BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 1577–1589 1577
DOI: 10.1021/bi901616z

The Basic Helix−Loop−Helix Region of Human Neurogenin 1 Is a Monomeric Natively Unfolded Protein Which Forms a "Fuzzy" Complex upon DNA Binding†

David Aguado-Llera,[‡,○] Erik Goormaghtigh,[§] Natalie de Geest,[∥,⊥] Xiao-Jiang Quan,[∥,⊥] Alicia Prieto,[#] Bassen A. Hassan,[∥,⊥] Javier Gómez,[‡] and José L. Neira*[,‡,△,○]

[‡]*Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, 03202 Elche (Alicante), Spain,* [§]*Centre for Structural Biology and Bioinformatics, Structure and Function of Biological Membranes, Université Libre de Bruxelles, B1050 Brussels, Belgium,* [∥]*Department of Molecular and Developmental Genetics, VIB, 3000 Leuven, Belgium,* [⊥]*Center for Human Genetics, KU Leuven School of Medicine, 3000 Leuven, Belgium,* [#]*Centro de Investigaciones Biológicas, CSIC, 28006 Madrid, Spain, and* [△]*Biocomputation and Complex Systems Physics Institute, 50009 Zaragoza, Spain.* [○]*These two authors contributed equally to this work*

Received September 16, 2009; Revised Manuscript Received November 30, 2009

ABSTRACT: Neuronal specification is regulated by the activity of transcription factors containing the basic helix−loop−helix motif (bHLH); these regulating proteins include, among others, the neurogenin (Ngn) family, related to the atonal family of genes. Neurogenin 1 (NGN1) is a 237-residue protein that contains a bHLH domain and is involved in neuronal differentiation. In this work, we synthesized the bHLH region of NGN1 (bHLHN) comprising residues 90−150 of the full-length NGN1. The domain is a monomeric natively unfolded protein with a pH-dependent premolten globule conformation, as shown by several spectroscopic techniques (namely, NMR, fluorescence, FTIR, and circular dichroism). The unfolded character of the domain also explains, first, the impossibility of its overexpression in several *Escherichia coli* strains and, second, its insolubility in aqueous buffers. To the best of our knowledge, this is the first extensive study of the conformational preferences of a bHLH domain under different solution conditions. Upon binding to two DNA E-boxes, the protein forms "fuzzy" complexes (that is, the complexes were not fully folded). The affinities of bHLHN for both DNA boxes were smaller than those of other bHLH domains, which might explain why the protein−DNA complexes were not fully folded.

In the initial stages of nervous system development in all animals, progenitor cells in the neuroepithelium generate cells that will differentiate into specific classes of neurons. This process is regulated by the activity of transcription factors containing the basic helix−loop−helix motif (bHLH),[1] such as the neurogenin (Ngn), atonal, and Mash1 proneural factors. The typical bHLH domain is approximately 60 residues long, comprising a DNA-binding basic region followed by two α-helices separated by a variable loop region (and, hence, the name HLH) (*1*). Through this region, the protein forms homo- or heterodimeric complexes with other members of the family (*2*). There are two basic subgroups of bHLH proteins within the bHLH superfamily of transcription factors: those containing the bHLH domain and those where the bHLH domain is contiguous to other regions, which additionally regulate dimerization (*3*). The two basic domains, brought together through the dimerization, bind specific hexanucleotide sequences, forming a stable, well-formed helix bundle (*4*, *5*). The consensus DNA sequence targeted by several bHLH proteins is the hexameric 5′-CANNTG-3′ element, referred to as "E-box" (*6*, *7*).

The bHLH motifs were first identified in the murine transcription factors E12 and E47 (*8*); the proneural genes of the bHLH class were first identified in *Drosophila* and then in vertebrates as key regulators of neural lineage development (*9*, *10*). In general, the bHLH domains play key roles in cell proliferation, determination, and differentiation in a wide variety of tissues at different stages of development in multicellular organisms (*11*). The three most studied genes related to bHLH domains in rodents are the mouse *achaete-scute* homologue (the so-called *Mash1*) and the members of the atonal-related family of genes, neurogenins (*Ngn*) 1 and 2. The generic neurogenic program in several progenitor cells is initiated by Mash1 and Ngns both *in vivo* (*12*) and *in vitro* (*13*). The proneural activity of Ngns involves the promotion of neurogenesis, concomitantly with the repression of the alternative glial fate (*13*). Furthermore, in the mammalian peripheral nervous system, Ngns promote sensory neuron identity (*14*), are involved in differentiation of the dorsal telencephalic glutamatergic neurons (*15*), and intervene in the specification of motoneurons in the ventral spinal cord (*12*). The promises held by stem cell therapies in neuronal diseases, such as Parkinson, have renewed the interest in gaining a deeper understanding of (i) the molecular mechanisms regulating the fate of neurons and their differentiation into specific neuronal populations, such as the dopaminergic neurons, and (ii) the DNA-recognition processes in which Ngn proteins are involved.

In the absence of DNA, the structure of the bHLH domains studied to date is disordered (*16*−*18*). However, bHLH domains tend to homo- or heterooligomerize with other bHLH proteins,

†This work was supported by grants from Spanish Ministerio de Ciencia e Innovación (SAF2008-05742-C02-01; CSD 2008-00005), FIPSE Private Foundation (Exp: 36557/06), and Generalitat Valenciana (ACOMP/2009/185) to J.L.N., as well as VIB and CREA grants from the KU Leuven to B.A.H. E.G. is the Research Director with the National Fund for Scientific Research.

*Corresponding author. Tel: +34 966658459. Fax: +34 966658758. E-mail: jlneira@umh.es.

[1]Abbreviations: ANS, 1-anilino-8-naphthalenesulfonate; bHLH, basic helix−loop−helix domain; CD, circular dichroism; FTIR, Fourier transform infrared spectroscopy; GdmCl, guanidinium chloride; Ngn, neurogenins; NGN1, neurogenin 1; NGN3, neurogenin 3; bHLHN, the bHLH of NGN1; NMR, nuclear magnetic resonance; TSP, sodium trimethylsilyl-[2,2,3,3-$^2H_4$]propionate.

Published on Web 01/26/2010

forming a stable, apparently well-folded structure (*19*); the conformational preferences of either isolated monomeric and/or homo- or heterooligomeric domains have not been studied in detail so far. In this work, we describe the conformational preferences of the isolated bHLH domain of human neurogenin 1 (NGN1), bHLHN, and its binding to two E-boxes. NGN1 is a 237-residue protein, with a basic helix−loop−helix region comprising 61 residues; this domain corresponds to the first subgroup (*3*). We synthesized this region using organic methods; attempts to express the isolated bHLHN domain were unsuccessful, since no recombinant protein was expressed at all, or in the most favorable expression trials, the resulting protein did not have the expected molecular weight. We observed that the isolated bHLHN is a monomeric natively unfolded protein, as shown by bioinformatic analysis and several biophysical and spectroscopic techniques, namely, fluorescence, CD, FTIR, DOSY-NMR, and 1D-NMR spectroscopies. There was, however, evidence of flickering helical structure as suggested by CD, which does not unfold in a sigmoidal fashion, as shown by thermal and chemical denaturations; furthermore, this structure appears to be pH-dependent. The protein forms homodimers (as suggested by fluorescence titrations) upon binding to the E-boxes, but in contrast to that observed in other bHLH domains, its affinity was smaller, and the protein did not fold completely; rather bHLHN formed "fuzzy" complexes with DNA, that is, a DNA−protein complex where the protein does not have a well-fixed rigid structure (*20*).

EXPERIMENTAL PROCEDURES

Materials. Ultrapure GdmCl and urea were purchased from ICN Biochemicals (USA). Exact concentrations of GdmCl and urea were calculated from the refractive index of the solution (*21*). Standard suppliers were used for all other chemicals. Water was deionized and purified on a Millipore system.

The protein was obtained by synthetic organic methods from Genscript (New Jersey, USA) (see below). Purity of the sample was tested by chromatography and mass spectra in that company and by SDS in our own laboratory. The sequence of the synthesized polypeptide was RRSRRVKANDRERNRMHLNAALDALRSVLPSFPDDTKLTKIETLRFAYNYIWALAETLRL, which comprises residues 90−150 of NGN1. This polypeptide contains regions 93−104 (the basic motif) and 105−145 (the helix−loop−helix motif). The bHLHN domain concentration was determined from the amino acid sequence by taking into account the extinction coefficient of tyrosines and tryptophan in model peptides (*22*) and the number of each of these residues in the sequence of bHLH (2 Tyr and 1 Trp). Due to the large number of arginine residues (10 out of 61), synthesis of the bHLHN domain was challenging, and finally, a total amount of 18 mg was produced. However, the synthesized bHLHN domain was insoluble in any buffer at concentrations above 60 μM (below that value no appreciable aggregation or precipitation was observed); conversely, in pure water, we were able to make stock solutions of 286 μM (whose final pH was close to 4.0). Low solubility has been previously observed in other proteins with no relationship to the bHLH superfamily (*23*) and in other members of the bHLH superfamily (*24*−*27*).

The assayed E-boxes correspond to E3-oligo, 5′-CTCTAACTGGCGACAGATGGGCCACTTTCT-3′, and E1-oligo, 5′-GGACCGGGAAGACCATATGGCGCATGCCGG-3′, which are the boxes recognized by NGN3, when forming heterodimers with the bHLH of the E47 protein (*28*). The oligos and their complementary strands were synthesized by Isogen (Barcelona, Spain) at the highest purity available and without salt. Single-stranded oligonucleotide concentration was calculated by using the molar extinction coefficient obtained from nucleotide composition.

DNA Duplexes. Annealing and formation of the double-stranded E3 and E1 oligonucleotides were performed by mixing equal amounts of the corresponding oligos in 10 mM sodium phosphate buffer at pH 7.0 containing 100 mM NaCl. The mixture was incubated for 5 min at 368 K and slowly cooled down to 298 K for 16 h, as described (*29*). This procedure yielded a double-stranded DNA, with no detectable single-stranded oligonucleotide being present as judged by 1% agarose gels and PAGE.

Protein Expression and Purification. DNA encoding the intracellular bHLHN (residues 90−150 of the intact protein) was amplified by PCR from a c-DNA template of a full-length human NGN1 clone. The PCR product was ligated between the *Nde*I and *Not*I sites of pET28a (Novagen, USA) and pGEX-4T-1 (GE Healthcare, Spain) expression vectors. We also tried protein expression in vectors kindly donated by Dr. Mark Bycroft and Mr. Mark Proctor (MRC Unit for Protein Function and Design, Cambridge, United Kingdom); these vectors use maltose binding protein, biotin, thioredoxin, or histidine as protein fusion tags. We were not able to obtain bHLHN expressed in any of those vectors (data not shown).

