# Site-Specific Phosphorylation of Lys-Ser-Pro Repeat Peptides from Neurofilament H by Cyclin-Dependent Kinase 5: Structural Basis for Substrate Recognition

Pushkar Sharma,[‡] Joseph J. Barchi, Jr.,[§,||] Xiaolin Huang,[§] Niranjana D. Amin,[‡] Howard Jaffe,[‡] and Harish C. Pant*[,‡]

*Laboratory of Neurochemistry, National Institute of Neurological Disorders and Stroke, and Laboratory of Medicinal Chemistry, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892*

*Received November 6, 1997; Revised Manuscript Received January 20, 1998*

ABSTRACT: Recent work has shown that high molecular weight neurofilament (NF) proteins are phosphorylated in their carboxy-terminal tail portion by the enzyme cyclin-dependent kinase 5 (CDK-5). The tail domain of neurofilaments contains 52 tripeptide repeats, viz. Lys-Ser-Pro, which mainly exist as KSPXK and KSPXXX motifs (X = amino acid). CDK-5 specifically phosphorylates the serine residues within the KSPXK sites. We probed the structural basis for this type of substrate selectivity by studying the conformation of synthetic peptides containing either KSPXK or KSPXXX repeats designed from native neurofilament sequences. Synthetic peptides with KSPXK repeats were phosphorylated on serine with a recombinant CDK-5/p25 complex whereas those with KSPXXX repeats were unreactive in this system. Circular dichroism (CD) studies in 50% TFE/$H_2O$ revealed a predominantly helical conformation for the KSPXXX-containing peptides, whereas the CD spectra for KSPXK-containing peptides indicated the presence of a high population of extended structures in water and 50% TFE solutions. However, detailed NMR analysis of one such peptide which included two such KSPXK repeats suggested a turn-like conformation encompassing the first KSPXK repeat. Restrained molecular dynamics calculations yielded an unusually stable, folded structure with a double "S"-like bend incorporating the central residues of the peptide. The data suggest that a transient reverse turn or loop-type structure may be a requirement for CDK-5-promoted phosphate transfer to neurofilament-specific peptide segments.

One of the major roles of cyclin-dependent kinase 5 (CDK-5)[1] is the phosphorylation of the tail portion of high molecular weight neurofilament protein (NF-H). The tail portion of NF-H contains 52 sites of a repeating tripeptide unit, Lys-Ser-Pro (KSP), the majority of which exist as either KSPXK or KSPXXX motifs. The number and distribution of these sites are species dependent (*1*). The majority of these repeats in mouse and rat NF-H are KSPXXX, and 40 contiguous KSPXXX repeats form a KSP "core". In contrast, human NF-H tails contain a higher number of KSPXK repeats, and the core is made up of 16 repeat units of this amino acid sequence. Most of the serines in these KSP motifs are heavily phosphorylated in native NF-H (*2*). There is evidence that the phosphorylation of neurofilament tails is important for their transport down the axon and for the potentiation of axonal caliber (*3*). In certain neurodegenerative disorders such as ALS, abnormal phosphorylation of NF-H leads to the accumulation of NFs in neuronal cell bodies (*4*). In vivo electron microscopy shows that the NF-H tail extends out radially from the filaments and forms a network with other cytoskeletal elements (*5*). Neuronal CDK-5 phosphorylates neurofilament tail domains at S/T residues, only in KS/TPXK motifs and not in any other KS/TP-containing segments such as KSPXXK or KSPXXX (*6*, *7*). Neuronal activators of this kinase, such as p35 or its truncated form p25, have been discovered which are responsible for the activation of this kinase in neurons, unlike other CDKs which are activated by cyclins in dividing cells (*8*, *9*). Although the structure of p35/p25 is not known, it has been predicted to be similar to the cyclins (*10*, *11*). However, recent crystal structures of CDK-2 (*12*) and its complex with cyclin (*13*, *14*) offered some insight into the mechanism of activation of CDKs by their regulatory cofactors. To date, there is no known structure of a CDK complexed with a peptide substrate, and hence the conformational requirements for phosphate transfer by these enzymes are still unclear. However, it has been shown that all CDKs specifically phosphorylate S/T residues in segments of the type S/TPXK/R (*15*), which implies a highly specific interaction between the CDKs and their substrates.

The only reports regarding the conformation of KSP-repeat peptides are based on CD studies performed on KSPXXX-containing peptides from NF-M (*16*). These CD studies suggested that an alternating $\beta$-turn or helical structure defined these peptides. They also showed that phosphorylation of the serines in these repeats disrupted the folded

* Address correspondence to this author at the Laboratory of Neurochemistry, NINDS, NIH, Building 36, Room 4D20, Bethesda, MD 20892. E-mail: hcp@codon.nih.gov. Phone: 301-402-2124.

[‡] Laboratory of Neurochemistry.

[§] Laboratory of Medicinal Chemistry.

[||] To whom inquiries regarding NMR should be addressed.

[1] Abbreviations: NF, neurofilaments; NF-H, high molecular weight NF protein; NF-M, medium molecular weight NF protein; ALS, amyotrophic lateral sclerosis; CDK-5, cyclin-dependent kinase 5; CD, circular dichroism; NMR, nuclear magnetic resonance; MS, mass spectrometry; LC, liquid chromatography; TFE, trifluoroethanol; NOE, nuclear Overhauser effect; PKA, cAMP-dependent protein kinase.

S0006-2960(97)02746-3 
Published on Web 03/17/1998

Table 1: CDK-5 Phosphorylation of KSP Peptides from NF-H[a]

| peptide | phosphorylation |
|---|---|
| VKSPAKEKAKSPEK (**1**) | + |
| KSPEKAKSPVKEEA (**2**) | + |
| EAKSPAEAKSPAEAK (**3**) | − |
| TKSRVKEEAKSPGEA (**4**) | − |

[a] Kinase assays were carried out as discussed in Materials and Methods.

structure of the peptides. The objective of the work herein was to probe the structural properties of KSPXK motifs which specifically enable them to be recognized and phosphorylated by CDK-5 and to compare their structural features with those of the KSPXXX constructs. To this end, peptides containing either KSPXK or KSPXXX repeats from mouse and human NF-H were synthesized. Circular dichroism spectra revealed that peptides with KSPXK motifs folded differently from the KSPXXX-containing peptides in 50% TFE/$H_2O$. We show that only KSPXK-type repeats are substrates for CDK-5 and describe the structural features of one of these peptides containing two KSPXK repeats using two-dimensional NMR techniques and molecular modeling. Results from mass spectral analysis of the phosphorylation of this peptide were consistent with CDK-5 having a higher affinity for the N-terminal serine within the KSPAK site. Our results emphasize that there is a unique structural specificity for neuronal substrates which are recognized by CDK-5.

