Characterization of Asx-turn types and their connate relationship with β-turns

Viola d'Mello, Gildas Goldsztejn, Venkateswara Rao Mundlapati, Valérie Brenner, Eric Gloaguen, Florence Charnay-Pouget, David J Aitken, Michel Mons

Characterization of Asx-turn types and their connate relationship with β-turns

Viola C. D'mello,[a,#] Gildas Goldsztejn,[a,‡] Venkateswara Rao Mundlapati,[a,†] Valérie Brenner,[a]

Eric Gloaguen,[a] Florence Charnay-Pouget,[b,§] David J. Aitken,[b,*] and Michel Mons[a,*]

a Université Paris-Saclay, CEA, CNRS, Laboratoire Interactions, Dynamiques et Lasers (LIDYL), 91191 Gif-sur-Yvette, France

b Université Paris-Saclay, CNRS, Institut de Chimie Moléculaire et des Matériaux d'Orsay (ICMMO), 91405 Orsay, France

present address: Graphene Research Labs, KIADB IT Park, Near Airport Bengaluru, 562149, India

‡ present address: Université Paris-Saclay, CNRS, Institut des Sciences Moléculaires d'Orsay (ISMO), 91405 Orsay, France

† present address: Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse (UPS), CNRS, CNES, 9 Avenue du Colonel Roche, 31028 Toulouse, France

§ present address: Université Clermont Auvergne, CNRS, SIGMA Clermont, ICCF, F-63000 Clermont-Ferrand, France.

ORCID		
	Viola C. D'mello	0000-0002-6634-846X
	Gildas Goldsztejn	0000-0002-9123-6419
	Venkateswara Rao Mundlapati	0000-0003-0559-9684
	Valérie Brenner	0000-0002-8004-1157
	Eric Gloaguen	0000-0002-1023-2791
	Florence Charnay-Pouget	0000-0002-3540-9573
	David J. Aitken	0000-0002-5164-6042
	Michel Mons	0000-0001-9930-831X

* Corresponding authors : E-mail : michel.mons@cea.fr ; david.aitken@universite-paris-saclay.fr

Abstract

Models of asparagine-containing dipeptides, specifically designed to favor intrinsic folding into an Asx-turn, were characterized both theoretically, using quantum chemistry, and experimentally, using laser spectroscopy in the gas phase. Both approaches provided evidence for the spontaneous folding of both the Asn-Ala and Asn-Gly dipeptide models into the most stable Asx-turn, a conformation stabilized by a C10 H-bond which was very similar to a type II' β-turn. In parallel, analysis of the Asx-turns implicating asparagine in crystallized protein structures in the Protein Data Bank revealed a sequence-dependent behavior. In Asn-Ala sequences, the Asx-turn was found in conjunction with a type I β-turn for which the first of the four defining residues was Asn. The observation that the Asx turn in these structures is mostly of type II' (*i.e.* its most stable innate structure) suggests that this motif may foster the formation and/or enhance the stability of the backbone β-turn. In contrast, the Asx turns observed in Asn-Gly sequences extensively adopted a type II Asx-turn structure, suggesting that their formation should be ascribed to other factors, such as hydration. The fact that the Asx-turn in a Asn-Gly sequence is also often found in combination with a hydrated β-bulge supports the premise that a Asn-Gly sequence might efficiently promote the formation of the β-bulge secondary structure.

1. Introduction

Among the secondary structures of proteins and peptides, turns occupy a prominent position. Their ubiquity stems from their unique capacity to enable the reversal of the protein chain direction.[1] Their conservation within protein families, together with their robustness *vis-à-vis* mutations, has countenanced the deduction that they play a key role in folding mechanisms.[1e, 2] A typology of turn secondary structures has been established, based on surveys of X-ray structures in the Protein Data Bank (PDB).[1, 3] Ranging over four residues, labelled *i* to *i+3*, β-turns (Figure 1, top) are characterized by specific H-bonding between backbone NH and CO groups four residues apart (referred to as a C10 H-bond, in accordance with the size of the pseudo-cycle formed by the H-bond). There are four different types of β-turn,[4] whose occurrence depends on whether side chains are present (*i.e.* glycine vs. other amino acids) and the configurations of the chiral residues involved.[1] Early PDB structural surveys also reported the occurrence of an alternative H-bonding pattern, not involving a backbone CO but implicating instead the side chain of either an asparagine (Asn) or an aspartic acid (Asp) residue (with the latter in its carboxylate form under physiological conditions) which acted as the hydrogen bond acceptor.[1a, c, 5] In these H-bonds, the carboxamide oxygen atom of the Asn side chain (or one of the carboxylate oxygen atoms of the Asp side chain) interacts with the next-but-one NH site along the backbone, giving rise to a C10 H-bond bearing a marked resemblance to a β-turn; these structures are referred to as Asx-turns (Figure 1, bottom). Detailed analyses, however, have shown that Asx-turns are rarely isolated structures and are most often found in combination with another secondary structure, essentially a β-turn,[1c, 5] with which they form an intertwined C10 H-bond network, labelled an Asx-β-turn.[6] The coexistence of these features, of course, raises the question of anteriority: which one precedes the other during the folding process and thereby serves as the folding nucleus of the structure.

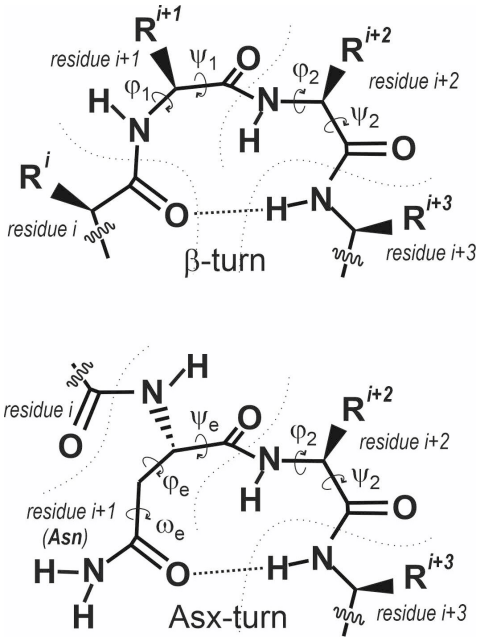


Figure 1: Geometrical representations of a β-turn (*top*) and an Asx-turn (*bottom*), with the latter illustrated for an Asn residue. The dihedrals describing the backbones are indicated by circular arrows (Ramachandran dihedrals in the β-turn case).

In this context, Asx-based secondary structures were suggested to account for the biased distribution of residues in the most frequently found β-turns, *viz.* type I,[2b, 5c, 7] and in particular for the over-representation (by a factor of 2-3) of the polar residues Asp and Asn in position *i* of a β-turn.[5c, 7b] It was thus surmised that if Asx-turns preexisted β-turns, they could foster the formation of the latter; alternatively, if formed subsequently

they could enhance β-turn stability. The over-representation (to a similar extent) of an Asx residue in a second location, namely position *i+2* of type I β-turns, was also reported[5c, 7b] but was not discussed in detail. It is worth noting here that the type I β-turn structures considered in these studies did not distinguish isolated turns from those embedded in larger secondary structures, such as β-hairpins[8] or β-bulges.[9] For these latter, the presence of a Gly residue in position *i+3* was considered as strongly favorable,[9b, 10] suggesting that the Asx-Gly (*i+2*, *i+3*) sub-sequence of the bulge may adopt a specific motif, again favoring formation of the β-turn or stabilizing it, depending on the folding chronology. Despite a detailed study of specific structures comprised of the most abundant residue at each of the four turn positions taken individually,[2c] such as the so-called consensus sequence Asx-Pro-Asx-Gly, no rationalisation has been proposed so far to the best of our knowledge.

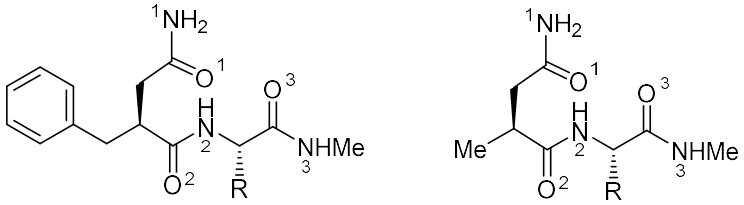
The structural diversity of Asx-turns was described in detail, based on a PDB survey,[11] which demonstrated a marked similarity between the typology of β-turns and Asx-turns; the latter closely resembled several types of the former, with a preference for type II', but again this study did not take account of the backbone or secondary structure environment of the Asx-turn considered.

