Supporting Information

N–H···X interactions stabilize intra-residue C5 hydrogen bonded conformations in heterocyclic α-amino acid derivatives

Venkateswara Rao Mundlapati, Zeynab Imani, Viola C. D'mello, Valérie Brenner, Eric Gloaguen, Jean-Pierre Baltaze, Sylvie Robin, Michel Mons, David J. Aitken

Table of Contents

S1. Synthesis

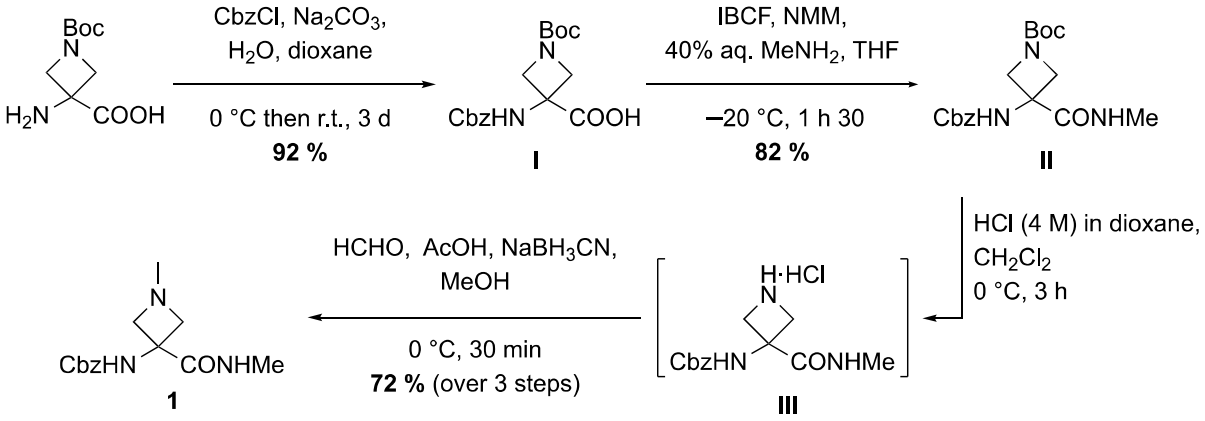
S1.1 General information: instrumentation and materials

Materials and methods. Compound **2** (Cbz-Attc-NHMe) was prepared as described previously.[1] 1-aminocyclobutanecarboxylic acid, 3-aminooxetane-3-carboxylic acid, 3-amino-1-(*tert*-butoxycarbonyl)-azetidine-3-carboxylic acid and isobutyl chloroformate (IBCF) (98%) were purchased from Fluorochem; benzyl chloroformate (CbzCl) (97 wt%) and $NaBH_3CN$ were purchased from Acros; magnesium sulfate (98%) was purchased from VWR Chemicals; *N*-methylmorpholine (NMM) (99%), aqueous methylamine (40%) and aqueous formaldehyde (37 %) were purchased from Sigma-Aldrich; 4 M HCl solution in dioxane was purchased from TCI. These reagents were used as supplied, except for NMM which was fractionally distilled (120 °C) from KOH. Hydrochloric acid (36 wt%) and sodium hydroxide (98.5%) were purchased from Sigma-Aldrich, $NaHCO_3$ (99%) was purchased from Acros; these chemicals were used to make aqueous solutions. THF was distilled from sodium/benzophenone under argon. Other solvents (>99% purity) were purchased from VWR Chemicals or Carlo Erba and used as received (abbreviation: PE = petroleum ether, boiling range 40-65 °C). Preparative flash chromatography was performed on silica gel columns (40-63 μm). Analytical thin-layer chromatography was carried out on commercial silica gel TLC plates of 0.25 mm thickness (Merck, Silica Gel $60F_{254}$); after elution, the plates were visualized by fluorescence at 254 nm then revealed using acidic *p*-anisaldehyde solution (5% in EtOH), ninhydrin solution (14 mM in EtOH), or $KMnO_4$ solution (7.5% in water). Retention factors (R_f) are given for such TLC analyses.

Routine compound characterization. Melting points were measured in open capillary tubes on a Büchi B-540 apparatus and are uncorrected. ^1H and ^{13}C NMR spectra were recorded at the indicated temperature on a Bruker spectrometer operating at 400 MHz for ^1H and at 100 MHz for ^{13}C. For ^1H NMR spectra, chemical shifts (δ) are reported in parts per million (ppm) using residual protonated solvent as the internal reference (δ = 7.26 ppm for chloroform, δ = 3.31 ppm for methanol). Splitting patterns for ^1H signals are designated as s (singlet), bs (broad singlet), d (doublet), or m (multiplet); coupling constants (*J*) are reported in hertz. For ^{13}C NMR spectra, chemical shifts (δ) are reported in parts per million (ppm) using the deuterated solvent as the internal reference (δ = 77.23 ppm for $CDCl_3$, δ = 49.15 ppm for CD_3OD). Fourier-transform infrared (IR) spectra were recorded on a Perkin Elmer Spectrum Two spectrometer for neat samples (whether liquids or solids) using the ATR diamond accessory; maximum absorbances (*v*) for significant bands are given in cm^{-1}. High-resolution mass spectrometry (HRMS) data were recorded on a Bruker MicroTOF-Q instrument using positive-mode electrospray ionization.

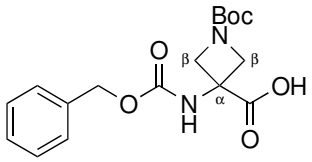
S1.2 Synthesis of Cbz-Aatc(Me)-NHMe, **1**

Compound **1** was synthesized according to Scheme S1.1.



Scheme S1.1.

3-Benzyloxycarbonylamino-1-(*tert*-butoxycarbonyl)azetidine-3-carboxylic acid I. To a suspension of 3-amino-1-(*tert*-butoxycarbonyl)azetidine-3-carboxylic acid (500 mg, 2.31 mmol, 1 eq.) in a 2:1 mixture of H_2O:dioxane (15 mL) was added Na_2CO_3 (735 mg, 6.93 mmol, 4.5 eq.). The suspension was stirred until all the solid had dissolved. The solution was cooled to 0 °C and CbzCl (170 µL, 1.16 mmol, 0.5 eq.) was added dropwise; the solution was then stirred for 24 h at room temperature. The solution was cooled to 0 °C and more CbzCl was added dropwise (170 µL, 1.16 mmol, 0.5 eq.). This procedure was repeated once more so that a total of 1.5 eq. of CbzCl had been added to the solution. After a further 24 h at room temperature, the reaction mixture was concentrated to a volume of 10 mL under reduced pressure. The residual solution was washed with hexane (3 × 30 mL) then slowly acidified at 0 °C with a 2 M aqueous HCl solution until pH 1 was reached. The solution was then extracted with $CHCl_3$ (6 × 30 mL) and the combined organic layers were dried over $MgSO_4$, filtered and concentrated under reduced pressure. Compound **I** was obtained as a light-yellow solid (744 mg, 92%) and used in the next step without further purification.



***R*$_f$** = 0.21 (CH_2Cl_2:MeOH:AcOH = 95:5:1).

Mp = 173–175 °C.

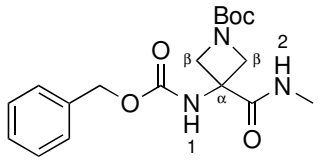
^{1}H NMR (400 MHz, CD_3OD, 300 K) δ7.47–7.17 (m, 5H, CH^{Ar}), 5.21 (s, 0.2H, CH_2^{Cbz} rotamer), 5.10 (s, 1.8H, CH_2^{Cbz} rotamer), 4.40–4.13 (m, 2H, $C^{\beta}H^{a}$), 4.40–3.86 (m, 2H, $C^{\beta}H^{b}$), 1.45 (s, 9H, *t*Bu).

^{13}C NMR (100 MHz, CD_3OD, 300 K) δ174.22 (CO acid), 157.98 (CO^{Boc}), 157.96 (CO^{Cbz}), 137.96 (C^{Ar}), 129.47, 129.04, 128.78 (CH^{Ar}), 81.56 (C^{tBu}), 67.69 (CH_2^{Cbz}), 59.50 and 58.37 (2 broad rotamer signals, $C^{\beta}H_2$), 57.17 (C^{α}), 28.56 (CH_3).

IR (neat) ν 3242, 2970, 1715, 1667, 1529 cm^{-1}.

HRMS [ESI(+)] *m/z* $[M+Na]^+$ calculated for $[C_{17}H_{22}N_2NaO_6]^+$: 373.1370, found: 373.1358.

***tert*-Butyl 3-benzyloxycarbonylamino-3-(methylcarbamoyl)azetidine-1-carboxylate II.** To a cold (– 20 °C) solution of **I** (200 mg, 0.57 mmol, 1 eq.) in THF (2 mL) in an argon-flushed 25 mL one-necked flask were added successively NMM (63 µL, 0.57 mmol, 1 eq.) and IBCF (74 µL, 0.57 mmol, 1 eq.). The activation period was 10 min at –20 °C. A 40% aqueous solution of $MeNH_2$ (0.5 mL, 5.71 mmol, 10 eq.) in THF (0.5 mL) was then added and the solution was stirred for 1.5 h at –20 °C, followed by addition of 5% aqueous $NaHCO_3$ (5 mL). The resulting mixture was stirred for 1 h at room temperature. CH_2Cl_2 (20 mL) was added and the two layers were separated. The aqueous layer was extracted with CH_2Cl_2 (5 × 20 mL). The combined organic layers was washed with 5% aqueous $NaHCO_3$ (2 × 5 mL), dried over $MgSO_4$, filtered and concentrated under reduced pressure. The residue was purified by flash chromatography (PE:EtOAc gradient from 7:3 to 0:1) to give **II** as a white solid (170 mg, 82%).



R_f = 0.09 (PE:EtOAc = 1:1).

Mp = 63–65 °C.

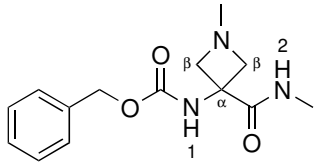
^{1}H NMR (400 MHz, $CDCl_3$, 300 K) δ7.48–7.11 (m, 5H, CH^{Ar}), 6.69 (bs, 1H, NH^2), 6.23 (bs, 1H, NH^1), 5.10 (s, 2H, CH_2^{Cbz}), 4.42–3.80 (m, 4H, $C^{\beta}H_2$), 2.82 (bs, 3H, NCH_3), 1.42 (s, 9H, *t*Bu).

^{13}C NMR (100 MHz, $CDCl_3$, 300 K) δ171.49 (CO amide), 156.60 (CO^{Boc}), 155.46 (CO^{Cbz}), 135.86 (C^{Ar}), 128.72, 128.51, 128.27 (CH^{Ar}), 80.49 (C^{tBu}), 67.38 (CH_2^{Cbz}), 57.78 ($C^{\beta}H_2$), 53.50 (C^{α}), 28.40 (NCH_3), 26.85 (CH_3).

IR (neat) ν 3300, 2976, 1658, 1518 cm^{-1}.

HRMS [ESI(+)] *m/z* $[M+Na]^+$ calculated for $[C_{18}H_{25}N_3NaO_5]^+$: 386.1686, found: 386.1674.

Benzyl (1-methyl-3-(methylcarbamoyl)azetidin-3-yl)carbamate, Cbz-Aatc(Me)-NHMe 1. To an ice-cooled solution of **II** (139 mg, 0.38 mmol, 1 eq.) in CH_2Cl_2 (10 mL) in an argon-flushed one-necked flask was added dropwise 4 M HCl solution in dioxane (4.8 mL, 19.1 mmol, 50 eq.). The resulting solution was stirred for 30 min at 0 °C then 3 h at room temperature, then concentrated under reduced pressure. Remaining volatiles were co-evaporated with $CHCl_3$ (4 × 10 ml) under reduced pressure to leave Cbz-Aatc-NHMe as its hydrochloride salt **III** which was used directly in the following step.
To an ice-cooled solution of **III** in MeOH (3 mL) in an one-necked flask was added acetic acid (109 µL, 1.9 mmol, 5 eq.). A 37% aqueous solution of formaldehyde (285 µL, 3.8 mmol, 10 eq.) was added dropwise and the resulting solution was stirred for 5 min at 0 °C. $NaBH_3CN$ (72 mg, 1.14 mmol, 3 eq.) was added carefully then the mixture was stirred overnight at room temperature. The solvent was removed under reduced pressure and the residue was partitioned between EtOAc (30 mL) and saturated aqueous $NaHCO_3$ solution (45 mL). The EtOAc layer was collected and the aqueous layer was extracted with EtOAc (9 × 30 mL). The combined organic layers were dried over $MgSO_4$, filtered and concentrated under reduced pressure. The residue was purified by flash chromatography (EtOAc:MeOH gradient from 95:5 to 50:50). The appropriate fractions were pooled and evaporated under reduced pressure. The residue was taken up in $CHCl_3$ and the turbid solution was filtered through a pad of Dicalite. The filtrate was evaporated to give **Cbz-Aatc(Me)-NHMe 1** as a beige solid (76 mg, 72%).



R_f = 0.30 (EtOAc:MeOH = 9:1).

Mp = 161–163 °C.

^{1}H NMR (400 MHz, $CDCl_3$, 300 K) δ8.16 (bs, 1H, NH^2), 7.45–7.19 (m, 5H, CH^{Ar}), 6.39 (bs, 1H, NH^1), 5.09 (s, 2H, CH_2^{Cbz}), 4.06–3.82 (m, 2H, $C^{\beta}H^a$), 3.60–3.36 (m, 2H, $C^{\beta}H^b$), 2.88 (d, ^3J = 4.5 Hz, 3H, NCH_3 amide), 2.45 (s, 3H, NCH_3 azetidine).