## MATERIALS AND METHODS

*Materials*. Two peptides of each type (KSPXK and KSPXXX) were synthesized commercially (Peptide Technologies Inc). The sequences were derived from the NF-H tail portion (C-terminal region) of mouse and human neurofilament proteins (Table 1). CDK-5 and p25 were expressed and purified as GST fusion proteins as reported earlier (*8*, *9*). All chemicals and reagents were purchased from Sigma.

*Phosphorylation Assay*. Standard assay mixtures contained 50 mM Tris, pH 7.4, 2.5 mM $MgCl_2$, 1 mM EGTA, 150−165 ng of CDK-5 complexed with a molar excess of p25 (preincubated enzyme preparation), 100 $\mu$M [$\gamma$-$^{32}$P]ATP, and 0.1−1.3 mM peptide substrate in a total volume of 30 $\mu$L. Reactions were initiated by adding [$\gamma$-$^{32}$P]ATP and were carried out at 30 °C for 60−80 min. Reactions were terminated by adsorption of the assay mixture onto phosphocellulose paper (Whatman). Phosphopeptide formation was measured by counting the radiolabeled $^{32}$P incorporation after several washings of the phosphocellulose paper with 75 mM phosphoric acid before a final wash with ethanol.

*Mass Spectrometry*. Peptide **1** was phosphorylated by the procedure described above. The kinase assay mixture containing the phosphorylated peptide was analyzed by LC/MS/MS after being desalted on line with a peptide trap cartridge (Michrom) on a 1.0 × 150 mm Monitor C18 column (Column Engineering, Ontario, Canada). The HPLC system was coupled to a Model LCQ mass spectrometer (Finnigan, San Jose, CA) equipped with an electrospray interface (ESI). The LCQ was operated in "triple play mode" in which the instrument was set up to automatically acquire full scan, zoom scan (high-resolution scan), and tandem (MS/MS) spectra from all ions above a present threshold resulting from peptide peaks eluting from the HPLC column. MS data were acquired on a Gateway 2000 computer (Gateway, N. Sioux City, SD) and after conversion to unix format analyzed on an Alpha Station 200 (Digital Equipment, Co., Maynard, MA) utilizing the Bioworks software package (Finnigan, San Jose, CA). MS/MS spectra were searched utilizing the PEPSEARCH (Finnigan, San Jose, CA) or SEQUEST (version B22, by J. Eng and J. Yates, Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA 98195-7730) programs against a database constructed from the published mouse NF-H sequence. Search parameters were set to consider possible phosphorylation (+80) at all serine residues. Fragment ions were labeled by utilizing the PEPMATCH program (Finnigan, San Jose, CA).

*Circular Dichroism*. Circular dichroic spectra were obtained on a Jasco J-720 spectropolarimeter. A constant dispersion of 1.0 nm was maintained by automatic slit width control. The spectra were measured from 180 to 240 nm with a time constant of 0.25 s and a scan speed of 10−50 nm/min. Signals were averaged 12 times before measurement. The sample concentrations were between 0.2 and 1.0 mM. The cell path length was either 0.2 or 1.0 cm.

*NMR Spectroscopy*. NMR spectra were collected on a Bruker AMX spectrometer operating at 500 MHz with an inverse, broad-band probe. Sample concentrations were between 6 and 8 mM dissolved in either 90% $H_2O/D_2O$, 50% TFE-$d_5$/$H_2O$, or 100% TFE-$d_5$. The pH of the aqueous sample was adjusted to 4.8 with phosphate buffer. Spectra were collected at 10 and 25 °C under the control of a Eurotherm variable temperature unit with an accuracy of ±0.1 °C. Two-dimensional spectra (DQF-COSY, TOCSY, ROESY, and NOESY) were recorded, employing standard pulse sequences with the number of acquisitions typically set to 64 for the NOESY (*17*), ROESY (*18*), and DQF-COSY (*19*) spectra and 32 for the TOCSY (*20*) spectra. Low-power presaturation was used to suppress the water resonance during the relaxation delay and the mixing period of the NOESY experiments. In general, spectra were recorded with 2K complex data points in $F_2$ for each of 480−512 $t_1$ increments with a sweep width of 5050 Hz in each dimension. TOCSY spectra were recorded with an isotropic mixing time of 65 ms and trim pulses of 2.5 ms immediately before and after the spin lock period. NOESY spectra were acquired with mixing times of 250 and 400 ms. The 400 ms NOESY experiments were used for integration of peak volumes to generate interproton distance constraints. The NOEs were classified into four groups of strong, medium, weak, and very weak with upper bounds set to 3.0, 3.5, 4.0, and 5.0 Å, respectively. The lower bound limit was set to 1.9 Å, and 0.5 Å was added to the upper bound for NOEs to methyl groups. The temperature coefficients of the amide protons were studied by collecting TOCSY spectra at seven different temperatures between 5 and 35 °C in 5 deg increments and are reported in −ppb/K. All spectra were processed with UXNMR on an X32 computer, and peak picking and volume integration were done with NMRCOMPASS (Molecular Simulations Inc.) after transfer to a Silicon Graphics INDIGO workstation. The data were zero-filled to 1024 points in $F_1$ prior to Fourier transform, multiplied

by a shifted (45−60°) sine function, and baseline corrected with a polynomial of order 5.

