

Maximum entropy reconstruction of joint ϕ, ψ -distribution with a coil-library prior: the backbone conformation of the peptide hormone motilin in aqueous solution from ϕ and ψ -dependent J-couplings

Tariq Massad · Jüri Jarvet · Risto Tanner ·
Katrín Tomson · Julia Smirnova · Peep Palumaa ·
Mariko Sugai · Toshiyuki Kohno · Kalju Vanatalu ·
Peter Damberg

Received: 9 May 2006/Accepted: 13 February 2007/Published online: 26 April 2007
© Springer Science+Business Media B.V. 2007

Abstract In this paper, we present a new method for structure determination of flexible “random-coil” peptides. A numerical method is described, where the experimentally measured $^3J^{H^N}H^z$ and $^3J^{H^z}N^{i+1}$ couplings, which depend on the ϕ and ψ dihedral angles, are analyzed jointly with the information from a coil-library through a maximum entropy approach. The coil-library is the distribution of dihedral angles found outside the elements of the secondary structure in the high-resolution protein structures. The method results in residue specific joint ϕ, ψ -distribution functions, which are in agreement with the experimental J-couplings and minimally committal to the information in the coil-library. The 22-residue human peptide hormone motilin, uniformly ^{15}N -labeled was studied. The $^3J^{H^z}N^{i+1}$ were measured from the E.COSY pattern in the sequential NOESY cross-peaks. By employing homodecoupling and an in-phase/anti-phase

filter, sharp H^z -resonances (about 5 Hz) were obtained enabling accurate determination of the coupling with minimal spectral overlap. Clear trends in the resulting ϕ, ψ -distribution functions along the sequence are observed, with a nascent helical structure in the central part of the peptide and more extended conformations of the receptor binding N-terminus as the most prominent characteristics. From the ϕ, ψ -distribution functions, the contribution from each residue to the thermodynamic entropy, i.e., the segmental entropies, are calculated and compared to segmental entropies estimated from ^{15}N -relaxation data. Remarkable agreement between the relaxation and J-couplings based methods is found. Residues belonging to the nascent helix and the C-terminus show segmental entropies, of approximately $-20\text{ J K}^{-1}\text{ mol}^{-1}$ and $-12\text{ J K}^{-1}\text{ mol}^{-1}$, respectively, in both series. The agreement between the two estimates of the segmental entropy, the agreement with the observed J-couplings, the agreement with the CD experiments, and the assignment of population to sterically allowed conformations show that the ϕ, ψ -distribution functions are indeed meaningful and useful descriptions of the conformational preferences for each residue in this flexible peptide.

Electronic Supplementary Material The online version of this article (doi:[10.1007/s10858-007-9150-1](https://doi.org/10.1007/s10858-007-9150-1)) contains supplementary material, which is available to authorized users.

T. Massad · J. Jarvet · P. Damberg (✉)
Department of Biochemistry and Biophysics, Stockholm
University, Svante Arrhenius v. 12, 106 91 Stockholm, Sweden
e-mail: peter.damberg@dbb.su.se

J. Jarvet · R. Tanner
National Institute of Chemical Physics and Biophysics,
Akadeemia tee 23, 12618 Tallinn, Estonia

K. Tomson · J. Smirnova · P. Palumaa · K. Vanatalu
Tallinn Technical University, Inst. of Gene Technology,
Akadeemia tee 15, 12618 Tallinn, Estonia

M. Sugai · T. Kohno
Mitsubishi Kagaku Institute of Life Sciences (MITILS), 11
Minamiooya, Machida-shi, Tokyo 194-8511, Japan

Keywords Random coil · Motilin · Maximum entropy · NMR · J-couplings

Abbreviations

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
IMAC	Immobilized metal ion affinity chromatography
TFA	Trifluoroacetic acid
SPE	Solid phase extraction
NMR	Nuclear magnetic resonance

ME	Maximum entropy
HSQC	Heteronuclear single quantum correlation spectroscopy
TOCSY	Total correlation spectroscopy
NOESY	Nuclear Overhauser enhancement spectroscopy
CD	Circular dichroism
PMSF	Phenylmethylsulfonyl fluoride
IPTG	Isopropyl β -D-thiogalactopyranoside
YUH	Yeast ubiquitin hydrolase
IPAP	In-phase/anti-phase
SMT-model	Swindell, MacArthur, and Thornton model
RDCs	Residual dipolar couplings

Introduction

The occurrence of unstructured polypeptides, especially polypeptide hormones, has been recognized for many years (Boesch et al. 1978; Daniels et al. 1978). There is no doubt nowadays that many unstructured polypeptides have well-defined biological functions. It is becoming increasingly clear that many functionally important protein segments occur outside globular domains (Dunker et al. 2002). Unstructured polypeptides play an important role in the regulation of transcription and translation, cellular signal transduction, the storage of small molecules, and regulation of self-assembly of large multiprotein complexes (Dunker et al. 2002; Tompa 2002; Uversky 2002; Wright and Dyson 1999). Non-globular segments of proteins act as sorting signals, mediate the post-translational modification processes such as proteolysis and phosphorylation (Linding et al. 2003). Moreover, unstructured polypeptides are related to DNA/RNA–protein interactions, they also function as protein ligands, inhibitors and scavengers (Eker et al. 2004). The structural characterization of the disordered polypeptides is important for understanding the protein function, folding pathways, receptor binding, and aggregation processes (Schweers et al. 1994). The amount of currently available information about the nature of the structure of the disordered polypeptides is limited, as they remain beyond the reach of classical structural biology.

Crystal structure analysis cannot provide information on unstructured states. Crystallography can only indicate the presence of unstructured regions through the absence of electron density in local regions (Dyson and Wright 2005). The unstructured state of flexible polypeptides is studied by many spectroscopic techniques. They include circular dichroism (CD) spectroscopy (Dukor and Keiderling 1991), Fourier transform infrared spectroscopy (Sosnick and Trewella 1992), fluorescence energy transfer (Buckler

et al. 1995), Tryptophan fluorescence transfer (Swaminathan et al. 1994), Nuclear magnetic resonance spectroscopy (NMR) (Shortle 1996; Wüthrich 1994), IR vibrational CD (Shi et al. 2002). Random-coil dimensions have also been described by NMR diffusion studies (Danielsson et al. 2002).

Flexible polypeptides are generally categorized as a “random-coil” structural state. The concept of a random-coil in statistical mechanics is a polymer where all the degrees of freedom are used in the conformational space, and there is no conformational restriction along the polypeptide chain (Smith et al. 1996b). For a peptide that corresponds to a situation where the population of different conformations of each amino acid residue is determined solely by the intrinsic preferences of the amino acid. The conformation of sequential neighbors in a random coil should be independent. Many pieces of evidence suggest that the “random-coil” polypeptides might not have completely random structure, and that there are some overall conformational preferences (Bochicchio and Tamburro 2002). Several different views on those conformational preferences have been put forward. Some claim that the flexible polypeptides may resemble the left handed β_1 -helix, also known as polyproline II conformation (PPII) (Bochicchio and Tamburro 2002; Rath et al. 2005). Recently, an analysis of NMR-data from short peptides using time-averaged restraints indicated that the preference for PPII conformation may be much less pronounced (Makowska et al. 2006). Residual dipolar coupling (RDCs) from denatured $\Delta 131\Delta$ staphylococcal nuclease display similarity to RDCs of native $\Delta 131\Delta$. This was interpreted as a preservation of native-like topology in denatured $\Delta 131\Delta$ (Shortle and Ackerman 2001). Mohana-Borges et al. (2004) put forward a quite different explanation for the RDCs in denatured apomyoglobin and denatured $\Delta 131\Delta$, which relates the RDCs to interactions between stretches of residues with the alignment media. A third explanation for the RDCs in denatured proteins has been provided by ensembles of structures generated from a coil library (Jha et al. 2005a; Bernadó et al. 2005). In this model, the distributions of dihedral angles of the 20 common amino acids observed in coil-regions of high resolution crystal structures are considered to be representative of the preferences of the 20 different residue types in flexible peptides. This model was first put forward by Swindells et al. (1995), and will be referred to as the SMT-model throughout this article. The SMT-model, and more sophisticated coil-library modeling approaches, where chemical nature and conformational preferences of adjacent residues are accounted for, have shown promising explanatory power for the residue specific average $^3J^{HNH}$ -couplings (Serrano 1995; Smith et al. 1996a; Avbelj and Baldwin 2003),

nearest neighbor effects (Griffiths-Jones et al. 1998), and RDCs for chemically denatured proteins (Jha et al. 2005a; Bernadó et al. 2005). Coil-library modeling appears as the leading idea for the understanding of flexible peptides today.

The conformational ensemble depends on conditions such as solvent composition, temperature and salt. Such changes in the conformational ensemble when changing the conditions are not explained by coil-library modeling in its current form. For example the coil-library model put forward by Jha et al. (2005b) explain the RDCs in urea denatured apo myoglobin fairly well while the same model show less promising agreement for the RDCs observed in the acid denatured form. An approach where the conformational ensemble can be described for a particular set of conditions is desirable.

The aim of this study is to further the insight into the local conformations of the residues in flexible peptides by analyzing homo and heteronuclear J-couplings which are obtained under a certain set of experimental conditions. The J-couplings are analyzed in a novel way together with the information from a coil-library. Residue specific joint ϕ, ψ -distribution functions are assigned for each residue in the polypeptide through a maximum entropy (ME) analysis (Jaynes 1963). The Shannon's entropy is defined relative to the ϕ and ψ dihedral angle distributions obtained from the coil library. Thereby the background information from the coil library is incorporated. The natural goal is to apply the outlined approach to a test case and scrutinize the results from the suggested analysis. The human peptide hormone motilin, with 22 residues, is studied as the test case. The success of the approach is judged from three independent criteria. First, the residue specific joint ϕ, ψ -distribution functions are in exact agreement with the observed J-couplings. Second, it is checked that the resulting joint ϕ, ψ -distributions assign the population to sterically allowed conformations. Third, the ϕ, ψ -distribution functions explain the CD-spectrum of the peptide.