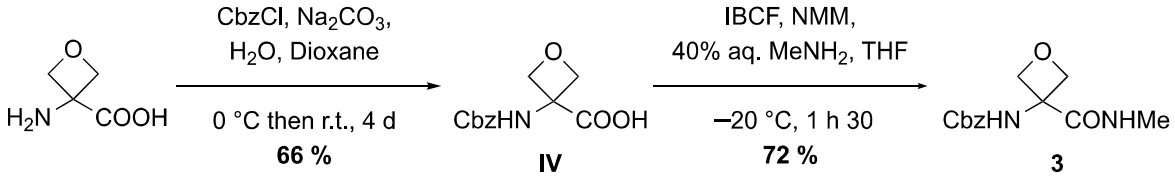
^{13}C NMR (100 MHz, $CDCl_3$, 300 K) δ173.18 (CO amide), 155.32 (CO^{Cbz}), 136.46 (C^{Ar}), 128.64, 128.23, 128.08 (CH^{Ar}), 66.53 (CH_2^{Cbz}), 63.14 ($C^{\beta}H_2$), 53.81 (C^{α}), 44.94 (NCH_3 azetidine), 26.63 (NCH_3 amide).

IR (neat) ν 3300, 2933, 2894, 2784, 1712, 1650, 1574, 1574, 1647 cm^{-1}.

HRMS [ESI(+)] *m/z* $[M+Na]^+$ calculated for $[C_{14}H_{19}N_3NaO_3]^+$: 300.1318, found: 300.1311, $[M_2+Na]^+$ calculated for $[C_{28}H_{38}N_6NaO_6]^+$: 577.2745, found: 577.2717.

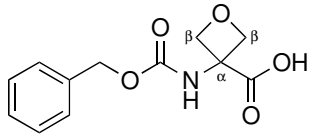
S1.3 Synthesis of Cbz-Aotc-NHMe, **3**

Compound **3** was synthesized according to Scheme S1.2.



Scheme S1.2.

3-(Benzyloxycarbonylamino)oxetane-3-carboxylic acid IV. To a suspension of 3-aminooxetane-3-carboxylic acid (1.00 g, 8.50 mmol, 1 eq.) in a 2:1 mixture of H_2O:dioxane (45 mL) was added Na_2CO_3 (4.05 g, 38.3 mmol, 4.5 eq.). The suspension was stirred until all the solid was dissolved. The solution was cooled to 0 °C and CbzCl (610 µL, 4.3 mmol, 0.5 eq.) was added dropwise; the solution was then stirred for 24 h at room temperature. The solution was then cooled to 0 °C and more CbzCl was added dropwise (610 µL, 4.3 mmol, 0.5 eq.). This procedure was repeated two more times so that a total of 2 eq. of CbzCl had been added to the solution. After a further 24 h at room temperature, the reaction mixture was concentrated to a volume of 15 mL under reduced pressure. The residual solution was washed with hexane (3 × 100 mL) then slowly acidified at 0 °C with a 2 M aqueous HCl solution until pH 1 was reached. The solution was then extracted with EtOAc (6 × 100 mL) and the combined organic layers were dried over $MgSO_4$, filtered and concentrated under reduced pressure. The residue was purified by flash chromatography (EtOAc:MeOH gradient from 98:2 to 90:10) to give **IV** as a sticky cream-coloured semi-solid (1.48 g, 66%).



***R*$_f$** = 0.32 (CH_2Cl_2:MeOH:AcOH = 95:5:1).

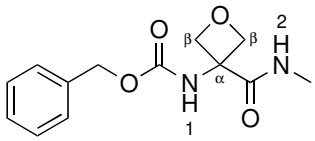
^{1}H NMR (400 MHz, $CDCl_3$, 300 K) δ8.14 (bs, 1H, OH), 7.41–7.23 (m, 5H, CH^{Ar}), 6.76 (bs, 0.2H, NH rotamer), 5.83 (bs, 0.8H, NH rotamer), 5.13 (s, 2H, CH_2^{Cbz}), 5.08–4.89 (m, 2H, $C^{\beta}H^{a}$), 4.86–4.65 (m, 2H, $C^{\beta}H^{b}$).

^{13}C NMR (100 MHz, $CDCl_3$, 300 K) δ 174.12 (CO acid), 155.76 (CO carbamate), 135.76 (C^{Ar}), 128.78, 128.60, 128.30 (CH^{Ar}), 77.72 ($C^{\beta}H_2$), 67.66 (CH_2^{Cbz}), 58.41 (C^{α}).

IR (neat) ν 3312, 2958, 1701, 1521 cm^{-1}.

HRMS [ESI(+)] *m/z* $[M+Na]^+$ calculated for $[C_{12}H_{13}NNaO_5]^+$: 274.0686, found: 274.0682.

Benzyl (3-(methylcarbamoyl)oxetan-3-yl)carbamate, Cbz-Aotc-NHMe 3. To a cold (– 20 °C) solution of **IV** (300 mg, 1.19 mmol, 1 eq.) in THF (5 mL) in an argon-flushed 25 mL one-necked flask were added successively NMM (131 µL, 1.19 mmol, 1 eq.) and IBCF (155 µL, 1.19 mmol, 1 eq.). The activation period was 10 min at –20 °C. A 40% aqueous solution of $MeNH_2$ (1.05 mL, 11.9 mmol, 10 eq.) in THF (2 mL) was then added and the solution was stirred for 1.5 h at –20 °C, followed by the addition of 5% aqueous $NaHCO_3$ (5 mL). The resulting mixture was stirred for 1 h at room temperature. CH_2Cl_2 (20 mL) was added and the two layers were separated. The aqueous layer was extracted with CH_2Cl_2 (5×20 mL). The combined organic layers were washed with 5% aqueous $NaHCO_3$ (2×5 mL), dried over $MgSO_4$, filtered and concentrated under reduced pressure. The residue was purified by flash chromatography (PE:EtOAc gradient from 7:3 to 0:1) to give **Cbz-Aotc-NHMe 3** as a cream-coloured solid (227 mg, 72%).



R_f = 0.38 (EtOAc).

Mp = 153–155 °C.

^{1}H NMR (400 MHz, $CDCl_3$, 300 K) δ 7.42–7.30 (m, 5H, CH^{Ar}), 6.85 (bs, 1H, NH^2), 5.92 (bs, 1H, NH^1), 5.14 (s, 2H, CH_2^{Cbz}), 5.10 (m, 2H, $C^{\beta}H^a$), 4.77 (m, 2H, $C^{\beta}H^b$), 2.91 (bs, 3H, NCH_3).

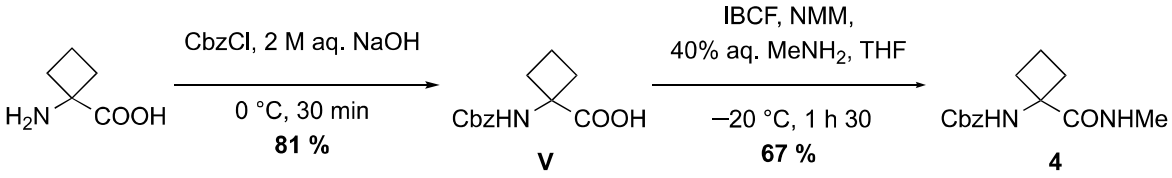
^{13}C NMR (100 MHz, $CDCl_3$, 300 K) δ 170.39 (CO amide), 154.46 (CO^{Cbz}), 134.99 (C^{Ar}), 128.78, 128.56, 128.29 (CH^{Ar}), 77.20 ($C^{\beta}H_2$), 66.41 (CH_2^{Cbz}), 58.02 (C^{α}), 25.82 (NCH_3).

IR (neat) ν 3313, 3227, 3036, 2947, 2881, 1712, 1653, 1532 cm^{-1}.

HRMS [ESI(+)] *m/z* $[M+Na]^+$ calculated for $[C_{13}H_{16}N_2NaO_4]^+$: 287.1002, found: 287.0997.

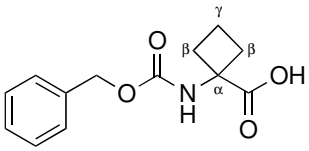
S1.4 Synthesis of Cbz-Ac4c-NHMe, **4**

Compound **4** was synthesized according to Scheme S1.3.



Scheme S1.3.

1-(Benzyloxycarbonylamino)cyclobutane-1-carboxylic acid V. To an ice-bath cooled solution of 1-aminocyclobutanecarboxylic acid (500 mg, 4.3 mmol, 1.0 eq.) in a 2 M aqueous NaOH solution (7.5 mL, 14.8 mmol, 3.40 eq.) in a 25 mL one-necked flask was added CbzCl (1.1 mL, 7.4 mmol, 1.7 eq.). The reaction mixture was stirred for 3 h at 0 °C. The resulting aqueous solution was washed with Et_2O (2×15 mL) and PE (15 mL). The aqueous layer was then acidified with a 2 M aqueous HCl solution to reach pH 1-2 and extracted with EtOAc (6×20 mL). The combined organic layers were dried over $MgSO_4$ and concentrated under reduce pressure to give **V** as a pale yellow solid (867 mg, 81%). This compound was used directly in the next step without further purification.



R_f = 0.42 (CH_2Cl_2:MeOH:AcOH = 95:5:1).

Mp = 79–81 °C.

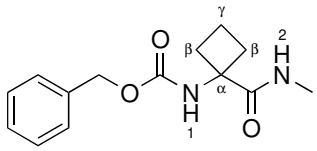
^{1}H NMR (400 MHz, $CDCl_3$, 300 K) δ8.48 (bs, 1H, OH), 7.42–7.24 (m, 5H, CH^{Ar}), 6.23 (bs, 0.2H, NH rotamer), 5.51 (bs, 0.8H, NH rotamer), 5.12 (s, 2H, CH_2^{Cbz}), 2.78–2.53 (m, 2H, $C^{\beta}H^{a}$), 2.52–2.23 (m, 2H, $C^{\beta}H^{b}$), 2.22–1.92 (m, 2H, $C^{\gamma}H_2$).

^{13}C NMR (100 MHz, $CDCl_3$, 300 K) δ178.47 (CO acid), 155.75 (CO^{Cbz}), 136.21 (C^{Ar}), 128.68, 128.35, 128.25 (CH^{Ar}), 67.12 (CH_2^{Cbz}), 58.38 (C^{α}), 32.10 and 31.34 (2 rotamer signals, $C^{\beta}H_2$), 15.14 ($C^{\gamma}H_2$).

IR (neat) *v* 3361, 3040, 3004, 2958, 1737, 1651, 1533 cm^{-1}.

HRMS [ESI(+)] *m/z* $[M+Na]^+$ calculated for $[C_{13}H_{15}NNaO_4]^+$: 272.0893, found: 272.0894.

Benzyl (1-methylcarbamoylcyclobutyl)carbamate, Cbz-Ac4c-NHMe 4. To a cold (– 20 °C) solution of **V** (190 mg, 0.76 mmol, 1 eq.) in THF (3 mL) in an argon-flushed 25 mL one-necked flask were added successively NMM (84 µL, 0.76 mmol, 1 eq.) and IBCF (99 µL, 0.76 mmol, 1 eq.). The activation period was 10 min at –20 °C. A solution of 40% aqueous $MeNH_2$ (660 µL, 7.62 mmol, 10 eq.) in THF (2 mL) was then added and the solution was stirred for 1.5 h at –20 °C, followed by the addition of 5% aqueous $NaHCO_3$ (5 mL). The resulting mixture was stirred for 1 h at room temperature. CH_2Cl_2 (20 mL) was added and the two layers were separated. The aqueous layer was extracted with CH_2Cl_2 (5×20 mL). The combined organic layers were washed with 5% aqueous $NaHCO_3$ (2×5 mL), dried over $MgSO_4$, filtered and concentrated under reduced pressure. The residue was purified by flash chromatography (PE:EtOAc gradient from 7:3 to 0:1) to give **Cbz-Ac4c-NHMe 4** as a white solid (134 mg, 67%).



R_f = 0.29 (PE:EtOAc = 1:1).

Mp = 139–141 °C.

^{1}H NMR (400 MHz, $CDCl_3$, 300 K) δ 7.43-7.26 (m, 5H, CH^{Ar}), 6.60 (bs, 1H, NH^2), 5.29 (bs, 1H, NH^1), 5.11 (s, 2H, CH_2^{Cbz}), 2.79 (bs, 3H, NCH_3), 2.75–2.58 (m, 2H, $C^{\beta}H^b$), 2.27–2.05 (m, 2H, $C^{\beta}H^a$), 2.10–1.88 (m, 2H, $C^{\gamma}H_2$).

