

NN08201, 956

G. VRIEND

Molecular Interactions
During the Assembly of
Cowpea Chlorotic Mottle Virus
Studied by Magnetic Resonance

Proefschrift

ter verkrijging van de graad van
doctor in de Landbouwwetenschappen,
op gezag van de rector magnificus
dr. C.C. Oosterlee,
in het openbaar te verdedigen
op vrijdag 14 oktober 1983
des namiddags te vier uur in de aula
van de Landbouwhogeschool te Wageningen.

CENTRALE LANDBOUWCATALOGUS



0000 0002 0947

BOEKTHEEK
DER
LANDBOUWHOGESCHOOL
WAGENINGEN

157 193155-03

Promotor : dr. T.J. Schaafsma,
hoogleraar in de molecuulfysica

Co-promotor : dr. M.A. Hemminga
wetenschappelijk hoofdmedewerker

STELLINGEN

- 1) De bepaling van lineaire stroomsnelheden met behulp van NMR is zonder veldgradiënt niet betrouwbaar.
Singer, J.R., and L.E. Crooks, (1985), Science 221, pp. 654-656.
- 2) Het gebruik van BPTI als testmolecuul voor nieuw ontwikkelde fysische meetmethoden wekt valse hoop bij potentiële gebruikers van deze methoden.
- 3) Barron en Wilson kennen ten onrechte enkele pieken in het ¹³C-MAS spectrum van Maungatua silt loam toe aan het meetobject in plaats van aan de gebruikte spinner.
Barron, P.F., and M.A. Wilson, (1981), Nature 289, pp. 275-276.
- 4) De factor 1.000.000 die Kusumi et al. te laag uitkomen bij het bepalen van de rotatiecorrelatietijd van rhodopsine in de met glutaardialdehyde behandelde visuele receptor membraan, wordt veroorzaakt doordat zij vergeten de anisotrope spinlabelrotatie bij hun interpretatie van de ST-ESR spectra te betrekken.
Kusumi, A., Ohishi, S., Ito, T., and T. Yoshizana, (1978), B.B.A. 507, pp. 539-543.
- 5) Hogan and Jardetzky concluderen ten onrechte dat de interne mobiliteit in DNA afneemt ten gevolge van de toevoeging van ethidiumbromide.
Hogan, M.E., and O. Jardetzky, (1980), Biochemistry 19, pp. 2079-2085.
- 6) De door Ribeiro, King, Restivo en Jardetzky bepaalde temperatuur tijdens de NOE metingen aan BPTI is een factor 10 minder nauwkeurig dan zij opgeven. Dezelfde auteurs geven de in BPTI bepaalde relaxatieparameters op met een nauwkeurigheid die 10.000 maal groter is dan uit de gepubliceerde spectra bepaald kan worden.
Ribeiro, A.A., King, R., Restivo, C., and O. Jardetzky, (1980), J.A.C.S. 102, pp. 4040-4051.
- 7) Angst voor computers wordt pas terecht indien kinderen op de lagere school octaal in plaats van decimaal moeten leren rekenen.

- 8) De ANWB zou de gebruikers van haar waterkaarten een dienst bewijzen indien zij campings waarop slechts caravans worden toegelaten op deze waterkaarten zou aangeven met een symbolische caravan in plaats van een symbolisch tentje.
- 9) Schakende computers zijn slecht voor "computerende" schakers.

hartelijk gefeliciteerd

==

CONTENTS:		pag.
	Acknowledgements.	5
1	General introduction.	7
1.1	Questions.	7
1.2	The icosahedron.	9
1.3	Structural aspects of spherical plant viruses.	11
1.4	Protein-RNA interaction.	14
1.5	Protein-protein interactions.	15
1.6	Cowpea chlorotic mottle virus.	16
1.7	Nuclear magnetic resonance.	18
1.8	Electron spin resonance.	20
1.9	Secondary structure predictions.	21
1.10	Energy calculations.	22
1.11	References.	23
2	Swelling of cowpea chlorotic mottle virus studied by $^1\text{H-NMR}$.	27
2.1	Introduction.	27
2.2	Materials and methods.	27
2.3	Results and discussion.	28
2.4	References.	30
3	EPR and ST-EPR on spin-labelled cowpea chlorotic mottle virus.	32
3.1	Introduction.	32
3.2	Materials and methods.	32
3.2.1	Preparation of spin-labelled virus.	32
3.2.2	Fixation with glutaraldehyde.	33
3.2.3	EPR and ST-EPR spectroscopy.	33
3.3	Results.	34
3.4	Discussion.	37
3.5	References.	39
4	Segmental mobility involved in protein-RNA interaction in cowpea chlorotic mottle virus.	40
4.1	Introduction.	40

4.2	Materials and methods.	40
4.2.1	Preparation of protein lacking the N-terminal amino acids.	40
4.2.2	Protein assembly.	41
4.2.3	Nucleoprotein assembly.	41
4.2.4	NMR measurements.	41
4.3	Results.	43
4.3.1	The role of the N-terminal amino acids in protein-RNA interaction.	43
4.3.2	Mobility of the N-terminal region of the protein.	45
4.4	Discussion.	46
4.5	References.	47
5	Mobility involved in protein-RNA interaction in spherical plant viruses studied by $^1\text{H-NMR}$.	48
5.1	Introduction.	48
5.2	Materials and methods.	48
5.2.1	Virus preparations.	48
5.2.2	NMR measurements.	49
5.3	Results.	49
5.3.1	Brome mosaic virus.	49
5.3.2	Belladonna mottle virus.	51
5.3.3	Cowpea mosaic virus.	51
5.4	Discussion.	51
5.5	References.	53
6	Interaction of the N-terminal arm of CCMV protein with nucleotides.	55
6.1	Introduction.	55
6.2	Materials and methods.	55
6.2.1	Preparation of viral coat protein.	55
6.2.2	Oligo-nucleotides.	56
6.2.3	NMR measurements.	56
6.3	Results.	56
6.3.1	The oligo-nucleotides.	56
6.3.2	Binding of oligo-nucleotides to protein dimers.	58
6.3.3	Binding of oligo-nucleotides to empty capsids.	61
6.4	Discussion.	64
6.5	References.	66

7	Secondary structure prediction of the N-terminal protein arm of CCMV.	68
7.1	Introduction.	68
7.2	Secondary structure prediction methods.	69
7.2.1	Chou and Fasman.	69
7.2.2	Burgess, Ponnuswamy and Scheraga.	70
7.2.3	Lim.	71
7.2.4	Modifications.	71
7.3	Results and discussion.	73
7.4	References.	78
8	Energy calculations on the N-terminal arm of CCMV.	79
8.1	Introduction.	79
8.2	The calculation method.	80
8.3	Results.	82
8.4	Discussion.	84
8.5	References.	88
9	Analysis of the structure and mobility of the N-terminal arm of CCMV protein.	89
9.1	A comparison of the N-terminal arm in dimers and empty capsids.	89
9.1.1	Introduction.	89
9.1.2	Materials and methods.	89
9.1.3	Results and discussion.	90
9.1.3.1	The number of mobile N-terminal arms in empty capsids.	90
9.1.3.2	Differences between arms in capsids and dimers.	92
9.1.3.3	Mobility in the protein subunits lacking the N-terminal arm.	94
9.2	The structure of the N-terminal arm.	96
9.2.1	Introduction.	96
9.2.2	Materials and methods.	97
9.2.3	Results and discussion.	97
9.2.3.1	2D-NOE measurements.	97
9.2.3.2	Conformational exchange in the N-terminal arm.	100
9.3	Assignments in the 0-5 ppm region in spectrum of empty capsids.	101
9.3.1	Introduction.	101
9.3.2	Materials and methods.	102
9.3.3	Results and discussion.	102
9.4	Carbon 13 relaxation measurements.	106
9.4.1	Introduction.	106

9.4.2	Materials and methods.	106
9.4.2.1	Carbon 13 enrichment.	106
9.4.2.2	NMR measurements.	107
9.4.3	Results and discussion.	107
9.5	Conclusions.	109
9.6	References.	110
	Summary.	111
	Samenvatting	113
	Curriculum Vitae	115
	Appendices:	
A	List of chemical shifts used for assignments.	116
B	Primary structure of the coat protein of CCMV.	118
C	Optimization method for secondary structure predictions.	119
D	List of abbreviations.	120

2. ERE WIE ERE TOEKOMT.

Toen ik op 2 september 1979 aan dit onderzoek begon, heb ik eerst opgezocht waar NMR een afkorting van is, en hoe een virus in elkaar zit. Dit leerproces heeft zich daarna nog vier jaar lang voortgezet. Het is aan de voortreffelijke samenwerking tussen de vakgroepen Virologie en Moleculaire Fysica te danken dat dit proefschrift het multidiciplinaire karakter heeft gekregen dat U zowel in de titel als in de inhoud terugvindt. Steeds zijn fysische technieken te hulp geroepen voor het beantwoorden van virologische vragen waarop andere technieken geen antwoord konden geven. Dat hiervoor af en toe de "black-box benadering" gekozen moest worden is door de uitermate plezierige samenwerking nooit een bezwaar geweest. Daar waar nodig werden de fysische resultaten virologisch ondersteund.

Alhoewel slechts mijn naam op de omslag van dit proefschrift staat, is een groot aantal mensen betrokken geweest bij de voorbereiding en tot stand koming.

Tjeerd Schaafsma, Marcus Hemminga en Dick Verduin ben ik zeer veel dank verschuldigd voor alle stimulerende discussies over en de correctie van zowel de over dit werk gepubliceerde artikelen als dit proefschrift. Als de aan de Landbouwhogeschool geldende regels van het promotie reglement dit zouden hebben toegelaten, zou mijn dankbaarheid tegenover Dick ook nog op een formelere manier geuit zijn.

Jan de Wit en Adrie de Jager hebben mij wegwijs gemaakt in de NMR wereld. Van Jan heb ik geleerd wat je met NMR (niet) kan doen en van Adrie hoe dat dan in de praktijk uitgevoerd moet worden.

Dick Verduin, Hanke Bloksma en Henk Hilhorst hebben mij steeds voorzien van veel gezuiverd virus en eiwit. Dick en Hanke hebben bovendien vier jaar lang al mijn preparaten gecontroleerd en gekarakteriseerd. Alle in dit proefschrift beschreven "centrifuge experimenten" zijn door Dick en Hanke uitgevoerd. Voor de samenwerking die hieraan ten grondslag lag, gaf de vakgroep Virologie in de persoon van professor J.P.H. v.d. Want, steeds bemoedigende steun.

Bij het gebruiken van de diverse NMR-spectrometers was de steun van Klaas Hallenga (270 MHz; VU Brussel), Joost Lohman (250 MHz; Bruker spectrospin BV Wormer), Klaas Dijkstra (360 MHz; SON/ZWO faciliteit Groningen), Pieter van Dael (200 en 500 MHz; SON/ZWO faciliteit Nijmegen) en Adrie de Jager (100 en 300 MHz; NMR faciliteit LHW) onmisbaar. Zonder Cees Haasnoot zou dit proefschrift geen 2D-NMR spectra hebben bevat.

Hugo Jongejan heeft geholpen bij de koolstof 13 bepalingen.

I like to thank professor M.G. Rossmann, professor S.C. Harrison and professor B. Strandberg for sending me pictures of respectively SBMV, TBSV and STNV.

De samenwerking met de vakgroep informatica, o.l.v. professor Elzas, is steeds erg plezierig geweest en heeft positief bijgedragen aan de tot stand koming van hoofdstuk 8 van dit proefschrift.

Dr. A. Heerschap, professor C.W. Hilbers en professor J.M. van Boom stelden oligo-nucleotiden beschikbaar.

A. van Baaren verzorgde op voortreffelijke wijze het fotowerk.

Het typewerk is keurig fotoready gemaakt door S. van Dijk + J.S. Key.

Verder gaat mijn dank om vele uiteenlopende redenen uit naar: mijn ouders, W. F. Pieters, Han Schilthuis, Mirjam Gerritsen, Victor Roos, U, het ACCU, de vakgroep Informatica, G.Looijen, Gerrit v.d. Velde, de R & Z V Naarden, H.A. Scheraga, alle medewerkers van het RCLH, De instrumentmakerij van de Gecombineerde Diensten Binnenhaven, de werkplaats van de vakgroep Fysische en Kolloid Chemie, al mijn collega's op Moleculaire Fysica, de glasblazerij van de vakgroep Organische Chemie, Gert van Eck, Cees de Groot en vele anderen.

Dit onderzoek werd financieel mogelijk gemaakt door de Stichting voor Biofysica. Tevens werd veelvuldig gebruik gemaakt van de NMR faciliteiten te Groningen en Nijmegen die financieel worden gesteund door de Stichting Scheikundig Onderzoek Nederland (SON), en van de door deze stichting gesubsidieerde accessoires bij de NMR faciliteit van Landbouwhogeschool.

1 GENERAL INTRODUCTION

1.1 QUESTIONS

A characteristic feature of viruses is that they are not capable to reproduce independently. Plant viruses need a plant as a host for reproduction. The genome of small viruses is only sufficient to code for one or a few different coat proteins. To build large structures, multiple copies of identical subunits are needed. Many plant viruses are therefore composed of nucleic acid encapsidated by a large number of identical protein subunits. Knowing these properties the following intriguing questions arise:

- How does virus penetrate the cell?
- How and where does disassembly occur?
- How does replication take place?
- How and where does virus assembly occur?

It is now almost 50 years ago that the first virus was isolated (1). In my opinion another 50 years will be needed to answer these questions. Even some fundamental molecular principles of the basic steps in the reproduction pathway of the virus are yet not fully understood. Only a thorough knowledge of the fundamentals of protein-protein, protein-nucleic acid and protein-membrane interactions can provide a basis for understanding the molecular mechanisms involved in virus reproduction. Magnetic resonance methods can contribute to the answering of some of these questions at the molecular level.

The X-ray structure of the regular part of tobacco mosaic virus (TMV) and its protein disk are known up to a resolution of 0.28 and 0.4 nm respectively (2,3). From combined electron microscopy and RNA binding studies a hypothesis for the assembly process of this virus has been developed (4). From nuclear magnetic resonance (NMR) studies it has been concluded that the TMV coat protein contains an internally mobile region which immobilizes upon interaction of the coat protein with the RNA (5,6). It has been suggested that this immobilizing protein part is the so called V-helix (6), but the evidence presented is far from convincing. Nevertheless, these studies indicate that internal mobility has some functional significance in the process of virus assembly. Up to now, TMV has been the only virus for which a detailed model for the assembly process has been proposed.

This thesis focusses on cowpea chlorotic mottle virus (CCMV), a spherical plant virus. This virus was chosen as the main subject of the investigations, since its coat protein has the property of forming well-defined aggregates in which either protein-protein or protein-RNA interactions can be studied (7). These aggregates can be produced in quantities sufficient for magnetic resonance studies.

The work described in this thesis has been initiated by the following question:

"Which dynamical and structural aspects of the coat protein are of importance for the molecular interactions during the assembly of spherical plant viruses?"

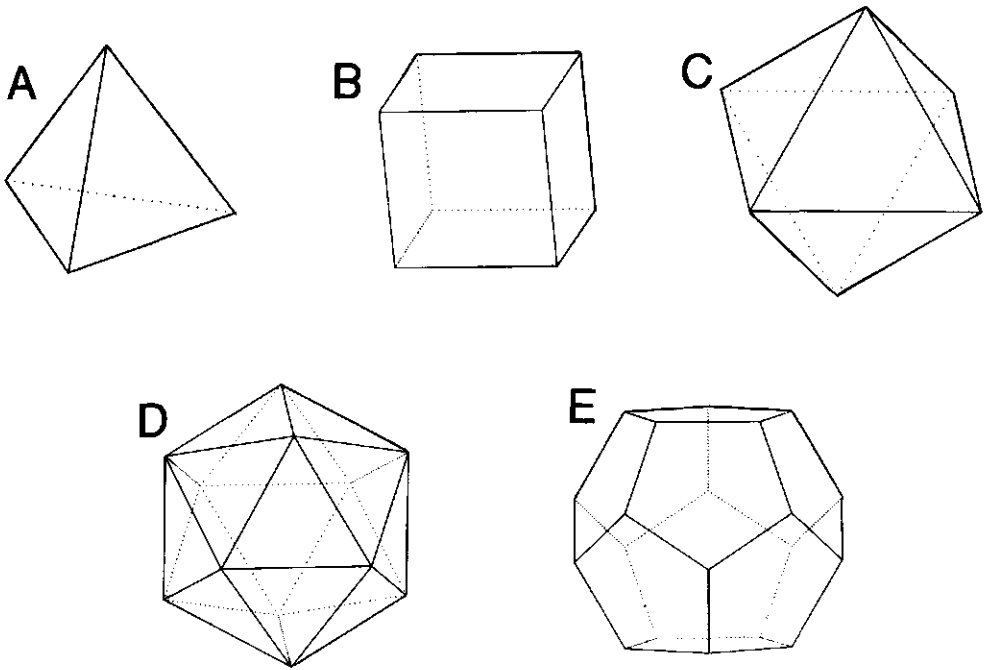


Figure 1.1. The five regular polyhedra:

A: Tetrahedron.

B: Hexahedron (cube).

C: Octahedron.

D: Icosahedron.

E: Dodecahedron.

The answer is in part obtained by the following techniques and methods: proton NMR, carbon 13 NMR, EPR, saturation transfer EPR, protein secondary structure predictions and protein energy calculations. The results of these studies have resulted in a rather detailed model for the assembly of CCMV that takes into account the observed structural and dynamical aspects of the virus protein.

The remainder of this chapter provides a general introduction into spherical plant viruses and the physical techniques and calculational methods which have been used in this thesis. A model for the assembly of CCMV is given in chapter 5, whereas the evidence for this model is provided by the chapters 2 to 9.

1.2 THE ICOSAHEDRON

Small spherical plant viruses consist of nucleic acid surrounded by a large number of identical protein subunits. Sometimes this protein coat consists of two or more different types of protein. In these cases clusters of the different proteins can be considered as identical structural units. In general, there are only a few capsid structures possible when only identical protein subunits are involved. These structures have tetrahedral, octahedral, icosahedral, cubic or dodecahedral symmetry (see Fig. 1.1). There must be one of these symmetries in the virus protein coat, since the size of the nucleic acid of small viruses is only sufficient to encode for one or a few different coat proteins. Caspar and Klug (8) predicted that almost all simple spherical viruses will possess icosahedral symmetry. Building an icosahedron requires 60 equivalent protein subunits. However, a large number of virus capsids is built from multiples of 60 identical proteins. Caspar and Klug suggested therefore a certain flexibility in the coat protein subunit. This flexibility enables the subunits to assume several different modes of interaction to their neighbours and gives rise to the phenomenon of "quasi-equivalence".

From the quasi-equivalence model for a capsid it follows that the multiplicity factor for the number of protein subunits per face of an icosahedron, the so called triangulation number T , can only have distinct values (see figure 1.2). T is given by the equation $T = Pf^2$ where P can be any number of the series 1,3,7, 13,19,21,31,..... ($=h^2+hxk+k^2$ for all pairs of integers h and k having no common factor) and f is any integer. The number of quasi-equivalent protein subunits in the coat is $60 T$. There are only a few exceptions to this rule (e.g. ref. 9). After the elucidation of the high resolution X-ray structure of several plant viruses, it was seen that larger differences between the protein subunits were observed than could be expected from the quasi-equivalence theory. Although this theory therefore had to be revisited (10), the major principles still hold for virus structures.

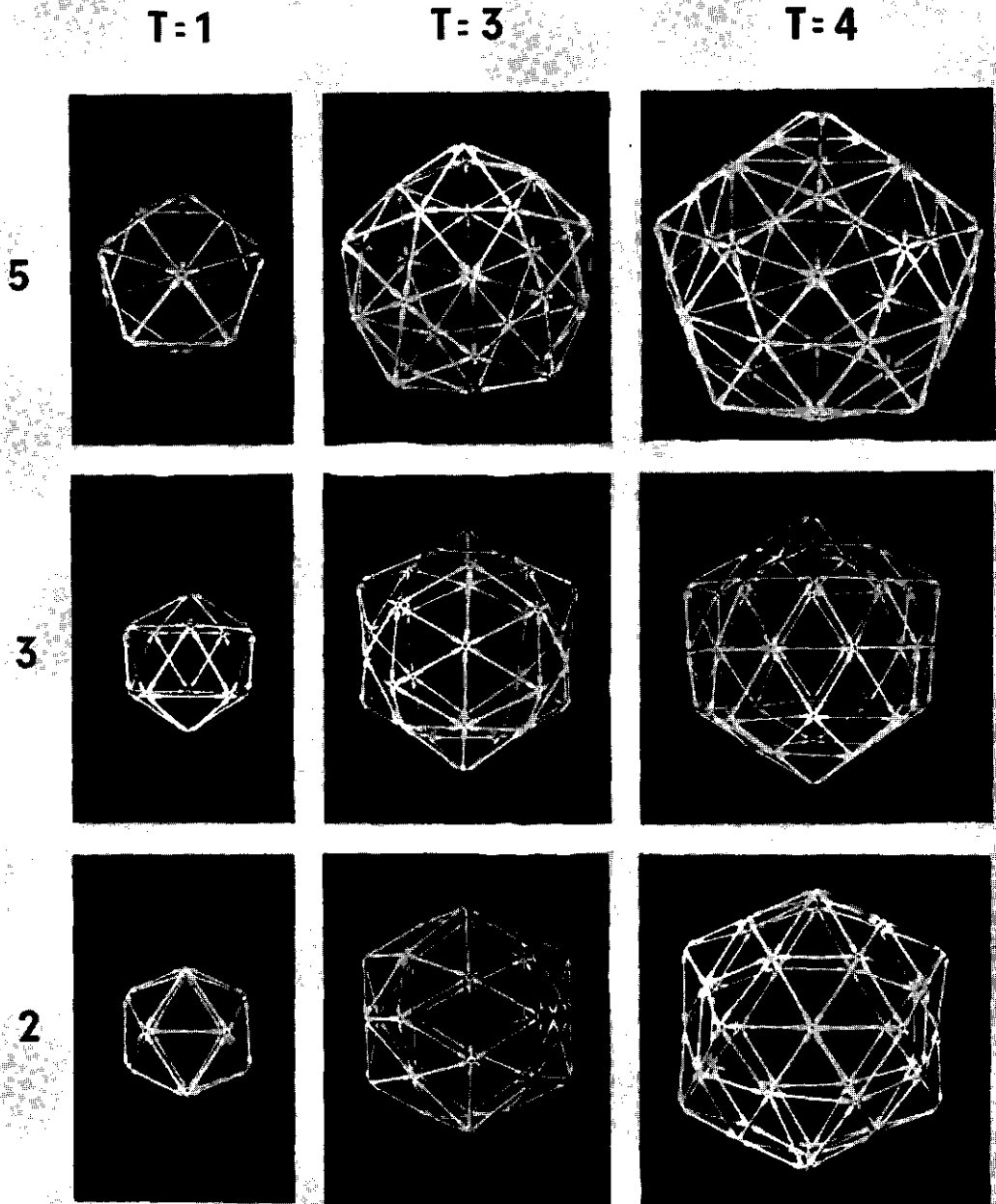


Figure 1.2. $T=1$, 3 and 4 icosahedrons, viewed almost along the 2, 3 and 5 fold icosahedral symmetry axes. The triangulation number T represents the number of triangles per icosahedral face. $T=1$ corresponds to the icosahedron of figure 1.1D. (From (7)).

1.3 STRUCTURAL ASPECTS OF SPHERICAL PLANT VIRUSES

A T=3 icosahedral virus coat contains 60 copies of each of the three quasi-equivalent protein subunits. These three types are called A-, B- and C- subunits (11). The A- subunits pack around a 5-fold axis. B- and C- subunits pack alternately around 3-fold axes (see Fig. 1.3).

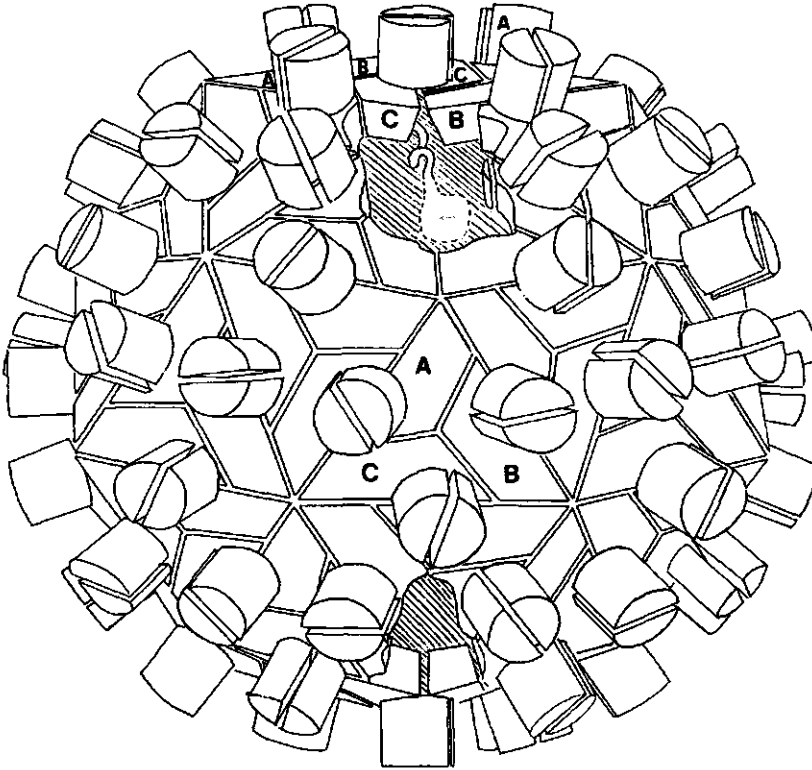
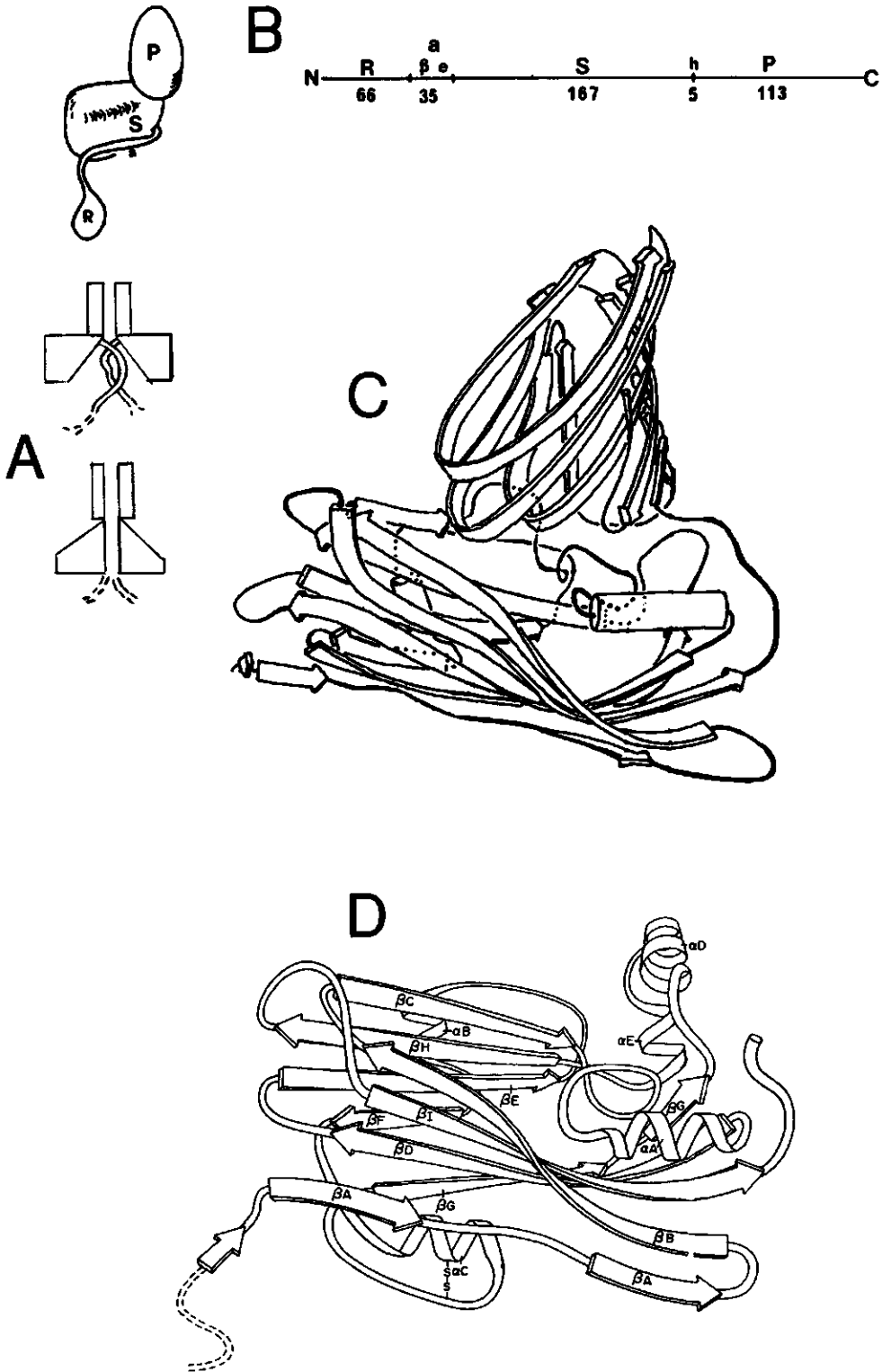


Figure 1.3. Architecture of the TBSV particle shell representative for T=3 icosahedral viruses. Arrangements of the subunits in the particle. A, B and C denote distinguishable packing environments for the subunit. S-domains (see figure 1.4) of A-subunits pack around five fold axes; S-domains of B and C alternate around three fold axes. P-domains of C subunits are paired. P-domains of A and B pair across quasi two-fold axis. The existence of a P-domain seems to be characteristic for TBSV.

At this moment there are only three spherical plant viruses for which the high resolution X-ray structure is solved (11,12,13). Therefore structural data obtained for CCMV can only be compared with these three viruses.

In tomato bushy stunt virus (TBSV) it was found that the protein contains an N-terminal region with a different structural packing in the three types of



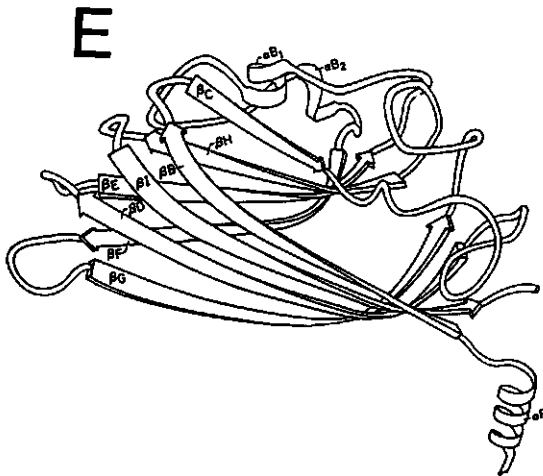


Figure 1.4. A: schematical representation of respectively a C-C dimer (upper) and a A-B dimer (lower). C-C dimers have a different hinge position, and part of the arm, designated β, ϵ in figure 1.4B, is folded in an ordered way along the bottom of the S-domain. In A-B dimers no part of this arm is ordered.

B: schematical representation of the subunit; left a sterical localization; right the linear localization of the parts of the protein. The P-domain is only seen in TBSV. The β, ϵ domain is absent in STNV.

C: Schematical representation of TBSV.

D: Schematical representation of SBMV.

E: Schematical representation of STNV.

The figures are drawn at different scale.

subunits (see figure 1.4A). The N-terminal regions of the A- and B- subunits can not be seen in the electron density map, whereas part of the N-terminal arm of the C- subunits is structured and forms links with the N-terminal part of other C-subunits. Each group of three C- subunits packed around a 3-fold axis had part of their N-terminal region clustered in a so-called ' β -annulus' (11).

The structure of southern bean mosaic virus (SBMV) (12) was found to possess a tertiary and quaternary structure similar to that of the TBSV shell domain (11) (see figure 1.4). The N-terminal protein regions of SBMV show the same order vs. disorder distribution as TBSV but the β -annulus structure observed in TBSV is not seen in the same way in SBMV. In SBMV a β -sheet is present in the N-terminal region of every C-subunit.

Satellite tobacco necrosis virus (STNV) is a T=1 virus, which implies 60 fully equivalent subunits in the protein envelope. The structure of the protein

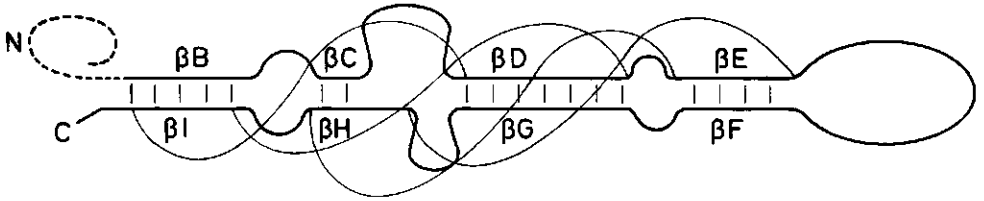


Figure 1.5. " β -role".

Schematical representation of the arrangement of the eight β -sheets B to I which are common for the three viruses in figure 1.4. The inserted loops are of various lengths and structures.

subunit of STNV is very similar to the shell domains of TBSV and SBMV (13). All three viruses have the same " β -role" topology (see figure 1.5) and contain two four-stranded antiparallel and twisted β -sheets. There are more similarities between the STNV protein and the coat protein of TBSV and SBMV (see figure 1.4).

The most striking difference lies in the N-terminal region. Since STNV is a T=1 virus, no differences between the subunits would be expected, and all N-terminal regions should in principle be equal. However, in STNV the first 11 to 13 amino acids are invisible in the electron density map, probably because they are disordered, resulting from binding to RNA with several non-equivalent local structures (13). The residues 14 to 24 adopt an α -helical configuration at a position close to a three-fold particle axis. The three helices from neighbouring subunits around this axis are held together by hydrophobic interactions of sidechains (isoleucine and leucine) pointing towards this axis. The basic residues in this helix, i.e. histidine, lysine and arginine, are exposed to the RNA.

1.4 PROTEIN-RNA INTERACTION

The molecular basis for the specificity of protein-RNA interaction in simple plant viruses is not yet solved. This interaction has been observed to be non-specific *in vitro* for CCMV (7). *In vivo* however, the assembly process is probably highly specific as can be concluded from the fact that purified virus only contains viral RNA. Unfortunately the assembly is too fast to study the protein-RNA interactions during this process by NMR or (ST-) ESR. These two techniques allow us only to look at stages in the assembly process that can be kept constant for at least several minutes. The *in vitro* assembly of CCMV is almost instantaneously after adding RNA to protein dimers.

For TYMV (14), SBMV (15) and STNV (13) it has been shown that many amino acids throughout the whole primary structure are in close proximity to, or even bind to the RNA. Assembly experiments on intact and tryptic digested CCMV protein demonstrate that the N-terminal protein region is the major RNA binding protein part (16). Since this arm is highly basic (it contains 6 arginine and 3 lysine residues), it has been suggested that interactions between positively charged amino acid side chains and negatively charged RNA phosphate groups are the driving force for protein-RNA interaction during the assembly of CCMV (17,18).

For STNV the N-terminal region is found to form partly an α -helix buried in the RNA. Argos carried out the secondary structure predictions of the N-terminal regions of several viruses (19), using the joint prediction method (20). He found a strong tendency for α -helix formation in the N-terminal region of the coat proteins of these viruses. Although SBMV has a basic N-terminal arm, it does not have the systematic separation of basic and hydrophobic residues to opposite sides of an α -helix that was seen for many other viruses. For TBSV the primary structure is not yet solved. The pronounced penetration of the N-terminal arm into the RNA (11) indicates that also for TBSV this arm is important for protein-RNA interaction.

Small synthetic oligo-nucleotides were used in this thesis to simulate an intermediate state during the assembly of CCMV. Interactions between arginine and lysine, and these oligo-nucleotides are indeed observed. Also it is concluded that the presence of secondary structure in the RNA is important for interaction.

