VTT PUBLICATIONS 342

Structural and functional studies of fungal cellulose-binding domains by NMR spectroscopy

Maija-Liisa Mattinen
VTT Chemical Technology

Laboratory of Analytical Chemistry
Department of Chemistry
University of Helsinki
Helsinki, Finland

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in Auditorium XII of the Main Building, Unioninkatu 34, Helsinki, on May 13th, 1998, at 12 o'clock noon.

TECHNICAL RESEARCH CENTRE OF FINLAND
ESPOO 1998

ISBN 951–38–5225–3
ISSN 1235–0621

ISBN 951–38–5226–1 (URL: http://www.inf.vtt.fi/pdf/) ISSN 1455–0849 (URL: http://www.inf.vtt.fi/pdf/)

JULKAISIJA – UTGIVARE – PUBLISHER

Valtion teknillinen tutkimuskeskus (VTT), Vuorimiehentie 5, PL 2000, 02044 VTT
puh. vaihde (90) 4561, telekopio 456 4374

Statens tekniska forskningscentral (VTT), Bergsmansvägen 5, PB 2000, 02044 VTT
tel. växel (90) 4561, telefax 456 4374

Technical Research Centre of Finland (VTT), Vuorimiehentie 5, P.O.Box 2000, FIN–02044 VTT, Finland
phone internat. +358 9 4561, telefax +358 9 456 4374

VTT Kemiantekniikka, Materiaalitekniikka, Molekyylien rakennetutkimus,
Biologinkuja 7, PL 1400, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 7026

VTT Kemiteknik, Materialteknik, Molekylers strukturforskning,
Biologgränden 7, PB 1400, 02044 VTT
tel. växel (09) 4561, fax (09) 456 7026

VTT Chemical Technology, Materials Technology, Molecular Structure Research,
Biologinkuja 7, P.O.Box 1400, FIN-02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 7026

Technical editing Maini Manninen

LIBELLA PAINOPALVELU OY, ESPOO 1998

Mattinen, Maija-Liisa. Structural and functional studies of fungal cellulose-binding domains by NMR spectroscopy. Espoo 1998, Technical Research Centre of Finland, VTT Publications 342. 72 p. + app. 63 p.

UDC 543.42:661.728

Keywords cellulose, cellulase, cellobiohydrolase, endoglucanase, cellulose-binding domain, fungi, NMR spectroscopy, *Trichoderma reesei*

Abstract

Cellulose has a important environmental role in the preservation of the global carbon cycle and commercial significance as a raw material for industry. To understand the biodegradation of cellulose on the atomic level, it is important to be able to relate enzyme activities with the three-dimensional (3D) structures of cellulases and cellulose. The cellulolytic system of the fungus *Trichoderma reesei* is one of the best understood of all cellulolytic systems. Most cellulases from *T. reesei* have a cellulose-binding domain (CBD) specialised for binding to cellulose.

As a means to understanding the interactions between cellulose and CBDs of cellobiohydrolase I (CBHI) and endoglucanase I (EGI) from *T. reesei*, 3D structures of these domains were studied by two-dimensional (2D) ^{1}H NMR (nuclear magnetic resonance) techniques. Structural effects of mutations of conserved amino acids in CBD_{CBHI}, Y5A, P16R, N29A, Y31A, Y32A and Q34A, were evaluated by comparing chemical shifts, coupling constants and NOEs of the backbone protons of the mutants and wild-type CBD_{CBHI}. In general, the substitutions did not alter significantly the secondary structures of these engineered peptides.

The adsorption experiments on cellulose showed that Y5A, Y31A and Y32A had lost nearly all their affinity to cellulose. For Y31A and Y32A the 3D structures revealed small local changes around the mutation on the flat face of CBD, which was expected to bind to cellulose. Therefore the structural roles of Tyr-31 and Tyr-32 must be minor, but their functional importance is clear since the mutants lacking these residues did not bind strongly to cellulose. In the case of Y5A the disruption of the structural framework at the N-terminus and the complete loss of affinity to cellulose implied that Tyr-5 has both structural and functional significance.
The 3D structure of a synthetic CBD_{EGI} was also determined by NMR spectroscopy. The structure was very similar to that of wild-type CBD_{CBHI}.

Within the precision of the structures, even the cellulose-binding face of CBD_{EGI} was similar to that of CBD_{CBHI}, apart from the place of attachment of the different side-chain. The determined NMR structure was also in agreement with an earlier modelled structure of CBD_{EGI}.

Finally, soluble cello-oligosaccharides were used as model compounds for cellobiose chains to investigate the interaction between CBD and cellulose by NMR spectroscopy. CBDs caused line broadening effects and decreasing T_2 relaxation times for certain sugar resonances, whereas there were no effects in the presence of a mutant that bound weakly to cellulose. Experiments showed that the interactions between CBD and cellobiose units of sugars are specific, supporting the model presented for the CBD binding to crystalline cellulose. It remained uncertain, however, how well the cello-oligosaccharides mimicked the binding of CBD to cellulose.

Preface

The work reported in this dissertation was carried out in the Laboratory of Chemical Technology of the Technical Research Centre of Finland (VTT) during the years 1993–1998. I wish first to express my sincere thanks to professors Torbjörn Drakenberg, Veikko Komppa and Markku Auer for providing excellent working facilities at VTT Chemical Technology. MSc. Arto Kiviranta is thanked for supporting me in the years it has taken to complete this work.

I warmly thank Professor Marja-Liisa Riekkola, the head of the Laboratory of Analytical Chemistry at the University of Helsinki, for her interest in my dissertation, and also for the encouragement she gave to continue my studies beyond the master's level. Professor Raimo Hiltunen at the Division of Pharmacognosy and Department of Pharmacy, University of Helsinki, is thanked for his help with grant applications. Professor Tuula Teeri at the Department of Biochemistry and Biotechnology, the Royal Institute of Technology in Stockholm, Sweden, and Docent Ilkka Kilpeläinen, the head of the NMR Laboratory of the Institute of Biotechnology at the University of Helsinki, are thanked for their critical comments on the manuscript.

I am also most grateful to my co-workers at VTT Biotechnology and Food Research Institute. In particular I would single out Dr. Tapani Reinikainen (currently Cultor Ltd. Technology Center), who initiated the CBD project, and Dr. Markus Linder, who synthesised all the CBD peptides studied in this work.

Special thanks go to Dr. Arto Annila for encouraging me to expand my knowledge from analytical chemistry to the field of biomolecular NMR spectroscopy, and time invested in instructive discussions concerning theory of NMR spectroscopy and its application to protein structure determination. My thanks go as well to Dr. Anita Teleman, MSc. Kimmo Pääkkönen and MSc. Vesa Harjunpää for their co-operation over the years.

Finally, I wish to thank my friends Minna Järvinen, Aila Maijanen, Jari Toiviainen, Kirsi Kataja and Eero Vuori for their caring attitude and for the non-scientific activities that we have enjoyed together, both in and outside the laboratory. The support of my mother and father is also most gratefully acknowledged.

The Academy of Finland and the Emil Aaltonen Foundation provided financial support during the years 1994–1998.

List of original publications

This thesis is based on the following original publications and one submitted manuscript, which are referred to in the text by the accompanying Roman numerals.

I Linder, M., Mattinen, M.-L., Kontteli, M., Lindeberg, G., Ståhlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G. and Annila, A. (1995). Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. Prot. Sci. 4, pp. 1056–1064.

II Mattinen, M.-L., Kontteli, M., Kerovuo, J., Linder, M., Annila, A., Lindeberg, G., Ståhlberg, J. and Drakenberg, T. (1995). Cellulose–cellulase interaction. Bull. Magn. Reson. 17, pp. 268–269.

III Mattinen, M.-L., Kontteli, M., Kerovuo, J., Linder, M., Annila, A., Lindeberg, G., Reinikainen, T. and Drakenberg T. (1997). Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from *Trichoderma reesei*. Prot. Sci. 6, pp. 294–303.

IV Mattinen, M.-L., Linder, M., Drakenberg, T. and Annila, A. (1998). Solution structure of the cellulose-binding domain of endoglucanase I from *Trichoderma reesei* and its interaction with cello-oligosaccharides. Submitted for publication in European Journal of Biochemistry.

V Mattinen, M.-L., Linder, M., Teleman, A. and Annila, A. (1997). Interaction between cellohexaose and cellulose-binding domains from *Trichoderma reese*i cellulases. FEBS Lett. 407, pp. 291–296.

Contents

Appendices of this publication are not included in the PDF version. Please order the printed version to get the complete publication (http://www.inf.vtt.fi/pdf/publications/1998/)

List of abbreviations

CBD	cellulose-binding domain
CBD_{N1}	N-terminal CBD of CenC
CBH	cellobiohydrolase
Cex	exoglucanase/xylanase from *Cellulomonas fimi*
CenC	β-1,4-glucanase from *Cellulomonas fimi*
Cip-CBD	CBD from *Clostridium thermocellum*
COSY	correlation spectroscopy
DG	distance geometry
1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
EG	endoglucanase
FID	free-induction decay
HOHAHA	homonuclear Hartmann–Hann
ISPA	isolated spin-pair approximation
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser enhancement spectroscopy
RELAYED-COSY	relayed coherence transfer correlation spectroscopy
rMD	restrained molecular dynamics
SA	simulated annealing
TOCSY	total correlation spectroscopy
T. reesei	*Trichoderma reesei*
TRNOE	transferred nuclear Overhauser enhancement
C. fimi	*Cellulomonas fimi*

1 Introduction

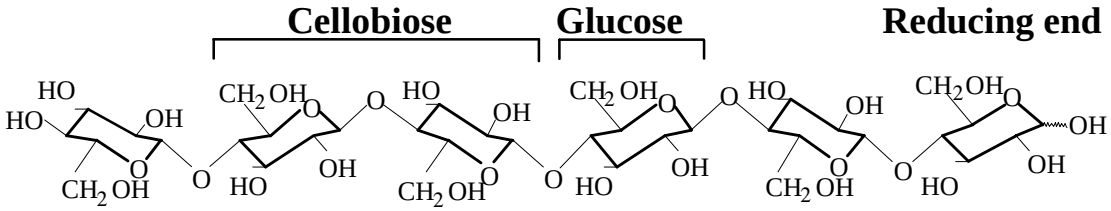
The hydrolytic degradation of cellulose is a complex phenomenon. It is carried out by cellulolytic micro-organisms, aerobic and anaerobic fungi and bacteria, which produce a variety of cellulolytic enzymes called cellulases. These enzymes generally have a cellulose-binding domain, which is specialised for binding to cellulose. One of the main objectives of this investigation was to determine the 3D structures of CBDs of the two major cellulases from *T. reesei* by NMR spectroscopy, to understand the phenomenon of cellulose–cellulase interaction on the atomic level.

1.1 Cellulose

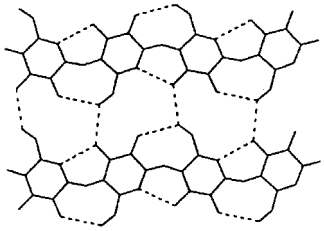
The chemical structure of natural cellulose is rather simple. It consists of long, parallel homopolymers of β-1,4-linked D-glucose monomers. Because every second glucose residue is rotated 180° with respect to the preceding residue, the successive glycosidic oxygens point in opposite directions. Hence, the smallest repeating unit of cellulose is cellobiose (Fig. 1a). The cellulose chains form long fibres, which are stabilised by a number of inter- and intramolecular hydrogen bonds (Fig. 1b). The length of the chains varies with the origin of the cellulose (Hon, 1994).

Cellulose may crystallise in four different forms, designated I, II, III and IV (Sarko, 1987), but only cellulose I occurs in nature. The other forms are obtained by chemical treatment of cellulose I. Native cellulose comprises two different allomorphs, a triclinic (I_α) and a monoclinic (I_β) phase, which have similar molecular conformations but different hydrogen-bonding patterns (Atalla and Vanderhart, 1984; Sugiyama *et al.*, 1991a–c). These two allomorphs may co-exist within a single microfibril of cellulose, but the relative amounts depend on the source of the cellulose (Atalla, 1993). Microfibrils contain crystalline and amorphous regions, and the size of the crystalline regions varies widely. The largest cross-section, 20 nm, is reported for algal *Valonia* cellulose and the smallest, 1.8 to 2.0 nm, for primary cell walls of plant cells (Sugiyama *et al.*, 1984, 1985; Kuga and Brown, 1987a–b; Chanzy, 1990). A model of the crystal faces of microfibril is shown in Fig. 1c (Gardner and Blackwell, 1974; Reinikainen *et al.*, 1995). Cellulose fibres are composed of several bunches of microfibrils (Fig. 1d).

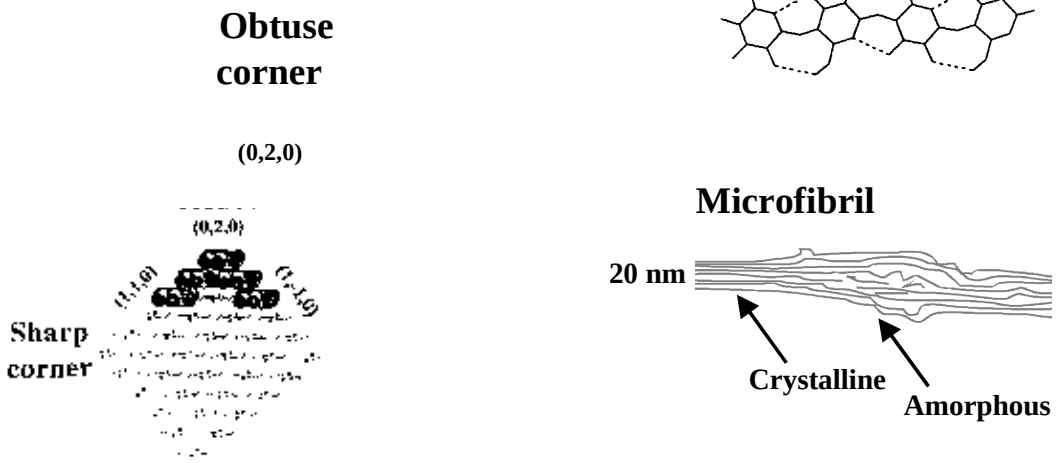
a.



b.



c.



d.

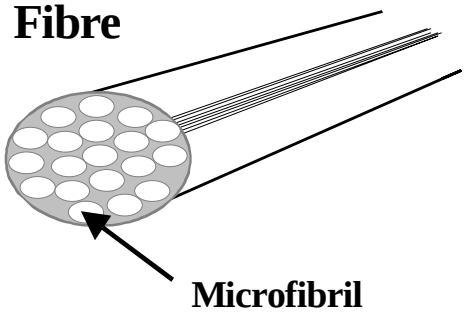


*Figure 1. Structure of cellulose. (**a**) Cellulose is a glucose polymer consisting of repeating units of cellobiose (Hon, 1994). (**b**) The parallel β-1,4-D-glucose chains form intra- and intermolecular hydrogen bonds (Ståhlberg, 1991). (**c**) These chains pack further into sheets, which form the microfibrils that contain both crystalline and amorphous regions. The cross-section of the microfibrils is drawn from the non-reducing end towards the reducing end. For clarity, only five strands at the top obtuse corner have been drawn with van der Waals radii. The model shows also the crystal faces of cellulose I (Kroon-Batenburg and Kroon, 1994; Reinikainen et al., 1995). (**d**) Cellulose fibre (Hon, 1994; Kroon-Batenburg and Kroon, 1994).*

1.2 Cellulases

In nature, cellulases are specialised for widely different environments. Some decompose plant litter in soil, whereas others act in anaerobic conditions in the rumen (Béguin and Aubert, 1994; Denman *et al.*, 1996).

A cellulolytic system typically consists of several enzymes, which can be classified into three categories on the basis of their activity. Exoglucanases degrade cellulose starting from the reducing and non-reducing chain ends. Because the main end-product of the reaction is cellobiose, the exoglucanases are also called cellobiohydrolases. These enzymes are usually active on the crystalline parts of cellulose. Endoglucanases cleave the cellulose chains at the interior sites and are active mainly on amorphous regions of cellulose. The endoglucanases can also hydrolyse substituted celluloses, such as carboxymethylcellulose and hydroxyethylcellulose. Together, the exo- and endogluganases act synergistically (Henrissat *et al.*, 1985; Nidetzky *et al.*, 1993; Irwin *et al.*, 1993; Divne *et al.*, 1994; Béguin and Aubert, 1994; Barr *et al.*, 1996). β-Glucosidases complete the decomposition process of cellulose by degrading small soluble oligosaccharides and cellobiose to glucose (Béguin, 1990). This last step is important because cellobiose is usually an inhibitor for cellulases. Thus the complete hydrolysis of cellulose to glucose demands the concerted action of exoglucanases, endoglucanases and β-glucosidases. In addition to these major cellulases, other enzymes, such as cellobiose oxidases and cellobiose dehydrogenases, may be operative in cellulose degradation (Coughlan, 1985). The detailed role of the oxidising enzymes is not clear, but perhaps they facilitate the action of cellulases by removing cellobiose from a reaction mixture when no β-glucosidases are present (Coudray *et al.*, 1982).

Fungal cellulolytic systems usually consist of several exo- and endoglucanases and one or two β-glucosidases. The number of the components depends on the fungus. The filamentous fungus *T. reesei* secretes a complete set of enzymes for the efficient hydrolysis of crystalline cellulose. The cellulolytic system consists of two cellobiohydrolases (CBHI and CBHII), at least four endoglucanases (EGI, EGII, EGIII and EGV) and one β-glucosidase. The biochemistry of these enzymes has been characterised in detail and they are usually active at acidic or neutral pH (Durand *et al.*, 1988; Nevalainen and Penttilä, 1995).

Bacterial cellulase systems are much more difficult to study than fungal cellulase systems and therefore much less is known about them. *Cellulomonas fimi* cellulases are among the most extensively studied bacterial enzymes (Coutinho *et al*., 1991, 1992; Din *et al*., 1991, 1994; Gilkes *et al*., 1991, 1992;

Meinke *et al*, 1991a−b, 1992, 1993; Pilz *et al*., 1990; Shen *et al*., 1991; Béguin, 1990; Wilson, 1992). Most bacterial cellulases cannot degrade crystalline cellulose effectively. Some organisms (for example the anaerobic bacterium *Clostridium thermocellum*) utilise large multi-enzyme complexes, known as cellulosomes, for cellulose degradation (Lamed *et al*., 1983; Fanutti *et al*., 1995; Dijkerman *et al*., 1996; Bayer *et al.*, 1995).

The complex and insoluble nature of cellulose requires some special properties of the enzymes degrading it. For example, the solid substrate cannot diffuse to the enzyme and thus the cellulase itself must be active and diffuse to the substrate. The crystallisation of an entire fungal cellulase for structure determination is difficult, probably due to the flexible and heterogeneously *O*-glycosylated linker peptide. Small-angle X-ray scattering (Abuja *et al*., 1988a−b) and dynamic light scattering (Boisset *et al*., 1995) have been used to study the spatial relationship between the individual domains of fungal cellulases. The results obtained with these two techniques indicate that both *T. reesei* cellobiohydrolases and *Humicola insolens* endoglucanase have a 'tadpole-like' organisation. The enzymes have two distinct domains, a large catalytic core domain and a small cellulose-binding domain, which are connected by a long linker peptide (Van Tilbeurgh *et al*., 1986; Teeri *et al*., 1987; Tomme *et al*., 1988) (Fig. 2). It is well established that the two-domain structure is a general feature of cellulases. Except for the small endoglucanase EGIII, all *T. reesei* cellulases have similar structural organisation, with two functionally and structurally distinct domains. Apparently some spatial separation between the domains is necessary for efficient function of cellulases on a solid crystalline cellulose substrate (Srisodsuk *et al*., 1993).

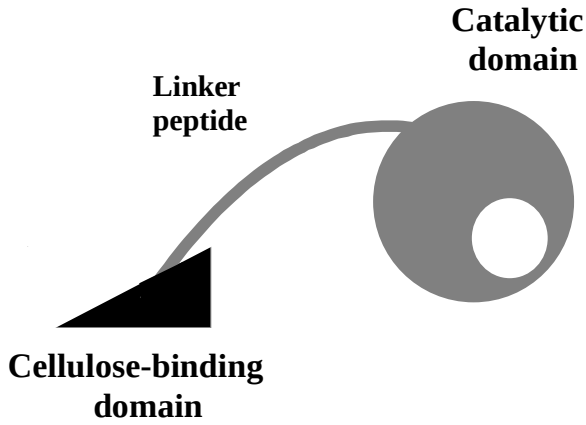


Figure 2. Schematic representation of the two-domain structure typical of cellulases. A large catalytic core domain is connected to a small cellulose-binding domain by a long O-glycosylated linker peptide.