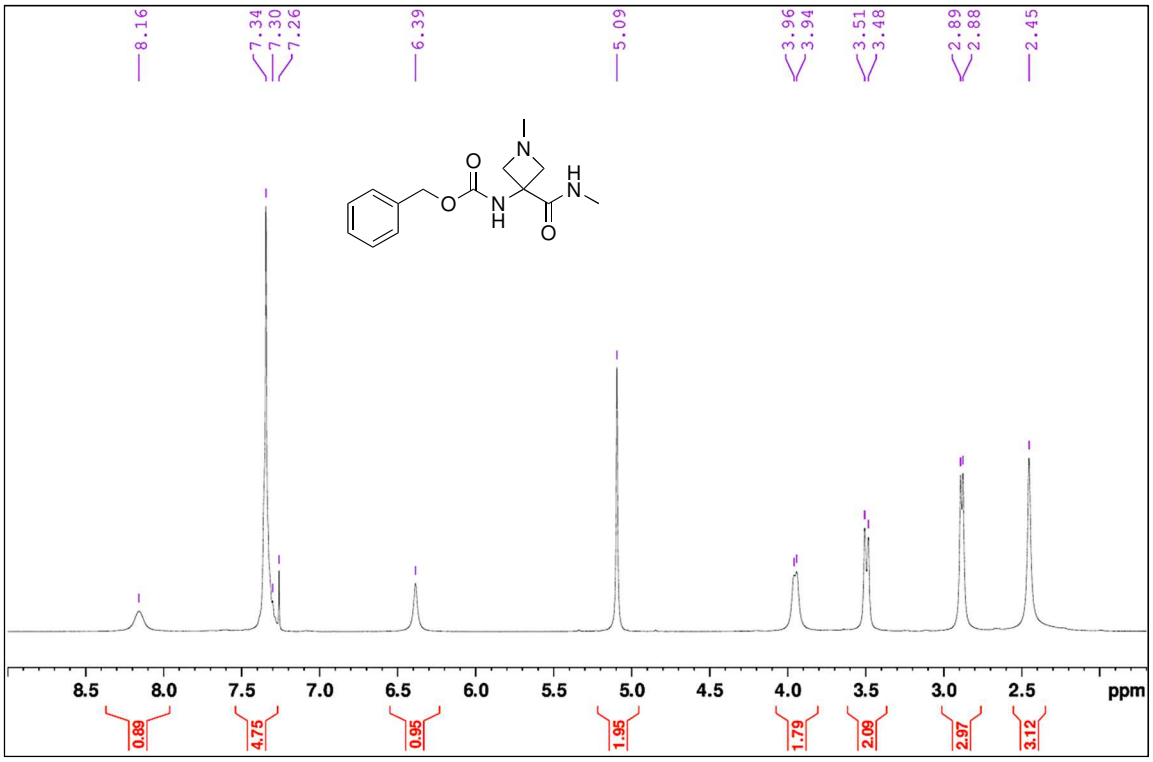
^{13}C NMR (100 MHz, $CDCl_3$, 300 K) δ 174.16 (CO amide), 155.64 (CO^{Cbz}), 136.14 (C^{Ar}), 128.67, 128.39, 128.20 (CH^{Ar}), 67.11 (CH_2^{Cbz}), 59.33 (C^{α}), 31.30 ($C^{\beta}H_2$), 26.51 (NCH_3), 15.36 ($C^{\gamma}H_2$).

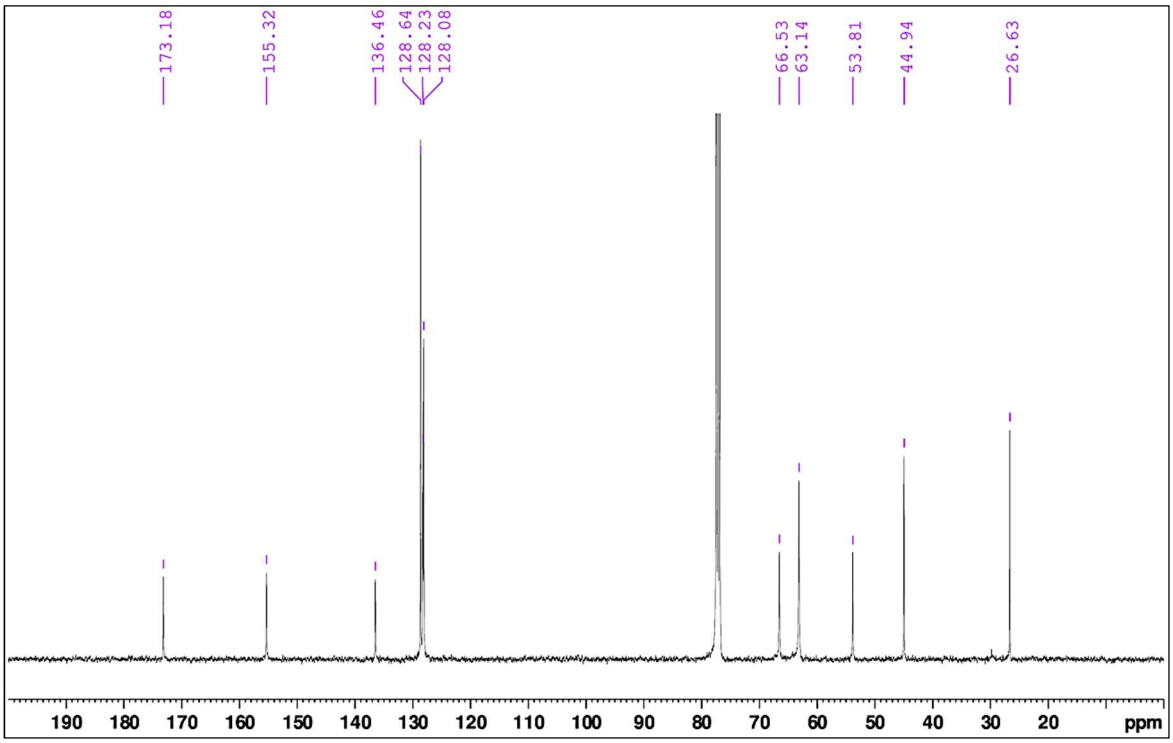
IR (neat) *v* 3325, 2942, 2880, 1682, 1666, 1645, 1515, 1454 cm^{-1}.

HRMS [ESI(+)] *m/z* $[M+Na]^+$ calculated for $[C_{14}H_{18}N_2NaO_3]^+$: 285.1210, found: 285.1201.

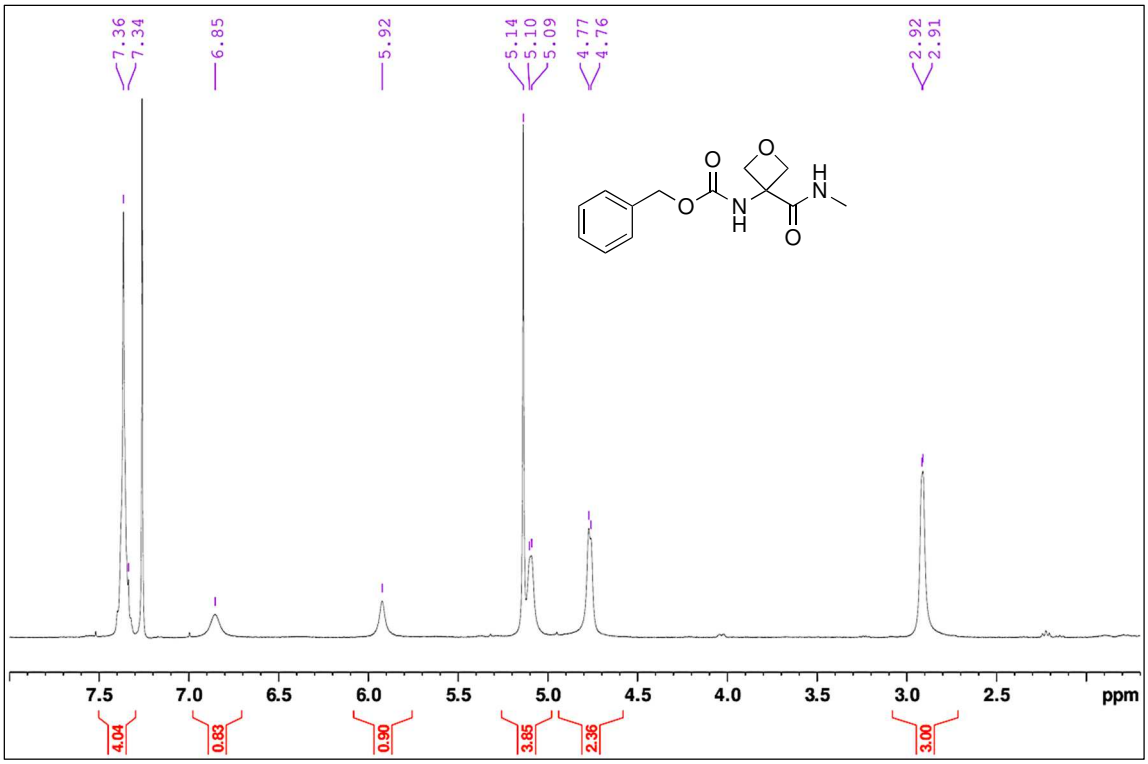
S1.5 Copies of ^{1}H and ^{13}C NMR spectra of new compounds