In addition to the structural insight provided by the ϕ, ψ -distribution functions their usefulness is exemplified by calculating the contribution to the conformational entropy from each fragment, i.e., the segmental entropy. Those segmental entropies are compared to segmental entropies evaluated from ^{15}N -relaxation data according to a published protocol (Yang and Kay 1996).

Theory

In a flexible peptide the conformation of a molecule changes constantly mainly through rotation around chemical bonds, i.e., changing dihedral angles. A

time-independent ensemble of conformations describes the conformational preferences at thermal equilibrium. Each dihedral angle, θ , populates values according to a distribution, $\rho(\theta)$. The J-couplings and the dihedral angles are interconnected via the well-known Karplus relation (Karplus 1959), $J(\theta)$. In a flexible peptide the observed couplings corresponds to ensemble averages

$$\langle J \rangle = \int J(\theta) \cdot \rho(\theta) d\theta. \quad (1)$$

When $\rho(\theta)$ is known it is straightforward to evaluate the ensemble average. The inverse problem, to find $\rho(\theta)$ from the observed J-coupling is an ill-posed problem because many different distributions fit the observed J-coupling equally well. Such ill-posed problems may be approached by employing the maximum entropy principle. The principle of maximum entropy stipulates that one should select the distribution that leaves the largest uncertainty of the distributions that are consistent with the observations (Stigler 1982). By selecting the distribution which is the most uncertain, over interpretation of the data is avoided since only as little information as necessary is inferred. The distribution, which maximizes the entropy while being consistent with the observations, is put forward as the new leading hypothesis. The Shannon entropy is used as a quantitative measure of the uncertainty. The Shannon entropy, S , of a dihedral angles distribution is given by Eq. 2

$$S = - \int \rho(\theta) \ln \frac{\rho(\theta)}{P(\theta)} d\theta, \quad (2)$$

where $\rho(\theta)$ is the distribution function. The term $P(\theta)$ is the so-called prior. It is the probability the investigator assigns to different values of θ before considering the experimental observations. Originally, the prior was introduced to avoid the mathematical invariance problem when continuous distributions were considered. If nothing is known about the distribution, equal probability should be assigned to all dihedral angles, i.e., a flat prior should be assigned where $P(\theta)$ equals a constant for all θ . However, for the ϕ and ψ -dihedral angles it is very well-established that certain combinations of ϕ and ψ are more populated than others. In addition, the preference for gauche or trans over eclipsed conformations is widely accepted. Crystal structures of proteins distinguish allowable ϕ and ψ angles in certain allowed regions in the Ramachandran map, while other regions are totally disallowed (Lovell et al. 2003; Ramachandran and Sasisekharan 1968). It appears reasonable to assign higher probability to such allowed regions. When analyzing two dihedral angles for which information exist about their interdependence it is beneficial to analyze the two angles jointly. For the ϕ

and ψ angles in peptides, useful information about the interdependence of the two angles exists. The Shannon entropy for $\rho(\phi, \psi)$ is given by:

$$S = - \iint \rho(\phi, \psi) \ln \frac{\rho(\phi, \psi)}{P(\phi, \psi)} d\phi d\psi. \quad (3)$$

In this paper the SMT-model is used to generate an “off the shelf” prior, $P(\phi, \psi)$, for non-globular peptides. The Shannon entropy is used as a tool for ranking all possible distributions according to how unexpected they were before considering the experimental data. The least unexpected $\rho(\phi, \psi)$ corresponds to the highest Shannon entropy while all other distributions have lower Shannon entropy.

The purpose of defining the Shannon entropy was to reach a situation where the dihedral angle distribution could be estimated from sparse J-couplings data as a pseudo inverse of Eq. 1. Of all distribution functions which correspond to an ensemble averaged J-coupling identical to the observation the one which maximize Eq. 3 is selected. Thereby the least unexpected distribution function which is consistent with the observed J-couplings is selected.

The standard procedure for the optimization of the entropy under equality constraints is the Lagrange method (Adams 1990). In this procedure, one Lagrange multiplicator per constraint is introduced. Thereby, it is possible to derive the closed form expression for the distribution function, $\rho(\phi, \psi)$, which is consistent with the experimental observations and maximizes the Shannon entropy. In the current work ${}^3J^{H^zN^{i+1}}(\psi)$ and ${}^3J^{H^zH^z}(\phi)$ constitute the experimental constraints. The same principle could be used for other ensemble averaged experimental observation like other J-couplings or distances etc.

The explicit derivation of $\rho(\phi, \psi)$ when ${}^3J^{H^zN^{i+1}}(\psi)$ and ${}^3J^{H^zH^z}(\phi)$ are the experimental constraints is given in the appendix. The derivation results in Eq. 4

$$\begin{aligned} \rho(\phi, \psi | \lambda_0, \lambda_1, \lambda_2) &= P(\phi, \psi) \exp(\lambda_0 + \lambda_1 {}^3J^{H^zH^z}(\phi) \\ &\quad + \lambda_2 {}^3J^{H^zN^{i+1}}(\psi)) \end{aligned} \quad (4)$$

The parameterization of Wang and Bax (1995, 1996) was used for the Karplus curves, i.e., ${}^3J^{H^zH^z}(\phi) = 6.98 * \cos(\phi - \pi/3)^2 - 1.38 * \cos(\phi - \pi/3) + 1.72$, and ${}^3J^{H^zN^{i+1}}(\psi) = -0.88 * \cos(\psi - 2\pi/3)^2 + 0.61 * \cos(\psi - 2\pi/3) - 0.27$.

The values of the Lagrange multiplicator, λ , are determined by varying them until the J-couplings calculated from the distribution agree with the observed couplings simultaneously as a normalized distribution is obtained. The values of the λ s consequently depend on the observations. The distribution function that is obtained by evaluating Eq. 4 using the correct λ -values is identified as a

posterior distribution obtained from the prior after considering the observations, cf. Bayes theorem (Stigler 1982). Hereafter the distribution function that optimizes the Shannon entropy while being consistent with the observations is referred to as the posterior.

Materials and methods

Expression and purification of ${}^{15}\text{N}$ -labeled motilin

Motilin was expressed as a fusion protein to his-tagged ubiquitin. The expression vector for the His10-ubiquitin-motilin fusion protein was constructed in the similar way as described earlier (Kohno et al. 1998) by using chemically synthesized oligonucleotide encoding motilin and transferred to *E. coli* BL21(DE3). Large-scale uniform ${}^{15}\text{N}$ labeling was carried out as a fed-batch culture (Paalme et al. 1990) by feeding the cells with glycerol and ${}^{15}\text{NH}_4\text{Cl}$ in 0.8 M M9 salts medium at 30°C in an Applikon 21 fermenter equipped with computer control. When $A_{600\text{ nm}} = 13$ was reached the induction of the recombinant protein synthesis was initiated by adding isopropyl- β -D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. To avoid overfeeding with glycerol the algorithm of adaptastat (Tomson et al. 2006) was applied after the addition of IPTG. Biomass was harvested at $A_{600\text{ nm}} = 18$ after 4 h of induction and the cell paste frozen at -70°C until use. The biomass was resuspended in 3–4 volumes of 50 mM Tris, 0.1 M NaCl buffer, pH 8.0 (buffer A) containing 0.5 mM phenylmethylsulfonyl fluoride (PMSF) and disrupted three times with a French press at 300 atm and 0°C. Cell debris was removed by centrifugation at 25,000g for 30 min at 4°C and the cell lysate was used for further protein purification. Cell lysates were applied to a 5 ml immobilized Ni^{2+} affinity column (Chelating Sepharose Fast Flow, GE Healthcare Life Sciences) equilibrated with buffer A. Column was washed with 50 ml of buffer A, then with 50 ml of buffer A containing 20 mM imidazole and adsorbed proteins were subsequently eluted with buffer A containing 250 mM imidazole. Fractions were analyzed for protein by SDS PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). Fusion protein-containing fractions were pooled and used for enzymatic cleavage of motilin from His10-ubiquitin-motilin fusion protein by recombinant yeast ubiquitin hydrolase (YUH) (Kohno et al. 1998). YUH was expressed in *E. coli* BL21(DE3) and purified by immobilized Ni^{2+} metal affinity chromatography (IMAC) as described above (except that buffer A contained 1 mM β -mercaptoethanol). Cleavage conditions were determined experimentally for each batch and usually 3 h of incubation of 4 mg/ml substrate and 0.1 mg/ml enzyme at 37°C were sufficient for complete cleavage of

motilin from the fusion protein. After the enzymatic cleavage, the pH of the reaction mixture was adjusted to 3.5 using 50% (v/v) trifluoroacetic acid (TFA), the mixture was centrifuged at 25,000g 4°C 20 min and the supernatant further purified using HPLC. The acidified reaction mixture was applied to OASIS HLB Vac RC solid phase extraction (SPE) cartridge (10 µm, 19 × 250 mm; Waters) and the motilin containing fraction eluted with a solution of 40% (v/v) methanol and 40% (v/v) acetonitrile in water. The eluent was diluted by adding two volumes of 0.1% (w/v) TFA in water and motilin was purified by reversed-phase HPLC via an XTerra Prep MS C18 OBD column (Waters) with solvents A (0.1% (w/v) TFA in water), B (0.1% TFA in methanol) and C (0.1% TFA in acetonitrile). The elution was carried out at 3 ml/min with linear gradient from 12% (v/v) to 34 % of both B and C over 60 min, with simultaneous monitoring of absorbance at 280 nm, and molecular ion masses from 400 to 2,000 amu via post-column splitter 1:30. Motilin was identified according to the characteristic mass spectrum containing 4+, 3+ and 2+ ions, collected and lyophilized.

