Biomolecular Recognition Mechanisms Studied by
NMR Spectroscopy and MD Simulations

# Biomolecular Recognition Mechanisms

## Studied by

# NMR Spectroscopy and MD Simulations

**Biomoleculaire herkenningsmechanismen bestudeerd
met NMR spectroscopie en MD simulaties**

(met een samenvatting in het Nederlands)

Proefschrift ter verkrijging van de graad van doctor aan de
Universiteit Utrecht op gezag van de Rector Magnificus, Prof.
Dr. W.H. Gispen, ingevolge het besluit van het College voor
Promoties in het openbaar te verdedigen op
dinsdag 8 juni 2004 des middags te 12.45 uur.

door

**Shang–Te Danny Hsu**

geboren op 23 november 1975,
te Taipei

獻給雙親

*aan mijn ouders*

# Preface

This thesis describes the use of solution Nuclear Magnetic Resonance (NMR) spectroscopy and Molecular Dynamics (MD) simulations to study the mechanism of biomolecular recognition with two model systems: i) lipid II–binding lantibiotics (lanthionine–containing antibiotics) and ii) the human immunodeficiency virus 1 (HIV-1) envelope protein (Env), gp120, and its receptor molecule, CD4. The first system concerns a group of unique antimicrobial peptides, which make use of an hitherto unknown mechanism of attacking bacteria by targeting the Achilles' heel of bacteria, the cell wall precursor, lipid II. In the light of antibiotic resistance, understanding of this recognition mechanism may lead to novel antibiotics. The second system focuses on the initiation step of the HIV-1 viral entry wherein the engagement of gp120 and CD4 switches on a cascade of conformational changes that are necessary for the membrane fusion between the virus and the host cell. The biological contexts of both systems are important to human health and numerous functional studies on both systems have been well documented. Yet, because of the underlying dynamics and the intricate assembly process of higher order complexes, a detailed structural description is currently lacking in both systems. We therefore applied advanced NMR and MD techniques to unravel the structure and dynamics of these complexes with the hope to facilitate the development of new antibiotics and vaccines for infectious diseases, such as AIDS. As biological functions are manifested by interactions at a molecular level, understanding of structural properties of these biomolecules may consolidate related biomedical research.

# Contents

# General Introduction

## Part I. General introduction of biomolecular recognition

Current advances in biological sciences are spurred by two major factors: our growing understanding of the nature of the basic elements involved in complex systems and the development of tools that provide higher performance. Seemingly recondite and impossibly complex, the physical and chemical nature behind the biological systems is, in fact, simple as stated by Linus Pauling and Max Delbrück in the early days of biological research when the development of Quantum Physics was at its apex (Pauling and Delbrück, 1940):

"It is our opinion that the processes of synthesis and folding of highly complex molecules in the living cell involve, in addition to covalent-bond formation, only the intermolecular interactions of van der Waals attraction and repulsion, electrostatic interactions, hydrogen-bond formation, etc., which are now rather well understood. These interactions are such as to give stability to a system of two molecules with complementary structure in juxtaposition.

... in order to achieve the maximum stability, the two molecules must have complementary surfaces, like die and coin, and also a complementary distribution of active group."

Biological function is the result of the interplay between two or more biomolecules, from duplication of genetic materials to immunogenic response or signal transduction. The effectiveness and correctness of the interactions determine the efficiency of these processes. Above all, recognition is the first step and is necessary in the cascade of events leading to functions. *"Recognition is assembly plus specificity"* (Janin, 1995). Because of the large number of biomolecules and the even larger number of possible interactions between them, it is crucial to first collect all the pieces of the puzzle. It is this necessity that initiated the exploration of *Genomics*. Databases of complete genome sequences are now available for many model systems, including *E. coli*, yeast, *C. elegans* and man. The complete sequence of the genome of an organism encrypts all possible products, that is, the genes defined by the open reading frames (ORF), which are to be translated into proteins. Like the lexicon of a language, genes are encoded with four types of nucleobases, namely adenine (A), thymine (T), guanine (G) and cytosine (C), unlike the many different alphabets of complex languages or the Chinese characters. Immediately after genomic sequencing, functional mapping of the ORFs takes place. Unless the function of each ORF can be ascribed, a genome sequence is nothing but a series of DNA base pairs, just like an alphabet and punctuation, which are needed to assemble ideas into words, sentences, paragraphs and chapters (Pollack and Iyer, 2002). In short, the functional annotation of proteins is the objective of *Functional Genomics*. To advance from Genomics to Proteomics one requires genome-wide knowledge at different levels from multicelluar organism, single cell to single molecule (Tyers and Mann, 2003). However, one should bear in mind that the functional annotation can be biased by the choice of technique. Functions identified *in vitro* may not be always present *in vivo* (Sprinzak *et al.*, 2003). Soon after completion and dissemination of the proteome, a complex molecular interaction network of the gene products will be constructed. The complexity of this is not only a result of the large number of interacting components but also the response profile in which time plays a crucial role. A field in the post-genomic era called *Interactomics* is now emerging (Ito *et al.*, 2001; Bader and Hogue, 2002; Valencia and Pazos, 2002; Li *et al.*, 2004). Ultimately, only when a comprehensive understanding of the components and a layout of the whole biological machinery is established, can we realise the idea of *Reverse Engineering* of systems biology (Csete and Doyle, 2002).

Along with the multidisciplinary effort on Proteomics, *Structural Genomics* (or *Structural Proteomics*) has proven to be amongst the most powerful approaches to provide insights into the functionalities of biomolecules (Burley *et al.*, 1999; Sali *et al.*, 2003). International efforts have been devoted to set up many Structural Genomics Consortia in North America (NIGMS), Japan (RIKEN) and Europe (SPINE)[‖]. The idea behind Structural Genomics is simple: to determine as many protein structures in a genome-wide manner as possible because of the belief that the universe of compact protein structures is spanned by a limited number of basic topology folds, between 1000 and 5000 (Holm and Sander, 1996; Brenner *et al.*, 1997). Considering cost and efficiency, the search for novel folds is the top priority of most Structural Genomics pipelines. This effort requires data mining of genomic databases using bioinformatics. On-line servers such as DALI (http://www.ebi.ac.uk/dali/) and SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/) are designated to categorise protein structures and folds when a new protein structure is determined. The significance of the knowledge of three-dimensional (3D) structures of biomolecules is that the functions are often conferred by the 3D structures (Branden and Tooze, 1998). As described by Pauling and Delbrück, the interaction interface requires complementarity of the two counterparts, not only geometrically but chemically; these functional structures, however, are not necessarily present all the time. Since the first characterisation of the 3D structure of nucleic acids (Watson and Crick, 1953) and proteins (Kendrew *et al.*, 1958; Perutz *et al.*, 1960) we now know that biomolecular structures are not static and that very often functionality requires substantial structural rearrangement upon recognition and complexation, *i.e.*, induced fit (Graham *et al.*, 2000; Williamson, 2000; Demarest *et al.*, 2002). Also important are the intrinsically disordered proteins that exist despite a lack of well-defined 3D structure. Intrinsic flexibility, in turn, makes these disordered proteins entropically more favourable (Dunker *et al.*, 2001; Dyson and Wright, 2002). Folding and binding can be expressed in energetic funnel models, which attempt to describe the energetic landscape of conformational space and indicate the pathways toward the energetically favourable states (Wolynes *et al.*, 1995; Xu *et al.*, 1997a; Dobson *et al.*, 1998; Tsai *et al.*, 1998; Tsai *et al.*, 1999; Dinner *et al.*, 2000; Kumar *et al.*, 2000; Vendruscolo *et al.*, 2001; Gruebele, 2002). One of the recent trends in Structural Biology is to understand how proteins fold into functional/native conformations; what are the causes that make proteins misfold, which can lead to many amyloid diseases, *e.g.* Alzheimer disease? How, if possible, can we divert the folding pathway into the desired ones (Dobson, 2003)?

Two proteins that share a high degree of primary sequence homology throughout evolution often posses similar folding topology and often similar functions. On the other hand, examples have shown that the similarity in 3D fold does not always imply a high primary sequence homology. How to decipher the structural information that is encrypted within the primary sequence of proteins is the ongoing *Protein Folding* problem (Dill and Chan, 1997; Fersht and Daggett, 2002; Plotkin and Onuchic, 2002a; Plotkin and Onuchic, 2002b). Recent progress in *ab initio* structure prediction has demonstrated the feasibility of predicting the tertiary protein structure from the primary sequence (Baker and Sali, 2001). This has become the objective of the Critical Assessment of methods of protein Structure Prediction (CASP) (Moult *et al.*, 2003). Homology modelling can be achieved, in general, provided that a homologous 3D structure is known with a critical threshold of 30% sequence homology. The reverse problem to *Protein*

---

[‖] Further information can be found in the following internet sites: NIGMS, http://www.nigms.nih.gov/psi/, RIKEN, http://www.rsgi.riken.go.jp/, SPINE, http://www.spineurope.org/

*Folding* is *Protein Design*: to design an artificial protein or even a new fold from scratch using the *de novo* knowledge-based approaches (Bryson *et al.*, 1995; Harbury *et al.*, 1998; Kuhlman *et al.*, 2003). Likewise, when protein structures are available, the next step will be the *ab initio* prediction of protein-protein interactions (Valencia and Pazos, 2002) *i.e.*, *Interactomics* in structural terms, as carried out in the Critical Assessment of PRedicted Interactions (CAPRI) (Janin *et al.*, 2003). New emerging algorithms are trying to improve the prediction accuracy and reduce the degree of complexity by using experimental data concerning the geometry or the chemistry of the interactions that are readily available in literature, or, relatively easy to obtain compared to the time-consuming full structural determination of the complexes (Morelli *et al.*, 2000; Dominguez *et al.*, 2003).

One immediate application of the knowledge of the 3D structure of a biomolecule is structure-oriented site-directed mutagenesis, which allows to correlate *in vitro* structural models with *in vivo* bioactivities. Without a 3D model, mutagensis scanning can be extremely lengthy and tedious. The interpretation of the observed functionality is also limited to the primary sequence. It is worth noting that critical residues in a protein are usually close in space but might be dispersed in the primary sequence. Knowledge of the 3D topology thus provides key information to facilitate biochemical manipulation as such. Furthermore, when the interaction sites are mapped, those that are disease-related readily become the vantage points for the structure-based drug design (SBDD). Structure-activity relationship (SAR) by NMR spectroscopy is an approach that uses the structural information about the ligand binding site to derive compounds that have high specificity and activity (Shuker *et al.*, 1996). It has now become widely used in the pharmaceutical industry for lead screening and rational drug design against potential targets instead of random screening of numerous small organic compounds.

### Biomolecular interactions

At the time of writing this thesis, the coordinates of about 25000 biomolecular structures have been deposited in the Protein Data Bank[#] (PDB, http://www.rcsb.org/pdb/). The number of binary or higher order biomolecular complexes represents only a small fraction (~4%). Nonetheless, the limited number of high resolution structures of biomolecular complexes is valuable for systematic surveys of biomolecular interaction modes between proteins (Thornton *et al.*, 1993; Mitchell *et al.*, 1994; Janin and Rodier, 1995; Jones and Thornton, 1996; Larsen *et al.*, 1998; Lo Conte *et al.*, 1999; Glaser *et al.*, 2001; Chakrabarti and Janin, 2002; Nooren and Thornton, 2003; Ofran and Rost, 2003), between protein and nucleic acid (Moodie and Thornton, 1993; De Guzman *et al.*, 1998; Jones *et al.*, 1999; Jones *et al.*, 2001; Luscombe *et al.*, 2001; Tolstorukov *et al.*, 2004), and between protein and carbohydrate (Taroni *et al.*, 2000). In contrast, structural understanding of the recognition of fatty acids is as yet limited to some special cases (Flower, 1996; Holmquist, 1998; Sinensky, 2000). Throughout the various surveys, a few general measures have been used for the analysis of the structures of biomolecular complexes, namely, the size of the complex interface, the atomic packing or the density of the structure of the complex. The chemical composition at the interface, electrostatic and van der Waals contacts and hydrogen bonding, are also subjected to statistics. The interface area *B* of a complex, also expressed as the buried surface area (BSA), is defined as the difference between the sum of the solvent accessible surface (SAS) of the free components (SAS[free])

---

[#] PDB is not limited to protein structures. It also contains 3D structures of other types of biomolecules, such as nucleic acids and carbohydrates, and their complexes.

and the surface of the complex (SAS$^{cplx}$) (Wodak and Janin, 2003):

$$BSA = \Delta SAS = \sum_i SAS_i^{free} - SAS^{cplx} \tag{1}$$

The contribution of each separate component is roughly equal for a binary complex and therefore the complex interface is sometimes quoted in literature as half the loss of SAS, *i.e.*, ($\Delta$SAS)/2 (Jones and Thornton, 1996). This, however, cannot account for complexes that consist of more than two components and therefore we choose to use the first convention in the following. There seems to be a correlation between the interface area and the specific function of biomolecules. For instance, the "standard size" protein-protein interfaces, as defined by Wodak and Janin, has an interface area of 1600±400 Å$^2$ and 9±4 intermolecular hydrogen bonds and the large surface, in turn, suffices to achieve very high stability. For protease-inhibitor or antibody-antigen complexes, which require higher specificity, the corresponding interface area is larger. Protein-nucleic acid complexes generally have smaller interface area (800±200 Å$^2$) with 24±6 amino acids and 12 ±3 nucleotides, while the interface area of enzymatic complexes is much larger (2200±250 Å$^2$). Throughout these surveys, some important motifs were also identified for recognition: arginine is the most commonly used amino acid type for nucleic acid backbone recognition and the 2' hydroxyl group of ribonucleic acid (RNA) plays an important role in the recognition process as well (Wodak and Janin, 2003). Depending on the functionalities of various systems, water is present in abundance at protein–protein and protein–nucleic acid interface. The bound water molecules often play an important role in the specificity of binding. Dehydration at the binding cavity is an important issue for the pharmaceutical design of high affinity ligands (Ladbury *et al.*, 1994; Morton and Ladbury, 1996; Tame *et al.*, 1996; Schwabe, 1997; Janin, 1999).

As described in the beginning of this chapter, intermolecular interaction modes are simply dictated by the non-bonded interactions (Leckband and Israelachvili, 2001): Coulombic interactions (Honig and Nicholls, 1995; Sheinerman *et al.*, 2000) (see review by Laberge, (1998) for a general introduction of the electrostatics of proteins), hydrophobic van der Waals interactions (Young *et al.*, 1994; DeLano *et al.*, 2000; Pitera and van Gunsteren, 2001) and hydrogen bonding (Borders *et al.*, 1994; Shimoni and Glusker, 1995; Xu *et al.*, 1997b; Sarkhel and Desiraju, 2004), whose partial covalent bond nature has, only recently, been confirmed experimentally by NMR spectroscopy (Dingley and Grzesiek, 1998; Pervushin *et al.*, 1998; Dingley *et al.*, 2001). A detailed description of the atomic structure of biomolecules and their complexes become necessary when one aims at dissecting the contribution of different modes of interaction, which would otherwise be impossible without structural knowledge. Conversely, structural insight needs to be complemented by biochemical assays in order to interpret it in the biological context.

Knowledge of biomolecular structure rendering possible the dissection of non-bonded contributions has facilitated the understanding of how well recognition can be achieved and how stable the structure of the complex is in order to determine the affinity between the two counterparts. The affinity ($K_D$) of any complex is given by a simple relation $K_D = k_{off}/k_{on}$, which holds for many diffusion limited reactions. The on- and off-rates are limited by two distinct mechanism: The rate constant of association (expressed as $k_{on}$) is limited by diffusion (Shoup *et al.*, 1981; Shoup and Szabo, 1982) and can be increased by a favourable electrostatic force (Schreiber and Fersht, 1996); the rate constant of dissociation (expressed as $k_{off}$) is determined by the strength of short range interactions, such as hydrophobic interactions, hydrogen bonding and salt-bridges. In conjunction with protein engineering, these principles have enabled us to

pinpoint the "hot spots" of binding interfaces (Bogan and Thorn, 1998), which serve as a blueprint for the rational design of a *super-protein* (Selzer *et al.*, 2000; Selzer and Schreiber, 2001).


## Part II. Lantibiotics and lipid II

*Antibiotic resistance and the need for a new magic bullet*

The emerging antibiotics-resistance problem, such as that of methicillin-resistant *Staphylococcus aureus* (MRSA), has underlined the urgent need for novel antimicrobial agents for infectious therapy. It has been reported that about 70% of bacteria found in hospital are resistant to at least one of the drugs most commonly used to treat infection. According to the Centres for Disease Control and Prevention (CDC) of the United States of America, the spread of nosocomial drug resistance amongst pathogens is growing annually at an alarming rate. Strains with some level of resistance to vancomycin, the last resort of clinical antibiotics, vancomycin-intermediate-resistant *S. aureus*, (VISA) have been reported since 1996 and the newly identified highly resistant strain (vancomycin-resistant *S. aureus*, VRSA) presents a new challenge to the battle against infectious pathogens (Hughes, 2003). In the search for future antimicrobial agents, a great deal can be learned from nature. Throughout evolution natural defence systems exploit a broad range of peptides that possess antimicrobial activities (Zasloff, 2002). Amongst the variety of antimicrobial peptides (AMPs), a prospective solution lies in a unique family of peptides called lantibiotics.

Lantibiotics (lanthionine-containing antibiotics) are an antimicrobial peptide family produced by bacteria as defence weapons (Kuipers *et al.*, 1996; Brötz and Sahl, 2000). The post-translationally modified lanthionine ring structures and dehydrated amino acids, such as dehydroalanine and dehydrobutyrine, which are the residual dehydration products in the ring-making process (Figure 1), are the hallmark of lantibiotics (Jack and Jung, 2000). Some additional modifications are also present in the N- and C-termini of various family members. To date, about 25 lantibiotics have been identified (Sahl and Bierbaum, 1998; Guder *et al.*, 2000). They are divided into two subfamilies, types A and B, based on the primary sequence, the lanthionine ring structure and, to a lesser extent, the functionality. Type A lantibiotics are generally longer than type B lantibiotics. Functionally, some type A lantibiotics can induce membrane permeability, disrupting the physiological membrane potential and causing the malfunction of the cell compartmentalisation. Type B lantibiotics, on the other hand, interfere with some enzymatic reactions, which are critical for the vitality of bacteria. Both type A and B lantibiotics can lead to cell death.

Novel lantibiotics are usually purified and identified from co-incubation of lantibiotic-producing strains and infectious pathogens, from which the presence of antimicrobial agent is indicated by the inhibition of pathogen growth. Characterisation of their chemical structure is more complicated than of conventional peptides because of the lanthionine linkages and modified amino acids. Solution NMR spectroscopy, which is commonly used for the characterisation of the 3D structures of lantibiotics, is an alternative tool for the elucidation of the primary sequences and the positions of the lanthionine linkages, when conventional fragmentation sequencing by mass spectrometry fails (Gross and Morell, 1970; Chan *et al.*, 1989; Mulders *et al.*, 1991; Chan *et al.*, 1992; Kuipers *et al.*, 1992; Sailer *et al.*, 1993; Smith *et al.*, 2000; Martin *et al.*, 2004).
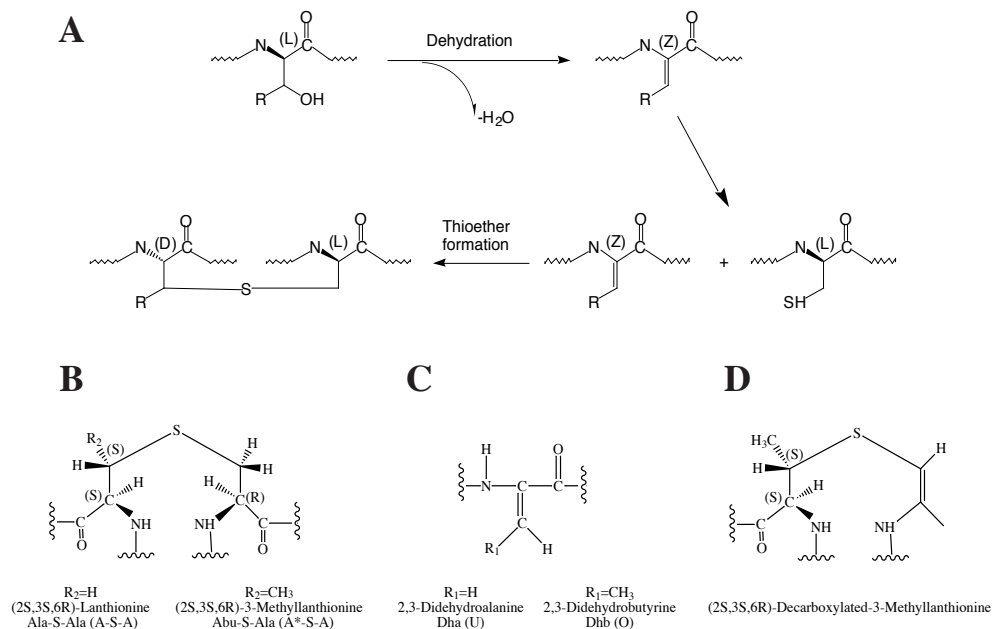
**Figure 1.** Formation of a lanthionine linkage. **A**. The L–form serine (R=H) or threonine (R=CH₃) is dehydrated forming the Z–form substrate for the following thioether formation with the proximal L–form cysteine residue, which leads to the lanthionine linkage with the D- (S) and L–form (R) conformations of the sulphur accepting and donating residues, respectively. **B**. Stereochemical structure of the lanthionine linkage. The three and one letter codes of the corresponding residue types are shown below. **C**. Z–form dehydrated amino acids. These are the excess dehydration products from the lanthionine formation process when fewer cysteine residues are available. **D**. Additional cyclised C–terminal ring structure that is present in mersacidin and epidermine.

Functional characterisation of lantibiotics is another important issue that is not well established in many cases. Over the years, an emerging view of lantibiotics' mechanistic mode of attacking bacteria involves of a common target molecule – the bacterial cell wall precursor, lipid II. Lipid II binding has been demonstrated for several lantibiotics, including nisin, one of the first identified (Brötz *et al.*, 1998a; Brötz *et al.*, 1998b; Brötz and Sahl, 2000). Nisin is nowadays commonly used in the food industry as a preservative. Nonetheless, understanding of the target recognition mechanism, which is essential for lantibiotic-based drug design, is limited and thorough structural insight is currently lacking.

*Nisin*

Nisin is a member of the type A lantibiotics (Figure 2A). It contains five lanthionine rings, namely rings A–E, and bears thirteen modified residues out of the thirty–four in total (38%). It is positively charged with three lysine residues and the amphipathic nature of nisin provides a potential to interact with the generally negatively charged bacterial membrane (Breukink and de Kruijff, 1999). Amongst several homologous type A lantibiotics (Figure 2B), nisin is the most extensively studied. It is also the only lantibiotic that is commercially used as preservative, known as E234, for cheese, milk products and beer. Nisin has been

discovered in 1928. In 1960, the antimicrobial activity of nisin was first attributed to its capability of inducing the leakage of intracellular compounds of *Clostridia* cells (Ramseier, 1960). Twenty years later, other evidence suggested that nisin forms a complex with the lipid-bound peptidoglycan precursor and interferes in cell wall synthesis, which might account for its antimicrobial activity (Reisinger *et al.*, 1980). However, later studies demonstrated that nisin and other type-A lantibiotics exhibit membrane permeablisation ability that leads to a rapid efflux of cytosolic content and hence causes a rapid stop of biosynthesis and eventually cell death (Ruhr and Sahl, 1985). Since then, the pore-forming activity became a preferred mode of action to account for the antimicrobial activity of nisin. This was supported by several *in vitro* biophysical studies, primarily based on membrane vesicle leakage experiments (Breukink *et al.*, 1997; Breukink *et al.*, 1998; van Kraaij *et al.*, 1998). The 3D structure of nisin was determined in aqueous solution (van de Ven *et al.*, 1991; Lian *et al.*, 1992), sodium docecylsulphate (SDS) and dodecylphosphocholine (DPC) micelles – membrane mimicking environments (van den Hooven *et al.*, 1996a; van den Hooven *et al.*, 1996b). Because of the modified amino acids and the lanthionine linkages, nisin possesses no canonical secondary structure element but a number of β-turns in the lanthionine rings, which are predefined by the linkages. Efforts have focused on the elucidation the membrane surface location and the transmembrane orientation of nisin in order to envision a model for the pore-forming activity using a combination of solution and solid state NMR spectroscopy, fluorescence spectroscopy in conjunction with mutagenesis and other techniques (Driessen *et al.*, 1995; van den Hooven *et al.*, 1996b; Dykes *et al.*, 1998; Lins *et al.*, 1999).

## A
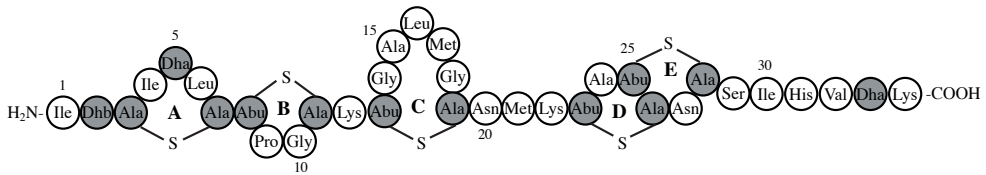


## B
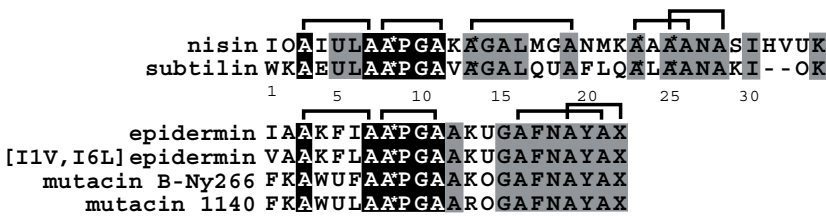


**Figure 2.** Nisin and other type A lantibiotics. **A**. Primary structure of nisin. The post-translationally modified residues are indicated in grey circles and the lanthionine rings are indicated (-S-) with labels A to E from N- to C-termini, accordingly. **B**. Primary sequence alignment of type A lantibiotics. Nisin and subtilin form a subfamily of type A lantibiotics. Another subfamily consists of epidermin, [I1V,I6L]–epidermin, mutacin B-Ny266 and mutacin 1140 (Guder *et al.*, 2000; Smith *et al.*, 2000) with a cyclised C-terminus (X: 2-aminovinyl-D-cysteine, AviCys). The universally conserved residues are highlighted in black and the conserved residues within the subfamilies are shaded in grey. The lanthionine linkages are indicated above each subfamily.

A significant discrepancy exists, however, between the pore-forming activity of nisin *in vitro* and the observed minimal inhibitory concentration (MIC) required for its antimicrobial activity *in vivo*. Nisin is much more active than many non-specific pore-forming peptides *in vivo*, such as magainin, but it is much less active *in vitro* when tested using model membrane systems. The cause of this puzzle was revealed when the peptidoglycan cell wall precursor, lipid II – identified by Reisinger *et al.* in 1960 as the target of nisin – was shown to act as the membrane receptor molecule and to facilitate the pore-forming activity of nisin. The incorporation of lipid II provides a specific anchoring point for nisin during membrane association and insertion. This targeted mode of action, in turn, lifts the required concentration of pore formation from a micromolar (µM) to a nanomolar (nM) range (Breukink *et al.*, 1999). This finding united the previously divergent views of the mode of actions of nisin. It also demonstrated a novel mode of action that is different from that of vancomycin and hence might be exploited for the development of novel antibiotics. The use of lipid II as an auxiliary molecule in the pore-forming activity of nisin is the first example of targeted pore formation and this finding has drastically changed the idea of the pore-forming mechanism of nisin.

## Lipid II

The bacterial cell wall is a cross-linked peptidoglycan matrix structure in the extracellular domain (Rogers and Perkins, 1968). It plays a central role in the vitality of cells because it is responsible for maintaining the shape of cells in resisting high osmotic pressure generated by the highly concentrated cytosolic contents and prevents the cell from bursting. Lipid II is the precursor of the monomeric peptidoglycan unit (Figure 3). The basic building block of the cell wall consists of two amino sugars, N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc), and a pentapeptide, often L-Ala-D-γ-Glu-L-Lys-D-Ala-D-Ala, which is attached to the carboxyl group of MurNAc (Figure 4A). These subunits are assembled in the cytosol on a membrane-anchoring carrier, undecaprenyl phosphate, yielding lipid II (GlcNAc-MurNAc-pentapeptide-pyrophosphoryl-undecaprenol, see Figure 3; for reviews see (van Heijenoort, 1994; van Heijenoort, 2001)). Lipid II is thereafter transported to the extracellular domain and the peptidoglycan is detached for polymerisation (Figure 4B).

The bacterial cell wall has been a major target for many antimicrobial agents (Katz and Caufield, 2003; Koch, 2003). More specifically, lipid II is amongst one of the most prominent targets for many antibiotics, such as vancomycin-like glycopeptides (Sheldrick *et al.*, 1978; Ge *et al.*, 1999), lipoglycodepsipeptide ramonplanin (Cudic *et al.*, 2002; McCafferty *et al.*, 2002; Hu *et al.*, 2003; Montecalvo, 2003) and various lantibiotics (Brötz *et al.*, 1997; Brötz *et al.*, 1998a; Brötz *et al.*, 1998b; Hsu *et al.*, 2002; Hsu *et al.*, 2003). Interfering with the lipid II synthesis machinery, which involves a series of enzymatic reactions with a broad spectrum of Mur enzymes (Figure 4), also has a great potential for novel antibiotics (Lazar and Walker, 2002; El Zoeiby *et al.*, 2003). The chemical complexity of lipid II is a major challenge for large-scale synthesis, which is required for various research approaches. Recently, a number of synthetic protocols have been reported for the production of lipid II and its derivatives (Lo *et al.*, 2001; Schwartz *et al.*, 2001; VanNieuwenhze *et al.*, 2001; Ye *et al.*, 2001; VanNieuwenhze *et al.*, 2002; Breukink *et al.*, 2003). 3D structural studies of lipid II, however, are still limited (Fermandjian *et al.*, 1987; Matter *et al.*, 1997; van Asselt *et al.*, 2000; Feher *et al.*, 2003).