In parallel with these investigations on crystallized proteins, the issue of the intrinsic formation of Asx-turn in two- or three-residue peptide chains has been studied in weakly polar solutions using IR and NMR spectroscopy, providing evidence for C10 H-bonding.[12] Crystal structures of these models were also obtained, providing evidence for the formation of Asx-turns, although very often in combination with other main chain–main chain or intermolecular interactions, which unfortunately blurred the inherent features; it transpired that the constitutive stability could only be demonstrated for the Asx-Pro sequence.[12] The authors also proposed, on the basis of theoretical considerations, that several types of Asx-turn might exist;[12] nonetheless, the assessment of the innate stability of these secondary features through molecular modelling using relevant methods ─ namely quantum chemistry ─ has not been addressed systematically to date. The formation of Asx-turns has been probed theoretically in a few studies using model peptides containing an Asn residue;[13] the turns were indeed present in stable conformations of these models, albeit invariably combined with other structural motifs. However, no experimental evidence for such structures was obtained, whether in solution[13a] or in the gas phase.[13b]

Over the last decade, gas phase isolation has been shown to provide useful insight into the folding properties of flexible molecules, including peptides, whether natural or synthetic.[14] Investigations conducted by a supersonic expansion followed by a laser study under isolated conditions provide invaluable information on the folding propensities of molecular systems. The secondary structures typical of protein chains ─ γ-turns, β-turns, β-sheets, β-hairpins, short 3_{10} helices ─ have all been detected in the gas phase using this approach, demonstrating the fundamental "folding" character of these features *in vacuo*. Comparison of experimental data (H-bond sensitive IR spectroscopy of each conformer together with relative abundances) with their theoretical counterparts, obtained at the relevant quantum chemistry level, enables cross-checking between the techniques used. So far, however, studies on the gas phase conformational preferences of Asn-containing peptides investigated using laser spectroscopy[13b, 15] have not provided experimental evidence for the formation of low energy structures based solely on an Asx-turn, presumably because the model peptide

sequences that were employed also bore H-bonding sites that enabled β-turn formation at the expense of Asx-turn structures.

The present work therefore aims at investigating the intrinsic folding behavior of an Asn residue by studying truncated Asn-containing peptide model sequences wherein the formation of β-turns or γ-turns is specifically excluded. Since the presence of a peptide bond at the N^{α}-position of the Asn residue would favor the formation of a C10 β-turn,[15] we reasoned that our objective could be achieved by replacing the Asn N^{α}-fragment with a benzyl group, while retaining an H-bond donor at the C-terminus in the form of a methylamide group. Two such sequences are addressed in this work, namely Asn-Ala and Asn-Gly, giving rise to the corresponding model compounds *N*-(2-benzylsuccinyl)alanine methylamide and *N*-(2-benzylsuccinyl)glycine methylamide (respectively referred to hereafter as Bn-NA (**1**) and Bn-NG (**2**); Figure 2). The benzyl group conveniently serves as the UV chromophore required by the laser spectroscopic techniques used.



1 (R = Me), Bn-NA
2 (R = H), Bn-NG

(R = Me), Me-NA
(R = H), Me-NG

Figure 2: *(left*) Model peptides **1** and **2** used for gas phase spectroscopy. *(right)* Model peptides used for turn typology evaluation by theoretical methods. Also indicated is the amide group numbering, and hence atom numbering, used in this paper.

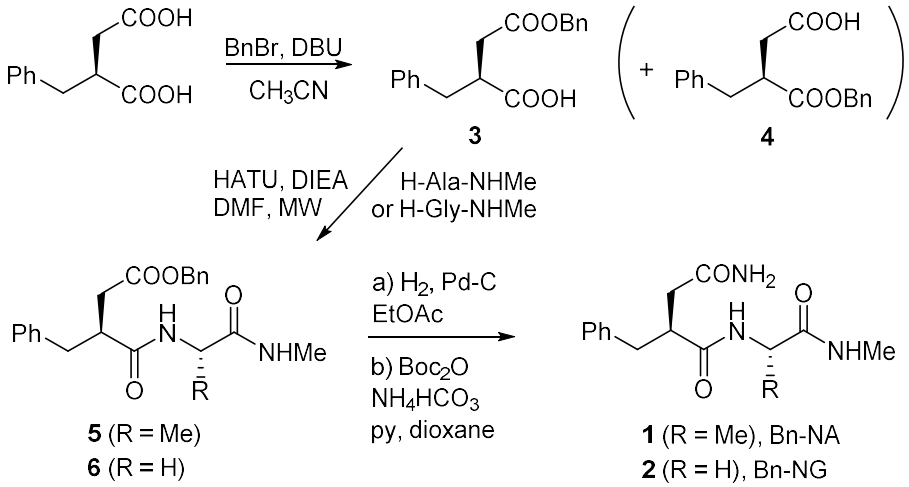
The experimental characterization of compounds **1** and **2** was carried out in the gas phase using UV and IR/UV double resonance spectroscopies. In parallel, the structures of these compounds and those of simplified analogs bearing a methyl group instead of the benzyl group ─ namely *N*-(2-methylsuccinyl)alanine methylamide and *N*-(2-methylsuccinyl)glycine methylamide (respectively referred to hereafter as Me-NA and Me-NG; Figure 2) ─ were investigated using relevant quantum chemistry techniques, with the objective of characterizing low-energy side chain–backbone Asx-turn structures and identifying their typology by analogy with the corresponding β-turns in a capped Ala-Ala or Ala-Gly sequence. These data were then used to interpret the experimental spectroscopic data and assign the experimentally-observed conformations. The H-bonds in the Asx-turn structures found using these procedures were also analyzed using a high-level theoretical characterization tool ─ namely Natural Bond Orbital (NBO) theory ─ which was recently applied successfully in an examination of amide–amide interactions in peptides.[16]

Finally, the observed stable structures were compared to the Asx-turn structures found in proteins, especially those combined with β-turns in which Asn occupies position *i* or *i+2*. The type of Asx-turn observed in proteins was found to depend upon the Asn environment, in particular the residue following Asn in the sequence, which leads us to propose different mechanistic roles for Asn-turns in the formation of the associated β-turn structures.

2. Materials and Methods

2.1. Synthesis

The synthesis of Bn-NA (**1**) and Bn-NG (**2**) is shown in Scheme 1. Treatment of commercial (*S*)-2-benzylsuccinic acid with benzyl bromide and DBU gave a ~2:1 mixture of the isomeric monoesters **3** and **4** which were separated chromatographically (38% and 16% isolated yields, respectively). Coupling of the former with the methylamide of either alanine or glycine (each prepared by deprotection of the corresponding Boc-protected derivative) using HATU/DIEA in DMF under microwave activation gave derivatives **5** and **6** (88% and 72% yields, respectively). Each of these derivatives was selectively deprotected by hydrogenolysis in the presence of a palladium catalyst and the resulting acid was converted into the corresponding carboxamide using di-*tert*-butyl dicarbonate and ammonium bicarbonate to furnish the target molecules **1** and **2**. While the yield for the last two steps was nearly quantitative for the alanine derivative, the transformation was considerably less efficient for the glycine derivative (33% yield), with reverse phase HPLC purification being deemed appropriate.



Scheme 1. Synthesis of the model peptides Bn-NA (**1**) and Bn-NG (**2**).

2.2. Quantum chemistry methods

A conformational search using the Macromodel Suite[17] was carried out on the Bn-NA and Bn-NG systems, wherein conformations were sought with a relative stability of up to 20 kJ·mol^{-1} with respect to the global minimum. Those conformations that mimicked any of the four types of β-turn (type I, type II, type I′, and type II′) were further submitted to a geometry optimization using quantum chemistry at the RI-B97-D3(BJ-abc)/def2-TZVPPD level of theory[18] using the Turbomole 7.3 suite.[19] The related compounds Me-NA and Me-NG were generated starting from any of the Bn-NA and Bn-NG conformations, substituting the phenyl ring by an H atom, then reoptimizing the structure. For all these conformations, harmonic vibrational frequencies were calculated at the same level of theory and were used to deduce the electronic energies corrected for zero-point vibrational energy ─ namely ΔH at 0 K and Gibbs free energies at 298 K ─ as well as the theoretical vibrational spectra.