Three-dimensional structures of isolated individual domains of many fungal and bacterial cellulases have been solved by X-ray crystallography and NMR spectroscopy. Crystal structures have been determined for the catalytic core domains of the *T. reesei* CBHI, CBHII and EGI (Divne *et al*., 1994; Rouvinen *et al*, 1990; Kleywegt *et al*., 1997) and the crystallisation of *T. reesei* EGIII has been reported (Ward *et al*., 1993). The 3D structures and properties of CBDs from different organisms are discussed separately in the following section (1.3).

1.3 Cellulose-binding domains

1.3.1 Classification and properties

The first CBD found was CBD_{CBHI} from *T. reesei* (Bhikhabhai *et al*., 1984; van Tilbeurgh *et al*., 1986). Since then, more than 120 CBDs have been identified and classified into ten families (I−X). Two of the larger families (II and III) have been divided into subfamilies (IIa, IIb, IIIa and IIIb). Several alternative classifications have been proposed (Gilkes *et al*., 1991; Coutinho *et al*., 1992; Béguin and Aubert, 1994), but only the most recent and most complete is presented in Table 1 (Tomme *et al*., 1995a). Most of the CBDs belong to families I, II and III; the remaining families contain only a few or in some cases only one member. Family I CBDs are all from fungi, while the CBDs of families II–V, VII, IX−X are all from bacteria.

The family I CBDs are small and compact peptides, each containing 32 to 36 amino acids. These fungal CBDs have very high sequence similarity. The best characterised CBDs in family I belong to cellulases from *T. reesei* (van Tilbeurgh *et al*., 1986; Kraulis *et al*., 1989; Ståhlberg *et al*., 1991; Reinikainen *et al*., 1992, 1995). The bacterial CBDs in family II are much larger than those in family I, and they contain 95−108 amino acids. The most studied family II CBDs are from *C. fimi* and *Pseudomonas fluorescens* (Gilkes *et al*., 1988, 1992; Ong *et al*., 1993; Poole *et al*., 1993; Din *et al*.; 1994; Hall *et al*., 1995). The members of family III are all found in cellulosome-producing bacteria. Some enzymes, typically ones belonging to families III, IV, VI or IX (Coutinho *et al*., 1992; Sakka *et al*, 1993; Winterhalter *et al*., 1995), have not just one but two CBDs. The likely advantage of having two adjacent CBDs is an increased affinity (Linder *et al*., 1996) or a broader substrate specificity of the enzyme (Coutinho *et al*., 1993). Multiple copies or internal CBDs have also been identified (Henrissat, 1994).

Table 1. Classification of CBDs from different organisms (Tomme et al., 1995a). 3D structures have been determined for bolded CBDs.

Family[1]	Organism	Enzyme	Location[2]	Residues[3]
I	*Agaricus bisporus*	Cel1	C	36
	Agaricus bisporus	Cel3a	N	36
	Agaricus bisporus	Cel3b	N	36
	Fusarium oxysporum	B homologue	N	33
	Fusarium oxysporum	C homologue 2	C	33
	Fusarium oxysporum	K homologue	C	37
	Fusarium oxysporum	Xyn	N	36
	Humicola grisea	CBHI	C	33
	Humicola insolens	EII	N	36
	Humicola insolens	CBHII	C	33
	Humicola insolens	A-1	-	33
	Humicola insolens	A-5	-	33
	Humicola insolens	A-8	-	33
	Humicola insolens	A-9	-	33
	Humicola insolens	A-11	-	33
	Humicola insolens	A-19	-	33
	Humicola insolens	43kDa	-	33
	Neocallimastix patriciarum	XylB	C	33
	Neurospora crassa	CBHI	C	33
	Penicillium janthinellum	CBHI	C	33
	Phanerochaete chrysosporium	CBHI	C	34
	Phanerochaete chrysosporium	CBHI-1	C	34
	Phanerochaete chrysosporium	CBHI-2	C	34
	Phanerochaete chrysosporium	CBHI-3	C	34
	Phanerochaete chrysosporium	CBHI-4	C	34
	Phanerochaete chrysosporium	CBHII	N	36
	Porphyra purpurea	PBP	x 4	33
	Trichoderma koningii	CBHI	C	33
	Trichoderma longibrachiatum	CBHI	C	33
	Trichoderma reesei	**CBHI**	**C**	**36[a]**
	Trichoderma reesei	CBHII	N	36
	Trichoderma reesei	**EGI**	**C**	**33[b]**
	Trichoderma reesei	EGII	N	36
	Trichoderma reesei	EGV	C	36
	Trichoderma reesei	Man	C	34
	Trichoderma viride	CBHI	C	33
IIa	*Butyrivibrio fibrisolvens*	End1	C	95
	Cellulomonas fimi	CenA	N	106
	Cellulomonas fimi	CenB	C	103
	Cellulomonas fimi	CenD	C	105
	Cellulomonas fimi	CbhA	C	106
	Cellulomonas fimi	CbhB	C	104
	Cellulomonas fimi	**Cex**	**C**	**106[c]**

(Table 1, continued)

Family[1]	Organism	Enzyme	Location[2]	Residues[3]
	Cellulomonas flavigenai	CflX	C	106
	Clostridium cellulovorans	EngD	C	108
	Clostridium longisporum	CelA	C	97
	Microbispora bispora	CelA	C	100
	Micromonospora cellulolyticum	McenA	C	100
	Pseudomonas fluorescens	EglA	C	100
	Pseudomonas fluorescens	CelB	N	102
	Pseudomonas fluorescens	CelC	N	99
	Pseudomonas fluorescens	CelE	C	100
	Pseudomonas fluorescens	XynA	N	101
	Pseudomonas fluorescens	XynB/C	N	99
	Pseudomonas fluorescens	XynD	N	102
	Streptomyces lividans	CelA	N	108
	Streptomyces lividans	CelB	C	106
	Streptomyces lividans	ChiC	N	105
	Streptomyces plicatus	Chi63	N	102
	Streptomyces rochei	EglS	C	103
	Thermomonospora fusca	E1	C	96
	Thermomonospora fusca	E2	C	96
	Thermomonospora fusca	E3	N	103
	Thermomonospora fusca	E4	C	104
	Thermomonospora fusca	E5	N	103
IIb	*Cellulomonas fimi*	XynD1	I	90
	Cellulomonas fimi	XynD2	C	90
	Streptomyces lividans	Axe	C	86
	Streptomyces lividans	XinB	C	86
	Thermomonospora fusca	XynA	C	86
IIIa	*Bacillus lautus*	CelA	C	150
	Bacillus lautus	ORF	C	150
	Bacillus subtilis BSE616	End	C	133
	Bacillus subtilis CK2	Cel	C	133
	Bacillus subtilis DLG	End1	C	132
	Bacillus subtilis N-24	End2	C	132
	Bacillus subtilis PAP115	End3	C	132
	Caldocellum saccharolyticum	CelA	I x 2	172/172
	Caldocellum saccharolyticum	CelB	I	172
	Caldocellum saccharolyticum	CelC	I x 2	172/172
	Caldocellum saccharolyticum	ManA	I x 2	172
	Clostridium cellulovorans	CbpA	N	161
	Clostridium stercorarium	CelZ	C	133
	Clostridium thermocellum	Cbh3	C	132
	Clostridium thermocellum	CelI	C	150
	Clostridium thermocellum	**CipA**	**I**	**156[d]**
	Clostridium thermocellum	CipB	I	167

(Table 1, continued)

Family[1]	Organism	Enzyme	Location[2]	Residues[3]
	Erwinia carotovora	CelV	C	156
IIIb	*Cellulomonas fimi*	CenB	I	131
	Clostridium cellulolyticum	CelCCG	I	138
	Clostridium stercorarium	CelZ	I	144
	Clostridium thermocellum	CelF	I	142
	Clostridium thermocellum	CelI	I	137
IV	***Cellulomonas fimi***	**CenC**	**N**	**148[e]**
	Cellulomonas fimi	CenC	I	148
	Clostridium cellulolyticum	CelCCE	N	168
	Myxococcus xanthus	CelA	N	139
	Streptomyces reticuli	Cel1	N	125
	Thermomonospora fusca	E1	N	141
V	***Erwinia chrysanthemi***	**EgZ**	**C**	**63[f]**
VI	*Bacillus polymyxa*	XynD	C	90
	Clostridium stercorarium	XynA	C x 2	87/92
	Clostridium thermocellum	XynZ	I	92
	Limulus sp.	Factor Gα	C x 2	87
	Microspora bispora	BglA	C	85
VII	*Clostridium thermocellum*	CelE	I	240
VIII	*Dictyostelium discoidum*	CelA	N	152
IX	*Clostridium thermocellum*	XynX	C x 2	174/189
	Thermoanaerobacterium saccharolyticum	XynA	C x 2	174/187
	Thermotoga maritima	XynA	C x 2	170/180
X	*Cellvibrio mixtus*	XynA	C	51
	Pseudomonas fluorescens	EglA	I	55
	Pseudomonas fluorescens	EglB	I	55
	Pseudomonas fluorescens	CelC	I	53
	Pseudomonas fluorescens	CelE	I	53
	Pseudomonas fluorescens	XynA	I	53
	Pseudomonas fluorescens	XynE	C	55

[1]Roman numerals indicate families and letters indicate subfamilies. [2]Location in the enzyme sequence: N = N-terminal, C = C-terminal, and I internal CBDs. 'x 2' and 'x 4' indicate two and four CBDs respectively. [3]Number of amino acids in the binding domain. References: [a]Kraulis et al., 1989; [b]publication IV; [c]Xu et al., 1995; [d]Tormo et al., 1996; [e]Johnson et al., 1996a; [f]Brun et al., 1998.

The specific properties of CBDs vary among members of the same family. Affinity is one parameter which changes from one CBD to another (Tomme *et al*., 1995b; Linder *et al*., 1995, 1996; Reinikainen *et al*., 1997). More qualitative differences, such as preference for different types of cellulose, have also been

reported (Coutinho *et al*., 1993). For example, in subfamily IIb, several CBDs appear to bind to xylan and some CBDs bind to both xylan and cellulose. In contrast, CBDs belonging to subfamily IIa appear to bind exclusively to cellulose (Irwin *et al*., 1994; Millward-Sadler *et al*., 1994).

1.3.2 Interaction with cellulose

Ever since the first CBD was discovered, a large number of biochemical studies have been carried out to understand how the two-domain structure of the cellulases works and to determine the exact role of CBD.

The effect of CBD on the enzyme function has been studied by separation of the core domain from the linker peptide and CBD by proteolytic cleavage or genetic engineering. The experiments have shown that, in isolation, the catalytic core domain has the same activity on soluble substrates as before, but the binding and activity towards insoluble cellulose are reduced. This has been demonstrated for both CBHI and CBHII (van Tilbeurgh *et al*., 1986; Tomme *et al*., 1988; Gilkes *et al*., 1988). Changing the length of the linker peptide between the core and CBD affects the enzymatic activity as well, which suggests that these two domains act in concert during the hydrolysis of cellulose (Shen *et al*., 1991; Srisodsuk *et al*., 1993; Wilson *et al*., 1995).

The presence of CBD is required for full activity of cellulases on crystalline cellulose. Even in the absence of CBD the catalytic domain is capable of hydrolysing solid cellulose to some extent. This indicates that CBD is not an essential part of the actual hydrolytic event, although it has been shown that an isolated CBD binds strongly to cellulose (Ståhlberg *et al*., 1991). The function of CBD has been proposed to enhance the activity of cellulases by increasing effective enzyme concentration on the cellulose surface (Ståhlberg *et al*., 1991). The general rule ‘the better the adsorption, the better the catalysis’ (Klyosov, 1990) seems to be true. It is also possible that cellulases with CBDs are required at early stages of cellulose degradation when most of the substrate is insoluble. At later stages of hydrolysis, when large amounts of soluble oligosaccharides are present, enzymes in the liquid phase are preferred. Catalytic core domains may be created, for example, by specific proteolysis of CBD (Linder and Teeri, 1997).

CBDs might also have an active role during the hydrolysis of cellulose. It has been suggested that CBDs release individual cellulose chains from the cellulose surface and supply them to the active site of the enzyme (Knowles *et al*., 1987;

Teeri *et al*., 1987; 1992). So far, only experiments with family II CBDs on cotton substrate have been reproducible.

A relevant question regarding CBD adsorption to cellulose is how tightly the peptide binds to cellulose. Because isolated core domains are easier to elute from the cellulose surface, it has long been thought that CBDs bind irreversibly to cellulose (Henrissat, 1994). Recently, however, it has been shown with CBD_{CBHI} that CBD binding to cellulose is fully reversible. This suggests a mode of action where the enzyme moves over the cellulose surface either progressively hydrolysing a cellulose chain or seeking new chain ends to initiate hydrolysis (Linder and Teeri, 1996). Also calculations of the cellulose surface coverage suggest that the CBDs likely bind over the entire length of cellulose fibres (Gilkes *et al*., 1992; Linder *et al*., 1995).

1.3.3 Three-dimensional structures

The first resolved 3D structure of a CBD was the family I *T. reesei* cellobiohydrolase I CBD (Kraulis *et al*., 1989). Recently several CBD structures from different families (II, III and IV) and organisms have been published (Xu *et al*., 1995; Tormo *et al*., 1996; Johnson *et al*., 1996a), and the structure of the family V *Erwinia chrysanthemi* EGZ will be presented in the near future (Brun *et al*., 1998).

Family I

The structure of 36-residue fungal CBD_{CBHI} has been determined by NMR spectroscopy (Kraulis *et al*., 1989). CBD_{CBHI} folds (Fig. 3) into a small wedge-shaped structure consisting of an irregular antiparallel triple-stranded β-sheet, which is stabilised by two disulfide bridges. The flat face of CBD_{CBHI} contains a planar region formed by one glutamine (Gln-34), one asparagine (Asn-29) and three aromatic residues (Tyr-5, Tyr-31 and Tyr-32), where the aromatic residues are spaced apart at a distance equal to the distance between every second glucose unit of the cellulose chain. Family I CBDs adsorb to crystalline and amorphous cellulose (Tomme *et al*., 1988; Reinikainen *et al*, 1992; Srisodsuk *et al*, 1993). The 3D CBD structures of CBHII, EGI, EGII and EGV from *T. reesei* have been homology modelled on the basis of the structure of CBD_{CBHI} (Hoffrén *et al*., 1995).

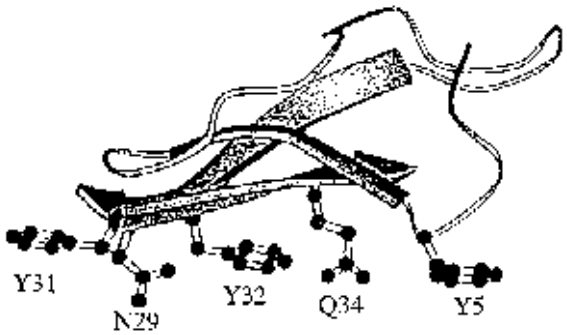


Figure 3. Schematic representation of CBD_{CBHI} with flat face (Kraulis et al., 1989).

Family II

The 3D structure of a 110-residue bacterial CBD from *C. fimi* Cex belonging to family II has been solved by NMR spectroscopy by Xu *et al*. (1995). CBD_{Cex} consists of a β-barrel fold of nine antiparallel β-strands (Fig. 4). A single disulfide bridge connects the N- and C-terminuses, and in solution CBD_{Cex} occurs as a dimer. One face of CBD_{Cex} forms a patch with three conserved tryptophans (Trp-17, Trp-54 and Trp-72) and one asparagine (Asn-87). Hydrophobic interactions between the cellulose surface and rings of tryptophans appear to be involved in the cellulose binding (Din *et al*., 1995). With the exception of CBD from xylanase D, the family II CBDs adsorb to both crystalline and amorphous cellulose.

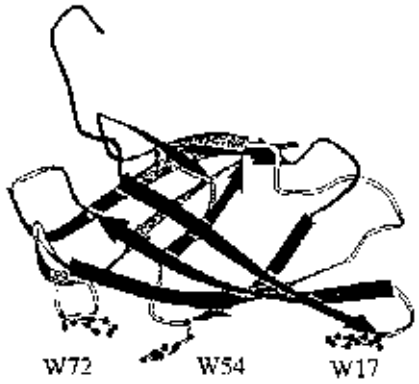


Figure 4. Schematic representation of the C. fimi CBD_{Cex}. Three tryptophan residues involved in the binding to cellulose and one disulfide bridge are also shown (Tomme et al., 1995a).

Family III

The 3D structure of a family III CBD from the cellulosomal scaffoldin subunit of *Clostrium thermocellum* has been determined by X-ray crystallography by Tormo *et al*. (1996). The scaffoldin CBD serves as a target of the entire cellulosome complex towards cellulose substrate (Shimon *et al*., 1997). The 155 residues of Cip-CBD peptide are folded into two antiparallel β-sheets, which form a β-sandwich with a jelly-roll topology. In addition to the nine major β-strands, the CBD contains loops connecting the secondary structure elements and a calcium binding site. On the face that is proposed to bind to crystalline cellulose, there is a planar area of aromatic residues (Tyr-67 and Trp-118), one histidine (His-57) and two polar side-chains of aspartic acid (Asp-56) and arginine (Arg-112) (Fig. 5).

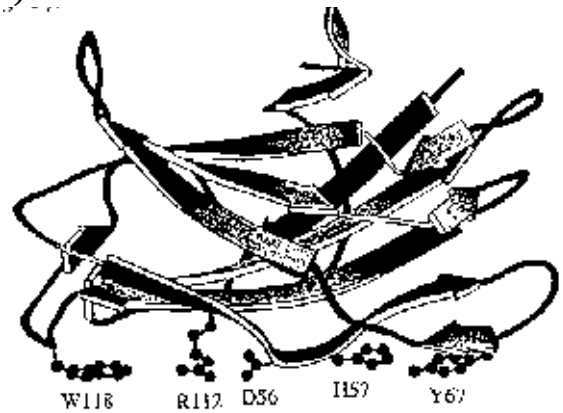


Figure 5. Structural organisation of Cip-CBD. The residues proposed to interact with a single glucose chain of cellulose are labelled (Tormo et al., 1996).

Family IV

Multidimensional heteronuclear NMR spectroscopy has been used to determine the structure of the 152-residue N-terminal CBD from *C. fimi* 1,4-β-glucanase CenC (Johnson *et al*., 1996a). The family IV CBD_{N1} is composed of ten β-strands folded into two antiparallel β-sheets, which form a β-sandwich with jelly-roll topology (Fig. 6). CBD_{N1} has a unique selectivity towards amorphous cellulose and soluble oligosaccharides (Johnson *et al*., 1996b). It does not bind to crystalline cellulose. This CBD does not have a flat surface as the other CBDs but a binding cleft containing hydrophobic amino acids (Val-17, Tyr-19, Val-48, Leu-77 and Ala-126). In addition to these residues there are numerous

hydrophilic groups (Asn-50, Arg-75, Asn-81, Thr-87, Asp-90, Gln-124 and Gln-128) which flank both sides of the binding site.

Figure 6. A ribbon diagram of CBD_{NI}. In addition to β-strands the binding cleft is shown with hydrophilic and hydrophobic residues (Johnson et al., 1996a).

Family V

2D ^{1}H NMR spectroscopy has been used to determine the 3D structure of the 62 amino acid C-terminal CBD of the endoglucanase Z (EGZ) secreted by *Erwinia chrysanthemi* (Brun *et al.*, 1998). The family V CBD_{EGZ} folds in an 'L' or 'ski boot' shape. The well-defined part of the molecule consists of several consecutive turns and a triple-stranded antiparallel β-sheet, which is perpendicular to a less-ordered loop (Fig. 7). Three exposed and aligned aromatic side-chains (Trp-18, Trp-43 and Tyr-44) are localised on one face of the molecule where they form a putative cellulose-binding site, which can cover five to six glucose units on the crystalline cellulose surface.

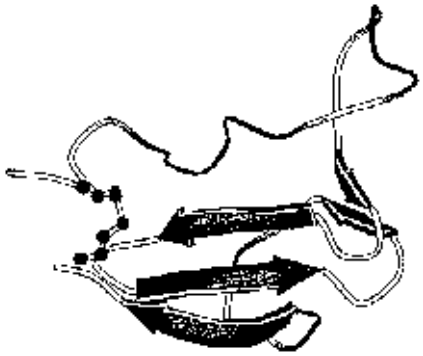


Figure 7. Schematic representation of CBD_{EGZ} with the secondary structure elements (Brun et al., 1998).

1.3.4 Use in biotechnology

The first systematic studies of cellulolytic enzymes were aimed at understanding and solving the problem of the decomposition of cellulose-based materials. In the future, CBDs may find importance in the biotechnology industry, but the success of many applications of cellulases requires a better understanding of how large-scale processes can be controlled and improved.

