Knowledge-based Homology Modeling and Experimental Determination of Amino Acid Side Chain Accessibility by the Laser Photo CIDNP (Chemically Induced Dynamic Nuclear Polarization) Approach in Solution: Lessons from the Small Sialidase of *Clostridium perfringens*

Hans-Christian Siebert[1,2], Emadeddin Tajkhorshid[3,4], Claus-Wilhelm von der Lieth[3], Reinhard G. Kleineidam[5,6], Susanne Kruse[6], Roland Schauer[6], Robert Kaptein[2], Hans-Joachim Gabius[1], and Johannes F. G. Vliegenthart[2,*]

[1] Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität, Veterinärstr. 13, D-80539 München, Germany

[2] Bijvoet Center for Biomolecular Research, University of Utrecht, P.O. Box 80075, NL-3508 TB Utrecht, The Netherlands Tel: ++31 30 2532184, Fax: ++31 30 2540980 (vlieg@cc.ruu.nl)

[3] Zentrale Spektroskopie, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

[4] Department of Medicinal Chemistry, School of Pharmacy, Tehran University of Medical Sciences, P.O. Box 14155/6451, Tehran, Iran

[5] Biochemisch Laboratorium, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands

[6] Biochemisches Institut, Christian-Albrechts-Universität, Olshausenstraße 40, D-24098 Kiel, Germany

Abstract

The success of knowledge-based homology modelling is critically dependent on the predictive potency of the program structure-based calculations, which attempt to translate homologous sequences into three-dimensional structures, and on the actual relevance of the crystal structure for the protein topology. As quality control, experimental data for selected parameters of the protein's conformation are required. Using the crystal structure of the sialidase of *Salmonella typhimurium* as framework for model building of the homologous enzyme from *Clostridium perfringens*, a set of energy-minimised conformers is derived. These proteins present e.g. Tyr, Trp and His residues with an assessable area on the surface, since the side chains of these amino acid residues are responsive to chemically induced dynamic nuclear polarization (CIDNP), monitored by NMR. Hence, as first lesson, a comparative analysis for model-derived and experimentally determined values can be performed. The second lesson of this study concerns the notable impact of single amino acid substitutions (Tyr/Phe, Cys/Ser) on the surface accessibility of the CIDNP-reactive amino acid side chains in mutant forms of the sialidase. Corroborating the predictions from the theoretical calculations, the spectra of the engineered mutants reveal marked and non-uniform alterations. Thus, the effect of apparently rather conservative amino acid substitutions on a distinct conformational aspect of this protein, even at distant sites, should not be underestimated.

Keywords: Sialidase, NMR, protein modelling, molecular dynamics

* To whom correspondence should be addressed

Introduction

Since the role of oligosaccharides in cellular glycoconjugates as information-storing structures mediating diverse physiological functions is well appreciated [1-3], efforts are warranted to describe in detail the properties of enzymes which are used for deliberately altering the sequence of the carbohydrate chains. Sialidases (N-acylneuraminosyl-glycohydrolases, EC 3.2.1.18) hydrolytically cleave the linkage between α–glycosidically bound *N*-acylneuraminic acid derivatives and the penultimate sugar (Fig. 1). N-acetylneuraminic acid and a wide variety of variants are often found as terminal constituents of oligosaccharide chains in glycoproteins and glycolipids serving as recognition or masking sites [4-8]. They are removed from the glycoconjugates by sialidases which are widely distributed not only throughout metazoan animals of the deuterostomate lineage, but also among viruses, bacteria and protozoa, many of which are unable to produce sialic acids [4]. Remarkably, the enzyme is frequently produced by microorganisms, which live in close contact with an animal host, whereby the enzyme may serve as a pathogenicity factor or as an important tool for processing of nutrients [4]. To gain insight into the structure/ function relationship the knowledge about the evolutionary pathway of sialidase phylogenesis and a comparison of the primary structure of this protein obtained from different sources can be instructive. Indeed, the alignment of the sequences of eight bacte-
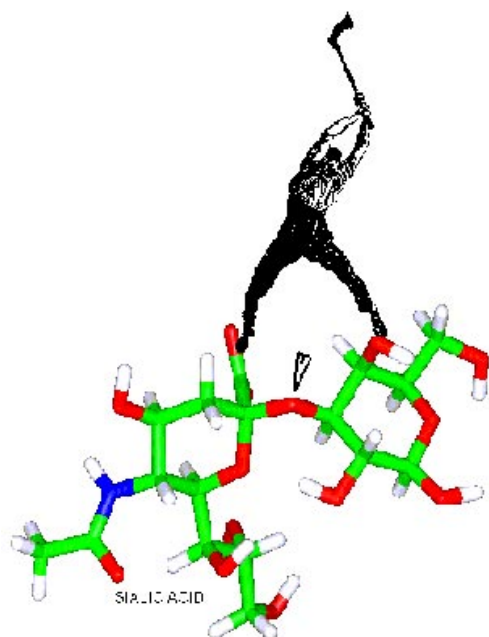

Figure 1. *Sialidases hydrolytically cleave the a-glycosidic linkage of glycosidically bound sialic acids which are mostly found as terminal constituents of oligosaccharide chains in glycoproteins and glycolipids. The subterminal sugar is galactose.*

rial sialidases indicates that they are most probably derived from a common ancestor by divergent evolution [9]. The complete conservation of 25 amino acid positions in these enzymes implies their involvement in a common functional and/ or structural role to be delineated by detailed inspection of mechanistic/ conformational aspects. At present, tertiary structures of three bacterial sialidases, namely from *Salmonella typhimurium* [10, 11],*Vibrio cholerae* [12] and *Micromonospora viridifaciens* [13], have been elucidated. Only the data set for the sialidase of *Salmonella typhimurium* [10] is stored in the Brookhaven data bank. Based on the mentioned evolutionary relationship within this family, the available structural information is employed as a framework to concoct a knowledge-based model of the small sialidase of *Clostridium perfringens* with a molecular mass of 43 kDa. Since regions of homology will most likely be centered in the core structures of related enzymes, the large sialidase of *Clostridium perfringens* with a molecular mass of 73 kDa is not considered in this study.

