Glycopeptides

DOI: 10.1002/anie.200602494

# Saccharide-Induced Peptide Conformation in Glycopeptides of the Recognition Region of LI-Cadherin**

*Axel Kuhn and Horst Kunz**

*Dedicated to Professor Dieter Hoppe on the occasion of his 65th birthday*

Glycoproteins of the outer cell membranes are components of the glycocalix involved in fundamental recognition processes.[1] Among them, the cadherins are crucial to cell adhesion and to the formation of tissues. The epithelial E-cadherin, having five extracellular consensus repeat domains,[2] also plays an important role in cell differentiation.[3] Its interaction is homotypic and homophilic. The homophilic recognition site identified by X-ray analysis and multidimensional NMR spectroscopy is located in the N-terminal domain and consists of three antiparallel β sheets—βC, βF, and βG—arranged around the central recognition motif HAV (His-Ala-Val, Figure 1 a).[4,5]

Glycopeptide partial sequences of the homophilic recognition region of E-cadherin were shown to induce differentiation in transformed but E-cadherin-expressing keratinocytes.[6] The effect proved dependent upon the conformation of the glycopeptide and was only observed if the saccharide forced the peptide backbone to adopt a turnlike conformation (Figure 1 a). Except for the amino acid sequence not much is known about the structure of LI-cadherin,[7] which was discovered in 1994 in rat liver and intestine. LI-cadherin mediates the $Ca^{2+}$-dependent cell adhesion[8] and is involved in intestine development as well as in differentiation processes.[9] In the N-terminal domain a recognition motif Ala-Ala-Leu was identified.[7] It can be concluded from sequence comparison that LI-cadherin has a homophilic recognition region similar to that in E-cadherin, in which a turn sequence (SQG) C-terminally follows the recognition motif AAL (Figure 1 b). To prove this hypothesis, glycopeptide sequences **I** ($L^{95}$AALDSQGAIV$^{105}$) of the LI-cadherin recognition region[7] were synthesized and subjected to conformational analysis by NMR spectroscopy.

The saccharide parts of these glycopeptides were systematically varied in the form of tumor-associated mucin antigens in order to determine the effect of type and size of the saccharide upon the peptide conformation. This appeared attractive, as no conclusions regarding a saccharide-dependant effect could be drawn from previous conformational studies on single mucin glycopeptides.[10–15]

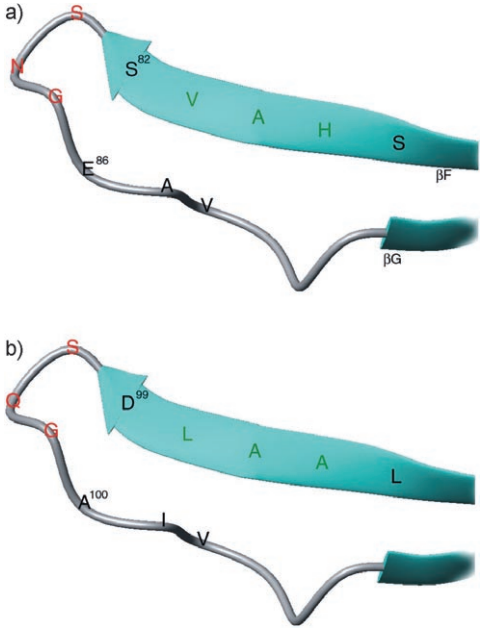



***Figure 1.*** Homophilic recognition region of a) E-cadherin (based on X-ray crystal structure) and b) LI-cadherin (postulated, based on sequence homologies).