Because CBDs fold independently and bind spontaneously to cellulose, they are ideal affinity tags for the specific immobilisation of proteins to a cellulose surface. Cellulose is a readily available, inexpensive, and biologically degradable matrix material and thus suitable for large-scale affinity purification or protein immobilisation. Cellulose is also chemically inert, which makes it suitable for the purification of food and pharmaceutical products. There are many potential applications for this technology and several examples of fusion proteins constructed for specific immobilisation on cellulose matrices have been reported (Greenwood *et al*., 1989; Ong *et al*., 1989, 1991; Phelps *et al*., 1994; Bayer *et al*., 1994, 1995; Reinikainen *et al*., 1997).

Recently, cellulases have been used for mild and selective modification of cellulosics, for example in the paper, textile and detergent industries to effect surface treatment without altering mechanical properties (Woodward *et al*., 1994; Pere *et al*., 1995; Rahkamo *et al*., 1996). In the textile industry, cotton cloth can be made softer by limited hydrolysis, and in paper manufacturing the drainability of fibre mass can be improved by surface modificatons. Cellulases can also be used for the purification of recycled newspaper (Woodward *et al*., 1994).

There has been some speculation that cellulases may be useful for the transformation of cellulose into soluble and fermentable sugars so that they become a source of renewable energy and chemicals. However, in total hydrolysis of cellulose to fermentable sugars, the efficient elution of the enzymes from the residual substrate that is required for their recycling is still a major cost problem (Jackson *et al*., 1996).

In all these applications, better understanding and control of the CBD adsorption could lead to the technical breakthroughs required to make them economically feasible. Furthermore, novel applications may arise from the genetic engineering of cloned enzymes.

2 Aims of the present study

Understanding of the hydrolysis of cellulose on an atomic level requires the relating of enzyme activity with the 3D structures of cellulases and cellulose. I have been studying the relationship by NMR spectroscopy, using synthetic family I CBDs from *T. reesei* cellulases. The specific aims of my investigation were the following:

1. To evaluate the possible structural changes, relative to the parent CBD, occurring in six engineered CBDs of CBHI (Y5A, P16R, N29A, Y31A, Y32A and Q34A). It was not known whether the functional changes in these peptides revealed in adsorption experiments were due to the altered side-chain functionality or to structural distortions induced by the mutations. The chemical shifts, coupling constants and NOEs of the backbone protons for the wild-type CBD_{CBHI} and the analogues were compared as a means of studying this problem. (*Publication I*)

2. To determine the high-resolution 3D structures of Y5A, Y31A and Y32A by 2D ^{1}H NMR. When the tyrosines on the flat face of CBD_{CBHI} were replaced by alanine, the affinity towards crystalline cellulose was dramatically reduced. It was impossible to evaluate satisfactorily the structural changes in these mutants on the basis of chemical shift, coupling constant and NOE comparisons. Only with the 3D structures did it become possible to examine in detail the structural changes on the binding face of these peptides. (*Publications II and III*)

3. To determine the high-resolution 3D structure of CBD_{EGI} by 2D ^{1}H NMR and compare it with the structure of CBD_{CBHI} and the model of CBD_{EGI}. (*Publication IV*)

4. To investigate the interaction between the soluble CBD and solid cellulose by NMR, by using soluble cello-oligosaccharides as model compounds for the cellobiose chain on the cellulose surface. The binding of enzyme to crystalline substrate is notable in that there are only a few biological systems in nature which involve reaction between two different phases, aqueous and solid. In practice, it is difficult to study the aqueous and solid phases of a the system simultaneously, since most of the presently available methods are applicable to one phase only. (*Publication V*)

3 Materials and methods

Several publications have reviewed 1D and 2D ^{1}H NMR methods in their theoretical (Sörensen *et al*., 1983; Harris, 1986; Ernst *et al*., 1989; Homans, 1992; Cavanagh *et al*., 1996) and practical aspects (Wüthrich, 1986; Bax, 1989; Wagner, 1990; Clore, 1991a–b; Roberts, 1993; Gronenborn and Clore, 1995). Thus the theory of the experiments requiring quantum mechanical formalism and a mathematical pulse operator theory is not described here in detail. Instead I present a short overview of the principles of the procedures that I applied in the protein structure determinations and the interaction studies of these CBDs with cellulose.

3.1 Introduction to NMR spectroscopy

NMR spectroscopy deals with the interaction between a magnetic field and the nuclear magnetic moments that arise from the nonzero spin angular momentum of a nucleus that is quantized. For NMR spectroscopy of proteins, the most important nuclei are the isotopes ^{1}H, ^{13}C and ^{15}N with nuclear spin quantum number (I) ½ and the deuteron (^{2}H) with 1. The nuclear spin angular momentum (**I**) is a vector quantity with magnitude given by

$$|\mathbf{I}^2| = \mathbf{I} \bullet \mathbf{I} = \hbar^2[I(I+1)]; \qquad I_z = \hbar m, \tag{1}$$

in which $\hbar$ is *Planck*'s constant, h, divided by 2π, and m is a magnetic quantum number. Only one of the *Cartesian* components of **I,** I_z**,** can be defined simultaneously with $\mathbf{I}^2$, due to the quantum mechanical restrictions. When a nucleus is placed in a magnetic field, the magnetic quantum number may take values m = ($-I$, $-I$ + 1,…, I - 1, I). Thus I_z has $2I$ + 1 possible values. For example, for a nuclei with I = ½, two values of m ($2I + 1 = 2 \cdot ½ + 1 = 2$) result in two equally spaced energy levels (α, β), which are known as the *Zeeman* levels. The nuclear magnetic moment ($\boldsymbol{\mu}$) is collinear with the nuclear spin angular momentum (**I**) and is defined by

$$\boldsymbol{\mu} = \gamma \mathbf{I}; \qquad \mu_z = \gamma I_z = \gamma \hbar m, \tag{2}$$

in which γ is the gyromagnetic ratio. In an external magnetic field, the spin states of the nucleus have energies given by

$$E = -\boldsymbol{\mu} \bullet \mathbf{B}, \tag{3}$$

in which **B** is the magnetic field vector. The *z* axis is usually defined along the magnetic field. Then the equation (3) can be written

$$E_m = -\gamma I_z B_0 = -\gamma \hbar m B_0, \qquad (4)$$

in which B_0 is the strength of the static magnetic field. At the equilibrium there is an excess of spins in the α state determined by the temperature and the strength of B_0 according to the *Boltzmann* distribution

$$N_\alpha/N_\beta = \exp(-\Delta E/kT) = \exp(-h\nu_0/kT) = \exp(-\hbar\gamma B_0/kT) \cong 1 - \gamma\hbar B_0/kT, \qquad (5)$$

in which *k* is the *Boltzmann* constant, N_α and N_β are the number of spins in the corresponding energy levels, and *T* is the absolute temperature. At room temperature with pre-set magnetic field strengths (expansion in equation 5), it is only the small population difference which gives rise to a net magnetisation of the sample, M_0, parallel to the static magnetic field, B_0. This is the reason for the low sensitivity of NMR as compared with infrared and ultraviolet spectroscopy.

Electromagnetic radiation at frequency (ν_0) can induce transitions between *Zeeman* levels according to the *Bohr* frequency condition. The selection rule for the magnetic dipole transitions is $\Delta m = \pm 1$ (Fig. 8). Thus the photon energy required to excite a transition between the *m* (-½) and *m* + 1 (+½) *Zeeman* states is

$$\Delta E = h\nu_0 = \hbar\gamma B_0; \qquad \nu_0 = \omega_0/2\pi = \gamma B_0/2\pi; \qquad \omega_0 = \gamma B_0. \qquad (6)$$

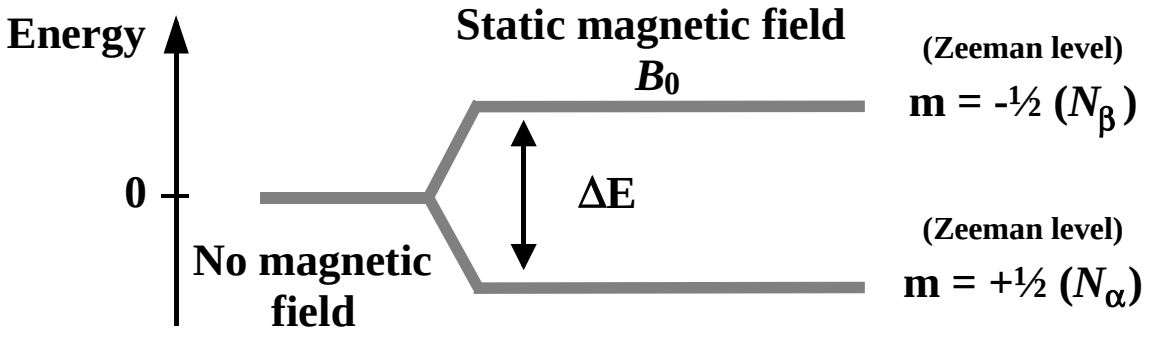


Figure 8. Nucleus with I = ½ in the magnetic field. Spin populations are not equal in the two energy levels. Adsorption of electromagnetic radiation induces transitions between the states (Field, 1989).

The net magnetisation precesses around the main static field axis (defined as the *z* direction) at its *Larmor* frequency (ω_0). The precessional frequency is identical with the frequency of electromagnetic radiation required to excite transitions between *Zeeman* levels (Fig. 9).

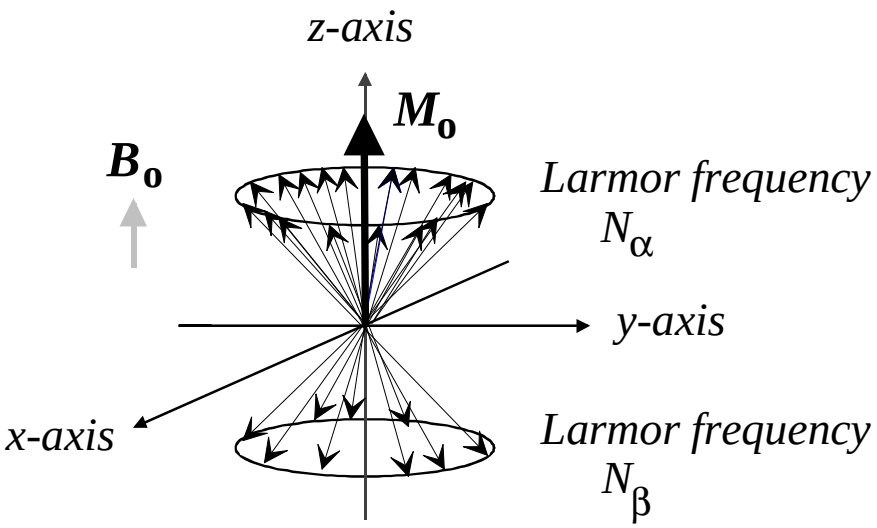


Figure 9. Distribution of the precessing nuclear spins, indicated as arrows, and the resultant net magnetisation (M_0) (Derome, 1987).

The nuclei in a molecule resonate at different frequencies because they experience a magnetic field somewhat different from B_0 depending on their local molecular surroundings. A nuclei can also sense the spin state of other nuclei via through-bond connectivities, if they are not too far apart in the molecule. Neighbouring magnetic dipoles can also interact through-space. The real frequency for each nucleus (chemical shift) is then

$$\nu = \gamma B_0/2\pi(1 - \sigma), \tag{7}$$

in which σ is the shielding factor. Instead of magnetic field-dependence of the absolute value of the chemical shift, a relative value is used:

$$\delta = [(\nu_i - \nu_{ref})/\nu_{ref}]10^6. \tag{8}$$

Chemical shift (δ) of a nucleus (i) is defined in parts per million (ppm) relative to some reference nucleus, often the methyl protons of tetramethylsilane (TMS).

3.2 One- and two-dimensional nmr experiments

In pulse NMR spectrometers the sample is irradiated with all frequencies at once and frequencies corresponding to the absorbed energy are sorted afterwards. 1D pulse NMR experiments can be explained rather simply by a vector model and rotating co-ordinate system (Fig. 10).

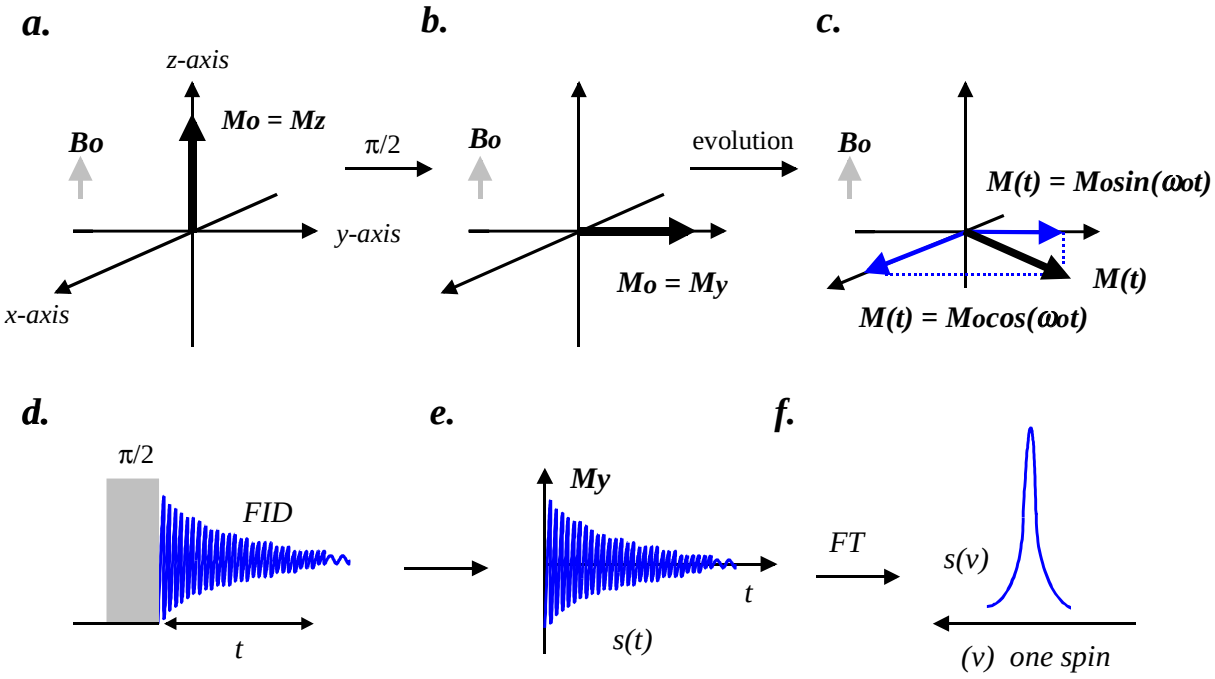


*Figure 10. A 1D pulse NMR experiment. (**a**) The net magnetisation (M_o) at the equilibrium along the z-axis. (**b**) Orientation of M_o along the y-axis after $\pi/2_x$-pulse. (**c**) Precession of M_o in the x-y plane. (**d**) A pulse sequence corresponding the 1D NMR experiment. (**e**) A sinusoidally decaying signal as a function time. (**f**) The detected signal after Fourier transform as a function of frequency (Derome, 1987).*

When the sample is placed in the magnetic field, the net magnetisation (M_0) is aligned parallel to the external field (B_0). To induce NMR transitions between the two energy levels, a radiofrequency field (B_1) is applied along the *x*-axis perpendicular to B_0. The pulse rotates M_z around the *x*-axis and the angle of the net magnetisation depends on the length of the pulse. For example, a π/2 pulse flips M_z along the *y*-axis; M_z is then zero and the two *Zeeman* levels are equally populated. After that, the transverse magnetisation (M_y) begins to precess around the *z*-axis with the *Larmor* frequency. However, the magnetisation

returns to the equilibrium position while still rotating around the *z*-axis. The rate at which the magnetisation is returning back along the *z*-axis is called the longitudinal or spin-lattice relaxation rate ($1/T_1$) and the rate at which the M_x and M_y components are decaying is called the transverse or spin–spin relaxation rate ($1/T_2$). A receiver coil in the *x*–*y* plane detects a current, which is proportional to the magnitude of the sinusoidally decaying signal of M_y, as a function of time. This time-domain signal (free-induction decay, FID) can be converted to a frequency-domain signal, i.e. the NMR spectrum, by *Fourier* transformation.

There are always several overlapping resonances in the 1D NMR spectra of proteins. However, these resonances can be often separated by introducing an additional dimension to the normal intensity vs. frequency representation of the 1D spectrum. This type of experiment was first suggested by Jeener (1971). A fundamental difference between 1D and 2D spectroscopy is that in 1D experiments only the *Zeeman* interaction and spin–spin interactions appear, but in 2D experiments pair-wise spin–spin interactions appear as well. 2D NMR experiments are divided into four periods (Fig. 11).

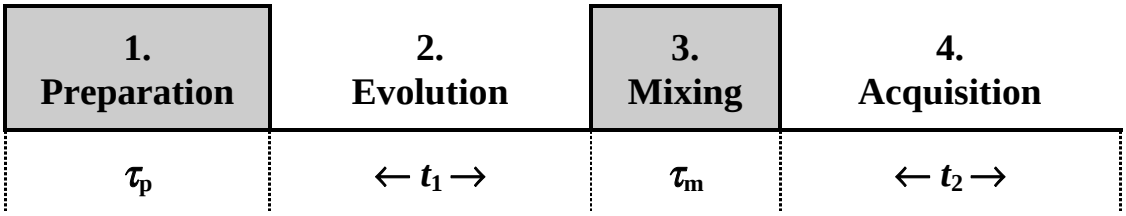


Figure 11. A basic scheme for the 2D NMR experiment. During the preparation period the spin system of the sample is changed (often by a π/2 pulse) to a non-equilibrium state involving some type of spin coherence. After evolution of the spin system, the coherence is transferred from one spin to another (mixing) and the decaying signal is detected during the acquisition period (Cavanagh et al., 1996).

A 2D NMR experiment comprises a set of 1D experiments (Fig. 12) in which t_1 is incremented in small steps (Δt_1). When one FID is recorded during t_2 for each value of t_1, an array of N individual FIDs is obtained. This forms a 2D data matrix $S(t_1, t_2)$. *Fourier* transformation of this time-dependent data matrix with respect to t_2 yields a series of spectra which vary in intensity and phase as a function of t_1. A second *Fourier* transformation with respect to t_1 gives a 2D

NMR spectrum as a function of two frequencies $S(\nu_1, \nu_2)$. A two-dimensional *Fourier* transform is given by

$$S(\nu_1, \nu_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(t_1, t_2)\exp(-i2\pi\nu_1 t_1)exp(-i2\pi\nu_2 t_2)dt_1 dt_2. \quad (9)$$

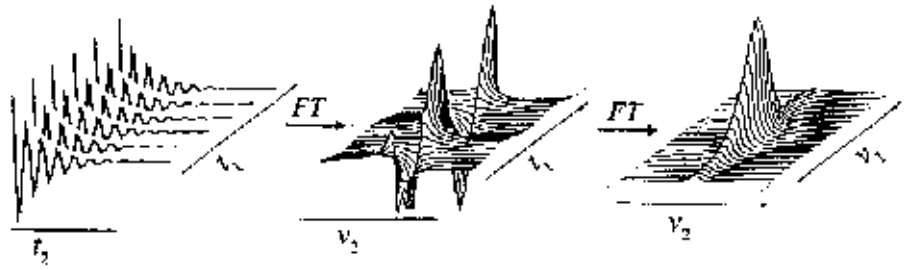


Figure 12. Stages involved in the Fourier transformation of a two-dimensional array of NMR data (Field, 1989).

Not all NMR experiments can be described with the simple vector model. The product operator formalism combines quantum-mechanical rigour with a formalism that is possible to visualise by vectors and rotations (Sorensen *et al*., 1983; van de Ven and Hilbers, 1983; Howarth *et al*., 1986). During an NMR experiment, different types of processes such as chemical shift evolution, scalar coupling evolution and r.f. pulses affect the spin states of the sample. Transformations of the initial spin states into new spin states can be described by a set of transformation rules of product operator formalism, which are illustrated geometrically in Fig. 13.