At least two sources of error may reduce the level of accuracy for the resulting model, namely the precision of the calculated predictions by modelling and the occurrence of differences between crystallographic and solution structures. Since the conditions for crystallization and the crystal packing order may affect the conformation of the protein, the solution structure will not necessarily be identical. Therefore, these inherent uncertainties within the modelling approach starting from data sets of the protein crystal call for collecting experimentally derived information. The actual solution structures are used to improve the reliability of the model, which is a demanding task for special NMR spectroscopic methods. Techniques which provide information on the positioning of distinct classes of amino acid side chains would be of assistance to contribute to address the defined issue.

Such a technique with focus on the side chains of tyrosine, tryptophan and histidine has been developed, taking advantage of a radical pair-generating dye [14, 15]. Upon laser irradiation a CIDNP (chemically induced dynamic nuclear polarization) radical reaction takes place in the presence of e. g. a flavin derivative [14, 15]. In glycosciences, this special technique has recently been instrumental to prove the occurrence of conformational changes after desialylation of human serum amyloid P component with only one invariant carbohydrate chain [16]. Moreover, the involvement of aromatic residues in the architecture of carbohydrate recognition domains in N-acetylglucosamine-binding lectins in solution has been documented with this approach [17]. These studies encourage to employ this technique as a quality control for knowledge-based homology modelled structures. In detail, the concomitant calculation of surface accessibilities of the respective types of side chains, which is possible by performing computer-assisted Connolly surface area assessment in the three-dimensional model [18, 19], enables a detailed comparison of these data with the obtained CIDNP results. To evaluate the potential of the combination of these two independent techniques, we herein report their applica-

```
              .    :    .    :    .    :    .    :    .    :    .    :   60
NANH_CLOPE   MCNKNNTFEKNLDISHKPEPLILFNKDNNIWNSKYFRIPNIQLLNDGTILTFSDIRYNGP   60
NANH_SALTY   -TVEKSVVFKAEGEHFTDQKGNTIVGSGSGGTTKYFRIPAMCTTSKGTIVVFADARHNTA   59


              .    :    .    :    .    :    .    :    .    :    .    :  120
NANH_CLOPE   DDHAYIDIASARSTDFGKTWSYNIAMKNNRIDSTYSRVMDSTTVITNT---GRIILIAGS  117
NANH_SALTY   SDQSFIDTAAARSTDGGKTWNKKIAIYNDRVNSKLSRVMDPTCIVANIQGRETILVMVGK  119


              .    :    .    :    .    :    .    :    .    :    .    :  180
NANH_CLOPE   WNTNGN-WAM-TTSTRRSDWSVQMIYSDDNGLTWSNKIDLTKDSSKVKNQPSNTIGWLGG  175
NANH_SALTY   WNNNDKTWGAYRDKAPDTDWDLVLYKSTDDGVTFS-KVETNIHDIVTKNGTISA--MLGG  176


              .    :    .    :    .    :    .    :    .    :    .    :  240
NANH_CLOPE   VGSGIVMDDGTIVMPAQISLRENNENNYYSLIIYSKDNGETWTMGNKVPNSNTSENMVIE  235
NANH_SALTY   VGSGLQLNDGKLVFPVQMVRTKNITTVLNTSFIYSTD-GITWSLPSGYCEGFGSENNIIE  235


              .    :    .    :    .    :    .    :    .    :    .    :  300
NANH_CLOPE   LDGALIMSTRYDYSGYRAAYISHDLGTTWEIYEPLNGKILTGKGSGCQGSFIKATTSNGH  295
NANH_SALTY   FNASLVN--NIRNSGLRRSFETKDFGKTWTEFPP-MDKKVDNRNHGVQGSTI-TIPSGNK  291


              .    :    .    :    .    :    .    :    .    :    .    :  360
NANH_CLOPE   RIGLISAPKNTKGEYIRDNIAVYMIDFDDLSKGVQEICIPYPEDGNKLGGGYSCLSFKNN  355
NANH_SALTY   LVAAHSSAQNKNNDYTRSDI--SLYAHNLYSGEVKLIDAFYPKVGNASGAGYSCLSYRKN  349


              .    :    .    :    .    :    .    :    .    :    .    :  420
NANH_CLOPE   ----HLGIVYEANGNIEYQDLTPYYSLINKQ-                              382
NANH_SALTY   VDKETLYVVYEANGSIEFQDLSRHLPVIKSYN                              381
```

Figure 2. *Sequence alignmnent of* Clostridium perfringens *and* Salmonella typhimurium *sialidases. Four Asp boxes and active site amino acids are shown in blue and red, respectively. (-) is used to present the gaps.*

tion with respect to the *Clostridium perfringens* small sialidase. As an internal control of both the sensitivity of the method and the validity of the interpretation, the experimental examination of deliberately engineered enzyme mutants has been included in the study. These calculations and the experimental data are also pertinent to answer the question to which extent introduction of single-site mutations can affect the monitored conformational parameter.

Materials and Methods

A. Knowledge-based homology modelling

The framework for modelling of the *Clostridium perfringens* small sialidase structure is provided by the crystallographic data for the sialidase of *Salmonella typhimurium* [entries 1SIL and 1SIM in the Brookhaven Protein Databank]. Both FastA [20] and BLAST algorithms [21] were employed to assess the scoring level of homology. The next steps of computa-

tions were performed at ExPASy Molecular Biology Server (Geneva University Hospital, University of Geneva, Geneva, Switzerland), using SwissModel Automated Protein Modelling Server (running at the Geneva Biomedical Research Institute, Glaxo Wellcome Research and Development S.A., Switzerland) which which makes use of ProMod (PROtein MODelling tool) [22-24]. The program is accessible through internet browsers like Netscape and e-mail. The alignments were introduced separately into the optimized mode of the program.

