Structure of human insulin-like peptide 5 and characterization of conserved hydrogen bonds and electrostatic interactions within the relaxin framework

Linda M Haugaard-Jönsson, Mohammed Akhter Hossain, Norelle L Daly, David J Craik, John D Wade, K Johan Rosengren

Structure of Human Insulin-Like Peptide 5 and Characterization of Conserved Hydrogen Bonds and Electrostatic Interactions within the Relaxin Framework

Linda M. Haugaard-Jönsson[*], Mohammed Akhter Hossain[†], Norelle L. Daly[‡], David J. Craik[‡], John D. Wade[†,§] and K. Johan Rosengren[*,¶]

[*]School of Pure and Applied Natural Sciences, University of Kalmar, SE-391 82 Kalmar, Sweden; [†]Howard Florey Institute and [§]School of Chemistry, The University of Melbourne, Parkville, VIC 3010, Australia; [‡]The University of Queensland, Institute for Molecular Bioscience, Brisbane, QLD 4072, Australia

[¶]Address correspondence to:
K. Johan Rosengren, School of Pure and Applied Natural Sciences, University of Kalmar, SE-391 82 Kalmar, Sweden, phone: +46 480 446152 fax: +46 480 446262 email: johan.rosengren@hik.se

Short Title: 3D structure of INSL5

SYNOPSIS

Insulin-like peptide 5 (INSL5) is a two-chain peptide hormone related to insulin and relaxin. It was recently discovered through searches of expressed sequence tag databases and although the full biological significance of INSL5 is still being elucidated, high expression in peripheral tissues such as the colon as well as in the brain and hypothalamus, suggests roles in gut contractility and neuroendocrine signaling. INSL5 activates the relaxin family peptide receptor 4 with high potency, and appears to be the endogenous ligand for this receptor, based on overlapping expression profiles and their apparent co-evolution. Here we have used solution state NMR to characterize the three-dimensional structure of synthetic human INSL5. The structure reveals an insulin/relaxin-like fold with three helical segments that are braced by three disulfide bonds and enclose a hydrophobic core. Furthermore, we characterized in detail the hydrogen bond network and electrostatic interactions between charged groups in INSL5 by NMR-monitored temperature and pH titrations and undertook a comprehensive structural comparison with other members of the relaxin family, thus identifying the conserved structural features of the relaxin fold. The B-chain helix, which is the primary receptor-binding site of the relaxins, is longer in INSL5 than in its close relative relaxin-3. As this feature results in a different positioning of the receptor activation domain Arg^{B23} and Trp^{B24} it may be an important contributor to the difference in biological activity observed for these two peptides. Overall the structural studies provide mechanistic insights into the receptor selectivity of this important family of hormones.

Keywords: INSL5, relaxin, peptide hormone, RXFP3, RXFP4, NMR, solution structure

INTRODUCTION

The relaxin family comprises seven peptides in humans, namely relaxins 1-3 [1-3] and the insulin-like peptides (INSL) 3-6 [4-7]. Although the mammalian hormone 'relaxin' (equivalent to relaxin-2 in humans) was identified 80 years ago and has long been known to be associated with pregnancy, both relaxin and its related hormones have, over recent years, been shown to be multifunctional peptides involved in numerous physiological processes. These processes include reproduction, neurosignaling, wound healing, collagen metabolism, cancer and others. Structurally the relaxins are related to insulin. In their mature form, after a connecting C-peptide in the precursor protein has been removed, they comprise two peptide chains (A and B) that are connected by two disulfide bonds, with the A-chain having one additional intra-molecular disulfide bond [8]. Together with insulin and the insulin-like growth factors I and II, the relaxins form the insulin/relaxin superfamily of signaling proteins (Figure 1).

INSL5 was first identified through database searches of expressed sequence tags [6] and although considerable efforts have been made to characterize its physiological role(s), the primary biological function remains unknown. Human INSL5 mRNA has been detected in various peripheral tissues, particularly in rectal and colon tissue [6]; however, INSL5 mRNA has also been located in the brain, primarily in the pituitary [9]. In mice, the highest level expression is found in the colon and kidneys [6, 10], but a recent study detected INSL5 mRNA in the hypothalamus. The immunohistochemical detection of INSL5-immunoreactive neurons in the paraventricular, supraoptic, accessory secretory, and supraoptic retrochiasmatic nuclei, as well as immunoreactive cell processes in the internal layer of the median eminence, together with its detection in the hypothalamus, suggests an additional neuroendocrine function for INSL5 in the mouse [11]. Intriguingly the genes for both INSL5 and its receptor are dysfunctional in the dog and rat genomes [9], despite the aforementioned findings all indicating an important role for INSL5 in mammalian biology. An important step towards the characterization of INSL5 function *in vivo* was made recently when an optimized synthesis strategy was presented, which allows the production of sufficient quantities of INSL5 and analogues for further biological studies [12].

In contrast to insulin and the insulin-like growth factors I and II, which interact with tyrosine kinase receptors, the relaxin peptides signal through G-protein coupled receptors (GPCRs). To date, four receptors have been identified and they are referred to as the relaxin family peptide receptors 1-4 (RXFP1-4). Interestingly, the relaxin-2 receptor, RXFP1 (LGR7) [13], and the INSL3 receptor, RXFP2 (LGR8) [14], are both characterized by a large N-terminal leucine rich-repeat containing ligand binding domain. In contrast, the relaxin-3 receptor, RXFP3 (GPCR135) [15], and the INSL5 receptor, RXFP4 (GPCR142) [9], are classic peptide ligand GPCRs. Considerable cross-reactivity between the ligands and the various receptors has been observed, with relaxin-2 being a potent agonist of both RXFP1 and RXFP2 [13], while relaxin-3, which is the ancestor of the relaxin family [16], has the unique feature of being able to activate relaxin receptors from both subgroups, namely RXFP1 [17], RXFP3 [15] and RXFP4 [18]. Despite relaxin-3 showing the same potency on the INSL5 receptor RXFP4 as INSL5 itself, INSL5 does not activate the relaxin-3 receptor RXFP3, although it is able to bind to RXFP3 with poor affinity, and act as a weak RXFP3 antagonist [9]. Thus, the process through which the relaxins activate their receptors is complex and relies on compatible structural features of both the ligands and the receptors.

Over recent years considerable progress has been made towards determination of the structure-activity relationships of relaxins; however, many aspects of the peptide-receptor interactions remain poorly understood. A clearer picture of the features guiding the interactions, and those which allow the activation of the receptors, is crucial for the design of selective agonists and antagonists that can be used both as pharmacological tools and as drug leads. The crystal structures of human relaxin-2 (H2 relaxin) [19] and the NMR solution structures of human relaxin-3 (H3 relaxin) [20], human INSL3 [21], and a relaxin chimera peptide consisting of the A-chain from human INSL5 and the B-chain from H3 relaxin (R3/I5) [22], have revealed that although the overall fold is well conserved, differences in amino acid sequence have crucial effects on the structure, particularly evident around the termini. Although evidence suggesting that both the A and B chains are involved in the receptor interaction is mounting, at least for the larger RXFP1 and RXFP2 [23, 24], the primary binding interface of the relaxin family peptides has been mapped to an exposed surface on the B-chain helix [21, 25, 26]. In relaxin-3 the binding site for RXFP3 comprises residues Arg^{B8}, Arg^{B12}, Ile^{B15}, Arg^{B16}, and Phe^{B20}, and in addition the C-terminal residues Arg^{B26} and Trp^{B27} are crucial for receptor activation [26]. Relaxin-3 also uses the same set of residues for activation of RXFP4, with the exception of Arg^{B12} which is not required, and given the high conservation of these residue in INSL5, it is likely that INSL5 binds to RXFP4 in a similar way [26].