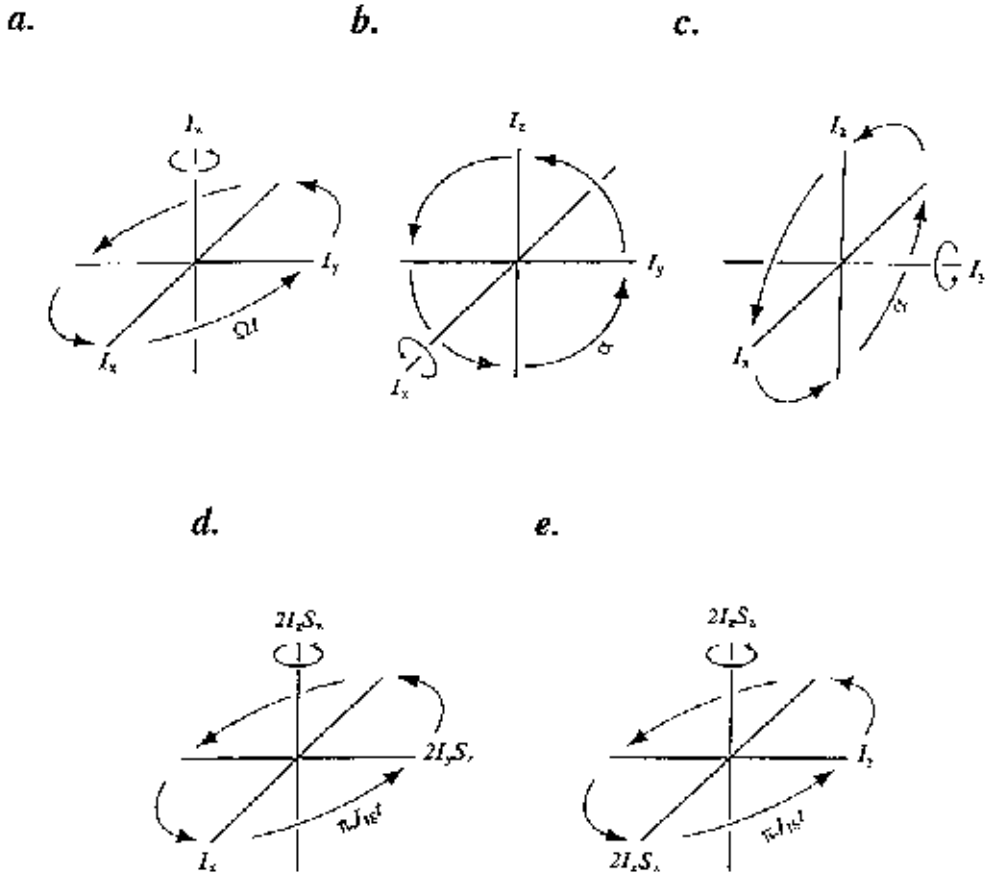


*Figure 13. Transformation of product operators. (**a**) The effects of chemical shift or z-pulse. (**b**, **c**) Rotations induced by x- and y-pulse. (**d**, **e**) Transformation under the scalar coupling evolution (Cavanagh et al., 1996).*

2D NMR experiments can be divided into the following three classes: (1) experiments that separate interactions into different dimensions (e.g. *J*-resolved NMR); (2) experiments that correlate signals that are scalar coupled (e.g. COSY, RELAY-COSY, TOCSY); (3) experiments that correlate spins that are connected through some dynamic process such as cross-relaxation or chemical exchange (e.g. NOESY). Four homonuclear 2D NMR experiments are described below. Together these ^{1}H NMR techniques, which were also applied in this work, form a basic set of experiments for the assignment and structure determination of proteins.

COSY (COrrelated SpectroscopY)

When two atoms are connected via chemical bond, the spin of one nucleus can sense the spin state of the other nucleus via the electrons of the bond. This is called scalar coupling, and it can be used to transfer magnetisation between two spins. COSY (Fig. 14) was the first of this type of 2D NMR experiment (Jeener, 1971; Aue *et al*., 1976). It consists of two $\pi/2$ pulses separated by an incrementable evolution period (t_1). The first pulse is the preparation pulse and the second pulse is the mixing pulse, which transfers coherence between *J*-coupled spins: for proteins, for example, between protons separated by two or three bonds. The FID is acquired during the detection time (t_2). A two-dimensional *Fourier* transform with respect to t_1 and t_2 gives the 2D spectrum, which is a map of all direct *J*-couplings in the molecule.

The COSY experiment is most readily described with the product operator formalism. For a *J* coupled two-spin system, *I* and *S*, with the *Larmor* frequencies Ω_I and Ω_S, the initial I_z magnetisation evolves through the COSY pulse sequence as follows:

$$I_z \xrightarrow{(\pi/2)_x\text{-}t_1\text{-}(\pi/2)_x} -I_z\cos(\Omega_I t_1)\cos(\pi J t_1) - 2I_xS_y\cos(\Omega_I t_1)\sin(\pi J t_1) + I_x\sin(\Omega_I t_1)\cos(\pi J t_1) - 2I_zS_y\sin(\Omega_I t_1)\sin(\pi J t_1). \quad (10)$$

Parallel evolution for S_z magnetisation is obtained by exchanging *I* and *S* labels. The first two terms do not lead to detectable magnetisation and can be ignored. However, the third term causes a diagonal peak and the fourth term a cross-peak modulated by Ω_I during t_1 and Ω_S during t_2. These two terms can also be written

Diagonal peak: $$\sin(\Omega_I t_1)\cos(\pi J t_1) = \tfrac{1}{2}[\sin(\Omega_I t_1 - \pi J t_1) + \sin(\Omega_I t_1 + \pi J t_1)] \quad (11)$$

Cross-peak: $$\sin(\Omega_I t_1)\sin(\pi J t_1) = \tfrac{1}{2}[\cos(\Omega_I t_1 - \pi J t_1) - \cos(\Omega_I t_1 + \pi J t_1)]. \quad (12)$$

The diagonal peak has an in-phase lineshape in f_1 dimension with the two multiplet components centred at Ω_I and separated by $2\pi J$. The cross-peak has an antiphase lineshape, but the two components of the peak are also centred at Ω_I and separated by $2\pi J$. The sinusoidal and cosinusoidal modulation means that the diagonal and the cross-peaks differ in phase by 90° and cannot be phased to adsorption simultaneously. If the evolution of the terms I_x and $2I_zS_y$ is considered during t_2, the results indicate that the lineshapes of the diagonal and cross-peaks in f_2 are the same as in the f_1 dimension.

RELAY-COSY (RELAYed coherence transfer COrrelated SpectroscopY)

This 2D NMR experiment is simply an extension of the COSY experiment (Fig. 14). Instead of acquisition of the NMR spectrum immediately after a coherence transfer from spin *I* to spin *S*, a delay period is placed in the pulse sequence prior to the acquisition. This allows antiphase magnetisation to develop for a second time. If spin *I* has other coupling partners than *S*, for example spin *T*, antiphase coherence develops between *S* and *T* during the delay period. Finally, a third $\pi/2$ pulse transfers coherence from *S* to *T*. Because chemical shifts are monitored only during t_1 and t_2, the cross-peak also appears in the spectrum between spins *I* and *T*, even when they are not directly coupled.

TOCSY (TOtal Correlated SpectroscopY)

TOCSY (Fig. 14), also known as HOHAHA (HOmo nuclear HArtmann-HAnn) spectroscopy, is a method for obtaining relayed connectivities (Braunschweiler and Ernst, 1983; Bax and Davis, 1985; Davis and Bax, 1985; Cavanagh and Rance, 1990). In TOCSY experiments, isotropic mixing is used to transfer in-phase magnetisation between several scalar coupled spins. During the mixing period, a strong spin-lock field is applied to remove the chemical shift differences of the spins. The cross-peaks are generated between all resonances within the same spin system.

NOESY (Nuclear Overhauser Effect SpectroscopY)

The experiments described so far are based on the coherence transfer via scalar couplings, and cross-peaks are formed only between protons belonging to the same spin system. But nuclear spins can also interact with each other via the dipolar coupling. In a NOESY (Fig. 14) experiment (Jeener *et al*., 1979; Anil Kumar *et al*., 1980; Macura and Ernst, 1980), a $\pi/2$-t_1-$\pi/2$ period frequency-labels the spins and returns the magnetisation back along the *z*-axis. During the mixing period (τ_m) the magnetisation is transferred via the dipolar coupling. The longer the mixing period, the more magnetisation can be transferred to the neighbouring spins. The cross-relaxation rate constant controls how quickly an NOE is transferred between *I* and *S*. For two isolated spins in a rigid molecule, the initial cross-relaxation rate (σ_{IS}) is proportional to the inverse sixth power of the internuclear distance (r). The detectable transverse magnetisation is created by the final $\pi/2$ pulse.

$$\sigma_{IS} \sim (1/r^6) \sim (V_{\text{cross-peak}}). \qquad (13)$$

Thus the cross-peaks in a NOESY spectrum result from the cross-relaxation of the longitudinal (I_z) magnetisation. The intensities of the peaks ($V_{\text{cross-peak}}$) depend on the cross-relaxation rate and are related to the distance between the correlated spins.

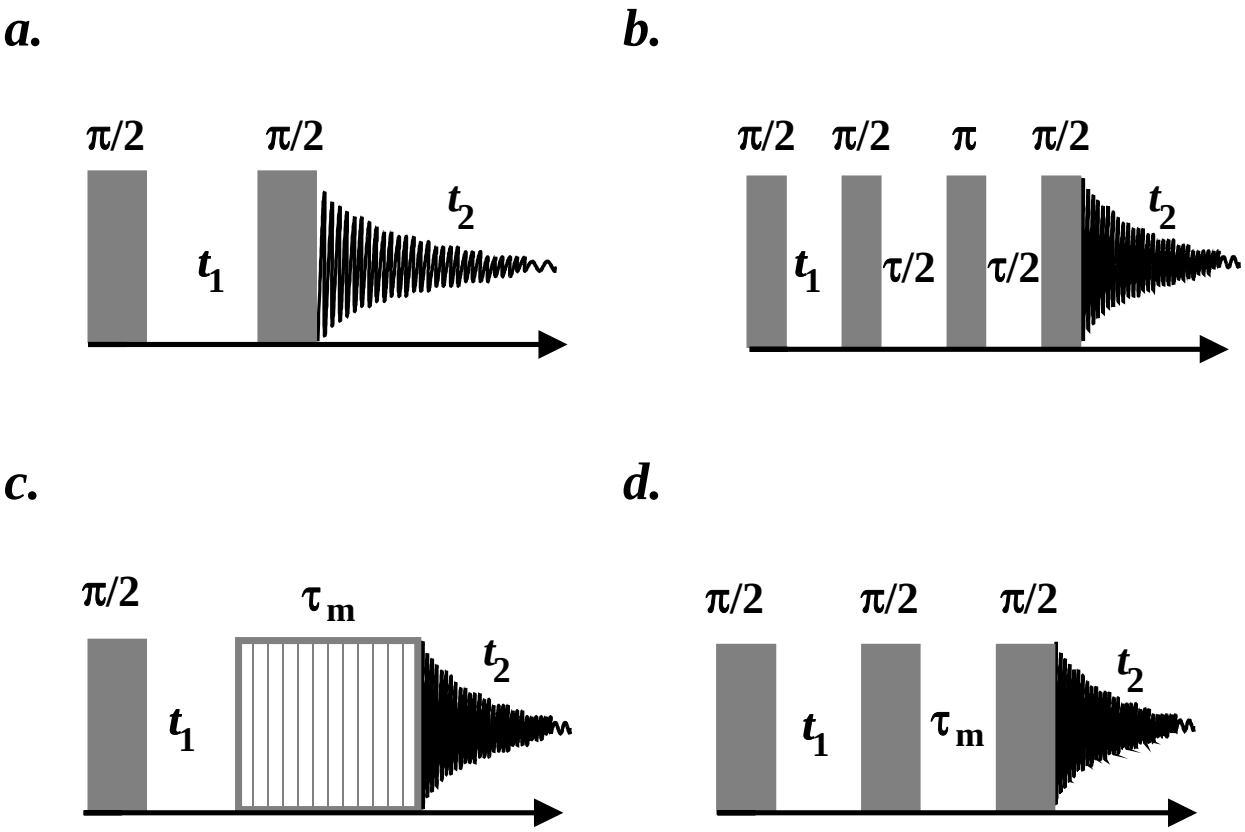


*Figure 14. The pulse sequences for (**a**) COSY, (**b**) RELAY-COSY, (**c**) TOCSY and (**d**) NOESY experiments (Cavanagh et al., 1996).*

NMR spectroscopy can be used to study protein–ligand interactions, especially to examine flexible protein surfaces through which proteins contact with the ligands (for example, polysaccharides). Changes in relaxation rates ($1/T_1$ or $1/T_2$) or line broadening effects of resonances are usually prerequisites for the observation of transferred NOEs (TRNOEs) between ligand and protein (Ni, 1994). Exchanged lineshapes are usually complicated functions of relative concentrations and kinetics of the interacting species. In the equilibrium a thermodynamic dissociation constant (K_d) is given by

$$L + E \leftrightarrow LE; \qquad K_d = k_{off}/k_{on} = [L][E]/[LE], \tag{14}$$

in which *LE* is the reversible complex of the ligand (*L*) and protein (*E*), and k_{on} and k_{off} are the specific rate constants for the association and dissociation of the ligand. In the presence of large excess of the ligand ($[L] \sim [L]_0$), K_d can be determined from any of these observable NMR parameters as a function of the total concentrations of the protein $[E]_0$ and ligand $[L]_0$. For example, for the line broadening effect

$$[E]_0 T = s(K_d + [L]) \sim s(K_d + [L]_0) \qquad T = 1/\pi\Delta\nu, \qquad (15)$$

$$[E]_0/\pi\Delta\nu \sim s(K_d + [L]_0), \qquad (\Delta\nu = \Delta\nu_{1/2,obs} - \Delta\nu_{1/2,free})$$

where *s* is a scaling constant depending on the experiment used (i.e. linewidth ($\Delta\nu_{1/2}$); relaxation time (T_1 or T_2)) and the nature of the chemical exchange for the binding interaction. K_d can be obtained from the horizontal intercept of the linear titration plot of equation 15 as a function of the concentration of the ligand.

For a slow-exchange process, the exchange rates are slow relative to the chemical shift difference ($\Delta\delta_i = \delta_{i,free} - \delta_{i,bound}$) for all ligand resonances. Thus the free and bound protons preserve their individual frequencies, but the resonances are broadened by amounts corresponding to the exchange rates. For a fast-exchange process, the exchange rates are much higher than $\Delta\delta_i$. Thus the resonances of the free and the bound ligands collapse into single peaks with average frequencies. Exchange conditions may vary for the different protons in the ligand.

The 2D transferred NOE experiment is used to study the conformation of a low molecular weight ligand bound to a fairly large protein (Albrand *et al*, 1979; Clore and Gronenborn, 1982, 1983). In this experiment, sizeable transferred TRNOEs are observed as long as ligands can associate with the protein and dissociate from the complex at least a few times during the NOE mixing. When the exchange between bound and free forms of ligand is fast, the binding of the ligand is relatively weak. This leads to a single set of ligand resonances with averaging occurring between bound and free values of the chemical shift and NOEs. Usually the protein is present in substoichiometric amount and its proton signals are broad (Ni, 1994).

3.3 Structure determination

The following section shortly summarises the sample preparation, the assignment of proton resonances in the 2D ^{1}H NMR spectra, the distance and dihedral angle retraints and the final structure determination.

3.3.1 Sample preparation

Even sophisticated NMR experiments cannot compensate for the spectral problems caused by an ill-behaved or ill-prepared sample. For structure determination studies the protein should be in a native, functional conformation, which means that the pH and composition of the solvent should be close to the physiological conditions of the molecule. Only then is the structure functionally relevant. NMR experiments should be performed under sample conditions that yield optimal spectra: that is, maximal resonance dispersion and minimal linewidths. The protein must be stable during the collection of the spectra, usually for at least one to two weeks; for it is advantageous to be able to use the same protein sample for all NMR experiments. The sample must also be sufficiently soluble at high concentrations. For example, as for the majority of NMR spectrometers, 0.3–0.6 ml of 1 mM protein solution in a 5-mm NMR tube is suitable for acquiring spectra in a reasonable time with satisfactory signal-to-noise ratios. Samples must also be free from contaminants originating for example from the NMR tube or protein preparation.

3.3.2 Assignment of proton resonances

Before the structure calculations can proceed, the 2D NMR spectra have to be assigned. The complete assignment of all ^{1}H resonances of a protein is the most time-consuming part of the structure determination. Often several NMR spectra have to be collected at different temperatures to allow resolution of overlapping resonances. Because the assignment step forms the basis for all further work, it is essential that it be performed carefully. Any errors in the assignment cause violations in the final 3D structure. In the initial stage of the structure determination, each resonance must be associated with a particular nucleus in the protein. Three types of information are available for the assignments: chemical environment (via chemical shifts), through-bond interactions (via scalar couplings) and through-space interactions (via dipolar couplings).

Tabulated random coil values of the chemical shifts of different amino acids facilitate the assignment (Bundi and Wüthrich, 1979; Wüthrich, 1986; Merutka *et al*., 1995; Wishart *et al*., 1995). For an example, Fig. 15 shows chemical shift ranges for the various types of ^{1}H resonances of protons.

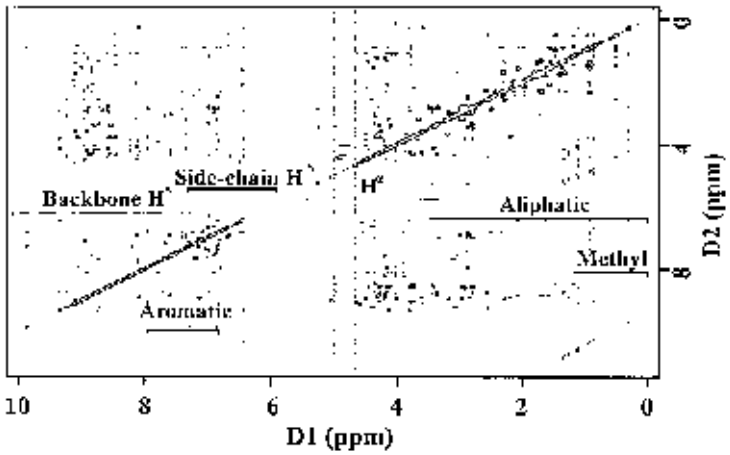


Figure 15. 2D NOESY spectrum of CBD_{EGI}.

Each spin system has a certain pattern of cross-peaks depending on the side-chain constitution. COSY, RELAY-COSY and TOCSY spectra are used to identify peaks belonging to the same spin system. First the picked NH-C$^{\alpha}$H cross-peaks are assigned from a COSY spectrum, and then the cross-peaks between NH and side-chain protons are identified from RELAY-COSY and TOCSY spectra acquired with several mixing times at different temperatures. Because of the weak ^{1}H-^{1}H scalar coupling over the peptide pond, magnetisation is not transferred between the backbone protons of the successive spin systems. Unless a severe overlap exists, the COSY, RELAY-COSY and TOCSY experiments are usually sufficient to assign all spin systems in a relatively short peptide (~ 40 residues). Once all or at least most of the spin systems have been identified, the sequential assignment process in an unlabelled protein is completed using the through-space connectivities (NOEs) between NH, C$^{\alpha}$H and C$^{\beta}$H protons that are close in space (< 5 Å) (Fig. 16). Sequential NOEs are also used to identify the secondary structure elements of protein (Wüthrich *et al*., 1982; Wüthrich, 1986; Dyson and Wright, 1991).

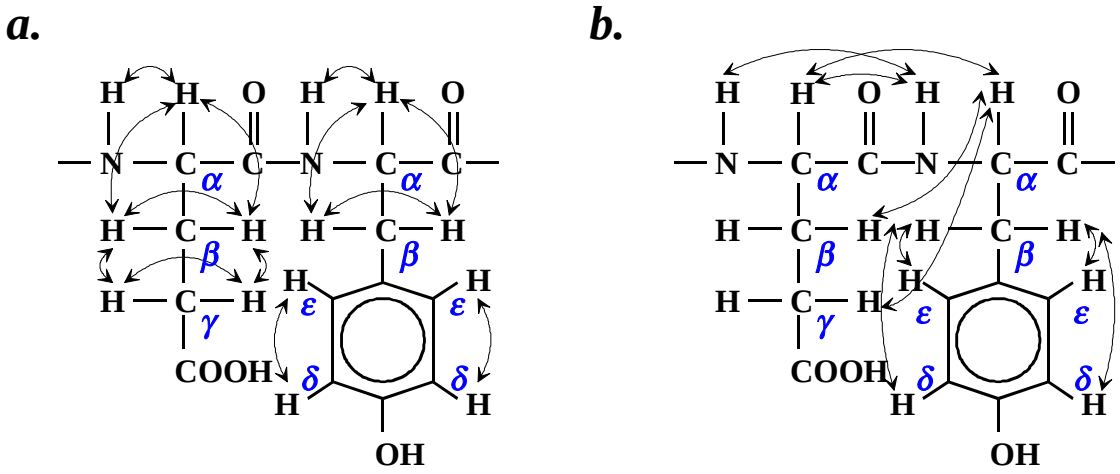


*Figure 16. Connectivities obtained from (**a**) COSY and (**b**) NOESY experiments. In the case of NOESY only additional connectivities are shown (Wagner, 1990).*

3.3.3 Distance and dihedral angle restraints

Scalar coupling interactions and dipolar cross relaxation rates are sensitive to molecular conformation. Quantification of these parameters permits structural analysis by NMR spectroscopy.