1.5 PROTEIN-PROTEIN INTERACTIONS

Kaper (21) arranged the simple plant viruses according to their dominating stabilizing interactions by fitting their characteristics to recognition profiles of TYMV-like and CMV-like viruses (see figure 1.6). From this list it can be seen that protein-protein interactions are of small importance for the stability of the CCMV particle.

Protein-protein interactions arise from hydrophobic and hydrophylic interactions. Carboxyl cages presumably give rise to strong hydrophylic pH dependent interactions (22). Divalent cations are observed to be present at the protein subunit interfaces (23). Variations of pH, salt concentration, divalent cations and temperature can be used to change the radius of the CCMV particle or the radial distribution of protein and RNA (24). These factors have been found to influence protein-RNA salt linkages and carboxyl cages (25).

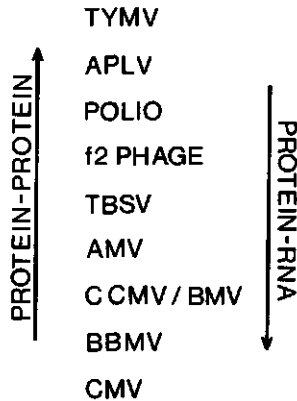


Figure 1.6. Arrangement of several simple viruses according to their dominating stabilizing interactions by fitting their characteristics to recognition profiles of TYMV-like (protein-protein stabilized) and CMV-like (protein-RNA stabilized) viruses (21).

TYMV: Turnip yellow mosaic virus.

APLV: Adee potato latent virus.

TBSV: Tomato bushy stunt virus.

AMV : Alfalfa mosaic virus.

BMV : Broam mosaic virus.

CCMV: Cowpea chlorotic mottle virus.

BBMV: Broad bean mottle virus.

CMV : Cucumber mosaic virus.

1.6 COWPEA CHLOROTIC MOTTLE VIRUS

CCMV is a member of the group of bromoviruses. This virus group is characterised by the multicomponent nature of the genome, divided over three different nucleoprotein particles (26). The encapsidation scheme for the four RNA molecules in the three particles of CCMV is shown in figure 1.7 (7). Except for the obvious differences in overall molecular mass and buoyant density, no physico-chemical differences between these three particles have been observed so far. Biological activity requires the presence of RNA -1, -2 and -3.

CCMV is stable around pH 5.0 and sediments around 88 S (7). Increasing the pH at low ionic strength ($\mu\text{v}0.2$) causes swelling of the virus (7) (78 S) and makes it sensitive to RNases and proteases (27). Full reversibility of this swelling process is only obtained in the presence of divalent cations (i.e. Mg^{2+}) (28).

Raising the pH at increased salt concentration ($\mu > 0.3$) leads to dissociation of the virus into protein dimers and RNA, almost free of protein (29). The isolated and separated constituting components of CCMV can be reassembled in vitro. The dependency of CCMV on pH and salt concentration is shown in figure 1.8 (7).

Depending on pH, salt concentration and temperature, the protein dimers can be assembled into several well-defined quaternary structures. Especially the formation of empty protein shells is of great interest, as it indicates that the information for the formation of icosahedral T=3 particles is contained in the protein subunit.

The changes in structure of CCMV during titration are solely due to protein effects since almost no conformational changes are observed in the RNA (30). Some specific RNA base interactions are eliminated upon swelling (30). In vitro reassembly experiments showed no specificity for RNA encapsidation. CCMV protein can not only encapsidate CCMV-RNA, but also other RNA molecules, or even polyanions like polyvinyl sulphate or sodium dextran sulphate (7). In vivo however, the assembly is probably highly specific.

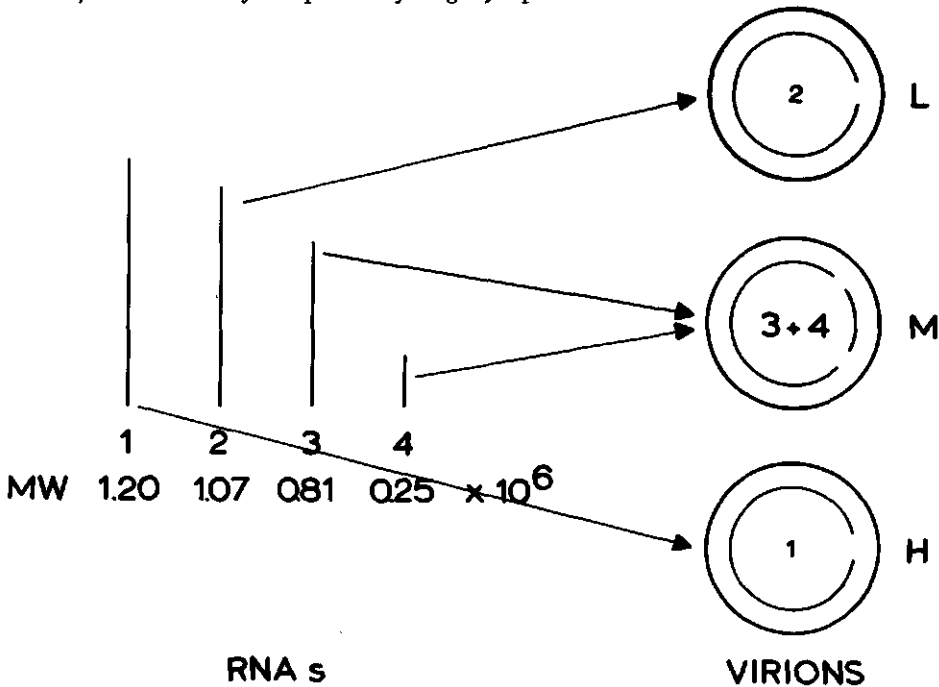


Figure 1.7. The distribution of CCMV RNA molecules over virus particles of different buoyant density. Three types of nucleo-protein particles are present, called light (L), medium (M) and heavy (H) according to their density in CsCl. These difference in buoyant density is the only observable difference between the particles.

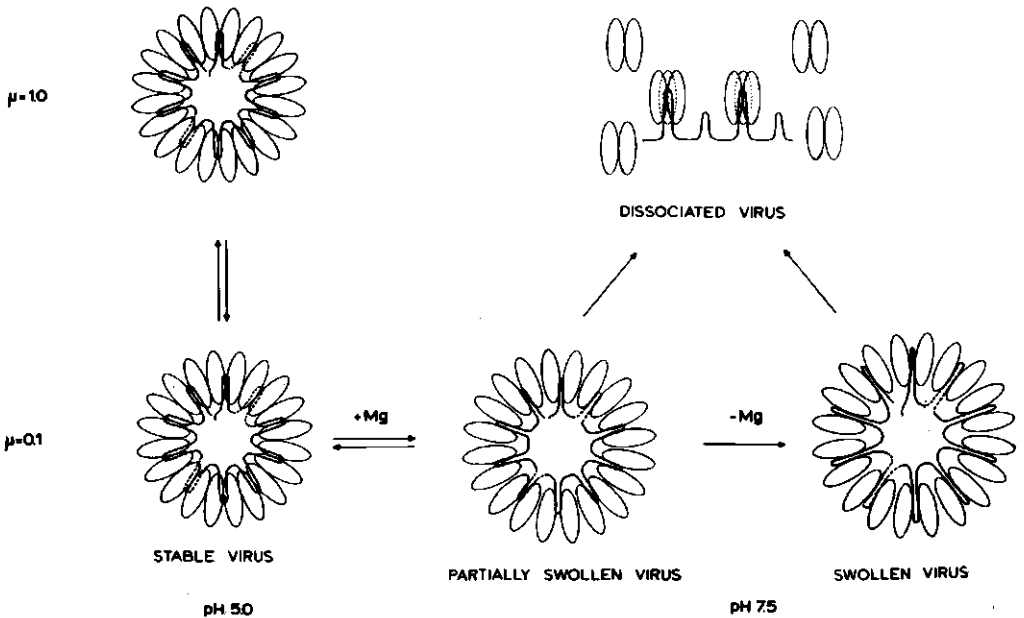


Figure 1.8. A schematical representation of the swelling and dissociation of CCMV. The ionic strength (μ) and the pH are varied respectively in the vertical and horizontal direction. At low ionic strength and pH 7.5, the influence of Mg^{2+} has been shown.

The dotted ellipses represent protein subunits bound to high-affinity sites on the RNA, observed with dissociated virus (From (7)).

1.7 NUCLEAR MAGNETIC RESONANCE

Nuclear magnetic resonance is at present the best possible technique to study the mobility of molecules or parts of molecules. Recent technical developments have made it possible to study a wide variety of nuclei in all kinds of molecular environments. The major advantage of NMR over a number of other physical techniques is the fact that NMR is non-destructive and non-invasive. In principle, it is possible to study every molecule containing nuclei with a non zero spin, in vitro as well as in vivo. Unfortunately there are also some drawbacks. The first problem is the relative low sensitivity of the NMR method. Therefore it is often necessary to have large amounts of highly concentrated material or use long measuring times. In the case of low abundant nuclei, enrichment can help to solve this problem. A second important problem is the

assignment of the numerous resonances that often appear in a spectrum of a large biomolecule. In the case of CCMV protein there are so many close lying resonances, that an unambiguous interpretation of the results is only partly possible.

For ^1H -NMR magnetic fields > 10 T are already routinely available. For other nuclei, technical problems have prevented up to now the application of these high magnetic fields for aqueous solution samples. New pulse sequences (31) and probe designs undoubtedly will solve the technical problems in the near future.

Measurement of spin-lattice (T_1) and spin-spin (T_2) relaxation times and nuclear overhauser effects (NOE) provide a way to study molecular motions (32). The theory describing the dependence of T_1 , T_2 and NOE upon isotropic molecular motion of rigid molecules has been well developed (32). For molecules with many modes of internal mobility a satisfactory interpretation method is not yet available. Several methods have been presented using combinations of two or more dependent or independent degrees of internal motional freedom (33), but these models are too crude to reflect internal motions in proteins. For most cases it is possible to use a rigid sphere model for interpreting T_1 , T_2 and NOE values, that gives qualitative information about the timescale of internal molecular motion. This procedure has been used throughout this thesis.

For a rigid sphere the following equations for the ^{13}C T_1 and T_2 values and the ^1H - ^{13}C nuclear overhauser effects can be derived (32):

$$\frac{1}{T_1} = \frac{h^2 \gamma_C^2 \gamma_H^2}{10r^6} \left\{ \frac{\tau_c}{1+(\omega_H - \omega_C)^2 \tau_c^2} + \frac{3\tau_c}{1+\omega_C^2 \tau_c^2} + \frac{6\tau_c}{1+(\omega_H + \omega_C)^2 \tau_c^2} \right\} \quad (1.1)$$

$$\frac{1}{T_2} = \frac{h^2 \gamma_C^2 \gamma_H^2}{20r^6} \left\{ 4\tau_c + \frac{\tau_c}{1+(\omega_H - \omega_C)^2 \tau_c^2} + \frac{3\tau_c}{1+\omega_C^2 \tau_c^2} + \frac{6\tau_c}{1+\omega_H^2 \tau_c^2} + \frac{6\tau_c}{1+(\omega_H + \omega_C)^2 \tau_c^2} \right\} \quad (1.2)$$

$$\text{NOE} = 1 + \frac{\gamma_H}{\gamma_C} \left\{ \frac{\frac{6\tau_c}{1+(\omega_H + \omega_C)^2 \tau_c^2} - \frac{\tau_c}{1+(\omega_H - \omega_C)^2 \tau_c^2}}{\frac{6\tau_c}{1+(\omega_H + \omega_C)^2 \tau_c^2} + \frac{3\tau_c}{1+\omega_C^2 \tau_c^2} + \frac{\tau_c}{1+(\omega_H - \omega_C)^2 \tau_c^2}} \right\} \quad (1.3)$$

Here it has been assumed that the relaxation process is dominated by dipole-dipole interactions. The rotational correlation time T_c is given by the Stokes-Einstein equation:

$$\tau_c = \frac{4\pi R_o^3 \eta}{3kT}$$

Where η is the viscosity and R_o is the radius of the particle. Using the equations for T_2 and τ_c , one can derive the following rule of thumb:

Line width \times 1/1000 \times Molecular mass

In CCMV protein the observed internal mobilities can be roughly divided in two classes: $\tau_c \sim 10^{-8}$ s and $\tau_c > 10^{-6}$ s. The former is observed for the CCMV protein dimer and the N-terminal protein region in the empty capsids and a few amino acid side chains in the protein envelope. The latter is observed for the major part of the protein, the RNA and the N-terminal arm in the presence of RNA.

1.8 ELECTRON SPIN RESONANCE

Electron paramagnetic resonance (EPR) can only be applied to molecules containing an unpaired electron. Most biomolecules do not contain a natural stable radical, which implies that a spin-label must be introduced before EPR spectra can be recorded. A comparison of the size of a spin-label with, for example, an amino acid or a nucleotide shows that the introduction of a spin-label in a biomolecule may cause a severe distortion of the native structure. Great care should therefore be taken when detailed conclusions are drawn from studies on spin-labelled biomolecules.

EPR is a sensitive tool when mobilities on the 10^{-7} to 10^{-12} s timescale are studied. The saturation transfer EPR (ST-EPR) technique is sensitive from 10^{-3} to 10^{-7} seconds (34). The theory behind EPR spectroscopy is well understood and EPR spectra reflecting all sorts of isotropic and anisotropic motions can be simulated in detail.

Serious difficulties are encountered with the interpretation of ST-EPR spectra. In cases of isotropic Brownian diffusion, reference spectra recorded under the same experimental conditions can be used to derive the rotational correlation time by means of comparison. In the presence of anisotropic motion, great care should be taken when making such comparisons, since anisotropic

motion is not included in the reference spectra. In this thesis an (ST-) EPR study on maleimide labelled CCMV will be presented. From this study it will be concluded that the spin-label is located on the protein in a cleft and performs a local anisotropic motion in this cleft.

1.9 SECONDARY STRUCTURE PREDICTIONS

The X-ray structure of CCMV and its protein are not known at this moment. Therefore, the only way to relate the NMR results presented in this thesis to the structure of CCMV is by assuming that its structure is related to the three spherical viruses for which the X-ray structure is elucidated (11-13). Because the N-terminal arm of the coat protein of CCMV is the only part of the virus that can be studied in detail by means of NMR, such a comparison could be fruitful, especially for the N-terminal arms. Unfortunately, X-ray diffraction is unable to resolve the complete structure of the N-terminal arm of spherical plant viruses (11-13). The major part of this arm has a disordered structure, because it is bound to RNA that possesses no icosahedral symmetry. In the virus the rotational correlation time of this arm is too large to enable NMR studies.

In the absence of RNA, the high mobility of this N-terminal arm prevents the observation of resolved X-ray reflections of this part of the protein, but high resolution NMR is able to extract some information about structural and dynamical properties of the N-terminal arm. The large number of overlapping resonances, however, makes it difficult to assign these resonances to individual amino acids and to extract information on a molecular level. However, it can be concluded from the NMR results (see chapter 9) that some secondary structure is present in the N-terminal arm of CCMV in the absence of RNA, but it does not become clear from these experiments which type of structure is observed.

For this reason the structural aspects of the N-terminal arm were examined by means of secondary structure predicting methods. In this thesis three methods have been used (35-37), two of which were modified to predict the protein structure when 1) all positive charges of lysine and arginine in the N-terminal arm are neutralized (as in the whole virus), and 2) when these charges are still present as in the absence of RNA.

There are many methods to predict the secondary protein structure starting from no more than the primary structure (19,20). Some of these methods can use additional chemical information to enhance the correctness of the prediction.

These methods (38) can roughly be divided in three classes:

- 1) Methods using statistically derived structure potentials for individual amino acids or groups of amino acids.

2) Methods using structural considerations, e.g. proline is not allowed to assume an α -helix conformation and glycine is a likely candidate for turns, since the absence of a bulky side chain eliminates steric hindrance.

3) Methods implicitly using extra information or other principles than mentioned for the first two classes.

Although no method gives a better result than 50% above a random prediction (20), a better result can be obtained by combining the predictions of several methods in a so-called joint prediction procedure.

Two methods were used from the first group i.e. the methods of Chou and Fasman (35) and the method of Burgess et al. (36), and one method from the second group, i.e. the method of Lim (37). The results fully confirm the findings of Argos (19) who predicted the secondary structure of BMV, as could be expected from the great similarities between the primary structures of this virus and that of CCMV.

The interactions between negatively charged phosphate groups of the RNA and positively charged lysine and arginine side chains, which results in a neutralization of these side chains, are likely to be important in the assembly process. Therefore two data sets were used for the methods of Lim, and Chou and Fasman. In one data set, parameters were incorporated to simulate charged side chains for arginine and lysine. In the second data set, these parameters were chosen to agree with neutral side chains. From the calculations it follows that charge neutralization is an important factor when studying the structural aspects of the N-terminal arm of CCMV protein. This is treated in more detail in chapter 7.

1.10 ENERGY CALCULATIONS

Secondary structure predictions revealed a tendency towards α -helix formation for the N-terminal protein arm of CCMV (see chapter 7). Therefore a different, more sophisticated, method has been used for elucidating the structural aspects of this arm in detail. The united atom model of Scheraga et al. (39) was chosen since it provides one of the best methods up to now for calculating the energy content of a given protein conformation. In addition a computerized algorithm for the use of this model is commercially available (40). This program, named UNICEPP, takes into account the following molecular interactions:

- 1) Electrostatic interactions,
- 2) Non bonded (Van der Waals) interactions,
- 3) Hydrogen bond energy,
- 4) Torsional energy,

- 5) Cystine loop closing energy,
- 6) Internal conformational energy of proline residues.

For obvious reasons, only the terms 1 to 4 contribute to the energy of the N-terminal protein arm of CCMV.

There are three major sources for errors, when this method is used. The main disadvantage is the complete neglect of interaction with solvent molecules. As has been concluded from molecular dynamics studies, the contribution of solvent molecules to the dynamic behaviour is more important than the contribution to the time averaged structure (41). In this thesis, there is only interest in a comparison of conformations at minimum energy. It is expected that for such a comparison the effect of water is smaller than for a molecular dynamics study. A second source of errors is inherent to the method itself. Most of the interaction parameters cannot be verified experimentally. Also these parameters can not be varied when the environmental conditions change during the process of energy minimization. A radial gradient for the dielectric constant (42) would probably be a much better model for a peptide in solution than the constant value of 2.0 chosen by Scheraga et al. for the UNICEPP program (39). The last major theoretical disadvantage is that in the calculations only one isolated peptide chain is considered, whereas in empty capsids the presence of the protein core and N-terminal arms of neighbouring subunits is felt. The main practical problem is the large amount of computer time required.

Chapter 8 describes a study on the structural aspects of the N-terminal protein arm of CCMV. The results obtained show that the charges in this N-terminal arm play an important role. Although many approximations are made in the approach chosen here, the results are clear enough to give insight in the structural aspects of this protein arm. The calculations described are the result of almost 500 hours of computer time.

1.11 REFERENCES

- (1) Stanley, W.M., (1936), *Phytopathology* 26, 305-307.
- (2) Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R., Klug, A., (1978), *Nature* 276, 362-368.
- (3) Stubbs, G., Warren, S., Holmes, K., (1977), *Nature* 267, 216-221.
Mandelkov, E., Stubbs, G., Warren, S., (1981), *J. Mol. Biol.* 152, 375-386.
- (4) See for a review:
Klug, A., *The Harvey Lectures*, series 74, 141-171.

- (5) Wit, J.L.de, Alma, N.C.M., Hemminga, M.A., Schaafsma, T.J., (1979), *Biochemistry* 18, 3973-3976.
Wit, J.L.de, Hemminga, M.A., Schaafsma, T.J., (1978), *J.Mag.Res.* 31, 97-107.
- (6) Jardetzky, O., Akasaka, K., Vogel, D., Morris, S., Holmes, K.C., (1978), *Nature* 273, 564-566.
- (7) Verduin, B.J.M., (1978), Thesis, Agricultural University Wageningen.
- (8) Caspar, D.L.D., Klug, A., (1962), *Symp. on Quant. Biol.* 27, 1-24. Cold Spring Harbor Laboratory of Quantitative Biology.
- (9) Rayment, I., Baker, T.S., Caspar, D.L.D., Murukami, W.T., (1982), *Nature* 295, 110-115.
- (10) Caspar, D.L.D., (1980), *Biophys.J.* 32, 103-138.
Harrison, S.C., (1980), *Biophys.J.* 32, 139-153.
- (11) Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, K.K., Bricogne, G., (1978), *Nature* 276, 368-373.
Harrison, S.C., (1980), *Biophys.J.* 10, 139-154.
- (12) Hermodson, M.A., Abad-Zapatero, C., Abdel-Meguid, S.S., Pundak, S., Rossmann, M.G., Tremain, J.H., (1982), *Virology* 119, 113-149.
Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., Leslie, A.G.W., Rayment, I., Rossmann, M.G., Suck, D., Tsukihara, T., (1980), *Nature* 286, 33-39.
- (13) Liljas, L., Unge, T., Jones, T.A., Fridborg, K., Lovgren, S., Skoglund, U., Strandberg, B., (1982), *J.Mol.Biol.* 159, 93-108.
Rossmann, M.G., Abad-Zapatero, C., Murthy, M.R.N., Liljas, L., Jones, T.A., Strandberg, B., (1983), *J.Mol.Biol.* 165, 711-736.
- (14) Ehresmann, B., Briand, J.P., Reinbolt, J., Witz, J., (1980), *Eur.J.Biochem.* 108, 123-129.
- (15) Rossmann, M.G., Abad-Zapatero, C., Hermodson, M.A., Erickson, J.W., (1983), *J.Mol.Biol.* 166, 37-83.
- (16) Vriend, G., Hemminga, M.A., Verduin, B.J.M., Wit, J.L.de, Schaafsma, T.J., (1981), *FEBS Lt.* 134, 167-171.
- (17) Vriend, G., Verduin, B.J.M., Hemminga, M.A., Schaafsma, T.J., (1982), *FEBS Lt.* 145, 49-52.
- (18) Kaper, J.M., (1975), *The Chemical Basis of Virus Structure, Dissociation and Reassembly*, North Holland Publishing Company, Amsterdam Oxford, Chapter 7.
Chidlow, J., Tremain, J.H., (1971), *Virology* 43, 267-278.
- (19) Argos, P., (1981), *Virology* 110, 55-62.
- (20) Argos, P., Schwartz, J., Schwartz, J., (1976), *B.B.A.* 439, 261-273.
Lenstra, J.A., (1977), *B.B.A.* 491, 333-338.

- (21) Kaper, J.M., (1975), *The Chemical Basis of Virus Structure, Dissociation and Reassembly*, North Holland Publishing Company, Amsterdam Oxford, 382-386.
- (22) Kaper, J.M., (1975), *The Chemical Basis of Virus Structure, Dissociation and Reassembly*, North Holland Publishing Company, Amsterdam Oxford, 328-330.
- Caspar, D.L.D., (1963), *Advan. Protein Chem.* 18, 37-121.
- (23) Bancroft, J.B., (1970), *Adv. Virus Res.* 16-136.
- Hzu, C.H. Sehgal, O.P., Pichet, E.E., (1976), *Virology* 69, 587-595.
- Abdel-Meguid, S.S., Yamane, T., Fukuyama, K., Rossmann, M.G., (1981), *Virology* 114, 81-85.
- Hull, R., (1978), *Virology* 89, 418-422.
- see also refs 10 and 12.
- (24) Zulauf, M., (1977), *J. Mol. Biol.* 114, 418-422.
- Jacrot, B., (1975), *J. Mol. Biol.* 95, 433-446.
- Pfeiffer, P., (1980), *Virology* 102, 54-61.
- (25) Kaper, J.M., (1975), *The Chemical Basis of Virus Structure, Dissociation and Reassembly*, North Holland Publishing Company, Amsterdam Oxford, Chapter 8.
- (26) Bancroft, J.B., Hiebert, E., Rees, M.W., Markham, R., (1968), *Virology* 34, 224-239.
- Bancroft, J.B., (1971), *Virology* 45, 830-834.
- (27) Pfeiffer, P., Hirth, L., (1975), *FEBS Lt.* 56, 144-148.
- (28) Chauvin, C., Pfeiffer, P., Witz, J., Jacrot, B., (1978), *Virology* 88, 138-148.
- (29) Bancroft, J.B., Hiebert, E., (1967) *Virology* 32, 354-356.
- (30) Verduin, B.J.M., Prescott, B., Thomas, G.J., (1983), *Biophys. J.* 41, abstr. 156.
- (31) Levitt, M.H., Freeman, R., (1981), *J. Mag. Res.* 43, 502-507.
- Levitt, M.H., Freeman, R., Frenkiel, T., (1982), *J. Mag. Res.* 47, 328-330.
- Levitt, M.H., Freeman, R., Frenkiel, T., (1982), *J. Mag. Res.* 50, 157-160.
- Shaha, A.J., Frenkiel, T., Freeman, R., (1983), *J. Mag. Res.* 52, 159-163.
- Shaha, A.J., Keeler, J., Frenkiel, T., Freeman, R., (1983), *J. Mag. Res.* 52, 335-338.
- (32) See any elementary Handbook on NMR eg:
- Farrar, T.C., Becker, E.D., *Pulse and Fourier Transform NMR*. Academic Press.
- Wuthrich, K., *NMR in Biological Research: Peptides and Proteins*. North Holland/ American Elsevier.

- (33) Woessner, D.E., Snowden, B.S., (1972), *Adv. Mol. Relaxation Processes* 3, 181.
Levine, Y.K., Partington, P., Roberts, G.C.K., (1973), *Mol. Phys.* 25, 497-514.
King, R., Jardetzky, O., (1978), *Chem. Phys. Lett.* 55, 15-18.
Tsutsumi, A., Chachaty, C., (1979), *Macromolecules* 12, 429-435.
- (34) Thomas, D.D., Dalton, L.R., Hyde, J.S., (1967), *J. Chem. Phys.* 65, 3006-3024.
- (35) Chou, P.Y., Fasman, G.D., (1974), *Biochemistry* 13, 211-221.
Chou, P.Y., Fasman, G.D., (1974), *Biochemistry* 13, 222-245.
- (36) Burgess, A.W., Ponnuswamy, P.K., Scheraga, H.A., (1974), *Isr. J. Chem.* 12, 239-286.
- (37) Lim, V.I., (1974), *J. Mol. Biol.* 88, 857-872.
Lim, V.I., (1974), *J. Mol. Biol.* 88, 873-894.
- (38) Some key references are:
Nagano, K., (1973), *J. Mol. Biol.* 75, 401-420.
Nagano, K., (1975), *J. Mol. Biol.* 84, 337-372.
Nagano, K., (1977), *J. Mol. Biol.* 109, 251-274.
Schulz, G.E., Barry, C.D., Friedman, J., Chou, P.Y., Fasman, G.D., Finkelstein, A.V., Lim, V.I., Ptitsyn, O.B., Kabat, E.A., Wu, T.T., Levitt, M., Robson, B., Nagano, K., (1974), *Nature* 250, 140-142.
Rose, G.D., Selzer, J.P., (1977), *J. Mol. Biol.* 113, 153-164.
See also refs. 19, 20, 35, 36 and 37.
- (39) Momany, F.A., McGuire, R.F., Burgess, A.W., Scheraga, H.A., (1975), *J. Phys. Chem.* 79, 2361-2381.
Dunfield, L., Burgess, A.W., Scheraga, H.A., (1975), *J. Phys. Chem.* 82, 2609-2616.
- (40) UNICEPP, Quantum Chemical Program Exchange Library, QCPE program nr. 361, Chemistry Department, Indiana University.
- (41) Rossky, P.J., Karplus, M., (1979), *J. Am. Chem. Soc.* 101, 1913-1937.
Gunsteren, W.F. van, Karplus, M., (1982), *Biochemistry* 21, 2259-2274.
- (42) Northrup, S.H., Pear, M.R., McCammon, J.A., Karplus, M., (1980), *Nature* 286, 304-305.
Lindner, B., (1967), *Adv. Chem. Phys.* 12, 225-282.

2 SWELLING OF COWPEA CHLOROTIC MOTTLE VIRUS STUDIED BY ^1H -NMR

2.1 INTRODUCTION

Cowpea chlorotic mottle virus is a spherical plant virus of the group of bromoviruses. An interesting feature of bromoviruses is the increase in hydrodynamic volume when the pH is increased from pH 5.0 to pH 7.5 at low ionic strength ($\mu \approx 0.2$) (1). A hysteresis around pH 6.5 is observed for the reversible proton release (2) and the Stokes' radius variation (3) during this pH-dependent swelling. At pH 5.0 the virus is not sensitive to peptidases and RNases. However, degradation of the virus occurs on adding peptidases or RNases at pH 7.5 (4,5). These effects have been interpreted in terms of rearrangements of the protein and the RNA in the virus (4,5).

In chapter 4 an NMR study is presented of CCMV and its protein (6). It is shown that the N-terminal region of CCMV protein is the RNA-binding part. This N-terminal region is very mobile in the absence of RNA. In the presence of RNA (as in native virus) the N-terminal region is immobilized.

In this chapter an NMR study of the pH dependent swelling of CCMV is presented. Only a few $-\text{CH}_2$ and $-\text{CH}_3$ groups of amino acid sidechains are found to be mobile on a timescale of 10^{-7} - 10^{-8} s. It is shown that neither the RNA nor the N-terminal protein region are mobile on this timescale at any pH during the titration. From these experiments, it is concluded that the protein-RNA interaction in CCMV is not altered significantly during the pH dependent swelling of the virus.

2.2 MATERIALS AND METHODS

The virus was purified as described by Verduin (7). After purification, the virus was dialyzed against 200 mM KCl, 10 mM MgCl_2 and 1 mM sodium phosphate (pH 5.0). H_2O in this solution was substituted by D_2O through three cycles of centrifugation and resuspension of the pellets in the above solution made up in D_2O . In D_2O solutions, pH meter readings were taken without correction for the presence of D_2O . The final virus concentration was ≈ 20 mg/ml. The pH of the solution was adjusted by adding very slowly 100 mM solutions of DCl or NaOD.

^1H -NMR spectra were recorded with a Bruker WM250 supercon spectrometer. 500 μl samples were measured at 7°C in the quadrature detection mode with D_2O lock and without ^1H -decoupling. The acquisition time was 0.41 s with a pulse delay of

0.6 s and 4000 scans were taken. The sensitivity enhancement was 10 Hz. The ppm scale was relative to sodium 2,2-dimethyl 2-silapentane-5-sulphonate (DSS). The vertical scale was corrected for (small) concentration differences. Spectral intensities were determined using a planimeter. The absolute spectral intensity at pH 5.0 was measured using spectra with a spectral width of 25 kHz. The relative spectral intensities were measured from peak areas in difference spectra at 10 kHz spectra width. The estimated error in the absolute spectral intensity is $\sim 30\%$. The error in the comparison of spectral intensities is much smaller.

2.3 RESULTS AND DISCUSSION

Since CCMV has a large molecular mass ($M \sim 4.6 \times 10^6$), the NMR spectral intensity of sharp peaks is a direct measure of the number of mobile nuclei (6,8,9). Rigid

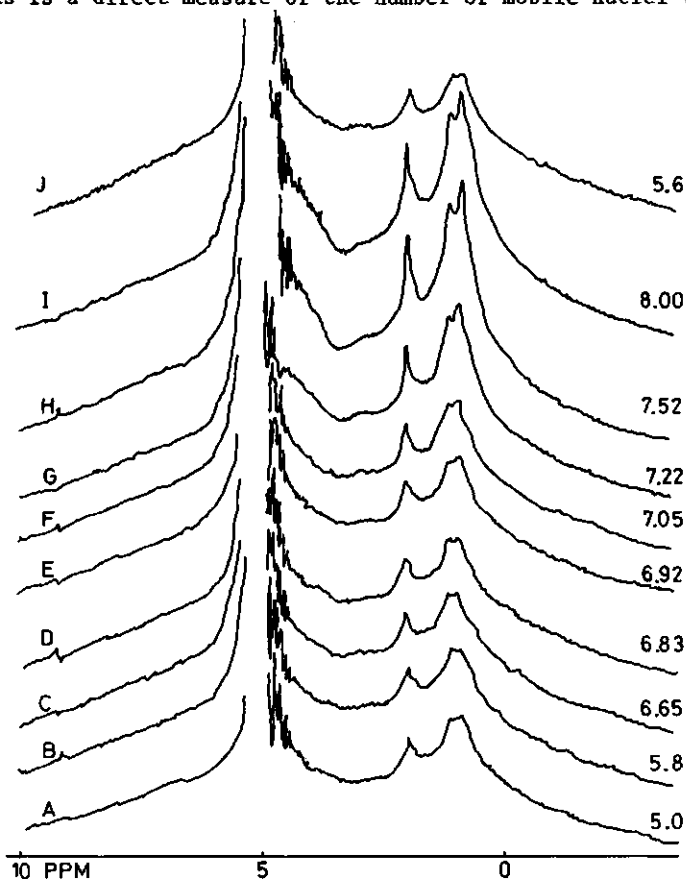


Figure 2.1. 250 MHz ^1H -NMR spectra of CCMV at several pH values between 5.0 and 8.0.

nuclei give rise to broad unresolved resonances. Figure 2.1 represents the 250 MHz $^1\text{H-NMR}$ spectrum of CCMV at several pH values between pH 5.0 and pH 8.0. More than 90% of the resonances are contained in a broad unresolved background signal, indicating that most of the protein and the RNA is immobile at the NMR timescale. The three sharp peaks in the 0.5-2.5 ppm region correspond to mobile $-\text{CH}_2$ and $-\text{CH}_3$ groups in amino acid side chains. The rotational correlation time of these groups is $\sim 10^{-8}$ s (6). Figure 2.2 shows the total spectral intensity of the sharp peaks as a function of pH. Since the T_1 values are almost equal ($T_1 \sim 0.1-0.2$ s) for all these peaks and independent of pH, an increase in spectral intensity is solely due to an increase in the number of mobile groups. Figure 2.1J is the spectrum of CCMV after back-titration from pH 8.0 to pH 5.6. This spectrum is identical to the original pH 5.8 spectrum. This proves that the observed increase in mobility is reversible. The pK value for the titration is 7.1, which is in good agreement with the results obtained by Jacrot (2).

The percentage mobile protons increases from $2.0 \pm 0.7\%$ at pH 5.0 to $8 \pm 3\%$ at pH 8.0. This mobility can solely be ascribed to $-\text{CH}_2$ and $-\text{CH}_3$ groups in amino acid side chains (11). The thickness of the protein shell is known to remain constant throughout the titration (4). The observed increase of mobility seems therefore to result from $-\text{CH}_2$ and $-\text{CH}_3$ groups of amino acids at the protein

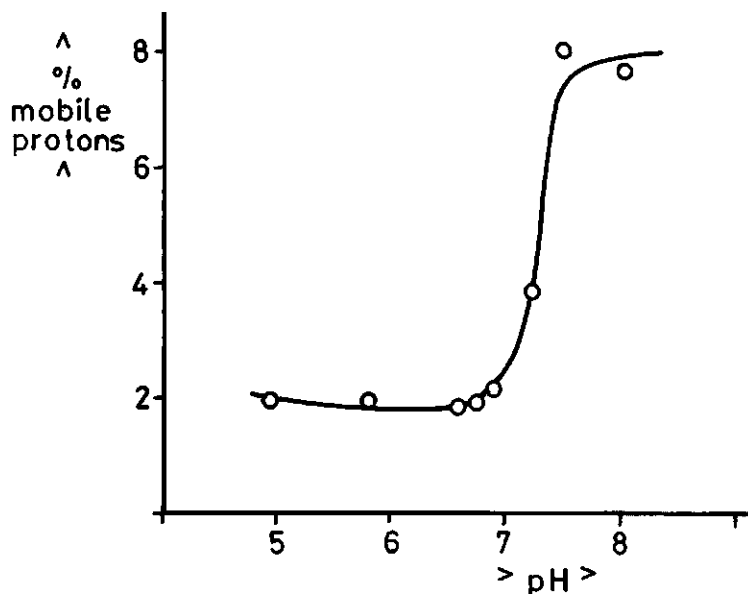


Figure 2.2. Percentage mobile protons in the 250 MHz $^1\text{H-NMR}$ spectra of figure 2.1 as a function of pH.

subunit contact faces that gain space to move when the virus protein shell is loosened by the radial expansion of the protein subunits.