Here we provide new mechanistic insights into the receptor selectivity of this family of peptides by presenting the solution structure of human INSL5, the last relaxin peptide for which the receptor has been identified. The structure has a conserved relaxin-like overall fold, but the B-chain α-helix, which is part of the crucial receptor binding and activation domains, is longer in INSL5 than in relaxin-3, and extends through to the C-terminal Trp^{B24}. Since both Arg^{B23} and Trp^{B24} are important for the activation of RXFP3 and RXFP4, the observed differences may play an important role in receptor selectivity. Furthermore, through NMR-monitored pH and temperature titrations, we provide a detailed analysis of the hydrogen bond network and electrostatic interactions that are conserved throughout the relaxin structural framework. In addition to the new insights into the receptor interaction of INSL5, these data will be important for guiding the design of relaxin analogues for further probing of the structure-activity relationships and in the development of potential drug leads.

Experimental Section

NMR Spectroscopy

The NMR sample for structure determination contained 1 mg lyophilized peptide diluted in 0.5 ml 90% H_2O/10% D_2O (v/v) at pH ~4. Two-dimensional ^1H data were recorded at 298 K on a Bruker Avance II spectrometer operating at 900 MHz and equipped with a cryogenic probe. All data, including TOCSY (with a MLEV17 mixing time of 60 ms), NOESY (with a mixing time of 100 and 150 ms) and DQF-COSY were recorded and processed using Topspin (Bruker, Germany). Generally, 4k data points were collected in the F2 dimension, with 512 increments in the F1 dimension over a spectral width of 12 ppm, with the F1 zero-filled to 1k data points and 90° phase-shifted sine bell window functions applied prior to transformation. Hydrogen bond donors were identified from temperature coefficients ($\Delta\delta_{HN}/\Delta T$): A series of TOCSY spectra were recorded at 600 MHz at 285, 295, 298, 305 and 310 K and the chemical shifts of the amide protons were plotted as a function of temperature, with the $\Delta\delta_{HN}/\Delta T$ being represented by the slope of the linear plot. The

pK_a values of all Asp and Glu residues were investigated by NMR monitored pH titrations. A series of TOCSY spectra were recorded at 600 MHz at pH 2.0, 2.5, 2.7, 3.0, 4.0, 5.0, and 6.0 and the chemical shift changes of all resonances were analyzed. All protons that could be positively identified at all pH values and showed a difference in chemical shift of more than 0.1 ppm throughout this interval, were fitted to sigmoidal curves using GraphPad Prism 5. The chemical shifts of all spectral data were calibrated to the resonance frequency of the solvent signal.

Structure Calculations

All spectral analysis was done using the program package CARA [27]. Resonance assignments were achieved using sequential assignment strategies and once the chemical shift assignments were confirmed, all NOE contacts were identified and converted to inter-proton distances using CYANA [28]. The backbone dihedral angle ϕ was restrained to -100 ±80º where a positive angle could be excluded based on an NOE pattern of a strong sequential $H\alpha_{i-1}$-HN_i in comparison to the intra-residual $H\alpha_i$-HN_i cross-peak. The side chain χ1 angles and the stereospecific Hβ proton assignment were determined by combining information about HN-Hβ and Hα-Hβ NOE peak intensities in a 100 ms mixing time NOESY with Hα-Hβ coupling constants from the DQF-COSY [29]. Based on this analysis the χ1 angles were restrained to -60 ±30º for a t^2g^3 conformation (Asp^{A16}, Cys^{A24}, Cys^{B7}, Glu^{B10}, Cys^{B19} and Trp^{B24}) and to -180 ±30º for a predicted g^2t^3 conformation (Met^{A14}). The χ1 angle of Asp^{A10} was restrained to -120 ±90º as a g^2g^3 conformation could be ruled out based on the coupling constant pattern. For Val^{B15}, the side chain conformation and stereo specific assignment of the two methyl groups could be determined by analysis of the pattern of NOEs. Hydrogen bond restraints were included for amide protons having a $\Delta\delta_{HN}/\Delta T$ >-4.6 ppm/K where suitable acceptors could be identified in the structure. Initial structure calculations were done using torsion angle dynamics in the program CYANA [28]. The final structure calculations were performed using simulated annealing and energy minimization protocols within the program CNS 1.2 [30], using the parameter and topology files parallhdg5.3.pro and topallhdg5.3.pro [31]. A set of 50 structures was generated by a two-part simulated annealing protocol. The first part is comprised of 4000 steps of 0.015 ps torsion angle dynamics, a cooling stage with 4000 steps of 0.015 ps during which the temperature is lowered from 50000 K to 0 K, and an energy minimization stage with 5000 steps of Powell minimization. The second part entails a refinement in explicit solvent and involves: (1) Heating to 500 K via steps of 100 K, each comprising 50 steps of 0.005 ps of Cartesian dynamics. (2) 3500 steps of Cartesian dynamics at 500 K. (3) A cooling phase in which the temperature is lowered in steps of 100 K, each comprising 3500 steps of Cartesian dynamics. (4) An energy minimization stage comprising 3000 steps of Powell minimization. From the 50 calculated the 20 lowest energy models were chosen to represent the INSL5 solution structure and analyzed using PROCHECK [32]. Structure figures were generated using MOLMOL [33]. The Ramachandran analysis showed that 92.6% of the residues are in the most favored region with the remaining 7.4% being found in the additionally allowed regions. The coordinates for the final set of structures, and the NMR restraints used to generate it, have been submitted to the PDB and given the accession code: 2kbc.

Results and Discussion

Peptide Synthesis

The human INSL5 A- and B-chain were separately synthesized by continuous flow Fmoc-solid phase peptide synthesis. Both chains were unusually difficult to make via conventional synthesis protocols but their successful assembly was achieved after the development of highly optimized conditions and through the use of a pseudoproline at a key position in the A-chain, as described in detail previously [12]. The two chains were subsequently combined by directed regioselective disulfide bond formation, and the final product was purified by RP-HPLC and characterized by MALDI-TOF MS and NMR analysis, confirming its identity and purity.