The program is comprised of the following model building steps :

I. Construction of the starting framework from the three-dimensional structure of *Salmonella typhimurium* sialidase.

II. Fitting of the *Clostridium perfringens* sialidase backbone onto this framework using primary sequence alignment optimized for three-dimensional similarity.

III. Reconstruction of loop regions from their 'stems' by structural homology searches through the Protein Databank, as described previously [25].

IV. Rebuilding of missing side chains using a library of allowed rotamers [26].

V. Optimization of bond geometries and compensation of unfavorable non-bonded contacts by 30 steps of steepest descent followed by 500 steps of conjugate gradient minimization using CHARMM (Chemistry at HARvard
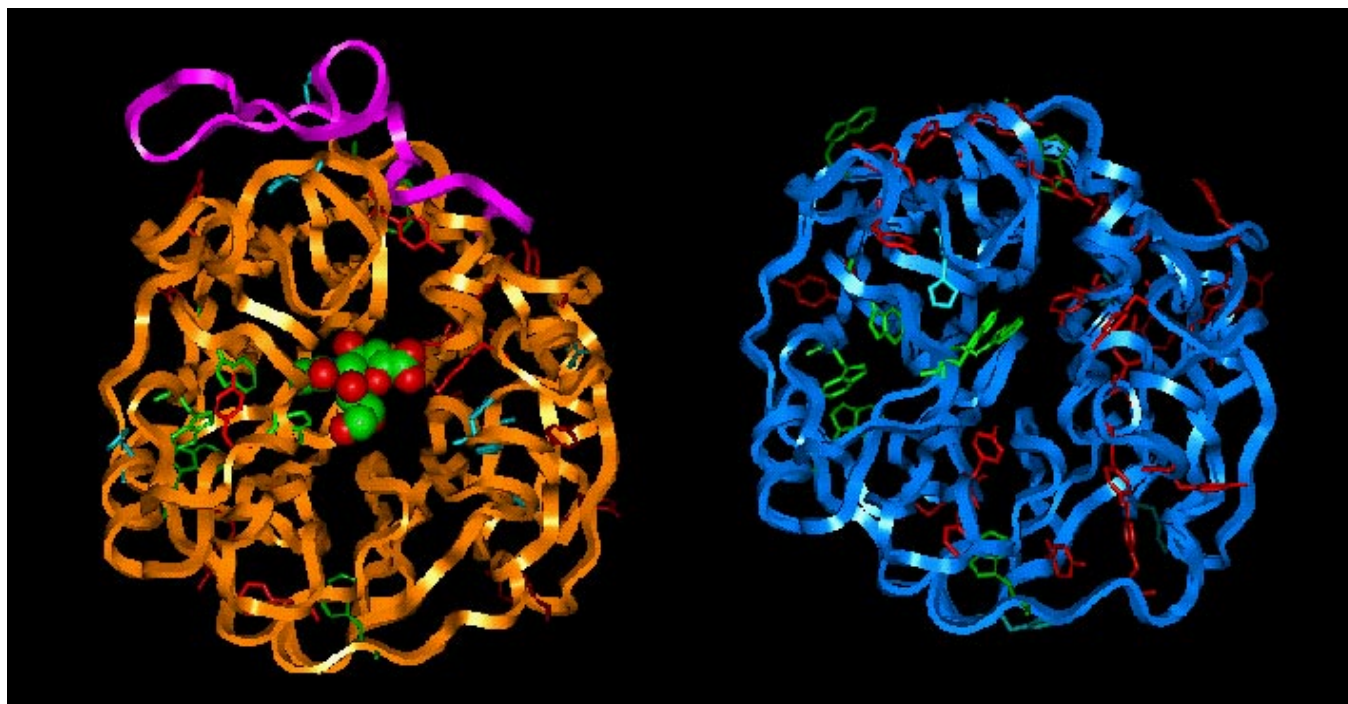
Figure 3. *Homology modelling-derived structure of the small sialidase of Clostridium perfringens (right) on the basis of the crystal structure of Salmonella typhimurium (left). The first 30 amino acids, which have not been incorporated in the modelling procedure, are shown in magenta. (tyrosine residues in red, tryptophan residues in green and histidine residues in blue).*

Macromolecular Mechanics) program [27] with the PARAM19 parameter set.

VI. Analysis of the resulting structure using the 3-D profile matching procedure [28].

The generation of hydrogen atoms, automatic assignment of partial charges of each atom and the solvation of the molecule were accomplished using the InsightII software. A molecular dynamics simulation was carried out for each molecule using the CVFF (Consistent Valence Force Field) at a temperature of 300 K with an equilibration time of 20 ps, a production period of 100 ps and an integration step of 1 fs. A cut-off distance of 10 Å and a dielectric constant of $\varepsilon = 4.0$ were used in all calculations.

Surface accessibilities of side chains of distinct amino acids of interest were calculated. The assessment of 'surface' command implemented in InsightII pinpoints the accessible exterior part of the relevant portions of the molecule by smoothening the van der Waals surface with a test sphere that displays the average radius of the solvent water (1.5 Å). The dot density of the spheres which represent the Connolly surface area was generally set to a value of 1. This dot density guarantees a sufficient distribution of an ensemble of calculated coordinate values which represent the surface area.

Based on the computationally generated model, two tyrosine residues with distinct differences in surface presentation were chosen for substitution with phenylalanine by site-directed mutagenesis. As a further control to evaluate the impact of Cys/Ser-substitution on the calculated/measured parameter, such a mutant was selected.

B. Site-directed mutagenesis, purification and characterization of recombinant enzymes

Site-directed mutagenesis, protein purification and enzymatic assays were performed, as described in detail elsewhere [29,30]. Substitutions are assumed to keep a potential impact on conformational aspects minimal. In detail, the following mutants were generated:

mutant 1: Y336F; mutant 2: Y347F; mutant 3: C349S.