RNA would give a strong resonance at 3.8 ppm. From the absence of sharp peaks at this position in the spectra, it is concluded that upon swelling no RNA is released from the virus and that no increase of mobility of RNA occurs.

In the absence of RNA, the N-terminal arm of CCMV protein gives a strong arginine signal at 3.2 ppm. From the absence of sharp peaks at 3.2 ppm in all the spectra, it is concluded that the immobility of the N-terminal protein region, due to the protein-RNA interaction, is not influenced by an increase in pH.

Although no structural information can be derived from these NMR results, they strongly suggest that the interactions between the N-terminal protein region and RNA remain unaltered during titration from pH 5.0 to pH 8.0 and back. This is in agreement with work on CCMV (10) and a similar virus BMV (4) wherein regions of RNA that interact with the protein were suggested not to reorganize during the swelling. Sensitivity of CCMV to proteases may also result from disclosure of sites hitherto hidden on the subunit contact faces, rather than from a partial unfolding of the the protein subunits themselves (4). This idea is experimentally confirmed by the NMR result presented in this chapter, since a partial unfolding of the protein subunits would give rise to NMR signals with a linewidth according to mobility much faster than observed in the spectra from figure 2.1.

The results presented in this chapter have also been published in: Vriend,G., Hemminga,M.A., Verduin,B.J.M., Schaafsma,T.J., (1982), FEBS Lt. 146, 319-321.

2.4 REFERENCES

- (1) Bancroft,J.B., Hiebert,E., (1967), Virology 32, 354-356.
- (2) Jacrot,B., (1975), J.Mol.Biol. 95, 433-446.
- (3) Zulauf,M., (1977), J.Mol.Biol. 144, 259-266.
- (4) Pfeiffer,P., (1980), Virology 102, 54-61.
- (5) Pfeiffer,P., Hirth,L., (1975), FEBS Lett. 56, 144-148.
- (6) Vriend,G., Hemminga,M.A., Verduin,B.J.M., De Wit,J.L., Schaafsma, T.J., (1981), FEBS Lett. 134, 167-171.
- (7) Verduin,B.J.M., (1978), J.Gen.Virol. 39, 131-147.

- (8) De Wit, J.L., Hemminga, M.A., Schaafsma, T.J., (1978), *J. Magn. Res.* 31, 79-107.
- (9) De Wit, J.L., Alma, N.C.M., Hemminga, M.A., Schaafsma, T.J., (1979), *Biochemistry*, 18, 3973-3976.
- (10) Adolph, K.W., (1975), *J. Gen. Virol.* 28, 137-146.
- (11) Wuthrich, K., (1976), *NMR in Biological Research: Peptides and Proteins*, Elsevier Biomedical, Amsterdam, New York.

3 EPR AND ST-EPR SPECTROSCOPY ON SPIN-LABELLED COWPEA CHLOROTIC MOTTLE VIRUS

3.1 INTRODUCTION

Cowpea chlorotic mottle virus (CCMV) is a spherical plant virus of the group of bromoviruses. It consists of RNA surrounded by 180 identical, icosahedrally arranged protein subunits (1). The virus is stable around pH 5.0. At pH 7.5 the virus is sensitive to proteases and RNases (2). At this pH an increase in ionic strength ($\mu > 0.3$) results in a dissociation of the virus into protein dimers and RNA almost free of protein (3). Each protein subunit in the virus contains two cysteine -SH groups. Krüse et al. have shown that one of these -SH groups is easily accessible for maleimide spin-labels when the virus is in its swollen state (4).

The same authors have measured electron paramagnetic resonance (EPR) and saturation transfer electron paramagnetic resonance (ST-EPR) spectra of several conformational stages of spin-labelled CCMV (SL-CCMV). To interpret their results they followed the ideas of Thomas et al. (5) and Marsh (6), and suggested an anisotropic motion of the whole spin-labelled protein subunit within the virus or an anisotropic motion of the spin-labelled part of the protein subunit in the core of the virus.

Here an EPR and ST-EPR study is presented on spin-labelled CCMV and spin-labelled CCMV treated with glutaraldehyde. From a combination of the results of these studies and previous NMR work on CCMV (7) it can only be concluded that the observed motional anisotropy in SL-CCMV is not the result of anisotropic protein mobility, but represents a local anisotropic motion of the spin-label with respect to the protein. From a comparison of our spectra with several previously published ST-EPR studies (8-10) on maleimide spin-labelled proteins it is concluded that this type of anisotropic spin-label motion may also be present in these systems, indicating that the interpretation of these ST-EPR studies deserves reconsideration.

3.2 MATERIALS AND METHODS

3.2.1 PREPARATION OF SPIN-LABELLED VIRUS

The preparation of spin-labelled CCMV has been described by Krüse et al. (4). Here a slightly modified procedure is presented that resulted in labelled

virus that had the same properties as native virus during all control experiments.

Virus was grown and isolated as described by Verduin (11). After purification the virus was dialysed against 50 mM Tris-HCl, 20 mM MgCl₂ and 50 mM NaCl, PH 7.5. The virus concentration was 2 mg/ml. The spin-labelling was carried out at 4°C under a nitrogen atmosphere. 10 µg of spin-label (3 maleimido-2,2,5,5-tetramethyl-1-pyrrolodinyloxyl) dissolved in ethanol (20 mg/ml) was added per milligram virus (this corresponds to 1 spin-label molecule/protein subunit). After 30 minutes unreacted spin-label was removed by extensive dialysis against 50 mM Tris-HCl, 10 mM MgCl₂ and 50 mM NaCl, pH 7.5. After removal of the unreacted spin-label the virus was dialysed to the measuring conditions and concentrated to ~30 mg/ml.

The integrity of the virus was investigated by analytical ultracentrifugation, rate zonal centrifugation analysis, electron microscopy and gel electrophoresis of protein and RNA as described by Verduin (12). 5,5'-dithiobis(2-nitrobenzoic acid) was used to determine the extent of spin-labelling (13). The amount of spin-labelling never exceeded 0.1 spin-label molecule per protein subunit. It was found that it was important to perform all manipulations at low concentrations of virus (2 mg/ml), spin-label (20 µg/ml) and glutaraldehyde (0.5% w/w). All samples were measured directly after concentrating.

3.2.2 FIXATION WITH GLUTARALDEHYDE.

The fixation of SL-CCMV was carried out with glutaraldehyde by dialysis for 6 hours against 50 mM Tris-HCl, 10 mM MgCl₂, 50 mM NaCl and 0.5% (w/w) glutaraldehyde, pH 7.5 and subsequent removal of residual fixative by extensive dialysis against the same buffer without glutaraldehyde. Through this procedure the fixation of the virus became complete as was proven by the inability of the virus to dissociate, even under extreme dissociating conditions. After fixation the integrity of the particle was checked by rate zonal centrifugation analysis and electron microscopy. No mutually cross-linked CCMV particles could be detected. Gel electrophoresis patterns of protein and RNA could not be obtained since the particle could not be dissociated.

3.2.3 EPR and ST-EPR SPECTROSCOPY

EPR spectra were recorded with a Varian E6 spectrometer in the X-band. The samples were contained in a glass capillary of 1.2 mm diameter. The sample height was 15 mm and the sample was carefully placed in the center of a

rectangular cavity (Varian E-231). The modulation frequency was 100 kHz. The microwave power was 5 mW and the modulation amplitude 0.5 mT at the center of the sample.

For the ST-EPR measurements the spectrometer was modified as described by De Jager and Hemminga (14). Second harmonic (200 kHz) quadrature ST-EPR spectra were recorded with 100 mW microwave power and 1.2 mT modulation amplitude at the center of the sample. Phase settings for the phase sensitive detection of the quadrature signal were the result of a linear interpolation of 6 spectra recorded closely around the final value. These spectra were recorded at 5 mW microwave power to ensure that saturation transfer effects were minimal. All spectra shown are the result of the summation of 30 scans of 2 minutes filtered with a time constant of 0.3 s. H'/H , C'/C and L'/L values (5,15) were compared with reference spectra of haemoglobine labelled with the same spin-label to obtain the isotropic values of the rotational correlation times (16). With the exception of the different salt and glycerol concentrations (which may influence the dielectrical constant), all measuring conditions were kept identical between the SL-CCMV and the haemoglobine spectra.

3.3 RESULTS

In figure 3.1 the EPR-spectra of SL-CCMV (figure 3.1A) and SL-CCMV protein dimers (figure 3.1B) are presented. The latter spectrum is characteristic for a rapid anisotropic motion (faster than 10^{-8} s) about the nitroxide z-axis (17,18). The lineshape of this spectrum corresponds well with computer simulations by Griffith and Jost (18) for a model that takes into account a rapid wobbling motion of the z-axis of the nitroxide group with a maximum half amplitude of 40° (corresponding to an order parameter $S=0.72$).

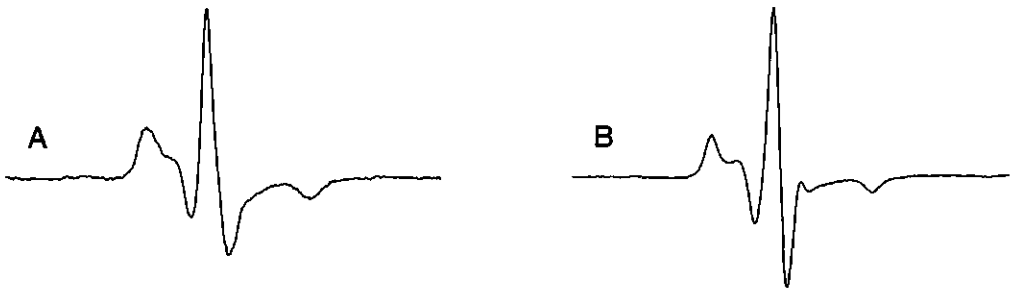


Figure 3.1. EPR spectra of:

A: Spin-labelled CCMV.

B: Spin-labelled CCMV protein dimers.

The spectrum of SL-CCMV (see figure 3.1A) is more difficult to interpret. All peaks are broadened, but no significant shifts are observed. This leads to the conclusion that the order parameter remains approximately constant, whereas the rate of anisotropic motion about the z-axis decreases below the limit of rapid motion (i.e. slower than 10^{-8} s). The rotational correlation time is still in an intermediate motional region ($\sim 10^{-8}$ - 5×10^{-8} s), since this spectrum significantly deviates from a rigid powder spectrum.

The overall rotational motion of the protein dimers or the virus particle does not influence the above conclusions. For the dimer the isotropic rotational correlation time, as calculated from the Stokes-Einstein equation, is 2.5×10^{-8} s, whereas the virus yields a value of 5×10^{-6} s. In both situations the system is in the slow motional region.

In the last decade ST-EPR has been applied to study and characterize the dynamic properties of a number of spin-labelled biomolecules. In most cases the values for H''/H , C'/C and L''/L from reference spectra of spin-labelled molecules undergoing isotropic Brownian diffusion have been used to derive the rotational correlation time of the molecule of interest (see (15) for an review). When the graphs of L''/L , C'/C and H''/H vs. the rotational correlation time are employed for the analysis of systems undergoing anisotropic diffusion, such as spin-labelled phospholipids in membranes, experimental values from a single spectrum often yield different values for the rotational correlation time. The analysis of anisotropic motion in ST-EPR spectra has been discussed by Marsh et al. (6, 19).

Figure 3.2A is the ST-EPR spectrum of SL-CCMV. This spectrum is very similar to ST-EPR spectra of spin-labelled phospholipids undergoing anisotropic motion in membranes (6). This supports the conclusion from the EPR spectra that the observed motional anisotropy corresponds to motion about the z-axis.

When the value of C'/C is compared with the values obtained from reference spectra for isotropic rotational motion, a value of $\sim 6 \times 10^{-7}$ s is found. This value represents the correlation time for the anisotropic motion about the z-axis. From H''/H and L''/L a value of $\sim 2 \times 10^{-5}$ s is obtained. This value represents the rotational correlation time of the motion of the z-axis itself. This value is approximately equal to the rotational correlation time for the overall motion of the virus particle, which is calculated to be 5×10^{-6} s.

Figure 3.2B gives the ST-EPR spectrum of SL-CCMV protein dimers. The observed spectral shape is rather unusual. This is probably the case because the mobility of the dimers (the calculated overall rotational correlation time is $\sim 2.5 \times 10^{-8}$ s) is rather fast on the ST-EPR timescale. It is therefore difficult to estimate

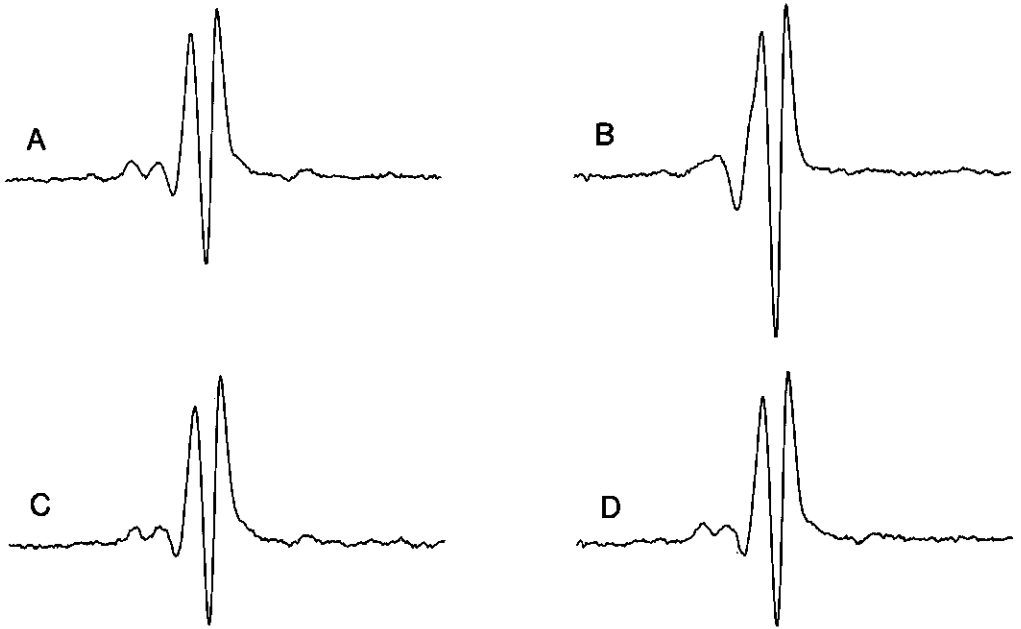


Figure 3.2. ST-EPR spectra of:

A: Spinlabelled CCMV.

B: Spin-labelled CCMV protein dimers.

C: Glutaraldehyde treated spin-labelled CCMV.

D: Glutaraldehyde treated spin-labelled under strong dissociating conditions.

the rotational correlation times from the spectral shape. The C'/C value indicates that the correlation time for the anisotropic motion about the z-axis is $\sim 10^{-7}$ s.

Krüse and Hemminga (4) also recorded EPR and ST-EPR spectra of SL-CCMV and reported anisotropic motion about the z-axis. They believed this to be the result of an anisotropic motion of the whole spin-labelled protein subunit within the virus or of an anisotropic motion of the spin-labelled part of the protein within the virus protein subunit. Figure 3.2C is the spectrum of glutaraldehyde treated SL-CCMV. Since no differences can be observed between this spectrum and that of untreated SL-CCMV, it is concluded that the observed anisotropic motion is not the result of rotations of the protein subunits within the virus core. An NMR study on CCMV (7) has shown that a fast mode of segmental protein motion can be excluded. Therefore the observed anisotropic mode of motion of the spin-label about its z-axis is a local anisotropic motion with respect to the protein at which the spin-label is attached.

Figure 3.2D is the spectrum of glutaraldehyde treated SL-CCMV at pH 8.0 and 1.0 M NaCl (strong dissociating conditions). No differences are observed between this spectrum and that of glutaraldehyde treated SL-CCMV in a non dissociating environment (see figure 3.2C), and thus it is concluded that the subunit fixation is complete.

3.4. DISCUSSION.

From the EPR and ST-EPR studies it clearly follows that the maleimide spin-label is not rigidly attached to the CCMV protein, but performs a local anisotropic motion about the z-axis in a cleft at the binding site. The spin-label can only be attached to the virus when the pH is above 7.2, the value at which the protein subunits are moved away from each other (20). It is therefore plausible to conclude that this cleft is located at the protein-protein subunit interface. The anisotropic motion about the z-axis is much slower in the whole virus than in the protein dimers. The most reasonable explanation for this observation is an increased steric hindrance due to the close proximity of the other protein subunits in the whole virus.

The observed anisotropic motion about the z-axis is surprising. From a projection scheme of the maleimide spin-label molecule (see figure 3.3A), one would expect the x-axis to be the rotation axis. Such a projection scheme however, can be very misleading, since from model building of the spin-label using space filling atoms, it can be shown that motion about the z-axis is more likely to occur than motion about the x- (or y-) axis.

The figures 3.3B-3.3D show some possible conformations of the spin-label in which the z-axis is almost parallel to the long molecular axis. From the EPR spectra of SL-CCMV protein dimers an order parameter $S=0.7$ is obtained. This order parameter suggests a wobbling motion of the z-axis within the binding cleft (17,18). The spin-label itself has an asymmetric carbon atom; a second asymmetric carbon atom is the result of binding to the protein. In addition the spin-label can occur in two conformations with different orientations of the rings with respect to each other. Between these two conformations there is a high rotational barrier (see figure 3). This large number of possible conformations allows several angles between the z-axis and the long axis of the spin-label, which makes it impossible to quantitatively interpret the observed order parameter. Altogether, these various modes of molecular motion lead to an anisotropic motion that is mainly about the z-axis.

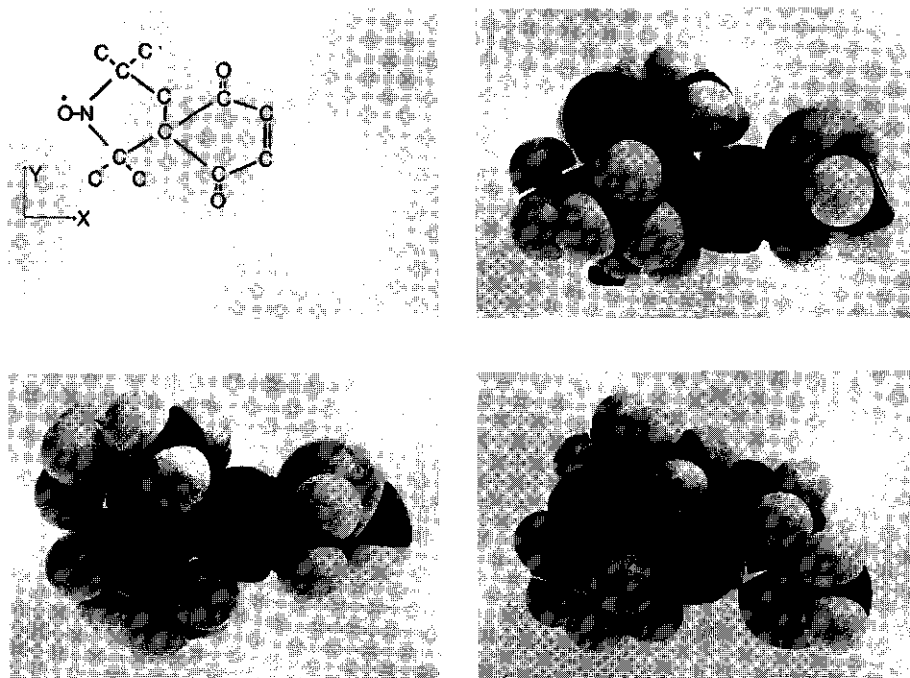


Figure 3.3. Projection scheme of the maleimide spin-label with the z-axis (N-O direction) perpendicular to the plane. The space filling models of three conformations are shown. These models indicate that the z-axis is most likely to be parallel to the long axes of the molecule.

In many other ST-EPR studies of maleimide spin-labelled proteins (8-10), ST-EPR lineshapes have been observed that are very similar to the ST-EPR spectrum of SL-CCMV. In all cases anisotropic protein mobility was invoked in order to interpret the ST-EPR lineshapes. From the present work it is suggested that the published ST-EPR studies on maleimide spin-labelled proteins need to be reconsidered in view of a local anisotropic spin-label motion at its binding site at the protein.

3.5. REFERENCES.

- (1) Bancroft, J.B., Hills, G.J., Markham, R., (1967), *Virology* 31, 354-379.
- (2) Pfeiffer, P., Hirth, L., (1975), *FEBS Lt.* 56, 144-148.
- (3) Bancroft, J.B., Hiebert, E., (1967), *Virology* 32, 354-356.
- (4) Kruse, J., Hemminga, M.A., (1981), *Eur. J. Biochem.* 113, 575-579.
- (5) Thomas, D.D., Dalton, L.R., Hyde, J.S., (1967), *J. Chem. Phys.* 65, 3006-3024.
- (6) Marsh, D., (1980), *Biochemistry* 19, 1632-1637.
- (7) Vriend, G., Hemminga, M.A., Verduin, B.J.M., Schaafsma, T.J., (1982), *FEBS Lt.* 146, 319-321.
- (8) Kusumi, A., Ohnishi, S., Ito, T., Yoshizawa, T., (1978), *B.B.A.* 507, 539-543.
- (9) Hidalgo, C., Thomas, D.D., Ikemoto, N., (1978), *J. Biol. Chem.* 253, 6879-6887.
- (10) Kirino, Y., Ohkuma, T., Shimizu, H., (1978), *J. Biochem. (Tokyo)* 84, 111-115.
- (11) Verduin, B.J.M., (1978), *J. Gen. Virol.* 39, 131-147.
- (12) Verduin, B.J.M., (1978), Ph.D. Thesis, Agricultural University Wageningen.
- (13) Means, G.E., Feeney, R.E., (1971), *Chemical Modifications of Proteins*, Holden-Day, San-Fransisco.
- (14) Jager, P.A. de, Hemminga, M.A., (1978), *J. Mag. Res.* 31, 491-496.
- (15) Hyde, J.S., Dalton, L.R., *Spin Labeling II, Theory and Applications*, Chapter 1, Academic Press, New York, San Fransisco, London.
- (16) Hemminga, M.A., (1983), *Chemistry and Physics of Lipids* 30, (In Press).
- (17) Israelachvili, J., Sjosten, J., Eriksson, L.E.G., Ehrstrom, M., Graslund, A., Ehrenberg, A., (1974), *B.B.A.* 339, 164-172.
- (18) Griffith, O.H., Jost, P.C., (1976), *Spin Labeling I, Theory and Applications*, Chapter 12, Academic Press, New York, San Fransisco, London.
- (19) Fajer, P., Marsh, D., (1983), *J. Mag. Res.* 51, 446-459.
- (20) Pfeiffer, P., (1980), *Virology* 102, 54-61.

4 SEGMENTAL MOBILITY INVOLVED IN PROTEIN-RNA INTERACTION IN COWPEA CHLOROTIC MOTTLE VIRUS

4.1 INTRODUCTION

Cowpea Chlorotic Mottle Virus is a spherical plant virus. It consists of three nucleoprotein particles containing 4 RNA molecules with sedimentation coefficients of 13 S, 18 S and 23 S (1,2). The molecular mass of each nucleoprotein particle is $\sim 4.6 \times 10^6$ g. All three particles have the same protein capsid, which consists of 180 identical polypeptide chains arranged icosahedrally around the RNA (3). CCMV is stable around pH 5.0, sedimenting at 88 S and increases in hydrodynamic volume at pH 7.5 and low ionic strength ($\mu \sim 0.2$).

In 1 M NaCl at pH 7.5 the virus dissociates into protein dimers and RNA almost free of protein (4). Both protein and RNA can be isolated in a native state and reassembled in vitro into nucleoprotein particles by lowering the ionic strength at pH 7.5 (4). In the absence of RNA and by lowering the pH to 5.0, protein dimers reassemble into empty protein capsids sedimenting at 52 S (5). The spatial arrangement of the polypeptides in these capsids is similar to that in the nucleoprotein capsid: a T=3 icosahedral surface lattice (6).

The coat protein contains a basic N-terminal region (7), which can be cleaved by tryptic digestion (8). It has been suggested that the N-terminal region is involved in protein-RNA interaction (8). In this chapter it is demonstrated that the N-terminal arm of CCMV is the RNA binding part of the protein.

Nuclear magnetic resonance is a suitable technique to study the internal mobility in large molecular systems (9-11). From the NMR experiments on the coat protein of CCMV presented in this chapter it is found that the N-terminal region is internally mobile with a correlation time of 10^{-9} - 10^{-8} s in the absence of RNA, whereas this mobility is lost upon interaction with RNA.

4.2. MATERIALS AND METHODS.

4.2.1. PREPARATION OF PROTEIN LACKING THE N-TERMINAL AMINO ACIDS.

Virus was purified as described by Verduin (12). Coat protein prepared from virus by the CaCl_2 method (13) was dialyzed against 300 mM NaCl, 1mM dithiothreitol (DTT) and 50 mM Tris-HCl (pH 7.5). To this solution trypsin was added at an enzyme-substrate ratio of 1/1000. The mixture was incubated for 1 h at 25°C. After incubation the pH of the mixture was adjusted to pH 5.0 with 1 M

sodium acetate. Thereafter, protein lacking the N-terminal region was separated from trypsin and oligopeptides on a Sepharose 6B column equilibrated with 300 mM NaCl and 50 mM sodium acetate (pH 5.0). The fractions containing empty protein capsids were concentrated by centrifugation and used for assembly or NMR experiments.

4.2.2. PROTEIN ASSEMBLY.

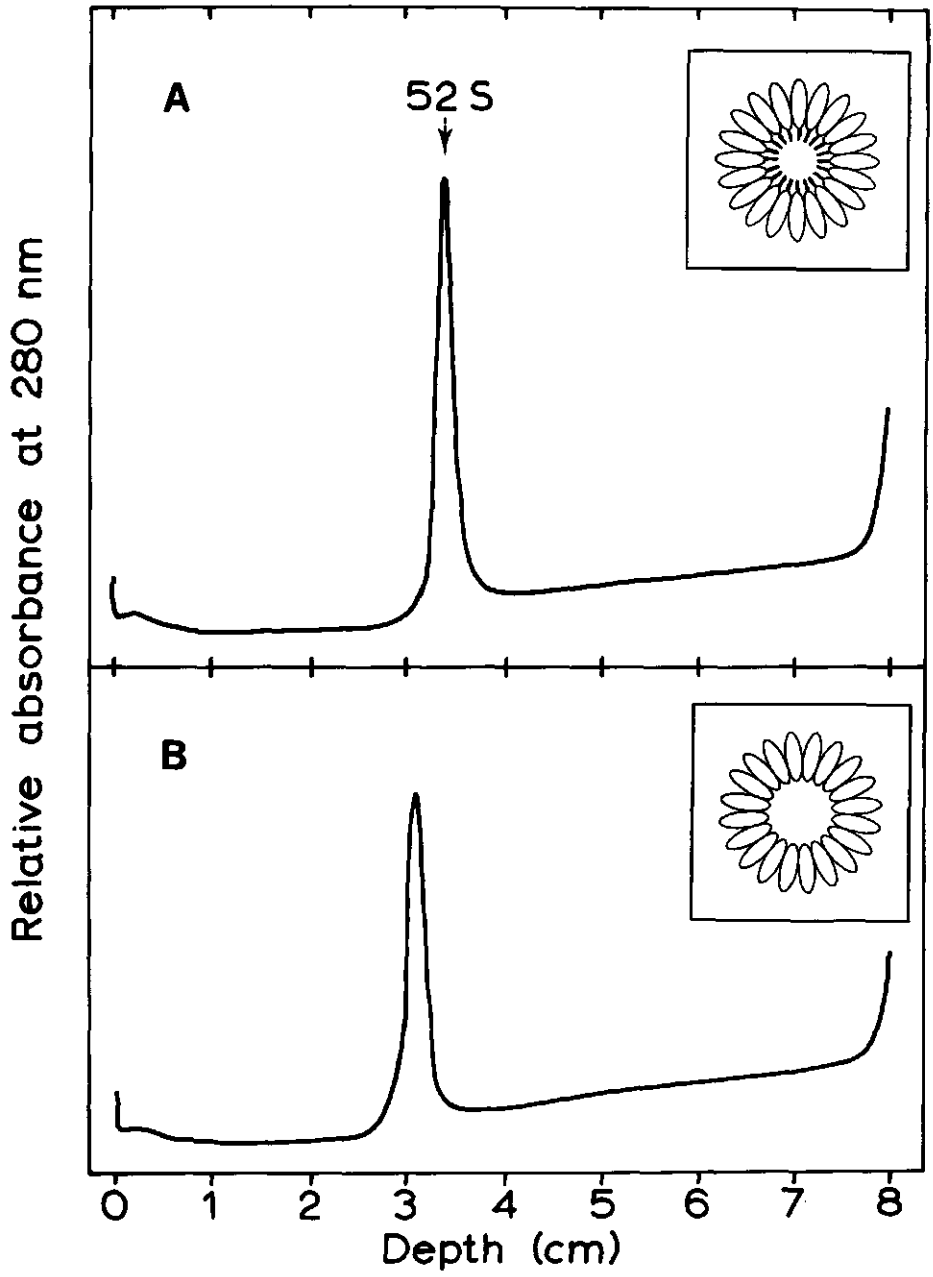
Association of coat protein was established by lowering the pH of a protein solution from pH 7.5 to pH 5.0. Both native protein and protein lacking the N-terminal region were dialyzed against dissociation buffer (1 M NaCl, 1 mM DTT and 50 mM Tris HCl pH 7.5) and the dimers were subsequently assembled into empty capsids by dialysis against 300 mM NaCl, 1 mM DTT and 50 mM sodium acetate (pH 5.0); 250 μ l containing 100 μ g protein was layered onto a linear 10%-50% (w/v) sucrose gradient made up in 300 mM NaCl, 1mM DTT and 50 mM sodium acetate (pH 5.0) and centrifuged for 15 h in a Beckman SW 41 rotor at 25000 rev./min and 5°C. The contents of the tubes were monitored with a LKB Uvicord III apparatus at 280 nm.

4.2.3. NUCLEOPROTEIN ASSEMBLY.

Association of protein and RNA was established by lowering the ionic strength at pH 7.5. Phenol extracted RNA was mixed either with native coat protein or with coat protein lacking the N-terminal region in dissociation buffer and dialyzed at 5°C for 2 h against 10 mM KCl, 5 mM MgCl₂, 1 mM DTT and 10 mM Tris-HCl (pH 7.5); 250 μ l of the mixture was layered onto a linear 10%-50% (w/v) sucrose gradient made up in 10 mM KCl, 5 mM MgCl₂, 1 mM DTT and 10 mM Tris-HCl (pH 7.5) and centrifuged for 15 h in a Beckman SW 41 rotor at 25000 rev./min and 5°C. The contents of the tubes were monitored with a LKB Uvicord III apparatus at 280 nm.

4.2.4. NMR MEASUREMENTS.

Virus was dialyzed against 200 mM KCl, 10 mM MgCl₂ and 0.5 mM sodium phosphate (pH 5.0). The protein preparations were dialyzed against 300 mM NaCl, 10 mM MgCl₂ and 0.5 mM sodium phosphate (pH 5.0). H₂O in the virus and protein solutions was substituted by D₂O through 3 cycles of centrifugation and resuspension of the pellets in the above mentioned solutions made up in D₂O. In D₂O solutions pH meter readings were taken without correction for the presence



of D_2O . The final concentration of nucleoprotein and protein varied from 3-18 mg/ml.

1H -NMR spectra were recorded with a Bruker WM250 spectrometer. Samples of 500 μ l were measured at 7°C in the quadrature detection mode with D_2O lock and without 1H decoupling. The acquisition time was 0.82 s and 4000 scans were taken. The sensitivity enhancement was 5 Hz. The ppm scale was relative to sodium 2,2-dimethyl 2-silapentane-5-sulphonate (DSS). The vertical scale was corrected for concentration differences.

4.3. RESULTS.

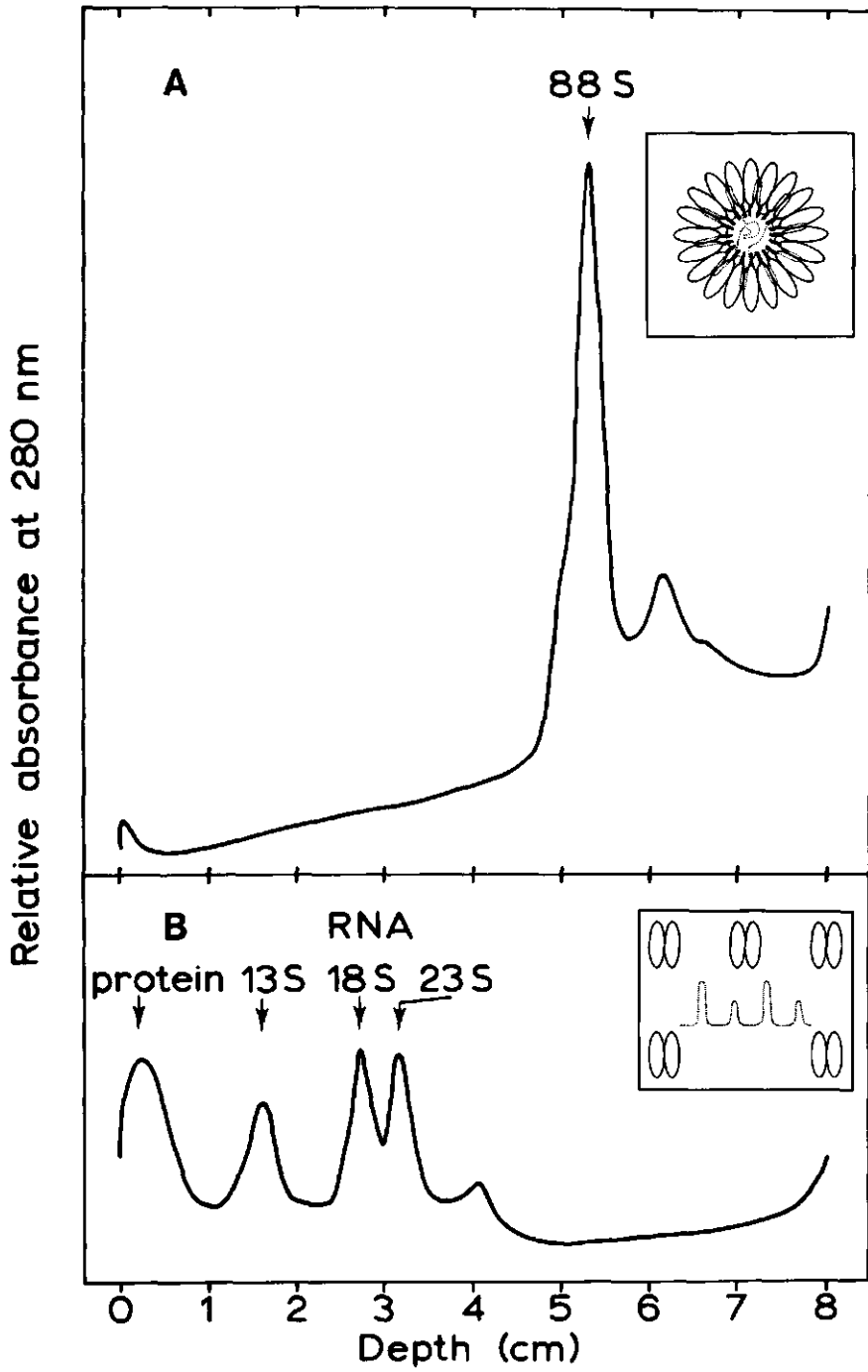
4.3.1. THE ROLE OF THE N-TERMINAL AMINO ACIDS IN PROTEIN-RNA INTERACTION.