NMR Spectroscopy and Resonance Assignments

For the structure determination of human INSL5, two-dimensional ^{1}H NMR spectra were recorded at 900 MHz on a sample containing ~0.2 mM peptide at pH 4. The spectra showed good signal dispersion, confirming that INSL5 is highly structured in solution, and allowed near complete backbone and side chain resonance assignments to be achieved through sequential assignment strategies [34]. As seen in previous studies of relaxins, significant line broadening was observed for a number of residues, including: Cys^{A7}, Cys^{A8}, Cys^{A12}, Leu^{A17}, Ser^{A18}, Cys^{A21}, Val^{B4}, Arg^{B5}, Leu^{B6}, Gly^{B8}, Leu^{B9}, Ile^{B12}, Ile^{B16}, Cys^{B19}, and Ala^{B20}. Particularly affected were the amide protons of Cys^{A8} and Cys^{A12}, which were broadened beyond detection under the conditions used for this study. Although this broadening complicated the sequential assignment process, the identity of the observed spin systems was established through the detection of key long-range contacts and by chemical shift comparison to other relaxins. The broadened resonances are centered around the Cys^{A7}-Cys^{A12} intra-chain disulfide bond, and are possibly the result of the disulfide bond switching between a pro-*R* and pro-*S* conformation, which is a feature that appears to be common throughout the relaxin family [21]. This disulfide bond is buried in the hydrophobic core and packs against the aromatic ring of Tyr^{B11}. Thus, small structural rearrangements can potentially affect many residues and lead to large chemical shift differences between conformers from ring-current effects, resulting in severe broadening. In INSL5 the broadening extends further from this region than in related family members and includes the amide protons of Cys^{A21}, Ile^{B16}, Cys^{B19}, and Ala^{B20}, which suggests that there may be further dynamic processes involving the Cys^{A21}-Cys^{B19} disulfide bond. INSL5 does not contain any proline residues, and hence minor conformations arising from cis/trans isomerisation of the imide-bond are not a factor in this case. Such isomerisation has previously been reported for relaxin-3 and INSL3. Secondary Hα chemical shifts (supplementary information), which are good indicators of secondary structure, clearly indicate that INSL5 adopts three helical segments and two short extended regions, consistent with a relaxin structural fold.

Structure Determination of INSL5

Structural constraints, including proton-proton distances, hydrogen bonds and dihedral angles, were deduced from the NMR data. Interproton distances were derived from the intensities of cross-peaks in a NOESY spectrum recorded at 298 K, with a mixing time of 100 ms. In addition to the intra-residual and sequential NOEs, which are expected from the covalent structure and form the basis of the sequential assignments, a large number of non-sequential contacts were identified. The majority of the medium range NOEs define the fold of two helical segments in the A-chain and one helical segment in the B-chain, whereas most long range contacts occur between the two A-chain helices, which lie anti-parallel to each other, or between the A-chain helices and the B-chain helix, placing the B-chain helix across the face of the two A-

chain helices. A summary of the complete distribution of non-sequential NOEs throughout the INSL5 sequence is available in the supplementary material.

Hydrogen bonds were identified by amide temperature coefficient ($\Delta\delta_{HN}/\Delta T$) analysis, i.e. by measuring the susceptibility of the amide proton chemical shifts to changes in temperature. Although the measurements were complicated by increased resonance broadening at lower temperatures, for most residues $\Delta\delta_{HN}/\Delta T$ values could be confidently determined. In general, amide protons that are involved in hydrogen bonds are protected from the solvent and are less affected by changes in temperature than free amides. $\Delta\delta_{HN}/\Delta T$ values of >-4.6 ppm/K are indicative (>85%) of hydrogen bond, with the probability increasing further to >93% if $-4.0 < \Delta\delta_{HN}/\Delta T < -1.0$ ppb/K [35]. However, phenomena such as partial unfolding with increasing temperature can affect the $\Delta\delta_{HN}/\Delta T$ values, and structural restraints based on such information should be used with caution [36]. Here we used preliminary structures to guide the analysis of the $\Delta\delta_{HN}/\Delta T$ data, and restraints for hydrogen bonds were only introduced if there was strong evidence from both the $\Delta\delta_{HN}/\Delta T$ and the structure calculations that a hydrogen bond was present. Strikingly, for all amides predicted to be hydrogen bonded based on the $\Delta\delta_{HN}/\Delta T$ data a suitable acceptor was identified in the structure. The majority of the hydrogen bonds are located within the three helical regions, with only the Val^{B4}/Cys^{A12} hydrogen bond pair being present between the two chains. The complete backbone hydrogen bond network is summarized in the supplementary information.

Information about backbone and side-chain dihedral angles is important for refining structures by defining the local geometry of the peptide chain. Generally dihedral angle restraints are derived from scalar coupling constants via the Karplus equation. Although in INSL5 the analysis was complicated by resonance broadening, which prevented accurate determination of many of the coupling constants in the DQF-COSY spectrum, a set of dihedral angle restraints for ϕ and $\chi 1$ angles was derived from a combination of apparent coupling constants, and intra-residual and sequential NOE patterns.

The experimental data were used to generate a family of 20 structures to represents the solution structure of INSL5. An overlay of these structures is shown in Figure 2, from which it is clear that the structure is well defined over most regions, with only the B-chain N-terminus, which is likely flexible in solution, being poorly defined. The structure is in excellent agreement with the experimental data and has good covalent geometry, as evident from Table 1, which summarizes the experimental and structural statistics.

Description of the Three-Dimensional Structure and Comparison to Other Relaxins

Overall the INSL5 structure reveals an insulin/relaxin fold with three helical segments that enclose a hydrophobic core in an arrangement that is held together by the three disulfide bonds. The A-chain is characterized by two α-helices comprising residues A2-A9 and A14-A22, which are separated by an extended region (residues A10-A13). The last part of the C-terminal helix (Leu^{A21} and Cys^{A22}) adopts 3_{10} helix characteristic hydrogen bonds and a similar pattern cannot be ruled out for the residues Cys^{A7}, Cys^{A8} and Thr^{A9}, as this region of the N-terminal helix is less defined due to the broadening and lack of restraints in this region. The B-chain is disordered at the N-terminus, which is followed by an extended region comprising residues B4-B7 and thereafter a longer helical segment comprising residues B8-B24. The two extended regions in the A- and B-chain adopt a short β-sheet stabilized by a pair of

hydrogen bonds between Cys^{A12} and Val^{B4}. The hydrophobic core of INSL5, which is enclosed by the elements of secondary structure comprises primarily the residues: Leu^{A3}, Cys^{A7}, Cys^{A12}, Leu^{A17}, Leu^{A20}, Leu^{B6}, Tyr^{B11}, and Val^{B15}, with residues Gln^{A4}, Cys^{A21}, Ile^{B12}, Ile^{B16} and Cys^{B19} providing additional hydrophobic interactions on the fringe of the core. This hydrophobic core is highly conserved between relaxins but the types of hydrophobic residues at these key positions can change, slightly altering the packing of the core and resulting in minor differences in the orientations of the helices. The conservation of the overall structural features is illustrated in Figure 3, which shows a comparison of the INSL5 structure with the NMR structure of H3 relaxin and the crystal structure of H2 relaxin. Figure 3 also illustrates the similar positioning of the residues found to be important for binding to the receptors RXFP3 and RXFP4, which are all exposed on the B-chain helix in INSL5 and H3 relaxin.