NMR-spectroscopy

The experiments were performed at 298 K using Varian INOVA spectrometers operating at 600 and 800 MHz proton resonance frequency. The $^3J^{H^{\alpha}N^{i+1}}$ couplings were measured from the E.COSY pattern observed for the

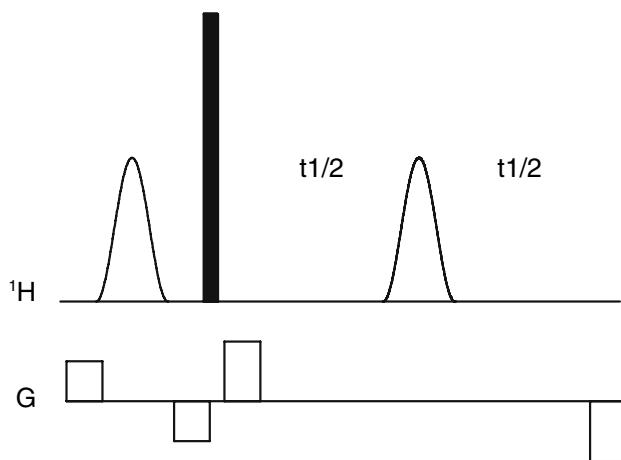


Fig. 1 The homodecoupling element that substitutes the regular evolution time, t_1 , in the NOESY and TOCSY experiments is displayed. For the low intensity shaped ^1H pulses, hyperbolic secant inversion pulses of 7.7 ms which invert H^N and H^β while leaving H^α resonances relatively unperturbed. The square pulse between the two shaped pulses is a 180° pulse used for refocusing. The gradient pairs, with opposite polarity surrounding the shaped pulses purges magnetization that is refocused by the shaped pulses. The gradient pulses were 200 µs and intensities of plus or minus approximately 3 Gauss/cm for the first pair and approximately 5 Gauss/cm for the second pair

sequential H^α to N^{i+1} cross-peaks in NOESY spectra. To obtain the sharpest possible signals in the H^α dimension the t_1 evolution time was modified to include a homodecoupling element displayed in Fig. 1.

In the middle of the t_1 time, a band-selective inversion pulse [a 7.7 ms long adiabatic hyperbolic secant pulse (Silver et al. 1984)], which was designed to invert the H^β and H^N protons without perturbing the H^α resonances severely, was applied. To compensate for chemical shift evolution and Bloch-Siegert shift during the band-selective pulse the t_1 time is preceded by an identical band-selective pulse followed by a hard refocusing pulse. The gradient pulses purges magnetization, which is refocused by the band-selective inversion pulse. The overlap was reduced further by introducing an in-phase/anti-phase filter in between the 90° read pulse, following the mixing time, and the 3-9-19 solvent suppression element (Sklenar et al. 1993) in the NOESY sequence. The in-phase/anti-phase filter is displayed in Fig. 2. Two data sets were acquired using either the filter element a or b, where the a element gives in-phase doublets and the b element gives anti-phase doublets. By displaying the sum and the difference between the a and b data sets two subspectra are obtained which contain either the right or the left component of the doublet due to the 92 Hz one-bond coupling between the H^N and ^{15}N . Alternatively, a spin-state selective filter could have been used. The phase of the 90° read pulse after the mixing time was shifted 90° when employing filter element b.

The relaxation rates were measured using HSQC-style inverse detected pulse sequences employing sensitivity enhancement (Farrow et al. 1994). The transverse relaxation rates were measured employing a spin-lock (i.e. $R_{1\rho}$) with 1.5 kHz spin-lock field strength. Eight different relaxation

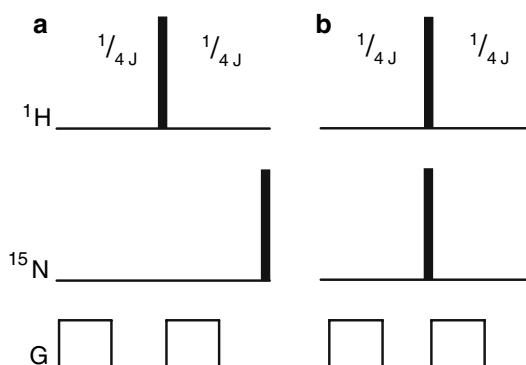


Fig. 2 In the in-phase/anti-phase filter employed here either the in-phase character of the doublet is preserved through the incorporation of the element a displayed to the left or the in-phase character of the doublet is changed into an anti phase doublet through the incorporation of the element b displayed to the right. All radio frequency pulses are 180° pulses and the gradient pulses purge magnetization that is not refocused by the ^1H refocusing pulse. The J coupling was set to 92 Hz when calculating the duration of the delay

delays were sampled ranging from 10 ms to 250 ms. For the measurement of the longitudinal auto relaxation rates seven different relaxation delays were sampled ranging from 10 ms to 1 s. The nuclear Overhauser enhancements from the protons to the nitrogen atoms were estimated from the ratio between the steady state ^{15}N intensities, observed when the protons were saturated for 5 s and the equilibrium magnetization, observed with 10 s recovery delays between the experiments. The spectra were Fourier transformed, baseline corrected and quantified within vnmr program. The relaxation rates were fitted using Mathematica (Wolfram 1991). For each residue, two correlation times and one order-parameter was fitted to the three relaxation rates, using an effective NH-distance of 1.02 Å and an effective ^{15}N CSA of –169 ppm (Damberg et al. 2005). The median of the long correlation time was used as the global correlation time when one internal correlation time and one order-parameter was fitted for each residue.

Formulating the priors from the SMT-model

The priors were formulated from a maximum entropy analysis of the precompiled coil-library described in detail in Fitzkee et al. (2005) which was downloaded in September 2004, with the criteria less than 20% sequence identity, refinement factor better than 0.25 and resolution better than 1.6 Å, i.e., 1.5 Å or better.

The dihedral angles of the segments were cleared from the context entries and residues with either the preceding or succeeding peptide plane in the *cis* configuration were removed. The data in the remaining coil-library was reformatted to 20 lists with the ϕ and ψ angles for each residue type. Those lists were analyzed employing maximum entropy formalism similar to the approach used in (Rowicka and Otwinowski 2004).

For each of the 20 residue types the average trigonometric moments were calculated according to the following equations:

$$\begin{aligned} m_{cc}^{kl} &= \frac{1}{N} \sum_{i=1}^N \cos(k\phi_i) \cos(l\psi_i) \\ m_{cs}^{kl} &= \frac{1}{N} \sum_{i=1}^N \cos(k\phi_i) \sin(l\psi_i) \\ m_{sc}^{kl} &= \frac{1}{N} \sum_{i=1}^N \sin(k\phi_i) \cos(l\psi_i) \\ m_{ss}^{kl} &= \frac{1}{N} \sum_{i=1}^N \sin(k\phi_i) \sin(l\psi_i) \end{aligned} \quad (5)$$

Here N is the number of occurrences of a certain residue type in the coil-library and the indices k and l run from zero or one to five. In principle, any number of trigonometric moments is possible to calculate. However, for higher trigonometric moments, where k or l are large, the uncertainty makes the information content negligible. Attempts using more trigonometric moments, with indices k and l up to 15 made no improvements. The 20 maximum entropy distributions, relative to flat priors that are consistent with the trigonometric moments were obtained by the method employing one Lagrange multiplicator per trigonometric moment, which results in the expression for the distribution function:

$$P(\phi, \psi) = \exp \left[\begin{array}{l} \sum_{k=0}^5 \sum_{l=0}^5 \lambda_{cc}^{kl} \cos(k\phi) \cos(l\psi) \\ + \sum_{k=0}^5 \sum_{l=1}^5 \lambda_{cs}^{kl} \cos(k\phi) \sin(l\psi) \\ + \sum_{k=1}^5 \sum_{l=0}^5 \lambda_{sc}^{kl} \sin(k\phi) \cos(l\psi) \\ + \sum_{k=1}^5 \sum_{l=1}^5 \lambda_{ss}^{kl} \sin(k\phi) \sin(l\psi) \end{array} \right]. \quad (6)$$

In all calculations $P(\phi, \psi)$ was digitalized on a grid of 128×128 points giving a angular resolution of better than 3° per point ($360^\circ/128$). The Lagrange multiplicators, λ , were found by a randomized search, starting from zero, which was stopped when the sum of squared errors (SoS) between the trigonometric moments calculated from the coil library and those calculated from $P(\phi, \psi)$ was smaller than 0.06. This is slightly lower than what is expected due to uncertainty in the trigonometric moments calculated from the coil-library. The expected value for the SoS due to the limited size of the coil-library was calculated as:

$$\begin{aligned} \text{SoS} \approx & \frac{1}{N} \sum_k \sum_l \text{Var}[\cos(k\phi_i) \cos(l\psi_i)] \\ & + \frac{1}{N} \sum_k \sum_l \text{Var}[\cos(k\phi_i) \sin(l\psi_i)] \\ & + \frac{1}{N} \sum_k \sum_l \text{Var}[\sin(k\phi_i) \cos(l\psi_i)] \\ & + \frac{1}{N} \sum_k \sum_l \text{Var}[\sin(k\phi_i) \sin(l\psi_i)]. \end{aligned} \quad (7)$$

Here the Variances (Var) of trigonometric moment for the N entries in the ϕ, ψ -lists are used to calculate the expected uncertainty in the average trigonometric moments according to the central limit theorem.