The bHLHN expression in commercial vectors was tried in BL21(DE3), Rosetta BL21, C41, and BL21pLys (Novagen, Germany). For the His-tagged protein (that is, in the pET28a vector), no overexpression was observed after induction with 1 mM IPTG (by using several temperatures ranging from 298 to 310 K at several IPTG concentrations). The use of the PGEX-4T-1 construction yielded an overexpressed band in the SDS gels in any of the *Escherichia coli* strains described above. Protein was purified according to resin manufacturer instructions (GE Healthcare, Spain); however, the protein band ran faster than expected from its molecular mass (data not shown), even though the broad-range protease inhibitor cocktail (Roche, Germany) was used during its purification. The mass analysis of the purified protein yielded a mass smaller than expected; furthermore, peptide mapping suggested that the obtained recombinant protein corresponded to bHLHN, with only 70% of coverage of the whole amino acid sequence (data not shown). Therefore, we were not able to obtain the intact bHLHN under any of the explored conditions. These findings do not agree with results observed in other bHLH domains, where overexpression of the recombinant proteins for structural and thermodynamic studies have been carried out (*5*, *19*, *30*); however, other bHLH examples have shown difficulties in protein overexpression (*24*, *27*).

Fluorescence. Fluorescence spectra were collected on a Cary Eclipse spectrofluorometer (Varian, USA) interfaced with a Peltier temperature-controlling system. Sample concentration was in the 2−4 μM range, expressed in monomer concentration. A 1 cm path length quartz cell (Hellma) was used. Unless it is stated, all of the experiments were acquired at 298 K.

(*a*) *Steady-State Fluorescence Measurements*. Samples of bHLHN were excited at 280 and 295 nm in the pH range from 2.0 to 12.0 to characterize a possible different behavior of either tryptophan or tyrosine residues. No differences, except in the spectral intensity, were observed either in the band shape or maximum wavelength, and then excitation at 280 nm was used

for all of the experiments. The slit width was equal to 5 nm for the excitation and emission lights. The fluorescence spectra were recorded between 300 and 400 nm. The signal was averaged for 1 s, and the wavelength increment was 1 nm. Blank corrections were made in all spectra.

GdmCl titrations were carried out at pH 7.0 (10 mM phosphate buffer). The proper amount of the denaturant from a 7 M stock solution was used, and samples were left overnight to equilibrate at room temperature. The salts and acids used in the pH titration were as follows: pH 2.0−3.0, 1 M phosphoric acid (formed by the corresponding amounts of H_3PO_4 and NaH_2PO_4); pH 3.0−4.0, 1 M formic acid (formed by the corresponding amounts of HCOOH and NaOOC); pH 4.0−5.5, 1 M acetic acid (formed by the corresponding amounts of H_3CCOOH and $NaOOCCH_3$); pH 6.0−7.0, 1 M NaH_2PO_4 (formed by the corresponding amounts of Na_2HPO_4 and NaH_2PO_4); pH 7.5−9.0, 1 M Tris acid (formed by the corresponding amounts of Tris-HCl and Tris-sodium salt); pH 9.5−11.0, 1 M Na_2CO_3 (formed by the corresponding amounts of $NaHCO_3$ and Na_2CO_3); pH 11.5−13.0, 1 M NaH_2PO_4 (formed by the corresponding amounts of Na_3PO_4 and NaH_2PO_4). The pH was measured with an ultrathin Aldrich electrode in a Radiometer (Copenhagen) pH meter. Those buffer stock solutions were diluted to give a final concentration in the sample of 10 mM either in the fluorescence or in the CD experiments.

To monitor the changes with the pH or the denaturant concentration, the fluorescence intensity and/or the average energy, which is defined as $\langle\lambda\rangle = \sum_1^n((1/\lambda_i)I_i)/\sum_1^n I_i$, was represented versus the denaturant concentration.

(*b*) *Quenching Experiments*. Quenching of intrinsic tryptophan and tyrosine fluorescence by iodide (*31*) was examined at different solution conditions. Excitation was at 280 nm, and emission was measured from 300 to 400 nm. In the absence of GdmCl, ionic strength was kept constant by addition of KCl, and $Na_2S_2O_3$ was added to a final concentration of 0.1 M to avoid formation of I_3^-. The slit width was set at 5 nm for both excitation and emission. We fit the data to the equation (*31*):

$$F_0/F = 1 + K_{sv}[X] \quad (1)$$

where K_{sv} is the Stern−Volmer constant for collisional quenching, F_0 is the fluorescence in the absence of KI, and F is the measured fluorescence at any KI concentration. The range of KI concentrations explored was 0−0.7 M.

Acrylamide quenching was also carried out at different pHs. The dynamic and static quenching constants for acrylamide were obtained by fitting the data from different wavelengths (in the range 330−340 nm) to the Stern−Volmer equation, which includes an exponential term to account for static quenching (*31*):

$$F_0/F = (1 + K_{sv}[X])e^{v[X]} \quad (2)$$

where K_{sv} is the Stern−Volmer constant for collisional quenching and v is the static quenching constant.

(*c*) *Thermal Denaturation Experiments*. Fluorescence thermal denaturation experiments were monitored by following the changes, after excitation at 280 and 295 nm, at 315, 340, and 350 nm. The scan rate was 60 K/h, with an average time of 1 s, and measurements were collected every 0.2 K. Experiments were carried out at several pHs.

(*d*) *ANS Binding*. Excitation wavelength was 380 nm, and emission was measured from 400 to 600 nm. Slit widths were 5 nm for excitation and emission lights. Stock solutions of ANS were prepared in water and diluted into the samples to yield a final 100 μM dye concentration. Dye concentrations were determined using an extinction coefficient of 8000 M^{-1} cm^{-1} at 372 nm. In all cases, blank solutions were subtracted from the corresponding spectra.

(*e*) *DNA Binding Experiments*. Increasing amounts of the E-box oligonucleotide duplexes, in the range 0.5−40 μM, were added to a solution of a fixed concentration of bHLHN in 10 mM phosphate buffer, pH 7.0; protein concentration was 5.6 μM (in monomer units) for the E1-box binding assay and 3.7 μM (in monomer units) for the E3-box binding assay. Fluorescence of the resulting samples was measured after a 2 h incubation time at 298 K. Experiments were carried out with excitation at 280 and 295 nm, and emission fluorescence was collected between 300 and 400 nm. The excitation and emission slits were 5 nm, and the data pitch interval was 1 nm. The dissociation constant of each complex was calculated by fitting the changes observed either in (i) the fluorescence intensity at a particular wavelength or (ii) the $\langle\lambda\rangle$ for a fixed concentration of bHLHN ([bHLHN]) versus the concentration of the added E-box ([E-box]) to

$$F_{meas} = F + \Delta F_{max}\left\{0.5\left(([\text{E-box}] + [\text{bHLHN}] + K_D) - \sqrt{([\text{bHLHN}] + [\text{E-box}] + K_D)^2 - 4[\text{bHLHN}][\text{E-box}]}\right)\right\} \quad (3)$$

where F_{meas} is the measured fluorescence after subtraction of the blank, ΔF_{max} is the change in the fluorescence measured at saturating E-box−oligonucleotide−duplex concentrations, F is the fluorescence intensity when no E-box−oligonucleotide duplex has been added, and K_D is the dissociation constant. The Job's method (*32*, *33*) was used to determine the stoichiometry of the reaction: for the E3-box−oligonucleotide duplex (see Results section) the stoichoimetry was 1:2 of [DNA]/[bHLHN], but for E1-box−oligonucleotide duplex the stoichiometry was 1:1; therefore, to obtain the affinity constant, the concentration of bHLHN in eq 3 was expressed in monomer units (for E1-box−oligonucleotide duplex) or in dimeric species units for the E3-box−oligonucleotide duplex (see Results). Inner-filter effects at 280 and 295 nm were corrected for the absorbance of the corresponding box (*34*). Absorbance measurements were carried out in a Shidmazu UV-1601 ultraviolet spectrophotometer using a 1 cm path length cell (Hellma).

We did not measure the binding constant by using CD titration experiments, because of the large amounts of protein required. That is why, together with the low solubility of the protein, we could not carry out isothermal titration calorimetry studies.

Circular Dichroism. Circular dichroism spectra were collected on a Jasco J810 spectropolarimeter fitted with a thermostated cell holder and interfaced with a Peltier heating unit. The instrument was periodically calibrated with (+)-10-camphorsulfonic acid. Unless it is stated, all the experiments were carried out at 298 K.

(*a*) *Steady-State Measurements*. Isothermal wavelength spectra at different pHs were acquired at a scan speed of 50 nm/min with a response time of 2 s and averaged over four scans. Far-UV measurements were performed using 10 μM protein (in monomer units) in 10 mM buffer (see above) in 0.1 cm path length quartz cells (Hellma). Near-UV spectra were acquired using 30−40 μM protein (in monomer units) in a 0.5 cm path

length cell. All spectra were corrected by subtracting the proper baseline. The molar ellipiticity, [Θ], was calculated according to

$$[\Theta] = \frac{\Theta}{(10lcN)} \quad (4)$$

where Θ is the measured ellipticity, l is the path length of the cell, c is the protein concentration, and N is the number of amino acids (61 for bHLHN).

The helical content of the domain was, at a first approach, approximated from the molar ellipticity at 222 nm according to (*31*)

$$f_h = [\Theta]_{222}/[\Theta]^{\infty}_{222}\left(1 - \frac{k}{n}\right) \quad (5)$$

where f_h is the helical fraction of the protein, $[\Theta]_{222}$ is the observed mean residue ellipticity, $[\Theta]^{\infty}_{222}$ is the mean ellipticity for an infinite α-helix at 222 nm (-34500 deg cm^2 dmol^{-1}), k is a wavelength-dependent constant (2.57 at 222 nm), and n is the number of peptide bonds. In addition to changes in secondary structure and aromatic side chains, far-UV CD is also sensitive to distortions of helices such as bends or tilting of the amide planes relative to the helical axis (*35*). Modern methods of CD analysis, such as CDNN, can take these factors into account and give reliable secondary structure contents. Therefore, we deconvoluted the CD spectrum of bHLHN at pH 7.0 by using the CDNN (*36*) and the DICHROWEB software (*37*, *38*).

In the GdmCl denaturation experiments, the far-UV spectra were corrected by subtracting the corresponding baseline in all cases. Every chemical denaturation experiment was repeated at least three times with fresh new samples; the chemical denaturation reactions were fully reversible (data not shown). In the pH-induced unfolding experiments of the domain, the pH was measured after completion of the experiments, and essentially no differences were observed with those pHs calculated from the buffer stock solutions; the proper blank solutions were subtracted in all cases.

For the binding experiments, far-UV CD spectra of the complexes formed by bHLHN (30 μM, in monomer units) and the double-stranded E1-box (30 μM) or E3-box (15 μM) were acquired. Far-UV spectra of each oligonucleotide and the isolated bHLHN were also acquired separately.