Although their general structural features are similar there are a number of interesting differences between INSL5 and H3 relaxin, including the conformation of the A-chain N-terminus and the B-chain termini. The A-chain N-terminus is three residues shorter in INSL5 than in H3 relaxin, H2 relaxin and INSL3 and as a result the helix in this region is also shorter. Recently it has been shown that this region of the hormones is not needed for receptor activity, and as evident from the structure of INSL5 presented here, it is not required for overall structural integrity either. The INSL5 B-chain N-terminus is also three residues shorter than in H3 relaxin and although poorly defined it folds around the extended region to form electrostatic interactions with the A-chain. In contrast, in H3 relaxin the tail interacts more with other parts of the B-chain, and in particular a number of hydrophobic interactions are seen between Pro^{B4} and Phe^{B20} and between Tyr^{B5} and Lys^{A17}. Even more significant are the differences around the B-chain C-terminus, which in INSL5 adopts a helix that extends towards the C-terminal Trp^{B24} residue. Interestingly this is similar to what has been seen in the crystal structure of H2 relaxin, but different to the NMR structures of H3 relaxin, INSL3 and the R3/I5 chimera, in all of which the helix ends at Cys^{B22}, with the five residue tail B23-B27 folding back on to the peptide core, as evident from a large number of NOEs between the core and the aromatic ring. No such NOEs are seen for INSL5.

Analysis of the Hydrogen Bond Network in Relaxins by Temperature Titrations

To further investigate these structural differences and to generate a detailed picture of the hydrogen bond network within the relaxin structural framework we recorded temperature coefficients also for H3 relaxin and a comparison between the two peptides is presented in Figure 4. The pattern of hydrogen bonds in the A-chain C-terminal helix is identical in both peptides, and also consistent with the crystal structure of H2 relaxin. The amide protons of Ser^{A13}(Ser^{A16} in H3 relaxin) and Asp^{A16}(Glu^{A19}) at the N-terminal part of the helix are hydrogen bonded to the side chain carboxyl of Asp^{A13}(Glu^{A16}). Residues A14/A17 and A15/A18 are not hydrogen bonded in the majority of the structures in the INSL5 and H3 relaxin ensembles, but in some models also interact with this carboxyl group. Residue Cys^{A12}(Cys^{A15}) hydrogen bonds to the B-chain in the short β-sheet region. The hydrogen bond pattern in the N-terminal A-chain helix is less defined in INSL5 and may differ slightly between the peptides, although this is difficult to confirm as both the $\Delta\delta_{HN}/\Delta T$ values and the structural details may be affected by the flexibility and broadening in this region. In the crystal structure of H2 relaxin an α-helical hydrogen bonding pattern is seen for residues corresponding to A1-A10 in INSL5, but this helix has been shown to be more flexible in H3 relaxin [20]. The B-chain N-terminal region is characterized

by a lack of hydrogen bonds, consistent with it being disordered. Val^{B4}(Val^{B7}) forms an inter-chain hydrogen bond to Cys^{A12}(Cys^{A15})and from Glu^{B10}(Glu^{B13}) onwards a stretch of α-helical hydrogen bonds is seen. However, the $\Delta\delta_{HN}/\Delta T$ values indicate two striking differences. First, Leu^{B9} in H3 relaxin appears to be hydrogen bonded and indeed is involved in an interchain hydrogen bond in the structure. The equivalent residue, Leu^{B6}, is not hydrogen bonded in INSL5. Second, the C-terminal residues in INSL5 all show $\Delta\delta_{HN}/\Delta T$'s consistent with hydrogen bonding, which is not the case in H3 relaxin. This finding confirms that the helical region extends further in INSL5 than in H3 relaxin. An interesting observation was however made for these five last residues. The data point at the highest temperature, 310 K, did not fit with the otherwise linear chemical shift/temperature relationship. This is likely the result of the helix being rather labile and starting to unfold at higher temperatures, leading to changes in shifts towards random coil values. As these amide protons are upfield-shifted, in this case the opposing shift changes due to unfolding vs. the usual hydrogen bonding temperature dependence would cancel each other, which would fit with the observations seen here. Thus for the values reported in Table 2, data points at 310 K were not included in the analysis of residues B20-B24 of INSL5.

Characterization of Electrostatic Interactions in INSL5 by pH Titrations

Salt bridges and electrostatic interactions can be confirmed by determination of the pK_a values of ionizable groups, which in turn can be derived from titration curves of chemical shifts as a function of pH. To analyze the structural roles of the five acidic residues in INSL5, Asp^{A2}, Asp^{A10}, Asp^{A16}, Glu^{B2} and Glu^{B10}, we studied the peptide in the range pH 2-6. The native peptide comprises additional acidic groups expected to titrate within this range at the two C-termini, but the synthetic peptide studied here, which is fully active, was prepared with amidated C-termini. Although most resonances were relatively unaffected by the changes in pH, apart from a general broadening of the amide protons with increasing pH as a result of increased solvent exchange, a number of resonances were significantly shifted. The titration curves for the resonances showing a chemical shift that changed by more than 0.1 ppm over the studied range are presented in Figure 5A.

The expected pK_a values for unperturbed Asp and Glu residues are 3.8-4.0 and 4.3-4.5 respectively [37]. However, the observed pK_a values of Asp^{A2}, Asp^{A10}, Asp^{A16}, Glu^{B2} and Glu^{B10} in INSL5 were 2.7, 2.6, 2.7, 2.8 and 2.9 respectively, indicating that interactions favor the negatively charged form and shift the equilibrium by >1 pH unit. Analysis of the INSL5 structure provides evidence for such interactions: Asp^{A2} functions as an N-terminal helix cap, and form a network of hydrogen bonds to the amide protons of Gln^{A4} and Thr^{A5}, as well as the side chain amide proton of Gln^{A4} and the side chain hydroxyl proton of Thr^{A5} (Figure 5B). Asp^{A10} is in close proximity to Lys^{B1}, which in addition to its positively charged side chain also carries the positively charged N-terminal amine group (Figure 5B). Asp^{A16} is also close to Lys^{B1} but also forms an N-terminal helix cap on the second A-chain helix and hydrogen bonds to both the amide proton and the side chain hydroxyl proton of Ser^{A13}, and its own amide proton (Figure 5C). The Glu^{B2} side chain interacts with Arg^{B5} in some structures, as seen in Figure 5B. In another set of structures the side chain hydrogen bonds to the hydroxyl side chains of Ser^{A13} and Thr^{A15}. It is not possible to rule out either of the conformations based on our experimental data. In addition to a possible interaction with Glu^{B2}, Arg^{B5} also interacts with the negative part of the N-terminal A-chain helix dipole. Finally, Glu^{B10} is situated at the N-terminal of the longer B-chain helix, interacting with the helix dipole and the positive

charge of Arg^{B13} (Figure 5D), although the latter interaction is unlikely to be important for the functional form of INSL5 as Arg^{B13} is crucial for receptor binding.