**A**



**B**
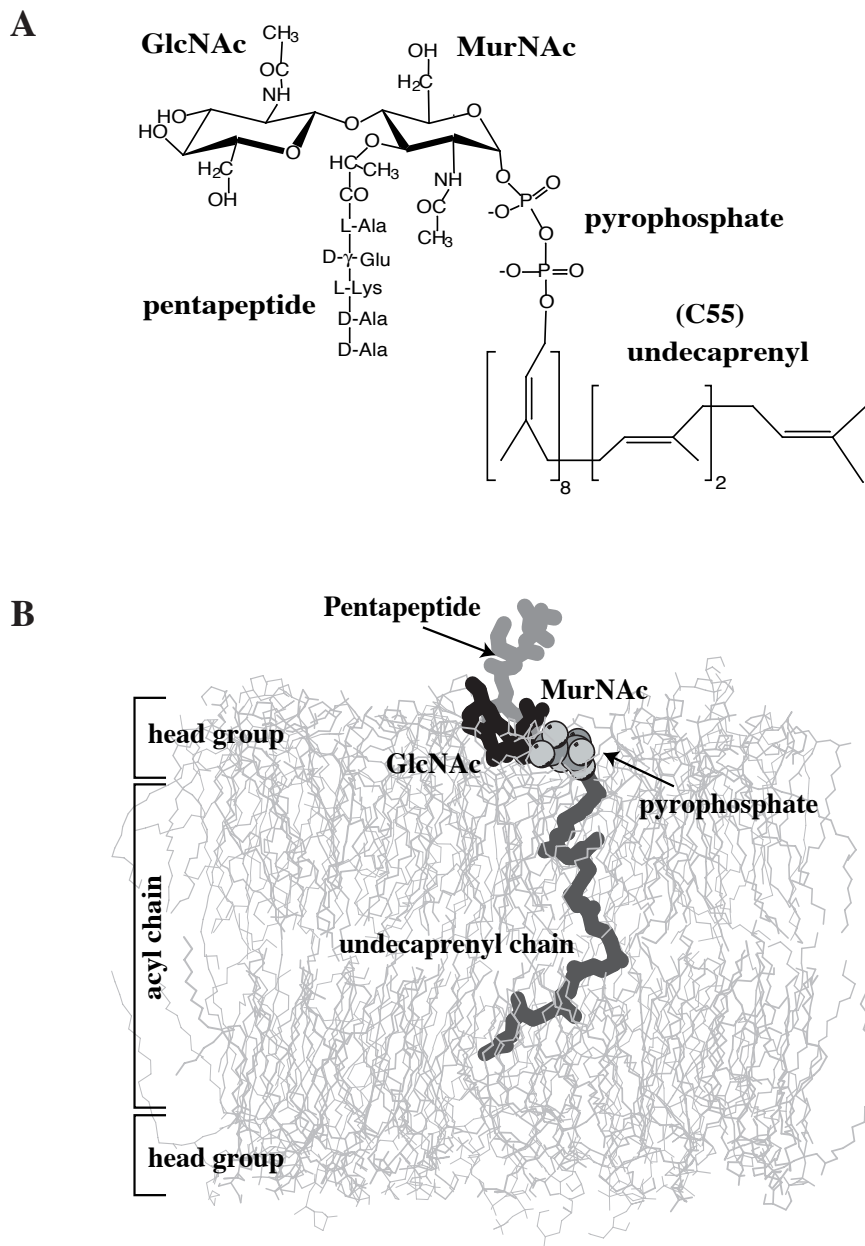


**Figure 3. A.** Chemical structure and **B.** 3D model of lipid II. The 3D model is a representative structure taken from a five nanosecond MD simulation trajectory of lipid II in a membrane with 128 1-palmitoyl-2-oleoyl-phosphocholine (POPC) molecules in an explicit water box. The regions of the head groups and the acyl chains of the POPC bilayer are indicated.

*Lipid II-mediated pore formation by nisin*

After the discovery of the targeting mechanism of nisin, a mutagenesis study revisited the structure-function relationship of individual amino acids of nisin with respect to lipid II binding affinity and pore formation *in vitro* and the *in vivo* antimicrobial activities (Wiedemann *et al.*, 2001). It was shown that the lipid II binding specificity of nisin is very sensitive to the structural composition and integrity of the lanthionine rings as well as to the flexibility at the hinge region that is required for conformational changes. For instance, a relatively small change as the introduction of a methyl group in ring A by a [S3T] point mutation resulted in a large reduction in the bioactivity. Similar functional disruption was observed when a monosulphide linkage was replaced by a disulphide bond with a [T13C] point mutation. In accordance with previous studies, the lipid II binding specificity of nisin was attributed to the N-terminal part, *i.e.*, rings A-C, while the C-terminal part was found to be responsible for the membrane insertion, which is required for pore formation (Breukink *et al.*, 1998). In order to describe the recognition mechanism in detail, we have mapped the lipid II binding interface onto the N-terminal part of nisin using NMR spectroscopy in SDS micelles (**Chapter 2**) (Hsu *et al.*, 2002). The most affected part of nisin upon lipid II binding coincides with the minimal structural element for residues 1–12, antagonising full-length nisin's activity of inhibiting bacterial growth, suggesting this fragment as the lipid II binding motif (Chan *et al.*, 1996). Additional perturbations in ring C also indicate that ring C is involved in lipid II or membrane interaction, without which the activity of nisin is basically abolished (Chan *et al.*, 1996). Upon lipid II binding, the C-terminal part of nisin inserts into the membrane bilayer in a perpendicular orientation with respect to the membrane surface. Finally, the assembly of the nisin/lipid II complex takes place to complete the formation of a metastable pore complex (van Heusden *et al.*, 2002). In this multi-step process, the formation of the nisin/lipid II complex reduces the three-dimensional diffusion of nisin into two-dimensional lateral diffusion on the membrane surface along with lipid II. Clustering of nisin/lipid II complexes occurs prior to the membrane permeabilization. The specific pore structure is assembled thereafter. The efficiency of nisin anchoring onto the membrane is an important factor. Reducing the number of isoprene repeats to less than four can substantially diminish the pore-forming activity of nisin whereas modifying the type of lipid chain from a unsaturated polyprenyl chain to a saturated dolichol chain does not affect nisin's activity (Breukink *et al.*, 2003).

What is important but remains elusive is the structural detail of the multi-step pore-forming process from target recognition to pore assembly. Preliminary data have indicated that the initial recognition of nisin and lipid II has a binding stoichiometry of 1:1. This is followed by the binding of another nisin molecule to a second association site on lipid II in order to complete pore formation. This hierarchy of lipid II binding for nisin is very different from that of another lipid II binding peptide, ramoplanin. Ramoplanin by itself exists as a dimer (Lo *et al.*, 2001) and it binds to lipid II with a 2:1 stoichiometry without the binding hierarchy of nisin (Hu *et al.*, 2003). A putative interaction mode between ramoplanin and lipid II, primarily electrostatic, was proposed based on chemical shift perturbations and a few intermolecular nuclear Overhauser effects (NOEs) but the structure of the complex is thus far unavailable (Cudic *et al.*, 2002). An important message of these studies is that lipid II has several potential binding sites for various types of antimicrobial peptides. Conjecturing from the stoichiometry determined by excimer fluorescence spectroscopy[*], a rough size estimate of the transient pore complex observed in atomic force microscopy (AFM)[§] and black lipid membrane study (Wiedemann *et al.*, 2004) in conjunction with an MD simulation of lipid II in a model membrane bilayer[†], a stable nisin/lipid II pore likely consists of
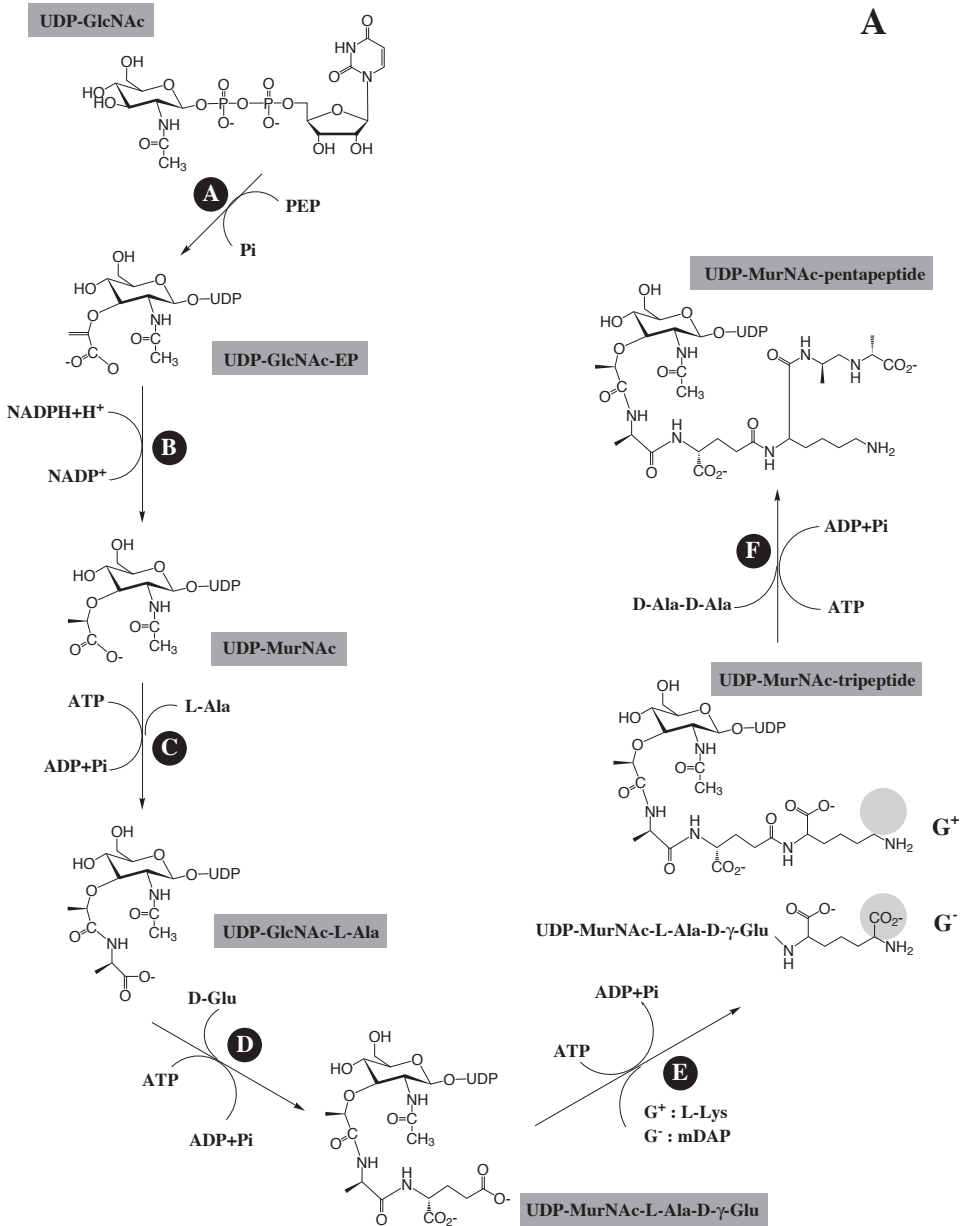
**Figure 4.** Chemical synthesis pathway of lipid II. **A**. Conversion of UDP–N–glucosamine to UDP–N–acetylmuramyl pentapeptide by the sequential action of MurA to MurF enzymes (labelled in black filled circles). UDP-GlcNAc, UDP–N–acetylglucosamine; UDP-GlcNAc-EP, UDP–N–acetylglucosamine enolpyruvate; UDP-MurNAc, UDP–N–acetylmuramic acid; mDAP, meso–diaminopimelic acid; G+, Gram–positive bacteria; G–, Gram–negative bacteria. The difference in peptide composition between G+ and G– is highlighted in grey circles. Figure adapted from (El Zoeiby *et al.*, 2003).

**B**
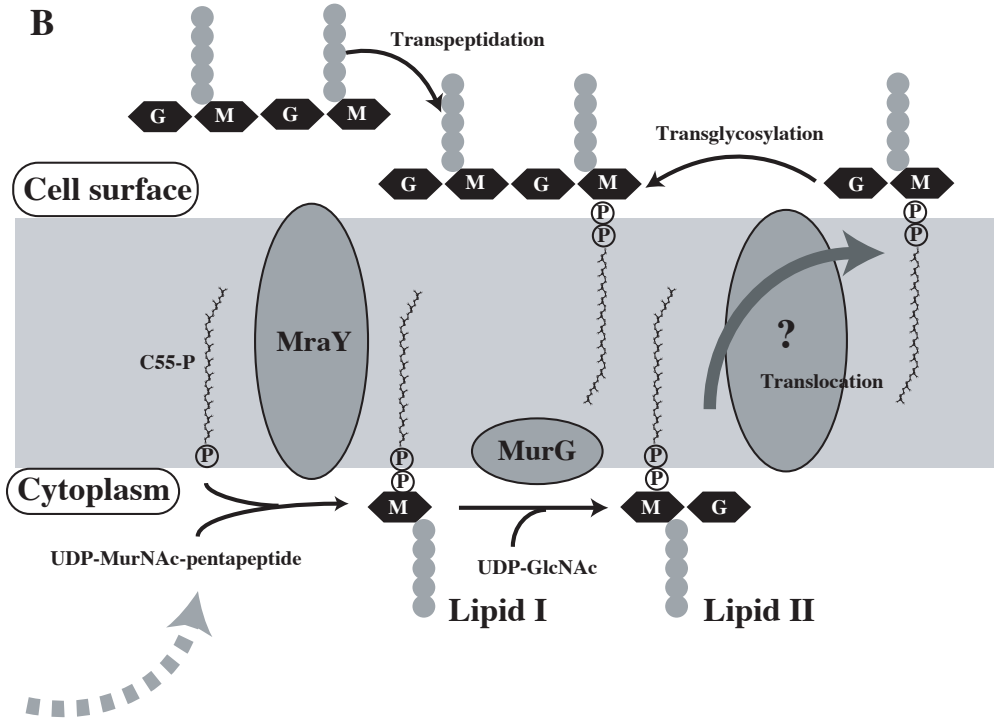


**Figure 4 (continued). B.** Attachment of UDP-MurNAc–pentapeptide onto the membrane (grey area) carrier, phosphoryl undecaprenyl (C55–P) to the formation of the extracellular peptidoglycan matrix. The pentapeptide is shown in five grey circles.

eight nisin and four lipid II molecules with a pore size of about 2-2.5 nm. However, a higher resolution picture is urgently needed to describe the pore complex in details. In **Chapter 3** the structural details of the initial stage of the pore–forming process, namely the recognition of lipid II by nisin, are addressed. The recognition of lipid II is mainly achieved *via* hydrogen bonding between the backbone of the N-terminal part of nisin and the pyrophosphate moiety of lipid II. The pyrophosphate cage, as proposed from the solution structure of the nisin/lipid II complex, is probably a common structural motif amongst the lipid II–binding type–A lantibiotics. It also provides a very prominent structural template for the development of future antibiotics because of the essential role played by the pyrophosphate moiety of lipid II in cell wall synthesis.

### Interactions between mersacidin and lipid II

Mersacidin is an example of type–B lantibiotics (Figure 5), which are generally shorter in primary sequence and exhibit compact and globular 3D structures. Mersacidin contains four lanthionine rings

---

⋆ Hasper, H.E. de Kruijff, B., and Breukink, E., submitted.

§ Breukink, E., personal communication.

† Hsu, S.–T.D. *et al.*, unpublished data.

in only twenty residues. It is highly hydrophobic with only one negatively charged residue, Glu17 (Bierbaum *et al.*, 1995). High resolution structures of mersacidin were solved in methanol both by NMR spectroscopy (Prasch *et al.*, 1997) and X-ray crystallography (Schneider *et al.*, 2000), the latter providing the only crystal structure amongst the lantibiotic family to date. Mersacidin shares some degree of primary sequence homology with other examples of type-B lantibiotics, actagardine and lacticin A1, especially in the conserved consecutive lanthionine rings and the glutamic acid residue (Figure 5). The primary sequence conservation also results in structural homology because of the confined ring structure (Zimmermann and Jung, 1997; Martin *et al.*, 2004). The conservation of this structural element, and of Glu17 in particular, has lead to the hypothesis that it is preserved as a target binding motif. Mersacidin recognises lipid II and, in contrast to the pore-forming activity of type-A lantibiotics such as nisin, interferes with transglycosylation in cell wall synthesis *via* complexation to lipid II (Brötz *et al.*, 1995; Brötz *et al.*, 1997; Brötz *et al.*, 1998a).

We have applied NMR spectroscopy to study the interaction between mersacidin and lipid II in DPC micelles (**Chapter 4**) (Hsu *et al.*, 2003). To our surprise initially, the results show large environment–
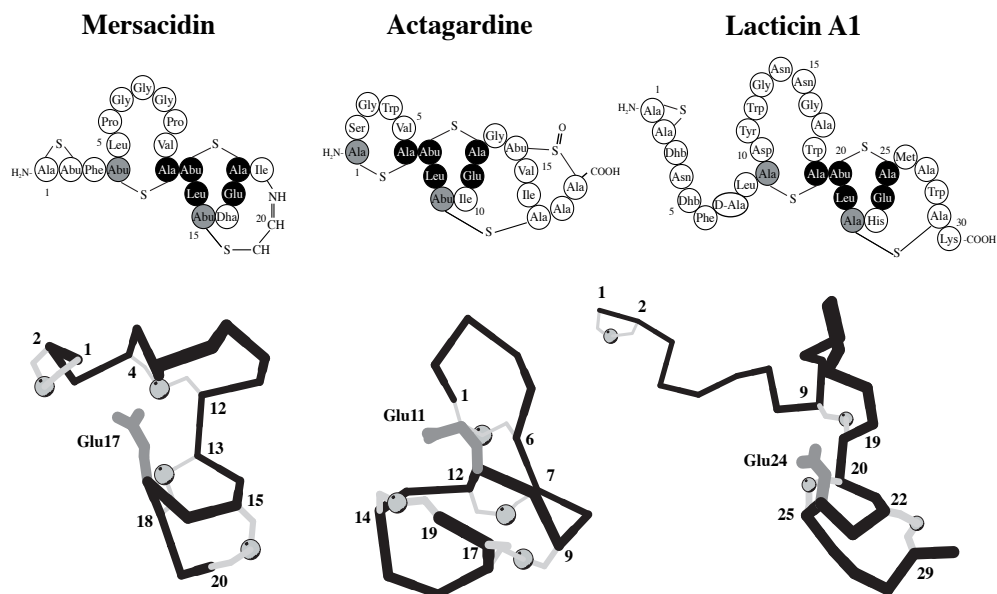


**Figure 5.** Mersacidin, actagardine and lacticin A1. The primary structures and lanthionine linkages are shown on top with the conserved residues in black and the residues that are conserved for the lanthionine linkages in grey. The 3D structures (bottom) are superimposed on the Cα atoms of the conserved residues (seven atoms) with respect to the structure of mersacidin (root-mean-squared deviations are 2.45 and 0.95 Å for actagardin and lacticin A1, respectively). The backbones are shown as Cα traces (black) with the lanthionine linkages (grey lines for side chains and spheres for the sulphur atoms) labelled with the corresponding residue numbers. The conserved glutamic acid side chains are shown in thick grey lines. The coordinates are taken from a representative structure of the structure ensemble of mersacidin (PDB entry: 1QOW, solved by X-ray crystallography with six isoforms in the unit cell), actagardine (1AJ1, NMR structure) and lacticin A1 (NMR structure, Vederas, C. J. personal communication). The figures were generated using Molscript (Kraulis, 1991).

dependent conformational changes, which demonstrate the structural variability of mersacidin in spite of the seemingly rigid lanthionine structure. In fact, the structural rigidity provided by the lanthionine linkages is limited. As is demonstrated in our NMR study, changing the backbone torsion angle at the hinge region can effectively modulate the charge distribution in the 3D structure in response to the hydrophobicity of the surrounding environment. A functional role for the charged Glu17 and the N-terminus is proposed, based on their increase in accessibility upon addition of lipid II: lipid II recognition is driven by electrostatic interactions. This structure-based hypothesis has been later confirmed by site-directed mutagenesis (Szekat *et al.*, 2003). Although little is currently known about the structure of the mersacidin/lipid II complex, partly due to the weak binding affinity, we have demonstrated that mersacidin is able to adapt different conformations that are functionally important.

## Use of NMR in the studies of lantibiotics and lipid II interactions

In the past three decades, advances of solution NMR spectroscopy have opened doors for biologists to investigate intermolecular interactions in complexes of large biomolecules (Fiaux *et al.*, 2002). The leap in the upper limit of NMR-reachable molecular sizes (current record is the complete backbone assignment of a protein with 723 residues (Tugarinov *et al.*, 2002)) was made possible by innovative isotopic labelling (Gardner and Kay, 1998), NMR methodology (Pervushin *et al.*, 1997; Tjandra and Bax, 1997; Tjandra *et al.*, 1997; Riek *et al.*, 1999; Sattler *et al.*, 1999; Pervushin, 2000) and cryo-probe technology. This section, however, will be limited to the description of methodologies applied in this thesis.

## NMR titration and chemical shift perturbation

A general feature of NMR spectroscopy is that the observed resonance frequencies depend on local environments of individual nuclei. The position of resonance frequencies in an NMR spectrum, called *chemical shift*, is very sensitive to the chemical environment. It is frequently used as an indicator of interaction in NMR titration experiments to pinpoint the interaction site. In a two-site exchange situation, a given nucleus exchanges with rate constant $k$ (expressed in $s^{-1}$) between two magnetically distinct sites, corresponding to the free and bound states, with resonance frequencies separated by a chemical shift difference $\Delta\Omega$ (expressed in Hz). The position and linewidth of the observed signal are determined by the relative size of the exchange rate and the chemical shift difference between the two states. In a fast exchange regime, $k >> \Delta\Omega$, a population weighted signal of the free and bound forms can be observed. This is particular useful for deriving the position of the bound form signal by following the population weighted signals during the course of a titration (*e.g.* mersacidin/lipid II interaction in **Chapter 4**). The binding affinity $K_d$ under NMR conditions, assuming a zero order reaction, can also be derived from the chemical shift perturbation $\Delta\delta$ as a function of the ligand/protein molar ratio. On the other hand, in a slow exchange regime, $k << \Delta\Omega$, two distinct resonances are observed, corresponding to the free and bound form signals. As the concentration of the interaction partners increases, the free form signal decreases and vanishes until the bound form is fully populated (*e.g.* nisin/lipid II interaction in **Chapters 2 and 3**). *Coalescence* occurs when the exchange rate is of the same order of the chemical shift separation of the two sites $k \sim \Delta\Omega$. This case, termed intermediate exchange, leads to line broadening and possible peak disappearance in a spectrum because of poor signal-to-noise ratio. Very often, this undesired situation can be circumvented by changing sample temperature so that the exchange rate is shifted to a fast or slow exchange regime.

In nature, biomolecular complexes are often found transiently because of the need for regulation. The lack of a stable complex is a common obstacle for structure determination by X-ray crystallography and NMR spectroscopy. However, obtaining chemical shift perturbations of transient complexes is rather straightforward (Crowley and Ubbink, 2003). After backbone assignment of the $^1$H-$^{15}$N heteronuclear single quantum coherence (HSQC) spectrum, which provides a finger print of each backbone amide group, the interaction site can be immediately identified by monitoring the shift of peak positions. Chemical shift perturbations combined with 3D structure of proteins are routinely used in NMR studies to map binding sites (Zuiderweg, 2002). The use of this information has been extended to the modelling of protein–protein complexes, which was proven to be advantageous compared to purely *ab initio* docking approaches (Palma *et al.*, 2000; Dominguez *et al.*, 2003).

### Temperature coefficient analysis

The chemical shift of a peptide backbone amide proton ($H^N$) is very sensitive to hydrogen bonding, which has a direct impact on the electron structure of this proton. The formation of a hydrogen bond with the amide proton is usually observed by a large downfield shift in NMR spectra along the proton dimension (Baxter and Williamson, 1997; Cierpicki and Otlewski, 2001). The presence of stable hydrogen bonds within a biomolecule is commonly inferred from hydrogen–deuterium (H/D) exchange experiments as the intramolecular hydrogen bonded amide protons have a lower exchange rate with the bulk solvent (Englander and Mayne, 1992; Englander and Krishna, 2001). The presence of a hydrogen bond can also be experimentally confirmed by measurement of scalar coupling across the hydrogen bond, which is originated from partial covalency (see below). For small peptides, however, the lack of secondary and tertiary structure leaves the backbone amide protons unprotected and therefore H/D exchange experiments are not useful: all amide protons will be exchanged by deuterium within the dead time (in a range of 10-20 minutes) of the experiment. Instead, the temperature coefficient of chemical shifts, $\Delta\delta/\Delta T$, can be used to describe the extent of involvement in hydrogen bonding of those amide protons in peptides.

The temperature coefficient gives the linear dependency of a chemical shift of the amide proton with respect to the sample temperature. The idea behind this is the assumption that the hydrogen bond length between the amide proton and hydrogen bond acceptor is temperature dependent. On increasing the temperature, the hydrogen bond weakens and lengthens as a result of thermal expansion of the system (Baxter and Williamson, 1997). The hydrogen bonded amide proton chemical shift moves relatively upfield with increasing temperature. The extent of hydrogen bond length increase differs between inter- and intramolecular hydrogen bonds. An amide proton that is hydrogen bonded to the bulk water oxygen is more sensitive to the temperature change and its temperature coefficient is of the same order as that of the water signal. An intramolecular hydrogen bond within a biomolecule, on the other hand, is more inert to the thermal expansion and hence its temperature dependency is smaller. However, the temperature coefficient is not definitive in indicating the formation of hydrogen bonds (Baxter and Williamson, 1997) because it is also sensitive to other effects such as the sequestering from bulk solvent due to hydrophobic contacts. Therefore, temperature coefficient analysis is usually used in a comparative manner to monitor the degree of protection, not as conclusive evidence of hydrogen bonding, in different environments or upon complex formation (*cf.* **Chapters 2 and 3**).

## Cross-hydrogen bond scalar coupling

Hydrogen bonds play an important role in biomolecular structure and function. They are the determinant of protein secondary structure (Pauling and Corey, 1953) and are the origin of base pairing in nucleic acids (Corey and Pauling, 1956). The concept of covalency of the hydrogen bond was first appreciated by Linus Pauling (Pauling, 1935; Pauling, 1967) but experimental proof only became available in recent years (Dingley and Grzesiek, 1998; Pervushin *et al.*, 1998; Isaacs *et al.*, 1999). Detection of a hydrogen bond by solution NMR exploits the underlying nature of scalar coupling across the hydrogen bond. Since then, various types of hydrogen bonds have been identified in biological systems (Dingley *et al.*, 2001).

In nisin/lipid II interactions, the pyrophosphate group of lipid II is essential for recognition as observed by solid state NMR spectroscopy[‡]. The advantage of phosphorous detection by NMR is that $^{31}P$ nuclei have a natural abundance of 100%, which obviate the need for isotope labelling of lipid II. Based on the concept of J-coupling modulated NMR spectroscopy, hydrogen bonds between phosphates and peptide amides have been previously observed (Löhr *et al.*, 2000; Mishima *et al.*, 2000). In a nisin/lipid II complex where the amide nitrogen ($^{15}N$) and the phosphorous ($^{31}P$) nuclei are in separate molecules, namely nisin and lipid II, a nitrogen to phosphorous scalar coupling can only be established across intermolecular hydrogen bonds. We employed the $^{31}P$-edited $^1H$-$^{15}N$ HSQC experiment for the detection of intermolecular hydrogen bonds and the result is clearly illustrated in the difference spectrum in which only the contribution of the nitrogen to phosphorous scalar coupling is observed (Figure 6). This has served as key evidence for the intermolecular interactions between nisin and lipid II.

## Pulse field gradient diffusion measurement

Another tool for monitoring the formation of biomolecular complexes in solution is the use of pulse field gradient diffusion NMR spectroscopy. This method measures the effective hydrodynamic radius $R_h$ of a molecule or complex of interest based on the translational diffusion rate $D_{trans}$ using the so-called PG–SLED experiment (Pulse Gradient Stimulated echo Longitudinal Encode-Decode) (Morris and Johnson, 1993; Altieri *et al.*, 1995). The translational diffusion rate $D_{trans}$ can be derived from the signal decay rate as a function of the applied gradient strength. It can also be derived from the relative diffusion rate with respect to an internal standard with a known hydrodynamic radius. The translational diffusion rate in turn gives the apparent hydrodynamic radius of the molecule $R_h$ by the Stoke-Einstein relation

$$D_{trans} = \frac{k_B T}{6\pi\eta R_h} \qquad (2)$$

where $k_B$ is Boltzmann's constant, $T$ the absolute temperature and $\eta$ the solvent viscosity. It is worth noting that the assumption of the molecule as a rigid and spherical particle is made in this derivation. Nevertheless, this approximation has been successfully tested in several studies in combination with small angle X-ray scattering (SAXS) (Jones et al., 1997; Wilkins et al., 1999). We applied this methodology in conjunction with $^{15}N$ relaxation measurements to discern the contribution between complex formation and solvent viscosity change during the complex formation between mersacidin and lipid II in the presence of DPC micelles. Further details are presented in **Chapter 4**.

---

‡ Bonev, B., Breukink, E., Swiezewska, E., de Kruijff, B. and Watts, A., submitted.
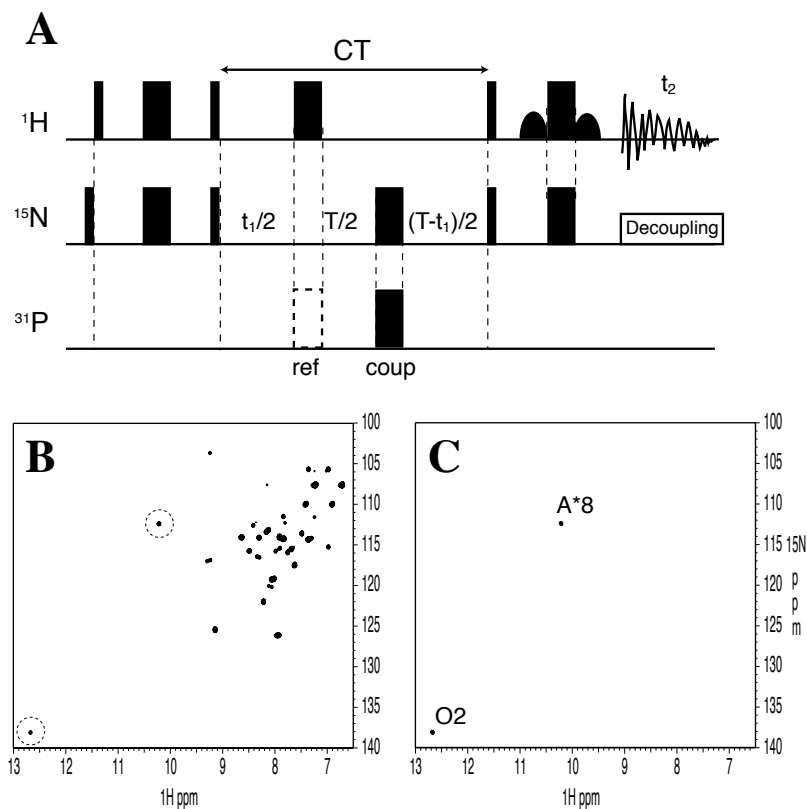
**Figure 6.** Detection of intermolecular hydrogen bond by NMR. **A.** Pulse sequence of the $^{31}$P-edited constant time (CT) $^{1}$H-$^{15}$N HSQC spectroscopy; **B.** reference spectrum; **C.** difference spectrum ($^{31}$P-coupled and decoupled) of $^{15}$N labelled nisin in complex with 3LII. The thin and thick bars correspond to 90° and 180° pulses for individual nuclei. The reference spectrum (the inversion pulse of $^{31}$P shown in dashed rectangular is synchronised along with the refocusing pulse of $^{1}$H during constant time) removes the cross hydrogen bond scalar coupling $^{3}$J$_{NP}$ during the t$_1$ evolution (CT=120 ms). The difference in peak intensity between the reference and coupled spectra implies hydrogen bond modulation during the constant time, shown in **c**. The amide groups of O2 and A*8 of nisin are hydrogen bonded to the pyrophosphate of 3LII (for details see **Chapter 3**).

## Structure determination of complexes by NMR

Biomolecular interactions can be best described when a stable complex is available for structure determination by NMR spectroscopy. Intermolecular NOEs provide the key information to describe the structure of a bimolecular complex. The NOE is the manifestation of spin dipole–dipole cross relaxation (Neuhaus and Williamson, 1989). The NOE intensity ($I_{ij}$) of an isolated spin pair $i$ and $j$ is inversely proportional to the sixth power of the spatial separation between the spin pair

$$I_{ij} \propto r_{ij}^{-6} \tag{3}$$

where $r_{ij}$ the distance between the two nuclei. However, in large systems with many protons close in space, like most biomolecules, spin diffusion is involved and hence accurate extraction of distances from NOEs

between a large set of atoms requires a relaxation matrix calculation (Bonvin *et al.*, 1991; Bonvin, 1993). Typically the distance information contained in homonuclear $^1$H-$^1$H NOEs is limited to a maximum of 5-6 Å, although longer distances up to 8 Å can be derived in a perdeuterated molecule (Koharudin *et al.*, 2003). A more serious problem in determining the solution structure of peptides is dynamical averaging, which stems from their inherent flexibility. For this reason, the derivation and use of distances from observed NOEs requires the consideration of ensemble averaging (Bonvin *et al.*, 1993; Bonvin and Brünger, 1995; Bonvin and Brünger, 1996) or time-averaging (Torda *et al.*, 1990; Nanzer *et al.*, 1994).