Results

J-couplings

The heteronuclear coupling ${}^3J^{H^{\alpha}N^{i+1}}$, which reports on the ψ -angle, was measured from the E.COSY pattern observed in the NOESY spectrum from ^{15}N -enriched peptide. It proved to be very beneficial to collapse the splittings of the α -proton resonances due to couplings with the amide and β -protons by the homodecoupling element described in the materials and methods section. Thereby, the line widths were reduced to approximately 5 Hz, making the determination of the peak position more precise as well as reducing overlap. The overlap was further reduced by applying the in-phase/anti-phase technique (IPAP), where two subspectra are obtained with 50% of the peaks in each subspectrum. An overlay of a small part of the two IPAP-NOESY subspectra, where the homo decoupling element was employed is shown in Fig. 3. The J-couplings between α -proton and the backbone ^{15}N of the succeeding residue are measured from the displacement of the H^{α} -resonance in the two subspectra. The heteronuclear coupling ${}^3J^{H^{\alpha}N^{i+1}}$ which reports on the ψ angle, could be measured for 15 residues of motilin polypeptide from the IPAP-NOESY. The precision of the measurements was estimated from the root mean square deviation (rmsd) between the IPAP-NOESY measurements and a non-IPAP predecessor. The RMSD of 0.044 Hz indicates that the variations along the

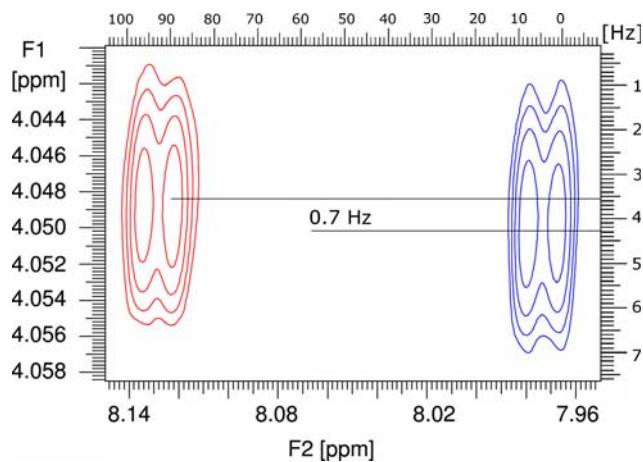


Fig. 3 Overlay of the same parts of the two subspectra (red and blue) from the IPAP-NOESY spectrum, employing homodecoupling in the H^{α} -dimension (F1). The E.COSY principle was utilized to measure the ${}^3J^{H^{\alpha}N^{i+1}}$ couplings. The difference between the resonance frequencies in the H^{α} -dimension of the two sequential cross-peaks corresponds to the coupling between the α proton and the ^{15}N atom of the succeeding residue. The difference between the doublet peaks in the amide proton dimension (F2) corresponds to the N-H coupling, which is 92 Hz

sequence must be caused by other factors than random noise.

The ${}^3J^{H^{\alpha}H^{\gamma}}$ couplings were measured from the splittings in one-dimensional traces of TROSY-HSQC (Pervushin et al. 1997). Signals from the backbone amide protons of residues which were overlapping in the TROSY-HSQC spectrum could be resolved in a homo-decoupled TOCSY experiment. The correlation coefficient between the couplings extracted from the TOCSY and TROSY is 0.99. The rmsd between the two data-sets is 0.12 Hz, indicating an uncertainty of 0.083 Hz in the average of the two.

The resonances from the amide protons of both glycine residues display triplet like patterns, because of couplings to the two H^{α} spins. The average of the two couplings was estimated from the outer components of the triplet. For 14 residues, both the homo and the heteronuclear couplings were measured successfully. For six of the residues, only one of the couplings could be measured. The measured J-couplings are compared to the predictions from the SMT-model in Fig. 4.

There is a noticeable offset between the predictions and the measurements, but locally there is significant correlation. The statistical verification of the local correlation is

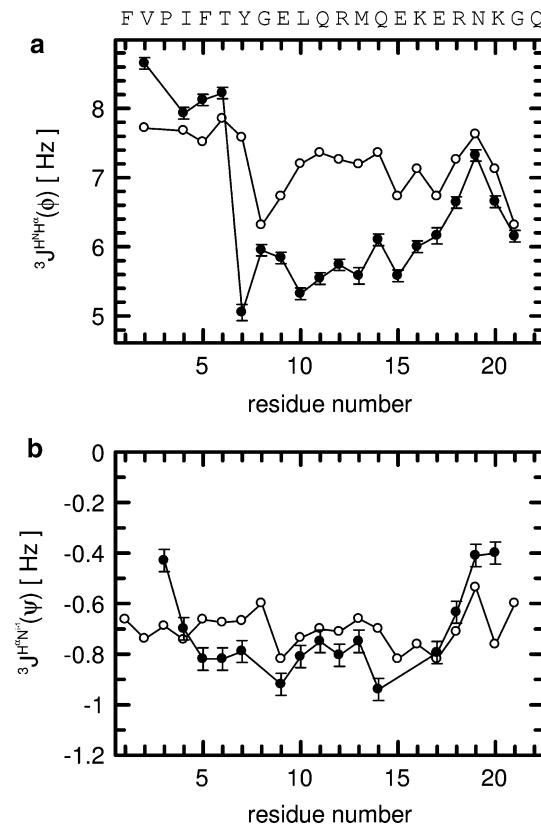


Fig. 4 Comparison between experimentally determined J-couplings (filled symbols) and the predictions from the SMT-model (open symbols)

obtained by considering sequential neighbors. The difference between the J-coupling observed for one residue and the J-coupling observed for the neighbor towards the C-terminus can be either positive or negative. For the homonuclear coupling, the SMT-model predicts the sign of the difference between sequential neighbors correctly for 12 pairs out of 18. For the heteronuclear coupling, the SMT-model correctly predicts the sign of the difference for 11 pairs out of 15. The SMT-model predicts the sign of the difference correct for 23 out of 32 pairs. In the absence of predictive power, the success for the sign prediction would follow a binomial distribution where the chance for 23 or more successes out of 32 is less than a percent. The SMT-model appears to have a certain predictive power although the model is not capable of predicting the couplings to within errors. In the central part of the sequence, the SMT-model predicts systematically to high homonuclear couplings (Fig. 4a). The smaller experimental couplings are consistent with a higher population of ϕ -angles corresponding to smaller couplings than stipulated by the SMT-model. Such small homonuclear couplings are found in α -helices where the ϕ -angle is close to -60° . The experimentally determined heteronuclear couplings show more negative numerical values in the central part than predicted by the SMT-model (Fig. 4b). Consequently, ψ -angles corresponding to the more negative couplings are populated more than stipulated by the SMT-model. Such couplings correspond to negative ψ -angles as found in α -helices. The maximum entropy approach provides a way to modulate the SMT-model to obtain updated dihedral angle distributions which are consistent with the experimental J-couplings.

Reconstruction of ϕ, ψ -distribution functions

The posterior, $\rho(\phi, \psi)$, for each residue was calculated from the experimental data according to Eq. 3. Residues where only one of the coupling constants is available were treated by setting the Lagrange multiplicator associated with the other coupling to zero. Figure 5 shows the posteriors for the 20 analyzed residues, out of the 22 residues in the motilin polypeptide.

The population is assigned to the allowed Ramachandran regions in the posteriors. The very good Ramachandran statistics is a consequence of the use of the SMT-model to formulate a prior. If the information from the SMT-model is suppressed by using a flat prior more than 60% of the population is assigned to the non-allowed regions according to the definitions of Lovell et al. (2003).

Structural trends can be seen directly from posteriors shown in Fig. 5. The N-terminal residues show mainly extended conformations where V2 has a high β -strand propensity with some tendency toward PPII conformation,

P3 and I4 has high PPII secondary structure. Residues F5, T6 both show increased population of negative ψ angles. The central part (Y7 to E17), with the exception of G8, is a nascent α -helix in equilibrium with PPII conformation, and the C-terminal (R18, N19, K20, G21) shows a high probability of having left-handed helix structure.

Circular dichroism

The CD-spectrum of motilin (Fig. 6) under the same conditions display two negative features around 205 nm and around 222 nm and a positive band below 196 nm.

The CD-spectrum of motilin is different from the CD-spectrum of so-called random-coil peptides. The positions of the minima and the maximum indicate the presence of α -helical secondary structure in motilin. The intensity is too low for a fully developed α -helix. Any fully developed secondary structure can be excluded. The CD-spectrum is very similar to a linear combination of the standard CD-spectra α -helix (32%), left-handed 3_1 helix (31%) and random-coil (37%), with a normalization coefficient of 0.25 to account for the lower intensity of the experimental CD-spectrum. This interpretation of the CD-spectrum is compatible with the posterior distribution functions where the central part of the peptide appears to form a nascent α -helix in equilibrium with a left-handed 3_1 helix, and parts of the peptide appears similar to a so-called random-coil.

Segmental entropies

Joint ϕ and ψ distributions allow for many types of analysis. A potential of mean force may be formulated and under the assumption that the conformation of the residues fluctuates independently it is possible to estimate the contribution to conformational entropy from the segment, i.e., the segmental entropy, S_p . The segmental entropy was calculated using Eq. 8

$$\Delta S_p = -R \iint \rho(\phi, \psi) \ln \frac{\rho(\phi, \psi)}{\langle \rho(\phi, \psi) \rangle} d\phi d\psi. \quad (8)$$

Here ΔS_p is the entropy difference relative to a flat distribution and R is the general gas constant. By taking the natural logarithm of the ratio between the probability density and the ensemble average of the probability density, $\langle \rho(\phi, \psi) \rangle$, the invariance problem is under control, i.e., the same segmental entropy results independent of the choice of units for the angles. The results are compared to ΔS_p -values for the priors from the coil-library, calculated using the same formula, in Fig. 7.