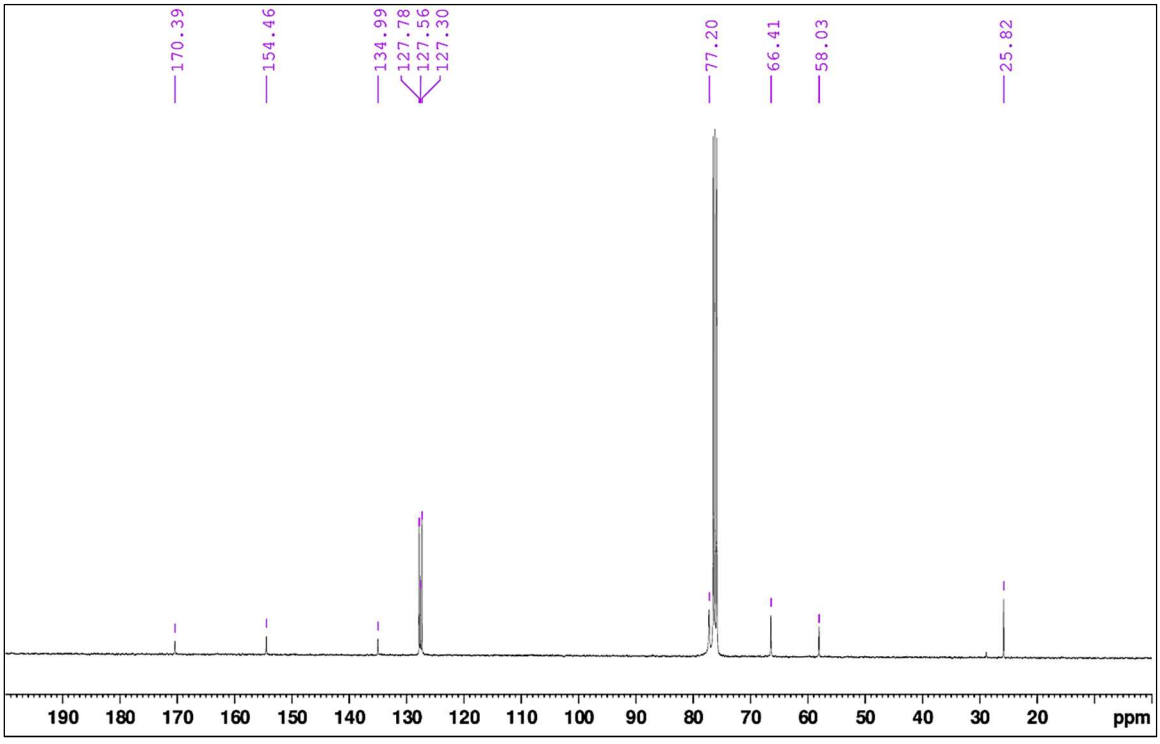
Cbz-Aatc(Me)-NHMe 1



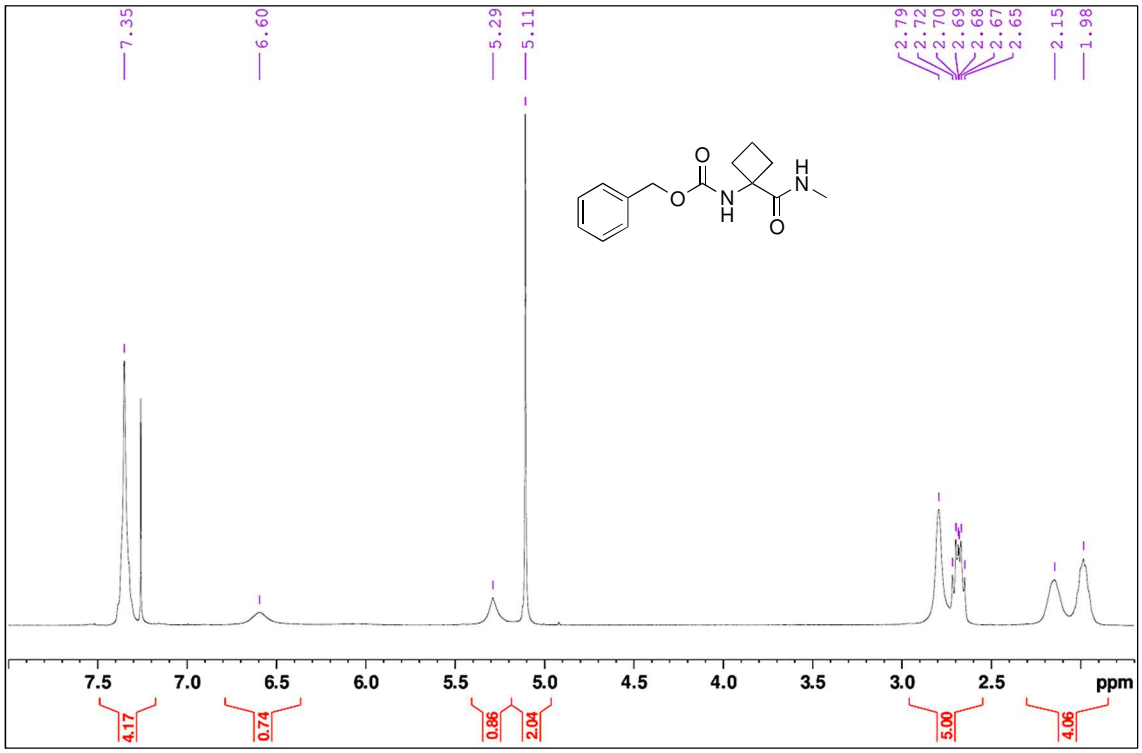


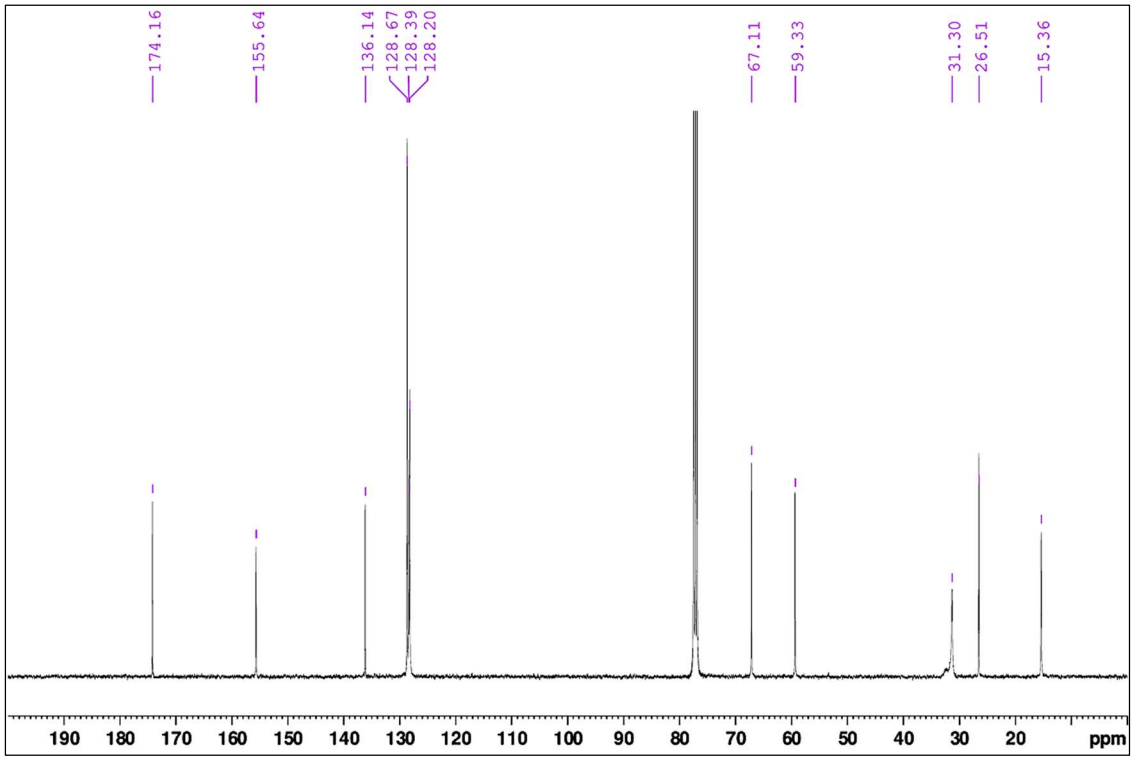
Cbz-Aotc-NHMe 3





Cbz-Ac4c-NHMe 4





S2. Theoretical chemistry

S2.1 Methods

An extensive and detailed description was obtained for compounds **1**-**4**, through a dual step procedure combining a force field exploration of the potential energy landscape of each compound in the gas phase and quantum chemistry structure optimizations. The exploration was done with the OPLS-2005 force field using the Monte-Carlo Multiple Minima procedure in the Schrödinger suite.[2] Subsequent geometry optimizations were done using the Density Functional Theory (DFT) at the RI-B97-D3(BJ-abc)/def2-TZVPPD level of theory[3-5] with the Turbomole 7.2 package[6] using parameters successfully used previously (gridsize m3; SCF convergence threshold 10^{-8} a.u.; gradient norm convergence threshold 10^{-5} a.u.). The structures were then refined and optimized at the MP4(SDQ)/def2-TZVPPD level (SCF convergence threshold 10^{-8} a.u.; gradient norm convergence threshold 4.5 10^{-4} a.u.) with the Gaussian 16-C.01 package[7] in an attempt to correct minor deficiencies in terms of structure and NH stretch frequencies in the sulphur-containing compound **2** (See Section S2.4).

S2.2 Structures and nomenclature

Three backbone families were found for compounds **1**-**4**, as illustrated in Figure S2.1: an extended 5-6γ form, a 7-folded form and an α-folded form. The nomenclature used to describe the conformations is based on a sequential analysis of the H-bonding status of the NH groups along the peptide chain. When a hydrogen bond is present, a number indicates the number of atoms involved in the ring formed by the H-bond between the NH being considered and its H-bond acceptor (referred to as the X atom). The "γ" label following this number indicates that the H-bond implicates a side chain acceptor heteroatom in a γ-position. An NH group free from any hydrogen bonding is designated by the label "f". Those interacting with the π electronic structure of the preceding amide group are indicated by the "π_{amide}" label. In this case the backbone adopts an α conformation (with ϕ, φ Ramachandran dihedrals ~ –30°, ~ –30° for α_L or ~ 30°, ~ 30° for α_D). The puckering of the 4-membered ring leads to the acceptor atom pointing towards the NH or the CO of the residue; this is indicated by a superscript N or O on the status of the NH or the residue.

For each backbone conformation, the Cbz group gives rise to three rotamers, a phenomenon that we refer to as Cbz-rotamerism. The rotamers are designated according to their C_{sp2}-C_{sp3}-O-C(=O) dihedral φ, namely *gauche+* (g^+; φ = +60°), *gauche–* (g^-; φ = –60°) and *trans* (*t*; φ = ± 180°). The carbamate can also be in a *trans* (often the most stable) or a *cis* conformation. Unless clearly stated, the carbamate will be considered in a *trans* conformation (*cf.* Fig. S2.1).

In conformations having a symmetry plane (as in the case of the 5-6γ motif), *gauche+* and *gauche–* forms of the Cbz cap are enantiomers and +/– signs are dropped in the terminology. In contrast, in the non-symmetrical conformations, g^+ and g^- forms can differ, in particular because of the presence/absence of interactions between the benzyl moiety and the rest of the molecule. In this case, the signs are given together with a label indicating the chirality of the backbone conformation: a subscript to the number of an N–H···O=C H-bond (*e.g.* 7_D) or to the α backbone conformation stands for the configurational stereochemistry type for a chiral L or D conformation.

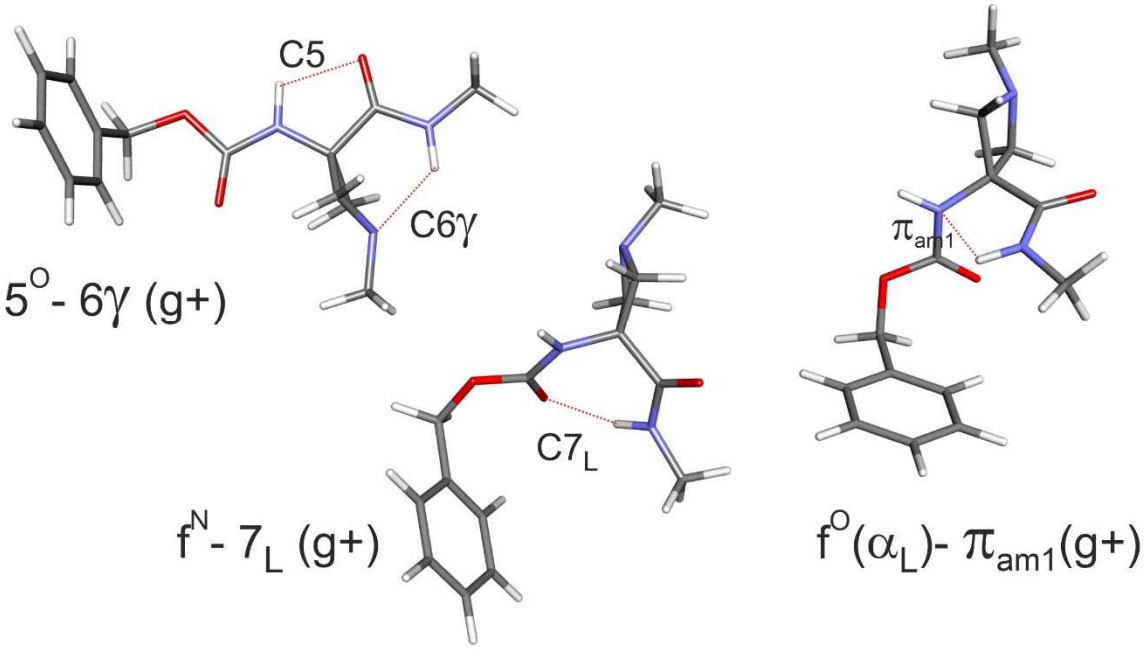


Figure S2.1: The three types of backbones encountered in the present work, illustrated on the capped derivatives of the Aatc(Me) residue (compound 1) as obtained from DFT-D3 level of theory: left: extended 5-6γ form; centre: 7 folded form and right α-folded form. In all these conformations the Cbz cap is given with a *gauche+* orientation and the carbamate with a *trans* conformation.

The extended 5-6γ structures were characterized at several levels of theory in order to assess the robustness of the calculations. Table S2.1 shows that the methoxycarbonyl and Cbz caps give rise to similar C6γ H-bonds and that the best level (MP4) distances tend to be longer than those obtained at the DFT-D level, suggesting that the latter method slightly overestimates the H-bond strength.

Table S2.1: Distances (pm) of the NH···X bonds in extended forms (5-6γ conformation) of amino acid derivatives with two types of N and C terminal caps, namely [Cbz (*gauche* rotamer); methylamide] (compounds 1-4), and [methoxycarbonyl; methylamide] obtained at the DFT-D3 et MP4 level of theory. This latter level could only be performed for the smaller type of caps [methoxycarbonyl; methylamide], due to computation times.

AA ; NH···X	Cbz-AA-NHMe *gauche* (RI-B97-D3(BJ-abc))	MeOCO-AA-NHMe (RI-B97-D3(BJ-abc))	MeOCO-AA-NHMe (MP4)
Aatc(Me) ; NH···N	212	212	213
Attc ; NH···S	241	240	244
Aotc ; NH···O	223	221	218
Ac4c ; NH···Cγ	258	256	253

S2.3 Conformational landscapes

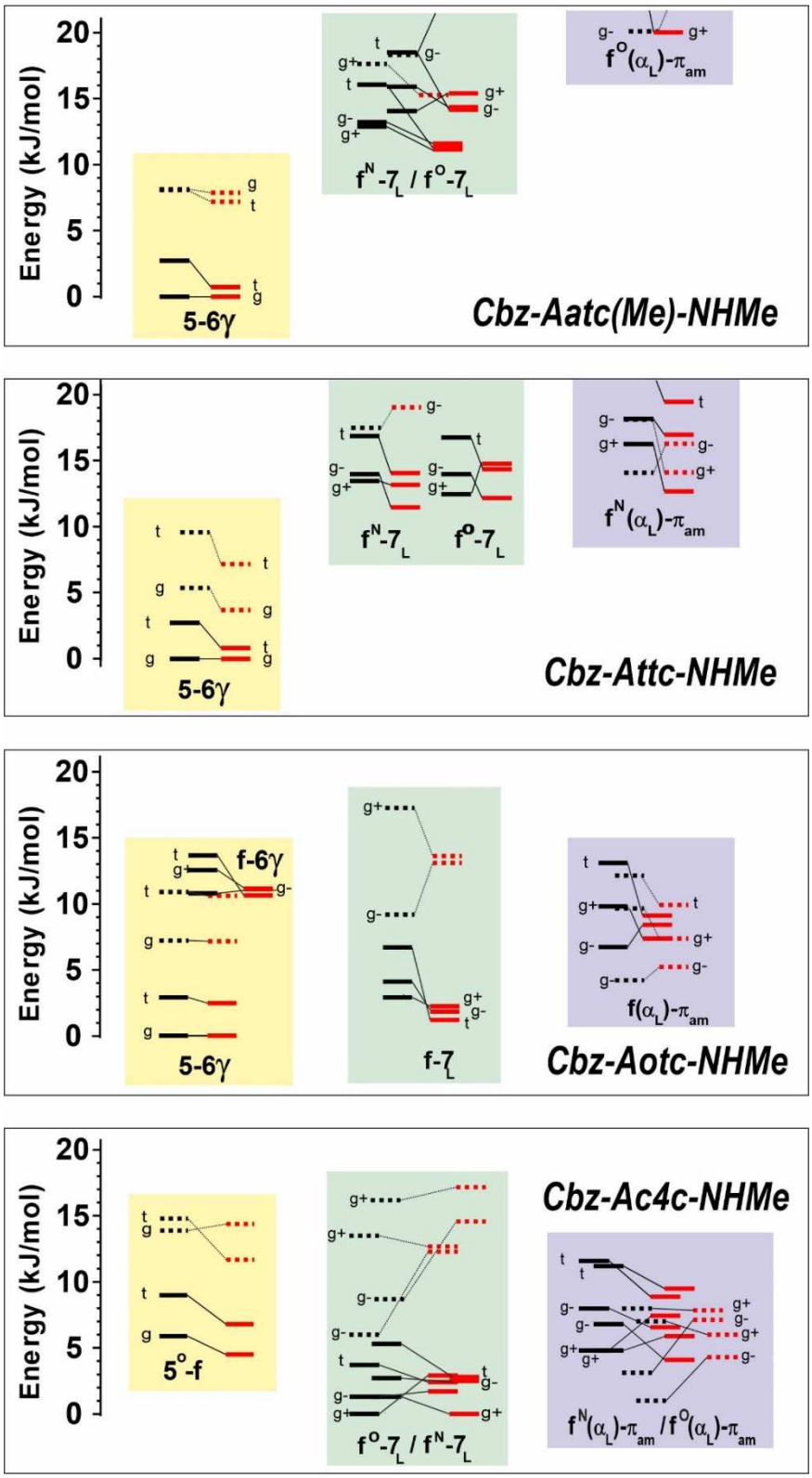


Figure S2.2: Gas phase energetic landscape of compounds 1-4, showing the relative stabilities of the three backbone families of Figure S2.1: ΔH at 0 K (black lines) and ΔG at 300 K (red lines). Full (resp. dotted) lines indicate *trans* (resp. *cis*) conformation of the carbamate N-terminal cap. The conformational terminology is detailed in the previous section. Previous work on peptides in which computationally calculated conformational landscapes were found to give a good match with the structures observed experimentally following a supersonic expansion[8-10] suggests that the accuracy of the Gibbs energies is of the order of a few (2-4) kJ/mol.

