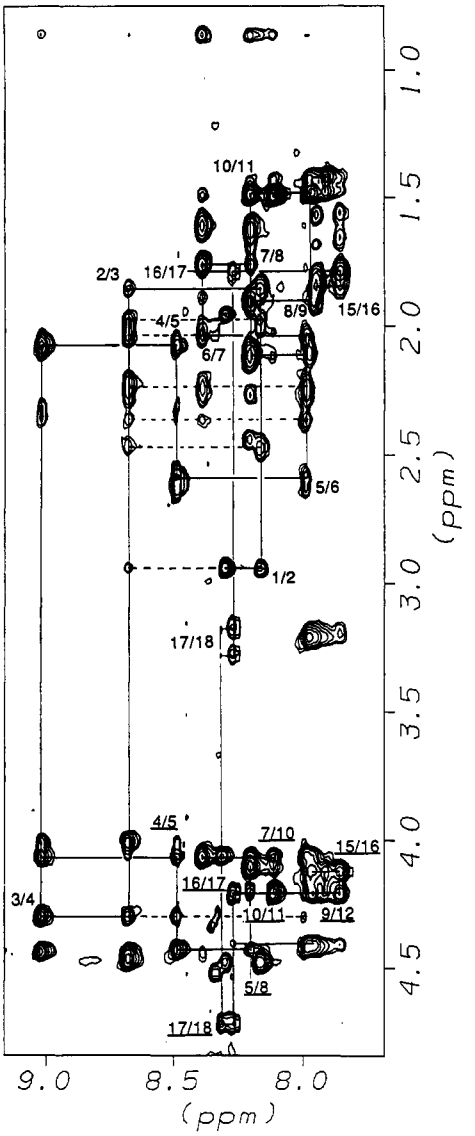

FIGURE 4: Section of the NOESY spectrum of S3, indicating cross peaks between backbone amide protons and C_α protons and side chains. Cross peaks between C_β protons and amide NH are indicated by the residue numbers involved. Underlined numbers indicate cross peaks between C_α protons and amide NH protons. The unusual cross peaks in the spectrum are connected by dashed lines; from top to bottom these are C_βH (Ser3)–NH (Asp5, Glu6), NH (Ser3)–C_γH (Met2), and C_γH–C_βH (Glu6).

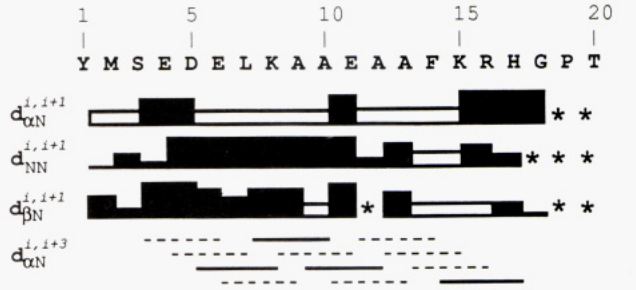

FIGURE 5: Schematic summary of NOE connectivities in the sequence of S3. Filled squares and solid lines indicate unambiguous NOE's. Empty squares and dashed lines indicate possible NOE's where chemical shift degeneracy interferes with identification. Asterisks inidicate the absence of a cross peak.

for the His at position 17 also reduces the helix content, as seen in Table I. These effects are distinct from the other positions examined, including mid-helical positions, and the N- and C-cap positions. It is not unlikely that each position in the capping region will exhibit a distinct order of stabilization imposed by specific side-chain interactions, as is seen in the frequencies of occupation of these positions by different side chains (Richardson & Richardson, 1988).

^{1}H NMR Resonance Assignments. To establish the structural basis for these interactions, we have performed a 2D NMR analysis of the structure in S3 and derived a detailed structural model for the serine capping interaction in an isolated α helix. Assignment of the ^{1}H spectrum of S3 followed the classical sequential procedure (Wuthrich, 1986) and is summarized in Table II. Figure 1 shows the region of the DQF-COSY spectrum that contains cross peaks between C_α protons and backbone amide proton resonances. There are 19 NH–C_α cross peaks corresponding to 19 of the 20 residues in S3. Two cross peaks at 8.34 and 8.56 ppm are observed for Thr20, which we attribute to the trans and cis conformations of Pro19, which lacks a cross peak in this region. Lys8 and Glu11 show an overlapping cross peak at 8.22 ppm. Intraresidue spin systems are identified in the fingerprint region of the TOCSY spectrum (Figure 2), and these confirm the Lys8-Glu11 overlap as well as the identity of Thr20 and Thr20′. The intensity of Thr20 is 3 times that of Thr20′, a value in the range expected for trans to cis isomerization of Pro19 (MacArthur & Thornton, 1991).