The NH stretch IR spectra were generated from a mode-dependent scaling procedure, using linear scaling factors (see details in Supporting Information Section 3) deduced from a previous fitting on a similar set of peptides, which provides a good agreement with experimental frequencies (within 20 cm^{-1} for NH stretches in NH···O bonds).[14a] In the amide I and II regions, the scaling factors used (1.004 and 1.008, respectively) were chosen due to the good level of prediction they provided in previous studies.[20] [21]

Detailed NBO analyses,[22] which are based on an expression of the electronic wavefunction in terms of a localized Lewis-type chemical bond depiction, were also performed for selected relevant conformations in order to provide a quantification of the stabilizing role of electron delocalization through H-bonding effects. Analyses were carried out on the RI-B97-D3(BJ-abc)/def2-TZVPPD structures using the NBO module[23] of the Gaussian 16 software.[24] The NBO analysis was carried out at both HF/TZVPP and MP2/TZVPP levels. The donor and acceptor NBO occupancies considered in the following are those obtained at the MP2/TZVPP level whereas the HF level was used to calculate, for each donor NBO(*i*) and acceptor NBO(*j*), the E(2) stabilization energy through the second order perturbation theory:

$$E(2) = \Delta E_{ij} = q_i \frac{F(i,j)^2}{|\varepsilon_i - \varepsilon_j|}$$

where q_i is the donor orbital occupancy, $F(i\,j)$ is the off-diagonal NBO Fock matrix element and ε_i (resp. ε_j) is the diagonal element, *i.e.* the donor (acceptor) orbital energy; the threshold for considering the interactions as significant being equal to 0.2 $kJ \cdot mol^{-1}$. Previous works have shown that the MP2/TZVPP level of theory used here provides a good compromise between accuracy and computation times.[16, 25]

2.3 Gas phase laser spectroscopic methods

The UV and IR spectroscopy of each isolated compound was recorded using a setup and procedures that have been described earlier.[14] Briefly, each compound was vaporized by laser desorption from a pressed pellet containing the sample and graphite powder. The vapor was introduced into a supersonic jet of a He/Ne mixture (18 bars) generated by a pulsed valve (10 Hz). The jet was skimmed and the cooled molecules entered the source of a time-of-flight mass spectrometer (TOF-MS) where they interacted with the UV laser (frequency-doubled Continuum Nd:YAG-pumped Radiant dye laser, 1 mJ, operating at 10 Hz). The ions formed by resonant two-photon ionization were then accelerated towards the TOF-MS detector. The time-selected signal was measured and processed by an in-house LabView program. Conformation-specific IR spectra were recorded using the IR/UV double resonance technique, in which an IR laser sent before the UV laser depopulated the vibrational ground state level of the molecules.[26] The UV laser being tuned on a transition of a well-defined conformer, scanning the IR laser and measuring the depletion of the UV-induced ion signal enabled us to record conformer-selective IR spectra. The IR beam was generated by a Continuum Nd:YAG-pumped LaserVision OPO/OPA, operating between 3100 and 4000 cm^{-1}, whose idler beam was used for the NH stretch region. The amide I and II regions (1400-1800 cm^{-1}) were covered by frequency difference mixing between signal and idler using a dedicated $AgGaSe_2$ crystal module. For a better signal-to-noise ratio the IR-induced ion signal depletion was normalized to an IR-off ion signal, obtained from a second passage of the UV beam into the jet.[27]

3. Theoretical conformational landscape of Me-NA and Me-NG

Quantum chemistry methods were first used to determine the Asx-turn structures adopted by the simplified Me-NA and Me-NG sequences together with their stability, in order to characterize their typology by comparison with relevant β-turns, namely the capped peptides Ac-Ala-Ala-NHMe and Ac-Ala-Gly-NHMe.

Structures

For both Me-based sequences, geometry optimization at the quantum chemistry level revealed three types of Asx-turns, with limited differences between the two sequences. These turns were all characterized by a C10 H-bond bridging the Asn side chain carbonyl and the C-terminal peptide NH bond, but differed by the backbone torsions along the sequence, as indicated by the sets of dihedrals defined in Figure 2 and presented in Tables S1 and S2. Comparison of these structures with those of β-turns in the capped dipeptides (Figure 1; Tables S1 and S2), obtained at the same level of theory, showed a striking resemblance, although less marked for type I (Figure 3), which led us to label the Asx-turns according to the β-turns they mimic.

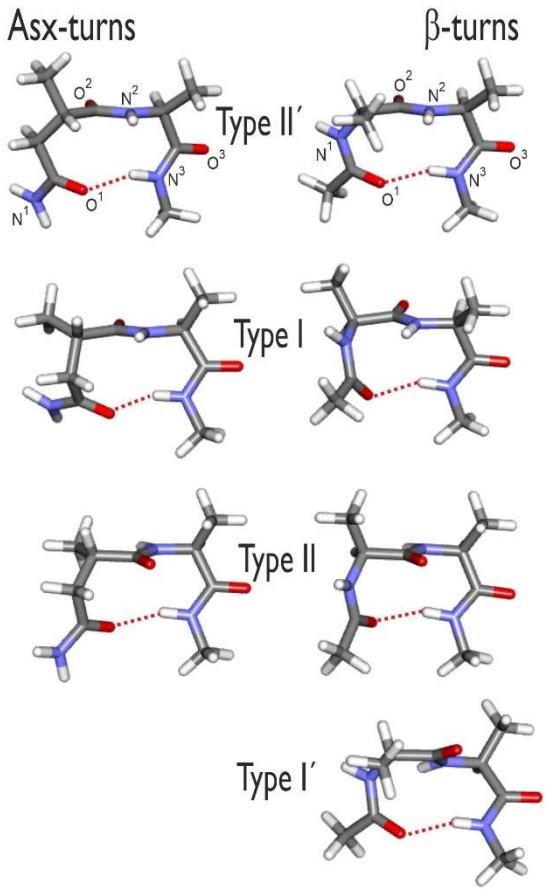


Figure 3: The three Asx-turns of Me-NA (*left*), compared to the four β-turn types of Ac-Ala-Ala-NHMe (*right*), as obtained by quantum chemistry at the DFT-D level of theory (B97-D3(BJ-abc)/def2-TZVPPD, see details in Methods section). The same types of structures are also found for the Me-NG compound and Ac-Ala-Gly-NHMe, respectively. The type I′ Asx-turn was not found as a stable conformation.

Energetics

In both Me-based sequences the most stable conformation was the type II′ turn (Table 1). It was followed by the types II and I, which were significantly higher in energy (more than 13 kJ·mol^{-1}, except for Me-NG for which the type II lowest form was only at 7 kJ·mol^{-1}). The energetic values were not drastically changed when temperature effects were taken into account due to the similar shapes of these forms. The type I' turn was not found to be a minimum in either sequence. Despite the structural resemblance, the energetic order of the structure types differs considerably between Asx-turns and β-turns: in the Ala-Ala sequence, the most stable β-turn is type I followed by type II', as expected from steric hindrance with α-substituted L-residues,[1d] in contrast to the Asx-turn model Me-NA, where the type I appears distorted (Table S1), and thus destabilized, relative to its β-turn counterpart due to a similar steric hindrance. The order is different with Ala-Gly where the most stable type II β-turn is challenged by type I, which is only 4 kJ·mol^{-1} higher in energy.

Table 1: Dependence of the energetics with the type of turn in the Asx-turn conformations of Bn-NA and Bn-NG, of Me-NA and Me-NG and of the β-turns formed by Ac-Ala-Ala-NHMe and Ac-Ala-Gly-NHMe. Zero-point energy corrected electronic relative energies (ΔH at 0 K) and Gibbs free energies (ΔG at 298 K), obtained at the B97-D3(BJ-abc)/def2-TZVPPD level of theory, are given in kJ·mol^{-1}. For each molecule, the values are given relative to the most stable conformation and types are ordered according to their relative stability at 298 K.