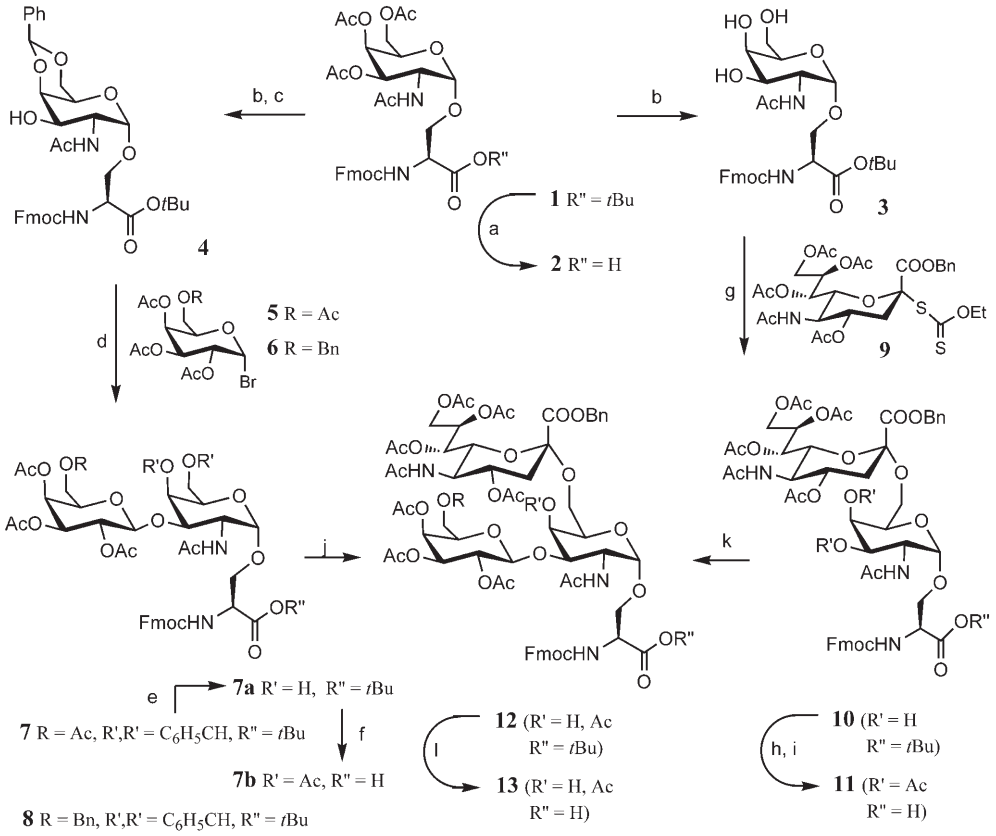
The 9-fluorenylmethoxycarbonyl (Fmoc)-protected *O*-glycosyl serine building blocks were synthesized from Fmoc-serine *tert*-butyl ester[16] via the $T_N$ antigen serine derivative **1**[17] (Schemes 1 and 2). Its selective deprotection led to compound **3**,[18] which served as the substrate for the construction of all extended saccharide side chains. All glycosylation reactions and manipulations of protecting groups must be carried out preserving both the base-labile Fmoc group and the acid-labile *tert*-butyl ester. The 4,6-benzylidene acetal **4** was formed under controlled acid catalysis. Subsequent Helferich glycosylation using the galactosyl bromides **5** and **6** afforded the T antigen serine derivatives **7** and **8**, respectively. Selective hydrolysis of the benzylidene acetal **7** to give **7a** was achieved by treatment with hot aqueous acetic acid. Acetylation and treatment with trifluoroacetic acid yielded the protected T antigen serine building block **7b** (Scheme 1).

Analogous acidolysis (step a in Scheme 1) converted **1** into the $T_N$ antigen serine derivative **2**. Xanthate **9** of sialic acid benzyl ester was coupled regio- and stereoselectively to

[*] Dr. A. Kuhn, Prof. Dr. H. Kunz
Institut für Organische Chemie
Universität Mainz
Duesbergweg 10–14, 55128 Mainz (Germany)
Fax: (+49) 6131-392-4786
E-mail: hokunz@mail.uni-mainz.de