The segmental entropy for the proline residue was found to be lowest of all residues, whereas the two glycine residues display higher segmental entropies. This relation



Fig. 5 Ramachandran plots of the posteriors of 20 out of 22 residue of motilin polypeptide. The horizontal and the vertical axis represent the ϕ and ψ angles, respectively, each axis ranges from -180° to $+180^\circ$. Each of the 10 colors represents 10% of the population

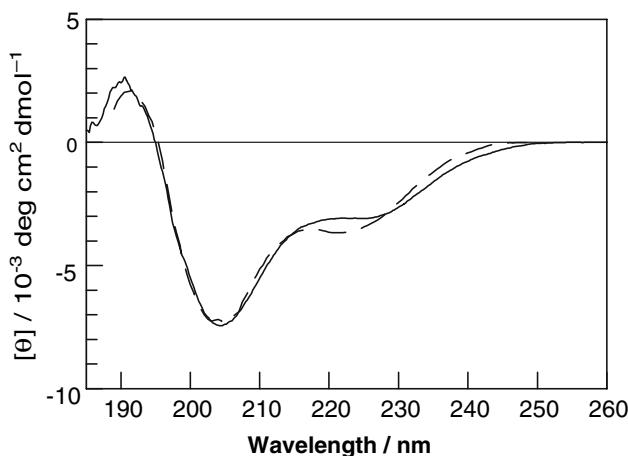


Fig. 6 The experimental CD-spectrum (solid line) and the linear combination between 32% α -helix, 37% random-coil and 31% left-handed 3₁-helix (dashed line)

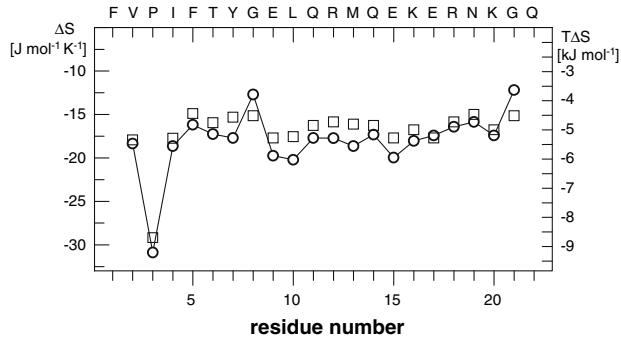


Fig. 7 Segmental entropies relative to a flat distribution estimated either from the posteriors from the ME-analysis of the J-couplings (circles) or from the priors from the coil-library (squares)

between the segmental entropy and different residue types in the posteriors is inherited from the priors. In the central part of the peptide, which appears to be a nascent helix, the segmental entropies are systematically overestimated from the SMT model. For the stretch of 10 residues with helical propensities, the entropic contribution to free energy from the freedom in the dihedral angles is underestimated by approximately 8 kJ if estimated from SMT model.

In addition to the approach to segmental entropies suggested here, a different NMR-approach to segmental entropies exists. In that approach (Yang and Kay 1996) NMR relaxation rates of the ¹⁵N-nuclei along the backbone are measured and interpreted according to the Lipari-Szabo model (Lipari and Szabo 1982), which provides squared generalized order-parameters. In the subsequent step, a simplistic parametric model is used that connects the order-parameter with an orientational distribution function of the NH-bond in some molecular frame. Segmental entropies are calculated from the orientational distribution functions for the NH-bond vectors along the backbone. In order to compare the segmental entropies

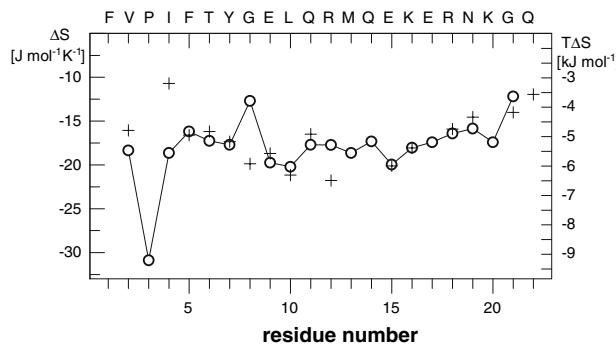


Fig. 8 The graph shows a comparison between segmental entropies estimated from the ME-analysis of the J-couplings (circles) and from ¹⁵N-relaxation data (crosses). The entropic contribution to the free-energy from each residue can be traced from the vertical axis on the right side

from the maximum entropy analysis of the J-couplings to the segmental entropies estimated from order-parameters the R_1 , R_2 and heteronuclear NOE of the backbone ¹⁵N-nuclei were measured. The Lipari-Szabo model was fitted to the ¹⁵N-relaxation data using a rotational correlation time of 2.58 ns, which was the median value of the long correlation time when individual Lipari-Szabo models were fitted to each site. The squared generalized order-parameters are around 0.55 for the central residues and decrease towards both the N and C-termini. The order-parameters were interpreted using the wobbling in a cone-model for which the entropy difference relative to a uniform orientational distribution function is given by Eq. 9

$$\Delta S_p^{\text{relax}} = R \ln \left[\pi \left(3 - \sqrt{1 + 8 S_{\text{LZ}}} \right) \right] - R \ln [4\pi]. \quad (9)$$

Here S_{LZ} is the Lipari-Szabo type generalized order-parameter for the NH-bond (i.e., not S^2_{LZ}). Due to spectral overlap (M13, Q14, E17 and K20) or absence of HSQC cross-peak (F1 and P3) relaxation rates, order-parameters and $\Delta S_p^{\text{relax}}$ were successfully estimated for 16 out of the 22 residues in motilin. The general agreement (Fig. 8) between the two estimates is good. In both series, the central residues display segmental entropies around -20 J K mol^{-1} and the segmental entropies increase for both the termini.

Discussion

General

The reconstruction of the joint ϕ and ψ distribution functions from limited data is an intrinsically ill-posed problem with more free parameters than observations. In the current work the problem is regularized by supplying an additional criterion. All possible dihedral angle distributions are

ranked such that more unexpected distributions are given a lower rank before/without considering the experimental data. When the experimental data becomes available, the distribution that has the highest ranking, i.e., the least unexpected, of the distributions that are consistent with the experimental data is selected and put forward as the posterior distribution, which is a new leading hypothesis. The Shannon entropy which is defined relative to an a priori assignment of the probability for different combinations of ϕ and ψ , i.e., the prior, is used as the tool to rank the distributions. Mathematically the selection of the ME-distribution is equivalent to modulating the prior by an exponential function with the scaled Karplus relations in the exponent (Eq. 4). As the modulating functions are smooth and flat the positions and shapes of the humps are not changed very much while the probability of each part is adjusted. By comparing the posteriors (Fig. 5) and the priors (Fig. 1S) it obvious that the ME-approach mainly modulates the relative populations of the local maxima.

The SMT-model for the dihedral angle distributions in so-called random-coil peptides is used to formulate the prior. If the expectation expressed by the prior is similar to the true distribution only a few observations should be needed to obtain meaningful posterior distribution.

In the current work a joint ϕ, ψ -distribution function for each residue in the motilin peptide have been assigned by the ME-approach with an informative prior. The posteriors assign the population to sterically possible conformations and describe the motilin structural ensemble better than the SMT-model. The J-couplings, which were used as constraints, are in perfect agreement with the posteriors, while the SMT-model fails to explain the J-couplings.

The observed CD-spectrum appears as a linear combination of the literature α -helix and “random-coil” or left-handed 3_1 -helical CD-spectra. Such a linear combination is consistent with the posteriors, while if motilin would obey the SMT-model the CD-spectrum would resemble the literature “random-coil” spectrum. Furthermore, the segmental entropies calculated from the posteriors give a consistent picture together with the segmental entropies calculated from the ^{15}N -relaxation rates. Taken together, the J-couplings, CD-spectrum, segmental entropies and the Ramachandran statistics indicate that the posteriors are meaningful descriptions of the dihedral angle preferences in motilin.

Why Shannon entropy?

Equation 1 cannot be formally inverted. The problem of finding $\rho(\theta)$ from the J-couplings is an ill-posed problem. In many cases it is meaningful to suggest a leading hypothesis for the distribution. Such a leading hypothesis has a value since it is testable and describes a current state

of knowledge. The leading hypothesis about $\rho(\phi, \psi)$ should be consistent with the experimental observations. In many cases, additional information exists that enables an experienced investigator to classify certain distributions as less likely to be true even if they are consistent with the experimental observations. For dihedral angle distributions an experienced investigator would claim that any distribution which assigns significant population to eclipsed conformations appears unrealistic. The Shannon entropy possesses a number of desired properties that make it an attractive tool to rank the dihedral angle distributions as more or less expected in a flexible peptide. The Shannon entropy will be maximal, i.e., zero, for the a priori expectation represented by the prior. It will be lower for all other distributions and decreases further if the population is assigned to unexpected dihedral angle combinations. The use of the Shannon entropy is also attractive because it results in a closed form expression which is very rapid to evaluate numerically. Finally, Shannon entropy has the property that it will give equivalent results if two independent dihedral angle distributions are evaluated jointly or separately. Thus, the approach does not introduce any non-warranted correlation between independent entities.