(*b*) *Thermal Denaturation*. Thermal denaturation experiments of bHLHN were performed at a constant heating rate of 60 K/h and a response time of 8 s. Thermal scans were collected in the far-UV region by following the ellipticity at 222 nm from 298 to 358 K in 0.1 cm path length quartz cells (Hellma) with a total protein concentration of 10 μM, in monomeric units. The reversibility of thermal transitions was tested by recording a new scan after cooling to 298 K the thermally denatured samples and comparing it with the spectrum obtained before denaturation; in all cases, both spectra were identical (data not shown). The possibility of drifting of the CD spectropolarimeter was tested by running two samples containing only buffer before and after the thermal experiments. No difference was observed between the scans. Every thermal denaturation experiment was repeated at least twice with new samples. In all cases, after the reheating experiment, the samples were transparent, and no precipitation was observed.

NMR Spectroscopy. The NMR experiments were acquired on a Bruker Avance DRX-500 spectrometer (Bruker GmbH, Karlsruhe, Germany) equipped with a triple resonance probe and z-pulse field gradients. Unless it is stated, all of the experiments were carried out at 298 K.

(*a*) *1D-NMR Spectroscopy*. Homonuclear 1D-NMR experiments were performed with bHLHN at concentrations of 50 μM (in monomer units) in 0.5 mL, pH 7.0 (uncorrected for deuterium isotope effects) and 10 mM phosphate buffer in H_2O/D_2O (90%/10% v/v). Spectra were also acquired in pure water at protein concentrations of 50 μM (in monomer units, at a final pH = 4.0). The spectra were acquired with 16K data points and averaged over 512 scans with 6000 Hz of spectral width (12 ppm); water suppression was achieved with the WATERGATE sequence (*39*). Baseline correction and zero-filling were applied. All spectra were processed and analyzed by using XWINNMR (Bruker GmbH, Karlsruhe, Germany) working on a PC computer. In the exchange experiments, the corresponding amount of bHLHN lipophilized powder was dissolved in 0.5 μL of D_2O. Sample temperature homogenization and shimming took 20 min. TSP was used as the external chemical shift reference.

In the binding experiments, 1D-NMR spectra of the isolated bHLH (60 μM in monomeric units) and double-stranded E1-box (60 μM) or E3-box (30 μM) were acquired.

(*b*) *Translational Diffusion Measurements* (*DOSY Experiments*). Translational self-diffusion measurements were performed with the pulsed-gradient spin-echo sequence. The following relationship exists between the translational self-diffusion parameter, D, and the delays used during acquisition (*40*):

$$\frac{I}{I_0} = -\exp\left(D\gamma_H{}^2\delta^2G^2\left(\Delta - \frac{\delta}{3} - \frac{\tau}{2}\right)\right) \quad (6)$$

where I is the measured peak intensity of a particular (or alternatively a group of) resonance(s), I_0 is the maximum peak intensity of the same resonance(s) at the smaller gradient strength, D is the translational self-diffusion constant (in cm^2 s^{-1}), δ is the duration (in seconds) of the gradient, G is the gradient strength (in T cm^{-1}), Δ is the time (in seconds) between the gradients, γ_H is the gyromagnetic constant of the proton, and τ is the recovery delay between the bipolar gradients (100 μs). Data were plotted as the $-\ln(I/I_0)$ versus G^2 and the slope of the resulting line is $D\gamma_H{}^2\delta^2(\Delta - \delta/3 - \tau/2)$, from where D can be obtained. The duration of the gradient was 3 ms, and the time between both gradients was 150 ms. The most upfield shifted methyl groups (those between 0.5 and 1 ppm) were used to measure the changes in intensity in the spectra of bHLHN.

To determine the D of the isolated bHLH (and then its corresponding hydrodynamic radius, R_S), we used the approach developed by Dobson and co-workers (*41*). The R_S of dioxane was 2.12 Å, and its experimentally determined D, under our conditions, was 8.17×10^{-6} cm^2 s^{-1}. The gradient strength was calibrated by using the D for the residual proton water line in a sample containing 100% D_2O in a 5 mm tube, as described (*40*).

(*c*) *Measurements of* T_2. Measurements of the T_2 (transverse relaxation time) provide a convenient method to determine the molecular mass of a macromolecule, since the correlation time, τ_c, is approximately equal to $1/(5T_2)$ (*42*). We measured the T_2 for isolated bHLHN with the 1-1 echo sequence (*43*); the calculation of the τ_c was carried out as described (see Results section) (*42*).

Analysis of the pH Titrations. The pH denaturation experiments were analyzed assuming that both species, protonated

and deprotonated, contribute to the CD and/or fluorescence spectra:

$$X = \frac{X_a + X_b 10^{pH - pK_a}}{1 + 10^{pH - pK_a}} \quad (7)$$

where X is the physical property being measured (ellipticity, average energy, and/or fluorescence intensity at any particular wavelength), X_a is that at low pHs (acidic form), X_b is that at high pHs (basic form), and pK_a is the apparent pK of the titrating group. The apparent pK_a reported was obtained from three different measurements, carried out with fresh new samples.

Fitting by nonlinear least-squares analysis to the equations described above was carried out by using the general curve fit option of Kaleidagraph (Abelbeck software) on a PC computer.

Fourier Transform Infrared Spectroscopy. Attenuated total reflection infrared (ATR-FTIR) spectra were obtained on a Bruker IFS55 FTIR spectrophotometer (Ettlingen, Germany) equipped with a MCT detector (broad band 12000−420 cm^{-1}, cooled with liquid N_2, and 24 h hold time) at a resolution of 2 cm^{-1} with an aperture of 3.5 mm and acquired in the double-sided, forward−backward mode. Two levels of zero filling of the interferogram prior to Fourier transform allowed encoding the data every 1 cm^{-1}. The spectrometer was continuously purged with dry air (Whatman 75-62, Haverhill, MA).

The internal reflection element was a 52 × 20 × 2 mm trapezoidal germanium ATR plate (ACM, Villiers St Frédéric, France) with an aperture angle of 45° yielding 25 internal reflections. The germanium crystals were washed in Superdecontamine (Intersciences, AS, Brussels, Belgium), a laboratory detergent solution at pH 13, rinsed with distilled water, washed with methanol and then with chloroform, and finally placed for 2 min in a plasma cleaner PDC23G (Harrick, Ossining, NY) working under reduced air pressure. Measurements were carried out at 298 K. Thin films were obtained by slowly evaporating a sample on the ATR crystal under a N_2 stream.

Sharp atmospheric water absorption lines were corrected for taking advantage of the line width difference existing between the atmospheric water and the solid sample bands (*44*). Our software computes the subtraction coefficient as the ratio of the atmospheric water band area between 1562 and 1555 cm^{-1} on the sample spectrum and on the reference atmospheric water spectrum. Finally, the corrected spectra were smoothed by apodization of its Fourier transform by using the Fourier transform with a function of 4 cm^{-1} Gaussian line shape.

Protein secondary structure was obtained as described (*45*). For hydrogen/deuterium exchange experiments, N_2 gas was saturated with D_2O by bubbling through a series of four vials containing D_2O; a flow rate of 50 mL/min was controlled by a flow tube (Fisher Bioblock Scientific, Illkirch, France). Sample deuteration started by connecting the measurement chamber to the D_2O-saturated N_2 flow from the output of the first sample chamber.

The areas of the amide I, II, and II′ were obtained by automatic integration. For each spectrum, the area of amide II was divided by the area of amide I to take into account the swelling of the sample layer due to the presence of D_2O (*46*). All kinetic curves were analyzed as multiexponential decays of amide protons using a nonlinear least-squares procedure.

Bioinformatic Analysis of Amino Acid Sequence of bHLHN. Protein sequence was submitted to the PONDR server (www.pondr.com), and analyses were performed using the neural network predictors VL-XT and CDF (*47*) (access to PONDR was provided by Molecular Kinetics, 6201 La Pas Trail-Ste 160, Indianapolis, IN 46268; main@molecularkinetics.com). We also used the IUPRED (*48*) and Foldindex (*49*) servers to allow for a comparison.

RESULTS

bHLHN Is a Monomeric Protein. We first tried to address the oligomeric state of the isolated bHLHN in solution by using analytical gel filtration chromatography. The domain was retained in the Superdex 16/60 HR column eluting at very large volumes (larger than the bed volume of the column, 19.1 mL) in a wide pH range (from 4.0 to 12.0). Therefore, we decided to determine the R_S from the measurements of the translational diffusion coefficient by DOSY-NMR at pH 7.0. The diffusion coefficient was $(1.0 \pm 0.5) \times 10^{-6}$ cm^2 s^{-1}, which yields assuming a spherical shape a R_S of 17.5 Å (*40*). The theoretical value for an unsolvated spherical molecule, R_0, is $R_0 = (3M\bar{V}/4N\pi)^{1/3}$, where $\bar{V}$ is the specific volume (0.73059 cm^3/g for bHLHN), N is Avogadro's number, and M the molecular mass, 7258.3 Da. For bHLHN, the above expression leads to 12.8 Å, which is smaller than that measured experimentally; however, the calculation of the R_0 for a spherical dimeric species according to the above expression is 16.1 Å. Thus, our results suggest that bHLH could be a monomer with a rather elongated shape or a spheric dimer. To distinguish between both species, we should use an additional measurement. The correlation time of the domain, τ_c, can be obtained from the T_2 measurements, and its calculation does not rely on a particular assumption about molecular shape, but rather it only depends on the molecular mass of the protein. The relationship between both parameters is (*42*, *43*) $\tau_c = 1/(5T_2)$. The T_2 for bHLHN is 33.6 ms, and then its τ_c is 4.17 ms, which leads to a molecular mass of 8.3 kDa, close to the expected molecular mass (7.3 kDa). These results, with the calculated translational diffusion coefficient described above, indicate that the isolated bHLH can be a monomer with a rather elongated shape.

bHLHN Is a Natively Unfolded Protein. We used different experimental and theoretical approaches to test the structural conformation of bHLHN. In the experimental approach, we employed several spectroscopic techniques.

(*a*) *Spectroscopic Measurements*.

Fluorescence. (1) Steady-State Fluorescence. We acquired the fluorescence spectra to map the tertiary structure of the domain (*50*). The bHLHN has a sole tryptophan at position 53 (in the numbering of the isolated bHLH) and two tyrosines at positions 49 and 51. The emission fluorescence spectrum of the domain at pH 7.0 is red shifted, with a maximum at 340 nm (Figure 1A, blank circles, continuous line). This indicates that the tryptophan is partially solvent exposed (in an aqueous environment the maximum should be at ca. 350 nm).

The pH dependence of $\langle\lambda\rangle$ showed a bell-shaped behavior, with a plateau between pH 4.0 and pH 9.0 (Figure 1C, filled squares, right axis), suggesting that the conformation of the protein was not modified in that pH range. At high and low pHs, however, the $\langle\lambda\rangle$ decreased probably because of structural rearrangements due to titration of acid (basic) residues or, alternatively, to acid (basic) hydrolysis.