The protease sensitivity of the protein at pH 7.5 (3) was utilized to prepare viral coat protein lacking 25 amino acids from the N-terminus. The N-terminal region is highly basic and contains 6 arginine and 3 lysine residues (7). In the absence of RNA, both native protein and protein lacking the N-terminal region assemble into spherical particles as observed in electron microscopy. This indicates that the N-terminal region is not required for interaction between protein subunits. Figure 4.1 represents the rate zonal centrifugation analysis of these RNA-free assembly products. As expected, the empty capsids assembled from protein lacking the N-terminal region sediment slower than 52 S. Figure 4.2 shows the rate zonal centrifugation analysis of the products obtained by mixing either native protein or protein lacking the N-terminal region with intact viral RNA and dialyzing to assembly conditions. In agreement with the results of Bancroft et al. (4), native protein reassembled with RNA into nucleoprotein particles sedimented around 88 S. The protein lacking the N-terminal region did not show interaction with RNA. These data demonstrate that the N-terminal region is the binding site for RNA.

Figure 4.1. Rate zonal centrifugation analysis of CCMV coat protein at pH 5.0:
A: CCMV native protein.

B: CCMV protein lacking the N-terminal region.

Inserts show schematic representations of the assembly products. Ellipses with tails represent coat protein with the N-terminal region intact.



4.3.2. MOBILITY OF THE N-TERMINAL REGION OF THE PROTEIN.

Since CCMV has a large molecular mass, the NMR spectral intensity of sharp peaks is a direct measure of the number of mobile nuclei (9,10). Rigid nuclei give rise to broad unresolved resonances. Figure 4.3A represents the 250 MHz

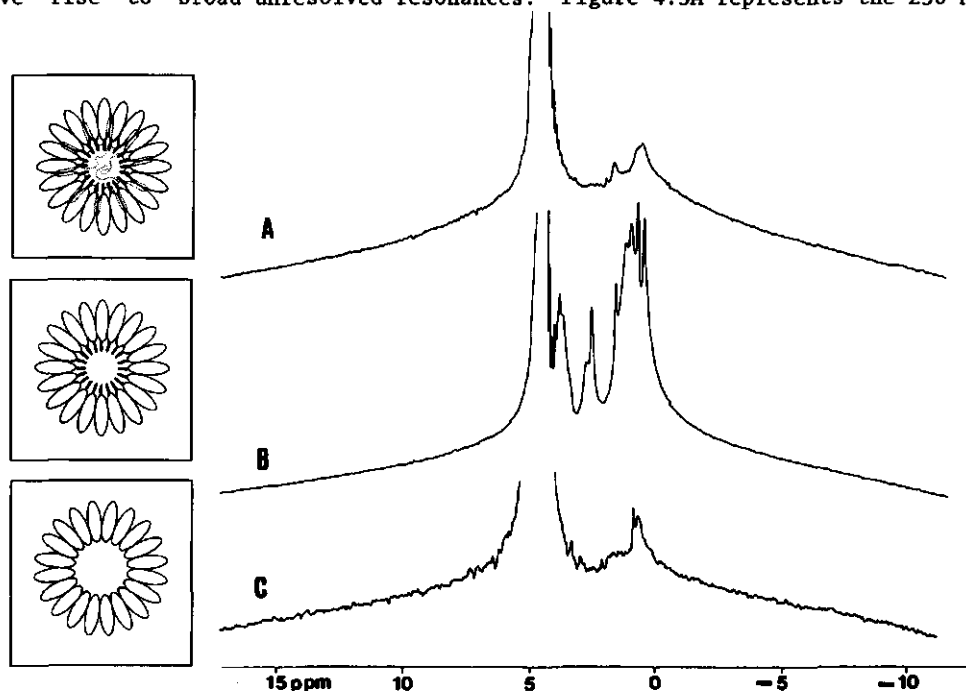


Figure 4.3. 250 MHz ^1H -NMR spectra of CCMV and its coat protein assembly products:

A: Native virus.

B: Empty protein capsids.

C: Empty protein capsids lacking the N-terminal region.

Inserts show schematic representations of CCMV and its coat protein assembly products. Ellipses with tails represent coat protein with the N-terminal region intact. Dotted lines represent the RNA.

Figure 4.2. Rate zonal centrifugation analysis of RNA-protein mixtures under assembly conditions:

A: 62.5 μg CCMV-RNA + 250 μg CCMV native protein.

B: 62.5 μg CCMV-RNA + 250 μg CCMV protein lacking the N-terminal region.

Inserts show schematic representations of the assembly products. Ellipses with tails represent coat protein with the N-terminal region intact. The dotted lines represent the RNA.

^1H -NMR spectrum of CCMV. A fraction of $\sim 98\%$ of the resonances is contained in a broad, unresolved background signal, indicating that most of the protein and the RNA is rigid. The few sharp peaks in the 0.5-2.5 ppm region, representing $\sim 2\%$ of the total spectral intensity, correspond to mobile side chains of amino acids at the protein contact faces (see chapter 2) with a rotational correlation time of $\sim 10^{-8}$ s (9). The spectrum of empty capsids of native protein is shown in figure 4.3B. The linewidth of the peaks in the 0.5-4.5 ppm region of this spectrum corresponds to an upper limit of the rotational correlation time of 3×10^{-9} s. The intensity of these peaks correspond to 10%-15% of the total NMR signal which is approximately equivalent to 15-25 amino acid residues. Figure 4.3C represents the spectrum of empty protein capsids lacking the N-terminal region. From the absence of most of the sharp peaks in the 0.5-4.5 ppm region it is concluded that the mobility observed in the empty capsids of native protein (figure 4.3B) is almost completely concentrated in the N-terminal region. Further evidence for the mobility of the N-terminal region is obtained from the observation that no aromatic resonances occur in the 5-7 ppm region of the spectrum in figure 4.3B. The N-terminal region does not contain any of the twelve aromatic amino acid residues present in native protein (7).

4.4. DISCUSSION.

From the absence of resonances of the N-terminal region in the native virus spectrum (figure 4.3A) it is concluded that immobilization of the N-terminal region occurs upon interaction with RNA. This mobility of the N-terminal region of the coat protein of CCMV is of great importance in the assembly process of the virus, since it provides a mechanism to enhance the probability of interaction of protein and RNA. Although no structural information can be derived from the NMR results presented in this chapter, these results are consistent with a model in which a flexible random coil configuration of the N-terminal region changes into a rigid α -helix on interaction with RNA. This model shows great resemblances with models proposed for other systems of which the stability mainly depends on protein-RNA interaction, such as brome mosaic virus (14), histones (15), or protamines (16).

Smearred out electron densities are found for the protein-RNA interaction regions in the high resolution X-ray structure determinations of tobacco mosaic virus (17), tomato bushy stunt virus (TBSV) (18) and southern bean mosaic virus (SBMV) (19). Internal mobility or disorder have been suggested to account for this effect (18). The results presented in this chapter on CCMV, which is very similar in structure to SBMV and TBSV, indicate that smearred out electron

densities can not be interpreted as mobility on a time scale faster than 10^{-6} s.

From the similarities of protein-RNA interaction in various biological systems, it is suggested that the phenomenon of immobilization of the protein region which interacts with RNA is a general feature throughout molecular biology.

The results presented in this chapter have also been published in:

Vriend, G., Hemminga, M.A., Verduin, B.J.M., Wit, J.L.de, Schaafsma, T.J., (1981)
FEBS Lt. 134, 167-171.

4.5. REFERENCES.

- (1) Bancroft, J.B., Hiebert, E., Rees, M.W., Markham, R., (1968), *Virology* 34, 224-239.
- (2) Bancroft, J.B., (1971), *Virology* 45, 830-834.
- (3) Bancroft, J.B., Hills, G.J., Markham, R., (1967), *Virology* 31, 354-379.
- (4) Bancroft, J.B., Hiebert, E., (1967), *Virology* 32, 354-356.
- (5) Bancroft, J.B., Wagner, G.W., Bracker, C.E., (1968), *Virology* 36, 146-149.
- (6) Finch, J.T., Bancroft, J.B., (1968), *Nature* 220, 815-816.
- (7) Rees, M.W., Short, M.N., Turner, D.S., (1980), *John Innes Inst. 70th Annu. Rep.*, 102-103.
- (8) Chidlow, J., Tremaine, J.H., (1971), *Virology* 43, 267-278.
- (9) De Wit, J.J., Hemminga, M.A., Schaafsma, T.J., (1978), *J. Mag. Res* 31, 97-107.
- (10) De Wit, J.L., Alma, N.C.M., Hemminga, M.A., Schaafsma, T.J., (1979), *Biochemistry* 18, 3973-3976.
- (11) Jardetzky, O., Akasaka, K., Vogel, D., Morris, S., Holmes, K.C., (1978), *Nature* 273, 564-566.
- (12) Verduin, B.J.M., (1978), *J. Gen. Virol* 39, 131-147.
- (13) Verduin, B.J.M., (1974), *FEBS Lt* 45, 50-54.
- (14) Argos, P., (1981), *Virology* 110, 55-62.
- (15) Sung, M.T., Dixon, G.H., (1970), *Proc. Natl. Acad. Sci.* 67, 1616-1623.
- (16) Warrant, R.W., Kim, S.H., (1978), *Nature* 271, 130-135.
- (17) Stubbs, G., Warren, S., Holmes, K.C., (1977) *nature* 267, 216-221.
- (18) Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K., Bricogne, G., (1978), *Nature* 276, 368-373.
- (19) Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., Leslie, A.G.W., Rayment, I., Rossmann, M.G., Suck, D., Tsukihara, T., (1980), *Nature* 268, 33-39.

5 MOBILITY INVOLVED IN PROTEIN-RNA INTERACTION IN SPHERICAL PLANT VIRUSES

5.1 INTRODUCTION

In the previous chapter it has been shown that the protein region in CCMV involved in protein-RNA interaction is very mobile in the absence of RNA (1). Upon binding RNA, immobilization of this protein region occurred. From a comparison with other systems, e.g. tobacco mosaic virus (2-4), tomato bushy stunt virus (5), southern bean mosaic virus (6), histones (7) and protamines (8) and using protein secondary structure predictions (9) (see chapter 7) a general model for the protein-RNA interaction in simple plant viruses has been suggested. This model is based on a random coil to α -helix transition for the protein region interacting with RNA upon binding.

In this chapter three other spherical plant viruses are considered. Brome mosaic virus (BMV), belladonna mottle virus (BdMV) and cowpea mosaic virus (CPMV) have been used, since their stability depends to a different extent upon protein-RNA and protein-protein interactions (10). The results obtained are in agreement with the general ideas about protein-RNA interaction which have been derived from experiments on CCMV (chapters 4 and 6).

5.2 MATERIALS AND METHODS

5.2.1 VIRUS PREPARATION

BMV was purified as described for CCMV (11). Dimeric coat protein prepared from virus by the CaCl_2 method (12) was assembled into empty protein capsids by dialysis against 300 mM NaCl, 1 mM dithiothreitol (DTT) and 50 mM sodium acetate (pH 5.0).

The strain of BdMV used in this study was obtained from Dr. P.Argos and was originally isolated from *Physalis heterophylla* (13). Virus was purified from tobacco *Nicotiana glutinosa* as described by Ghabrial et al. (14) with the modification as described by Heuss et al. (15). The naturally occurring empty protein capsids (BdMV-T) were separated from the bottom component (BdMV-B) by isopycnic centrifugation in 40% (w/w) cesium chloride in 50 mM potassium phosphate buffer (pH 5.5).

The Sb isolate of CPMV was grown in *Vigna unguiculata* L. var. Blackeye Early Ramshorn and purified as described by Van Kammen (16). The top, middle and bottom components (CPMV-T, CPMV-M and CPMV-B) were separated by centrifugation in a linear 15%-30% (w/w) zonal sucrose gradient in 10 mM sodium phosphate

buffer at pH 7.0. The gradient was centrifuged for 16 h at 23000 rev./min at 5°C in an MSE B XV Ti rotor. CPMV was also purified using chloroform-butanol clarification of the leaf homogenate and isoelectric precipitation of the virus in order to avoid contamination of the virus with polyethylene glycol and sucrose.

5.2.2 NMR MEASUREMENTS

Preparations of BdmV and CPMV were dialyzed against 200 mM KCl, 10 mM MgCl₂, and 5 mM sodium phosphate (pH 7.5). BMV was dialyzed against the same buffer at pH 5.0 and its protein against 300 mM KCl, 10 mM MgCl₂ and 5 mM sodium phosphate (pH 5.0). H₂O in the solutions was substituted by D₂O by extensive dialysis against the above mentioned solutions made up in D₂O. The final concentration of protein or nucleoprotein varied from 5-20 mg/ml. ¹H-NMR spectra were recorded with a Bruker WM250 or a Bruker WM500 supercon spectrometer. Samples of 500 µl were measured in the quadrature detection mode with D₂O lock, and without proton decoupling. The acquisition time plus pulse delay was 1 s and 4000 scans were taken. The sensitivity enhancement was 5 Hz. The ppm scale was relative to sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS). The verticale scale was corrected for concentration differences between the components of each virus. The spectral width was 40 ppm.

5.3 RESULTS

5.3.1 BROME MOSAIC VIRUS

Figure 5.1A represents the 250 MHz ¹H-NMR spectrum of BMV. The few peaks in the 1.0-2.5 ppm region contain about 2% of the total spectral intensity. These peaks can only be assigned to -CH₂ and -CH₃ groups in hydrophylic amino acids (17), probably located at or near the protein surface. The linewidth of these peaks corresponds to an upper limit of the rotational correlation time of $\sim 10^{-8}$ s (1-4). The spectrum of BMV empty capsids (BMV-PT) is shown in figure 5.1B. The sharp peaks in the 0.5-4.5 ppm region of this spectrum contain $\sim 15\%$ of the total spectral intensity. The position and relative intensities of these peaks are in good agreement with what is predicted from the primary structure of the 25 amino acid N-terminal arm of BMV protein (17,18). The linewidth of these peaks cooresponds to an upper limit for the rotational correlation time of 10^{-9} s.

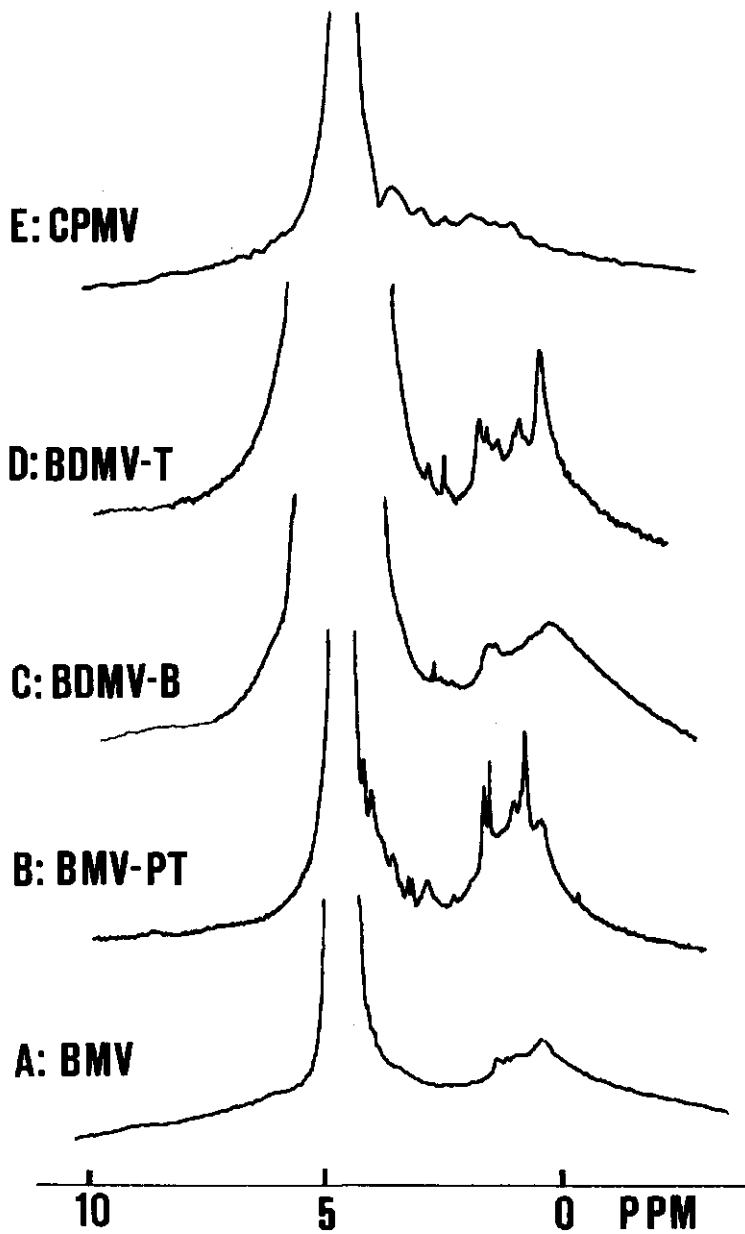


Figure 5.1. $^1\text{H-NMR}$ spectra:

A: BMV at 250 MHz.

B: BMV-PT at 250 MHz.

C: BdMV-B at 500 MHz.

D: BdMV-T at 500 MHz.

E: CPMV combined T,M and B fractions at 250 MHz.

Only the 0-10 ppm region is shown.

5.3.2 BELLADONNA MOTTLE VIRUS

Figure 5.1C represents the 500 MHz ^1H -NMR spectrum of BdmV-B. Except for one small peak at ~ 3 ppm, which probably arises from a mobile lysine sidechain at the protein surface, no mobility on a timescale faster than 3×10^{-8} s can be observed. The spectrum of BdmV-T is shown in figure 5.1D. The sharp peaks in the 1-4 ppm region contain $\sim 10\%$ of the total spectral intensity.

The linewidth of these peaks corresponds to an upper limit of the rotational correlation time of 3×10^{-9} s. In this case experiments have been carried out at 250 MHz (spectra not shown) and 500 MHz. This enables us to conclude that the observed linewidth is due to isotropic chemical shift differences (3).

The real rotational correlation time is therefore probably much smaller than the above mentioned upper limit of 3×10^{-9} s. The observed resonances may be assigned to a few lysine residues, some threonine residues and probably one or two aspartic acid, glutamic acid, glutamine and asparagine residues.

No aromatic residues nor cysteine or cystine are present among the mobile amino acids.

5.3.3 COWPEA MOSAIC VIRUS.

Figure 5.1E represents the 250 MHz ^1H -NMR spectrum of the unseparated CPMV-T, CPMV-M and CPMV-B components of CPMV purified by the chloroform-butanol method. Since no sharp peaks are observed in this spectrum, it is concluded that neither one of the CPMV components contains internal mobile groups. The spectra of the individual components are not shown since they could not be obtained without large sugar and polyethylene glycol resonances. Except for these resonances, no sharp peaks were observed in the spectra of the individual components.

5.4 DISCUSSION

In the chapters 4 and 6, protein-RNA binding studies on CCMV are presented. For this virus the 25 N-terminal amino acids are shown to be very mobile in the absence of RNA. In the presence of RNA no mobility is observed. Protein lacking these 25 amino acids is not capable of binding RNA. From the great genetical and morphological resemblance of CCMV and BMV, a similar behaviour is expected for these two viruses. The spectra in figure 1A and B prove this idea to be correct. The 25 N-terminal amino acids of BMV protein are almost completely predicted to be in an α -helix conformation (9). The observed rotational correlation time of $\sim 10^{-9}$ s makes it hard to believe that this protein region is completely

α -helical. Therefore, it is suggested that the protein region which binds the RNA adapts a flexible random coil configuration in the absence of RNA. Such a conformation is also found from secondary structure predictions, where the positive charges on the arginine and lysine residues are retained (G.Vriend unpublished results).

The mobility of the N-terminal protein region is of great importance since it provides a mechanism to enhance the probability of interaction between protein and RNA. The large number of basic residues in this protein region (6 arginine and 3 lysine residues in CCMV (19); 7 arginine and 1 lysine residues in BMV (18)) suggests that interaction between the negative phosphate groups in the RNA and the positive basic amino acids is the primary driving force for nucleoprotein assembly. It has been proven (6,8,2-4,21) or suggested (1,7,9) that many nucleic acid binding proteins provide the best protein-nucleic acid interaction when the protein adopts an α -helix conformation. For this reason a random coil - α -helix transition for the N-terminal protein region upon binding RNA is suggested. Figure 5.2 is the schematical representation of the model.

The observed spectra of BdmV fit this model very well. The mobile lysine residues in the BdmV protein provide the positive charges for binding the RNA phosphate groups. Figure 5.1D proves that immobilization of the mobile protein region occurs upon binding RNA. Interactions between protonated cytosine residues and dicarboxyl acidic acids have been found in turnip yellow mosaic virus (a virus closely related to BdmV) as a secondary mode of protein-RNA interaction (10,21). The presence of glutamic acid and aspartic acid in the mobile protein region of BdmV suggest that these interactions also take place in BdmV.

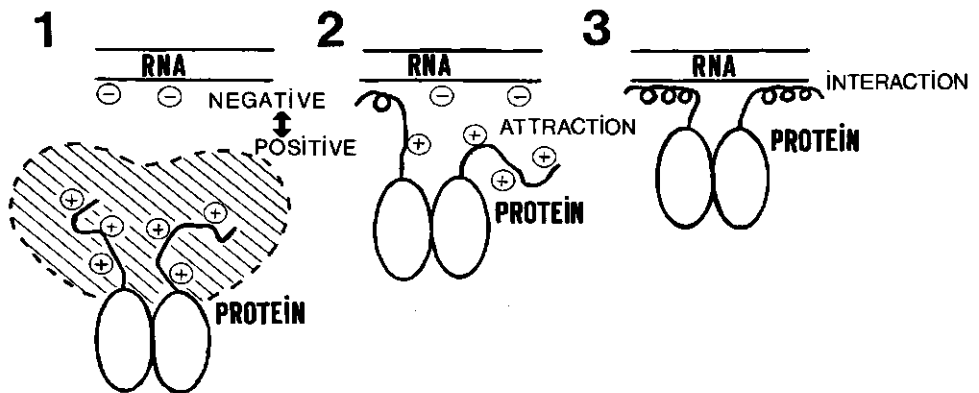


Figure 5.2. Schematical representation of the protein-RNA interaction model.

Protein-RNA interactions in CPMV strongly differ from those in all other viruses mentioned (10,22). No internal mobile amino acid residues are observed in any CPMV component, indicating that the assembly process is directed by other than phosphate-basic amino acid interactions. Interactions between virus protein and genome linked protein have been suggested to be important in the assembly process (22). The absence of a mobile RNA binding protein region supports this idea.

The results presented in this chapter have also been published in:

Vriend,G., Verduin,B.J.M., Hemminga,M.A., Schaafsma,T.J., (1982), FEBS Lt. 145, 49-52.

5.5 REFERENCES

- (1) Vriend,G., Hemminga,M.A., Verduin,B.J.M., Wit,J.L.de, Schaafsma,T.J., (1981), FEBS Lt. 134, 167-171.
- (2) Jardetzky,O., Akasaka,K., Vogel,D., Morris,S., Holmes,K.C., (1978), Nature 273, 564-566.
- (3) Wit,J.L.de, Hemminga,M.A., Schaafsma,T.J., (1978), J.Mag.Res. 31, 97-107.
- (4) Wit,J.L.de, Alma,N.C.M., Hemminga,M.A., Schaafsma,T.J., (1979), Biochemistry 18, 3973-3976.
- (5) Harrison,S.C., Olson,A.J., Schutt,C.E., Winkler,F.K., Bricogne,G., (1978) Nature 276, 368-373.
- (6) Abad-Zapatero,C., Abdel-Meguid,S.S., Johnson,J.E., Leslie,A.G.W., Rayment,I., Rossmann,M.G., Suck,D., Tsukihara,T., (1980), Nature 286, 33-39.
- (7) Sung,M.T., Dixon,G.H., (1970), Proc.Natl.Acad.Sci. 67, 1616-1623.
- (8) Warrant,R.W., Kim,S.H., (1978), Nature 271, 130-135.
- (9) Argos,P., (1981), Virology 110, 55-62.
- (10) Kaper,J.M., (1975), The Chemical Basis of Virus Structure, Dissociation and Reassembly, Elsevier, Amsterdam.
- (11) Verduin,B.J.M., (1978), J.Gen.Virology 39, 131-147.
- (12) Verduin,B.J.M., (1974), FEBS Lt. 45, 50-54.
- (13) Moline,H.E., Fries,R.E., (1974), Phytopathology 64, 44-48.
- (14) Ghabrial,S.A., Shepherd,R.J., Grogan,R.G., (1967), Virology 33, 17-25.
- (15) Heuss,K.L., Murthy,M.R.N., Argos,P., (1981), J.Mol.Biol. 146, 629-633.
- (16) Kammen,A.van, (1967), Virology 31, 633-642.
- (17) Wuthrich,K., (1976), NMR in Biological Research: Peptides and Proteins, North-Holland/American Elsevier, New York.

- (18) Moosic, J.P., (1978), Ph.D. Thesis, University of Wisconsin.
- (19) Rees, M.W., Short, M.N., Turner, D.S., (1980), John Innes Inst. 70th Annu. Rep., 102-103.
- (20) Stubbs, G., Warren, S., Holmes, K., (1977), Nature 267, 216-221.
- (21) Ehresman, B., Briand, J.P., Reinbolt, J., Witz, J., (1980), Eur. J. Biochem. 108, 123-129.
- (22) Winner, E., (1982), Cell 28, 199-201.

6 INTERACTION OF THE N-TERMINAL ARM OF CCMV PROTEIN WITH OLIGO-NUCLEOTIDES

6.1 INTRODUCTION

Cowpea chlorotic mottle virus (CCMV) is a spherical plant virus of the group of bromoviruses. It consists of RNA surrounded by 180 identical, icosahedrally arranged protein subunits (1). The virus is stable around pH 5.0 and low ionic strength ($\mu \sim 0.2$) (2). When the pH is raised to 7.5 at high ionic strength ($\mu > 0.3$), the virus dissociates into protein dimers and RNA almost free of protein (2). Both RNA and protein can be isolated and reassembled in vitro (2). In the absence of RNA the protein subunits can be assembled into empty protein capsids by lowering the pH to 5.0 (3). These empty protein capsids have the same structure as the native virus: a T=3 icosahedral surface lattice (4).

The protein subunit contains a basic N-terminal arm of 25 amino acids that can be cleaved by tryptic digestion (6). This arm has been shown to be the RNA binding part of the protein (7). In the absence of RNA, this arm is very mobile whereas immobilization occurs upon binding the RNA (7). The N-terminal arm contains 6 arginine and 3 lysine residues (5), and it has been suggested that interactions between the positive sidechains of these amino acids and the negative phosphate groups in the RNA are responsible for the protein-RNA interaction (7,8). The in vivo assembly process is very specific. In vitro however, no specificity has been observed and CCMV protein tends to associate with any poly-nucleotide, or even polyanions like polyvinyl sulphate or sodium dextran sulphate (9). Using oligo-uridylic acids of several lengths, Bancroft et al. (10) demonstrated that at least a minimum requirement of ~ 23 nucleotides is needed to induce capsid formation.

In this thesis the objective is to study molecular interactions during the assembly process of CCMV. In this chapter the binding of several oligo-nucleotides to CCMV protein is studied by $^1\text{H-NMR}$. From these binding experiments it is concluded that the secondary structure of the oligo-nucleotide plays an role in the process of binding. Also it is shown that interactions between the nucleotides and arginine and lysine can be observed. The part of the N-terminal protein arm which primarily interacts with the oligo-nucleotides is found to range from approximately residue 10 to 25.

6.2 MATERIALS AND METHODS

6.2.1 PREPARATION OF VIRAL COAT PROTEIN

Virus was purified as described before (11). Coat protein prepared from virus by the CaCl_2 method (12) was dialyzed against 300 mM NaCl, 10 mM MgCl_2 and 10 mM

sodium phosphate pH 5.0. H_2O in this solution was substituted by D_2O by extensive dialysis against the above solution made up in D_2O . In D_2O pH meter readings were taken without correction for the presence of D_2O . The protein concentration was initially 2.0 mg/ml.

To obtain protein lacking the N-terminal arm, coat protein was dissolved in 300 mM NaCl, 50 mM Tris-HCl pH 7.5 and treated for 24 h at 5°C with 5 mg Trypsin 30 Enzygel (Boehringer) per 100 mg of protein. After incubation the gel was removed by low speed centrifugation and the supernatant was dialyzed against 1M NaCl, 50 mM sodium acetate pH 5.0 to remove the small cleavage products and induce the formation of empty capsids.

6.2.2 OLIGO-NUCLEOTIDES

The oligo-ribonucleotide $(Ap)_8A$ was obtained from Boehringer Mannheim GmbH.

A mixture of oligo-A ribonucleotides ranging in length from ~ 14 to ~ 16 nucleotides was a kind gift of Dr. A. Heerschap and professor C.W. Hilbers. This mixture is referred to as $(Ap)_{14}A$ since the average size is ~ 15 nucleotides.

$(AT)_5$ (deoxy form) was a kind gift of professor J.H. van Boom.

6.2.3 NMR MEASUREMENTS

1H -NMR spectra were recorded with a Bruker WM500 spectrometer. Samples of 400-600 μl were measured at 7°C in the quadrature detection mode with D_2O lock. The acquisition time was 0.4 s and 2048 scans were taken. The sensitivity enhancement was 5 Hz. The ppm scale was relative to sodium 2,2-dimethyl 2-silapentane-5-sulphonate (DSS). The vertical scale was corrected for concentration differences between the samples of each series of experiments. The HDO peak was presaturated 2.0 sec with 1.8 W decoupling power. The samples were dissolved in D_2O containing 300 mM NaCl, 10 mM $MgCl_2$ and 10 mM sodium phosphate at pH 5.0 or pH 7.5.

6.3 RESULTS

6.3.1 THE OLIGO-NUCLEOTIDES

Figure 6.1A shows the 1H -NMR spectrum of $(Ap)_8A$ at pH 5.0. It has been shown that oligo-A can adopt a double stranded helix conformation at low pH (13,14). The wide range of resonance positions of the H2 and H8 protons (see figure 6.2) around 8 ppm, which is the result of ring-current effects due to base stacking,

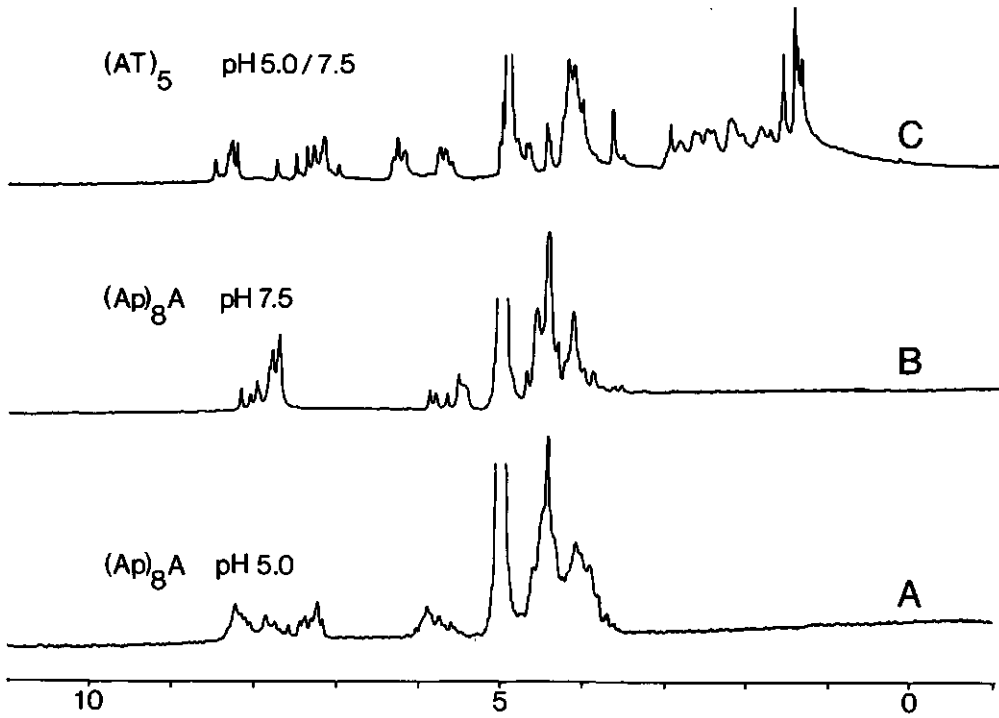


Figure 6.1. 500 MHz ^1H -NMR spectra of:
 A: $(\text{Ap})_8\text{A}$ at pH 5.0 (double stranded).
 B: $(\text{Ap})_8\text{A}$ at pH 7.5 (single stranded).
 C: $(\text{AT})_5$ at pH 5.0 and pH 7.5 (no differences observable).

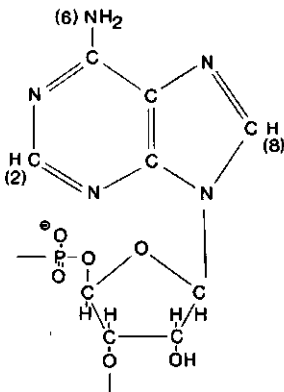


Figure 6.2. Schematic structure of the adenine group (2) and (8) indicate the protons observable around 8 ppm. (6) indicates the amino group.

confirms that this is also the case for $(Ap)_8A$. Figure 6.1B shows the spectrum of $(Ap)_8A$ at pH 7.5. At this pH the deprotonation of the amino group at position 6 (see figure 6.2) prevents the base pairing needed for forming a double stranded structure (13), but $(Ap)_8A$ shows base stacking at this pH as is concluded from an NMR melting study (G.Vriend, unpublished results). The structure of $(Ap)_8A$ at pH 7.5 and low temperature is concluded to be mainly a single stranded helix.

The same effects of formation of a double stranded helix at pH 5.0 and a single stranded helix at pH 7.5 as observed for $(Ap)_8A$, have been observed by Heerschap for the $(Ap)_{14}A$ oligo-nucleotide mixture (14).

The spectrum of $(AT)_5$ is shown in figure 6.1C. No differences could be observed between the spectra of $(AT)_5$ at pH 5.0 and pH 7.5. It is therefore assumed that $(AT)_5$ adopts a double stranded conformation at both pH values.

6.3.2 BINDING OF OLIGO-NUCLEOTIDES TO PROTEIN DIMERS

Figure 6.3A represents the spectrum of dimers of CCMV protein. The result of the addition of $(Ap)_8A$ and $(Ap)_{14}A$ is shown in the figures 6.3B and 6.3C respectively. The inserts in these two figures give the lowfield regions of an equal amount of oligo-nucleotide in the absence of protein dimers. It is seen that the linewidth, the relative positions and the intensities of the free observable resonances of the oligo-nucleotides are similar in the presence and in the absence of protein dimers. It is also observed that the arginine and the lysine resonances (around 3.0 ppm) are not influenced by the addition of $(Ap)_8A$ or $(Ap)_{14}A$. From these two facts it is concluded that neither $(Ap)_8A$ nor $(Ap)_{14}A$ binds to protein dimers at pH 7.5.

Figure 6.3D shows the result of the addition of $(AT)_5$ to protein dimers at pH 7.5. The total concentration of $(AT)_5$ (on monomer basis) is approximately 0.6 times the concentration of protein subunits. The insert in figure 6.3D shows the lowfield region of an equal amount of $(AT)_5$ in the absence of protein. Rate zonal centrifugation analysis of the assembly products of $(AT)_5$ and protein dimers at pH 7.5, although assembled at lower nucleoprotein concentrations and lower ionic strength, revealed the presence of nucleoprotein particles sedimenting around 80 S (see figure 6.4). Due to the formation of nucleoprotein particles with a molecular weight of $\sim 4 \times 10^6$, all resonances of the $(AT)_5$ and the protein molecules that are assembled, are broadened beyond detection. Figure 6.4 shows the results of a rate zonal sedimentation analysis of $(AT)_5$ -CCMV coat protein products. $(AT)_5$ and coat protein were assembled in molar ratios of 1:0.3 (A), 1:0.6 (B) and 1:1.5 (C). In figure 6.4 A and B free $(AT)_5$ remains at

the top of the gradients. Assembly with molar excess of $(AT)_5$ (see figure 6.4C) results in formation of both "virus" particles and "virus" doubles at low ionic strength. The "virus" doubles loose their extra protein coat at higher ionic strength.

