

NMR-based determination of the binding epitope and conformational analysis of MUC-1 glycopeptides and peptides bound to the breast cancer-selective monoclonal antibody SM3

Heiko Möller¹, Nida Serttas¹, Hans Paulsen¹, Joy M. Burchell², Joyce Taylor-Papadimitriou² and Bernd Meyer¹

¹Institute of Organic Chemistry, University of Hamburg, Germany; ²Imperial Cancer Research Fund Breast Cancer Biology Group, Guy's Hospital, London, UK

Mucin glycoproteins on breast cancer cells carry shortened carbohydrate chains. These partially deglycosylated mucin 1 (MUC-1) structures are recognized by the monoclonal antibody SM3, which is being tested for its diagnostic utility. We used NMR spectroscopy to analyze the binding mode and the binding epitope of peptide and glycopeptide antigens to the SM3 antibody. The pentapeptide PDTRP and the glycopentapeptide PDT(*O*- α -D-GalNAc)RP are known ligands of the monoclonal antibody. The 3D structures of the ligands in the bound conformation were determined by analyzing trNOESY build-up rates. The peptide was found to adopt an extended conformation that fits into the binding pocket of the antibody. The binding epitopes of the ligands were determined by saturation transfer difference (STD) NMR spectroscopy. The peptide's epitope is predominantly

located in the N-terminal PDT segment whereas the C-terminal RP segment has fewer interactions with the protein. In contrast, the glycopeptide is interacting with SM3 utilizing all its amino acids. Pro1 shows the strongest binding effect that slightly decays towards Pro5. The GalNAc residue interacts mainly via the N-acetyl residue while the other protons show less interactions similar to that of Pro5. The glycopeptide in the bound state also has an extended conformation of the peptide with the carbohydrate oriented towards the N-terminus. Docking studies showed that peptide and glycopeptide fit the binding pocket of the mAb SM3 very well.

Keywords: glycopeptide antibody complex; STD NMR; breast cancer; MUC-1; binding epitope.

The extracellular part of the epithelial glycoprotein MUC-1 consists of tandem repeats of 20 amino acids (PDTRPAPGSTAPPAHGVTS_A, where the start of the tandem repeat peptide sequence varies. We follow here the definition by Gendler *et al.* who defined the start at PDTRP [1]. Residues of peptides, that were elongated at the N-terminus, are designated by an apostrophe, e.g. Ala20'-Pro1-Asp2-Thr3-Arg4-Pro5.) [2]. Each repeat can carry up to five O-glycosyl chains at Ser and Thr residues that account for the high carbohydrate content of the mucins [3]. Usually, 70–100 repeats are found in mucins. The clustering of O-linked glycans on MUC-1 leads to an extended protein core. Membrane-bound mucins extend several hundred nanometers into the lumen and thus represent a first barrier to the environment. They have important functions in cell-

cell recognition and shield the cell from microorganisms, toxins and proteolytic attack [4].

Many diseases affect the production of mucus. Both the amount and the characteristics of the mucus can be altered. In cystic fibrosis, for example, due to changes of the ionic environment dramatic alterations in rheological properties correlate with changes in carbohydrate composition [2]. Modified oligosaccharides are also found in mucins of patients with Crohn's disease [2].

Epithelial cells express the membrane-bound MUC-1 at their apical surface. In carcinomas, the localization at the apical surface is lost. High concentrations of MUC-1 spread out over the whole cell surface. This may protect the cells against low pH and may interfere with immune surveillance by causing steric hindrance of surface antigen presentation [2,4].

In breast cancer, the MUC-1 glycoprotein is overexpressed and aberrantly glycosylated. Thus, in contrast to the mucin produced by normal breast epithelial cells, which carry core2 based structures [5], MUC1 from breast cancer cells carries highly truncated, mainly core 1 based oligosaccharide structures [6,7]. In some cases, the first sugar added, *N*-acetyl galactosamine is not extended, or is sialylated to form the cancer-related sialyl Tn epitope. Because of the shorter side chains, the peptide core of the cancer mucin is more exposed, and antibodies have been developed which recognize epitopes exposed in the cancer mucin, which are normally masked by large oligosaccharide side chains. These antigenic peptide sequences therefore constitute cancer-associated epitopes which are also found in the

Correspondence to B. Meyer, Institute of Organic Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany.

Fax: + 49 (0)40 42838 2878, Tel.: + 49 (0)40 42838 5913,

E-mail: bernd.meyer@sg1.chemie.uni-hamburg.de

Abbreviations: MUC-1, mucin 1 glycoprotein; SM3, breast cancer-selective monoclonal antibody; STD NMR, saturation transfer difference NMR; trNOE, transferred nuclear Overhauser enhancement; SPR, surface plasmon resonance; SAR, structure activity relationship; MD, molecular dynamics.

short sugar side chains (e.g. T, ST and TF antigens) [8,9]. The monoclonal antibody SM3 was raised in mice against partially deglycosylated human MUC-1. It shows a high specificity to the MUC-1 of breast cancer cells. SM3 is being tested for its diagnostic value [10,11] and also has a high therapeutic potential.

The minimum peptide antigen epitope to SM3 was identified by the pepsan technique using ELISA detection. Using heptapeptides the resulting binding epitope was identified as Asp2-Thr3 [12], using octapeptides the resulting epitope was Pro1-Asp2-Thr3-Arg4-Pro5 [13] and using nona- and 20mer peptides the resulting epitope was identified as Ala20'-Pro1-Asp2-Thr3-Arg4-Pro5 and Pro1-Asp2-Thr3, respectively [14,15]. For SM3 reacting with pentamers and dimers of the MUC-1 tandem repeat the binding constants were determined by surface plasmon resonance to be $K_d = 6.25 \times 10^{-8}$ to 4.5×10^{-7} M [16].

Previous NMR studies of peptide and glycopeptide fragments of MUC-1 containing the amino-acid sequence PDTRP in the central part which were carried out in solution without an antibody present reveal that this sequence motif seems to adopt a knob-like or bent structure [17] (J. Dojahn, C. Diotel, M. Paulsen and B. Meyer, unpublished results). It was postulated that this knob-like structure renders this region especially accessible to protein interactions necessary for stimulation of immune responses. Also in this context, the oligosaccharides attached to Thr3 are most accessible for interaction with the cells of the immune system.

It is not clear against what actual epitope SM3 was developed. The antibody binds more strongly to a MUC-1 that has only part of its carbohydrate chains removed [10]. It was later shown on a molecular level that a small oligosaccharide attached to Thr3 enhances binding affinity of glycopeptides to the antibody [18].

Conventional pepsan analysis however, does not allow easy analysis of the contribution of the carbohydrate portion. To assess the involvement of carbohydrates in antibody recognition of glycosylated structures numerous glycopeptides would have to be synthesized and even this approach would not directly reveal what part of the oligosaccharides interacts with the protein.

Dokurno *et al.* published an X-ray structure analysis of SM3 complexed with the MUC-1 peptide TSA PDTRPAPGST [19]. At each end of the antigenic peptide PDTRP two additional amino acids are resolved in the X-ray crystal structure, while Thr18', Gly8-Ser9-Thr10 and the side chain of Ser19' are disordered in the crystal. The amino acids of the peptide sequence (S)APDTRPAP have many interactions with the antibody's surface. The covered surface of the individual amino acids varies strongly. While Ser19'-Ala20', Thr3 and Pro5-Ala6 have relatively small contact areas to the protein, Pro1-Asp2 and Arg4 are much more buried by the antibody.

NMR spectroscopy can be used to assess binding properties of ligands under near physiological conditions in solution by a variety of methods, e.g. trNOEs [20], STD NMR [21–23], and SAR by NMR [24]. TrNOE spectra can also be used to elucidate the 3D structure of the bound ligand. Saturation transfer difference (STD) NMR is a technique that can be used to characterize and identify binding [21–23]. It can also be used to identify the binding epitope of ligands to a protein receptor [21]. This feature can

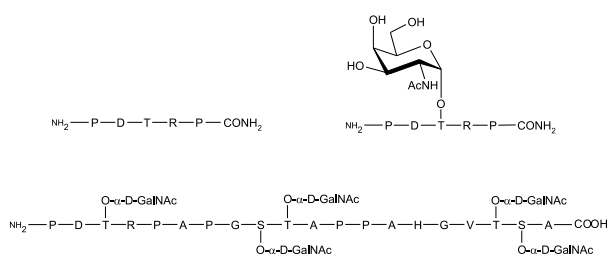


Fig. 1. Pentapeptide and glycopentapeptide used in NMR studies and completely glycosylated MUC-1 repetitive unit.

be used to quickly identify the binding contribution from either peptide or carbohydrate, especially in the case of glycopeptides. In contrast to conventional methods only one substrate is necessary to obtain that information.