S2.4 NH stretch frequencies

For each structure resulting from an optimized geometry at the DFT-D quantum chemistry level, a frequency calculation was performed at the same level of theory within the harmonic approximation using the NumForce module of the Turbomole 7.2 package,[6] with the "-central" option. These calculations enabled us to check that the conformations obtained corresponded to "true" minima and the harmonic frequencies derived were then scaled in order to account for anharmonicity and potential deficiencies of the level of theory used, leading to theoretical data directly reconciled with experimental data. In the NH stretching (amide A) region, a linear scaling factor was applied to the harmonic frequencies ($f^{scaled} = a + b.f^{harmonic}$), based on a fit of frequency data on a set of capped peptides studies in the gas phase.[11] The a and b parameters have been recently recalculated (a = 372.8 cm^{-1} and b = 0.86953) for the level of theory presently used.[12] This strategy generally provides theoretical frequencies in good agreement with experimental frequencies (within 20 cm^{-1} for NH stretches in NH···O bonds).[11]

For NH···S H-bonds, however, the results were found to be of a lesser quality, with discrepancies reaching 30 cm^{-1}.[1, 13] For this reason, structures were then reoptimized at the MP4(SDQ) level, leading to similar geometries although with slightly shorter H-bonds (Table S2.1). Frequencies were then estimated at the same level, following the protocol detailed below. Curvatures (k) of the potential energy surface along the two NH stretch modes were obtained from the average of the energy changes due to a displacement according to each of the purely local NH modes, *i.e.* through a shift of the H atom along the NH bond of +1 and –1 pm; these shifts being compensated by a corresponding anticolinear shift of the N atom, whose length was in the inverse mass ratio. The harmonic frequency f was obtained from f (in cm^{-1}) = $2\pi/c\sqrt{k/\mu}$, where μ is the reduced mass of the N-H system and *c* (the speed of light) is expressed in cm/s. A fixed scaling factor was then determined by a best fit of the 8 frequencies experimentally observed for the four compounds (Fig. 2). Its value 0.878 provided an average unsigned error of 11 cm^{-1}. These scaled frequency values were used to build up MP4 simulated spectra which are compared with DFT-D spectra in Figure S2.3. This shows that MP4 frequencies provide an alternative match all along the series of 5-6γ conformations, where the discrepancy with regard to the experiment is comparable but is more uniformly distributed among the several bands.

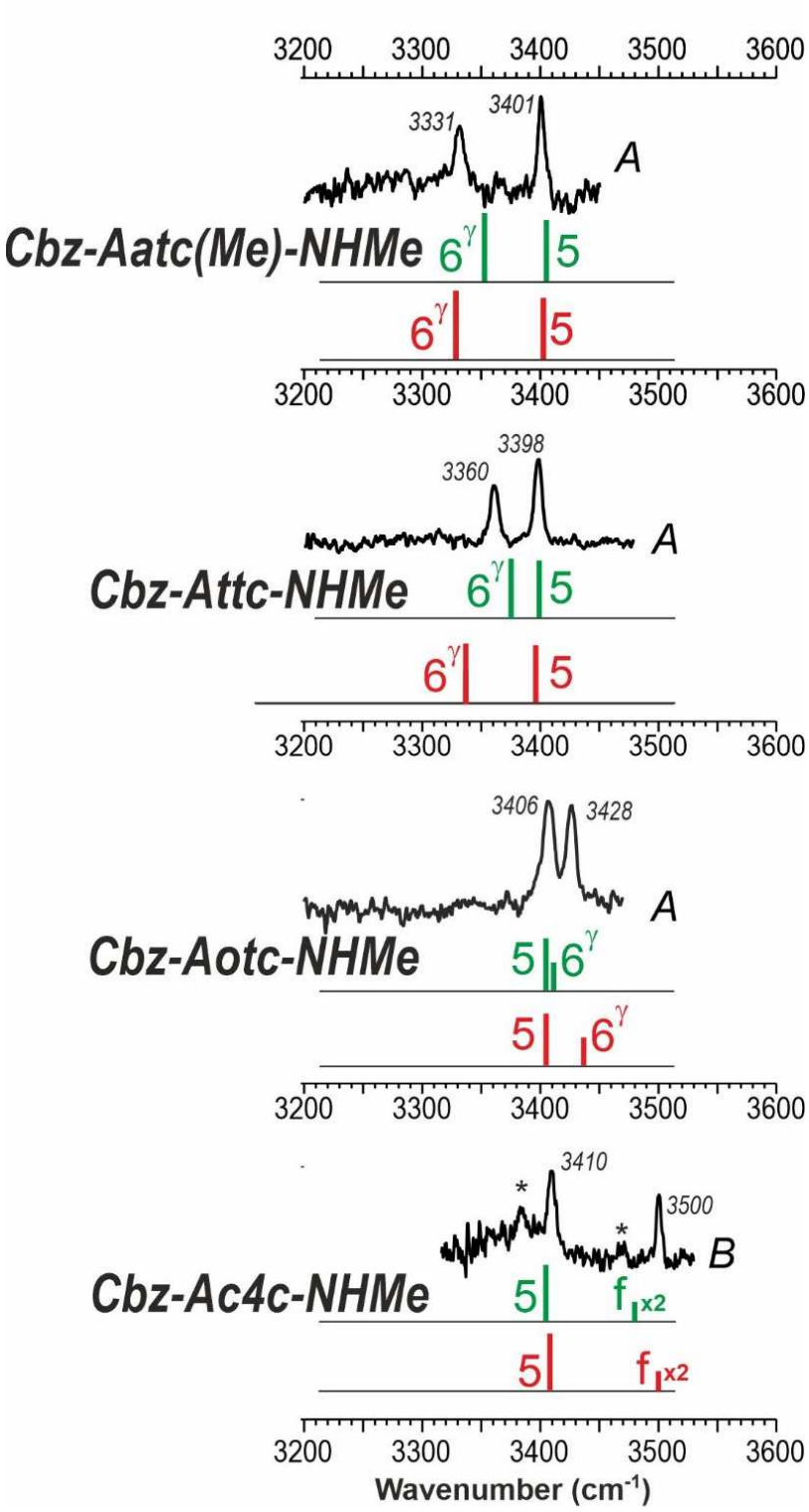


Figure S2.3: Simulated amide A IR spectrum of the g^+ rotamer of the 5-6γ extended form of compounds 1-4 obtained at the DFT-D; RI-B97-D3-BJ-abc/def2-TZVPPD level of theory (red sticks) and of the corresponding methoxycarbonyl-capped models of 1-4, obtained at the MP4 level of theory (green sticks). See Sections S2.1 and S2.4 for theoretical details. The intensities of the MP4 spectrum have been taken from the DFT-D level of theory. For the sake of comparison, the corresponding experimental IR/UV gas phase spectrum is shown as a black trace.

S2.5 NH stretch signature of C5 H-bonding: a quantum chemical investigation

The presence of a NH stretch spectral signature around 3380 cm^{-1} for the series of compounds **1**-**3** in weakly polar solutions (Fig. 5 and Fig. S5.1) led us to deduce the presence of an extended form stabilized by a C5 interaction involving NH(1). This band appears at approximately the same frequency as the C5 NH stretch in capped derivatives of α,α-disubstituted amino acids such as 2-aminoisobutyric acid (Aib)[14-16] and diethylglycine (Deg),[17] but is significantly red-shifted compared to the C5 NH stretch in capped derivatives of α-monosubstituted amino acids[18-20] or glycine.[21, 22] This observation deserved a specific examination. Given the satisfactory modeling of the C5 H-bonds at the DFT-D quantum chemistry level presently used for the systems studied in the gas phase, both here (Fig. 3 and Fig. S2.2) and previously,[12, 23, 24] a theoretical comparative investigation was performed in order to rationalize this spectroscopic feature. For this purpose, frequency calculations were carried out at the DFT-D level (unscaled harmonic frequencies; see Section S2.1) for the extended form of a series of capped amino acids, comprising cyclic (Aatc(Me), Attc, Aotc, Ac4c) or non-cyclic (Aib, Deg) α,α-disubstituted amino acids and the Gly and Ala mono α-substituted (Gly, Ala) taken for reference, with two types of N-terminal cap (corresponding to amide or carbamate, resp.): acetyl or benzyloxycarbonyl (in a g^+ conformation), and methylamide N-terminal caps. The results are plotted in Figure S2.4, as a function of the C5 distance.

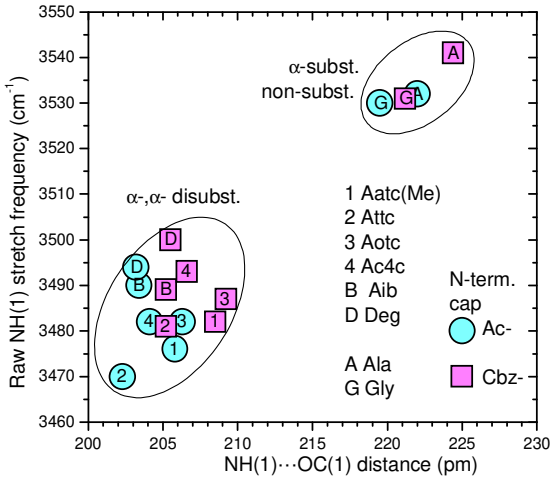


Figure S2.4: Gas phase NH(1) stretch frequency (DFT-D level, scaled harmonic data, in cm^{-1}) as a function of the NH(1)···OC(1) distance (in pm) for a series of capped compounds of cyclic (Aatc(Me), Attc, Aotc, Ac4c) or non-cyclic (Aib, Deg) α,α-disubstituted amino acids, as well as α-substituted (Ala) and non-substituted (Gly) amino acids. Two types of caps were considered for the N-terminal: acetyl (blue disks) and benzyloxycarbonyl (magenta squares), representing amide and carbamate groups, respectively. A methylamide cap was considered for the C-terminal.

The salient result is the presence of two spectroscopic clusters, corresponding on the one hand to non-/mono-substitution (Gly and Ala residues resp.) and on the other hand to di-substitution of the amino acid Cα. Compared to the Ala and Gly references, the di-substituted amino acids constitute a well-separated cluster, red-shifted by about ~50 cm^{-1}. Noticeably, within each cluster the Cbz derivatives tend to be blue shifted compared with their acetyl counterparts, typically by ~6 cm^{-1}, demonstrating that the type of N-cap employed is barely significant and that the Cbz-capped amino acids constitute good models for the study of intramolecular interactions. The presence of clusters also indicates the minor role of the side chain for determining the C5 frequency, compared to the influence of the degree of substitution of the Cα atom. The theoretical analysis of such vibrational specificities was recently analyzed by Brenner et al.,[25] who showed that in di-substituted amino acids the existence of specific close contacts between the two $C\beta H_2$ methylene groups and the vicinal peptide links distorts the backbone locally and forces the NH(1)···OC(1) distance to shrink by typically ~15-20 pm, leading to the observed red-shifted NH stretch frequencies.

S2.6 Dipole moments of remarkable conformations

Table S2.2: Dipole moments (Debye) of selected conformations of compounds 1-4, depending on the backbone type, the *trans/cis* carbamate conformation, and the *gauche±/trans* Cbz conformation, obtained at the RI-B97-D3(BJ-abc)/def2-TZVPPD level of theory.

Molecule	5-6γ		f^N-7_L		f^O-7_L		$f^O(\alpha_L)$-π_{am}		$f^N(\alpha_L)$-π_{am}	
	trans	*cis*	*trans*	*cis*	*trans*	*cis*	*trans*	*cis*	*trans*	*cis*
1 Cbz-Aatc(Me)-NHMe	2.61 *g* 2.15 *t*	5.68 *g* 6.22 *t*	2.40 *g+* 2.88 *g-* 3.00 *t*	- [a] 3.61 *g-* 3.73 *t*	2.94 *g+* 3.57 *g-* 3.64 *t*	3.85 *g+* 3.76 *g-* -	5.24 *g+* 5.00 *g-* 5.65 *t*	1.47 *g+* 1.84 *g-* -	not calc.	not calc.
2 Cbz-Attc-NHMe	1.68 *g* 1.55 *t*	4.20 g 4.67 *t*	3.57 *g+* 3.74 *g-* 3.86 *t*	4.48 *g+* 3.79 *g-* -	4.20 *g+* 4.11 *g-* 4.52 *t*	3.72 *g+* 4.15 *g-* -	not calc.	not calc.	5.21 *g+* 4.36 *g-* 5.16 *t*	1.99 *g+* 3.36 *g-* -
3 Cbz-Aotc-NHMe	1.75 *g* 1.59 *t*	4.51 *g* 4.34 *t*	4.30 *g+* 4.00 *g-* 4.38 *t*	3.68 *g+* 4.33 *g-* -	$f^N = f^O$	$f^N = f^O$	5.91 *g+* 5.01 *g-* 5.84 *t*	2.24 *g+* 3.47 *g-* 2.80 *t*	$f^N = f^O$	$f^N = f^O$
4 Cbz-Ac4c-NHMe	3.22 *g* 2.87 *t*	5.90 *g* 6.38 *t*	2.81 *g+* 3.40 *g-* 3.54 *t*	3.43 *g+* 3.60 *g-* -	2.36 *g+* 3.21 *g-* 3.35 *t*	3.32 *g+* 3.58 *g-* -	4.48 *g+* 4.32 *g-* 4.97 *t*	0.98 *g+* 1.54 *g-* -	4.47 *g+* 4.28 *g-* 4.94 *t*	0.96 *g+* 1.49 *g-* -

a) not a stable structure; optimization converges towards a *cis* $f^N(\alpha_L)$-π_{am} structure.

S3. UV Spectroscopy

Figure S3.1 shows the UV spectra of compounds **1-4** in the spectral region of the origin transition of the first ππ* transition of the phenyl ring in the near UV.