*Molecular Modeling and Dynamics*. Model building and molecular dynamics simulations were performed on an SGI Indigo workstation in the context of the QUANTA molecular modeling package (version 4.1, Molecular Simulations, Inc.). The CHARMm force field (*21*) was used for all calculations coupled either with QUANTA or with a stand-alone version on a DEC AlphaServer 2100 (Digital Equipment Corp., Marlboro, MA). Energy minimizations were typically computed until convergence (defined as an energy gradient of 0.001 kcal $mol^{-1}$), using the adopted basis Newton Raphson (ABNR) algorithm as implemented in QUANTA. Molecular dynamics simulations were carried out on structures either in vacuo using a distance-dependent dielectric constant ($D = r$) or in a system hydrated with a 10 Å sphere of preequilibrated water molecules. A shifted potential was used to a distance of 12 Å with a nonbonded cutoff of 14 Å. The nonbonded lists were updated every 25 steps of dynamics. A typical dynamic run was carried out by first heating the system from 0 to 300 K over 10 ps with a time step of 0.001 ps. The system was then equilibrated for 20 ps, and a constant-temperature dynamic simulation was then performed for 100 ps. The simulation trajectory was recorded every 1 ps and subsequently analyzed to determine the behavior of the molecule. A SHAKE algorithm was used to constrain bonds to hydrogen to within $10^{-8}$ Å.

The restrained molecular dynamics simulations were also carried out following a protocol similar to that above, where the energy term for the distance restraints were added to the total potential energy of the system as a harmonic potential function. An iterative process was applied where additional NOEs were assigned from partially overlapped regions of the spectra based on the models generated from the simulations in water with unambiguous restraints. The best structures with lowest violation from this stage of the calculations were again energy minimized and subjected to restrained dynamic simulations at 278 K for 100 ps. The conformations with lowest violation and lowest energy were extracted on the basis of analyzing the final simulation trajectories.

## RESULTS

*Phosphorylation of Peptides **1**−**4***. Peptides **1** and **2** were derived from the NF-H sequence of mouse and human species, respectively, while peptides **3** and **4** were from the KSP repeat region of mouse NF-H. In vitro phosphorylation assays on peptides **1**−**4** were carried out using the recombinant CDK-5/p25 complex (see Materials and Methods) to discern which motif would act as a substrate for this kinase. The results obtained from these reactions are summarized in Table 1. Peptides **1** and **2** were phosphorylated rapidly while peptides **3** and **4** showed no signs of phosphate incorporation under identical reaction conditions, which is consistent with previous studies (*22*). The data suggest that CDK-5 specifically phosphorylates only KSPXK motifs and not analogous $KSP(X)_n$-containing sequences. It is known that CDK-5 prefers X to be a basic residue in the SPXK consensus sequence and the affinity for this motif is reduced dramatically when X is an acidic residue like E or D (*23*). The kinase has intermediate affinity for substrate if X is a small hydrophobic residue like A, G, or V. Of the two KSPXK sites contained in peptide **1**, X is either A or E. In order to compare the affinity of CDK-5 for these sites, LC/MS analysis of the product obtained from the reaction of peptide **1** with CDK-5 was performed as described in Materials and Methods. As expected, the more polar phosphorylated peptide eluted before the native sequence. Only one phosphorylated product was observed. The phosphopeptide was identified by examination of the positive mode MS/MS spectra for a dominant ion resulting from the neutral loss of 98 ($H_3PO_4$) for a singly charged and 49 ($H_3PO_4/2$) for a doubly charged ion for each phosphoserine. Figure 1 shows the tandem MS/MS spectrum from a doubly charged ion at *m*/*z* 803.9. The spectrum is dominated by a strong ion at *m*/*z* 755.5, which is consistent with a loss of 49 ($H_3PO_4/2$) from the parent ion at *m*/*z* 803.9 and flags the spectrum resulting from a monophosphopeptide (Figure 1). The tandem MS/MS spectrum was searched against a database constructed from the reported mouse NF-H sequence (Genbank) utilizing the PEPSEARCH (*24*) or SEQUEST program. The automated output (Table 2) of the search using PEPSEARCH unequivocally identified (rank = 1, correlation score = 1) the MS/MS spectrum as derived from the monophosphopeptide VKS*PAKEKAKSPEK. The normalized correlation score for the second ranked peptide was 0.624, which is indicative of a high degree of confidence in the assignment of the peptide with rank = 1. The failure to detect either a doubly phosphorylated peptide or a peptide phosphorylated at the second repeat is consistent with CDK-5

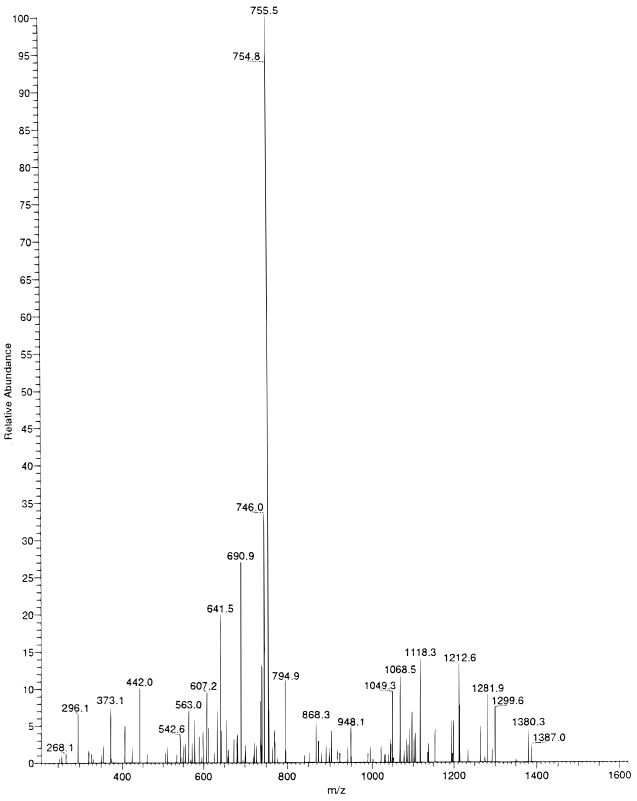


FIGURE 1: Full scan MS/MS spectrum from triple play analysis (Materials and Methods) of the *m*/*z* 804.5 doubly charged ion from the kinase assay mixture containing phosphorylated peptide **1** (VKS*PAKEKAKSPEK) (Table 1). The spectrum is dominated by a strong ion at *m*/*z* consistent with the loss of 98/2 ($H_3PO_4/2$) from the doubly charged parent ion at *m*/*z* 755.5, which flags this spectrum as resulting from a doubly charged monophosphopeptide.