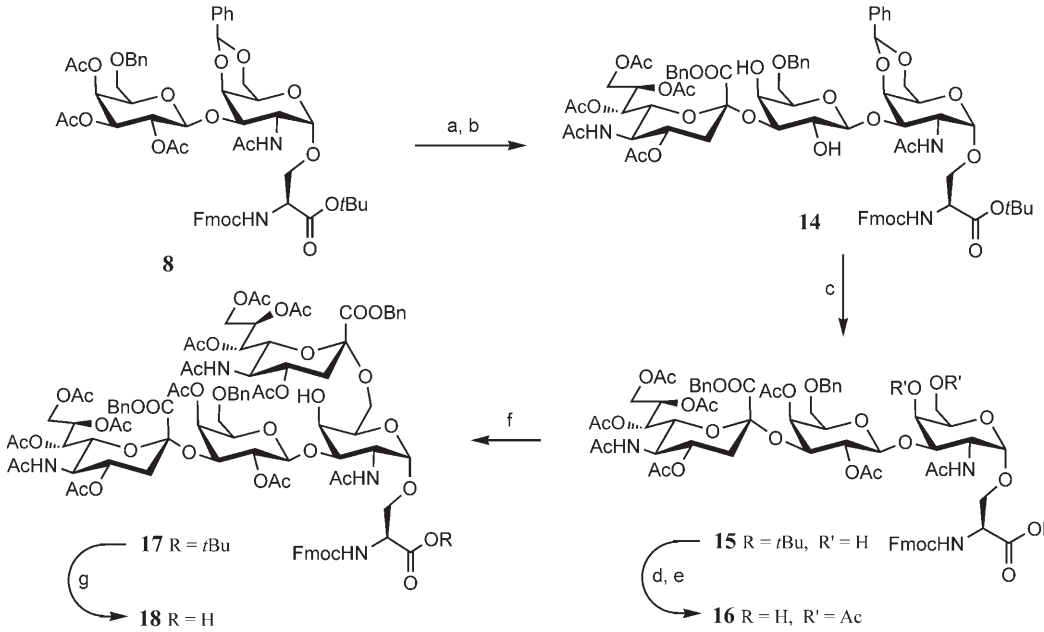
[**] This work was supported by the Volkswagen-Stiftung.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.



***Scheme 1.*** a) TFA/$CH_2Cl_2$ 1:1, anisole, 6 h, RT, 73% (HPLC); b) $NaOCH_3$, $CH_3OH$, pH 8.5, RT, 72%; c) benzaldehyde dimethyl acetal, *p*-toluenesulfonic acid, pH 4, $CH_3CN$, RT, 63–77%; d) **5**/**6**, $Hg(CN)_2$, $CH_2Cl_2$/$CH_3NO_2$, 14 h, RT, 66% (**7**), 91% (**8**); e) AcOH/$H_2O$ 4:1, 85 °C, 1 h, 64–90%; f) 1. $Ac_2O$/pyridine, 0 °C/4 h→14 h, RT, 97%; 2. TFA, anisole, 1.5 h, RT, 53% (HPLC); g) **9**, AgOTf, $CH_3SBr$, $CH_3CN/CH_2Cl_2$, −65 °C, 38–44%; h) $Ac_2O$/pyridine, 0 °C/4 h→14 h, RT, 99%; i) TFA/$CH_2Cl_2$ 1:1, anisole, 6 h, RT, 99%; j) **9**, AgOTf, $CH_3SBr$, $CH_3CN/CH_2Cl_2$, −65 °C, 20% (R = Ac), 71% (R = Bn); k) **5**, $Hg(CN)_2$, $CH_2Cl_2/CH_3NO_2$, 14 h, RT, 44%; l) R = Ac: 1. $Ac_2O$/pyridine, 14 h, RT, 99%; 2. TFA, anisole, 1.5 h, RT, 70% (HPLC); R = Bn: TFA, anisole, 1.75 h, RT, 61% (HPLC). Bn = benzyl, OTf = trifluoromethanesulfonate, TFA = trifluoroacetic acid.