Distance restraints for the protein structure determination are readily generated from integrated NOE cross-peak volumes by the simplified model called isolated spin-pair approximation (ISPA) (Neuhaus and Williamson, 1989). The requirement for the model is that the two interacting spins are separated by less than 5 Å (or so) in space. When several NOESY spectra with different mixing times (for example, 30-200 ms) are acquired, intensity build-up curves are obtained for each peak as a function of mixing time. The initial cross-relaxation rate ($\sigma_i \sim V_i$, equation 13) can then be estimated from the initial slope of the NOE build-up curve. Thus, if one interproton distance (r_{ref}) is known from the covalent geometry (for example, distances between geminal methylene protons or vicinal protons in aromatic rings), the unknown interproton distance (r_{ij}) can be determined from the relationship

$$r_{ij} \sim r_{ref}(\sigma_{ref}/\sigma_{ij})^{1/6} \sim r_{ref}(V_{ref}/V_{ij})^{1/6} \qquad (16)$$

in which V_{ref} and V_{ij} are the integrated cross-peak intensities. In practice, for proteins, the ISPA is valid only for short mixing times, where the dependence of cross-peak intensities on the mixing time is close to linear (Barsukov and Lian,

1993). For longer mixing times, magnetisation may be transferred between interacting spins in multiple steps via spin diffusion. In addition to this, ISPA also ignores anisotropic tumbling of the molecule. Thus ^{1}H-^{1}H distances cannot be determined very precisely from the NOE cross-peak intensities. To account for imprecision, distances are treated as distance ranges rather than precise distances. The lower bounds are usually determined from the sum of the van der Waals radii (1.8 Å) and the upper bounds are determined from the NOE intensities. NOE cross-peaks can also be grouped on the basis of their intensities. Usually three different categories are distinguished: strong (< 2.5 Å), medium (< 3.5 Å) and weak (< 5.0 Å) (Wütrich, 1986). NOEs that originate from groups of protons with degenerate chemical shifts, for example methyl groups, are commonly referred to pseudoatoms located at the geometrical centre of the protons. The upper bounds of the pseudoatoms are increased by a correction factor (Wüthrich *et al.*, 1983). In practice, the number of NOEs is more critical for the quality of the final structure than is the precision of the upper and lower bounds.

Additional structure information is obtained from the vicinal (over three covalent bonds) scalar coupling constants (^{3}J). The scalar coupling depends on the dihedral angle (ϕ). The minimum separation that can be measured between the antiphase components of a COSY cross-peak equals to approximately 0.58 times the linewidth at half height (Wütrich, 1986). After $^{3}J_{\mathrm{NH}\alpha}$ coupling constants have been measured (Marion and Wüthrich, 1983), dihedral angle restraints for the corresponding ϕ-torsion angles (NH-N-C$^{\alpha}$-C$^{\alpha}$H) are obtained using the *Karplus* equation (Karplus, 1963)

$$^{3}J = A\cos^{2}\phi + B\cos\phi + C, \qquad (17)$$

in which the constants A, B and C depend on the type of nuclei involved in the covalent bonds. Theoretical attempts to determine these constants in such a way that the chemical properties of the atoms involved in the coupling are taken in to account have led to complex relationships that are not generally applicable. Instead, semiempirical methods of obtaining these values have proved successful. For example, values A = 6.4, B = -1.4 and C = 1.9 can be used to determine ϕ-torsion angles from $^{3}J_{\mathrm{NH}\alpha}$ coupling constants (Pardi *et al.*, 1984).

As can be seen from the *Karplus* curve (Fig. 17) as many as four different conformations can yield the same value of $^{3}J_{\mathrm{NH}\alpha}$ (Wüthrich, 1986). However, ϕ is usually restricted in the region of $-90° < \phi < -40°$ when $^{3}J_{\mathrm{HN}\alpha} < 6.0$ Hz, and in the region of $-160° < \phi < -80°$ when $^{3}J_{\mathrm{HN}\alpha} > 8.0$ Hz (Case *et al.*, 1994). For β-strands, $^{3}J_{\mathrm{NH}\alpha}$ is about 9 Hz and for helices it is about 4 Hz. Sometimes

intramolecular motions give rise to values for $^3J_{HN\alpha}$ that are averaged over distributions of dihedral angles. This means that useful dihedral restraints cannot be derived from accurately measured coupling constants. For example, in short and unstructured linear peptides, extensive backbone motion results in a coupling constant of about 7 Hz. Thus $^3J_{NH\alpha}$ coupling constants 6.0–8.0 Hz are usually not used in protein structure determinations because they may reflect motional averaging of multiple conformations (Cavanagh *et al*., 1996).

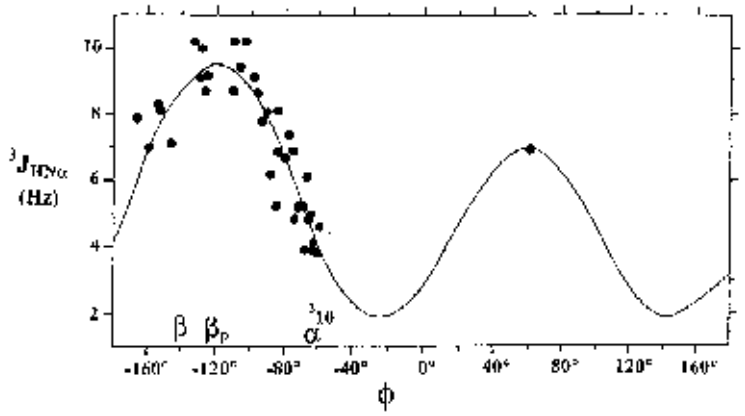


Figure 17. The Karplus curve describing the variation of $^3J_{NH\alpha}$ as a function of backbone dihedral angle (ϕ). For all amino acids in proteins except glycine, the ϕ-angle is concentrated in the range -30° to -180° (Richardson, 1981). Regions of ϕ angles for different secondary structure elements are also shown (Wüthrich, 1986).

Stereospecific assignments of β-methylene protons, γ-methyl protons of valines and δ-methyl protons of leucines are important for the protein structure determination. $^3J_{\alpha\beta}$ coupling constants can be measured from the COSY spectrum and they are closely related to the χ_1 side-chain torsion angles ($C^{\alpha}H$-C^{α}-C^{β}-$C^{\beta}H$). The use of intraresidue NOEs together with $^3J_{\alpha\beta}$ coupling constants makes it possible to distinguish different side-chain conformations (Wagner, 1990) as shown in Fig. 18.

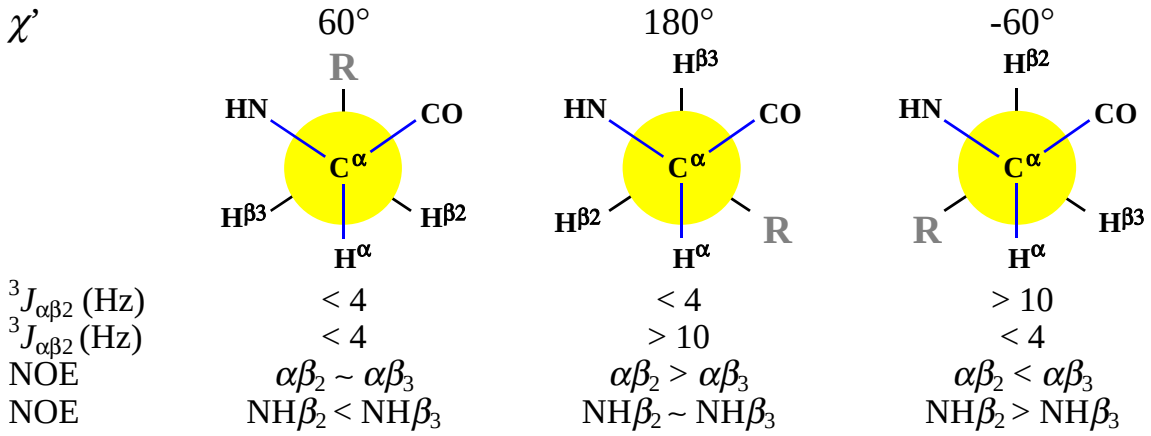


Figure 18. Stereospecific identification of β-methylene protons from $^3J_{\alpha\beta}$ coupling constants and intraresidue NOEs (Wagner, 1990).

3.3.4 Structure calculations

The structure calculations of a protein can be started when the majority of the distance restraints and a considerably smaller number of dihedral angle restraints have been determined. Bond lengths, bond angles and other standard elements of covalent geometry are obtained from the covalent structure of protein. The data is not sufficient to uniquely define the 3D structure of a protein, because the constraints are included as ranges of allowed values, the NOEs are limited to distances less than approximately 5 Å, the data contain experimental uncertainties, and only a small subset of all possible distance and dihedral constraints are observable. In practice, several rounds of structures are calculated. The first structures are calculated with a subset of well-defined NOEs. The uncertain or unknown NOEs, which are identified between the successive structure calculations, are included later in the calculations. Often, further stereospecific assignments can be made during the refinement process. This procedure reduces the violations and restrains the final set of low-energy structures more tightly to a better converged ensemble.

Several methods are available for calculating protein structures (Crippen, 1977; Havel *et al.*, 1983; Wagner *et al.*, 1987; Havel, 1990, 1991). The two most common approaches for generation of structures are restrained molecular dynamics (rMD) and distance geometry (DG). The computationally efficient implementations of the rMD method, which use a simplified force field (i.e. the electrostatic and attractive van der Waals terms are ignored), are referred to as dynamical simulated annealing (SA). Many laboratories employ a hybrid

method (Nilges *et al*., 1988a–c), in which initial structures are generated by DG (Fig. 19) and refined (annealed) by rMD. These protocols aim to determine co-ordinates for the atoms that satisfy the experimental and theoretical input in an unbiased way, while exploring all regions of conformational space compatible with the observed NMR restraints.

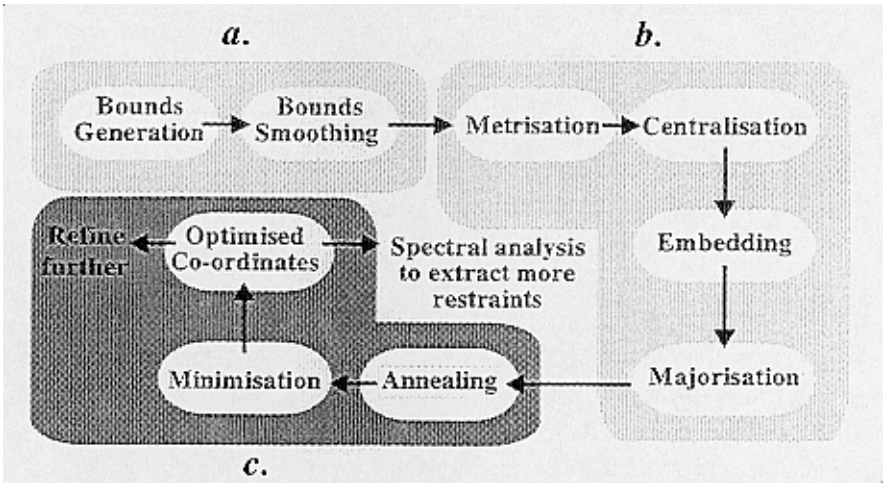


*Figure 19. Various steps in DG structure generation. (**a**) Bounds generation: DG database is created from the incomplete and imprecise set of experimental restraints; Bound smoothing: A complete and more precise set of bounds, i.e. Euclidean limits, are calculated using the triangle or tetrangle inequalities. (**b**) Metrisation: A matrix of random (i.e. trial) distances obeying the triangle inequality are chosen; Centralisation: Central metric matrix is generated; Embedding: The random distance matrix is converted to Cartesian co-ordinates; Majorisation: Distances measured from the embedded co-ordinates are least-squares fitted to the 'trial distances' in the metrisation. (**c**) Annealing: SA refinement; Minimisation: Conjugate gradients minimisation (NMRchitect, 1995).*

Distance geometry is a powerful technique for the determination of global folds of proteins (Blumenthal, 1970; Havel and Wüthrich, 1984, 1985; Kuntz *et al*., 1989; Havel, 1991). Because this method relies solely on the distances, the resulting structures have the correct global fold but poor local geometry. Even new and efficient DG algorithms which sample large regions of conformational space, the structures may end up in local energy minima (Kuszewski *et al*., 1992). Accordingly, structures are first calculated, for example, with distance geometry and then refined by the rMD or SA (Scheek *et al*., 1989; Nilges *et al*., 1988b). The steps of simulated annealing are shown in Fig. 20.

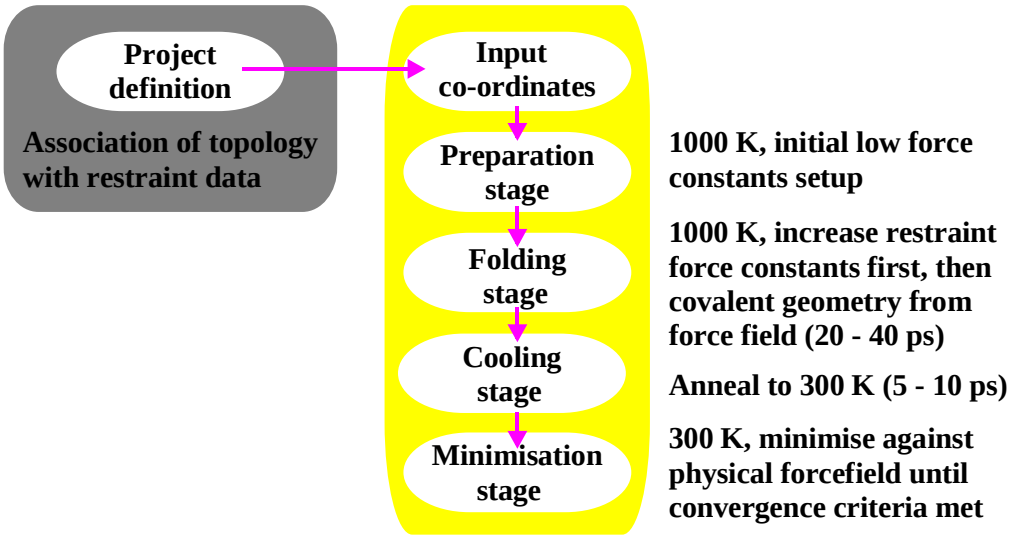


Figure 20. Simulated annealing methodology. Project definition: The initial and final scaling factors are defined independently for dihedral, distance and chiral restraints. For NOE and ^{3}J derived distances the force constants are typically on the order of 10–50 kcal/molÅ^{2}; Input co-ordinates: Co-ordinates of initial DG structures are loaded; Preparation stage: Internal force constants are adjusted to values such that the potential energy of the system equals the kinetic energy at 1000 K. The force constants are typically on the order of 0.1 in the energy units for the bonds, NOE restraints, angles, torsions and improper terms; Folding stage: This stage is carried out at 1000 K, during which the internal and NOE restraint force constants are increased until they reach their full value. During the later part of the dynamics phase the nonbond force constant is increased to regularise simultaneously the geometry and nonbond interactions; Cooling stage: The molecule is cooled to room temperature without trapping the molecule in a local minimum conformation; Minimisation stage: The annealed conformations are minimised with respect to the restraints and physical forcefield (NMRchitect, 1995).

The precision of the structure directly depends on the number of experimental restraints. Low resolution structures typically have as few as five restraints per residue, whereas the most precise structures obtained from ^{1}H data may have up to 15 . In the latter case, the mean rms deviation of backbone atoms about the mean structure may be as low as 0.5 Å for well-defined regions of structure. Of course, if it is possible to use ^{15}N or ^{13}C labelling, more ^{1}H-^{1}H NOEs can be unambiguously identified and the number of restraints per residue is then between 20 and 25. In that case the rms deviation for the backbone atoms may be as low as 0.3–0.5 Å. The number of stereospecifically assigned prochiralgroups also affects the precision of structures (Güntert *et al.*, 1989).

4 Results and discussion

4.1 Choice of CBDs

Efficient hydrolysis of cellulose requires tight binding of cellulases to cellulose mediated by CBD. The role of CBD during the hydrolysis is not clear on the atomic level: CBD might anchor the whole enzyme to the cellulose surface or it might actively facilitate the breakdown of cellulose chains (Teeri *et al.*, 1992). The 3D structure of CBD_{CBHI} from *T. reesei* (Kraulis *et al.*, 1989) is wedge-shaped with a flat and a rough face. As a means of determining the key residues for the hydrolysis, all amino acids on the flat face of the wedge (Tyr-5, Asn-29, Tyr-31, Tyr-32 and Gln-34) were replaced one by one by functionally neutral alanines. Because there are no obvious interaction sites for cellulose on the rough face of the wedge, only one residue, Pro-16, was replaced, by a bulky, positively charged arginine, to break this surface. The mutation sites are shown in Fig. 21.

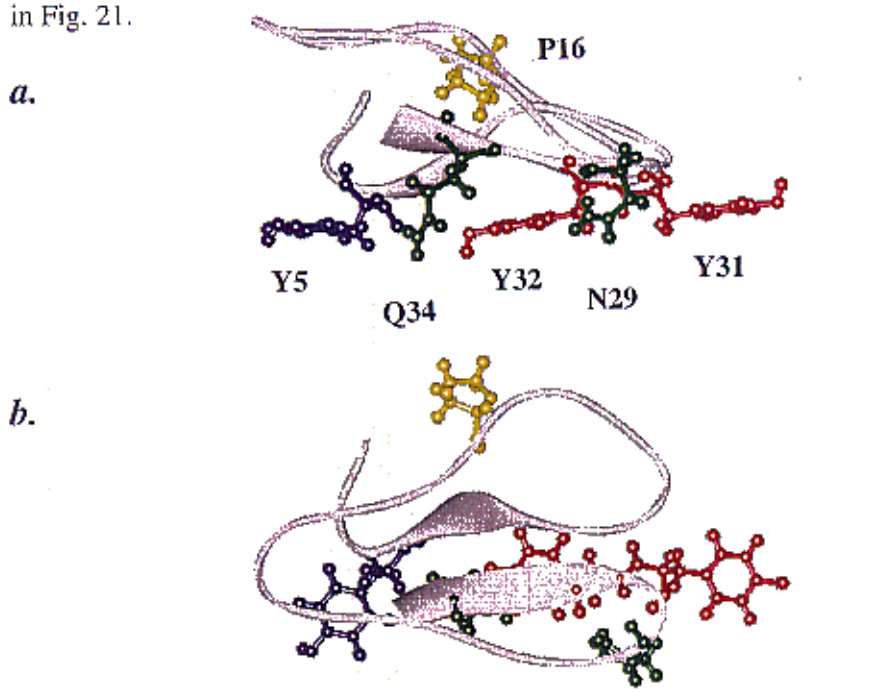


*Figure 21. C^{α}-trace of the intact CBD_{CBHI} with the side-chains (Tyr-5, Pro-16, Asn-29, Tyr-31, Tyr-32 and Gln-34) studied in this work (Kraulis et al., 1989). Structure (**a**) is a view from the side and structure (**b**) is from the top of the wedge, looking down towards the flat face.*

The binding of the engineered peptides to crystalline cellulose was determined by measuring the equilibrium concentration after adsorption at different starting concentrations of the peptide. Effects of the amino acid substitutions were greatest for the peptides Y5A and Y32A, which totally lost their affinity to cellulose (Fig. 22).

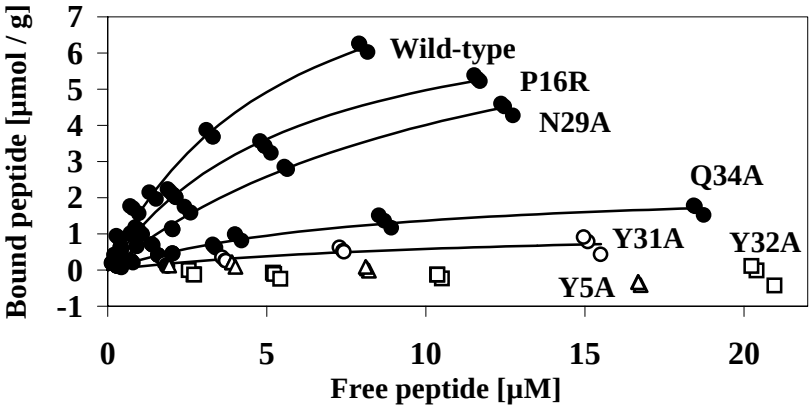


Figure 22. Adsorption isotherms of the wild-type CBD_{CBHI} and six engineered peptides (Y5A, P16R, N29A, Y31A, Y32A and Q34A).