Here, we present the STD NMR epitope mapping and trNOE-based conformational analysis of the MUC-1 peptide PDTRP and the MUC-1 glycopeptide PDT(*O*- α -D-GalNAc)RP (cf. Figure 1) bound to the monoclonal antibody SM3.

MATERIALS AND METHODS

Chemicals

Chemicals for peptide synthesis were obtained from PerSeptive Biosystems (Wiesbaden, Germany), acetonitrile from Alfa (Karlsruhe, Germany), triisopropylsilane and D₂O from Sigma Aldrich (Steinheim, Germany), all other chemicals of analytical grade were obtained from Merck (Darmstadt, Germany). The glycopeptides [29] and the monoclonal antibody SM3 [1] were prepared as described.

NMR Experiments

All spectra were recorded on Bruker DRX 500 spectrometer with a triple resonance 5 mm inverse probe head. For trNOE experiments with PDTRP the sample contained 3.6 mg of SM3 (M_r 156 kDa, 23 nmol, 38 μ M) and 270 μ g of PDTRP (M_r 583.64 g \cdot mol⁻¹, 460 nmol, 760 μ M) in 600 μ L NaCl/P_i (buffer concentration 20 mM, H₂O/D₂O = 9 : 1, 0.05% NaN₃) at pH 7.0. This corresponds to a ligand to protein ratio of 20 : 1. For STD experiments the ligand to protein ratio was raised to 200 : 1 (4.6 μ mol, 7.6 mM PDTRP). TrNOE studies with the glycopeptide were carried out with a sample containing 3 mg of SM3 (19.2 nmol, 32 μ M) and 300 μ g of PDT(*O*- α -D-GalNAc)RP (384 nmol, 640 μ M) in 600 μ L NaCl/P_i buffered solution (buffer concentration 20 mM) at pH 7.0. This corresponds to a ligand to protein ratio of 20 : 1. For STD experiments the ligand to protein ratio was raised to 150 : 1 (2.88 μ mol, 4.8 mM PDT(*O*- α -D-GalNAc)RP).

Peptide or glycopeptide were added to the protein solution using \approx 22 mM stock solutions. At the highest excess, this resulted in a sample dilution of 25% for the peptide and 18% for the glycopeptide. As no titration experiments were carried out, this dilution was not important for the data analysis.

Solute exchange was achieved by ultrafiltration of the 156-kDa SM3 antibody with a Centricon (Millipore) membrane having a cutoff value of 50 kDa.

All spectra were measured at 280 K. All chemical shifts are referenced to the HDO signal at 4.90 p.p.m. for ^1H . Water suppression was achieved using the WATERGATE sequence in all experiments. NMR chemical shifts of peptide and glycopeptide are listed in Tables 1 and 2, respectively. All spectra of samples containing protein were recorded with a 30 ms spin lock pulse, or so called $T_{1\rho}$ filter ($\gamma B_1 = 4680$ Hz) after the $\pi/2$ pulse, which eliminates the background protein resonances to facilitate analysis. Interpretation of the spectra were carried out with the XWINNMR (Bruker, v. 2.5) and the AURELIA program (Bruker, v. 2.1.5) on Silicon Graphics O_2 workstations. 1D STD NMR spectra were multiplied by an exponential line broadening function of 5 Hz prior to Fourier transformation. The irradiation power in all STD NMR experiments was set to ≈ 0.15 W. Selective presaturation of the protein was achieved by a train of 40 Gaussian shaped pulses of 50 ms length, each separated by a 1 ms delay, leading to a total saturation time of 2.04 s. The pulse scheme is as follows: relaxation delay, presaturation pulse train, ($\pi/2$), spin lock (where applicable), acquisition. Subtraction of the 1D STD spectra was performed internally via phase cycling after every scan to minimize temperature and magnet instability artefacts. The so called on resonance irradiation of the protein was performed at a chemical shift of -2 ppm. Off resonance irradiation was applied at 40 p.p.m., where no protein signals are present. Between 256 and 1024 total scans were collected, using 10 ppm spectral widths for the 1D STD NMR spectra.

2D STD TOCSY spectra were recorded with 40 scans per t_1 increment. A total of 256 t_1 increments were collected in an interlaced mode for the on and off resonance spectra. Prior to subtraction both spectra were processed and phased identically. A MLEV (composite pulse decoupling used for TOCSY spin lock) mixing time of 100 ms was applied in all TOCSY spectra. The acquisition times for the 2D experiments were typically around 22 h. 2D spectra were multiplied with a squared cosine bell function in all dimensions and zero filled two times. The pulse sequence for the 2D NOESY spectra included a filter to suppress zero quantum coherence. The spectra were recorded with mixing times of 50, 100, 150, 300 and 500 ms and 80 scans for each of the 205 t_1 increments. The 2D ROESY spectrum was recorded with a mixing time of 300 ms and 80 scans for each of the 205 t_1 increments using a spin lock field of $\gamma B_1 = 1967$ Hz at 4.9 p.p.m.

Distance geometry calculations

The starting structures were generated with distance range constraints obtained from the NOE distances by adding or subtracting 5% for upper and lower limit, respectively. The conformation of PDTRP bound to SM3 was described by 13 distance range constraints (cf. Table 3). 500 structures were calculated using the Redac strategy implemented in the DYANA package [26]. The conformation with lowest target function was used for the following molecular dynamics (MD) simulation. The structure of PDT(*O*- α -D-GalNAc)RP in the binding site of SM3 was defined by 16

Table 1. ^1H -NMR chemical shifts of PDTRP in p.p.m. Spectra were recorded at 280 K with HDO resonance at 4.9 p.p.m. Resonances of protons marked by – were not visible.

	NH	α	β	β'	γ	γ'	δ	δ'
Pro8	–	4.347	2.389	1.978	1.978	1.978	3.336	3.336
Asp9	–	4.666	2.707	2.543				
Thr10	8.386	4.252	4.155		1.134			
Arg11	8.484	4.574	1.798	1.724	1.629	1.629	3.160	3.160
Pro12		4.327	2.255	1.978	1.978	1.879	3.782	3.582
CONH	7.761							
CONH'	7.043							

Table 2. ^1H -NMR chemical shifts of PDT(*O*- α -D-GalNAc)RP in p.p.m. Spectra were recorded at 280 K with HDO resonance at 4.9 p.p.m. Resonances of protons marked by – were not visible.

	NH	α	β	β'	γ	γ'	δ	δ'
Pro8	–	4.347	2.397	1.983	1.983	1.983	3.368	3.317
Asp9	–	4.779	2.747	2.558				
Thr10	8.914	4.445	4.297		1.217			
Arg11	8.510	4.504	1.813	1.678	1.645	1.645	3.160	3.160
Pro12		4.295	2.270	1.990	1.990	1.879	3.705	3.591
CONH	7.788							
CONH'	7.017							

	NH	CH3	1	2	3	4	5	6a	6b
GalNAc	7.803	1.975	4.775	4.016	3.850	3.914	3.972	3.715	3.691

Table 3. Constraints for PDTRP derived from trNOE build-up rates.

For distances between Protons of Asp2, Thr3, Arg4 (including intra-residue contacts of Arg4) and between Thr3 and Pro5 the trNOE build-up of the β -protons of Asp2 was taken as reference. For contacts between Arg4 and Pro5 the δ -protons of Pro5 acted as reference.

Proton pair	Lower limit (Å)	Upper limit (Å)
Asp2- α /Thr3- γ	3.58	3.96
Asp2- α /Thr3-NH	2.35	2.59
Pro5- $\delta\delta'$ /Arg4- $\delta\delta'$	3.41	3.77
Pro5- $\delta\delta'$ /Arg4- $\gamma\gamma'$	2.76	3.05
Arg4- α /Arg4- $\delta\delta'$	3.10	3.42
Arg4- α /Arg4-NH	2.57	2.84
Arg4- α /Pro5- $\delta\delta'$	2.30	2.55
Arg4- $\gamma\gamma'$ /Arg4-NH	2.97	3.28
Arg4-NH/Arg4- $\beta\beta'$	2.75	3.04
Thr3- γ /Thr3- α	3.02	3.34
Thr3- α /Thr3-NH	3.26	3.60
Thr3- α /Arg4-NH	3.38	3.74
Pro5- $\delta\delta'$ /Thr3- γ	3.97	4.39

distance range constraints (cf. Table 4). The distance geometry calculations were performed by an internal algorithm in SYBYL (v. 6.3, Tripos). A total of 100 structures were generated and energetically optimized. The lowest energy conformation acted as starting structure for the following MD simulation.