The chemical shift of the amide proton of Val^{B4} has significant pH dependence. Interestingly, this amide forms an inter-chain hydrogen bond with Cys^{A12}, and thus the change in chemical shift reflects a change in this interaction. Strongly hydrogen bonded amides are deshielded and as the chemical shift of Val^{B4} moves downfield with increasing pH, this supports a strengthening of the hydrogen bond, and thus a tighter association between the chains at high pH. Furthermore, the Hα protons of Asp^{A13}, Glu^{B5}, Gly^{A11}, Val^{B4} and Leu^{B6} all move downfield with increasing pH, confirming that the β-sheet interaction between the two chains in this region is strengthened. The small difference in pK_a values between the carboxylic groups makes it difficult to pin-point one electrostatic interactions that is responsible for this stabilizing effect, but the ones involving Lys^{B1} and Glu^{B2}, both tie the B-chain N-terminal region closer to the molecular core.

In the light of the finding that all acidic residues in human INSL5 form interactions in the three-dimensional fold, it is interesting to note that Asp^{A2}, Asp^{A16} and Glu^{B10} at the N-terminus of each of the three helical segments are highly conserved throughout INSL5 sequences from different species and in H3 relaxin. However, in other peptides throughout the family these residues are not as well conserved, indicating that they are not an absolute requirement for a stable relaxin fold.

INSL5 interactions with the receptors RXFP3 and RXFP4

The observation that INSL5, like H3 relaxin and INSL3, appears to have a flexible fold, is interesting given that a significant degree of flexibility has previously been suggested to be crucial for the biological activity of insulin. The insulin analogues des-PheB25 insulin and PT insulin both have increased activity at the insulin receptor, despite having destabilized folds with increased flexibility [38, 39]. It is possible that for INSL5 and the other relaxins a flexible fold is required for adopting an optimal conformation for interacting with their receptors. Structure-activity studies on H3 relaxin have established that it interacts with both the RXFP3 and the RXFP4 receptors in a similar fashion, using a series of residues exposed on the surface of the B-chain. Specifically, Arg^{B8}, Ile^{B15}, Arg^{B16} and Phe^{B20} are important for the binding to both receptors, and in addition Arg^{B12} is important for binding to RXFP3 [26]. Based on the close relationship between H3 relaxin and INSL5, as well as between their receptors RXFP3 and RXFP4, it would seem likely that INSL5 interacts with RXFP4 in a similar way. This is further supported by the high conservation of the residues identified as receptor contacts in H3 relaxin, with the only difference between the peptides being a conservative Phe^{B20} to Tyr^{B17} substitution (Figure 1). In contrast, at the position equivalent to Arg^{B12} in H3 relaxin, which was found to be important only for binding to RXFP3, INSL5 hosts Leu^{B9}, and it is possible that this non-conservative change accounts for much of the decrease in affinity of INSL5 for this receptor.

More intriguing is the difference between the two peptides in terms of potency on the RXFP receptors. For H3 relaxin the activation domain for both RXFP3 and RXFP4 is located at the B-chain C-terminal residues Arg^{B26} and Trp^{B27}. However, despite both these residues being fully conserved in INSL5, INSL5 acts as an RXFP4 agonist but an RXFP3 antagonist. We show here that there are important structural differences between H3 relaxin and INSL5 in this region. The INSL5 B-chain helix extends to the C-terminal Trp^{B24}, whereas in H3 relaxin the helical segments end with

Cys^{B22}, five residues earlier (Figure 5). In H3 relaxin, rather than extending away from the core the C-terminal tail folds back onto the molecule and the aromatic side chain of Trp^{B27} forms a large number of contacts with residues Val^{B18}, Ile^{B19} and the Cys^{A24}-Cys^{B22} disulfide bond, which are evident from NOEs [20]. A similar arrangement of the tail is seen in the solution structures of both the R3/I5 chimera and INSL3 [21, 22]. Despite the Trp in these molecules clearly preferring to interact with the peptide core, the tail appears generally quite flexible, with limited backbone NOEs and small deviations from random coil shifts, and it is likely that once bound to the receptor it can adopt a different conformation if needed for activity. Thus, if the RXFP3 and RXFP4 receptors require different positioning of the C-terminal tail for activation, this may explain why H3 relaxin can activate both receptors. In contrast in INSL5 the preferred positioning of Arg^{B21} and Trp^{B22} is clearly defined by the helical conformation and is further away from the molecular core. This conformation may not but suitable for the interaction needed for activation of RXFP3, thereby giving INSL5 its antagonistic activity. The reason for the structural differences between INSL5 and H3 relaxin is likely the presence of the sequence Gly^{B23}-Gly^{B24} immediately after Cys^{B22}, which would be expected to be strongly destabilizing for a helical conformation. In contrast, INSL5 has the considerably more favorable Ala^{B20}-Ser^{B21} pair in the equivalent position, which could support an extension of the helix.

CONCLUSIONS

In this study we presented the solution NMR structure of the human peptide hormone INSL5 and characterized the hydrogen bond and electrostatic interactions that stabilize the relaxin structural framework. The structure is tightly folded and characterized by three helical segments that enclose a hydrophobic core. Despite the stabilizing hydrophobic interactions in the core, the extensive hydrogen bond network, favorable electrostatic interactions involving charged side chain groups and the covalent cross bracing of the disulfide array, the structure is rather flexible as evident from significant line broadening of many resonances. This flexibility appears to be a common characteristic of the relaxin family, and may well be important for function. The main structural differences between INSL5 and H3 relaxin occur around the B-chain termini, which adopts a helical conformation in INSL5. These differences are particularly interesting considering that the C-terminal Arg-Trp residues are crucial for H3 relaxin's activation of both the RXFP3 and RXFP4 receptors. Future mutational studies should be directed at this region to explore the role of this Arg-Trp motif and the helical conformation in the pharmacological profile of INSL5.

ACKNOWLEDGEMENT

KJR receives financial support from the Faculty of Natural Sciences & Technology, University of Kalmar and Åke Wiberg's Foundation. JDW is supported by the National Health & Medical Research Council (NHMRC), Australia (JDW). DJC is an NHMRC Principal Research Fellow. NLD is a Queensland Smart State Fellow. We acknowledge the use of the 900 MHz NMR within the Queensland NMR Network (QNN).

