

**The Membrane-Bound Form
of Gene 9 Minor Coat Protein of
Bacteriophage M13**

M.C. Houbiers

promotor:

prof. dr. T. J. Schaafsma

hoogleraar in de moleculaire fysica, Wageningen Universiteit

co-promotor:

dr. M. A. Hemminga,

universitair hoofddocent, Laboratorium voor Biofysica, Wageningen Universiteit

Promotiecommissie:

prof. dr. J.M. Vlak,

Wageningen Universiteit

prof. dr. A. Kuhn,

Universiteit van Hohenheim, Stuttgart

dr. E. Goormaghtigh,

Vrije Universiteit, Brussel

prof. dr. J.A. Killian,

Universiteit Utrecht

The Membrane-Bound Form of Gene 9 Minor Coat Protein of Bacteriophage M13

M.C. Houbiers

Proefschrift

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
prof. dr. ir. L. Speelman
in het openbaar te verdedigen
op maandag 30 september 2002
des namiddags te vier uur in de aula

Houbiers, M.C.

The Membrane-Bound Form of Gene 9 Minor Coat Protein of Bacteriophage M13

Thesis, Wageningen University

ISBN: 90-5808-692-5

Dankwoord

Dit proefschrift is tot stand gekomen met de hulp en steun van velen, die ik hier wil bedanken. Ook alle anderen die op een of andere manier een bijdrage geleverd hebben en hier niet genoemd worden wil ik bedanken.

Het begon als een NMR project. Ik heb veel hulp gehad uit Nijmegen van Tineke Papavoine. Bij haar begeleider Frank van de Ven, aan wie ik een dankbare herinnering heb, kon ik altijd terecht voor hulp. Groot was de schok over zijn plotselinge overlijden. In Wageningen werd ik op weg geholpen door Adrie en Enrico, en later ook door Carlo van Mierlo. Ook aan het NMR-clubje heb ik veel gehad.

Cees Versluijs, bedankt voor jouw belangeloze hulp bij de lastige massaspectrometrie metingen. Uiteindelijk kwamen we terecht bij Unilever in Vlaardingen, waar Anita Oosterlaken-Buijs ons geholpen heeft met de MALDI-TOF metingen.

Rudy, hartstikke bedankt voor alle hulp met de monolaag metingen, maar ook voor de recepten en tuintips, en ook voor de plantjes die ik van je kreeg voor de moestuin. Anne en Wim, bedankt voor alle hulp en het me wegwijs maken op het lab. Jullie hielpen me om me thuis te voelen op jullie vakgroep en in het Utrechtse. En tof dat ik met jullie een stukje mee kon lopen op de GR20 in Corsica! Ben de Kruijff, bedankt voor je interesse en suggesties tijdens mijn drie-maands verblijf op jouw vakgroep.

Harmen de Jongh, bedankt voor al je ideeën en de hulp bij de georiënteerde CD-metingen. Frank de Lange, jammer dat de metingen niet doorgingen vanwege wat technische probleempjes. Toch bedankt.

Erik Goormaghtigh, thank you for your kind hospitality in Brussels and all your help with the ATR-FTIR measurements. I appreciate very much your taking time to discuss about the results. Jean-Marie Ruyschaert, thanks for your interest when I was at your department.

Folkert Hoekstra, Wim Wolkers en Mark Alberda, bedankt voor alle hulp bij FTIR metingen in Wageningen.

Antoinette Killian, Anja Ridder, Andreas Kuhn en de anderen van de Utrecht-Wageningen overleggroep wil ik ook bedanken. Dit overleg was erg nuttig en gaf me steeds weer nieuwe ideeën en inspiratie.

Jan, Enrico, Kitty, Sander, Shirley, Tom & Tom, Frank, Ullrich en alle anderen van de vakgroep: bedankt voor alle gezelligheid, de terrasbezoekjes, de gezellige pauzes, het samen eten op de mensa, de langlaufuitjes, etc. Marcus, bedankt dat je me de kans gaf dit project te doen en af te ronden. En bedankt dat je me de gelegenheid gaf om cursussen en congressen te bezoeken die erg stimulerend waren en waarvan ik weer vol goede moed en met ideeën terug kwam.

Speciaal wil ik mijn kamergenoten bedanken. Eerst zat ik bij David. Hvala! Dank voor de inspirerende discussies over de viruseiwitten. Ik oefende met jou het Nederlands. En tijdens het gastvrije verblijf in jouw land leerde ik een beetje Sloveens van jou en je familie. We hebben heerlijke wijn gemaakt en hoge bergen beklommen. Jouw aansporing als het even moeilijk ging: “The only way is up!”.

Later deelde ik met Ivon een kamer. Bedankt voor de gezelligheid, gesprekken over werk en ook niet-werk, en het kwijt kunnen van alle frustraties die we tegen kwamen, voor alle dropjes, spelletjesavonden, en vele leuke wandelweekenden. En wie had dat nu gedacht in het begin: is het ons toch allebei gelukt!

Gelukkig mocht ik ook nog een student begeleiden: Yves, bedankt voor je werk. Het was leuk om met jou na gedane arbeid de toerist uit te hangen in Brussel en te genieten van al het lekkers daar (mosselen, wafels, bonbons).

In het bijzonder wil ik Cor en Ruud bedanken voor de bergen werk die ze hebben verricht op het lab, maar vooral ook voor alle discussies, kritische bijdragen en motiverende belangstelling. Dankzij jullie kon ik de motivatie houden om door te gaan. Ik heb veel technieken van jullie geleerd op het lab. Altijd weer spannend wat er uit een eiwit-gel of een blot zou komen! Die eerste blot had toch een prachtig kleurtje: helemaal paars!

Petra, tijdens onze pizza-lunches op donderdagen konden we altijd lief en leed van ons leven als onderzoekers in opleiding delen.

Catharien, bedankt voor je steun en je kritische kijk op alles. Zelfs vanaf de andere kant van de wereld kon jij me door je brieven moed geven. Ik herinner me nog steeds onze discussie over de zin(loosheid) van “bergen verzetten”.

Baraksgenoten van Droevendaal, elke dag was het weer heerlijk om thuis te komen en samen te genieten en lekker te eten in onze tuin. De gezelligheid hielp mij de tegenslagen bij het onderzoek te relativeren, zodat ik telkens weer frisse moed kon putten.

Wim, bedankt voor je digitale bijdrage aan de kافت van dit proefschrift.

Makandra, bedankt dat ik de kans kreeg om op het land te werken met zoveel lieve mensen, toen heel mijn lichaam protesteerde tegen het werk aan het proefschrift. In de tuin kon ik weer aarden en genieten van het leven.

Mijn ouders wil ik bedanken voor alle kansen en steun die ze mij hebben gegeven. Familie en vrienden, bedankt voor alles buiten het werk om.

Emiel, ook jij hebt bijgedragen aan dit boekje zonder dat je er iets van wist. Tijdens het wachten op jouw komst vond ik de rust en de tijd om dit af te maken.

Het allermeest wil ik Cor bedanken voor alle liefdevolle hulp. Jouw onvoorwaardelijke steun was essentieel bij het afronden van dit proefschrift. Bedankt! (Kunnen we toch eindelijk die kast opruimen!).

Contents

Abbreviations

Chapter 1:	General Introduction	1
Chapter 2:	Conformational and Aggregational Properties of the Gene 9 Minor Coat Protein of Bacteriophage M13 in Membrane-mimicking Systems. <i>(Biochemistry 38 (1999), 1128-1135)</i>	29
Chapter 3:	Conformation and Orientation of the Gene 9 Minor Coat Protein of Bacteriophage M13 in Phospholipid Bilayers. <i>(Biochim. Biophys. Acta 1511 (2001), 224-235)</i>	55
Chapter 4:	Spontaneous Insertion of Gene 9 Minor Coat Protein of Bacteriophage M13 in Model Membranes. <i>(Biochim. Biophys. Acta 1511 (2001), 309-316)</i>	79
Chapter 5:	Summarizing Discussion	95
	Summary	108
	Samenvatting	111
	Curriculum vitae	115
	List of Publications	117

Het in dit proefschrift beschreven onderzoek werd mogelijk gemaakt met financiële steun van de Stichting Levenswetenschappen (SLW) van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Keywords in this thesis:

**Bacteriophage M13 / Minor Coat Protein / Membrane Protein / CD / ATR-FTIR/
Conformation / Orientation / Aggregation / Membrane Insertion**

Abbreviations

ATR-FTIR	attenuated total reflection Fourier transform infrared
CD	circular dichroism
C16S	g9p with mutation Cys ₁₆ to Ser
C16S/S32C	g9p with mutations Cys ₁₆ to Ser and Ser ₃₂ to Cys
CF	5,6-Carboxyfluorescein
DMPC	1,2-dimyristoyl- <i>sn</i> -glycero-3-phosphocholine
DMPG	1,2-dimyristoyl- <i>sn</i> -glycero-3-phosphoglycerol
DOPC	1,2-dioleoyl- <i>sn</i> -glycero-3-phosphocholine
DOPG	1,2-dioleoyl- <i>sn</i> -glycero-3-phosphoglycerol
DPPC	1,2-dipalmitoyl- <i>sn</i> -glycero-3-phosphocholine
DSC	differential scanning calorimetry
EDT	ethane dithiol
Fmoc	9-fluorenylmethoxycarbonyl
FTIR	Fourier transform infrared
G7p, G8p, G9p	respectively gene 7-, gene 8- and gene 9-protein
HPSEC	high performance size exclusion chromatography
L/P	lipid to protein molar ratio
LUV	large unilamellar vesicle
PMSF	phenylmethylsulfonyl fluoride
R	dichroic ratio
RT	room temperature
SDS	sodium dodecyl sulfate
TFA	trifluoroacetic acid
TFE	trifluoroethanol

Chapter 1

General introduction

This thesis describes our attempts to gain knowledge about an intriguing protein: the gene 9 minor coat protein (g9p) of bacteriophage M13. Although it is called a 'minor' protein because of the few copies present in the phage particles it may play a major role in initiating the phage assembly process, which occurs in the *Escherichia coli* membrane. Therefore, it is worth learning about the properties of this small, but important protein. Because g9p is a membrane protein, this introduction starts with a short description of membranes, membrane proteins and protein-lipid interactions. Also several useful methods which are available to reveal properties of this protein in the membrane will be briefly described, with a focus on infrared spectroscopy. Then the introduction continues with the current insights in bacteriophage M13, with special attention for the g9p and its role in the phage and in the assembly process. In the final part the aim of this study and the outline of this thesis are presented.

ABOUT MEMBRANES AND THEIR PROTEINS

Biomembranes

Function

Biological membranes form the barriers that separate the aqueous contents of a cell from the outside world. The membrane regulates communication across the barrier. This can happen by passage of ions or molecules across the membrane, or by passage of information via structural changes in membrane components. In eukaryotic organisms membranes also separate compartments within one cell. Compartmentalization separates different biochemical processes and therefore allows for structural and functional differentiation. Furthermore, membranes can also accommodate multi-enzyme complexes, which perform subsequent steps. Thereby the membranes are important in facilitating metabolic conversions.

Architecture

Biological membranes consist of lipids and proteins which are held together by non-covalent interactions. The basic structure of the membrane is formed by the lipids, which are arranged as a continuous double layer, 4 to 5 nm thick and relatively

impermeable to most water-soluble molecules [1]. Proteins present in this double layer are responsible for numerous biochemical functions in the membrane. Membranes are dynamic structures where lipids and proteins can diffuse laterally to some degree, as visualized in the fluid mosaic model of Singer and Nicholson [2]. However, diffusion of some membrane proteins is limited and not random, and even specialized domains may exist [3].

Lipids

There is a wide variety of lipids, of which the most abundant are phospholipids. Phospholipids contain a headgroup, which consists of a glycerol backbone, a negatively charged phosphate group, and an alcohol moiety. The pH determines the net charge of the headgroup. The most common alcohol moieties of phospholipids in cells are choline, glycerol and ethanolamine. Also two acyl chains are attached to the glycerol backbone via an ester linkage. The length (varying from 14 to 24 carbon atoms, but most commonly 16 or 18 carbon atoms) and unsaturation of the chains vary and affect the membrane fluidity. The combination of a polar headgroup and hydrophobic chains gives the lipid an amphipathic character, which leads to self-assembly into a bilayer structure in an aqueous environment. However, if the overall shape of a lipid deviates from a cylindrical shape and is more cone shaped due to a small ratio of the headgroup and chain cross-sectional areas, it has a tendency to form a non-bilayer structure, such as a hexagonal phase, instead of a bilayer [4].

Polarity profile

The lipid molecules are organized into a planar bilayer structure in which the hydrophobic parts of the lipids are in the core of the bilayer, whereas the hydrophilic headgroups are at the outside in contact with the aqueous environment. This results in a hydrophobic core region and a polar head group region, which forms the interface between the apolar core and the aqueous environment. The boundaries between these two regions are not sharp, resulting in a heterogeneous interface region [5]. Going from the inside to the outside of the membrane there is a polarity gradient, which is steepest around the chain carbonyl groups just under the lipid headgroups. The cross-sectional

polarity profile of a 1,2 dioleoyl-sn-glycero-3-phosphocholine (DOPC) bilayer and the time-averaged distribution of the groups are depicted in Figure 1.

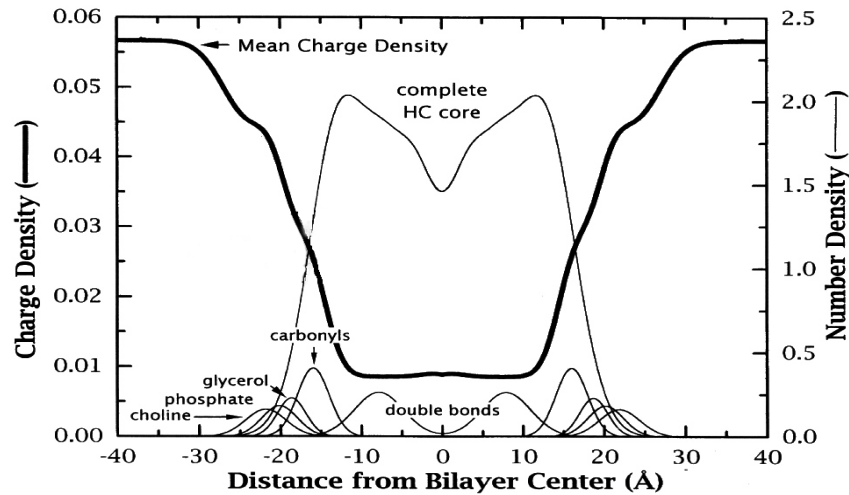


Figure 1: Group distribution of 1,2 dioleoyl-sn-glycero-3-phosphocholine (DOPC) assembled in a fluid bilayer. The thick line indicates the polarity profile from the outside to the inside of the membrane (after White and Whimley [6]).

***Escherichia coli* membrane**

The gram-negative bacteria *E. coli* is surrounded by two membranes: the inner and the outer membrane. This results in three compartments: the cytoplasm, the periplasm, and the extracellular medium. Lipids present in the inner membrane of *E. coli* cells are the zwitterionic phosphatidylethanolamine, 75-80% of the total phospholipids [7], and the at physiological pH negatively charged phosphatidylglycerol (20%) and cardiolipin (1-5%). The lipids in the inner membrane are organized in a liquid-crystalline bilayer [8].

Membrane proteins

Membrane proteins can be divided into two classes: peripheral and integral membrane proteins. Peripheral membrane proteins are bound to the surface of the membrane, whereas integral membrane proteins span the membrane with one or more trans-membrane segments. These transmembrane segments show some common structural

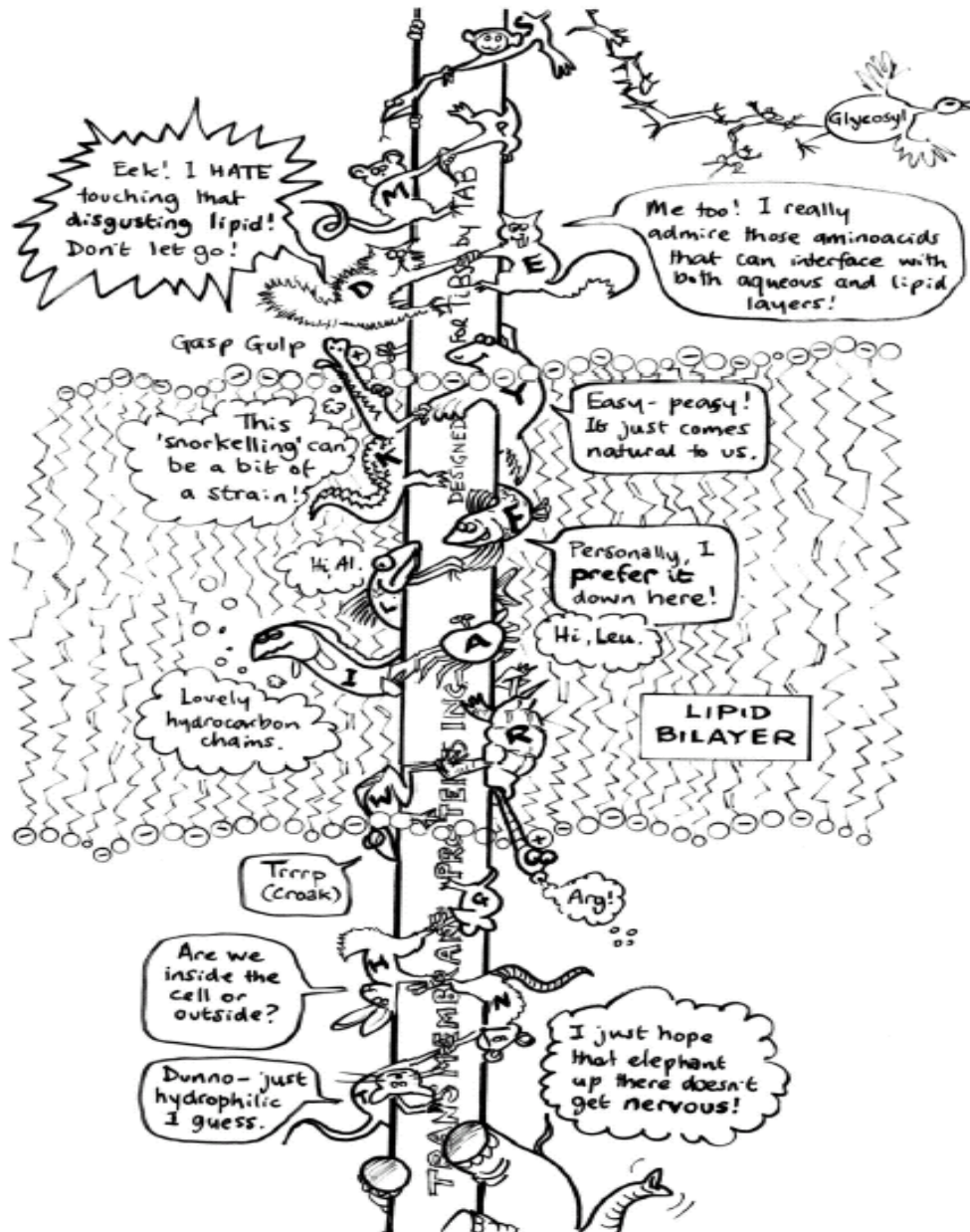
features, as the conformation is mainly determined by the hydrophobic environment of the membrane.

Main chain

Since water is excluded from the hydrophobic membrane core and the lipid fatty acyl chains are unable to establish hydrogen bonds, the polar main chain hydrogen bonding groups (NH and CO) of the transmembrane protein segment form hydrogen bonds with each other instead of with water. Satisfaction of (most) main chain hydrogen bonds forms the major constraint and induces formation of two stable periodic secondary structures, α -helices or β -sheets: Hydrogen bonding between groups of adjacent transmembrane segments gives β -sheets, whereas hydrogen bonding between residues within a single transmembrane segment gives α -helices. For the hydrogen bonds at the edges of the sheet to be satisfied, the sheet must close upon itself to form a cylinder (β -barrel) with a hydrophobic outside facing the lipids and a more polar inside lining a central pore as is the case for bacterial porins [9]. In gram negative bacteria and mitochondria these β -barrel structures seem to be restricted to the outer membrane. Proteins with transmembrane α -helical segments are present in the inner membrane of bacteria, in the endoplasmic reticulum (ER) membrane of eukaryotes, in the inner membrane of mitochondria and in the chloroplast thylakoid membrane. A single α -helix can be accommodated in the membrane, but many membrane proteins are found to span the membrane with several α -helices. Many α -helix forming peptides assume an unordered structure in aqueous buffer, and the α -helices are only induced upon binding and insertion into lipid membranes. A significant contribution to the driving force of this conformational change is most likely a change in hydrogen bonding from peptide backbone-water hydrogen bonds in the unordered conformation to intramolecular peptide backbone hydrogen bonding in the α -helical structure.

Side chains

Also the positioning of specific amino acid side chains within the transmembrane segments shows some general characteristics. Hydrophobic side chains tend to be expelled from the water and prefer the hydrophobic acyl chain environment. Hydrogen



Therefore they prefer the polar phase. Side chains from tryptophan or tyrosine (also hydrogen bonding) are more easy to bury into the lipids. These amino acids prefer to position in the interface between the polar headgroup and the acyl chains of the membrane bilayer. This leads to the general feature that a transmembrane α -helical segment consists of a hydrophobic stretch, flanked by short segments enriched in aromatic residues and broader bands containing charged residues [10-12]. This is cartoonically depicted in Figure 2.

For an α -helix the hydrophobic stretch contains approximately 20 residues [13]. By prediction algorithms such hydrophobic segments can be searched for [13, 14]. In this way, transmembrane segments and their topology can be predicted from sequence data.

Protein-lipid interactions

Protein-lipid interactions are important for both peripheral and integral membrane proteins.

Peripheral membrane proteins can interact with the membrane surface in various ways: (1) by electrostatic interactions between charged phospholipid headgroups and charged amino acid side chains, (2) by hydrophobic interactions without traversing the membrane, (3) by binding to another protein, which is present in the membrane, (4) by anchoring to the membrane by a covalently attached protein acyl chain. Also a combination of these interactions is possible.

Integral membrane proteins can interact with the membrane by hydrophobic as well as electrostatic interactions. The hydrophobic stretch interacts with the hydrophobic core of the membrane, and therefore, the length of this stretch should approximately match the thickness of the membrane core. The polar residues can interact with the polar lipid headgroups [9, 11-13, 15]. In this region a statistical preference is found for charged amino acid residues [13, 15], which are thought to function as an 'anchor' in the interface region [10, 16, 17]. In addition, charged residues can also function as topological determinants (see below). Also aromatic amino acids like tryptophans and tyrosines show up more frequently in the interface region as is found by statistical

analysis [12, 15]. Probably aromatic residue side chains prefer a precise location in the interface region [10]. In this way aromatic residues may also function as 'anchors'.

Matching of the length of the hydrophobic protein segment with the thickness of the hydrophobic membrane core plays also an important role in protein-lipid interactions. If the transmembrane helix is too short (e.g. 15 residues in one helix in leader peptidase from *E. coli*) or too long (e.g. 24 residues in lactose permease) the protein or membrane can react by several mechanisms to avoid this energetically unfavourable mismatch situation, for example: protein segregation in domains, tilting of protein segments, protein backbone adaptation, preferential surrounding of proteins by lipids with matching acyl chain lengths, formation of non-bilayer structures, or by failure the incorporation of the protein in the membrane [10, 16]. Such adaptations will influence the membrane organization and consequently membrane function [16].

Protein insertion and translocation in the *E. coli* inner membrane

Topological signals

Membrane proteins contain topogenic determinants in their sequence, which direct their translocation across the membrane. Membrane proteins can in fact be regarded as partially translocated, since they have got stuck and permanently stay in the membrane by the effect of additional signals. Topological signals are: cleavable signal sequences, uncleavable signal sequences, stop-transfer sequences and reverse signal sequences. In general, these topological signals are hydrophobic regions surrounded by hydrophilic amino acids. Signal sequences consist of three portions: a positively charged N-terminus, a hydrophobic core of 7-15 amino acid residues, and a C-terminal polar region containing the signal peptidase cleavage site to remove the signal sequence after translocation [18]. Cleavable signal sequences initiate translocation of C-terminal hydrophilic regions of proteins and are cleaved during translocation. An example is the major coat protein of bacteriophage M13, which is synthesized in a precursor form. Uncleaved signal sequences also initiate C-terminal translocation, but remain as a membrane anchor region, as for example in the *E. coli* leader peptidase [19, 20]. These cleavable and uncleavable signal sequences orient in the membrane with the N-terminus in the cytoplasm and the C-terminus out (N_{cyto}, C_{out}). Stop-transfer domains function to hold

the translocating polypeptide chain in the membrane and keep the opposite orientation (N_{out}, C_{cyto}). Reverse signals initiate translocation in the N-terminal direction and remain as transmembrane anchors [21, 22]. There are also sequences functioning as an insertion domain, where two hydrophobic domains separated by a polar region may act in concert to translocate the central polar domain [23, 24]. These processes are reviewed by Dalbey and Kuhn [25].

Insertion pathways in E. coli

Several pathways are known for the insertion of proteins into the *E. coli* membrane. Four known pathways, or targeting routes, will be mentioned here:

1. The Sec-dependent pathway, which exports a large number of newly synthesized outer membrane and periplasmic proteins across the inner membrane [26]. These proteins have a common signal peptide of 18-26 amino acids in their N-terminal region [27]. (For the characteristics of signal peptides, see above). The Sec machinery transports polypeptides in an unfolded state [28]. The integral membrane protein complex SecYE is believed to function as a protein channel through which the exported protein moves [29]. The membrane proteins SecG, YajC, Sec D and SecF assist in this process and enhance the efficiency [30]. It has been shown that SecYE together with SecA constitute the minimal requirement for protein translocation [31, 32]. The ATPase SecA is also part of the translocase. SecA, which cycles between a cytosolic and a membrane inserted state, drives the translocation of a preprotein in a stepwise way [33]. This requires the proton motive force [34] and ATP hydrolysis [35]. SecA is the receptor for the cytosolic molecular chaperone SecB. SecB associates with fully translated preprotein or with nascent polypeptide [29, 36] and maintains the translocation-competent state while targeting it to the plasma membrane. Recently a new component YidC has been shown to interact with SecY and nascent proteins [37]. YidC depletion appeared not only to affect membrane integration of Sec-dependent proteins, but also inhibited insertion of the Sec-independent M13 procoat protein. Proteins to be exported to the periplasm were not affected by YidC depletion. Thus, YidC seems to be a translocation component specific for membrane proteins. YidC may be an important component of the Sec translocon, but its precise function is still under investigation. [38]

More detailed information on this insertion pathway can be found in recent reviews on protein insertion and translocation [25, 39].

2. The signal recognition particle (SRP) pathway, which is mainly involved in translocation of inner membrane proteins [40]; also some secreted proteins follow this pathway [41]. This pathway takes place cotranslationally. SRP acts as a molecular chaperone for highly hydrophobic signal sequences of nascent polypeptides [42] and is likely to keep the polypeptide in a transport-competent state. SRP has been recently found to be composed of the Ffh protein and a 4.5S RNA, which appeared to be homologues of the eukaryotic SRP subunits (Ffh stands for fifty-four homologue) [43, 44]. SRP is recognized by the membrane-attached SRP receptor FtsY [45], which was also discovered by homology to a subunit of the eukaryotic SRP receptor [43, 44]. After delivery of the polypeptide at the plasma membrane by the SRP target system, the nascent polypeptide further follows the Sec translocation route by inserting in the SecYEG translocon [26, 46]. Thus the SRP and Sec pathways share the same translocon. However, some proteins require the SRP membrane protein targetting system but not the translocase [28]. It is possible that alternative translocase complexes exist.

3. The so-called 'twin arginine translocation' (TAT) pathway, which translocates proteins with a signal peptide containing a typical Arg-Arg motif via a Δ pH-dependent but Sec-translocon independent mechanism [47, 48]. Unusual is that this TAT machinery translocates proteins in a folded state [49], such as metalloenzymes and protein complexes.

4. There is also the possibility that certain membrane proteins insert spontaneously into the lipid bilayer, independent of any translocation machinery [50]. Most inner membrane proteins in *E. coli* appear to be targeted to the Sec machinery [40]. Until recently only the phage M13 procoat and Pf3 coat protein were thought to insert spontaneously without the need of SRP or the Sec machinery [28, 51]. However, very recently the YidC component was found to be required for the insertion of some proteins, such as the major coat protein of bacteriophage M13 [38]. Whether other proteins can insert into membranes without the aid of other proteins remains to be seen.

Topology

By statistical analysis there seems to be a correlation between the topology of transmembrane helices in membranes and the distribution of positive charges. Non-translocated hydrophilic segments, which stay inside, contain more positively charged amino acid residues than the translocated hydrophilic segments on the outside. This phenomenon is called the "positive-inside rule" and seems to apply for all kinds of membrane proteins [27]. Mutagenesis studies have confirmed this rule: By changing the distribution of positive charges, the orientation of the protein under investigation could be inverted [52, 53].

Study of a membrane protein by membrane model systems

Because of the complexity of biomembranes, model systems have to be used in order to be able to study the membrane protein and the protein-lipid interactions in an unambiguous way. This leads to the use of simplified model membrane systems of which the composition can be controlled and systematically varied.

Such a model system consists at least of the protein of interest and a membrane mimicking environment. As a membrane mimicking environment detergent micelles have been widely used [54]. These consist of single-chain amphiphiles, which form globular aggregates with the polar groups exposed to the aqueous phase and the hydrophobic parts in the micelle interior [55, 56]. However, a lipid environment is more appropriate to mimic the typical membrane structure. A lipid environment can be represented by a bilayer or one half of a bilayer (monolayer). Dehydrated lipid bilayers can be deposited on surfaces from a lipid suspension. Rehydration of dry lipid films produces large multilayered vesicles. These can be converted into single layered vesicles by sonification or extrusion techniques. Besides, extrusion techniques yield vesicles of defined sizes. Various protocols exist for mixing the protein and lipid components to obtain liposomes in which the protein is incorporated (e.g. [57, 58]). All these model systems are very useful to study various properties of protein-lipid interactions by different techniques.

The choice of which model system is used depends on the kind of experiment. In this study we use the following model systems for various experiments:

(1) Detergent micelles are useful to study aggregation properties of the membrane protein e.g. by size exclusion chromatography, since the micellar solution can serve as elution buffer.

(2) Monolayers are used in experiments to determine whether proteins have an affinity for lipids. In these type of experiments the changes in surface pressure of a monolayer of lipids, representing one half of the bilayer, can be measured. When a protein is able to penetrate the lipid monolayer, this will be measured as an increase in surface pressure [59].

(3) Large vesicles are useful for centrifugation experiments in which binding of protein to the vesicles can be determined. Vesicles of defined size can be used to study spontaneous insertion of the protein in the bilayer. By filling the interior of the vesicles with a fluorescent dye, leakage of the vesicle membrane upon protein incorporation can be monitored by measuring changes in fluorescence of the diluted dye [60].

(4) Small unilamellar vesicles are required for Circular Dichroism (CD) measurements, since these minimize scattering artifacts. CD yields secondary structure information. It measures the differential absorption of left- and right-handed circularly polarized light by the optically active protein molecule. This results in distinct spectral shapes due to the distinct electronic environments sensed by the amide groups in different secondary structures [61].

(5) Deposition of lipid-protein bilayers on crystals is used for Fourier Transform Infrared (FTIR) experiments or Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) experiments. By FTIR knowledge about the secondary structure can be obtained as well.