MD simulations

Constrained MD simulations were carried out with the SYBYL program on Silicon Graphics Octane (R12000) computers, using the Tripos force field. A harmonic potential was employed at the edges of the distance range constraints. The force field constants were set to 2 kcal·(mol Å²)⁻¹. Constraints to pseudoatoms, generated by SYBYL, were used for nonstereospecifically assigned methylene groups and methyl groups. The starting structures were placed in water boxes (PDTRP: 931 water molecules, 30 × 30 × 30 Å³, PDTRP/SM3 complex: 1708 water molecules, 40 × 41 × 40 Å³, PDT(*O*- α -D-GalNAc)RP: 1152 water molecules, 33 × 33 × 33 Å³, PDT(*O*- α -D-GalNAc)RP/SM3 complex: 2521 water molecules, 45 × 48 × 42 Å³).

Before starting the MD simulation the box was energy optimized over 200 steps. The constrained simulation was performed at 300 K. The charges were calculated with the Gasteiger Marsili method and a dielectric constant of four was used. A cutoff radius of 8 Å was used for the nonbonded interactions. The initial velocities for the atoms were taken from a Boltzmann distribution at 300 K and the step size for the integration of Newton's equation was 1 fs. The coupling to the temperature bath was set to 100 fs and the nonbonded interactions were updated every 25 fs. The MD simulations ran for 100 ps at constant volume and temperature.

The final structures were energy minimized over 1000 steps and overlaid to the PDTRP fragment of the ligand of the X-ray structure (RCSB PDB entry 1SM3). After small manual corrections, the ligands were docked into the binding site of SM3 using the FLEXIDOCK module of the SYBYL software package. The docking structures after

Table 4. Constraints for PDT(*O*- α -D-GalNAc)RP derived from trNOE build-up rates and trROESY data. *The NOEs marked with an asterisk are overlapping and are assumed to have equal intensity.

Proton pair	Lower limit (Å)	Upper limit (Å)
Asp2- α /Thr3-NH*	2.35	2.60
GalNAc-H1/Thr3-NH*	2.35	2.60
Pro5- $\delta\delta'$ /Arg4- $\gamma\gamma'$	3.03	3.35
Arg4- α /Arg4- $\delta\delta'$	3.17	3.51
Arg4- α /Arg4- $\gamma\gamma'$	2.48	2.74
Arg4- α /Pro5- $\delta\delta'$	2.33	2.58
Arg4- $\beta\beta'$ /Arg4- $\delta\delta'$	2.99	3.30
Arg4- $\beta\beta'$ /Arg4-NH	2.83	3.12
Arg4- $\gamma\gamma'$ /Arg4-NH	2.56	2.83
Thr3- α /Thr3- γ	2.46	2.72
Thr3- α /Thr3-NH	2.52	2.78
Thr3- α /Arg4-NH	2.44	2.69
Thr3- γ /Thr3-NH	2.99	3.31
Thr3- γ /GalNAc-H5	2.97	3.28
Thr3-NH/GalNAc-NH	3.13	3.46
GalNAc-NH/GalNAc-H3	2.65	2.93

100 000 generations were subjected to final MD simulations in the binding site of the antibody with flexible protein residues in a perimeter of 10 Å from the ligand.

RESULTS

The small PDTRP peptide and its glycosylated derivative were used because larger peptides did not show measurable trNOE effects. This is most likely due to slow exchange between the bound and the free state. For dimers and pentamers of the MUC-1 tandem repeat [16], i.e. 40mer and 60mer peptides, the dissociation constant was determined to $K_d = 10^{-7}$ by SPR. At an on-rate of $k_{on} = 10^6 \text{ M}^{-1}\text{s}^{-1}$ typical for antibody interactions one would have an off-rate $k_{off} = 0.1 \text{ s}^{-1}$, which is too slow for obtaining measurable trNOE effects. The exact kinetic constants were not published. More importantly, the larger peptides decomposed in the presence of the antibody within a few days (N. Serttas, H. Möller, J.M. Burchell, J. Taylor-Papadimitriou, B. Meyer and H. Paulsen, unpublished results). To overcome these problems with large peptides, we used short peptides to utilize their faster dissociation rates [25] and their stability in the presence of SM3. With the pentapeptide and glycopentapeptide we obtained strong STD effects and weak trNOEs.

SM3 in complex with the peptide PDTRP

STD Experiments. In contrast to the larger peptides and glycopeptides, the pentapeptide PDTRP is stable in the presence of SM3 and possesses a favourable off-rate on the NMR time scale to yield good trNOE spectra. Figure 2A shows the 1D STD spectrum (red) and a normal ¹H-spectrum (black) of the complex of PDTRP with SM3. For comparison, the signals of the Pro1 β -methylene protons are adjusted to have the same height. As evident from Fig. 2, proton resonances of Pro1 and Asp2 have the highest intensities in the STD spectrum, signals of Thr3 are of medium intensity, while the signals of Arg4 and Pro5 have the lowest intensity. The δ -protons of Pro5 have only

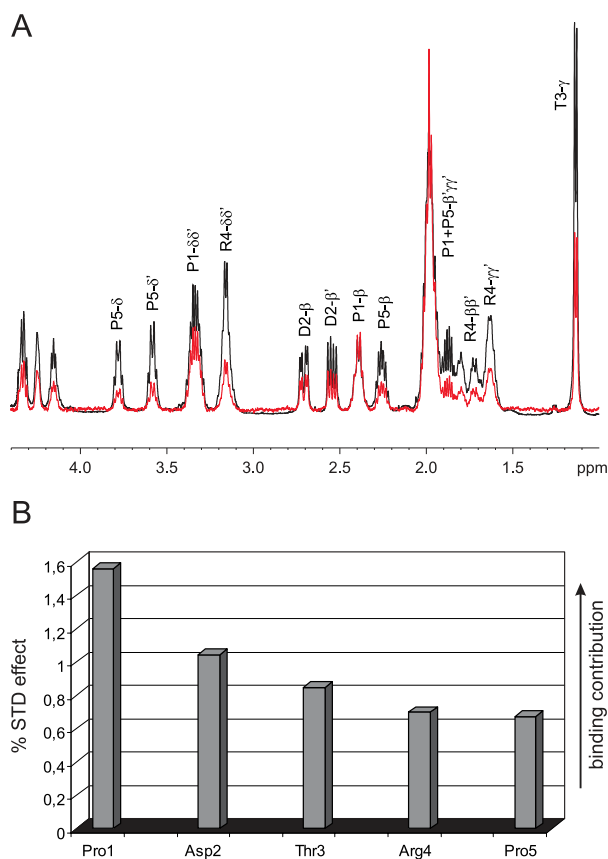


Fig. 2. STD data. (A) Superposition of a 1D STD spectrum (red) and a reference ¹H-spectrum (black) of PDTRP in complex with the antibody SM3. The intensity is adjusted, so that the Pro1- β -methylene signal is of same height in both spectra. Clearly visible are strong STD effects for protons of Pro1 and Asp2 whereas Arg4 and Pro5 show weaker signals in the STD spectrum. (B) Mean STD values (in percent) of the protons of the individual amino acids calculated for each amino acid of PDTRP from the 1D spectrum.

25% relative intensity in the STD spectrum. Obviously, Pro1 and Asp2 get more saturation from the protein than the remaining residues of the ligand and therefore have more and tighter contacts to the antibody's surface. The mean STD intensities of each residue are summarized in Fig. 2B. Here, it is evident that the mean intensities of signals of Pro5 have only 40% intensity relative to those of Pro1. Overall, there is a continuous drop in intensity from the N-terminus to the C-terminus with a 50% value being reached at Thr3.