One of the time-limiting steps in solution structure determination is the assignment of a large number NOEs. Intermolecular NOEs are usually difficult to identify amongst the dominating intramolecular NOEs. With the help of isotope labelling, the combination of labelled and unlabelled components of the complex permits the extraction of intermolecular NOEs from isotope-filter experiments (Otting and Wüthrich, 1989; Slijper *et al.*, 1996). Other approaches such as saturation transfer of magnetisation across the complex interface have been proposed (Takahashi *et al.*, 2000). These are useful tools for separating the intra- and intermolecular NOEs. Yet, as the molecular size increases, the total number of NOEs increases accordingly. Typically a well-defined protein structure yields 15-30 NOEs per residue, *i.e.*, >1000 NOEs for a moderate protein domain of about 100 residues. Automated assignment methods for NOESY-type spectra have been devised to shorten the labour-intensive process and reduce the potential errors originating from biased manual assignment (Nilges *et al.*, 1997; Linge *et al.*, 2001; Herrmann *et al.*, 2002). Nevertheless, complete assignment of side chain resonance frequencies is required and spectral overlap can limit the performance of the assignment. Although the complexity of NOEs is usually not a main concern for small peptides, the advantage of automation is at any rate favoured for structure determination. Furthermore, several parameters have been developed, such as residual dipolar coupling (RDC) (Tjandra and Bax, 1997) and backbone relaxation anisotropy (Tjandra *et al.*, 1997), in order to provide long-range order to circumvent the limitation of short-range NOE-based structure calculation. These, however, are usually not applicable for small peptides that lack well defined globular structure. Perhaps one useful application of these orientational restraints is the use in defining the orientation and curvature of a membrane binding peptide. The structure of these peptides can sometimes be affected by the interaction with various membrane mimicking environments, such as micelles and bicelles, but these changes are difficult to be discerned by NOE-based approaches (Chou *et al.*, 2002). Finally, the development of better force fields and refinement protocols in an explicit solvent model also contributes to the improvement of the quality of solution structures by NMR (Linge *et al.*, 2003).

## Part III. HIV-1 Env gp120 and CD4

*HIV and AIDS*

The epidemic of Acquired Immune Deficiency Syndrome (AIDS) has become a major challenge of modern medicine since the first reported case in 1982. While current treatments are focusing on containing the disease, at best, the development of an anti-HIV vaccine is a top priority in laboratory research (Baltimore, 2002; Gaschen *et al.*, 2002). AIDS is attributed to the infection of type-1 Human Immunodeficiency Virus (HIV-1), which deteriorates the natural defence system against infection. According to a report undertaken by the World Health Organisation (WHO, http://www.who.int) in December 2003, there are currently more than forty million people living with HIV/AIDS worldwide.

In the year 2003 alone, five million new cases of HIV infection have been reported and three million deaths have been attributed to AIDS. Southern Africa is the worst hit area in the world: 65% of the cases occur in this area and 40% of the population of Bostswana and Swaziland is infected.

A vaccine, by definition, is a substance put into the body to elicit immune response by which pathogen-specific antibodies can be produced to neutralise the pathogen. Thereby the infection can be inhibited. To devise an effective vaccine, we need to know the structure of the pathogen – because vaccines are structural fragments or mimics of pathogens that are recognised by the immune system (Wyatt *et al.*, 1998; Letvin *et al.*, 2002).

HIV-1 is a type 1 RNA retrovirus, which hijacks the bio-machinery of host cells for its reproduction. This requires the delivery of the viral genetic material into the host by direct fusion of the viral and host plasma cell membranes so that the viral genome can be integrated in that of the host cell (Figure 7) (Eckert and Kim, 2001a; Colman and Lawrence, 2003; Gallo *et al.*, 2003). Two decades of the best efforts have yielded a wealth of structural data on HIV-1 proteins (Figure 8) (Turner and Summers, 1999). The structural knowledge on HIV-1 viral proteins forms the basis of vaccine design efforts against HIV-1. Inhibitors of some important enzymes, such as HIV-1 protease, have been developed from structural information (Wlodawer *et al.*, 1989; Prabu-Jeyabalan *et al.*, 2002); some of these inhibitors are now in clinical use as drugs (Cameron *et al.*, 1999). In the context of vaccine design, the envelop proteins (Env), gp120 and gp41 (Figures 7 and 8), have received the most attention because they are located on the surface of the virion and are exposed to immune attack. Furthermore, the recognition of the envelope protein gp120 and the human helper T-cell receptor, CD4, is the initial step that triggers subsequent co-receptor binding (Davis *et al.*, 1992; Poignard *et al.*, 2001), insertion of the membrane fusion peptide of gp41 and finally fusion of the viral and host cell membranes (Melikyan *et al.*, 2000; Gallo *et al.*, 2001; Colman and Lawrence, 2003). Thus, a vaccine that can block the initiation process can effectively prevent viral entry and therefore inhibit infection (Chan and Kim, 1998; Eckert *et al.*, 1999; Eckert and Kim, 2001a; Eckert and Kim, 2001b; Root *et al.*, 2001; Barbato *et al.*, 2003).

*The envelope glycoprotein of HIV-1*

HIV-1 Env is synthesized as a polyprotein (gp160). It folds and trimerizes in the endoplasmic recticulum (ER) where ten disulphide bonds are formed and ~30 N-linked glycans are attached (Leonard *et al.*, 1990). Env is subsequently cleaved in the Golgi complex into a soluble part, gp120 (SU), and a transmembrane part, gp41 (TM) (Moulard and Decroly, 2000) (Figure 8). These subunits are assembled on the virion surface in a trimeric spike structure with three gp120 non-covalently attached to a stalk structure consisting of three gp41 (Wyatt and Sodroski, 1998). Both gp120 and gp41 are heavily glycosylated, the glycans representing a third of the total mass. Correct disulphide bonding and glycosylation are required to pass the quality control machinery of the ER and errors in these modifications will lead to retention in the ER lumen (Land *et al.*, 2003). The 3D structures of gp120 (Kwong *et al.*, 1998; Kwong *et al.*, 2000a) and gp41 (Chan *et al.*, 1997; Weissenhorn *et al.*, 1997) were determined by X-ray crystallography. The information obtained was, however, limited to the core regions. A large portion of both proteins was truncated for crystallisation due to the intrinsic flexibility of several loops and of the termini. Most glycans were also absent in the structures. From immunogenic (Sattentau and Moore, 1991; Moore and Sodroski, 1996) and thermodynamic studies (Myszka *et al.*, 2000), we know that the structure of gp120 undergoes a substantial rearrangement upon binding to CD4 and to the co-receptor, primarily CCR5 or
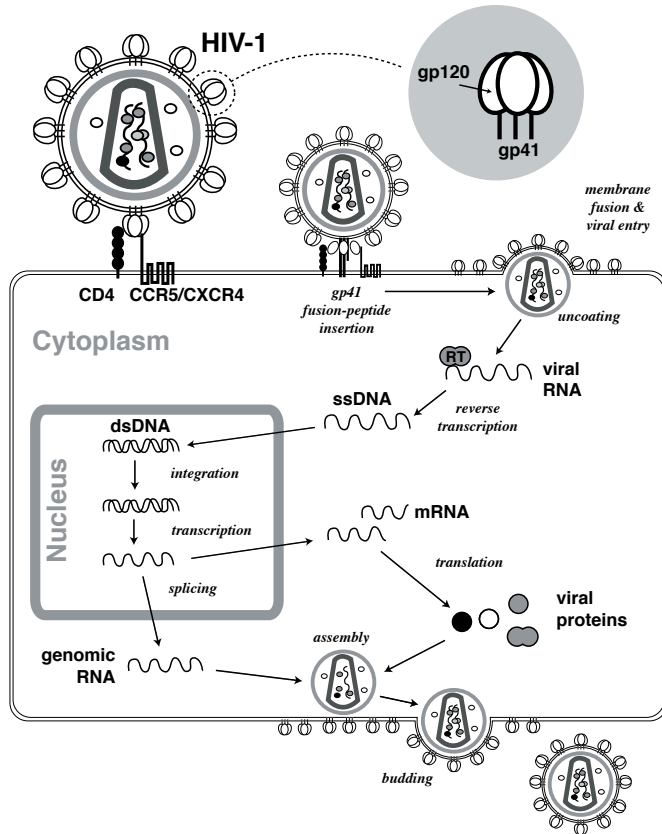
**Figure 7.** Life cycle of HIV-1. The first step of viral entry is the engagement of the trimeric envelope proteins gp120 and gp41 (indicated in dashed circle) with the receptor CD4 and subsequently the chemokine co-receptor CCR5 or CXCR4.

CXCR4 (Moore and Binley, 1998; Berger *et al.*, 1999). The structural change in gp120 initiates a more dramatic change in gp41 from a metastable, pre-fusogenic state, to an extended coiled-coil, pre-hairpin state, in which the N-terminal fusion peptide is inserted into the host cell membrane, and finally to a six-hairpin bundle, which drives the membrane fusion (Chan and Kim, 1998; Melikyan *et al.*, 2000; Eckert and Kim, 2001a). Unfortunately the mechanism of the whole fusion process is still poorly understood. We can only intrapolate models based on the structures corresponding to the two end states of the fusion process, the gp120/CD4 complex providing the starting structure and the ectodomain of gp41 the end state.

The third hypervariable loop (V3) in gp120 (Figure 8) is another focus in the field of vaccine design because it is found to contain the principle neutralising domain (PND) (Gorny *et al.*, 1993). It is also crucial for the recruitment of the co-receptor, most likely because of its highly basic nature that confers an attraction gradient to the generally acidic co-receptors (Rizzuto *et al.*, 1998). Despite its variable sequence composition, a well-conserved region of the V3 loop, GPRG, is found between two antiparallel
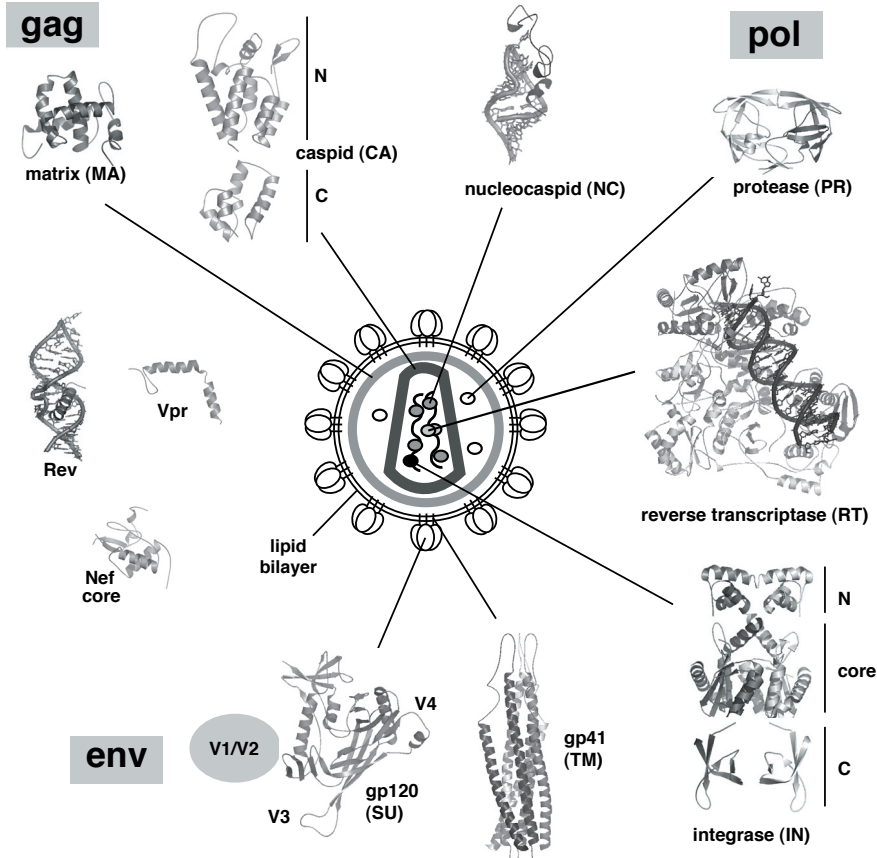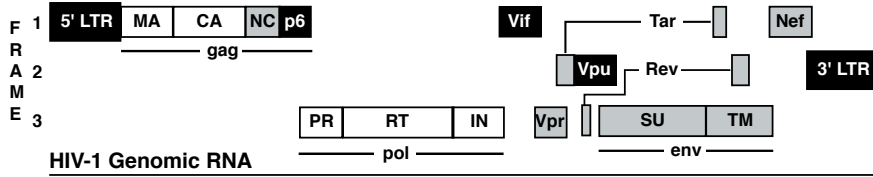
**Figure 8.** Genome and encoded proteins of HIV-1. Three reading frames of the viral genome (Top panel) encode the viral structural capsid proteins (*gag*), functional enzymes (*pol*), envelope proteins (*env*) and other accessory and regulatory proteins: Vpr, Vpu, Vif, Rev and Tar. A schematic viral particle is shown below with the corresponding 3D structures. Those whose structures are not available or are only partially available are highlighted in black and grey boxes in the corresponding reading frames (Top). The figures are generated using Pymol with PDB entries of 1TAM (matrix, **MA**), 1AK4 (N-terminal caspid, **CA**), 1AFV (C-terminal caspid, **CA**), 1A1T (nucleocaspid, **NC**, in complex with its recognition RNA element), 1HXW (protease, **PR**), 1RTD (reverse transcriptase, **RT**, in complex with the DNA template), 1KGY (N-terminal and core domains of integrase, **IN**), 1IHV (C-terminal integrase), 1ENV (trimeric ectodomain of gp41, **TM**, in the post-fusion state), 1G9M (core gp120, **SU**, with the V3 and V4 loop modelled by Swiss–Model (Guex and Peitsch, 1997) and the position of the truncated V1/V2 in grey circle), 1CEU (**Vpr**), 1AUZ (**Nef** core) and 1ETF (**Rev** fragment in complex with its response RNA element).

beta strands. Its structure is heterogeneous, forming various types of turn structures. Solution NMR studies have shown that the V3 loop in isolation is flexible and with no defined structure (Tolman *et al.*, 1993; Vranken *et al.*, 1995; Zvi *et al.*, 1995; Catasti *et al.*, 1996). Because of the sequence heterogeneity and the structural polymorphism, the monoclonal antibodies (MAbs) that have been elicited against V3 loops of various clinical isolates usually lack a general neutralisation capacity. This is presumably because of the structural polymorphism of the V3 loop in complex with antibodies as has been demonstrated by NMR spectroscopy (Tugarinov *et al.*, 1999; Tugarinov *et al.*, 2000; Sharon *et al.*, 2003) and X-ray crystallography (Ghiara *et al.*, 1994; Ghiara *et al.*, 1997; Stanfield *et al.*, 1999). Only recently was an antibody found that recognises the backbone of the conserved GPRG region of the V3 loop in an unique triple strand configuration, leading to broad neutralisation against HIV-1 (Stanfield *et al.*, 2004).

Monomeric gp120 is commonly used as a model system for experiments *in vitro*. In fact, it also exists *in vivo* since shedding of gp120 from the virion surface can occur spontaneously (McKeating *et al.*, 1991; Layne *et al.*, 1992) or upon binding of soluble CD4 to Env (Moore *et al.*, 1990; Dimitrov *et al.*, 1992). The structural difference between the monomeric and trimeric gp120 most likely provides a *conformational camouflage* for HIV-1 to evade immune neutralisation (Jardetzky, 2002; Kwong *et al.*, 2002). The importance of the trimeric state of gp120 for efficient membrane fusion has enlightened the development of gp120-based vaccines. Nonetheless, many pieces of the puzzle are still missing. What does free gp120 look like? How are the three monomeric gp120 molecules organised into a trimeric structure? What are the roles of those hypervariable loops that are missing in the solved structures? And how do gp120 and gp41 interact and form the heterohexameric spike structure? Some of these issues have been addressed by molecular modelling of trimeric gp120 based on the structure of the monomeric gp120 (Kwong *et al.*, 2000b), by cooperative binding of two antibodies that separately recognise the CD4-binding epitope and the V1-V3 loops (Zwick *et al.*, 2003), and by disulphide bond engineering between gp120 and gp41 (Sanders *et al.*, 2000). However, little is understood about the role of dynamics in this envelope-mediated membrane fusion process so far.

### CD4-mediated HIV-1 viral entry

The modes of membrane fusion of influenza virus, paramyxovirus and HIV-1 belong to type 1 viral membrane fusion (Colman and Lawrence, 2003). These viruses use the common spike structure of the envelope proteins, all in trimeric form – the haemagglutinin (HA) for influenza virus and the Env for HIV-1 – for host cell recognition followed by insertion of the hydrophobic fusion peptides. Membrane fusion requires a large conformational change in at least one of the envelope proteins. The fusion of influenza virus is triggered by a local pH decrease as it is being internalised by endosome, whereas HIV-1 fusion is pH independent and solely mechanical. Furthermore, the kinetics of the fusion process conducted by the envelope of HIV-1 is quite slow, with a half time of about 20 minutes to reach maximum fusion, as inferred from a study using Simian Immunodeficiency Virus (SIV) (Gallo *et al.*, 2003). In this regard HIV-1 and influenza virus are very different, as the kinetics of membrane fusion of the latter is less than half a minute.

CD4 is a receptor molecule that is primary presented on the surface of the helper T-cell. It increases the cooperative association between a T cell and an antigen-presenting cell by interacting with non-polymorphic portions of the complex between class II major histocompatibility complex (MHC) and T-cell receptor molecules. CD4 is also the target of HIV-1. Its binding to gp120 initiates viral entry

into the host cell. CD4 is anchored onto the membrane with four extracellular immunoglobulin–like domains, namely D1-D4. Both class II MHC and HIV-1 gp120 bind to the distal D1 domain with the critical residue of CD4, Phe43, centred in the interface (Wang *et al.*, 1990; Kwong *et al.*, 1998; Kwong *et al.*, 2000a). The association of the viral envelope protein with CD4 is more efficient than that of the physiological receptor by more than four orders of magnitude. This strong binding affinity between gp120 and CD4 overwhelms the generally weak physiological binding necessary for cellular regulation (Wang, 2002). Hence, viral entry can be efficiently achieved without much frustration from competitors. The strong viral binding affinity is due to the large buried surface area (BSA) between gp120 and CD4 (Kwong *et al.*, 1998; Kwong *et al.*, 2000a), which is twice as large as that in the structure of the class II HMC and CD4 complex (Wang *et al.*, 1990) (Figure 9).

### Conformational change during gp120/CD4 complexation

Recently, a thermodynamic study indicated a large gain in enthalpy ($\Delta H$), accompanied by a large loss in entropy ($\Delta S$) upon complex formation. This results in a small net free energy change ($\Delta G = \Delta H - T\Delta S$)



| 2107 | Total BSA (Å$^2$) | 1009 |
|---|---|---|
| 10$^{-8}$ | Affinity (M) | 2x10$^{-4}$ |
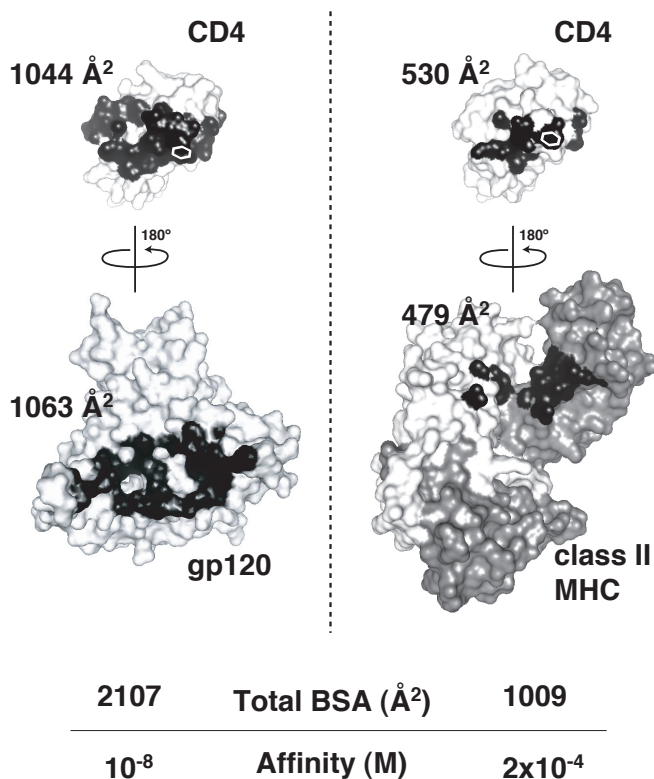
**Figure 9.** Comparison between buried surface area (BSA) and binding affinity. The atoms in close contact with the counterpart (interatomic distance < 5Å) are coloured in black. The phenyl ring of CD4-Phe43 is indicated in white lines. The class II MHC consists of two separate chains, coloured in white and grey. The corresponding PDB entries of gp120/CD4 and the class II MHC/CD4 complexes are 1G9M and 1JL4, respectively. The BSAs were calculated from the crystal structure using NACCESS with a probe radius of 1.4 Å.

after cancellation of two large contributions (Myszka *et al.*, 2000). The removal of as much as 40% of the total sequence, including the large flexible loops V1–V3 and the N- and C-termini of gp120, has little effect on the kinetics and thermodynamics of CD4 binding. The entrapment of coordinated water is also unlikely to be involved in the loss of entropy (Myszka *et al.*, 2000). Hence, the energetically unfavourable entropy loss, indicative of loss of flexibility, *i.e.*, the formation of rigid structures, must occur within the core of gp120 and CD4. Based on these observations in conjunction with other biochemical data, it has been suggested that an extensive structural rearrangement must occur during the CD4 binding process. A generally accepted model is that a four-stranded *bridging sheet* is formed or stabilised upon complexation, pulling the inner and outer domains of gp120 together to form a recessed Phe43 receptive cavity to accommodate the hemisphere of the D1 domain of CD4 (Figure 10). These structural rearrangements may, in part, account for the observed changes in enthalpy ($\Delta H$) and heat capacity ($\Delta C_p$).
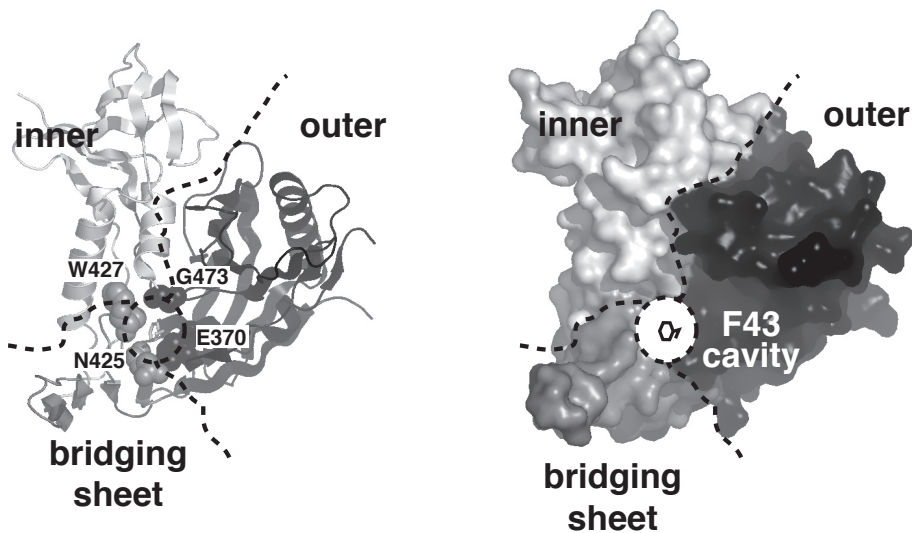


**Figure 10.** Structure and sub-domains of core gp120. The ribbon presentation (left) and van der Waals surface (right) of core gp120 are coloured in light grey, black and dark grey for the inner and outer domains and the bridging sheet, respectively. Four residues in gp120 that show hydrophobic contacts with the CD4-Phe43 are shown in spheres in the ribbon representation.

## MD simulations of gp120 and CD4

To date, the structure of the complex of core gp120, CD4 and the CD4-induced antibody (CD4i) constitutes the primary structural information on gp120 (Kwong *et al.*, 1998; Kwong *et al.*, 2000a). A number of structural studies limited to the V3 loop fragments and their neutralising antibodies have been reported (Ghiara *et al.*, 1994; Ghiara *et al.*, 1997; Stanfield *et al.*, 1999; Tugarinov *et al.*, 1999; Tugarinov *et al.*, 2000; Sharon *et al.*, 2003; Stanfield *et al.*, 2004). Little is known, however, about the structure and dynamics of gp120 in the free form. We therefore carried out MD simulations in order to access structural and dynamical differences between the free form of gp120 and its complex with CD4.

MD simulations solve Newton's laws of motion for a system of $N$ interacting particles

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i \ , i = 1\ldots N. \tag{4}$$

where $m_i$ is the mass of particle $i$ , $r_i$ its position vector and $F_i$ the vector describing the forces acting on particle $i$.

The force $F$ is given by the negative of the derivative of a potential function $V(r_1, \ r_1, \ldots, \ r_N)$ describing the interactions between particles in the system.

$$F_i = -\frac{\partial V}{\partial r_i} \tag{5}$$

Newton's equations of motions are solved by numerical integration over time steps $\Delta t$ of typically two to four femtoseconds, which leads to the new positions of the particles in the system at time $t+\Delta t$, $r(t+\Delta t)$. By repeating this integration process a large number of times (typically several millions) a trajectory containing the atom coordinates as a function of time can be generated. Analysis of such a trajectory allows to describe the structural and dynamical properties of the system under study.

The potential energy function, $V(r)$, contains energy terms describing the bonded and non–bonded interactions in the system. Bonded interactions typically involve bond stretching, angle bending and rotations around bonds (dihedral angles). Additional energy terms, the so–called improper dihedral angles energy terms, are also often used to maintain the planarity of particular groups, such as aromatic rings and the chirality of some atoms such as the C$\alpha$ atom of L– and D–form.

$$\begin{aligned} V_{bonded} &= V(b)_{bond} + V(\theta)_{angle} + V(\zeta)_{improper} + V(\varphi)_{dihedral} \\ &= \sum_{bond} \frac{1}{2} k_b [b - b_0]^2 + \sum_{angle} \frac{1}{2} k_\theta [\theta - \theta_0]^2 + \sum_{improper} \frac{1}{2} k_\theta [\zeta - \zeta_0]^2 + \sum_{dihedral} k_\varphi [1 + \cos(n\varphi - \delta)] \end{aligned} \tag{6}$$

The potential energy functions describing bond lengths, bond angles and improper dihedrals are usually expressed as harmonic functions with *equilibrium* or *reference* values $b_0$, $\theta_0$ and $\zeta_0$ for ideal bond lengths, bond angles and improper angles, respectively. The potential energy terms describing the rotations around bonds, the dihedral angle energy terms, are typically expressed as trigonometric functions with a periodicity $n$ and a phase shift $\delta$ to allow for multiple minima. Energy penalties are derived based on these functions with the respective force constants, $k_b$, $k_\theta$, $k_\zeta$ and $k_\varphi$.

The non–bonded energy terms describe the electrostatic and van der Waals interactions between atoms. They are usually expressed as Coulombic and Lennard–Jones potential functions, respectively.

$$\begin{aligned} V_{non-bonded} &= V_{Coulomb} + V_{L-J} \\ &= \sum_i \sum_{j \neq i} \left[ \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right] + \sum_i \sum_{j \neq i} 4\varepsilon \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] \end{aligned} \tag{7}$$

where $r_{ij}$ is the distance between atom $i$ and $j$, $\varepsilon_0$ the dielectric constant, $q_i$ and $q_j$ the partial charge of atoms $i$ and $j$, respectively, $\varepsilon$ the well depth and $\sigma$ the collision diameter (the separation distance at which the Lennard–Jones energy is zero) (Leach, 2001).

Several force fields have been developed and parametrized over the years, which are commonly used for biomolecular MD simulations, such as GROMOS (van Gunsteren *et al.*, 1996), AMBER (Pearlman *et al.*, 1991), CHARMM (MacKerell *et al.*, 1998) and OPLS (Jorgensen and Tirado-Rives, 1988). They might differ in their parameters, such as equilibrium values, force constants, partial charges, van der Waals parameters, *etc.*, and in the exact functional form of the potential energy function.

Current MD simulations cannot compute macroscopic properties of a given system based on Avogadro's numbers of atoms. Instead, most classical MD simulations only give access to microscopic properties of one or a few solute molecules and its surrounding environment. Periodic boundary conditions allow to simulate systems in a fluid-like condition where individual molecules can freely diffuse in all directions as they are in fact replicated in a periodic array with a given symmetry. Interactions between solute and solvent, such as protein–water interactions, play an important role in defining the thermodynamic and structural properties of a solute molecule. It is therefore crucial to properly describe solvent effects. All MD simulations described in this thesis have been performed under periodic boundary conditions with explicit solvent using the simple particle charge (SPC) water model (Berendsen *et al.*, 1981).

In **Chapter 5** we describe MD simulations of free gp120, its receptor CD4, and their complex (Hsu and Bonvin, 2004). Substantial conformational changes with a concerted loop contraction were observed in gp120 upon CD4 binding, accompanied by a substantial displacement of the V3 loop. Translating atom positional root-mean-squared fluctuations (RMSF) derived from the MD simulations into temperature factors (B-factors), commonly used for the evaluation of structural mobility and disorder in the crystalline state, agreement could be reached between simulations and experimental findings. A significant result of this study is the description of the differentiated and dynamical modes of interaction between gp120 and CD4 involving long range electrostatic attractions, hydrogen bonds and short range van der Waals contacts. The various interactions are spatially distributed on the complex interface in perfect accordance with the "hot spot" model (Clackson and Wells, 1995). The centre of the gp120/CD4 complex interface is the hydrophobic "knob-and-socket" CD4-Phe43 binding patch surrounded by a large number of residues involved in hydrophilic interactions, hydrogen bonding and salt bridges that holds the two molecules together (Figure 11). The same recognition mechanism is utilised by a potent gp120-specific IgG antibody, b12: a protruding tryptophan residue in the finger-like H3 loop is likely the CD4-Phe43 mimic that binds into the receptive cavity in gp120 (Saphire *et al.*, 2001). Modelling of the gp120/b12 interface suggests a BSA similar to that of gp120/CD4. Not surprisingly, the binding affinity of the gp120/b12 complex (in terms of free energy) is similar to that of the gp120-CD4 complex (Kwong *et al.*, 2002).