Besides the techniques mentioned here a large number of other techniques is used in the study of membrane proteins. Biophysical techniques, such as high-resolution nuclear

magnetic resonance (NMR) and solid state NMR, electron spin resonance (ESR), fluorescence spectroscopy provide useful information.

Thus, a variety of techniques can yield detailed information about these protein-lipid systems. Since the FTIR and ATR-FTIR techniques played a dominant role in our characterization of the g9p in model membrane systems the next section will provide a brief explanation of these techniques.

FTIR and ATR-FTIR as a tool to study protein-lipid systems

Background

The infrared region of the electromagnetic spectrum ranges from 200-5000 cm⁻¹ (50000-2000 nm). Infrared spectroscopy detects the absorbance of infrared radiation with a frequency that matches that of an oscillating permanent electric dipole moment, which accompanies an infrared active molecular vibration. In a simplified way a molecule can be seen as a set of atomic masses held together by springs that represent the bonds. On the basis of molecular mechanics, a simple harmonic oscillator approximation can be applied to a chemical bond between atoms A and B of mass m_A and m_B respectively. The vibration frequency is then given by:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}$$

where k is the Hooke's law force constant, and μ is the reduced mass of the system: $\mu = m_A \cdot m_B / (m_A + m_B)$. Thus, the frequency of this vibration is directly related to the atomic masses and molecular geometry, since k depends on characteristics like bond strength, length, and torsion angles. Since a bond can expand as well as bend, there are vibrational bending modes (designed as ν) as well as stretching modes (designed as δ). Since bending is more easy than stretching, bending modes are lower in energy and absorb at a lower frequency than stretching modes.

FTIR as a tool for study of protein-lipid systems

The sensitivity of infrared spectroscopy was greatly enhanced by the introduction of Fourier transform infrared spectroscopy [62]. This caused IR to be widely applied for molecular analysis. An advantage of infrared spectroscopy is that samples of almost any substance, in almost any morphological form, without the direct need for labeling, can be analyzed. This is facilitated by the availability of a wide range of sampling methods, e.g. transmittance spectroscopy or reflectance spectroscopy [62]. Therefore, it is a suitable technique to study membranes. Protein secondary structure as well as lipid characteristics can be measured simultaneously. Biomembranes are preferentially studied in thin films, formed by drying an aqueous suspension onto a suitable infrared substrate in either transmission or attenuated total reflectance (ATR) mode [63, 64].

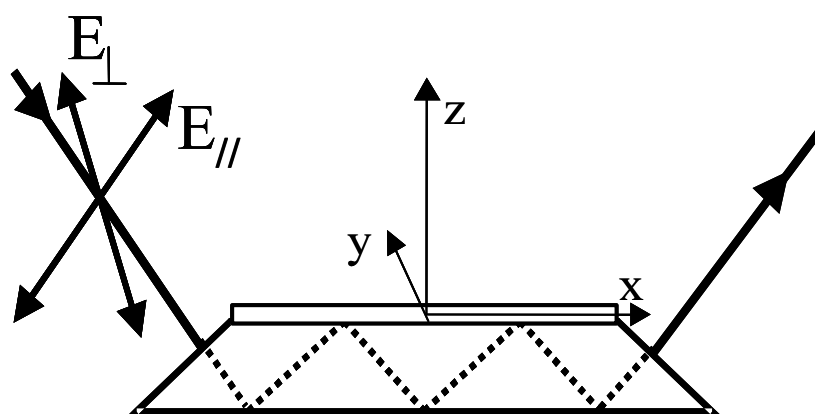


Figure 3: Schematic presentation of the experimental configuration of an ATR experiment. Incident radiation reflects within the internal reflection element and partially penetrates the sample area. The reflected beam therefore is attenuated by the absorption of the sample. When using plane-polarized light, this technique can also be applied to determine the molecular orientation.

In ATR-FTIR the sample is dried onto an infrared transparent crystal, and the infrared light is directed into the crystal under conditions where total internal reflectance is obtained. At the surface of the crystal, the reflected beam partially penetrates into the surrounding medium, and is attenuated by the absorption of the sample. The penetration depth of the infrared beam into the surrounding medium is typically a few micrometers. Thus, ATR is especially suited for the analysis of biomembranes, since it allows

penetration into the sample, without introducing too large a disturbance caused by the strong infrared water bands. Furthermore, using linearly polarized infrared light, ATR spectroscopy can be applied to obtain information on the relative orientation of specific components in the membrane film (Chapter 3).

Information that can be extracted

Complex molecular structures, such as proteins, are expected to exhibit a large number of infrared active vibrations. However, particular modes show considerable overlap and absorb at so called characteristic group frequencies. Thus, most absorption bands can be found at well-defined positions in the spectrum, resulting in a limited set of spectral bands.

Table 1: Band positions of molecular groups commonly found in biological samples.

Frequency (cm ⁻¹)	Assignment*	Main Source
± 3600-3000	ν (OH)	water
± 3000-2800	$\nu_{as,s}$ (CH _{2,3})	lipids
± 1740	ν (C=O)	lipids (esters)
± 1650	HOH scissoring	water
± 1650	ν (C=O)	protein (amide I)
± 1540	δ (N-H)	protein (amide II)
± 1450	δ (C-H)	lipids
± 1240	ν_{as} (PO ₂ ⁻)	phospholipids

* *Stretching vibrations are denoted by ν , bending modes by δ*

The subscripts (as, s) denote anti-symmetric and symmetric vibrations, respectively.

These band positions are not precise, but can shift with molecular structure and environmental parameters, such as solvent, pH, and temperature. The amide I (1710-1590 cm⁻¹) and amide II (1590-1480 cm⁻¹) vibrations of proteins are very useful for structural analysis. These bands arise from the amide backbone vibrations of the

polypeptide chain, and are sensitive to hydrogen bonding and hence the secondary structure of the protein. Hydrogen bonding causes a down-shift in frequency and a broadening of the stretching bands of both donor (e.g. carbonyl oxygen: amide I) and acceptor (e.g. amide nitrogen), because the donor can be stretched more easily towards the acceptor. Bending becomes more difficult upon hydrogen bonding (which is linear), so bending vibrations (e.g. amide II) are shifted to higher energy upon hydrogen bonding. Specific positions in these amide bands have been correlated with specific secondary structure elements [65, 66]. Lipids also exhibit discrete bands that can provide information about the lipid conformation.

A complicating factor in FTIR is the strong absorbance of water, since it obscures important bands in biological samples. This can be partly overcome by drying the sample, or by using D₂O instead of H₂O. Another complicating factor is that many bands show considerable overlap. Thus, assignment of spectral components to specific secondary structure elements requires careful decomposition of the amide I band [63, 67]. Component bands can be identified using second derivative analysis or deconvolution techniques [68]. The resulting component bands can then be used in curve-fitting analysis of the measured band shape to give fractional areas of these bands. This provides an estimate of the fractions of different secondary structures present in the protein.

BACTERIOPHAGE M13 AND THE GENE 9 PROTEIN

Bacteriophage M13 was found for the first time in 1963 in Munich in waste water [69]. It belongs to the family of filamentous phages (Inoviridae) that infect different gram-negative bacteria. Bacteriophage M13 belongs to the F-specific filamentous phages (Ff group), which only infect male *Escherichia coli* cells carrying F-pili. Closely related phages are the fd and fl phages.

Filamentous bacteriophages differ from most other bacterial viruses in that they do not kill their host. Rather, they are continuously extruded across the bacterial inner and outer membrane without causing cell lysis. The infected cells continue to grow and divide, but at a decreased rate [70].

Since bacteriophage M13 is a relatively simple molecular complex, it has been extensively characterized. The DNA and major coat proteins are relatively easy to isolate. The information gained about its structure and reproductive cycle and its ability to produce large numbers of progeny, providing an ample amount of biological material, is very important for biological questions, applied as well as basic. On one hand, the bacteriophage is a useful workhorse for DNA sequencing, site-directed mutagenesis, and foreign peptide display on the bacteriophage particle [71-73]. On the other hand, since the phage genome is so small and relatively easy to analyse genetically and biochemically, it has been used as an experimental system for elucidating many molecular mechanisms, which might be translated to other more complex biological systems. The study of protein-lipid interactions, protein translocation, structure-function relationships of the various phage-encoded proteins, as well as the mutual interactions in the formation of macromolecular assemblies, makes it an interesting field of research. Especially interesting is the question how viruses and bacteriophages assemble and disassemble, thereby using the properties of the plasma membrane of biological cells.

Structure of M13 bacteriophage

The virus particle is a flexible rod of about 900 nm length (the length is actually dependent on the length of the enclosed genome [74]) and about 6.5 nm in diameter [75]. The DNA of the phage M13 is a single stranded molecule consisting of 6408 nucleotides. It comprises 12% of the molecular mass of the phage particle [76]. The genome encodes 11 proteins within 9 open reading frames, of which five are structural coat proteins (g3p, g6p, g7p, g8p and g9p), three are required for phage DNA replication (g2p, g5p and g10p), and two morphogenetic proteins serve assembly and secretion functions (g1p, g4p and g11p) [77-79]. All these proteins are required for phage production. In addition, there is an intergenic region that does not code for proteins, but contains signals for the initiation of DNA synthesis, the initiation of capsid formation (packaging signal, see below), and the termination of RNA synthesis.

The DNA is extended along the axis of the particle and encapsulated in a 1.5-2.0 nm thick tube formed by about 2700 copies of the product of gene 8. Since this protein is the most abundant product of the genome it is referred to as the major coat protein. A few copies each of four different minor coat proteins are located at the ends of the phage

filament: The distal end, which is assembled first, contains gene 7 and gene 9 minor coat proteins (products of gene 7 and 9) [80, 81]. Labeling studies indicated the presence of three to four copies of each [82], whereas considerations based on symmetry propose five copies [83]. In the absence of either g7p or g9p, almost no phage particles are formed [84]. The proximal end, which enters the host cell first, contains approximately five copies of the gene 3 and gene 6 minor coat proteins (products of gene 3 and 6) [82]. The former two (g7p and g9p) are necessary for the assembly of a new phage particle, the latter two (g3p and g6p) are required for particle stability and phage infectivity. The single stranded circular DNA contains an imperfect 32-bp hairpin structure with a small bulge, responsible for encapsidation of the DNA into phage particles [85]: the packaging signal (PS). The orientation of the DNA within the particle is determined by the packaging signal, which is located at the distal end (with g7p and g9p) [84]. For reviews see also [77, 78, 86-88].

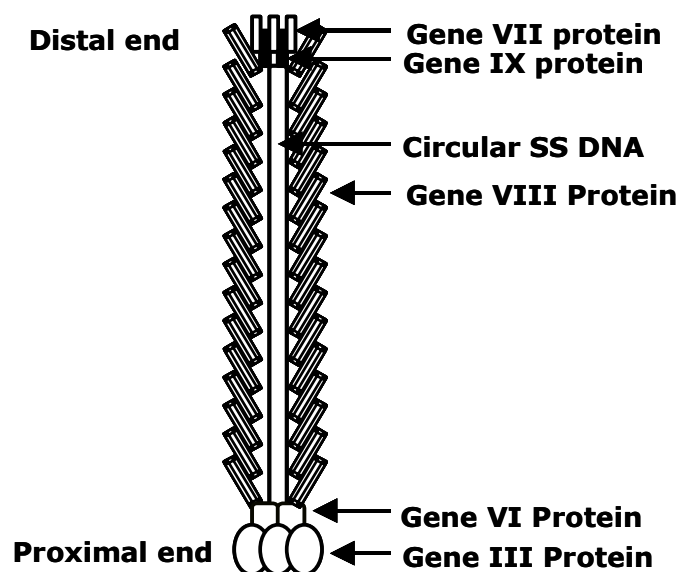


Figure 4: Schematic picture of bacteriophage M13.

The reproductive cycle

The reproductive life cycle of bacteriophage M13 starts with the adsorption of the phage particle at the F-pili of *E. coli*, specifically mediated by g3p [89, 90]. During the infectious entry of the phage the hydrophobic g6p, which functions as a kind of sealing agent to stabilize the phage particle [84], is lost resulting in a destabilized nucleoprotein particle. The major coat protein is stripped off this destabilized particle and dissolves in the inner membrane [91]. The DNA is released in the cytoplasm and new DNA and protein molecules are produced using the host cell machinery [92]. Within the cytoplasm newly synthesized DNA is protected by g5p. Newly synthesized coat proteins are stored in the inner membrane until they are used to form a new phage particle in the assembly process, thereby replacing the g5p [93]. These processes are described in a number of review articles [79, 87, 88].

The assembly process

An important feature of the Ff phage is that assembly and simultaneous secretion occur in or on the membrane. This aspect sets these phages apart from most bacterial viruses, which assemble in the cytoplasm. The assembly is initiated by the packaging signal (PS), which is exposed in the g5p-ssDNA complex [94]. The PS probably interacts with g1p [85], which triggers the assembly process. The proteins g1p and g11p are both associated with the inner membrane, and mixed multimers of them probably constitute the part of the extrusion channel that passes through the inner membrane [95, 96]. G7p and g9p might also directly or indirectly initiate assembly by interacting with the PS. This is supported by the finding that deletions in the PS could be compensated for by mutations in g7p, g9p, or g1p [85]. The fact that g7p and g9p are located at the end of the phage particle which emerges first [97] also suggests an important role early in assembly. Activation of g1p triggers the formation and opening of the extrusion channel in the outer membrane, which is formed by approximately 14 copies of g4p [98]. Furthermore, the association of thioredoxin of the host is required for phage assembly [99, 100]. During the assembly process g5p monomers in the g5p-ssDNA complex are replaced by g8p major coat protein molecules, which wrap around the DNA forming a protecting coat for the phage genome outside the cell. The assembly is completed by binding of g3p and g6p to the end of the DNA, upon which a new bacteriophage particle

is released into the medium. A detailed description of the assembly can be found in reviews [78, 87]. The assembly process is schematically illustrated in Figure 4.

Detailed knowledge of phage assembly is rudimentary. The structures of phage proteins can be addressed prior to assembly in the membrane-bound state, and after integration in the intact phage particle. An understanding of the intermediate process of assembly requires knowledge of both. Together with information about interactions from biochemical, molecular biological, and genetic analysis the dynamic process of assembly is partly being revealed.

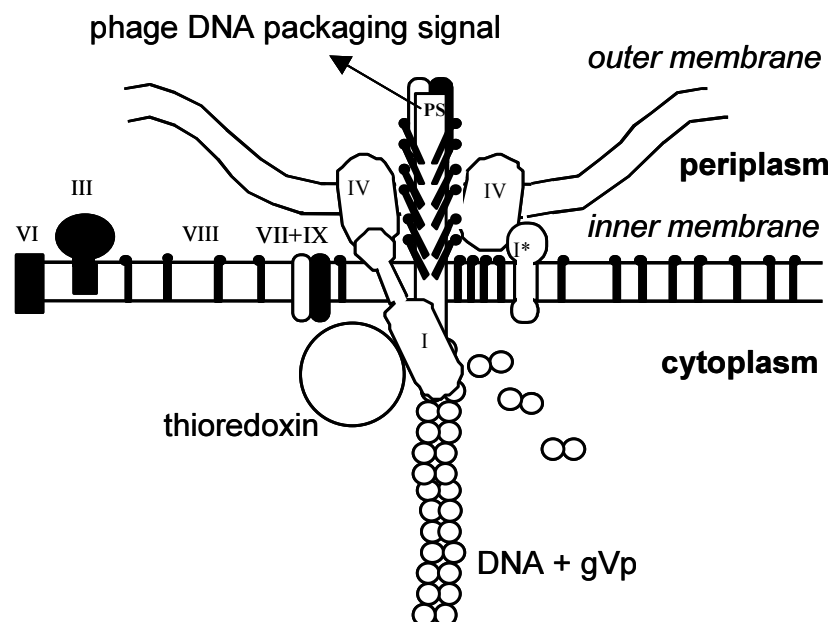


Figure 5: Assembly of bacteriophage is a membrane-bound process, involving viral as well as host cell encoded proteins. Gene 9 and gene 7 proteins might play an important role in initiation of assembly by interaction with the packaging signal of the phage genome.

Features of the major coat protein

Prior to assembly, the coat proteins are stored in the inner membrane. The 50-amino-acid residue g8p has been subjected to many studies in the membrane-bound form, since it is relatively easy to isolate the protein and obtain it in ample amounts. It inserts into

the membrane as a preprotein with a 23-amino-acid leader sequence, which is then clipped off by the bacterial leader peptidase [77]. G8p appears to consist of several domains in the membrane bound state: a hydrophilic C-terminal domain with positively charged residues inside in the cytoplasm, a hydrophobic transmembrane α -helix spanning the membrane, followed by a hinge and an amphipathic helix more in line with the plane of the membrane [101-104] and a short disordered acidic N-terminal region in the periplasm. Upon incorporation in bacteriophage, it changes to an aligned configuration with a single long helix [75, 105, 106]. In bacteriophage the major coat proteins overlap each other in a tilted helical array, with the flexible amphipathic N-termini located outside and the basic C-termini interacting with the DNA at the inside of the coat. The hydrophobic domains interlock the proteins with their neighbouring coat proteins [107].

Gene 7 and gene 9 minor coat proteins

In contrast to g8p, what is known about the even smaller g7p and g9p (33 and 32 amino acids, respectively) is mainly based on assumptions [84, 108]. These two proteins were proposed to be inserted into the membrane directly after or during synthesis, based on the finding that they retain their N-terminal formyl group [82]. This also suggests that they are incorporated in the phage as primary translational products, without being proteolytically processed (in contrast to g8p). Endemann and Model showed by sub-cellular location experiments that g7p and g9p are fractionated with the *E. coli* inner membrane [109]. The orientations of these proteins in the membrane are not clear, but based on speculations. G9p is likely to be anchored in the membrane by its hydrophobic N-terminal half. The presence of two positively charged residues in the hydrophilic carboxy-terminal half suggests that this part of the protein is cytoplasmic. The carboxy-terminal portion of g7p is highly hydrophobic, suggesting that this side is anchored in the membrane [87]. It remains a question whether these two proteins associate with each other in the membrane prior to their incorporation in phage. Endeman and Model detected no complex of g7p and g9p [109].

As mentioned above g7p and g9p are thought to play a role early in assembly, based on the finding that deletions in the packaging signal could be compensated for by mutations

in g7p and g9p [85]. This is supported by the finding that they are located on one end of the phage [81], which appeared to be the end that emerges first [80].

Analysis of the sequences of g7p and g9p indicates that they have hydrophobic regions of length similar to those in g8p and positive charges. However, the position of g7p and g9p at the distal end of bacteriophage is controversial. Makowski [83] constructed a model for the distal end, assuming that these similarities with g8p are indicative of structural homologies. He calculated that when g7p and g9p are extended α -helices, this results in an energetically unstable model for the virus end. Instead, L-shaped proteins consisting of two α -helices were proposed. The g9p was assumed to be closest to the DNA, interacting with the C-terminal side. Makowski also assumed the presence of 5 copies of each protein based on symmetry arguments. This is in contrast with the finding of Endeman and Model [109], who showed by using sera against g7p and g9p that g9p is accessible in intact phage and g7p is much less accessible.

The gene 9 minor coat protein

The primary structure of the 32-residue g9p (molecular weight is 3681 Da) has been deduced from the nucleotide sequence of the M13 DNA genome [110]. The primary sequence, including the N-terminal formyl group is given in figure 6.

G9p has no charged residues at the N-terminus, whereas the C-terminus contains mainly charged and polar residues. Hydropathy analysis indicates the presence of a hydrophobic stretch at the N-terminal side. The structure in the membrane is not known, but based on assumptions [87]. For the determination of the structure in the phage particle, the abundance of g9p is too low compared to the many major coat proteins present. G9p is a multifunctional protein: it should associate with the host membrane by protein-lipid interactions, it plays a role in the assembly process, possibly by specifically interacting with the phage DNA packaging signal and other proteins present in the assembly site, and finally it becomes part of one end of the phage particle involving also protein-protein interactions with neighbouring coat proteins. It is an intriguing question how such a small protein can fulfill all these functions.

1 5 10 15 20 25 30

MSVLVYSFASFVLGWCLRSGITYFTRLMETSS

HOC-Met-Ser-Val-Leu-Val-Tyr-Ser-Phe-Ala-Ser-Phe-Val-Leu-Gly-Trp-Cys-Leu-Arg-Ser-Gly-Ile-Thr-Tyr-Phe-Thr-Arg-Leu-Met-Glu-Thr-Ser-Ser-COOH

Figure 6: Primary sequence of gene 9 protein.

AIM OF THE STUDY AND OUTLINE OF THIS THESIS

Until now, an assembly model, requiring detailed knowledge of both cellular and ultimate location of phage coat proteins, was based on assumptions regarding the minor coat proteins [83, 84, 108]. This study wants to contribute information on the gene 9 minor coat protein before assembly. Therefore, the aim of this study is to characterize the g9p in the membrane-bound state, prior to assembly. For this purpose we used chemically synthesized protein, which we incorporated in several membrane-mimicking systems.

Chapter 2 describes the incorporation of g9p in several membrane-mimicking systems starting from organic solvents and detergent micelles and proceeding to phospholipid vesicles. The conformational and aggregational properties were characterized by combining secondary structure determination by CD and FTIR on one hand, and determination of aggregation size by size exclusion chromatography on the other hand.

In chapter 3 the conformation of g9p in membranes is described in more detail by FTIR and subsequent spectral band analysis. This results in a quantification of the secondary structure elements of the protein. In addition, the orientation of the protein with respect to the membrane plane was addressed by ATR-FTIR, resulting in a better understanding of the membrane-bound state of the protein.

In chapter 4 the question of how g9p gets into the membrane is addressed. The interaction of externally added g9p with lipid membranes is characterized by sucrose density centrifugation, insertion experiments in lipid monolayers, and leakage

experiments of bilayers. These experiments suggest that g9p is able to penetrate into membranes. Finally cleavage experiments show which side of the protein inserts in the membrane, and which side remains accessible for protein digestion.

Chapter 5 summarizes these results and discusses the biological implications for the *in vivo* situation in the *E. coli* inner membrane.

References

- 1 Gorter, E. and Grendel, F. (1925) J. Exp. Med. 41, 439-443.
- 2 Singer, S.J. and Nicholson, G.L. (1972) Science 175, 720-731.
- 3 Jacobson, K., Sheets, E.D. and Simson, R. (1995) Science 268, 1441-1442.
- 4 de Kruijff, B. (1987) Nature 329, 587-588.
- 5 Wiener, M.C. and White, S.H. (1992) Biophys. J. 61, 428-433.
- 6 White, S.H. and Wimley, W.C. (1998) Biochim. Biophys. Acta 1376, 339-352.
- 7 Raetz, C.R. (1978) Microbiol. Rev. 42, 614-659.
- 8 Burnell, E., van Alphen, L., Verkleij, A. and de Kruijff, B. (1980) Biochim. Biophys. Acta 597, 492-501.
- 9 Schulz, G.E. (1993) Curr. Opin. Cell Biol. 5, 701-707.
- 10 Killian, J.A. and von Heijne, G. (2000) Trends Biochem. Sci. 25, 429-434.
- 11 Sakai, H. and Tsukihara, T. (1998) J. Biochem. 124, 1051-1059.
- 12 Wallin, E., Tsukihara, T., Yoshikawa, S., von Heijne, G. and Elofsson, A. (1997) Protein Sci. 6, 808-815.
- 13 Arkin, I.T. and Brunger, A.T. (1998) Biochim. Biophys. Acta 1429, 113-128.
- 14 Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
- 15 Landolt-Marticorena, C., Williams, K.A., Deber, C.M. and Reithmeier, R.A.F. (1993) J. Mol. Biol. 229, 602-608.
- 16 Killian, J.A. (1998) Biochim. Biophys. Acta 1376, 401-416.
- 17 Spruijt, R.B., Wolfs, C.J.A.M., Verver, J.W.G. and Hemminga, M.A. (1996) Biochemistry 35, 10383-10391.
- 18 von Heijne, G. (1985) J. Mol. Biol. 184, 99-105.
- 19 Dalbey, R.E., Kuhn, A. and Wickner, W. (1987) J. Biol. Chem. 262, 13241-13245.
- 20 Dalbey, R.E. and Wickner, W. (1987) Science 235, 783-787.
- 21 Kuhn, A., Rohrer, J. and Gallusser, A. (1990) J. Struct. Biol. 104, 38-43.
- 22 Dalbey, R.E. (1990) Trends Biochem. Sci. 15, 253-257.
- 23 Kuhn, A., Kreil, G. and Wickner, W. (1987) EMBO J. 6, 501-505.
- 24 Engelman, D.M. and Steitz, T.A. (1981) Cell 23, 411-422.
- 25 Dalbey, R.E. and Kuhn, A. (2000) Ann. Rev. Cell Dev. Biol. 16, 51-87.
- 26 Scotti, P.A., Valent, Q.A., Manting, E.H., Urbanus, M.L., Driessen, A.J., Oudega, B. and Luijck, J. (1999) J. Biol. Chem. 274, 29883-29888.
- 27 Von Heijne, G. (1986) EMBO J. 5, 3021-3027.

- 28 Cristobal, S., Scotti, P., Luirink, J., von Heijne, G. and de Gier, J.W. (1999) *J. Biol. Chem.* 274, 20068-20070.
- 29 Hartl, F.U., Lecker, S., Schiebel, E., Hendrick, J.P. and Wickner, W. (1990) *Cell* 63, 269-279.
- 30 Duong, F. and Wickner, W. (1997) *EMBO J.* 16, 4871-4879.
- 31 Murphy, C.K., Stewart, E.J. and Beckwith, J. (1995) *Gene* 155, 1-7.
- 32 Oliver, D.B. and Beckwith, J. (1982) *Cell* 30, 311-319.
- 33 Economou, A. and Wickner, W. (1994) *Cell* 78, 835-843.
- 34 Nishiyama, K., Fukuda, A., Morita, K. and Tokuda, H. (1999) *EMBO J.* 18, 1049-1058.
- 35 Economou, A. (1998) *Mol. Microbiol.* 27, 511-518.
- 36 Randall, L.L. and Hardy, S.J.S. (1995) *Trends Biochem. Sci.* 20, 65-69.
- 37 Scotti, P.A., Urbanus, M.L., Brunner, J., de Gier, J.W., von Heijne, G., van der Does, C., Driessen, A.J., Oudega, B. and Luirink, J. (2000) *EMBO J.* 19, 542-549.
- 38 Samuelson, J.C., Chen, M.Y., Jiang, F.L., Moller, I., Wiedmann, M., Kuhn, A., Phillips, G.J. and Dalbey, R.E. (2000) *Nature* 406, 637-641.
- 39 Agarraberes, F.A. and Dice, J.F. (2001) *Biochim. Biophys. Acta* 1513, 1-24.
- 40 de Gier, J.W., Valent, Q.A., Von Heijne, G. and Luirink, J. (1997) *FEBS Lett.* 408, 1-4.
- 41 Luirink, J. and Dobberstein, B. (1994) *Mol. Microbiol.* 11, 9-13.
- 42 Luirink, J., High, S., Wood, H., Giner, A., Tollervey, D. and Dobberstein, B. (1992) *Nature* 359, 741-743.
- 43 Bernstein, H.D., Poritz, M.A., Strub, K., Hoben, P.J., Brenner, S. and Walter, P. (1989) *Nature* 340, 482-486.
- 44 Romisch, K., Webb, J., Herz, J., Prehn, S., Frank, R., Vingron, M. and Dobberstein, B. (1989) *Nature* 340, 478-482.
- 45 Fekkes, P. and Driessen, A.J. (1999) *Microbiol. Mol. Biol. Rev.* 63, 161-173.
- 46 Valent, Q.A., Scotti, P.A., High, S., de Gier, J.W., von Heijne, G., Lentzen, G., Wintermeyer, W., Oudega, B. and Luirink, J. (1998) *EMBO J.* 17, 2504-2512.
- 47 Creighton, A.M., Hulford, A., Mant, A., Robinson, D. and Robinson, C. (1995) *J. Biol. Chem.* 270, 1663-1669.
- 48 Chaddock, A.M., Mant, A., Karnauchov, I., Brink, S., Herrmann, R.G., Klosgen, R.B. and Robinson, C. (1995) *EMBO J.* 14, 2715-2722.
- 49 Berks, B.C., Sargent, F. and Palmer, T. (2000) *Mol. Microbiol.* 35, 260-274.
- 50 Von Heijne, G. (1984) *EMBO J.* 3, 2315-2318.
- 51 Kiefer, D. and Kuhn, A. (1999) *EMBO J.* 18, 6299-6306.
- 52 von Heijne, G. (1989) *Nature* 341, 456-458.
- 53 Nilsson, I. and von Heijne, G. (1990) *Cell* 62, 1135-1141.
- 54 Tanford, C. and Reynolds, J.A. (1976) *Biochim. Biophys. Acta* 457, 133-170.
- 55 Helenius, A. and Simons, K. (1975) *Biochim. Biophys. Acta* 415, 29-79.
- 56 Helenius, A., McCaslin, D.R., Fries, E. and Tanford, C. (1979) *Methods Enzymol.* 56, 734-749.
- 57 Killian, J.A., Trouard, T.P., Greathouse, D.V., Chupin, V. and Lindblom, G. (1994) *FEBS Lett.* 348, 161-165.
- 58 Spruijt, R.B., Wolfs, C.J.A.M. and Hemminga, M.A. (1989) *Biochemistry* 28, 9158-9165.
- 59 Demel, R. (1974) *Methods Enzymol.* 32, 539-544.