Table 2: Result of PEPSEARCH Output from Phosphorylation of Peptide **1**[a]

| no. | rank | $C_n$ | $\Delta C_n$ | protein | mult/ ions | subsequence[c] |
|---|---|---|---|---|---|---|
| 1 | 1 | 1.000 | 0.000 | mNF-H[b] | 12/26 | (Q)VKS*PAKEKAKSPEK |
| 2 | 2 | 0.624 | 0.376 | mNF-H | 11/26 | (Q)VKSPAKEKAKS*PEK |
| 3 | 3 | 0.568 | 0.432 | mNF-H | 8/24 | (T)T*KTAEDT*KAKEP |
| 4 | 4 | 0.537 | 0.463 | mNF-H | 8/24 | (T)TKTEAEDTKAKEP |
| 5 | 5 | 0.536 | 0.464 | mNF-H | 12/26 | (E)QVKSPAKEKAKS*PE |
| 6 | 6 | 0.500 | 0.500 | mNF-H | 9/26 | (E)QVKS*PAKEKAKSPE |
| 7 | 7 | 0.465 | 0.534 | mNF-H | 8/28 | (E)VKS*PAEAKSPAEVKS |
| 8 | 8 | 0.437 | 0.563 | mNF-H | 8/28 | (P)VKEGAKSPAEAKS*PE |
| 9 | 9 | 0.422 | 0.578 | mNF-H | 8/34 | (F)GSADALLGAPFAPLHGG |
| 10 | 10 | 0.387 | 0.623 | mNF-H | 6/24 | (T)T*KT*EAEDTKAKEP |

[a] Table parameters as previously described (*38*, *39*). Rank signifies the result of analysis of the sequences using a cross-correlation function; $C_n$ is the normalized score from the cross-correlation function; $\Delta C_n$ is the difference between the cross-correlation parameter of the top-scoring sequence and the listed sequence; mult is the number of additional times a subsequence occurs in protein(s) in the database; ions is the number of ions of the type y, $y_o$, y*, b, $b_o$, b*, or a (*40*, *41*) observed in the MS/MS spectrum versus the number predicted. [b] mNF-H is mouse high molecular weight neurofilament protein. [c] S* and T* indicate phosphorylated residues.

specifically phosphorylating the KSPAK repeat in **1**. Similar results were obtained using the SEQUEST program (data not shown.

*Circular Dichroism Studies of NF Peptides.* The CD spectra for the four synthetic peptides are shown in Figure 2. All the peptides show typical random coil structures in water as evidenced by a large negative ellipticity at 200 nm. It is well-known that the addition of trifluoroethanol to an aqueous solution of an otherwise structurally random peptide may stabilize nascent, secondary structural elements by mimicking a membrane or intracellular environment (*25*, *26*). For KSPXXX peptides **3** and **4** (Figure 2b), it is evident that, on increasing the concentrations of TFE, there is a slight increase in negative ellipticity at 224 nm while the peak at 200 nm shifts to around 208 nm with a concomitant increase in ellipticity. As the concentration of TFE is increased to 50%, a positive peak at 190 nm appears along with a concomitant increase in the ellipticity at 222 nm, indicating the stabilization of transient helical character (*27*). Interestingly, this jump in TFE concentration also caused a major decrease in negative ellipticity for the peak at ~208 nm. The spectra of peptides **1** and **2** were only slightly different than those for peptides **3** and **4** in water (smaller negative peak at 222 nm) but behaved differently upon addition of small amounts of TFE. The shift in the negative peak at 200 nm was also observed with a corresponding increase in the magnitude of the negative ellipticity (Figure 2a). However, on raising the TFE concentration to 50%, a substantial increase in negative ellipticity is seen for the peak at 222 nm, and in contrast to peptides **3** and **4**, a further increase in negative ellipticity at the ~201 nm peak was observed with no corresponding positive peak at 190 nm. These data suggested that an extended conformation predominates for these peptides **1** and **2**. On the basis of the spectra, one could speculate that either a small population of helical folds is in rapid equilibrium with extended conformers or there were $\beta$-turn structures present in solution. Peptides containing $\beta$-turns are known to show negative ellipticity at 222 nm (*27*). The spectra for all the peptides studied were independent of the concentration in the range 0.1−1 mM.

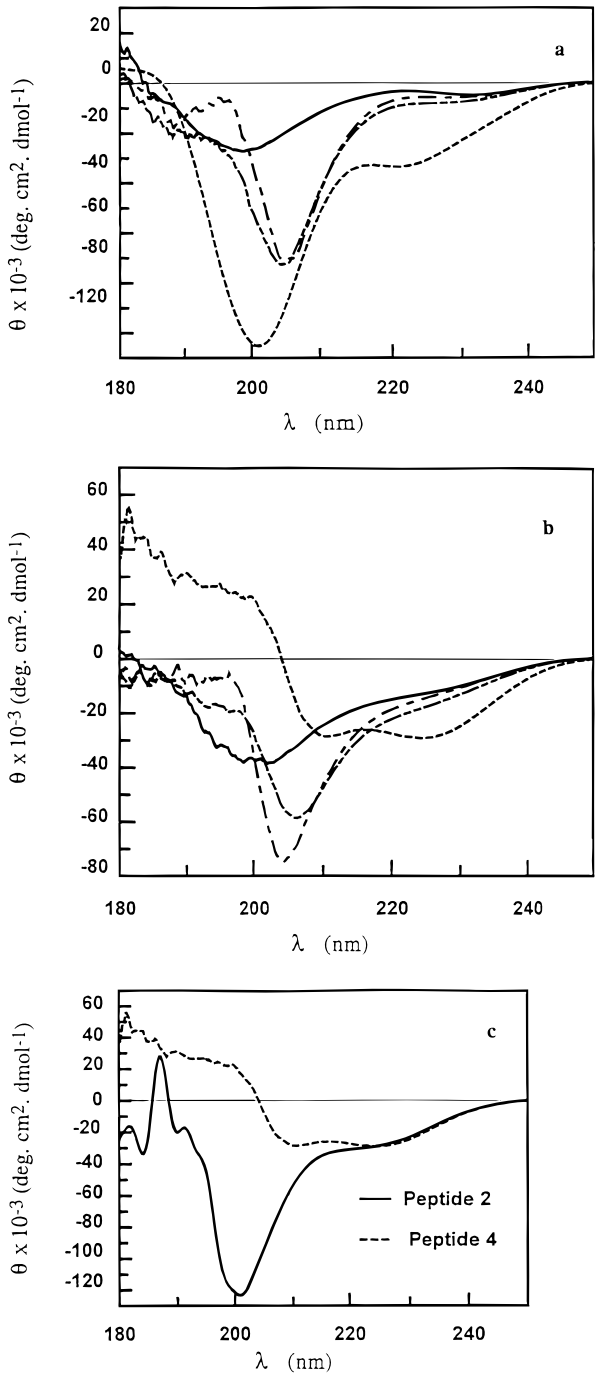


FIGURE 2: Circular dichroism spectra of (a) peptide **1** and (b) peptide **3** in different TFE concentrations: (—) water; (— - — ) 10% TFE; (— - - — ) 20% TFE; (- - -) 50% TFE. (c) CD spectra of peptides **2** and **4** in 50% TFE.