The functional changes occurring when amino acid substitutions are introduced into proteins may be caused either by the altered side-chain functionality or by a structural change induced by the mutation. Thus, it was necessary to assess the structural effects of the substitutions by 2D NMR spectroscopy. Because it would be a rather tedious task to determine 3D structures for all six engineered peptides, the structural effects of the mutations were first studied for all the peptides together by comparing chemical shifts, coupling constants and NOEs of the backbone protons for the wild-type CBD_{CBHI}. As a means of examining the structural changes on the flat faces of the mutants in detail, the high resolution 3D structures were determined for Y5A, Y31A and Y32A by 2D ^{1}H NMR spectroscopy.

T. reesei secretes a cellulolytic system consisting of two cellobiohydrolases and at least four endoglucanases. EGIII lacks CBD, but the CBDs of the other enzymes show a marked similarity in sequence (Shoemaker *et al.,* 1983; Teeri *et al.*, 1987; Penttilä *et al.*, 1986; Saloheimo *et al.*, 1988, 1994), as is clear from Table 2. The residues on the flat face are identical except for those at positions 5 and 31, which can be either tyrosine or tryptophane.

The rough face exhibits greater variation in the amino acid composition. The most interesting residues on this face are in the positions 16 and 30, which in CBDs of the four main cellulases (CBHI, CBHII, EGI and EGII) may be either proline or cysteine for position 16 and proline or aspartic acid for position 30. In CBD_{CBHI}, Pro-16 is situated in the middle of the rough face and Pro-30 is at the tip of the wedge. Comparison of the sequences of CBD_{CBHI} and CBD_{EGI} shows the most interesting differences to lie in the positions 5, 16 and 30. In CBD_{EGI}, position 5 is occupied by tryptophane and 30 by aspartic acid. Cysteine in position 16 is participating in a third disulphide bridge.

To find out whether these differences could explain the different roles of the cellobiohydrolases and endoglucanases during the hydrolysis of cellulose, I determined the 3D structure of CBD_{EGI} by ^{1}H 2D NMR spectroscopy. Cellobiohydrolases degrade cellulose from the chain ends and are active on the crystalline parts of cellulose, whereas endoglucanases are active on the interior sites and amorphous regions of cellulose.

Table 2. Sequence alignment of CBDs from T. reesei cellulases. Grey boxes indicate amino acids on the flat face of CBD_{CBHI} and CBD_{EGI}. White boxes show residues that differ in the sequences of CBD_{CBHI} and CBD_{EGI} and are situated mainly on the rough face. Numbering in the table corresponds to the residues of CBD_{CBHI}.

Source			**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
CBHI			T	Q	S	H	**Y**	G	Q	C	G	G	I	G
EGI	S	C	T	Q	T	H	**W**	G	Q	C	G	G	I	G
CBHII	Q	A	C	S	S	V	**W**	G	Q	C	G	G	Q	N
EGII			Q	Q	T	V	**W**	G	Q	C	G	G	I	G
EGV			Q	Q	T	L	**Y**	G	Q	C	G	G	A	G
			13	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
CBHI			Y	S	G	**P**	**T**	**V**	C	**A**	S	G	T	T
EGI			Y	S	G	**C**	**K**	**T**	C	**T**	S	G	T	T
CBHII			W	S	G	**P**	**T**	**C**	C	**A**	S	G	S	T
EGII			W	S	G	**P**	**T**	**N**	C	**A**	P	G	S	A
EGV			W	T	G	**P**	**T**	**T**	C	**Q**	A	P	G	T
			25	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
CBHI			C	Q	**V**	**L**	**N**	**P**	**Y**	**Y**	S	**Q**	C	L
EGI			C	Q	**Y**	**S**	**N**	**D**	**Y**	**Y**	S	**Q**	C	L
CBHII			C	V	**Y**	**S**	**N**	**D**	**Y**	**Y**	S	**Q**	C	L
EGII			C	S	**T**	**L**	**N**	**P**	**Y**	**Y**	A	**Q**	C	I
EGV			C	K	**V**	**Q**	**N**	**Q**	**W**	**Y**	S	**Q**	C	L

4.2 Structural analysis of CBDs

4.2.1 Secondary structures of the mutants

(Publication I)

Y5A, Y31A and Y32A

The chemical shifts of Y5A, Y31A and Y32A differed relatively little from the shifts of the wild-type. This is not necessarily an indication of merely small structural changes, however, because in a small domain such as this CBD, the majority of ^{1}H shifts deviate little from the random coil shifts. The difference with the wild-type was greatest for the $^{3}J_{HN\alpha}$ coupling constants of Y5A. In particular, the coupling constants of residues 3–6 and 10–13 were no longer characteristic for type II turn, indicating structural changes in the N-terminus and in the loop between the *β1*- and *β2*-strands. The coupling constants of Y31A and Y32A differed on only a few counts with those of the wild-type. Many of the resonances in the NOESY spectra of Y5A were clustered, and the number of cross-peaks was small. Some new short-range NOEs were found between protons in the N-terminus, and no new long-range NOEs were observed. The NOESY spectra of Y31A and Y32A were very similar to those of the wild-type; only some new long-range NOE were observed between the *β1*- and *β3*-strands, indicating local structural changes.

N29A and Q34A

For N29A the changes in chemical shifts were largest near the mutated residue and the greatest differences in $^{3}J_{HN\alpha}$ coupling constants were observed in the N-terminus. Also, several new NOEs were found for N29A. All of these corresponded to short distances in the wild-type structure as well, except the NOE between Thr-24 and Cys-19 for which the distance in the wild-type was longer. Possibly the loop containing Thr-24 and Cys-19 in N29A may be somewhat flattened. The presence of additional cross-peaks that could not be unambiguously assigned suggested that the Ala-29 to Pro-30 peptide bond might adopt the *cis* conformation leading to increased flexibility in the type I turn. In the case of Q34A the NMR data indicated only small conformational perturbations.

P16R

The changes in the chemical shifts near the N-terminus were greater for P16R than the other peptides. Also, differences between the $^3J_{HN\alpha}$ coupling constants of this and the wild-type peptide were large near the N-terminus. From the observed new NOEs, the new cross-peaks of Gly-9 indicated some change on the rough face. Some new NOEs to the side-chain of arginine were also identified. The β- and γ-protons showed NOEs to NH of Thr-17 and Val-18, but no NOEs were observed to the protons of other side-chains. Evidently, then, the side-chain of arginine points outward from the molecule in the direction of the N-terminus.

4.2.2 Structures of Y5A, Y31A and Y32A

(Publications II and III)

Comparison of the Y5A, Y31A and Y32A structures with the structure of the wild-type revealed an unmistakable resemblance (Fig. 23). The backbones of Y31A and Y32A followed closely the fold of the wild-type, and the β-strands and turns were preserved. Furthermore, the entire backbones and most of the side-chains were well defined. Thus the mutations of Tyr-31 and Tyr-32 to alanine neither altered nor decreased the compactness of the CBD fold. This means that these residues were not part of the structural framework.

In contrast, the mutation of Tyr-5 to alanine had clear structural consequences. Although the segments corresponding to the β-strands and the turns remained rather well defined, the N-terminus and the spatially close segment Gln-15–Cys-25 were poorly defined. In fact, the definition of the structure for this part of the peptide was so poor that the positions of the N-terminus and the β2-strand remained ambiguous. In about half of the structures accepted for Y5A the N-terminus was above the loop Ser-14–Cys-19, whereas in the wild-type the N-terminus was consistently below the loop. This implies that the overall fold of the structure is compromised. Clearly, Tyr-5 had a key role in maintaining the structural integrity.

The structural consequences of the mutations to the flat face of the wedge are shown in Fig. 24. For clarity, only the C^α-traces and the side-chains of the residues on the flat face are drawn for all members of the families . The flat face of the wild-type is well preserved in Y31A: Ala-31 is only slightly moved towards Asn-29 compared with Tyr-31, and there are only minor other

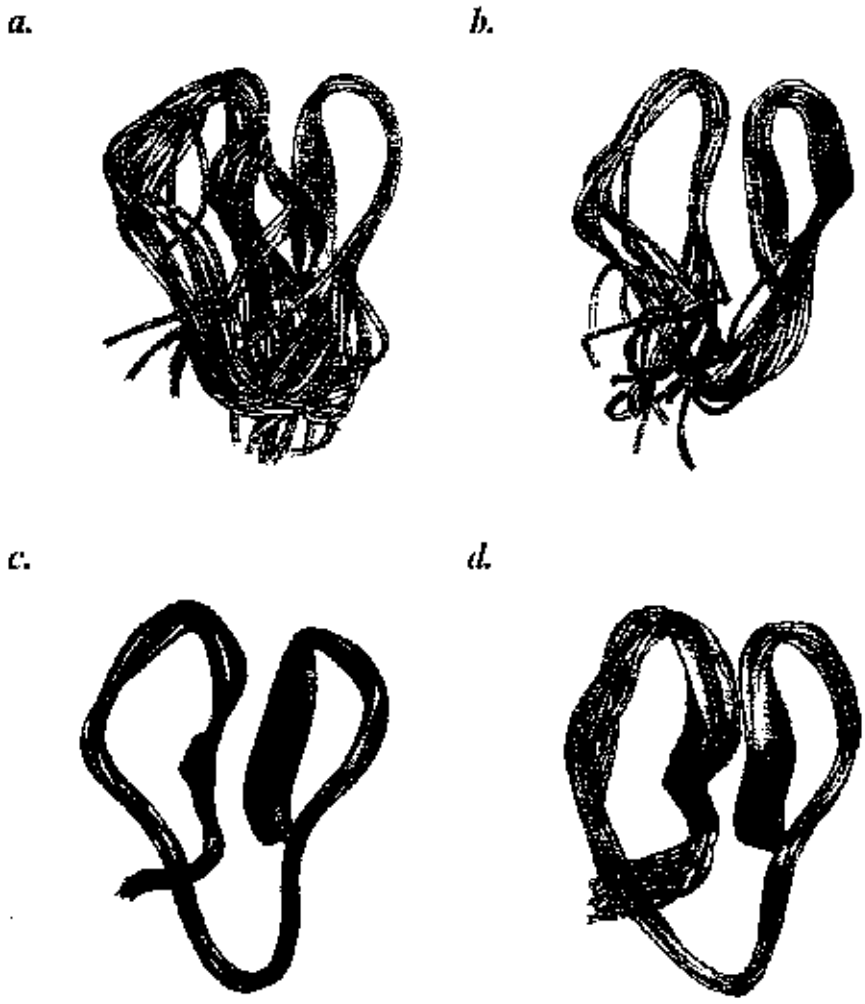


*Figure 23. C^{α}-traces of Y5A, Y31A and Y32A. (**a**) Twenty-three and (**b**) fourteen structures of Y5A belonging to the different structural families. (**c**) Eighteen and (**d**) nineteen structures of Y31A and Y32A, respectively. Average atomic rms deviations of the individual structures about the mean structure for the backbones are $Y5A^{a}$ (2.2 ± 0.81 Å), $Y5A^{b}$ (2.0 ± 0.83 Å), Y31A (0.53 ± 0.15 Å) and Y32A (0.81 ± 0.21 Å).*

rearrangements in the remaining residues on the flat face. This explains the small but detectable binding affinity of Y31A to cellulose. In the case of Y32A the flat face has become concave. Ala-32 is buried in the interior and the

residues neighbouring Tyr-32 in the wild-type structure have acquired more space. Thus the residues on the flat face of Y32A have neither the planarity nor the periodicity of those on the flat face of the wild-type. Clearly the reorganisation of the side-chains on the flat face caused a complete loss of affinity towards cellulose. The flat face is less precisely defined for Y5A than for Y31A and Y32A. When the N-terminus is below the plane of the loop Ser-14–Cys-19, Ala-5 is approximately in the plane of the flat face, but when the N-terminus is above, Ala-5 is far above. For both families of structures the remaining residues of the flat face have average positions similar to those in the wild-type. It is not surprising, therefore, that Y5A has no affinity towards cellulose.

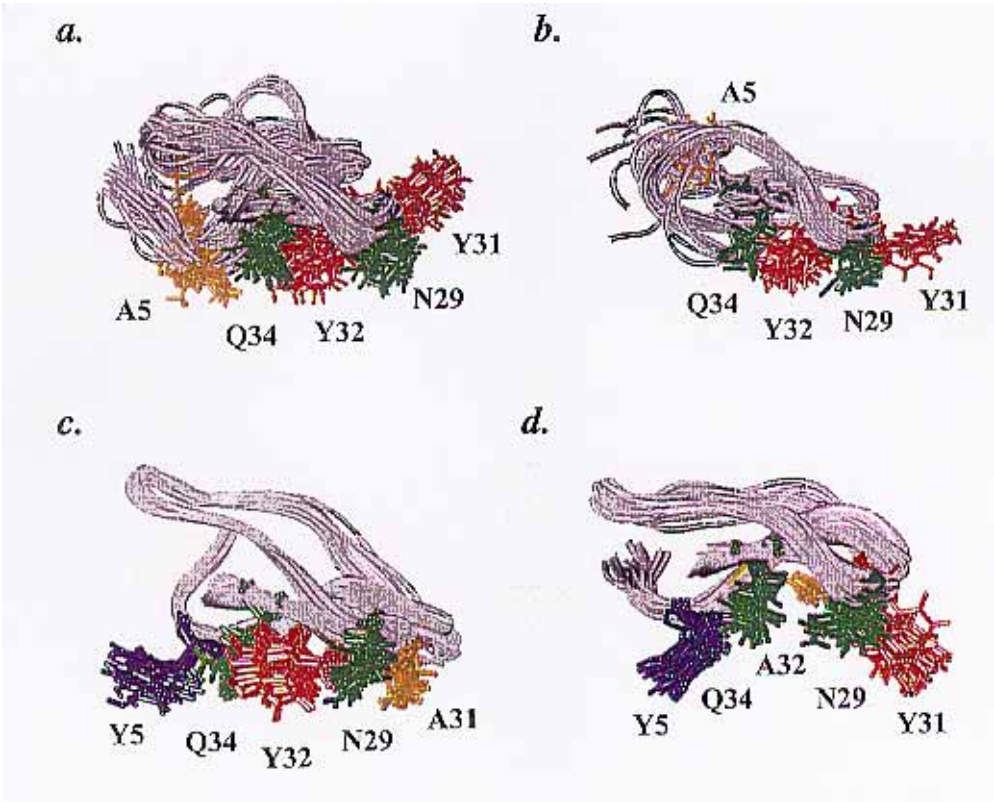


*Figure 24. Side-view of the flat face. (**a**) Twenty-three and (**b**) fourteen structures of Y5A, (**c**) eighteen and (**d**) nineteen structures of Y31A and Y32A respectively. Average atomic rms deviations of individual structures about the mean structure for all atoms are Y5A[a] (2.8 ± 0.73 Å), Y5A[b] (2.5 ± 0.76 Å), Y31A (1.0 ± 0.13 Å) and Y32A (1.3 ± 0.26 Å).*

4.2.3 Structure of CBD_{EGI}

(Publication IV)

The C^{α}-trace and disulphide bridges of C-terminal CBD_{EGI} are shown in Fig. 25. The overall fold of the peptide consists of three antiparallel β-strands. When the backbone atoms of CBD_{EGI} and CBD_{CBHI} were superimposed, the main chains matched very closely. In fact, the differences were within the precision of the structures. In the core of all CBDs, four cysteine residues form two disulphide bridges. CBD_{EGI} contains a third disulphide bridge near the N-terminus, but this did not seem to affect the backbone conformation of the N-terminus. When the backbone atoms of the model of CBD_{EGI} and the NMR structure of CBD_{EGI} were superimposed, the rms deviation between the backbones was of the same order of magnitude as for CBD_{CBHI} and CBD_{EGI}.

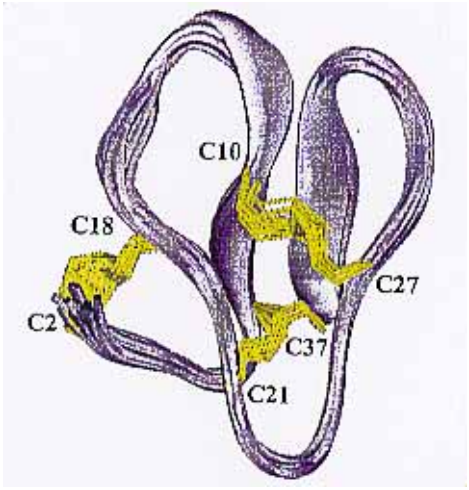


Figure 25. Nineteen structures of CBD_{EGI} with the three disulphide bridges. The backbone is well defined having an average rms deviation of 0.50 $\pm$ 0.22 Å about the mean structure.

From the side-view, CBD_{EGI} is wedge-shaped. For clarity, Fig. 26 depicts only some of the functionally important side-chains and those residues that differ in the CBD_{EGI} and CBD_{CBHI} sequences. Comparison of the overall structures of CBD_{EGI} and CBD_{CBHI} shows them to differ in shape only slightly. The rough face of CBD_{EGI}, where most of the amino acid differences between these two peptides are situated, is somewhat flatter than that of CBD_{CBHI}. It is impossible, therefore, to discern any significant structural differences between CBD_{EGI} and CBD_{CBHI}.

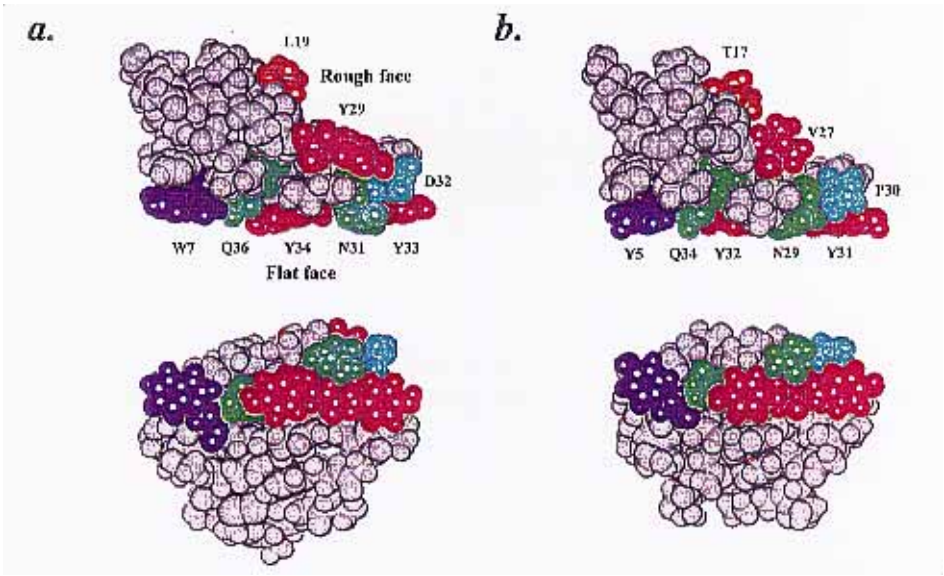


*Figure 26. Side-view and top-view of (**a**) CBD_{EGI} and (**b**) CBD_{CBHI}. Functionally important side-chains on the flat faces, and three residues on the rough faces which differ in sequences, are coloured. For all atoms of CBD_{EGI} the rms deviation about the mean structure is 0.83 $\pm$ 0.26 Å.*

If the difference in affinity between these two peptides is to be explained, at least some of the residues that differ in the sequences must be functionally important. The side-chains Ser-3/Thr-5, Val-18/Thr-20, Ala-20/Thr-22 and Lys-28/Ser-30 of CBD_{CBHI} and CBD_{EGI}, respectively, correspond to each other well. In the case of the substitution Thr-17 to Lys-19, the Lys-19 side-chain is considerably longer and occupies more space than Thr-17. Structural differences for the residue pairs were greater where one of the amino acids contained a ring in the side-chain, for example for tyrosines or prolines. The difference between Tyr-5 in CBD_{CBHI} and Trp-7 in CBD_{EGI} did not result in a significant difference in the structure of the binding face, because the other residues are common. It is known that the substitution Y5W in CBD_{CBHI} significantly increases the affinity to cellulose, although not to the level of the wild-type CBD_{EGI}. In contrast, the substitution P30D decreases the affinity to cellulose (Linder *et al*., 1995). In the 3D structures of CBD_{EGI} and CBD_{CBHI}, Asp-32 and Pro-30 correspondingly occupied closely similar positions. The differences between the model of CBD_{EGI} and the solution structure of CBD_{EGI} were, of course, very similar to those found between this peptide and CBD_{CBHI}, because the model of CBD_{EGI} was generated from the NMR structure of CBD_{CBHI} (Hoffrén *et al*., 1995).

4.3 Model for CBD binding to cellulose (Publications III, IV and V)

Although the 3D structures of several CBDs are known, studies on the interaction with cellulose are complicated by the unknown structure of cellulose. In addition to the crystalline regions of cellulose, there are crystal defects, dislocations, chain ends, twists etc., as well as amorphous regions, all of which may be initial target sites for cellulases (Hon, 1994). So far it has not been possible by NMR to observe CBD binding to the solid cellulose surface directly. Accordingly, I used long, soluble cello-oligosaccharides, cellopentaose and cellohexaose, as model substrates for glucose chains, even though cello-oligosaccharides, being flexible, cannot provide a distinct rigid surface where CBD could attach.