Table 1. *Kinetic parameters for the cleavage of methyl-umbelliferyl-N-acetylneuraminic acid by the* Clostridium perfringens *wild-type sialidase and its three mutants with single amino acid substitution.*

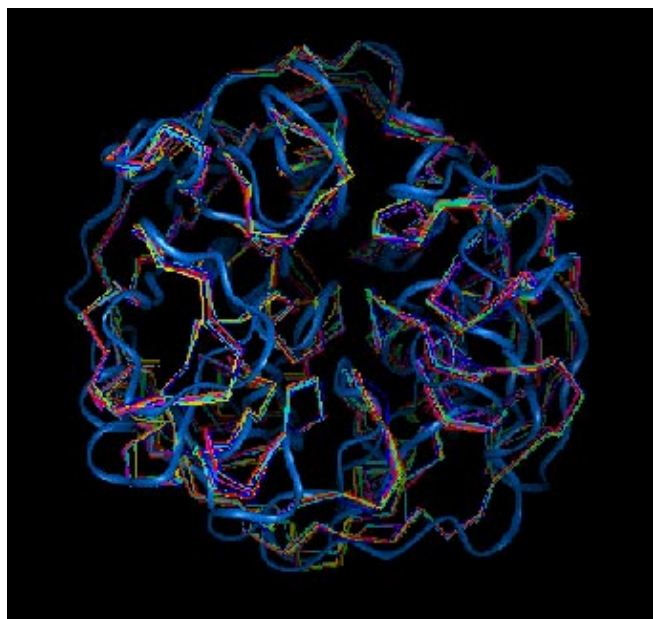|	K_M (mM)	V_{max} (%)
wild type	0.17	100
Y336F	0.19	11
Y347F	0.30	0.009
C349S	0.16	37

Figure 4. *Superposition of the starting structure and the energy-minimized conformations of ten structures obtained by molecular dynamics simulation of Clostridium perfringens small sialidase.*
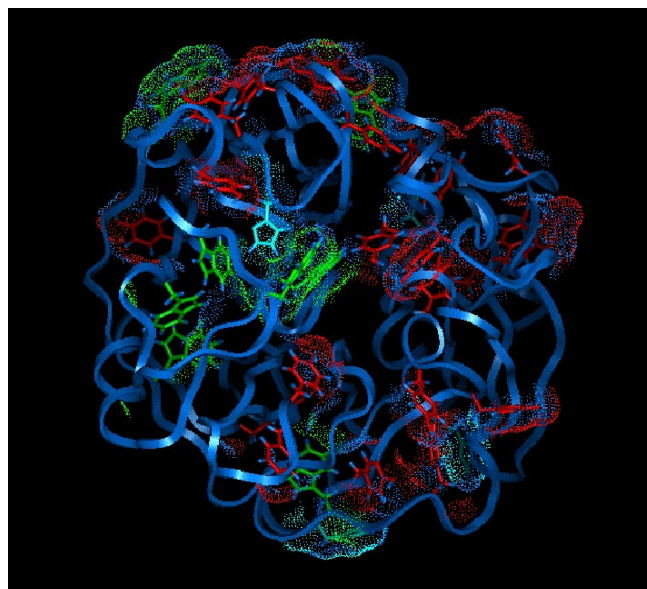

Figure 5. *The Connolly surface areas of aromatic amino acid side chains for the wild-type Clostridium perfringens small sialidase (tyrosine residues in red, tryptophan residues in green and histidine residues in blue).*

C. CIDNP Method

CIDNP experiments were performed at 360 MHz on a Bruker AM-360 NMR spectrometer, as described in detail elsewhere [15-17]. CIDNP was induced by using flavin I mononucleotide as radical pair-generating dye. Briefly, the light of a continuous-wave argon ion laser (Spectra Physics, Mountain View, USA) that operates in the multiline mode with principal wavelengths of 488.0 and 514.5 nm was directed to the sample by an optical fiber and chopped by a mechanical shutter to avoid harmful sample heating. Its operation was controlled directly by the spectrometer. Typical operation conditions were: 1 s presaturation pulse for water suppression, 0.5 s light pulse (5 W), 5 ms RF pulse (90 degree flip angle), 1 s acquisition time, and 5 s delay. A number of 16 or 32 light scans gave an adequate signal-to-noise ratio for the tested samples. The CIDNP experiments were carried out at pH 6.5 in 10 mM phosphate buffer. The CIDNP effect caused by the Tyr-residues corresponds to a spin-density distribution of the intermediate phenoxy radical with strong negative signals of the ε_1, ε_2 protons and less intense positive signals for the δ_1, δ_2 protons. The CIDNP signals of tryptophan are generated by an intermediate radical with strong spin density at the δ_1, ε_3 and h_2 positions of the indole unit and very small spin density at the ζ_2 and ζ_3 positions, which all lead to positive CIDNP signals. The CIDNP difference spectra of histidine displays positive singlet signals for proton ε_1 and proton δ_2 [15].

Results and Discussion

Application of both FastA [20] and BLAST [21] algorithms with default set of functions implemented in PROMOD yield relatively high scores of similarity between the two bacterial enzymes (Fasta optimised score of 558; BLAST score of 202 and Poison probability P(N) of 4.3e-21). This result is illustrated by the alignment of the sequences of these two enzymes, as shown in Fig. 2. Owing to an insufficient extent of similarity, the N-terminal stretch of amino acid residues and two C-terminal amino acids of the small *Clostridium perfringens* sialidase have been excluded from further modelling (Fig. 3). The lack of conservation is judged to be an indicator that these parts are not crucial for functional aspects common to both enzymes. Notably, the alignment conserves the amino acids which are located in the four Asp boxes and the active site of the enzyme. Considering the extent of similarity and the homology score as meaningful indicators for the decision to perform knowledge-based modelling, these results justify to proceed to the calculations, implementing the program steps described in materials and methods. Modelling was performed with the wild-type enzyme and the mutants derived from single amino acid substitutions. The individual sequence alterations in mutant enzymes primarily affected their V_{max} values, as shown in Table 1. The K_M-values for the substrate methylumbelliferyl-N-acetylneuraminic acid is increased for the Y347F sialidase and is only slightly affected for the other two mutant en-