Bands A of **1-3** and B of **4** are all in the same narrow spectral range (37580–37590 cm^{-1}). This feature together with the assignments of the IR spectra recorded using these bands in the IR/UV spectroscopic procedure allows us to assign them to a same conformation of the backbone and of the Cbz rotamer, namely an extended backbone structure. From previous work on Attc and on its capped trimer, it has been shown that the Cbz rotamers were the *gauche* species (+ and –, not distinguished spectroscopically since enantiomers). The satellite A1 bands, located typically 15 cm^{-1} from the corresponding A band, were shown to yield the same IR/UV spectrum (see Section S4) and are thus assigned to a vibronic band of the same species; the low frequency modes associated to this Franck-Condon activity corresponds to the torsion of the Cbz cap, as shown by quantum chemistry frequency calculations.

Bands A, A1-6, C, D and D' of **4** are assigned to several folded structures (See Section S4).

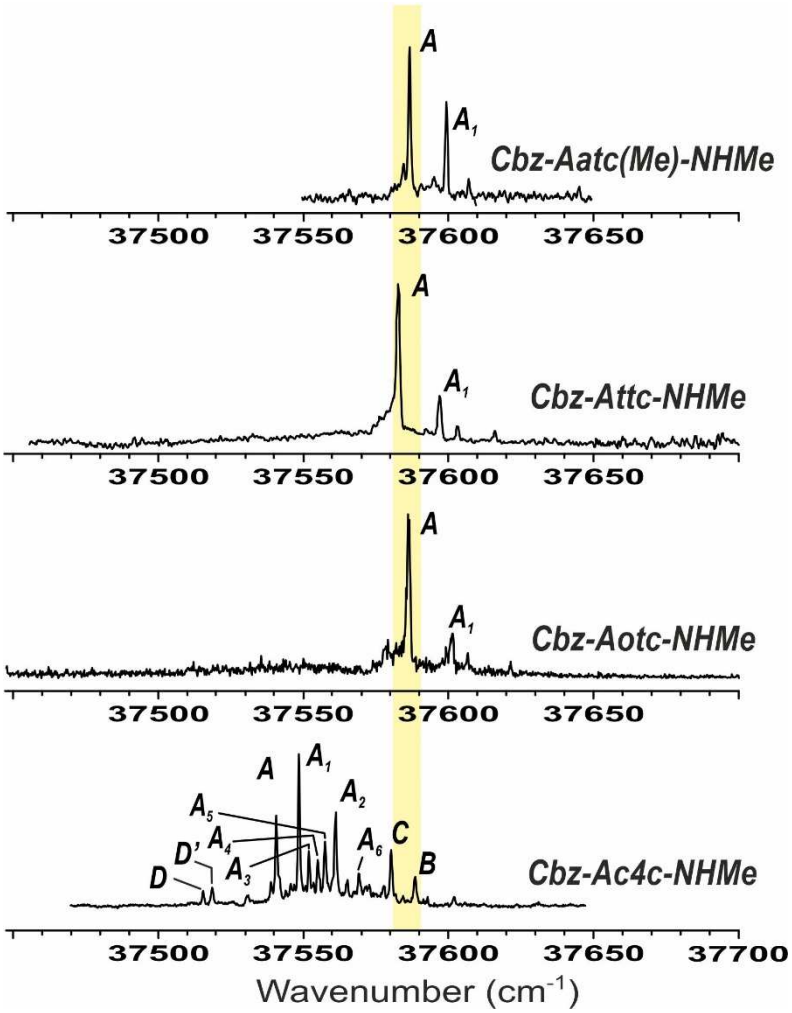


Figure S3.1: Mass-selected near UV spectra of compounds 1-4 in the spectral region of the origin band of the first ππ* transition of the phenyl ring.

S4. IR spectroscopy in the gas phase

S4.1 Extended forms

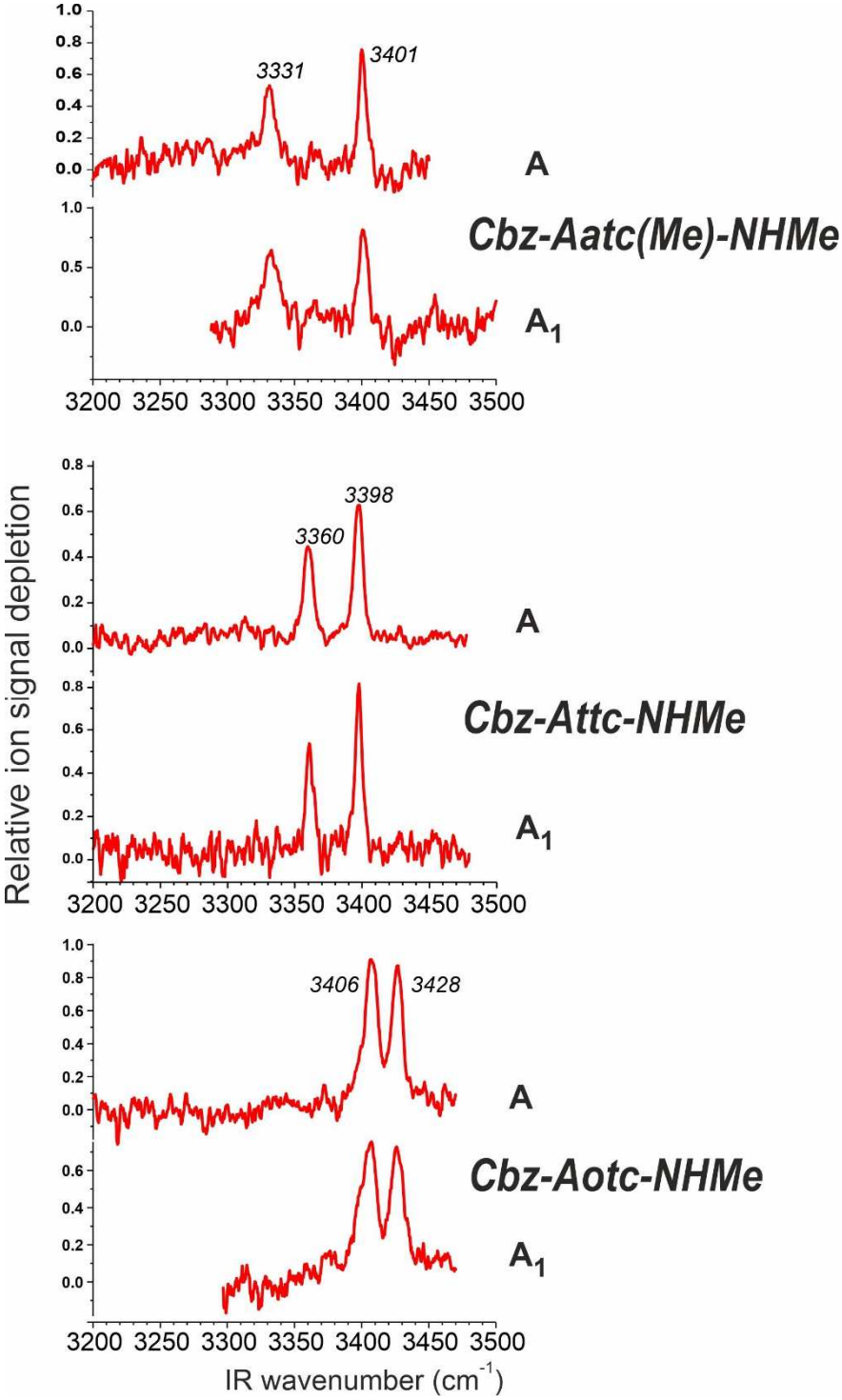


Figure S4.1: IR Spectra of compounds 1-3 recorded by IR/UV spectroscopy when probing the conformational population on UV bands from A and A1 bands, as indicated in Fig. S3.1. Identical spectra reveal that the same population is probed and therefore that bands A and A1 are assigned to the same conformer, A being the origin and A1 a vibronic band.

S4.2 IR spectroscopy of Cbz-Ac4c-NHMe

The origin of the various spectroscopic UV features of compound **4** were determined from their IR/UV spectra. The bands labelled A1-A6 in the UV spectrum (Fig. S4.2) all exhibited the same IR/UV spectrum, demonstrating that they belonged to the same conformer, or at least to conformers having the same backbone, since conformers differing by the orientation of the Cbz cap are known to exhibit very similar (nearly identical) spectra.[1]

In contrast, bands labelled B, C, D and D' showed different IR/UV spectra (Fig. S4.3). Despite the selectivity usually achieved by the IR/UV spectroscopy, traces of the IR spectrum of A could be detected in the spectra of B and D', because of the presence of a nearby weak UV band due to conformer A (a hot band in the low frequency - 8 cm^{-1} - active mode in the latter case). Taking this point into account, it was noticeable that the IR spectra obtained from bands D and D' strongly overlap. From this, we concluded the presence of at least four conformers, labelled A-D according to their UV bands.

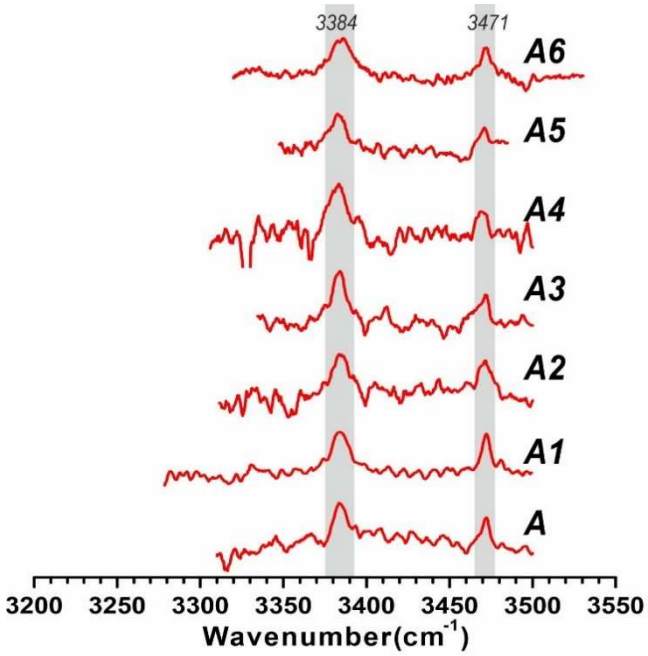


Figure S4.2: IR/UV spectra of compound 4 recorded from bands A-A6 of Fig. S3.1.

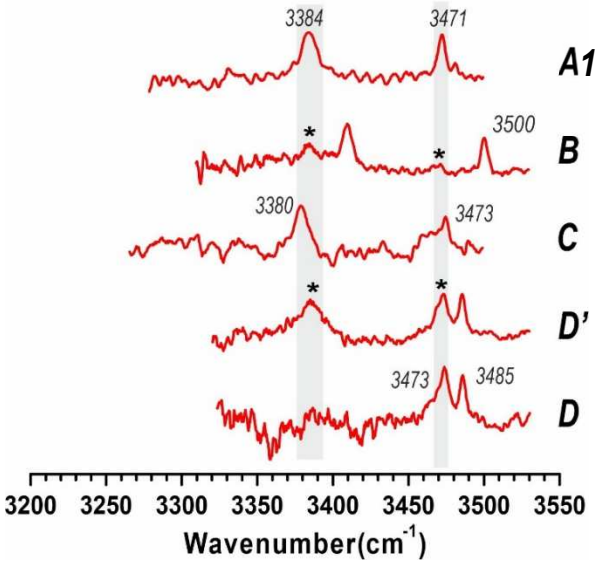


Figure S4.3: IR/UV spectra of compound 4 recorded from bands A1, B, C, D and D' of Fig. S3.1. Asterisks indicate contribution of conformer A spectrum (grey regions), due to a poor selectivity of the UV probe process.

The comparison (Fig. S4.4) of the experimental data (band systems A1-D) with simulated spectra (DFT-D3 level of theory; see Section 2.4) of the most stable forms of each backbone family at 300 K (red levels in Fig. S2.2) suggests to assign conformers A and C to a C7 folded backbone, and the D band to an α backbone configuration, B being assigned to an extended C5-C6γ backbone (see main text). A more precise assignment would be speculative. As shown in Fig. S4.4, the spectra of the most stable C7 forms were very similar and could therefore account for A and C. Similarly, several α-folded conformers were able to account for the D form.

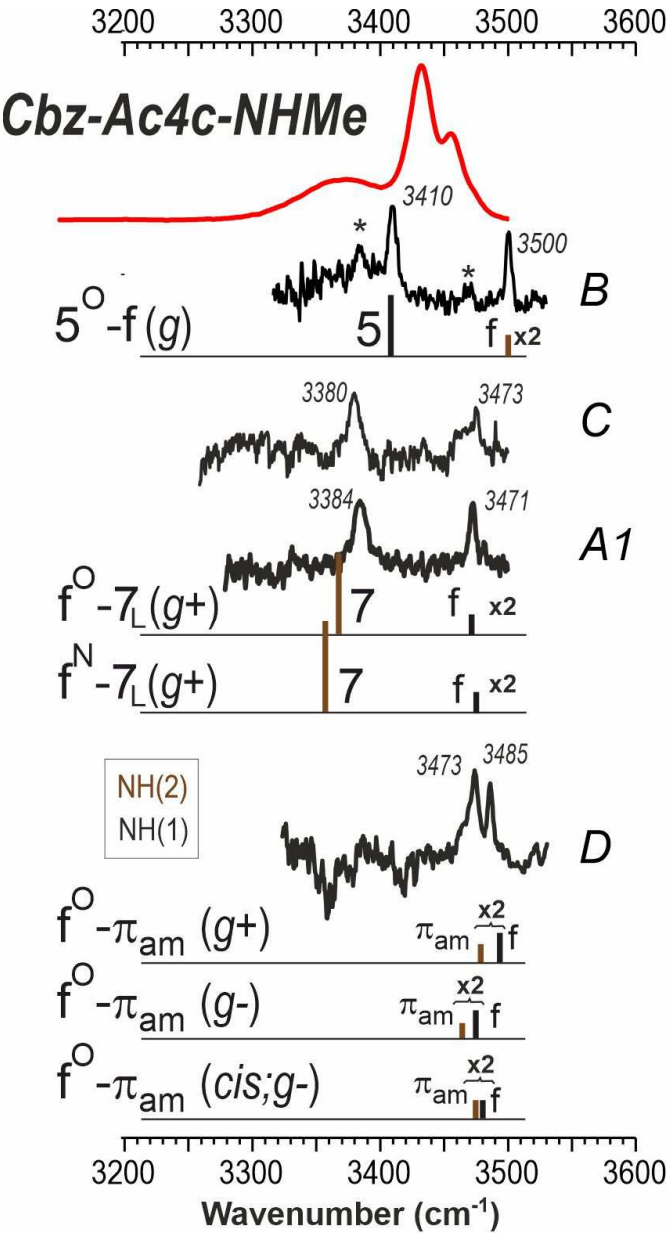


Figure S4.4: Conformer-selective gas phase IR spectra recorded from UV bands A1-D of compound 4 (see Fig. 2, center panel, left), together with simulated spectra (DFT-D level of theory; see details in Section 4.2) for the most stable form of each backbone family.