*NMR Structure of Peptide **1** in 50% TFE*. Only the KSPXK-containing peptides **1** and **2** were shown to be substrates for CDK-5. To investigate the structure of the KSPXK-containing peptide **1** in more detail, one- and two-dimensional NMR techniques were used. Since only minor differences were observed in the CD spectra of peptides **1** and **2** upon addition of TFE, we attempted to assign the NMR structure of peptide **1** in water. The NMR spectra of peptide **1** collected in 90% $H_2O/D_2O$, pH 4.8, showed a mainly random coil type conformation, even at low temperatures. Addition of 50% TFE increased the peak dispersion and resulted in the acquisition of additional intensity in many

Table 3: $^1$H NMR Chemical Shifts for Peptide **1**[a]

| residue | NH | αH | βH | γH | δH | ϵH |
|---|---|---|---|---|---|---|
| V1 | | 3.784 | 2.178 | 1.000 | | |
| K2 | 8.499 | 4.394 | 1.794 | 1.407 | 1.675 | 2.943 |
| | | | 1.730 | 1.407 | 1.675 | 2.943 |
| S3 | 8.210 | 4.815 | 3.910 | | | |
| | | | 3.784 | | | |
| P4 | | 4.406 | 2.269 | 1.998 | 3.784 | |
| | | | 1.937 | 1.998 | 3.784 | |
| A5 | 8.003 | 4.184 | 1.344 | | | |
| K6 | 7.945 | 4.205 | 1.828 | 1.449 | 1.680 | 2.964[b] |
| | | | 1.765 | 1.386 | 1.680 | 2.964 |
| E7 | 7.989 | 4.282 | 2.101 | 2.438 | | |
| | | | 1.996 | 2.438 | | |
| K8 | 8.081 | 4.226 | 1.828 | 1.428 | 1.695 | 2.964 |
| | | | 1.744 | 1.428 | 1.695 | 2.964 |
| A9 | 7.991 | 4.256 | 1.386 | | | |
| K10 | 7.991 | 4.340 | 1.828 | 1.427 | 1.659 | 2.964 |
| | | | 1.723 | 1.427 | 1.659 | 2.964 |
| S11 | 8.034 | 4.752 | 3.847 | | | |
| | | | 3.847 | | | |
| P12 | | 4.446 | 2.248 | 1.996 | 3.742 | |
| | | | 1.937 | 1.996 | 3.742 | |
| E13 | 8.073 | 4.352 | 2.080 | 2.459 | | |
| | | | 1.933 | 2.459 | | |
| K14 | 8.134 | 4.331 | 1.870 | 1.428 | 1.659 | 2.964 |
| | | | 1.765 | 1.428 | 1.659 | 2.964 |

[a] In 50% TFE/$H_2O$. [b] The ϵ proton methylene envelope of the four lysine residues resonates at 2.964 ppm.

correlations which could be attributed to the presence of transient secondary structural elements in this solvent system. Assignment of the NMR spectra of peptide **1** in 50% TFE/$H_2O$ was carried out by standard spin pattern identification using data from DQF-COSY and TOCSY spectra, while sequential connectivities were assigned from both NOESY and ROESY data (*28*). The chemical shifts of peptide **1** in 50% TFE/$H_2O$ are listed in Table 3.

Temperature coefficients for the amide protons of peptide **1** were generally greater than 6 ppb/K, which suggested that no strong hydrogen bonds or shielded NH groups were present under the experimental conditions. However, residues S3, K6, and K10 had coefficient values in the intermediate range (5.3), suggesting a more compact disposition of atoms in the vicinity of these residues. The assessment was confirmed by molecular modeling where these amides were shown to be within hydrogen-bonding distance to an acceptor atom (vide infra).

The NMR spectra of peptide **1** in 50% TFE/$H_2O$ exhibited predominantly *trans* X-Pro linkages as evidenced by the strong NOEs observed between the Hα protons of both serine residues with their neighboring proline Hδ protons (*29*). A very small percentage (<5%) of a minor isomer was detected at 288 K although the low concentration of this isomer precluded its assignment. The N-terminal valine methyl protons showed strong NOEs to a lysine amide proton, which implied that this resonance was K2NH. In spite of some spectral overlap, the sequential NOEs correlating with this NH and the following residues allowed nearly full assignment of the remaining amide, Hα, and side chain protons. The NOESY spectra exhibited strong $d\alpha N(i,i+1)$ NOEs between all residues (Figure 3a), suggestive of a predominantly random coil conformation in this solvent system. However, medium to strong $dNN(i,i+1)$ connectivities along with characteristic medium-range correlations were observed, suggesting that defined secondary structural elements were