Table 2. *Surface accessibilities (Å²) of CIDNP-reactive amino acid side chains of wild-type Clostridium perfringens sialidase and its three mutants*

	residue	wild type area	wild type SD	Y336F area	Y336F SD	Y347F area	Y347F SD	C349S area	C349S SD
Tyr	35	24.7	2.0	12.2	13.3	13.6	4.3	16.0	5.4
	57	73.2	9.4	70.2	7.1	65.7	8.0	74.4	9.2
	65	54.3	6.6	49.7	9.5	49.2	15.0	55.7	4.5
	82	63.9	6.4	43.3	4.5	65.6	5.5	61.6	6.2
	95	58.7	1.8	56.1	4.6	87.6	8.4	56.9	8.9
	141	19.1	2.5	29.1	5.7	29.9	5.4	26.5	5.4
	203	61.5	4.8	63.0	8.7	41.5	5.8	59.3	5.2
	204	41.4	4.7	4.5	5.1	0.6	1.9	22.3	2.7
	209	23.4	9.1	26.1	9.7	34.1	14.0	1.8	3.0
	246	0.0	0.0	0.0	0.0	1.8	2.3	0.0	0.0
	248	25.1	5.3	30.4	4.2	15.6	4.9	17.7	4.0
	251	64.2	4.2	52.7	3.9	74.7	13.2	67.7	4.7
	255	0.0	0.0	2.1	3.4	0.9	2.7	34.9	5.9
	267	17.0	3.7	68.6	5.9	79.6	6.0	34.5	5.6
	310	56.3	5.2	101.2	3.9	34.4	13.5	49.4	10.6
	318	25.4	3.3	22.6	3.4	51.1	18.0	24.6	11.0
	336	64.8	6.3	****	****	122.0	8.7	96.5	7.0
	347	0.0	0.0	0.0	0.0	****	****	1.1	3.4
	361	36.6	3.6	27.3	6.6	36.7	9.0	22.1	6.1
	369	40.9	1.0	3.9	3.7	3.1	3.5	14.0	6.1
	375	38.8	4.5	91.6	7.1	98.8	9.8	5.3	4.0
	376	8.3	8.7	48.8	6.3	71.9	9.6	74.7	5.1
Trp	80	47.4	7.5	26.6	8.9	44.9	7.0	35.1	6.9
	118	0.0	0.0	0.2	0.8	2.8	2.5	0.0	0.0
	124	59.9	6.1	85.4	7.4	38.5	13.0	106.8	6.1
	135	0.0	0.0	14.0	10.9	18.4	4.9	13.8	2.7
	149	55.3	3.4	67.2	5.2	45.0	5.2	50.4	4.3
	172	3.4	4.1	14.9	4.2	0.0	0.0	10.6	9.8
	217	9.4	10.7	1.8	4.0	20.8	4.5	36.4	18.1
	264	4.3	4.7	19.9	9.1	16.0	4.5	13.8	6.5
His	63	35.9	4.4	27.9	4.4	33.2	5.0	1.8	3.2
	258	64.8	10.9	83.3	6.2	84.3	10.9	83.9	14.9
	285	77.1	14.3	75.3	8.6	80.5	6.5	87.5	10.5
	356	17.4	9.3	22.8	7.9	7.6	5.7	13.3	5.8

Table 3. *Calculated surface accessibility changes (Å²) of CIDNP-reactive amino acid side chains of three Clostridium perfringens sialidase mutants in comparison to the respective residues of the modelling-derived wild-type structure*

	residue	wild type	Δ-area Y336F	Δ-area Y347F	Δ-area C349S
Tyr	35	24.7	-12.5	-11.1	-8.7
	57	73.2	-3.0	-7.5	+1.2
	65	54.3	-4.6	-5.1	+1.4
	82	63.9	-20.6	+1.7	-2.3
	95	58.7	-2.6	+28.9	-1.8
	141	19.1	+10.0	+10.8	+7.4
	203	61.5	+1.5	-20.0	-2.2
	204	41.4	-36.9	-40.8	-19.1
	209	23.4	+2.7	+10.7	-21.6
	246	0.0	+0.0	+1.8	+0.0
	248	25.1	+5.3	-9.5	-7.4
	251	64.2	-11.5	+10.5	+3.5
	255	0.0	+2.1	+0.9	+34.9
	267	17.0	+51.6	+62.6	+17.5
	310	56.3	+44.9	-21.9	-6.9
	318	25.4	-2.8	+25.7	-0.8
	336	64.8	****	+57.2	+31.7
	347	0.0	0.0	****	+1.1
	361	36.6	-9.3	+0.1	-14.5
	369	40.9	-37.0	-37.8	-26.9
	375	38.8	+52.8	+60.0	-33.5
	376	8.3	+40.5	+63.6	+66.4
Trp	80	47.4	-20.8	-2.5	-12.3
	118	0.0	+0.2	+2.8	0.0
	124	59.9	+25.5	-21.4	+46.9
	135	0.0	+14.0	+18.4	+13.8
	149	55.3	+11.9	-10.3	-4.9
	172	3.4	+11.5	-3.4	+7.2
	217	9.4	-7.6	+11.4	+27.0
	264	4.3	+15.6	+11.7	+9.5
His	63	35.9	-8.0	-2.7	-34.1
	258	64.8	+18.5	+19.5	+19.1
	285	77.1	-1.8	+3.4	+10.4
	356	17.4	+5.4	-9.8	-4.1

zymes (Table 1). The maximal rate of cleavage, however, is decreased in all mutants, i. e. to a level of 37 % (C349S), 11 % (Y336F) and 0.009 % (Y347F).