- 60 Weinstein, J.N., Blumenthal, R. and Klausner, R.D. (1986) *Methods Enzymol.* 128, 657-668.
- 61 Fasman, G.D. (1996) in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G.D., ed.), pp. 381-412, Plenum Press, New York.
- 62 Perkins, W.D. (1987) *J. Chem. Ed.* 64, A296-A305.
- 63 Goormaghtigh, E., Raussens, V. and Ruyschaert, J.-M. (1999) *Biochim. Biophys. Acta* 1422, 105-185.
- 64 Fringeli, U. and Gunthard, H. (1981) in *Membrane Spectroscopy* (Verlag, S., ed.), Vol. 31, pp. 270-332, Grell, E, Berlin and Heidelberg.
- 65 Bandekar, J. (1992) *Biochim. Biophys. Acta* 1120, 123-143.
- 66 Arrondo, J.L. and Goni, F.M. (1999) *Prog. Biophys. Mol. Biol.* 72, 367-405.
- 67 Surewicz, W.K. and Mantsch, H.H. (1988) *Biochim. Biophys. Acta* 952, 115-130.
- 68 Surewicz, W.K., Mantsch, H.H. and Chapman, D. (1993) *Biochemistry* 32, 389-394.
- 69 Hofschneider, P.H. (1963) *Z. Naturforsch. B* 18.
- 70 Marvin, D.A. and Hohn, B. (1969) *Bacteriol. Rev.* 33, 172-209.
- 71 Makowski, L. (1994) *Curr. Opinion Struct. Biol.* 4, 225-230.
- 72 Kuhn, A., Wickner, W. and Kreil, G. (1986) *Nature* 322, 335-339.
- 73 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A laboratory manual*, Cold Spring Harbor Laboratory, New York.
- 74 Marvin, D.A. (1990) *Int. J. Biol. Macromol.* 12, 125-138.
- 75 Glucksman, M.J., Bhattacharjee, S. and Makowski, L. (1992) *J. Mol. Biol.* 226, 455-470.
- 76 Berkowitz, S. and Day, L. (1976) *J. Mol. Biol.* 102, 531-547.
- 77 Model, P. and Russel, M. (1988) in *The Bacteriophages* (Calender, R., ed.), Vol. 2, pp. 375-456, Plenum Press, New York.
- 78 Makowski, L. and Russel, M. (1997) in *Structural Biology of Viruses* (Chui, W., Burnett, M. and Garcea, R.L., ed.), pp. 352-380, Oxford University Press, Inc., New York.
- 79 Rasched, I. and Oberer, E. (1986) *Microbiol. Rev.* 50, 401-427.
- 80 Lopez, J. and Webster, R.E. (1983) *Virology* 127, 177-193.
- 81 Grant, R.A., Lin, T.-C., Webster, R.E. and Konigsberg, W. (1981) *Prog. Clin. Biol. Res.* 64, 413-428.
- 82 Simons, G.F.M., Beintema, J., Duisterwinkel, F.J., Konings, R.N.H. and Schoenmakers, J.G.G. (1981) *Prog. Clin. Biol. Res.* 64, 401-411.
- 83 Makowski, L. (1992) *J. Mol. Biol.* 228, 885-892.
- 84 Webster, R.E. and Lopez, J. (1985) in *Virus structure and assembly* (Casjens, S., ed.), pp. 235-267, Jones and Bartlett Publishers Inc., Boston MA.
- 85 Russel, M. and Model, P. (1989) *J. Virol.* 63, 3284-3295.
- 86 Rasched, I. and Oberer, E. (1986) *Microbiol. Rev.* 50, 401-427.
- 87 Webster, R.E. (1996) in *Phage Display of peptides and proteins* (Kay, B., Winter, J. and McCafferty, J., eds.), pp. 1-20, Academic Press, San Diego.
- 88 Marvin, D. (1998) *Curr. opinion Struct. Biol.* 8, 150-158.
- 89 Riechmann, L. and Holliger, P. (1997) *Cell* 90, 351-360.
- 90 Holliger, P. and Riechmann, L. (1997) *Structure* 5, 265-275.
- 91 Ohkawa, I. and Webster, R.E. (1981) *J. Biol. Chem.* 256, 9951-9958.

-
- 92 Pratt, D., Tzagoloff, H. and Beaudoin, J. (1969) *Virology* 39, 42-53.
- 93 Mandel, G. and Wickner, W. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 236-240.
- 94 Bauer, M. and Smith, G.P. (1988) *Virology* 167, 166-175.
- 95 Guy-Caffey, J.K., Rapoza, M.P., Jolley, K.A. and Webster, R.E. (1992) *Journal of Bacteriology* 174, 2460-2465.
- 96 Lopez, J. and Webster, R.E. (1982) *J. Virol.* 42, 1099-1107.
- 97 Lopez, J. and Webster, R.E. (1985) *Journal of Bacteriology* 163, 1270-1274.
- 98 Linderoth, N.A., Simon, M.N. and Russel, M. (1997) *Science* 278, 1635-1638.
- 99 Russel, M. and Model, P. (1986) *J. Biol. Chem.* 261, 14997-15005.
- 100 Russel, M. and Model, P. (1985) *Proc. Natl. Acad. Sci. U. S. A.* 82, 29-33.
- 101 Spruijt, R.B., Meijer, A.B., Wolfs, C.J.A.M. and Hemminga, M.A. (2000) *Biochim. Biophys. Acta* 1508, 311-323.
- 102 Meijer, A.B., Spruijt, R.B., Wolfs, C.J.A.M. and Hemminga, M.A. (2000) *Biochemistry* 39, 6157-6163.
- 103 Meijer, A.B., Spruijt, R.B., Wolfs, C. and Hemminga, M.A. (2001) *Biochemistry* 40, 8815-8820.
- 104 Meijer, A.B., Spruijt, R.B., Wolfs, C. and Hemminga, M.A. (2001) *Biochemistry* 40, 5081-5086.
- 105 Opella, S.J., Cross, T.A., DiVerdi, J.A. and Sturm, C.F. (1980) *Biophys. J.* 32, 531-548.
- 106 Opella, S.J., Stewart, P.L. and Valentine, K.G. (1987) *Q. Rev. Biophys.* 19, 7-49.
- 107 Marvin, D.A., Hale, R.D., Nave, C. and Citterich, M.H. (1994) *J. Mol. Biol.* 235, 260-286.
- 108 Russel, M. (1991) *Mol. Microbiol.* 5, 1607-1613.
- 109 Endemann, H. and Model, P. (1995) *J. Mol. Biol.* 250, 496-506.
- 110 Van Wezenbeek, P., Hulsebos, T. and Schoenmakers, J. (1980) *Gene* 11, 129-148.

Chapter 2

Conformational and aggregational properties of the gene 9 minor coat protein of bacteriophage M13 in membrane-mimicking systems

M. Chantal Houbiers, Ruud B. Spruijt, Cor J.A.M. Wolfs, Marcus A. Hemminga

Published in: Biochemistry 38 (1999), 1128-1135.

Abstract

The membrane-bound state of the gene 9 minor coat protein of bacteriophage M13 was studied in various membrane-mimicking systems, including organic solvents, detergent micelles, and phospholipid bilayers. For this purpose we determined the conformational and aggregational properties of the chemically synthesized protein by CD, FTIR, and HPSEC. The protein appears to be in a monomeric or small oligomeric α -helical state in TFE, but adopts a mixture of α -helical and random structure after subsequent incorporation into SDS or DOPG. When solubilized by sodium cholate, however, the protein undergoes a transition in time into large aggregates, which contain mainly β -sheet conformation. The rate of this β -polymerization process was decreased at lower temperature and higher concentrations of sodium cholate. This aggregation was reversed only upon addition of high concentrations of the strong detergent SDS. By reconstituting the cholate-solubilized protein into DOPG, it was found that the state of the protein, whether initially α -helical monomeric/oligomeric or β -sheet aggregate, did not change. Based on our results we propose that the principal conformational state of membrane-bound gene 9 protein *in vivo* is α -helical.

Introduction

M13 is a filamentous bacteriophage which consists of a circular single stranded DNA molecule surrounded by a cylinder of coat proteins. The particles are 880 nm in length and 6 to 7 nm in diameter and infect strains of *E. coli* bearing F pili. The molecular mass of a particle is about 1.6×10^7 Da, of which 88% is protein and 12% is DNA [1]. There are about 2700 copies present of the major coat protein, the product of gene 8. The minor coat proteins, present in 3-5 copies each, are located at the ends: one end contains the gene 7 protein (3630 Da) and gene 9 protein (3681 Da). The other end contains the gene 3 and gene 6 proteins (for reviews see [2-6]). During the membrane-bound phage assembly process, the coat proteins as well as three phage encoded non-structural proteins are involved: the proteins encoded by genes 1, 4, and 11. Bacterial proteins are also involved, one of which is thioredoxin. During the process of assembly the phage DNA sheds the gene 5 proteins, and picks up the coat proteins while extruding through the cellular envelope membranes (for reviews see [2-8]). Whereas a lot of information has become available about the structure of the major coat protein in the phage particle [9-11] and in its membrane-bound state [12-19], the information about the minor coat proteins encoded by genes 7 and 9 is scarce and incomplete. The primary structures were deduced from the nucleotide sequence of the M13 DNA genome [20]. Grant [21] suggested that these proteins are located at one end of the phage particle. This end appears to be the protein part that leaves the cell first [22]. Based on the finding that the gene 7 and gene 9 proteins retain their amino-terminal formyl group [23] they were proposed to be inserted into the membrane during or directly after synthesis. This suggests that the gene 7 and gene 9 proteins are incorporated in phage as primary translational products. These proteins, therefore, behave differently from the gene 8 protein, which is synthesized as a pre-protein [24] and is finally processed to yield the membrane-bound coat protein [25]. It was shown by subcellular location experiments that both gene 7 and gene 9 proteins are fractionated with the *E. coli* inner membrane [26].

The finding that the gene 7 and gene 9 proteins are located at the end of the phage particle that emerges first [22] shows that they must play a role early in assembly. It was found that deletions in the packaging signal could be compensated by mutations in gene

7, gene 9, and gene 1 [27]. This suggests an important role for the products of these genes in initiating the assembly process by interacting with the packaging signal of the phage genome. Furthermore, absence of gene 7 or gene 9 proteins almost completely abolishes the production of phage [2].

A complete description of assembly requires a detailed knowledge of both cellular localization and ultimate location of the coat proteins in the phage, and functioning of all components during the assembly process. Until now, an assembly model was based on assumptions regarding the minor coat proteins [2, 7, 28].

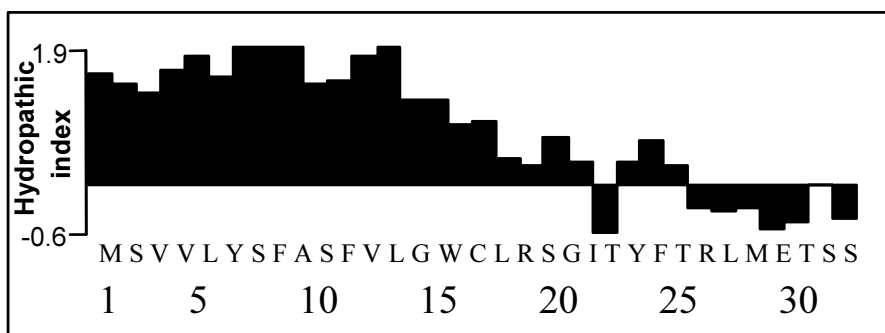


Figure 1. Hydropathy analysis of the gene 9 amino acid sequence according to the hydropathy scale of Kyte and Doolittle [55], using a sliding window of 9. The N-terminal methionine is assumed to be formylated [23].

In this chapter we will focus on the properties of the gene 9 protein. The primary structure of the gene 9 protein [20] and its hydropathy pattern, shown in Figure 1, indicate the presence of a hydrophobic stretch, long enough to span the membrane in an α -helical way. The gene 9 protein has no charged residue at the N-terminus, whereas the C-terminus contains mainly charged and polar residues.

Our work, therefore, is aimed to characterize the membrane-bound state of the gene 9 protein. We studied the conformational and aggregational properties of the protein in several membrane mimicking systems, proceeding from organic solvents to detergent micelles and phospholipid bilayers.

Materials and methods

Materials.

Fmoc-amino acids and the solid support p-alkoxybenzyl alcohol resin (Wang-resin) were purchased from Bachem, Bubendorf (Switzerland). The compounds 1-hydroxy-benzotriazole, di-isopropyl-carbodiimide, and ethanedithiol (EDT) were obtained from Fluka (Buchs, Switzerland). Dichloromethane (DCM), dimethylformamide (DMF), isopropyl alcohol, piperidine, trifluoroacetic acid (TFA), and sodium dodecylsulfate (SDS) were obtained from Merck (Darmstadt, Germany), and 2,2,2-trifluoroethanol (TFE) was obtained from Acros (New Jersey, USA). 1,2-dioleoyl-sn-glycero-3-phosphoglycerol (DOPG) was obtained from Avanti Polar Lipids (Alabaster, AL). Sodium cholate was from Sigma (St. Louis, MO). The Superose 12 and the Superdex 75 columns were obtained from Pharmacia (Uppsala, Sweden), the P6-DG column from BioRad (Hercules, CA).

Chemical synthesis and purification of the protein.

The gene 9 protein was synthesized by Mr. P.J.H.M. Adams at the Peptide Department of the Laboratory for Organic Chemistry of the University of Nijmegen. The cysteine residues in the protein were protected as disulfides with S-tert-butyl groups. The protein was prepared on a Wang resin using a solid-phase technique with a semi-automatic peptide synthesizer (Labortec SP640) according to the Fmoc-protocol. Apart from the carrier, the steps in each cycle were performed as described previously [29]. The removal of α -amino protective groups and the efficiency of each new acylation was controlled with the Kaiser-test [30]; raisin grains were visually judged by observing them after the test against a white background. Even a slight coloration could be detected in this way, and the pertinent cycle was repeated, if coloration was observed. Each cycle, irrespective of the result of the Kaiser-test, was terminated by treatment with a capping mixture (Ac_2O /di-isopropylamine). After completion of the required number of acylations, each protein-resin adduct was suspended in the mixture TFA-EDT-water (38:1:1) for hours and filtered. The filtrate was added to the tenfold volume excess of absolute ether to precipitate the crude S-protected peptide trifluoroacetate. It

should be noted that disulfides are not attacked by EDT in acidic media [31]. The product was subjected to amino acid analysis, confirming that all amino acids were present. The protein was further checked by gel electrophoresis, HPLC and mass spectrometry.

Solubilization of protein in organic solvent.

The protein was first dissolved in a small volume of TFA (10-20 μ l per mg of protein), and dried under a stream of nitrogen. Next, the protein was washed with TFE, followed by evaporation of the TFE under a stream of nitrogen, to remove residual TFA. Finally the protein was dissolved in a known volume of TFE. The protein concentration was determined by the procedure of Peterson [32] with bovine serum albumin as a standard.

Solubilization of protein in detergent.

To solubilize the gene 9 protein in detergent (sodium cholate or SDS), the method described by Killian [33] was used. Briefly, the desired amount of gene 9 protein dissolved in TFE was added to an equal volume of water, containing the desired amount of detergent. The mixture was vortexed for a few seconds. Next, this mixture was diluted to yield a 16:1 ratio of water to TFE by volume. The samples were mixed by vortexing for 2 s and lyophilized. The sample was rehydrated in the desired volume.

Determination of the aggregation state of the protein.

The aggregational state of the gene 9 protein was determined using HPSEC on a Superose 12 column (1.0 x 30 cm) or a Superdex 75 column (1.0 x 30 cm). A flow of 0.4 or 0.5 ml/min was applied, eluting in 50 mM cholate, 10 mM sodium phosphate, pH 8.0 for all samples but the SDS sample. The latter was eluted with 25 mM SDS, 10 mM sodium phosphate, pH 8.0. On-line detection was done by measuring the absorption and fluorescence. The absorption was measured at 280 nm and emission at 340 nm, with an excitation wavelength of 280 nm. For molecular weight calibration of the columns, alcohol dehydrogenase (150000 Da), bovine serum albumin (67000 Da), myoglobin (18800 Da), A-protein (6500 Da), and bacitracin (1400 Da) were used. The aggre-

gational state of the gene 9 protein in TFE was checked by passage over a nanosep filter, with a molecular weight cut-off of 3 kDa, from Pall Filtron (Northborough, MA, USA). The filtrate and residue were tested for the presence of protein by tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis [34].

Fourier Transform Infrared Spectroscopy.

For the FTIR experiments the gene 9 protein stock solution in TFE was treated by an additional step to remove traces of TFA, since TFA contributes to the signal in the amide I region [35]. This was done by adding a 2 mM HCl solution to the TFE stock in a 1/1 (v/v) ratio, to yield a 1 mM HCl concentration in the protein solution. The solution was subsequently lyophilized, redissolved into TFE, and used for further sample preparation. Samples for FTIR typically contained 0.1-0.2 mg protein at a cholate/protein molar ratio of 60, or a lipid/protein ratio of 50. Dehydrated films were prepared on a CaF₂ window (13 mm diameter) in a cabin continuously purged with dry air. FTIR spectra were recorded on a Perkin-Elmer 1750 FTIR spectrometer at room temperature. The optical bench was purged with dry air (Balston, Maidstone Kent, UK) at a flow rate of 25 L min⁻¹. The acquisition parameters were: 4 cm⁻¹ resolution; 128 co-added interferograms, 3500-900 cm⁻¹ wave number range. For the sake of comparison of the samples with cholate, the spectral intensities were normalized with the cholate peak intensity at 1560 cm⁻¹.

Circular Dichroism.

CD spectra were recorded from 190 to 260 nm on a JASCO 715 spectrometer at 25 °C, using a 1-mm path length cell, 1-nm bandwidth, 0.1-nm resolution, 1-s response time. Spectra were corrected for the background with an equally prepared sample without the protein. Protein concentrations were within the range of 25 to 46 µM, and are mentioned in the Figure legends.

Reconstitution of protein into lipid vesicles.

Reconstitution of the protein into lipid bilayers was performed by two different methods: 1) using cholate and subsequent dialysis, 2) via cosolubilization in TFE/water. In both cases the desired amount of DOPG was taken from a chloroform solution. The chloroform was evaporated under a stream of nitrogen, and the remaining lipid film was kept overnight under vacuum to remove all traces of chloroform. In the first method, the lipid film was dissolved in 100 mM cholate, and mixed with cholate solubilized protein (prepared according to the method described above under "Solubilization of protein in detergent"). Reconstitution was carried out by removal of cholate by dialysis for 48 h (4 times 12 h) at room temperature, according to Spruijt et al. [36], using a dialysis membrane with a molecular weight cutoff of 12-14 kDa. The second method is equal to the method of solubilization of protein into detergent, as described above, except that instead of detergent, now DOPG is used. After rehydration the vesicles were sonicated on ice for 3 minutes with a Branson B15 sonifier to minimize scattering by the vesicles in the CD measurements. Titanium particles of the sonifier tip were removed by centrifugation.

Results

Characterization of gene 9 protein in TFE, detergents, and lipids.

The gene 9 protein could be dissolved in TFE and a mixture of TFE/water (1/1, v/v), but the protein was insoluble in water and strictly hydrophobic solvent, such as chloroform (as observed from visual turbidity, due to the presence of large precipitates). Completely dissolved gene 9 protein in TFE was used to be solubilized in detergents or phospholipids via cosolubilization in TFE/water (1/1, v/v), and subsequent dilution, lyophilization and rehydration [33]. In Figure 2, CD spectra are presented of gene 9 protein in TFE, sodium cholate, SDS, and DOPG. Spectra were recorded immediately after sample preparation via the cosolubilization method. The CD spectrum of protein in TFE/water (1/1, v/v) was equal to the spectrum in pure TFE (data not shown). The spectrum in TFE features two minima at 208 and 222 nm, indicating the presence of α -helical structure [37]. Also the spectra in SDS, cholate, and DOPG are characterized

by these two minima. However, the minima are not so well-pronounced, slightly merging into a single broad band. Furthermore, the intensity is much lower than in TFE. Additionally, the FTIR spectrum of gene 9 protein in DOPG was recorded. The amide I band, shown in the insert of Figure 2, contains a single band at 1656 cm^{-1} . A band around this position is generally found for α -helical structures [38], but also random structure gives rise to a band around this position. The decreased CD intensity in detergents and phospholipids suggests that the amount of α -helix is lowered in these systems, as compared to TFE. Probably part of the α -helical conformation, present in TFE, has converted into another structure. However, from the FTIR spectrum in DOPG it is clear that no β -sheet structure is present. Therefore, it is likely that the remaining structure is random structure, in agreement with the observed FTIR spectrum. Thus, the conformation of gene 9 protein in detergents or phospholipids is a mixture of α -helix and random coil. Assuming that in TFE the protein is 100% α -helical, we roughly estimated the amount of α -helix in these systems, resulting in $50\text{-}60\% \pm 10\%$ α -helix in SDS, cholate, and DOPG. The remaining part is supposed to be random coil. Such a large contribution of random coil is not surprising, since in such a small protein a relatively large part of the amino acids is close to one of the two ends, and terminal amino acid residues are generally less ordered.

The aggregational states of the protein in the samples, shown in figure 2, were checked by HPSEC or ultrafiltration. The protein dissolved in TFE was observed to pass a 3 kDa filter, leaving no residue on the filter, indicating the predominant presence of protein monomers. Gene 9 protein solubilized by SDS and analyzed using HPSEC in 25 mM SDS, showed a single peak, corresponding to monomeric protein (data not shown). Gene 9 protein solubilized by cholate and reconstituted into DOPG, subjected to HPSEC in 50 mM cholate, eluted as a single broadened peak at a position indicating an average size of trimers. Since estimation of the protein size by this method is not sufficiently accurate to discriminate between monomers or small oligomers, we will refer to this state as "monomeric/oligomeric state", bearing in mind that the protein may be present as monomer or small oligomers.

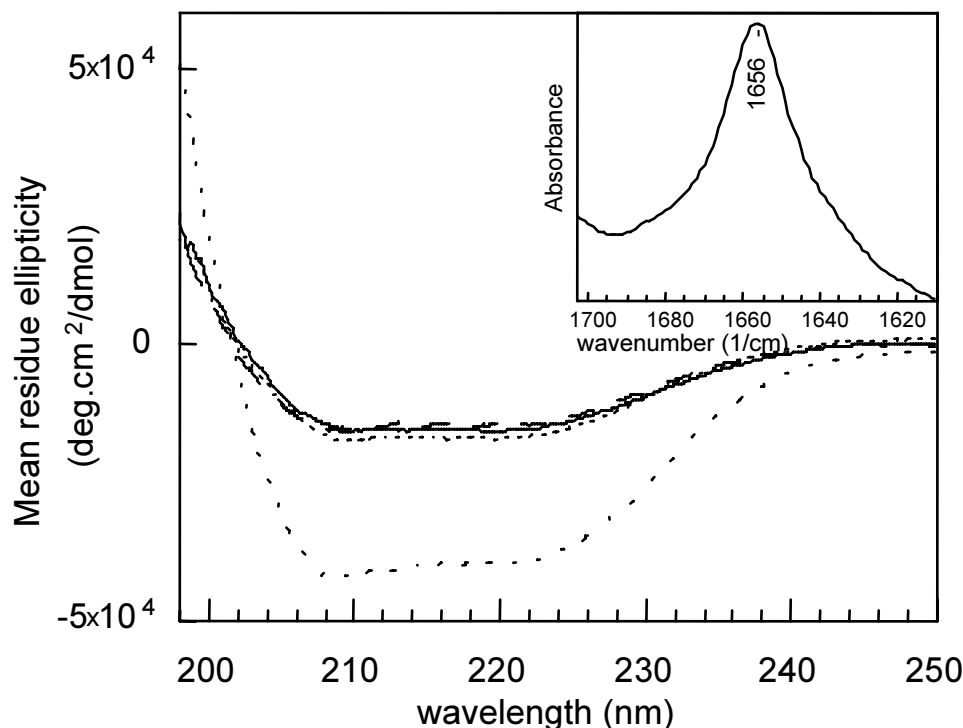


Figure 2. CD spectra of the gene 9 protein in TFE (- -), and incorporated via the cosolubilization method in TFE/H₂O into cholate (-----), SDS (— — —), and DOPG (solid line), directly after sample preparation. Protein concentrations were 46 μ M in pure TFE, 46 μ M in 50 mM SDS/10 mM phosphate/pH 8.0, 35 μ M in 26 mM cholate/10 mM phosphate/pH 8.0, 24 μ M in 12 mM DOPG/10 mM phosphate/pH 8.0. Insert: FTIR absorbance of dehydrated film of gene 9 protein incorporated into DOPG, at an L/P ratio of 50.

Change in conformational and aggregational state of the cholate-solubilized protein.

After 3 days, the conformational and aggregational state of the gene 9 protein in TFE, cholate, SDS, and DOPG was checked again by CD and HPSEC, to determine the stability of the protein. For the protein in TFE, SDS, and DOPG the results were similar (data not shown), indicating a stable state of the protein in these systems. It should be mentioned that the CD spectra of the protein in TFE or TFE/water (1/1, v/v) did not even change after weeks (data not shown). In contrast, the CD spectrum of the cholate-solubilized gene 9 protein was significantly changed after 3 days (Figure 3A). Instead of two minima at 208 and 222 nm, a single minimum around 218 nm was observed, while the zerocrossing and intensity hardly changed.

To analyse the protein conformation in more detail, also FTIR spectra of the cholate-solubilized gene 9 protein were recorded. In Figure 3B, the amide I bands of the corresponding samples are shown. Directly after preparation the most intense absorbance is at 1657 cm^{-1} , suggesting a predominant mixture of α -helical and random structure [38], in agreement with the CD result. There is also some intensity present around 1680 cm^{-1} , possibly originating from turn-like structures, or distorted α -helices [38]. The spectral changes after 3 days are shown in Figure 3B: the absorbance at 1657 cm^{-1} has diminished, in favour of a new peak at 1624 cm^{-1} . Also a new small shoulder at 1697 cm^{-1} is present. In the literature the two peaks around 1624 and 1697 cm^{-1} have been assigned to anti-parallel β -sheet structure [38-40].

The aggregational state of the protein in cholate was checked by HPSEC (data not shown). Whereas the protein was found to be monomeric/oligomeric directly after sample preparation, after 3 days over 90% of the protein eluted in the void volume, indicating the predominant presence of large aggregates. Thus, the conformational change, as observed by CD and FTIR, is concomitant with a state of increased aggregation of the gene 9 protein.

Reversibility of the aggregated state of the gene 9 protein.

Apparently, the protein can exist in two states: as monomers/oligomers containing α -helical and random structure, and as large aggregates containing β -sheet structure. By addition of SDS, we tested whether the aggregated state in cholate was reversible. The CD and FTIR spectra of aggregated protein in cholate, to which SDS has been added, are shown in Figure 3A and 3B, respectively. By comparison of the spectra before and after addition of SDS, it is clear that SDS completely restores the α -helical conformation as measured by CD, but that the FTIR absorbance at 1657 cm^{-1} is only partly restored. It should be noted that the final SDS/protein ratio in the FTIR experiment was much smaller as compared to CD (see legend Figure 3), explaining the partial disruption of the aggregates.

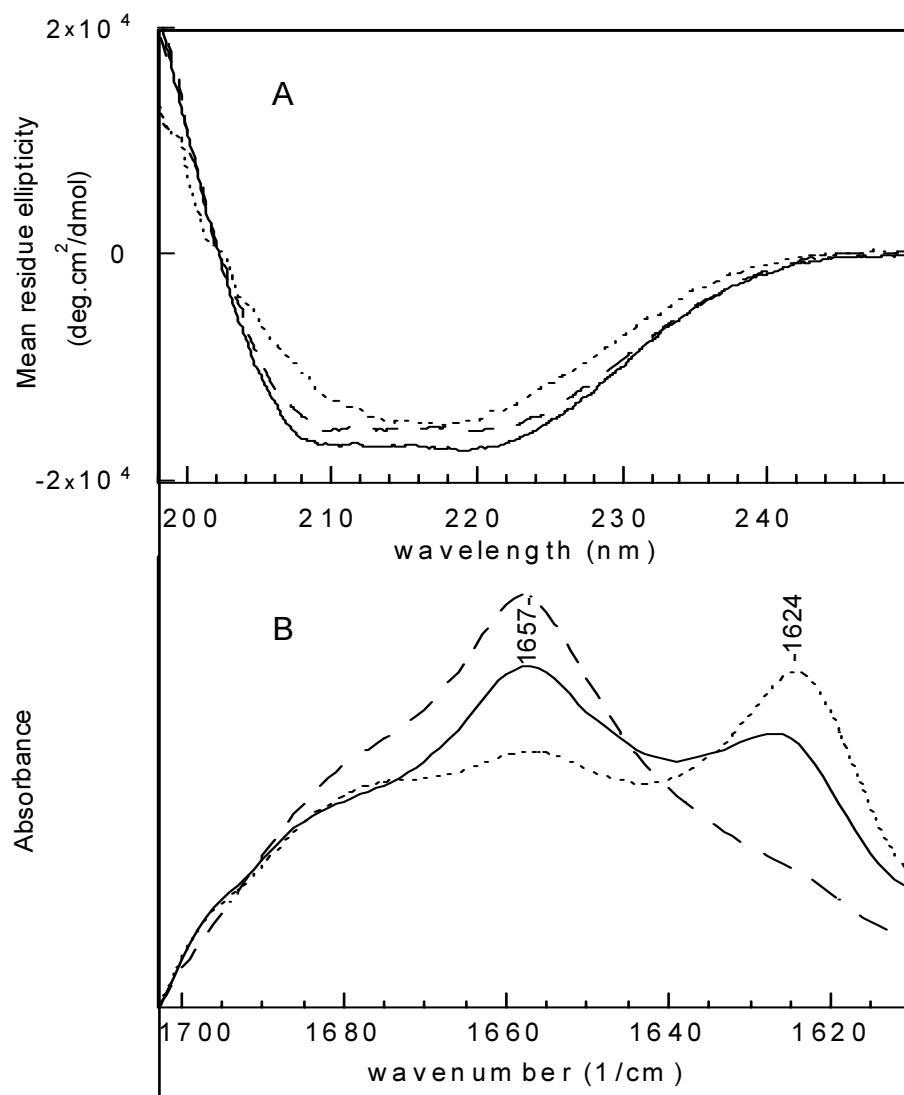


Figure 3. A: CD spectra of gene 9 protein in cholate at times zero (— — —) and 3 days after sample preparation (- - - -), and after addition of SDS to the sample at time 3 days (solid line). These samples were made from the (more concentrated) FTIR samples of Figure 3B (protein concentration 1.4 mM, cholate 83 mM), but diluted with a cholate solution just before the CD measurement, resulting in 35 μ M protein in 26 mM cholate. The FTIR sample with 50 mM SDS added was diluted with an SDS solution, resulting in 26 mM SDS. B: FTIR spectra of dehydrated film of gene 9 protein in cholate at times zero (— — —) and 3 days (- - - -) after sample preparation, and after addition of SDS to the sample at time 3 days (solid line). The cholate/protein ratio was 60, the SDS/protein ratio was 36. The spectra were not corrected for background of cholate, giving rise to a broad band underneath protein peaks.

Since the aggregation of the gene 9 protein is reversible, it should be possible to influence the aggregation state, as determined by HPSEC, to some extent by variation of

conditions. A typical HPSEC chromatogram is presented in Figure 4, showing the void volume peak at 15 min and a peak at 28 min, representing aggregated and monomeric/oligomeric protein, respectively. As is shown in Table 1, the aggregational state is influenced by several parameters.

Table 1. Relative percentages of aggregated gene 9 protein after incubation at several conditions, as measured by HPSEC

Sample ^a	Final concentration (mM) after addition of:	Temperature(°C) ^c	Incubation time	Relative area of void volume (%) ^d
1	-	RT	no	0
2	-	RT	5 hr	44
3	-	RT	1 day	88
4	-	RT	2 weeks	100
5	-	4	1day	20
6	-	4	2 weeks	94
7	-	-20	1 day	0
8	-	-20	2 weeks	0
9	-	75	20 sec	57
10 ^b	cholate, 250 mM	RT	no	79
11 ^b	cholate, 250 mM	RT	5 hr	38
12 ^b	cholate, 250 mM	RT	> 1 day	>38
13 ^b	SDS, 250 mM	RT	< 1 hr	3

a Sample contains 0.56 mM gene 9 protein and 50 mM cholate in sodium phosphate pH 8.0.

b Already aggregated protein, as pre-incubated at RT during 1 day (=sample 3). Incubation times mentioned are after addition in column 2 to pre-incubated sample.

c RT = room temperature.

d Values are $\pm 5\%$.

From lines 1 to 4 in Table 1 it is clear that the protein aggregates with time, and that it is a slow process (hours to days). Lines 5 to 9 show that the aggregation rate depends on the temperature. Lowering the temperature to 4 °C slows down the aggregation process, whereas storage at -20 °C completely stops aggregation. Increasing the temperature to

75 °C results in a very fast aggregation (seconds). Lines 10 and 11 show that an increase of the cholate concentration in an already aggregated sample can disrupt the aggregates slowly to some extent, but eventually aggregation will proceed (line 12). The strong detergent SDS can disrupt aggregates almost completely in a short time, as is clear from line 13.