## MD analysis beyond the atomic coordinates

Benefiting from the exploding computation power and the continuous improvements in force fields, modern biomolecular simulations can reach longer time scales, and larger systems and, more importantly, provide a more realistic description of the dynamical and structural properties of interest (van Gunsteren and Berendsen, 1990; van Gunsteren *et al.*, 2001; Hansson *et al.*, 2002; Karplus, 2002; Karplus and McCammon, 2003). Examples like spontaneous membrane or vesicle formation (Marrink *et al.*, 2001; Marrink and Mark, 2003) and atomic description of protein folding pathways (Duan and Kollman, 1998; Mayor *et al.*, 2003) have been reported. MD simulations permit us to describe the atomic trajectory
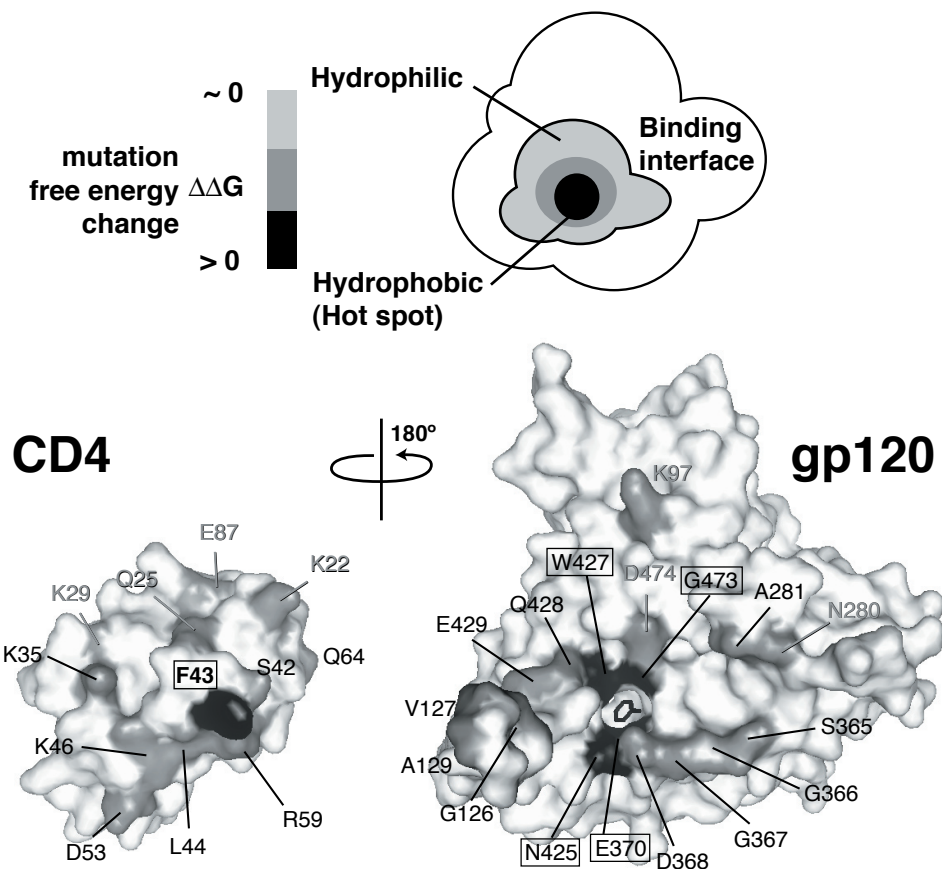
**Figure 11.** "Hot spot" model. A cartoon shows the hot spot for target recognition displayed on the binding site with corresponding double free energy change (ΔΔG) upon mutation coloured from light grey to black. The representative structures of gp120 and CD4 are shown in van der Waals surface with residues involved in hydrophobic and hydrophilic interactions in black and grey, respectively. The residues involved in side chain-side chain hydrogen bonding are labelled with grey letters; those that are involved in side chain-backbone or backbone-backbone hydrogen bonding are labelled with black letters. The labels of the four residues that form the receptive CD4-Phe43 binding cavity are boxed.

of a molecule of interest over time scales ranging from nanoseconds to microseconds. This means that the conformations of the system during this time, *i.e.*, the microstates, can be precisely characterised. In general terms, the trajectory contains the information needed to derive the classical Hamiltonian $H(\boldsymbol{p},\boldsymbol{q})$ of generalised coordinates $\boldsymbol{q}$ and their conjugate momenta $\boldsymbol{p}$. For instance, the Hamiltonian for a classical system of $N$ atoms with coordinate $\boldsymbol{r} \equiv (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$ and momenta $\boldsymbol{p} \equiv (\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_N)$ has the form

$$H(\boldsymbol{p},\boldsymbol{r}) = \sum_i^N \frac{p_i^2}{2m_i} + V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) \tag{8}$$

where $m_i$ is the atomic mass of atom $i$, and $V(\mathbf{r})$ is the potential energy function describing the interactions between atoms. The two terms represent the kinetic and potential energies, respectively. In the canonical ensemble with constant number of particles $N$, constant volume $V$ and temperature $T$ (constant NVT), the Helmholtz free energy $F$ is given by

$$F_{NVT} = -k_B T \ln Z \tag{9}$$

where $k_B$ is Boltzmann's constant and $Z$ the partition function. The partition function for a system of $N$ identical particles is given by

$$Z = \left[ \frac{1}{(h^{3N}N!)} \iint e^{-H(\mathbf{p},\mathbf{r})/k_B T} d\mathbf{p} d\mathbf{r} \right] \tag{10}$$

where $h$ is Planck's constant. It is thus possible, in principle, to derive free energies from simulated trajectories provided that sufficient sampling has been achieved and that complete integration for the sum of states, *i.e.*, calculation of the partition function $Z$, is feasible (van Holde *et al.*, 1998). In practice, however, achieving a *complete sampling* over all energy states is difficult for large complex systems because, usually, classical MD simulations only sample conformations in local energy minima. They do not adequately sample higher energy regions of phase space that make important contributions to the free energy. Instead, the calculation of free energy *differences* between two systems or states is a more feasible task because free energy is a state function, which is independent of path. With appropriate choice of thermodynamic integration or simply by counting different configurations, the relative free energy between two states can be accurately estimated (Leach, 2001). A number of computational methods to derive free energies have been developed (Northrup *et al.*, 1982; van Gunsteren and Berendsen, 1987; Gerber *et al.*, 1993; Kollman, 1993; van Gunsteren *et al.*, 1993; McCarrick and Kollman, 1994; Simonson *et al.*, 2002; van Gunsteren *et al.*, 2002).

The Gibbs free energy $G$ is often used to describe the chemical potential of biological systems instead of the Helmholtz free energy $F$ because most biochemical experiments are conducted at constant temperature and constant pressure (Kittel and Kroemer, 1980). The change in Gibbs free energy $\Delta G$ between states A and B is given by two components: the change in enthalpy $\Delta H$ and the change in entropy $\Delta S$

$$\Delta G = \Delta H - T\Delta S = U - P\Delta V - T\Delta S \tag{11}$$

where $U$ is the energy of the system, $P$ the pressure, $V$ the volume and $T$ the absolute temperature of the system. The inherent large fluctuations in the calculation of the enthalpy as well as the requirement of a complete conformational sampling in order to obtain the partition function or the number of states of the system, make the evaluation of enthalpy and entropy from simulations much more difficult than the calculation of free energy differences (Kollman, 1993; Reinhardt *et al.*, 2001). Yet, sometimes an estimate of entropic changes of a particular system is highly desirable. For this purpose, Schlitter introduced a heuristic formula to derive an upper bound on the entropy based on the mass weighted covariance matrix of atom–positional fluctuations (Schlitter, 1993). This approach has been tested on several biomolecular systems to derive the *configurational entropy* $S_{cf}$ and account for the entropic contribution in the folding process (Schäfer *et al.*, 2001; Schäfer *et al.*, 2002).

An attempt to derive the configurational entropy of the gp120/CD4 complex based on Schlitter's formula is described in **Chapter 6**. A new "combined" approach is introduced in which additional information on the degree of overlap of various states can be extracted by *combining* two trajectories corresponding to two different states of the system, in this particular case, free gp120 and its complex with CD4. The microscopic observations in the MD simulations are in qualitative agreement with the macroscopic experimental findings.

### *When simulations meet experiments*

Our original aim in simulating HIV-1 gp120 was to provide a rationale for the discovery of an *in vitro* revived mutant of gp120. It has been demonstrated that the nine disulphide bonds in gp120 are not equivalent in their contributions to the folding efficiency as well as to the infectivity (van Anken, 2003). For some disulphide bonds, a cysteine to alanine mutation prohibits gp120 from passing the quality control machinery during the secretory pathway. The misfolded proteins are retained within the ER leaving no envelope protein available for viral infection. After prolonged incubation of these dysfunctional mutants, some viral activity could be restored from an initially dysfunctional C385A/C418A gp120 mutant. Through *in vitro* evolution, the virus first replaced the alanine at position 418 by a valine (A418V) and subsequently introduced another threonine to isoleucine mutation at position 415 (T415I). The gp120 *revertant* C385A/C418V/T415I shows an improvement in its folding efficiency. As a result, the infectivity is partially restored as well. Put into the context of biomolecular recognition, the questions to be addressed by simulation would then be: How does the revertant mutation of two amino acids (A418V/T415I) affect the folding property? How does the ER folding machinery recognise the difference between wild type, mutant and revertant?

Given the size of gp120 and the auxiliary molecules involved in proper folding, such as chaperones and disulphide bond isomerases, simulation of the folding process from an extended structure to the native fold is beyond the reach of current computational power. We therefore limited ourselves to simulating the effects of those mutations on the stability of the native state. Comparison of the simulations of wild type gp120, of the dysfunctional C385A/C418A mutant and of the partially functional C385A/C418V/T415I revertant showed no distinguishable change in global structure and dynamics (**Chapter 7**). Yet, to pass through the protein folding quality control in the ER and restore infectivity (Sitia and Braakman, 2003), a significant fraction of the conformations must populate the native state (Dinner *et al.*, 2000). The threshold set by the quality control system is unknown. In other words, how much structural difference can be tolerated and to what extent can a misfolded protein be identified by the quality control machinery remains unclear. The revertant gp120 is a particularly intriguing example in the sense that the loss of a covalent bond, *i.e.*, a disulphide bond, can be compensated by the introduction of two β-branched hydrophobic side chains. In fact, the perturbations introduced by these small changes in sequence may be local. An analysis of the local structural integrity in the vicinity of the mutation site reveals that the revertant improves the side chain packing that was lost in the mutant and that several inter-strand hydrogen bonds become stabilised. The increase in stability of the β-strands in this region may be sufficient to lift the revertant across the threshold of the ER quality control while the mutant fails to pass. This MD simulation study provides a perspective to assist the rationalisation of the experimental observations.

# Mapping the targeted membrane pore formation mechanism by solution NMR:

## the nisin Z and lipid II interaction in SDS micelles[#]

Shang-Te Hsu[1], Eefjan Breukink[2], Ben de Kruijff [2], Robert Kaptein[1],
Alexandre M.J.J. Bonvin[1] and Nico A.J. van Nuland[1]

*[1]Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research and*
*[2]Institute of Biomembranes, Centre of Biomembranes and Lipid Enzymology,*
*Utrecht University, 3584CH Utrecht, The Netherlands*

## Abstract

**Nisin is an example of type A lantibiotics that contain cyclic lanthionine rings and unusual dehydrated amino acids. Amongst the numerous pore-forming antimicrobial peptides, type A lantibiotics form an unique family of post-translationally modified peptides.** *Via* **the recognition of cell wall precursor lipid II nisin has the capacity to form pores against Gram-positive bacteria with an extremely high activity in the** *nanomolar* **(nM) range. Here we report a high resolution NMR spectroscopy study of nisin/lipid II interactions in SDS micelles as a model membrane system in order to elucidate the mechanism of molecular recognition at residue level. The binding to lipid II was studied through $^{15}N$-$^{1}H$ HSQC titration, backbone amide proton temperature coefficient analysis and heteronuclear $^{15}N\{^{1}H\}$-NOE relaxation dynamics experiments. Upon the addition of lipid II significant changes were monitored in the N-terminal part of nisin. An extremely low amide proton temperature coefficient ($\Delta\delta/\Delta T$) was found for the amide proton of Ala3 (> –0.1 ppb/K) in the complex form. This suggests tight hydrogen bonding and/or isolation from the bulk solvent for this residue. Large chemical shift perturbations were also observed in the first two rings. In contrast, the C-terminal part of nisin was almost unaffected. This part of the molecule remains flexible and solvent exposed. Based on our results, a multi-step pore forming mechanism is proposed. The N-terminal part of nisin first binds to lipid II and a subsequent structural rearrangement takes place. The C-terminal part of nisin is possibly responsible for the activation of the pore formation. In light of the emerging antibiotic resistance problems, an understanding of the specific recognition mechanism of nisin with lipid II at the residue specific level may therefore aid in the development of novel antibiotics.**

## Introduction

Nisin Z is a 34–residue peptide, which is produced by and primarily acts against Gram-positive bacteria. It belongs to the type A lantibiotics (lanthionine–containing antibiotics) family (Breukink and de Kruijff, 1999). The common features of lantibiotics are the unique post-translationally modified amino acids dehydroalanine (*Dha*), dehydrobutyrine (*Dhb*) and the lanthionine rings that are formed by thioether bonds (see **Chapter 1**). Typical type A lantibiotics are elongated, flexible and amphipathic peptides that possess pore forming abilities. In contrast, type B lantibiotics are compact, globular and hydrophobic peptides, some of which kill bacteria by blocking cell wall synthesis (Brötz and Sahl, 2000). Nisin Z contains three positively charged lysine residues and five intramolecular lanthionine rings, namely rings A–E (Figure 1A). Nisin Z differs from its natural variant nisin A only by a single residue at position 27, a histidine instead of an asparagine, without change in bacterial activity (Kuipers *et al.*, 1996). In the following we will refer to nisin Z simply by nisin unless specific comparisons between nisin A and nisin Z are discussed.
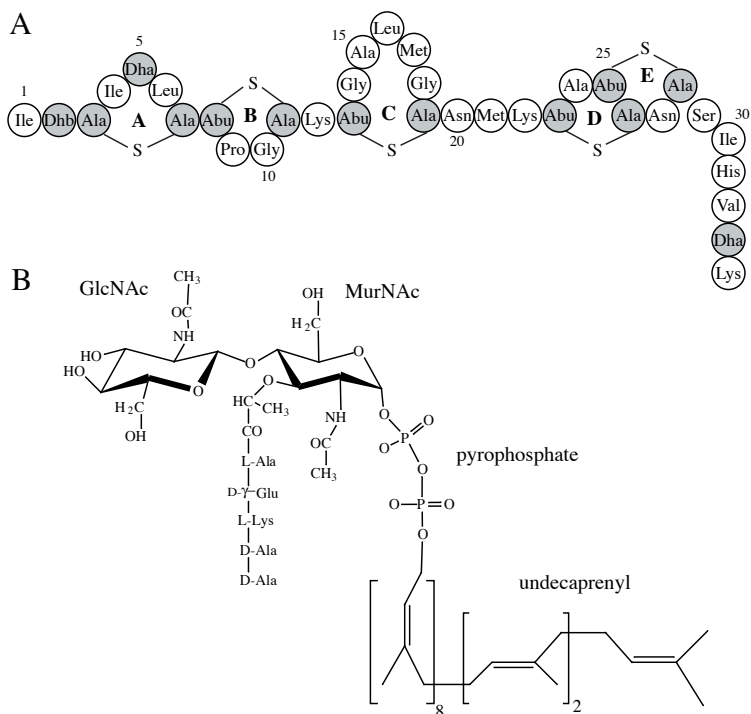
**Figure 1.** Primary structure of nisin Z and lipid II. **A.** *Dha* (**U**), dehydroalanine; *Dhb* (**O**) dehydrobutyrine; *Ala-S-Ala* (***A-S-A***), lanthionine; *Abu-S-Ala* (***A⋆-S-A***), β-methyllanthionine; **B.** *GlcNAc, N*-acetylglucosamine; *MurNAc, N*-acetylmuramic acid.

The pore forming function of nisin has been studied in various model membrane systems *in vitro* by vesicle binding, carboxyflurescein (CF) leakage, potassium ion leakage experiments (for review, see (Breukink and de Kruijff, 1999)) and solid state nuclear magnetic resonance (NMR) spectroscopy (Bonev *et al.*, 2000). High resolution NMR solution structures of nisin A have been determined in water (van de Ven *et al.*, 1991) and in membrane-mimicking environments of dodecylphosphocholine (DPC) and sodium dodecylsulphate (SDS) solutions (van den Hooven *et al.*, 1996a). These studies showed a very flexible, extended structure, with rather well defined local ring substructures. The orientation of nisin on the membrane surface was also studied by fluorescence (Breukink *et al.*, 1998), solution NMR (van den Hooven *et al.*, 1996b) and solid–state NMR (Jastimi *et al.*, 1999; Bonev *et al.*, 2000). The amphipathic nature of nisin provides a preferential orientation on the membrane interface: both N– and C–terminal parts of nisin were found to be involved in membrane binding but the N-terminal part was more deeply embedded (van den Hooven *et al.*, 1996b; Breukink *et al.*, 1998). Unlike the highly specific activity (*nM* range) obtained from *in vivo* studies, the activities of nisin that were previously obtained from *in vitro* experiments were only in the *micromolar* (μM) range.

This discrepancy has recently been solved by the discovery that lipid II (Figure 1B), the membrane-bound peptidoglycan precursor for cell wall synthesis, is used as a receptor/docking molecule by nisin

(Brötz *et al.*, 1998; Breukink *et al.*, 1999). The recognition of lipid II facilitates and enhances the pore forming function of nisin *in vitro* by a factor of $10^3$ in model membrane systems (Breukink *et al.*, 1999; Wiedemann *et al.*, 2001). Lipid II is the key element for the synthesis of the protective cell wall for bacteria. The undecaprenyl tail of lipid II acts as a carrier, which transports the peptidoglycan subunit from the cytoplasm to the extracellular domain. Because of the importance of lipid II in the bacterial cell wall synthesis pathway, it is also an unique target for some antibiotics, *e.g.*, vancomycin (Sheldrick *et al.*, 1978). The addition of vancomycin inhibits the activity of nisin in the presence of lipid II to some extent, but the binding motif of lipid II for nisin is probably different from that of vancomycin (Breukink *et al.*, 1999). However, the exact binding mechanism for nisin is still unclear. The discovery of the specific and high affinity interaction between nisin and lipid II provides the first known example of targeted pore formation. The results suggest that nisin, or derivatives of it, could one day be used as an alternative antibiotic to overcome the emerging antibiotic resistance problems. The nisin/lipid II interaction is also a useful model system to understand this novel pore forming mechanism and to provide a blueprint for further peptide engineering and antibiotics development. The elucidation at a structural level of the targeted pore forming mechanism will be the most comprehensive approach.

Here we present a NMR study of the interaction between nisin and lipid II in SDS micelles as a membrane mimicking environment. Various types of high resolution NMR experiments have been performed to provide insight into the binding mechanism at residue level. We will show that only the N-terminal part of nisin strongly interacts with lipid II whereas the C-terminal part remains unaffected by the formation of the complex. These results suggest that the pore forming mechanism is a multi-step process in which nisin possesses different functionalities throughout several segments: *i)* initial recognition of the N-terminal part followed by subsequent aggregation and *ii)* at the later stage, activation of the pore formation by the C-terminal part.

## Material and methods

### Sample preparation

[15]N-labelled nisin Z was isolated and purified as described (Kuipers *et al.*, 1992) using a ^{15}N-enriched growth medium as nitrogen source, and lipid II was prepared as described (Brötz *et al.*, 1997). Nisin was dissolved in 500 μL, 10% $D_2O$, 90% $H_2O$ with 25 mM sodium phosphate buffer adjusted to pH 6.0, resulting in sample concentrations of 1.8 mM for the unlabelled peptide and 1 mM for the ^{15}N-labelled peptide. 4% and 2% perdeuterated $d_{25}$-SDS were added into the unlabeled and ^{15}N-labelled samples, respectively, to bring the SDS concentration to approximately 100-fold excess with respect to the peptide concentration. Since the 1D NMR spectrum showed no significant changes when the nisin-to-SDS ratio was changed from 1:20 to 1:60 (van den Hooven, 1995), further changes were therefore solely due to the incorporation of lipid II. This also ensured a ratio of approximately one peptide per micelle in order to form a 1:1 complex with lipid II.

### NMR Spectroscopy and titration experiments

NMR spectra were recorded on Varian INOVA 750 MHz and 500 MHz spectrometers. 2D NOESY (Jeener *et al.*, 1982) with mixing times of 100, 150, and 200 ms and 2D clean-TOCSY (Griesinger *et al.*, 1988) with mixing times of 60 and 100 ms were performed for the backbone resonance assignment

of nisin in SDS solution with and without the presence of lipid II. All 2D spectra were recorded with 2048 complex *t2* points and 1024 complex *t1* points. 3D $^{15}$N NOESY-HSQC (Marion *et al.*, 1989b) and $^{15}$N TOCSY-HSQC with DIPSI-2 spin-lock sequence (Marion *et al.*, 1989a) were recorded with mixing times of 100 ms and 60 ms, respectively. For both 3D experiments, the spectra size was set to 1600★280★144 complex points for the direct $^1$H dimension and the indirect $^1$H and $^{15}$N dimensions, respectively. In the $^{15}$N-$^1$H HSQC titration experiments, 25 μL of 3mM lipid II in 2% $d_{25}$-SDS was added successively five times into 400μL of 0.75 mM $^{15}$N-labelled nisin. This led to 24% dilution of the nisin concentration at the final step and a final molar ratio of nisin to lipid II of 1:1.25. $^{15}$N-$^1$H HSQC spectra (Kay *et al.*, 1992) were recorded with 1600 complex *t2* points and 180 complex *t1* points. All spectra were recorded at 20 °C with the WATERGATE water suppression protocol (Piotto *et al.*, 1992). Steady-state heteronuclear $^{15}$N{$^1$H}-NOE experiments with and without saturation irradiation during the 3s relaxation delay period were recorded on the nisin/lipid II complex at 40 °C as described previously (Meunier *et al.*, 2000). The spectra size was the same as in the titration experiments. For amide proton temperature coefficient measurements, $^{15}$N-$^1$H HSQC spectra were recorded at 10, 20, 30, and 40 °C. Additional 2D and 3D NOESY experiments were performed at 40 °C on the nisin/lipid II complex after titration. All spectra were processed with the NMRPIPE software package (Delaglio *et al.*, 1995) and analyzed with NMRView (Johnson and Blevins, 1994).

## Results

### Backbone resonance assignments of nisin Z

Although the small sequence difference of nisin Z from its natural variant, nisin A, (H27N) does not affect its bacterial activity (Kuipers *et al.*, 1996), its backbone resonances in SDS micelles turned out to be rather different. Apparently, the difference in pH (for nisin A pH = 2.1 (Sailer *et al.*, 1993) and 3.5 (van den Hooven, 1995); for nisin Z in our current report pH = 6) affects the chemical shifts substantially. This is unlikely caused by the single residue variation, H27N, which would probably only induce localised chemical shift changes due to the difference in electrostatic interaction. The backbone assignment of nisin Z in SDS solution was performed using the standard sequential assignment procedure by combination of $^{15}$N NOESY-HSQC and $^{15}$N TOCSY-HSQC spectra. Most residues could be assigned in the free form except for the residues flanked by two glycine residues, *i.e.*, Ala15, Leu16, and Met17, in ring C. During the course of lipid II titration, many residues showed slow exchange binding patterns, *e.g.* Ala3. At the final step of the titration, a similar sequential assignment approach was attempted; however, only few residues were clearly resolved. Most residues showed weak NOE cross-peaks that hampered the complete assignment of nisin in the lipid II-bound form. For those residues that did not show significant shift upon complexation, mainly in the C-terminal part, the assignments were taken from the free form. These assignments were confirmed by the similarity in the spin system patterns in the NOESY spectrum. The backbone amide proton NOE connectivity HN$_{(i-1)}$-HN$_{(i)}$ was initially used for the sequential assignments of those residues that showed large chemical shift perturbations. They were then checked by comparison of the spin system patterns. Figure 2 shows the assigned HSQC spectrum of nisin in the free and the lipid II-bound form. Note that there are still some unassigned residues because of the lack or the low intensity of NOE signals.
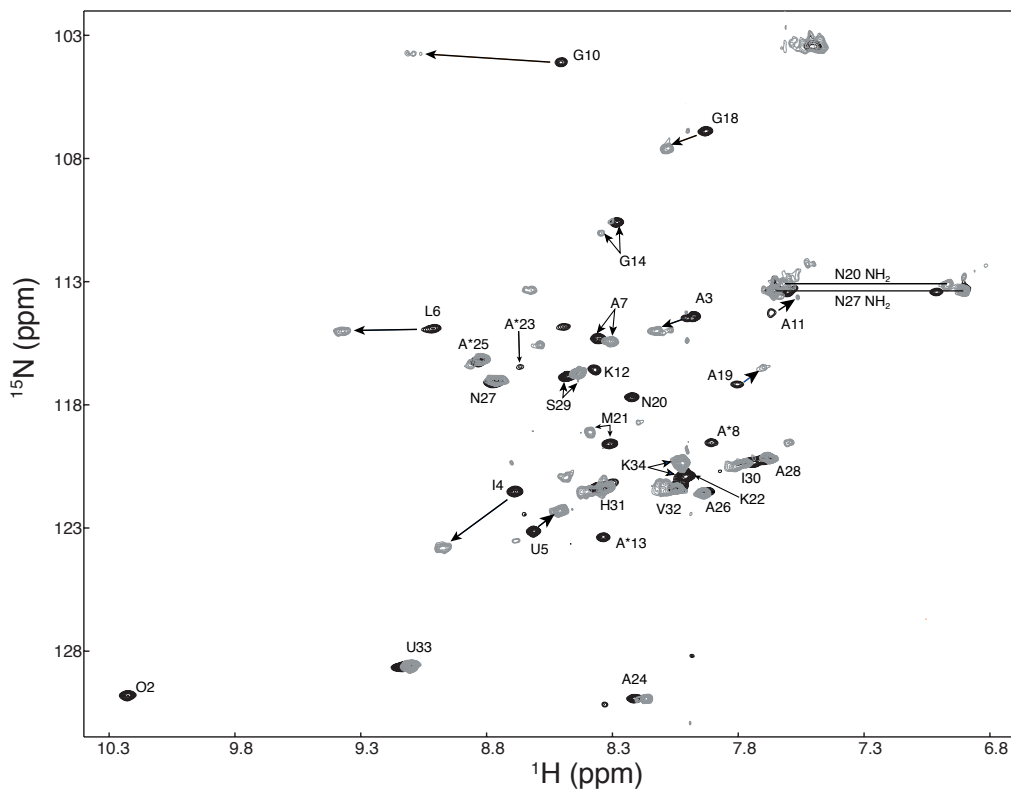
**Figure 2.** $^{15}$N-$^{1}$H HSQC titration spectra of $^{15}$N-labelled nisin before (black) and after the addition of lipid II (grey). The grey cross-peaks correspond to a molar ratio of nisin/lipid II of 1:1.25. Significant chemical shift changes are indicated by arrows. The unlabelled cross-peaks could not be assigned due to the lack of signals in the TOCSY and NOESY spectra. A*8, K12, A*13, N20, K22 could only be assigned in the free form (See text for details).

### Lipid II titration experiments show strong binding affinity

The titration of lipid II to $^{15}$N-labelled nisin in the presence of SDS shows that the binding is in the slow exchange regime. Several residues undergo large chemical shift perturbation and show typical slow exchange sets of peaks corresponding to the free and the bound form. These peaks coexist during the titration process until 25% excess of lipid II is added, after which all nisin molecules are in the complex form. According to the titration profile, the stoichiometry corresponds to a 1:1 nisin to lipid II complex (data not shown). Figure 3 shows the chemical shift changes upon binding as a function of the nisin sequence. The largest changes are observed in the first three N-terminal ring systems, especially Ile4, Leu6 and Gly10. The chemical shifts of the C-terminal part of nisin, starting from Ala24 up to Lys34, are hardly affected upon the addition of lipid II. We were not able to complete the assignment of the HSQC spectrum at the last point of the titration study; some peaks disappeared due to line broadening after 50% addition of lipid II and some peaks showed overlap. In contrast to the fast exchange process, in which the

average signals of the bound and the free forms can be easily followed during the titration steps, it was not possible to track the chemical shifts in the slow exchange regime. This is particularly the case for Lys12 and Asn20, for which no suitable cross-peaks close to the original positions could be identified. These missing peaks possibly overlap with others or simply disappear due to line broadening. The cause of line broadening is most likely due to slow conformational rearrangement of the relative orientations of the ring systems. Other residues like Ile4, Leu6, and Gly10 shifted by more than 0.3 ppm downfield along the $^1$H dimension (Figure 3A), suggesting that ring A and B are in the vicinity of the nisin/lipid II interface. These large downfield shifts are most likely caused by the deshielding effect upon the formation of intra or intermolecular hydrogen bonds with lipid II.
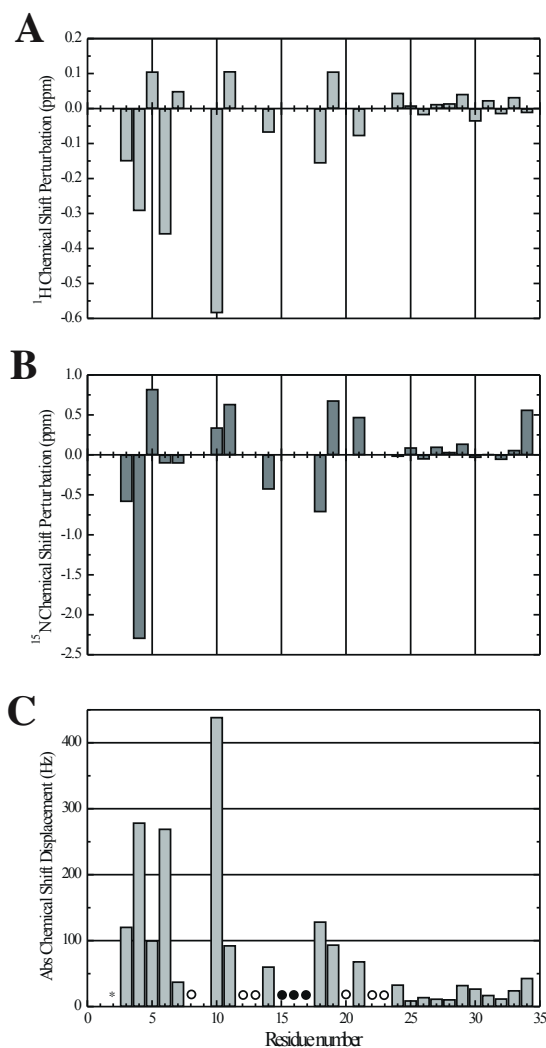


**Figure 3.** Chemical shift perturbation along the sequence of nisin upon the addition of lipid II. **A.** $^1$H chemical shift perturbations are displayed as the difference in chemical shift between the free and the bound forms, respectively, $\Delta\delta_{HN} = \delta_{HN}^{nisin} - \delta_{HN}^{nisin/lipid\ II}$. **B.** $^{15}$N chemical shift perturbations $\Delta\delta_N = \delta_N^{nisin} - \delta_N^{nisin/lipid\ II}$. **C.** Absolute chemical shift displacements in Hz. The values are calculated as $[(\Delta\delta_{HN})^2 + (\Delta\delta_N)^2]^{1/2}$ from spectra recorded at 750 MHz. Dhb2 is labelled with an asterisk (⋆) for the loss of signal after the addition of lipid II. *Closed circles* (●): unassigned residues in the absence of lipid II due to spectral overlap or ambiguities; *Open circles* (○): unassigned residues in the complex form.