By 2D STD TOCSY experiments one can use the increased dispersion for a more detailed epitope mapping. In Fig. 3 the STD and normal TOCSY spectrum of the PDTRP/SM3 complex are shown. The peaks of Arg4 and Pro5 are so low in intensity that they do not appear at the intensity cutoff shown. Signals of Pro1, Asp2 and Thr3 are clearly visible confirming the results from the 1D STD experiments. The strongest signals are again cross peaks from Pro1.

trNOE experiments with PDTRP and SM3

The conformation of the peptide ligand PDTRP bound to SM3 was obtained from transferred NOE spectra. In a

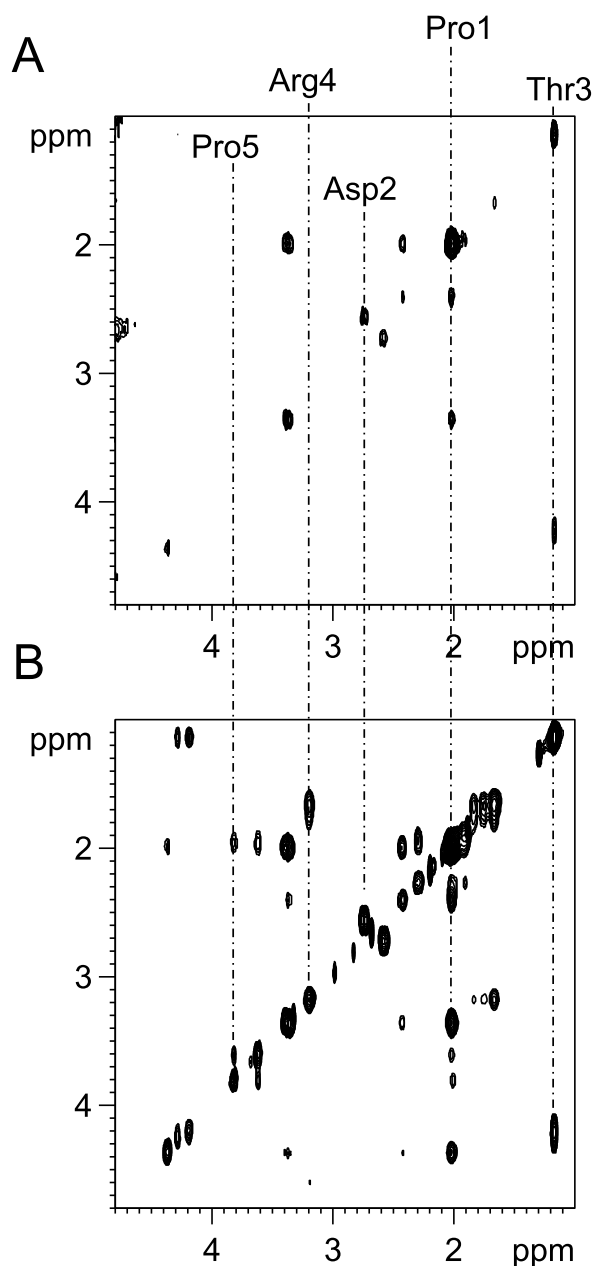


Fig. 3. TOCSY spectra. (A) 2D STD TOCSY and (B) conventional TOCSY spectrum of PDTRP in complex with SM3. The strongest STD signals originate from protons of Pro1, Asp2 and Thr3. The signals of Arg4 and Pro5 visible in the reference TOCSY (B) vanish completely or have low intensity in the STD experiment (A).

trNOESY spectrum of PDTRP in presence of SM3 (data not shown) all crosspeaks are of the same sign as the diagonal signals and have relatively weak intensity. Thus, these negative NOEs originate from the bound conformation. In absence of the antibody PDTRP shows exclusively positive NOEs. Most contacts are sequential or intraresidue NOEs, which is in agreement with the elongated conformation presented below. Pro5- δ /Thr3- γ is the only long range interaction that can be detected in the trNOE spectrum.

The trNOE spectra were recorded as a function of the mixing time with intervals of 50, 100, 150, 300, and

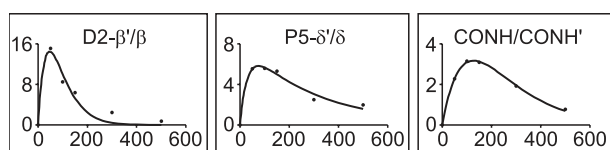


Fig. 4. trNOE build-up rates of PDTRP in presence of SM3 plotted as percentage trNOE vs. mixing time (ms). The ligand to protein ratio is 20 : 1. The three curves represent geminal protons with a fixed distance of 1.8 Å that are normally used as reference pairs. The build-up of the trNOE between Asp2-β and β' is much faster than that of the other two reference points indicating differences of the rotational correlation times of these proton pairs.

500 ms. Shorter mixing times did not give spectra with an interpretable signal/noise ratio. Inter proton distances were calculated from the extrapolated slope at mixing time zero using a biexponential fitting algorithm on the trNOE build-up curves. The distance is obtained by comparing the trNOE build-up of an interesting proton pair with that of a reference proton pair which has a known fixed distance, i.e. geminal protons. PDTRP offers three well resolved reference points, each of which forms a pair of geminal protons with a proton/proton distance of 1.8 Å: Asp2-β/β', Pro5-δ/δ' and the C-terminal carboxamide NH protons. As can be seen from the three panels in Fig. 4, the initial slopes of the build-up curves of the geminal proton pairs are very different. The build-up rate of the β protons of Asp2 is about 2.2-fold as big as that of the Pro5 δ protons. This big difference in NOEs converts to about 10% difference in the corresponding distances because of the r^{-6} dependence of the NOEs on the distances. As a result, using Asp2-β/β' as reference the distance of Pro5-δ/δ' was calculated to be 2.06 Å while with Pro5-δ/δ' as reference the Asp2 methylene protons should have a distance of 1.57 Å.

There are two possible explanations for this behavior: (a) the two segments of the peptide have a very different rotational correlation time, i.e. have very different degrees of freedom, or (b) the NOE between the Asp2 β protons is relayed by a protein proton. Neither explanation can easily be proven.

It is, however, unlikely that a transfer through protein protons is responsible for the enhanced cross relaxation of the Asp2 β protons. Assuming a binding mode as found in the X-ray structure analysis the closest distance of a protein proton to the Asp2-β/β' proton is 2.7 and 3.4 Å, respectively. This relay proton contributes to the observed cross relaxation rate of Asp2-β with β' with 2% only. All other protons are further away and thus have less contributions. The observed difference of more than 100% compared with the cross relaxation rate of Pro5-δ/δ' can consequently not

be explained by a relay phenomenon. We find on the other hand that the differences in segment flexibility are in perfect agreement with the STD-based epitope mapping presented above.

To accommodate these variations throughout the molecule we chose to reference distances to reference atom pairs in the same segment, i.e. distances between protons of Asp2,

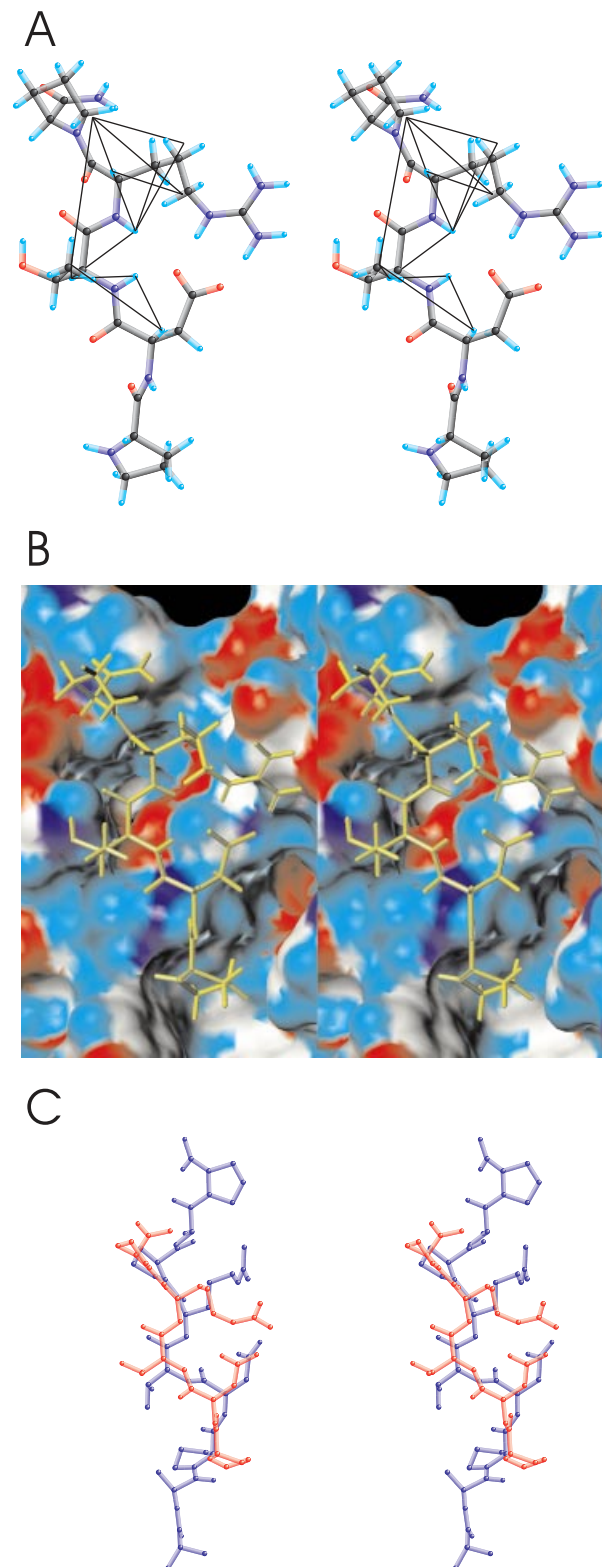


Fig. 5. PDTRP structures. (A) TrNOE-derived structure of PDTRP. The constraints (black lines) lead to a good conformational definition from Asp2 to Pro5. There were no NOE contact from Pro1 to Asp2 such that this segment was adjusted to fit the binding site of SM3. (B) PDTRP (yellow) in the binding site of SM3 (atom colored surface) (RCSB PDB entry 1SM3). This image shows the peptide antibody complex after 100 ps constrained MD simulation and minimization over 200 steps. Both the ligand and the binding site were kept flexible during the simulation. (C) Superposition of PDTRP (red) with the ligand of the X-ray structure analysis AAPDTRPAP (blue).