The effect of lipids on the initial state of cholate-solubilized protein.

Cholate-solubilized protein in the monomeric/oligomeric α -helix containing state or in the aggregated β -sheet state, as characterized in Figure 3, was reconstituted into DOPG by cholate dialysis.

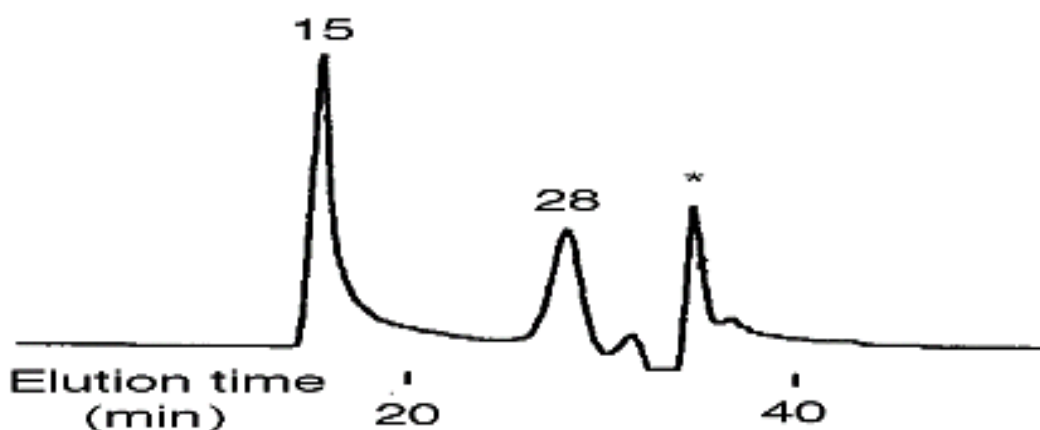


Figure 4. Typical HPSEC elution pattern of gene 9 protein, solubilized in cholate. The void volume at 15 min and the monomeric/small-oligomeric peak at 28 min are indicated. The peak indicated by an asterisk is due to background of the buffer. The column was Superose 12 (1.5 x 30 cm), at a flow of 0.5 ml/min, with fluorescence detection at 340 nm, exciting at 280 nm, the elution buffer was 50 mM cholate, 10 mM phosphate, pH 8.0.

The aggregational state of the protein after transfer to DOPG was determined using HPSEC. The initially monomeric/oligomeric protein retained this state upon reconstitution. Also when initially aggregated protein was reconstituted into DOPG, the aggregated state was retained (data not shown). In both cases the initial aggregational state of the protein appeared to be maintained as the elution patterns remained similar during several days. Apparently, the aggregational state of the protein is not affected by reconstitution, or upon subsequent aging. The corresponding CD spectra of gene 9 protein after transfer to DOPG are shown in Figure 5.

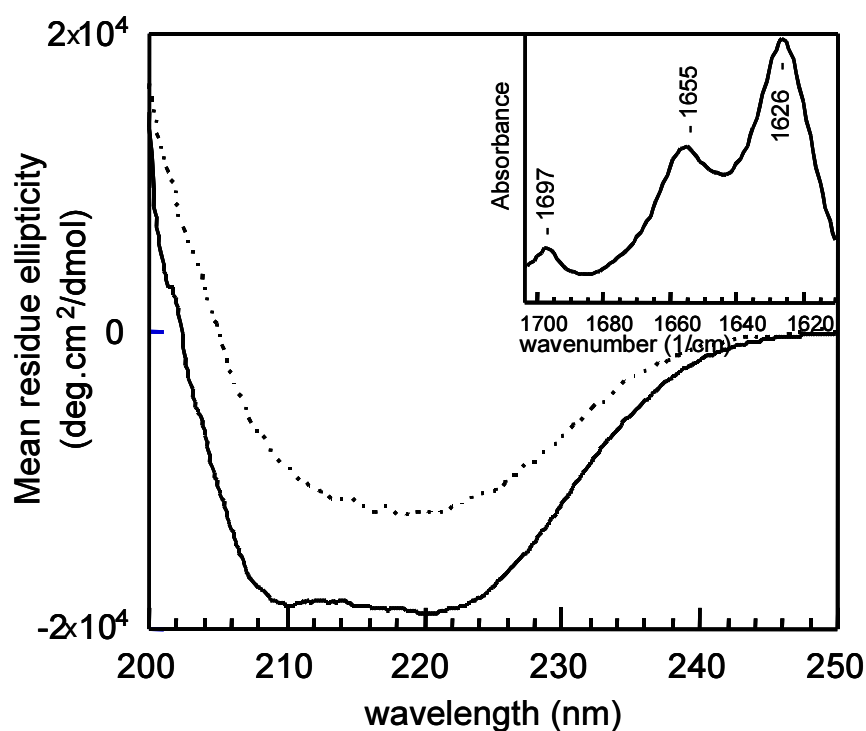


Figure 5. CD spectra of gene 9 protein containing vesicles, reconstituted via cholate dialysis. The solid line represents reconstitution started from monomeric/small-oligomeric protein in cholate; the dashed line represents reconstitution started from aggregated protein in cholate. The protein concentration was 30 μ M, the concentration DOPG was 1.7 mM, resulting in an L/P of 57. Insert: FTIR absorbance of dehydrated film of gene 9 protein containing DOPG vesicles, reconstituted from aggregated protein in cholate, at an L/P ratio of 50.

When α -helix containing monomeric/oligomeric protein in cholate was transferred to DOPG by the cholate dialysis protocol, a similar CD-spectrum was observed in DOPG. This CD spectrum was comparable to the spectrum in DOPG as shown in Figure 2, in

which the protein was directly transferred from TFE to DOPG, and identical to the CD spectrum of the α -helix containing protein in cholate prior to reconstitution (Figure 3A). In contrast, when initially aggregated protein was reconstituted into DOPG, the CD spectrum still showed a single minimum at 219 nm, but showed, as compared to the CD spectrum of the aggregated protein in cholate (Figure 3A), an additional red-shifted zero-crossing and a lower intensity.

The CD spectrum of the aggregated protein in DOPG appeared to be indicative for the presence of β -sheet structures [37]. The corresponding infrared absorbance by the amide I band of the same aggregated sample in DOPG is shown in the insert of Figure 5. Three peaks are visible: a major band at 1626 cm^{-1} , a peak at 1655 cm^{-1} , and a small peak at 1697 cm^{-1} . The 1626 and 1697 cm^{-1} peaks are in accordance with the presence of antiparallel β -sheet structure [38, 39]. The 1655 cm^{-1} band, however, corresponding to α -helix or random coil, indicates that also these secondary structures, as observed in the initial state of the aggregated protein in cholate (Figure 3B), were preserved.

Discussion

From the literature the gene 9 protein is known to be located in the bacterial inner membrane prior to assembly [26]. However, no detailed information is available about the nature of the interaction with the membrane, and about the conformation of the protein in the membrane-bound state prior to assembly. The hydropathy plot (Figure 1) suggests the possibility for the protein to span the membrane in an α -helical way, which requires the presence of a hydrophobic segment of about 20 amino acids. Since the protein has not been characterized before, in this work the conformational and aggregational properties of the protein in various model systems are investigated, moving towards reconstitution and determination of the state of the protein in model phospholipid membranes.

Characterization of gene 9 protein in TFE, detergents, and lipids.

After dissolving protein in TFE, it was transferred to cholate, SDS, or DOPG by cosolubilization in a TFE/water mixture, and subsequent dilution, lyophilization, and

rehydration as described by Killian et al. [33]. In all systems, including TFE, we found the protein to be in an α -helix containing monomeric or small-oligomeric state, with no large aggregates present.

For CD measurements of gene 9 protein in pure TFE two minima at 208 and 222 nm were observed, typical for an α -helical conformation [37, 41]. These two minima were also present in the other systems, but less distinct and less intense. In pure TFE and TFE/water (1/1, v/v), the negative CD bands are more intense and the minima are better defined, as compared to the gene 9 protein in cholate, SDS, and DOPG. Two effects could explain the intensity increase in TFE. Firstly, the helix in TFE may be elongated, compared to the helix in the other systems. In cholate, SDS, and DOPG, possibly not the full peptide, but only the hydrophobic stretch of about 20 residues (Figure 1) is immersed in a hydrophobic environment. The remaining charged part, likely to be exposed to the aqueous phase in micelles or vesicles, may contain other secondary structures, like random structure. This part may also become α -helical in TFE, since TFE induces native-like helices and segments that have an α -helix propensity [42-44], thus increasing the length of the α -helical strand. According to the literature, the intensity of the CD band for the helix is dependent on the average number of residues per strand [45]. Secondly, the dielectric constant of the medium could be different in TFE, affecting the transition probabilities [46], but this cannot account for the large intensity difference.

The protein in DOPG, measured by FTIR, resulted in a single amide I band at 1656 cm^{-1} , which may be another indication for α -helix structure [38, 47]. However, a band at 1656 cm^{-1} can also be interpreted as random structure [38, 48]. Combining this with the CD data, we propose that the conformation in DOPG is a mixture of α -helix and random structure. FTIR clearly indicates that no β -structure is present, so the decreased CD intensity is probably due to a conversion into random structure, which also agrees with the observed FTIR spectrum. From the observations we roughly estimated the percentage of α -helix to be $50\text{-}60\% \pm 10\%$, and the remaining part is supposed to be random coil. The protein in SDS shows a similar CD spectrum, also indicating the presence of this mixture of α -helical and random structure. The FTIR spectrum of the protein in cholate (Figure 3B) shows a low intensity band around 1680

cm^{-1} . Often peaks in this area are assigned to turns [38]. However, definitive assignment of turns and discrimination from unordered structures is difficult [49], since data on turns are scarce. In addition, several types of turns exist, giving rise to a broad distribution of frequencies in the 1695-1660 cm^{-1} region. Therefore, the gene 9 protein in cholate might contain turn-like structures or distorted helical parts, but the main part is a mixture of α -helical and random structure.

From the literature it is known that close proximity of aromatic amino acid residues can give rise to a significant contribution in the CD spectrum [50, 51]. This effect could make secondary structure estimations unreliable. In the case of the gene 9 protein in lipid systems, we do not expect the aromatic residues of the protein to be strongly immobilized or clustered. Moreover, we have investigated the near-UV region (250-300 nm) of the CD spectrum of the protein in lipid bilayers (data not shown), where we could find no detectable contribution of aromatic side chains. This observation also suggests that the contribution of the aromatic amino acid residues in the far-UV region (190-250 nm) is negligible.

The aggregational state of the protein, determined by a cutoff filter or HPSEC, in TFE or detergents was determined to be in between monomers and small oligomers. We also used HPSEC analysis in cholate elution buffer, to determine the aggregational state of the gene 9 protein in DOPG vesicles. When applying the sample to the cholate-eluted column, the conditions are actually changed, because the sample is diluted and mixed with cholate. This may result in a change of the aggregation state. However, with the protein concentrations used, it was found that the aggregational state hardly changed, since both elution patterns with almost all protein in the void volume peak, as well as patterns with all protein in the monomeric/oligomeric peak were observed. This suggests that during the time of elution no disruption of aggregates nor formation of aggregates occurs.

Stability of the protein in TFE, detergent micelles, and lipids.

In TFE and TFE/water (1/1, v/v) the protein can be stored for prolonged times, without any change in conformational or aggregational state. Also in SDS and DOPG, the CD

spectra and HPSEC elution profiles did not change during several days, indicating conformational and aggregational stability upon ageing. In cholate, however, a conformational change was observed with both CD and FTIR. The CD spectrum after 3 days showed a single minimum at 218 nm (instead of the two minima at 208 and 222 nm directly after preparation), although the zerocrossing and intensity hardly changed (Figure 3A). This is not a typical β -sheet spectrum as deduced from the CD spectra of poly-L-lysine in the β -sheet form [37], which has a red-shifted zerocrossing and a decreased intensity compared to the α -helical form. However, the appearance of a main band at 1624 cm^{-1} and a small accompanying shoulder at 1697 cm^{-1} in the FTIR spectrum (Figure 3B), clearly indicate the presence of antiparallel β -sheet structure [38-40].

Combining the FTIR and CD data, the altered CD spectrum of cholate-solubilized gene 9 protein is supposed to represent mainly antiparallel β -sheet structure, with some additional α -helix, random and turn structure. More recent literature shows that the CD spectrum of β -sheet structure may result in a considerable variation of spectra, all with a single minimum between 210 and 225 nm [41]. Spectra with a single minimum and zerocrossing around 200-203 nm have for example been observed for the all- β protein prealbumin, the α/β protein subtilisin, and the $\alpha+\beta$ protein ribonuclease [52]. (See for definition of these protein classes [53]). The spectral changes, indicating a conformational change, are concomitant with the presence of large aggregates, since all protein eluted in the void volume in the HPSEC analysis. Therefore, we conclude that in cholate the protein can exist as α -helix containing monomers/oligomers, or as mainly β -sheet-containing aggregates. The cholate-solubilized protein undergoes a transition towards the aggregated state in time. Apparently, cholate is not suitable to solubilize the gene 9 protein.

Reversibility of the aggregation.

Addition of the strong detergent SDS to the aggregated β -sheet protein resulted in restoration of the α -helix containing conformation, as shown by CD and FTIR, and in a concomitant disruption of the large aggregates, as shown by the disappearance of the void volume peak and return of the monomeric/oligomeric peak with HPSEC. Thus,

aggregation of the cholate-solubilized gene 9 protein is a reversible process in SDS, affected by several parameters mentioned in Table 1.

The effect of lipids on the initial aggregational state of cholate-solubilized protein.

The cholate-solubilized protein can be reconstituted into vesicles of DOPG. The state of the protein, whether α -helix containing monomeric/oligomeric or β -sheet aggregate, does not change upon incorporation into phospholipids. Monomeric/oligomeric protein, solubilized in cholate, was incorporated into DOPG using the cholate dialysis protocol. The CD spectrum in DOPG was α -helical [37] and remained identical during several days. The aggregational state, as determined by HPSEC, was monomeric/oligomeric, and stable for days as well. When comparing the state of the protein before reconstitution (Figure 3), and after reconstitution (Figure 5), CD, as well as HPSEC, yield similar results. This indicates that transfer from cholate to DOPG does not change the conformational and aggregational state of the protein. After incorporation in DOPG, the state of the protein is stable, whereas it is not stable in cholate. The state of the protein was similar to directly transferred protein into DOPG from TFE, as shown with CD in Figure 2.

Also large aggregates, solubilized in cholate, were incorporated into DOPG using the cholate dialysis protocol. The CD spectrum in DOPG was a typical β -sheet spectrum [37], with a red-shifted zero crossing, decreased intensity, and a single minimum at 219 nm. The presence of β -sheet in this sample was additionally confirmed by FTIR, showing bands at 1624 and 1697 cm^{-1} , characteristic for antiparallel β -sheet (38-40]. Also some intensity at 1655 cm^{-1} was observed, showing that still some α -helix or random coil was present.

Small differences in the zero crossing and intensity of the CD spectra before (Figure 3A) and after reconstitution (Figure 5) suggest, however, that the protein might be organized somewhat differently in the planar DOPG-bilayers, as compared to the more random and spherical organization in cholate micelles. This might influence the homogeneity of the structures, and protein-protein interactions. Also, in cholate some turn structure might be present, as concluded from the presence of a peak at 1680 cm^{-1} (Figure 3B),

which is not present in DOPG. Despite these minor differences, the aggregational and conformational state of the protein remains very similar before and after reconstitution. The protein remains present as large aggregates, maintaining a predominantly β -sheet conformation, which cannot be disrupted by DOPG. From these observations it follows that when using the cholate-dialysis protocol as a method for reconstitution, care has to be taken to avoid aggregation during cholate exposure, since the aggregational state is preserved in DOPG afterwards. Thus in both cases, whether the protein was reconstituted as monomeric/oligomeric protein or as large aggregates, a conserving effect of the lipids was observed.

The conformational and aggregational properties of the gene 9 protein in membrane-mimicking systems.

Summarizing the conclusions, we found that the gene 9 protein is predominantly α -helical in TFE and TFE/water (1/1, v/v) and can be incorporated in detergent micelles and phospholipid vesicles, the system most close to the *in vivo* situation, by the method of cosolubilization via TFE. An alternative method of reconstitution via cholate dialysis requires precaution, because cholate is not sufficiently strong to prevent the protein from aggregating. It is surprising that this aggregation can be reversed by changing conditions. This in contrast with the gene 8 protein of bacteriophage M13, which forms irreversible aggregates [36].

Apparently, the tendency to aggregate is a property of the protein. However, it is not conceivable that this feature is important *in vivo*, since the aggregation-related conformational change appeared not to be reversible with phospholipids. In addition, the observed aggregation occurred at a time scale of days, which is very slow for a process to be relevant *in vivo*. The β -structures present in detergent are anti-parallel strands, as shown by FTIR. If anti-parallel strands are present *in vivo*, then one expects the gene 9 proteins to be oriented in two directions before conversion. This is not very likely with two positive charges on the C-terminal end. Another indication is that, similar to the gene 8 protein, also gene 9 and gene 7 protein of infecting bacteriophages may enter the membrane and can be re-used for synthesis of new particles [2]. This means that they must be inserted in the membrane in the same way as newly synthesized proteins. It is

questionable that this could be done with recycled proteins in a β -conformation. Thus, the observed aggregation probably is a non-native artifact. This conclusion means that it is important to control the aggregation of the protein during reconstitution or purification.

The gene 9 protein adopts an α -helical conformation in TFE and in TFE/water (1/1, v/v). TFE, although said to disrupt tertiary interactions [54], can be a good solvent for single spanning membrane peptides. TFE is known to induce α -helical structure in proteins or peptides, which have a propensity to form α -helices in the native state. However, at very high percentages of TFE also other secondary structures become α -helical [54]. From these results, TFE seems a suitable solvent from which to incorporate the protein into more membrane-like systems such as micelles and vesicles.

Although the α -helical content is lowered upon solubilization by detergents or reconstitution into DOPG, the rough estimate of 50-60% α -helical conformation in these systems is an indication for the ability of the protein to form an α -helix in membrane-mimicking systems. SDS is often used as a model system to determine membrane protein structures, although it has been argued recently that it does not mimic well the typical nature of a phospholipid membrane structure [17, 18]. For the gene 8 protein of M13 and the related fd bacteriophage the conformation has been studied in SDS by high resolution NMR [13-16], and it appears very similar to the conformation in phospholipids as studied by solid state NMR by Opella [9]. Moreover, the transmembrane domain determined from cysteine scanning in phospholipid membranes [17] followed exactly the determined transmembrane helix in SDS.

Biological implications.

The observation of an α -helix containing structure of the gene 9 protein in different hydrophobic environments (TFE, detergents, phospholipids), suggests that this state is the functional state *in vivo*, and effective in the assembly process. This is consistent with the finding that the protein is located in the inner membrane prior to assembly [26]. In general, transmembrane domains of inner membrane proteins are expected to be α -helical. Up to now the conformation of the gene 9 protein was unknown. Previously a

model was described in which the products of gene 7 and 9 were assumed to be α -helices, interacting with each other by their hydrophobic domains to cover the end of the phage particle [28]. Thus, an α -helical conformation present in the gene 9 protein may be suitable in different stages of the life cycle of the bacteriophage, where it must be able to exist in various environments. Firstly it must insert rapidly from the cytoplasm into the membrane phase, secondly it must recognize and associate with the assembly site and thirdly it might be essential for initializing the phage assembly process [27]. Finally, it is part of the viral particle. The finding that the α -helix containing protein can be reconstituted in lipids in an α -helical conformation may be biologically relevant, and opens the way for more detailed studies of gene 9 protein in lipid bilayers in the future.

References

- 1 Berkowitz, S., and Day, L. (1976) *J. Mol. Biol.* 102, 531-547.
- 2 Webster, R. E., and Lopez, J. (1985) in *Virus structure and assembly* (Casjens, S., Ed.) pp 235-267, Jones and Bartlett Publishers Inc., Boston MA.
- 3 Rasched, I., and Oberer, E. (1986) *Microbiol. Rev.* 50, 401-427.
- 4 Model, P., and Russel, M. (1988) in *The Bacteriophages* (Calender, R., Ed.) pp 375-456, Plenum Press, New York.
- 5 Webster, R. E. (1996) in *Phage Display of peptides and proteins* (Kay, B., Winter, J., and McCafferty, J., Eds.) pp 1-20, Academic Press, San Diego.
- 6 Marvin, D. (1998) *Curr. Opinion in Struct. Biol.* 8, 150-158.
- 7 Russel, M. (1991) *Mol. Microbiol.* 5, 1607-1613.
- 8 Russel, M. (1995) *Trends Microbiol.* 3, 223-228.
- 9 Opella, S. J., and McDonnell, P. A. (1993) in *NMR of Proteins* (Clare, G. M., and Gronenborn, A. M., Eds.) pp 160-189, The MacMillan Press Ltd, London, GB.
- 10 Glucksman, M. J., Bhattacharjee, S., and Makowski, L. (1992) *J. Mol. Biol.* 226, 455-470.
- 11 Marvin, D. A., Hale, R. D., Nave, C., and Citterich, M. H. (1994) *J. Mol. Biol.* 235, 260-286.
- 12 Spruijt, R. B., and Hemminga, M. A. (1991) *Biochemistry* 30, 11147-11154.
- 13 Henry, G. D., and Sykes, B. D. (1992) *Biochemistry* 31, 5284-5297.
- 14 McDonnell, P. A., and Opella, S. J. (1993) *J. Magn. Reson., Ser. B* 102, 120-125.
- 15 Van de Ven, F. J. M., Van Os, J. W. M., Aelen, J. M. A., Wymenga, S. S., Remerowski, M. L., Konings, R. N. H., and Hilbers, C. W. (1993) *Biochemistry* 32, 8322-8328.
- 16 Papavoine, C. H. M., Aelen, J. M. A., Konings, R. N. H., Hilbers, C. W., and Van de ven, F. J. M. (1995) *Eur. J. Biochem.* 232, 490-500.

- 17 Spruijt, R. B., Wolfs, C. J. A. M., Verver, J. W. G., and Hemminga, M. A. (1996) *Biochemistry* 35, 10383-10391.
- 18 Stopar, D., Spruijt, R. B., Wolfs, C. J. A. M., and Hemminga, M. A. (1996) *Biochemistry* 35, 15467-15473.
- 19 Wolkers, W. F., Haris, P. I., Pistorius, A. M. A., Chapman, D., and Hemminga, M. A. (1995) *Biochemistry* 34, 7825-7833.
- 20 Van Wezenbeek, P., Hulsebos, T., and Schoenmakers, J. (1980) *Gene* 11, 129-148.
- 21 Grant, R. A., Lin, T.-C., Webster, R. E., and Konigsberg, W. (1981) *Prog. Clin. Biol. Res.* 64, 413-428.
- 22 Lopez, J., and Webster, R. E. (1983) *Virology* 127, 177-193.
- 23 Simons, G. F. M., Beintema, J., Duisterwinkel, F. J., Konings, R. N. H., and Schoenmakers, J. G. G. (1981) *Prog. Clin. Biol. Res.* 64, 401-411.
- 24 Sugimoto, K., Sugisaki, H., Okamoto, T., and Takanami, M. (1977) *J. Mol. Biol.* 111, 487-507.
- 25 Wickner, W. (1983) *Trends Biochem. Sci.* 8, 90-94.
- 26 Endemann, H., and Model, P. (1995) *J. Mol. Biol.* 250, 496-506.
- 27 Russel, M., and Model, P. (1989) *J. Virol.* 63, 3284-3295.
- 28 Makowski, L. (1992) *J. Mol. Biol.* 228, 885-892.
- 29 Rietman, B., Smulders, R., Eggen, I., van Vliet, A., van de Werken, G., and Tesser, G. (1994) *Int. J. Pept. Prot. Res.* 44, 199-206.
- 30 Kaiser, B., Collescot, R., Bossinger, C., and Cook, P. (1970) *Anal. Biochem.* 34, 595-598.
- 31 Seidel, C., Klein, C., Empl, B., Bayer, H., Lin, M., and Batz, H. (1991) *Peptides Proc. 21st EPS*, 236-237.
- 32 Peterson, G. L. (1977) *Anal. Biochem.* 83, 346-356.
- 33 Killian, J. A., Trouard, T. P., Greathouse, D. V., Chupin, V., and Lindblom, G. (1994) *FEBS Lett.* 348, 161-165.
- 34 Schägger, H., and Von Jagow, G. (1987) *Anal. Biochem.* 166, 368-379.
- 35 Haris, P. I., and Chapman, D. (1995) *Biopolymers* 37, 251-263.
- 36 Spruijt, R. B., Wolfs, C. J. A. M., and Hemminga, M. A. (1989) *Biochemistry* 28, 9158-9165.
- 37 Greenfield, N. J., and Fasman, G. D. (1969) *Biochemistry* 8, 4108-4116.
- 38 Goormaghtigh, E., Cabiaux, V., and Ruysschaert, J.-M. (1994) in *Physicochemical Methods in the Study of Biomembranes* (Hilderson, H. J., and Ralston, G. B., Eds.) pp 329-450, Plenum Press, New York.
- 39 Chirgadze, Y. N., Shestopalov, B. V., and Venyaminov, S. Y. (1973) *Biopolymers* 12, 1337-1351.
- 40 Chirgadze, Y. N., and Nevskaya, N. A. (1976) *Biopolymers* 15, 607-625.
- 41 Venyaminov, S. Y., and Yang, J. (1996) in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G., Ed.) pp 69-105, Plenum Press, New York.
- 42 Segawa, S., Fukuno, T., Fujiwara, K., and Noda, Y. (1991) *Biopolymers* 31, 497-509.
- 43 Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., and Wright, P. E. (1992) *J. Mol. Biol.* 226, 795-817.
- 44 Kemmink, J., and Creighton, T. E. (1995) *Biochemistry* 34, 12630-12635.

- 45 Chen, Y. H., Yang, J. T., and Chau, K. H. (1974) *Biochemistry* 13, 3350-3359.
- 46 Fasman, G. D. (1995) *Biopolymers* 37, 339-362.
- 47 Nevskaya, N. A., and Chirgadze, Y. N. (1976) *Biopolymers* 15, 637-648.
- 48 Miura, T., and Thomas, G. J., Jr. (1995) in *Introduction to Biophysical Methods for Protein and Nucleic Acid Research* (Glaser, J. A., and Deutscher, M. P., Eds.) pp 261-315, Academic Press, San Diego.
- 49 Torii, H., and Tasumi, M. (1992) *J. Chem. Phys.* 96, 3379-3387.
- 50 Arnold, G. E., Day, L. A., and Dunker, A. K. (1992) *Biochemistry* 31, 7948-7956.
- 51 Manning, M. C., and Woody, R. W. (1989) *Biochemistry* 28, 8609-8613.
- 52 Hennessey, J. P., Jr. , and Johnson, W. C., Jr. (1981) *Biochemistry* 20, 1085-1094.
- 53 Levitt, M., and Chothia, C. (1976) *Nature* 261, 552-558.
- 54 Sivaraman, T., Kumar, T. K. S., and Yu, C. (1996) *Int. J. Biol. Macromol.* 19, 235-239.
- 55 Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* 157, 105-132.

Chapter 3

Conformation and Orientation of the Gene 9 Minor Coat Protein of Bacteriophage M13 in Phospholipid Bilayers

M. Chantal Houbiers, Cor J.A.M. Wolfs, Ruud B. Spruijt, Yves J.M. Bollen,
Marcus A. Hemminga, Erik Goormaghtigh

Published in: Biochim. Biophys. Acta 1511 (2001), 224-235.

Abstract

The membrane-bound state of the gene 9 minor coat protein of bacteriophage M13 was studied in model membrane systems, which varied in lipid head group and lipid acyl chain composition. By using FTIR spectroscopy and subsequent band analysis a quantitative analysis of the secondary structure of the protein was obtained. The secondary structure of the gene 9 protein predominantly consists of α -helical (67%) and turn (33%) structures. The turn structure is likely to be located C-terminally where it has a function in recognizing the phage DNA during bacteriophage assembly. ATR-FTIR spectroscopy was used to determine the orientation of gene 9 protein in the membrane, revealing that the α -helical domain is mainly transmembrane. The conformational and orientational measurements result in two models for the gene 9 protein in the membrane; a single transmembrane helix model and a two-helix model consisting of a 15 amino acid long trans-membrane helix and a 10 amino acid long helix oriented parallel to the membrane plane. Potential structural consequences for both models are discussed.

Introduction

The filamentous bacteriophage M13 consists of a circular single-stranded DNA molecule of 6408 nucleotides encapsulated in a long cylindrical protein coat. The coat is composed of about 2700 copies of the major coat protein, the product of gene 8. One end of the phage contains about 5 copies each of the protein products of gene 7 and gene 9. The primary structures of the phage-encoded proteins were deduced from the nucleotide sequence of the DNA genome. Detailed descriptions of the phage particle, the reproductive cycle, and assembly can be found in several reviews [1-7]. New phage particles are assembled and extruded from its host *Escherichia coli* at the membrane bound assembly site, without lysis of the host cell. In the assembly site, coat proteins as well as nonstructural proteins are present, and the viral DNA is extruded through the cellular membranes while picking up the coat proteins.

Gene 7 and gene 9 proteins were proposed to be associated with the membrane, prior to assembly, based on the finding that they retain their amino-terminal formyl group [8]. Simons [8] suggested that gene 7 and gene 9 proteins are incorporated in phage as primary translational products, so no proteolytic processing occurs. Endeman and Model showed indeed that the gene 7 and gene 9 proteins are localized in the *E. coli* inner membrane prior to assembly [9].

Gene 7 and gene 9 proteins might play an important initiation role by interacting with the packaging signal of the phage genome [10]. This is supported by the fact that gene 7 and gene 9 proteins are located at the end of the phage particle that emerges first [1]. It was also found that deletions in the packaging signal could be compensated for by mutations in gene 1, gene 7, and gene 9 proteins [10], suggesting a key role for these proteins early in the assembly process. Furthermore, absence of gene 7 or gene 9 protein almost completely abolishes the production of phage particles [1].

In this investigation we focus on the gene 9 protein. The amino acid sequence of this 32-residue protein is the following [11]: HCO-Met-Ser-Val-Leu-Val-Tyr-Ser-Phe-Ala-Ser-Phe-Val-Leu-Gly-Trp-Cys-Leu-Arg⁺-Ser-Gly-Ile-Thr-Tyr-Phe-Thr-Arg⁺-Leu-Met-Glu⁻-Thr-Ser-Ser-COOH. The N-terminal methionine contains the formyl group [8]. To

understand more about the role of this protein in assembly, knowledge about its membrane-bound state prior to assembly is of key importance. Hydropathy analysis of the amino acid sequence shows that there is a hydrophobic stretch at the N-terminal side, which runs from residue 1 to around 16, as has been shown before [12]. In addition, we calculated a prediction for an amphipathic helix by Eisenberg's method [13] and found a stretch running from residue 16 to 31. We also searched for predicted turns, using the method of Chou and Fasman [14]. This resulted in two areas with a very high turn potential, which were from residue 17 to 20, and from residue 29 to 32. These predictions are depicted in Figure 1.