The obtained four modelled structures (wild type and three mutants) were subjected to the molecular dynamics simulation and further minimization prior to the measurement of surface accessibilities of the three kinds of CIDNP-reactive amino acid side chains. The applied value for the dielectric constant guarantees a reasonable consideration of the solvent effect (e.g. dumping hydrogen bond forces), as already shown e.g. in the cases of the oligosaccharide chain of the ganglioside 9-O-acetyl-GD1a and a galectin-binding disaccharide (Galb1-2Galb1-R) [31-35]. A production period of 100 ps and an integration step of 1 fs was necessary to produce reliable trajectories within each simulation. Following a series of 250 integrations, the actual conformational parameters were stored. Based on 400 acquired conformations within the production time, those ten data sets with the lowest potential energy level were automatically selected and processed for energy minimization of the complete ensemble using the conjugate gradient method. The starting structure with the indication for spatial flexibility, as inferred by the calculations, is shown in Fig. 4.

As already outlined, model building is based on the assumption that sequence similarities of homologous proteins will translate into equivalent three-dimensional structures. Moreover, their elucidation by X-ray crystallography anchoring knowledge-based homology modelling is assumed to be of relevance for structural aspects in solution. Since the obtained structural models will yield detailed predictions on distinct conformational aspects such as Connolly surface areas of side chains, the degree of validity of such theoretical conclusions should be assessable by an adequate experimental approach such as laser photo CIDNP-technique. By individually calculating the surface accessibilities of the principally reactive residues on the basis of the modelling-derived data set, compiled in Table 2 and exemplarily illustrated for the wild-type enzyme in Fig. 5, testable assumptions for the shape of the CIDNP-spectra are envisaged. It is remarkable that the calculations for the three mutant enzymes unveil obvious differences in this conformational aspect (Table 3). Attributing a similar essential role in the catalytic mechanism for Tyr347, as inferred for the sialidase of *Salmonella typhimurium* [10], these data suggest that Phe347 is less suitable for this role. Furthermore, substitutions may trigger conformational changes affecting the active-site topology. However, evident changes in surface accessibility are not only seen for the Y347F mutant but also for the other Tyr/Phe- and the Cys/Ser-substitutions. Therefore, these substitutions do not seem to be close to neutral for this modelling derived surface accessibilities, which is tested experimentally in the next step. The analysis of the CIDNP-spectra will allow to infer the degree of validity for these theoretically attained conclusions.
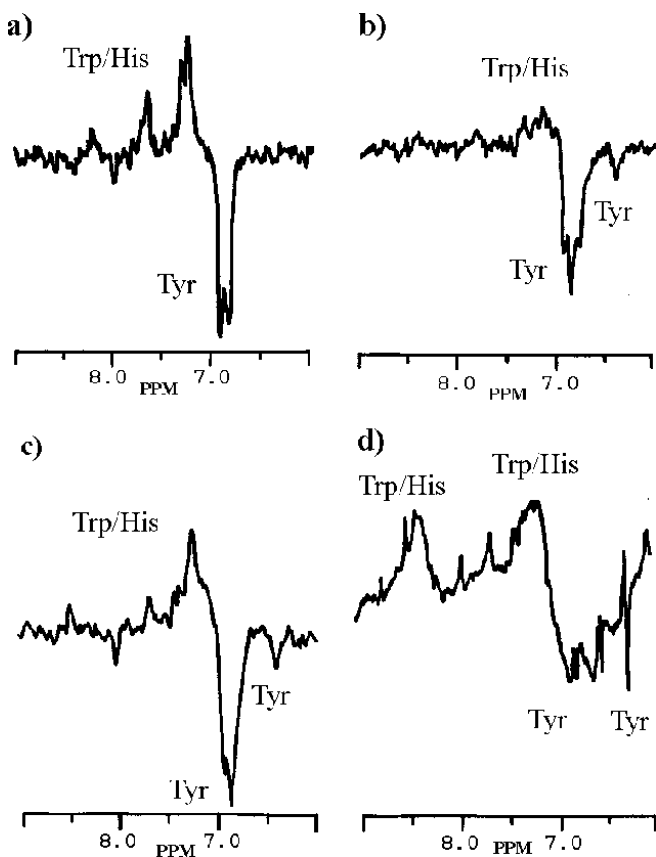
Figure 6. *Laser photo CIDNP difference spectra (aromatic part) of the Clostridium perfringens small sialidase: a. Wild-type enzyme; b. Mutant 1 (Y336F); c. Mutant 2 (Y347F); d. Mutant 3 (C349S).*

First of all, signals for the targeted side chains should be observable for the wild-type enzyme (Table 2). Therefore, *Clostridium perfringens* small sialidase and the three mutant forms with single-site substitutions are suitable for a CIDNP-NMR study directed towards the surface-exposed aromatic amino acids, tyrosine, tryptophan and histidine as sensitive sensors. According to our CIDNP-studies on hevein-like lectins one can expect to monitor a CIDNP-signal, when the surface accessibility area reaches a level of above 80 Å^2 [17]. Residues in the range between 60 and 80 Å^2 should be considered to be potentially feasible to generate or contribute to the intensity of the CIDNP signals. When more than one side chain of the same type is reactive, the overall signal will be shaped by the contributions of the individual moieties. Since five Tyr-residues are predicted to have a high surface exposition (Table 2), we can presume that the Tyr-signal in the spectrum of the wild-type enzyme, shown in Fig. 6a, is generated by an overlap. The CIDNP spectra of mutant 1 (Y336F) and mutant 2 (Y347F) display no major intensity changes and the occurrence of a new small Tyr-signal, as shown in Fig.

6b, c. These results point to alterations in the surface accessibility of a distinct Tyr-residue by conformational rearrangements induced by the amino acid replacement. In mutant 1 (Y336F) a tyrosine residue with a remarkably large surface accessibility area is substituted by a CIDNP-inert phenylalanine. The detailed comparison of the data sets reveals that this substitution is predicted to affect the positioning of several side chains, i.e. Tyr82, Tyr204, Tyr267, Tyr310, Tyr375 and Trp124 (Table 2, Table 3). A similar situation occurs for mutant 2 (Y347F). In this case a completely buried Tyr-residue has been chosen for the replacement, the resulting pattern of changes being non-identical. The impact on several residues hampers unambiguous interpretation of the spectra.