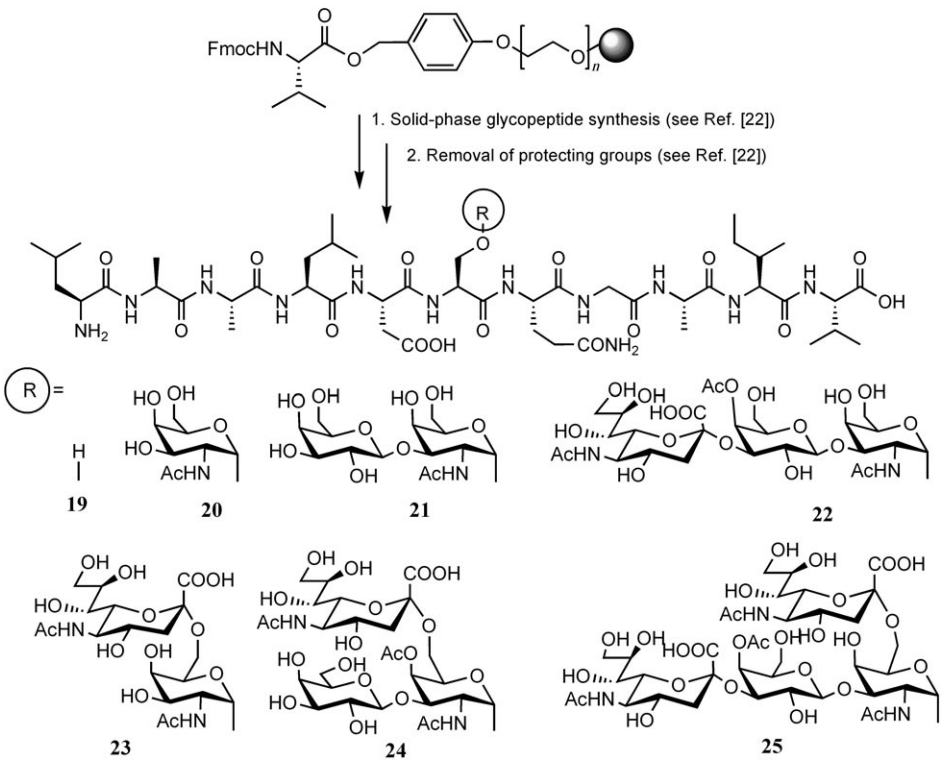
the partially unprotected $T_N$ antigen and T antigen serine derivatives **3** and **7a** at −65 °C to give the sialyl-$T_N$ antigen (**10**) and 2,6-sialyl-T antigen (**12**) structures, respectively. After O-acetylation and acidolysis of the *tert*-butyl esters the sialyl-$T_N$ (**11**) and 2,6-sialyl-T antigen (**13**) serine building blocks were obtained (Scheme 1). Alternative formation of the 2,6-sialyl-T structure by Helferich glycosylation of **10** (step k in Scheme 1) afforded **12** with lower yield than that obtained for the sialylation of **7a**. The 2,3-sialyl-T derivative required for the complete set of tumor-associated mucin saccharide antigens was synthesized from precursor **8** (Scheme 2). Its O-deacetylation with catalytic sodium methoxide in methanol proceeded very slowly even at pH 9.5,[19] showing that more rapid transesterification reactions (e.g. of **1**) proceed first by O-acetyl migration to the 6-position.



***Scheme 2.*** a) $NaOCH_3$, $CH_3OH$, pH 9.5, 24 h, RT, 39%; b) **9**, AgOTf, $CH_3SBr$, $CH_3CN/CH_2Cl_2$, 3 h/−50 °C→14 h/−30 °C, 49%; c) 1. $Ac_2O$/pyridine, 14 h, RT, 95%; 2. AcOH/$H_2O$ 4:1, 85 °C, 1 h, 76%; d) $Ac_2O$/pyridine, 14 h, RT, 99%; e) TFA, anisole, 1 h, RT, 85% (HPLC); f) **9**, AgOTf, $CH_3SBr$, $CH_3CN/CH_2Cl_2$, 3 h/−50 °C→14 h/−30 °C, 36%; g) TFA, anisole, 1.5 h, RT, 93%.

The subsequent regio- and stereoselective sialylation with the xanthate **9** at −50 °C stereoselectively produced the 2,3-sialyl-T serine derivative **14**. After acetylation of the free hydroxy groups of the galactose portion and solvolysis of the benzylidene acetal, the partially unprotected galactosamine derivative **15** was furnished. Its O-acetylation and acidolysis gave the 2,3-sialyl-T serine building block **16**. By selective sialylation of **15**, again using xanthate **9** at −50 °C, the most complex structure, the glycophorin antigen **17**, was formed. Subsequent acidolysis of the *tert*-butyl ester resulted in the formation of building block **18**. The solid-phase syntheses of LI-cadherin glycopeptides **I** were carried out using a Wang anchor[20] on Tentagel resin[21] preloaded with Fmoc valine according to published procedures[22] (Scheme 3).




***Scheme 3.*** Solid-phase syntheses of glycopeptides **20**–**25** containing the amino sequence LAALDSQGAIV.

For comparison, the non-glycosylated LI-cadherin peptide **19** was also synthesized using the same procedures. The yields of products isolated after acidolytic cleavage from the solid support as well as the yields of peptide **19** and glycopeptides **20**–**25**[23] after cleavage of the remaining benzyl groups by hydrogenolysis, removal of the O-acetyl groups by transesterification, and purification by HPLC are listed in Table 1.