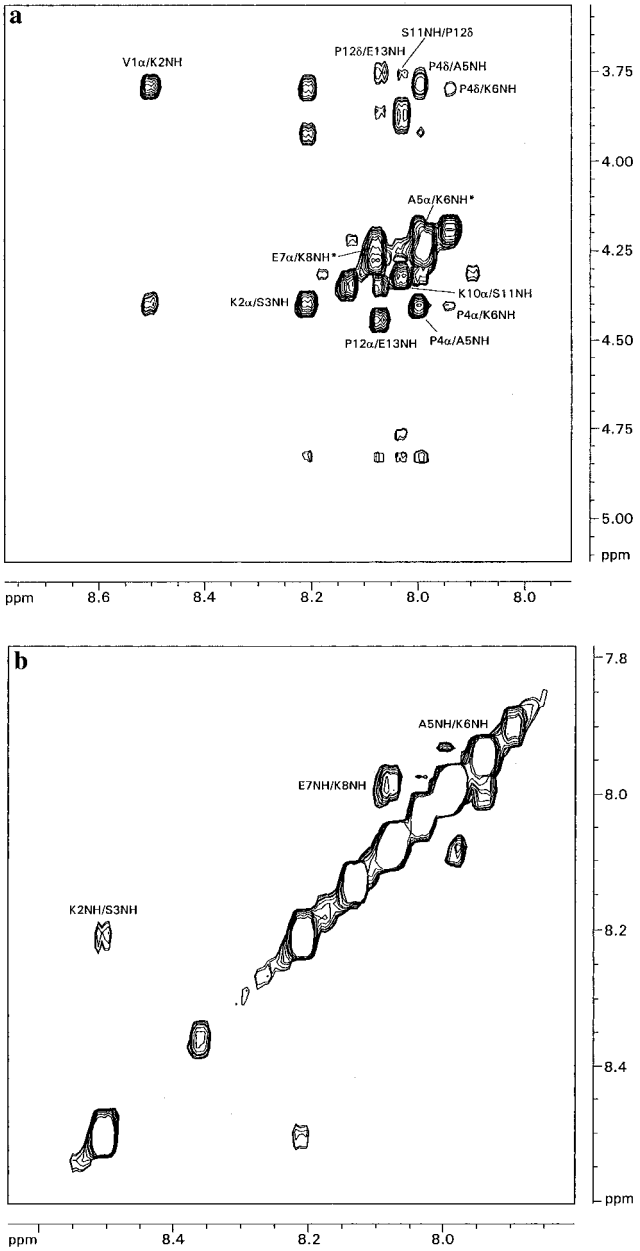


FIGURE 3: NOESY spectrum (400 ms mixing time) of (a) the fingerprint and (b) the NH−NH regions of peptide **1** in 50% TFE. Relevant peaks are labeled. Peaks labeled with an asterisk were assigned by examination of the β region of the spectra and by variable temperature measurements.

also populated in solution. A strong $d\alpha N(i,i+1)$ NOE was observed between P4 and A5 along with a strong $dNN(i,i+1)$ correlation from A5 to K6. This combination, along with a medium-range $d\alpha N(i,i+2)$ NOE between P4Hα and K6NH suggested that this SPXK repeat populates a β type II turn conformation (*28, 29*). A weak NOE between P4Hδ and A5NH indicated that a small percentage of a β type I turn may also be present. NOEs involving the second SPXK repeat located at the C-terminus of the peptide were observed between S11Hα−P12Hδ and P12δ−E13NH (Figure 3b). This could be interpreted as a transient bend in the C-terminal portion of the peptide. Since there were no other diagnostic or long-range NOEs near this end of the molecule, estimates of secondary structure at the second repeat are qualitative at best. In addition, due to the lack of long-range NOEs between the two SPXK repeats, two important structural units

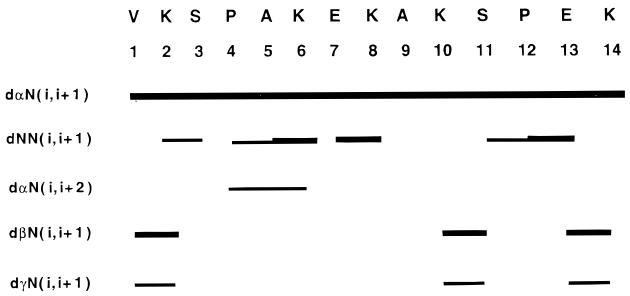


FIGURE 4: Schematic representation of the NOEs observed for peptide **1**. The thickness of the bars indicates the intensity of correlation.

in this small peptide were independent of each other under these conditions. An NOE between E7NH and K8NH, residues which effectively "link" the two repeat segments, indicated a kink in the center of peptide backbone. This could possibly be due to salt bridge formation between oppositely charged E and K residues (vide infra) or an effect of two turn-like structural units or bends being present at each peptide terminus. It was suggested by the data that the SPXK repeats which are phosphorylated by CDK-5 may be portions of structually flexible exposed loops. These results compare with studies done by Suzuki (*30, 31*), who showed that DNA binding motifs comprised of SPKK segments appear mostly as $\beta$-turns in proteins.

*Restrained Molecular Dynamics of Peptide **1***. Since peptide **1** was largely unstructured in solution even in a hydrophobic environment, the conformation of this peptide was studied by molecular modeling either alone or on the basis of data generated by NMR. Initial calculations were carried out with the CHARMm force field without initial restraints to explore the conformational space sampled by the peptide. Starting structures were generated by high-temperature dynamics in vacuo. A series of folded conformations were obtained as initial models. Some of these structures contained turn-like dispositions about the SPXK repeats while others were looped at different positions on the peptide backbone. These structures were used as input for a thorough molecular dynamics simulated annealing protocol using a set of 58 distance restraints generated from the analysis of the NOESY data. This consisted of heating the structures to 1000 K and slowly cooling back to 300 K while increasing the force constants on the distance potentials during cooling. In order to obtain a more realistic view of the lowest energy conformations of peptide **1** and to quench any artifactual ionic interactions which may be stabilized in vacuo, the calculations were performed in a 10 Å box of preequilibrated water molecules. Starting from structures with very different conformations, the calculations converged to a well-defined peptide fold in spite of the small number of NOE constraints. An overlay of 24 structures that were calculated and showed the lowest violations and energies is shown in Figure 5a. Figure 5b shows the structure which best fit the NMR data as calculated with this protocol. The convergence of the calculated structures to one family of conformations was very high; analysis of final restrained molecular dynamics trajectory revealed that all conformers were in stationary state after 10 ps simulation. The total energies of respective conformers were within a fluctuation of 4%. The rmsd for the 24 best structures selected was 0.97 Å for the backbone atoms and 1.13 Å for all heavy