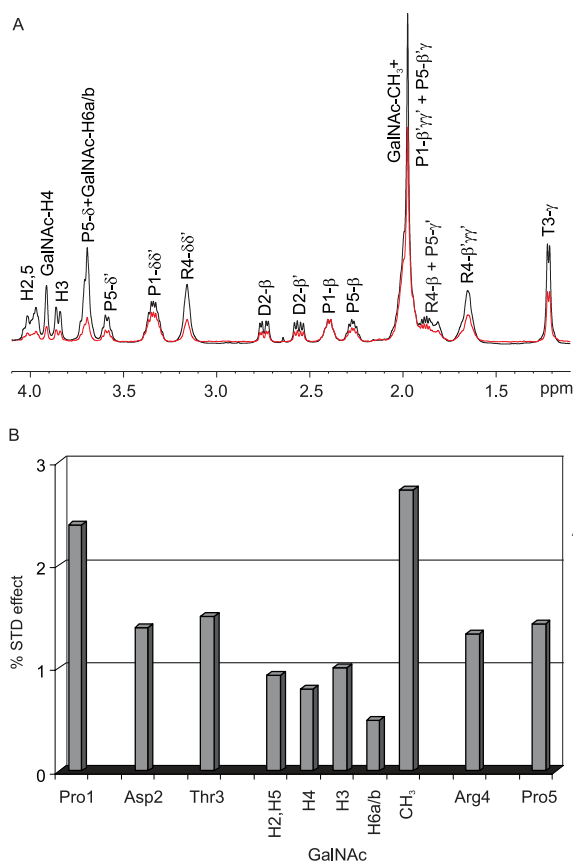


Fig. 6. PDT(*O*- α -D-GalNAc)RP spectra. (A) Superposition of a 1D STD spectrum (red) and a reference ^1H -spectrum (black) of PDT(*O*- α -D-GalNAc)RP in complex with the antibody SM3. The intensity is adjusted such that the Pro1- β -methylene signal is of same height in both spectra. In this 1D experiment Asp2, Thr3, Arg4 and Pro5 have signals of about the same intensity. Pro1 and the GalNAc *N*-acetyl methyl group show stronger STD effects while the signals of the GalNAc ring protons are of lower intensity. (B) Percent STD effects calculated from the 1D spectrum of PDT(*O*- α -D-GalNAc)RP in presence of SM3. Mean values are shown for the amino acids. STD effects of GalNAc protons are presented in detail. Only the *N*-acetyl methyl group obtains significant saturation at about the same level as Pro1.

Thr3, Arg4 (including intra residue contacts of Arg4) and between Thr3 and Pro5 were referenced to the β -protons of Asp2. The δ -protons of Pro5 were used as reference for contacts between Arg4 and Pro5. This approach inherently carries the possibility of up to 10% error of the distances in either segment. We also carried out structure calculations with exclusively referencing on Asp2- β/β' or Pro5- δ/δ' . This produced similar conformations as in the mixed referencing approach but with more constraint violations (data not shown). The carboxamide protons were not used as a reference pair because their initial slope was even lower than that of Pro5- δ/δ' which is probably due to exchange with the solvent. The final constraints that went into the distance geometry/molecular dynamics simulation are summarized in Table 3.

The calculations for the bound structures were performed in several steps. (a) We generated conformations by distance geometry calculations using the constraints

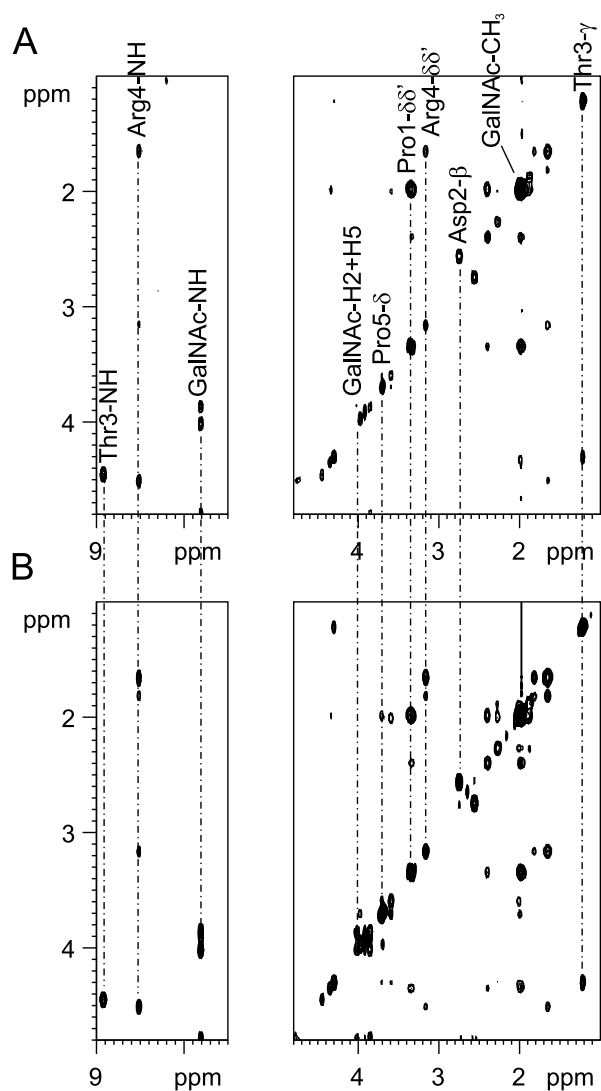


Fig. 7. 2D STD TOCSY (A) and a conventional TOCSY spectrum (B) of PDT(*O*- α -D-GalNAc)RP in complex with SM3. The strongest STD signals originate from protons of Pro1, Asp2 and Thr3. The protons of Arg4 and Pro5 give weak signals in the STD experiment. Most of the GalNAc resonances disappear completely. Only the *N*-acetyl methyl group shows a huge diagonal signal.

from NOE experiments with the program DYANA [26]. (b) The structures with the lowest target function were then subjected to constrained MD simulations over 100 ps. (c) The resulting structures were superimposed on the peptide from X-ray crystallography [19]. Due to the relatively small number of constraints we could not obtain a high resolution structure. Conformations that could not be fitted into the protein of the X-ray structure analysis were not followed further. (d) After small manual corrections to avoid clashes with the protein, we docked the ligand into the binding site with the software tool FLEXIDOCK within the software package SYBYL. (e) We carried out another constrained MD over 100 ps in the binding pocket with ligand and protein flexible. All MD simulations were carried out in water boxes.