To analyze their preferred conformation, peptide **19** and glycopeptides **20**–**25** were investigated by NOESY and ROESY NMR spectroscopy in [$D_6$]DMSO solution. For discussion of the most important NOE contacts, the amino acids of the LI-cadherin sequence **I** are designated from the N terminus $L^1$ to the C terminus $V^{11}$. Based on the observed

***Table 1:*** Yields of solid-phase synthesis of peptide **19** and glycopeptides **20**–**25**.[a]

| Building block | – | **2** | **7b** | **11** | **13** | **16** | **18** |
|---|---|---|---|---|---|---|---|
| (Glyco-)Peptide | **19** | **20** | **21** | **23** | **24** | **22** | **25** |
| Yield [%][b] | 99 | 78 | 46 | 61 | 84 | 23 | 22 |
| Yield (overall) [%] | 99 | 57 | 23 | 24 | 57 | 17 | 21 |

[a] The glycopeptides were prepared by solid-phase synthesis using the glycosyl serine building blocks **2**, **7b**, **11**, **13**, **16**, and **18**. [b] After cleavage from solid phase.

NOE contacts,[24] a three-dimensional structure of each (glyco)peptide (Figure 2) was determined by molecular dynamics calculations and energy minimization by an MM2 force field method.[25,26] According to published procedures, the found dihedral angles within the turn region (Table 2) were compared to the ideal dihedral angles reported for the known turn types.[27] The deviations from these turn types are given in Table 2 as the sums of the differences in these angles.

From these values the following conclusions were drawn (Figure 2). Peptide **19** adopts a looplike conformation (NH contacts in the sequence DSQ) in the vicinity of the turn sequence DSQG. It does not resemble any ideal turn conformation (contacts from NH-$A^9$ to α- and γ-CH-$L^4$). The structure is reminiscent of a beginning helix that is not continued towards the C terminus. The C terminus folds back to the turn sequence (α-CH-$Q^7$ to α-CH-$V^{11}$), resulting in a kind of double loop. The situation for the $T_N$-antigen glycopeptide **20** is similar. Nevertheless, comparison of the dihedral angles (Table 2) suggests a β III′ turn formed within the sequence LDSQ, which, however, cannot assume a stable β-sheet structure (contacts between α-CH-$L^4$ and α-CH-$V^{11}$, and between γ-CH-$L^4$ and α-CH-$I^{10}$).

The conformation of the sequence DSQG of T-antigen glycopeptide **21** is most like a β-II turn (Table 2, Figure 2). The saccharide apparently stabilizes this conformation (contacts between CH-NAc and α-CH-$D^5$, α-CH-$S^6$, and NH-$D^5$). The N-terminal sequence and the C-terminal sequence adopt a conformation similar to a β sheet (γ-CH-$L^4$ and γ-CH-$I^{10}$, β-CH-$L^4$ and NH-$V^{11}$, and NH-$D^5$ and β-CH-$I^{10}$).

In sialyl-T antigen LI-cadherin glycopeptides **22**, **24**, and **25** the larger saccharides exhibit a marked influence on the conformation of the peptide backbone, which according to NOE contacts apparently forms an increasingly closer loop

***Table 2:*** Dihedral angles of (glyco)peptides within the amino acid sequence LAALDSQGAIV.[a]

| | Antigen | Turn sequence | Calculated dihedral angles | | | | Deviation from the ideal turn (type β) | | | | | | (type γ) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\Phi_{i+1}$ | $\Psi_{i+1}$ | $\Phi_{i+2}$ | $\Psi_{i+2}$ | I | I′ | II | II′ | III | III′ | γ | γ′ |
| **20** | $T_N$ | LDSQ | 43.8 | 39.1 | 69.7 | 28.7 | 361 | 74 | 224 | 354 | 362 | **41** | 115 | 125 |
| **23** | $ST_N$ | DSQ | 77.1 | −63.0 | | | | | | | | | **20** | 260 |
| **21** | T | DSQG | −19.3 | 121.9 | 66.2 | −5.3 | 354 | 200 | **62** | 473 | 345 | 218 | 119 | 277 |
| **24** | (2→6)ST | DSQG | 160.4 | −2.9 | 59.0 | −0.7 | 397 | **165** | 365 | 357 | 397 | 168 | **158** | 283 |
| **22** | (2→3)ST | LDSQ | 81.9 | −36.8 | −98.0 | −168.8 | 326 | 446 | 646 | 292 | 321 | 449 | **45** | 239 |
| **25** | $S_2T$ | DSQG | 45.1 | −79.5 | −159.5 | 0.9 | 225 | 375 | 545 | **136** | 272 | 376 | **34** | 245 |

[a] The smallest sums of deviation from the ideal dihedral angles are printed in bold.