(2) ANS-Binding Experiments. ANS is used as a fluorescence probe which binds to spatially close solvent-exposed hydrophobic patches (*51*), shifting the fluorescence maxima from 520 (not

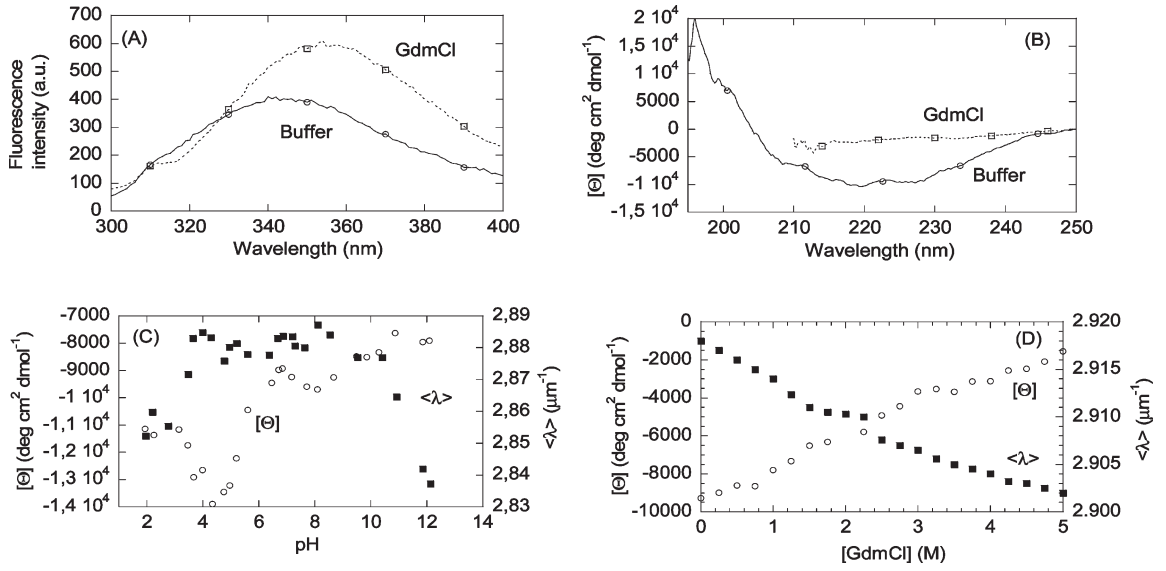

FIGURE 1: Fluorescence and far-UV CD experiments of isolated bHLHN. (A) Fluorescence spectrum of bHLHN in aqueous solution and 10 mM phosphate buffer (continuous line, blank circles) and in 5 M GdmCl (dotted line, blank squares) at pH 7.0. (B) Far-UV spectra of bHLHN in aqueous solution and 10 mM phosphate buffer (continuous line, blank circles) and in 5 M GdmCl, pH 7.0 (dotted line, blank squares). (C) Changes in fluorescence ($\langle\lambda\rangle$) (right axis, filled squares) and in the molar ellipticity at 222 nm (left axis, blank circles) upon pH. (D) Changes in fluorescence ($\langle\lambda\rangle$) (right axis, filled squares) and ellipticity at 222 nm (left axis, blank circles) upon GdmCl concentration. The conditions were 2 μM protein (in monomer units), at 298 K, for fluorescence experiments and 10 μM (in monomer units) for the CD spectra; buffer concentration was 10 mM in all cases. Spectra were acquired in either 1 cm (fluorescence) or 0.1 cm (CD) path length cells.

Table 1: Quenching Parameters of bHLHN in KI and Acrylamide[a]

	280 nm		295 nm	
conditions	K_{sv} (M^{-1}) (KI)	K_{sv} (M^{-1}) (acrylamide)	K_{sv} (M^{-1}) (KI)	K_{sv} (M^{-1}) (acrylamide)
pH 1.4[b]		12 ± 1		7 ± 1
pH 7.3	0.8 ± 0.1	4.7 ± 0.4	0.62 ± 0.07	3 ± 1
pH 12.0	0.9 ± 0.1	4.9 ± 0.8	0.7 ± 0.1	4.6 ± 0.7
6 M GdmCl	4.0 ± 0.4	14 ± 1	3.3 ± 0.5	7 ± 1

[a]Errors are data fit errors to eq 1 (KI) or eq 2 (acrylamide). The constants were obtained from data at 338 nm (similar constants were obtained by fitting the intensities at 335, 336, and 337 nm). Experiments were carried out at 298 K. [b]At acidic pH, in the presence of KI, the protein precipitated.

bound isolated ANS) to 480 nm (ANS bound to a hydrophobic patch). In the presence of bHLHN, at any of the pHs explored, the maxima of the ANS fluorescence spectra were 520 nm; these data suggest that the protein did not bind ANS to a large extent (data not shown).

(3) Examination of Tryptophan and Tyrosine Solvent Exposure by Fluorescence Quenching. To further examine whether there is tertiary structure around Tyr49, Tyr51, and Trp53, which could hamper solvent accessibility to the aromatic moieties, we studied iodide and acrylamide quenching in the presence and in the absence of denaturants (Table 1). The K_{sv} values were larger in acrylamide than in KI, as observed in other proteins (*31*). For acrylamide, the K_{sv} parameters in the presence of GdmCl or at acidic pHs were similar, suggesting that the tryptophan and the tyrosine residues were fully exposed. However, at physiological and basic pHs, the K_{sv} parameters (either in KI or acrylamide) were smaller, indicating that the tryptophan and tyrosine residues were solvent-exposed (as suggested from the maximum wavelength; see above), but they were not fully accessible.

(4) Chemical Denaturation Monitored by Intrinsic Fluorescence. In the presence of GdmCl (or urea; data not shown), the fluorescence spectra showed an increase in the intensity as the concentration of denaturant varied, with small gradual changes in the maxima wavelengths (Figure 1A, dotted line, blank squares). A linear behavior was observed in the $\langle\lambda\rangle$ as [GdmCl] increased (Figure 1D, filled squares, right axis), which is expected for solvent-exposed aromatic rings (*50*).

(5) Thermal Denaturation Monitored by Intrinsic Fluorescence. The changes in the emission fluorescence were examined at 315, 340, and 350 nm at several pHs. No sigmoidal behavior was observed at any pH (Figure 1 of Supporting Information).

Circular Dichroism Experiments. (1) Far-UV CD. We used far-UV CD in the structural analysis of bHLHN as a spectroscopic probe sensitive to protein secondary structure (*35*). The CD spectrum of the domain at pH 7.0 in aqueous solution showed a minimum negative ellipticity at ca. 225 nm and at 210 nm, which are characteristic of α-helix- or turn-like structures (Figure 1B, continuous line, blank circles); however, contribution of aromatic signals at this wavelength cannot be ruled out (*35*). The estimated population of α-helix- or turn-like structures from eq 5 is 34% (the molar ellipticity of bHLHN at 222 nm is −9297.2 deg cm^2 $dmol^{-1}$). Deconvolution by CDNN and DICROWEB led to the finding that there is a high percentage of helical structure in the domain (Table 2). The FTIR and deconvolution programs (Table 2) predict a higher content of α-helix than that calculated by eq 5; this is probably due to the fact that aromatic residues also present absorption bands at 222 nm (*35*).

(2) Chemical and pH Denaturations. Two transitions were observed in the pH-dependent behavior of the ellipticity

Table 2: Secondary Structural Analysis of bHLHN As Determined by CD[a] and FTIR

	circular dichroism[b]					FTIR
structural assignment	CDNN	k2D	Selcon 3	Contin	CDSSTR	% of total secondary structure[c]
α-helix	95.6	100	81.80	95.6	61	67.0
β-sheet	0.2	0	0.10	0.2	15	3.0
β-turns	3.2	0	5.40	3.2	6	16.0
random coil	0.1	0	12.70	0.1	18	17.0

[a]CD experiments were acquired at pH 7.0 (phosphate buffer) at 10 μM protein concentration (in monomer units). Experiments were carried out at 298 K. [b]Access provided by DICROWEB for all the CD deconvolution procedures except CDNN. For the programs Selcon and Contin 3, the indicated percentages of helical structure comprise the regular and distorted components. [c]FTIR experiments were acquired in pure water. In the presence of buffer at pH 7.0, the percentages of β-sheet were altered on the expense of helical- and turn-like populations (see text for details).

at 222 nm (Figure 1C, blank circles, left axis). The first with $pK_a = 5.4 \pm 0.3$ could be ascribed to a glutamic acid residue or at His17 (the sole histidine in bHLHN) (*52*). The second one at pH ~9 can be attributed to lysine and/or tyrosine residues (*52*), although we could not determine precisely the titration midpoint due to the lack of experimental points in the sigmoidal region of the curve and at very basic pHs.

Conversely, the molar ellipticity at 222 nm decreased gradually (in absolute values) as the concentration of GdmCl was raised (Figure 1D, blank circles, left axis). These results further suggest that there is residual structure in bHLHN, which does not unfold cooperatively; we do not know, however, if this structure is local or involves short-lived long-range interactions.

(3) Near-UV. We used near-UV CD to detect possible changes in the asymmetric environment of aromatic residues (*35*). The near-UV of bHLHN was very weak with no intense bands (data not shown). The absence of a near-UV signal could be due to (i) the lack of an asymmetric environment for the aromatic residues or (ii) the low content of aromatic residues in the protein (1 Trp, 2 Tyr, 1 His, and 2 Phe). We favor the first explanation because of the agreement with the findings from the other spectroscopic probes.

(4) Thermal Denaturations. To further investigate the possible presence of residual secondary structure in bHLHN, we carried out thermal denaturations by following the changes in ellipticity at 222 nm. The ellipticity at this wavelength shows a linear behavior (Figure 1 of Supporting Information). This result suggests the absence of cooperativity, as expected for a noncompact structure, and agrees with the behavior observed in thermal denaturations followed by fluorescence (see above).

Fourier Transform Infrared Spectroscopy. FTIR is a powerful method for investigation of protein secondary structure. The main advantage in comparison with CD and fluorescence is that FTIR is much more sensitive to the presence of β-structure or random-coil conformations. In the case of proteins, structural information can be obtained by analyzing the amide I region of the spectrum (1700−1600 cm^{-1}). The absorbance of this band is mainly due to the stretching vibration of the carbonyl peptide bond, whose frequency is highly sensitive to hydrogen bonding and thus to protein secondary structure (*53*). The amide I of the FTIR spectrum of bHLHN in pure water (final pH = 4.0) is centered at 1654 cm^{-1}, and the amide II is at 1545 cm^{-1} (Figure 2A). The presence of both bands at those wavelengths is characteristic of α-helix structure. The blind application of the methods described in the Experimental Procedures section yields the percentages of secondary structure of Table 2 (*45*).