In 1D ^{1}H NMR experiments, resonances of the cellohexaose broadened when CBD_{CBHI} was added to the solution. Similar line broadenings were observed for CBD_{EGI} and CBD_{CBHII}. The experiments were reproducible with CBD_{EGI} in cellopentaose solution, but the effects were less pronounced. CBD binding to the long cello-oligosaccharides must have been specific because no broadenings were observed for Y31A and the reference compounds, i.e. pure tyrosine and tryptophan. The effects were most significant for the a1β-protons in the reducing end of the cello-oligosaccharide and for the internal and terminal 1 protons. The broadening of lines was also observed in the spectral region of the 3, 4 and 5 protons. For protons 2, 6 and 6' the increase in the linewidth was less than 3 Hz. No line broadenings were observed for a1α-protons in the reducing end of the sugar. From the cellohexaose titration experiments the equilibrium binding constant (K_d) was estimated to be on the order of 350 $\pm$ 90 μM for all studied CBDs at 5°C. T_2 relaxation measurements confirmed the linewidth measurements. When the intact CBHII enzyme was added to the mixtures of CBD and cellohexaose, the resonances narrowed. This showed that CBD interacted with the long cello-oligosaccharides, but it did not bind to the degradation products, i.e., cellobiose and cellotriose. However, the specific proof for the binding remained to be obtained, because no unambiguous transferred NOEs were observed between the tyrosine protons of CBD and the certain known protons of cellohexaose.

The experiments showed that cellohexaose may serve at least to some extent as a model compound for glucose chains on the cellulose surface, to mimic the CBD–cellulose interaction. On this basis a model was proposed for the binding, in which the three aligned aromatic rings of CBD_{CBHI} and CBD_{EGI} stack on

every other glucose unit of the glucose polymer (Fig. 27). It is conceivable that the aromatic rings preferably align with the hydrophobic patches of C1, C3 and C5 of glucose rings, which are larger than the patches of C2 and C4 rings. Such face-to-face stacking of aromatic residues with glucose units of cello-oligosaccharides has been observed in several protein–carbohydrate interactions. The planar polar amides of the glutamines and asparagines and hydroxyls of the tyrosines are potential candidates for hydrogen bonding with oxygens of glucose (Din *et al.*, 1995; Vyas, 1991; Quiocho, 1986, 1993). The amides that are off the alignment may assist in the inter-polymer hydrogen bond breaking.

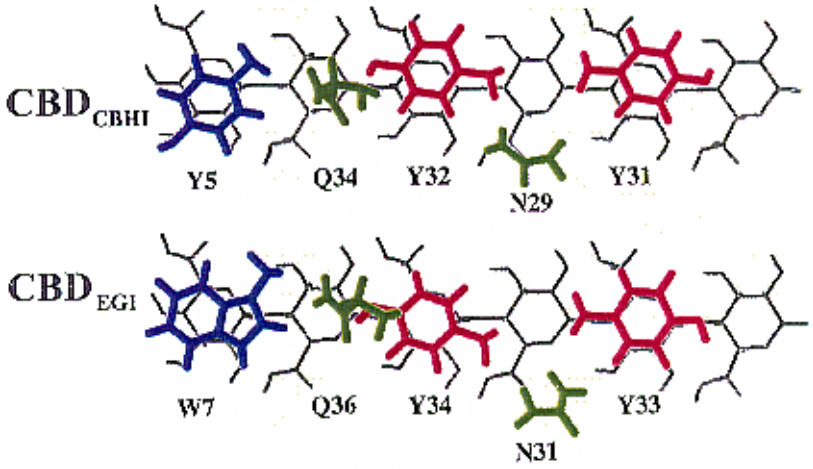


Figure 27. Binding model for CBD_{CBHI} and CBD_{EGI} on cellulose. CBD_{CBHI} and CBD_{EGI} are positioned on the glucose chain (only six units are shown) with the aromatic rings stacked on the top of the glucose rings. The alignment is within the precision of structure determination.

In natural cellulose, only in the obtuse corner (020) of the crystal are the glucose rings fully exposed (Gardner and Blackwell, 1974; Chanzy *et al.*, 1984; Reinikainen *et al.*, 1995). Exposed rings are a prerequisite for the binding model. There is experimental evidence to suggest, nevertheless, that most of the surface of a cellulose crystal is covered by CBD at saturation (Linder *et al.*, 1995), but it is unclear if CBDs bound on 110 or 1-10 surfaces actually contribute to the crystal breakdown. Face-to-face stacking of aromatic rings on the sugar rings may be required for a productive binding. When the polymer of the obtuse corner becomes solvated, more (020) surface is exposed.

5 Conclusions and future perspectives

The primary objective of this work was to investigate the relationship between the 3D structures and function of isolated CBDs from family I. Structural analyses were performed for six engineered peptides of CBD_{CBHI} (Y5A, P16R, N29A, Y31A, Y32A and Q34A) and for the intact peptide CBD_{EGI}. Understanding of the function of CBD is of fundamental importance, among other things, for the rational utilisation of cellulases in industrial applications. The main experimental method applied was ^{1}H 2D NMR spectroscopy in combination with distance geometry and simulated annealing calculations.

Studies involving amino acid substitutions showed which parts of CBD bind to cellulose and which parts do not. Except for Y5A, the conformation of the backbone was not essentially different in the engineered peptides and the wild-type CBD. In general, the N-terminus seemed to be the region most sensitive to changes. Together with the adsorption measurements, the secondary structure analyses of the mutants allowed the conclusion that the residues Tyr-31, Tyr-32 and Gln-34 on the flat face of CBD are functionally important in the binding to cellulose, while the rough face is not directly involved. Because of the clear structural changes in Y5A, unambiguous conclusions could not be drawn about the function of Tyr-5. Nevertheless, it was likely to be involved in the binding because of its position on the flat face.

The 3D structures of Y5A, Y31A and Y32A showed that the number of tyrosines and their precise spatial arrangement on the flat face are critical for the binding to crystalline cellulose. It seems to be a common denominator among the various CBD families that, in spite of completely different folds, a well-defined and rigid framework structure permitting very precise positioning of the aromatic side-chains is required for the specific interaction with the solid carbohydrate surface. Experiments with soluble cello-oligosaccharides supported this conclusion. The structure of the binding site is, of course, different for the family IV CBD than for the other CBDs since it binds to soluble sugars.

Determination of the NMR structure of CBD_{EGI} did not suggest an obvious reason for the different side-chains in CBD_{EGI} and CBD_{CBHI} nor a reason for the third disulphide bridge. Perhaps the third disulphide bridge is important for the connection of CBD to the linker peptide and thus for the function of CBD_{EGI}. Comparison of the model and NMR structure of CBD_{EGI} showed that peptide structures can be reliably modelled when the amino acid identity is high in the determined and template structure.

Knowledge of the 3D structure is the beginning, not the endpoint of structural biology and biophysics of protein function. Because cellulases, like other enzymes, are not rigid isolated molecules, understanding their function requires in addition knowledge of their dynamics and interactions with other biomolecules. NMR spectroscopy is not limited to structure determinations, however, and perhaps in the future other NMR-derived parameters can be used to study the intramolecular dynamical properties of cellulases. In particular, recent developments in multidimensional NMR techniques have increased the use of NMR spectroscopy for generating dynamical information on small- to medium-size biological macromolecules. Moreover, the use of NMR active ^{13}C and ^{15}N isotopes has resulted in dramatic advances in design and implementation of multidimensional heteronuclear NMR techniques. Consequently, the maximum size of protein that can be solved by NMR spectroscopy today has increased from ~ 10 kDa with ^{1}H NMR techniques to ~ 30 kDa with ^{13}C and ^{15}N NMR techniques, and perhaps to ~ 40 or ~ 50 kDa with ^{13}C and ^{15}N NMR techniques combined with ^{2}H enrichment. In the future, continued advances in sample preparation, magnet and r.f. technology and computational methods, in combination with the development of NMR spectroscopic techniques themselves, will make it possible to exploit the quantum mechanical properties of nuclear spin systems in a more sophisticated way.

References

Abuja, P., Pliz, I., Claeyssens, M. and Tomme, P. 1988a. Domain structure of cellobiohydrolase II as studied by small x-ray scattering: Close resemblance to cellobiohydrolase I. Biochem. Biophys. Res. Comm. 156, pp. 180–185.

Abuja, P., Schmuck, M., Pliz, I., Tomme, P., Claeyssens, M. and Esterbauer, H. 1988b. Structural and functional domains of cellobiohydrolase I from *Trichoderma reesei.* Eur. Biophys. J. 15, pp. 339–342.

Albrand, J.-P., Birdsall, B., Feeney, J., Roberts, G. and Burgen, A. 1979. The use of transferred nuclear Overhauser effects in the study of the conformations of small molecules bound to proteins. Int. J. Biolog. Macromolecules 1, pp. 37–41.

Anil, K., Ernst, R. and Wüthrich, K. 1980. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. Biochem. Biophys. Res. Comm. 95, pp. 1–6.

Atalla, R. and Vanderhart, D. 1984. Native cellulose: a composite of two distinct crystalline forms. Science 223, pp. 283–285.

Atalla, R. 1993. The structures of native celluloses. In: Suominen, P. and Reinikainen T. (Eds.). *Tricoderma reesei* cellulases and other hydrolases. Foundation for Biotechnical and Industrial Fermentation Research, Helsinki. Vol. 8, pp. 25–39.

Aue, W., Bartholdi, E. and Ernst, R. 1976. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. J. Chem. Phys. 64, pp. 2229–2246.

Barr, B., Hsieh, Y.-L., Ganem, B. and Wilson D. 1996. Identification of two functionally different classes of exocellulases. Biochemistry 35, pp. 589–592.

Barsukov, I. and Lian, L.-Y. 1993. Structure determination from NMR data I. Analysis of NMR data. In: Roberts, G. (Ed.). NMR of Macromolecules. A Practical Approach. IRL Press, Oxford. Pp. 315–357.

Bax, A. and Davis, D. 1985. MLEV-17-based two-dimensional homonuclear magnetisation transfer spectroscopy. J. Magn. Reson. 65, pp. 355–360.

Bax, A. 1989. Two-dimensional NMR and protein structure. Annu. Rev. Biochem. 58, pp. 223–256.

Bayer, E., Morag, E. and Lamed, R. 1994. The cellulosome–a treasure trove for biotechnology. TIBTECH 12, pp. 379–386.

Bayer, E., Morag, E., Wilchek, M., Lamed, R., Yaron, S. and Shoham, Y. 1995. Cellulosome domains for novel biotechnological application. In: Petersen, S., Svensson, B. and Pedersen, S. (Eds.). Carbohydrate Bioengineering. Elsevier Science, Amsterdam. Pp. 251–259.

Béguin, P. 1990. Molecular biology of cellulose degradation. Ann. Rev. Microbiol. 44, pp. 219–248.

Béguin, P. and Aubert, J. 1994. The biological degradation of cellulose. FEMS Microbiol. Rev. 13, pp. 25–58.

Bhikhabhai, R., Johansson, G. and Pettersson, G. 1984. Cellobiohydrolase from *Trichoderma reesei*: internal homology and prediction of secondary structure. Int. J. Peptide Protein Res. 25, pp. 368–374.

Blumenthal, L. 1970. Theory and applications of distance geometry. Chelsea, New York. 347 p.

Boisset, C., Borsali, R., Schülein, M. and Henrissat, B. 1995. Dynamic light scattering of the two-domain structure of *Humicola insolens* endoglucanase V. FEBS Lett. 376, pp. 49–52.

Braunschweiler, L. and Ernst, R. 1983. Coherence transfer by isotropic mixing: Application to correlation spectroscopy. J. Magn. Reson. 53, pp. 521–528.

Brun, E., Moriaud, F., Gans, P., Blackledge, M., Barras, F. and Marion, D. 1998. Solution structure of the cellulose-binding domain of the endoglucanase Z secreted by *Erwina chrysanthemi*. Biochemistry, in press.

Bundi, A. and Wüthrich, K. 1979. 1H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. Biopolymers 18, pp. 285–297.

Case, D., Dyson, H. and Wright, P. 1994. Use of chemical shifts and coupling constants in nuclear magnetic resonance structural studies on peptides and proteins. Methods Enzymol. 239, pp. 392–416.

Cavanagh, J. and Rance, M. 1990. Sensitivity improvement in isotropic mixing (TOCSY) experiments. J. Magn. Reson. 88, pp. 72–85.

Cavanagh, J., Fairbrother, W., Palmer III, A. and Skelton, N. 1996. Protein NMR Spectroscopy. Principles and Practice. Academic Press, Inc., San Diego. 587 p.

Chanzy, H., Henrissat, B. and Vuong, R. 1984. Colloidal gold labeling of 1,4-β-D-glucan cellobiohydrolase on *Valonia* cellulose microcrystals. FEBS Lett. 172, pp. 193–197.

Chanzy, H. 1990. Aspects of cellulose structure. In: Kennedy, J., Phillips, G. and Williams, P. (Eds.). Cellulose sources and exploitation: industrial utilization, biotechnology and physico-chemical properties. Ellis Horwood, New York. Pp. 3–12.

Clore, M. and Gronenborn, A. 1982. Theory and application of the transferred nuclear Overhauser effect to study of the conformations of small ligands bound to proteins. J. Magn. Reson. 48, pp. 402–417.

Clore, M. and Gronenborn, A. 1983. Theory of the time dependent transferred nuclear Overhauser effect: Applications to structural analysis of ligand–protein complexes in solution. J. Magn. Reson. 53, pp. 423–442.

Clore, G. and Gronenborn, A. 1991a. Two-, three- and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution. Annu. Rev. Biophys. Biochem. 20, pp. 29–63.

Clore, G., Wingfield, P. and Gronenborn, A. 1991b. High resolution three-dimensional structure of interleukin-1β in solution by three- and four-dimensional nuclear macnetic resonance spectroscopy. Biochemistry 30, pp. 2315–2323.

Coudray, M., Canevascini, G. and Meier, H. 1982. Characterization of a cellobiose dehydrogenase in the cellulolytic fungus *Sporotrichum Chrysosporium* thermophile. Biochem. J. 203, pp. 277–284.

Coughlan, M. 1985. The properties of fungal and bacterial cellulases with comment on their production and application. In: Russel, G. (Ed.). Biotechnology and genetic reviews. Intercept Ltd., Newcastle upon Tyne. Vol 3, pp. 39–109.

Coutinho, J., Moser, B., Kilburn, D., Warren, R. and Miller, R. Jr. 1991. Nucleotide sequence of the endoglucanase C gene *cenC* of *Cellulomonas fimi*, its high-level expression in *Escherichia coli*, and characterization of its products. Mol. Microbiol. 5, pp. 1221–1233.

Coutinho, J., Gilkes, N., Warren, R., Kilburn, D. and Miller, R. Jr. 1992. The binding of *Cellulomonas fimi* endoglucanase C CenC to cellulose and Sephadex is mediated by the *N*-terminal repeats. Mol. Microbiol. 6, pp. 1243–1252.

Coutinho, J., Gilkes, N., Kilburn, D., Warren, R. and Miller, J. 1993. The nature of the cellulose-binding domain affects the activities of a bacterial endoglucanase on different forms of cellulose. FEMS Microbiol. Lett. 113, pp. 211–218.

Crippen, G. 1977. A novel approach to the calculation of conformation: distance geometry. J. Comp. Phys. 24, pp. 96–107.

Davis, D. and Bax, A. 1985. Assignment of complex ^{1}H NMR spectra via two-dimensional homonuclear Hartmann–Hann spectroscopy. J. Am. Chem. Soc. 107, pp. 2820–2821.

Denman, S., Xue, G. and Patel, B. 1996. Characterisation of a *Neocallomastix patriciarum* cellulase cDNA (CelA) homologous to *Tricoderma reesei* cellobiohydrolase II. Appl. Environ. Microbiol. 62, pp. 1889–1896.

Derome, A. 1987. Describing pulse NMR. In: Baldwin, J. (Ed.). Modern NMR Techniques for Chemistry Research. Pergamon Press, Great Britain. 280 p.

Dijkerman, R., Vervurem, M., Camp, H. and Drift, C. van der 1996. Adsorption characteristics of cellulolytic enzymes from the anaerobic fungus *Piromyces* sp. strain E2 on microcrystalline cellulose. Appl. Environ. Microbiol. 62, pp. 20–25.

Din, N., Gilkes, N., Tekant, B., Miller, R. Jr., Warren, R. and Kilburn, D. 1991. Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. Bio/Technol. 9, pp. 1096–1099.

Din, N., Forsythe, I., Burtnick, L., Gilkes, N., Miller, R., Warren, R. and Kilburn D. 1994. The cellulose-binding domain of endoglucanase A (CenA) from *Cellumonas fimi*: Evidence for involvement of tryptophan residues in binding. Mol. Microbiol. 11, pp. 747–755.

Din, N., Coutinho, J., Gilkes, N., Jervis, E., Kilburn, D., Miller, R. Jr., Ong, E., Tomme, P. and Warren, R. 1995. Interactions of cellulases from *Cellulomonas fimi* with cellulose. In: Petersen, S.B., Svensson, B. and Pedersen, S. (Eds.). Carbohydrate Bioengineering. Elsevier Science, Amsterdam. Pp. 261–270.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J., Teeri, T. and Jones, A. 1994. The three-dimensional structure of cellobiohydrolase I from *Trichoderma reesei* reveals a lysozyme-like active site on a lectin-like framework. Science 265, pp. 524–528.

Durand, H., Baron, M., Calmels, T. and Tiraby, G. 1988. Classical and molecular genetics applied to *Trichoderma reesei* for the selection of improved cellulolytic industrial strains. In: Aubert, J.-P., Béguin, P. and Millet, J., (Eds.). Biochemistry and Genetics of Cellulose Degradation. Academic Press, San Diego. Pp. 135–151.

Dyson, H. and Wright, P. 1991. Defining solution conformations of small linear peptides. Annu. Rev. Biophys. Biophys. Chem. 20, pp. 519–538.

Ernst, R., Bodenhausen, G. and Wokaun, A. 1987. Principles of nuclear magnetic resonance in one and two dimensions. Oxford Science Publications. 610 p.

Fanutti, C., Ponyi, T., Black, G., Hazlewood, G. and Gilbert, H. 1995. The conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from anaerobic fungi functions as a protein docking domain. J. Biol. Chem. 270, pp. 29314–29322.

Field, L. 1989. Fundamental Aspects of NMR Spectroscopy. In: Field, L. and Sternhell, S. (Eds.). Analytical NMR. John Wiley & Sons Ltd, Chichester. 250 p.

Gardner, K. and Blackwell, J. 1974. The structure of native cellulose. Biopolymers 13, pp. 1975–2001.

Gilkes, N., Warren, R., Miller, R. and Kilburn, D. 1988. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. 263, pp. 10401–10407.

Gilkes, N., Henrissat, B., Kilburn, D., Miller, R. Jr. and Warren, R. 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function and enzyme families. Microbiol. Rev. 55, pp. 305–315.

Gilkes, N., Jervis, E., Henrissat, B., Tekant, B., Miller, R. Jr., Warren, R. and Kilburn, D. 1992. The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. J. Biol. Chem. 267, pp. 6743–6749.

Greenwood, J., Gilkes, N., Kilburn, D., Miller, R. and Warren, R. 1989. Fusion to an endoglucanase allows alkaline phosphatase to bind to cellulose. FEBS Lett. 244, pp. 127–131.

Gronenborn, A. and Clore, G. 1995. Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy. Crit. Rev. Biochem. Mol. Biol. 30, pp. 351–358.

Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. 1989. Automated stereospecific ^{1}H NMR assignments and their impact on the precision of protein structure determinations in solution. J. Am. Chem. Soc. 111, pp. 3997–4004.

Hall, J., Black, G., Ferreira, L., Millward-Saddler, S., Ali, B., Hazlewood, G. and Gilbert, G. 1995. The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluoresceens* subsp. *celluloa* is important for the efficient hydrolysis of Avicel. Biochem. J. 309, pp. 749–756.

Harris, R. 1986. Nuclear Magnetic Resonance spectroscopy. Longman Scientific and Technical, Longman Harlow Group UK Limited. 268 p.

Havel, T., Kuntz, I. and Crippen, G. 1983. The theory and practice of distance geometry. Bull. Math. Biol. 45, pp. 665–720.

Havel, T. and Wüthrich, K. 1984. A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^{1}H-^{1}H proximities in solution. Bull. Math. Biol. 46, pp. 673–698.

Havel, T. and Wüthrich, K. 1985. An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. J. Mol. Biol. 182, pp. 281–294.