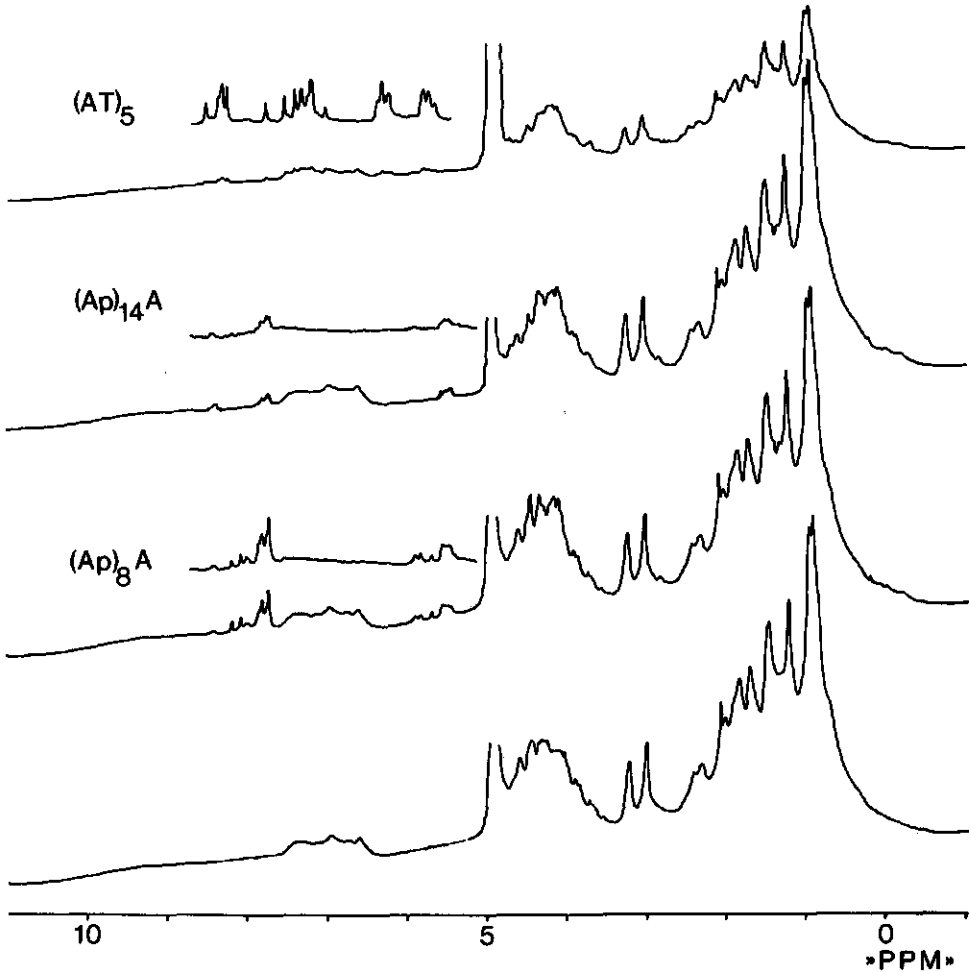


Figure 6.3. Binding of oligo-nucleotides to protein dimers at pH 7.5:

- A: Protein dimers.
- B: Protein dimers + $\sim 200 \mu\text{g}$ $(Ap)_8A$
- C: Protein dimers + $\sim 100 \mu\text{g}$ $(Ap)_{14}A$
- D: Protein dimers + $\sim 200 \mu\text{g}$ $(AT)_5$

The amount of protein was 2 mg. The sample size was 400 μl . Inserts show the lowfield regions of the spectra of the same amount of oligo-nucleotide as used in the binding experiment, but in the absence of protein.

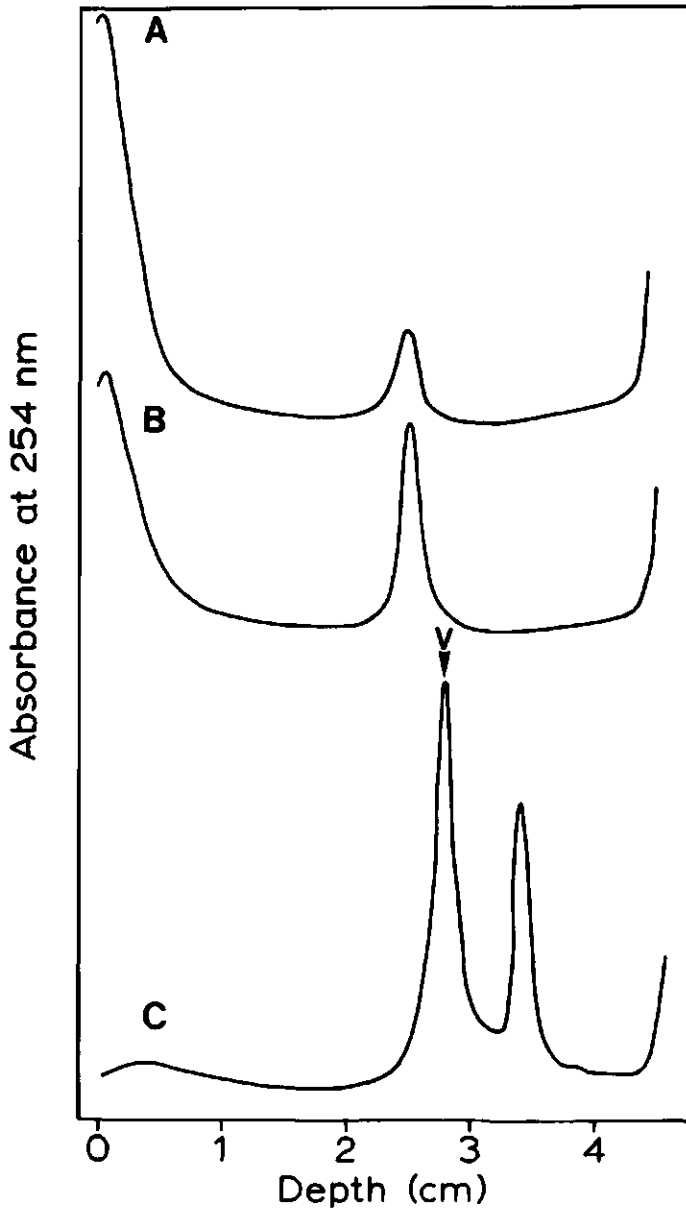


Figure 6.4. Rate zonal sedimentation analysis of $(AT)_5$ -CCMV coat protein assembly products. 20 μg of $(AT)_5$ was mixed with increasing concentrations of CCMV protein: 40 μg (A), 80 μg (B) and 200 μg (C) and the solutions (300 μl) were dialyzed against assembly buffer: 10 mM KCl, 5 mM MgCl_2 , 1 mM dithiothreitol and 10 mM Tris-HCl pH 7.5. 150 μl of each solution was layered on 10-50% (w/w) sucrose gradients in assembly buffer and centrifuged for 2h at 50,000 rpm and 5°C in a Beckman SW 55 rotor. The contents of the tubes were analysed as described in ref. 11. V represents the position of intact CCMV.

strength resulting in a solution of nucleoprotein particles sedimenting like virus and free coat protein dimers (20).

From these experiments it can be estimated that one double stranded $(AT)_5$ molecule binds to 1.5-3 protein subunits.

6.3.3 BINDING OF OLIGO-NUCLEOTIDES TO EMPTY CAPSIDS

Figure 6.5 shows some titration experiments of empty protein capsids at pH 5.0 with $(Ap)_8A$. At this pH the nucleotide adapts a double stranded helix conformation. The NMR results indicate that now one double stranded $(Ap)_8A$ binds

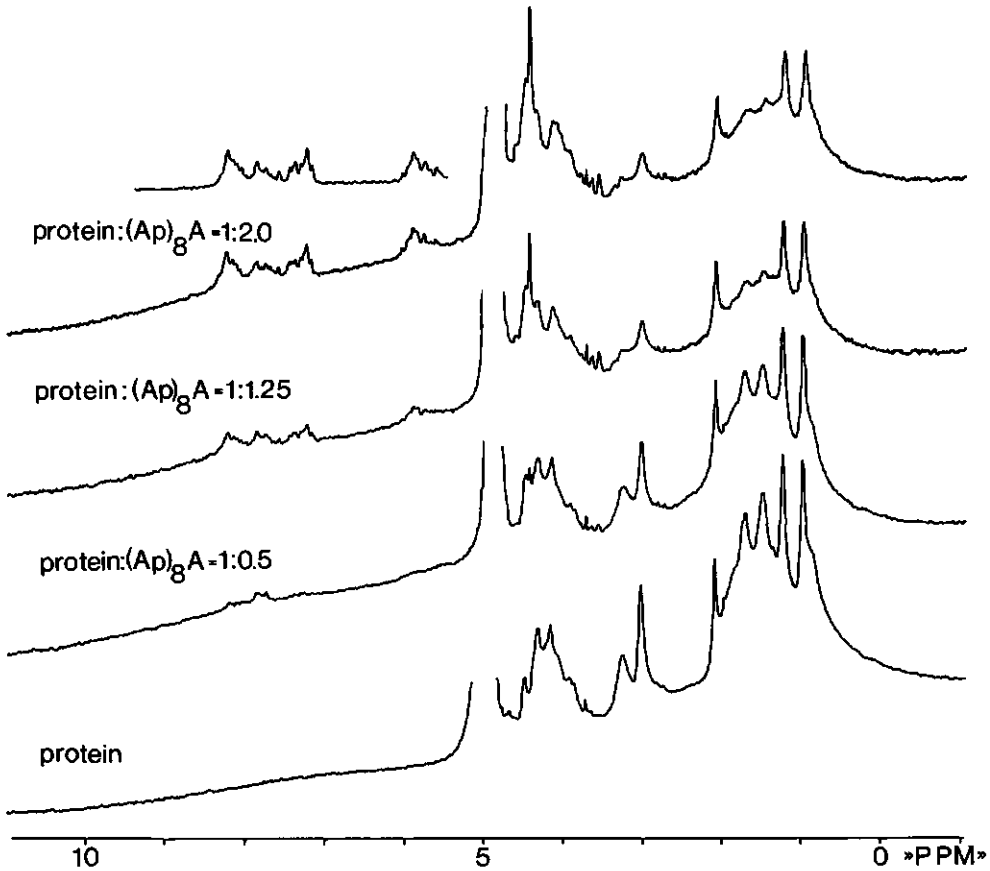


Figure 6.5. Titration of empty protein capsids with $(Ap)_8A$. The indicated ratios are between protein subunits and oligo-nucleotide on a monomer basis. The protein concentration is $\sim 0.25 \mu\text{Mol/ml}$. The insert shows the lowfield region of a $\sim 0.25 \mu\text{Mol/ml}$ $(Ap)_8A$ solution (monomer concentration) at pH 5.0.

bases could remain mobile after binding of the oligo-nucleotide to the empty capsids.

Figure 6.7 shows the result of the titration of empty protein capsids with $(AT)_5$ at pH 5.0. Although many protein peaks in the 1-5 ppm region overlap with $(AT)_5$ peaks, it can be seen that the same amino acids remain mobile as is observed for the titration with $(Ap)_8A$.

Figure 6.8 shows the result of the addition of $(Ap)_8A$ and $(AT)_5$ to protein lacking the N-terminal arm, both in the dimer form at pH 7.5 and in the form of empty capsids at pH 5.0. From the fact that no effect on the nucleotides is ob-

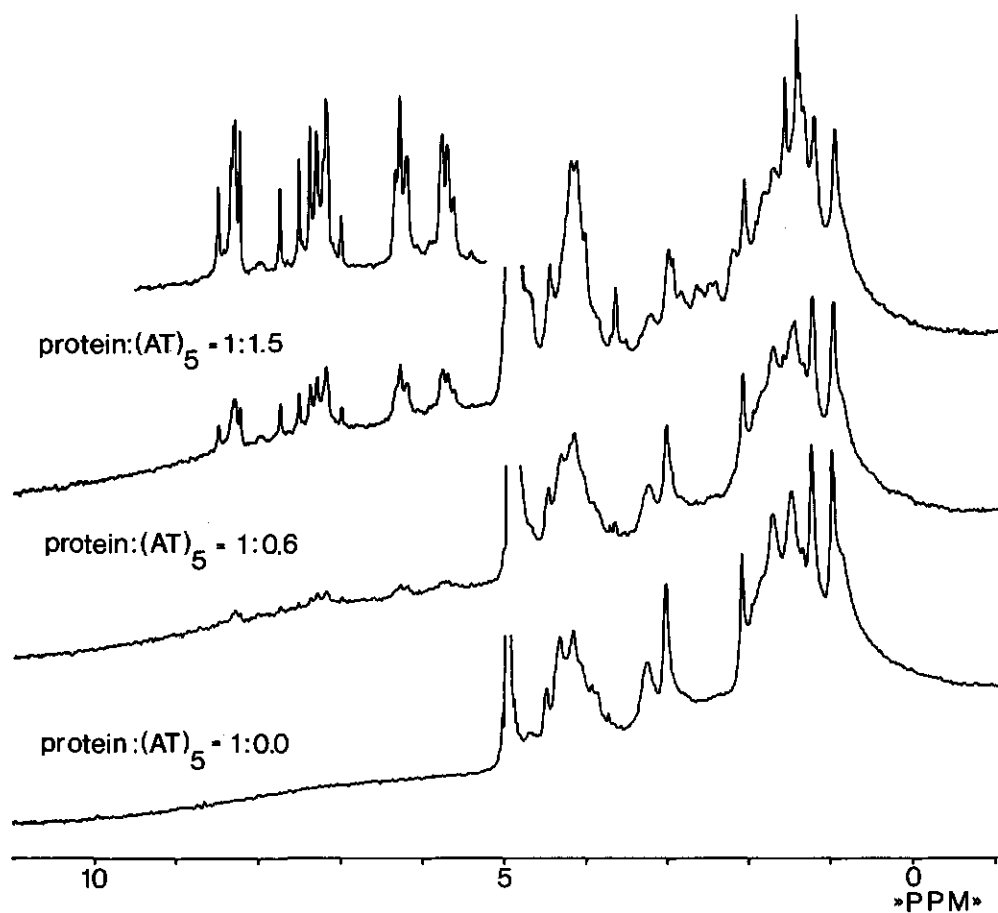


Figure 6.7. Titration of empty protein capsids with $(AT)_5$. The indicated ratios are between protein subunits and oligo-nucleotide on a monomer basis. The protein concentration is $\sim 0.25 \mu\text{Mol/ml}$. The insert shows the lowfield region of a $\sim 0.25 \mu\text{Mol/ml}$ $(AT)_5$ solution (monomer concentration).

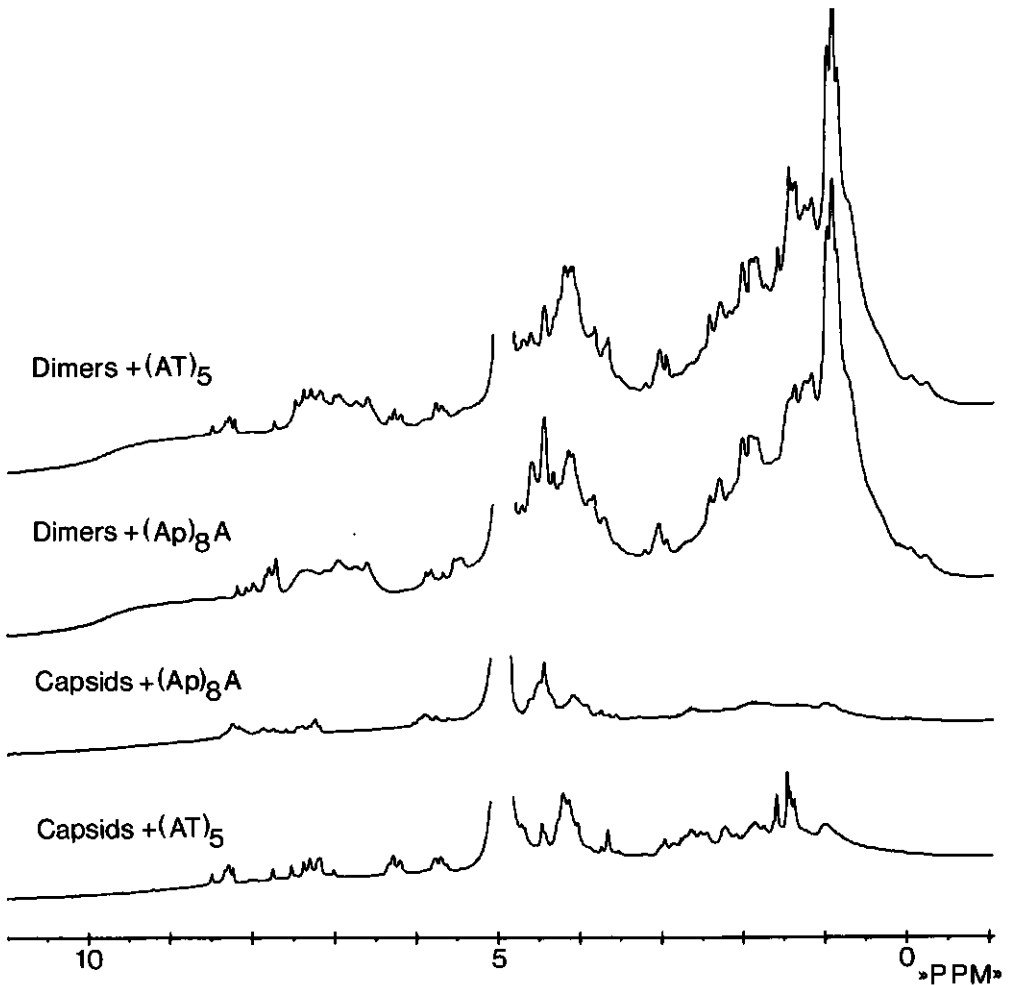


Figure 6.8. $^1\text{H-NMR}$ spectra showing the effect of the addition of $(\text{Ap})_8\text{A}$ or $(\text{AT})_5$ to empty protein capsids or protein dimers, both lacking the N-terminal regions.

served in any of the four cases, it is concluded that the nucleotides only bind to the N-terminal arm as could be expected from the assembly studies presented in chapter 4.

6.4 DISCUSSION

Upon binding of the double stranded oligo-nucleotides $(\text{AT})_5$ and $(\text{Ap})_8\text{A}$, immobilization of the N-terminal arm or a part of it occurs. The mode of interaction between the oligo-nucleotide seems to be different at pH 5.0 and

pH 7.5. At pH 5.0 binding of $(AT)_5$ or $(Ap)_8A$ to empty protein capsids results in immobilization of approximately the residues 10-25 in the N-terminal arm. When $(AT)_5$ binds to protein dimers at pH 7.5 the whole N-terminal arm is immobilized. Single stranded $(Ap)_8A$ does not bind to protein dimers at pH 7.5. The fact that these oligo-nucleotides and even poly-nucleotides like poly-A or viral RNA can bind to the empty capsids (B.J.M. Verduin unpublished results) suggests that the mobile N-terminal arms are located at the "outside" of the capsid. In the capsids formed from $(AT)_5$ and protein dimers at pH 7.5, the oligo-nucleotides and the N-terminal arms are most likely at the "inside" just as in the native virus. In this hypothesis "outside" stands for accessible from the outside, whereas "inside" means non-accessible. It is not certain whether the 7 N-terminal amino acids are immobilized in the capsids formed of dimers and $(AT)_5$ because they lack the space to move, or that this immobilization results from a different mode of interaction as compared with the interaction at pH 5.0.

In this study the main interest was the investigation of the structure and the mobility of the N-terminal arm upon binding oligo-nucleotides. From the present results however, also some conclusions can be drawn about structural aspects of the oligo-nucleotides which are important for the protein - oligo-nucleotide interaction. Double stranded $(AT)_5$ binds to protein dimers, whereas single stranded $(Ap)_{14}A$ and $(Ap)_8A$ do not bind. Double stranded $(Ap)_8A$ and $(AT)_5$ bind to protein capsids. Therefore it might be concluded that oligo-nucleotides only bind to the coat protein of CCMV if they possess a double stranded structure. This can only be a cautious conclusion since single strandedness versus double strandedness is not the only variable in the experiments. The other variables are: 1) The types of double stranded structures are different for the two oligo-nucleotides (13,15); 2) The binding experiments are performed at two different pH values, which could influence the adenine groups in the nucleotides; 3) The size of the nucleotides is different; 4) The number of phosphate groups in a double stranded oligo-nucleotide is twice the number present in the single stranded oligo-nucleotide. This may be important since Bancroft et al. already indicated that the size of the nucleotide could be important for interaction (10); 5) $(AT)_5$ is a deoxyribo-nucleotide, whereas $(Ap)_8A$ and $(Ap)_{14}A$ are ribonucleotides; 6) At pH 5.0 the protein forms empty capsids, whereas at pH 7.5 the protein is in the dimeric form.

For alfalfa mosaic virus (AMV) it has been found that the RNA contains some high-affinity sites for the coat protein (16). In CCMV some protein subunits remain attached to the RNA after dissociation of the virus (9). This suggests that the RNA of CCMV also contains some high-affinity sites for the coatprotein. Binding of some coat protein subunits to the high-affinity sites on the RNA is

required for infectivity of the RNA of AMV (17), but not for the RNA of CCMV (9). It is not known whether the coat protein subunits, which are bound to the high-affinity sites, are primers for the encapsidation of the RNA, as could be expected from a thermodynamical point of view (18). From the fact that CCMV-RNA which is completely devoid of coat protein subunits is infectuous, it might be concluded that the high affinity sites on the RNA are only needed for the specificity of the assembly process. Both the primary and the secondary structure of the 3'terminal regions of the 4 RNA's of CCMV are nearly identical (19). For all these 3'terminal regions t-RNA like structures are predicted (19). Therefore it is suggested that at least one high-affinity site is located in this 3'terminal region.

Future studies using different oligo-nucleotides should solve the remaining questions about the role of the base sequence and the secondary structure of the oligo-nucleotide in the process of protein-RNA interaction. Once the high affinity site at the RNA's of CCMV is found, NMR studies on the binding of this oligo-nucleotide or parts of it to the coat protein of CCMV can provide the basis for understanding the molecular principles of protein-RNA interaction which are at work during the assembly of CCMV.

6.5 REFERENCES

- (1) Bancroft, J.B., Hills, G.J., Markham, R., (1967), *Virology* 31, 454-379.
- (2) Bancroft, J.B., Hiebert, E., (1967), *Virology* 32, 354-356.
- (3) Bancroft, J.B., Wagner, G.W., Bracker, C.E., (1968) *Virology* 36, 146-149.
- (4) Finch, J.T., Bancroft, J.B., (1968), *Nature* 220, 815-816.
- (5) Rees, M.W., Short, M.N., Turner, D.S., (1980), *John Innes Inst. 70th Annu. Rep.*, 102-103.
- (6) Chidlow, J., Tremain, J.H., (1971), *Virology* 43, 267-278.
- (7) Vriend, G., Hemminga, M.A., Verduin, B.J.M., Wit de, J.L., Schaafsma, T.J., (1981), *FEBS Lt.* 134, 167-171.
- (8) Kaper, J.M., (1975), *The Chemical Basis of Virus Structure, Dissociation and Reassembly*, North Holland Publishing Company, Amsterdam Oxford, Chapter 7.
- (9) Verduin, B.J.M., (1978), *Thesis*, Agricultural University, Wageningen.
- (10) Bancroft, J.B., Hiebert, E., Bracker, C.E., (1969), *Virology* 39, 924-930.
- (11) Verduin, B.J.M., (1978), *J.Gen.Virol.* 39, 131-147.
- (12) Verduin, B.J.M., (1974), *FEBS Lt.* 45, 50-54.
- (13) Geerdes, H.A.M. (1979), *Thesis*, Catholic University, Nijmegen.
- (14) Heerschap, A., Unpublished results.

- (15) Vorlickova, M., Kypr, J., Sklenar, V., (1983), *J. Mol. Biol.* 166, 85-92.
- (16) Zuidema, D., Bierhuizen, M.F.A., Cornelissen, B.J.C., Bol, J.F., Jaspers, E.M.J., (1983), *Virology* 125, 361-369.
- (17) Smit, C.H., Roosier, J., Vloten-Doting van, L., Jaspars, E.M.J., (1981), *Virology* 112, 169-173.
- (18) Srinivasan, S., Jaspars, E.M.J., Hinz, H.J., (1977), *FEBS Lt.* 80, 288-290.
- (19) Ahlquist, P., Dasgupta, R., Kaesberg, P., (1981), *Cell* 23, 183-189.
- (20) Bancroft, J.B., (1971), *Adv. Virus Res.* 16, 99-134.

7 SECONDARY STRUCTURE PREDICTION OF CCMV PROTEIN

7.1 INTRODUCTION

In the previous chapters it has been suggested that the N-terminal arm of CCMV protein adopts an α -helix conformation upon interaction with nucleotides. This idea is supported by the fact that the N-terminal arm STNV is found to be partly in an α -helix conformation (1).

Chapter 9 contains experimental evidence for the presence of some secondary structure elements in the N-terminal arm of CCMV protein, but it is not clear which type of structure is observed. Also these structural elements cannot be attributed to a particular section of the N-terminal arm. In this chapter three different prediction methods are used to get an indication about the presence and location of α -helices, β -sheets and turns in the coat protein of CCMV. The methods used are those described by Chou and Fasman (2,3), Burgess et al. (4), and Lim (5,6). The first two methods use a protein data set consisting of many proteins for which both the primary protein structure and the X-ray structure are solved, to derive so-called structural potentials for the 20 common amino acids. These structural potentials are related to the probability of occurrence of the amino acids in one of the three main regular structures (α -helix, β -sheet and turn). It has been shown that these structural potentials can be improved for the secondary structure prediction of the protein of interest by extension of the data set on which the prediction method is based, with proteins related to that protein of interest (7). Therefore the parts of the structures of the coat proteins of SBMV and STNV that can be seen in X-ray electron density maps have been used to enhance the performance of these two secondary structure prediction methods for spherical plant viruses. The method of Lim can not be optimized in the same way since this method is based on general principles of protein folding rather than on a statistical analysis of a protein data set.

The methods of Chou and Fasman (2,3) and Lim (5,6) are modified to enable the simulation of charge neutralization of the arginine and lysine sidechains. This was done since in the previous chapter it has been shown that interaction between these amino acids and the nucleotide play an important role in the process of protein-RNA interaction.

7.2 SECONDARY STRUCTURE PREDICTING METHODS

7.2.1 CHOU AND FASMAN

The method described by Chou and Fasman (2,3) predicts secondary structure from the primary structure alone. Although the method is very simple, its results are as good as for any other method (8-11). The basic principle of the method is an indirect comparison of an unknown protein with a large number of proteins for which both the primary and secondary structure are known. The frequency of every amino acid in the protein data set in any of the three main conformations (α -helix, β -sheet and turn) is compared with the overall abundance of this amino acid. From these frequencies structural potentials are derived. These structural potentials are then used to predict the secondary structure of the protein of interest. The rules for using these potentials have been formulated by Chou and Fasman (3).

Robson et al. (12) and Klapper (13) studied the statistical relevance of the use of a too small protein data set. From their results it can be concluded that in general, extension of the protein data set does not implicitly enhance the reliability of methods derived from it. It has been shown by Argos et al. (14) that extension of the protein data set used to derive the structural potentials for the Chou and Fasman method does not automatically lead to an increase of the reliability of this method. However, it has been shown that extension of the protein data set with proteins related to the protein of interest, will enhance the reliability of the prediction (4,7). Therefore the structures of the coat proteins of SBMV and STNV are added to the the protein data set of Chou and Fasman, and the structural potentials were then calculated again.

The reliability of the prediction was further enhanced by optimizing the structural potentials calculated in this study for the prediction of the secondary structure of these two proteins (the details of this optimization are presented in appendix C).

Argos et al. (14) doubled the protein data set of Chou and Fasman. This resulted in changes up to 25% for some of the structural potentials for α -helix or β -sheet. The predictions using these new potentials did not differ very much from the predictions using the original potentials. This indicates that the succes of the prediction method does not significantly depend on non-systematic changes in the structural potentials. In this study the structural potentials were kept within 5% from their initial value during the process of optimization for the prediction of the secondary structure of SBMV and STNV.

Since it is found that the coat proteins of spherical plant viruses have a

remarkable high amount of β -sheets, the cutoff value for the decision whether a residue is predicted to be part of a β -sheet or not (the β -sheet boundary) has been decreased by 2%. A larger decrease did not result in an increased number of correctly predicted residues.

The optimization process resulted in general in a decrease of the α -helix structural potentials except for histidine, methionine and serine. The β -sheet structural potentials were in general increased. The strongest increase was found for the β -sheet structural potentials of alanine, glutamic acid and serine.

Despite the addition of two proteins to the protein data set and the subsequent optimization of the potentials, no potential differed more than 12% from the value originally derived by Chou and Fasman (2,3).

7.2.2 BURGESS, PONNUSWAMY AND SCHERAGA

Burgess et al. (4) have used a different procedure to derive the structural potentials for α -helix and β -sheet. They counted the occurrence of every amino acid in the protein data set within the areas in the Ramachandran plot corresponding to an α -helix or a β -sheet. The turn potentials have the same function as in the method of Chou and Fasman, the only difference being that nine positions around the centre of the turn are included, whereas Chou and Fasman include only four positions. The Chou and Fasman method defines arbitrary α -helix and β -sheet boundaries, whereas in the method of Burgess et al. these boundaries are made a function of the overall amino acid content of the protein for which the secondary structure is predicted.

Since neither the Ramachandran plots of SBMV and STNV nor the full protein data set used by Burgess et al. are immediately suitable for practical application, the method of Burgess et al. is only optimized for the prediction of the secondary structure of the coat proteins of SBMV and STNV using the following procedure:

- 1) The structural potentials for α -helix and β -sheet are optimized for the prediction of the secondary structure of the coat proteins of SBMV and STNV in a similar way as done to the method of Chou and Fasman. The maximum change that was allowed for any structural potential was increased to 10%.
- 2) After this optimization of the structural potentials the mean structural tendency (defined as the lowest of the α -helix and the β -sheet boundary) was optimized. It was found that the maximum number of correctly predicted residues was obtained when this mean structural tendency was increased by 4%.

The optimization process showed in general the same tendency of decreasing α -helix potentials and increasing β -sheet potentials. The increase of performance due to the increase of the mean structural tendency was mainly the result of less overpredictions of α -helices.

7.2.3 LIM

From a study on the architectural principles of the packing of polypeptide chains and the influence of water on this process, Lim has derived a large set of rules for the prediction of the secondary structure of proteins (5,6). Lim divides the amino acids roughly in four categories: large hydrophobic, large hydrophylic, small hydrophobic and small hydrophylic. Lim explicitly warns that his rules can not be applied to residues in quaternary structure areas, near cofactors or around active centers of proteins. Since this method is based on general principles rather than on the use of a protein data set, it is not well possible to optimize the method for the prediction of viral coat proteins in the same way as done for the methods of Chou and Fasman, and Burgess et al.

7.2.4 MODIFICATIONS

A close spatial proximity of equally charged residues is hardly ever observed in proteins. Especially in an α -helix the occurrence of equally charged residues separated by 0, 2 or 3 other residues (implying a close spatial proximity) is much less than statistically expected (5,6,15). In figure 7.1 a double α -helical net (16) of the N-terminal arm of CCMV protein is presented. As can be seen, an α -helical fold of this arm would result in the occurrence of many pairs of positive charged arginine and lysine residues separated by 0, 2 or 3 other residues. Such a structure will therefore be very unlikely due to the strong repulsion of positive charges. The protein data set of Chou and Fasman contains only very few pairs or clusters of positively charged amino acids. Therefore it may be expected that the structural potentials of arginine and lysine following from this method are not correct when applied to the N-terminal arm of CCMV protein. The prediction using the normal potentials for these two amino acids is representative for the case when the charges of arginine and lysine are neutralized as in globular proteins or in the presence of nucleic acid phosphate groups. In order to simulate the α -helix breaking character of arginine and lysine in the N-terminal arm of CCMV protein in the absence of charge neutralising nucleotides, the secondary structure of this arm is also predicted

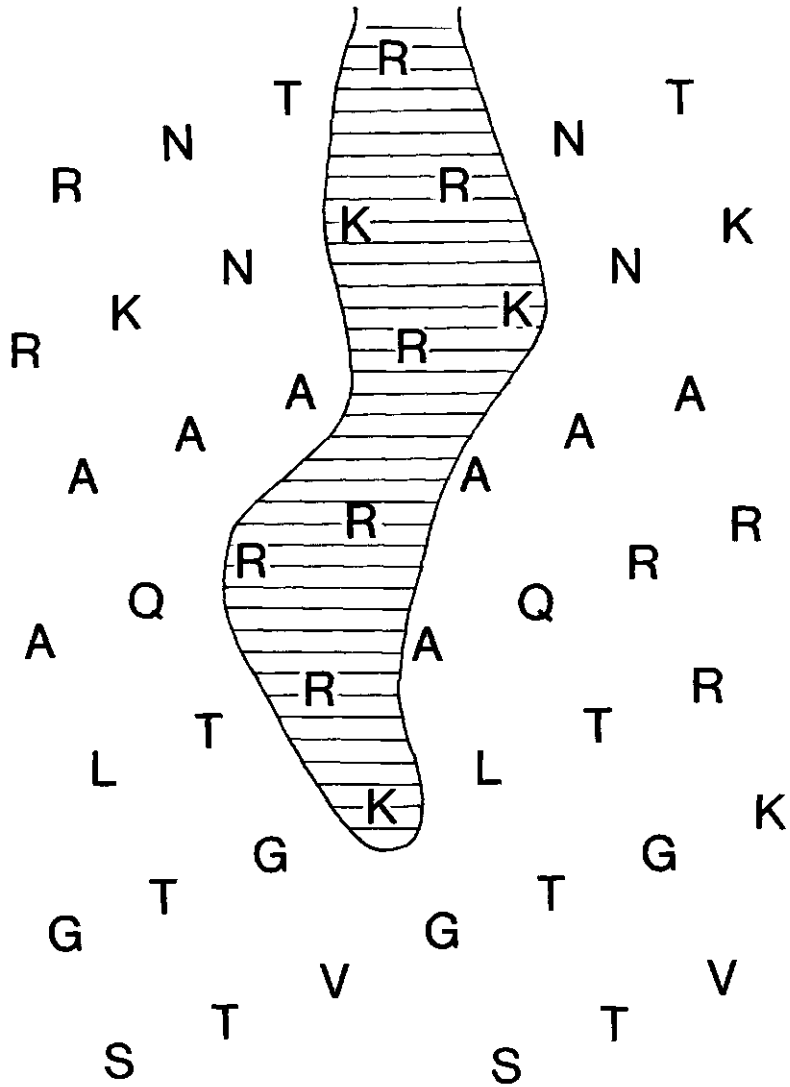


Figure 7.1. A double α -helical net representation (16) of the N-terminal arm of CCMV. In this net two arms are placed next to each other, in order to enable the observation of the α -helical conformation of the N-terminal arm from all directions. The distances between adjacent residues is proportional to the distances between their $C\alpha$ atoms in an α -helix. The region containing all adjacent arginine (R) and lysine (K) residues is shaded.

using α -helix structural potentials for arginine and lysine that were lowered by 20%. In this case arginine changes from a helix indifferent residue into a weak helix breaker, whereas lysine changes from a weak helix former into a helix indifferent residue.

Using the same motivation, the method of Lim was used in two different ways. First the normal classification of large hydrophylic for arginine and lysine was used. Second, charge neutralization was simulated by classifying arginine and lysine as large hydrophobic.

7.3 RESULTS AND DISCUSSION

The methods of Chou and Fasman, and Burgess et al. have been optimized for the prediction of the secondary structure of the coat proteins of SBMV and STNV. In tabel 7.1 the results of the prediction of the secondary structure of these two proteins, using the optimized methods, are compared with the stuctural data derived from X-ray studies. From this tabel it can be seen that 75% of all secondary structural elements are predicted with only some faults in the borders of these elements. Both methods tend to underpredict α -helices. A general increase of the prediction of α -helices by lowering the α -helix boundary, or by increasing all α -helix potentials with a certain factor did increase the number of predicted α -helices, but unfortunately there were more incorrectly than correctly predicted new helices. On the average, 58% of all residues is predicted correctly, which is rather satisfying when compared to the results of other secondary structure prediction studies (8-11).