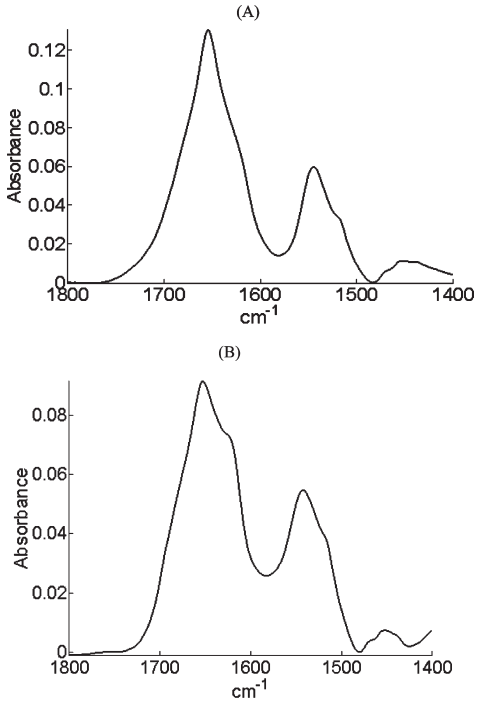

FIGURE 2: FTIR spectra of bHLHN. (A) Amide I and amide II regions of the FTIR spectra of the domain in pure water (no buffer added) at pH 4.0. (B) Amide I and amide II regions of the FTIR spectra of the domain at pH 7.0 in phosphate buffer (30 mM). In both cases 10 μg of protein (in monomer units) was used. Experiments were acquired at 298 K.

FTIR experiments were also carried out at pH 7.0 in phosphate buffer (50 mM). They show the presence of a shoulder at 1622 cm^{-1} (Figure 2B), suggesting protein aggregation. At this pH, the percentages of secondary structure are as follows: α-helix, 47%; β-sheet, 20%; β-turns, 15%; random coil conformations, 17%. That is, the population of helical structure is decreased, whereas the amount of β-sheet and disordered structures is substantially increased, when compared with experiments at pH 4.0. Probably, FTIR is monitoring the aggregation and precipitation that we observed to occur in the presence of buffers (see Experimental Procedures section).

Nuclear Magnetic Resonance. NMR can give information about the general fold of a polypeptide chain in solution at residue level. The 1D-NMR spectrum of bHLHN at 298 K did not show a chemical shift dispersion in the amide (Figure 3A) nor in the methyl regions (Figure 2 of Supporting Information). In those regions all of the resonances were clustered as expected for random-coil proteins (*54*), namely, between 8.0 and 8.7 ppm

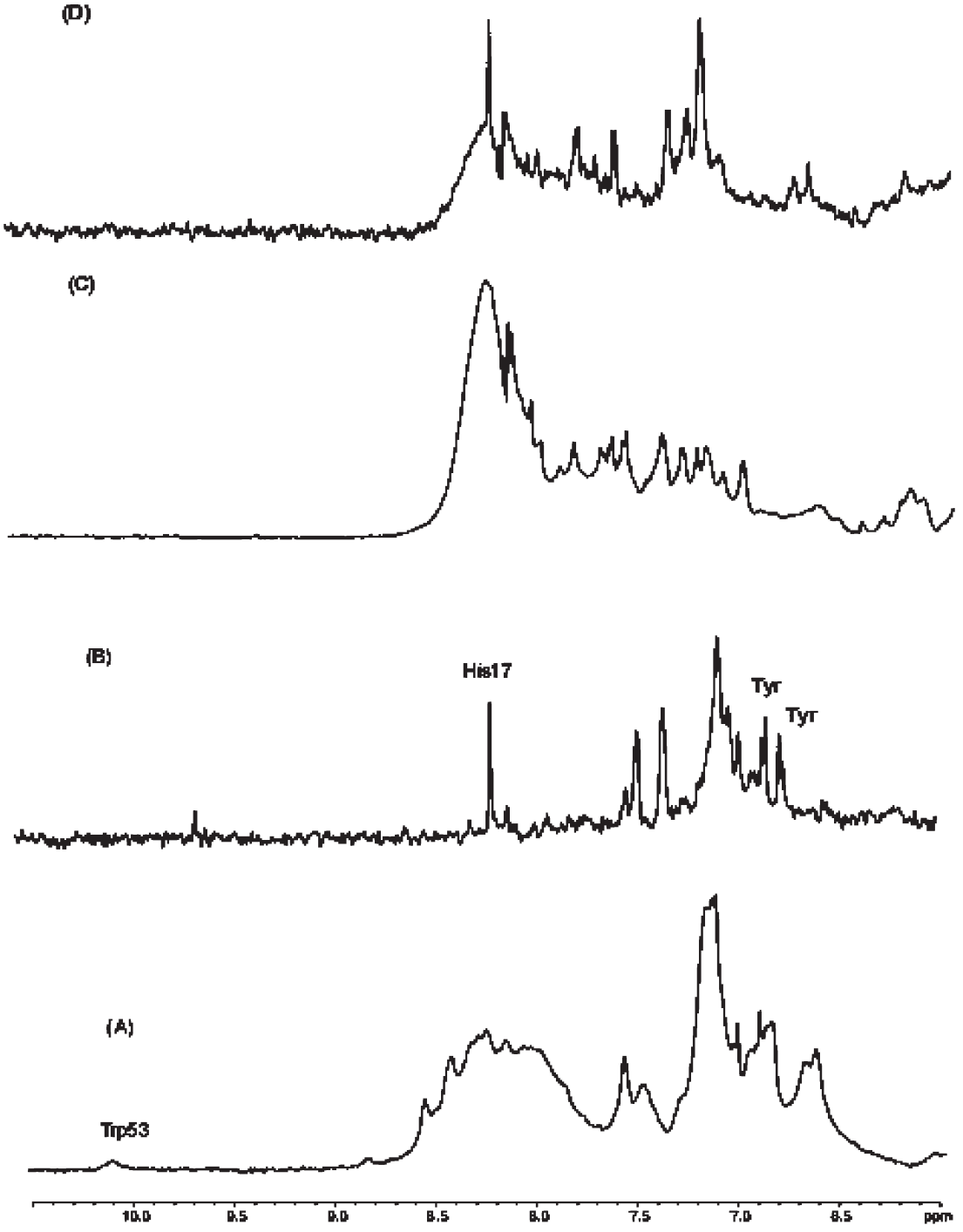

FIGURE 3: NMR spectra of bHLHN in the absence and in the presence of the E-boxes. (A) Amide region of isolated bHLHN at pH 7.0 (10 mM, phosphate buffer), 298 K, in aqueous solution; the position of the indole proton of Trp53 is indicated. (B) Amide region of isolated bHLHN at pH 7.0 in D_2O (no corrections for isotope effects were applied); the positions of the meta protons of Tyr49 and Tyr51 and the C_2H of His17 are indicated. (C) Amide region of the complex between the E1-box and bHLHN at pH 7.0 in aqueous solution (10 mM phosphate buffer). (D) Amide region of the complex between the E3-box and bHLHN at pH 7.0 in aqueous solution (10 mM phosphate buffer). Protein concentration was 60 μM (in monomer units) in the absence of the E-boxes; in the presence of the E-boxes, the concentrations were 60 μM bHLHN (in monomer units), 30 μM for the E3-box, and 60 μM for the E1-box; phosphate buffer concentration was 20 mM, and temperature was 298 K.

(for the amide signals) (Figure 3A,B) and between 0.8 and 1.0 ppm (for the methyl protons) (Figure 2 of Supporting Information). However, we can observe that there is a small shoulder at 0.72 ppm, suggesting the presence of some upfield-shifted protons, due probably to the presence of residual local interactions (Figure 2 of Supporting Information). The presence of residual structure is further confirmed by two pieces of evidence: (i) the random-coil value for the indole proton of a tryptophan is 10.22 ppm (*54*), but the indole proton of bHLHN appeared at 10.11 ppm (Figure 3A), and (ii) the presence of two doublets in the D_2O spectrum, which can be assigned to the meta protons of tyrosine rings, at 6.80 and 6.85 ppm; the chemical shifts for the meta protons of a random-coil tyrosine appear at 6.86 ppm (*54*). Thus, as suggested by FTIR and CD, there is NMR evidence of residual structure, although the protein is devoid of the tertiary structure.

In the hydrogen/deuterium exchange experiments all of the amide protons disappeared at pH 7.0 (Figure 3B) and at pH 4.0

(data not shown). Only the aromatic protons of the protein and the signals of the proton of the C_2H of the His (at 8.25 ppm) could be observed. These results suggest that no stable hydrogen-bonding structure is present in the protein. We also followed the hydrogen/deuterium exchange by monitoring the decrease of the FTIR amide II band at pH 4.0 (data not shown). We observed an exponential behavior, where the value of the exchange rate (14.37 min) agrees with that theoretically calculated (13.5 min) from random-coil models (*55*); these results further support the absence of hydrogen bonding in bHLHN.

Then, all of the spectroscopic results show that there is evidence of residual structure in bHLHN (possibly helical, as suggested by CD and FTIR), but it is not hydrogen-bonded.

(*b*) *Theoretical Analysis of the bHLHN Sequence*. The domain is significantly enriched in alanine, arginine (which are disordered-promoting residues), and leucine; further, it is depleted in tryptophan (one tryptophan, 1.6%), tyrosine (two tyrosines, 3.3%), phenylalanine (two residues, 1.6%), and other order-promoting residues (Figure 4A).

The use of the different theoretical predictors was, however, puzzling. The VL-XT program predicts only a disordered region: that comprising the basic patch of residues at the N-terminus (residues 1−17) (Figure 4B). Similar results were obtained by using the IUPRED web server (Figure 4C), showing the agreement between both theoretical approaches. Foldindex predicts that bHLHN is disordered, with an "unfoldability" value of −0.070, a charge of 0.098, and a phobic value of 0.423 (data not shown). Further, the charge-hydrophobicity phase-space analysis of PONDR indicates the unfolded nature of the domain (Figure 3A of Supporting Information), and the CDF (cumulative distribution function) analysis of PONDR (predictor of natural disordered regions) predicts that the domain is above the established limit, indicating a well-folded protein (Figure 3B of Supporting Information).

To sum up, the theoretical results suggest that bHLHN is not a completely disordered protein, but rather there are several regions of ordered structure. Then, these theoretical results seem to support the spectroscopic findings described above.

bHLHN Binds to the E1- and E3-Boxes with Low Affinity. Since bHLHN is a basic helix domain we wondered whether (i) it was able to recognize and bind any E-box and (ii) its structure and oligomerization state were altered upon DNA binding. We chose the E1- and E3-boxes which are recognized by the heterodimer formed by NGN3 and the bHLH protein, E47 (*28*); interestingly enough, the NGN3 homodimer does not bind any of the E-boxes.