S5. IR and NMR spectroscopy in solution

S5.1 IR spectroscopy

Solution state infrared spectra of compounds **1-4** were recorded at 293 K on a Fourier-transform Perkin Elmer Spectrum Two spectrometer, using 10 mM solutions in $CHCl_3$ held in Omni-cell Specac 1 mm path-length NaCl plates. Figure S5.1 shows the absorbance bands in the NH stretch (amide A) region.

No changes in the profile of the relevant absorption bands of each compound was observed over the concentration range 5 mM to 30 mM (Figure S5.2), indicating that only intramolecular interactions are at play at 10 mM.

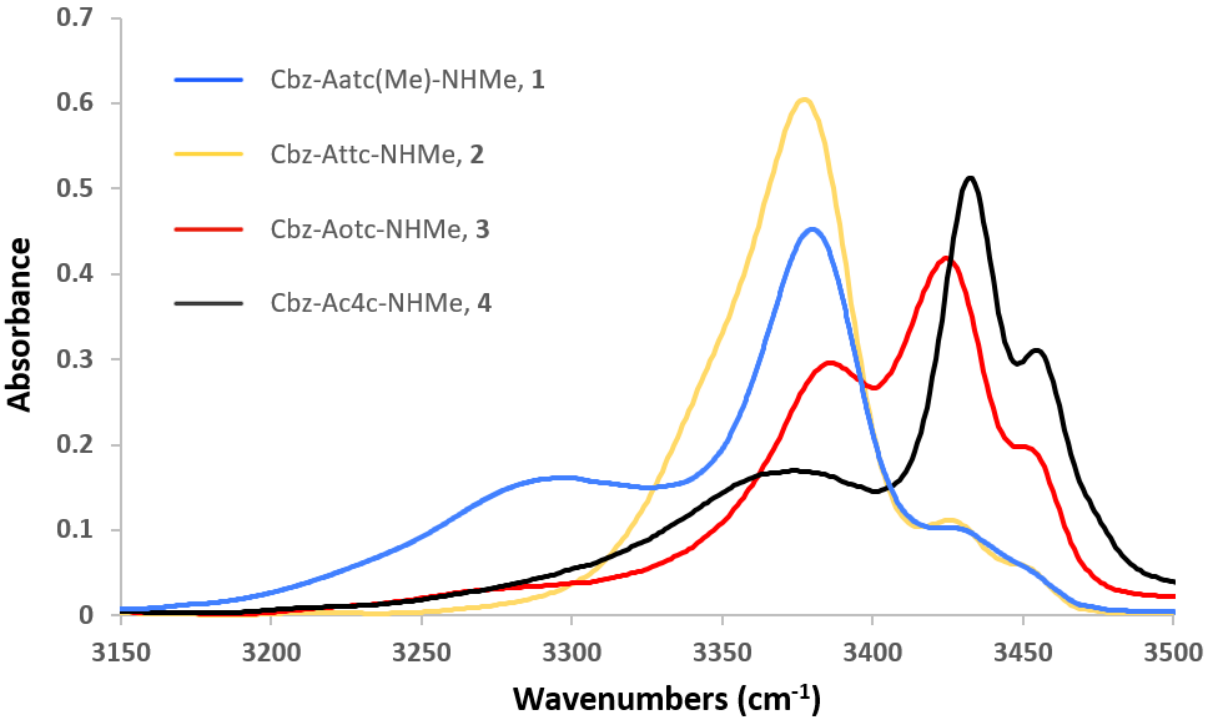
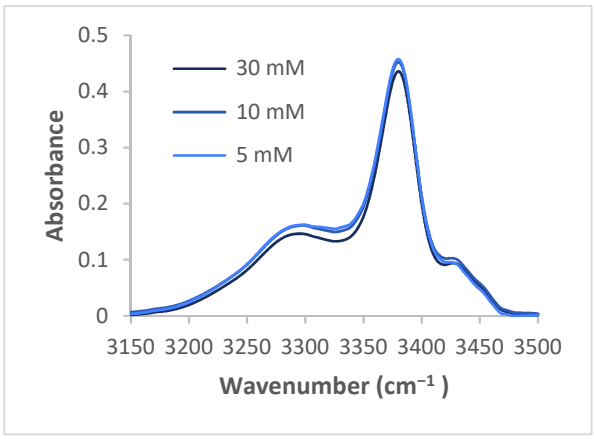
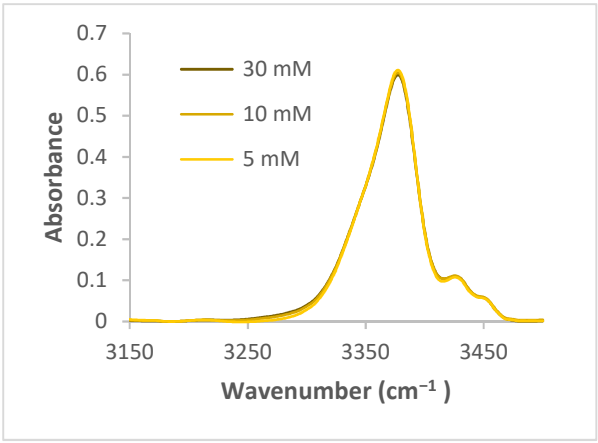


Figure S5.1: Comparison of solution state IR absorption spectra of compounds 1-4 in the NH stretch region (10 mM solutions in $CHCl_3$).

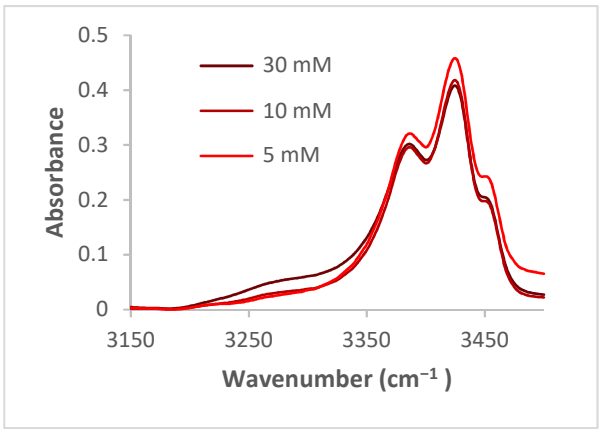
Cbz-Aatc(Me)-NHMe, 1



Cbz-Attc-NHMe, 2



Cbz-Aotc-NHMe, 3



Cbz-Ac4c-NHMe, 4

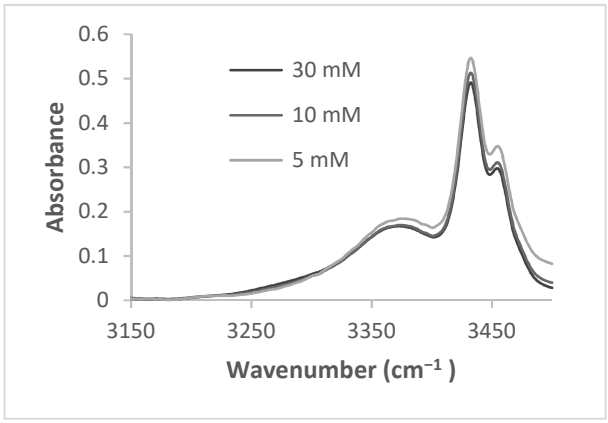


Figure S5.2: Solution state IR absorption spectra of each compound 1-4 in the NH stretch region at three different concentrations.

S5.2 ^{1}H NMR chemical shifts of NH signals

Figure S5.3 shows an expansion of the spectral window where the NH proton signals of compounds **1-4** are located. Solutions were 5 mM in $CDCl_3$; spectra were recorded at 300 K on a Bruker 400 MHz spectrometer.

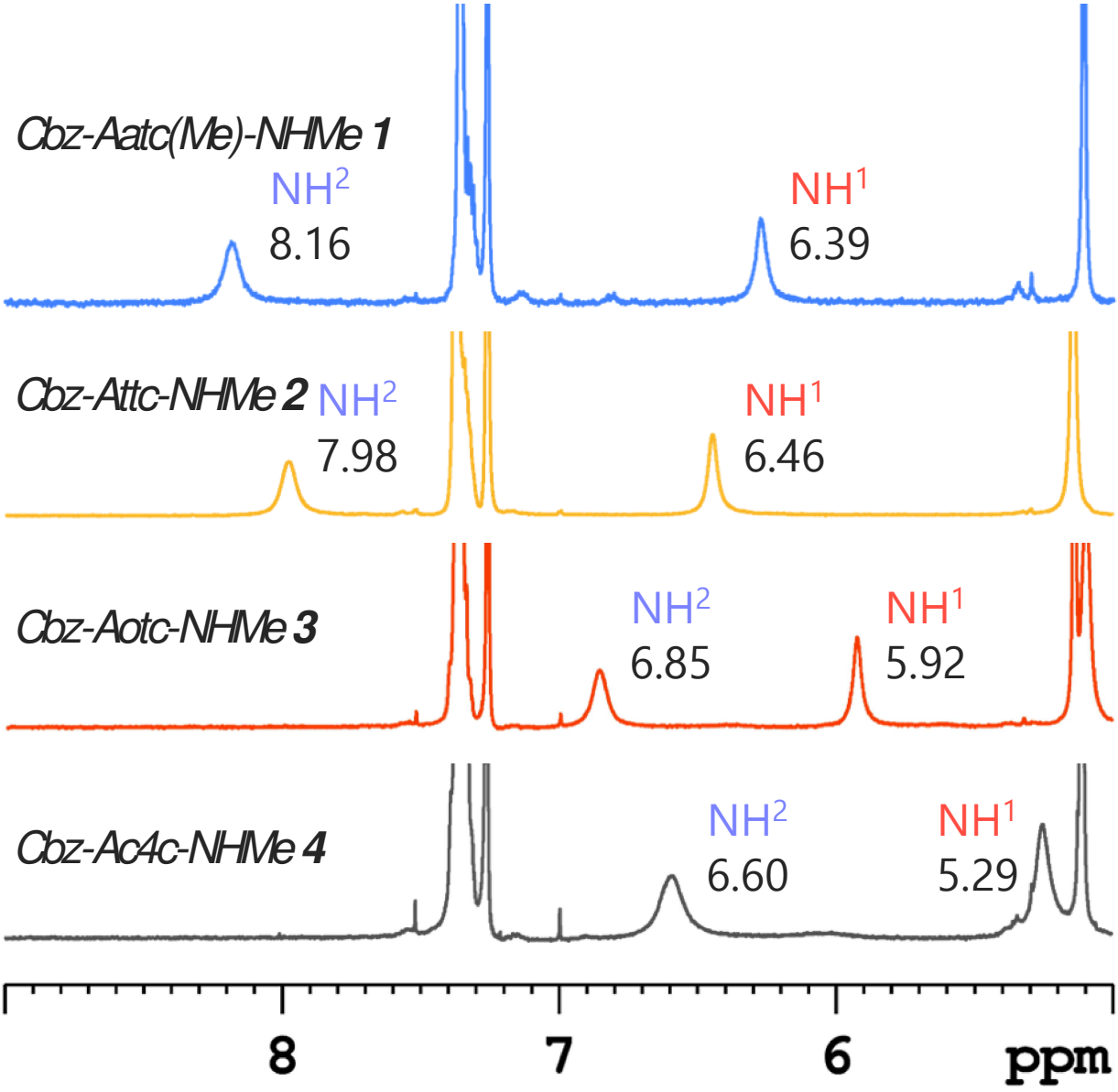
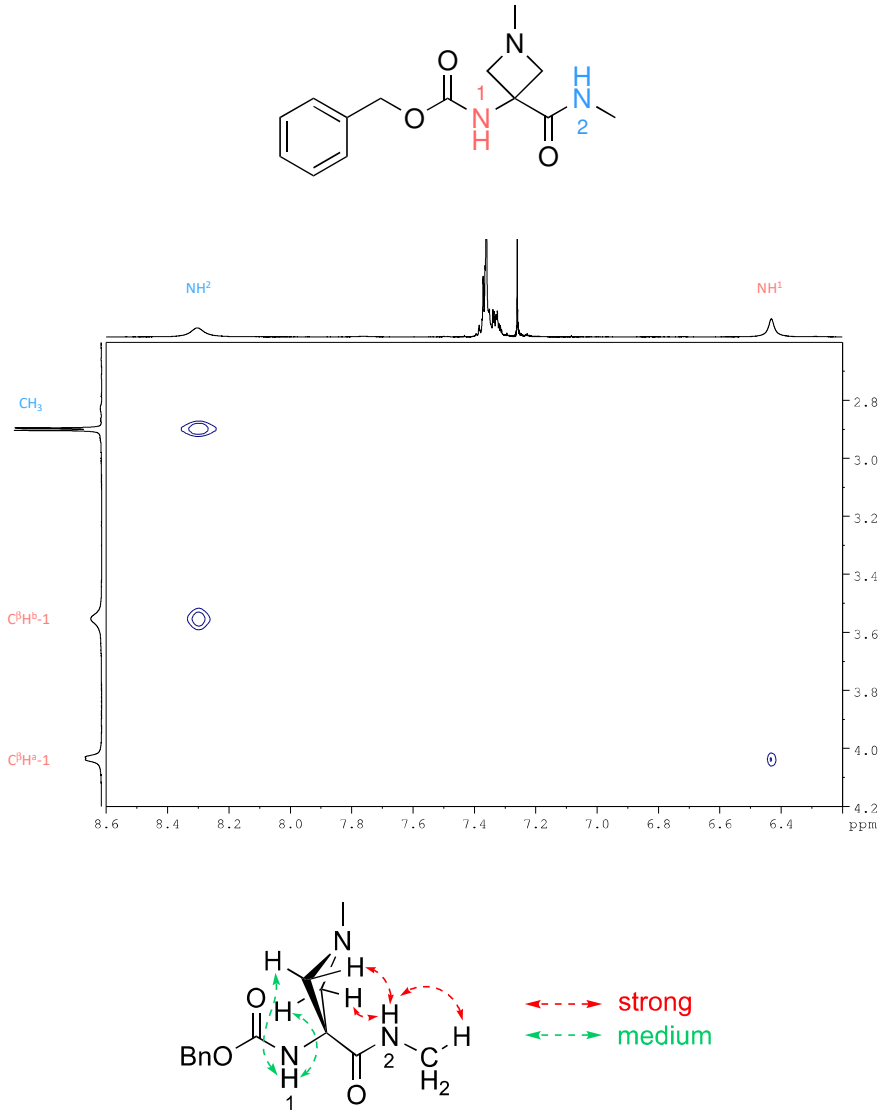


Figure S5.3: ^{1}H NMR chemical shifts of NH signals of compounds 1-4.