The new small Tyr-signal may be attributed to an increase of surface accessibility areas in both mutants Y336F and Y347F (Fig. 7). Such a common area–increase has been predicted for Tyr267 and even more distinct for Tyr375, as seen in Table 3. The residue Tyr267 with an surface area of 17.0 Å^2 in the wild type displays an area enlargement of 51.6 Å^2 for mutant 1 (Y336F) and 62.6 Å^2 for mutant 2 (Y347F). Remarkably, the corresponding Tyr-residue of mutant 3 (C349S) has an area increment of only 17.5 Å^2 and shall not give a contribution to a CIDNP-signal as similarly inferred for Tyr375 (Table 3). The surface area of this residue increases from 38.8 Å^2 to about 52.8 Å^2 in mutant 1 and about 60 Å^2 in mutant 2 which raises the expectation for a clearly visible Tyr-signal.

The detailed analysis of alterations in surface area above the relevant threshold value suggests that the overall Tyr-signal intensity should not vary markedly, as actually can be seen in Fig. 6a-c. Among the group of tyrosine residues, increases and compensatory reduction (or vice versa) are fairly balanced. However, the shape of Tyr-signals in the CIDNP-spectrum of the C349S mutant vary (Fig. 6d). The surface accessibility area for Tyr375, e.g. in mutant 3, decreases about 33.5 Å^2 (Table 3).

Relative to the interpretation of the tyrosine part of the CIDNP spectra, the analysis of the Trp/His-relevant part is aggravated by signal overlap to which weak Tyr signals of the δ_1, δ_2 protons can even contribute [17].

Likewise, the CIDNP spectrum of mutant 3 shows peculiarities with two rather strong Trp/His-signals one at 8.4 ppm and the other enhanced one at 7.2 ppm (Fig. 6d). The enhancement in Trp/His signal intensity for mutant 3 can be reconciled with an increase above the threshold value for clearly visible CIDNP signals (80 Å^2) as occurring for His285. The surface accessibility area of this residue enlarges from 77.1 Å^2 to 87.5 Å^2. Furthermore, Trp124 has a very high surface accessibility area of 106.8 Å^2 only in mutant 3. Although an unambiguous assignment for the altered Trp/His signals of mutant 3 is not possible, we have nonetheless found evidence that a Cys/Ser replacement indeed leads to distinct changes in the spectrum, which are reflected in the calculations (Table 2, Table 3).
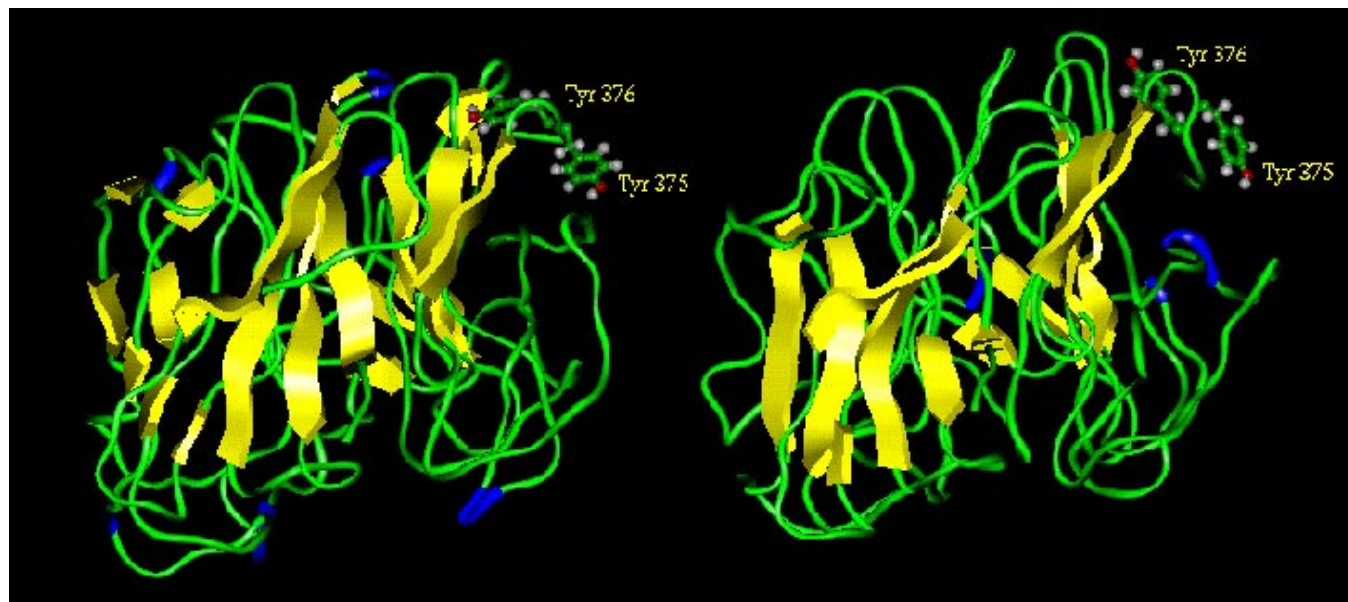
Figure 7. *Side views of the three-dimensional model structures of the Y336F (left) and Y347F (right) mutants of* Clostridium perfringens *small sialidase. Only side chains of Tyr375 and Tyr376, which have apparently enlarged their surface accessibility area in both mutants, are shown. Coloring is on the basis of secondary structure; yellow for sheets, green for random coils and blue for turns.*

out reliably, whether such rather global alterations can affect carbohydrate binding properties.