## Amide proton temperature coefficients and solvent accessibility

Hydrogen bond formation can stabilise the exchangeable backbone amide protons and therefore reduce the exchange rate with bulk solvent. The stability can also arise from a reduction of solvent accessibility, such as burial into a complex interface. The amide proton temperature coefficient analysis has been performed to verify whether the largest chemical shift changes of nisin in the titration experiment are indeed due to hydrogen bond formation or due to exclusion from solvent. The temperature dependency of amide proton chemical shifts is a simple probe that provides information on the involvement in hydrogen bond formation or sequestering from the solvent (Cierpicki and Otlewski, 2001). A low dependency of the amide chemical shift on temperature in aqueous environment, *i.e.*, temperature coefficient $\Delta\delta/\Delta T >$ –2.5 ppb/K is usually indicative of the presence of a stable hydrogen bond. For solvent accessible amide protons, which are more sensitive to a temperature change, the slope is generally more negative than –5 ppb/K (Ballardin *et al.*, 1978; Graham *et al.*, 1992). We have collected several $^{15}$N-$^{1}$H HSQC spectra of nisin/lipid II complex under different temperatures ranging from 10 °C to 40 °C. Figure 4 shows the temperature coefficients of nisin and examples of the amide chemical shift changes for some of the residues. While the amide proton chemical shift of Ala3 hardly changes upon temperature increase, Ala7 and Ser29 are much more affected. Based on the above criteria, the stable amide protons with low solvent accessibility either *via* hydrogen bond formation or structurally protected from the solvent exchange process are identified for Ala3, Ile4, Dha5, Gly10, Gly14 and Ala19.
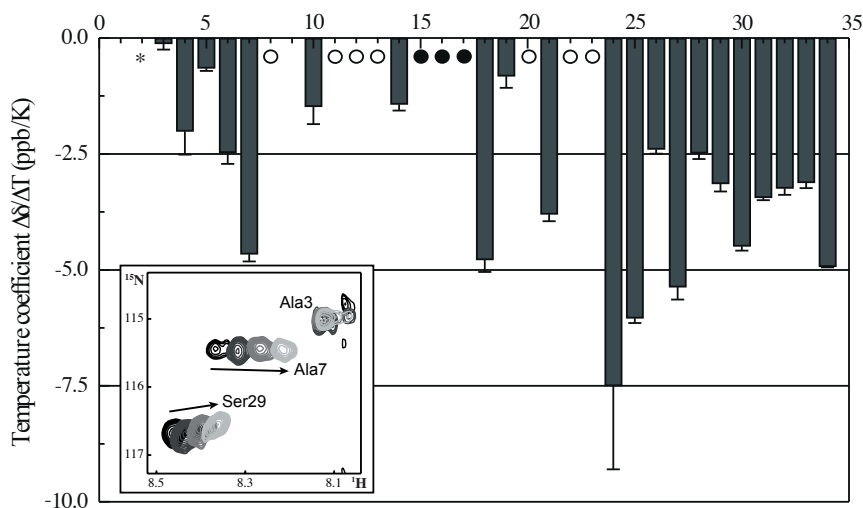


**Figure 4.** Amide proton temperature coefficients ($\Delta\delta/\Delta T$) of nisin in the complex form. The temperature coefficients were obtained by linear fitting of the amide chemical shift $\delta_{HN}$ as a function of the temperature measured at 10 °C, 20 °C, 30 °C, and 40 °C. The unassigned residues are labelled as described in Figure 3. Inset: representative region of $^{15}$N-$^{1}$H HSQC spectra of nisin in the complex. The cross-peaks are coloured in grey scale from black (10 °C) to light grey (40 °C). The direction of the change of the cross-peak position of each residue as a function of temperature is indicated by an arrow. Note that Ala3 is almost unaffected throughout a wide temperature range.

*Dynamic properties of nisin/lipid II complex*

Unlike most proteins that possess rigid and well-defined globular structures, nisin shows flexibility that is functionally important. This creates, however, a major challenge for our current NMR study: the dynamics results in an averaging of structural information, which makes characterisation of the complex difficult. Steady-state heteronuclear $^{15}N\{^1H\}$-NOE experiments have been performed on the nisin/lipid II complex to investigate its dynamic properties. The heteronuclear NOE is a very sensitive probe for fast backbone motion in the *picosecond* to *nanosecond* time scale (Kay *et al.*, 1989; Kay, 1998; Palmer III, 2001). It gives a measure of the rigidity of the backbone and is defined as the ratio of peak intensities with and without saturation irradiation on the amide protons. Well-defined and rigid structure elements usually give positive NOEs close to unity. For the nisin/lipid II complex at 40 °C, negative NOEs were observed for Dha33 and Lys34 and either weak or zero NOEs for residues 30-32: an evidence of a highly dynamic motion. In contrast, the N-terminal part was relatively rigid showing positive NOEs (data not shown). These findings are in agreement with the chemical shift perturbation and amide temperature coefficient analysis and confirm that the C-terminal part of the molecule is not involved in the binding of lipid II.

## Discussion

In the present report, we have provided a residue specific analysis of nisin/lipid II interactions in a model membrane system by solution NMR spectroscopy. Figure 5 summarises the results of the HSQC titration experiment and amide proton temperature coefficient analysis. It should be noted that the observed chemical shift changes of nisin upon the addition of lipid II could in principle arise from either the direct interactions of the affected residues with lipid II or the indirect effect from changes in nisin-SDS interactions on forming the complex. The incorporation of lipid II can potentially induce a morphological change of SDS micelles and therefore change the mode of interaction between nisin and micellar surface. It has been demonstrated that the difference in the surface curvatures of different membrane mimicking media, *e.g.*, micelles and bicelles, can induce subtle backbone conformational changes of membrane interacting peptides (Chou *et al.*, 2002). These structural differences are clearly reflected by substantial changes in the $^1H$-$^{15}N$ HSQC spectra. In our case, however, no peak shift was observed for nisin during the titration process; only the signal population decreases at the free form position and increases at the bound form position until all nisin molecules are bound to lipid II. This suggests that lipid II integrates into SDS micelles and interacts with nisin in such a way that nisin and lipid II form a specific complex in SDS micelles. This does not affect the interaction mode between the remaining nisin molecules and the lipid II-free SDS micelles. A previous SDS titration study of nisin in the absence of lipid II has shown that the progressive excess of SDS concentration, which could also result in larger micelle size, does not change the spectrum of nisin beyond a nisin-to-SDS ratio of 1:20 (van den Hooven, 1995). This suggests that the conformation of nisin as reflected in NMR spectra is not sensitive to the morphology of SDS micelles. Although hampered by the lack of intermolecular NOE data from the nisin/lipid II complex, all data strongly support the idea that the addition of lipid II induces a substantial change in the chemical environment of the N-terminal part of nisin, mainly through direct binding. The interaction between nisin and SDS micelles is therefore affected as well. The micelle morphology may alter due to the insertion of lipid II, although this effect is relatively minor.
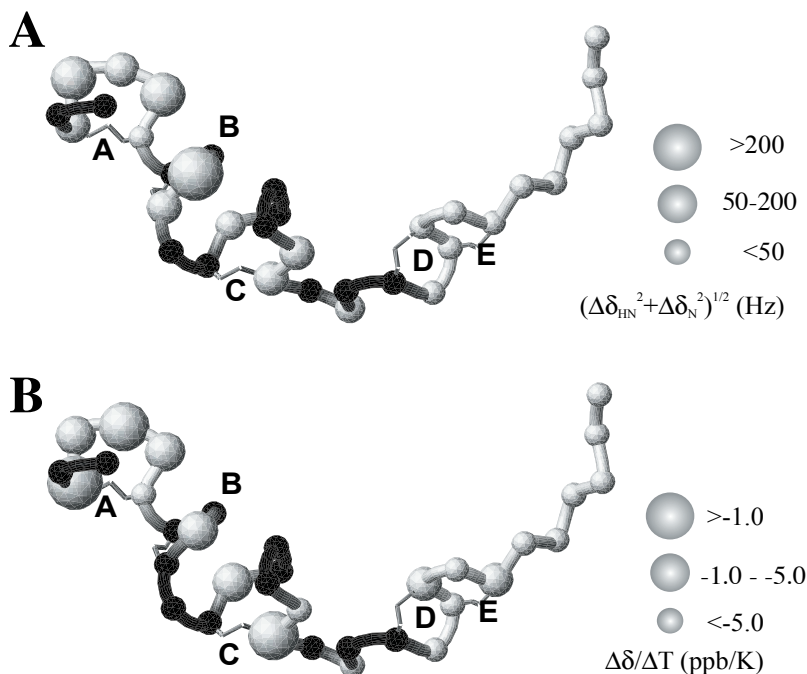
**Figure 5.** N-terminal part of nisin is responsible for lipid II binding. **A.** The chemical shift perturbations after the addition of lipid II to nisin. **B.** The temperature dependency of nisin in the complex form with nisin. The thioether linkages are shown in sticks. The ring systems of nisin are labelled from A to E according to the definition in Figure 1A. Backbone Cα atoms are shown in spheres with radii from small (unaffected) to large (significantly affected) as indicated on the right. The unassigned residues are coloured in black. The nisin coordinates correspond to the representative structure of nisin in the presence of DPC micelles (van den Hooven *et al.*, 1996a). These were kindly provided by Dr. H. W. van den Hooven. The figures were generated by MOLMOL (Koradi *et al.*, 1996).

Although the intrinsic flexibility of nisin and a higher sample pH value leave some residues unassigned in our current study, high pH (> pH 5) is required because the nisin/lipid II complex cannot be formed at pH 3-4. A signature circular dichroism (CD) absorption that is indicative of the nisin/lipid II complex formation can only be observed in higher pH. The CD spectra of nisin in the presence of lipid II and SDS micelles in low pH are similar to those observed without lipid II (data not shown). In the absence of lipid II, the CD spectra of nisin in SDS micelles do not change significantly between pH 5 and 8. These are also similar to the previously reported CD data of nisin in SDS micelles at lower pH (pH 3-4) in spite of their difference in absorption intensities (Dykes *et al.*, 1998). This suggests that the conformation of nisin in SDS micelle solution without lipid II is not sensitive to changes in pH. In contrast, the CD spectra of nisin are very sensitive to the pH changes in aqueous solution. Differences in the sample pH values and the single residue variation between nisin A and Z (H27N) might contribute to the observed spectral difference between our current study and those previously observed without the incorporation of lipid II. Nevertheless, in the current study we are focused on the effect of lipid II-binding on nisin in order to map the lipid II binding site. In addition, a 1:1 nisin to lipid II binding stoichiometry can be

obtained from the lipid II titration experiments in SDS micelles, which is in agreement with proposed stoichiometry obtained from the direct binding assay (Breukink *et al.*, 1999).

### N-terminal part of nisin is responsible for the first recognition step

The binding to lipid II perturbs mainly the chemical shifts of residues localized in the first two rings, A and B. The specific binding process is in the slow exchange regime and some residues show large chemical shift changes, for example, Ile4, Leu6 and Gly10 (Figure 2). Other residues show broad lines in the free and/or in the complex form, mostly those in ring C and the following hinge region. This probably reflects some structural inhomogeneity in this part of the molecule.

The complex formation also enhances the stability of several backbone amide protons in the N-terminal part of the structure shown by the temperature coefficient analysis. The extremely small amide proton temperature coefficient of Ala3 (> -0.1 ppb/K) suggests that it is highly protected from the exchange process with the bulk solvent. The downfield shifting of the amide protons in rings A and B along the $^{1}$H dimension can be explained by the deshielding effect due to the involvement in a hydrogen bond network in the structure of nisin/lipid II complex. For Dha5, although slightly upfield shifted upon binding, its amide proton might be already protected prior to the complex formation so that no additional stability enhancement is observed. It is therefore reasonable to hypothesise that the first three rings, especially rings A and B, form a binding pocket and bind tightly to lipid II. Another aspect in this study, compared with the previous reports in aqueous solution, SDS and DPC micelles, is that the binding of lipid II reduces the amide proton temperature coefficients of the amide protons of ring A and B approximately two fold (those of nisin A in SDS micelles are about −4 to -8 ppb/K throughout the sequence except for Dha5 and Ala11(van den Hooven *et al.*, 1996b)). Nevertheless, care should be taken in the comparison since our sample pH value is different from the previous study. Despite the difference in pH condition, a significant change of the temperature coefficient from about -7 ppb/K to −0.1 ppb/K is observed for Ala3 while the C-terminal part of nisin is basically unaffected (generally < -4 ppb/K). This suggests that the complex interface is formed by the N-terminal part of nisin and lipid II and therefore results in the low solvent accessibility particularly in this part of nisin.

We also note the excellent correlation amongst the chemical shift perturbation, amide proton temperature coefficient analysis and relaxation studies: they all indicate different structural behaviour along the sequence of nisin, with the hinge region defining the border between these structural elements. The C-terminal part of nisin, starting from ring D and E, does not seem to interact with lipid II. Although the backbone geometry of these two intertwined rings is tightly confined by the thioether linkages, the corresponding backbone amide protons are still highly solvent accessible and show high mobility. This indicates that these two rings are neither buried into the SDS micelles nor are they in complex with lipid II. Our observations suggest that only the N-terminal part of nisin is responsible for the recognition of lipid II, the C-terminal part being probably more important for the pore formation at a later stage. This model is in line with two other studies: *i)* the N-terminal fragment nisin[1-12] can act as an antagonist to inhibit the bacterial activities of the full-length nisin (Chan *et al.*, 1996b) and *ii)* the introduction of a positively charged lysine at position 32 (V32K) does not seriously disrupt nisin activity in the presence of lipid II (Wiedemann *et al.*, 2001). Recently, Wiedemann and co-workers have performed a series of mutagenesis studies, which have pointed out the importance of structure integrity of the lanthionine rings A and C and of the flexibility of the hinge region for the lipid II-mediated pore formation (Wiedemann

*et al.*, 2001) . From these studies, it has been shown that there exists a high structural selectivity at position 3 (Ala3) while the preceding position (Dhb2) has more tolerance. Introduction of an extra methyl group at position 3 of the prepeptide (S3T) modifies the lanthionine linkage to a 3-methyllanthionine, and this substantially reduces its activity (MIC for *M. flavus*: S3T = 39 nM versus wild type = 3.3 nM); while the modification of Dhb2 into other residue types has less effects. In addition, hydrolytic cleavage at Dha5, by which the ring A is opened, results in substantial loss of biological activity, whereas the same cleavage at Dha33, removing the last two residues, has very little effect (Chan *et al.*, 1989; Chan *et al.*, 1996a).

As discussed above, the low amide proton temperature coefficients and the localised chemical shift perturbations provide the first direct proof that the N-terminal part of nisin is tightly bound to lipid II. The stable nisin/lipid II complex is a result of a highly complementary binding interface. Any additional chemical modification at the interface would abolish the complementarity and therefore disrupt its activity. S3T mutation causes a structural clash because of the extra methyl group in the ring linkage position, which makes it difficult for nisin to adopt a suitable geometry for the binding of lipid II. The same explanation may also apply to Gly14 and Ala19. These two amide protons show similar patterns of chemical shift perturbation and amide proton temperature coefficient during complex formation. This suggests that ring C is also involved in the binding process but maybe to a lesser extent. Our NMR results above have located the binding interface of nisin, essentially rings A and B, and explained the selectivity of residue types obtained from the mutagenesis study, especially for Dhb2 and Ala3 in ring A.

The involvement of ring C has been studied by other mutagenesis studies. The removal of the C-terminal fragment, which keeps the first 20 residues of nisin, nisin[1-20], increases the minimal inhibitory concentration (MIC) values for *L. lactis* and *M. luteus* by 115-fold and 50-fold, respectively. A further removal of ring C, nisin[1-12], results in increases of MIC values for *L. lactis* and *M. luteus* by more than 2500-fold and 200-fold, respectively, with respect to the full-length nisin[*] (Chan *et al.*, 1989). This suggests that ring C is important for nisin's bacterial growth inhibitory function. Introduction of a charged residue by replacing Met17 to a lysine residue (M17K) can reduce the activity of nisin by 50% (Wiedemann *et al.*, 2001). Furthermore, a single T13C point mutation of the prepeptide, which extends the ring C linkage by replacing the thioether bond to a disulphide bond, results in a mutant that is almost inactive *in vivo* and in the liposome CF leakage assay (Wiedemann *et al.*, 2001). These results explain the involvement of ring C, in particular Gly14 and Ala19, during the lipid II binding process that is observed in our study.

It is important to note that the dual dynamic property (stable N-terminus versus flexible C-terminus) of nisin can be critical for its bacterial activity. The previously proposed wedge model was based on results obtained in the absence of lipid II. Most of the rings were found to be involved in the membrane binding process (Driessen *et al.*, 1995; Moll *et al.*, 1996). Without lipid II, the electrostatic attraction between the positively charged residues of nisin and the negatively charged head groups of membranes is dominant. It is therefore no surprise that both termini were found to interact with the model membrane systems in the previous studies; the contributions of the three lysine residues are rather equivalent. The enhancement of activity observed after the incorporation of lipid II however indicated that there must exist another dominating interaction between lipid II and nisin. The functional role of the C-terminal part in pore formation is no longer equivalent to that of the N-terminal part. In contrast

---

[*] The MIC values (μg/ml) against *L. latic* and *M. luteus* are 0.08 and 0.07 for full-length nisin[1-34], 9.21 and 3.5 for nisin[1-20], and >200 and 16.5 for nisin[1-12], respectively.

to the recognition role of the N-terminal part, it is more important for the pore formation at a later stage (van Heusden *et al.*, 2002).

### The hinge region bridges the two dynamic structural elements

In a previous NMR study of nisin A the hinge region was found to be relatively structured forming a type II β-turn in the presence of SDS and DPC micelles. This turn is stabilised by backbone hydrogen bonds, in particular between Met21 and Abu24 (van den Hooven *et al.*, 1996a). In the case of nisin Z in the presence of lipid II, however, high mobility and/or structural inhomogeneity is still present. This might be functionally important: shortening of the hinge region by the removal of two residues at position 20 and 21, or introduction of proline residues at these positions to enhance rigidity causes a severe loss of the pore forming activity (Wiedemann *et al.*, 2001). Reduction of flexibility in the hinge region can disable further pore formation by which the translocation of the C-terminal part across the membrane is necessary (van Kraaij *et al.*, 1998). Since there exist two very different dynamic motions along the structure of nisin, the hinge region therefore plays the transitional role to bridge the two termini together. This intermediate motion can also be the reason of the line broadening and the missing peaks in this part of the structure, for example, Asn20.

### Aggregation of the nisin/lipid II complex after the binding process

Elucidation of the nisin-lipid II complex structure was not possible at this stage due to a dramatic loss of cross-peaks in NOESY spectra. After the titration of lipid II into the nisin/SDS solution, slow aggregation was observed and the increase of size was such that only the C-terminal part of nisin exhibiting high flexibility showed resolvable cross-peaks. Increasing the temperature substantially improved the line widths in the HSQC spectra, but the NOESY spectra were still poor even at the elevated temperature of 40 ˚C. Not being able to penetrate SDS gel in the standard electrophoresis experiment when nisin is bound to lipid II is also a supplementary indication of the formation of a sizeable aggregate (data not shown). This suggests that the subsequent aggregation of the nisin/lipid II complex is essential for the pore formation. The loss of NOE signals after the complex formation is not only due to the slowing down of the molecular tumbling but also to the structural heterogeneity during the aggregation process.

### Multi-step lipid II targeted pore formation mechanism

Based on the information in our current study and previously reported mutagenesis results, we propose a multi-step model for lipid II–mediated nisin pore formation. The first step involves the recognition of the lipid II head group by the N-terminal part of nisin. A tight nisin/lipid II complex is thereby formed, anchored into the membrane by the undecaprenyl tail of lipid II. Subsequent conformational rearrangement takes place by means of the aggregation of the nisin/lipid II complexes. The minimum number of nisin molecules for pore formation is however still unclear. Finally, the insertion of the C-terminal part of nisin into the membrane completes the pore forming process.

Here we have observed the first recognition step in the complete mechanism. The mapping of the binding interface of nisin has provided insight into the functionality of each amino acid along the nisin sequence. The conserved residues amongst the type A lantibiotics might also share similar functionalities. We think that the work presented here will benefit to the future studies of this target mediated membrane pore formation mechanism and might aid the future development of novel antibiotics.

## Acknowledgement

# The pyrophosphate cage: the structure of the nisin/lipid II complex provides a blueprint for novel antibiotics[#]

Shang-Te D. Hsu[1], Eefjan Breukink[2], Eugene Tischenko[1], Mandy A. G. Lutters[2], Ben de Kruijff[2], Robert Kaptein[1], Alexandre M.J.J. Bonvin[1] and Nico A.J. van Nuland[1]

*[1]Department of NMR Spectroscopy and [2]Department of Biochemistry of Membranes, Bijvoet Center for Biomolecular Research, Utrecht University, 3584CH Utrecht, The Netherlands*

**The emerging antibiotics–resistance problem such as that caused by methicillin–resistant _Staphylococcus aureus_ (MRSA) has underlined the urgent need for novel antimicrobial agents for infectious therapy. The family of lantibiotics (lanthionine-containing antibiotics) (Sahl and Bierbaum, 1998; Guder _et al._, 2000) harbours promising candidates to alleviate this problem. Nisin, a member of this family, possesses an unique pore-forming activity against bacteria. It binds to lipid II, the essential precursor of cell wall synthesis. Thereby the membrane permeabilization activity of nisin is increased by three orders of magnitude (Breukink _et al._, 1999). Here we report the solution structure of the complex of nisin and lipid II using nuclear magnetic resonance (NMR) spectroscopy. The structure shows that the N-terminal part of nisin forms a cage wherein the pyrophosphate moiety of lipid II is bound primarily _via_ intermolecular hydrogen bonds between nisin backbone amides and the phosphate groups. This cage structure is a novel lipid II binding motif that provides a rationale for the conservation of the lanthionine ring structures amongst several lipid II-binding lantibiotics. The structure of the pyrophosphate cage offers a template for the structure-based design of novel antibiotics.**

The cell wall is essential for the vitality of bacteria as its strong extracellular peptidoglycan matrix resists the high osmotic pressure of the cytoplasm. It is well conserved throughout evolution and is therefore a prominent target for many antibiotics (for review, see (Koch, 2003)). The monomeric peptidoglycan unit, the basic building block of the cell wall, consists of two amino sugars, N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc), and a pentapeptide, often L-Ala-D-γ-Glu-L-Lys-D-Ala-D-Ala, which is attached to the carboxyl group of MurNAc. These subunits are assembled in the cytosol on a membrane-anchoring carrier, undecaprenyl phosphate, yielding lipid II (GlcNAc-MurNAc-pentapeptide-pyrophosphoryl-undecaprenol, Figure 1A; for review, see (van Heijenoort, 1994)). Lipid II is thereafter transported to the extracellular domain for polymerisation of the peptidoglycan moiety.

The essential role of lipid II in cell wall synthesis makes it a target for many antimicrobial peptides (Brötz _et al._, 1998). This includes vancomycin, the clinical antibiotic of last resort (Sheldrick _et al._, 1978), and ramoplanin, which is now in phase three clinical trial (McCafferty _et al._, 2002). We note, however, that vancomycin resistance is now emerging (Hughes, 2003). Lipid II is also targeted by nisin, a 34-residue post-translationally modified peptide containing five lanthionine rings and three dehydrated amino acids (Figure 1B). Lipid II serves not only as an anchoring receptor for nisin but also facilitates the "targeted" pore forming activity of nisin thereby lifting nisin's activity on bilayers to the nanomolar range (Breukink _et al._, 1999). Vancomycin-resistant bacteria with the _vanA_-type gene cluster evade the attack of vancomycin by changing the binding epitope from D-Ala-D-Ala to D-Ala-D-Lac (Hughes, 2003). In contrast, nisin is still fully active against the _vanA_-type resistant strain indicating that it binds to lipid II in a different way (Breukink _et al._, 1999). The importance of lipid II in cell wall synthesis implicates that any molecule that binds lipid II with high affinity is a potential antibiotic. Therefore, understanding of the recognition mechanism between nisin and lipid II at a molecular level could ultimately lead to the development of novel antibiotics.

**Table 1.** Structural statistics for the nisin/3LII complex★

| | |
|---|---|
| Distance restraints | |
| Intramolecular | 619 |
| Intermolecular | 36 |
| Hydrogen bonds | 2 |
| Torsion angles | |
| Backbone φ | 6 |
| Side chain χ1 | 7 |
| R.m.s. deviation from experimental restraints | |
| All distance restraints (Å)† | 0.049±0.003 |
| All torsional restraints‡ | 0.40±0.19 |
| Coordinate precision (Å)‖ | |
| Nisin | |
| N, Cα, C' (residues 1-12) | 0.25±0.09 |
| Heavy atoms (residues 1-12) | 0.49±0.16 |
| N, Cα, C' (all) | 6.23±2.30 |
| Heavy atoms (all) | 6.93±2.35 |
| 3LII (all heavy atoms) | 2.67±0.73 |
| Nisin/3LII complex interface¶ | 0.56±0.16 |
| CNS intermolecular energies after DMSO refinement (kcal mol$^{-1}$)§ | |
| Total E$_{total}$ | –151±42 |
| Van der Waals E$_{vdW}$ | –28±5 |
| Electrostatic E$_{elec}$ | –124±42 |
| Interface buried surface area (Å$^2$)# | 842±86 |

★ The statistics is obtained from an ensemble of twenty lowest-energy solution structures of the largest nisin/3LII complex structure cluster. The structures were refined in a shell of explicit DMSO solvent model (Linge *et al.*, 2002). Root mean square (r.m.s.) deviations for bond lengths, bond angles and improper dihedral angles are 0.00060±0.00002Å, 1.69±0.03° and 1.01±0.04°, respectively.

No distance restraint was violated by more than 0.42 Å in any of the final structures.

No dihedral angle restraint was violated by more than 5°.

‖ Coordinate precision is given as the average pairwise Cartesian coordinate r.m.s. deviations over the ensemble, <r.m.s.d.>$_{pairwise}$.

¶ The interface of the complex is defined as the residues 1-12 of nisin and MurNAc, pyrophosphate moiety and the first isoprene unit of 3LII (110 heavy atoms).

§ The non-bonded energies were calculated using OPLS parameters (Jorgensen and Tirado-Rives, 1988) with an 8.5Å cutoff.

# The buried surface area was calculated with CNS (Brünger *et al.*, 1998) using a 1.4Å water probe radius and a 0.005Å precision. It is defined as the solvent accessible surface difference (ΔSAS) when nisin and 3LII were subjected to calculation separately (SAS$^{nisin}$ and SAS$^{3LII}$) and as a whole (SAS$^{nis/3LII}$)

The solution structure of the 1:1 complex of nisin and lipid II was determined in dimethyl sulfoxide (DMSO) using a lipid II variant (3LII) with a shortened prenyl tail of three isoprene units (Figure 1C and Table 1). The interface between nisin and 3LII is defined by 36 intermolecular NOEs between residues 1-10 of nisin and MurNAc and the isoprene units of 3LII. In addition, two intermolecular hydrogen bonds were experimentally identified and were used as distance restraints (Table 1). These involve the amide groups of Dhb2 (dehydrobutyrine) and Abu8 (α-aminobutyric acid) of nisin and the pyrophosphate moiety of 3LII, as identified by the presence of the cross hydrogen bond nitrogen-
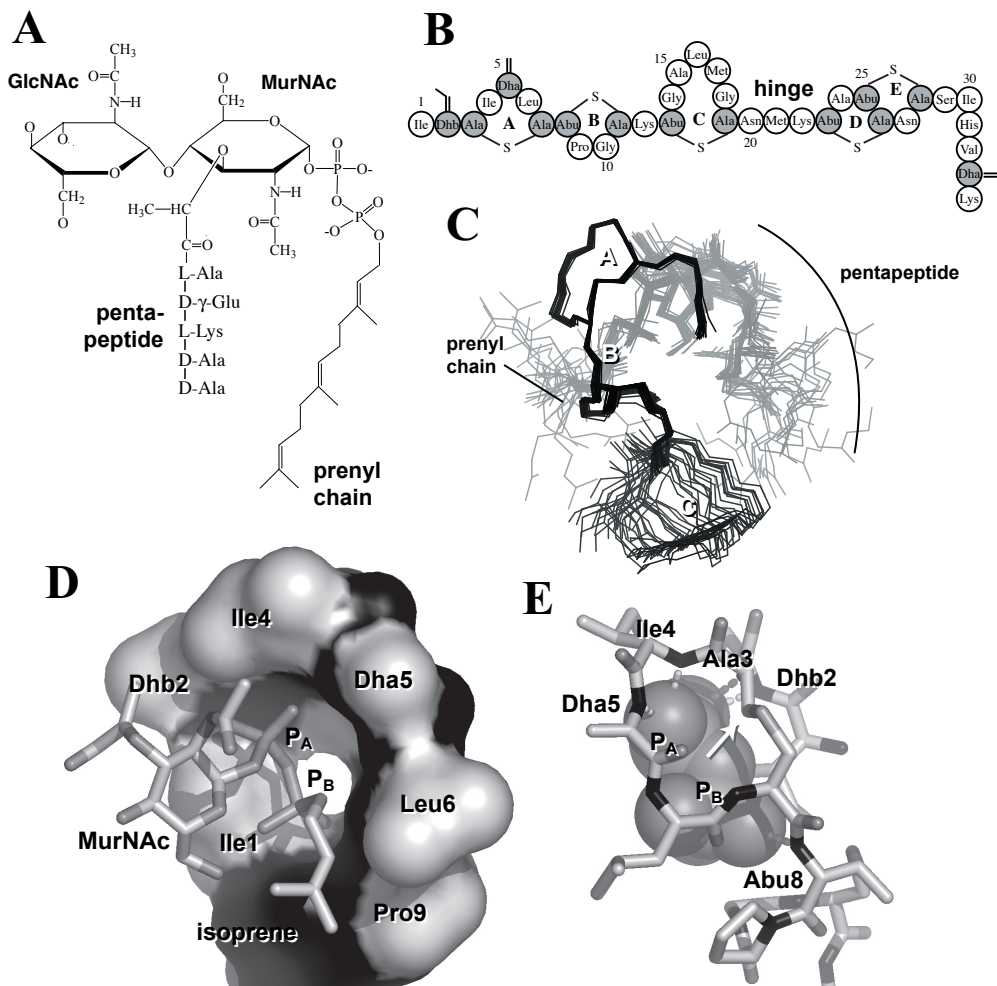
**Figure 1.** Solution structure of the nisin/3LII complex. **A.** Chemical structure of 3LII. **B.** Primary structure of nisin. The lanthionine rings, A-E, and the side chain structures of the dehydrated residues (dehydroalanine, Dha and dehydrobutyrine, Dhb) are shown. **C.** Ensemble of the twenty lowest-energy NMR structures superimposed on the heavy atoms of the interface of the complex (*cf.* Table 1). The backbone of nisin (residues 1-19) including lanthionine linkages is coloured black. The unstructured C-terminal part is omitted for clarity. The pentapeptide, MurNAc, GlcNAc, pyrophosphate moiety and the prenyl chain of 3LII are coloured grey. **D.** The N-terminal part of nisin (residues 1-12 shown in van der Waals surface) encages the pyrophosphate moiety of 3LII. The side chains (light grey) of nisin are labelled. The MurNAc, pyrophosphate and the first isoprene unit of 3LII are shown is sticks and labelled, respectively. **E.** Nisin backbone-3LII pyrophosphate intermolecular hydrogen bond network. Hydrogen bonds with high occurrence in the ensemble of structures (as defined by LIGPLOT (Wallace *et al.*, 1995), see Table 2) are indicated by dashed lines and the corresponding residues are labelled (pyrophosphate group in spheres). The carbon, oxygen and nitrogen atoms of nisin are coloured light grey, dark grey and black, respectively. A white arrow indicates the position of Ala3-Cβ, at which addition of a methyl group can disrupt the bioactivity (see text for details). The figures were generated using Pymol (http://www.pymol.org/).
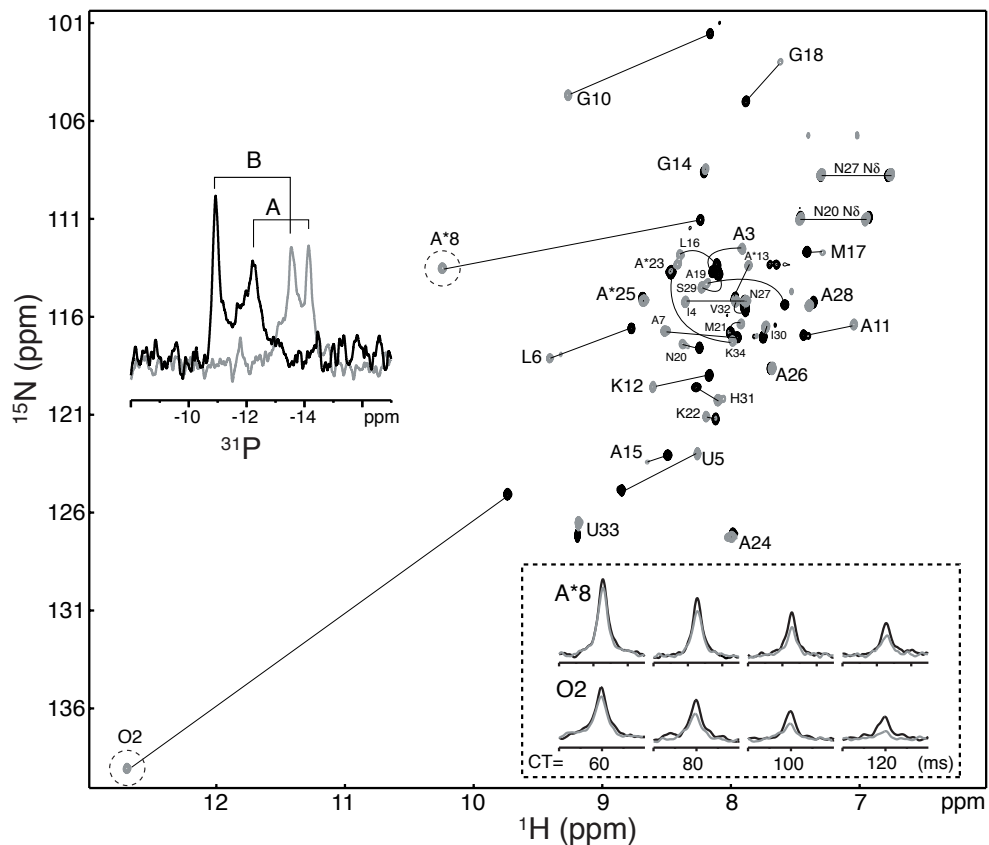
**Figure 2.** Nisin/3LII complex formation. Overlaid ¹H-¹⁵N HSQC spectra of the free (black) and 3LII bound nisin (grey). Changes in chemical shifts are indicated by cyan lines with corresponding residue labels (U: dehydroalanine; O: dehydrobutyrine; A★: α-aminobutyric acid). The amide protons of O2 and A★8 that are hydrogen bonded to the pyrophosphate group of 3LII are indicated by dashed circles. Cross sections of the O2 and A★8 signals along the ¹H dimension in the ³¹P-edited constant time (CT) ¹H-¹⁵N HSQC (Löhr *et al.*, 2000; Mishima *et al.*, 2000) are shown in the lower right inset (dashed box). The intensities of these hydrogen bonded amides are modulated by the cross hydrogen bond scalar coupling to the phosphorous ($^{3h}J_{NP}$) and are attenuated due to transverse relaxation decay during the CT delays (grey). The reference experiment with ³¹P decoupling during CT delays removes the $^{3h}J_{NP}$ coupling modulation (black). Upper left inset shows the free (black) and nisin bound (grey) ³¹P NMR spectra of the pyrophosphate group in 3LII ($\Delta\delta(P_A)$ = -1.91 ppm and $\Delta\delta(P_B)$ = -2.62 ppm). Signals A and B correspond to the individual phosphates that are attached to the MurNAc moiety and the prenyl chain, respectively.