	ΔH (0 K) (kJ·mol^{-1})	ΔG (298 K) (kJ·mol^{-1})		ΔH (0 K) (kJ·mol^{-1})	ΔG (298 K) (kJ·mol^{-1})
Asx-turns Bn-NA			Asx-turns Bn-NG		
type II′ t	0	0	type II′ t	0	0
type II′ g+	6.3	3.8	type II′ g+	4.0	3.5
type II′ g-	13.4	13.5	type II' g-	11.8	13.9
type I g+	16.7	18.3	type II g+	8.3	8.9
type I t	20.8	23.2	type II t	12.9	14.4
type I g-	21.1	27.1	type II g-	25.3	28.1
type II g+	21.5	22.6	type I g+	15.4	20.3
type II t	24.4	26.8	type I t	18.5	22.2
type II g-	38.0	40.8	type I g-	19.1	25.8
Asx-turns Me-NA			Asx-turns Me-NG		
type II′	0	0	type II′	0	0
type I	13.6	15.5	type II	7.1	7.1
type II	17.1	20.3	type I	13.8	16.6
β turns Ac-Ala-Ala-NHMe			β turns Ac-Ala-Gly-NHMe		
type I	0	0	type II	0	0
type II′	6.7	9.0	type I	3.6	4.8
type II	7.5	10.5	type I′	10.3	12.6
type I′	17.5	21.1	type II′	9.8	13.1

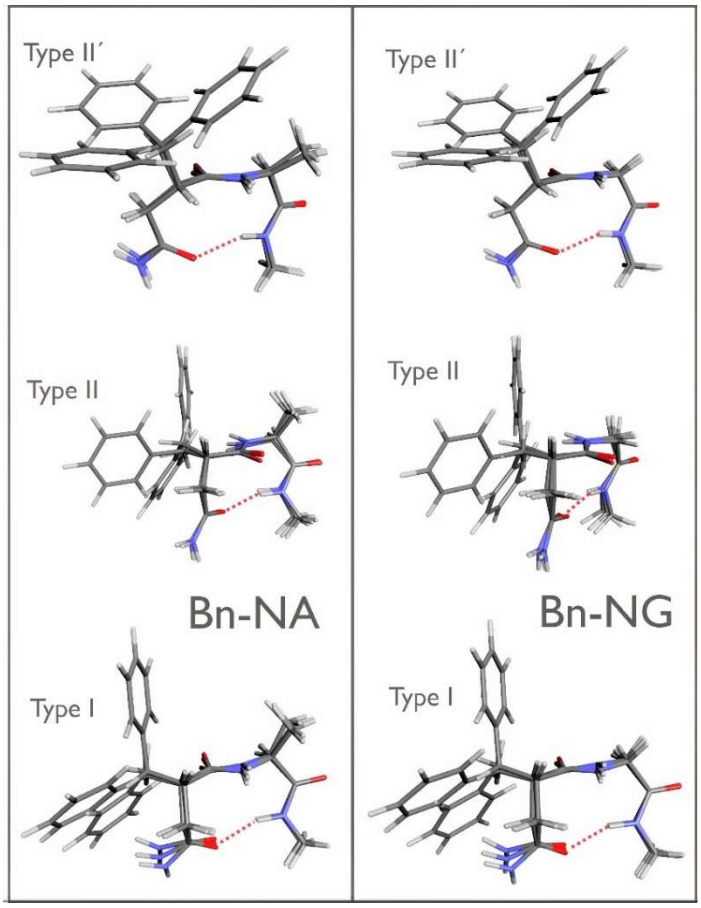


Figure 4: Quantum chemistry optimized structures of the Asx-turns formed by Bn-NA (*left*) and Bn-NG (*right*) mimicking the respective β-turn conformations. For each turn type, the three rotamers of the benzyl moiety are displayed simultaneously, with their backbone structures overlaid. Dotted lines indicate the C10 H-bond.

Effect of the benzyl group

The presence of three rotamers of the benzyl group for each turn type (Figure 4) introduced a spread in the energetic distribution, due to the dependence of the interactions between the benzyl moiety and the backbone upon the orientation of the former. The energetic ordering between the types was respected, however, with the type II' rotamers being the most stable in most cases; the exception was the Bn-NG molecule where the highest type II' rotamer was challenged by the lowest type II form. These calculations qualified the Bn-NA and Bn-NG molecules as good models to observe isolated Asn-turns in the gas phase.

4. Gas phase characterization of Bn-NA and Bn-NG

The excitation spectra of the two peptide model systems were recorded using one-color R2PI (Figure 5). The sharp peaks attested to the efficient cooling of the molecules in the jet. The excitation spectra of Bn-NA and Bn-NG each showed one intense band in the red part of the spectrum, accompanied by several weaker bands which can be *a priori* assigned to either vibronic features of the same conformation or to other conformations populated in the jet. For further assignment, IR spectra were recorded using the ion depletion technique.

IR/UV spectra were recorded for Bn-NA and Bn-NG in both NH stretch, amide I, and amide II regions (Figure 6, Table S3). For each molecule, the IR NH stretch spectra recorded for the labelled UV bands of Figure 5 were all identical (See Figs S1 and S2), demonstrating the existence of a single predominant conformation populated in the jet.

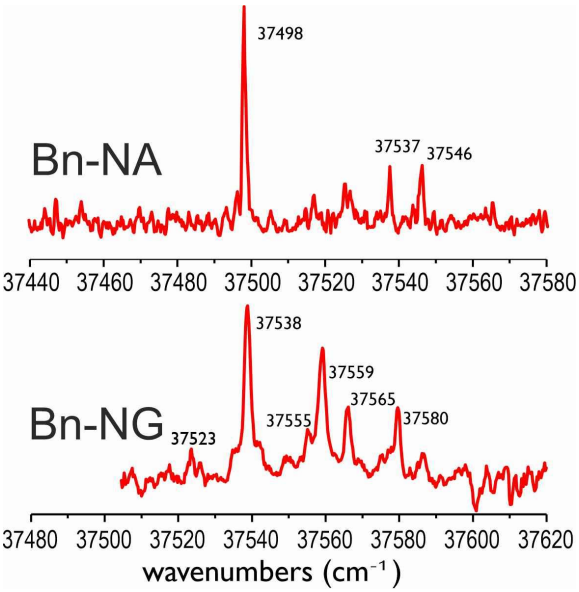


Figure 5: Near-UV mass-selected resonant two-photon ionization spectrum of the Bn-NA and Bn-NG molecules, in the spectral region of the origin of the first ππ* transition of the benzyl moiety.

Moreover, when the IR spectra of Bn-NA and Bn-NG were compared, they were found to be qualitatively similar in the NH stretch region, although some differences were observed in the amide I and II ranges.

The assignments were first guided by general spectroscopic rules and confirmed by theoretical calculations.[14] The free amide NH stretch absorptions are expected to appear in the 3450-3500 cm^{-1} range, while weak interactions like a C5 or an NH-π interaction can cause this feature to red-shift to 3420-3450 cm^{-1}.[28] A broad and intense peak that was significantly red-shifted to around 3400-3420 cm^{-1} was an indication of the presence of a hydrogen bond. The coupled NH oscillators of a free C-terminus carboxamide group were expected to give rise to a doublet with a splitting of 117 cm^{-1}, with the blue asymmetric stretch appearing around 3500-3560 cm^{-1}. [29]

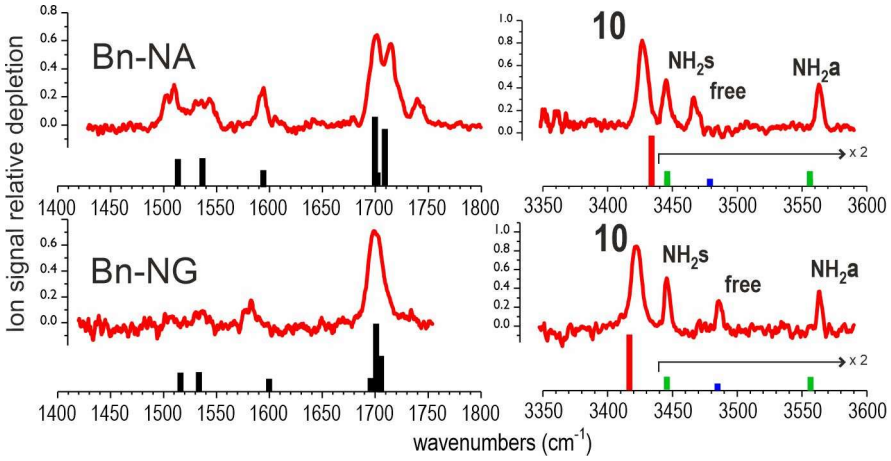


Figure 6: NH stretch (*right*) and amide II and II (*left*) regions of the IR spectrum of the Bn-NA (upper panel) and Bn-NG (lower panel) molecules, as obtained from IR/UV double resonance spectroscopy, using their most intense UV band as a probe. Absorption spectra appear as depletions, plotted as positive signals. In the lower part of each panel the theoretical spectrum of the most stable form calculated for each molecule is given, namely the type II' *trans* form. In the NH stretch region, the assignment of the NH stretch bands is color-coded: green for the symmetric and antisymmetric components of the Asn NH_2 group, blue for $NH_{Ala/Gly}$ and red for the C-terminus NHMe group.