Havel, T. 1990. The sampling properties of some distance geometry algorithms applied to unconstrained polypeptide chains: A study of 1830 independently computed conformations. Biopolymers 29, pp. 1565–1585.

Havel, T. 1991. An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. Prog. Biophys. Mol. Biol. 56, pp. 43–78.

Henrissat, B., Driguez, H., Viet, C. and Schülein, M. 1985. Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. Bio/Technol. 3, pp. 722–726.

Henrissat, B. 1994. Cellulases and their interaction with cellulose. Cellulose 1, pp. 169–196.

Hoffrén, A.-M., Teeri, T. and Teleman, O. 1995. Molecular dynamics simulation of fungal cellulose-binding domains: differences in molecular rigidity but a preserved cellulose binding surface. Protein Eng. 8, pp. 443–450.

Homans, S. 1990. A Dictionary of concepts in NMR. Clarendon press, Oxford. 343 p.

Hon, D. 1994. Cellulose: a random walk along its historical path. Cellulose 1, pp. 1–25.

Howarth, M., Lian, L., Hawkes, G. and Sales, K. 1986. Formalism for the description of multiple-pulse NMR experiments. J. Magn. Reson. 68, pp. 433–452.

Irwin, D., Spezio, M., Walker, L. and Wilson, D. 1993. Activity studies of eight purified cellulases: specificity, synergism and binding domain effects. Biotechnol. Bioeng. 42, pp. 1002–1013.

Irwin, D., Jung, E. and Wilson, D. 1994. Characterization of a *Thermonospora fusca* xylanase. Appl. Environ. Microbiol. 60, pp. 763–770.

Jackson, L., Joyce, T., Heitmann, J. and Giesbrecht, F. 1996. Enzyme recovery from secondary fiber treated with cellulase and xylanase. J. Biotechnol. 45, pp. 33–44.

Jeener, J. 1971. Ampére Summer School, Basko Polje, Yugoslavia.

Jeener, J., Meier, B., Bachmann, P. and Ernst, R. 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. J. Chem. Phys. 71, pp. 4546–4553.

Johnson, P., Joshi, M., Tomme, P., Kilburn, D. and McIntosh, L. 1996a. Structure of the N-terminal cellulose-binding domain of *Cellumonas fimi* CenC determined by nuclear magnetic resonance spectroscopy. Biochemistry 35, pp. 14381–14394.

Johnson, P., Tomme, P., Joshi, M. and McIntosh, L. 1996b. Interaction of soluble cello-oligosaccharides with the N-terminal cellulose-binding domain of *Cellumonas fimi* CenC. 2. NMR and ultraviolet absorption spectroscopy. Biochemistry 35, pp. 13895–13906.

Karplus, M. 1963. Vicinal proton coupling in nuclear magnetic resonance. J. Am. Chem. Soc. 85, pp. 2870–2871.

Kleywegt, G., Zou, J.-Y., Divne, C., Davies, G., Sinning, I., Ståhlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T. and Jones, A. 1997. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. J. Mol. Biol. 272, pp. 383–397.

Klyosov, A. 1990. Trends in biochemistry and enzymology of cellulose degradation. Biochemistry 29, pp. 10577–10585.

Knowles, J., Lehtovaara, P. and Teeri, T. 1987. Cellulase families and their genes. Trends. Biotechnol. 5, pp. 255–261.

Kraulis, P., Clore, M., Nilges, M., Jones, A., Petterson, G., Knowles, J. and Gronenborn, A. 1989. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. Biochemistry 28, pp. 7241–7257.

Kroon-Batenburg, L. and Kroon, J. 1994. The crystal and molecular structures of cellulose. Carbohydrates in Europe, pp. 15–19.

Kuga, S. and Brown, R. Jr. 1987a. Practical aspects of lattice imaging of cellulose. J. Electr. Microsc. Tech. 6, pp. 349–356.

Kuga, S. and Brown, R. Jr. 1987b. Lattice imaging of ramie cellulose. Polymer comm. 28, pp. 311–314.

Kuntz, I., Thomason, J. and Oshiro, C. 1989. Distance geometry. Methods Enzymol. 177, pp. 159–204.

Kuszewski, J., Nilges, M. and Brünger, A. 1992. Sampling and efficiency of metric matrix distance geometry: A novel partial metrization algorithm. J. Biomol. NMR 2, pp. 33–56.

Lamed, R., Setter, E. and Bayer, E. 1983. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. J. Bacteriol. 156, pp. 828–836.

Linder, M., Lindeberg, G., Reinikainen, T., Teeri, T. and Pettersson, G. 1995. The difference in affinity between two fungal cellulose-binding domains is dominated by single amino acid substitution. FEBS Lett. 372, pp. 96–98.

Linder, M., Salovuori, I., Ruohonen, L. and Teeri, T. 1996. Characterization of a double cellulose-binding domain: synergistic high-affinity binding to cellulose. J. Biol. Chem. 271, pp. 21258–21272.

Linder, M. and Teeri, T. 1996. The cellulose-binding domain of the major cellobiohydrolase of *Trichoderma reesei* exhibits true reversibility and a high exchange rate on crystalline cellulose. Proc. Natl. Acad. Sci. 93, pp. 12251–12255.

Linder, M. and Teeri, T. 1997. The roles and function of cellulose-binding domains. J. Biotechnol. 57, pp. 15–28.

Macura, S. and Ernst, R. 1980. Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. Mol. Phys. 41, pp. 95–117.

Marion, D. and Wüthrich, K. 1983. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants. Biochem. Biophys. Res. Commun. 113, pp. 967–974.

Meinke, A., Braun, C., Gilkes, N., Kilburn, D., Miller, R. Jr. and Warren, R. 1991a. Unusual sequence organization in CenB, an inverting endoglucanase from *Cellulomonas fimi*. J. Bacteriol. 173, pp. 308–314.

Meinke, A., Gilkes, N., Kilburn, D., Miller, R. Jr. and Warren, R. 1991b. Multiple domains in endoglucanase B (CenB) from *Cellulomonas fimi*: functions and relatedness to domains in other polypeptides. J. Bacteriol. 173, pp. 7126–7135.

Meinke, A., Schmuck, M., Gilkes, N., Kilburn, D., Miller, R. Jr. and Warren, R. 1992. The tertiary structure of endo-β-1,4-glucanase B (CenB), a multidomain cellulase from the bacterium *Cellulomonas fimi*. Glycobiology 2, pp. 321–326.

Meinke, A., Gilkes, N., Kilburn, D., Miller, R. Jr. and Warren, R. 1993. Cellulose-binding polypeptides from *Cellulomonas fimi*: Endoglucanase D (CenD), a family A β-1,4-glucanase. J. Bacteriol. 175, pp. 1910–1918.

Merutka, G., Dyson, H., and Wright, P. 1995. 'Random coil' 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. J. Biomol. NMR 5, pp. 14–27.

Millward-Sadler, S., Poole, D., Henrissat, B., Hazlewood, G., Clarke, J. and Gilbert, H. 1994. Evidence for a general role for high-affinity non-catalytic cellulose-binding domains in microbial plant cell wall hydrolases. J. Mol. Microbiol. 11, pp. 375–382.

Neuhaus, D. and Williamson, M. 1989. The nuclear Overhauser effect in structural and conformational analysis. VHC Publishers Inc., New York. 522 p.

Nevalainen, H. and Penttilä, M. 1995. Molecular biology of cellulolytic fungi. In: Kück, U. (Ed.). The Mycota II. Genetics and Biotechnology. Springer-Verlag, Berlin, Heidelberg. Pp. 303–319.

Ni, F. 1994. Recent developments in transferred NOE methods. Prog. NMR Spectrosc. 26, pp. 517–606.

Nidetzky, B., Hayn, M., Macarron, R. and Steiner, W. 1993. Synergism of *Trichoderma reesei* cellulases while degrading different celluloses. Biotechnology Lett. 15, pp. 71–76.

Nilges, M., Clore, G. and Gronenborn, A. 1988a. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. FEBS Lett. 239, pp. 129–136.

Nilges, M., Clore, G. and Gronenborn, A. 1988b. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry–dynamical simulated annealing calculations. FEBS Lett. 229, pp. 317–324.

Nilges, M., Gronenborn, A., Brünger, A. and Clore, G. 1988c. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. Protein Eng. 2, pp. 27–38.

NMRchitect 1995. Structure generation, refinement and analysis. User guide. Biosym/MSI, San Diego, California, USA.

Ong, E., Greenwood, J., Gilkes, N., Kilburn, D., Miller, R. and Warren, R. 1989. The cellulose-binding domains of cellulases: tools for biotechnology. Trends. Biotech. 7, pp. 239–243.

Ong, E., Gilkes, N., Miller, R. and Warren, R. 1991. Enzyme immobilization using a cellulose-binding domain: properties of a β-glucosidase fusion protein. Enzyme Microb. Technol. 13, pp. 59–65.

Ong, E., Gilkes, N., Miller, R., Warren, R. and Kilburn, D. 1993. The cellulose-binding domain (CBD_{Cex}) of an exoglucanase from *Cellulomonas fimi*: production in *Escherichia coli* and characterization of the polypeptide. Biotechnol. Bioeng. 42, pp. 401–409.

Pardi, A., Billeter, M. and Wüthrich, K. 1984. Calibration of the angular dependence of the amide proton-Cα proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein. J. Mol. Biol. 180, pp. 741–751.

Penttilä, M., Lehtovaara, P., Nevalainen, H., Bhikhabhai, R. and Knowles, J. 1986. Homology between cellulase genes of *Trichoderma reesei:* Complete nucleotide sequence of the endoglucanase I gene. Gene 45, pp. 253–263.

Pere, J., Siika-aho, M., Buchert, J. and Viikari, L. 1995. Effects of purified *Trichoderma reesei* cellulases on the fiber properties of craft pulp. Tappi J. 78, pp. 71–78.

Phelps, M., Hobbs, J., Kilburn, D. and Turner, R. 1994. Technology for regenerable biosensor probes based on enzyme-cellulose binding domain conjugates. Biotechnol. Prog. 10, pp. 433–440.

Pilz, I., Schwarz, E., Kilburn, D., Miller, R. Jr., Warren, R. and Gilkes, N. 1990. The tertiary structure of a bacterial cellulase determined by small-angle X-ray scattering analysis. Biochem. J. 271, pp. 277–280.

Poole, D., Hazlewood, G., Huskisson, N., Virden, R. and Gilbert, H. 1993. The role of conserved tryptophan residue in the interaction of a bacterial cellulose binding domain with its ligand. FEMS Microbiol. Lett. 106, pp. 77–84.

Quiocho, F. 1986. Carbohydrate-binding proteins: tertiary structures and protein–sugar interactions. Ann. Rev. Biochem. 55, pp. 287–315.

Quiocho, F.A. 1993. Probing the atomic interactions between proteins and carbohydrates. Biochemical Society Transactions 21, pp. 442–448.

Rahkamo, L., Siika-aho, M., Vehviläinen, M., Dolk, M., Viikari, L., Nousiainen, P. and Buchert, J. 1996. Modification of hardwood dissolving pulp with purified *Trichodermaa reesei* cellulases. Cellulose 3, pp. 1–11.

Reinikainen, T., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T., Knowles, J. and Teeri, T. 1992. Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase I. Proteins 14, pp. 475–482.

Reinikainen, T., Teleman, O. and Teeri, T. 1995. Effects of pH and high ionic strength on the adsorption and activity of native and mutated cellobiohydrolase I from *Trichoderma reesei*. Proteins 22, pp. 392–403.

Reinikainen, T., Takkinen, K. and Teeri, T. 1997. Comparison of the adsorption properties of a single-chain antibody fragment fused to a fungal or a bacterial cellulose-binding domain. Enzyme Microb. Technol. 20, pp. 143–149.

Richardson, J. 1981. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34, pp. 167–339.

Roberts, G. 1993. NMR of macromolecules. A practical approach. Oxford University Press, IRL. 399 p.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. and Jones, T. 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science 249, pp. 380–386.

Sakka, K., Kojioma, Y., Kondo, T., Karita, S., Ohmiya, K. and Shimada, K. 1993. Nucleotide sequence of the *Clostridium stercoriarium* xynA gene encoding xylanase A: Identification of catalytic and cellulose binding domains. Biosci. Biotech. Biochem. 57, pp. 273–277.

Saloheimo, M., Lehtovaara, P., Penttilä; M., Teeri, T., Ståhlberd, J., Johansson G., Pettersson, G., Claeyssens, M., Tomme, P. and Knowles, J. 1988. EGIII, a new endoglucanase from *Trichoderma reesei*: The characterization of both gene and enzyme. Gene 63, pp. 11–21.

Saloheimo, A., Henrissat, B., Hoffrén, A., Teleman, O. and Penttilä, M. 1994. A novel, small endoglucanase gene, *egl5*, from *Trichoderma reesei* isolated by expression in yeast. Mol Microbiol 12, pp. 219–228.

Sarko, A. 1987. Cellulose–How much do we know about its structure? In: Kennedy, J.F., Phillips, G.O. and Williams, P.A. (Eds.). Wood and cellulose industrial utilization, biotechnology, structure and properties. John Wiley and Sons, New York, pp. 55–69.

Scheek, R. and Gunsteren, W. van 1989. Molecular dynamics simulation techniques for determination of molecular structures from nuclear magnetic resonance data. Methods Enzymol., pp. 204–218.

Shen, H., Schmuck, M., Pliz, I., Gilkes, N., Kilburn, D., Miller, R. Jr. and Warren, R. 1991. Deletion of the linker connecting the catalytic and cellulose-binding domains of endoglucanase A (CenA) of *Cellulomonas fimi* alters its conformation and catalytic activity. J. Biol. Chem. 17, pp. 11335–11340.

Shimon, L., Bayer, E., Morag, E., Lamed, R., Yaron, S., Shoham, Y. and Frolow, F. 1997. A cohesion domain from *Clostridium thermocellum*: The crystal structure provides new insights into cellulosome assembly. Structure 5, pp. 381–390.

Shoemaker, S., Schweickart, V., Lander, M., Gelfand, D., Kwok, S., Myambo, K. and Innis, M. 1983. Molecular cloning of exo-cellobiohydrolase I derived from *Trichoderma reesei* strain L27. Biotechnology 1, pp. 691–696.

Sorensen, O., Eich, G., Levitt, M., Bodenhausen, G. and Ernst, R. 1983. Product operator formalism for the description of NMR pulse experiments. Prog. NMR Spectrosc. 16, pp. 163–192.

Srisodsuk, M., Reinikainen, T., Penttilä, M. and Teeri, T. 1993. Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. J. Biol. Chem. 268, pp. 20756–20761.

Ståhlberg, J., Johansson, G. and Pettersson, G. 1991. A new model for enzymatic hydrolysis of cellulose based on the two-domain structure of cellobiohydrolase I. Biotechnology 9, pp. 286–290.

Ståhlberg, J. 1991. Functional organization of cellulases from *Trichoderma reesei*, Ph.D. Thesis. University of Uppsala. Sweden. 45 p.

Sugiyama, J., Harada, F., Fujiyoshi, Y. and Uyeda, N. 1984. High resolution observations of cellulose microfibrils. Mokuzai Gakkaishi 20, pp. 98–99.

Sugiyama, J., Harada, F., Fujiyoshi, Y. and Uyeda, N. 1985. Lattice images from ultrathin sections of cellulose microfibrils in the cell wall of *Valonia macrophysa*. Planta 166, pp. 161–168.

Sugiyama, J., Okano, T., Yamamoto, H. and Horii, F. 1991a. Transformation of *Valonia* cellulose crystals by an alkaline hydrothermal treatment. Macromolecules 23, pp. 2461–2466.

Sugiyama, J., Vuong, R. and Chanzy, H. 1991b. An electron diffraction study on the two phases occurring in native cellulose from algal cell wall. Macromolecules 24, pp. 4168–4175.

Sugiyama, J., Persson, J. and Chanzy, H. 1991c. Combined infrared and electron diffraction study of polymorphism of native celluloses. Macromolecules 24, pp. 2461–2466.

Sörensen, O., Eich, G., Levitt, M., Bodenhausen, G. and Ernst, R. 1983. Product operator formalism for the description of NMR pulse experiments. Progress in NMR Spectroscopy 16, pp. 163–192.

Teeri, T., Lehtovaara, P., Kauppinen, S., Salovuori, I. and Knowles, J. 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: Gene sequence and expression of cellobiohydrolase II. Gene 51, pp. 43–52.

Teeri, T., Reinikainen, T., Ruohonen, L., Jones, T. and Knowles, J. 1992. Domain function in *Trichoderma reesei* cellobiohydrolases. J. Biotechnol. 24, pp. 169–176.

Tilbeurgh, H. van, Tomme, P., Claeyssens, M., Bhikhabhai, R. and Pettersson, G. 1986. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei.* FEBS Lett. 204, pp. 223–227.

Tomme, P., Tilbeurgh, H. van, Pettersson, G., Damme, J. van, Vandekerckhove, J., Knowles, J., Teeri, T. and Claeyssens, M. 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem. 170, pp. 575–581.

Tomme, P., Warren, R., Miller, R., Kilburn, D. and Gilkes, N. 1995a. Cellulose-binding domains: Classification and properties. In: Saddler, J. and Penner M. (Eds.). Enzymatic degradation of insoluble carbohydrates. American Chemical Society, Washington DC. Vol 618, pp. 143–163.

Tomme, P., Driver, D., Amandoron, E., Miller, R., Warren, R. and Kilburn, D. 1995b. Comparison of a fungal (family I) and bacterial (family II) cellulose-binding domain. J. Bacteriol. 177, pp. 4356–4363.

Tormo, J., Lamed, R., Chirino, A., Morag, E., Bayer, E., Shoham, Y. and Steitz, T. 1996. Crystal structure of a bacterial family III cellulose-binding domain: a general mechanism for attachment to cellulose. The EMBO J. 15, pp. 5739–5751.

Ven, F. van de and Hilbers, C. 1983. A simple formalism for the description of multiple-pulse experiments. Application to a weakly coupled two-spin (I = ½) system. J. Magn. Reson. 54, pp. 512–520.

Vyas, N. 1991. Atomic features of protein-carbohydrate interactions. Curr. Opin. Struct. Biol. 1, pp. 723–740.

Wagner, G., Braun, W., Havel, T., Schaumann, T., Go, N. and Wüthrich, K. 1987. Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. J. Mol. Biol. 196, pp. 611–639.

Wagner, G. 1990. NMR investigations of protein structure. Progress in NMR Spectroscopy 22, pp. 101–139.

Ward, M., Wu, S., Dauberman, J., Weiss, G., Larenas, E., Bower, B., Rey, M., Clarkson, K. and Bott, R. 1993. Cloning, sequence and preliminary structural analysis of a small, high pI endoglucanase (EGIII) from *Trichoderma reesei*. In: Suominen, P. and Reinikainen, T. (Eds.). *Trichoderma reesei* cellulases and other hydrolases. Foundation for Biotechnical and Industrial Fermentation Research, Helsinki. Vol 8, pp. 153–158.

Wilson, D. 1992. Biochemistry and genetics of *Actinomycete* cellulases. Critical reviews in biotechnology 12, pp. 45–63.

Wilson, D., Spezio, M., Irwin, D., Karplus, A. and Taylor, J. 1995. Comparison of enzymes catalyzing the hydrolysis of insoluble polysaccharides. In: Saddler, J. and Penner M. (Eds.). Enzymatic degradation of insoluble carbohydrates. American Chemical Society, Washington DC. Vol 618, pp. 1–12.

Winterhalter, C., Heinrich, P., Candussio, A., Wich, G. and Liebl, W. 1995. Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*. Mol. Microbiol. 15, pp. 431–444.

Wishart, D., Bigham, C., Holm, A., Hodges, R. and Sykes, B. 1995. ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J. Biomol. NMR 5, pp. 67–81.

Woodward, J., Stephan, L., Koran, L., Wong, K. and Saddler, J. 1994. Enzymatic separation of high-quality uninked pulp fibers from recycled newspaper. Biotechnology 12, pp. 905–908.

Wüthrich, K. 1986. NMR of proteins and nucleic acids. Wiley and Sons, Inc. 292 p.

Wüthrich, K., Wider, G., Wagner, G. and Braun, W. 1982. Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. J. Mol. Biol. 155, pp. 311–319.

Wüthrich, K., Billeter, M. and Brown, W. 1983. Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J. Mol. Biol. 169, pp. 949–961.

Xu, G.-Y., Ong, E., Gilkes, N., Kilburn, D., Muhandiram, D., Harris-Brandts, M., Carver, J., Kay, L. and Harvey, T. 1995. Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. 1995. Biochemistry 34, pp. 6993–7009.

Appendices of this publication are not included in the PDF version.
Please order the printed version to get the complete publication
(http://www.inf.vtt.fi/pdf/publications/1998/)