phosphorous (¹⁵N-H...O-³¹P) scalar coupling $^{3h}J_{NP}$ (see Material and Methods) (dashed box inset, Figure 2). The nisin/3LII complex is well-defined at the interface with a rather disordered tail (Figure 1C). This is supported by the fact that a reduction of the intrinsic flexibility of nisin upon 3LII binding is only observed in the N-terminal part, as indicated by the heteronuclear ¹⁵N{¹H}-NOE data (Figure 3). Due to the large number of modified residues, nisin exhibits no canonical secondary structure, but does contain type I β-turns in rings B and C and two consecutive type II and II' β-turns in the intertwined
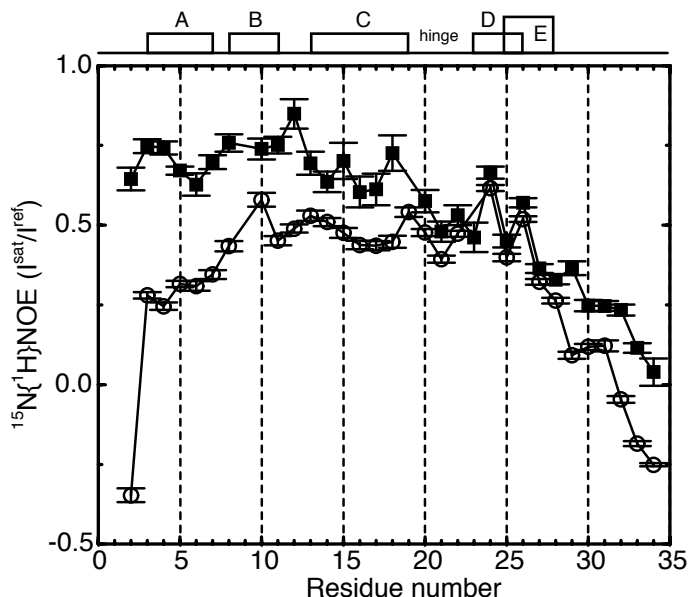
**Figure 3.** $^{15}$N{$^1$H} heteronuclear NOEs of nisin in the free (open circles) and complex (filled squares) forms. $^{15}$N{$^1$H} NOE experiment was carried out using the on- or off-resonance proton irradiations during the last three seconds of the five-second relaxation delays in an interleaved way to avoid differential heating effect. $^{15}$N{$^1$H} NOE is given by the ratio of the $^1$H-saturated (on-resonance) and reference (off-resonance) peak intensities ($I^{sat}/I^{ref}$). An amide group in a structured region will have a value close to unity while an amide group in an unstructured, flexible region will have a value close to zero, or even below.

rings D and E. In contrast to the extended and flexible structure in the absence of 3LII (not shown), upon binding to 3LII, the N-terminus of nisin folds back onto the first two lanthionine rings, A and B, forming a cage-like structure. This cage encompasses the pyrophosphate group of 3LII with a buried surface area of *ca* 850 Å$^2$ (Table 1). This represents more than a third of the 3LII total solvent accessible surface. Except the side chains of Ala3 and Ala7, which are part of the lanthionine ring linkage, all other side chains reside at the rim of the binding cleft (Figure 1D). The unique backbone architecture of the pyrophosphate cage allows the formation of five intermolecular hydrogen bonds between the backbone amides of nisin and the pyrophosphate group (Figure 1E and Table 2). By contrast, another polyprenyl-pyrophosphate binding molecule, the mammalian farnesyltransferase, coordinates the pyrophosphate group solely *via* hydrogen bonding with conserved side chains (Long *et al.*, 2002). As pre-organisation with lanthionine linkages provides a lower entropy loss upon binding, this may be the dominant factor in the recognition mechanism of the pyrophosphate group of lipid II.

The formation of intermolecular hydrogen bonds and close contacts to the highly electronegative phosphate moiety induces large chemical shift perturbations in the N-terminal part of nisin, in line with those previously observed in the lipid II-containing sodium dodecylsulphate (SDS) micelles (Hsu *et al.*, 2002) (Figures 2 and 4). The hydrogen bond formation is responsible for the large downfield shifts along the proton dimension for Dhb2 and Abu8 (dashed circles in Figure 2) and the upfield shifts of the

**Table 2.** Statistics of intermolecular interactions★

*Intermolecular hydrogen bonds*

| nisin | | 3LII | | | |
|-------|---|------|---|---|---|
| Donor | | Acceptor | | Occurrence (%)† | Donor–acceptor distance (Å) |
| Dhb2 | N | $P_A$ | O1 | 100 | 2.81±0.03 |
| Ala3 | N | $P_A$ | O1 | 100 | 3.07±0.08 |
| Ile4 | N | $P_A$ | O2 | 100 | 2.60±0.03 |
| Dha5 | N | $P_A$ | O2 | 100 | 2.87±0.12 |
| Abu8 | N | $P_B$ | O2 | 100 | 3.00±0.09 |

*Non-bonded intermolecular contacts*

| nisin | | 3LII | | Occurrence (%)† | Atom pair distance (Å) |
|-------|---|------|---|-----------------|------------------------|
| Leu6 | Cδ1 | Isoprene1 | C2 | 55 | 3.73±0.07 |
| Gly10 | Cα | Isoprene1 | C2 | 100 | 3.44±0.13 |
| Gly10 | Cα | Isoprene1 | C3 | 100 | 3.48±0.10 |
| Pro9 | C | Isoprene1 | C4 | 100 | 3.57±0.14 |
| Pro9 | Cβ | Isoprene1 | C4 | 100 | 3.46±0.14 |
| Gly10 | Cα | Isoprene1 | C4 | 100 | 3.58±0.14 |
| Abu13 | Cγ2 | Isoprene1 | C4 | 70 | 3.68±0.13 |
| Pro9 | Cβ | Isoprene2 | C1 | 85 | 3.43±0.08 |

★ The intermolecular hydrogen bonds and non-bonded contacts were analysed by LIGPLOT (Wallace *et al.*, 1995) using the default settings. 2.7Å and 3.35Å proton-acceptor and donor-acceptor distance cutoffs, respectively with minimum 90° angles (D-H-A, H-A-AA, D-A-AA) for hydrogen bonds; 3.9Å heavy atom pair distance cutoff for non-bonded contacts.

† The minimum occurrence cutoff was set to 50% (10 out of 20 structures in the final DMSO refined ensemble).

two phosphorous nuclei (inset, Figure 2). Apart from the two experimentally identified intermolecular hydrogen bonds, additional ones are present in the complex structure (Figure 1C), consistent with the reduction of the amide proton chemical shift temperature dependency (temperature coefficient) (Hsu *et al.*, 2002) upon complex formation (Figure 5). Because the scalar coupling $^{3h}J_{NP}$ is highly dependent on the hydrogen bond geometry (Czernek and Brüschweiler, 2001), some of these couplings may be too weak to be observed in our experimental setting.

The binding mode involving the pyrophosphate moiety as revealed here shows that nisin does not target the same motif of lipid II as does vancomycin (Breukink *et al.*, 1999), and possibly ramoplanin (McCafferty *et al.*, 2002). It also explains why nisin does not differentiate between lipid II and lipid I, which differ in the GlcNAc moiety (Brötz *et al.*, 1998). The chemical shift changes in D-Ala-D-Ala and the terminal isoprene units of 3LII upon nisin binding were negligible (Figure 4); only a localised perturbation was observed in the β1-4 linkage between MurNAc and GlcNAc. These leave the pyrophosphate, MurNAc and the first isoprene in lipid II and the N-terminal part of nisin as the determinants for nisin-lipid II recognition. The structure of the complex explains several earlier observations. First, extension of the N-terminus, [ITPQ]-nisin, dramatically diminishes the antimicrobial activity (Kuipers *et al.*, 1993). Second, the introduction of an extra methyl group at the side chain of Ala3, [S3T]-nisin, decreases the lipid II binding affinity of nisin 50-fold (Wiedemann *et al.*, 2001). In contrast, similar modifications in the second position, Dhb2, had little effect (Wiedemann *et al.*, 2001). Third, [ΔAla5]-nisin, with ring A nicked open by hydrolytic cleavage, is biologically inactive (Chan
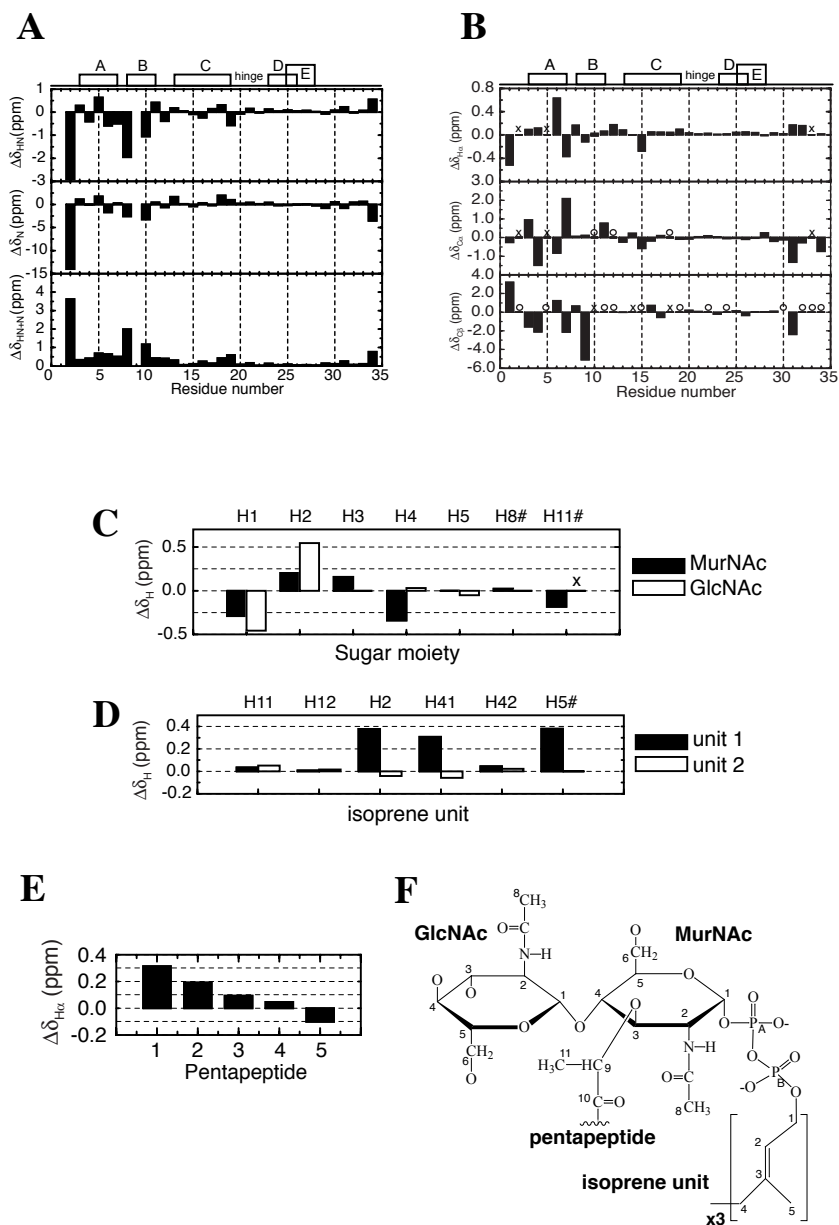
**Figure 4.** Chemical shift perturbations ($\Delta\delta = \delta_{free} - \delta_{complex}$) upon nisin/3LII complexation. **A.** Nisin backbone amide proton (top), amide nitrogen (middle) and weighted sum (bottom) ($\Delta\delta_{HN+N} = [(\Delta\delta_{HN})^2 + (0.154 \star \Delta\delta_N)^2]^{1/2}$). **B.** H$\alpha$ (top), C$\alpha$ (middle) and C$\beta$ (bottom) of nisin. The unassigned resonances are labelled with open circles (O) and the non–existing resonances are labelled with cross (x); **C**. Protons of the sugar moieties of 3LII (the atom nomenclatures are labelled on top and are defined in **F**; **D.** Protons of the isoprene units. The last isoprene unit was not fully assigned due to fast motions and the lack of NOEs; **E.** H$\alpha$ of the pentapeptide of 3LII; **F.** The nomenclature of the sugars and the isoprene units of 3LII.
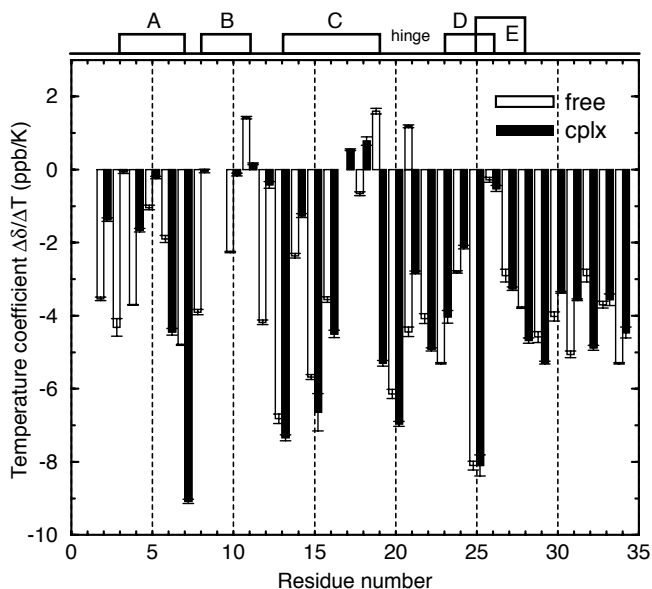
**Figure 5.** Temperature coefficients of nisin amide protons (HN) in the free (open bars) and complex (filled bars) forms. The temperature coefficients were derived from the temperature dependency of the amide proton chemical shifts ($\Delta\delta/\Delta T$) using a linear fitting of individual amide proton resonances obtain from $^1$H-$^{15}$N HSQC spectra recorded at 290, 295, 300, 305, 310 and 315K. The unit of the chemical shift is given in part per billion (ppb).

*et al.*, 1989). All these observations can be rationalised when examining the interface of the complex structure: i) since side chain of the N-terminal Ile1 is part of the interface (Figure 1C), extension of the N-terminus may interfere with the interfacial complimentarity; ii) the additional methyl group in [S3T]–nisin occupies the encaging cavity (white arrow, Figure 1E) and therefore hampers the formation of intermolecular hydrogen bonds; iii) the ring opening loses the rigidity of the ring structure and hence destabilises the intermolecular hydrogen bond between Dha5 and the pyrophosphate moiety.

Throughout evolution, the lanthionine ring structure of nisin, in particular, rings A and B, are well-conserved amongst several lipid II-binding lantibiotics (Sahl and Bierbaum, 1998; Guder *et al.*, 2000). Efforts in creating more effective antibiotics by changing the side chain compositions of nisin-like peptides have only achieved a marginal improvement of bioactivity to date (Liu and Hansen, 1992). This can now be understood since binding to lipid II is dictated mainly by the backbone scaffold of the ring structures. Using the nisin/3LII complex as a template, we performed homology modelling of several related lantibiotics containing conserved lanthionine rings A and B (Figure 6). This reveals that the pyrophosphate cage can accommodate a variety of side chain compositions. Apart from the conserved positions 3 and 7-11, all observed differences amongst the homologs reside at the circumference of the pyrophosphate cage. The occurrence of various amino acid types, both hydrophilic and hydrophobic, suggests that the side chain interactions are of minor importance in lipid II binding. Notably, the hydrophobic character at position 6 remains preserved. This part interacts with the prenyl chain (Table 2) and possibly also the membrane bilayer in which lipid II is anchored. This notion of a common motif

Table 2. **Statistics of intermolecular interactions★**

| *Intermolecular hydrogen bonds* | | | | | |
|---|---|---|---|---|---|
| nisin | | 3LII | | | |
| Donor | | Acceptor | | Occurrence (%)† | Donor–acceptor distance (Å) |
| Dhb2 | N | $P_A$ | O1 | 100 | 2.81±0.03 |
| Ala3 | N | $P_A$ | O1 | 100 | 3.07±0.08 |
| Ile4 | N | $P_A$ | O2 | 100 | 2.60±0.03 |
| Dha5 | N | $P_A$ | O2 | 100 | 2.87±0.12 |
| Abu8 | N | $P_B$ | O2 | 100 | 3.00±0.09 |
| *Non-bonded intermolecular contacts* | | | | | |
| nisin | | 3LII | | Occurrence (%)† | Atom pair distance (Å) |
| Leu6 | Cδ1 | Isoprene1 | C2 | 55 | 3.73±0.07 |
| Gly10 | Cα | Isoprene1 | C2 | 100 | 3.44±0.13 |
| Gly10 | Cα | Isoprene1 | C3 | 100 | 3.48±0.10 |
| Pro9 | C | Isoprene1 | C4 | 100 | 3.57±0.14 |
| Pro9 | Cβ | Isoprene1 | C4 | 100 | 3.46±0.14 |
| Gly10 | Cα | Isoprene1 | C4 | 100 | 3.58±0.14 |
| Abu13 | Cγ2 | Isoprene1 | C4 | 70 | 3.68±0.13 |
| Pro9 | Cβ | Isoprene2 | C1 | 85 | 3.43±0.08 |

★ The intermolecular hydrogen bonds and non-bonded contacts were analysed by LIGPLOT (Wallace *et al.*, 1995) using the default settings. 2.7Å and 3.35Å proton-acceptor and donor-acceptor distance cutoffs, respectively with minimum 90º angles (D-H-A, H-A-AA, D-A-AA) for hydrogen bonds; 3.9Å heavy atom pair distance cutoff for non-bonded contacts.

† The minimum occurrence cutoff was set to 50% (10 out of 20 structures in the final DMSO refined ensemble).

for lipid II recognition is strengthened by the observation that lipid II is also the target of epidermin (Brötz *et al.*, 1998) and mutacin 1140 (unpublished data) (Figure 6B). We therefore propose that the pyrophosphate cage is the general lipid II-binding architecture for these lantibiotics.

The formation of a 1:1 nisin/lipid II complex is the first step towards pore formation. Once nisin is anchored onto lipid II on the membrane surface, subsequent conformational changes take place that lead to the assembly of a higher order oligomeric complex (Breukink *et al.*, 2003). Pore formation requires a transmembrane orientation of nisin involving insertion of its C-terminal part (van Heusden *et al.*, 2002) and a flexible hinge region (residues 20-22) (Wiedemann *et al.*, 2001). Perhaps one can conjecture the early stage of this multi-step process based on the current nisin/3LII complex structure. The positioning of the N-terminal part of nisin with respect to the prenyl chain in the nisin/3LII complex structure may already pave the way for subsequent membrane insertion. Assuming that the interface structure of nisin/3LII complex is preserved in the membrane bilayer we propose a model for the membrane insertion of nisin based on the current nisin/3LII complex structure (Figure 7). This shows that initially ring C of nisin would readily be embedded into the membrane at a comparable depth as the first few isoprene repeats (Figure 7). The modelling demonstrates that nisin is capable of spanning the membrane bilayer with the pyrophosphate cage as the anchoring point. However, this requires conformational flexibility in the hinge region, which is depicted by the three conformations taken from the ensemble of calculated structures (Figure 7). Lack of a flexible hinge by truncation or proline replacement would hamper this insertion process, explaining the lack of pore-formation activity of these mutants (Wiedemann *et al.*, 2001).
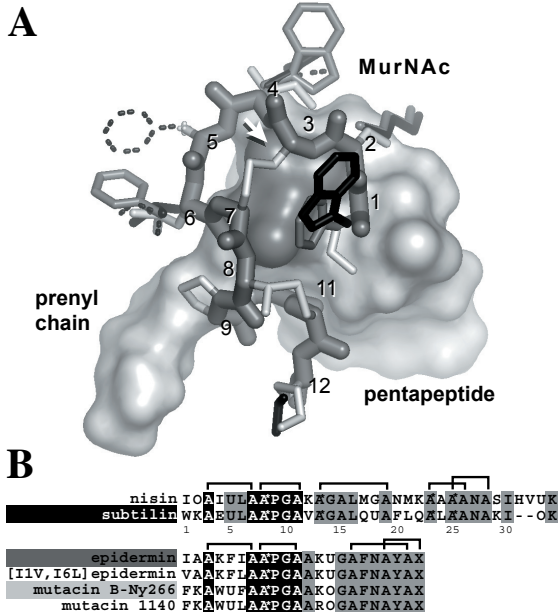
**Figure 6.** General binding model of lipid II-binding lantibiotics. **A.** *In silico* mutagenesis was performed to model the side chains (see below) of subtilin (black), epidermin (dark grey) and mutacin B–Ny266 (light grey) based on the nisin structure (white). The backbone is shown in black with the conserved side chains in white. Only the first twelve residues are shown; 3LII is shown in van der Waals surface. The position of the Ala3 side chain is indicted by a white arrow. The structure has the same orientation as in Figure 1C. **B.** Primary sequence alignment of type A lantibiotics. Nisin and subtilin form a subfamily of type A lantibiotics. Another subfamily consists of epidermin, [I1V,I6L]–epidermin, mutacin B-Ny266 and mutacin 1140 (Guder *et al.*, 2000; Smith *et al.*, 2000) with a cyclised C-terminus (X: 2-aminovinyl–D–cysteine, AviCys). The universally conserved residues are highlighted in black and the conserved residues within the subfamilies are shaded in grey. The lanthionine linkages are indicated above each subfamily.
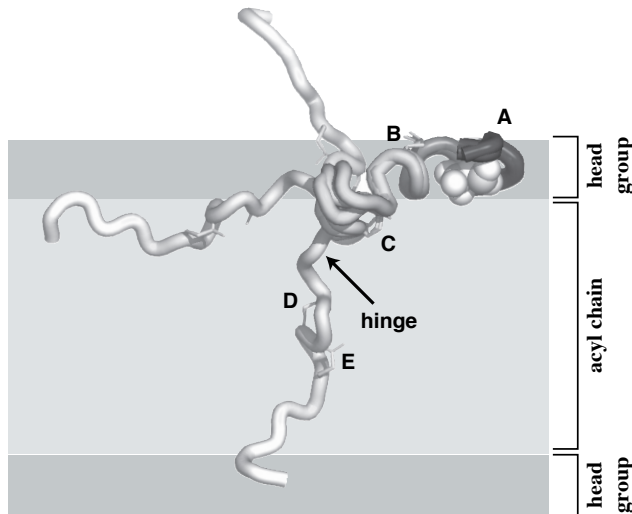


**Figure 7.** Model of the nisin/lipid II complex in a membrane bilayer. The backbone of three representative structures of nisin are shown and coloured from black to white from N- to C-termini. Each conformation, taken from the ensemble of nisin/3LII complex structures, corresponds to a possible orientation of the C-terminal part of nisin outside, on the surface, or inserted into the membrane bilayer, respectively. They are orientated in such a way that the pyrophosphate group of 3LII (spheres) is at the same depth as head groups of lipid molecules (dark grey) in a membrane with a thickness of ~40Å.

In conclusion, this study has revealed a novel lipid II recognition motif for nisin and related lantibiotics. The pyrophosphate moiety plays an essential role in the bacterial cell wall synthesis. Its chemical entity is the Achilles' heel of bacteria that cannot be altered or replaced by simple mutations as opposed to alteration of the attached pentapeptide that can lead to vancomycin–resistance. Since its discovery in 1928, nisin is used as a food preservative and in animal health products, *e.g.*, treatments of mastitis. Penicillin, on the other hand, discovered at the same time as nisin, is no longer effective for the treatment of hospital–acquired infections, as reported by the World Health Organisation (WHO, http://www.who.int/). In essence, the structure nisin/3LII complex as described here can provide a basis for future antibiotic design.

## Materials and Methods

### Sample preparation

Uniformly ^{15}N-labelled nisin and unlabelled 3LII were produced as described previously (Hsu *et al.*, 2002; Breukink *et al.*, 2003). The nisin/3LII complex was prepared by mixing equal molar amounts of lyophilised [$U$-^{15}N]–nisin and 3LII in water followed by centrifugation. The supernatant was removed and the insoluble complex pellet was rinsed with water several times. The pellet was lyophilised and dissolved in 99.9% perdeuterated $d_6$–DMSO (Cambridge Stable Isotope, Inc.) reaching a final concentration of *ca* 1.2 mM. Free [$U$-^{15}N]–nisin and 3LII samples were prepared by dissolving the lyophilised sample into $d_6$–DMSO to reach similar sample concentrations.

### NMR spectroscopy

NMR spectra were recorded at 300K on Bruker DRX500, DRX600, DRX750 and DRX900 spectrometers equipped with a triple resonance or a cryogenic probe (DRX600) for ^{1}H, ^{15}N and ^{13}C detection and a multinuclear (DRX500) or a broadband probe (DRX600) for ^{31}P detection. Spectra were processed with NMRPipe (Delaglio *et al.*, 1995) and analysed using NMRView (Johnson and Blevins, 1994). The ^{1}H and ^{15}N chemical shifts were assigned using standard methods (Cavanagh *et al.*, 1996). [13]C chemical shifts of the well-resolved peaks were assigned from ^{1}H-^{13}C HSQC and [1]H-^{13}C heteronuclear multiple bond correlation (HMBC) (Cavanagh *et al.*, 1996) recorded with natural abundance. The two ^{31}P chemical shifts of the pyrophosphate groups in 3LII were distinguished by ^{1}H-^{31}P HMBC. The cross hydrogen bond scalar coupling $^{3h}J_{NP}$ was identified by ^{31}P–edited constant time (CT) ^{1}H-^{15}N HSQC (Löhr *et al.*, 2000; Mishima *et al.*, 2000) with and without ^{31}P decoupling during the constant time delays of 60, 80, 100 and 120 ms. The assigned chemical shift data sets were deposited into the BioMagResBank under accession codes 6144, 6145 and 6146 for free nisin, free 3LII and their complex, respectively.

### Structure calculation and modelling

Distance restraints were derived from three-dimensional ^{15}N NOESY-HSQC and two-dimensional homonuclear [1]H-^{1}H NOESY. Intermolecular restraints were obtained from ^{15}N-edited filter experiments (Cavanagh *et al.*, 1996). Two experimentally identified intermolecular hydrogen bonds (Dhb2-NH and Abu8-NH to pyrophosphate group in 3LII) were introduced in the later stages of the calculations as ambiguous distance restraints from the amide groups to any of the oxygen atoms in the pyrophosphate

group with bounds of 1.6–2.4Å (H–O) and 2.5-3.3Å (N-O). Backbone $\phi$ dihedral angle restraints were derived from $^3$J(HN,H$\alpha$) coupling constants determined by three-dimensional HNHA (Cavanagh *et al.*, 1996). The side chain rotameric states ($\chi$1) were derived from the intensity-based comparison of NOESY/DQF-COSY spectra (Basus, 1989). The experimentally determined distance and dihedral restraints (Table 1) were applied in a simulated annealing protocol using CNS (Brünger *et al.*, 1998) and HADDOCK (Dominguez *et al.*, 2003) with nisin and 3LII treated as fully flexible for docking. The structures were refined using an explicit DMSO solvent model (Linge *et al.*, 2002). Homology modelling was achieved using Pymol (http://www.pymol.org/) from the representative nisin/3LII complex structure (the closest to the mean). Side chains were mutated *in silico* followed by manual adjustment of the rotamer conformations and subsequent energy minimization. The atomic coordinates of the nisin/3LII complex (PDB entry 1UZT) have been deposited in the Protein Data Bank at the Macromolecular Structure Database site at the European Bioinformatics Institute, Cambridge, UK.