On this basis, a primary assignment of the bands observed in the IR/UV spectra could be proposed. The broad and red-shifted feature at around 3420 cm^{-1} in the IR spectra of both compounds indicated the presence of a hydrogen-bonded NH group. The doublet formed by the peaks at 3445 cm^{-1} and 3562 cm^{-1} was assigned to the symmetric and asymmetric stretches respectively of a free C-terminus carboxamide group, in agreement with previously observed free or nearly carboxamide group.[15, 30] The band at 3465 cm^{-1} (3485 cm^{-1}) in Bn-NA (Bn-NG) indicates a free NH oscillator. All these features were in full agreement with expectations for an Asx-turn.

A more refined assignment, both in terms of Bn rotamer and turn type, was carried out by comparison with calculated IR spectra, taking into account the calibrations carried out in previous studies.[31] In order to assign the type of turn in the observed conformer, we compared the experimental IR spectral data with the computed IR spectra of the various conformations (Figure 6 and Figures S4-S5). An overall consideration of the calculated spectra for the most stable types of turn showed that the amide I region was significantly dependent upon the turn type and was therefore able to provide a good indication as to the structure, when combined with energetic data. The same spectral considerations (Figures S4 and S5) showed that rotamers of the same turn type (which exhibit similar backbones) present similar IR spectra in the amide I and II regions, whereas the amide A region showed marked differences, especially in terms of red shift ─ and hence strength ─ of the C10 H-bond. This property illustrates the flexibility of the Asx-turn feature and the sensitivity of its strength to its environment, and it was used advantageously to support the assignments made from the amides I and II. A careful inspection of Figs S4 and S5 shows that type II' conformations provide a good fit with the experimental data in the amide I region, with a two-component CO stretch motif for Bn-NA (Figure 6 top) and a broad band exhibiting a blue shoulder for Bn-NG (Figure 6 bottom). In both cases, the most stable rotamer (*trans*) was the one that best accounted for the amide A spectrum, in particular the C10 band position. For these reasons, we confidently assign the forms observed to type II' Asx-turns, with a *trans* Bn rotamer orientation.

5. Discussion of the isolated structures in the gas phase

The good agreement between gas phase experimental data and theory validated the theoretical methods used for both the structure and the energetics and provided evidence for the formation of an Asn(*i*) carboxamide C=O···NH(*i*+2) C10 H-bond and hence for an Asx-turn structure in Bn-NA and Bn-NG. The H-bond strengths, as judged from spectral shifts (Figure 6) or intermolecular distances (Table 2), are similar to those found in comparable β-turns,[32] albeit slightly diminished. The C10 H-bond strengths remain relatively modest within the H-bonding scale of peptides; the C7 H-bonds of γ-turns, for example, are much stronger. This was confirmed by NBO analysis of the related Me-NA molecule, which showed that the stabilization induced by the type II' Asx-turn formation is of the order of 16 $kJ \cdot mol^{-1}$ for an intermolecular distance of 208 pm, whereas the corresponding values for a regular type II' β-turn are 21 $kJ \cdot mol^{-1}$ and 203 pm, respectively. Despite being characterized by a similar geometrical disposition, the H-bond in the Asx-turn is weaker and has a larger intermolecular distance, epitomizing the frustration of the H-bonding due to the constraints imposed by the covalent backbone along the Asn side chain and the main chain around Ala. Interestingly, the opposite conclusion is drawn for type II turns (Table 2), where the frustration is greater in the regular β-turn (stabilization

energy of 16 kJ·mol^{-1} vs. 21 kJ·mol^{-1} in the Asx-turn), demonstrating the dependence of the constraints on the turn type.

Besides C10 H-bonding, other (minor) interactions have been evoked previously that may further stabilize a β-turn structure: i) an NH$\cdots\pi_{amide}$ interaction,[20, 33] where the N^{3}H bond not only provides the donor for the C10 H-bond but also interacts with the π system of the central amide of the turn, and ii) a carbonyl–carbonyl interaction[34] involving the N-terminus and the central amide of a turn. The same NBO analysis that was used to assess H-bonding was also employed to quantify these interactions in Asx-turns as well as in their β-turn counterparts (Table 2). The first type of interaction, corresponding to the $\pi_{amide\,2} \rightarrow \sigma^*N^3H$ NBO interaction, was found to be modest, at around 1 kJ·mol^{-1}; the second was not detected at all and was therefore lower than the NBO analysis threshold value of 0.2 kJ·mol^{-1}.

Table 2: Interactions between donor (*i*) and acceptor (*j*) NBOs and corresponding stabilization energies (noted E*ij*$^{(2)}$, given in kJ·mol^{-1}) in Me-NA Asx-turn types II' and II and in their Ac-Ala-Ala-NHMe β-turn counterparts. Several NBOs are found as active donors contributing to the σ^*N^3H NBO of the NH group for the C10 H-bond: the lone pairs (lp(1) and lp(2)) of the carbonyl O^1 atom and the carbonyl C^{1}O^1 bonds (π and marginally σ) belonging to the Asn side chain in Asx-turns or to the acetyl group in the β-turns. An interaction between the NBO located on the lp(1) of the N^2 atom of the central amide group and the σ^*N^3H NBO, made possible by the short N^{3}H$\cdots$N^2 distance, is also found but remains negligible compared to the C10 H-bonding.

		MeNA type II'	MeNA type II	Ac-Ala-Ala-NHMe type II'	Ac-Ala-Ala-NHMe type II
		N^{3}H ··· O^{3}C C10 H-bond			
E*ij*(2)	lp(1)O^1 → σ*N^{3}H	8.2	11.3	12.8	8.2
(kJ·mol^{-1})	lp(1)O^1 → σ*N^{3}H	6.6	2.0	4.3	5.9
	lp(2)O^1 → σ*N^{3}H	1.0	7.3	4.0	2.4
	σC^{1}O^1 → σ*N^{3}H			0.3	
Sum E*ij*(2) (kJ·mol^{-1})		15.8	20.6	21.4	16.5
O$^1\cdots$HN^3 distance (pm)		208.5	204.7	202.6	208.6
		N^{3}H→ π_{amide2} H-bond			
E*ij*(2) (kJ·mol^{-1})	lp(1)N^2 → σ*N^{3}H	1.6	1.0	1.1	1.0
N$^2\cdots$HN^3 distance (pm)		233.1	236.3	235.0	236.5

6. Comparison with crystallized protein data

Abundance of Asx-turns

A statistical search within a set of 430 proteins was carried out using the Motivated Proteins on-line facility.[35] It revealed 1319 Asx-turns, leading to an average number of 3.1 Asx-turns per protein. The type distribution of these Asx-turns was found to be as follows: type II′, 67 %; type I, 20 %, type II, 13 %, while type I′ was marginal at only 0.2%. This was in line with the statistical findings of Duddy et al.[11] When particularized to "Asn-turns" (wherein Asx is represented explicitly by Asn) these abundances remained similar, namely: type II′, 64 %; type I, 26 %, type II, 12 %, and type I′ only 0.3 %. It is of note that these abundances were in accordance with the

relative stability of the Asx-turns of the Asn-Ala peptide models (as listed in Table 1): the type II′ Asx-turn was found to be the most stable Asx-turn conformation, followed by type II and type I conformations; type I′ was not found as a stable conformation. This correlation between protein crystallographic data and the calculated gas phase stabilities clearly suggests that the abundance of the type II′ Asx-turn in proteins may be explained by its intrinsic stability.

Dependence upon the sequence

A detailed analysis of the nature (Ala or Gly) of the residue adjacent to Asn (*i.e.* following Asn in the sequence) in the Asn-turn revealed some significant differences. Figure 7 shows the type distribution of Asn-turns in Asn-Ala and Asn-Gly sequences resulting from the Motivated Proteins facility search. Whereas the Asn-Ala sequence was essentially associated to types II' and I, in line with the general trend mentioned above when the sequence was not particularized, the Asn-turns with a Asn-Gly sequence were predominantly of type II. This suggests that the underlying mechanism at play in the appearance of Asn-turns in Asn-Gly sequences differs radically from the other Asn-Xaa sequences, where Xaa stands for any residue other than Gly.