REFERENCES

1 Hudson, P., Haley, J., John, M., Cronk, M., Crawford, R., Haralambidis, J., Tregear, G., Shine, J. and Niall, H. (1983) Structure of a genomic clone encoding biologically active human relaxin. Nature **301**, 628-631

2 Hudson, P., John, M., Crawford, R., Haralambidis, J., Scanlon, D., Gorman, J., Tregear, G., Shine, J. and Niall, H. (1984) Relaxin gene expression in human ovaries and the predicted structure of a human preprorelaxin by analysis of cDNA clones. Embo J. **3**, 2333-2339

3 Bathgate, R. A., Samuel, C. S., Burazin, T. C., Layfield, S., Claasz, A. A., Reytomas, I. G., Dawson, N. F., Zhao, C., Bond, C., Summers, R. J., Parry, L. J., Wade, J. D. and Tregear, G. W. (2002) Human relaxin gene 3 (H3) and the equivalent mouse relaxin (M3) gene. Novel members of the relaxin peptide family. J. Biol. Chem. **277**, 1148-1157

4 Burkhardt, E., Adham, I. M., Hobohm, U., Murphy, D., Sander, C. and Engel, W. (1994) A human cDNA coding for the Leydig insulin-like peptide (Ley I-L). Hum. Genet. **94**, 91-94

5 Chassin, D., Laurent, A., Janneau, J. L., Berger, R. and Bellet, D. (1995) Cloning of a new member of the insulin gene superfamily (INSL4) expressed in human placenta. Genomics **29**, 465-470

6 Conklin, D., Lofton-Day, C. E., Haldeman, B. A., Ching, A., Whitmore, T. E., Lok, S. and Jaspers, S. (1999) Identification of INSL5, a new member of the insulin superfamily. Genomics **60**, 50-56

7 Lok, S., Johnston, D. S., Conklin, D., Lofton-Day, C. E., Adams, R. L., Jelmberg, A. C., Whitmore, T. E., Schrader, S., Griswold, M. D. and Jaspers, S. R. (2000) Identification of INSL6, a new member of the insulin family that is expressed in the testis of the human and rat. Biol. Reprod. **62**, 1593-1599

8 Schwabe, C. and McDonald, J. K. (1977) Relaxin: a disulfide homolog of insulin. Science **197**, 914-915

9 Liu, C., Kuei, C., Sutton, S., Chen, J., Bonaventure, P., Wu, J., Nepomuceno, D., Kamme, F., Tran, D. T., Zhu, J., Wilkinson, T., Bathgate, R., Eriste, E., Sillard, R. and Lovenberg, T. W. (2005) INSL5 is a high affinity specific agonist for GPCR142 (GPR100). J. Biol. Chem. **280**, 292-300

10 Hsu, S. Y. (1999) Cloning of two novel mammalian paralogs of relaxin/insulin family proteins and their expression in testis and kidney. Mol. Endocrinol. **13**, 2163-2174

11 Dun, S. L., Brailoiu, E., Wang, Y., Brailoiu, G. C., Liu-Chen, L. Y., Yang, J., Chang, J. K. and Dun, N. J. (2006) Insulin-like peptide 5: expression in the mouse brain and mobilization of calcium. Endocrinology **147**, 3243-3248

12 Akhter Hossain, M., Bathgate, R. A., Kong, C. K., Shabanpoor, F., Zhang, S., Haugaard-Jönsson, L. M., Rosengren, K. J., Tregear, G. W. and Wade, J. D. (2008) Synthesis, conformation, and activity of human insulin-like peptide 5 (INSL5). Chembiochem **9**, 1816-1822

13 Hsu, S. Y., Nakabayashi, K., Nishi, S., Kumagai, J., Kudo, M., Sherwood, O. D. and Hsueh, A. J. (2002) Activation of orphan receptors by the hormone relaxin. Science **295**, 671-674

14 Kumagai, J., Hsu, S. Y., Matsumi, H., Roh, J. S., Fu, P., Wade, J. D., Bathgate, R. A. and Hsueh, A. J. (2002) INSL3/Leydig insulin-like peptide activates the LGR8 receptor important in testis descent. J. Biol. Chem. **277**, 31283-31286

15 Liu, C., Eriste, E., Sutton, S., Chen, J., Roland, B., Kuei, C., Farmer, N., Jornvall, H., Sillard, R. and Lovenberg, T. W. (2003) Identification of relaxin-3/INSL7 as an endogenous ligand for the orphan G-protein-coupled receptor GPCR135. J. Biol. Chem. **278**, 50754-50764

16 Wilkinson, T. N., Speed, T. P., Tregear, G. W. and Bathgate, R. A. (2005) Evolution of the relaxin-like peptide family. BMC Evol. Biol. **5**, 14

17 Sudo, S., Kumagai, J., Nishi, S., Layfield, S., Ferraro, T., Bathgate, R. A. and Hsueh, A. J. (2003) H3 relaxin is a specific ligand for LGR7 and activates the receptor by interacting with both the ectodomain and the exoloop 2. J. Biol. Chem. **278**, 7855-7862

18 Liu, C., Chen, J., Sutton, S., Roland, B., Kuei, C., Farmer, N., Sillard, R. and Lovenberg, T. W. (2003) Identification of relaxin-3/INSL7 as a ligand for GPCR142. J. Biol. Chem. **278**, 50765-50770

19 Eigenbrot, C., Randal, M., Quan, C., Burnier, J., O'Connell, L., Rinderknecht, E. and Kossiakoff, A. A. (1991) X-ray structure of human relaxin at 1.5 A. Comparison to insulin and implications for receptor binding determinants. J. Mol. Biol. **221**, 15-21

20 Rosengren, K. J., Lin, F., Bathgate, R. A., Tregear, G. W., Daly, N. L., Wade, J. D. and Craik, D. J. (2006) Solution structure and novel insights into the determinants of the receptor specificity of human relaxin-3. J. Biol. Chem. **281**, 5845-5851

21 Rosengren, K. J., Zhang, S., Lin, F., Daly, N. L., Scott, D. J., Hughes, R. A., Bathgate, R. A., Craik, D. J. and Wade, J. D. (2006) Solution structure and characterization of the LGR8 receptor binding surface of insulin-like peptide 3. J. Biol. Chem. **281**, 28287-28295

22 Haugaard-Jonsson, L. M., Hossain, M. A., Daly, N. L., Bathgate, R. A., Wade, J. D., Craik, D. J. and Rosengren, K. J. (2008) Structure of the R3/I5 chimeric relaxin peptide, a selective GPCR135 and GPCR142 agonist. J. Biol. Chem. **283**, 23811-23818

23 Hossain, M. A., Rosengren, K. J., Haugaard-Jönsson, L. M., Zhang, S., Layfield, S., Ferraro, T., Daly, N. L., Tregear, G. W., Wade, J. D. and Bathgate, R. A. (2008) The A-chain of human relaxin family peptides has distinct roles in the binding and activation of the different relaxin family peptide receptors. J. Biol. Chem. **283**, 17287-17297

24 Park, J. I., Semyonov, J., Yi, W., Chang, C. L. and Hsu, S. Y. (2008) Regulation of receptor signaling by relaxin A chain motifs: derivation of pan-specific and LGR7-specific human relaxin analogs. J. Biol. Chem. **283**, 32099-32109

25 Bullesbach, E. E., Yang, S. and Schwabe, C. (1992) The receptor-binding site of human relaxin II. A dual prong-binding mechanism. J. Biol. Chem. **267**, 22957-22960