## Acknowledgement

# NMR study of mersacidin and
# lipid II interaction in DPC micelles:

## conformational changes are a key to antimicrobial activity[#]

Shang–Te D. Hsu[1], Eefjan Breukink[2], Gabriele Bierbaum[3], Hans–Georg Sahl[3],
Ben de Kruijff [2], Robert Kaptein[1], Nico A.J. van Nuland[1] and Alexandre M.J.J. Bonvin[1]

[1]Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research and
[2]Institute of Biomembranes, Center of Biomembranes and Lipid Enzymology,
Utrecht University, 3584CH Utrecht, The Netherlands
[3]Institute of Medical Microbiology and Immunolgy, University of Bonn, D-53105 Bonn, Germany
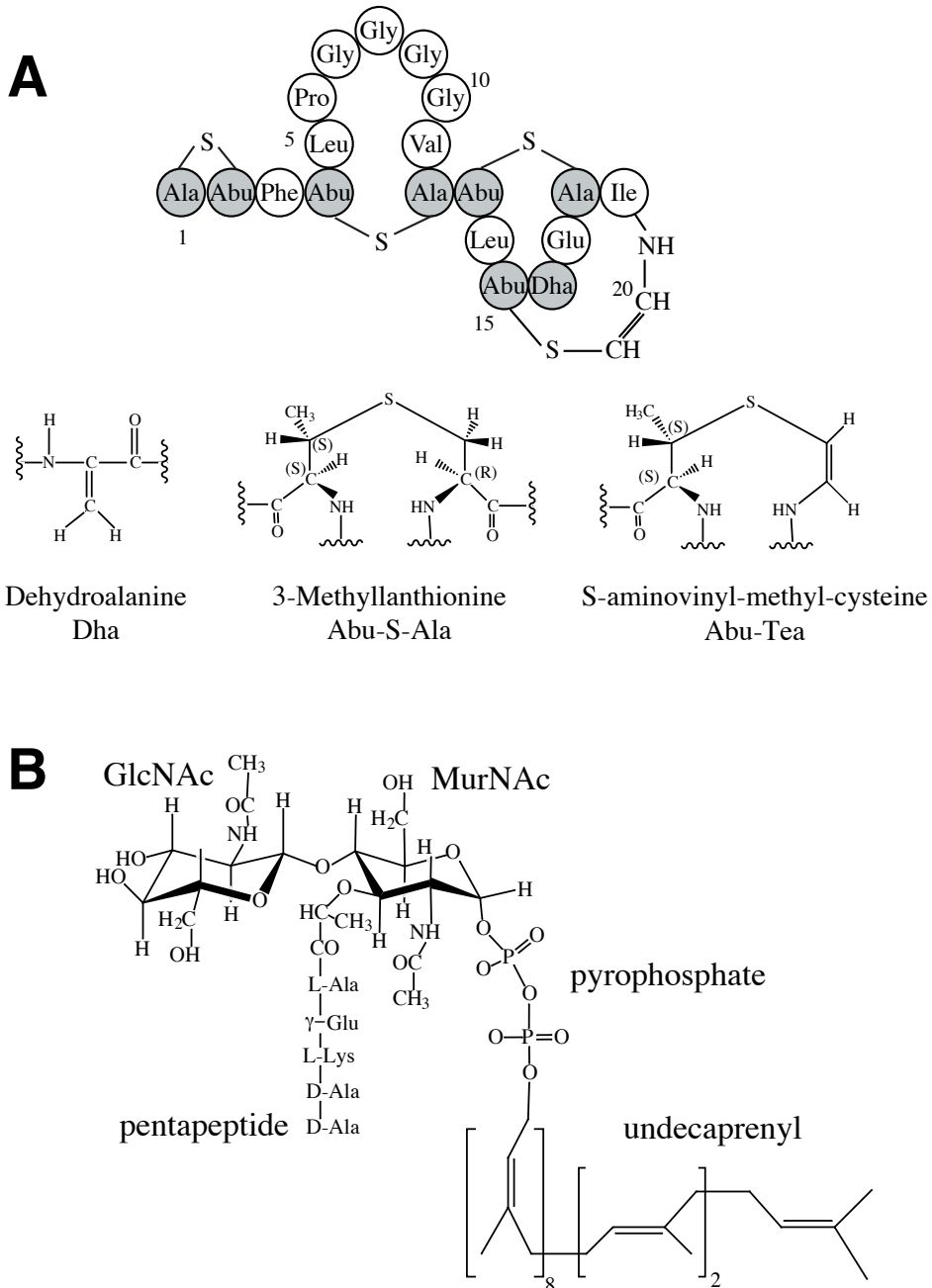
## Abstract

**Mersacidin belongs to the type B lantibiotics (lanthionine-containing <u>antibiotics</u>) that contain post-translationally modified amino acids and cyclic ring structures. It targets the cell wall precursor lipid II and thereby inhibits cell wall synthesis. In light of the emerging antibiotics resistance problem, the understanding of the antibacterial activity on a structural basis provides a key to circumvent this issue. Here we present solution NMR studies of mersacidin-lipid II interaction in DPC micelles. Distinct solution structures of mersacidin were determined in three different states: in a water/methanol solution and in DPC micelles with and without lipid II. The structures in various sample conditions reveal remarkable conformational changes, in which the junction between Ala12 and Abu13 effectively serves as the hinge for the opening and closure of the ring structures. The DPC micelle-bound form resembles the previously determined NMR and X-ray crystal structures of mersacidin in pure methanol, but substantially deviates from the other two states in our current report. The structural changes delineate the large chemical shift perturbations observed during the course of a two-step $^{15}$N-$^1$H HSQC titration. They also modulate the surface charges distribution of mersacidin suggesting that electrostatics play a central role in the mersacidin-lipid II interaction. The observed conformational adaptability of mersacidin might be a general feature of lipid II-interacting antibiotics/peptides.**

## Introduction

Many antimicrobial peptides act against microorganisms through pore formation on the cell membrane. The permeability originates in principle from a nonspecific assembly that results in a pore-like structure, where the amphipathic nature of the amino acid composition facilitates the clustering process (for review, see (Zasloff, 2002)). Apart from this ubiquitous mechanism, some antimicrobial peptides, such as ramoplanin, enduramycin and janiemycin and the glycopeptides vancomycin and teicoplanin, use specific targets that plays a central role in cell wall synthesis, namely lipid II, to achieve their bioactivity with much higher efficiency (Nagarajan, 1993; Brötz *et al.*, 1998b; Cudic *et al.*, 2002). Lipid II also serves as a target for the pore forming peptides, nisin and epidermin, which belong to the type A lantibiotic (lanthionine-containing <u>antibiotic</u>) family (Brötz *et al.*, 1998b; Breukink *et al.*, 1999). Some type B lantibiotics also possess functions of targeted cell wall synthesis inhibition; one of the best studied peptide is mersacidin (Figure 1A) (Brötz *et al.*, 1997; Brötz *et al.*, 1998a). Although these lipid II-targeting antimicrobial peptides share a common binding molecule, the recognition epitopes among these peptides are somewhat different: co-incubation of vancomycin or other inhibitors of transglycosylases or transpeptidases with mersacidin does not impede its lipid II binding capacity (Brötz *et al.*, 1998a). As the antibiotic resistance is becoming more and more severe, the diversity of such a targeting action is of great interest. The sophisticated chemical composition of lipid II provides the complexity that can be targeted in various ways. It consists of a peptidoglycan headgroup that serves as the building block for the cross-linked cell wall matrix and of a pyrophosphate-undecaprenyl lipid tail that functions as the carrier for the transport of the peptidoglycan moiety from the cytoplasm to the extracellular domain (Figure 1B). Although

**A**



**Dehydroalanine
Dha**

**3-Methyllanthionine
Abu-S-Ala**

**S-aminovinyl-methyl-cysteine
Abu-Tea**

**B**

GlcNAc    MurNAc

pyrophosphate

pentapeptide    D-Ala

undecaprenyl

**Figure 1.** Primary structures of mersacidin (**A**) and lipid II (**B**). **A.** Post-translationally modified amino acids are highlighted in grey and their chemical structures are depicted below. **B.** GlcNAc, *N*-acetylglucosamine; MurNAc, *N*-acetylmuramic acid.

the targeted antimicrobial activity of mersacidin is evident, little detail is known, however, about its mechanism of recognition and inhibition.

Mersacidin is a 20-residue peptide with nine post-translationally modified amino acids and a single negatively charged residue, Glu17 (Figure 1A) (for reviews, see (Brötz and Sahl, 2000; Jack and Jung, 2000)). It contains four ring structures: two separate ones in the N-terminal part and two intertwined ones in the C-terminal part. The three-dimensional structure of mersacidin has been solved both by solution NMR spectroscopy (Prasch *et al.*, 1997) and X-ray crystallography (Schneider *et al.*, 2000). Unlike type A lantibiotics, which are mostly extended and flexible, the structure of mersacidin is globular and compact. In both the crystalline and solution states the local ring structures are tightly confined by the lanthionine linkages. The overall conformations obtained from the two different methods are similar, except for a minor difference in the orientation of the glycine-rich ring (residues 5-11). These structures were however both solved in pure methanol due to the poor solubility of mersacidin in aqueous solution, and in the absence of lipid II, which is required for its bioactivity. In order to understand this mechanism, knowledge of its structure upon binding to lipid II under physiological conditions is crucial.

We report here high resolution NMR spectroscopy studies of the interaction between mersacidin and its binding target lipid II in DPC micelles, which were used as a membrane mimic. $^{15}N$-$^{1}H$ HSQC titration experiments provide the residue-specific insight into the interactions with lipid II. The structures of mersacidin in various sample environments, namely free in methanol/water solution and DPC bound in the absence and presence of lipid II, were determined by NOE-based structure calculations with explicit solvent refinements. The effect of binding to DPC micelles and lipid II on the mersacidin dynamics are characterised by means of $^{15}N$-relaxation together with gradient-edited diffusion experiments. Despite the large number of published solution NMR structural and relaxation dynamics studies of membrane proteins and peptides (Williams *et al.*, 1996; Opella, 1997; Bader *et al.*, 2001; Neidigh *et al.*, 2001; Opella *et al.*, 2001), only a few examples of high-resolution NMR studies of protein/peptide-ligand interactions in the presence of membrane-like environments are available to date (Kutateladze and Overduin, 2001; Hsu *et al.*, 2002). We will show that the differences in sample environments result in substantial conformational changes that modulate charge accessibility. These changes in charge distributions most likely play a crucial role in the mechanism of mersacidin bioactivity.

## Material and methods

### Sample preparation

For the overproduction and purification of $^{15}N$-mersacidin, the producer strain *Bacillus* spec. HIL Y-85,54728 was inoculated into 10 ml of tryptone soy broth and incubated overnight. The preculture served as inoculum for 10 ml of synthetic production medium (Bierbaum *et al.*, 1995) which contained 50 mM $^{15}NH_4Cl$ (98 %+) (Cambridge Isotope Laboratories. Inc., Cambridge, UK), and 10 mM $K_2SO_4$, and was incubated at 30°C to an optical density of 0.1 at 600 nm. 0.1 ml of this culture served as inoculum for ten 100 ml cultures in the same medium, which were incubated for 72 h at 30°C with vigorous agitation in 1 l Erlenmeyer flasks. The culture supernatant was sterilized by filtration and loaded onto a 50 ml Serdolit PAD-I column (Serva Electrophoresis GmbH, Heidelberg, Germany), which had been washed with methanol and equilibrated with distilled water. The column was washed with 10 bed volumes of distilled water and 10 bed volumes of 50 % methanol in 50 mM potassium phosphate, pH

7 and the peptide was eluted with 500 ml acetonitrile, 0.1 % trifluoric acid. The antibacterial activity of the fractions was checked in a bioassay employing *Micrococcus luteus* ATCC 4698 as an indicator strain. Active fractions were pooled, concentrated by rotary evaporation and precipitated proteins were removed by centrifugation. Aliquots of the concentrate were applied to a POROS 20 R2 HPLC column (Applied Biosystems, Weiterstadt, Germany) using the following gradient (eluent A: 0.1% trifluoric acid in water, eluent B: 0.1% trifluoric acid in acetonitrile): 0 min 5% B, 12 min 30% B, 20 min 40% B, 22 min 100% B, at a flow of 5 ml/min. Active fractions were pooled, concentrated by evaporation and rechromatographed on the POROS column. After lyophilisation of the active fractions, the peptide was applied to a reversed phase HPLC column (RP18) employing the following gradient: 0 min 5% B, 30 min 50% B, 44 min 67.5% B, 47 min 100% B. The mass of the purified peptide, which eluted at 55% B, was checked by mass spectrometry. Lipid II was prepared as described earlier (Brötz *et al.*, 1997).

## NMR Spectroscopy for spectral assignments

Due to the solubility problem of mersacidin, freeze-dried mersacidin was first dissolved in perdeuterated $d_3$-methanol (Cambridge Isotope Laboratories. Inc., Cambridge, UK) as a 10 mg/ml stock solution. It was then diluted with sodium phosphate buffer and dodecylphosphocholine (DPC) solution and water. Lipid II was taken from a stock solution (in $CHCl_3$:MeOH = 1:1) and vacuum-dried before mixing. Mersacidin-containing DPC solution was then added to dissolve lipid II and the resulting sample solution was transferred into a NMR tube for measurements. A short sonication was applied after each mixing step to ensure uniform mixing and proper micelle formation. The typical sample concentration was 2 mM mersacidin in 10 mM sodium phosphate buffer at pH 6.0 with a total volume of 500 µl. The sample hence contained 37% methanol and 63% $H_2O$. For structure calculation purposes, 4% (~100 mM) perdeuterated $d_{38}$-DPC (Cambridge Isotope Laboratories. Inc., Cambridge, UK) with or without 2 mM lipid II was added to obtain lipid II embedded in DPC micelles and a control sample of DPC micelles alone. For simplicity, *MeOH/H2O*, *lipid II* and *DPC_{bound}* are defined here as the three sample conditions of free mersacidin in methanol/water mixture and in the DPC micelle solution with and without lipid II, respectively.

All NMR experiments were carried out on Varian UnityPlus 500 and Bruker DRX600 and DRX750 spectrometers at 293 K. Spectra including 2D NOESY with mixing times of 50, 100, 200 ms, a TOCSY with mixing time of 70 ms and a 3D NOESY-HSQC with mixing time of 100 ms were collected to obtain complete backbone and side chain proton resonance assignments (Cavanagh *et al.*, 1996). A 2D DQF-COSY was recorded to extract the backbone $^3J_{HN-H\alpha}$ coupling constants in combination with a short mixing time (50 ms) NOESY as described earlier (Ludvigsen *et al.*, 1991). Stereospecific assignment of side chain methylene groups was achieved based on the intensity correlations of the spin systems (Basus, 1989; Nilges *et al.*, 1990; Xu *et al.*, 1992). All spectra were processed using the NMRPipe software package (Delaglio *et al.*, 1995) and analysed with NMRView (Johnson and Blevins, 1994) except for those of diffusion measurements (see below). The assigned chemical shift data sets of mersacidin in three different conditions, *MeOH/H2O*, *DPC_{bound}* and *Lipid II*, were deposited in BioMagResBank (BMRB) under accession codes 5581, 5582, and 5580, respectively.

## Two-step $^{15}N$-$^1H$ HSQC titration experiments

For the starting point corresponding to the free mersacidin in 37%/63% MeOH/$H_2O$ mixture, 1 µmole

of the $^{15}$N-labelled mersacidin stock solution in d$_3$-methanol was diluted with 50 μl of 100 mM sodium phosphate buffer and water was added to reach a total volume of 450 μl. In the first titration step, aliquots of a 40% DPC stock solution with 37% methanol content were successively added resulting in DPC concentration of 0.2, 1, 2, 3, 4, 5 and 6%. This is respectively equivalent to a range of ratios of mersacidin versus DPC micelles of 20:1 up to 1:1.5, assuming that each DPC micelle consists of 50-55 monomers (Wymore *et al.*, 1999; Marrink *et al.*, 2000). At the end of the first titration step (*DPC$_{bound}$*) the total volume was 525 μl resulting in a 16% sample dilution but with identical methanol content. For the second titration step, portions of 0.05, 0.25, 0.5, 0.75, 1, 1.25 and 1.5 μmole of vacuum-dried lipid II were first prepared in separate containers and then successively dissolved in the sample taken from the NMR tube to reach the same concentration ratios as those of the DPC titration steps. The final sample (*lipid II*) therefore contained 1 μmole of $^{15}$N-mersacidin, ~1.5 μmole of DPC micelles and 1.5 μmole of lipid II. In both titration steps, the chemical shift changes of mersacidin were all saturated by excess of DPC micelles and lipid II at a 1:1.25 molar ratio and higher.

## $^{15}$N relaxation experiments for backbone dynamics

$^{15}$N longitudinal relaxation times T$_1$ and transverse relaxation times T$_2$ and $^{15}$H-$^1$N steady-state NOEs of mersacidin were obtained from series of two-dimensional experiments with coherence selection achieved by pulse field gradients (PFG) (Farrow *et al.*, 1994). For T$_1$ measurements, seven spectra were recorded with relaxation delays T set to 10, 20, 100, 300, 500, 750 and 1000 ms. For the rotating frame relaxation times T$_{1\rho}$, spectra were first obtained using a $^{15}$N-spin lock continuous wave radio frequency (rf) with a field strength $\nu_1$ of 1 kHz for seven relaxation delays T of 20, 50, 70, 100, 120, 200 and 350 ms. The relaxation times T$_1$ and T$_{1\rho}$ were then derived from a single exponential decay fitting of the peak intensities using *xcrvfit* (http://www.pence.ca/ftp/xcrvfit). The T$_2$ of each residue was subsequently derived from the observed relaxation time T$_{1\rho}$ by correcting for the offset $\Delta\nu$ of the rf field to the resonance by use of the relation $1/T_{1\rho} = (1/T_1) \cos^2\theta + (1/T_2) \sin^2\theta$, where $\theta = \tan^{-1}(\nu_1 / \Delta\nu)$. $^{15}$N-$^1$H heteronuclear NOEs were determined from the ratio of peak intensities with and without the saturation (I$^{on}$/I$^{off}$) of the amide protons for 3 sec. All $^{15}$N relaxation experiments were carried out on a Varian UnityPlus 500 spectrometer at 293 K.

Since the structure of mersacidin exhibits high flexibility, quantitative analysis of the relaxation data in terms of spectral density functions is not applicable (Dayie *et al.*, 1996). Therefore, the relaxation data in the three different sample conditions were only compared in a qualitative way. A further simplification allows us to estimate the correlation time from the averaged T$_2$ value (Anglister *et al.*, 1993):

$$\tau_c \sim 1/(5T_2) \tag{1}$$

The overall correlation time can also be expressed as a function of the molecular size and the bulk solvent viscosity $\eta$. For a spherical molecule of radius *a* rotating in a liquid of viscosity $\eta$, the rotational correlation time $\tau_c$ is given by the Stokes relation (Wand *et al.*, 1998):

$$\tau_c = 4\pi a^3\eta/3kT = V\eta/kT = 1/6D_{rot} \tag{2}$$

where *V* is the volume of the molecule, *k* is the Boltzman constant, *T* is the absolute temperature, and $D_{rot}$ is the rotational diffusion constant for a spherical molecule. The molecular radius *a* is actually the effective hydrodynamic radius $R_h$ of the molecule with the hydration shell. The overall correlation time estimated

from NMR relaxation measurements can thus be compared with the effective molecular size and sample viscosity that are measured from PFG-NMR diffusion experiments.

### Pulse field gradient NMR diffusion experiment

The hydrodynamic radius $R_h$ can be calculated from the translational diffusion coefficient $D_{trans}$ of the particles through another Stokes relation:

$$D_{trans} = kT/6\pi\eta R_h \tag{3}$$

PFG-NMR diffusion measurement with the PG-SLED (Pulse Gradient Stimulated Echo Longitudinal Encode-Decode) sequence enables us to obtain $D_{trans}$, which is proportional to the decay rate $d$ of the NMR signal attenuation as a function of gradient strength $g$ (Wilkins *et al.*, 1999):

$$I(g) = I_o \exp[-dg^2] \tag{4}$$

$I_o$ is the NMR peak intensity in the absence of gradient pulses and $g$ is the field strength of the bipolar gradient pulse pair. Changes in the solvent viscosity $\eta$ in different sample environments can be monitored using the methanol signal as an internal standard, assuming that methanol does not interact with other solutes and thus that its hydrodynamic radius $R_h$ is invariable in analogy to the use of dioxan for protein folding studies as described earlier (Wilkins *et al.*, 1999). The ratio of the rate constants of methanol in different conditions gives the relative change in bulk solvent viscosity $\eta$. Knowing this, the relative hydrodynamic radii $R_h$ of mersacidin with respect to methanol in different environments can thus be extracted. In practice, each diffusion data set consists of a series of 40 1D $^1$H spectra with an increment 2.5% of the gradient strength from 2.5 to 100% collected at 750 MHz with a three-axis gradient probe (x-axis for bipolar gradient pulse pair and y- and z-axis for residual signal crushing). Data processing was performed with Felix from Biosym Technologies (San Diego, CA) and Origin7.0 from OriginLab (North Hampton, MA) was used for non-linear fitting to obtain the translational diffusion coefficients $D_{trans}$.

### Structure calculation and analysis

All structure calculations were performed with CNS (Brünger *et al.*, 1998) using the ARIA setup and protocols (Nilges and Donoghue, 1998). Semi-automated NOE assignment was used to assist the spectral assignment (Linge *et al.*, 2001). This was done from a partially assigned NMRView peaklist. The initially unassigned cross-peaks were defined as ambiguous distance restraints with a lower weighing factor. The calibration of the cross-peak intensities against distances was done automatically at the beginning of each iteration. The additional unambiguously assigned cross-peaks were interactively re-examined with NMRView and the checked peaklist was then used as the input for the next calculation in an iterating way until all cross-peaks were assigned. Each semi-automated assignment step with ARIA consisted of eight iterations with successive reduction of the violation tolerance and a final refinement in explicit solvent using default ARIA parameters unless otherwise stated. The Parallhdg5.3 force field with the PROLSQ parameters was used (Linge *et al.*, 2002). The topologies of dehydroalanine (Dha), aminobutyric acid (D-Abu), 3-methyllanthionine (MeLan) and the cyclized C-terminus of the S-aminovinyl-methyl-cysteine (Tea) were constructed based on alanine, threonine, cysteine and comparison of available databases. Four thioether bridges were introduced. Nine backbone $\phi$ and four side chain $\chi_1$ torsion angle restraints obtained from the stereospecific assignments of the methylene groups of the thioester-linked Ala12 and

Ala18 and of Leu14, and Glu17, were used in the structure calculations. A torsion angle dynamics (TAD) simulated annealing (SA) protocol was performed, initially at 10000 K (8000 steps), followed by a first cooling stage to 50 K (50K/step); Cartesian space refinement was used for the second stage cooling from 2000K to 1000 K in 16000 steps and the subsequent third cooling stage from 1000 K to 50 K in 4000 steps followed by 200 steps of energy minimisation. The slow cooling process at the second stage ensures a better convergence of the calculated structures. 50 structures with the lowest restraint energy were further subjected to explicit solvent refinement (OPLS water and DMSO models) as described earlier (Bonvin *et al.*, 2001), and 20 of these were kept for clustering and structural analysis. A cluster is defined as a group of at least four structures with pairwise backbone (N, Cα and C' of residues 3-4,12-20) positional root mean square deviations (RMSD) lower than 0.3 Å. Structures were visualized and analyzed with MOLMOL (Koradi *et al.*, 1996). DynDom was used to identify conformational changes and to define domains and effective hinge regions of the structures obtained under the three different sample conditions (Feenstra *et al.*, 1999). The coordinates of the three structure ensembles, *MeOH/H₂O*, *DPC*$_{bound}$ and *Lipid II*, were deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) under accession codes 1MQX, 1MQY, and 1MQZ, respectively.

## Results

### *Substantial chemical shift changes indicate a strong dependency of the mersacidin conformation on its environment*

The assignment of the resonances of mersacidin in various sample environments was achieved by a standard multidimensional NMR protocol using the ¹⁵N labelled sample (Cavanagh *et al.*, 1996). A two-step ¹⁵N-¹H HSQC titration of perdeuterated DPC micelles followed by unlabelled lipid II was then used to investigate the interaction of mersacidin and lipid II. The chemical shifts are excellent probes for biomolecular interaction studies because of their sensitivity to the changes of surrounding chemical environments. They are commonly used for the mapping of ligand binding site in proteins (Zuiderweg, 2002) and yet, as will be discussed later, the perturbations upon ligand binding can be so large that structural rearrangements could govern the chemical shift changes. The first step was used as a control to monitor how mersacidin was influenced by the membrane mimic environment. The titration experiments showed significant chemical shift perturbations of mersacidin in both titration steps (Figure 2). The addition of DPC micelles strongly affects the backbone amide protons of Gly7, Abu13, Abu15 and Glu17 in the ¹H dimension, mainly upfield shifts, and most of the C-terminal part in the ¹⁵N dimension (Figure 3). Along the ¹⁵N dimension, a downfield shift occurs at Dha16, which is flanked by progressively increased, almost symmetric, upfield shifts. Subsequent addition of lipid II gives rise to large downfield shifts for the amide protons of Gly7, Abu13 and Glu17 in the ¹H dimension; a similar chemical shift perturbation profile is observed in the ¹⁵N dimensions. The general direction of changes is, however, inverted compared to the first DPC micelles titration step. In both steps there are clearly localised structural influences. Unlike the strong binding affinity between nisin and lipid II, resulting in slow exchange cross–peak patterns in similar titration experiments (Hsu *et al.*, 2002), mersacidin follows a fast exchange profile indicating a weaker binding under these conditions. The gradual changes in position can easily be followed during the course of both titration processes (Figure 2). Unexpectedly, the addition of DPC micelles seems to affect mersacidin more than what its specific target, lipid II, does. The titration data have suggested an
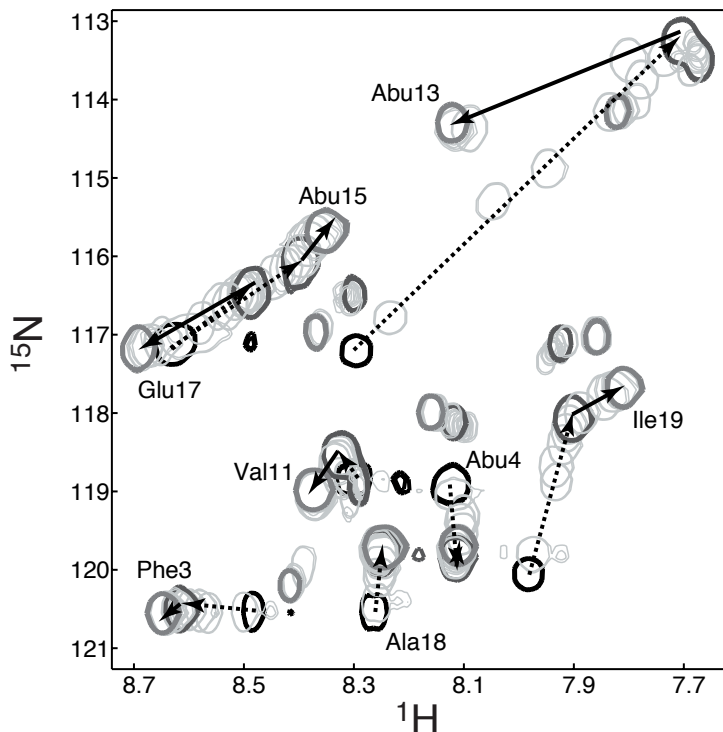
**Figure 2.** Representative region of the two-step $^{15}$N-$^1$H HSQC titration experiments of DPC micelles and lipid II at 293 K, 500 MHz. Starting point, *MeOH/H$_2$O*: black; the end point after DPC titration, *DPC$_{bound}$* and the second end point after lipid II titration, *lipid II* in dark grey. Directions of chemical shift changes are shown in dashed (*MeOH/ H$_2$O → DPC$_{bound}$*) and solid lines (*DPC$_{bound}$* → *lipid II*). The intermediate titration steps (5, 25, 50, 75, 100, 125% of mersacidin concentration) are shown in light grey for DPC micelles and lipid II

interesting interconversion process. For Gly7, Gly8, Ala12, Abu13 and Glu17, the $^1$H chemical shifts first move upfield and then conversely shift downfield. A similar behaviour is monitored along the $^{15}$N dimension for most residues, *e.g.* Glu17. Since chemical shifts are closely related to the surrounding chemical environment primarily defined by the structure, the direction and displacement of the cross-peaks suggest that the structure of mersacidin undergoes a substantial change upon the addition of DPC micelles and is somehow restored close to its initial conformation, if not identical, upon the subsequent addition of lipid II. Yet, it could also be an indication that mersacidin falls off the DPC micelles after the addition of lipid II and restores a state similar to the initial condition.

### Binding-induced overall backbone relaxation dynamics changes

To confirm that mersacidin stays bound to the DPC micelles upon titration of lipid II, we performed $^{15}$N relaxation experiments in conjunction with diffusion experiments to examine the dynamics and molecular size of the complexes. NMR $^{15}$N relaxation experiments have been extensively applied to proteins to study the dynamics of protein motions within a wide range of time scales (Palmer III, 1997;
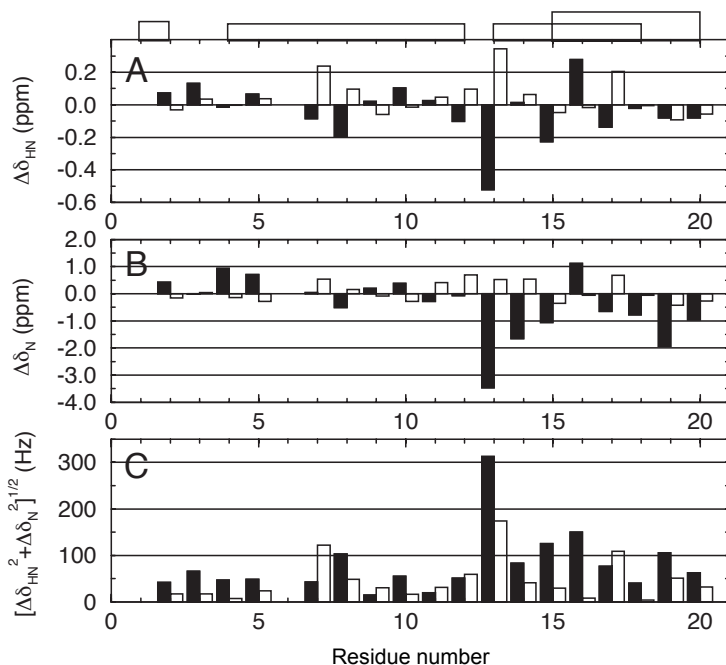
**Figure 3.** Chemical shift perturbation as a function of the primary sequence of mersacidin upon successive addition of DPC micelles (filled bars) and lipid II (open bars). **A.** $^1$H chemical shift perturbation ($\Delta\delta_{HN}$). **B.** $^{15}$N chemical shift perturbation ($\Delta\delta_N$). **C.** Absolute displacements $[(\Delta\delta_{HN})^2+(\Delta\delta_N)^2]^{1/2}$ in Hz from HSQC spectra collected at 500 MHz. The lanthionine linkages are depicted in lines on the top panel.