The priors

The Shannon entropy is defined relative to a prior to avoid the invariance problem. The prior should express the investigators state of knowledge before considering the experimental observations to be analyzed. Different investigators may have different knowledge and assign different priors. The suggested ME-approach is therefore subjective. An infinitely insightful investigator would be able to suggest the true distribution and use it as the prior. Since the true distribution must be consistent with all the experimental observations, the experimental data would never exclude the true distribution. If the true distribution was used as the prior, then, the true distribution would show the highest Shannon entropy. The true distribution would consequently be put forward as the ME-solution for the infinitely insightful investigator. In classical ME-approaches a flat prior is employed, which corresponds to a situation of complete lack of insight before considering the experiments. For the joint ϕ, ψ -distribution in flexible peptides such a flat prior is naïve and ignorant since background information exist. Further evidence that flat priors are no good is that the posteriors evaluated from flat priors are unrealistic and assigns more than 60% of the population to sterically disallowed conformations. The poor result when the background information is suppressed highlight the importance of incorporating all pieces of information, experimental observations as well as background information, into the analysis. In general, better

priors that are closer to the true distributions will yield better posteriors.

The SMT-model

In the current work, the priors are derived from the SMT-model (Swindells et al. 1995). In this section, the strengths and weaknesses of the SMT-model are discussed. The random coil model that is used to describe the unstructured polypeptides (SMT-model, (Swindell et al. 1995)), has two intrinsic properties. First, it assumes that the residues that are not within secondary structure elements in the protein database are able to represent the random-coil state conformation. Second, it assumes that the secondary structure of the amino acid residues is completely independent on its neighbors.

Coil-library models similar to the current implementation of the SMT-model have previously been shown to have predictive power for residue type specific average ${}^3J^{H^N H^Z}$ and for RDCs in unstructured peptides and denatured proteins. For Motilin, the SMT-model has predictive power for the J-coupling difference between sequential neighbors. This can be rationalized as follows. The intrinsic preferences for each residue as well as the environment influence the dihedral angle distributions. If sequential residues reside in similar environments, the influence of the environment will approximately cancel. Thus, the difference in intrinsic preference for sequential neighbors may dominate the difference in the dihedral angle distributions. Although the SMT-model has predictive power, it does not provide quantitatively correct predicted J-couplings. The model needs improvement but the predictive power shows that it is a good starting point.

There are clear trends along the motilin peptide sequence with stretches of residues behaving similarly. Such trends are observed both in the J-couplings (Fig. 4) and in the ${}^{15}N$ -relaxation rates (supporting information). It is obvious from Fig. 4 that the error in the prediction from the SMT-model of the coupling constant for one residue is similar to the error in the prediction for the succeeding residue. Consequently, the observations from one residue have predictive power for the observations from neighboring residues. The predictive power means that the neighbors are not independent as was stipulated in the SMT-model. The observation that the J-couplings in sequential neighbors are dependent shows that the conformational preferences of one residue influence the conformational preferences of the neighboring residues. The observation that the conformation of sequential neighbors are dependent is in agreement with a recent coil-library (Jha et al. 2005b) analysis where it was found that also in coil-regions in globular proteins the conformations of sequential residues are correlated. However, there is a

problem when analyzing the influence from the conformation of the neighbor in a coil-library. Since regular secondary structures are excluded, the selection procedure may introduce a bias against consecutive residues in similar conformation. In addition to the influence from the conformation of the neighbor observed in the current work and in the coil-library analysis of Jha et al. there is a documented influence on the conformational preferences from the chemical nature of the neighboring residues (Jha et al. 2005b). Those correlations indicate that it should be possible to assign better priors by considering the chemical nature of neighboring residues and J-couplings observed in neighboring residues.

As the structural preferences of flexible peptides are further explored it will in the future be possible to use more insightful models to assign better priors. One example of such more insightful models that may provide better priors has recently been suggested by Jha et al. In that work, the influence from the conformation and chemical identity of the neighboring residues are characterized in a coil-library. From the observed characteristics a model where those effects are accounted for was suggested. The more sophisticated model appear to explain the observed RDCs in urea denatured myoglobin better than the SMT-model. This more sophisticated model challenges the independence of the residues in the traditional “random-coil” models.

Motilin

The N-terminal part of the motilin peptide is responsible for receptor binding (Macielag et al. 1992; Peeters et al. 1992). In three different solvent mixtures, 30% hexa-fluoro-isopropanol (Kahn et al. 1990) (Edmondson et al. 1991) in the presence of sodium dodecyl sulfate micelles (Jarvet et al. 1997) and isotropic bicelles (Andersson and Måler 2002, 2003), there are two similar turns in the vicinity of the proline residue in position 3. The inverse γ -turn found at P3 in the three solvent mixtures is not populated largely in water, according to the ME-analysis of the J-couplings. There is no population of ψ -angles between 50° and 70° , as found in inverse γ -turns for P3 in Fig. 5. The only experimental observation related to this angle, ${}^3J^{H^Z N^{+1}}$, is -0.43 Hz. For ψ -angles between 50° and 70° couplings between -0.24 and -0.15 Hz are expected, which can be excluded with the current precision. There is an indication that the turn, resembling a β -turn type I, which extends over residues P3 to T6, as found for motilin in the presence of isotropic bicelles, may be also populated to some extent in aqueous solution. This is evident since the ϕ, ψ -combinations found in the structure in bicelles for residues I4 ($\phi = -55^\circ \pm 19^\circ, \psi = -58^\circ \pm 8^\circ$) and F5 ($\phi = -122^\circ \pm 6^\circ, \psi = -30^\circ \pm 7^\circ$) are both populated to some ex-

tent according to the ME-analysis of the J-couplings in aqueous solution, cf. Fig. 5. The central α -helix, which is not necessary for receptor interaction, observed in the three solvent mixtures is significantly populated also in water.

Circular dichroism

The CD-spectrum of motilin is different from literature CD-spectra for fully developed secondary structures. The positions of the positive and negative bands display similarity to the literature α -helical spectrum. The shape may indicate the presence of helical structure, consistent with the posteriors. The experimental CD-spectrum has low intensity compared to literature CD-spectra for fully developed secondary structures. (Shi et al. 2002) studied the capped GGXGG pentapeptides by CD. Also the pentapeptides display low CD-intensity compared to the expected CD in the literature “random-coil” (Reid 2000). We speculate that the basis CD-spectra from globular proteins and from fully developed secondary structures may be less representative for flexible peptides.

From the posteriors presented in Fig. 5 one would expect that the CD-spectrum of motilin may resemble a linear combination of the CD-spectra for left-handed 3_1 -helix and α -helix and random-coil. Indeed such a linear combination fit the experimental CD-spectrum very closely. However, the coefficients do not sum up to unity.

Segmental entropy

The segmental entropies evaluated from the J-couplings employing the suggested ME-approach show similarity to segmental entropies (i.e., entropy per residue) evaluated here from relaxation data according to Yang and Kay (1996). There is a trend with decreased segmental entropies in the central nascent helix and increased segmental entropies toward the ends. The similarity between the two estimates of the segmental entropies is noteworthy, since fundamentally different distribution functions are analyzed. In the ME-analysis of J-couplings, the joint distributions of ϕ and ψ are analyzed and distributions of orientations of the NH-bond vector in some molecular frame are analyzed in the relaxation based approach. A motional mode may influence the two different distribution functions differently and neither J-couplings nor NMR-relaxation is expected to capture all modes. Experimental evidence exists which indicates that a joint analysis of relaxation rates for carbonyl carbons together with ^{15}N -relaxation may capture modes which escape detection when ^{15}N -relaxation only is analyzed (Wang et al. 2003, 2005). In addition, the J-couplings and ^{15}N -relaxation data are sensitive to conformational fluctuations occurring on different time-scales. While fluctuations occurring on the pico to nano-

second time-scale govern ^{15}N -relaxation, the time-scale window is much wider for J-couplings which are sensitive also to slower fluctuations down to 100 milli-seconds. Thus, exact equivalence between ΔS_p^J and $\Delta S_p^{\text{relax}}$ is not expected.

Some of the assumptions underlying the relaxation-based approach may be questioned. First, it is assumed that generalized order-parameters are at all possible to estimate from spin-relaxation data from flexible peptides. The estimation of such order-parameters is known to be very sensitive to whether models with two or three Lorenzian components in the spectral density function are used (Damberg et al. 2002). In motilin there are residues for which the correlation times and order-parameters are nearly confounded, which further stress the issue of the validity of the dynamic model. From a theoretical point of view, it is very difficult or impossible (Prompers and Brüschweiler 2002), to define a molecular frame in which the order-parameter is defined since a common overall structure is missing in flexible peptides. It should also noted that the partitioning into components from internal motions and globular tumbling depends on the assumption that the two processes are independent, which is not generally true for flexible peptides.

Second, the relation between entropy and the order-parameters depends on non-warranted assumptions regarding the shape of the orientational distribution function. To be able to link the order-parameter and the entropy a simplistic parametric function for the orientational distribution function is suggested, for example square-well, wobbling on the surface of a cone or wobbling in a cone etc. (Yang and Kay 1996). Those models have only one parameter which determines the spread in the distribution. For each value of spread parameter it is possible to calculate both the order-parameter and the entropy of the distribution. Hence, there is a link between the order-parameter and the segmental entropy via the spread parameter. The problem is that the relation is different for different models. For several models it was found that the entropy vs. order-parameter graphs are very similar and related by an offset (Yang and Kay 1996). However, the comparison included mainly symmetric unimodal distributions. For multimodal distributions a change in the order-parameter would correspond to a smaller entropy change compared to unimodal distributions. Thus the $\Delta S_p^{\text{relax}}$ values depend on the implicit assumption of unimodal orientational distributions with the same parametric shape. There is no evidence that the orientational distribution functions behave that way.