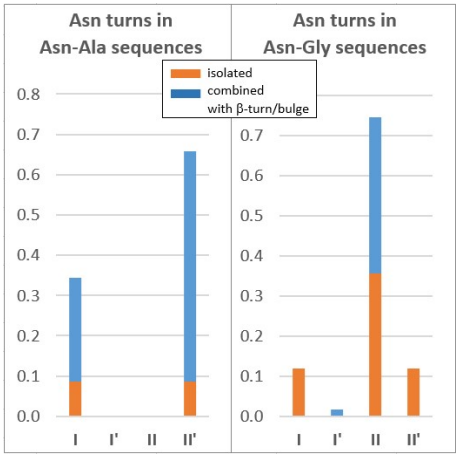


Figure 7: Distribution of the Asn-turn types for two adjacent residues of Asn in the Asn-turn, namely Ala and Gly, as obtained from a direct examination of the turns detected by the Motivated Proteins facility. The character of the Asn-turns (isolated vs. combined with β-turn/bulge structures) is indicated by a color code (orange vs. blue, resp.).

Overlap with other secondary structures

The same analysis showed that many of these Asn-turns actually partially overlapped with another turn structure, namely a β-turn or a β-bulge (in blue in Figure 7). However, as noted above for the type of Asn-turn, the overlap with these β-turn or β-bulge structures and the fraction of Asn-turns combined with these structures was also found to depend dramatically on the nature of the residue following Asn in the sequence.

When the residue following Asn in an Asn-turn is alanine, this gives rise mainly to type II' and I Asn-turns; most of these turns (83%) are combined with a type I β-turn (Figure 7) and in these cases the Asn-Ala sequence occupies the *i, i+1* positions of the β-turn (Figure 8). This turn combination has been described as an Asx-β-turn by Milner-White and coworkers.[6] Our present studies now show that it comes with two structural variants, which differ by the orientation of the Asn side chain (and hence the type of Asx-turn).

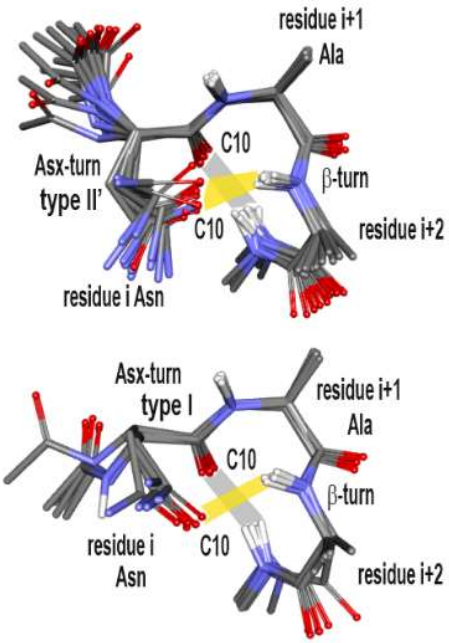


Figure 8: Asn-turns in Asn-Ala sequences combined with a type I β-turn, as found in the PDB partial survey carried out using Motivated Proteins. O and N peptide atoms are colored in red and blue, respectively. Unless present in high resolution PDB structures, H atoms of the amide groups have been added on a force field basis. Two types of Asn-turns are observed, type II' (*top*) and type I (*bottom*). For each type the Asn- and β-turn backbones have been overlaid. The yellow bands indicate the C10 H-bond of the Asn-turn and the grey ones indicate the C10 H-bond of the β-turn. Errors in the Asn side chain orientation deduced from the X-ray data analysis cause a few Asn carbonyl O atoms to appear as colored in blue.

In contrast, when the residue following Asn in an Asn-turn is glycine, a radically different picture emerges. The majority (59 %) of the Asn-turns in these Asn-Gly sequences are found to be isolated, not combined with any nearby secondary structure, and type II is most prominent (Figs. 7 and 9). When the Asn-turn is in combination with a β-turn or a β-bulge (*i.e.* in the remaining 41% of the Asn-turns), the Asn-Gly sequence is found to occupy positions *i+2, i+3* of the β-turn/β-bulge and to adopt nearly exclusively a type II Asx-turn structure (Figure 10), which has never been detected before, to the best of our knowledge.

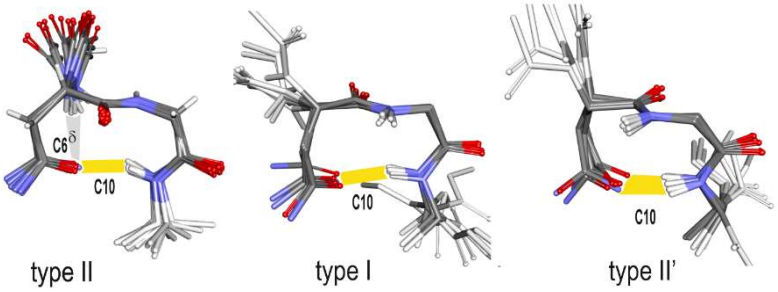


Figure 9: Isolated Asn-turns in Asn-Gly sequences, as found in the PDB survey carried out using Motivated Proteins. O and N peptide atoms are colored in red and blue respectively. For each main type of Asn-turn observed, the backbones are overlaid. The yellow bands indicate the C10 H-bond of the Asn-turn, together with, in the type II case, an elongated intraresidue $C6^{\delta}$ H-bond (NH···O(carboxamide) distance ~260 pm) bridging the NH bond of the previous residue and the Asn side chain carbonyl.

It transpires that the type distribution in proteins of Asx-turns implicating Asn depends significantly on the nature of the adjacent residue (Figure 7).

The Asn-Ala sequence is representative of the general behavior observed for all Asn-Xaa sequences, as mentioned at the beginning of this section. The fact that the Asn-turns observed in proteins correspond to the most stable intrinsic structures found in our model studies, and that their abundance follows the energetic order, vivify two extreme-case localized mechanisms: i) Asn-turns may form spontaneously and play the role of stable intermediate structures serving as precursors of β-turns, or ii) once a β-turn is formed it can be stabilized by an Asn-turn, provided that an Asn residue occupies the position *i* of the β-turn. This would simply arise from the fact that the Ramachandran dihedrals at position *i+1* imposed by the formation of a β-turn roughly match those corresponding to the second residue of an intrinsic Asn-turn. Table S1 provides a clear example: position *i+1* occupied by an Ala residue in a type I β-turn gives rise to (φ_1, ψ_1) = (–72°, –16°), not excessively different from the dihedrals of an Ala residue in position *i+2* of an Asn-turn : (φ_2, ψ_2) = (–91°, 1°) and (–112°, 28°), for types II' and I Asn-turns respectively. Both mechanisms might be active in turn formation. Asn residues may indeed be quite effective in backbone dehydration[36] since they can replace the hydrating water molecule attached to a backbone NH site and simultaneously accommodate it, due to the capacity of their side chain to serve as both an H-bond donor and acceptor.

Asx-turns implicating an Asn-Gly sequence present a quite different portfolio. Firstly, the ubiquity of isolated Asn-turns of all types (Figure 7) suggests that these structures can form spontaneously, presumably aided by flexibility in the absence of a side chain on the glycine residue, a postulate which was made previously to account for the large probability of finding Gly in the *i+3* position of a β-turn.[5c] Secondly, the most frequent type in the isolated turns, type II, is not the most intrinsically stable form (Table 1), which suggests that the formation mechanism does not rely on innate properties but instead depends on additional interactions, for example with the water molecules that solvate the Asn side chain. Thirdly, Asx-turns in Asn-Gly sequences are also observed in combination with a type I β-turn or β-bulge (which also contains a β-turn). In this case, the Asx-turn is almost exclusively of type II, which fits well the structural arrangement of the β-turn/bulge (Figure. 10) and of its network of hydration molecules.