We first tested whether there was binding by measuring the changes in spectroscopic probes (namely, CD and 1D-NMR) when the double-stranded E1- or E3-boxes were mixed with the protein. The CD spectrum of the complex formed by bHLHN and the E1-box was different from the sum spectrum obtained by the addition of the spectra of each biomolecule separately (Figure 5A); the same occurs for the E3-box (Figure 5B), where the differences between the sum and complex spectra were larger than in the E1-box; it is important to indicate at this stage that it has been suggested that the larger the changes in a spectroscopic property (or other biophysical feature) which allows to map binding, the larger the affinity (that is, the smaller the K_D) (*26*, *56*). If we assume that the E-boxes do not change their conformation upon binding to bHLHN (as has been shown in X-ray studies of other bHLH−DNA complexes (*5*)), we conclude that the domain must alter its structure upon binding.

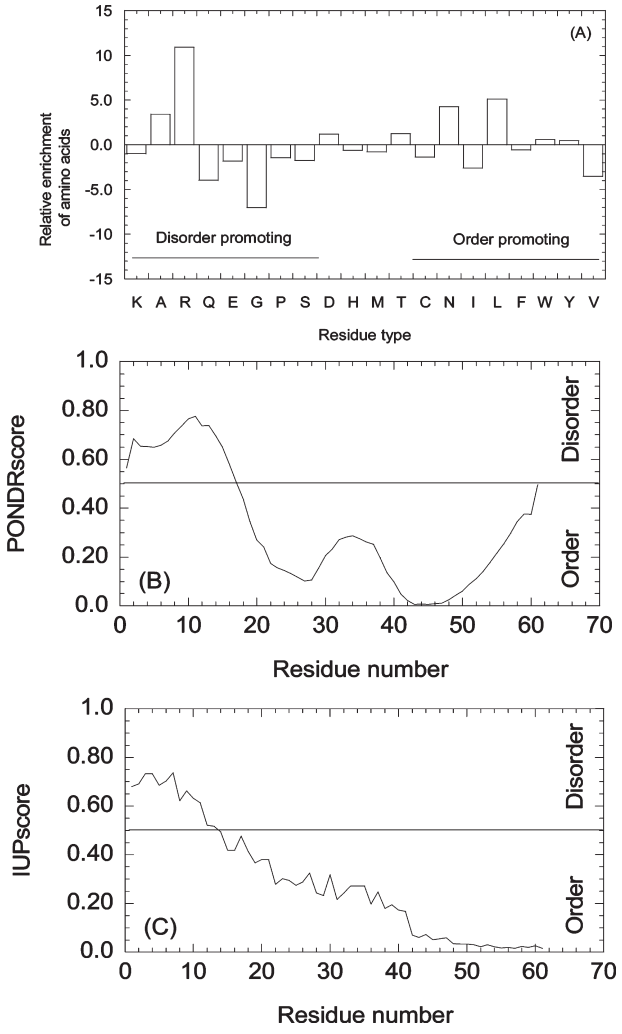

FIGURE 4: *In silico* analysis of the bHLHN sequence. (A) Distribution of order- and disorder-promoting amino acids in the bHLHN sequence. The distribution shows deviation in amino acid composition of bHLHN from the average values in the Swiss-Prot database (as obtained from the World Wide Web at http://www.expasy.org/sprot/relnotes/relstat.html). (B) PONDR prediction of unstructured regions of the bHLHN sequence with the use of VL-XT predictor. The predictor score is plotted against the residue number. The threshold is 0.5, and residues with a higher score are considered disordered. (C) IUPred prediction of unstructured regions (long-disorder prediction) of bHLHN. The threshold is also 0.5.

The most intense changes in the 1D-NMR spectra of the complexes occurred at the indole and tyrosine aromatic protons (Figure 3C,D), which disappeared in the presence of the boxes; this result suggests that the tryptophan and the aromatic rings of both tyrosines are involved in binding, but they have a slow exchange within the NMR time scale, probably due to the presence of several exchanging conformations. However, we did not observe signal dispersion in the amide (Figure 3C,D) or in the methyl region (data not shown), indicating that the DNA binding, although it altered the structure of the protein (as shown by far-UV CD), did not cause a dramatic increase in its structure.

Next, we tried to determine quantitatively the value of the affinity constants for both E-boxes. We decided to follow the changes in the intrinsic fluorescence of the domain upon binding, due to the smaller amount of protein required during the experiments and the fact that the tryptophan and tyrosine residues (as shown by NMR experiments; see above) are involved

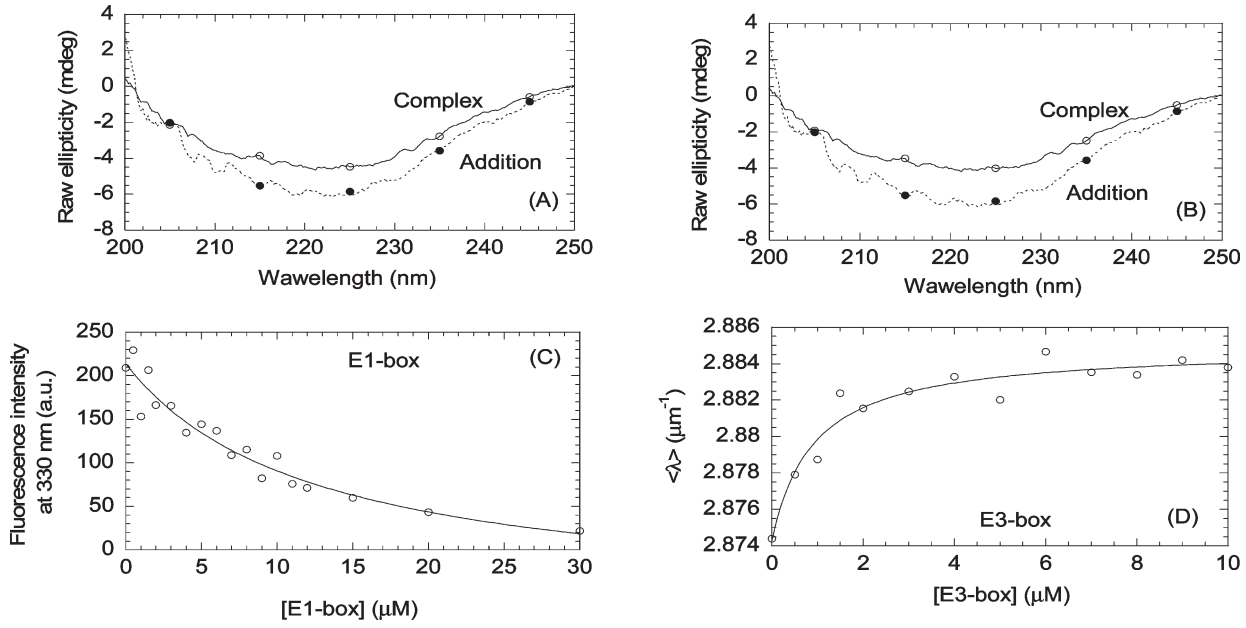

FIGURE 5: DNA binding to bHLHN. (A) Far-UV CD spectra of the complex E1-box−bHLHN (blank circles, continuous line) and the sum spectrum obtained by addition of the spectra of each biomolecule (filled circles, dotted line). (B) Far-UV CD spectra of the complex E3-box−bHLHN (blank circles, continuous line) and the sum spectrum obtained by addition of the spectra of each biomolecule (filled circles, dotted line). (C) Fluorescence titration curves of E1-box from the changes in the intensity at 330 nm after excitation at 280 nm (similar titration curves were obtained from excitation at 295 nm and data collection at 315 or 350 nm and from $\langle\lambda\rangle$ changes). (D) Fluorescence titration curves of the E3-box from the changes in $\langle\lambda\rangle$ after excitation at 280 nm (similar titration curves were obtained from excitation at 295 nm and data collection at 315 or 350 nm). Protein concentrations (in protomer units) in the fluorescence experiments for the E1- and E3-boxes were 5.3 and 3.6 μM, respectively. Experiments were carried out at 298 K.

in the binding region. The affinity for the E1-box was 8.1 ± 3.4 μM (Figure 5C), and that for the E3-box was 0.9 ± 0.3 μM (Figure 5D), which explains why the changes in the CD spectra were larger for this E-box (Figure 5A,B) (*26*, *56*). However, the use of the Job's method (*30*, *31*) suggests that the stoichiometry is not 2:1 (protein:DNA) for both boxes (Figure 4 of Supporting Information).

DISCUSSION

bHLHN Is a Natively Unfolded Protein with a Premolten Globule Conformation. Natively unfolded proteins (or intrinsically disordered proteins) are proteins that accomplish their functions in the absence of well-folded secondary and tertiary structures under physiological conditions (*57*, *58*). As it has been stated, to assess intrinsic disorder, it is necessary to use a combination of theoretical predictions and experimental methods (*58*).

In this work, we show experimentally that isolated bHLHN in solution lacks a well-defined tertiary structure, and it is mainly a disordered protein, with neither stable hydrogen bonds nor a well-formed core. The lack of dispersion of the amide signals and methyl groups in the NMR spectrum and the absence of hydrogen bonds, as shown by hydrogen-exchange experiments (followed by either FTIR or NMR), also pinpoint to the absence of a stable tertiary structure. Furthermore, the absence of cooperativity during the thermal and chemical denaturations (monitored by CD and fluorescence) also suggests that the tertiary structure of the domain, if any, is very weak and no hydrophobic core is formed. The presence of such noncooperative transitions are typically observed in chemical or thermal denaturations of partially folded states devoid of persistent long-range tertiary contacts (*59*).

However, there is evidence of residual secondary structure as concluded from FTIR and CD (Figure 1B) (Table 2). Far-UV CD spectra show the presence of helical- or turn-like secondary structures, which are further supported by CD deconvolution methods (Table 2). Moreover, FTIR also indicates the existence of turn-like conformations, although the percentage (~16%) is higher than that observed by CD (Table 2). The differences between the values of the populations of secondary structure reported by both techniques could be due to (i) the deconvolution procedures, (ii) the presence of aromatic residues (CD), which also absorb at 222 nm, or (iii) the empirical character of eq 5 used in determining the percentage of secondary structure (CD) (*31*). Whatever the reason, the amount of residual secondary structure in the domain could be due to transient turn- or helical-like conformations, which are flickering and disappear at high GdmCl concentrations. The fact that NMR did not show evidence of those secondary structures (except for the shifts in random-coil values in some resonances; Figure 3B), whereas the other spectroscopic techniques do, is due to their different spectroscopic time scales, as has been shown in other partially folded or completely unfolded proteins (*60*). The lack of compactness of the secondary structure would also explain the high susceptibility to proteases during our attempts to obtain the recombinant protein.