Figure 7.2 presents the results of the prediction of the secondary structure of the coat protein of CCMV using the optimized methods. For this prediction the normal potentials for arginine and lysine were used, which represents the case of charge neutralization. In general it is seen that both methods agree rather well. For most structural elements only small differences in the borders are observed. For these cases the total prediction is taken as the average of the two predictions. In case of more severe disagreement between these two methods, the method of Lim has been applied to make the desiscion for the total prediction. For the residues 165-189 no total prediction is given, since in this region there is a complete disagreement between the two methods.

When the relative positions of the structural elements in CCMV in figure 7.2 are compared with the tertiary structure of SBMV and STNV, it is seen that the so called " β -role" (see figure 1.5) observed in these viruses is probably also present in CCMV. Note that SBMV and STNV are added to the protein data set of Chou and Fasman. The method of Burgess et al. is also optimized using these two

	110														120										
	S	G	T	V	K	S	C	V	T	E	T	Q	T	T	A	A	A	S	F	Q	V	A	L	A	V
B	T	T	T				β	β	β	β	β								α	α	α	α	α	α	α
C+F	T	T	β	β	β	β	β	β	β	β	β	β	β	β	β	β	β		α	α	α	α	α	α	α
Total	T	T	β	β	β	β	β	β	β	β	β	β	β	β	β	β	β		α	α	α	α	α	α	α

	130										140							150							
	A	D	N	S	K	D	V	V	A	A	M	Y	P	E	A	F	K	G	I	T	L	E	Q	L	A
B	α	T	T	T	T	T					T	T									α	α	α	α	α
C+F	α	α	T α	T α	T α	T α	α	α	α	α	α					β	β	β	β	β	β	α	α	α	α
Total	α	T	T	T	T	T	α	α	α	α	α											α	α	α	α

	160											170													
	A	D	L	T	I	Y	L	Y	S	S	A	A	L	T	E	G	D	V	I	V	α	L	E	V	E
B	α	α	α	β	β	β	β	β	β	β					T	T	T	T			α	α	α	α	α
C+F	α	α		β	β	β	β	β									β	β	β	β	β				
Total	α	α	α	β	β	β	β	β	β																

	180										189				
	α	V	R	P	T	F	D	D	S	F	T	P	V	Y	
B	α		T	T	T	T	T	T	T						
C+F		β	β	β	β	β	T	T	T	β /T	β	β	β	β	
Total															

proteins. The prediction of a large number of β -sheets in CCMV protein may partly arise from the fact that SBMV and STNV also contain a high percentage of β -sheet. The close structural resemblance of SBMV, STNV and CCMV makes the secondary structure prediction of CCMV more reliable. Indeed, a high percentage of β -sheet has been found in the coat protein of CCMV using the laser Raman spectroscopy technique (17).

Figure 7.3. shows the prediction of the secondary structure of the N-terminal arm of CCMV protein using the prediction methods both in the normal way and after modification to take into account charge effects. It is clearly seen that an α -helix is predicted from approximately residue 10 to residue 20 only in the absence of positive charges.

The double helical net of the N-terminal arm of CCMV protein, presented in figure 7.1, indicates that when this arm adopts an α -helix conformation, the positive charges are located at one site of the helix. For other spherical plant viruses this charged part of the α -helix in the N-terminal arm has been observed or predicted (1,18) to be exposed to the RNA. In STNV the other, hydrophobic, site of the N-terminal α -helix is found to interact with the hydrophobic sites of the N-terminal helices of neighbouring protein subunits (1). The prediction for the N-terminal arm of CCMV protein suggests that this may be also occurring for CCMV.

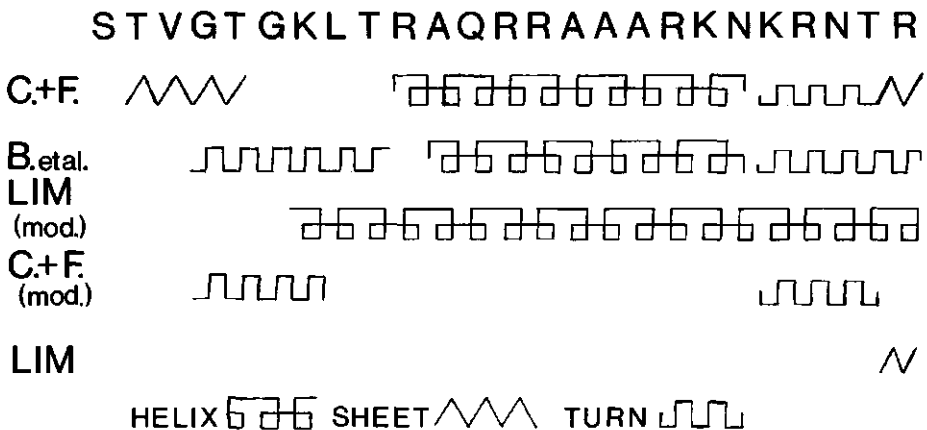


Figure 7.3. Secondary structure predictions for the N-terminal arm of CCMV protein according to:

C + F : the optimized method of Chou and Fasman.

B. et al.: the optimized method of Burgess et al.

LIM (mod): the method of Lim after neutralizing arginine and lysine sidechains.

C+F (mod): the optimized method of Chou and Fasman after lowering the α -helix potentials for arginine and lysine.

LIM : the method of Lim.

The first three predictions use neutralized sidechains. The last two predictions use charged sidechains.

TABEL 7.1. Results of the predictions of the secondary structure of SBMV and STNV using the optimized methods of Chou and Fasman, and Burgess et al. The nomenclature for the structural elements is taken from figure 1.4. A prediction of a structural element is called correct when both boundaries are predicted within 2 residues from the boundary found by X-ray crystallography.

Structure	Chou and Fasman	Burgess et al.
SBMV β A (43-51)	Correct	Correct
β A (58-64)	Correct	Correct
β B	9 Residues C-terminal not predicted; turn 64-65 predicted β -sheet	Correct β -sheet
β C	Under predicted at both ends	Correct (Prediction continued 7 residues in C-terminal direction)
α A	Correct	Missed
β D	Correct	Correct
β E	Correct	Missed
α B	Missed	Missed
β F	Correct	Missed
α C	Predicted β -sheet	Missed
β G (185-189)	3 Residues shifted N-terminal	Correct (Overpredicted N-terminal)
β G (197-199)	Missed	Correct
α D	Missed	Underpredicted N-terminal
α E	Correct	Missed
β H	4 Residues shifted N-terminal	3 Residues shifted N-terminal
β I	Underpredicted at both ends	3 Residues shifted C-terminal
Overpredictions:	254-260 Predicted α -helix	140-144 Predicted β -sheet
STNV N-terminal		
α -helix	Correct	Correct
β B	Underpredicted at both ends	Missed
β C	Correct	Missed
β D	Correct (69-73 Missed)	Correct
β E	Correct	Correct
α (101-106)	4 Residues shifted N-terminal	Correct
α (118-123)	Predicted β -sheet	Predicted β -sheet
β F	Correct	Correct
β G	Correct	Underpredicted C-terminal
β H	Correct	Correct
β I	Correct 185-188 Predicted turn	Correct
Overpredictions:	55-60 Predicted α -helix	None

7.4 REFERENCES

- (1) Liljas, L., Unge, T., Jones, T.A., Fridborg, K., Lovgren, S., Skoglund, U., Strandberg, B., (1982), *J.Mol.Biol.* 159, 93-108.
- (2) Chou, P.Y., Fasman, G.D., (1974), *Biochemistry* 13, 211-221.
- (3) Chou, P.Y., Fasman, G.D., (1974), *Biochemistry* 13, 222-245.
- (4) Burgess, A.W., Ponnuswamy, P.K., Scheraga, H.A., (1974), *Isr.J.Chem.* 12, 239-286.
- (5) Lim, V.I., (1974), *J.Mol.Biol.* 88, 857-872.
- (6) Lim, V.I., (1974), *J.Mol.Biol.* 88, 873-894.
- (7) Nagano, K., (1973), *J.Mol.Biol.* 75, 401-420.
- (8) Matthews, B.W., (1975), *B.B.A.* 405, 442-451.
- (9) Schultz, G.E., Barry, C.D., Friedman, J., Chou, P.Y., Fasman, G.D., Finkelstein, A.V., Lim, V.I., Ptitsyn, O.B., Kabat, E.A., Wu, T.T., Levitt, M., Robson, B., Nagano, K., (1974), *Nature* 250, 140-142.
- (10) Lenstra, J.A., (1977), *B.B.A.* 491, 333-338.
- (11) Wu, T.T., Szu, S.C., Jernigan, R.L., Bilofsky, H., Kabat, E.A., (1978), *Biopolymers* 17, 555-572
- (12) Robson, B., Suzuki, E., (1976), *J.Mol.Biol.* 107, 327-356.
- (13) Klapper, M.H., (1977), *Bioch.Biophys.Res.Comm.* 78, 1018-1024.
- (14) Argos, P., Hanei, M., Garavito, R.M., (1978), *FEBS Lt.* 93, 19-24.
- (15) Maxfield, F.R., Scheraga, H.A., (1975), *Macromolecules* 8, 491-493.
- (16) Lim, V.I., (1978), *FEBS Lt.* 89, 10-14.
- (17) Verduin, B.J.M., Prescott, B., Thomas, G.J., (1983), *Biophys.J.* 41, abst. 156.
- (18) Argos, P., (1981), *Virology* 110, 55-62.

8 ENERGY CALCULATIONS ON THE N-TERMINAL ARM OF CCMV PROTEIN

8.1 INTRODUCTION

In this thesis one of the objectives is to elucidate the structural aspects of the N-terminal arm of CCMV protein. In the virus, the mobility of the arm is too slow to give observable resonances. In chapter 9 evidence is presented for the presence of some secondary structure in this N-terminal arm in empty protein capsids. Also it is concluded that this secondary structure is metastable and that conformational exchange between at least two states is possible (see chapter 9).

Unfortunately, the NMR technique only provides limited information about the structural properties of the N-terminal arms. The main limitation of this technique is caused by the lack of spectral resolution, due to the relative low mobility of the N-terminal arm ($\tau_c \sim 10^{-8} - 10^{-9}$ s) and the large number of almost equivalent groups that give rise to a strong overlap of resonances. With the present state of the NMR technique, the structure of the N-terminal arm can hardly be studied in greater detail than in the study presented in this thesis.

An alternative way to obtain information about the structural and dynamical aspects of the N-terminal arm is by calculational methods. A secondary structure prediction of the N-terminal arm is presented in chapter 7. This analysis shows that an α -helical conformation of a large part of the N-terminal arm is favoured in the virus, whereas this arm does not adopt a regular structure in the absence of RNA. These secondary structure prediction methods, however, only provide the relative changes for the occurrence of a peptide or a part of it in one of the regular structures (α -helix, β -sheet and turn). A much better way to obtain information about the N-terminal arm would be the application of molecular dynamics. This method can be used to derive the time average structure(s) and the frequencies of all internal motions (which then could be checked by NMR). Unfortunately, available computers are not fast enough to perform a molecular dynamics simulation for a sufficiently long timescale to derive conclusions about structure and mobility of a peptide with a length of 25 amino acids. Therefore in this chapter only an energy calculation of several conformational states of the N-terminal arm of the coat protein of CCMV is performed. This approach is certainly not ideal, but it is the best that can be done within a reasonable amount of computer time. The results obtained are compared with the 2D-NOE results presented in chapter 9.

For the study presented in this chapter the UNICEPP program (1) is used, which calculates the total molecular energy of a given conformation of the N-terminal

arm. This program is combined with a minimization procedure according to Powell (2,3), to find the conformation with the lowest energy. Structures of the N-terminal arm were evaluated on an interactive computer graphics system of the Agricultural University.

8.2 THE CALCULATIONAL METHOD

Since even the computer time needed for a reduced grid search in the conformational space is still far beyond the practical possibility, the following method was used for the calculation of the structure of the N-terminal arm of the coat protein of CCMV:

An interactive computer graphics system was used to derive an initial structure for the N-terminal arm. Three starting conformations were chosen in the β -sheet region and three in the α -helix region. From these starting points energy minimizations were proceeded until the nearest local minimum was found. This minimization was done three times, once with the partial charges of arginine and lysine according to a situation in which the overall net charge is zero, once with the overall positive charge retained as described by Momany et al. (4), and once with the positive charges retained, but reduced by a factor of two, as seems more reasonable in an aqueous solution containing negative counterions for the positively charged arginine and lysine sidechains.

The potential energy functions used as a basis for the program UNICEPP have been described in a series of articles (4-10). The characteristics of the potential energy functions that are of relevance for the study of the N-terminal protein arm of CCMV can be summarized as follows (1):

- 1) All $-\text{CH}_3$, $-\text{CH}_2$, aliphatic $-\text{CH}$ and aromatic $-\text{CH}$ groups are treated as single united atoms. The partial charges of the atoms that are combined in an united atom are added up to obtain the net partial charge of the united atom.
- 2) The potential energy of a pairwise electrostatic interaction is computed as a Coulomb potential between (united) atom-centered monopole partial charges Q_i and Q_j :

$$U = 332.0 \times Q_i \times Q_j / D \times R_{ij}$$

where Q_i and Q_j are in electrostatic charge units, the dielectric constant D is taken as 2.0, R_{ij} is the internuclear distance and 332.0 is the factor needed to give the energy in units of Kcal/mole.

- 3) A modified Lennard-Jones 6-12 potential is used to compute the nonbonded interaction energy of an atom pair i,j :

$$U = A \times R_{ij}^{-12} - B \times R_{ij}^{-6}$$

where R_{ij} is the internuclear distance and A and B are constants for the repulsion and attraction between each pair of (united) atoms (1).

Table 8.1. Partial charges of the atoms in the sidechains of arginine and lysine used in the structure calculations.

} means: combined in an united atom.

Charges are in electronic charge units.

The atoms are indicated according to the IUPAC-IUB nomenclature (11).

All other charges had the values given in (8).

Atom	total charge		
	Zero	Half retained	Retained
Lys C β }	-0.030	-0.030	-0.030
H β }	0.015	0.015	0.015
C γ }	-0.025	-0.025	-0.025
H γ }	0.020	0.020	0.020
C δ }	-0.030	-0.072	-0.115
H δ }	0.020	0.060	0.100
C ϵ }	0.110	0.075	0.050
H ϵ }	-0.005	0.067	0.120
N ζ	-0.395	-0.357	-0.320
H ζ	0.150	0.235	0.320
Arg C β }	-0.030	-0.030	-0.030
H β }	0.015	0.015	0.015
C γ }	-0.030	-0.030	-0.030
H γ }	0.030	0.030	0.030
C δ }	0.100	0.110	0.120
H δ }	0.010	0.010	0.010
N ϵ	-0.350	-0.325	-0.300
H ϵ	0.170	0.200	0.230
C ζ	0.480	0.530	0.580
N η 1	-0.420	-0.405	-0.390
H η 1	0.090	0.185	0.280
N η 2	-0.410	-0.400	-0.390
H η 2	0.160	0.220	0.280

4) Hydrogen bond energy is calculated by:

$$U = C \times R_{ij}^{-12} - D \times R_{ij}^{-10}$$

where C and D are the constants for the repulsion and the attraction between all pairs of proton donors and acceptors (1), and R_{ij} is the internuclear distance.

5) A torsional contribution is calculated for every dihedral angle in the united atom structure.

The values for the partial charges of the lysine and arginine sidechains used in this study are presented in tabel 8.1.

8.3 RESULTS

An energy minimization was performed three times for each of the six starting conformations used in this study. The backbone dihedral angles of the 25 amino acids of the N-terminal arm all had the same value in a starting conformation. The starting values for the dihedral angles and the values in the local minimum reached by the the minimization procedure are shown in the Ramachandran plots in figure 8.1. The relative energy contents of the structures in the local minima are summarized in tabel 8.2.

Tabel 8.2. Relative energy contents of the structures presented in figure 8.1. The values are expressed in Kcal/mol relative to an arbitrary reference point. The codes representing the structures refer to figure 8.1.

Structure (,)	Partial charges on arginine and lysine		
	Zero	Half retained	Retained
β I (-30,150)	-188	-71	+87
β II (-50,150)	-169	-80	+66
β III (-150,150)	-155	-71	+67
α I (-118,-57)	-134	- 96	+94
α II (-57,-57)	-227	-126	+63
α III (-50,-50)	-238	-127	+51

From the relative energy values presented in this tabel it is seen that in terms of the energy content calculated, the normal right handed helix is always the most favorable conformation, as could be expected from general ideas on protein structure (11). However, it is seen that the energy differences between

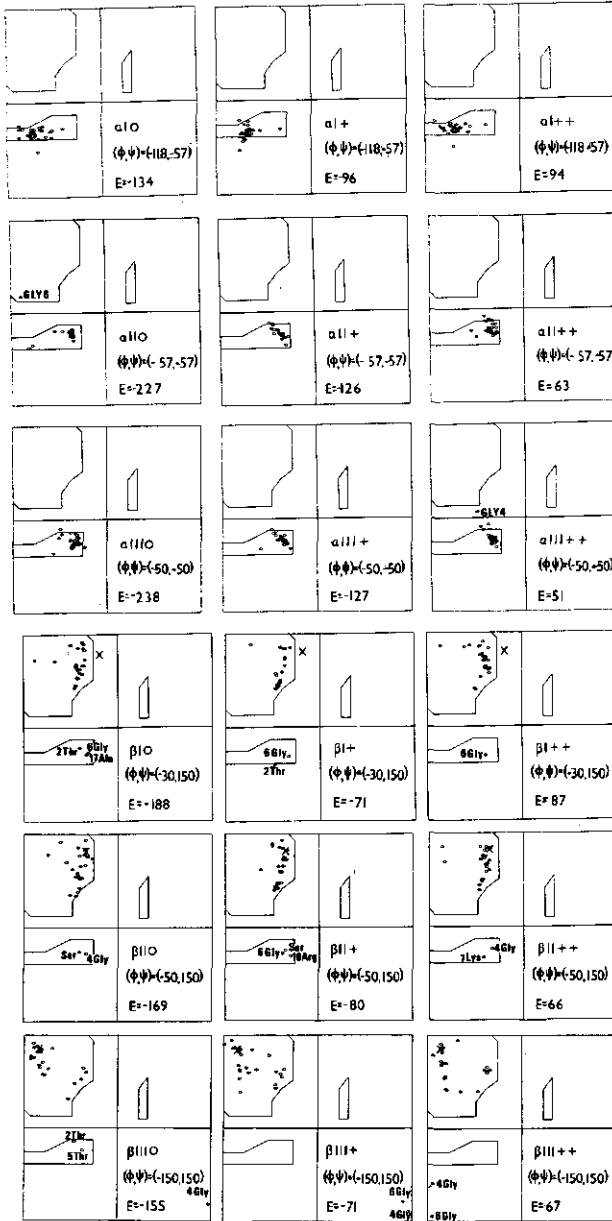


Figure 8.1. Ramachandran plots of the 18 structures of the N-terminal arm of CCMV at their local energy minimum.

Circles indicate backbone dihedral angles.

Crosses indicate the starting point for all dihedral angles.

Several residues for which the dihedral angles differ strongly from all other dihedral angles and from the starting values are indicated. As can be seen in most cases these residues lie in the region from residue 1 to residue 7.

the several structures become less when a larger positive charge is retained on the the arginine and lysine side chains. This observation does not result from the fact that different local minima were obtained in the presence or the absence of charges. This was checked by introducing the charges on the structures obtained by minimizing the energy in the absence of charges, and subsequent continuation of the minimization procedure until the new local minimum was reached. The values for the energy content obtained by this method could be compared to the values found by calculating the structure from the initial starting conformation with the positive charges present all the time. The same procedure was also performed the other way around. In this case the differences with the values obtained with the corresponding uncharged structure were somewhat larger (i.e. up to +31 Kcal for β I), but the differences between the α -helix and the β -sheet structures remained significant.

In chapter 9 a 2D-NOE study on empty capsids of the coat protein of CCMV is presented. From this study a few pairs of residues were found to have a close spatial proximity. These pairs are 8-leucine H δ - 9-threonine H γ and 2-threonine H γ - 3-valine H γ . Two other close spatial proximities were tentatively assigned: 4-glycine H α - 5-threonine NH and 6-glycine H α - 7-lysine NH. In tabel 8.3 the distances between these four pairs of protons in the 18 structures at minimal energy are presented.

From this tabel it is seen that the distance between 8-leucine H δ and 9-threonine H γ is 4.1 Ångstrom in the uncharged α I structure. In all other structures this distance is even larger. These distances are too long to give measurable NOE effects (see chapter 9). The 2-threonine H γ - 3-valine H γ distance is only short enough to give a measurable NOE effect in the uncharged structure α II. All distances calculated between 4-glycine H α and 5-threonine NH, and 6-glycine H α and 7-lysine NH are smaller than 3.2 Ångstrom and are therefore in good agreement with the 2D-NOE results from chapter 9. Note however, that these two distances can never become longer than 3.2 Ångstrom in any structure.

8.4 DISCUSSION

The best method for calculating the structure of a protein molecule is the application of molecular dynamics in every local energy minimum in the conformational space of this protein molecule. These calculations must be performed in the presence of water and all other solute molecules. When the solutions of the equations of motion for all of the atoms in the molecule are integrated over a long enough period of time, the resulting trajectory for a

Table 8.3. Distances between groups in the N-terminal arm (in Ångstrom units).

A = d(8-leucine H δ , 9-threonine H γ)

B = d(2-threonine H γ , 3-valine H γ)

C = d(4-glycine H α , 5-threonine NH)

D = d(6-glycine H α , 7-lysine NH)

In case of equivalent protons, the shortest distance is presented.

Structure	A	B	C	D
Uncharged β I	5.8	5.8	2.4	2.1
β II	5.9	5.8	2.4	2.2
β III	7.2	5.3	2.1	2.2
α I	4.1	5.0	2.2	2.1
α II	5.4	3.2	2.2	2.2
α III	4.6	4.2	2.5	2.6
Charges β I	6.2	5.6	2.2	2.2
half β II	5.8	5.0	2.4	2.1
retained β III	6.5	5.7	3.2	2.2
α I	4.4	5.7	2.7	2.6
α II	4.3	5.7	2.7	2.8
α III	4.4	3.8	2.1	2.1
Charges β I	5.9	6.0	2.2	2.3
retained β II	6.0	5.7	2.2	2.2
β III	5.8	6.0	2.9	2.2
α I	4.4	5.2	2.6	2.6
α II	5.3	4.7	2.6	2.6
α III	5.1	4.1	2.6	3.1

molecule may be used to compute the average space- and time-dependent properties of the system. Unfortunately, the amount of computer time needed for such an approach is far beyond the limit of what is practically possible. To overcome this problem several approximations must be introduced. The following approximations are generally accepted:

1) A large amount of computer time is saved by completely neglecting the solvent, in our case water. This neglect may be serious for the biological relevance of the result, since water has an essential role in biological systems and is ultimately involved in the structure and function of biomolecules.

Rosky and Karplus (13,14) describe the results of molecular dynamics simulations of an alanine dipeptide both in the presence and in the absence of water molecules. They find that neither the average structure, nor the magnitude of the local fluctuations of the dipeptide is strongly affected by the solvent environment. The observed small structural differences in the presence of water that were significant, can not be interpreted in terms of a simple description in which a water molecule is considered to be either bonded or non bonded to the dipeptide (13). Van Gunsteren and Karplus (15) applied molecular dynamics to bovine pancreatic trypsin inhibitor. They calculated the behaviour of this molecule in vacuum, in the presence of an apolar solvent, and in a crystalline environment. They started their molecular dynamics simulations from the X-ray structure and found that the equilibrium structure in the presence of solvent molecules or the crystal field is significantly closer to the X-ray structure than is the vacuum result. Large local deviations from the X-ray structure are observed in vacuum for amino acid sidechains that point out of the molecule. Although the differences are significant when compared to the experimental errors, they remain small in an absolute sense and it may be possible to model certain environmental effects without taking explicit account of the individual solvent atoms (16). In the calculations presented in this chapter, there is only interest in the structures of the N-terminal arm at local energy minima in the conformational space. In view of what has been noted above it is clear that no absolute conclusions can be drawn from the results presented in this chapter, since the calculated energy only contains an enthalpy contribution and no contribution from the entropy of the solute, nor from the entropy of the solvent is included.

2) Van Gunsteren and Berendsen describe different algorithms for integrating the equations of motion (17). They propose a scheme by which constraints can be incorporated in a molecular dynamics simulation. They found that the application of bond-length constraints reduce the required computer time by a factor of three. The major disadvantage of the introduction of these constraints is the loss of information of the high frequency motions (17). In the approach chosen for the calculation of the structure of the N-terminal arm of CCMV protein, the only effect of the introduction of bond-length constraints is that during the minimization process the pathway to the nearest local minimum is influenced. The same authors also discuss the introduction of bond-angle constraints, but this is not very important for the UNICEPP method since this method uses the so called united atom approximation for most groups present in a protein (1,10).

3) The number of interatomic interactions in the molecule of interest is reduced by the introduction of the united atom approximation (1). Dunfield et al. (10)

show that the differences between the results obtained by using the method in which every atom is treated individually and their method are very small. The calculational time is reduced by a factor of 2.5 by the introduction of this approximation.

4) Using the united atom approximation the N-terminal arm of CCMV contains 156 variable dihedral angles. Thus the energy surface is 156-dimensional, and is very complex. This energy surface contains many local minima and a minimization procedure will lead to the minimum in the same potential energy well as the conformation from which the computation was started. The main problem is that in order to reach the minimum of lowest energy (which even does not implicitly have to be the native structure) a procedure is required to surmount potential energy barriers. But even then, one can never be sure that the lowest energy minimum is found before a grid search through the whole conformational space is performed. Scheraga has suggested that the multiple minima problem can be overcome by taking all possible combinations of low lying minima for the single residues, and minimizing their energies (18). This method has been applied to the computation of the cyclic decapeptide gramicidin S, for which 10541 starting conformations were needed (19,20). Such a procedure however is too time-consuming for the N-terminal arm.

From the fact that no systematic correlation between the calculated structures and the results from the 2D-NOE study presented in chapter 9 can be observed, it is concluded that no detailed structural conclusions can be drawn from the results presented in this chapter. However, the results presented here support the conclusion drawn in the previous chapter that no structural preference is observed when the positive charges are retained on the arginine and lysine side chains, whereas in the absence of positive charges the α -helix conformation is favorable. The fact that no structural preference is observed when the positive charges are retained is also in agreement with the observed conformational exchange in the N-terminal arm (see chapter 9.2), since it was found that the conformations participating in the exchange process are equally populated. (assuming a two-site exchange model). Such an equality of populations can only be achieved when the structures participating in the exchange process possess an equal chance of occurrence, and thus an equal energy content. In this case however, both the enthalpy and the entropy should be regarded, which is not done in this study.

8.5 REFERENCES

- (1) UNICEPP, Quantum Chemical Program Exchange Library, QCPE Program Nr 361, Chemistry Department, Indiana University.
- (2) Powell, M.J.D., (1964), *Comp.J.* 7.
- (3) VAO4A, Harwell Subroutine Library, QCPE Program Nr 60, Indiana University.
- (4) Momany, F.A., McGuire, F.R., Burgess, A.W., Scheraga, H.A., (1975), *J. Phys. Chem.* 79, 2361-2381.
- (5) Yan, J.F., Momany, F.A., Hoffmann, R., Scheraga, H.A., (1970), *J. Phys. Chem.* 74, 420-433.
- (6) Momany, F.A., McGuire, R.F., Yan, J.F., Scheraga, H.A., (1970), *J. Phys. Chem.* 74, 2424-2438.
- (7) Momany, F.A., McGuire, R.F., Yan, J.F., Scheraga, H.A., (1971), *J. Phys. Chem.* 75, 2286-2297.
- (8) McGuire, R.F., Momany, F.A., Scheraga, H.A., (1972), *J. Phys. Chem.* 76, 375-393.
- (9) Lewis, P.N., Momany, F.A., Scheraga, H.A., (1973), *Isr. J. Chem.* 11, 121-152.
- (10) Dunfield, N., Burgess, A.W., Scheraga, H.A., (1975), *J. Phys. Chem.* 82, 2609-2616.
- (11) IUPAC-IUB Commission on Biochemical Nomenclature, (1970), *Biochemistry* 9, 3471-3479.
- (12) Lim, V.I., (1974), *J. Mol. Biol.* 88, 857-872.
- (13) Karplus, M., Rossky, P.J., (1980), *ACS symp.* 127, 23-42.
- (14) Rossky, P.J., Karplus, M., (1979), *J.A.C.S.* 101, 1913-1937.
- (15) Gunsteren van, W.F., Karplus, M., (1982), *Biochemistry* 21, 2259-2274.
- (16) Northrup, S.H., Pear, M.R., McCammon, J.A., Karplus, M., (1980), *Nature* 286, 304-305.
- (17) Gunsteren van, W.F., Berendsen, H.J.C., (1977), *Mol. Phys.* 34, 1311-1327.
- (18) Scheraga, H.A., (1983), *Biopolymers* 82, 2609-2616.
- (19) Dygert, M., Go, N., Scheraga, H.A., (1975), *Macromolecules* 8, 750-761.
- (20) Rachovsky, S., Scheraga, H.A., (1980), *Proc. Natl. Acad. Sci.* 77, 6965-6967.

9 ANALYSIS OF THE STRUCTURE AND MOBILITY OF THE N-TERMINAL ARM OF CCMV PROTEIN

In chapter 5 a model for the protein-RNA interaction in CCMV (and other viruses) is presented. Evidence for the successive steps of this interaction model is presented throughout the chapters 2-8. The model includes a random coil conformation for the N-terminal arm of the protein in the absence of RNA. In this chapter some results are presented modifying this idea. Although the studies presented in this chapter are still incomplete they already provide sufficient data to draw some conclusions about the structure of the N-terminal arm of CCMV in the absence of RNA.

9.1 A COMPARISON BETWEEN N-TERMINAL ARMS IN DIMERS AND EMPTY CAPSIDS

9.1.1 INTRODUCTION

In chapter 4 a $^1\text{H-NMR}$ study is presented on CCMV and its empty protein capsids both of intact protein and of protein lacking the N-terminal region. The spectrum of empty capsids of intact protein revealed the presence of internal protein mobility. From the fact that this mobility is not present in the empty capsids of protein lacking the N-terminal arm, it is concluded that the observed mobility is located in the N-terminal arm. Although the observed sharp peaks in the spectrum of empty capsids of native protein, indicating the internal mobility, result from the N-terminal arms only, it has not become clear from the results presented in chapter 4 whether every N-terminal arm is mobile or only a fraction of the arms. In this section it is shown that not every N-terminal arm is mobile in empty capsids of native protein. Also it will be shown that the N-terminal arm possesses extra mobility in the protein dimers and that some residues (e.g. the valine residue in the core that lies directly adjacent to the N-terminal arm) gain mobility upon removal of the N-terminal arm.

9.1.2 MATERIALS AND METHODS

The four protein samples used in this study were prepared as described in chapter 4. The protein preparations were dialyzed against 300 mM NaCl, 10 mM MgCl_2 and 10 mM sodium phosphate pH 5.0 (empty capsids) or pH 7.5 (dimers). In these protein solutions H_2O was substituted by D_2O by extensive dialysis against the same solutions made up in D_2O . The final concentration of protein varied from 5-8 mg/ml. $^1\text{H-NMR}$ spectra were recorded with a Bruker CXP300 spectrometer. Samples of 500 μl were measured with D_2O lock and without proton decoupling. The

temperature was kept at 7°C. The acquisition time plus pulse delay was 2.4 s and 400 scans were taken. The sensitivity enhancement was 5 Hz. The ppm scale was relative to DSS. In all figures the vertical scale was corrected for differences in protein subunit concentrations.

9.1.3 RESULTS AND DISCUSSION

9.1.3.1 THE NUMBER OF MOBILE N-TERMINAL ARMS IN EMPTY CAPSIDS

Figure 9.1 shows the spectra of empty capsids and dimers of both native protein and protein lacking the N-terminal arm. Figure 9.2A shows the highfield region of the difference spectrum of the dimers of intact protein and dimers

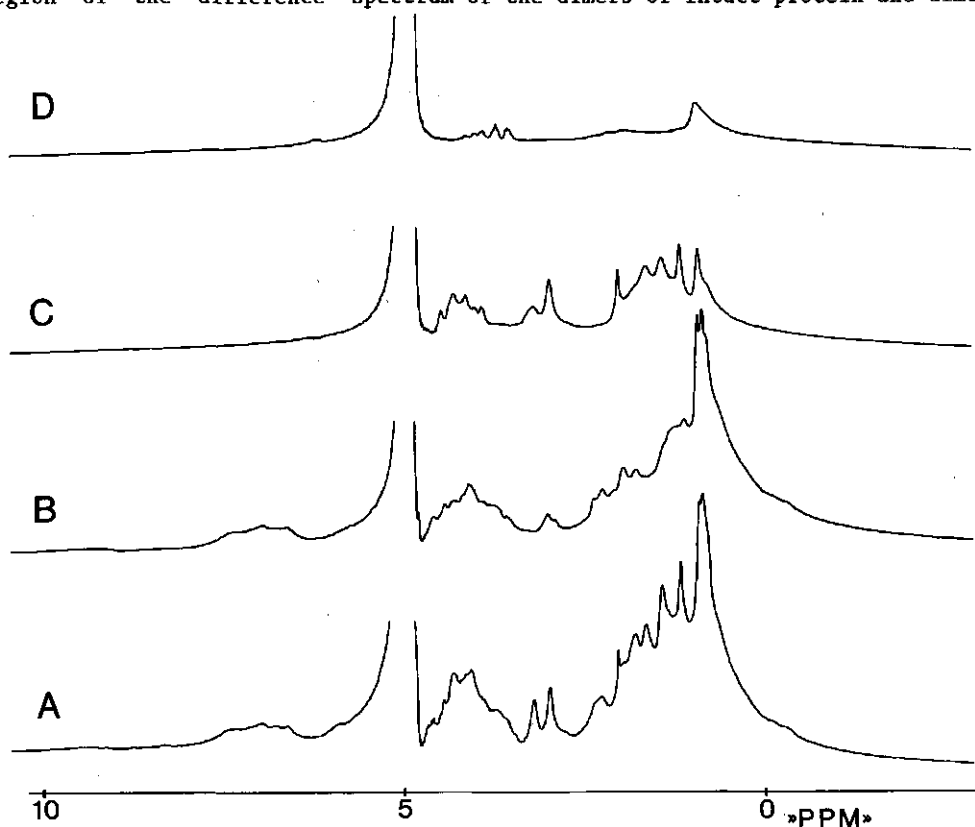


Figure 9.1. ^1H -NMR spectra of empty capsids and dimers.

A: dimers of intact protein.

B: dimers lacking the N-terminal region.

C: empty capsids of intact protein.

D: empty capsids lacking the N-terminal region.