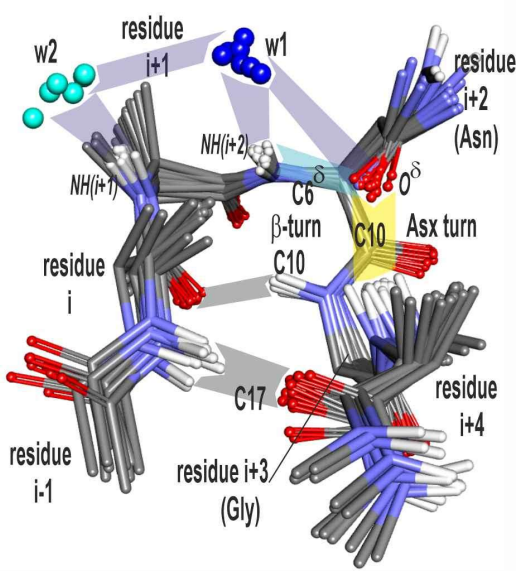


Figure 10: Asn-turns in Asn-Gly sequences found in combination with type I β-turns or β-bulges, as found in the PDB survey carried out using Motivated Proteins. For both structures (β-turns and β-bulges) the central parts of the structures were overlaid (see Section S4.1 for separate plots of turns and bulges). When detected (in high resolution structures of the PDB), O atoms of the hydration waters (labelled w1 and w2) were also indicated (blue and cyan balls, resp.). The H-bonds present in the structure are identified by a color code: the C10 H-bond of the Asn-turn in yellow; bands show the β-turn C10 H-bonds and C17 H-bonds in β-bulges in grey, the hydration H-bond network in blue, and the elongated $C6^{\delta}$ H-bond linking the central amide to the Asn side chain in cyan. The NH···O distances of the C10 and C17 bonds are of the order of 200 pm, as well as the NH(*i+1*)···w2 and NH(*i+2*)···w1 bonds. The distance between water clusters is of the order of 300 pm, as well as the w1···O(carboxamide) distance, providing evidence for hydration of the backbone NH(*i+1*) and NH(*i+2*) groups together with the O(carboxamide) atoms of the Asn side chain.

This observation suggests that a type II Asx-turn in a Asn-Gly sequence may specifically foster the formation of a hydrated β-turn/bulge structure, with a C10 H-bond located just *before* the Asn-Gly sequence in the peptide chain. This postulate is supported by the fact that the Asn ψ Ramachandran dihedral (~0°) in the sequences of Figure 10 and in the type II Bn-NA and Me-NA models of Table S1 fits the ψ_{i+2} value required for the formation of a type I β-turn (see for instance Ac-Ala-Ala-NHMe in Table S1). A previous study by us,[15a] in which Asn-containing β-turns in both isolated model peptides and crystallized proteins were compared, has been reexamined (Section S4.2) and further supports this contention.

Interestingly, these peculiarities of the Asx-turns in Asn-Gly sequences echo the specific abundances of Asn (and Asp) residues in β-turns, as collected by data mining studies,[5c, 7b] in particular in types I and I' β-turns. Regarding type I turns, Asx residues exhibit an overrepresentation (by a factor of 2-3) in positions *i* and *i+2*. More precisely, correlations of residue presence in a type I β-turn, detected using an on-line survey facility,[37] showed that when Asn occupies an *i+2* position in a type I β-turn, Gly as the residue in position *i+3* is overrepresented by a factor of 3.5 in comparison with its average abundance in proteins. Conversely, in the same type I β-turns, when Asn is in position *i*, Gly in position *i+1* is underrepresented by a factor of 4.2, well beyond the average penalty that represents a flexible Gly residue in position *i+1* of a β-turn (characterized by an underrepresentation of 2.3, taking the probability of having a Gly residue in position *i+1*, whatever the residues in the other position and normalizing it to the average abundance of Gly in proteins).

A similar specificity also seems to occur in type I' β-turns. The Asn-Gly sequence is indeed known to be overrepresented in positions *i+1, i+2* of these turns in proteins[5c, 7b] and to foster the folding of short peptides into type I' β-turn structures in solution.[13a, 38] It is interesting to note that the Ramachandran dihedrals of a type I' β-turn (for example see Table S1 in Supp. Info.: (φ_1, ψ_1) = (62°, 30°) and (φ_2, ψ_2) = (95°, -11°), for an Ala-Ala sequence Table S1) coincide to those of a type II Asx-turn (ψ_{Asn} = 3° and (φ_{Gly}, ψ_{Gly}) = (94°, -12°)), suggesting

that interconversion from the latter to the former occurs mainly through a simple torsion of the φ_{Gly} dihedral, presumably mediated by the water environment.[39]

These observations are entirely consistent with the hypothesis made above whereby, once an Asn-Gly sequence folds into an Asn-turn, it has a great probability to locally induce a β-turn, either of type I, where the Asn residue occupies the position *i+2*, or of type I', with Asn in the *i+1* position.

7. Conclusion

The present work reports the synthesis of Asn-containing peptide models that were specifically designed to inhibit the formation of a β-turn and therefore allow intrinsic folding into an Asx-turn. Two model systems were considered, which differed by the identity of the residue following Asn in the sequence, namely Ala and Gly. A theoretical analysis showed that the most stable Asx-turns mimic a type II' β-turn, whatever the next neighbor residue of Asn. The gas phase experiments, carried out in a supersonic expansion, provided spectroscopic evidence for the spontaneous folding of the model compounds into the most stable form, namely a type II' Asx-turn, confirming the validity of the theoretical approach.

A survey of Asn-turns in PDB proteins was also carried out. It confirmed earlier data, which showed that Asn-turns are often associated with a β-turn where Asn occupies the position *i*, the whole structure being referred to as a Asx-β-turn. In this case the Asx-turn is mainly of type II', corresponding to its most stable innate structure. When the next neighbor residue along the sequence is a Gly, however, the survey showed that i) a majority of Asx turns are found as isolated secondary structures, ii) most of them are of type II, *i.e.* differ from that observed in the gas phase, and iii) when associated to another structure, this latter is a type I β-turn/β-bulge type, where Asn exclusively occupies the position *i+2* in the turn. The structures observed in crystallized proteins suggest that the Asn-Gly sequence exhibits a propensity to form a type II Asn-turn, which is therefore not controlled by intrinsic features but presumably by other factors such as hydration. In all cases, Asn-turns might be efficient promotors of type I β-turns and β-bulges. Finally, the turn-controlling behavior of Asn is expected to be shared by the aspartic acid residue, whose side-chain carboxylate function (in physiological conditions) exhibits similar properties to the carboxamide group of asparagine, namely good H-bond accepting behavior, in line with the resemblance of their codons in the genetic code (which only differ by their first letter), which tends to minimize the harmful effects of Asp/Asn mutations.

Acknowledgements

Support from the French National Research Agency (ANR; Grant ANR-17-CE29-0008 "TUNIFOLD-S") and from the "Investissements d'Avenir" Funding program (LabEx PALM; grant ANR-10-LABX-0039-PALM; DIRCOS) are acknowledged. This work was granted access to the HPC facility of [TGCC/CINES/IDRIS] under the Grants 2020-A0070807540 and 2021-A0090807540 awarded by GENCI (Grand Equipement National de Calcul Intensif) and to the CCRT High Performance Computing (HPC) facility at CEA under the Grants CCRT2020-p606bren and CCRT2021-p606bren.

Conflict of Interest

The authors declare no conflict of interest.

Data availability statement

Details of organic synthesis, quantum chemistry, UV and IR spectroscopy and data mining are provided in a Supporting Information file available on the web. Additional data that support the findings of this study are available from the corresponding authors upon reasonable request.

Keywords

Conformation analysis · Peptides · Asx-Turn · Laser spectroscopy · Quantum chemistry

References

[1] a) P. Y. Chou, G. D. Fasman, *J. Mol. Biol.* **1977**, *115*, 135-175; b) J. A. Smith, L. G. Pease, *CRC Crit. Rev. Biochem.* **1980**, *8*, 315-399; c) E. N. Baker, R. E. Hubbard, *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97-179; d) G. D. Rose, L. M. Gierasch, J. A. Smith, *Adv. Protein Chem.* **1985**, *37*, 1-109; e) A. M. C. Marcelino, L. M. Gierasch, *Biopolymers* **2008**, *89*, 380-391.

[2] a) P. Y. Chou, G. D. Fasman, *Biophys. J.* **1979**, *26*, 385-399; b) H. J. Hsu, H. J. Chang, H. P. Peng, S. S. Huang, M. Y. Lin, A. S. Yang, *Structure* **2006**, *14*, 1499-1510; c) J. Lee, V. K. Dubey, L. M. Longo, M. Blaber, *J. Mol. Biol.* **2008**, *377*, 1251-1264.

[3] A. G. de Brevern, *Sci. Rep.* **2016**, *6*, 33191.

[4] C. M. Wilmot, J. M. Thornton, *J. Mol. Biol.* **1988**, *203*, 221-232.