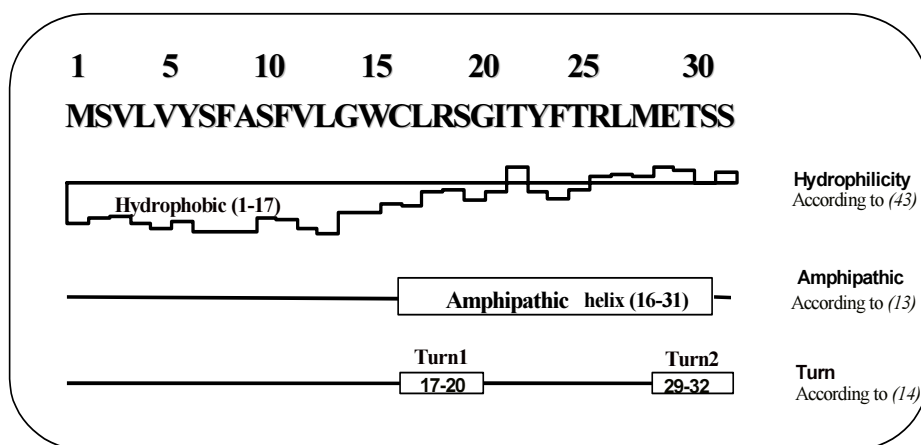


Figure 1: Hydropobicity analysis of gene 9 protein according to [43], using a sliding window of 9. Amphipathicity analysis of the gene 9 amino acid sequence by Eisenberg's method [13]. Turn prediction for the gene 9 protein by the method of Chou and Fasman [14].

Recently the secondary structure of synthetic gene 9 protein was found to be mainly α -helical in TFE, detergent micelles, and in vesicles of DOPG, provided that care was taken to prevent aggregation into β -sheet structures [12]. Based on this information, the membrane-bound conformation of the gene 9 protein was proposed to be predominantly α -helical. In this work we attempt to gain information about how the gene 9 protein is assembled in membranes composed of various lipids. By FTIR spectroscopy we determined the orientation of the α -helical domain of gene 9 protein with respect to the membrane plane. In addition, FTIR band analysis revealed new features about the

secondary structure in the non-helical domain. Altogether, these conformational and orientational measurements result in new models of the gene 9 protein embedded in membranes.

Materials and Methodes

Materials.

Fmoc-amino acids and the solid support *p*-alkoxybenzyl alcohol resin (Wang resin) were purchased from Bachem (Bubendorf, Switzerland). The compounds 1-hydroxybenzotriazole, diisopropylcarbodiimide, and ethanedithiol (EDT) were obtained from Fluka (Buchs, Switzerland). Dichloromethane (DCM), dimethylformamide (DMF), isopropyl alcohol, piperidine, trifluoroacetic acid (TFA) were obtained from Merck (Darmstad, Germany), and 2,2,2-trifluoroethanol (TFE) was obtained from Acros (Pittsburgh, PA). 1,2-Dioleoyl-*sn*-glycero-3-phosphoglycerol (DOPG), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC), 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), and 1,2-dimyristoyl-*sn*-glycero-3-phosphoglycerol (DMPG) were obtained from Avanti Polar Lipids (Alabaster, AL). The Superdex 75 column was obtained from Pharmacia (Uppsala, Sweden). The gene 9 protein was prepared by solid phase synthetic techniques and characterized as described before [12]. The N-terminus was synthesized with a formyl group.

Solubilization of protein in TFE

The protein was first dissolved in a small volume of TFA (10-20 μ l/mg of protein) and dried under a stream of nitrogen. Next, the protein was washed two times with TFE, followed by evaporation of the TFE under a stream of nitrogen to remove residual TFA. Finally the protein was dissolved in a known volume of TFE. To remove residual traces of TFA (which can contribute to the signal in the amide I region in FTIR measurements [15]), a 2 mM HCl solution was added to the TFE stock in a 1/1 (v/v) ratio, to yield a 1 mM HCl concentration in the protein solution. The solution was subsequently lyophilized, re-dissolved into TFE, and used for further sample preparation. The protein

concentration was determined by the procedure of [16] with bovine serum albumin as a standard.

Reconstitution of protein into lipid vesicles

A desired amount of phospholipid was taken from a chloroform solution in the case of dioleoyl lipids. The chloroform was evaporated under a stream of nitrogen, and the remaining lipid film was kept overnight under vacuum to remove all traces of chloroform. In the case of dimyristoyl lipids the desired amount was weighed as a powder. To reconstitute the gene 9 protein in the lipids, the method described by [17] was used. Briefly, the desired amount of gene 9 protein dissolved in TFE was added to an equal volume of water, containing the desired amount of lipids. The mixture was vortexed for 2 seconds at room temperature. Next, this mixture was diluted to yield a 16:1 ratio of water to TFE by volume. The samples were mixed by vortexing for 2 s and lyophilized. The sample was rehydrated in the desired volume of water.

Sucrose Gradient Centrifugation

Samples of reconstituted proteins in DOPC or DOPG were placed on top of a linear 0-40% w/w sucrose gradient, and centrifuged for 18 hours at 110000g at 4 °C in a Beckman SW41 rotor. To enable visualization of the lipid-protein complexes, octadecyl Rhodamine B (0.16 % mol/mol lipid) was added, which dissolves into the membranes with high specificity.

Differential Scanning Calorimetry

Gene 9 protein in DPPC vesicles was prepared by mixing a known amount of protein solubilized in TFE with 5 mg of DPPC dissolved in chloroform. After evaporation of the solvent the samples were dried under vacuum for 3 hours and resuspended in 750 µl 50 mM Tris, pH 8.0. After homogenization the liposomes were collected by low speed centrifugation and redissolved in 30 µl of buffer. 20 µl of this sample was used in a 25-µl cup for DSC. Samples were measured using a Pyris 1 DSC (Perkin-Elmer) at a

heating rate of 10 °C/min. From the heating scans the phase transition temperature and peak width were determined.

Determination of the aggregation state of the protein

The aggregation state of the gene 9 protein samples was checked after FTIR measurements. The samples were diluted 15 times with cholate elution buffer (50 mM cholate, 10 mM sodium phosphate, 20 mM NaCl, pH 8.0) to dissolve the lipid vesicles and immediately analyzed after addition of the cholate. High-performance size exclusion chromatography (HPSEC) was performed on a Superdex 75 column (1.0 x 30 cm). A flow of 0.5 ml/min was applied in the cholate elution buffer mentioned above. On-line detection was done by fluorescence measurement. For molecular weight calibration of the columns, alcohol dehydrogenase (150 000 Da), bovine serum albumin (67 000 Da), myoglobin (18 800 Da), A-protein (6500 Da), and bacitracin (1400 Da) were used.

Fourier Transform Infrared Spectroscopy

Samples for FTIR typically contained 0.04 mg protein at a lipid/protein ratio of 25 (mol/mol). Films were prepared on a CaF₂ window (13 mm diameter). FTIR spectra were recorded on a Perkin-Elmer 1750 FTIR spectrometer at room temperature. The optical bench was purged with dry air (Balston, Maidstone Kent, U.K.) at a flow rate of 25 L min⁻¹. The acquisition parameters were the following: 4 cm⁻¹ resolution; 32 co-added interferograms, 3500-900 cm⁻¹ wave number range. For the sake of comparison, spectral intensities were normalized with the maximum peak in the amide I region, after subtraction of a straight baseline passing through the two minima bordering the amide I region. Partially dehydrated films were prepared by keeping the samples, spread as a suspension on the CaF₂ window, for several hours in a cabin continuously purged with dry air. Samples were mounted in the FTIR spectrometer without exposure to the room air by using a rubber spacer and a second CaF₂ window to cover the sample film. Deuterated samples were prepared similarly, but were immersed in 40 µl D₂O before drying for several hours in the cabin purged with dry air. Before measurement 3 µl of D₂O was added on top of the dried film, which was enough to fully immerse the

film in D₂O during the measurement. A second CaF₂ window was used to cover the sample.

Attenuated Total Reflection Fourier Transform Infrared Spectroscopy.

Oriented multilayers were formed by slow evaporation of a 15 μ l sample, containing 0.03 mg protein and the lipids, on one side of the ATR plate, which was placed into a sample holder (Perkin Elmer 186-0354). Spectra were recorded at room temperature on a Perkin-Elmer infrared spectrophotometer 1720X equipped with a Perkin-Elmer microspecular reflectance accessory and a polarizer mount assembly with a gold wire grid element. The internal reflection element was a germanium ATR plate (50x20x2 mm, Harrick EJ2121) with an aperture angle of 45° yielding 25 internal reflections. 128 accumulations were performed to improve the signal/noise ratio. The spectrophotometer was continuously purged with dried air (FT-IR purge gas generator 75-62, Whatman, Haverhill, MA, USA). Spectra were recorded at a nominal resolution of 2 cm⁻¹. Remaining spikes due to H₂O vapor were subtracted as described before [18]. Subsequently, the spectra were smoothed by apodization of the Fourier transformed spectrum by a Gaussian lineshape with a full width at half height of 4 cm⁻¹.

Orientation of the secondary structures

In an α -helix, the main transition dipole moment lies approximately parallel to the helix axis. It is therefore possible to determine the mean orientation of the α -helix structure from the orientation of the peptide bond C=O group [19]. Spectra were recorded with parallel (\parallel) and perpendicular (\perp) polarized incident light with respect to the incidence plane. The dichroic spectra are the difference between the spectra recorded with parallel and perpendicular polarizations. The perpendicular spectra were multiplied by a factor R_{ISO} (i.e., the dichroic ratio for an isotropically oriented sample) before subtraction to take into account the difference in the relative power of the evanescent fields [20]. This is allowed since the C=O groups of the phospholipids are assumed to have no net orientation [20]. Polarization was expressed as the dichroic ratio R . R is defined as the ratio A_{\parallel}/A_{\perp} , where A_{\parallel} and A_{\perp} are the absorbances (measured as band areas) with the parallel and perpendicular orientations of the polarizer. The dichroic ratio was converted

into an order parameter, which can be related to a maximum tilt angle θ of an orientational distribution of dipoles [19, 20]. The angle θ is defined as the maximum angle of the molecular (helix) axis with respect to the normal of the ATR plate. For the amount of material used throughout this work, film thickness (0.1-0.3 μm) is small when compared to the IR wavelength (6.06 μm at 1650 cm^{-1}), but close to the penetration depth d_p (0.39 μm). This does not allow the "thin film" approximation [19] to be used, but instead the film thickness has to be taken into account for the establishment of the equations describing the dichroic ratio as a function of the orientational order parameter [20]. This can be achieved by computing an apparent film thickness from R_{ISO} (i.e., the dichroic ratio observed for an isotropically oriented dipole), using $n_1 = 4.0$ (germanium), $n_2 = 1.7$ (protein), $n_3 = 1.0$ (air) in order to evaluate correctly the electric field component perpendicular to the ATR plate [20]. The angle between the long axis of the α -helix and the (C=O) dipole moment was taken as 27° [19, 21].

Secondary structure determination

Curve fitting with Lorentzian lines was performed with the software Origin 4.0. The position of the bands was chosen on the basis of the shape of the deconvoluted spectra. Fourier self-deconvolution was used using a Lorentzian line shape (full width at half height = 30 cm^{-1}) and a Gaussian line shape (full width at half height = 15 cm^{-1}) for the apodization. This procedure has been detailed elsewhere [19, 22, 23]. Intensity, width, and frequency of the Lorentzian bands were adjusted by the program. To avoid contributions of the lipid C=O vibration in the amide I region, we removed the lipid C=O band around 1740 cm^{-1} by subtracting the pure lipid spectrum after scaling on the lipid C=O band prior to curve fitting. A straight baseline was adjusted as an additional parameter to obtain the best fit. After band fitting each component was assigned to a secondary structure type, and then the integrated areas of these components were expressed as a percentage of the total amide I area and were used to determine the secondary structure elements [22].

Results

Reconstitution of protein in lipid vesicles

Gene 9 protein was reconstituted in phospholipid vesicles via the method of cosolubilization with organic solvent [17], which was shown to be a proper method for the gene 9 protein [12]. Binding of the gene 9 protein to DOPC and DOPG vesicles was checked by sucrose density gradient centrifugation. All protein was bound to lipids (data not shown). This was concluded from the observation of a lipid-protein band at intermediate position in the centrifugation tube, whereas no bands were observed at the top of the tube where, in control experiments, pure lipid vesicles end up, or at the bottom of the tube where the pure proteins were located. In order to check whether the lipid acyl chains were influenced by the gene 9 protein, we studied gene 9 protein reconstituted in lipids by Differential Scanning Calorimetry. As a membrane model system we used DPPC, which has a phase transition at 42 °C. When DPPC vesicles contained gene 9 protein the phase transition shifted to lower temperature and broadened (data not shown). Upon increasing the gene 9 protein concentration the effect was more pronounced. These results indicate that the lipid acyl chains are influenced by the presence of protein. However, by using FTIR we found that the acyl chains at room temperature were not greatly disturbed by the presence of protein in DMPC, DMPG, DOPC, and DOPG phospholipids. This is concluded from the observation that the acyl chain stretch vibrations of the lipids around 2925 and 2854 cm^{-1} show only a very small shift in the lipid C=O stretching, which could result from a slightly different hydration of the carbonyl of the lipid heads, with or without gene 9 protein reconstituted in the membrane (data not shown).

Conformation

Figure 2 (curve A) shows the 1800–1400 cm^{-1} region of the FTIR spectrum of gene 9 protein in DMPG. The lipid (C=O) band is found at 1738 cm^{-1} , the protein amide I at 1657 cm^{-1} and amide II at 1546 cm^{-1} . The lipid contribution was removed by subtraction of the FTIR spectrum of pure lipid contribution in figure 2C. The resulting difference spectrum (curve C) shows a wiggle with a negative and a positive intensity centered

around the position of the lipid C=O band. This indicates that in the presence of protein the lipid (C=O) band experience a slight shift to higher wave numbers.

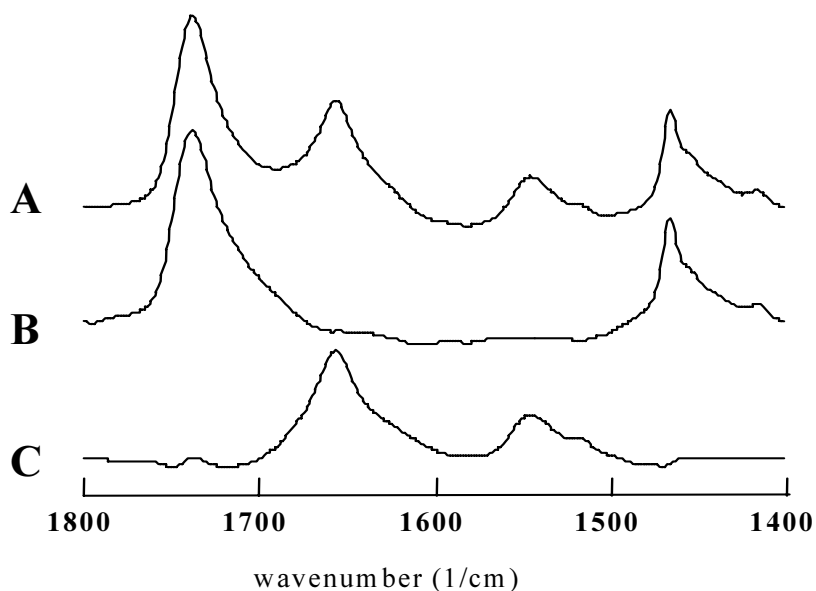


Figure 2: FTIR absorption spectra of partially dehydrated films of gene 9 protein reconstituted in DOPG at an L/P ratio of 25 (mol/mol) (A) and of identically treated DOPG without protein (B). In the lower spectrum (C) spectrum (B) is subtracted from (A) after normalization on the areas of the lipid C=O band, to remove the contribution of the lipids.

Since the amide I band ($1600\text{--}1700\text{ cm}^{-1}$) is most useful for structure determination, Figure 3 compares the amide I regions of the FTIR spectra of gene 9 protein in three different lipid systems with dimyristoyl acyl chains: DMPC, DMPC/DMPG (3/1, mol/mol), and DMPG (the latter as taken from graph A of Figure 2). Similar bandshapes were obtained for three different lipid systems with dioleoyl acyl chains: DOPC, DOPC/DOPG, and DOPG (data not shown). All these lipid systems, varying in head group composition (neutral PC versus negatively charged PG head group) and in acyl chain composition (saturated dimyristoyl versus unsaturated dioleoyl acyl chains), result in similar amide I bands of gene 9 protein reconstituted in these systems.

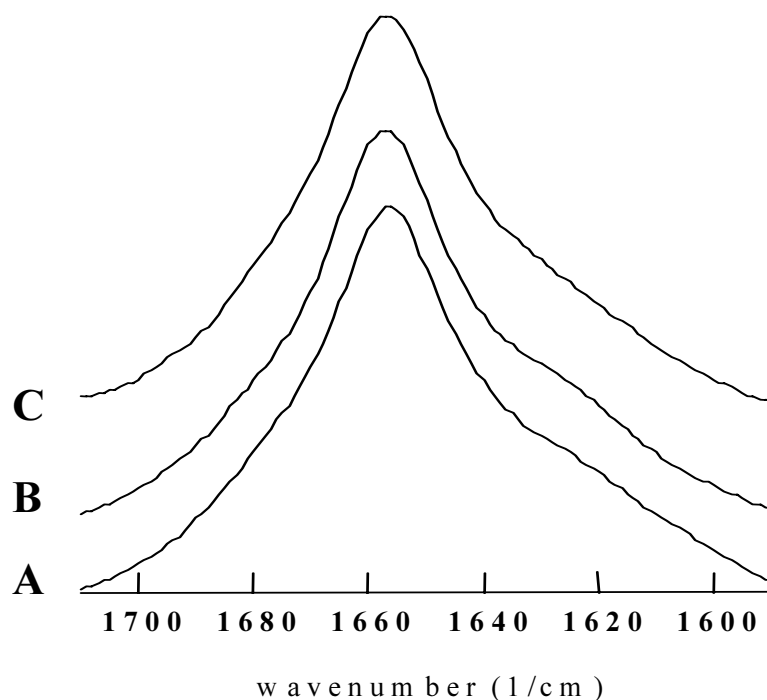


Figure 3: Amide I region of the FTIR absorption spectra of partially dehydrated films of gene 9 protein reconstituted in DMPC (A), DMPC/DMPG (3/1, mol/mol) (B), DMPG (C). Lipid contributions have been subtracted as shown in Figure 2. L/P ratios were 25 (mol/mol).

Deconvolution of the spectra, as shown in Figure 4 (C) for gene 9 protein in DMPC/DMPG (3/1, mol/mol), identifies the presence of three components located around 1679, 1657, and 1627 cm^{-1} . Curve fitting (e.g. Figure 4 A), was used to quantify the relative contributions of the three bands. The difference spectrum between the original spectrum and the fit is presented in Figure 4 (B).

Curve fitting results obtained for the gene 9 protein in the dimyristoyl-lipid series and the assignments are presented in Table 1. The results for the dioleoyl-lipids (not shown) are very similar. The main band at 1657 cm^{-1} is assigned to α -helix [18]. The relative narrow bandwidth agrees with a major contribution of α -helical absorption. Since it is known that random coil contribution overlaps the α -helical absorption in films of H_2O suspension [24,25] we also measured the samples in dioleoyl lipids in D_2O suspension

after exposure to D₂O for several hours. Random coil structure is known to exchange

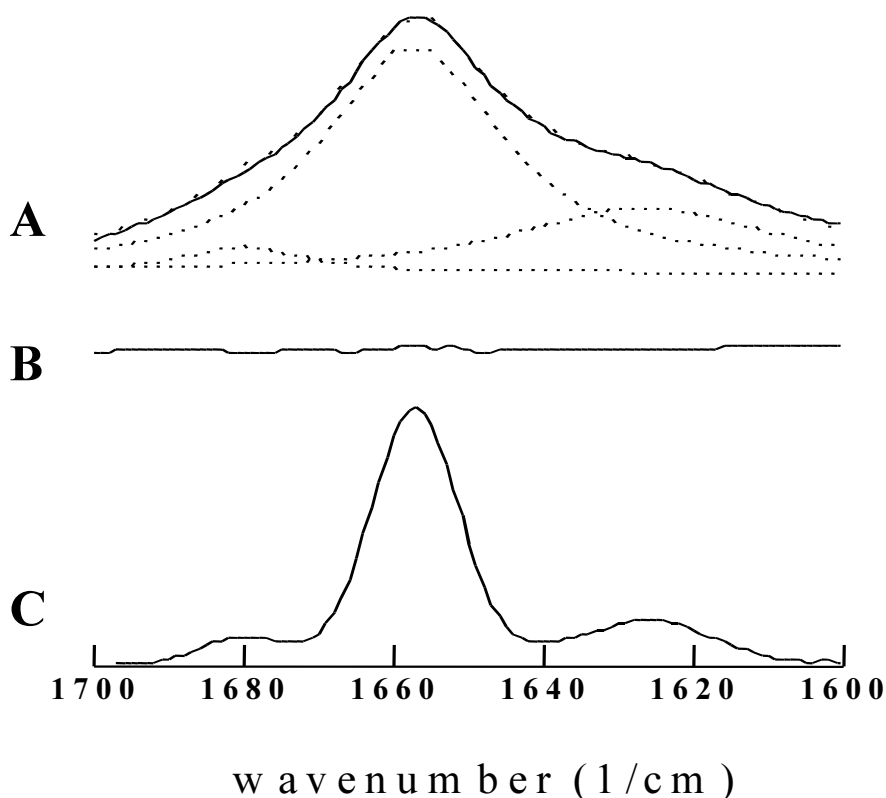


Figure 4: (A) FTIR absorption spectrum of a partially dehydrated film of gene 9 protein reconstituted in DMPC/DMPG (3/1, mol/mol) after lipid subtraction (solid line) and the component spectra (dashed lines) that are used to fit the spectrum according to the parameters given in Table 1. The total fit is represented by dashes. (B) Difference of the absorption spectrum and the sum of the fitted component spectra in (A). (C) Deconvoluted spectrum.

hydrogen for deuterium on a fast timescale, resulting in a shift of the amide I band to approximately 1640 cm⁻¹ [18, 24-26]. We did not observe such a shift. The amide I band was shifted only 1 to 3 cm⁻¹ to lower wavenumber and the loss of amide II intensity was very small confirming the α -helix assignment. Furthermore, the very slow exchange rate indicates that the helix structure is either extraordinarily stable or that it is embedded in the hydrophobic region of the lipid membrane. In addition, we measured these samples by Circular Dichroism (data not shown). All spectra were similar to the CD spectrum of gene 9 protein in DOPG [12], indicating the secondary structure of gene 9 protein is mainly α -helical. Neither did CD measurements indicate any β -sheet structure, showing

Table 1: Analysis of the Amide I Region of Gene 9 Protein in Various Lipid Systems^a

lipid system	unpolarized spectra				dichroic spectra (\parallel - \perp)		
	γ (cm^{-1})	$\Delta\gamma$ (cm^{-1}) ^b	area (%) ^c	assignment	γ (cm^{-1})	$\Delta\gamma$ (cm^{-1}) ^b	area (%) ^c
DMPC	1679	20	6	hydrogen- bonded turn	1680	13	4
	1657	29	61	α -helix	1657	25	90
	1625	49	33	non- hydrogen- bonded turn	1635	18	6
DMPC/DMPG (3/1, mol/mol)	1680	18	5	hydrogen- bonded turn	1680	18	8
	1657	29	67	α -helix	1659	22	86
	1627	42	28	non- hydrogen- bonded turn	1637	20	6
DMPG	1678	17	5	hydrogen- bonded turn	1678	22	11
	1657	29	69	α -helix	1659	23	84
	1627	43	26	non- hydrogen- bonded turn	1636	17	5

^a The error on the frequency γ and linewidth $\Delta\gamma$ is $\pm 2 \text{ cm}^{-1}$.

^b Full width at half height

^c Relative area in amide I. The error in the calculated area is $\pm 5\%$.

a zerocrossing around 202-203 nm for all samples. The presence of β -sheet would shift the zero crossing up to 210 nm [27].

The bands around 1627 and 1679 cm^{-1} are assigned to turn structure. HPSEC analysis of these samples did not show the presence of any aggregated protein (data not shown), which has been shown to correspond with β -sheet structure [12]. Thus, the presence of only monomeric protein excludes β -sheet structure, since the monomeric/oligomeric state corresponds to a predominant α -helical structure [12]. Therefore, the FTIR bands around 1627 and 1679 cm^{-1} can be assigned to turn structure, and not to β -sheet structure. Based on the work of [28, 29] the low wave number band around 1627 cm^{-1} is assigned to hydrogen-bonded groups present in a turn structure, whereas the high wave number band around 1679 cm^{-1} is assigned to the non-hydrogen bonded C=O groups present in a turn structure [28-30]. On the basis of this assignment, the total amount of turn structure then amounts to $35 \pm 5\%$. Quantification in this way supposes that the extinction coefficients do not vary with secondary structure type. Literature, however, shows that this may not be the case [31]. In that case the band at 1685 cm^{-1} would be underestimated, and the band at 1630 cm^{-1} would be similarly to some extent overestimated with respect to the α -helix content. The assignment of the bands around 1627 and 1679 cm^{-1} to turn structure is in agreement with the results found for bacteriophage M13 major coat protein. This protein yielded similar FTIR bands, which were assigned to turn structures because the ratio of the two bands did not correspond to the ratio of 10 expected for β -sheet [30]. Byler and Susi assigned similar bands at 1675 and 1638 cm^{-1} for several highly helical proteins (hemoglobin, myoglobin, cytochrome c, and ferritin) and assigned these to the short extended chains connecting the helical cylinders (and not to β -sheet) [24]. Haris observed similar bands in rhodopsin in bovine rod outer segment membranes, which were assigned to turn structures [32].

We checked for amino acid side chain contributions in our spectra by calculating the contributions from the tyrosine, arginine, glutamic acid, and phenylalanine residues, present in gene 9 protein (data not shown). These side chains contribute in the amide I and amide II regions [33]. Subtraction of these contributions from our spectrum does not significantly influence spectral intensities and line shapes.

Orientation

To determine the orientation of the gene 9 protein in lipid bilayers by ATR-FTIR spectroscopy the phospholipid membrane should be well ordered. For this reason we used the dimyristoyl-lipids instead of the dioleoyl-lipids. In addition, the saturated acyl chains of the dimyristoyl-lipids allow an accurate orientation determination of the lipids by means of the $\gamma_{\text{W}}(\text{CH}_2)$ progression [34].

For gene 9 protein reconstituted in DMPC, DMPC/DMPG (3/1, mol/mol), and DMPG vesicles spectra were recorded with incident light polarized parallel (\parallel) or perpendicular (\perp) to the incidence plane. The dichroic spectra are the difference between the spectra recorded with parallel and perpendicular polarizations, as described in Materials and Methods. Figure 5 shows the ATR-FTIR spectra of the gene 9 protein reconstituted in DMPC/DMPG (3/1, mol/mol) for parallel and perpendicular orientations of the polarizer (lower spectra) and the dichroic spectrum (top). The dichroic spectrum in Figure 5 demonstrates a positive deviation, indicating that the dipole is oriented mainly parallel to the normal of the plate. Since in an α -helix the amide I dipole is close to the helix long axis [19] this result demonstrates that the helix axis is oriented mainly parallel to the normal of the ATR plate. Nearly identical results were obtained for the gene 9 protein reconstituted in DMPC or in DMPG (spectra not shown).

From Figure 5 it can be noted that bandwidth in the dichroic spectrum is smaller than the bandwidth of the unpolarized spectra. Analysis of the component bands (Table 1) shows that the bands assigned to turn contribute much less in the dichroic spectra, whereas the α -helix band is more predominant (up to 90% contribution). The relatively lower prevalence of the turn bands in the dichroic spectrum indicates that these structures are less oriented, whereas the α -helical structure adopts a strong orientation. The absence of the non-helical structures (turns, random) in the polarized spectra explains why the dichroic spectra are more narrow. The orientation of the lipid acyl chains was assessed by means of the $\gamma_{\text{W}}(\text{CH}_2)$ progression between 1200 and 1350 cm^{-1} . This band structure is originating from a resonance between the CH_2 groups and the ester group of the all-*trans* hydrocarbon chain in the α -position (peaks at 1202, 1227, 1254, 1278, 1304, and 1327 cm^{-1}) [34]. The strong \parallel polarizations of these peaks, and

therefore their positive deviations in the dichroic spectrum (data not shown) indicate that the all-*trans* hydrocarbon chains of DMPC are nearly normal to the germanium surface. This implies that the membranes are highly oriented parallel to the ATR plate and that the helix of the gene 9 protein is oriented preferentially perpendicular relative to the membrane plane.

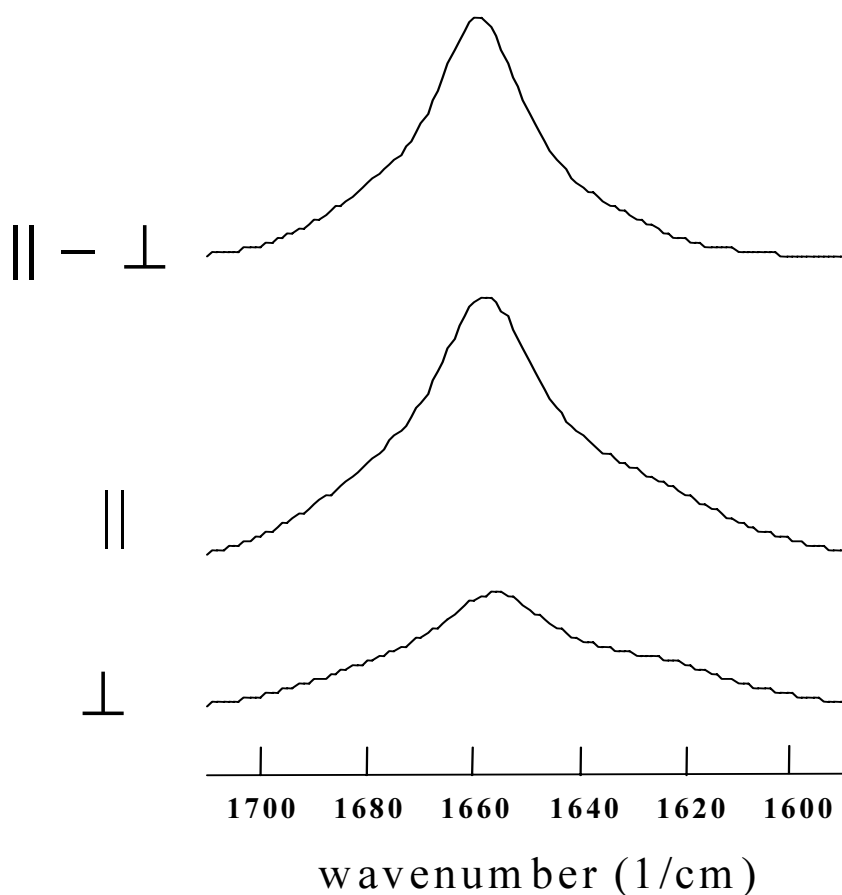


Figure 5: ATR-FTIR spectra of gene 9 protein reconstituted in DMPC/DMPG (3/1, mol/mol), partially dehydrated, for parallel and perpendicular orientations of the polarizer. The dichroic spectrum (top) was obtained as $||$ minus \perp spectrum after multiplication of the \perp spectrum by R_{iso} (see Table 2). It should be noted that the intensity of the dichroic spectrum (top) has been expanded twice for better visualization. The L/P ratio was 25 (mol/mol).