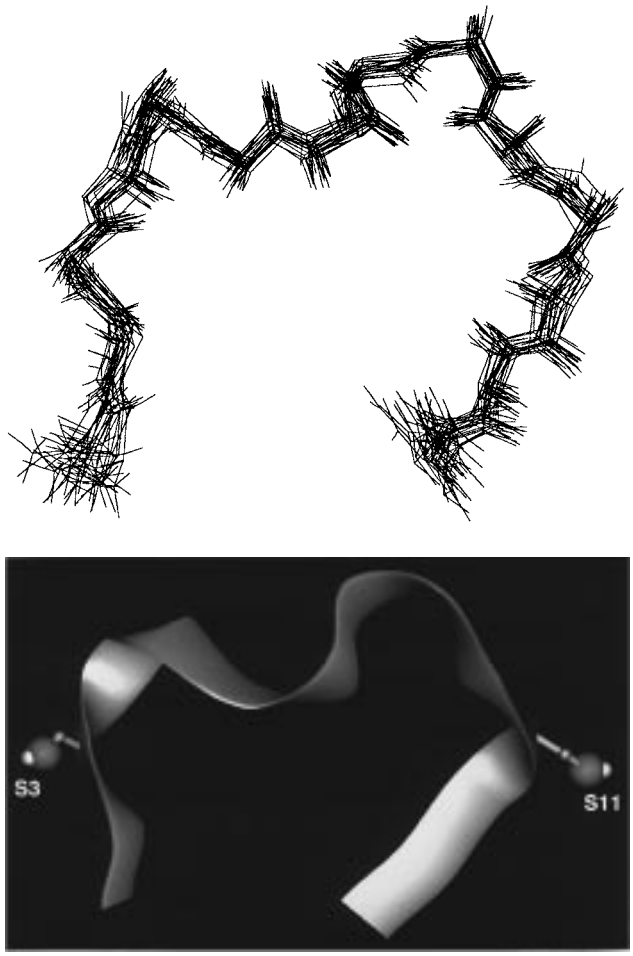


FIGURE 5: (a, top) Overlay of the backbone atoms of the 24 best structures (rmsd = 0.97 Å) calculated using several restrained molecular dynamics simulations based on NOE restraints using the protocol described in Materials and Methods. (b, bottom) Ribbon diagram of the best fit average structure of peptide **1** from those shown in (a). Serine side chains in KSPXK motifs are drawn with van der Waals volumes in white (H) and red (O).

atoms. For residues 3−10 the rmsd was 0.36 Å. The structure reveals an "S"-like shape in the central portion of the peptide (Figure 5) where the ends of the molecule loop together in a salt bridge involving the side chains of K2 and E13. This ionic interaction was very stable even in a water environment and remains intact even when the production dynamics were run for several nanoseconds (Steinbach, unpublished results). Hence, the fold depicted is very stable under these conditions.

The KSPXK repeats are part of the loops comprising the S-like curve in the backbone fold. Examination of these loops reveals that neither end of the peptide populates a standard $\beta$-turn in the calculated structures. Rather the first proline (P4) unit nucleates the onset of this double loop, and a more standard $\beta$ type I turn (III) is seen for residues E7-K8-A9-K10 (Figure 5). This turn seems to facilitate the double loop structure which encompasses the two repeats. This central $\beta$-bend is stabilized by the close proximity of the side chains of E7 and K10 (possible salt bridge) and by a hydrogen bond between the E7 carbonyl and the K10 NH. The second KSPXK repeat is more extended in this conformation. However, it is clear that the hydroxyl groups of the serine residues in these repeats (S3 and S11) are well exposed to the medium in the geometry of this fold (Figure

5). This may be a prerequisite for recognition and phosphorylation of these groups by specific kinases.

## DISCUSSION

The differences in the CD spectra of peptides **1** and **2** from peptides **3** and **4** confirmed that each motif populated a unique structural ensemble. We used TFE to stabilize the structures of these peptides as they showed mainly random structures in water. The CD spectra for the KSPXXX peptides in TFE suggested the presence of helical content and looked qualitatively similar to the data obtained for closely related peptides derived from NF-M KSPXXX repeats (*16*). They concluded that helical structures were present in peptides of similar size to those described herein. In contrast, the CD spectra of our KSPXK peptides revealed that only a small population of folded conformers was present in solution. The NMR data suggested the presence of a transient loop or $\beta$ type II turn at the SPAK motif within the first KSPXK repeat. In comparison, the second KSPXK repeat (X = E) did not present spectral data indicative of stable secondary structures under the conditions studied. These two repeats were joined by a $\beta$-turn in the center of the peptide (E7-K8-A9-K10) that gives rise to an S-shaped structure. Since the conformations of the repeats were independent of each other, it is possible that the KSPXK core region of human NF-H consists of these loops. Phosphorylation at these sites in NF-H might facilitate extensions of the modified side chains and formation of cross bridges with other cytoskeleton elements. These cross bridges have been observed previously by electron microscopy (*5*). A recent study performed in our laboratory revealed that a human NF-H KSPXK-containing core comprised of 16 contiguous KSPXK repeats could be fragmented by trypsin (Jaffe et al., unpublished results). However, the complementary rat KSP core which contains 16 continuous KSPXXX repeats was resistant to most proteases (*2*). It is possible that the human KSPXK core is comprised of continuous loops as was observed in our NMR structure. In these loops, the lysine side chains may be more accessible as compared to the more highly folded conformations suggested for the KSPXXX repeat core based on CD data on KSPXXX-containing peptides.

The mass spectrometric analysis of the phosphorylation of peptide **1** by CDK-5 showed the presence of only a monophosphorylated peptide. The site for phosphorylation was determined to be the serine in the first KSPAK motif. There was no detectable double phosphorylated peptide or a peptide phosphorylated at the serine of the second KSPEK repeat. These data confirm that CDK-5 has a higher affinity for an SP segment followed by a neutral small hydrophobic residue as compared to an acidic residue (*22*). Interestingly, the NMR data revealed that the KSPAK repeat seems to be slightly more well-defined structurally in peptide **1** than the second, KSPEK segment. It is possible that the proximity of E and K destabilizes the proper positioning of the substrate hydroxyl group by interfering in the formation of an active turn/loop disposition of this segment.

It has been suggested that secondary and tertiary structures are important factors in substrate recognition by protein kinases (*32*, *33*). The crystal structure of PKA (*34*) and its complex with an inhibitor/substrate peptide (*35*) revealed the importance of structural and sequential elements in substrate recognition. A recent report on the NMR structure of the inhibitor peptide alone showed that the portion of the peptide which contains the phosphoacceptor site adopts a conformation in solution different from that depicted in the crystal structure of the complex (*36*). This suggested that there is an induced conformational change upon binding to the enzyme. The inherent flexibility of our peptides in either water or partially hydrophobic solutions implies that the rigidity of the phosphoacceptor segments must be limited to adapt to the catalytic binding pocket of the enzyme.