[5] a) J. S. Richardson, J. A. Tainer, D. C. Richardson, *Biophys. J.* **1980**, *32*, 211-213; b) D. C. Rees, M. Lewis, W. N. Lipscomb, *J. Mol. Biol.* **1983**, *168*, 367-387; c) E. G. Hutchinson, J. M. Thornton, *Protein Sci.* **1994**, *3*, 2207-2216; d) D. R. Wilson, B. B. Finlay, *Protein Eng.* **1997**, *10*, 519-529; e) N. Eswar, C. Ramakrishnan, *Protein Eng.* **1999**, *12*, 447-455; f) N. Eswar, C. Ramakrishnan, *Protein Eng.* **2000**, *13*, 227-238.

[6] W. Y. Wan, E. J. Milner-White, *J. Mol. Biol.* **1999**, *286*, 1633-1649.

[7] a) P. Y. Chou, G. D. Fasman, *Biophys. J.* **1979**, *26*, 367-383; b) K. Guruprasad, S. Rajkumar, *J. Biosci.* **2000**, *25*, 143-156; c) J. M. Anderson, B. Jurban, K. N. L. Huggins, A. A. Shcherbakov, I. Shu, B. Kier, N. H. Andersen, *Biochemistry* **2016**, *55*, 5537-5553.

[8] E. J. Milner-White, R. Poet, *Biochem. J* **1986**, *240*, 289-292.

[9] a) E. J. Milner-White, R. Poet, *Trends Biochem. Sci.* **1987**, *12*, 189-192; b) B. L. Sibanda, T. L. Blundell, J. M. Thornton, *J. Mol. Biol.* **1989**, *206*, 759-777; c) A. W. E. Chan, E. G. Hutchinson, D. Harris, J. M. Thornton, *Protein Sci.* **1993**, *2*, 1574-1590; d) J. D. Watson, E. J. Milner-White, *J. Mol. Biol.* **2002**, *315*, 171-182.

[10] B. L. Sibanda, J. M. Thornton, *J. Mol. Biol.* **1993**, *229*, 428-447.

[11] W. J. Duddy, J. W. M. Nissink, F. H. Allen, E. J. Milner-White, *Protein Sci.* **2004**, *13*, 3051-3055.

[12] A. Abbadi, M. McHarfi, A. Aubry, S. Premilat, G. Boussard, M. Marraud, *J. Am. Chem. Soc.* **1991**, *113*, 2729-2735.

[13] a) Y. K. Kang, I. K. Yoo, *Biopolymers* **2016**, *105*, 653-664; b) E. J. Cocinero, E. C. Stanca-Kaposta, D. P. Gamblin, B. G. Davis, J. P. Simons, *J. Am. Chem. Soc.* **2009**, *131*, 1282-1287.

[14] a) E. Gloaguen, M. Mons, *Top. Curr. Chem.* **2015**, *364*, 225-270; b) E. Gloaguen, M. Mons, K. Schwing, M. Gerhards, *Chem. Rev.* **2020**, *120*, 12490–12562.

[15] a) S. Habka, W. Y. Sohn, V. Vaquero-Vara, M. Geleoc, B. Tardivel, V. Brenner, E. Gloaguen, M. Mons, *Phys. Chem. Chem. Phys.* **2018**, *20*, 3411-3423; b) K. N. Blodgett, J. L. Fischer, J. Lee, S. H. Choi, T. S. Zwier, *J. Phys. Chem. A* **2018**, *122*, 8762-8775.

[16] V. Brenner, E. Gloaguen, M. Mons, *Phys. Chem. Chem. Phys.* **2019**, *21*, 24601-24619.

[17] Macromodel, Schrödinger Release 2019-2013, Schrödinger, LLC, New York, NY.

[18] a) K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, *Theor. Chem. Acc.* **1997**, *97*, 119-124; b) S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104; c) D. Rappoport, F. Furche, *J. Chem. Phys.* **2010**, *133*, 134105.

[19] Turbomole V7.2 A Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, Turbomole GmbH, since 2007; available from http://www.turbomole.com. **2017**

[20] Z. Imani, V. R. Mundlapati, G. Goldsztejn, V. Brenner, E. Gloaguen, R. Guillot, J. P. Baltaze, K. Le Barbu-Debus, S. Robin, A. Zehnacker, M. Mons, D. J. Aitken, *Chem. Sci.* **2020**, *11*, 9191-9197.

[21] G. Goldsztejn, V. Mundlapati, V. Brenner, E. Gloaguen, M. Mons, I. León, C. Cabezas, J. L. Alonso, *Phys. Chem. Chem. Phys.* **2020**, *22*, 20284-20294.

[22] a) I. V. Alabugin, G. dos Pasos Gomes, M. A. Abdo, *WIREs Comput. Mol. Sci.* **2019**, *9*, e1389; b) I. V. Alabugin, K. M. Gilmore, P. W. Peterson, *WIREs Comput. Mol. Sci.* **2011**, *1*, 109-141; c) A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **1988**, *88*, 899-926; d) F. Weinhold, *J. Comput. Chem.* **2012**, *33*, 2363-2379.

[23] F. Weinhold, A. E. Reed, J. E. Carpenter, E. D. Glendening, NBO version 3.1.

[24] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. R. Gomperts, B. B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, N. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 16, **2016**, Gaussian, Inc., Wallingford CT.

[25] E. Gloaguen, V. Brenner, M. Alauddin, B. Tardivel, M. Mons, A. Zehnacker-Rentien, V. Declerck, D. J. Aitken, *Angew. Chem. Int. Ed.* **2014**, *53*, 13756-13759.

[26] A. M. Rijs, J. Oomens in *IR Spectroscopic Techniques to Study Isolated Biomolecules*, *Vol. 364* Eds.: A. M. Rijs and J. Oomens), **2015**, pp. 1-42.

[27] E. Gloaguen, H. Valdes, F. Pagliarulo, R. Pollet, B. Tardivel, P. Hobza, F. Piuzzi, M. Mons, *J. Phys. Chem. A* **2010**, *114*, 2973-2982.

[28] W. Y. Sohn, V. Brenner, E. Gloaguen, M. Mons, *Phys. Chem. Chem. Phys.* **2016**, *18*, 29969-29978.

[29] W. Chin, F. Piuzzi, I. Dimicoli, M. Mons, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1033-1048.

[30] P. S. Walsh, J. C. Dean, C. McBurney, H. Kang, S. H. Gellman, T. S. Zwier, *Phys. Chem. Chem. Phys.* **2016**, *18*, 11306-11322.

[31] H. S. Biswal, Y. Loquais, B. Tardivel, E. Gloaguen, M. Mons, *J. Am. Chem. Soc.* **2011**, *133*, 3931-3942.

[32] Y. Loquais, E. Gloaguen, S. Habka, V. Vaquero-Vara, V. Brenner, B. Tardivel, M. Mons, *J. Phys. Chem. A* **2015**, *119*, 5932-5941.

[33] R. Chaudret, B. de Courcy, J. Contreras-García, E. Gloaguen, A. Zehnacker-Rentien, M. Mons, J.-P. Piquemal, *Phys. Chem. Chem. Phys.* **2014**, *16*, 2285-2288.

[34] B. Khatri, P. Majumder, J. Nagesh, A. Penmatsa, J. Chatterjee, *Chem. Sci.* **2020**, *11*, 9480-9487.

[35] D. P. Leader, E. J. Milner-White, *BMC Bioinformatics* **2009**, *10*, 60.

[36] a) G. D. Rose, P. J. Fleming, J. R. Banavar, A. Maritan, *PNAS* **2006**, *103*, 16623-16633; b) T. Kimura, A. Maeda, S. Nishiguchi, K. Ishimori, I. Morishima, T. Konno, Y. Goto, S. Takahashi, *PNAS* **2008**, *105*, 13391-13396.

[37] K. Guruprasad, M. S. Prasad, G. R. Kumar, **2011**, Database of Structural Motifs in Proteins; DSMP-0 (V1.0), interrogated in 2017.

[38] a) E. deAlba, M. A. Jimenez, M. Rico, *J. Am. Chem. Soc.* **1997**, *119*, 175-183; b) D. A. Mitsikas, N. M. Glykos, *Plos One* **2020**, *15*, e0243429 and references therein.

[39] a) Y. Levy, J. N. Onuchic, *Ann. Rev. BioPhys. BioMol. Str.* **2006**, *35*, 389-415; b) S. Busch, C. D. Bruce, C. Redfield, C. D. Lorenz, S. E. McLain, *Angew. Chem. Int. Ed.* **2013**, *52*, 13091-13095.