The orientation of dipoles can be analyzed in a more quantitative way by considering dichroic ratio values. The dichroic ratio R , defined as $A_{||}/A_{\perp}$, can be converted into an

order parameter, which can be converted into the so-called maximum tilt angle θ between the molecular (helix) axis and the normal of the ATR plate [20]. In our calculations we assume perfect membrane ordering, resulting in an order parameter of 1.0. This assumption is justified by the observation of a very high degree of orientation of the lipid molecules. Dichroic ratios of the $\gamma_{\text{W}}(\text{CH}_2)$ progression bands of the all-*trans* hydrocarbon chains of the lipids between 1300 and 1200 cm^{-1} yielded values in the range from 3 to 6. This indicates that the all-*trans* hydrocarbon chains were oriented nearly normal to the germanium surface [34, 35], irrespective of whether protein was present or not. This observation implies an almost perfect membrane order. The dichroic ratios calculated from the amide I bands of the ATR-FTIR spectra of gene 9 protein reconstituted in DMPC, DMPC/DMPG (3/1, mol/mol), and DMPG for parallel and perpendicular orientations of the polarizer are presented in Table 2. The third column shows values for R_{ISO} that were used to calculate the real film thickness, which is required for the orientation determination [20]. The fourth column shows the calculated maximum tilt angle θ of the helix according to Raussens et al. [36].

Discussion

Gene 9 protein is known to be located in the inner membrane of *E. coli* before incorporation into bacteriophage M13 [9]. A first characterization of the gene 9 protein in the membrane-bound state was performed in model membrane systems including organic solvent, detergent micelles, and phospholipid vesicles [12]. The gene 9 protein was found to be mainly α -helical and the percentage of α -helix was estimated based on CD measurements. In this paper we aim at determining how the gene 9 protein is located in the membrane by measuring the orientation of the α -helical domain in the bilayer, using FTIR spectroscopy. In addition, we performed a quantitative analysis of the components of the FTIR amide I band, resulting in new information on secondary structure of the gene 9 protein.

Reconstitution of protein in lipid vesicles

First we established that the gene 9 protein was incorporated in the phospholipid bilayers. The gene 9 protein was reconstituted in phospholipid vesicles by using a protocol of cosolubilization with organic solvent [17], which has also been shown to be applicable for the gene 9 protein [12]. After reconstitution protein-lipid complexes have been analyzed by sucrose density centrifugation to determine whether all protein was complexed to the lipids. The observation of one band at intermediate density, indicating a mix of proteins and lipids, and no protein band at high density or lipid band at low density, indicated that all protein was complexed with the phospholipid vesicles, and that no free protein was present. By using DSC the effect of protein incorporation on the phase transition temperature of DPPC acyl chains was monitored. The observation that gene 9 protein causes a shift and a broadening of this phase transition indicates that the protein must be influencing the lipid acyl chains. We concluded that the protein is not just bound to the surface of the membrane, but must be penetrating in the interior of the membrane [37]. These observations indicate that in the experiments carried out after reconstitution the gene 9 protein was at least partly incorporated in the bilayer interior, and not just peripherally bound to the surface of the membrane.

Conformation

Analysis of the amide I regions of the infrared spectra of gene 9 protein in DOPC, DOPC/DOPG (3/1, mol/mol), DOPG, DMPC, DMPC/DMPG (3/1, mol/mol), and DMPG (Figure 3 and Figure 4) shows that it consists of three bands. The main band, representing 67% of the area, at 1657 cm^{-1} is assigned to α -helix [18, 24]. This assignment is supported by measurement of the samples in D_2O , and by CD measurements. Apparently, the protein contains similar amounts of α -helix irrespective of the head group charge or acyl chain saturation. The percentage of α -helix found here is slightly higher than reported before [12].

The dichroic spectra (Figure 5), resulting from the orientation measurements, and the analysis of their component bands (Table 1) provide additional information. The bands around 1627 and 1679 cm^{-1} have a net positive, but much smaller, deviation in the

dichroic spectrum than in the unpolarized spectrum. This agrees with turn structure, since turns are expected to have a less defined orientation than α -helices. HPSEC and CD measurements support the conclusion that the presence of β -sheet can be ruled out. On the basis of this assignment the total amount of turn structure then mounts up to 31-39%. Gene 9 protein contains 32 amino acids, so it is possible that gene 9 protein might contain two turns, each of about 5 amino acids, as already suggested by the turn prediction [14].

Table 2: Dichroic Ratio Values of Amide I and corresponding Maximum Tilt Angles for Gene 9 Protein in Various Lipid Systems

lipid system ^a	R^b	R_α^b	R_{iso}^b	$\theta(^{\circ})^c$	x^d
DMPC	1.72	1.96	1.18	33	0.61
DMPC/DMPG (3/1, mol/mol)	2.06	2.43	1.24	30	0.70
DMPG	2.06	2.48	1.20	27	0.77

^a The L/P ratios were 25 (mol/mol). The lipid contribution was subtracted from both spectra with the parallel and perpendicular orientations of the polarizer as was performed in Figure 2.

^b R is the dichroic ratio; R_α is the dichroic ratio for the helix component and R_{iso} the dichroic ratio for an isotropically oriented sample.

^c One-helix model assuming 100% helix. The error in the calculated angles is $\pm 5^{\circ}$.

^d Two-helix model. x is the fraction of transmembrane helix relative to the total helix content.

Orientation

To our knowledge the only detailed information available about the membrane-bound state of the gene 9 protein describes conformational properties [12]. Nothing is known about how gene 9 protein is assembled in the lipid membrane. Here we provide additional information by measuring the orientation of gene 9 protein in lipid bilayers, deposited on germanium plates by ATR FTIR spectroscopy. In an α -helix the transition dipole moment of the amide I band lies approximately parallel to the helical axis, so it is possible to determine the mean orientation of an α -helix from the orientation of the C=O group [19]. Dichroic spectra, allowing a qualitative analysis of the orientation of the gene 9 protein in DMPC, DMPC/DMPG (3:1), and DMPG show a positive band at 1657 cm^{-1} , indicating that the α -helix is oriented preferentially parallel to the membrane normal. The turn structure, assigned to the peaks around 1680 and 1630 cm^{-1} , is less well oriented as judged from their relatively weaker contributions in the dichroic spectra (Table 1).

Assuming that the membranes are perfectly ordered, the dichroic data were analyzed quantitatively by calculating the maximum tilt angle (Table 2) [20]. These results indicate that in the different lipid systems the gene 9 protein deviates by a maximum tilt angle of 27 to 33° from the membrane normal. In the presence of membrane disorder, this angle is reduced, so on the average the helix axis may be closer to the membrane normal. Until now we have assumed in the calculations of the orientation that the protein is fully α -helical. Therefore we also calculated the maximum tilt angle in the case that a fraction of 67% of the molecule is α -helical, as found in Table 1 for the DMPC/DMPG system. In this calculation it is assumed that the remaining protein part has no net orientation. This is a reasonable assumption, since turn structures hardly show up in the dichroic spectrum (Figure 5 and Table 1), and therefore are not well oriented. By this correction the dichroic ratios of the α -helix component (R_α), as given in Table 2, show a reduction of the maximum tilt angle by about 5° . This leads to a maximum tilt angle of approximately 28° .

Model of gene 9 protein in membranes and biological implications

Summarizing, we conclude from Table 1 that the gene 9 protein conformation is about 67% α -helical, and that the remaining part consists mainly of turn structure. The α -helical part adopts a preferentially transmembrane orientation with calculated maximum tilt angles as shown in Table 2. If we interpret this result assuming a single α -helical strand in the gene 9 protein, this implies that the helix has a length of about 21 residues and is maximally tilted up to 28° from the membrane normal. This value compares well with the results found for the single membrane spanning major coat protein of bacteriophage M13 [38]. Combining this model with hydrophobicity data (Figure 1), it appears that the helix is located at the N-terminus. This implies that Trp₁₅ is not at the interface, but is located more to the center of the membrane, and Arg₁₈ is positioned within the transmembrane helix. Similar to the lysine residues in gene 8 protein of bacteriophage M13 [39], Arg₁₈ might function as an anchor of one end of the helix. The turn structure must be located at the C-terminus, facing the aqueous phase.

An alternative interpretation can be made when the measured dichroic ratio is assumed to be the average of the dichroic ratios of two highly oriented helices within one protein: one transmembrane helix, and a second helix parallel to the membrane plane. The fraction (x) of transmembrane helix contribution to the total helix content, indicating the relative length of this helix, can be calculated from the experimental dichroic ratio and the dichroic ratios of perfectly parallel and perpendicular oriented helices (Table 2) [40]. This calculation results in a longer helix (about 15 residues) oriented transmembrane, and a shorter helix (about 6 residues) oriented parallel to the membrane surface.

This two-helix model agrees remarkably well with the prediction of a hydrophobic helix at the N-terminus, which shows one turn connecting the two helices, and another turn at the C-terminus (Figure 1). This model suggests favourable locations for Trp₁₅ and Gly₂₀: Trp₁₅ is located around the membrane interface, and that Gly₂₀ may be involved in a turn structure, which connects the transmembrane and amphipathic helix. However, it also implies that the transmembrane helix is about 15 residues long, which would be too short to fully span a membrane in an α -helical conformation. This problem can partly be relieved by having the amphipathic helix deeply inserted in the polar head group region,

anchoring in the hydrophobic region with its hydrophobic residues Phe₂₄ and Leu₂₇. In this case the helix may just reach the lipid carbonyl groups on the opposite site of the membrane by its polar formyl group at the N-terminus.

The present spectroscopic results cannot discriminate between the one-helix or the two-helix model. In comparing both models, it can be seen that the largest fraction of the helical region adopts a transmembrane orientation and a turn structure shows up at the C-terminus. The presence of turn structure at the C-terminus may be of importance for the biological function of the gene 9 protein, since turns frequently have been suggested as the bioactive conformation in recognition processes [41, 42]. This is supported by the finding that the C-terminal end plays an important role in the formation of the complex with the DNA hairpin loop of the packaging signal of the phage DNA [10].

References

- 1 Webster, R.E. and Lopez, J. (1985) in *Virus structure and assembly* (Casjens, S., ed.), pp. 235-267, Jones and Bartlett Publishers Inc., Boston MA.
- 2 Rasched, I. and Oberer, E. (1986) *Microbiol. Rev.* 50, 401-427.
- 3 Model, P. and Russel, M. (1988) in *The Bacteriophages* (Calender, R., ed.), Vol. 2, pp. 375-456, Plenum Press, New York.
- 4 Webster, R.E. (1996) in *Phage Display of peptides and proteins* (Kay, B., Winter, J. and McCafferty, J., eds.), pp. 1-20, Academic Press, San Diego.
- 5 Marvin, D. (1998) *Curr. Opinion in Struct. Biol.* 8, 150-158.
- 6 Russel, M. (1991) *Mol. Microbiol.* 5, 1607-1613.
- 7 Russel, M. (1995) *Trends Microbiol.* 3, 223-228.
- 8 Simons, G.F.M., Beintema, J., Duisterwinkel, F.J., Konings, R.N.H. and Schoenmakers, J.G.G. (1981) *Prog. Clin. Biol. Res.* 64, 401-411.
- 9 Endemann, H. and Model, P. (1995) *J. Mol. Biol.* 250, 496-506.
- 10 Russel, M. and Model, P. (1989) *J. Virol.* 63, 3284-3295.
- 11 Van Wezenbeek, P., Hulsebos, T. and Schoenmakers, J. (1980) *Gene* 11, 129-148.
- 12 Houbiers, M.C., Spruijt, R.B., Wolfs, C.J.A.M. and Hemminga, M.A. (1999) *Biochemistry* 38, 1128-1135.
- 13 Eisenberg, D., Weiss, R. and Terwilliger, T. (1984) *Proc. Natl. Acad. Sci.* 81, 140-144.
- 14 Chou, P. and Fasman, G. (1978) *Annu. Rev. Biochem* 47, 251-276.
- 15 Haris, P.I. and Chapman, D. (1995) *Biopolymers* 37, 251-263.
- 16 Peterson, G.L. (1977) *Anal. Biochem.* 83, 346-356.
- 17 Killian, J.A., Trouard, T.P., Greathouse, D.V., Chupin, V. and Lindblom, G. (1994) *FEBS Lett.* 348, 161-165.

- 18 Goormaghtigh, E., Cabiaux, V. and Ruyschaert, J.-M. (1994) in *Physicochemical Methods in the Study of Biomembranes* (Hilderson, H.J. and Ralston, G.B., eds.), Vol. 23, pp. 329-450, Plenum Press, New York.
- 19 Goormaghtigh, E. and Ruyschaert, J.-M. (1990) in *Molecular Description of Biological membranes by Computer Aided Conformational Analysis* (Brasseur, R., ed.), Vol. 1, pp. 285-329, CRC Press, Boca Raton, FL.
- 20 Bechinger, B., Ruyschaert, J.-M. and Goormaghtigh, E. (1999) *Biophys. J.* 76, 552-563.
- 21 Rothschild, K. and Clark, N. (1979) *Biophys. J.* 25, 473-487.
- 22 Goormaghtigh, E., Cabiaux, V. and Ruyschaert, J.-M. (1990) *Eur. J. Biochem.* 193, 409-420.
- 23 Kauppinen, J.K., Moffat, D.J., Mantsch, H.H. and Cameron, D.G. (1981) *Appl. Spectrosc.* 35, 271-276.
- 24 Byler, D. and Susi, H. (1986) *Biopolymers* 25, 469-487.
- 25 Surewicz, W.K. and Mantsch, H.H. (1988) *Biochim. Biophys. Acta* 952, 115-130.
- 26 Haris, P., Lee, D. and Chapman, D. (1986) *Biochim. Biophys. Acta* 874, 255-265.
- 27 Greenfield, N.J. and Fasman, G.D. (1969) *Biochemistry* 8, 4108-4116.
- 28 Bandekar, J. (1992) *Biochim. Biophys. Acta.* 1120, 123-143.
- 29 Mantsch, H., Perczel, A., Hollosi, M. and Fasman, G. (1993) *Biopolymers* 33, 201-207.
- 30 Wolkers, W.F., Haris, P.I., Pistorius, A.M.A., Chapman, D. and Hemminga, M.A. (1995) *Biochemistry* 34, 7825-7833.
- 31 de Jongh, H.H., Goormaghtigh, E. and Ruyschaert, J.-M. (1996) *Anal. Biochem.* 242, 95-103.
- 32 Haris, P., Coke, M. and Chapman, D. (1989) *Biochim. Biophys. Acta.* 995, 160-167.
- 33 Venyaminov, S. and Kalnin, N. (1990) *Biopolymers* 30, 1243-1257.
- 34 Fringeli, U. and Gunthard, H. (1981) in *Membrane Spectroscopy* (Verlag, S., ed.), Vol. 31, pp. 270-332, Grell, E, Berlin and Heidelberg.
- 35 Raussens, V., Narayanaswami, V., Goormaghtigh, E., Ryan, R.O. and Ruyschaert, J.M. (1995) *J. Biol. Chem.* 270, 12542-12547.
- 36 Raussens, V., Ruyschaert, J.-M. and Goormaghtigh, E. (1997) *J. Biol. Chem.* 272, 262-270.
- 37 Azpiazu, I., Gomez-Fernandez, J.C. and Chapman, D. (1993) *Biochemistry* 32, 10720-10726.
- 38 Glaubitz, C., Grobner, G. and Watts, A. (2000) *Biochem. Biophys. Acta* 1463, 151-161.
- 39 Stopar, D., Spruijt, R.B., Wolfs, C.J.A.M. and Hemminga, M.A. (1996) *Biochemistry* 35, 15467-15473.
- 40 Goormaghtigh, E., Raussens, V. and Ruyschaert, J.-M. (1999) *Biochim. Biophys. Acta* 1422, 105-185.
- 41 Smith, J. and Pease, L. (1980) *CRC. Crit. Rev. Biochem.* 8, 315-399.
- 42 Perczel, A. and Hollosi, M. (1996) in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G., ed.), pp. 285-379, Plenum Press, New York.
- 43 Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132.

Chapter 4

Spontaneous insertion of gene 9 minor coat protein of bacteriophage M13 in model membranes

M. Chantal Houbiers, Ruud B. Spruijt, Rudy A. Demel, Marcus A. Hemminga and
Cor J.A.M. Wolfs

Published in: Biochim. Biophys. Acta 1511 (2001), 309-316.

Abstract

Gene 9 minor coat protein from bacteriophage M13 is known to be located in the inner membrane after phage infection of *E. coli*. The way of insertion of this small protein (32 amino acids) into membranes is still unknown. Here we show that the protein is able to insert in monolayers. The limiting surface pressure of 35 mN/m for DOPC and DOPG lipid systems indicates that this spontaneous insertion can also occur *in vivo*. By carboxyfluorescein leakage experiments of vesicles it is demonstrated that protein monomers, or at least small aggregates, are more effective in releasing carboxyfluorescein than highly aggregated protein. The final orientation of the protein in the bilayer after insertion was addressed by Proteinase K digestion, thereby making use of the unique C-terminal location of the antigenic binding site. After insertion the C-terminus is still available for the enzymatic digestion, while the N-terminus is not. This leads to the overall conclusion that the protein is able to insert spontaneously into membranes without the need of any machinery or transmembrane gradient, with the positively charged C-terminus remaining on the outside. The orientation after insertion of gene 9 protein is in agreement with the "positive inside rule".

Introduction

Cleavable leader peptides or uncleaved signals can function to initiate translocation of proteins across membranes [1]. However, translocation of short domains with no or a small number of charged residues often does not require such signals. Examples of short proteins that translocate in this way include the Pf3 and M13 major coat proteins [2]. Whether gene 9 minor coat protein (g9p) of bacteriophage M13 is also able to translocate across membranes will be addressed in this paper.

G9p is a minor coat protein of bacteriophage M13. The amino acid sequence of this 32-residue protein (molecular weight is 3681 Da) is the following [3]:

HCO-Met-Ser-Val-Leu-Val-Tyr-Ser-Phe-Ala-Ser-Phe-Val-Leu-Gly-Trp-Cys-Leu-Arg⁺-Ser-Gly-Ile-Thr-Tyr-Phe-Thr-Arg⁺-Leu-Met-Glu⁻-Thr-Ser-Ser-COOH.

A few copies of the g9p are located together with g7ps at one end of this 880 nm long cylindrically shaped filamentous phage [4]. The main constituents of the bacteriophage are the single stranded DNA, surrounded all along the cylinder with approximately 2700 copies of g8ps, also known as major coat proteins. Five copies each of the minor coat proteins, which are the products of gene 3 and gene 6, cover the other end.

Bacteriophage M13 uses *E. coli* cells as a host, entering by the F-pilus. Coat proteins, as well as non-structural viral proteins, such as gene 1, gene 4, gene 11 proteins, and host proteins (thioredoxin) play a role in the assembly process of newly formed bacteriophages (for reviews see [5-9]). The gene 9 and gene 7 minor coat proteins were shown to interact with a hairpin loop in the viral DNA [10], thereby possibly playing an important role in initiating the assembly process. Because of this interaction with the packaging signal the phage end containing g9ps and g7ps is the first to emerge from the cell [11]. Prior to incorporation in the bacteriophage g9p is known to be located in the *E. coli* inner membrane [12], probably mainly α -helical, as has been demonstrated in model membranes [13].

The protein is synthesized and inserted into the inner membrane [12] without any signal sequence. The finding that the N-terminus of the g9p contains a formyl group, and therefore is not processed [14], gives rise to the idea that the protein must insert rapidly into the membrane, before proteolytic processing can occur. G9p does not contain a cleavable signal peptide. However, when looking at the sequence there are similarities with the general features of signal peptides, although the sequence is in a reversed order (positive charges are located C-terminally from the hydrophobic region). Similar to signal peptides [15, 16], g9p contains a hydrophobic stretch followed by a hydrophilic basic region with two arginine residues. The structure of signal peptides seems to be designed for interaction with lipids, as demonstrated from studies of the interaction with and translocation into phospholipid vesicles [17-19]. The N-terminal side of g9p is the most likely side to insert, since the N-terminus remains formylated after membrane insertion and is still formylated after assembly in the viral particle [14].

In the present paper we contribute to the model presented above showing the penetrative strength of g9p by monolayer and vesicle leakage experiments. The orientation of g9p, after reconstitution and insertion, is addressed by Proteinase K digestion, thereby making use of the C-terminal location of the antigenic site.

Materials and Methods

Materials

TFA was obtained from Merck (Darmstadt, Germany), and TFE was obtained from Acros (Pittsburgh, PA). DOPG and DOPC were obtained from Avanti Polar Lipids (Alabaster, AL). Proteinase K was from Boehringer (Mannheim, Germany). Alkaline phosphatase was from Boehringer (Mannheim, Germany). Polycarbonate filters (200 nm) were from Avestin (Ottawa, Canada). The mini-extruder device was from Milsch Equipment (Laudenbach, Germany). 5,6-Carboxyfluorescein was obtained from Kodak, and used after purification. The M13mp18rf template was from Pharmacia (Uppsala, Sweden). The Quikchange protocol was from Stratagene (La Jolla, CA). The g9p was prepared by solid phase synthetic techniques and characterized as described before [13]. The N-terminus was synthesized with a formyl group.

Solubilization of g9p in TFE

The protein was first dissolved in a small volume of TFA (10-20 μ l/mg of protein) and dried under a stream of nitrogen. Next, the protein was washed two times with TFE, followed by evaporation of the TFE under a stream of nitrogen, to remove residual TFA. Finally the protein was dissolved in a known volume of TFE. The protein concentration was determined by the procedure of Peterson [20] with bovine serum albumin as a standard.

Monolayer experiments

For the monolayer experiments, the pressure was measured by the Wilhelmy balance method using filter paper in a thermostatically controlled box, using a Cahn D202 microbalance [21]. This method is based on the determination of the weight of the filter paper, which is partially submersed in the aqueous sub phase. Differences in plate weight are directly related to surface pressure. The measurements were performed as follows. The lipid monolayers were spread from a chloroform solution to give an initial surface pressure between 15 and 40 mN/m on a sub phase solution of 10 mM Tris hydrochloride, 0.2 mM EDTA, pH 8.0 (TE8). During the measurements the temperature was kept at 22.5 ± 1 °C, and the sub phase was continuously stirred with a magnetic bar. Instead of a teflon dish a glass dish with a teflon coated rim was used, which allows more rigorous cleaning of the dish between the separate experiments. The glass dish had a volume of 4 ml and a surface of 4.41 cm². Protein, dissolved in 30 μ l of TFE/TE8 (1:1) was added through a hole on one side of the dish. Unless stated otherwise, a saturating amount of protein was injected into the sub phase. The pressure changes were followed until the surface pressure increase had reached a maximal value, usually within 30 minutes.

Reconstitution of g9p into lipid vesicles

To reconstitute the protein into lipid bilayers, desired amounts of lipids (in chloroform) and protein (in TFE) were mixed, dried under a stream of nitrogen and lyophilized overnight to remove residual solvent. The sample was rehydrated in the desired volume of buffer and freeze-thawed several times. To prepare LUVs the sample was extruded 31 times through a polycarbonate filter with 200 nm pores. The filters were mounted in a mini-extruder device as described in literature [22]. Vesicles without protein were prepared the same way.

Sucrose Gradient Centrifugation

Samples were brought on top of a linear 5-40% (w/w) sucrose gradient in 10 mM Tris, 0.2 mM EDTA, 150 mM NaCl, pH 8.0, and centrifuged for 18 hours at 110000g at 4 °C in a Beckman SW41 rotor. To enable visualization of the lipid-protein complexes, the gradient was fractionated and analyzed by gel electrophoresis and western blotting as described below.

Preparation of carboxyfluorescein loaded vesicles and determination of release of carboxyfluorescein

An amount of 12.5 μ mol of lipid (DOPC or DOPG), dissolved in chloroform, was dried under a stream of nitrogen followed by high vacuum for several hours. CF (121 mM) in 2 ml of a solution of 10 mM Tris hydrochloride, 0.2 mM EDTA, pH 8.0 was used to rehydrate the lipids. The solution was vortexed and subjected to several cycles of freezing in liquid nitrogen and thawing in hot water to form multilamellar liposomes. Subsequently, the vesicles were converted into large unilamellar vesicles (LUVs) by extrusion as described above. External CF was removed by passing the vesicles over a P10 column (1.5 x 20 cm), pre-equilibrated with 10 mM Tris hydrochloride, 0.2 mM EDTA, 50 mM NaCl, pH 8.0. After collection of the CF-loaded vesicles, they were diluted with the same buffer to give a maximal optical density at 490 nm between 0.09 and 0.1 in a 1-cm cuvette. Protein was added in a 1:1 mixture of TFE:buffer to 1 ml of the vesicle solution and stirred. The release of CF was followed in time on a Perkin-

Elmer LS-5 fluorescence spectrometer (excitation slit 5 nm, emission slit 5 nm) at an excitation wavelength of 490 nm, and an emission wavelength of 520 nm. The measurements were performed at room temperature. Maximal (100%) fluorescence was obtained by addition of 20 μ l of a 10% (v/v) of Triton X-100. To check for background leakage, an equal volume of solvent without protein was added. The L/P ratios of the samples were calculated from the known amount of protein added, and the phospholipid concentration determined by total fluorescence intensity of CF after addition of cholate (final concentration 50 mM) to release all CF from the vesicles. An average diameter of the vesicles of 200 nm is assumed.

Overexpression and mutagenesis of g9p

Overexpression of g9p was achieved by introducing M13 gene 9 into the pT7-7 vector [23]. By PCR, using M13mp18rf as a template, NdeI and BamHI sites were introduced respectively preceding and following the gene 9 protein-coding gene. After insertion in the respective sites of pT7-7, the plasmid was introduced in *E. coli* BL21(DE3) strain and expression of g9p was obtained in LB medium after induction by IPTG to a final concentration of 0.25 mM. Cells were harvested 1 hour after induction. The presence of g9p before and after induction was analysed as described below. Mutant g9ps were produced using the Quikchange protocol. By sequencing the plasmid DNA the presence of the mutation was checked.

Protein analysis by SDS Gel Electrophoresis and Western Blotting

To verify the presence of the g9p after overexpression, gel electrophoresis on a tricine SDS polyacrylamide gel was performed [24]. Western blotting [25] and immuno detection were used to verify the reaction with antibody, using Goat-anti-Rabbit IgG conjugated with alkaline phosphatase as the secondary antibody. Primary antibodies were raised against synthetic protein in female New Zealand white rabbits. Spot intensities were determined by image analysis after white light scanning of the blot, thereby using a calibration curve from a series created by blotting of known relative amounts of protein.

Cleavage of g9p by proteinase K

Samples of reconstituted g9p in 200 nm lipid vesicles, as well as the 200 nm lipid vesicles used in the insertion experiments, were produced as described above. For insertion, g9p, dissolved in a small volume of TFE, was added to the lipid vesicles and left at room temperature for 1 hour. 1 ml of the mixtures contained 1.6 μmol lipid and 3.2 nmol g9p, resulting in an L/P ratio of 500. Vesicles with protein attached were purified from non-interacting protein by sucrose gradient. The clear visible band was isolated and dialyzed for 16 hours against buffer. After dialysis Proteinase K was added, at a concentration of 50 $\mu\text{g/ml}$, and the samples were incubated at room temperature during 1 hour. The reaction was stopped by adding PMSF (final concentration 5 mM), an equal volume of gel electrophoresis sample buffer (double concentration) and immediately putting in a boiling water bath for 3 minutes. The samples were then further analyzed by gel electrophoresis and western blotting as described above.

Results

Monolayer experiments

The interaction of g9p with lipid monolayers was monitored by measurement of the surface pressure increase in a constant area setup. Since extrinsic proteins and other non-penetrative agents do not induce an increase of the monolayer surface pressure [26], this effect can be related to the degree and nature of penetration. The highest initial pressure allowing penetration, the limiting surface pressure, is often used as a measure of the penetrative force. The g9p is surface active itself, spontaneously forming a protein monolayer to a surface pressure of 23 mN/m after introduction into the sub phase. The injection of g9p under a phospholipid monolayer gave rise to a surface pressure increase, which reached a stable end level after 20-30 minutes. Injection of the solvent alone had a small effect on the surface pressure of 0.5-1.0 mN/m. The maximal pressure increase in DOPC is dependent on the amount of protein added in the sub phase up to a saturating level. For the negatively charged DOPG similar results were found (data not shown). The saturating level was determined (data not shown) and for further experiments an injected amount of 11.6 nmol of protein was chosen, which is sufficient

to saturate the monolayer and produce a stable end pressure within 30 minutes. Figure 1 shows the dependence of the final pressure increase on the initial pressure with a DOPC monolayer.

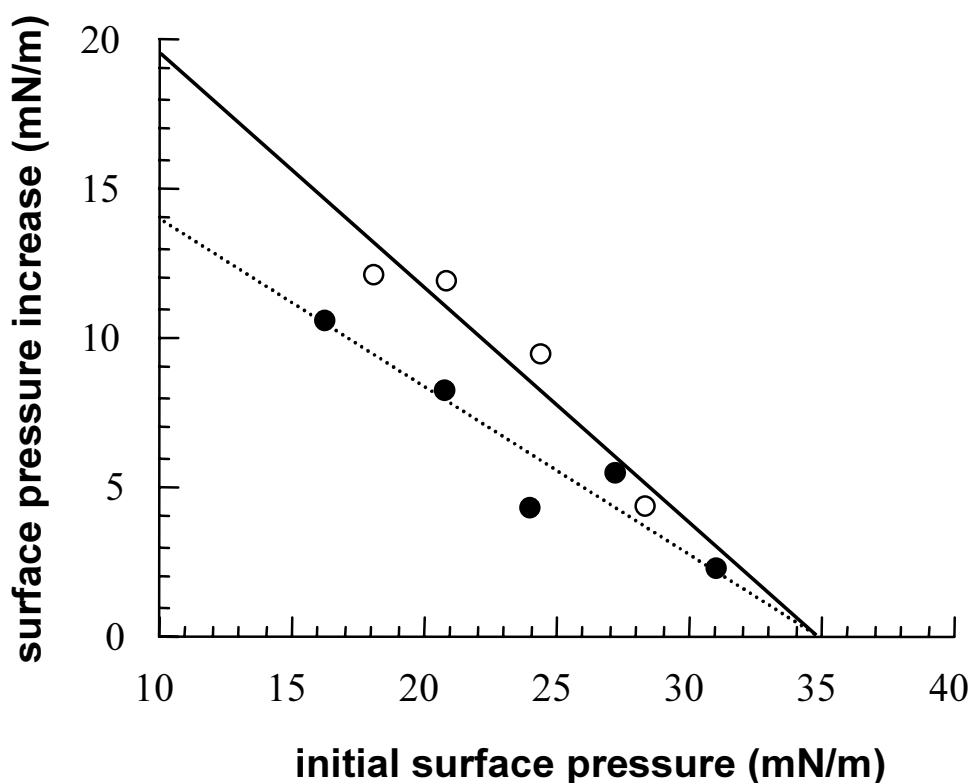


Figure 1: Surface pressure increase after injection of 11.6 nmol (resulting in a 2.9 μM concentration in the sub phase) *g9p* in the sub phase of monolayers of DOPC (filled circles, dotted line) and DOPG (open circles, solid line) as a function of the initial pressure.

A higher initial surface pressure results in a smaller pressure increase, indicating a more restricted penetration. From this figure, a limiting surface pressure of about 35 mN/m could be derived. Measurements performed with DOPG monolayers, also shown in Figure 1, resulted in the same limiting surface pressure of about 35 mN/m, although the slope of the curve was steeper.

Leakage of Carboxyfluorescein

In addition to the insertion of g9p into phospholipid monolayers the effect of the g9p on bilayers was investigated.