The presence of that small amount of secondary structure causes the chain to adopt a partially collapsed form. The expected radius, R, for a 61 residue long unfolded chain should be 22.5 Å, which was obtained by using the Flory's power law $R = R_h N^{\nu}$, being N the number of residues, ν the exponential scaling factor ($\nu = 0.598$), and R_h a constant which depends on the polypeptide length ($R_h = 1.927$ Å (*61*)); this calculated theoretical value is larger than that measured experimentally for bHLHN (17.4 Å; see above). We can further elaborate the issue of the protein

compactness caused by the residual structure by using the empirical equations developed by Uversky and co-workers (*18*, *62*). Our calculations suggest that the size of bHLHN is halfway between that of a molten globule (16.9 Å) and that of a premolten globule (21 Å). Since we have tested that bHLHN does not bind ANS (see above), as happens in other premolten globule structures (*18*), we suggest that the domain acquires a slightly collapsed premolten globule conformation. Further support to this type of partially folded structure comes from the fact that some partially unfolded proteins are only soluble in salt-free water (*23*, *63*); indeed, bHLHN was soluble in pure water at any concentration, but in the presence of buffers it precipitated at concentrations above 60 μM. This low solubility has hampered the measurement of the bHLHN−DNA affinity by isothermal titration calorimetry and other techniques. Moreover, low solubility and difficulty of expression have also been observed in other natural DNA-binding proteins (*24*, *25*), DNA-binding chimeras (*64*, *65*) (where studies have been only possible in the presence of low urea concentrations), and even in some examples, protein expression has been only possible in the presence of the natural protein partner (*27*). Taken together, these findings from different laboratories suggest that a majority of DNA-binding proteins (and not only bHLH domains) can be unfolded in the cells in the absence of the nucleic acid or other natural protein partners.

Finally, it has also been suggested that intrinsically unfolded proteins show an intriguing pH dependence (*66*). In bHLHN, the ellipticity shows two titrations (Figure 1C, blank circles, left axis), whereas the fluorescence did not change except at the extremes of pH (Figure 1C, filled squares, right axis). The acidic transition, which increases the percentage of helical structure (because the ellipticity at 222 nm decreased in absolute values, Figure 1C, blank circles), appears to be associated with a glutamic residue or the sole histidine; since such titration is not monitored by fluorescence, it must involve local conformational changes far away from the tryptophan and tyrosine residues. The most plausible explanation is that the possible conformational changes involve the histidine, which also absorbs at 222 nm (*35*).

The findings above raise the question: how does this natively unfolding behavior compare with that of other bHLH domains? Most of the isolated bHLH proteins characterized to date are natively unfolded proteins, and other DNA-binding proteins, either natural (*24*−*27*, *56*) or man made constructed (*64*, *65*), have been suggested to be unfolded; but, in any example, an extensive pH-unfolding study has been carried out to ascertain whether the unfolding features were pH-dependent. Therefore, to the best of our knowledge, this is the first extensive study of the behavior of a bHLH domain under different solution conditions. Furthermore, bHLHN is one of the few bHLH examples which remain monomeric (*19*, *67*, *25*) and unfolded up to 60 μM. For instance, the MyoD-bHLH (a master regulator of skeletal muscle development) at low concentrations is unfolded and, therefore, monomeric, but the isolated E47-bHLH protein form at low protein concentrations, a highly stable dimer (*19*); and similarly, it happens with the oncogenic bHLH proteins c-Myc, Max, and Mad1 (*25*).

DNA-Mediated Dimerization and "Fuzzy" Folding of bHLHN. Upon binding to the E3-box, bHLHN dimerizes as shown by fluorescence measurements. However, we do not know whether the association of particular DNA-bound dimers results from more favorable protein−DNA contacts in the complex or, alternatively, from better protein−protein interactions. In other bHLH examples, where the dimerization constant (usually in the low micromolar range) of the domains has been measured, it has been argued that DNA binding precedes to dimerization (*19*, *26*); since the self-oligomerization in bHLHN is very weak (larger than 60 μM), we conclude that in this particular domain homodimerization and binding occur concomitantly.

In other bHLH proteins, dimerization and DNA binding induce folding of the domain, as monitored by far-UV CD (*17*, *19*). In bHLHN, the far-UV CD spectrum of the protein also changes upon DNA binding, but the 1D-NMR did not show a spreading of the amide signals upon binding to any of the E-boxes (Figure 3C,D). This lack of signal dispersion in the NMR spectra, even though there is an increase of structure (as reported by CD, Figure 5A,B), could be due to the different spectroscopic time scales of far-UV CD and NMR, but it could be also reflecting the presence of a "fuzzy" complex between DNA and bHLHN (*20*, *68*), where the increased amount of secondary structure (as monitored by CD) is flickering and not hydrogen-bonded, and there are still long polypeptide patches which remain disordered. Although they have not been recognized as "fuzzy" DNA−protein complexes, there are other DNA-binding proteins in the literature which show high protein mobility upon binding to a specific DNA site or displacements (and then, mobility) in the recognition of the corresponding DNA-binding site (*56*, *69*, *70*); then, it is tempting to suggest that "fuzziness" might be a common feature among DNA−protein complexes than already thought.

However, if upon DNA binding we obtain a more ordered structure (even though the complex remains "fuzzy"), where does that structure come from? We suggest that the bound state should be chosen through selection of a conformer, which was sampled in the residual secondary structure adopted by the isolated bHLHN in solution and predicted by the bioinformatic programs (Figure 4): that is, the helical region at the C-terminus of the domain. This, in turn, suggests that the conformation of the bound form of bHLHN is more dependent on the local sequence (intrinsic local preferences of the monomer), and it is not determined by DNA binding. The restricted ensemble of available conformations explored by the isolated monomeric bHLHN would serve to reduce the entropic cost of binding and dimerization and, in addition, to increase affinity, at least for the formation of the first encounter complex, if the predominant population resembles that of the bound form. This "conformational selection" of states has been suggested to occur in other natively unfolded proteins upon binding to their partners (see ref *71* and references therein). However, recent kinetic studies with unfolded proteins do not support the "conformational selection" model, but rather the alternative "fly casting" model, by which a disordered protein accelerates its binding to its partner by making nonspecific and non-native interactions, which serve as anchoring points to induce the folding (*72*, *73*). Probably the selection of any of the two limiting models relies on the intrinsic sequence of the protein.

DNA Affinity and Comparison with Other bHLH Domains. The DNA binding constants of the majority of bHLH complexes are in the range 1−30 nM (*17*, *19*, *25*, *26*, *65*) or even smaller (from 10^{-13} to 10^{-9} M) (*24*); those measured in this work are in the range of micromolar. These results lead to the conclusion that the mode of binding can differ widely among the different DNA-binding proteins of the same family (*27*) or even in protein chimeras containing the DNA-binding region and dimerization modules from different members of the family

(*64*, *65*); we suggest, on the basis of the results of this work and those of other laboratories, that the affinity might be related to the grade of "fuzziness" in the complex formed. If this hypothesis is further tested in other DNA−protein examples solved in the future, it should be taken into account when transcriptional protein-screening studies are carried out.

Interestingly enough, for the E-boxes explored in this work, the homodimer NGN3 was not able to bind either of the boxes used here, and it was only in the presence of E47 (and, then, forming a heterodimer) when binding occurred. Therefore, it might be that although the affinity of bHLHN for the DNA was larger than that of NGN3, it would not be high enough; thus, the measured low affinity could be monitoring the poor folding (and, then, the "fuzziness" of the complex). In this sense, the small differences in the sequences among both bHLH domains are enough to promote binding but not to achieve a large specificity.

Conclusions. We have shown that the bHLH domain of NGN1 is a natively unfolded protein, with a premolten globule conformation. The domain shows a low affinity for DNA boxes and forms "fuzzy" complexes with the duplexes.

ACKNOWLEDGMENT

We deeply thank both referees for their constructive comments, ideas, and suggestions. J.L.N. thanks Dr. Francisco N. Barrera (Yale University, USA) for careful reading and suggestions on the manuscript and Dr. Lee Whitmore for providing access to DICHROWEB. We deeply thank May García, María del Carmen Fuster, Raquel A. Jorquera, and Javier Casanova for excellent technical assistance.