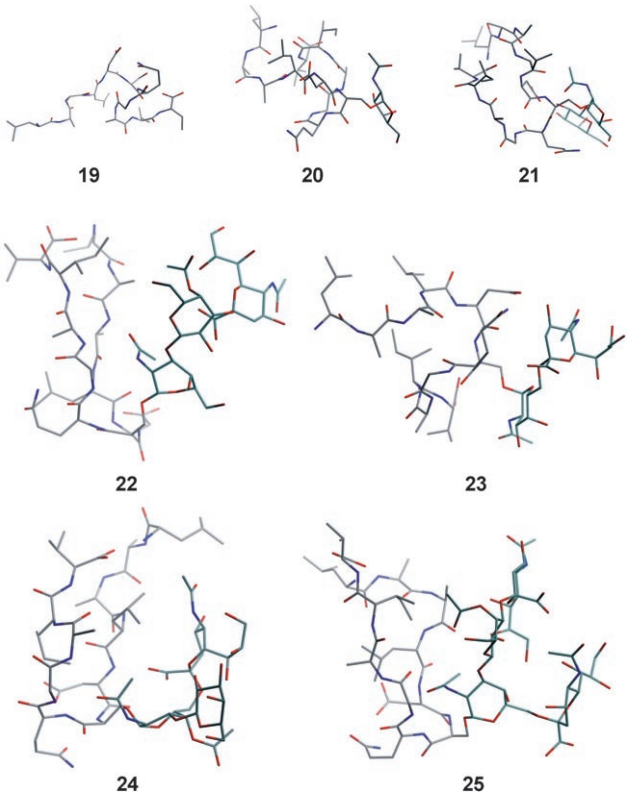



***Figure 2.*** Preferred conformations of peptide **19** and glycopeptides **20**–**25** of LI-cadherin determined by NOESY/ROESY experiments in [$D_6$]DMSO. The saccharide (more intensively colored) is positioned to the right of the corresponding peptide (C gray, O red, N blue).

with antiparallel alignment of the N- and C-terminal sequences. Contacts between NHAc and $A^9$ (in **22** and **24**) and between H-2 (GalNAc) and NH-$G^8$ (in **25**) suggest that the saccharide is oriented parallel to this loop (in Figure 2, right). From the deviations in the dihedral angles (Table 2) one can conclude that a γ-turn-like conformation that adopts an expanded and twisted form is most likely. This especially applies to 2,6-sialyl antigen glycopeptide **24**. A γ-turn conformation can also be assumed for the sequence DSQ of sialyl-$T_N$ antigen glycopeptide **23**, with backfolding of the C terminus (NOE contacts between α-CH-$Q^7$ and γ-CH-$I^{10}$, and between NH-$G^8$ and β-$A^9$) similar to **20**. For the sialyl-T glycopeptides, the interactions between the peptide sequences in the N- and C-terminal regions are typical (**22**: between α-CH-$A^3$ and NH-$I^{10}$/NH-$V^{11}$; **24**: between NH-$L^4$ and α-CH-$V^{11}$, between β-CH-$L^4$ and γ-CH-$I^{10}$; **25**: between β-CH-$L^4$ and γ- and δ-CH-$I^{10}$). It is characteristic for the conformations of these compounds that only few direct interactions between the saccharide and peptide parts are observable in most cases and restricted to the GalNAc group (**22**: between AcNH-GalNac and α-CH-$A^9$, between H-2-GalNAc and β-CH-$S^6$; **24**: between CHAc-GalNAc and α-CH-$D^5$/α-CH-$A^9$; **25**: between H-2-GalNAc and NH-$G^8$).