The amide–amide region of the NOESY spectrum is shown in Figure 3. A series of cross peaks connecting adjacent NH protons, characteristic of α-helical structure (Wuthrich, 1986), can be unambiguously identified from residues 2 to 13 and from residues 15 to 17. This is consistent with the extent of helical structure detected by CD (Table I). Cross peaks between NH protons and α or β protons three residues away in the fingerprint region (Figure 4) of the NOESY spectrum confirm the presence of an α helix. Figure 5 summarizes all the NOE connectivities observed between backbone protons in S3.

Unusual Cross Peaks in the NOESY Spectrum. Up to this point, the structure of S3 appears to be a classical α helix, based on the pattern of connectivities summarized in Figure 5 (Wuthrich, 1986). However, several unusual cross peaks are observed in the NOESY that do not correspond to standard α-helix patterns. First, we note the amide–amide cross peak between Ser3 and Glu6 indicated by the dashed line in Figure 3. Second, the amide of Ser3 shows additional cross peaks with side-chain protons: C_βH of Tyr1, C_βH and C_γH of Met2, and C_βH and C_γH of Glu6. Third, additional cross peaks can be detected between the C_βH's of Ser3 and backbone amides of Glu4, Asp5, and Glu6 (Figure 4). Introducing these constraints into a structural model for the N terminus of S3 leads to the conformation illustrated in Figure 6. This model has the following features. The amide of Ser3 points toward the C-terminal direction rather than the N terminus as in a regular α helix. This is required by the short distance (<3.5 Å; Wuthrich, 1986) between the amide NH protons of Ser3 and Glu6. Thus Ser3 has its CO positioned to H-bond with the NH of Leu7 at N4 (see legend to Table I), while its own NH is not in a helical configuration, as should be true for the N-cap residue (Richardson & Richardson, 1988). Formation of a hydrogen bond between the O_γ of Ser3 and the NH of Glu6 (at N3; see legend to Table I) accounts for all the short distances imposed by the NOE cross peaks described above. As can be seen in Figure 6, C_βH of Ser3 is close to the NH's of Glu4, Asp5, and Glu6. Since the NH of Asp5 lies within H-bonding distance of Ser O_γ, it seems reasonable to suggest that the O_γ shares a bifurcated H-bond with this group as well.

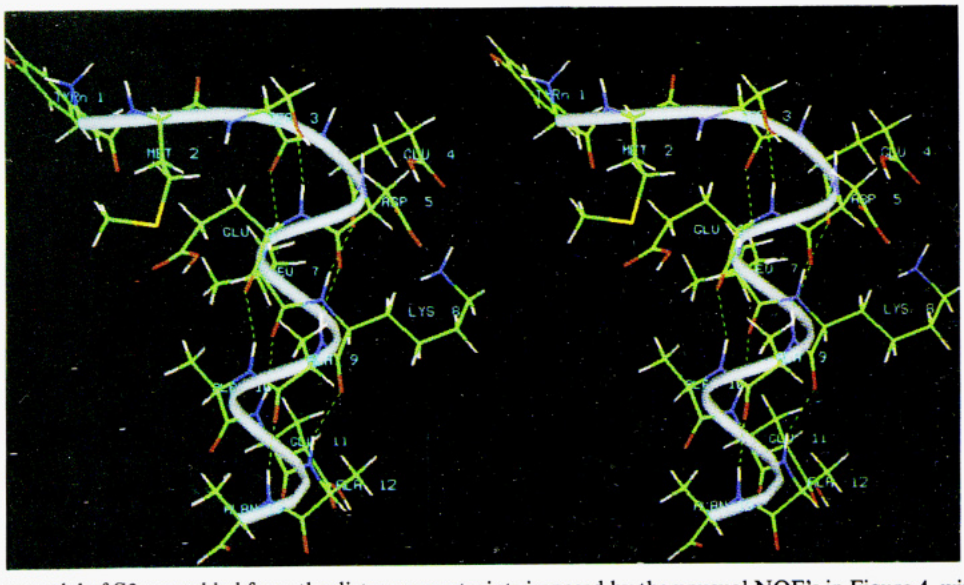

FIGURE 6: Molecular model of S3 assembled from the distance constraints imposed by the unusual NOE's in Figure 4, with the connectivities in Figure 5. Only the N terminus of the chain is shown for clarity.