SUPPORTING INFORMATION AVAILABLE

Four figures containing the thermal denaturations of isolated bHLHN (Figure 1), the methyl region of the NMR spectrum of isolated bHLHN (Figure 2), the theoretical predictions of the natively unfolded features of bHLHN by using PONDR (Figure 3), and the Job's plots of the binding of bHLHN for both E-boxes (Figure 4). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature 363*, 38–45.
2. Kadesch, T. (1993) Consequences of heteromeric interactions among helix-loop-helix proteins. *Cell Growth Diff. 4*, 49–55.
3. Massari, M. E., and Murre, C. (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol. 20*, 429–440.
4. Ma, P. C., Rould, M. A., Wintraub, H., and Pabo, C. O. (1994) Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell 77*, 451–459.
5. Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997) Crystal structure of PHO4 bHLH domain DNA complex: flanking base recognition. *EMBO J. 16*, 4689–4697.
6. Blackwell, T. K., and Weintraub, H. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science 250*, 1104–1110.
7. Baxevanis, A. D., and Vinson, C. R. (1993) Interactions of coiled coils in transcription factors: where is the specificity? *Curr. Opin. Genet. Dev. 3*, 278–285.
8. Murre, C., McCaw, P. S., and Baltimore, D. (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding in, daugtherless, MyoD and Myc proteins. *Cell 56*, 777–783.
9. Jan, Y. N., and Jan, L. Y. (1993) HLH proteins, fly neurogenesis and vertebrate myogenesis. *Cell 75*, 827–830.
10. Guillemot, F. (1999) Vertebrate bHLH genes and the determination of neuronal fates. *Exp. Cell Res. 253*, 357–364.
11. Hassan, B. A., and Bellen, H. J. (2000) Doing the math: is the mouse a good model for fly development? *Genes Dev. 14*, 1852–1865.
12. Mizuguchi, R., Sugimori, M., Takebayashi, H., Kosako, H., Nagao, M., Yoshida, S., Nabeshima, Y., Shimamura, K., and Nakafuku, M. (2001) Combinatorial roles of olig2 and neurogenin2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. *Neuron 31*, 757–771.
13. Sun, Y., Nadal-Vicens, M., Misono, S., Lin, M. Z., Zubiaga, A., Hua, X., Fan, G., and Greenberg, M. E. (2001) Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell 104*, 365–376.
14. Lo, L., Dormand, E., Greenwood, A., and Anderson, D. J. (2002) Comparison of the generic neuronal differentiation and neuron subtype specification functions of mammalian achaete-scute and atonal homologs in cultured neural progenitor cells. *Development 129*, 1553–1567.
15. Schuurmans, C., Armant, O., Nieto, M., Stenman, J. M., Britz, O., Klenin, N., Brown, C., Langevin, L. M., Seibt, J., Tang, H., Cunningham, J. M., Dyck, R., Walsh, C., Campbell, K., Polleux, F., and Guillemot, F. (2004) Sequential phases of cortical specification involve neurogenin-dependent and -independent pathways. *EMBO J. 23*, 2892–2902.
16. Meierhans, D., el Ariss, C., Neuenschwander, M., Sieber, M., Stackhouse, J. F., and Allemann, R. K. (1995) DNA binding specificity of the basic-helix-loop-helix protein MASH-1. *Biochemistry 34*, 11026–11036.
17. Künne, A. G., and Allemann, R. K. (1997) Covalently linking BHLH subunits of MASH-1 increases specificity of DNA binding. *Biochemistry 36*, 1085–1091.
18. Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci. 11*, 739–756.
19. Wendt, H., Thomas, R. M., and Ellenberger, T. (1998) DNA-mediated folding and assembly of MyoD-E-47 heterodimers. *J. Biol. Chem. 273*, 5735–5743.
20. Tompa, P., and Fuxreiter, M. (2002) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci. 33*, 2–8.
21. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol. 131*, 266–280.
22. Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem. 182*, 319–326.
23. Li, M., Liu, J., Ran, X., Fang, M., Shi, J., Qin, H., Goh, J.-M., and Song, J. (2006) Resurrecting abandoned proteins with pure water: CD and NMR studies of protein fragments solubilized in salt-free water. *Biophys. J. 91*, 4201–4209.
24. Jung, K. C., Rhee, H. S., Park, C. H., and Yang, C.-H. (2005) Determination of the dissociation constants for recombinant c-Myc, Max, and DNA complexes: the inhibitory effect of linoleic acid on the DNA-binding step. *Biochem. Biophys. Res. Commun. 334*, 269–275.
25. Hu, J., Banerjee, A., and Goss, D. J. (2005) Assembly of b/HLH/Z proteins c-Myc, Max, and Mad1 with cognate DNA: importance of protein-protein and protein-DNA interactions. *Biochemistry 44*, 11855–11863.
26. Meier-Andrejszki, L., Bjelic, S., Naud, J.-F., Lavigne, P., and Jelesarov, I. (2007) Thermodynamics of b-HLH-LZ protein binding to DNA: the energetic importance of protein-DNA contacts in site-specific E-box recognition by the complete gene product of the Max p21 transcription factor. *Biochemistry 46*, 12427–12440.
27. Chapman-Smith, A., and Whitelaw, M. L. (2006) Novel DNA binding by a basic helix-loop-helix protein: the role of the dioxin receptor PAS domain. *J. Biol. Chem. 281*, 12535–12545.
28. Huang, H.-P., Liu, M., El-Hodiri, H., Chu, K., Jamrich, M., and Tsai, M.-J. (2000) Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. *Mol. Cell. Biol. 20*, 3292–3307.
29. Cicero, D. O., Nadra, A. D., Eliseo, T., Dellarole, M., Paci, M., and de Prat-Gay, G. (2006) Structural and thermodynamic basis for the enhanced transcriptional control by the human papillomavirus strain-16 E2 protein. *Biochemistry 45*, 6551–6560.
30. Künne, A. G., Sieber, M., Meierhans, D., and Allemann, R. K. (1998) Thermodynamic of the DNA binding reaction of transcription factor MASH-1. *Biochemistry 37*, 4217–4223.
31. Muro-Pastor, M. I., Barrera, F. N., Reyes, J. C., Florencio, F. J., and Neira, J. L. (2003) The inactivating factor of glutamine synthetase, IF7, is a "natively unfolded" protein. *Protein Sci. 12*, 1443–1454.

32. Facchiano, A., and Ragone, R. (2003) Modification of Job's method for determining the stoichiometry of protein-protein complexes. *Anal. Biochem. 313*, 170–172.
33. Otzen, D. E., Lundvig, D. M., Wimmer, R., Nielsen, L. H., Pedersen, J. R., and Jensen, P. H. (2005) p25alpha is flexible but natively folded and binds tubulin with oligomeric stoichiometry. *Protein Sci. 14*, 1396–1409.
34. Birdsall, B., King, R. W., Wheeler, M. R., Lewis, C. A., Jr., Goode, S., Dunlap, R. B., and Roberts, G. C. K. (1983) Correction for light absorption in fluorescence studies of protein-ligand interactions. *Anal. Biochem. 132*, 353–361.
35. Kelly, S. M., and Price, N. C. (2000) The use of circular dichroism in the investigation of protein structure and function. *Cur. Protein Pept. Sci. 1*, 349–384.
36. Bohm, G., Muhr, R., and Jaenicke, R. (1992) Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng. 5*, 191–195.
37. Whitmore, L., and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers 89*, 392–400.
38. Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res. 32*, 668–673.
39. Piotto, M., Saudek, V., and Sklenar, V. (1993) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR 2*, 661–665.
40. Czypionka, A., de los Paños, O. R., Mateu, M. G., Barrera, F. N., Hurtado-Gómez, E., Gómez, J., Vidal, M., and Neira, J. L. (2007) The isolated C-terminal domain of Ring 1B is a dimer made of stable, well-structured monomers. *Biochemistry 46*, 12764–12776.
41. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Smith, L. J. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry 38*, 16424–16431.
42. Anglister, J., Grzesiek, S., Ren, H., Klee, C. B., and Bax, A. (1993) Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS. *J. Biomol. NMR 3*, 121–126.
43. Sklenar, V., and Bax, A. (1987) Spin echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson. 74*, 469–479.
44. Goormaghtigh, E., and Ruysschaert, J. M. (2004) Subtraction of atmospheric water contribution in Fourier transform infrared spectroscopy of biological membranes and proteins. *Spectrochim. Acta 50A*, 2137–2144.
45. Goormaghtigh, E., Ruysschaert, J. M., and Raussens, V. (2006) Evaluation of the information content in infrared spectra for protein secondary structure determination. *Biophys. J. 90*, 2946–2957.
46. Raussens, V., Ruysschaert, J. M., and Goormaghtigh, E. (2004) Analysis of H-1/H-2 exchange kinetics using model infrared spectra. *Appl. Spectrosc. 58*, 68–82.
47. Romero, P., Obadovic, Z., Li, X., Garner, E., Brown, C., and Dunker, A. K. (2001) Sequence complexity of disordered proteins. *Proteins: Struct., Funct., Genet. 42*, 38–48.
48. Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics 21*, 3433–3444.
49. Priluski, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., and Sussman, J. L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics 21*, 3435–3438.
50. Schmid, F. X. (1997) Optical spectrosocopy to characterize protein conformation and conformational changes, in Protein Structure (Creighton, T. E., Ed.) 2nd ed., pp 261−297, Oxford University Press, Oxford.
51. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., and Gilmanshin, R. I. (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers 31*, 119–128.
52. Thurlkill, R. L., Grimsley, G. R., Scholtz, J. M., and Pace, C. N. (2006) pK values of the ionizable groups of proteins. *Protein Sci. 15*, 1214–1218.
53. Surewicz, W. K., and Mantsch, H. H. (1988) New insight into protein secondary structure from resolution-enhanced infrared spectra. *Biochim. Biophys. Acta 952*, 115–30.
54. Wüthrich, K. (1986) NMR of proteins and nucleic acids, John Wiley & Sons, New York.
55. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins 17*, 75–86.
56. Chan, I.-S., Shahravan, S. H., Fedorova, A. V., and Shin, J. A. (2008) The bZIP targets overlapping DNA sub-sites within a half site resulting in increasing binding affinities. *Biochemistry 47*, 9646–9652.
57. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol. 6*, 197–208.
58. Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci. 27*, 527–529.
59. Calloni, G., Lendel, C., Campioni, S., Giannini, S., Gliozzi, A., Reini, A., Vendruscolo, M., Dobson, C. M., Salvatella, X., and Chitti, F. (2008) Structure and dynamics of a partially folded protein are decoupled from its mechanism of aggregation. *J. Am. Chem. Soc. 130*, 13040–13050.
60. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, G., Sato, S., Jas, G. S., Freund, S. M. V., Alonso, D. O. V., Daggett, V., and Fersht, A. R. (2003) The complete folding pathway of a protein from nanosecond to microsecond. *Nature 42*, 863–867.
61. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S., and Plaxco, K. W. (2004) Random-coil behaviour and the dimensions of proteins. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12491–12496.
62. Tcherkasskaya, O., Davidson, E. A., and Uversky, V. N. (2003) Biophysical constraints of protein structure prediction. *J. Proteome Res. 2*, 37–42.
63. Song, J. (2009) Insight into "insoluble proteins" with pure water. *FEBS Lett. 583*, 953–959.
64. Chow, H.-K., Xu, J., Shahravan, S. H., De Jong, A. T., Chen, G., and Shin, J. A. (2008) Hybrids of the bHLH and bZIP protein motifs display different DNA-binding activities *in vivo* vs. *in vitro*. *PLoS ONE 3*, e3514.
65. Xu, J., Chen, G., De Jong, A. T., Shahravan, S. H., and Shin, J. A. (2009) Max-E47, a designed minimalist protein that targets the E-box DNA site in vivo and in vitro. *J. Am. Chem. Soc. 131*, 7839–7848.
66. Receveur-Brechot, V., Bourhis, J. M., Uversky, V. N., Canard, B., and Longhi, S. (2006) Assessing protein disorder and induced folding. *Proteins 62*, 24–45.
67. Fairman, R., Beran-Steed, R. K., Anthony-Cahill, S. J., Lear, J. D., Stafford, W. F. I., De Grado, W. F., Benfield, P. A., and Brenner, S. L. (1993) Multiple oligomeric states regulate the DNA binding of helix-loop-helix peptides. *Proc. Natl. Acad. Sci. U.S.A. 90*, 10429–10433.
68. Sigalov, A. B., Kim, W. M., Slaien, M., and Stern, L. J. (2008) The intrinsically disordered citoplasmatic domain of the T cell receptor ζ chain binds to the Nef protein of simian immunodeficiency virus without a disorder to order transition. *Biochemistry 47*, 12942–12944.
69. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G., III (1999) Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. *J. Mol. Biol. 285*, 2133–2146.
70. Saudek, V., Pasley, H. S., Gibson, T., Gausepohl, H., Frank, R., and Pastore, A. (1991) Solution structure of the basic region from the transcriptional activator GCN4. *Biochemistry 30*, 1310–1317.
71. Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K., and Uversky, V. N. (2006) Analysis of molecular recognition features (MoRFs). *J. Mol. Biol. 362*, 1043–1059.
72. Narayanan, R., Ganesh, O. K., Edison, A. S., and Hagen, S. J. (2008) Kinetics of folding and binding of an intrinsically disordered protein: the inhibitor of yeast aspartic proteinase YprA. *J. Am. Chem. Soc 130*, 11477–11485.
73. Rea, A. M., Thurston, V., and Searle, M. S. (2009) Mechanism of ligand induced folding of a natively unfolded helix-like variant of rabbit I-BABP. *Biochemistry 48*, 7556–7564.