Figure 5 A shows the resulting peptide conformation with constraints depicted as lines. The peptide in the binding

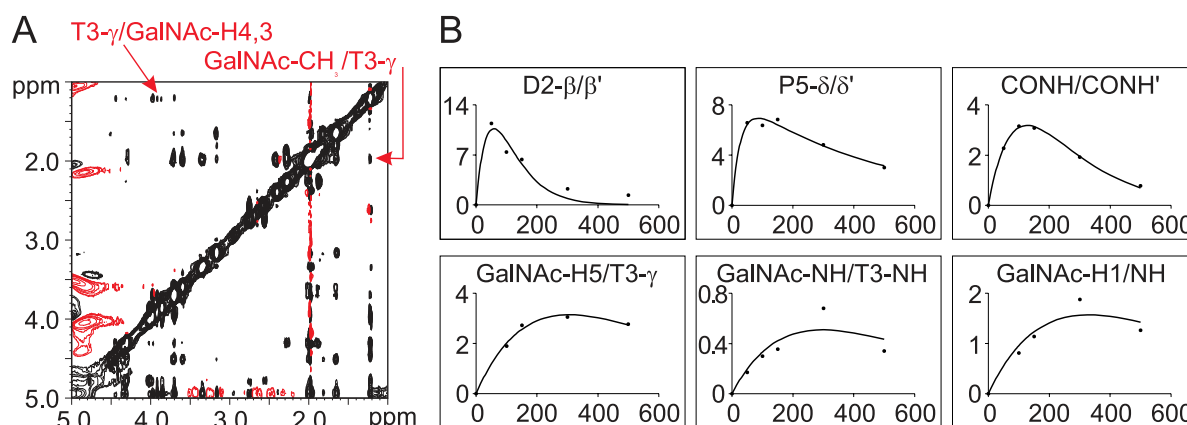


Fig. 8. trNOE data. (A) trNOESY of the glycopeptide PDT(*O*- α -D-GalNAc)RP in presence of SM3. Due to spin diffusion some peaks vanish in the trROESY or have negative sign (marked by an arrow) and were subsequently not used for distance calculation. (B) trNOE build-up rates plotted as percentage trNOE vs. mixing time (ms). The upper three curves come from geminal protons with a fixed distance of 1.8 Å that are normally used as reference pairs. In case of the glycopeptide the difference between Asp2- β and Pro5- δ is much smaller than for the peptide indicating similar rotational correlation times of these proton pairs. The lower three diagrams show examples of build-ups of structurally relevant NOE contacts.

site of SM3 is shown in Fig. 5B. For comparison of X-ray and NMR structure a least square superposition of both conformations can be seen in Fig. 5C. It is obvious that the NMR based structure determination of the bound conformation of the pentapeptide agrees with that obtained in the crystal. Probably because of fast exchange with the solvent H₂O at a pH of 7.0 the amide proton of Asp2 was invisible. The normal remedy for this is lowering the pH, which cannot be used here because we wanted to preserve near physiological conditions in the sample. As a consequence of this exchange phenomenon there are no NOE contacts between Pro1 and Asp2. This segment of the ligand is thus ill defined and was manually adjusted to fit the X-ray structure of the peptide.

SM3 in complex with the glycopeptide PDT(*O*- α -D-GalNAc)RP

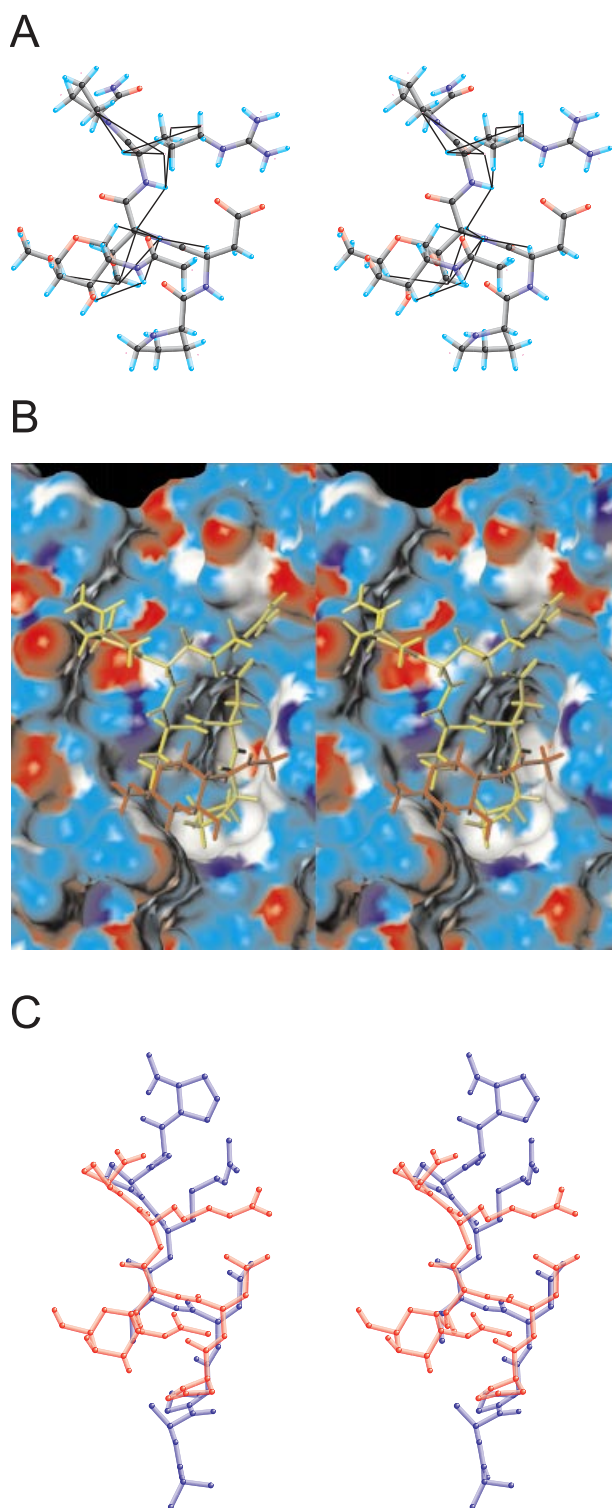
STD Experiments. MUC-1 peptides with O-glycosylation at Thr3 show increased binding to SM3 [18]. As there is no published X-ray structure of a glycopeptide binding to SM3 it is not known how a sugar moiety contributes to binding energy and whether peptide and glycopeptide bind in a similar way. From STD NMR spectra (cf. Figures 6 and 7) it is obvious that the ring protons of the GalNAc residue of PDT(*O*- α -D-GalNAc)RP receive overall less saturation than each of the amino acids. Only the N-acetyl methyl group has a strong STD NMR signal. It is very unlikely that this effect is due to different relaxation rates of the methyl group because we have shown earlier that carbohydrates interacting with a protein through their ring protons do not show an STD effect on the N-acetyl methyl group [21]. The mean of all STD values in each residue is presented in Fig. 6B confirming strong interactions of Pro1 and the GalNAc N-acetyl methyl group with the antibody. The differences between amino acids are less pronounced than in case of the unglycosylated peptide indicating that either the glycopeptide is less flexible or that the binding contributions are more evenly distributed within the glycopeptide. It has

been established in literature that the O-type glycosylation in peptides introduces a stabilization of that particular peptide fragment [27,28].

trNOE experiments with PDT(*O*- α -D-GalNAc)RP and SM3

Conformational analysis of the glycopeptide in the bound state was not as straightforward as in the peptide case, because the free glycopeptide gives already negative NOEs due to solvation of the GalNAc moiety which in turn produces a relatively long correlation time. By comparing build-up rates of the glycopeptide NOEs with and without SM3 we could prove that we had in fact real trNOEs. The maximum of the build-up curve moved from about 600 ms without protein (data not shown) to 150 ms in presence of SM3 (cf. Figure 8B). In the first attempt to calculate a structure some constraint inconsistencies occurred. Therefore, a trROESY spectrum was recorded (data not shown) to identify cross peaks with a high fraction of spin diffusion. Peaks that vanish or even change sign in the trROESY were subsequently not used for distance calculations. (cf. Figure 8A).

In contrast to PDTRP binding to SM3, large differences in segment flexibility were not observed in the case of the glycopeptide. As one can estimate from trNOE build-up rates shown in Fig. 8B there are significantly less differences in segment correlation time. This is evidence for a conformational stabilization by the GalNAc moiety. The structure calculation basically followed the same scheme presented above for the peptide. As the program DYANA cannot handle glycopeptides it was substituted by the DG algorithm of the SYBYL software package. Again, the amide proton of Asp2 was invisible which led to an ill-defined N-terminal part of the ligand. The glycopeptide did not show long range NOEs, therefore only sequential and intra residue contacts were used for constraint generation. The glycopeptide structure which is the result of the DG calculation, constrained MD simulation of the ligand alone and constrained MD simulation in the binding site of SM3



is shown in Fig. 9. A superposition of NMR glycopeptide structure and X-ray peptide ligand can be seen in Fig. 9C.