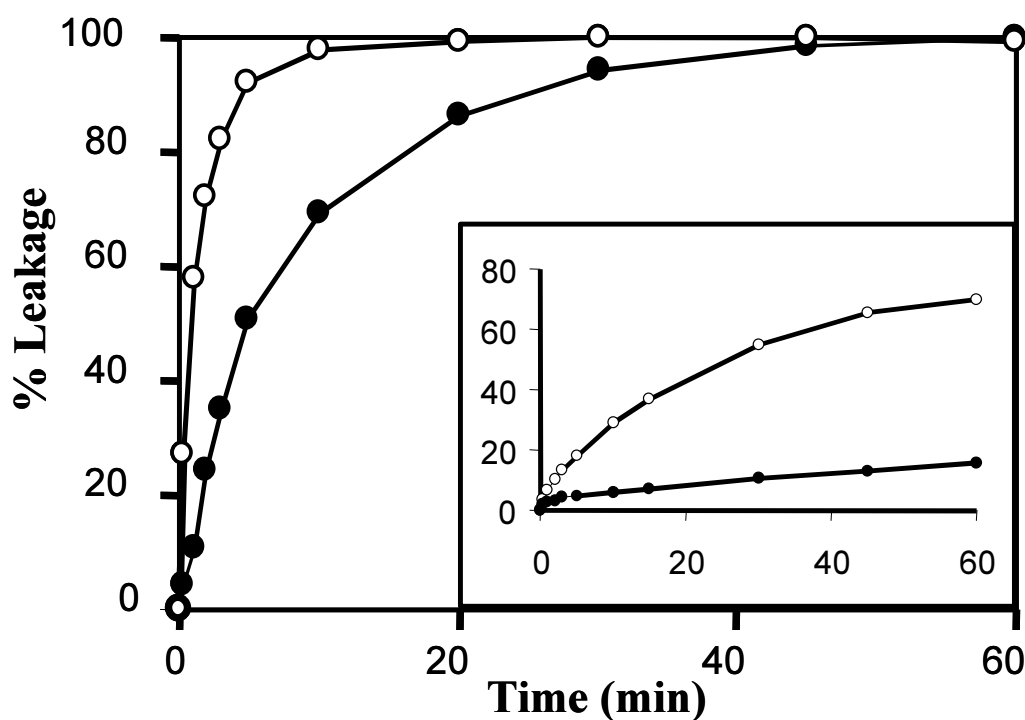


Figure 2: G9p (1 nmol) added in respectively 4 μ l (filled circles) and 40 μ l (open circles) TFE to CF-loaded DOPC at an L/P ratio of 25. Inset: G9p (0.8 nmol) added in respectively 50% (v/v) (open circles) and 20% (v/v) TFE/water (filled circles) to CF-loaded DOPC at an L/P ratio of 38.

Figure 2 shows the influence of the TFE/g9p ratio on the rate of leakage of CF-loaded 200 nm DOPC vesicles. The initial leakage rate is higher at a higher TFE/g9p ratio. No leakage was observed upon addition of TFE alone. Decreasing the percentage TFE (v/v) in the protein solution, before adding, also decreases the leakage rate as shown in Figure 2 (inset).

Location of the antigenic site

The antigenic site of the g9p was predicted to be located at the C-terminal side (Figure 3A).

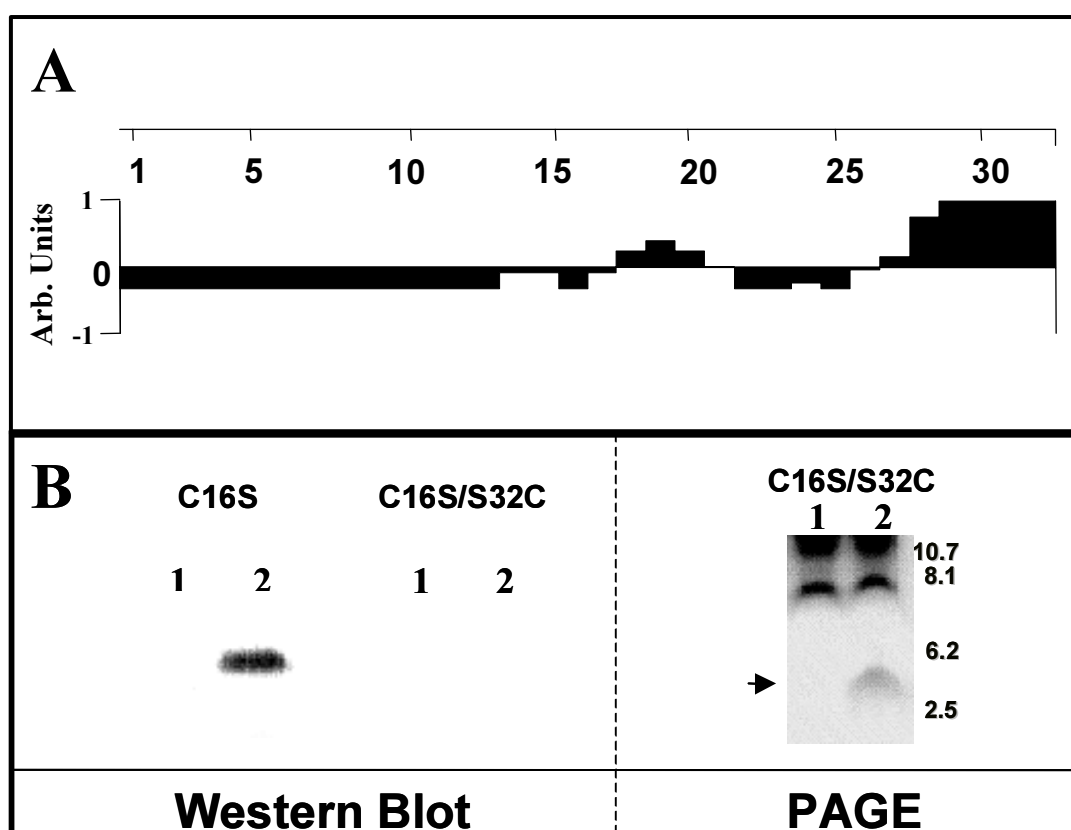


Figure 3: (A) Antigenic index prediction for g9p, according to Jameson-Wolf [32]. (B) On the left side is a western blot of the C16S and C16S/S32C mutated proteins before (lane 1) and after overexpression (lane 2). On the right side is a protein coloured SDS tricine gel. Lane 1 shows the situation before and lane 2 after overexpression of the mutant protein C16S/S32C. Marker protein positions are also indicated (kDa). The arrow indicates the position of the g9p.

We found evidence for this, based on the expression and detection of two mutant proteins. In both mutant proteins the Cys₁₆ was changed to Ser. Additionally, one mutant protein also contained the Ser₃₂ to Cys mutation (C16S/S32C). Although both mutant proteins were present after overexpression, as shown for the C16S/S32C mutant protein by protein coloured SDS gel electrophoresis (Figure 3B), only the protein with the C16S

mutation is detectable by western blotting (Figure 3B). This indicates that changing Ser₃₂ to Cys completely abolishes antibody binding to the protein. This confirms the C-terminal location of the antigenic site.

Proteinase K digestion experiments

The orientation of g9p in membrane vesicles after insertion was addressed by Proteinase K digestion experiments. Cleavage of the C-terminal side, which contains the antigenic site, was followed by the disappearance of intensity on a western blot. In these experiments vesicles with inserted g9p were compared with vesicles in which g9p was reconstituted. The experiment was performed with 200 nm vesicles consisting of DOPC or DOPG. After incubation, sucrose density centrifugation followed by western blot analysis was performed. Comparing the initial blot intensities of the inserted and reconstituted system before digestion shows that the efficiency for insertion is low. Not inserted protein, which is most likely aggregated, ended up at the bottom of the tube after centrifugation (data not shown). Figure 4 shows the western blot of the samples before and after addition of Proteinase K to sucrose gradient purified g9p containing DOPC and DOPG vesicles. In comparing DOPC and DOPG, it can be seen that after reconstitution about 69% of the protein in DOPC and 63% of the protein in DOPG can not be cleaved during the time of the experiment. After insertion of g9p in DOPG and DOPC respectively 12% and 37 % of the initial amount of protein was left after digestion.

Discussion

Interaction with lipids

Monolayer measurements show that g9p has the ability to penetrate into lipid monolayers of DOPC and DOPG. Figure 1 shows that upon addition of protein to a lipid monolayer the surface pressure increases. Protein molecules that are known to interact only with the phospholipid head group do not induce such an increase in surface pressure [27]. Therefore this increase can be interpreted as the result of protein insertion in the hydrophobic part of the monolayer. We found for monolayers of DOPC as well as

DOPG a limiting surface pressure of about 35 mN/m. The limiting pressure value is indicative of the potency of the protein to penetrate into membranes. The measured limiting pressure for the g9p is in the range of pressures assumed for biological membranes, for which an "equivalence pressure" of 25 to 35 mN/m has been estimated [28]. This indicates that the ability of the g9p to insert into lipid monolayers might as well occur *in vivo*. The limiting surface pressure is comparable with values found for other penetrating peptides, such as PhoE and M13 major coat protein signal sequences [29].

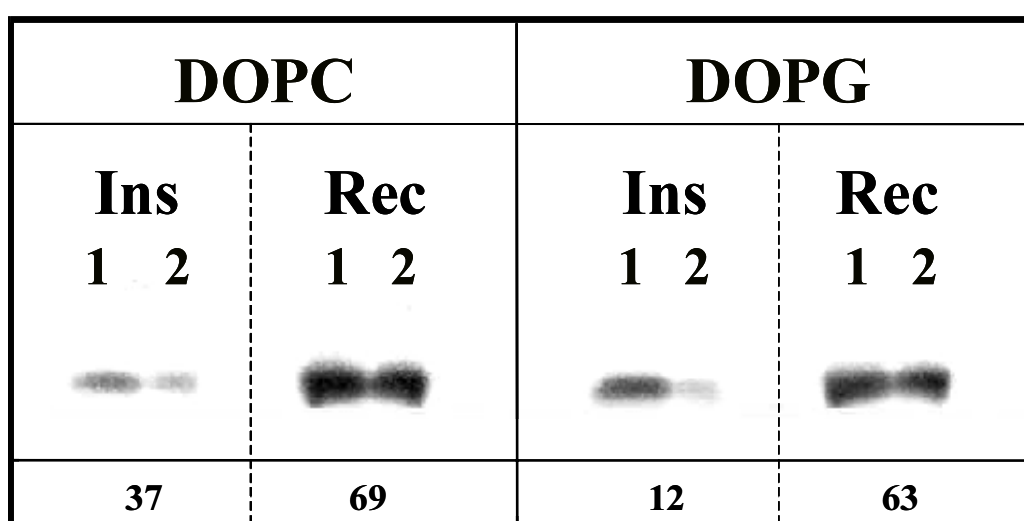


Figure 4: Western blot of Proteinase K digestion experiments of g9p incorporated in vesicles by reconstitution (Rec), or incorporated in vesicles by spontaneous insertion (Ins) after sucrose gradient purification. The left four lanes represent DOPC, the right four lanes represent DOPG. In each lipid system the left two lanes represent inserted (Ins) g9p, and the right two lanes represent reconstituted (Rec) g9p. The samples without Proteinase K are marked by 1, the same samples after incubation with Proteinase K are marked by 2. The figures at the bottom indicate the percentage intensity left at the spots after Proteinase K digestion.

With either lipid, the limiting surface pressure is about 35 mN/m. Apparently, the insertion is mainly based on hydrophobic forces, and less on electrostatic attraction. Although these experiments show insertion into monolayers, they do not provide information on how much and what region of the protein inserts. Nevertheless, it can be

concluded from these monolayer experiments that after binding to the membrane the protein spontaneously inserts into the membrane.

Since the monolayer is only half the thickness of the bilayer and therefore cannot represent the actual insertion process, we also studied the interaction with bilayers. Information on protein penetration into bilayers was obtained from CF leakage experiments (Figure 2). Although the mechanism by which leakage is induced is unknown, from Figure 2 it can be seen that the initial leakage rate increases with increasing TFE/protein ratios. Because TFE has a solubilizing effect on the protein, this suggests that smaller protein aggregates and monomers are more effective for leakage as compared to large protein aggregates. These experiments show that g9p is able to interact with bilayers similar as in the monolayer experiments. This finding also agrees with the insertion experiments in Figure 4, which show that the gene 9 protein is able to integrate with the membrane.

Orientation after insertion

The finding that the antigenic site is located at the C-terminal end enabled us to specifically demonstrate cleavage in the C-terminal region with the help of antibodies. For the reconstituted g9p the protein is expected to be randomly oriented, so that 50% of the C-termini is predicted on the outside of the vesicle. This fairly well agrees with the 37% reduction of the spot intensity, observed for DOPG, on the western blot. Thus, roughly half of the C-termini are accessible for Proteinase K. For inserted protein, however, a unique orientation is expected if the N-terminal side penetrates the membrane, leaving the C-terminus out. In this case, all antigenic sites can be cleaved, leaving no intensity on blot. If it inserts in the opposite way, no antigenic site will be cleaved off, resulting in no reduction of the intensity on blot. The band on the western blot of inserted g9p in DOPG disappears almost completely, leaving only 12% of the initial intensity after incubation with Proteinase K (Figure 4). This indicates that almost all C-terminal ends were accessible and therefore located outside the vesicle. The band positions of reconstituted g9p before and after cleavage are the same. Therefore, the size of the g9p that retains the C-terminus equals the size of the uncleaved protein. This indicates that the N-terminal region is unaffected by Proteinase K, since a size reduction

in molecular weight is expected if (part of) the N-terminal would be clipped off. Therefore, it is concluded that the N-terminus is inserted into the membrane. Comparison of the reconstituted with the inserted g9p, leads to the conclusion that the protein inserts its N-terminal part into the DOPG membrane, whereas the C-terminal region is exposed to the aqueous vesicle exterior. In the case of DOPC vesicles, the difference between insertion and reconstitution is less apparent, since after insertion not all antigenic sites can be cleaved, leaving a clear band (37 %). This effect is probably due to steric hindrance as a result of vesicle aggregation. This finding is supported by the observation that addition of only a low mol percentage of DOPG, thereby introducing vesicle repulsion, resulted in a much better digestion (data not shown).

In conclusion we believe that the g9p inserts with a specific orientation in phospholipid vesicles, namely with the N-terminus inserted in the membrane bilayer and the C-terminus remaining at the exterior. Insertion into monolayers occurs at pressures that are of biological significance, thereby allowing extrapolating to the *in vivo* situation in the *E. coli* inner membrane. *In vivo* this would mean that the C-terminus retains in the cytoplasm. This orientation agrees with what would be expected based on the finding that the N-terminal formyl group is protected in the *E. coli* membrane [14]. Furthermore, the topology of the protein is in agreement with the "positive inside rule" [30, 31] and the role of the protein during phage assembly [10].

References

- 1 Cao, G., Cheng, S., Whitley, P., von Heijne, G., Kuhn, A. and Dalbey, R.E. (1994) J. Biol. Chem 269, 26898-903.
- 2 Kuhn, A. (1995) FEMS Microbiol. Reviews 17, 185-90.
- 3 Van Wezenbeek, P.M.G.F., Hulsebos, T.J.M. and Schoenmakers, J.G.G. (1980) Gene 11, 129-48.
- 4 Grant, R.A., Lin, T.-C., Konigsberg, W. and Webster, R.E. (1981) J. Biol. Chem. 256, 539-46.
- 5 Rasched, I. and Oberer, E. (1986) Microbiol. Rev. 50, 401-427.
- 6 Model, P. and Russel, M. (1988) Filamentous Bacteriophage, Plenum Press, New York.
- 7 Webster, R.E. (1996) in Phage Display of Peptides and Proteins (Kay, B.K., Winter, J. and McCafferty, J., eds.), pp. 1-20, Academic Press Inc, San Diego.
- 8 Marvin, D. (1998) Curr. opinion Struct. Biol. 8, 150-158.
- 9 Russel, M. (1995) Trends Microbiol. 3, 223-8.

- 10 Russel, M. and Model, P. (1989) *J. Virol.* 63, 3284-95.
- 11 Lopez, J. and Webster, R.E. (1983) *Virology* 127, 177-93.
- 12 Endemann, H. and Model, P. (1995) *J. Mol. Biol.* 250, 496-506.
- 13 Houbiers, M.C., Spruijt, R.B., Wolfs, C.J.A.M. and Hemminga, M.A. (1999) *Biochemistry* 38, 1128-1135.
- 14 Simons, G.F.M., Konings, R.N.H. and Schoenmakers, J.G.G. (1981) *Proc. Natl. Acad. Sci. U. S. A.* 78, 4194-8.
- 15 Perlman, D. and Halvorson, H.O. (1983) *J. Mol. Biol.* 167, 391-409.
- 16 Von Heijne, G. (1984) *EMBO J.* 3, 2315-8.
- 17 Geller, B.L. and Wickner, W. (1985) *J. Biol. Chem.* 260, 13281-5.
- 18 Ohno Iwashita, Y. and Wickner, W. (1983) *J. Biol. Chem.* 258, 1895-900.
- 19 Shinnar, A.E. and Kaiser, E.T. (1984) *J. Am. Chem. Soc.* 106, 5006-7.
- 20 Peterson, G.L. (1977) *Anal. Biochem.* 83, 346-56.
- 21 Demel, R.A. (1994) in *Subcellular Biochemistry, Physicochemical Methods in the Study of Biomembranes* (Hilderson, H.J. and Ralston, G.B., eds.), pp. 83-120, Plenum Press, New York.
- 22 MacDonald, R.C., MacDonald, R.I., Menco, B.P., Takeshita, K., Subbarao, N.K. and Hu, L.R. (1991) *Biochim-Biophys-Acta* 1061, 297-303.
- 23 Studier, F.W. and Moffatt, B.A. (1986) *J. Mol. Biol.* 189, 113-130.
- 24 Schagger, H. and Von Jagow, G. (1987) *Anal. Biochem* 166, 368-79.
- 25 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A laboratory manual*, Cold Spring Harbor Laboratory, New York.
- 26 Demel, R.A., London, Y., Geurts van Kessel, W.S., Vossenbergh, F.G. and Deenen, L.L.v. (1973) *Biochim Biophys Acta* 311, 507-19.
- 27 Demel, R. (1974) *Methods in Enzymology* 32, 539-544.
- 28 Demel, R.A., Geurts van Kessel, W.S., Zwaal, R.F., Roelofsen, B. and van Deenen, L.L. (1975) *Biochim. Biophys. Acta* 406, 97-107.
- 29 Batenburg, A.M., Demel, R.A., Verkleij, A.J. and De Kruijff, B. (1988) *Biochemistry* 27, 5678-5685.
- 30 Von Heijne, G. (1986) *EMBO J.* 5, 3021-3027.
- 31 Von Heijne, G. (1988) *Biochim. Biophys. Acta* 947, 307-33.
- 32 Jameson, B.A. and Wolf, H. (1988) *CABIOS* 4, 181-6.

Chapter 5

Summarizing Discussion

Introduction and outline

Gene 9 minor coat protein of bacteriophage M13, present in the *E. coli* inner membrane prior to incorporation bacteriophage [1], has been assumed to reside in the membrane as a helical structure with the charged C-terminal end in the cytoplasm [2]. However, there is no direct experimental support for the structure and topology of this protein in the membrane. Also spontaneous insertion of the protein into the membrane was suggested [3], but not proven experimentally. The objective of this work, therefore, was to contribute experimental results to verify these assumptions. In this work, we aimed to characterize the g9p in the membrane-bound state, prior to assembly. For this we incorporated and studied the protein in model membrane systems. This final chapter summarizes and discusses the results that were obtained, and an attempt is made to indicate the biological relevance of these results.

Gene 9 protein, a sticky protein

G9p, which was already predicted to be hydrophobic by hydropathy analysis (Chapter 3), appeared to be difficult to handle due to its hydrophobic properties. In all experiments, precaution had to be taken to avoid aggregation and it appeared difficult to recover this "sticky protein" from columns, monolayer dishes, and grids for mass spectrometry. In order to get this under control, we looked for suitable solvents and we characterized the aggregation properties of the g9p. The organic solvent TFE was found to be a good solubilizer for the protein. Detergent solutions, used as elution media in HPSEC to characterize the aggregation state of the protein, required precaution. In a weak detergent, like cholate, the g9p appeared to aggregate, which was concomitant with a conformational change from α -helix to β -sheet (Chapter 2). The strong detergent SDS was able to restore the initial non-aggregated state of the protein. Phospholipids appeared to be unable to change the aggregational state. Thus, upon incorporation of the protein into phospholipids from cholate, the aggregational state of the protein was preserved. When using cholate in the procedure of reconstitution, care had to be taken to avoid aggregation in cholate. Another method of reconstitution, starting by cosolubilizing protein and lipids in a TFE/water mixture [4], kept the protein in the α -helical non-aggregated state. From this characterization of the aggregational proper-

ties of the protein we were able to proceed further to produce reproducible protein-lipid systems, which were essential to study the protein in a lipid environment.

Towards a structural model for g9p in the membrane

To obtain secondary structure information of the g9p in model membranes, we performed CD and FTIR measurements (Chapter 2 and 3). G9p was found to be predominantly α -helical when reconstituted properly into phospholipid membranes. A more detailed analysis by quantification of secondary structure elements from the amide I band, obtained by FTIR measurement, revealed the presence of a substantial amount of turn structure besides an α -helical structure (Chapter 3). The presence of β -sheet was excluded. We found that about 67% of g9p consists of an α -helix conformation, and that there may be two turns present. In addition, we incorporated g9p into oriented bilayers. By measuring dichroic spectra of these bilayers with ATR FTIR, we determined that the α -helix of g9p is oriented preferentially parallel to the membrane normal. Thus, g9p adopts a mainly transmembrane orientation. The turn structure is less well oriented. Quantitative analysis of the dichroic data yielded an estimate of the maximum tilt angle of the helix from the membrane normal. This angle appeared to be around 28° , assuming the presence of 67% α -helix and perfect membrane order. In the case of membrane disorder, this angle becomes smaller, so the helix may be oriented closer to the membrane normal. These results, indicating that the α -helix is about 21 residues long and maximally tilted up to 28° from the membrane normal, are comparable with the results found for the major coat protein [5]. However, the results obtained by ATR FTIR cannot discriminate the average orientation of several helices (within one protein) from the orientation of a single helix. So the calculated tilt angle may as well be the result of a combination of two helices: one oriented transmembrane and another oriented parallel to the membrane plane. To consider this possibility, we calculated the relative lengths of such helices. We found that one transmembrane helix of about 15 residues and a second helix, oriented parallel to the membrane plane, of about 6 residues can also account for the observed results.

Since the orientation results cannot discriminate between one helix and the average of more helices, we present here two possible models for g9p: a one-helix model and a

two-helix model. These are tentative models, based on our results as well as on prediction data (Chapter 3, Figure 1). Hydrophobicity analysis predicts a hydrophobic helix to be located at the N-terminus. The C-terminal region is predicted to have an amphipathic character. In addition two turns are predicted: one around residue 17-20, the other at the C-terminal end from residue 29-32. These prediction results agree well with our experimental results. Further information comes from the proteinase cleavage experiments, which show that the C-terminal end must be accessible for proteinase K and thus must reach the outside of the membrane. These experiments also show that the N-terminal side is not accessible and therefore should be fully buried in the membrane. Combining all this information, we propose two possibilities for the structure and position of g9p in the membrane, as depicted in Figure 1.

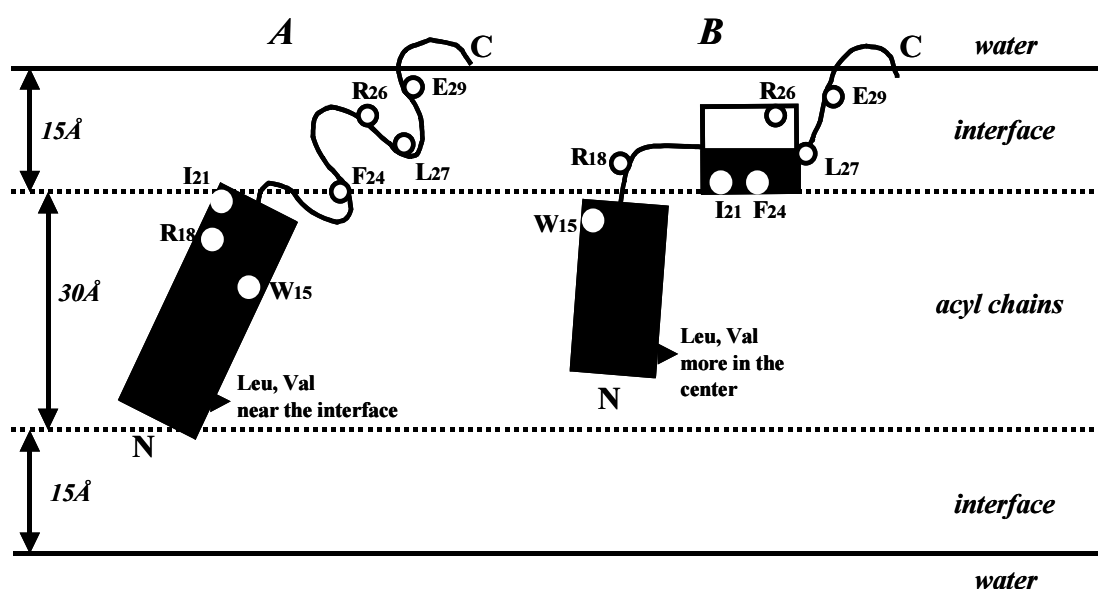


Figure 1: Possible schematic models of g9p in a membrane. The size of the DOPC bilayer is according to the structure as described by White & Wimley [6], which is depicted in Figure 1 of Chapter 1. The bars represent α -helical structure, where the black area represents hydrophobic domains and the white area represents hydrophilic domain. The width of the α -helix represents the size of the backbone as well as the side chains. (A) One-helix model, (B) Two-helix model.

In the one-helix model in Figure 1 the N-terminal helix is tilted with respect to the membrane normal. The helix contains very hydrophobic side chains as (iso)leucines and

valines, positioned in the center as well as near the edges of the hydrophobic core region. Trp₁₅, known to have a preference for a small girdle in the membrane-water interface region [7, 8], ends up more towards the center of the hydrophobic core, which might not be the most favourable position. Also the charged Arg₁₈ is positioned rather deep in the hydrophobic core. It is known that the positively charged group of this residue, attached to the backbone chain by several carbon atoms, can 'snorkle' towards the interface region [8]. Following the helical part is a stretch mainly consisting of turn structure, in which hydrophobic side chains (like Phe₂₄ and Leu₂₇) are likely to be more positioned towards or in the hydrophobic core region, and the positively charged Arg₂₆ and hydroxyl-bearing side chains (Thr and Ser) are likely to be in the interfacial region. The negatively charged Glu₂₉ is probably near the water phase, since negative charges are said to be repelled by the negatively charged phosphate groups [8]. The three C-terminal amino acids are placed near the water phase, since they bear hydroxyl groups that can form hydrogen bonds, and because cleavage experiments indicated that this region is accessible for proteinase K (Chapter 4).

In the two-helix model (Figure 1) the same kind of interactions are present. However, because the N-terminal transmembrane helix is shorter, Trp₁₅ can be positioned more favourably near the membrane-water interface region, where it may serve as an anchor [7, 8]. The very hydrophobic side chains (Leu, Val) in the transmembrane helix are now more towards the center of the hydrophobic core region. These side chains are known to be most prevalent in the middle of the bilayer [9]. The reason might be that in the center part the lipid acyl chains are more flexible and therefore might accommodate these bulky side chains more easily. The formylated Met₁, and Ser₂ might function to anchor the N-terminal side of the helix towards the interfacial region. However, it is likely that Trp₁₅ anchors the helix more strongly in the opposite interfacial region. Tryptophan has been characterized to be a strong anchor in the interfacial region, which involves that it wins the competition of other weaker anchors [10]. Thus, the N-terminal side might not completely reach the opposite interfacial region. Progressing C-terminally, the helix is followed by a hinge region (turn structure) that might be positioned by the opposing effects of hydrophobic and hydrophilic side chains: Leu₁₇ may anchor in the hydrophobic phase, while Arg₁₈ may anchor in the interfacial region. After Gly₂₀, which might serve as a helix breaker, a second short helix is depicted parallel to the membrane

plane. This helix is clearly amphipathic with hydrophobic side chains (Ile₂₁, Phe₂₄, Leu₂₇) on the side of the hydrophobic core, and with more hydrophilic side chains (Thr₂₂, Tyr₂₃, Thr₂₅) in the interfacial region. Arg₂₆ is positioned most close to the water side. This short helix is followed by another turn region, where Glu₂₉ can be positioned favourably near the water phase as well as the remaining three hydroxyl-bearing residues (see above as in one-helix model).

In summary: It is possible to construct models for g9p in a membrane, which agree with our experimental results and which also fulfill general principles known to govern the positioning of protein domains in a membrane, as for example has been demonstrated for the gene 8 major coat protein [11-13]. These general principles, as reviewed by Killian and von Heijne [8], are mainly based on anchoring interactions of specific residues to specific membrane regions (e.g. Trp in the aromatic girdle, charged residues in the interfacial region, hydrophobic residues in the hydrophobic core), possibly accompanied by a 'snorkling' effect when the backbone part of a residue is not in the suitable environment. From these principles the two-helix model seems somewhat more favourable than the one-helix model. The two-helix model shows remarkable resemblance with the model that is proposed for the g8p of bacteriophage M13. This protein is present in the membrane as a transmembrane helix, linked by a hinge region to an amphipathic region that is more in line with the plane of the membrane [11, 14-16]. This seems to be a common motive in coat proteins of filamentous bacteriophages [14, 17-19].

One point that remains to be addressed about the two-helix model is that the transmembrane helix (approximately 15 residues) is too short to fully span the hydrophobic membrane core. However, the N-terminal side does not contain charged or very polar residues, so it might remain favourably buried in the hydrophobic membrane core. In general, when a transmembrane helix is too short to span the membrane (hydrophobic mismatch) reactions of the membrane are induced to avoid this unfavourable situation [20]. As suggested by de Planck et al., mismatch of strong membrane anchors (Trp, Tyr, Lys) will result in a drastic membrane adaptation (e.g. formation of hexagonal phase), whereas weak anchors (Phe, Arg) will only have minor effects on the membrane [21]. For g9p, where Trp, Tyr, Phe and Arg are present,

membrane adaptation might occur in some cases. Our observation that g9p can cause leakage of membranes (Chapter 4) may agree with membrane disturbing properties of g9p. One can speculate that these properties might serve a function at some stage of the phage assembly process.

It should be stressed that these one- and two-helix models for the g9p are tentative models, based on our results and what is known from literature. It is possible that also other models agree. Perhaps both proposed models are possible, in that the amphipathic helix of the two-helix model might be flexibly able to move towards a more transmembrane orientation, resulting in a model more alike the one-helix model. For g8p such a model has been proposed [22].

A remarkable feature of the g9p is the high number of aromatic residues: with one tryptophane, three phenylalanines and two tyrosines 19% of all residues are aromatic. What the function is of so many aromatic residues for g9p remains unsolved. G8p has similar aromatic residues, comprising 12% of all amino acids. One function of these is to provide protein-protein interactions in the coat of intact phage by stacking interactions with aromatic residues of neighbouring proteins, as has been proposed after structure analysis of intact phage by X-ray [23].