In light of the potential problems with the ^{15}N -relaxation based approach, the segmental entropy analysis of J-couplings has a great value as an additional independent measure.

In addition to the trend with lower segmental entropies in the central part and increased segmental entropies towards the termini, the ME-analysis of the J-couplings indicate that glycine residues have higher segmental entropies and the proline residue has decreased segmental entropy. The increased entropy of the glycine residues can be attributed to the increased conformational freedom due to the absence of steric clashes with a side chain and the decreased segmental entropy of the proline is attributed to the decreased conformational freedom due to the covalent ring structure of the side-chain. The pattern indicates that both the position within the molecule and residue type has an impact on the segmental entropy. Also in the $\Delta S_p^{\text{relax}}$ data, the trend with decreased segmental entropies in the central part and increased segmental entropies towards the termini is observed. The increased segmental entropy for G8 is not apparent, for G21 $\Delta S_p^{\text{relax}}$ is high, but that can equally well be ascribed to the position close to the C-terminus. For P3 no relaxation data were obtained, and consequently no $\Delta S_p^{\text{relax}}$ value was obtained. In contrast to the ΔS_p^J -data, the residue type variations are not observed in the ΔS_p^J -data. This discrepancy is either or both due to a lack of J-coupling constraint, in particular for G8, resulting in an overestimation of ΔS_p^J , and/or that rotation around the dihedral angles also in adjacent residues govern the ^{15}N -relaxation.

The observation that both for the J-couplings and for the relaxation rates a measure from one residue has predictive power for adjacent residues shows that the behavior of the residues is not independent. Therefore, the partitioning of the conformational entropy into contributions from different segments is not rigorously valid. The assumed independence, when in fact the segments are dependent, will lead to an overestimation of the total conformational entropy.

The difference in the segmental entropies between the priors from the coil-library and the posteriors correspond to a contribution to the free energy difference of 8 kJ mole⁻¹ for the central nascent helix. The difference indicates that the coil-library may overestimate the conformational entropy of an unstructured peptide.

The agreement between the segmental entropies estimated from both the J-couplings and the relaxation data suggests that J-couplings and ^{15}N -relaxation rates are connected via a common factor, i.e., segmental entropy, in flexible peptides. The similarity indicates that the motional modes that contribute to ^{15}N -relaxation are closely related to the fluctuations in the most nearby dihedral angles. The existence of a common factor indicates that by some, not yet known, joint analysis of both J-couplings and relaxation rates it may become possible to understand the dynamic behavior of flexible peptides beyond modeling the polypeptide chain as a “random-coil” homo-polymer. In such

models an effective persistence length and an intrinsic relaxation rate are the adjustable parameters, and deviations from the simple model are accounted for by introducing ad hoc terms for branching, disulfide bonds and hydrophobic clusters etc (Klein-Seetharaman et al. 2002).

Conclusion

The suggested maximum entropy analysis of J-couplings accounts for the ensemble character of a flexible polypeptide and incorporates the crucial background information provided by the coil-library analysis. The results are detailed and useful descriptions of the conformational preferences for each residue in terms of to what extent different combinations of ϕ and ψ are populated. The resulting distribution functions are consistent with the observed J-couplings and explain the CD-spectrum of motilin. The detailed and useful description allows for the estimation of the contribution to the conformational entropy from the conformational freedom in the backbone of each residue. The motilin peptide adopts preferentially extended conformations in the receptor binding N-terminal part and form a nascent α -helix in equilibrium with a left-handed 3_1 -helix in the central part.

Acknowledgements We thank Britt-Marie Olsson for help with the peptide purification, Joshua Hicks for his valuable comments on the manuscript and Astrid Gräslund for continuous support. This work was supported by the Estonian Science Foundation for grant no. 5138 and The Wellcome Trust Grant no. 060781 and the Carl-Trygger foundation and the Magn.Bergvall foundation (PD).

Appendix

In this appendix the derivation of Eq. 4 is given. The goal here is to find a closed form expression for the normalized distribution function, $\rho(\phi, \psi)$ which maximizes the Shannon entropy while being consistent with the observed J-couplings. Thus, it is maximization under equality constraints, which is solved by Lagrange’s method. The constraint equations which $\rho(\phi, \psi)$ must satisfy are:

$$\begin{aligned} \int J^{H^z H^x}(\phi) \int \rho(\phi, \psi) d\psi d\phi - J_{\text{obs}}^{H^z H^x} &= 0 \\ \int J^{H^z N^{i+1}}(\psi) \int \rho(\phi, \psi) d\phi d\psi - J_{\text{obs}}^{H^z N^{i+1}} &= 0 \\ \iint \rho(\phi, \psi) d\phi d\psi - 1 &= 0. \end{aligned}$$

The objective function to be maximized is the Shannon entropy

$$S(\rho(\phi, \psi)) = - \iint \rho(\phi, \psi) \frac{\rho(\phi, \psi)}{P(\phi, \psi)} d\phi d\psi.$$

With this objective function and the constraint equations the Lagrangian, L, becomes:

$$\begin{aligned} L(\rho(\phi, \psi)) &= S(\rho(\phi, \psi)) + \lambda'_0 \cdot \left(\iint \rho(\phi, \psi) d\phi d\psi - 1 \right) \\ &\quad + \lambda_1 \cdot \left(\int J^{H^N H^x}(\phi) \int \rho(\phi, \psi) d\psi d\phi - J_{obs}^{H^N H^x} \right) \\ &\quad + \lambda_2 \cdot \left(\int J^{H^x N^{i+1}}(\psi) \int \rho(\phi, \psi) d\phi d\psi - J_{obs}^{H^x N^{i+1}} \right). \end{aligned}$$

Kuhn-Tucker theorem stipulates that the derivative of the Lagrangian with respect to the probability density at any combination of ϕ and ψ must be zero when the objective function is maximized subject to the constraints. Thus, the derivative of the Lagrangian is evaluated and set to zero.

$$\begin{aligned} \frac{\partial}{\partial \rho(\phi, \psi)} (L(\rho(\phi, \psi))) &= \frac{\partial}{\partial \rho(\phi, \psi)} (S(\rho(\phi, \psi))) \\ &\quad + \frac{\partial}{\partial \rho(\phi, \psi)} \left(\lambda'_0 \cdot \left(\iint \rho(\phi, \psi) d\phi d\psi - 1 \right) \right) \\ &\quad + \frac{\partial}{\partial \rho(\phi, \psi)} \left(\lambda_1 \cdot \left(\int J^{H^N H^x}(\phi) \int \rho(\phi, \psi) d\psi d\phi - J_{obs}^{H^N H^x} \right) \right) \\ &\quad + \frac{\partial}{\partial \rho(\phi, \psi)} \left(\lambda_2 \cdot \left(\int J^{H^x N^{i+1}}(\psi) \int \rho(\phi, \psi) d\phi d\psi - J_{obs}^{H^x N^{i+1}} \right) \right) \\ &= -1 - \log \frac{\rho(\phi, \psi)}{P(\phi, \psi)} + \lambda'_0 + \lambda_1 \cdot J^{H^N H^x}(\phi) \\ &\quad + \lambda_2 \cdot J^{H^x N^{i+1}}(\psi) \\ &= 0. \end{aligned}$$

As the derivative of the Lagrangian is zero an equation to solve for $\rho(\phi, \psi)$ is arrived at. Solving for $\rho(\phi, \psi)$ gives:

$$\begin{aligned} \rho(\phi, \psi) &= P(\phi, \psi) \cdot \\ &\quad \exp \left(\lambda'_0 - 1 + \lambda_1 \cdot J^{H^N H^x}(\phi) + \lambda_2 \cdot J^{H^x N^{i+1}}(\psi) \right). \end{aligned}$$

Substituting $\lambda_0 = \lambda'_0 - 1$ gives

$$\begin{aligned} \rho(\phi, \psi) &= P(\phi, \psi) \cdot \\ &\quad \exp \left(\lambda_0 + \lambda_1 \cdot J^{H^N H^x}(\phi) + \lambda_2 \cdot J^{H^x N^{i+1}}(\psi) \right). \end{aligned}$$

This is exactly Eq. 4.