Conclusion

Starting from the crystal structure of a bacterial sialidase, knowledge-based homology modelling produces a set of energy-minimized conformations for the small sialidase of *Clostridium perfringens.* Although the number of CIDNP-reactive amino acids is too large to unequivocally assign signals to defined residues, what has been possible for hevein [17], the signal intensity can be reconciled with the model-derived expectations. To obtain clear-cut conclusions about the validity of the assumed solution structure, a more restricted size of the group of surface-presented amino acids is indispensable, defining the limits of applicability of the method. In addition to the description of certain properties of the wild-type enzyme the impact of introducing amino acid substitutions by site-directed mutagenesis has been theoretically delineated and experimentally tested with respect to CIDNP-reactive side chains. Intriguingly, the changes in surface accessibilities of widely separated residues affected by a Tyr/Phe- or a Cys/Ser-substitution intimate that disparate conformational changes of mutant enzymes relative to the wild-type form should not be underestimated. Extending the combination of computational and NMR techniques, as illustrated in recent reports and reviews with emphasis on sugar-binding proteins [31-35], it will technically amenable to figure

References

1. Gabius H.-J. and Gabius S. (eds.) (1993) *Lectins and Glycobiology* , Springer Verlag,Heidelberg-New York.
2. Gabius H.-J., Kayser K. and Gabius S. (1995) *Naturwissenschaften* **82**, 533-543.
3. Gabius H.-J. and Gabius S. (eds.) (1997)*Glycosciences: Status and Perspectives* , Chapman & Hall, Weinheim.
4. Schauer R. (1982) *Adv. Carbohydr. Chem. Biochem.* **40**, 131-234.
5. Schauer R. (1982) *Sialic Acids: Chemistry, Metabolism and Function*, Springer Verlag, Wien-New York.
6. Corfield A. (1992) *Glycobiology* **2**, 509-521.
7. Schauer R., Kelm S., Reuter G., Roggentin P. and Shaw L. (1995) in *Biology of the Sialic Acids* (Rosenberg A., ed.) pp. 7-67, Plenum Press, New York.
8. Reuter G. and Gabius H.-J. (1996) *Biol. Chem. Hoppe-Seyler*, **377**, 325-342.
9. Roggentin P., Schauer R., Hoyer L. L. and Vimr E. R. (1993) *Mol. Microbiol.* **9**, 915-921.
10. Crennell S. J., Garman E. F., Laver W. G., Vimr E. R. and Taylor G. L. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 9852-9856.

11. Crennell S. J., Garman E. F., Philippon C., Vasella, A., Laver W. G., Vimr E. R. and Taylor G. L. (1996) *J. Mol. Biol.* **259**, 264-280.

12. Crennell S. J., Garman E. F., Laver W. G., Vimr E. R. and Taylor G. L. (1994) *Structure* **2**, 535-544.

13. Gaskell A., Crennell S. J., and Taylor G. (1995) *Structure* **3**, 1197-1205.

14. Kaptein R., Dijkstra K. and Nicolay K. (1978) *Nature* **274**, 293-294.

15. Kaptein R. (1982) in *Biological Magnetic Resonance* **4** (Berliner, L. J., ed.) pp. 145-191, Plenum Press, New York.

16. Siebert H.-C., André S., Reuter G., Gabius H.-J., Kaptein R. and Vliegenthart J. F. G. (1995) *FEBS Lett.* **371**, 13-16.

17. Siebert H.-C., von der Lieth C.-W., Kaptein R.,Soedjanaatmadja U. M. S., Vliegenthart J. F. G., Wright C. S. and Gabius H.-J. (1996) *J. Mol. Model.* **2**, 351-353.

18. Connolly M. L. (1983) *J. Appl. Cryst.* **16**, 548-558.

19. Connolly M. L. (1983) *Science* **221**, 709-713.

20. Pearson W. R. and Lipman D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444-2448.

21. Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. (1990) *J. Mol. Biol.* **215**, 403-410.

22. Peitsch M. C. and Jongeneel C. V. (1993) *Int. Immunol.* **5**, 233-238.

23. Peitsch M. C. (1995) *Bio/Technology* **13**, 658-660.

24. Peitsch M. C. (1996) *Biochem. Soc. Trans.* **24**, 274-279.

25. Greer J. (1990) *Proteins* **7**, 317-334.

26. Ponder J. W. and Richards F. M. (1987) *J. Mol. Biol.* **193**, 775-791.

27. Brooks B. R., Bruccoleri R. E., Olafson B. D., States D. J., Swaminathan S. and Karplus M. (1983) *J. Comp. Chem.* **4**, 187-217.

28. Luthy R., Bowie J. U. and Eisenberg D. (1992) *Nature* **356**, 83-85.

29. Roggentin T., Kleineidam R. G., Schauer R. and Roggentin P. (1992) *Glycoconjugate J.* **9**, 235-240.

30. Kruse S., Kleineidam R. G., Roggentin P. and Schauer R. (1996) *Protein Express. Purific.* **7**, 415-422.

31. Siebert H.-C., von der Lieth C.-W., Dong X., Reuter G., Schauer R., Gabius H.-J. and Vliegenthart J. F. G. (1996) *Glycobiology*, in press.

32. Siebert H.-C., Gilleron M., Kaltner H., von der Lieth C.-W., Kozár T., Bovin N. V., Korchagina E. Y., Vliegenthart J. F. G. and Gabius H.-J. (1996) *Biochem. Biophys. Res. Commun.* **219**, 205-212.

33. Siebert H.-C., Kaptein R. and Vliegenthart J. F. G. (1993) in *Lectins and Glycobiology* (Gabius H.-J. and Gabius S., eds.) pp. 105-116, Springer Verlag, Heidelberg-New York.

34. Siebert H.-C., von der Lieth C.-W., Gilleron M., Reuter G., Wittmann J., Vliegenthart J. F.G. and Gabius H.-J. (1997) in *Glycosciences: Status and Perspectives* (Gabius H.-J. & Gabius S., eds.) Chapman & Hall, Weinheim, pp. 291-310.

35. von der Lieth C.-W., Kozár T., Hull W. E. (1996) *J. Mol. Struct. (Theochem)*, in press.