Another intriguing question that remains unsolved is why there are so many hydroxyl groups in the amino acid side chains (Ser, Thr, Tyr), also in the transmembrane part. Their function is still unclear. Gray and Matthews [24] found that these polar residues can be present in a hydrophobic membrane environment because serine and threonine have the ability to form an intrahelical hydrogen bond with their backbone i-3 or i-4 oxygen atom, thus satisfying the hydrogen bonding potential of the side chains. It was shown that this can promote bends in the helix backbone, possibly facilitating conformational changes in membrane proteins [25]. Another possible function of these residues is that motifs of several serines and threonines can drive strong helix-helix interactions in a cooperative way by interhelical hydrogen bonding of these side chains [26]. Thus, these polar residues might play a role in promoting transmembrane helix interactions.

Spontaneous insertion of gene 9 protein

G9p appeared to be able to penetrate phospholipid monolayers at biologically relevant lateral membrane pressures. Upon addition to dye-loaded vesicles, g9p was also able to induce vesicle leakage. When g9p was added externally to DOPG vesicles, externally added proteinase K was able to degrade most C-terminal ends. The N-terminal ends remained unaffected, indicating that this protein part is protected in the membrane and therefore inaccessible for proteinase K. When proteinase K was added to DOPG vesicles in which g9p was already incorporated by reconstitution (and therefore randomly oriented), more than 50% of the C-terminal ends remained intact.

These observations (Chapter 4) indicate that g9p is able to insert spontaneously in phospholipid monolayers and bilayers, without the presence and aid of other proteins or the proton motive force. Moreover, these results indicate that after insertion g9p adopts an orientation with the N-terminus inserted in the vesicle membrane and the C-terminus outside (N_{in} - C_{out}). This is in agreement with the fact that the N-terminus of g9p *in vivo* remains formylated [3]. The finding that the N-terminal side of g9p inserts in the vesicle membrane and the positively charged C-terminal side remains outside the vesicle agrees with the positive inside rule [27], which states that the non-translocated part of a protein is enriched in positive charges.

Most proteins need other proteins, as chaperones and a secretion machinery, for insertion and translocation (Chapter 1). Only few exceptions are thought to insert spontaneously without the help of other proteins. This appears to be restricted to short polypeptides up to 38 amino acids [28]. Also Pf3 mutant coat proteins have been shown to insert into a bilayer without the help of any other membrane protein [29]. Like the short polypeptides that are able to insert spontaneously [28], g9p is also a very short polypeptide with a hydrophobic region, so it is well possible it does not need any other proteins to insert. In addition, it seems well probable that no proton motive force is needed for insertion of the N-terminal side, as was shown for Pf3 mutant protein [30]. For M13 coat protein and leader peptidase it was suggested that the electrical component of the proton motive force pulls the acidic protein region electrophoretically

across the membrane [31, 32]. Since the g9p has no acidic residues on the N-terminal side, it might not need the proton motive force for insertion.

The difference in the protein degradation effect between vesicles with reconstituted g9p and vesicles with inserted g9p, observed upon addition of proteinase K, was less pronounced with DOPC vesicles than with DOPG vesicles. With DOPG vesicles nearly 90% of the inserted g9p could be cleaved at the C-termini, indicating that the C-termini are accessible and thus located at the vesicle exterior. With DOPC a higher fraction of the C-termini of inserted g9p remained intact (63% of the C-termini were cleaved). This decreased cleavage effect was explained by steric hindrance due to DOPC vesicle aggregation upon addition of g9p (Chapter 4). This decreased cleavage effect could be restored by addition of small amounts of DOPG, introducing vesicle repulsion by the negatively charged headgroups. However, an alternative explanation is also possible, which was not mentioned in Chapter 4: It is possible that the insertion process and the adoption of a N_{in} - C_{out} orientation is less efficient in DOPC vesicles than in DOPG vesicles due to a lack of anionic lipids. Negatively charged PG head groups may play a role in the insertion and may determine the orientation of g9p in a similar way as has been demonstrated for M13 major coat protein [33] and constructs of LEP [34]. For a review on the role of anionic lipids in protein insertion and translocation in bacterial membranes, see [35]. In principle, both the N-terminus and C-terminus can partition into the membrane. But if the positive charges of the C-terminus interact with PG headgroups they will remain there and only the N-terminus will insert. The absence of PG may therefore explain the passage of positive charges into the membrane. For g9p, the absence of PG in the vesicles may thus explain a deviation from the N_{in} - C_{out} topology.

In analogy with earlier proposed mechanisms as summarized by Dalbey and Kuhn [36] we propose the following mechanism for the insertion of g9p. The first step is binding of the protein to the membrane. This may be accomplished by hydrophobic (for both DOPG and DOPC vesicles) as well as by electrostatic interactions between the basic C-terminal residues and the acidic phospholipid head groups (in the case of DOPG vesicles). The second step is partitioning of the hydrophobic region into the lipid bilayer, driven by hydrophobic forces [31]. The anionic DOPG phospholipid may

contribute as a determinant to orient the protein by holding the positive charges of the protein in the interface region.

Biological implications

Our results from experiments with simplified model systems cannot be directly translated to the *in vivo* system. The *in vivo* membrane is much more complex and many other proteins are present, e.g. translocation machineries and the bacteriophage assembly site. However, simplified model systems offer the possibility to study g9p properties unambiguously in detail and provide new information. Therefore, such studies may contribute to new clues about the *in vivo* situation. Especially when the results agree with other pieces of information already obtained from the *in vivo* system, they can contribute to a better understanding of the role of g9p *in vivo*. In this section an attempt will be made to indicate the relevance of our findings for the *in vivo* situation.

Our studies showed that g9p is predominantly α -helical with some turn structure. The α -helical conformation seems also likely to be present *in vivo*, since g9p has been shown to reside in the inner membrane prior to assembly [1] and since inner membrane proteins are expected to be α -helical. The α -helical conformation might serve as well in different stages of the phage life cycle. A model for intact phage was proposed in which g7ps and g9ps were assumed to be α -helices, interacting with each other by their hydrophobic domains to cover the end of the phage particle [37].

Our conclusions that g9p can insert spontaneously and adopts a mainly transmembrane orientation in the vesicle membrane with the N-terminus in the membrane and the C-terminus outside may also be relevant for the *in vivo* situation. This is supported by the finding that the N-terminal formyl group is protected in the *E. coli* membrane [3], which also suggests rapid insertion of the N-terminus. However, additional experiments will be needed to check whether g9p also inserts spontaneously *in vivo*, without the help of other proteins. Translation of our topological results to the situation in the *E. coli* membrane means that the positively charged C-terminal end of g9p would remain in the cytoplasm. Moreover, the C-terminal end, containing turn structure and a positive charge, seems a good candidate to interact with the viral DNA when it is present in the

cytoplasm. This agrees with the finding that the C-terminal end of g9p plays a role in the formation of a complex with the DNA hairpin loop of the packaging signal of the phage DNA [38]. In addition, turn structure has frequently been suggested to be active in recognition processes [39, 40].

Our proposed conformation and positioning of g9p in the membrane show similarities with g8p. G8p is known to be incorporated in intact phage with the positively charged C-terminal side interacting with the DNA, while the hydrophobic domains interact with neighbouring coat proteins. A similar mechanism of incorporation in intact phage would be possible for g9p if it is residing in the membrane with the C-terminus sticking out in the cytoplasm.

In conclusion, our findings in model membranes regarding the conformation and topology of g9p provide consistent insights with other findings about g9p *in vivo*. This indicates that our results are of relevance for the *in vivo* situation. By using model systems we were able to provide evidence to what was mainly based on assumptions regarding the conformation and topology in the membrane. This result may open the way to further research, for example mutagenesis studies, in order to elucidate more about this intriguing protein, which is so small that probably every amino acid plays a crucial role. It is fascinating that such a small protein can fulfill so many roles such as: insertion into the membrane, interaction with other protein present in the assembly site, specific interaction with the DNA hairpin loop, and being part of the viral coat. A future challenge remains to get a better insight in the assembly process and the role of g9p therein.

References

- 1 Endemann, H. and Model, P. (1995) J. Mol. Biol. 250, 496-506.
- 2 Webster, R.E. (1996) in Phage Display of peptides and proteins (Kay, B., Winter, J. and McCafferty, J., eds.), pp. 1-20, Academic Press, San Diego.
- 3 Simons, G.F.M., Beintema, J., Duisterwinkel, F.J., Konings, R.N.H. and Schoenmakers, J.G.G. (1981) Prog. Clin. Biol. Res. 64, 401-411.
- 4 Killian, J.A., Trouard, T.P., Greathouse, D.V., Chupin, V. and Lindblom, G. (1994) FEBS Lett. 348, 161-165.
- 5 Glaubitz, C., Grobner, G. and Watts, A. (2000) Biochem. Biophys. Acta 1463, 151-161.

- 6 White, S.H. and Wimley, W.C. (1998) *Biochim. Biophys. Acta* 1376, 339-352.
- 7 Sakai, H. and Tsukihara, T. (1998) *J. Biochem.* 124, 1051-1059.
- 8 Killian, J.A. and von Heijne, G. (2000) *Trends Biochem. Sci.* 25, 429-434.
- 9 Arkin, I.T. and Brunger, A.T. (1998) *Biochim. Biophys. Acta* 1429, 113-128.
- 10 De Planque, M.R.R., Kruijtzter, J.A.W., Liskamp, R.M.J., Marsh, D., Greathouse, D.V., Koeppe, R.E., de Kruijff, B. and Killian, J.A. (1999) *J. Biol. Chem.* 274, 20839-20846.
- 11 Meijer, A.B., Spruijt, R.B., Wolfs, C. and Hemminga, M.A. (2001) *Biochemistry* 40, 8815-8820.
- 12 Spruijt, R.B., Wolfs, C.J.A.M., Verver, J.W.G. and Hemminga, M.A. (1996) *Biochemistry* 35, 10383-10391.
- 13 Stopar, D., Spruijt, R.B., Wolfs, C.J.A.M. and Hemminga, M.A. (1996) *Biochemistry* 35, 15467-15473.
- 14 Meijer, A.B., Spruijt, R.B., Wolfs, C.J.A.M. and Hemminga, M.A. (2000) *Biochemistry* 39, 6157-6163.
- 15 Meijer, A.B., Spruijt, R.B., Wolfs, C. and Hemminga, M.A. (2001) *Biochemistry* 40, 5081-5086.
- 16 Spruijt, R.B., Meijer, A.B., Wolfs, C.J.A.M. and Hemminga, M.A. (2000) *Biochim. Biophys. Acta* 1508, 311-323.
- 17 Shon, K.J., Kim, Y.G., Colnago, L.A. and Opella, S.J. (1991) *Science* 252, 1303-1304.
- 18 Papavoine, C.H.M., Christiaans, B.E.C., Folmer, R.H.A., Konings, R.N.H. and Hilbers, C.W. (1998) *J. Mol. Biol.* 282, 401-419.
- 19 McDonnell, P.A. and Opella, S.J. (1993) *J. Magn. Reson., Ser. B* 102, 120-125.
- 20 Killian, J.A. (1998) *Biochim. Biophys. Acta* 1376, 401-416.
- 21 de Planck, M.R.R. (2000) Thesis, Utrecht University.
- 22 Meijer, A.B. (2001) Thesis, Wageningen University.
- 23 Marvin, D.A., Hale, R.D., Nave, C. and Citterich, M.H. (1994) *J. Mol. Biol.* 235, 260-286.
- 24 Gray, T.M. and Matthews, B.W. (1984) *J. Mol. Biol.* 175, 75-81.
- 25 Ballesteros, J.A., Deupi, X., Olivella, M., Haaksma, E.E.J. and Pardo, L. (2000) *Biophys. J.* 79, 2754-2760.
- 26 Dawson, J.P., Weinger, J.S. and Engelman, D.M. (2002) *J. Mol. Biol.* 316, 799-805.
- 27 Von Heijne, G. (1986) *EMBO J.* 5, 3021-3027.
- 28 Cao, G., Cheng, S., Whitley, P., von Heijne, G., Kuhn, A. and Dalbey, R.E. (1994) *J. Biol. Chem.* 269, 26898-26903.
- 29 Ridder, A., Morein, S., Stam, J.G., Kuhn, A., de Kruijff, B. and Killian, J.A. (2000) *Biochemistry* 39, 6521-6528.
- 30 Kiefer, D. and Kuhn, A. (1999) *EMBO J.* 18, 6299-6306.
- 31 Kuhn, A., Kreil, G. and Wickner, W. (1986) *EMBO J.* 5, 3681-3685.
- 32 Cao, G., Kuhn, A. and Dalbey, R.E. (1995) *EMBO J.* 14, 866-875.
- 33 Kusters, R., Breukink, E., Gallusser, A., Kuhn, A. and de Kruijff, B. (1994) *J. Biol. Chem.* 269, 1560-1563.
- 34 van Klompenburg, W., Nilsson, I., von Heijne, G. and de Kruijff, B. (1997) *EMBO J.* 16, 4261-4266.
- 35 Van Klompenburg, W. and De Kruijff, B. (1998) *J. Membrane Biology* 162, 1-7.

- 36 Dalbey, R.E. and Kuhn, A. (2000) *Ann. Rev. Cell Dev. Biol.* 16, 51-87.
- 37 Makowski, L. (1992) *J. Mol. Biol.* 228, 885-892.
- 38 Russel, M. and Model, P. (1989) *J. Virol.* 63, 3284-3295.
- 39 Smith, J. and Pease, L. (1980) *CRC Crit. Rev. Biochem.* 8, 315-399.
- 40 Perczel, A. and Hollosi, M. (1996) in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G., ed.), pp. 285-379, Plenum Press, New York.

Summary

Introduction

Bacteriophage M13 is a virus that infects the bacteria *Escherichia coli* (*E. coli*), a single cell organism that resides in our intestines. It consists of the cytoplasm (contents) and a double membrane that keeps the contents together (the barrier to the outside world). The membrane is formed by lipid molecules which consist of a head group that very much likes water (hydrophilic) and two fatty tails that dislike water (hydrophobic). In order to avoid contact with water the fatty tails group together in such a way that a planar double layer (bilayer) of the lipid molecules results. In this lipid bilayer also proteins are present to fulfill different functions, e.g. regulate transport across this barrier. The lipids and proteins together form the membrane. However, this membrane is not a complete barrier for certain intruders in the *E. coli* cell. A virus, such as M13 bacteriophage, can enter the *E. coli* cell. It uses the *E. coli* cell as a host by using the cell's components to multiply itself. And finally the bacteriophage is able to cross the membrane and put itself together on its way out. During this latter process, called the assembly, the new bacteriophage is formed out of its components. It is a special feature of bacteriophage M13 that the assembly takes place in the membrane. Even more special is that bacteriophage M13 does not kill its host. Normally, many other viruses kill their hosts by lysing the cells. Bacteriophage M13 consists of a single DNA molecule, which codes the genomic information, and a "coat" of proteins, which surrounds and protects the DNA. Bacteriophage M13 looks like a rod of almost 1 micrometer length and about 6.5 nanometer diameter. The coat proteins along the long side are all identical and present in many copies (about 2700) and therefore are called major coat proteins. On either end of the bacteriophage are a few different coat proteins, called minor coat proteins. One of these is the topic of this work: the gene 9 protein (g9p). G9p consists of only 32 amino acids. It is located on the end of the bacteriophage that emerges first from the *E. coli* cell during the assembly. Therefore, g9p is expected to play a role early in the assembly process. Genetic experiments also showed that g9p is a possible candidate to interact with the tip of the DNA molecule. This makes g9p an interesting object of study, since it might lift up some of the veil over the assembly process. The assembly process is still like a black box: the components that go in and the outcome (intact bacteriophage) are

known, but what happens in between is still a mystery, although parts of it have been revealed. By taking a closer look of g9p we hope to contribute a small piece in unraveling this mystery. Up to now the structure and position of g9p before assembly were based on speculation. In this thesis we aimed to characterize g9p in more detail by analysis of the structure it adopts in a membrane. In addition, we aimed to determine how and which side enters the membrane.

Research

To study g9p we used chemically synthesized protein and since biological membranes are very complex structures we studied g9p in simpler artificial membranes.

G9p has been predicted to be hydrophobic (Chapter 2) and is therefore difficult to handle. It is difficult to remove the protein once it sticks to a surface, and it has a tendency to form aggregates by sticking to each other. It can be dissolved in the organic solvent TFE. In an aqueous solution, detergents are necessary to solubilize the protein. In Chapter 2 conditions for solubilizing the protein were found. In a weak detergent g9p appeared to aggregate. This was accompanied by a conformational change from α -helix to β -sheet, as measured by CD and FTIR. The protein was incorporated in phospholipid membranes via TFE or detergent. When using detergent, care had to be taken to avoid aggregation into β -sheets. Phospholipid membranes appeared to preserve the state of the protein, either aggregated or non-aggregated. The work described in Chapter 2 resulted in the production of reproducible protein-lipid systems, which were necessary to study g9p in the membrane-bound state.

Information about the structure of g9p is described in Chapter 2 and 3. CD and FTIR revealed that g9p is predominantly α -helical in the phospholipid membrane. By analysis of the amide I band of the FTIR spectrum secondary structure elements were quantified. About 67% of the protein structure was α -helix, and the remaining part was assigned to be turn structure. In addition, the orientation of the α -helix axis with respect to the membrane normal was determined by measuring ATR-FTIR spectra of the protein incorporated in oriented bilayers. G9p appeared to be oriented preferentially parallel to the membrane normal, which means that the orientation is mainly transmembrane

(Chapter 3). Based on these results and general principles that govern the positioning of protein domains in bilayers, we proposed possibilities for a tentative structural model of g9p in a membrane (Chapter 3 and 5). One model assumes a single N-terminal transmembrane helix of 21 amino acid residues long and a C-terminally located turn structure. An alternative model assumes a 15 amino acid trans-membrane helix and a second shorter amphipathic helix, which is more in line with the membrane plane. Turn structure links the two helices. The remaining turn structure is located C-terminally according to this model. This part was shown to contain the antigenic site (Chapter 4). The latter helix-turn-helix model remarkably resembles models of the major coat proteins of bacteriophage M13 and related bacteriophages.

In Chapter 4 g9p was found to be able to insert spontaneously into phospholipid monolayers and bilayers. In addition, the N-terminal side of the protein was determined to be the part to insert, whereas the C-terminal part was determined to remain outside the membrane (N_{in} - C_{out} orientation). This was shown by cleavage experiments. After cleavage, the presence or absence of the antigenic site, which was found to be located at the C-terminal end of the protein, functioned as a determinant of the topology. Negatively charged phospholipids may enhance the efficiency of the adoption of a N_{in} - C_{out} orientation, which is in agreement with our experimental results and with literature (Chapter 5).

The relevance of these experimental results with model systems is that they may provide indications for the *in vivo* situation. Up to now, no evidence was available regarding the conformation of g9p. The predominantly α -helical conformation we observed seems likely to be present *in vivo* also (Chapter 5). The finding that g9p can insert spontaneously in model membranes may be relevant for the *in vivo* situation. Our orientational and topological results show what part of the protein is buried in the membrane and what part is accessible. Translated to the *in vivo* situation this implies that the C-terminal side is sticking out in the cytoplasm, thus making it a good candidate to interact with the viral DNA in the assembly process.

Samenvatting

Inleiding

Bacteriofaag M13 is een virus dat de bacterie *Escherichia coli* (*E. coli*) infecteert. *E. coli* is een eencellig organisme dat te vinden is in onze darmen. Het bestaat uit cytoplasma (de inhoud) en een dubbele membraan die de inhoud bij elkaar houdt (de barrière voor de buitenwereld). De membraan wordt gevormd door lipide moleculen die bestaan uit een kopgroep die zich graag in water bevindt (hydrofiel) en twee vetzuurstaarten die waterafstotend zijn (hydrofoob). Om contact met water te vermijden groeperen deze vetzuurstaarten zich zodanig dat er een vlakke dubbele laag (bilaag) van lipide moleculen ontstaat. In de lipide bilaag zijn ook eiwitten aanwezig die verschillende functies vervullen, b.v. het reguleren van transport door deze barrière. De lipiden en eiwitten vormen samen de membraan. Echter, deze membraan is geen onoverkomelijke barrière voor bepaalde indringers in de *E. coli* cel. Een virus, zoals M13 bacteriofaag, kan de *E. coli* cel toch binnen komen. Het gebruikt de *E. coli* cel als een gastheer door de componenten van de cel te gebruiken om zichzelf te vermenigvuldigen. En na afloop is de bacteriofaag in staat de membraan weer te passeren op weg naar buiten. Tijdens deze passage wordt de nieuwe bacteriofaag vanuit zijn componenten in elkaar gezet. Dit heet dan ook de ‘assemblage’. Het is een speciaal kenmerk van bacteriofaag M13 dat deze assemblage plaats vindt in de membraan. Nog specialer is dat bacteriofaag M13 zijn gastheer intact laat. Normaliter worden gastheercellen kapot gemaakt en gaat de gastheer dood. Bacteriofaag M13 bestaat uit een enkel DNA molecuul, dat codeert voor de erfelijke informatie, en een ‘mantel’ van eiwitten, die het DNA omringt en beschermt. Bacteriofaag M13 ziet eruit als een staaf van bijna 1 micrometer lang en ongeveer 6.5 nanometer doorsnee. De manteleiwitten rond de lange kant zijn allemaal identiek en aanwezig in vele kopieën (ongeveer 2700) en worden daarom ook wel ‘meest voorkomende’ manteleiwitten genoemd. Aan elk uiteinde van de bacteriofaag bevinden zich een aantal verschillende mantel eiwitten. Van elk zijn er slechts enkele exemplaren. Eén van deze is het onderwerp van dit werk: het gen 9 eiwit (afgekort ‘g9p’ van gene 9 protein). G9p bestaat uit slechts 32 aminozuren. Het bevindt zich aan het uiteinde van de bacteriofaag dat het eerste uit de *E. coli* cel naar buiten komt bij de assemblage. Daarom is het aannemelijk dat g9p al vroeg in

het assemblage proces een rol speelt. Genetische experimenten hebben ook aangetoond dat g9p een mogelijke kandidaat is om interactie te hebben met een deel van het DNA molecuul dat uitsteekt. Dit maakt g9p een interessant onderwerp om te bestuderen, aangezien het misschien een tipje van de sluier over het assemblage proces zou kunnen oplichten. Het assemblage proces is nog als een 'black box': wat erin gaat (de componenten) en wat eruit komt (intacte bacteriofaag) zijn bekend, maar wat daar tussen in gebeurt is nog steeds een mysterie, hoewel delen ervan al bekend zijn. Door wat meer in detail naar g9p te kijken hopen we een klein stukje bij te dragen aan het ontrafelen van dit mysterie. Tot nu toe waren de structuur en de positie van g9p voorafgaand aan de assemblage gebaseerd op speculatie. In dit proefschrift wilden we g9p in meer detail karakteriseren door de structuur die het in de membraan aanneemt te analyseren. Bovendien wilden we weten hoe en met welke kant het de membraan ingaat.

Onderzoek

Om g9p te bestuderen hebben we chemisch gesynthetiseerd eiwit gebruikt. Aangezien biologische membranen zeer complexe structuren zijn hebben we g9p bestudeerd in eenvoudigere artificiële membranen.

Volgens de voorspelling is g9p hydrofoob (Hoofdstuk 2) en daardoor moeilijk te hanteren. Het is moeilijk om het eiwit te verwijderen als het eenmaal aan een oppervlak vastplakt, en het heeft een neiging om aggregaten te vormen door samen te klonteren. Het kan worden opgelost in het organisch oplosmiddel TFE. In water zijn detergentia (zeepachtige moleculen) nodig om het eiwit in oplossing te houden. Conditie om het eiwit in de juiste vorm in oplossing te houden worden beschreven in Hoofdstuk 2. In een zwak detergent blijkt g9p te aggregeren. Dit gaat gepaard met een verandering van α -helix naar β -sheet conformatie, zoals gemeten is met CD en FTIR. Het eiwit werd geïncorporeerd in fosfolipide membranen via TFE of detergents. In het geval er detergents werd gebruikt moest er opgelet worden dat het eiwit niet in β -sheets geaggregeerde. Het bleek dat in fosfolipide membranen de aggregatietoestand, hetzij geaggregeerd hetzij niet-geaggregeerd, behouden blijft. Het werk dat in Hoofdstuk 2 beschreven staat resulteerde in de productie van reproduceerbare eiwit-lipide systemen, die nodig waren om g9p in de membraan-gebonden toestand te bestuderen.

Informatie over de structuur van g9p is beschreven in Hoofdstuk 2 en 3. CD en FTIR resultaten wezen uit dat g9p voor het grootste deel α -helix is in de fosfolipide membraan. Door de amide I band in het FTIR spectrum te analyseren werden secundaire structuur elementen gekwantificeerd. Ongeveer 67% van de eiwit structuur was α -helix, en turn structuur werd toegekend aan het overige deel. Ook werd de oriëntatie van de α -helix ten opzichte van de membraan-normaal bepaald door ATR-FTIR spectra van het eiwit in georiënteerde bilagen te meten. G9p bleek voornamelijk parallel aan de membraan-normaal georiënteerd te zijn, wat betekent dat de oriëntatie vooral transmembraan is (Hoofdstuk 3). Op basis van deze resultaten en algemene principes bij de plaatsing van eiwit domeinen in membranen zijn er twee modellen voorgesteld voor de structuur van g9p in de membraan (Hoofdstuk 3 en 5). Het ene model veronderstelt één enkele transmembrane helix van 21 aminozuren lang aan de N-terminale zijde en turn structuur aan de C-terminale zijde. Het andere model veronderstelt een transmembrane helix van 15 residuen lang en een tweede kortere amphipatische helix die meer parallel aan het membraanvlak loopt. Turn structuur verbindt de twee helices. De resterende turn structuur bevindt zich bij de C-terminus volgens dit model. Dit deel bleek ook de enige herkenningsplaats te zijn voor antilichamen (Hoofdstuk 4). Het laatstgenoemde helix-turn-helix model lijkt opmerkelijk veel op modellen van het meest voorkomende manteleiwit (g8p) van bacteriofaag M13 en gerelateerde bacteriofagen.

In Hoofdstuk 4 werd gevonden dat g9p in staat is om spontaan in fosfolipide monolagen en bilagen te inserteren. Bovendien werd bepaald dat de N-terminus van het eiwit de zijde is die inserteert en dat de C-terminus buiten de membraan blijft ($N_{\text{binnen}}-C_{\text{buiten}}$ oriëntatie). Dit werd aangetoond met protease knip-experimenten. De aan- of afwezigheid van de bindingsplaats voor antilichamen, die gelokaliseerd bleek te zijn aan de C-terminale zijde van het eiwit, functioneerde als een aanwijzing voor de topologie na de protease behandeling. Het is mogelijk dat negatief geladen fosfolipiden de $N_{\text{binnen}}-C_{\text{buiten}}$ oriëntatie bevorderen, wat in overeenstemming is met onze resultaten en met wat er bekend is (Hoofdstuk 5).

Het belang van deze experimentele resultaten met model systemen is dat ze aanwijzingen kunnen geven voor de *in vivo* situatie. Tot nu toe was er geen bewijs voor de conformatie van g9p. Hier is gevonden dat de conformatie voor het grootste deel

α -helix is. Het is waarschijnlijk dat dit *in vivo* ook het geval is (Hoofdstuk 5). Het gegeven dat g9p spontaan in model membranen kan inserteren zou ook relevant kunnen zijn voor de *in vivo* situatie. Onze resultaten m.b.t. oriëntatie en topologie van het eiwit, vertaald naar de *in vivo* situatie, betekenen dat de C-terminus uitsteekt in het cytoplasma. Daarmee is het een goede kandidaat om interactie te hebben met virale DNA tijdens het assemblage proces.

Curriculum vitae

Marie Chantal Houbiers werd op 24 oktober 1968 in Groningen geboren. In 1987 behaalde zij met lof haar Gymnasium β diploma op het Collegium Marianum in Venlo. Daarna volgde ze als uitwisselingsstudente nog het laatste jaar van de Dodge City Senior Highschool in Kansas, USA. In 1988 begon zij haar studie Moleculaire Wetenschappen aan de Landbouwniversiteit Wageningen. Zij deed een afstudeervak bij de vakgroep Biochemie in Wageningen in de groep van prof. dr. A.J.W.G. Visser. In dit werk werden eiwit-lipide systemen bestudeerd met behulp van fluorescentie technieken. Daarna vertrok ze naar Portugal waar zij met behulp van magnetische resonantie onderzoek verrichtte aan moleculen die wijn rood kleuren. Dit onderzoek werd uitgevoerd bij het Instituto de Tecnologia Química e Biológica van de Universidade Nova te Lissabon, onder supervisie van dr. H. Santos. In 1994 behaalde ze het doctoraaldiploma met lof. Op 1 april 1994 begon haar aanstelling als Onderzoeker in Opleiding bij de vakgroep Moleculaire Fysica in Wageningen, waar het in dit proefschrift beschreven onderzoek werd uitgevoerd. Omdat zij zich de laatste jaren naast het natuurwetenschappelijke ook steeds meer ging interesseren in het sociale volgt zij nu een opleiding Sociaal Pedagogische Hulpverlening aan de Hogeschool van Arnhem en Nijmegen en verrichtte zij stages op verschillende therapeutische landbouwbedrijven. Momenteel is zij werkzaam in zorginstellingen voor verstandelijk gehandicapten.

List of Publications

Pap, E.H.W., Houbiers, M.C., Santema, J.S., van Hoek, A., Visser, A.J.W.G. (1996), "Quantitative fluorescence analysis of the adsorption of lysozyme to phospholipid vesicles", *European Biophysical Journal* 24, p. 223-231.

Chantal Houbiers, João C. Lima, António L. Maçanita, Helena Santos (1998), "Color Stabilization of Malvidin 3-Glucoside: Self-aggregation of the Flavylum Cation and Copigmentation with the Z-Chalcone Form", *Journal of Physical Chemistry-B* 102, p. 3578-3586.

C.J.A.M. Wolfs, M.C. Houbiers, R.B. Spruijt and M.A. Hemminga (1998), "Characterization of the Gene VII and Gene IX minor coat proteins from Bacteriophage M13". *Proceedings of the NATO Study Institute "Molecular Mechanisms of Lipid and Protein Traffic"*, held at Cragès, France, June 16-28, 1997. NATO ASI Series H: Cell Biology, 106 "Lipid and Protein Traffic, Pathways and Molecular Mechanisms" (Ed. J.A.F. Op den Kamp, Springer-Verlag Berlin Heidelberg), p. 105-114.

Houbiers, M.C., Spruijt, R.B., Wolfs, C.J.A.M., Hemminga, M.A. (1999), "Conformational and aggregational properties of the gene 9 minor coat protein of bacteriophage M13 in membrane-mimicking systems", *Biochemistry* 38, p. 1128-1135.

Houbiers, M.C., Spruijt, R.B., Demel, R.A., Hemminga, M.A., Wolfs, C.J.A.M. (2001), "Spontaneous insertion of gene 9 minor coat protein of bacteriophage M13 in model membranes", *Biochimica et Biophysica Acta* 1511, p. 309-316.

Houbiers, M.C., Wolfs, C.J.A.M., Spruijt, R.B., Bollen, Y.J.M., Hemminga, M.A. Goormaghtigh, E. (2001), "Conformation and orientation of the gene 9 minor coat protein of bacteriophage M13 in phospholipid bilayers", *Biochimica et Biophysica Acta* 1511: p. 224-235.