References

- Adams RA (1990) Calculus: a complete course. Addison-Wesley, Canada
- Andersson A, Mäler L (2002) NMR solution structure and dynamics of motilin in isotropic phospholipid bicellar solution. *J Biomol NMR* 24:103–112
- Andersson A, Mäler L (2003) Motilin-bicelle interactions: membrane position and translational diffusion. *FEBS Lett* 545:139–143
- Avbelj F, Baldwin L (2003) Origin of neighboring residue effect on peptide backbone conformation. *Proc Natl Acad Sci USA* 100:5742–5747
- Bernadó P, Bertoncini CW, Griesinger C, Zweckstetter M, Blackledge M (2005) Defining long-range order and local disorder in native α -Synuclein using residual dipolar couplings. *J Am Chem Soc* 127:17968–17969
- Bochicchio B, Tamburro AM (2002) Polyproline II Structure in Proteins: Identification by Chiroptical Spectroscopies, Stability, and Functions. *Chirality* 14:782–792
- Boesch C, Bundi A, Oppiger M, Wüthrich K (1978) 1H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur J Biochem* 91:209–214
- Buckler DR, Haas E, Scheraga HA (1995) Analysis of the Structure of Ribonuclease A in Native and Partially Denatured States by Time-Resolved Nonradiative Dynamic Excitation Energy Transfer between Site-Specific Extrinsic Probes. *Biochemistry* 34:15965–15978
- Damberg P, Jarvet J, Allard P, Mets Ü, Rigler R, Gräslund A (2002) $^{13}C-^1H$ NMR relaxation and fluorescence anisotropy decay study of tyrosine dynamics in Motilin. *Biophys J* 83:2812–2825
- Damberg P, Jarvet J, Gräslund A (2005) Limited Variations in ^{15}N CSA Magnitudes and Orientations in Ubiquitin Are Revealed by Joint Analysis of Longitudinal and Transverse NMR Relaxation. *J Am Chem Soc* 127:1995–2005
- Daniels AJ, Williams RJP, Wright PE (1978) The character of the stored molecules in chromaffin granules of the adrenal medulla: a nuclear magnetic resonance study. *Neuroscience* 3:573–585
- Danielsson J, Jarvet J, Damberg P, Gräslund A (2002) Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer $\text{A}\beta$ (1–40) peptide. Determination of hydrodynamic radii of random coil peptides of varying length. *Magn Reson Chem* 40:S89–S97
- Dukor RK, Keiderling TA (1991) Reassessment of the random coil conformation. Vibrational CD study of proline oligopeptides and related polypeptides. *Biopolymers* 31:1747–1761
- Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41:6573–6582
- Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nature Rev* 6:197–208
- Edmondson S, Khan N, Shriver J, Zdunek J, Gräslund A (1991) The solution structure of motilin from NMR distance constraints, distance geometry, molecular dynamics, and an iterative full relaxation matrix refinement. *Biochemistry* 30:11271–11279
- Eker F, Griebenow K, Cao X, Nafie LA, Schweitzer-Stennet R (2004) Preferred peptide backbone conformations in the unfolded state revealed by the structure analysis of alanine-based (AXA) tripeptides in aqueous solution. *Proc Natl Acad Sci USA* 101:10054–10059
- Farrow NA, Muhandiram R, Singer AU, Pascale SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and a phosphopeptide-complexed SRC Homology-2 domain studied by N-15 NMR relaxation. *Biochemistry* 33:5984–6003
- Fitzkee NC, Fleming PJ, Rose GD (2005) The protein coil library: a structural database of nonhelix, nonstrand fragments derived from the PDB. *Proteins: Struct, Funct, Bioinform* 58:852–854
- Griffiths-Jones SR, Sharman GJ, Maynard AJ, Searle MS (1998) Modulation of intrinsic ϕ, ψ propensities of amino acids by

- neighbouring residues in the coil regions of protein structures: NMR analysis and dissection of a β -hairpin peptide. *J Mol Biol* 284:1597–1609
- Jarvet J, Zdunek J, Damberg P, Gräslund A (1997) Three-dimensional structure and position of porcine motilin in sodium dodecyl sulfate micelles determined by ^1H NMR. *Biochemistry* 36:8153–8163
- Jaynes ET (1963) Information theory and statistical mechanics. In Ford K (ed) Statistical physics. Benjamin, New York, pp 181–182
- Jha AK, Colubri A, Zaman MH, Koide S, Sosnick TR, Freed KF (2005a) Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* 44:9691–9702
- Jha AK, Colubri A, Freed KF, Sosnick TR (2005b) Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc Natl Acad Sci USA* 102:13099–13104
- Kahn N, Gräslund A, Ehrenberg A, Shriver J (1990) Sequence-specific ^1H NMR assignments and secondary structure of porcine Motilin. *Biochemistry* 29:5743–5751
- Karplus M (1959) Contact electron-spin interactions of nuclear magnetic moments. *J Chem Phys* 30:11–15
- Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. *Science* 295:1719–1722
- Kohno T, Kusunoki H, Sato K, Wakamatsu K (1998) A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: an application to the production of mastoparan-X uniformly enriched with ^{15}N and $^{15}\text{N}/^{13}\text{C}$. *J Biomol NMR* 12:109–121
- Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Protein disorder prediction: implications for structural proteomics. *Structure* 11:1453–1459
- Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* 104:4546–4559
- Lovell SC, Davis IW, Arendall III WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by $C\alpha$ geometry: ϕ, ψ and $C\beta$ deviation. *Proteins: Struct, Funct, Genet* 50:437–450
- Macielag MJ, Peeters TL, Konteatis ZD, Florance JR, Depoortere I, Lessor RA, Bare LA, Cheng Y, Galdeas A (1992) Synthesis and in vitro evaluation of [Leu13]porcine motilin fragments. *Peptides* 13:565–569
- Makowska J, Rodziewicz-Motwidlo S, Baginska K, Vila JA, Liwo A, Chmurzynski L, Scheraga HA (2006) Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc Natl Acad Sci USA* 103:1744–1749
- Mohana-Borges R, Goto NK, Kroon GJA, Dyson HJ, Wright PE (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J Mol Biol* 340:1131–1142
- Paalme T, Tiisma K, Kahru A, Vanatalu K, Vilu R (1990) Glucose-limited fed-batch cultivation of *Escherichia coli* with computer-controlled fixed growth rate. *Biotechnol Bioeng* 35:312–319
- Peeters TL, Macielag MJ, Depoortere I, Konteatis ZD, Florance JR, Lessor RA, Galdeas A (1992) D-Amino acid and alanine scans of the bioactive portion of porcine motilin. *Peptides* 13:1103–1107
- Pervushin K, Riek R, Wider G, Wüthrich KW (1997) Attenuated T_2 relaxation by mutual cancellation of dipole-dipole couplings and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94:12366–12371
- Prompers JJ, Brüschweiler PR (2002) Dynamic and Structural Analysis of Isotropically Distributed Molecular Ensembles. *Proteins: Struct, Funct, Genet* 46:177–189
- Ramachandran GN, Sasisekharan V (1968) Conformation of poly-peptides and proteins. *Adv Prot Chem* 23:283–438
- Rath A, Davidson AR, Deber CM (2005) The structure of ‘Unstructured’ regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. *Biopolymers (Peptide Science)* 80:179–185
- Reid RE (2000) Peptide and protein drug analysis. Dekker, New York
- Rowicka M, Otwinowski Z (2004) Application of maximum entropy principle to modeling torsion angle probability distribution in proteins. *AIP Conf Proc* 707:359–370
- Schweers O, Schonbrunn-Hanebeck E, Marx A, Mandelkow E (1994) Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *J Biol Chem* 269:24290–24297
- Serrano L (1995) Comparison between the ϕ distribution of the amino acids in the protein database and NMR data indicates that amino acids have various ϕ propensities in the random coil conformation. *J Mol Biol* 254:322–333
- Shi Z, Woody RW, Kallenbach NR (2002) Is polyproline II a major backbone conformation in unfolded protein? *Adv Prot Chem* 62:163–240
- Shortle DR (1996) Structural analysis of non-native states of proteins by NMR. *Curr Opin Struct Biol* 6:24–30
- Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293:487–489
- Silver MS, Joseph RI, Hoult DI (1984) Highly selective $\pi/2$ and π pulse generation. *J Magn Reson* 59:347–351
- Sklenar V, Piotto M, Leppik R, Saudek V (1993) Gradient-Tailored water suppression for $^1\text{H}-^{15}\text{N}$ HSQC Experiments optimized to retain full sensitivity. *J Magn Reson A* 102:241–245
- Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM (1996a) Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J Mol Biol* 255:494–506
- Smith LJ, Fiebig KM, Schwalbe H, Dobson CM (1996b) The concept of the random coil-Residual structure in peptides and denatured proteins. *Fold Des* 1:R95–R106
- Sosnick TR, Trewella J (1992) Denatured states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry* 31:8329–8335
- Stigler SM (1982) Thomas Bayes's Bayesian Inference. *J R Statist A* 145:250–258
- Swaminathan R, Krishnamoorthy G, Periasamy N (1994) Similarity of fluorescence lifetime distributions for single tryptophan proteins in the random coil state. *Biophys J* 67:2013–2023
- Swindell MB, MacArthur MW, Thornton JM (1995) Intrinsic ϕ, ψ propensities of amino acids derived from the coil region of known structures. *J Natur Struct Biol* 2:596–603
- Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527–533
- Tomson K, Barber J, Vanatalu K (2006) Adaptastat- a new method for optimising of bacterial growth conditions in continuous culture: Interactive substrate limitation based on dissolved oxygen measurement. *J Microbiol Meth* 64:380–390
- Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11:739–756
- Wang AC, Bax A (1995) Reparametrization of the karplus relation for $^3\text{J}(\text{H}^{\alpha}-\text{N})$ and $^3\text{J}(\text{H}-\text{C}')$ in peptides from uniformly $^{13}\text{C}/^{15}\text{N}$ -enriched human ubiquitin. *J Am Chem Soc* 117:1810–1813
- Wang AC, Bax A (1996) Determination of the backbone dihedral angles ϕ in human ubiquitin from reparametrized empirical karplus equations. *J Am Chem Soc* 118:2483–2494

- Wang T, Cai S, Zuiderweg ER (2003) Temperature dependence of anisotropic protein backbone dynamics. *J Am Chem Soc* 125:8639–8643
- Wang T, Frederick KK, Igumentova TI, Wand AJ, Zuiderweg ER (2005) Changes in calmodulin main-chain dynamics upon ligand binding revealed by cross-correlated NMR relaxation measurements. *J Am Chem Soc* 127:828–829
- Wolfram S (1991) Mathematica: a system for doing mathematics by computer. Addison-Wesley, New York
- Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J Mol Biol* 293:321–331
- Wüthrich K (1994) NMR assignments as a basis for structural characterization of denatured states of globular proteins. *Curr Opin Struct Biol* 4:93–99
- Yang D, Kay EL (1996) Contributions to Conformational Entropy Arising from Bond Vector Fluctuations Measured from NMR-Derived Order Parameters: Application to Protein Folding. *J Mol Biol* 263:369–382