26 Kuei, C., Sutton, S., Bonaventure, P., Pudiak, C., Shelton, J., Zhu, J., Nepomuceno, D., Wu, J., Chen, J., Kamme, F., Seierstad, M., Hack, M. D., Bathgate, R. A., Hossain, M. A., Wade, J. D., Atack, J., Lovenberg, T. W. and Liu, C. (2007) R3(BDelta23 27)R/I5 chimeric peptide, a selective antagonist for GPCR135 and GPCR142 over relaxin receptor LGR7: in vitro and in vivo characterization. J. Biol. Chem. **282**, 25425-25435

27 Keller, R. (2004) The Computer Aided Resonance Assignment Tutorial. CANTINA Verlag

28 Guntert, P., Mumenthaler, C. and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. **273**, 283-298

29 Wagner, G. (1990) NMR Investigations of Protein-Structure. Prog. Nucl. Magn. Reson. Spectrosc. **22**, 101-139

30 Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system. Nat. Protoc. **2**, 2728-2733

31 Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Guntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R. and Bonvin, A. M. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins **59**, 662-672

32 Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR **8**, 477-486

33 Koradi, R., Billeter, M. and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. **14**, 51-55, 29-32

34 Wuthrich, K. (1986) NMR of proteins and nucleic acid. John Wiley and Sons Inc., New York

35 Cierpicki, T. and Otlewski, J. (2001) Amide proton temperature coefficients as hydrogen bond indicators in proteins. J. Biomol. NMR **21**, 249-261

36 Andersen, N. H., Neidigh, J. W., Harris, S. M., Lee, G. M., Liu, Z. H. and Tong, H. (1997) Extracting information from the temperature gradients of polypeptide NH chemical shifts .1. The importance of conformational averaging. J. Am. Chem. Soc. **119**, 8547-8561

37 Matthew, J. B., Gurd, F. R. N., Garciamoreno, E. B., Flanagan, M. A., March, K. L. and Shire, S. J. (1985) Ph-Dependent Processes in Proteins. CRC Crit. Rev. Biochem. **18**, 91-197

38 Jorgensen, A. M. M., Olsen, H. B., Balschmidt, P. and Led, J. J. (1996) Solution structure of the superactive monomeric Des-[Phe(BP5)] human insulin mutant: Elucidation of the structural basis for the monomerization of Des-[Phe(B25)] insulin and the dimerization of native insulin. J. Mol. Biol. **257**, 684-699

39 Keller, D., Clausen, R., Josefsen, K. and Led, J. J. (2001) Flexibility and bioactivity of insulin: an NMR investigation of the solution structure and folding of an unusually flexible human insulin mutant with increased biological activity. Biochemistry **40**, 10732-10740

Table 1. NMR and refinement statistics for INSL5

	INSL5
NMR distance & dihedral constraints	
Distance constraints	
Total Inter-residue NOE	454
Sequential (|i-j| = 1)	184
Medium-range (|i-j| ≤ 4)	146
Long-range (|i-j| ≥ 5)	124
Hydrogen bonds	40 (for 20 H-bonds)
Total dihedral angle restraints	
Φ	16
χ-1	9
Structure Statistics	
Violations (mean ± S.D.)	
Distance constraints (> 0.2 Å)	0.2 /structure
Max. dist. constraint violation (Å)	0.29
Dihedral angle constraints (> 2º)	0.05 /structure
Max. dihedral angle violation (º)	2.8
Mean Energies (kcal/mol)	
$E_{overall}$	-1835 ±9.1
E_{bond}	7.23 ±0.51
E_{angles}	77.2 ±3.5
$E_{improper}$	8.78 ±1.1
E_{vdw}	-132 ±7.1
E_{NOE}	7.81 ±1.4
E_{cdih}	0.036 ±0.11
E_{dih}	252 ±5.0
E_{elec}	-2056 ±13
Deviations from idealized geometry	
Bond lengths (Å)	0.00323 ± 0.00011
Bond angles (º)	0.637 ± 0.014
Impropers (º)	0.401 ± 0.026
Average pairwise r.m.s.d.* (Å)	
Heavy	1.22 ± 0.20
Backbone	0.64 ± 0.13
Ramachandran Statistics	
Most Favored Regions	92.6%
Additionally Allowed	7.4%
Generously Allowed	0%
Disallowed	0%

*Pairwise root mean square deviation was calculated over residues A1-A21 and B4-B24.

FIGURE LEGENDS

Figure 1. Sequence comparison of human INSL5 (boxed) to INSL5 from other species (top part), and to the remaining nine members of the human insulin/relaxin family (bottom part). The conserved cysteine residues, which form three disulfide bonds are highlighted in grey and their connectivity indicated by connecting lines. After the processing of the propeptide the N-terminal Gln^{A1} in both INSL5 and relaxin-2 is converted into a pyroglutamic acid.

Figure 2. The solution structure of human INSL5. Stereoview of the 20 lowest energy structures superimposed over residues A1-A21 and B4-B24. The N- and C-terminal residues are labeled with chain name and residues number. Disulfide bonds are shown in grey.

Figure 3. Structural comparison of human INSL5 (A), H3 relaxin (B) and H2 relaxin (C) illustrating the presence and positioning of the secondary structure and the cross-bracing disulfide array. (D) Comparison of the 'active site' of H3 relaxin (dark grey) and INSL5 (light grey). The heavy side chains are shown for Arg^{B5}/Arg^{B8}, Leu^{B9}/Leu^{B12}, Ile^{B12}/Ile^{B15}, Arg^{B13}/Arg^{B16}, Tyr^{B17}/Phe^{B20}, Arg^{B23}/Arg^{B26} and Trp^{B24}/Trp^{B27}, which have been shown to be important for the affinity and activation of the receptors RXFP3 and RXFP4.

Figure 4. Temperature coefficient data for INSL5 and H3 relaxin. The amino acid sequence of INSL5 is shown in bold with the sequence variation in H3 relaxin being indicated below. The cut-off value used to identify likely hydrogen bonds (-4.6 ppb/K) is represented by a dashed line. The helical regions are marked by schematic helices in the lower part of the figure. Missing data points indicate that the $\Delta\delta_{HN}/\Delta T$ values could not be determined due to poor signal intensity due to linebroadening or resonance overlap.

Figure 5. NMR monitored pH titrations of INSL5. (A) Titration curves for protons that could be identified at all conditions and that showed chemical shift changes of >0.1 ppm within the studied pH range. (B) Structural illustration of the hydrogen bond network of Asp^{A2}, and the electrostatic interactions between Asp^{A10}-Lys^{B1} and Glu^{B2}-Arg^{B5}. (C) Structural illustration of the hydrogen bond network of Asp^{A16}. (D) Structural illustration of the electrostatic interaction between Glu^{B10} and the B-chain helix dipole and Arg^{B13}.

Figure 1.