DISCUSSION

PDTRP in complex with SM3

Using STD NMR spectra it is possible to perform a detailed epitope mapping of the peptide bound to SM3. Pro1 gives

Fig. 9. PDT(*O*- α -D-GalNAc)RP with trNOE-derived constraints (black lines) (A), (B) PDT (*O*- α -D-GalNAc)RP (yellow, red) in the binding site of SM3 (RCSB PDB entry 1SM3) (atom color), and (C) Superposition of PDT(*O*- α -D-GalNAc)RP (red) with AAPDTRPAP (blue) as found in the X-ray crystal structure analysis. In (A), the glycopeptide is well defined from Asp2 to Pro5. There were no NOE contacts from Asp2 to Pro1 such that this segment was adjusted to fit into the binding site of SM3. In (B), the glycopeptide antibody complex is shown after 100 ps constrained MD simulation and minimization over 200 steps. Both the ligand and the binding site were kept flexible during the simulation.

most intensive STD signals corresponding to a tight contact to the protein. We see strong STD signals also for Asp2 and Thr3. Arg4 and Pro5 are only weakly bound resulting in smaller integrals.

The trNOE data suggests that there is a different flexibility in the N-terminal and the C-terminal parts of the molecule due to interactions with the protein. This is in full agreement with the STD determination of the binding epitope that is located on the N-terminal side of the molecule with Pro1, Asp2 and Thr3 as the major interacting residues. Pro5 of the peptide has a much shorter segment correlation time compared to Asp2 which means more flexibility and less contact to the protein. In the 3D structure of PDTRP docked into the binding site of SM3 Pro1 and Asp2 fill a deep cavity of the antibody while Arg4 and Pro5 have less contact to the surface of SM3 (cf. Figure 5B). Also, the conformation of the peptide as determined from the trNOE study fits perfectly into the binding cavity of the protein. During constrained MD simulation of the peptide/antibody complex in a water box the peptide remains in the binding pocket and does not change its conformation significantly.

The trNOE derived structure has a salt bridge between the Asp2-carboxyl and the Arg4 guanidino group. Such an electrostatic interaction was also found by Fontenot *et al.* [17] in the corresponding structure of the 60mer triple repeat of the MUC1 peptide. This salt bridge is not present in the X-ray structure of the peptide SM3 complex. Crystal contacts may however, be responsible for this because of interactions of the arginine with glutamate 126 and asparagine 128 at the bottom of the next protein molecule. As the salt bridge was also found by Fontenot *et al.* [17] in a solution structure we do not believe that it is induced by binding.

The X-ray structure of a complex of a 13mer peptide TSAPDTRPAPGST and the antibody SM3 was determined by Dokurno *et al.* [19]. They report a significant contribution of Arg4 to binding by analyzing the surface of the residues covered by the protein. They also see an interaction of the C-terminal part RPAP with the surface of the antibody. In the shorter peptide that we used we found a high flexibility of the segment Arg4-Pro5 and heavily reduced STD intensity for these two amino acids. The binding of the 13mer peptide in the X-ray crystal structure is, however, stabilized at the C-terminus by significant interactions between the bound peptide and the bottom of the next Fab segment in the crystal (cf. below and Fig. 10).

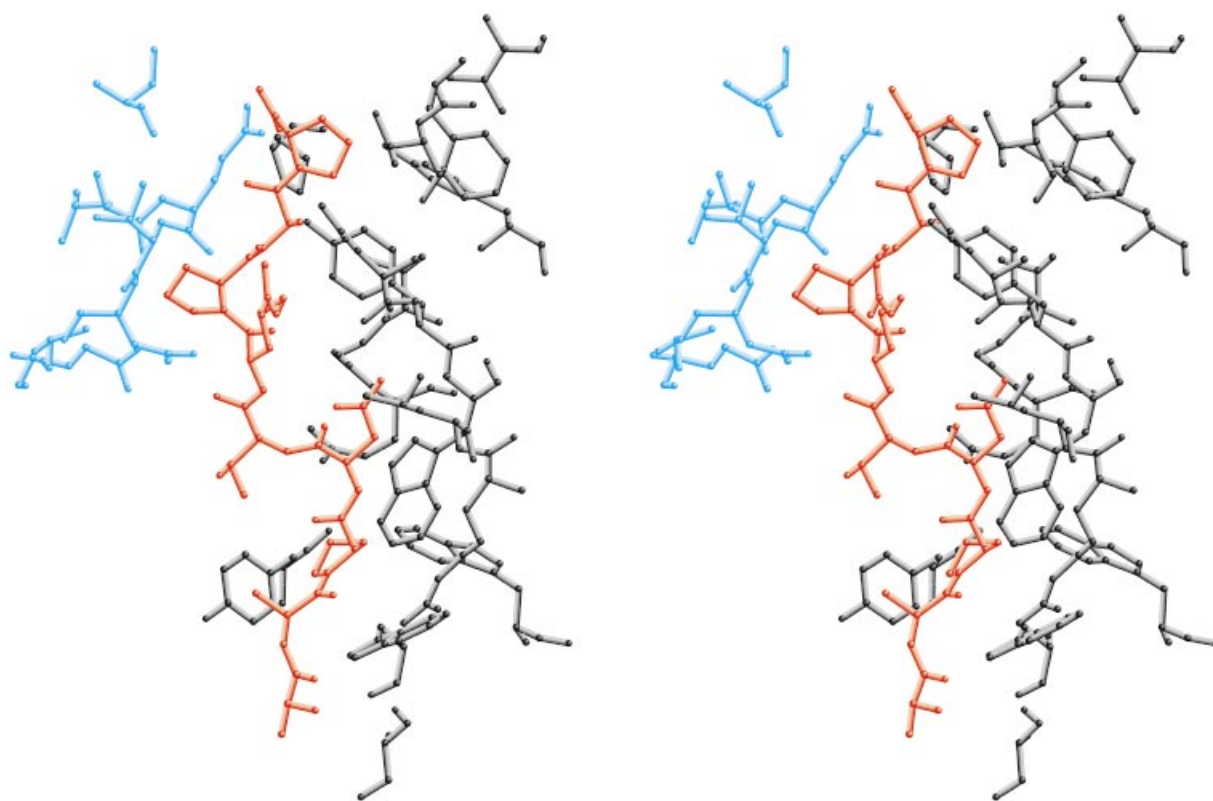


Fig. 10. X-ray crystal structure of SM3 in complex with TSAPDTRPAPGST (RCSB PDB entry 1SM3). Resolved residues of the peptide ligand are colored red (AAPDTRPAP). The binding site of the antibody is colored black. Residues colored blue are parts of a neighboring Fab fragment in the crystal cell that have a distance of 4 Å or less to the ligand. The C-terminal part of the peptide (RPAP) has close contacts to the next protein in the crystal cell and is therefore stabilized on the surface of SM3. It is unclear whether this part of the peptide represents the situation in solution.

PDT(α -D-GalNAc)RP in complex with SM3

The amino acids of the glycopeptide give STD signals of about equal intensity from Asp2 to Pro5 with only Pro1 and the methyl group of the GalNAc residue being significantly stronger. According to that the GalNAc ring has less contact to the surface of SM3 than the other amino acids. Looking at trNOE build-up rates the glycopeptide possesses a uniform correlation time. This is in contrast to the behavior of the peptide when bound to SM3 where we found differing segmental correlation times. This is probably due to a conformational stabilization caused by the GalNAc residue and/or by the binding contribution from the GalNAc residue.

Docking the glycopeptide to the antibody the sugar ring has little contact to the protein surface. Only the N-acetyl methyl group is positioned in close proximity to the antibody, which is in agreement with the STD NMR data (cf. Figures 6,7 and 9). The overall agreement between the contacts obtained from docking the trNOE derived structure into the binding cavity of SM3 and the binding epitope obtained from STD NMR is very good.

Comparison with the X-ray and previous NMR structures

As mentioned above, in the crystal cell the peptide shows contacts with two Fab residues. Most of these contacts are with the binding cavity in the Fv domain. However, there

are significant additional contacts to the bottom of the next Fab (cf. Figure 10). The surface covered by the nonbinding bottom portion of the next Fab is formed by the C-terminal segment RPAP. The size of the interaction surface between the peptide and the non binding bottom of the Fab is 173 Å² which corresponds to about 36% of the interaction of the full peptide with the binding cavity of SM3. The sizes of the surfaces were determined using distances between ligand atoms and protein atoms of less than 3 Å. These additional interactions are solely due to crystal packing and are certainly contributing to the stabilization of the C-terminal portion of the peptide as observed in the X-ray crystal structure. It is unclear whether this segment of the peptide would also show the same stable arrangement when in solution the additional interactions are not present. In light of the NMR data presented here it seems more likely that there is no or little contribution to the binding of the peptide from the C-terminal part.