DISCUSSION

This structural model for the cap region in an isolated helix shows that back-bonding from the strongly helix-stabilizing Ser3 side chain at N-cap to the main chain at N3 or possibly to both the main chains at N3 and N2 (see legend to Table I) is involved, as postulated by Presta and Rose (1988). Comparable models can be built with Asn, Asp, Gln, or Glu at N-cap, so that the interaction is not specific for Ser. However, the relative efficiency of a side chain in this position will depend on the donor group as well as the steric properties of the side chain. The detailed model in Figure 6 further suggests that the capping effect of one side chain will be influenced by neighboring side chains in the end positions. For example, the short distances observed between Ser at N-cap and the NH of residues at N1, N2, and N3 (see legend to Table I) indicate that adjacent side chains can alter the detailed arrangement, whether or not they are capable of capping. This implies that there will be interactions among side chains in the end positions. Just as vicinal effects have been proposed to arise in mid-helical positions (Ptitsyn, 1992), we can anticipate that structural and energetic differences at the ends of a helix will involve next-neighbor or even more remote interactions. The orders of stabilization reported here for different positions in our peptide models probably reflect such effects already; the apparent absence of capping effects in other model peptides has been noted also (Chakrabartty et al., 1992). Rose has pointed out that the high frequency of Ser at N-cap and of Glu at N3 reflects a reciprocal interaction in which the side chain of Ser interacts with the backbone amide of Glu while that of Glu can simultaneously bond to the amide of Ser. This two-bond interaction, which can exert a strong influence on nucleation of the helix and folding of the N terminus, can be demonstrated directly in this series of peptide models and will be reported elsewhere. The present models thus should make it possible to determine both individual and cooperative effects of capping side chains at each end position and, hence, provide a useful reference free energy for interpreting substitution data in proteins, much of which involves capping positions (Serrano & Fersht, 1989; Serrano et al., 1992), as well as for improving helix prediction schemes (Fasman, 1989).

ACKNOWLEDGMENT

We are grateful to George Rose for a number of stimulating discussions throughout this work.

REFERENCES

Barskaya, T. V., & Ptitsyn, O. B. (1971) *Biopolymers 10*, 2181–2197.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Bell, J. A., Becktel, W. J., Sauer, U., Baase, W. A., & Matthews, B. M. (1992) *Biochemistry 31*, 3590–3596.
Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.
Bychkova, V. E., Ptitsyn, O. B., & Barskaya, T. V. (1971) *Biopolymers 10*, 2161–2180.
Chakrabartty, A., & Baldwin, R. L. (1992) Abstracts, Sixth Symposium, Protein Society, San Diego, S57.
Chen, G. C., & Yang, J. T. (1977) *Anal. Lett. 10*, 1195–1207.
Fasman, G. D. (1989) *Prediction of protein structure and the principles of protein conformation*, Plenum, New York.
Gans, P. J., Lyu, P. C., Manning, M. C., Woody, R. W., & Kallenbach, N. R. (1991) *Biopolymers 31*, 1605–1614.
Levitt, M. (1978) *Biochemistry 17*, 4277–4285.
Liff, M. I., Lyu, P. C., & Kallenbach, N. R. (1991) *J. Am. Chem. Soc. 113*, 1014–1019.
Lyu, P. C., Liff, M. I., Marky, L. A., & Kallenbach, N. R. (1990) *Science 250*, 669–673.
Lyu, P. C., Sherman, J. C., Chen, A., & Kallenbach, N. R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5317–5320.
Lyu, P. C., Gans, P. L., & Kallenbach, N. R. (1992a) *J. Mol. Biol. 223*, 343–350.
Lyu, P. C., Zhou, H. X., Jelveh, N., Wemmer, D. E., & Kallenbach, N. R. (1992b) *J. Am. Chem. Soc. 114*, 6560–6562.
MacArthur, M. W., & Thornton, J. M. (1991) *J. Mol. Biol. 218*, 397–412.
Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95–117.
O'Neil, K. T., & DeGrado, W. F. (1990) *Science 250*, 646–651.
Otting, G., Widmer, H., Wagner, G., & Wuthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.
Padmanabhan, S., Marqusee, S., Ridgeway, T., Lane, T. M., & Baldwin, R. L. (1990) *Nature 344*, 268–270.
Pauling, L., & Corey, R. B. (1951) *Proc. Natl. Acad. Sci. U.S.A. 37*, 251–256.
Presta, L. G., & Rose, G. D. (1988) *Science 240*, 1632–1641.
Ptitsyn, O. B. (1992) *Curr. Opin. Struct. Biol. 2*, 13–20.
Rance, M., Sorenson, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648–1652.
Schellman, C. (1981) in *Protein Folding* (Jaenicke, R., Ed.) pp 53–61, Elsevier/North-Holland, Amsterdam.
Serrano, L., & Fersht, A. R. (1989) *Nature 342*, 296–299.
Serrano, L., Neira, J.-L., Sancho, J., & Fersht, A. R. (1992) *Nature 356*, 453–455.
Stickle, D. F., Presta, L. G., Dill, K. A., & Rose, G. D. (1992) *J. Mol. Biol.* (in press).
Sueki, M., Lee, S., Powers, S. P., Denton, J. B., Koniishi, Y., & Scheraga, H. A. (1984) *Macromolecules 17*, 148–155.
Wuthrich, K. (1986) *NMR of proteins and nucleic acids*, pp 162–175, Wiley, New York.