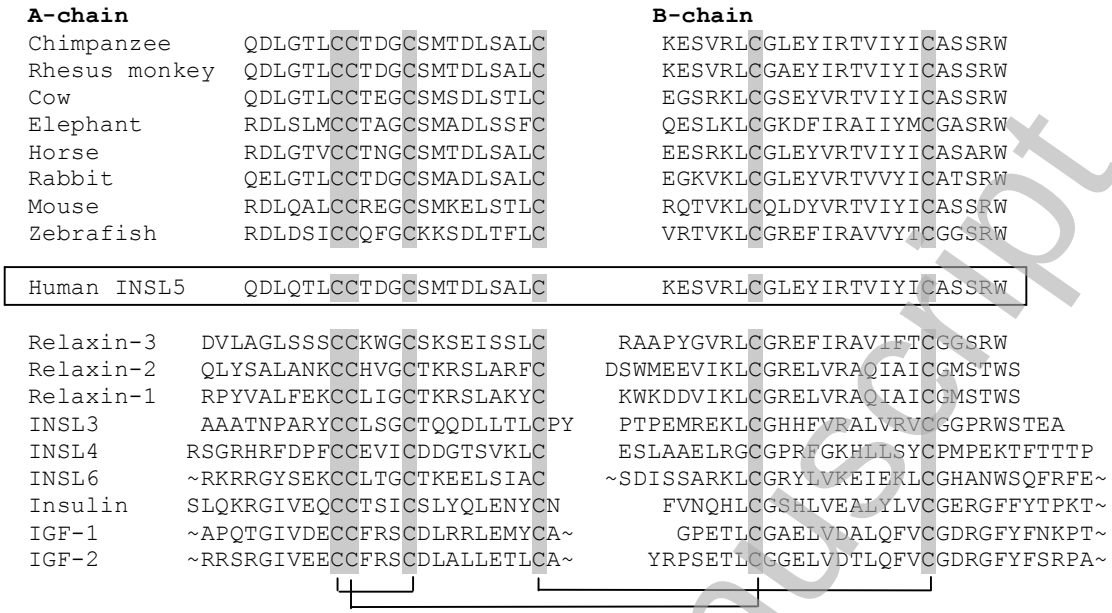

Figure 2.

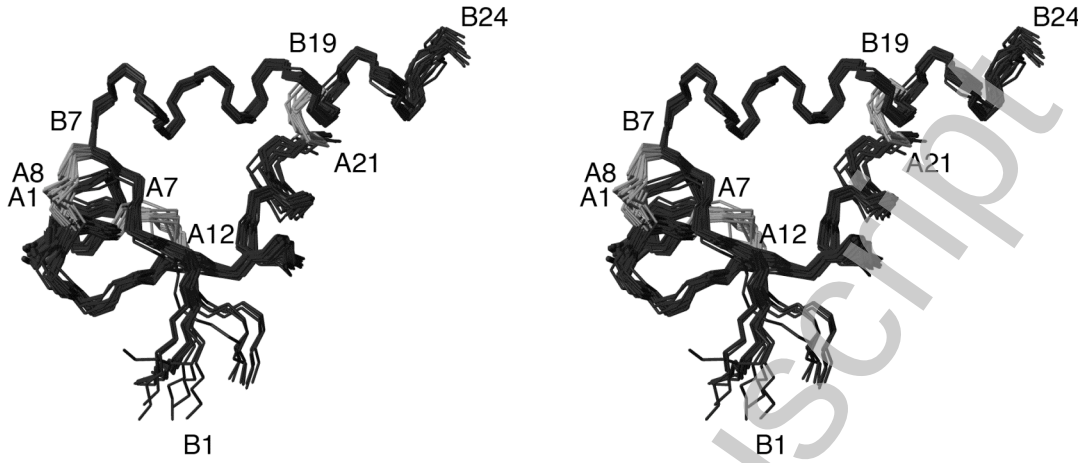

Figure 3.

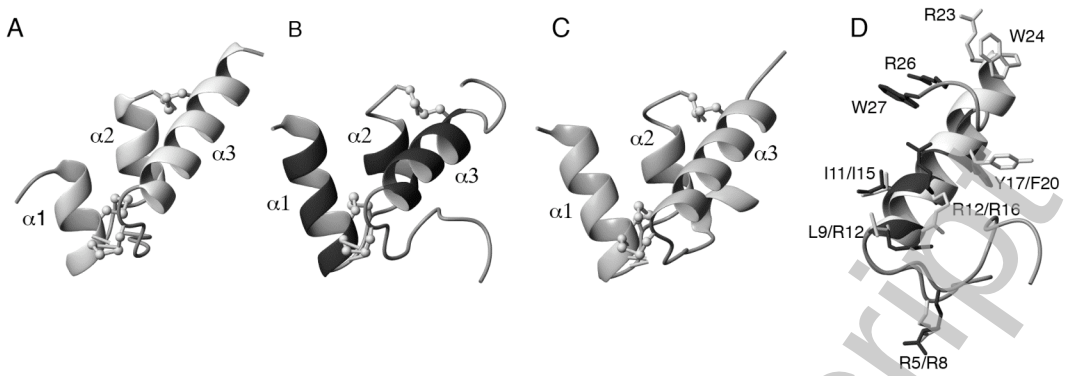

Figure 4.

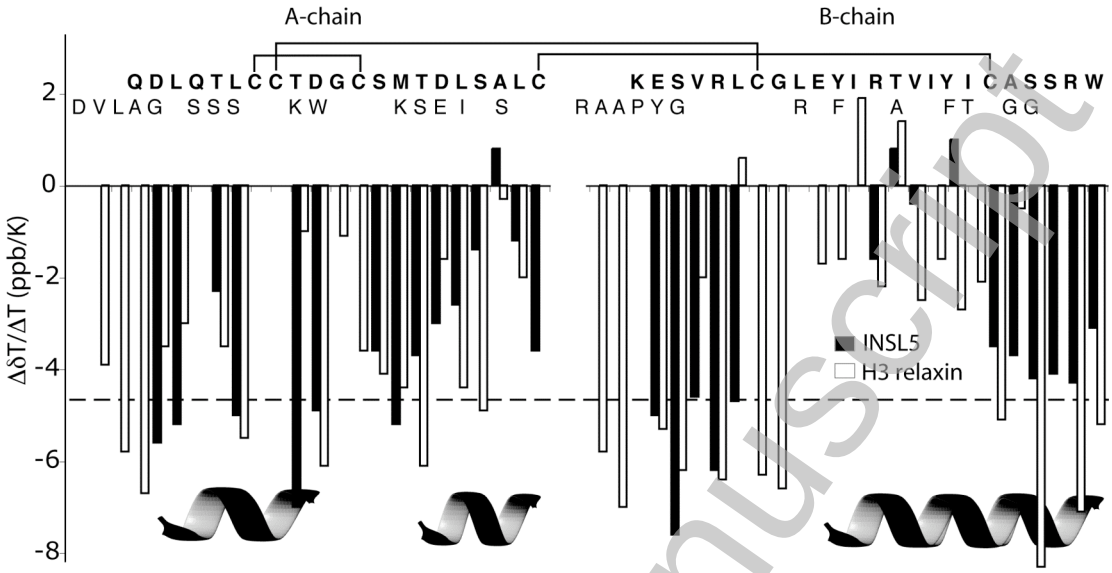

Figure 5.