Vertical scales are comparable (Acq. time $> 4xT_1$)

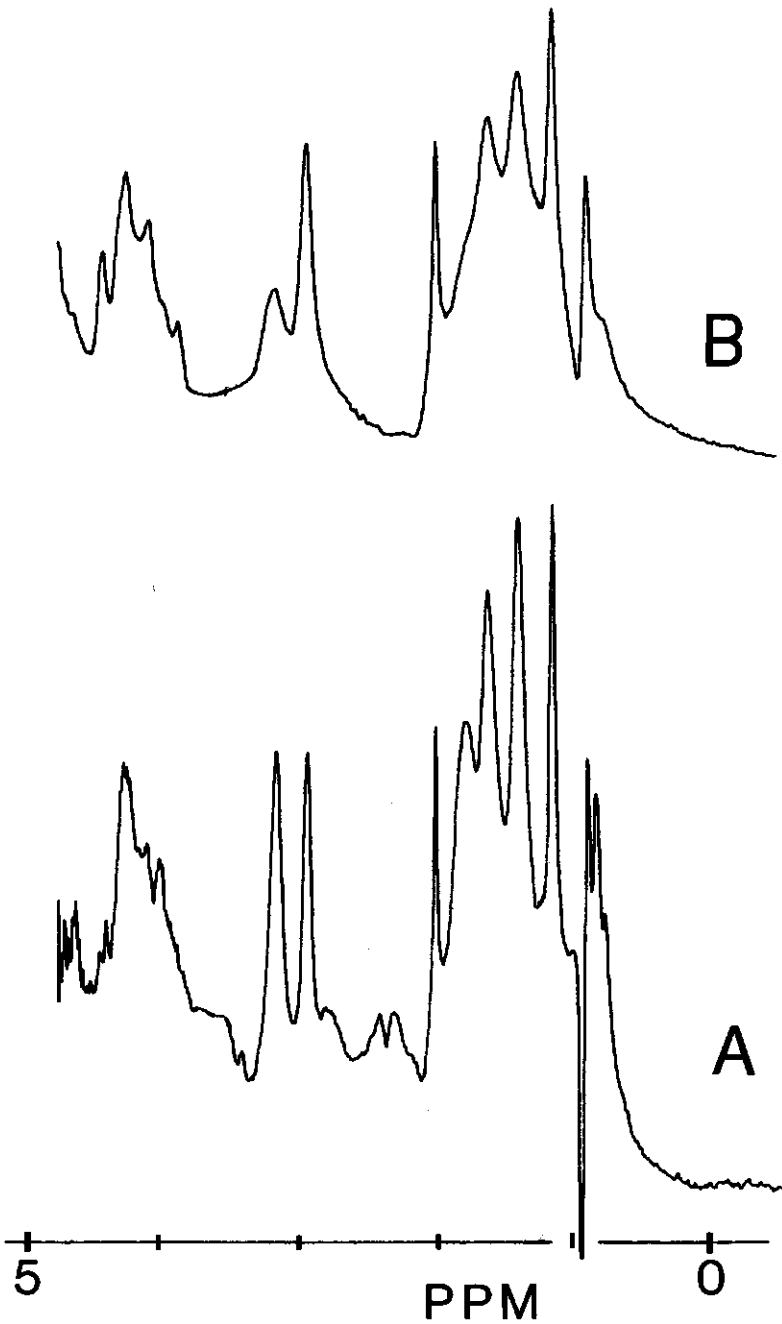


Figure 9.2. Difference spectra of dimers and empty capsids.

A: difference spectrum of intact dimers minus dimers lacking the N-terminal region.

B: difference spectrum of empty capsids of intact protein minus empty capsids lacking the N-terminal regions.

lacking the N-terminal arms. When it is assumed that the core of the protein does not show differences in these two types of dimers, figure 9.2A represents the spectrum of only N-terminal arms of protein dimers, since all other resonances are removed by subtraction. Figure 9.2B shows the difference spectrum of empty capsids with and without N-terminal arms. Using the same reasoning as for figure 9.2A, figure 9.2B approximately represents the spectrum of the N-terminal arms of the empty capsids only. In spectrum 9.2A all 25 residues of every N-terminal arm present in the sample contribute to the spectral intensity, since the rotational correlation time of the dimers is such that no peaks are broader than 40-100 Hz (1) ($T_c \sim 2.5 \times 10^{-8}$ (2)). Besides some differences between figure 9.2A and 9.2B which will be discussed below, it is seen that the overall spectral intensity is less for the N-terminal arms in capsids than for the arms in dimers. Except for the glutamine resonances around 2.4 ppm and the asparagine resonances at ~ 2.9 ppm, which are broadened beyond detection or shifted strongly in figure 9.2B, it is seen that the integrated spectral intensity of all lines decreases approximately with the same extent going from figure 9.2A to 9.2B. It is therefore concluded that a certain fraction of the N-terminal arms is completely immobile in the empty capsids. The results shown in figure 9.2 indicate that approximately 30% of the arms is immobile in the empty capsids. This amount of "immobile" arms varied from 20-55% throughout a series of successive experiments. Although the reason for this variation is unknown, it clearly demonstrates that at least two types of N-terminal arms are present in empty protein capsids: mobile arms and immobile arms. Some N-terminal arms may be immobile because they are involved in protein-protein interaction in a similar way as observed for parts of the N-terminal arms of the C-subunits of TBSV (3) or SBMV (4).

9.1.3.2 DIFFERENCES BETWEEN ARMS IN CAPSIDS AND DIMERS

An inspection of the spectrum of dimers of intact protein (figure 9.1A) reveals that this spectrum contains sharp lines (linewidth ~ 10 -40 Hz) and broad lines (linewidth ~ 40 -100 Hz). The estimated linewidth in the absence of internal motions for the protein dimers is 40-100 Hz (1). Internal motions can decrease this linewidth. The sharp peaks most probably arise from internally mobile amino acids. To select the sharpest lines in the spectrum, the convolution difference method was applied. The result of this convolution difference is shown in figure 9.3A. Using the random coil positions for amino acid resonances presented in appendix A, the sharpest peaks in the dimer spectrum can be assigned to arginine (3.2 ppm), lysine (3.0 ppm), valine (1.0 ppm), leucine (0.95 ppm) and glutamine

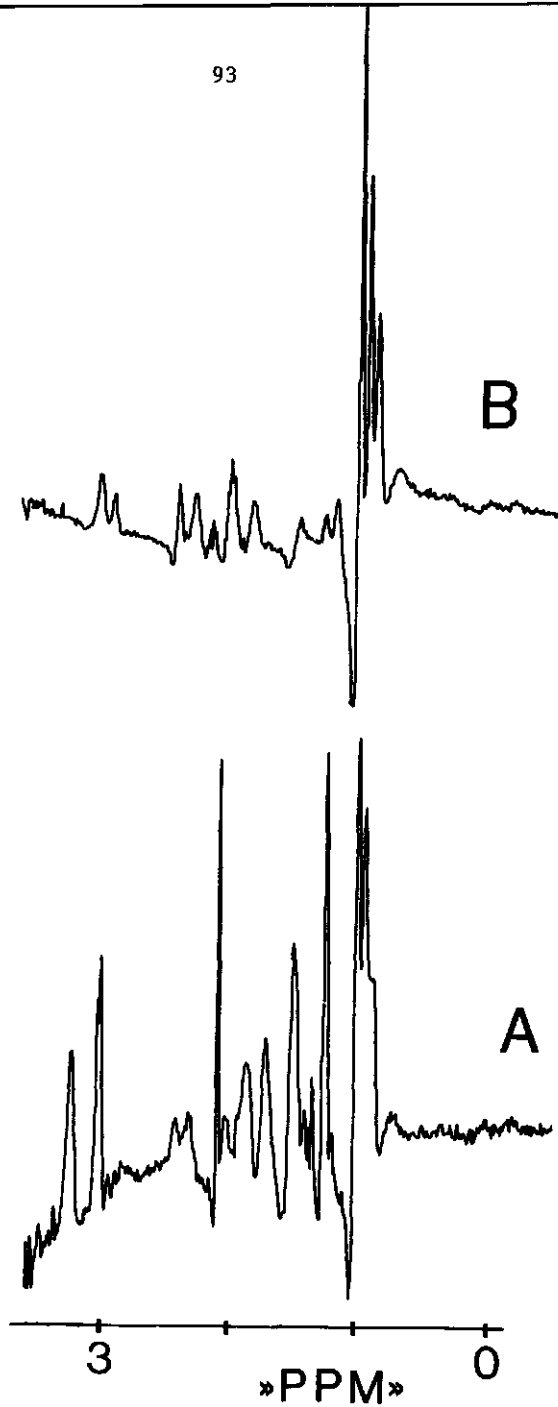


Figure 9.3. Convolution difference spectra of dimers.

A: convolution difference spectrum of intact dimers; 2Hz broadened minus 5Hz broadened.

B: convolution difference spectrum of dimers lacking the N-terminal arm; 4Hz broadened minus 9Hz broadened.

Vertical scales cannot be compared.

(2.3, 2.5 ppm). Due to the overlap with other peaks, the resonances of alanine (1.2 ppm) and the acetyl group of serine (2.1 ppm) can not be assigned unambiguously. From the resemblance between this list of amino acids and the primary structure of the N-terminal arm, it is concluded that the N-terminal arm is mainly responsible for these sharp peaks in the dimer spectrum and indicates that in the dimers the N-terminal arm possesses extra mobility with respect to the core of the protein subunit.

When the spectra of figure 9.2 A and B are compared, a broadening is observed of all lines of the N-terminal arms in capsids. It is also seen that some lines are broadened more than other lines. This is especially the case for arginine Hd (~ 3.2 ppm), arginine H β and lysine H β (~ 1.85 ppm) and very strongly for the asparagine H β resonances around 2.9 ppm and glutamine H γ resonances around 2.4 ppm.

Figure 9.4 shows the spectrum of empty protein capsids at several levels of resolution enhancement. The Lorentz-Gauss transformation used to obtain this resolution enhancement, in principle decreases all linewidths with the same amount. Lines with a relatively short T_2 can not be sharpened unlimited, since the process of resolution enhancement decreases the signal to noise ratio of these lines more than lines with a relative long T_2 . Also lines that are heterogeneously broadened due to small differences in the chemical shift of the protons contributing to that line cannot be sharpened unlimited. When the successive steps of resolution enhancement are followed, it is seen that the resonances of glycine (3.9 ppm), serine (3.75, 2.1 ppm), threonine (1.2 ppm), valine (1.0 ppm), leucine (0.95, 0.93 ppm), and the lysine H ϵ (3.0 ppm) can easily be sharpened to a few Hz. This indicates that these peaks are not heterogeneously broadened. For several other resonances (e.g. arginine (3.2, 1.9, 1.7 ppm), the other resonances of lysine (1.5, 1.7, 1.9 ppm), alanine (1.4 ppm) and asparagine (2.8 ppm)) the lines remain broad at increasing levels of resolution enhancement, or are split up in several individual resonances. This can only be the case if these resonances consists of close lying peaks, arising from chemical shift differences of equivalent protons in different chemical surroundings. This observation strongly suggest that some secondary structure is present in the N-terminal arms in empty capsids.

9.1.3.3 MOBILITY IN THE PROTEIN SUBUNITS LACKING THE N-TERMINAL ARM

The N-terminal protein arm consists of 25 amino acids. In the previous section the mobility of these 25 amino acids in empty capsids and protein dimers has been discussed. In the spectrum of empty capsids (see figure 9.1D) and dimers

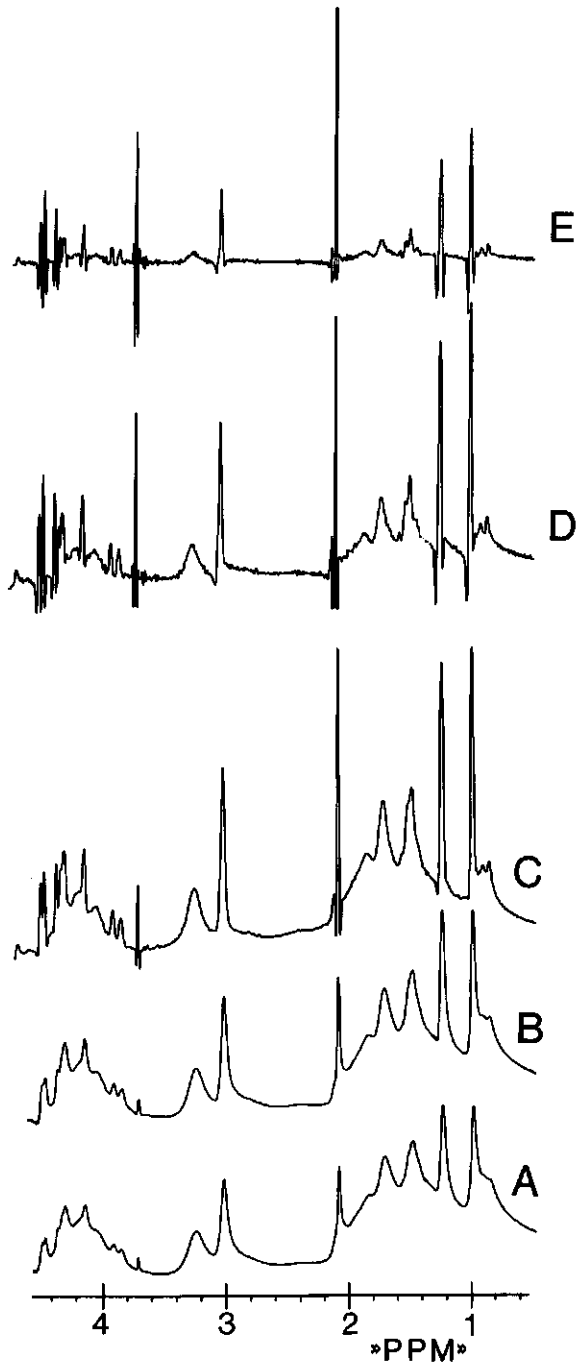


Figure 9.4. Resolution enhanced 0-5 ppm region of the $^1\text{H-NMR}$ spectrum of empty capsids of intact protein. A-E: five successive steps of further resolution enhancement.

(see figure 9.1B) lacking the N-terminal arm some internal mobility remains observable. In the spectrum of empty capsids (see figure 9.1D), this internal mobility arises from $-\text{CH}_3$ groups of valine, leucine or isoleucine.

For the protein dimers lacking the N-terminal arm the convolution difference spectrum is shown in figure 9.3B. When the observed resonance positions are compared with the random coil positions presented in appendix A, it is concluded that the peaks arise from lysine (3.0, 1.5, 1.7, 1.9 ppm), glutamine (2.3, 2.5 ppm), asparagine (2.8 ppm) and a few amino acids from the group: valine, leucine and isoleucine (0.9-1.0 ppm). It is not certain whether all these residues are also mobile in the presence of the N-terminal arm or that this mobility is the result of the removal of this arm. The fact that the two residues directly adjacent to the N-terminal arm (the residues 26 and 27) are both valine suggests that the very sharp lines around 0.96 ppm which are observed in the spectra 9.1B and 9.3B arise from these valine residues. These residues most likely gain mobility upon the removal of the N-terminal arm. These peaks show a very small downfield shift (0.01-0.02 ppm) with respect to the resonance position of the valine in the N-terminal arm. This could explain the occurrence of the strong negative peak at 0.97 ppm in spectrum 9.2A.

9.2 THE STRUCTURE OF THE N-TERMINAL ARM

9.2.1 INTRODUCTION

In the previous section evidence has been presented that on the average, some secondary structure is present in the N-terminal arm of the coat protein of CCMV in empty capsids. In order to study this secondary structure in more detail a 2D-NOE spectrum of empty capsids was recorded, since it has been shown that such a spectrum can provide structural information (5-7). Because the pulse sequence used for obtaining the 2D-NOE spectrum generally does not discriminate between NOE effects and exchange (8,9), a frequency dependent study of the amide protons in the N-terminal arm was performed as well.

From the results obtained it is concluded that the N-terminal arm can occur in at least two metastable structures. Between these structures, conformational exchange is observed with a rate constant of $\sim 2 \times 10^{-3}$ s. Again, the results suggest that some α -helical structure is present in the N-terminal arm. Although no detailed conclusions can be drawn about the type and the location of the observed secondary structure, the results presented in this chapter unambiguously prove that a part of the N-terminal arm of the coat protein of CCMV is structured in empty capsids.

9.2.2 MATERIALS AND METHODS

Empty capsids of CCMV protein were prepared as described in chapter 4. The capsids were dialyzed against 300 mM NaCl, 10 mM MgCl₂ and 10 mM sodiumphosphate pH 5.0, dissolved in H₂O. A pellet of empty capsids, produced from this solution by centrifugation for 4 h at 40000 rpm in a Beckman Ti60 rotor, was gently stirred for 48 h in 1 ml of the above solution made up in H₂O : D₂O = 85 : 15. By this procedure a "saturated" protein solution containing ~150 mg/ml, was obtained. From this solution 400 µl was used for the 2D-NOE measurements. The protein concentration of the sample used for conventional one dimensional NMR measurements was ~5 mg/ml. The 2D-NOE spectrum was recorded with a Bruker WM500 spectrometer using the (90-τ1-90-τm-90-FID-delay)n pulse sequence (5-9). The mixing time (τm) was 100 ms and 128 spectra were recorded with a τ1 increment of 80 µs. The number of scans was 162. The spectrum is resolution enhanced in both dimensions.

Conventional one dimensional NMR spectra were recorded with a Bruker CXP300 and a Bruker WM500 spectrometer. 512 scans of 0.35 s were taken. The sensitivity enhancement was 5 Hz.

In both experiments the temperature was kept at 7°C. Also in both experiments the water peak was saturated by 2 s pre-irradiation. The ppm scales are relative to DSS.

9.2.3 RESULTS AND DISCUSSION

9.2.3.1 2D-NOE MEASUREMENTS

Figure 9.5 shows the 2D-NOE spectrum of empty capsids of CCMV coat protein. The contour plot of this spectrum is presented in figure 9.6. All peaks in these spectra arise from the residues in the N-terminal arm (10). The conventional NMR spectrum is seen along the diagonal of these figures. The peaks in the 6-9 ppm region arise from amide proton resonances. These resonances are discussed here since they are very sensitive to secondary structure effects (6). The resonances in the 0-5 ppm region will be discussed in the next section (see section 9.3).

The cross peaks in the spectra of figure 9.5 and 9.6 can arise from dipolar cross-relaxation (NOE), chemical or conformational exchange, and from remaining J-coupling effects (zero-quantum coherence). Although it is possible to suppress remaining J-coupling effects (8,9), this was not done in this study, since both J-coupling and dipolar cross-relaxation are an indication for a close spatial proximity.

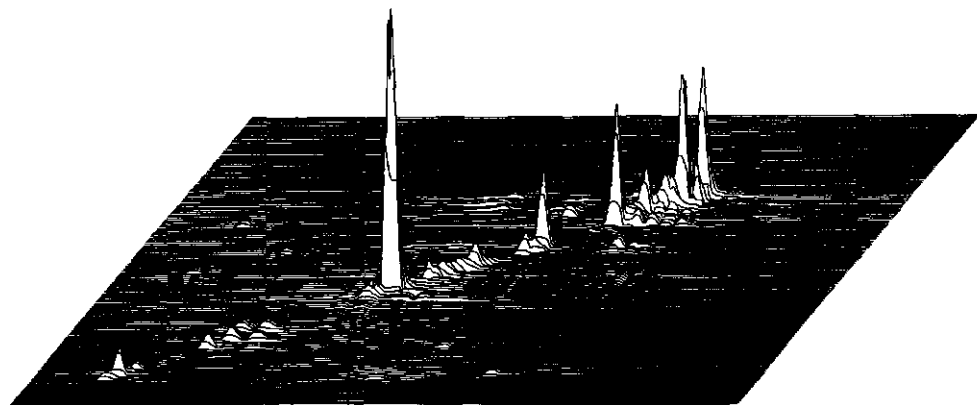


Figure 9.5. 2D-NOE spectrum of empty capsids of intact protein, dissolved in H_2O

Two facts are implicitly used in the assignments of cross peaks in the 2D-NOE spectra presented in this section and in section 9.3:

1) If a cross peak can both be assigned to an intra-residue and an inter-residue interaction, the former is more likely to be the cause of this cross peak. 2) If a cross peak can both arise from a connectivity between two adjacent residues and a connectivity between two residues further away from each other in the primary structure, the former is more likely to apply (5-7).

In the 6-9 ppm region, the following resonances can be assigned:

"Cross peak a" indicates an interaction between the amide proton at 8.5 ppm and the peak at 2.1 ppm. The peak at 2.1 ppm can arise from the acetyl group of serine, the valine $H\beta$ or the glutamine $H\beta$. Since the amide proton at 8.5 ppm has a relative long T_2 value, just as has been observed for all serine resonances, (see section 9.1) the peak at 8.5 ppm is assigned to the amide proton of serine.

"Cross peak b" indicates an interaction between the arginine $H\delta$ and the peak at 7.3 ppm. Therefore the peak at 7.3 ppm is assigned to the arginine $H\epsilon$. The peak at 7.3 ppm also shows interaction with other amide protons. This is very unlikely for an arginine $H\epsilon$ and therefore it is concluded that this peak cannot be assigned to arginine $H\epsilon$ alone. Hence the peak at 7.3 ppm is assigned to the arginine $H\epsilon$ and at least one other unidentified amide proton.

The two cross peaks marked "c" vertically coincide with the glycine α -protons at 3.9 ppm. It is hard to believe that both cross peaks arise from intra residue $\alpha H-NH$ interactions, since this would imply that both glycine NH protons differ 0.4 ppm in their resonance position. It is therefore concluded that at least one of the two peaks at 8.3 and 8.7 ppm can be assigned to a neighbour residue of glycine in the N-terminal arm. Therefore the two peaks at 8.3 and 8.7 ppm are

assigned to the amide protons of 7-lysine, 5-threonine or 4- or 6- glycine.

When using the random coil positions for amino acid resonances (see appendix A) it is seen that "cross peak d" could very well vertically coincide with the α -proton of serine. Therefore it is concluded that "cross peak d" indicates an intra. residue α H-NH interaction. Hence the resonance at 8.5 ppm is assigned to serine-NH.

For the 2D-NOE NMR spectra in figure 9.5 and 9.6 a mixing time of 100 ms is used. It has been shown by Wagner and Wuethrich (7) that for BPTI such a mixing time is adequate at 500 MHz. This mixing time is just long enough that NOE's

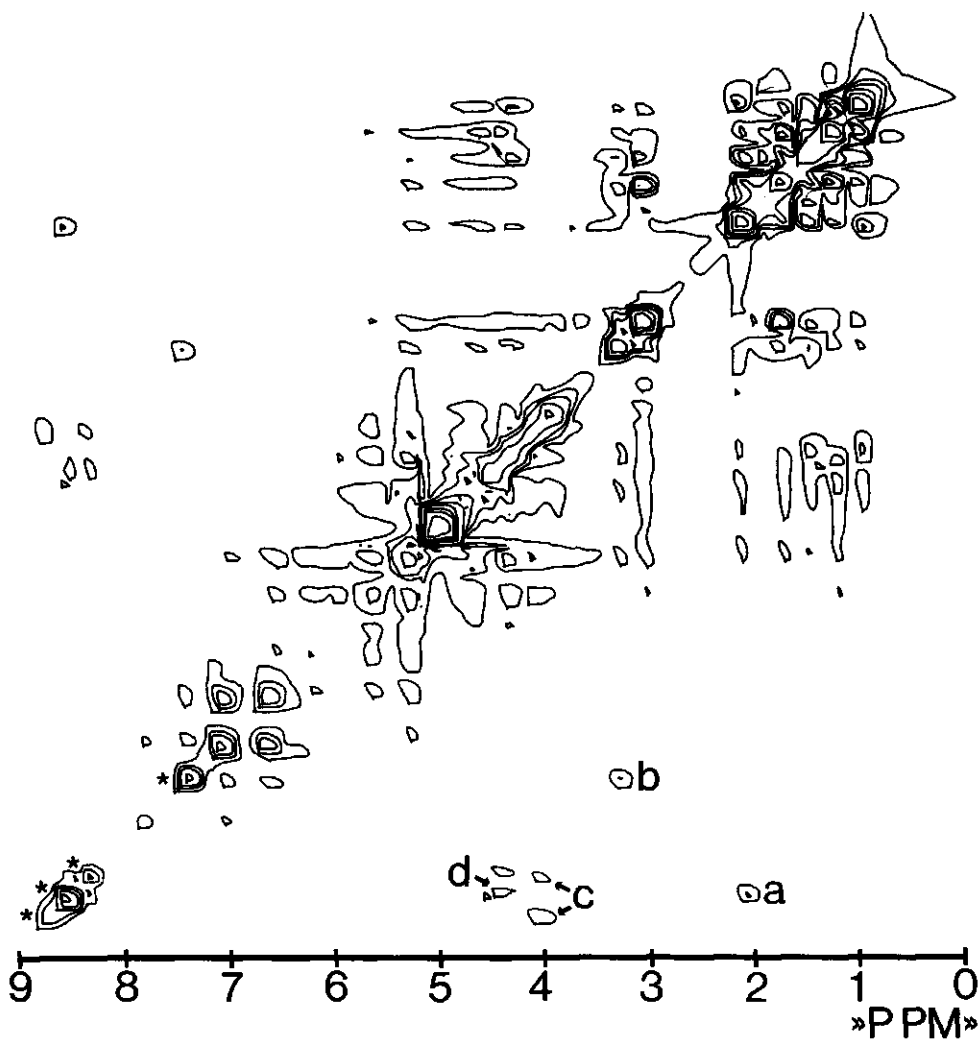


Figure 9.6. Contour plot of the 2D-NOE spectrum of empty capsids of intact protein presented in figure 9.5.

manifesting distances ~ 3 Å between different protons in the polypeptide are observable. However, since the N-terminal arm has a high degree of flexibility, it is impossible to decide whether an observed NOE cross peak indicates a fixed or a time-averaged close spatial proximity.

All diagonal resonances in the 6 - 9 ppm range in figure 9.5 and figure 9.6 arise from backbone NH-protons. An exception must be made for the arginine H ϵ resonance at 7.3 ppm. This is the only non-exchangable side chain NH-proton. These resonances share quite a number of cross peaks, located especially in the 6 - 8 ppm region. These cross peaks can have two origins: NH-NH interactions or exchange. In the next section (see 9.2.3.2) it will be shown from a frequency dependent study, that the strong cross peak between the NH-protons at 6.5 and 7.0 ppm can be explained to be the result of conformational exchange. For the other resonances no evidence for exchange processes is found. The only other explanation besides exchange is inter residu NH-NH interactions. These interactions indicate that the polypeptide is in an α -helix conformation (7). The cross peaks in the 6-8 ppm region (except the cross peak between the resonances at 6.5 and 7.0 ppm) suggest that part of the N-terminal arm of CCMV protein has an α -helix structure in empty protein capsids.

9.2.3.2 CONFORMATIONAL EXCHANGE IN THE N-TERMINAL ARM

Figure 9.7 shows the amide proton region of the spectra of empty capsids measured at 300 and 500 MHz. The most striking difference between these two spectra is the fact that the two separate peaks at 6.5 and 7.0 ppm in the 500 MHz spectrum collaps into a broad hump around 6.7 ppm at 300 MHz. These two peaks also give rise to a strong NOE cross peak at 500 MHz (see figures 9.5 and 9.6). These effects can only be explained by assuming a slow exchange between at least two distinct, metastable conformations. Using common exchange theory (12) the rate constant is calculated to be $\sim 2 \times 10^{-3}$ s. In this calculation the following assumptions are made: The process is a two site exchange, no spin-spin coupling is present and the T_2 's and the lifetimes are equal in both states.

Several other differences in linewidth can be observed between the two spectra of figure 9.7, that cannot be explained by the difference in measuring frequency. This is especially the case for the serine amide proton at 8.5 ppm and the unassigned peaks around 7.7 ppm. These extra line broadening effects may indicate the presence of exchange for these NH-protons, but no conclusions can be drawn about the rate of these processes.

It can be concluded that exchange is present between at least two different conformations which are stable on the millisecond timescale. This proves that

500 MHz

300 MHz

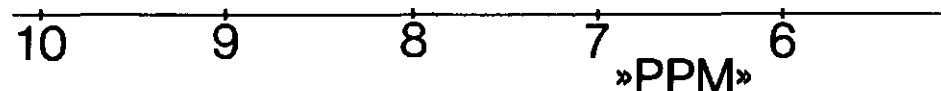


Figure 9.7. ^1H -NMR spectra at 300 and 500 MHz of empty capsids of intact protein dissolved in H_2O . Only the 6-10 ppm region is shown.

part of the N-terminal arm possesses secondary structure. In the 6-9 ppm region of the spectra of figure 9.7 31 NH-protons are expected to give rise to peaks (25 backbone NH-protons and 6 arginine side chain protons). Since the spectral intensity of the peaks involved in the exchange process contains 40-50% of the spectral intensity observed between 6 and 9 ppm. This indicates that approximately half of the residues in the N-terminal arm participates in this process of conformational exchange.

9.3 ASSIGNMENTS IN THE 0-5 PPM REGION IN THE SPECTRUM OF EMPTY CAPSIDS

9.3.1 INTRODUCTION

Throughout this thesis spectra of empty capsids of the coat protein of CCMV are presented recorded under different conditions. This section aims at a complete assignment for the resonances of the N-terminal arm observed in the 0-5 ppm region. In order to be able to do so, the resolution enhanced spectra of empty capsids presented in figure 9.4 are used in combination with a 2D-NOE spectrum of empty capsids, and the list of random coil resonance positions presented in appendix A. It is possible to assign every peak in the spectrum to one or more types of protons in the N-terminal arm. It is not possible to distinguish between resonances of equal amino acids at different positions in the primary structure.

9.3.2 MATERIALS AND METHODS

The resolution enhanced spectra of empty capsids of the coat protein of CCMV have been discussed in section 9.1. The 2D-NOE spectrum of the empty capsids was obtained as described in the previous section. The sample was prepared as described in section 9.2.2, with the exception that the protein concentration was slightly smaller. H_2O in the sample was substituted by D_2O through extensive dialysis against 300 mM NaCl, 10 mM $MgCl_2$ and 10 mM sodium phosphate pH 5.0, made up in D_2O . All measuring conditions are identical to those used in section 9.2 to obtain the 2D-NOE spectrum of empty capsids dissolved in H_2O .

9.3.3 RESULTS AND DISCUSSION

Table 9.1 shows all the protons present in the N-terminal arm that contribute to the spectral intensity in the 0-5 ppm region and their random coil resonance positions.

Table 9.1. Amino acids present in the N-terminal arm of the coat protein of CCMV and their random coil resonance positions. Numbers in brackets indicate the number of equivalent protons in the N-terminal arm. Amide protons are not presented as they are not observable in D_2O . The second column indicates the frequency of occurrence in the N-terminal arm.

Amino acid	occurrence	H α	H β	H γ	H δ	H ϵ
Alanine	4	4.35 (4)	1.43 (12)			
Arginine	6	4.40 (6)	1.85 (6)	1.70 (12)	3.26 (12)	
			1.70 (6)			
Asparagine	2	4.76 (2)	2.92 (2)			
			2.90 (2)			
Glutamine	1	4.30 (1)	2.13 (1)	2.38 (2)		
			2.01 (1)			
Glycine	2	3.97 (4)				
Leucine	1	4.39 (1)	1.65 (2)	1.65 (1)	0.88 (6)	
Lysine	3	4.36 (3)	1.85 (3)	1.47 (6)	1.70 (6)	3.02 (6)
			1.73 (3)			
Serine	1	4.50 (1)	3.89 (2)	acetyl group 2.10 (3)		
Threonine	4	4.35 (4)	4.18 (4)	1.17 (12)		
Valine	1	4.18 (1)	2.13 (1)	0.97 (3)		
				0.95 (3)		

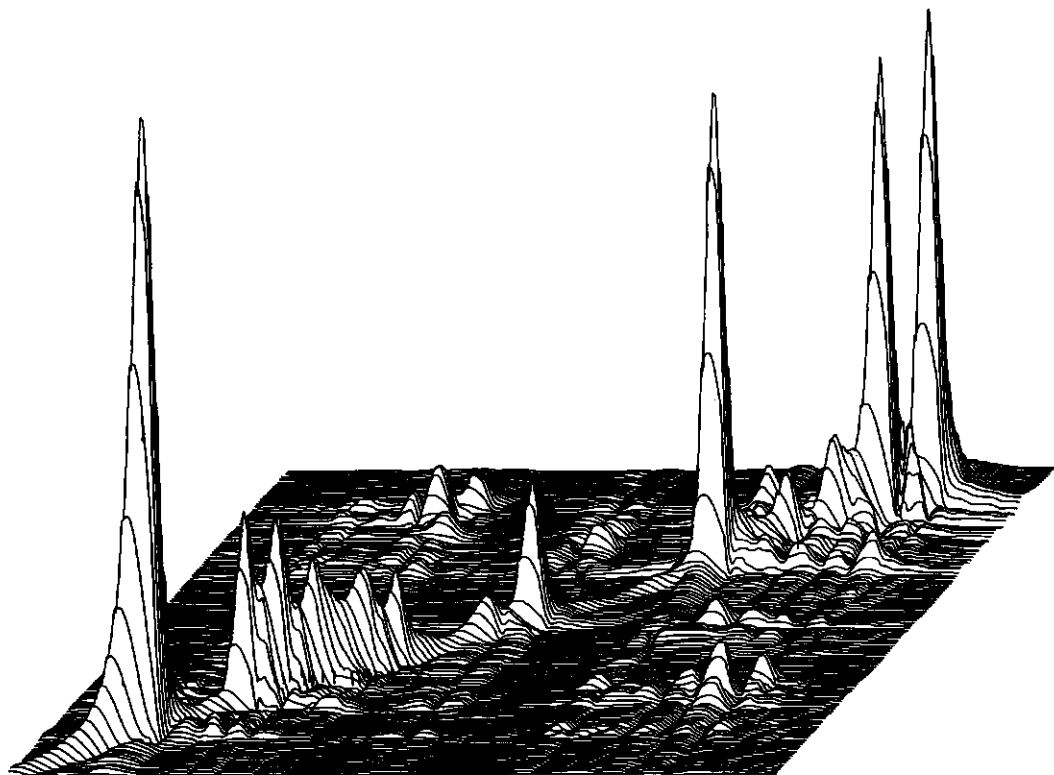


Figure 9.8. 2D-NOE spectrum of empty capsids of intact protein dissolved in D_2O . Only the 0-5 ppm region is shown.

Figure 9.8 shows the 1-5 ppm region of the 2D-NOE spectrum of empty capsids in D_2O . The contour plot of this spectrum is shown in figure 9.9. A comparison between the observed peak positions in the projection of this spectrum and the (random coil) resonance positions from tabel 9.1 leads to the following assignments:

"a": asparagine $H\alpha$

"b": serine $H\alpha$

"c": lysine $H\alpha$, threonine $H\alpha$, alanine $H\alpha$, arginine $H\alpha$, glutamine $H\alpha$, leucine $H\alpha$

"d": threonine $H\beta$

"e": serine $H\beta$

"f": arginine $H\delta$

"g": lysine $H\epsilon$

"h": asparagine $H\beta$

"i": glutamine $H\gamma$

"j": Valine $H\beta$, glutamine $H\beta$, N-terminal acetyl group of serine

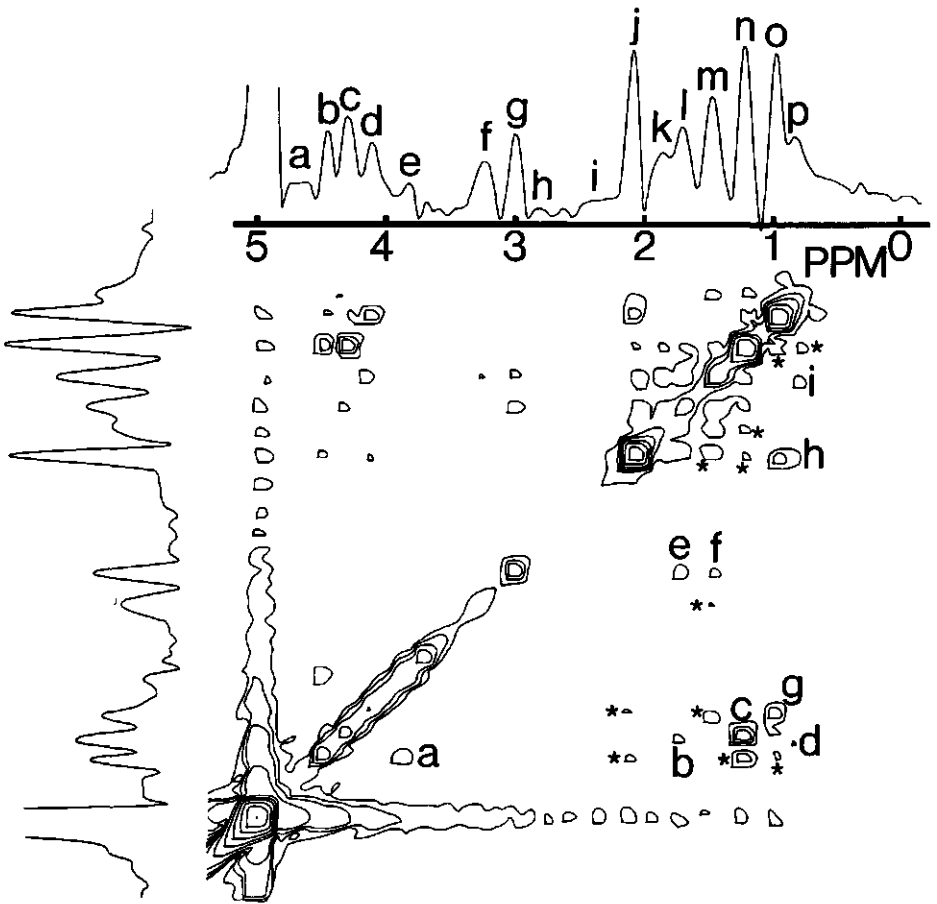


Figure 9.9. Contour plot of the 2D-NOE spectrum of empty capsids of intact protein presented in figure 9.8.