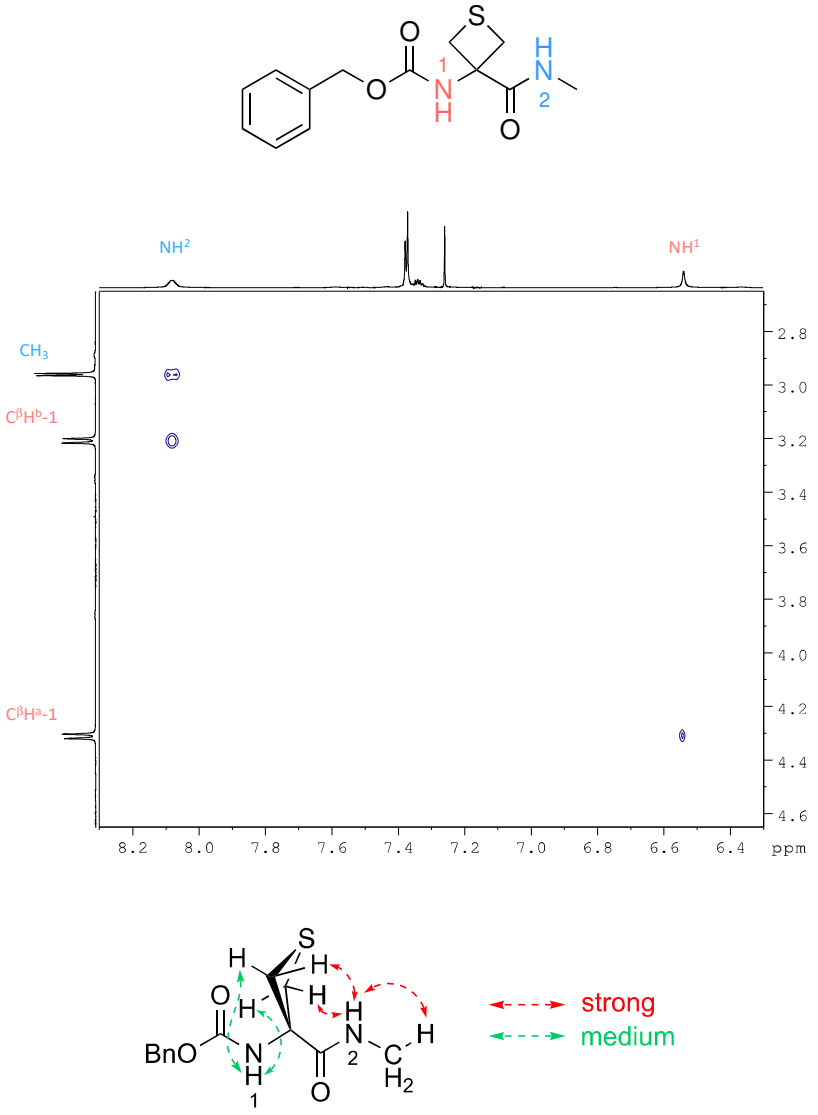
S5.3 NOESY experiments

A NOESY experiment was performed on each compound **1-4** (20 mM solution in $CDCl_3$); data were acquired at 273 K on a Bruker 600 MHz spectrometer. The pulse sequence was noesygpph, and the mixing time was 600 ms. The experiment was performed by collecting 1024 points in f2 and 512 points in f1. Data are presented in Figures S5.4 to S5.7.



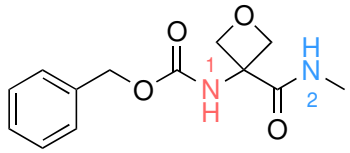
NH	NOE correlation	Strength (by volume) of NOE correlation
NH^1	$C^\beta H^a$	medium
NH^2	$C^\beta H^b$	strong
	CH_3	strong

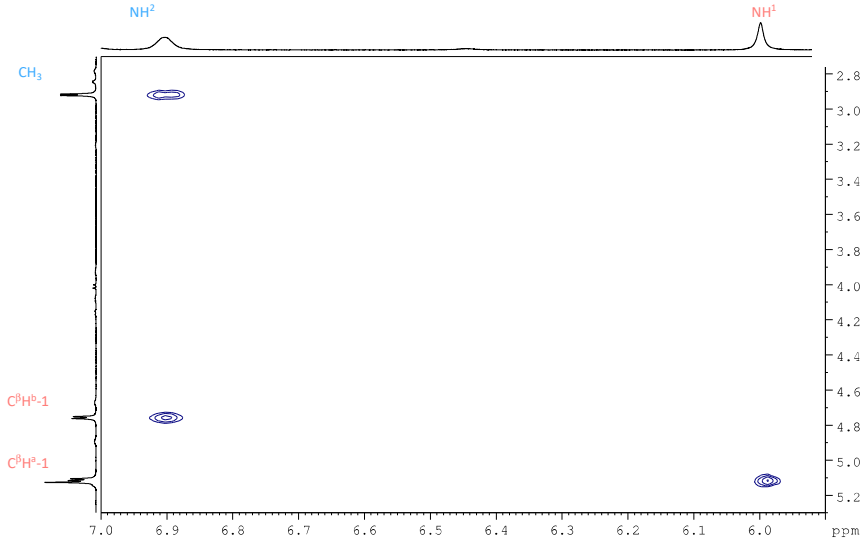
Figure S5.4: NOESY data for Cbz-Aatc(Me)-NHMe, compound 1.

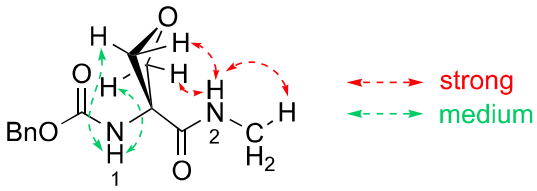


NH	NOE correlation	Strength (by volume) of NOE correlation
NH^1	$C^\beta H^a$	medium
NH^2	$C^\beta H^b$	strong
	CH_3	strong

Figure S5.5: NOESY data for Cbz-Attc-NHMe, compound 2.

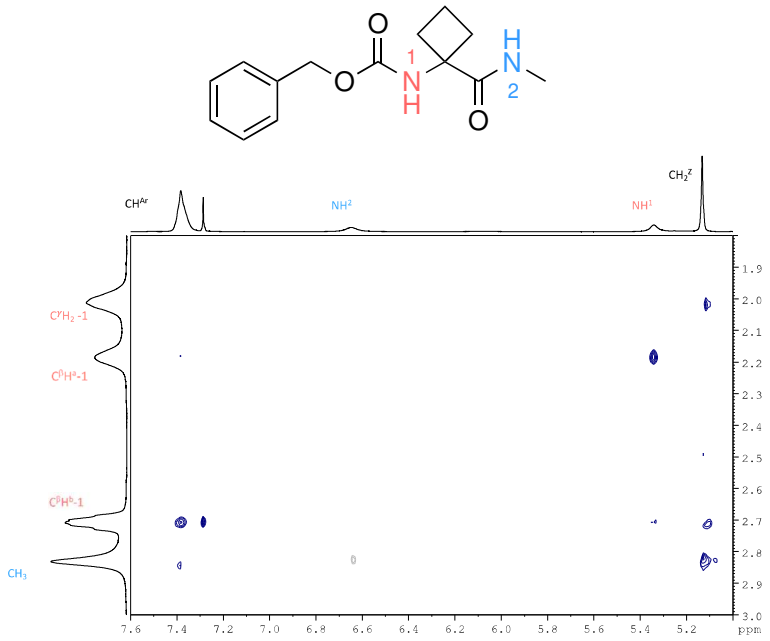






NH	NOE correlation	Strength (by volume) of NOE correlation
NH^1	$C^\beta H^a$	medium
NH^2	$C^\beta H^b$	strong
	CH_3	strong

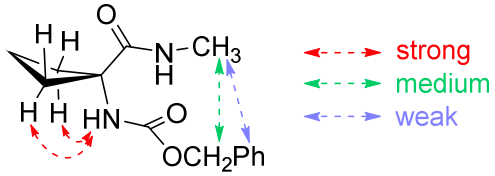
Figure S5.6: NOESY data for Cbz-Aotc-NHMe, compound 3.



proton	NOE correlation	Strength (by volume) of NOE correlation
NH^1	C$^{\beta}$H^a	strong
	C$^{\beta}$H^b	weak
NH^2	CH$_3$	weak
CH$_2^{Cbz}$	C$^{\gamma}$H$_2$	medium
	C$^{\beta}$H^b	weak
	CH$_3$	medium
CH^{Ar}	C$^{\beta}$H^b	medium
	CH$_3$	weak

Figure S5.7: NOESY data for Cbz-Ac4c-NHMe, compound 4.

The NOESY data for compound **4** are presented here for completeness, although interpretation should be made with caution since IR data suggests that at least two interconverting conformer families are expected to be present. Nonetheless, the lack of correlation between NH^2 and cyclobutane C$^{\beta}$ protons, as well as non-negligible correlations between the benzyl moiety protons and the methyl group protons (depicted below), suggest contributions from folded conformations possibly including (but not necessarily requiring) a C7 form or an α-form, such as those indicated by the IR studies.



S6. References

[1] 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 and D. J. Aitken, *Chem. Sci.*, 2020, **11**, 9191-9197.

[2] Macromodel Release 2019-3, Schrödinger, LLC: New York, NY.

[3] K. Eichkorn, F. Weigend, O. Treutler and R. Ahlrichs, *Theor. Chem. Acc.*, 1997, **97**, 119-124.

[4] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

[5] D. Rappoport and F. Furche, *J. Chem. Phys.*, 2010, **133**, 134105.

[6] Turbomole V7.2 2017, Turbomole, a development of University of Karlsruhe and Forschungszentrum Karlsruhe, 1989-2007, Turbomole GmbH, since 2007; available from http://www.turbomole.com:

[7] 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 and D. J. Fox Gaussian 16, Revision C.01; Gaussian, Inc., Wallingford CT.:

[8] M. Alauddin, H. S. Biswal, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2169-2178.

[9] E. Gloaguen, B. de Courcy, J. P. Piquemal, J. Pilme, O. Parisel, R. Pollet, H. S. Biswal, F. Piuzzi, B. Tardivel, M. Broquier and M. Mons, *J. Am. Chem. Soc.*, 2010, **132**, 11860-11863.

[10] E. Gloaguen, M. Mons, K. Schwing and M. Gerhards, *Chem. Rev.*, 2020, **120**, 12490–12562.

[11] E. Gloaguen and M. Mons, *Top. Curr. Chem.*, 2015, **364**, 225-270.

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

[13] H. S. Biswal, E. Gloaguen, Y. Loquais, B. Tardivel and M. Mons, *J. Phys. Chem. Lett.*, 2012, **3**, 755-759.

[14] A. Aubry, J. Protas, G. Boussard, M. Marraud and J. Néel, *Biopolymers*, 1978, **17**, 1693-1711.

[15] Y. Paterson, E. R. Stimson, D. J. Evans, S. J. Leach and H. A. Scheraga, *Int. J. Pept. Protein Res.*, 1982, **20**, 468-480.

[16] C. P. Rao, R. Nagaraj, C. N. R. Rao and P. Balaram, *Biochemistry*, 1980, **19**, 425-431.

[17] R. W. Newberry and R. T. Raines, *Nat. Chem. Biol.*, 2016, **12**, 1084-1089.

[18] M. Avignon, P. V. Huong, J. Lascombe, M. Marraud and J. Néel, *Biopolymers*, 1969, **8**, 69-89.

[19] M. A. Broda, B. Rzeszotarska, L. Smelka and M. Rospenk, *J. Pep. Res.*, 1997, **50**, 342-351.

[20] M. T. Cung, M. Marraud and J. Néel, *Anal. Chim. (France)*, 1972, **7**, 183-209.

[21] R. R. Gardner and S. H. Gellman, *J. Am. Chem. Soc.*, 1995, **117**, 10411-10412.

[22] J. H. Yang and S. H. Gellman, *J. Am. Chem. Soc.*, 1998, **120**, 9090-9091.

[23] S. Habka, W. Y. Sohn, V. Vaquero-Vara, M. Geleoc, B. Tardivel, V. Brenner, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3411-3423.

[24] W. Y. Sohn, S. Habka, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2017, **19**, 17128-17142.

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