The described conformational analyses of LI-cadherin glycopeptides not only prove that the saccharide side chains exert a marked influence upon the conformation of the peptide chain (Figure 2) but also show that this effect is dependent on the type and size of the saccharide. Beyond this case,[7] it is important to note that this conformational influence was investigated for the tumor-associated mucin saccharide antigens. With a glycopeptide vaccine from tumor-associated mucin MUC1 a specific immune response in mice could be induced. The induced antibody reacted neither with the non-glycosylated peptide of identical sequence nor with the tumor-associated saccharide antigen (sialyl-$T_N$), but only with the glycopeptide,[28] probably indicating that the peptide chain adopts the recognized conformation only within the glycopeptide.

Received: June 21, 2006
Revised: September 20, 2006
Published online: December 8, 2006

**Keywords:** cadherins · conformational analysis · glycopeptides · solid-phase synthesis

[1] a) N. Jentoft, *Trends Biochem. Sci.* **1990**, *15*, 291; b) J. Dekker, J. W. A. Rossen, H. A. Büller, A. W. C. Einerhand, *Trends Biochem. Sci.* **2002**, *27*, 126.
[2] M. Takeichi, *Science* **1991**, *251*, 1451.
[3] a) O. Huber, C. Bierkamp, R. Kemler, *Curr. Opin. Cell Biol.* **1996**, *8*, 685; b) A. W. Koch, S. Pokutta, A. Lustig, J. Engel, *Biochemistry* **1997**, *36*, 3507.
[4] M. Overduin, T. S. Harvey, S. Bagby, K. I. Tong, P. Yan, M. Takeichi, M. Ikura, *Science* **1995**, *267*, 386.
[5] B. Nagar, M. Overduin, M. Ikura, J. M. Rini, *Nature* **1996**, *380*, 360.
[6] J. Habermann, K. Stüber, T. Skripko, T. Reipen, R. Wieser, H. Kunz, *Angew. Chem.* **2002**, *114*, 4423; *Angew. Chem. Int. Ed.* **2002**, *41*, 4249.
[7] D. Berndorff, R. Gessner, B. Kreft, N. Schnoy, A.-M. Lajous-Petter, N. Loch, W. Reutter, M. Hortsch, R. Tauber, *J. Cell Biol.* **1994**, *125*, 1353.

[8] B. Kreft, D. Berndorff, A. Böttinger, S. Finnemann, D. Wedlich, M. Hortsch, R. Tauber, *J. Cell Biol.* **1997**, *136*, 1109.

[9] a) D. G. Silberg, G. P. Swain, E. R. Suh, P. G. Traber, *Gastroenterology* **2000**, *119*, 961; b) C. Grötzinger, J. Kneifel, D. Patschan, N. Schnoy, I. Anagnostopoulos, S. Faiss, R. Tauber, B. Wiedenmann, R. Geßner, *Gut* **2001**, *49*, 73; c) M. Takamura, M. Sakamoto, Y. Ino, T. Shimamura, T. Ichida, S. Hirohashi, *Cancer Sci.* **2003**, *94*, 425.

[10] a) M. J. Scanlon, S. D. Morley, D. E. Jackson, M. R. Price, S. B. J. Tendler, *Biochem. J.* **1992**, *284*, 137; b) P. Braun, G. M. Davies, M. R. Price, P. M. Williams, S. B. J. Tendler, H. Kunz, *Bioorg. Med. Chem.* **1998**, *6*, 1531.

[11] a) L. Kinarsky, M. Nomoto, Y. Ikematsu, H. Hassan, E. P. Bennett, R. L. Cerny, H. Clausen, M. A. Holtingworth, S. Sherman, *Biochemistry* **1998**, *37*, 12811; b) S. Mensdorff-Pouilly, L. Kinarsky, K. Engelmann, S. E. Baldus, R. H. Verheien, M. A. Hollinworth, V. Pisarev, S. Sherman, F.-G. Hanisch, *Glycobiology* **2005**, *15*, 735.

[12] D. M. Coltart, A. K. Royyuru, L. J. Williams, P. W. Glunz, D. Sames, S. D. Kuduk, J. B. Schwarz, X.-T. Chen, S. J. Danishefsky, D. H. Live, *J. Am. Chem. Soc.* **2002**, *124*, 9833.

[13] a) J. Schuman, A. P. Campbell, R. R. Koganly, B. M. Longenecker, *J. Pept. Res.* **2003**, *61*, 91; b) J. S. Grénstead, J. T. Schuman, A. P. Campbell, *Biochemistry* **2003**, *42*, 14293.