The recent crystal structure of the cyclin-dependent kinase 2 (CDK-2) revealed a bilobal structure common to the protein kinase family (*13*, *14*). There is more than 80% homology between CDK-2 and CDK-5. Homology modeling of CDK-5 revealed high structural similarity with CDK-2 in the catalytic cleft (*37*). Since to date there are no crystal structures of a CDK/substrate complex, the atomic level interactions between this enzyme family and their substrates remain unclear. However, there have been suggestions that the important lysine of the SPXK consensus sequence for CDKs might be involved in an interaction with a triad of acidic residues (*13*). From our studies, it appears that CDK-5 seems to select an SPAK motif contained in a peptide which may form one or more turns or loops in solution. It is possible that such a conformation mediates the interaction between an essential lysine side chain of the substrate with some key acidic residues of the enzyme. The structure of a CDK/peptide substrate complex would be very useful in unravelling the specific interactions involved in the phosphate transfer process and in turn foster the design of specific inhibitors for particular CDKs.

## REFERENCES

1. Pant, H. C., and Veeranna (1995) *Biochem. Cell Biol. 73*, 575−592.
2. Elhanany, E., Jaffe, H., Link, W. T., Sheeley, D. M., Gainer, H., and Pant, H. C. (1994) *J. Neurochem. 63*, 2324−2335.
3. Dewaegh, S. M., Lee, V. M.-Y., and Brady, S. T. (1992) *Cell 68*, 35−46.
4. Julien, J. P. (1997) *Trends Cell Biol. 7*, 243−249.
5. Hisanaga, S., and Hirokawa, N. (1988) *J. Mol. Biol. 202*, 297−305.
6. Shetty, K. T., Link, W. T., and Pant, H. C. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6844−6848.
7. Sun, D., Leung, C. L., and Liem, R. K. H. (1996) *J. Biol. Chem. 271*, 14245−14251.
8. Lew, J., Huang, Q. Q., Qi, Z., Winkfein, R. J., Aebersold, R., Hunt, T., and Wang, J. H. (1994) *Nature 371*, 423−426.
9. Tsai, L. H., Delalle, I., Caviness, V. S., Chae, T., and Harlow, E. (1994) *Nature 371*, 419−423.
10. Tang, D., Chun, A. C. S., Zhang, M., and Wang, J. H. (1997) *J. Biol. Chem. 272*, 12318−12327.
11. Brown, N. R., Noble, M. E., Endicott, J. A., Garman, E. F., Wakatsuki, S., Mitchell, E., Rasmussen, B., Hunt, T., and Johnson, L. N. (1995) *Structure 3*, 1235−1274.
12. De Bondt, H. L., Rosenblatt, J., Jancarik, Jones, H. D., Morgan, D. O., and Kim, S. H. (1993) *Nature 363*, 595−602.
13. Jeffery, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N. P. (1995) *Nature 376*, 313−320.
14. Russo, A. A., Jeffrey, P. D., and Pavletich, N. P. (1996) *Nat. Struct. Biol. 3*, 696−700.
15. Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F., Piwnica, and Cantley, L. C. (1994) *Curr. Biol. 4*, 973−982.
16. Ortvas, L., Jr., Hollosi, M., Perczel, A., Dietzschold, B., and Fasman, G. D. (1988) *J. Protein Chem. 7*, 365−376.

17. Kumar, A., Wagner, G., Ernst, R. R., & Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun. 96*, 1156−1163.
18. Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., and Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106*, 811−813.
19. Rance, M., Sovensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
20. Bax, A., & Davies, D. G. (1985) *J. Magn. Reson. 65*, 355−360.
21. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
22. Veeranna, Shetty, K. T., Amin, N., Grant, P., Albers, R. W., and Pant, H. C. (1996) *Neurochem. Res. 21*, 629−636.
23. Beaudette, K. N., Lew, J., and Wang, J. H. (1993) *J. Biol. Chem. 268*, 20825−20830.
24. Wheeler, K., Mylchreest, I., Shoftstahl, J. R., Yates, III, and Eng, J. (1995) *TSQ Application Report No. 247*, Finnigan MAT, San Jose, CA.
25. Nelson, J. W., and Kallenbach, N. R. (1986) *Proteins 1*, 211−217.
26. Sonnichsen, F. D., Van Eyk, J. E., Hodges, R. S., and Sykes, B. D. (1992) *Biochemistry 21*, 8790−8798.
27. Woody, R. (1995) *Methods Enzymol. 246*, 34−71.
28. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons Inc., New York.
29. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., and Wright, P. E. (1988) *J. Mol. Biol. 201*, 161−200.
30. Suzuki, M. (1989) *J. Mol. Biol. 207*, 61−84.
31. Suzuki, M. (1989) *EMBO J. 8*, 797−814.
32. Geahlen, R. L., and Harrison, M. L. (1990) in *Peptides and Protein Phosphorylation* (Kemp, B. E., Ed.) pp 239−253, CRC Press, Boca Raton, FL.
33. Pearson, R. B., and Kemp, B. E. (1991) *Methods Enzymol. 200*, 62−81.
34. Knighton, D. R., Zheng, J., Tenzyck, L. F., Xuong, G. N., Taylor, S. S., and Sowadski, J. M. (1991) *Science 253*, 407−414.
35. Knighton, D. R., Zheng, J., Tenzyck, L. F., Ashport, V. A., Xuong, G. N., Taylor, S. S., and Sowadski, J. M. (1991) *Science 253*, 414−420.
36. Padilla, A., Hauer, J. A., Tsigelny, I., Parello, J., and Taylor, S. S. (1997) *J. Pept. Res. 49*, 210−220.
37. Sharma, P., Barchi, J. J., Huang, X., Amin., N. D., Albers, R. W., Pearlstein, R., and Pant, H. C. (1997) *FASEB J. 11*, 1184.
38. Eng, J. K., McCormak, A. L., and Yates, J. R., III (1994) *J. Am. Soc. Mass Spectrom. 5*, 976−989.
39. Yates, J. R. (1996) *Methods Enzymol. 271*, 351−377.
40. Biemann, K., and Martin, S. A. (1987) *Mass Spectrom. Rev. 6*, 1.
41. Roepstorff, P., and Fohlman, J. (1984) *J. Biomed. Mass Spectrom. 11*, 601.

BI972746O