It was postulated that the mucin forms a knob-like structure that helps expose the immunogenic epitope around the glycosylation site at Thr3. This knob is mainly built by the flanking peptide segments of DTR while the DTR-motif itself has a more or less elongated structure [17] (J. Dojahn, C. Diotel, H. Paulsen and B. Meyer, unpublished results). Our constraints are compatible with an elongated conformation of the pentapeptide PDTRP. However, as the knob becomes only evident at the amino acids beyond the two flanking prolines, we

cannot reject or confirm the presence of a knob-like structure based on our data from the pentapeptide. Our conclusion is that the peptide segment recognized by the antibody SM3 is basically elongated while the adjacent amino acids may still form the knob (cf. Fontenot *et al.* [17]).

CONCLUSIONS

Here, we demonstrate that binding studies of peptides and glycopeptides with large proteins can easily be performed with STD NMR. Combined with trNOE information one can obtain a 3D picture with the binding residues of the ligand identified in their relative orientation in space and thus efficiently optimize on the structure of the ligand for further immunization studies.

In contrast to the X-ray structure analysis NMR spectroscopy does not show artefacts from crystal packing. Thus, NMR spectroscopy was able to obtain the information on binding from a sample that is close to the natural physiological environment.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 470/B2, the Graduate College GRK 464 and a grant from the Fonds der chemischen Industrie (FCI) to H. M.

REFERENCES

- Gendler, S.J., Burchell, J.M., Duhig, T., Lampert, D., White, R., Parker, M. & Taylor-Papadimitriou, J. (1987) Cloning of partial cDNA encoding differentiation and tumor-associated mucin glycoproteins expressed by human mammary epithelium. *Proc. Natl Acad. Sci. USA* **84**, 6060–6064.
- Strous, G.J. & Dekker, J. (1992) Mucin-type glycoproteins. *Crit. Rev. Biochem. Mol. Biol.* **27**, 57–92.
- Muller, S., Goletz, S., Packer, N., Gooley, A., Lawson, A.M. & Hanisch, F.G. (1997) Localization of O-glycosylation sites on glycopeptide fragments from lactation-associated MUC1. All putative sites within the tandem repeat are glycosylation targets *in vivo*. *J. Biol. Chem.* **272**, 24780–24793.
- Devine, P.L. & McKenzie, I.F. (1992) Mucins: structure, function, and associations with malignancy. *Bioessays* **14**, 619–625.
- Hanisch, F.G., Uhlenbruck, G., Peter-Katalinic, J., Egge, H., Dabrowski, J. & Dabrowski, U. (1989) Structures of neutral O-linked poly-lactosaminoglycans on human skim milk mucins. A novel type of linearly extended poly-N-acetyl-lactosamine backbones with Gal beta (1–4) GlcNAc beta (1–6) repeating units. *J. Biol. Chem.* **264**, 872–883.
- Hull, S.R., Bright, A., Carraway, K.L., Abe, M., Hayes, D.F. & Kufe, D.W. (1989) Oligosaccharide differences in the DF3 sialo-mucin antigen from normal human milk and the BT-20 human breast carcinoma cell line. *Cancer Commun.* **1**, 261–267.
- Lloyd, K.O., Burchell, J., Kudryashov, V., Yin, B.W. & Taylor-Papadimitriou, J. (1996) Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines. Demonstration of simpler and fewer glycan chains in tumor cells. *J. Biol. Chem.* **271**, 33325–33334.
- Jerome, K.R., Barnd, D.L., Bendt, K.M., Boyer, C.M., Taylor-Papadimitriou, J., McKenzie, I.F., Bast, R.C. & Finn, O.J. (1991) Cytotoxic T-lymphocytes derived from patients with breast adenocarcinoma recognize an epitope present on the protein core of a mucin molecule preferentially expressed by malignant cells. *Cancer Res.* **51**, 2908–2916.
- Baekstrom, D., Nilsson, O., Price, M.R., Lindholm, L. & Hansson, G.C. (1993) Discrimination of MUC1 mucins from other sialyl-Le(a)-carrying glycoproteins produced by colon carcinoma cells using a novel monoclonal antibody. *Cancer Res.* **53**, 755–761.
- Burchell, J., Gendler, S., Taylor-Papadimitriou, J., Girling, A., Lewis, A., Millis, R. & Lampert, D. (1987) Development and characterization of breast cancer reactive monoclonal antibodies directed to the core protein of the human milk mucin. *Cancer Res.* **47**, 5476–5482.
- Biassoni, L., Granowska, M., Carroll, M.J., Mather, S.J., Howell, R., Ellison, D., MacNeill, F.A., Wells, C.A., Carpenter, R. & Britton, K.E. (1998) ^{99m}Tc-labelled SM3 in the preoperative evaluation of axillary lymph nodes and primary breast cancer with change detection statistical processing as an aid to tumour detection. *Br. J. Cancer* **77**, 131–138.
- Petrakou, E., Murray, A. & Price, M.R. (1998) Epitope mapping of anti-MUC1 mucin protein core monoclonal antibodies. *Tumour Biol.* **19**, 21–29.
- Burchell, J., Taylor-Papadimitriou, J., Boshell, M., Gendler, S. & Duhig, T. (1989) A short sequence, within the amino acid tandem repeat of a cancer-associated mucin, contains immunodominant epitopes. *Int. J. Cancer* **44**, 691–696.
- Blockzijl, A., Nilsson, K. & Nilsson, O. (1998) Epitope characterization of MUC1 antibodies. *Tumour Biol.* **19**, 46–56.
- Schol, D.J., Meulenbroek, M.F., Snijdwint, F.G., von Mensdorff-Pouilly, S., Verstraeten, R.A., Murakami, F., Kenemans, P. & Hilgers, J. (1998) 'Epitope fingerprinting' using overlapping 20-mer peptides of the MUC1 tandem repeat sequence. *Tumour Biol.* **19**, 35–45.
- Karanikas, V., Patton, K., Jamieson, G., Pietersz, G. & McKenzie, I. (1998) Affinity of antibodies to MUC1 antigens. *Tumour Biol.* **19**, 71–78.
- Fontenot, J.D., Mariappan, S.V., Catasti, P., Domenech, N., Finn, O.J. & Gupta, G. (1995) Structure of a tumor associated antigen containing a tandemly repeated immunodominant epitope. *J. Biomol. Struct. Dyn.* **13**, 245–260.
- Karsten, U., Diotel, C., Klich, G., Paulsen, H., Goletz, S., Muller, S. & Hanisch, F.G. (1998) Enhanced binding of antibodies to the DTR motif of MUC1 tandem repeat peptide is mediated by site-specific glycosylation. *Cancer Res.* **58**, 2541–2549.
- Dokurno, P., Bates, P.A., Band, H.A., Stewart, L.M., Lally, J.M., Burchell, J.M., Taylor-Papadimitriou, J., Snary, D., Sternberg, M.J. & Freemont, P.S. (1998) Crystal structure at 1.95 Å resolution of the breast tumour-specific antibody SM3 complexed with its peptide epitope reveals novel hypervariable loop recognition. *J. Mol. Biol.* **284**, 713–728.
- Ni, F. (1994) Recent developments in transferred NOE methods. *Prog. NMR Spectrosc.* **26**, 517–606.
- Mayer, M. & Meyer, B. (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **123**, 6108–6117.
- Mayer, M. & Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **38**, 1784–1788.
- Klein, J., Meinecke, R., Mayer, M. & Meyer, B. (1999) Detecting binding affinity to immobilized receptor proteins in compound libraries by HR-MAS STD NMR. *J. Am. Chem. Soc.* **121**, 5336–5337.
- Shuker, S.B., Hajduk, P.J., Meadows, R.P. & Fesik, S.W. (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534.
- Anglister, J. & Zilber, B. (1990) Antibodies against a peptide of cholera toxin differing in cross-reactivity with the toxin differ in

- their specific interactions with the peptide as observed by ^1H NMR spectroscopy. *Biochemistry* **29**, 921–928.
26. Günthert, P., Mumenthaler, C. & Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
 27. Paulsen, H., Pollex-Kruger, A. & Sinnwell, V. (1991) Conformational analysis of N-terminal O-glycopeptide sequences of interleukin-2. *Carbohydr. Res.* **214**, 199–226.
 28. Pieper, J., Ott, K.H. & Meyer, B. (1996) Stabilization of the T1 fragment of glycophorin A (N) through interactions with N- and O-linked glycans. *Nat. Struct. Biol.* **3**, 228–232.
 29. Klich, G., Paulsen, H., Meyer, B., Meldal, M. & Bock, K. (1997) Synthesis and characterisation of highly glycosylated glycopeptides with Tn-antigenic structures corresponding to human glycophorin AN. *Carbohydr. Res.* **299**, 33–48.