[14] M. Liu, B. Acress, J.-M. Balloul, N. Bizonarne, S. Paul, P. Slos, P. Squiban, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14567.

[15] a) S. Dziadek, C. G. Espinola, H. Kunz, *Aust. J. Chem.* **2003**, *56*, 519; b) S. Dziadek, C. Griesinger, H. Kunz, U. M. Reinscheid, *Chem. Eur. J.* **2006**, *12*, 4981.

[16] H. Kunz in *Preparative Carbohydrate Chemistry* (Ed.: S. Hanessian), Marcel Dekker, New York, **1997**, pp. 265–281, in particular pp. 274–275.

[17] H. Paulsen, K. Adermann, *Liebigs Ann. Chem.* **1989**, 751.

[18] T. Reipen, H. Kunz, *Synthesis* **2003**, 2487.

[19] C. Brocke, H. Kunz, *Synlett* **2003**, 2052.

[20] S. S. Wang, *J. Am. Chem. Soc.* **1973**, *95*, 1328.

[21] E. Bayer, W. Rapp, *Chem. Pept. Proteins* **1986**, *3*, 3.

[22] C. Brocke, H. Kunz, *Synthesis* **2004**, 525.

[23] $[\alpha]$ values ($c$ in DMSO) and $^{13}C$ NMR (103.0 MHz, $[D_6]$DMSO): **20**: $[\alpha] = -1.1\ \mathrm{deg\,cm^3\,g^{-1}\,dm^{-1}}$ ($c = 2\ \mathrm{g\,cm^{-3}}$), $\delta = 97.6$ ppm (C1); **21**: $[\alpha] = +8.1\ \mathrm{deg\,cm^3\,g^{-1}\,dm^{-1}}$ ($c = 1.5\ \mathrm{g\,cm^{-3}}$), $\delta =$ 104.2 ppm (C1″, Gal), 97.8 ppm (C1); **22**: $[\alpha] = -9.5\ \mathrm{deg\,cm^3\,g^{-1}\,dm^{-1}}$ ($c = 0.9\ \mathrm{g\,cm^{-3}}$); $\delta = 57.3$ ($Val^{\alpha}$), 57.2 ($Ser^{\alpha}$), 22.5 (NAc, GalNAc), 22.4 ppm (NAc, NeuNAc). **23**: $[\alpha] = -0.5\ \mathrm{deg\,cm^3\,g^{-1}\,dm^{-1}}$ ($c = 1.2\ \mathrm{g\,cm^{-3}}$), $\delta = 97.5$ (C1), 56.8 ppm ($Val^{\alpha}$); **24**: $[\alpha] = +6.3\ \mathrm{deg\,cm^3\,g^{-1}\,dm^{-1}}$ ($c = 3.6\ \mathrm{g\,cm^{-3}}$), $\delta = 103.7$ (C1, Gal), 97.5 ppm (C1, GalNAc); **25**: $[\alpha] = 8.3\ \mathrm{deg\,cm^3\,g^{-1}\,dm^{-1}}$ ($c = 1\ \mathrm{g\,cm^{-3}}$), $\delta = 101.3$ (C1; Gal), 97.7 ppm (C1, GalNAc).

[24] Contacts used for calculations are listed in the Supporting Information. The influence of the solvent ($D_2O$/$[D_6]$DMSO) is also discussed. Since the NMR spectra of the glycopeptide Ac-AALDSC(GalNAc)QGAI-OH (similar to compound **20**) in $D_2O$ and in $[D_6]$DMSO showed no important differences, all NMR experiments of **19**–**25** were carried out in $[D_6]$DMSO solution.

[25] U. Burkert, N. L. Allinger, *Molecular Mechanics American Chem. Soc.*, Washington, DC, **1982**.

[26] Chem3D. Version 9.0, CambridgeSoft **1986**–**2004**, Cambridge, MA, USA.

[27] a) J. S. Richardson, *Adv. Protein Chem.* **1981**, *34*, 167; b) D. S. Wishart C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J. Biomol. NMR* **1995**, *5*, 67.

[28] S. Dziadek, A. Hobel, E. Schmitt, H. Kunz, *Angew. Chem.* **2005**, *117*, 7803; *Angew. Chem. Int. Ed.* **2005**, *44*, 7630.