- "k": Lysine H β , arginine H β , glutamine H β
- "l": lysine H δ , arginine H β , arginine H γ , lysine H β
- "m": lysine H γ , alanine H β
- "n": threonine H γ
- "o": valine H γ
- "p": leucine H δ

Using the resonance positions of tabel 9.1, the glycine Ha resonances should occur between the peaks "d" and "e". Therefore the right shoulder of peak "d" is assigned to glycine H α . The leucine H β and H γ resonances should occur as a left shoulder of peak "m".

Additional evidence for these assignments is given below:

- Peaks "b" and "e" are respectively assigned to serine H α and serine H β . These peaks are coupled in the 2D spectrum (cross peak A).
- In figure 9.4 it is seen that resolution enhancement sharpens and increases the intensity all resonances assigned to serine (serine H α =peak "b", serine H β =peak "e", acetyl group of serine=peak J), indicating a relative long T₂ for all protons in this residue.
- Peak "c" contains lysine H α and arginine H α . This peak is coupled in the 2D spectrum to peak "l" ("cross peak B"), which contains arginine H β and lysine H β .
- Peak "c" contains threonine H α and is coupled to threonine H γ (peak "n") by "cross peak c".
- The right shoulder of peak "d" is assigned to glycine. In figure 9.4 it is seen that these peaks can be increased in intensity and sharpened by a resolution enhancement procedure just as the other 5 residues in the group of 7 N-terminal residues. In the 2D-NOE spectrum presented in the previous section, these peaks are observed to be coupled to NH protonen.
- "Cross peak e" is assigned to the lysine H ϵ - Lysine H δ coupling (peak "g" - peak "l").
- "Cross peak f" couples the lysine H ϵ (peak "g") with the lysine H γ (peak "m").
- "Cross peak g" couples the valine H γ (peak "o") with the valine H α (peak "d").
- "Cross peak h" couples the valine H γ (peak "o") with the valine H β (peak "j").
- "Cross peak i" couples the leucine H δ (peak "p") to the lowfield shoulder of peak "m", which is assigned to leucine H γ and H β . In figure 9.4E it is seen that the lowfield shoulder of peak "m" consists of an independent line.
- Peak "h" is assigned to asparagine H β . The fact that these peaks are rather broad can be explained by assuming a reduced mobility due to the fact that the two asparagine residues in the N-terminal arm lie close to the immobile protein core.

Most of the assignments for cross peaks in the above list are made using the fact that the probability of observing an intra residue coupling is much larger than the probability of observing an inter residue coupling.

Several cross peaks in the 2D spectrum presented in the figures 9.8 and 9.9 are marked with an asterix. These peaks can not be assigned to intra residue interactions. Hence, these cross peaks are a measure for short inter residue distances. Unfortunately, there are too many different assignments possible for these cross peaks to use them for elucidation of the structure of the N-terminal arm. The only assignments that can be made are the couplings between leucine H δ (peak "p") and threonine H γ (peak "n"), and between threonine H γ (peak "n") and valine H γ (peak "o"). The chance that these inter residue interactions result

from two adjacent residues is much larger than the chance that they results from an interaction between two residues far away from each other in the primary structure (5-7). Therefore, these interactions are concluded to be the result of a close spatial proximity between 8-leucine H δ and 9-threonine H γ , and between 2-threonine H γ and 3-valine H γ .

9.4 ¹³C RELAXATION MEASUREMENTS

9.4.1 INTRODUCTION

One of the objectives of this thesis is the quantification of the mobility of the N-terminal arm of the coat protein of CCMV. The ¹H-NMR experiments have only given a rough estimate of the rotational correlation time of this arm, being 10⁻⁸ - 10⁻⁹ s. The ¹H-NMR experiments do not allow a more quantitative analysis of the observed mobility. For this reason a ¹³C-NMR relaxation study was carried out, which allows a much better analysis of the relaxation times.

The signal-to-noise problem, which is inherent to natural abundance ¹³C-NMR, is overcome by biosynthetical enrichment of the virus. Well-resolved ¹³C-NMR spectra are obtained of the empty protein capsids, in which the sharp resonances can be assigned to the N-terminal arm. The relaxation times T₁, as well as the NOE parameters are obtained of these resonances at different measuring frequencies. The results indicate a rotational correlation of approximately 3x10⁻⁹ s for almost all amino acids in the N-terminal arm, without large differences in mobility between different groups in this arm.

9.4.2 MATERIALS AND METHODS

9.4.2.1 ¹³C ENRICHMENT

The method used to produce ¹³C-enriched CCMV was similar to the method described for the enrichment of TMV (1). The enrichment proceeds through assimilation of 90% ¹³C enriched carbon dioxide by the primary leaves of vigna unguiculata infected with CCMV. ~8 g of 90% ¹³C enriched BaCO₃ (for generating ¹³CO₂) was used per 100 g of leaves, so that the final enrichment of CCMV always ended up between 8 and 12%. Vigna plants of about ten days were infected with CCMV and after 8 hours of recuperation the plants were cut off and placed in water in a completely air-tight mini-greenhouse connected to a Kipp apparatus. The ¹³CO₂ was then slowly added (in ~24h) and the plants were kept at 23°C until the first signs of starvation appeared (after ~5 days). The plants were illuminated with

two 40 W TL lights placed 1 m above the greenhouse, and two Philips SON 400 W lamps placed 50 cm above the greenhouse. The two 400 W lamps were switched on and off twice every hour to avoid overheating in the greenhouse. The $^{13}\text{C}/^{12}\text{C}$ ratio in the virus was determined by burning completely ~ 0.5 mg of CCMV in an oven at 1100°C under a continuous oxygen flow and measuring the $^{13}\text{C}/^{12}\text{C}$ ratio in the resulting mixture of gasses using a Mat Finigan 250 mass spectrometer.

9.4.2.2 NMR MEASUREMENTS

The protein sample was prepared as described in chapter 4. ~ 70 mg of 9.6% ^{13}C enriched protein was dissolved in 2 ml 300 mM NaCl, 10 mM MgCl_2 and 10 mM sodium phosphate pH 5.0, containing 10% D_2O . The spectra were recorded with a Bruker CXP300, a Bruker WM250 and a Bruker WM360 spectrometer. T_1 values were measured using the progressive saturation method. Typically 40000 scans were recorded per spectrum. The sensitivity enhancement was 5 Hz. Peak heights were used as a measure for the spectral intensity used to derive the T_1 value. Broadband 1H noise decoupling of 2 W (300 MHz) or 4 W (360 MHz) was employed continuously. The temperature was kept at $7 \pm 2^\circ\text{C}$.

NOE values were measured using a two-level decoupling scheme: 4 W (360 MHz) and 5 W (250 MHz) decoupling power was used during the acquisition time, whereas 0.5-1 W was used during the delay. In the spectra without NOE of course no decoupling was used during the delay. The acquisition time was 0.2 s and the delay time was 2.0 s. The ppm scale is referred to CS_2 , assuming 125.8 ppm for the threonine $\text{C}\beta$ position.

9.4.3 RESULTS AND DISCUSSION

Figure 9.10 shows the 360 MHz ^{13}C -NMR spectrum of empty protein capsids. The sharp peaks in this spectrum arise from the mobile N-terminal region as was checked by removal through tryptic digestion of this arm. Tabel 9.3 shows the measured relaxation times and NOE values of the peaks indicated in figure 9.10, together with the assignments made by comparing the peak positions with the list of ^{13}C resonances presented in appendix B. From tabel 9.3 it is seen that within the limits of experimental error almost no significant differences are observable between the T_1 values of ^{13}C nuclei bearing protons. When the observed T_1 values are interpreted using equation 1.1 the rotational correlation times are found to range from 10^{-10} - 3×10^{-10} s or from 3×10^{-9} - 10^{-8} s. Using equation 1.3 the NOE values indicate $\tau_c > 10^{-9}$ s (only the NOE values found for the peaks 10, 14 and 15 give rise to smaller τ_c values). When the observed

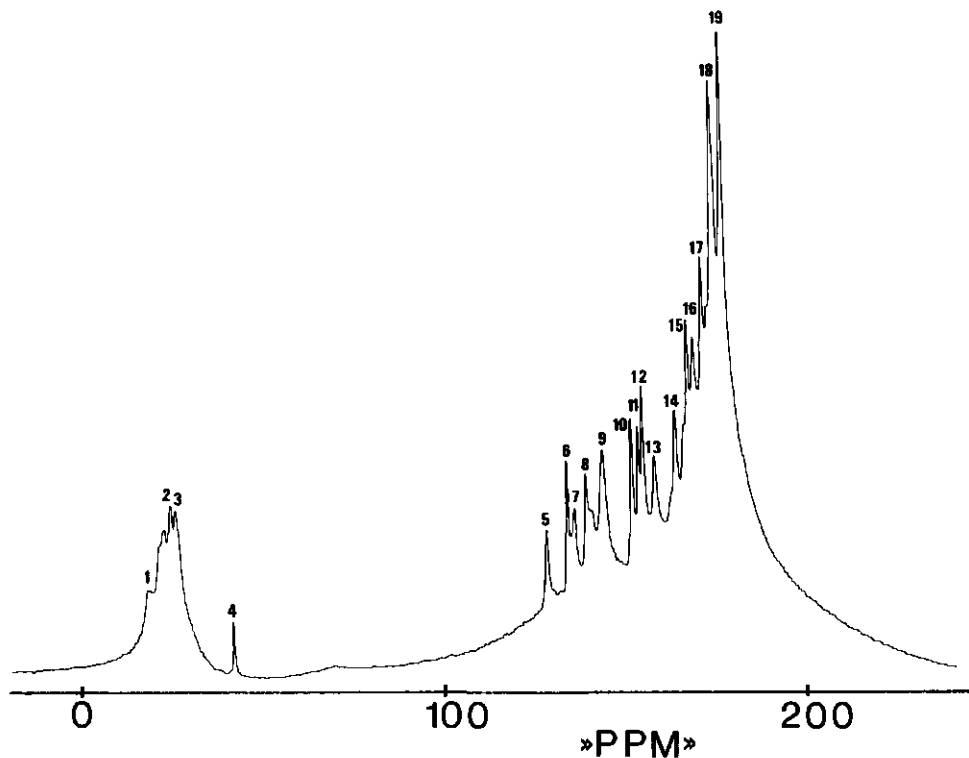


Figure 9.10. 75 MHz ^{13}C -NMR spectrum of empty capsids of 9.6% ^{13}C enriched protein.

linewidths are supposed to represent T_2 and are interpreted in terms of equation 1.2, $\tau_c \sim 10^{-8}$ s is calculated.

Due to the fact that most lines are superpositions of many resonances, it is not possible to obtain precise values for individual T_1 , T_2 or NOE values. Most values presented in tabel 9.3 are therefore the average of a group of nuclei. Due to the low signal-to-noise ratio the errors in the values presented are rather large (25 - 50%). The fact that NOE values at 250 MHz are on the average lower than the NOE values at 360 MHz strongly suggest the presence of some systematic instrumental errors, since from equation 1.3 just the opposite is expected.

It should be noted that this analysis of the τ_c values is based on the relaxation properties of a rigid sphere. Although this is probably not the case for the N-terminal arm, the errors in the observed relaxation times do not allow a more precise interpretation.

From the results it is seen that the rotational correlation time of almost all the amino acids in the N-terminal arm is approximately 3×10^{-9} s. No significant spread is observed, indicating the absence of large differences in mobility between groups in the N-terminal arm.

Table 9.3. T_1 and NOE values for the resonances indicated in figure 9.10.

Peak	Assignment	T_1 360 MHz	T_1 300 MHz	NOE 360 MHz	NOE 250 MHz
1		1.4	1.5 (s)	1.0 (s)	1.1
2	C=O	1.6	1.3	1.0	1.1
3		2.0	1.2	1.0	1.1
4	Arg C ζ	13.3	1.2	1.0	1.4
5	Thr C β	0.4	0.3	1.3	1.0
6	Ser C β	0.3	0.2	1.3	1.6
7	Thr C α	0.4	0.4	1.5	1.0
8	Ser C α	0.5	0.5	1.4	1.2
9	Ala C α	0.5	0.5	1.4	1.2
10	Gly C α / Ac-Ser	0.4	0.6	2.0	1.5
11	Arg C δ	0.3	0.4	1.5	1.5
12	Lys C ϵ / Leu C β	0.4	0.5	1.5	1.4
13	Ile C β / Asn C β	0.4	0.3	1.5	1.0
14	Lys C β / Val C β	0.4	0.6	2.3	1.0
15	Lys C δ	0.4	0.5	3.0	1.3
16	Leu C γ / Arg C γ	0.4	0.5	1.1	1.0
17	Lys C γ / Leu C δ 1	0.4	0.4	1.9	1.2
18	Thr C γ / Val C γ 1	0.5	0.5	1.9	1.3
19	Ala C β / Val C γ 2	0.4	0.4	1.6	1.6
	estimated errors	25%	50%	25%	50%

9.5 CONCLUSIONS

In this chapter several NMR studies, some of them having a preliminary character, are presented, which shed more light upon the molecular structure and dynamics of the N-terminal arm of the coat protein of CCMV. When combining all results, it can be concluded that the N-terminal arm of the coat protein of CCMV is internally mobile both in protein dimers and in empty protein capsids. In the empty capsids this mobility is in the order of 3×10^{-9} s. In the empty capsids not all the N-terminal arms are mobile: a fraction of the arms is not seen in

the NMR spectrum, indicating that these arms possess a rotational correlation time $>10^{-6}$ s. Furthermore it is concluded that in the empty capsids the N-terminal arms can occur in at least two metastable conformations between which conformational exchange on the millisecond timescale is observed. The N-terminal arm in empty capsids is partly structured, and probably forms an α -helix with a lifetime of $\sim 2 \times 10^{-3}$ s.

From the fact that a part of the arm is structured on the millisecond timescale but still has an overall mobility on the 10^{-8} - 10^{-9} s timescale, it is concluded that this overall mobility is caused by the flexibility of the C-terminal amino acids of the arm. These amino acids may form a flexible hinge between the rigid protein core and the region in the N-terminal arm containing the secondary structure. This model is visualized in figure 9.11.

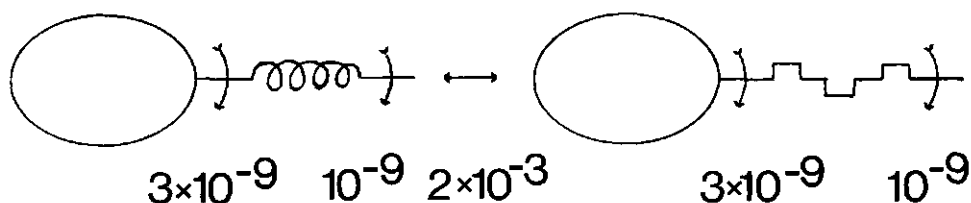


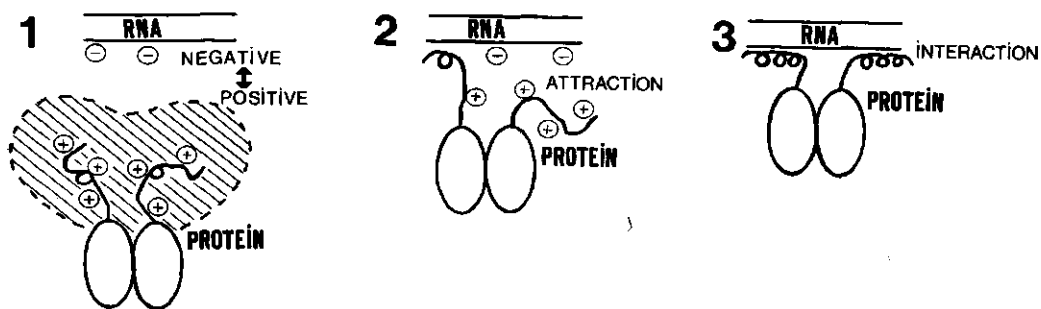
Figure 9.11. Schematical representation of the observed mobilities in the protein subunit in empty capsids.

9.6. REFERENCES.

- (1) Wit de, J.L., (1978), Thesis, Agricultural University, Wageningen.
- (2) Chapter 3 of this thesis.
- (3) Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, K.K., Bricogne, G., (1978), Nature 276, 368-373.
- (4) Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., Leslie, A.G.W., Rayment, I., Rossmann, M.G., Suck, D., Tsukihara, T., (1980), Nature 286, 33-39.
- (5) Wuethrich, K., Wider, G., Wagner, G., Braun, W., (1982), J.Mol.Biol. 155, 311-319.
- (6) Billeter, M., Braun, W., Wuethrich, K., (1982), J.Mol.Biol. 155, 321-346.
- (7) Wagner, G., Wuethrich, K., (1982), J.Mol.Biol. 155, 321-346.
- (8) Macura, S., Huang, Y., Suter, D., Ernst, R.R., (1981), J.Magn.Res. 43, 259-281.
- (9) Macura, S., Wuethrich, K., Ernst, R.R., (1982), J.Magn.Res. 46, 269-282.
- (10) Vriend, G., Hemminga, M.A., Verduin, B.J.M., Wit de, J.L., Schaafsma, T.J., (1981), FEBS Lt. 134, 167-171.
- (11) Farrar, T.C., Becker, E.D., Pulse and Fourier Transform NMR. Academic Press.

SUMMARY

This thesis describes the application of ^1H - and ^{13}C -NMR, EPR, ST-EPR and calculational methods to study cowpea chlorotic mottle virus. This virus consists of RNA encapsidated by 180 identical protein subunits, arranged icosahedrally. The isolated coat protein of cowpea chlorotic mottle virus can be brought into several well defined states of aggregation. This study could be carried out, because these stages can be produced in quantities sufficient to allow magnetic resonance measurements. All the results obtained are combined in the following model for the assembly process of the virus:



In this model a rigid protein core with a highly mobile, basic, N-terminal arm is invoked. This arm is the RNA binding part of the protein. The high mobility of this arm enhances the probability of interaction with the RNA and enables the protein to exhibit different modes of bonding to different local structures of the RNA. Upon binding the RNA, the N-terminal arm adopts a rigid α -helical conformation.

In chapter two it is shown that virtually no mobility on a timescale faster than 10^{-7} s can be observed in the virus.

In chapter three an EPR and ST-EPR study is presented on spin-labelled virus. From the results it is concluded that no anisotropic subunit mobility is present in the virus on the 10^{-5} - 10^{-7} s timescale. Also in this chapter it is suggested that many ST-EPR results on maleimide-labelled proteins, in which anisotropic protein mobility was invoked for the interpretation of the spectra, are based on artefacts.

Chapter four shows that the N-terminal arm is the RNA binding part of the protein. This arm is found to be very mobile in the absence of RNA, whereas immobilization occurs upon binding RNA.

Chapter five is a small excursion to other plant viruses. It is shown that brome mosaic virus and belladonna mottle virus show the same behaviour as cowpea chlorotic mottle virus, whereas cowpea mosaic virus behaves completely different.

Chapter six deals with the molecular interactions during the assembly process. In this chapter it is shown that the arginine and lysine containing part of the N-terminal arm is responsible for the binding of oligo-nucleotides. It is suggested that a double stranded nucleotide conformation is required for proper interaction.

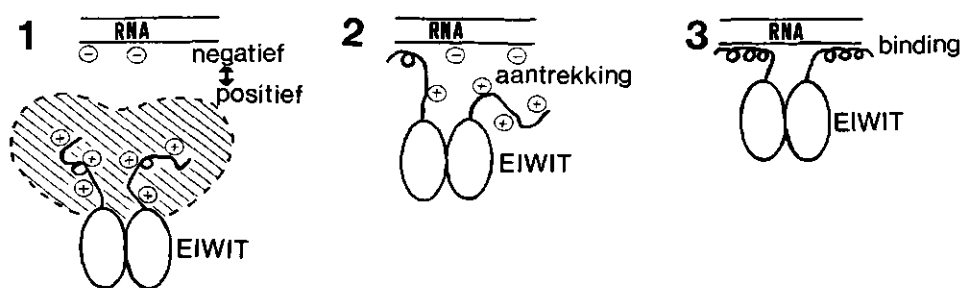
In chapter 7 a secondary structure prediction of the coat protein of CCMV is presented. The results indicate that the N-terminal arm shows no structural preference when the positive charges of arginine and lysine are retained, whereas in the absence of these positive charges there is a tendency towards α -helix formation. Also it is concluded that cowpea chlorotic mottle virus possesses the same β -role topology as southern bean mosaic virus, satellite tobacco necrosis virus and tomato bushy stunt virus.

Chapter eight presents the results of an energy calculation study on the N-terminal arm. Although the method used contains too many approximations to allow detailed conclusions, the results of chapter seven are fully confirmed by these calculations.

In the model for the protein-RNA interaction in CCMV which is presented in chapter five, a random coil conformation for the N-terminal protein arm is invoked. In chapter nine an extension of this part of the model is presented. It is shown that the N-terminal possesses some time-averaged secondary structure (presumably an α -helix) in the absence of RNA. A conformational exchange process is observed in the N-terminal arm, in which about half of the residues participate. In empty capsids not all the N-terminal arms turn out to be mobile. 20-55% of the N-terminal arms is immobile on the NMR timescale in the empty capsids of the coat protein of CCMV.

SAMENVATTING.

Dit proefschrift beschrijft de toepassing van ^1H - en ^{13}C NMR, ESR, ST-ESR en computer simulaties voor het bestuderen van cowpea chlorotic mottle virus. Dit virus bestaat uit RNA ingekapseld in 180 gelijke eiwit moleculen. Deze eiwit moleculen zijn icosaedrisch gerangschikt. Na isolatie kan dit manteleiwit in een aantal goed gedefinieerde aggregatie toestanden gebracht worden. Dit onderzoek was mogelijk omdat voldoende grote hoeveelheden van deze aggregatie toestanden geproduceerd kunnen worden om er magnetische resonantie metingen aan te kunnen doen. Alle verkregen resultaten zijn gecombineerd in het volgende model voor het assemblage proces:



Dit model gaat uit van een starre eiwiteenheid met daaraan vast een zeer beweeglijke, negatief geladen N-terminale arm. Deze arm is het RNA bindende stuk van het eiwit. De grote beweeglijkheid bevordert de kans dat het eiwit en het RNA kunnen reageren, bovendien kan het eiwit door deze beweeglijkheid op verschillende manieren reageren met andere locale structuren in het RNA. Als het eiwit aan het RNA bindt vormt het een starre α -helix.

In hoofdstuk twee wordt aangetoond dat geen beweeglijkheid sneller dan 10^{-7} s waargenomen kan worden in het intacte virus.

In hoofdstuk drie worden de resultaten gepresenteerd van een ESR en ST-ESR studie aan gespinlabeld virus. De resultaten tonen aan dat in het virus geen anisotrope beweeglijkheid van de eiwiteenheden waargenomen kan worden op de 10^{-5} - 10^{-7} s tijdschaal. In dit hoofdstuk wordt voorts gesuggereerd dat een aantal ST-ESR studies aan maleimide gelabelde eiwitten, waaruit anisotrope eiwit beweeglijkheid is afgeleid, op artefacten berusten.

In hoofdstuk vier wordt aangetoond dat de N-terminale arm het RNA bindende deel van het eiwit is. Bovendien wordt bewezen dat deze arm zeer beweeglijk is in de afwezigheid van RNA en dat de arm star is in de aanwezigheid van RNA.

In hoofdstuk vijf wordt een uitstapje naar andere virussen beschreven. Er wordt aangetoond dat brome mosaic virus en belladonna mottle virus zich net zo gedragen als cowpea chlorotic mottle virus, terwijl cowpea mosaic virus compleet afwijkend gedrag laat zien.

In hoofdstuk zes worden de moleculaire interacties gedurende het assemblage proces besproken. In dit hoofdstuk wordt aangetoond dat het arginine en lysine bevattende deel van de N-terminale arm verantwoordelijk is voor de binding van oligo-nucleotiden. Bovendien wordt de hypothese geopperd dat een dubbelstrengs nucleotide structuur nodig is voor binding.

In hoofdstuk zeven wordt de voorspelling van de secundaire structuur van de eiwiteenheid gepresenteerd. De resultaten duiden erop dat de N-terminale arm geen preferentie voor enige regelmatige structuur vertoont wanneer de arginine en de lysine zijketens geladen zijn, terwijl de α -helix veruit het gunstigst is na neutralisatie. Verder wordt geconcludeerd dat de eiwiteenheid van cowpea chlorotic mottle virus dezelfde " β -role" topologie bezit als southern bean mosaic virus, tomato bushy stunt virus en satellite tobacco necrosis virus.

Hoofdstuk acht beschrijft de resultaten van energie berekeningen aan de N-terminale arm. Ondanks het feit dat de gebruikte methode teveel benaderingen bevat om gedetailleerde conclusies te kunnen trekken, bevestigen deze resultaten de conclusies van het vorige hoofdstuk.

Het model voor de eiwit-RNA interactie in cowpea chlorotic mottle virus dat in hoofdstuk vijf is gepresenteerd, bevat een "random-coil" conformatie voor de N-terminale arm. In hoofdstuk negen wordt dit deel van het model uitgebreid. In dit hoofdstuk wordt aangetoond dat de N-terminale arm gemiddeld in de tijd enige secundaire structuur (vermoedelijk een α -helix) bevat indien geen RNA aanwezig is. Ongeveer de helft van de residuen in de N-terminale arm is betrokken in een uiwisselings proces tussen meerdere conformaties. In lege eiwit bollen zijn niet alle N-terminale armpjes beweeglijk. 20-55% van deze armpjes is star op de NMR tijdschaal.

CURRICULUM VITAE.

Gerrit Vriend is geboren op 26 april 1956 te Borne. In 1974 behaalde hij het Atheneum B diploma aan het Goois Lyceum te Bussum. In datzelfde jaar begon hij zijn studie chemie aan de Rijks Universiteit Utrecht. In 1977 behaalde hij zijn kandidaats examen S2 (biologische orientatie van de chemie). Op 23 mei 1979 behaalde hij zijn doctoraal examen (met genoegen), met als hoofdvakken Biochemie en Informatica.

Op 1 september 1979 werd hij vanwege de Stichting voor Biofysica (ZWO) wetenschappelijk medewerker bij de vakgroep Moleculaire Fysica van de Landbouwhogeschool Wageningen. Dit proefschrift is het resultaat van vier jaar onderzoek aldaar.

Appendix A. Chemical shifts.

Chemical shifts of protons in peptides.

amino acid	H α	H β	H γ	H δ	H ϵ	H2	H3	H4
Ala	4.35	1.43						
Arg	4.4	1.85	1.70	3.26				
		1.70						
Asn	4.76	2.92						
		2.90						
Asp	4.77	2.84						
		2.75						
Cys	5.69	3.28						
		2.96						
Gln	4.37	2.13	2.38					
		2.01						
Glu	4.29	2.09	2.31					
		1.96	2.28					
Gly	3.97							
His	4.63	3.26				8.12		7.14
		3.19						
Ile	4.22	1.89	1.47	0.88				
			1.19					
			0.94					
Leu	4.39	1.65	1.65	0.88				
Lys	4.36	1.85	1.47	1.70	3.02			
		1.73						
Met	4.51	2.16	2.63	2.12				
		2.00						
Phe	4.66	3.22					aromatics	7.34
		2.99						
Pro	4.47	2.30	2.03	3.65				
		1.98						
Ser	4.50	3.89						
Thr	4.35	4.18	1.17					
Trp	4.70	3.22				7.24		
		3.20						
Tyr	4.60	3.13				6.88	7.15	
		2.92						
Val	4.18	2.13	0.97					
			0.95					

Chemical shifts of carbon nuclei relative to CS₂

Ile C δ	181.3	Cys C β	155.6
Ile C γ 2	177.7	Asp C β	155.2
Ala C β	176.3	Lys C ϵ	153.4
Val C γ 2	176.0	Leu C β	153.1
Val C γ 1	175.0	Arg C δ	152.1
Thr C γ	174.1	Gly C α	150.3
Leu C δ 2	171.7	Pro C δ	146.9
Lys C γ	170.9	Ala C α	143.5
Leu C δ 1	170.6	Cys C α	142.0
Arg C γ	168.7	Asn C α	141.9
Pro C γ	168.5	Asp C α	141.4
Leu C γ	168.5	Leu C α	140.0
Ile C γ 1	167.7	Arg C α	139.6
Lys C δ	166.6	Gln C α	139.3
Trp C β	166.6	Lys C α	138.9
Gln C β	166.0	Glu C α	138.9
Arg C β	165.0	Trp C α	139.6
Glu C β	164.8	Phe C α	137.6
Pro C β	163.6	Tyr C α	137.4
Lys C β	162.7	Ser C α	137.3
Val C β	162.6	Ile C α	134.5
Gln C γ	161.6	Val C α	134.1
Glu C γ	159.1	Thr C α	133.7
Asn C β	157.6	Pro C α	131.9
Tyr C β	156.8	Ser C α	131.6
Phe C β	156.0	Thr C β	125.8
Ile C β	155.8	Trp C γ	84.0
Trp C ζ 2	80.9	Tyr C γ	64.9
Tyr C ϵ 1/ ϵ 2	77.4	Phe C δ 1/ δ 2	64.0
Trp C ϵ 3/n2	74.4	Phe C ϵ 1/ ϵ 2	63.5
Trp C ϵ 3/n2	73.4	Tyr C δ 2	62.4
Trp C δ 3/ δ 1	70.8	Tyr C δ 1	62.4
Trp C δ 1/ δ 3	67.7	Phe C γ	56.4
Trp C δ 2	66.2	Trp C ϵ 2	56.4
Phe C ζ	65.6	Tyr C ζ	38.2
		Arg C ζ	36.1

Proton Resonances are from Bundi, A., (1977), Thesis, ETH, Zurich.

Carbon resonances are from James, T.J., (1975). NMR in Biochemistry, 246-247, Academic Press, New York.

Appendix B. Primary structure of CCMV in one and three letter code.

1 S SER 37 S SER 73 L LEU 109 T THR 145 T THR 181 F PHE
 2 T THR 38 G GLY 74 P PRO 110 E GLU 146 L LEU 182 D ASP
 3 V VAL 39 Q GLN 75 N ASN 111 T THR 147 E GLU 183 D ASP
 4 G GLY 40 G GLY 76 E GLU 112 Q GLN 148 Q GLN 184 S SER
 5 T THR 41 K LYS 77 L LEU 113 T THR 149 L LEU 185 F PHE
 6 G GLY 42 A ALA 78 S SER 114 T THR 150 A ALA 186 T THR
 7 K LYS 43 I ILE 79 S SER 115 A ALA 151 A ALA 187 P PRO
 8 L LEU 44 K LYS 80 E GLU 116 A ALA 152 D ASP 188 V VAL
 9 T THR 45 A ALA 81 R ARG 117 A ALA 153 L LEU 189 Y TYR
 10 R ARG 46 W TRP 82 N ASN 118 S SER 154 T THR
 11 A ALA 47 T THR 83 K LYS 119 F PHE 155 I ILE
 12 Q GLN 48 G GLY 84 Q GLN 120 Q GLN 156 Y TYR
 13 R ARG 49 Y TYR 85 L LEU 121 V VAL 157 L LEU
 14 R ARG 50 S SER 86 K LYS 122 A ALA 158 Y TYR
 15 A ALA 51 V VAL 87 V VAL 123 L LEU 159 S SER
 16 A ALA 52 S SER 88 G GLY 124 A ALA 160 S SER
 17 A ALA 53 K LYS 89 R ARG 125 V VAL 161 A ALA
 18 R ARG 54 W TRP 90 V VAL 126 A ALA 162 A ALA
 19 K LYS 55 T THR 91 L LEU 127 D ASP 163 L LEU
 20 N ASN 56 A ALA 92 L LEU 128 N ASN 164 T THR
 21 K LYS 57 S SER 93 W TPP 129 S SER 165 E GLU
 22 R ARG 58 C CYS 94 L LEU 130 K LYS 166 G GLY
 23 N ASN 59 A ALA 95 G GLY 131 D ASP 167 D ASP
 24 T THR 60 A ALA 96 L LEU 132 V VAL 168 V VAL
 25 R ARG 61 A ALA 97 L LEU 133 V VAL 169 I ILE
 26 V VAL 62 E GLU 98 P PRO 134 A ALA 170 V VAL
 27 V VAL 63 A ALA 99 S SER 135 A ALA 171 H HIS
 28 Q GLN 64 K LYS 100 V VAL 136 M MET 172 L LEU
 29 P PRO 65 V VAL 101 S SER 137 Y TYR 173 E GLU
 30 V VAL 66 T THR 102 G GLY 138 P PRO 174 V VAL
 31 I ILE 67 S SER 103 T THR 139 E GLU 175 E GLU
 32 V VAL 68 A ALA 104 V VAL 140 A ALA 176 H HIS
 33 E GLU 69 I ILE 105 K LYS 141 F PHE 177 V VAL
 34 P PRO 70 T THR 106 S SER 142 K LYS 178 R ARG
 35 I ILE 71 I ILE 107 C CYS 143 G GLY 179 P PRO
 36 A ALA 72 S SER 108 V VAL 144 I ILE 180 T THR

Appendix C. Optimization procedure for the parameters in the structure prediction methods.

1. Boundaries for each potential were established 5% (10%) above and below the initial potential.
2. A minimization procedure according to Powell [ref. 2 in chapter 8] was performed. In this method inverse of the number of correctly predicted residues was minimized by varying the structure potentials.
3. After obtaining a local minimum one parameter was randomly perturbed, whereafter the minimization procedure was restarted.
4. Step three was repeated until all parameters were at least five times perturbed.
5. No potential was ever allowed to cross the boundaries 5% (10%) above or below the initial potential.

Appendix D. List of abbreviations.

AMV	alfalfa mosaic virus
APLV	adean potato latent virus
B	secondary structure prediction method of Burgess et al.
BBMV	broad bean mottle virus
BdMV	belladonna mottle virus
BdMV-B	bottom component of belladonna mottle virus
BdMV-T	top component of belladonna mottle virus
BMV	broam mosaic virus
BMV-PT	artificial top component of broam mosaic virus
CCMV	cowpea chlorotic mottle virus
C+F	secondary structure prediction method of Chou and Fasman
CPMV	cowpea mosaic virus
CPMV-B	bottom component of cowpea mosaic virus
CPMV-M	middle component of cowpea mosaic virus
CPMV-T	top component of cowpea mosaic virus
D	dielectric constant
DSS	2,2-dimethyl 2-silapentane-5-sulphonate
DTT	dithiothreitol
EPR	electron paramagnetic resonance
FID	free induction decay
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
oligo-A	oligo-adenosine
poly-A	poly-adenosine
ppm	parts per million
Qi	partial charge of atom i
Qj	partial charge of atom j
Rij	distance between atoms i and j
S	Svedberg constant
ST-EPR	saturation transfer electron paramagnetic resonance
SBMV	southern bean mosaic virus
SL-CCMV	spin-labelled cowpea chlorotic mottle virus
STNV	satellite tobacco necrosis virus
T	triangulation number or Tesla
TBSV	tomato bushy stunt virus
TMV	tobacco mosaic virus

TYMV	turnip yellow mosaic virus
U	energy
(Ap) ₈ A	ribonucleic acid form of (adenosine-phospho) ₈ -adenosine
(Ap) ₁₄ A	a mixture of the ribonucleic acid form of (adenosine-phospho) _x -adenosine molecules with x approximately ranging from 13 to 15
(AT) ₅	deoxy-ribonucleic acid form of (adenosine-phospho-thymine-phospho) ₄ -adenosine-phospho-thymine
D ₂ O	deuterated water
¹³ C	carbon 13
2D-NOE	two dimensional nuclear Overhauser effect
¹ H	proton
T ₁	spin-lattice relaxation time
T ₂	spin-spin relaxation time
τ _c	rotational correlation time
τ ₁	waiting time
τ _m	mixing time
μ	ionic strength

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V