Kay, 1998; Palmer III, 2001). Changes in protein structure dynamics upon ligand binding can be extracted from the relative differences before and after binding. For example, the rigidification of a flexible loop upon ligand binding can be monitored when $T_2$ is shortened and $^{15}$N{$^1$H} NOE is increased. It should be noted that the relaxation parameters are also closely associated to the molecular weight. In the case of binding interactions, the increased molecular weight of the overall complex would predominantly affect the relaxation parameters of each component globally. This effect can be particularly pronounced when a small peptide binds to large micelles (Bader *et al.*, 2001). In our case, the molecular weight of DPC micelles is roughly ten times higher than that of mersacidin and lipid II. During the titration process, mersacidin binds to DPC micelles and lipid II and hence its apparent molecular weight is considerably increased. From the result of the relaxation dynamics analysis in three different states, namely *MeOH/ H$_2$O*, *DPC$_{bound}$* and *lipid II*, the plateau of the $^{15}$N relaxation parameters indeed uniformly increases for the $^{15}$N{$^1$H} heteronuclear NOEs and longitudinal relaxation times $T_1$ and decreases for the transverse relaxation times $T_2$ as a result of large complex formation (Figure 4). These relaxation data also reflect the dynamics of the local structures. The similar sequential patterns observed in the various states indicate no particular stabilization of local structure upon binding; there still exists local motions. In all states, the glycine-rich ring exhibits higher flexibility as indicated by the lower $^{15}$N{$^1$H} heteronuclear NOEs and longer $T_2$ with respect to the average values. In contrast, the residues that are linked by lanthionine linkages, Abu4, Ala12, Abu13, Abu15, Ala18 and Tea20, exhibit higher $^{15}$N{$^1$H} heteronuclear NOE and
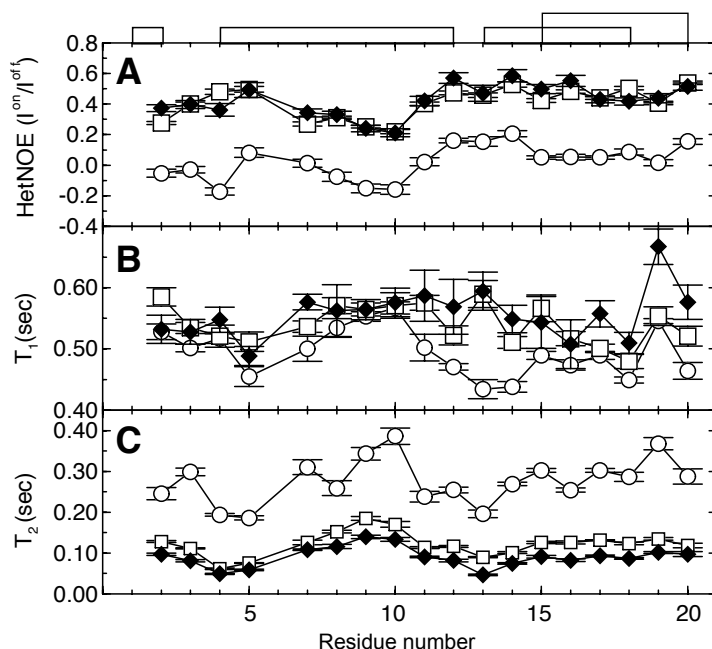
**Figure 4.** $^{15}$N relaxation dynamics parameters of mersacidin. **A.** $^{15}$N{$^{1}$H} heteronuclear NOE. **B.** Longitudinal relaxation time $T_1$. **C.** Transverse relaxation time $T_2$. Open circles (O): *MeOH/H$_2$O*; open squares (□): *DPC$_{bound}$*; filled diamonds (◆): *lipid II*. The standard deviations are shown in bars. The lanthionine linkages are depicted on the top panel.

shorter $T_2$ (Figure 4). Overall, the most stable structure elements are the segments that are connected by the glycine-rich ring (residues 4–5 and 12–14), their backbone motion being obviously restricted by their ring structures.

The increase of effective molecular size and bulk solvent viscosity can both contribute to the global change in relaxation parameters (Wand *et al.*, 1998). Therefore, PFG-NMR diffusion experiments were used to discriminate between those. In such experiments, the change of solution viscosity can be monitored independently by using a separate internal probe, in our case, methanol (see Material and Methods) (Wilkins *et al.*, 1999). The addition of DPC micelles and lipid II increased the solvent viscosity by 7.4% and 6.7%, respectively. This is significantly less than the 22.9% and 26.1% increases of the translational diffusion coefficient ratio of mersacidin and methanol upon addition of DPC micelles and subsequently lipid II, respectively (Table 1). This can only be explained by an increase of the effective molecular size due to the binding to DPC micelles and lipid II. A good correlation is obtained between the overall tumbling rates $\tau_c$ estimated from $T_2$ relaxation and from diffusion experiments ($R^2 = 0.86$). This confirms that mersacidin binds DPC micelles and stays bound when lipid II is present, which is at the origin of the observed global effect on the relaxation parameters due to the decrease of the overall tumbling rate.

**Table 1.** Translational diffusion parameters obtained from PFG-NMR experiments

| Decay rate constant[a] | $MeOH/H_2O$ | $DPC_{bound}$ | Lipid II |
|---|---|---|---|
| $d^{methanol}$ | 15.04 | 14.00 | 14.09 |
| $d^{mersacidin}$ | 2.04 | 1.55 | 1.52 |
| Relative increase of bulk solvent viscosity (%)[b] | | | |
| $(\eta^i/\eta^{MeOH/H2O}) - 1$ | 0 | 7.4 | 6.7 |
| Relative hydrodynamic radius | | | |
| $R_h^{mersacidin}/R_h^{methanol}$ | 7.35 | 9.03 | 9.27 |

a.  The decay rate constant $d$ is obtained from Gaussian fitting of the attenuated NMR signals as a function of the fractional gradient strength $g$ using Equation 4.
b.  The increase of bulk solvent viscosity $\eta$ can be derived from the observed decay rate constants for methanol and is inversely proportional to their ratios (see Material and Methods).

### Solution NMR structures reveal dramatic conformational changes

Figure 5 summarizes the observed proton-proton contacts of mersacidin derived from NOESY spectra recorded under different sample conditions. The NOE contact maps show similar patterns in the three different conditions with only a few variations, suggesting that the topology of mersacidin does not alter much upon changes in the environment. Some long range NOEs are absent in $MeOH/H_2O$, *e.g.* between Phe3 and Glu17, and some variations occur in the glycine-rich ring under all three conditions. These small differences in NOE contacts reflect, however, remarkable conformational changes; neglecting some of those for structure calculations can substantially alter the relative orientation of the connected rings. Figure 6 shows the ensemble of structures obtained from structure calculation based on the NOEs derived from the NOESY spectra recorded in different environments. The intertwined ring structure
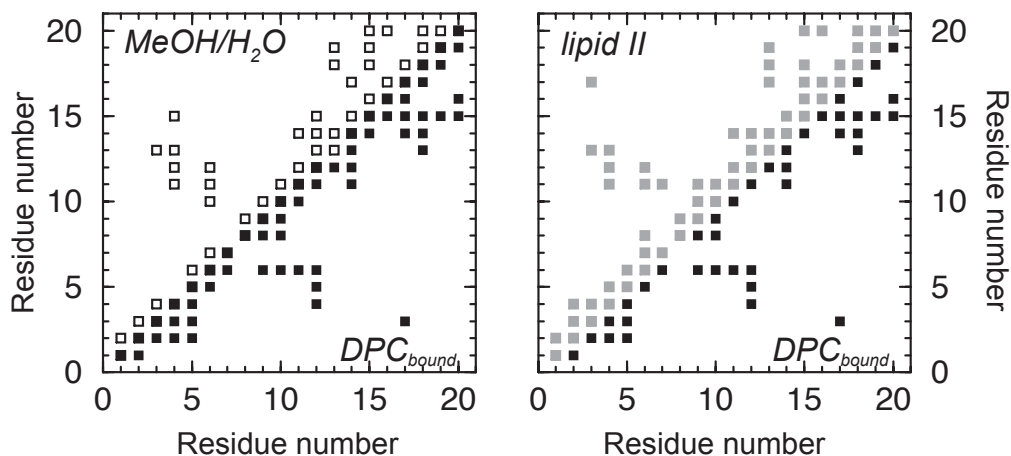


**Figure 5.** NOE contact maps of mersacidin in different sample conditions. Each square corresponds to at least one observed NOE between the two corresponding residues in the $MeOH/H_2O$ mixture (upper panel in the left), in lipid II – containing DPC micelles (upper panel in the right) and in DPC micelles only (both lower panels).

**MeOH/H₂O**       **DPCbound**       **L2a**       **L2b**
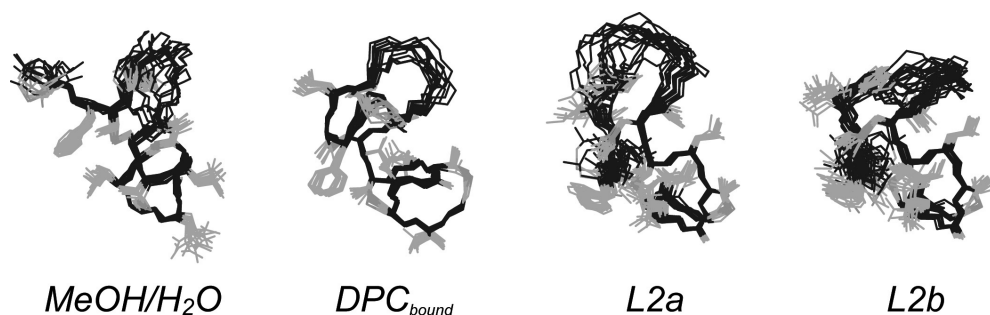
**Figure 6.** Solution NMR structures of mersacidin in various sample conditions. Only heavy atoms are shown for clarity. Backbone and side chains are shown in black and grey, respectively. *MeOH/H₂O*: free mersacidin in methanol/water mixture; *DPCbound*: DPC micelle bound form; *L2a* and *L2b*: two conformations in the lipid II bound form with the only difference in the glycine-rich loop (see text for detail). The structures were fitted on backbone heavy atoms within each ensemble and the various ensembles were aligned using the backbone heavy atoms of residues 13-20.

provides a well-defined backbone structure in the C-terminal part (residue 13-20) in all conditions although *MeOH/H₂O* deviates slightly from *DPCbound* and *lipid II*. The N-terminus and the glycine-rich ring are less defined due to the fact that the glycine repeat is prone to low NOE density as well as local flexibility. Two populations of the glycine-rich ring are found after clustering for the lipid II bound form (named *L2a* and *L2b*). These two clusters have identical backbone conformations except for the glycine residues: two distinct ring conformations are observed with their planes flipped by almost 180° with respect to each other. Since they both satisfy all distance restraints, they will be both kept for the following structural comparisons.

Amongst all three conditions, the DPC micelles bound form (*DPCbound*) is the best defined in terms of RMSD of the ensemble and the most favourable in terms of total energy after explicit solvent refinement (Table 2). The structures contain no conventional secondary structure element because of their unusual post-translationally modified amino acids. In order to evaluate the effect of solvent hydrophobicity onto the structure – methanol and DPC are both hydrophobic and this can affect especially the side chain packing – two types of solvent models (water: hydrophilic; DMSO: hydrophobic) were used for structure refinement. The results show no distinguishable difference both in structure and in energy (data not shown). In both cases, the electrostatic energy term is dominant and the *DPCbound* structures have the lowest electrostatic energies. The main structural difference occurs in the relative orientation of the glycine-rich ring with respect to the intertwined rings. The addition of DPC micelles induces a large-amplitude twisting of 162° of those rings with Abu13 and Leu14 as the effective hinge (Figure 7). The subsequent addition of lipid II in the presence of DPC micelles induces then an inversed twist (100° for *L2a* and 167° for *L2b*) that results in a conformation close to the initial state. The hinge region, centred at Abu13, experiences a dramatic change in its chemical environment that is reflected in the chemical shift changes during the course of the $^{15}$N–$^1$H HSQC titrations. The same conformational change directly affects the electrostatic energy since the distance between the two sole charged groups (N-terminus and Glu17 side chain) is modulated by this twisting motion. This becomes clear when looking at the surface electrostatic potential (Figure 8). *DPCbound* highly resembles the structure that was previously solved in the

**Table 2.** Structural statistics of mersacidin in different sample environments

| | Environments[a] | | Lipid II | |
|---|---|---|---|---|
| | MeOH/$H_2O$ | DPC$_{bound}$ | L2a | L2b |
| Number of structures[b] | 12 | 13 | 9 | 8 |
| *RMSD (Å) with respect to the average structure of each ensemble (backbone/heavy)* | | | | |
| Residue 3–4, 12–20 | 0.27±0.08/0.59±0.06 | 0.23±0.04/0.44±0.06 | 0.41±0.11/0.69±0.12 | 0.44±0.14/0.72±0.16 |
| All | 0.90±0.24/0.99±0.15 | 0.58±0.15/0.64±0.11 | 0.72±0.11/0.89±0.11 | 0.79±0.35/0.96±0.27 |
| *Number of experimental restraints for structural calculation* | | | | |
| Total NOEs | 200 | 248 | 241 | |
| Intraresidue NOEs | 101 | 97 | 111 | |
| Sequential NOEs | 59 | 102 | 86 | |
| Medium range NOEs | 10 | 16 | 13 | |
| Long range NOEs | 30 | 33 | 31 | |
| Dihedral restraints[c] | 13 | 13 | 13 | |
| *Restraint statistic[d]* | | | | |
| NOE RMSD ($10^{-2}$ Å) | 3.30±0.28 | 2.39±0.75 | 3.67±0.31 | 3.76±0.24 |
| Dihedrals RMSD (°) | 1.28±0.19 | 0.71±0.39 | 0.69±0.21 | 0.91±0.30 |
| *RMSD from idealized covalent geometry* | | | | |
| Bonds (Å) | 0.0045±0.0002 | 0.0036±0.0002 | 0.0046±0.0003 | 0.0051±0.0002 |
| Angles (°) | 0.78±0.04 | 0.56±0.02 | 0.74±0.03 | 0.78±0.04 |
| Impropers (°) | 2.1±0.1 | 1.3±0.1 | 2.0±0.2 | 2.2±0.2 |
| Dihedrals (°) | 14.3±0.8 | 13.2±0.8 | 14.5±0.8 | 14.7±0.7 |
| *CNS energies after water refinement (kcal/mol)* | | | | |
| $E_{total}$ | –386±27 | –638±15 | –434±25 | –407±20 |
| $E_{vdw}$ | –83±19 | –96±78 | –94±10 | –96±4 |
| $E_{elec}$ | –483±13 | –642±10 | –521±11 | –514±7 |

**a.** MeOH/$H_2O$: free mersacidin in methanol/water mixture solution; DPC$_{bound}$: mersacidin in DPC micelles; L2a and L2b: two distinct conformations in lipid II bound form obtained with the same structural restraints.

**b.** The number of structures in each condition is obtained from clustering analysis (see Material and Method).

**c.** Nine backbone $\phi$ angles and four side chain $\chi_1$ angles (Ala12, Leu14, Glu17 and Ala18).

**d.** No structure had violation larger than 0.5 Å for NOE restraints and 5° for dihedral angle restraints.

crystalline state in pure methanol. The N–terminus and Glu17 side chain in *MeOH/$H_2O$* point towards two opposite sides while the two charged groups come close to each other as a result of the twist induced by DPC micelles, which obviously lowers the electrostatic energy (Table 2).

Comparison of the structures of mersacidin with and without lipid II in DPC micelles has revealed intriguing conformational changes. While the hydrophobicity of DPC micelles induces the closure of
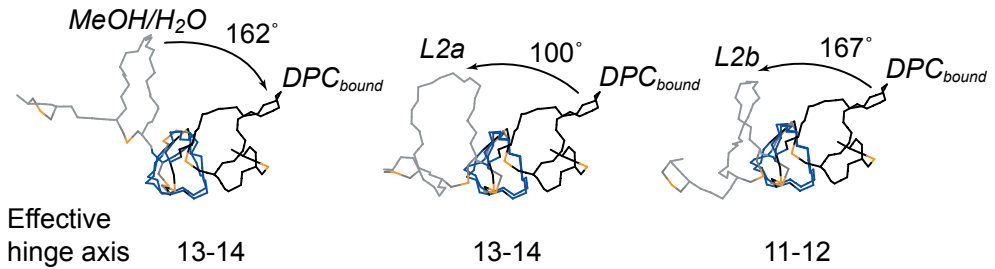
**Figure 7.** Conformational change analysis using the program DynDom (Feenstra *et al.*, 1999). The first two rings are coloured in black for *DPC_bound* and in grey for *MeOH/H₂O*, *L2a* and *L2b*. The rotation of the fragments with respect to the hinge is indicated by an arrow. The structures were fitted on backbone heavy atoms of residues 13–20.
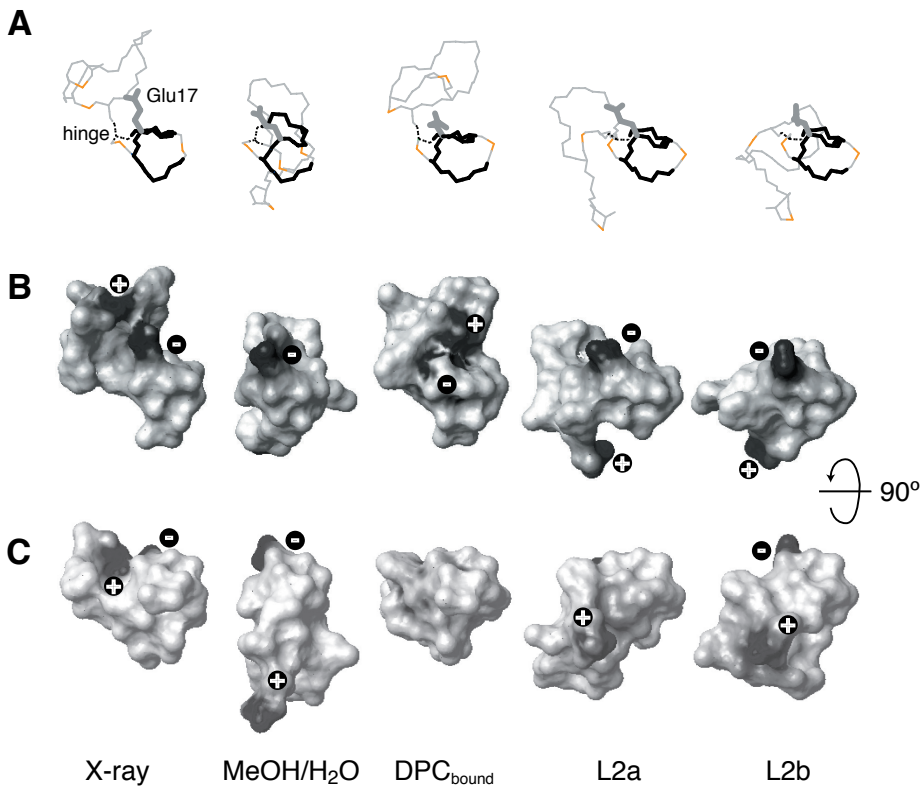


**Figure 8.** Conformational changes of mersacidin in response to changes in environments. **A.** Representative structures (closest to average) of each structure ensemble. The Glu17 side chain is shown in thick grey lines. The hinge residue Abu13 is shown in dashed lines. The structures were fitted on backbone heavy atoms of residues 13-20 (thick black lines). The remaining of the structure is shown in grey. **B.** Surface electrostatic potential of mersacidin in different sample conditions calculated with MOLMOL (Koradi *et al.*, 1996). Positive and negative potentials are indicated in dark grey with labels indicating corresponding signs. The structures are in the same orientation as in **A**. **C.** 90° rotation along the x-axis of the above structures. The charge distributions reveal the increase of charge accessibility after the addition of lipid II in the hydrophobic DPC micelles solution. The structure of mersacidin in DPC micelles resembles the X-ray structure that was solved in pure methanol.

its structure, in contrast, the addition of lipid II results in the exposure of the charged groups (Figure 8). The exposure of the Glu17 side chain can be quantitatively monitored from the change in its solvent accessible surface (SAS) area. The addition of DPC micelles reduces the SAS area from 155 Å$^2$ (*MeOH/ H$_2$O*) to approximately 55 Å$^2$ (*DPC$_{bound}$*) and the addition of lipid II increases again the SAS area to about the same value of 155 Å$^2$ (*L2a* and *L2b*). The positively charged N-terminus, Ala1, undergoes similar changes in SAS area with almost identical values (data not shown). Another significant change occurs at Ile19 with a 60% increase in SAS area upon the addition of DPC micelles (from 100 Å$^2$ to 160 Å$^2$) and a drop to 80 Å$^2$ when lipid II is present. In addition to the dramatic changes in side chain packing and charge distribution, the backbone structure is also affected when DPC micelles are added to free mersacidin. Notable differences in backbone torsion angles for Abu2, Phe3, Abu4, Ala12, Abu13, Leu14 and Abu15 are monitored between the free (*MeOH/H$_2$O*) and the DPC micelle bound form (*DPC$_{bound}$*) while subsequent addition of lipid II (*L2a* and *L2b*) does not induce much more changes (data not shown). These differences are consistent with the observed $^{15}$N chemical shift changes, which are indicative for backbone conformational changes. Large changes occur in the N- and C-terminal parts when DPC micelles are added and much smaller changes are found when lipid II is added (Figure 3). Even though the opening of the structure increases the electrostatic energy of mersacidin *per se*, the binding interface and in particular the interaction with its counter part, lipid II, must compensate the energetic cost of exposing the two charges (Table 2). This suggests that the binding mechanism is based on electrostatic interactions.

## Discussion

We have shown that the structure of mersacidin can be substantially modulated by sample conditions. The structures of type B lantibiotics have long been thought to be compact and rigid given the lanthionine ring linkages and their relatively small size in length (Brötz and Sahl, 2000). While the intrinsic flexible nature of the glycine-rich ring gives rise to a higher variability in structure as monitored by $^{15}$N relaxation experiments (Figure 4), which was even observed in the crystal structure (Schneider *et al.*, 2000), mersacidin was, to our surprise, found to possess a minute hinge region (Ala12 and Abu13) that can effectively alter the overall backbone geometry by only localised backbone torsion angle changes. Structures obtained from various sample conditions show distinct backbone torsion angle distributions in this region. Furthermore, our results have demonstrated that the chemical shift changes of mersacidin during the course of titration experiments were basically governed by conformational changes, which is rather unusual in the case of protein-protein or protein-ligand interaction studies. This is one of the few examples where large chemical shift perturbations primarily originate from conformational changes, which makes the conventional binding interface mapping approaches inapplicable (Zuiderweg, 2002). Multi-step titrations and structure determination have revealed the flexible nature of mersacidin and its adaptation to the environment in terms of its three dimensional structure. Upon binding to lipid II, the structural changes directly affect the exposure of the charge groups suggesting that electrostatic interactions govern the binding mechanism despite the rather hydrophobic nature of mersacidin.

As discussed above, changes in the charge distribution of mersacidin likely play a crucial role in the mersacidin-lipid II interaction. Due to the lack of intermolecular NOEs, there is no sufficient evidence to assess which charge group is mainly responsible for the binding thus far. However, derivatisation studies

have shown that the methylation of the N-terminus increases the minimal inhibitory concentration (MIC) of mersacidin by two fold, and biotinylation makes it inactive. The side chain of Glu17 has been proposed to be the lipid II binding motif of mersacidin and a replacement of Glu17 by Ala in mutagenesis studies makes it inactive. Inactivation is also achieved by amidation of the carboxyl group of Glu17*. These results strongly indicate that the binding requires these charges. Furthermore, the bioactivity is also ion dependent: The presence of calcium ions can enhance the activity *in vivo* whereas magnesium ions do not have any effect (not shown). Since lipid II is mostly negatively charged (Figure 1B), it is likely that the calcium ion is needed to bridge the Glu17 side chain to other negatively charged groups of lipid II, unless it forms a direct salt bridge with the positively charged side chain of Lys3 of lipid II. Should the calcium ion indeed provide a bridge between mersacidin and lipid II, this would explain the lack of intermolecular NOEs (even with long NOESY mixing times up to 500 ms) due to the increased distance between the two, which would exceed a detectable NOE range (>5 Å). Although mersacidin is strongly affected during the course of the titration, little chemical shift changes were observed for lipid II, except for the amide proton of γ-Glu2 (data not shown), suggesting that the non-exchangeable protons of lipid II (Figure 1B) may not be directly involved in binding whereas the abundant exchangeable hydroxyl protons and/or the pyrophosphate group could play a more important role. It has also been shown that mersacidin can discriminate lipid I from lipid II, indicating that GlcNAc, the only difference between these two, contains part of the its recognition motif (Brötz *et al.*, 1998a).

Proteins in hydrophobic solvent are thought to retain their native structure as a result of kinetic trapping, due to stronger intramolecular hydrogen bonding and a more rigid structure in the absence of water (Mattos and Ringe, 2001). For small peptides like mersacidin, however, their structures can vary remarkably with different environments as have been demonstrated in early studies by Kessler and coworkers (Kessler *et al.*, 1992; Kessler *et al.*, 1993). A compact structure of mersacidin was observed in the DPC micelle bound form ($DPC_{bound}$; our study), in pure methanol solution (Prasch *et al.*, 1997) and in the crystalline state (Schneider *et al.*, 2000) due to the environmental hydrophobicity that is known to facilitate the formation of the intramolecular hydrogen bonds. In the X-ray structure, two hydrogen bonds were identified: from the N-terminus and the Gly7 backbone amide to the Glu17 side chain (Schneider *et al.*, 2000). This is consistent with our finding that the two oppositely charged groups are directed towards each other only in a highly hydrophobic environment in the absence of lipid II. In contrast, in the presence of lipid II, the bound form is clearly distinct from the free form structure as has been shown in this work. Recently, the significance of flexibility for bioactivity has been highlighted in the case of the nisin–lipid II interaction (Breukink *et al.*, 1999): a truncation or rigidification of the flexible hinge region substantially abolishes its bioactivity (Wiedemann *et al.*, 2001). Here, mersacidin uses its hinge to adjust the exposure of charges when it binds to its "docking molecule", lipid II. We have thus shown in the case of mersacidin that conformational versatility provides a mean to adapt to the surrounding environments. It is therefore important, when using static structures of such peptides for structure–based peptide engineering, that the sample conditions be carefully considered.

---

* W. Dückheimer (Hoechst AG,) personal communication. This is later published in Szekat, C., Jack, R.W., Skutlarek, D., Farber, H. and Bierbaum, G. (2003) *Appl. Environ. Microbiol.*, **69**, 3777-3783.

Many enzymatic functions of proteins require conformational changes during substrate binding and reaction process. The same is true for the mersacidin–lipid II interaction, as observed when mersacidin binds to DPC micelles and lipid II. Similar structural changes have been reported by McCafferty and co-workers for ramoplanin (Cudic *et al.*, 2002), a lipoglycodepsipeptide antibiotic, that undergoes a flattening of its structure when it binds to a lipid II analog, UDP-MurNAc-peptapeptide. Ramoplanin is known to be involved in the inhibition of transglycosylation during the biosynthesis of lipid II. It does compete, although not efficiently, with mersacidin for the binding to lipid II (Brötz *et al.*, 1998b). Their functionalities are similar in the sense that both peptides recognise specifically parts of lipid II in order to block cell wall biosynthesis. Notably, ramoplanin also shares a topological similarity in structure with mersacidin and actagardine, another type B lantibiotic, although its has no sequence homology with these two lantibiotics (Zimmermann and Jung, 1997; Cudic *et al.*, 2002). From this point of view, conformational changes are evidently a common feature of the lipid II-binding peptides. The structural insight gained here can, therefore, be extended to the type B lantibiotics actagardine (Zimmermann and Jung, 1997) and Ala(0)-actagardine (Vertesy *et al.*, 1999) since they possess a high degree of sequence homology (the hinge of Ala12 and Abu13 is also present in actagardine) (Zimmermann and Jung, 1997; Brötz and Sahl, 2000), and probably to other lipid II binding peptides, like ramoplanin, as well.

## Acknowledgements

# Atomic insight into the CD4 binding-induced conformational changes in HIV-1 gp120[#]

Shang-Te D. Hsu and Alexandre M.J.J. Bonvin

*Bijvoet Center for Biomolecular Research, Utrecht University,
3584CH Utrecht, The Netherlands*

# Abstract

**The entry of HIV-1 into a target cell requires gp120 and receptor CD4 as well as co-receptor CCR5/CXCR4 recognition events associated with conformational changes of the involved proteins. The binding of CD4 to gp120 is the initiation step of the whole process involving structural rearrangements that are crucial for subsequent pathways. Despite the wealth of knowledge about the gp120/CD4 interactions, details of the conformational changes occurring at this stage remain elusive. We have performed molecular dynamics simulations in explicit solvent based on the gp120/CD4/CD4i crystal structure in conjunction with modeled V3 and V4 loops to gain insight into the dynamics of the binding process. Three differentiated interaction modes between CD4 and gp120 were found, which involve electrostatics, hydrogen bond and van der Waals networks. A "binding funnel" model is proposed based on the dynamical nature of the binding interface together with a CD4-attraction gradient centered in gp120 at the CD4-Phe43-binding cavity. Distinct dynamical behaviors of free and CD4-bound gp120 were monitored, which likely represent the ground and pre-fusogenic states, respectively. The transition between these states revealed concerted motions in gp120 leading to: *i*) loop contractions around the CD4-Phe43-insertion cavity; *ii*) stabilization of the four-stranded "bridging sheet" structure; and *iii*) translocation and clustering of the V3 loop and the bridging sheet leading to the formation of the co-receptor binding site. Our results provide new insight into the dynamic of the underlying molecular recognition mechanism that complements the biochemical and structural studies.**

# Introduction

The entry of Human Immunodeficiency Virus type-1 (HIV-1) into a target cell is an extremely intricate process (for reviews, see (Chan and Kim, 1998; Wyatt and Sodroski, 1998; Berger *et al.*, 1999; Poignard *et al.*, 2001)), which requires the recognition of the host cell receptor CD4 and co-receptors, mainly CCR5 and CXCR4, by the viral envelope glycoprotein (Env), consisting of gp120 and gp41. Env is organised into trimeric spike structures on the surface of the virion (Kwong *et al.*, 2000b) anchored in the membrane by the transmembrane subunit gp41 while gp120 is presented at the surface as the frontier for the recognition process. The initiation step involves binding of CD4 to gp120, which induces conformational changes in gp120 (Sattentau and Moore, 1991; Thali *et al.*, 1993; Moore and Binley, 1998) leading to the exposure and/or formation of antigenic epitopes recognised by the co-receptors (Trkola *et al.*, 1996; Wu *et al.*, 1996). Binding to the co-receptors takes place subsequently and gp41 undergoes conformational changes that mediate the fusion process by formation of the fusogenic structure bringing viral and target cell membranes into close vicinity. This complex mechanism involves a series of structural rearrangements in which conformational dynamics plays a crucial role.

Over the past decades, a large number of crucial elements involved in this system, in particular gp120, have been unraveled using a variety of approaches (Rizzuto *et al.*, 1998; Wyatt and Sodroski, 1998). The recent determination of the crystallographic structure of the core of gp120 in complex with CD4 and

the CD4-induced antibody (CD4i), which recognises the co-receptor binding site, provided tremendous structural insights (Kwong *et al.*, 1998; Kwong *et al.*, 2000a). Despite the highly variable sequence of HIV-1 gp120 throughout evolution, the topology of the core has been retained with some degree of sequence diversity (Kwong *et al.*, 1998; Kwong *et al.*, 2000a). It contains several disulphide bridges that buckle the flexible hypervariable loops to form knots at the periphery of the rigid core and basically define their boundaries. The heavy glycosylation at the "silent face" and the mobile variable loops together provide a shielding umbrella to evade immune system neutralization (Wyatt and Sodroski, 1998). CD4 is bound to gp120 predominantly *via* electrostatic interactions with a large, but mismatched interface. An unusually protuberant Phe43 of CD4 inserts into a receptive hole of gp120 (Kwong *et al.*, 1998) making up a "knob-and-socket" interaction (Moore and Binley, 1998). Evidence from a thermodynamics study has suggested that the binding of CD4 to gp120 induces substantial structural rearrangements primarily in the core structure, a truncated form of gp120 that was used in the crystallographic studies where the flexible N- and C-termini as well the variable loops, V1-V3, are absent (Myszka *et al.*, 2000) (see Figure 1). It is now generally accepted that the gp120/CD4 complex formation leads to the accessibility of the co-receptor binding site. The third variable (V3) loop, in particular, plays a central role here since it contains not only the principle neutralising determinant (PND) (Rusche *et al.*, 1988) but also the main co-receptor determinant (CXCR4 versus CCR5 usage) for HIV-1 at this stage (Berger *et al.*, 1999). X-ray crystallographic and Nuclear Magnetic Resonance (NMR) studies of various types of V3 loop fragments have shown a high degree of structural heterogeneity (Chandrasekhar *et al.*, 1991; Rini *et al.*,

```
                  β1              α1                          β2      V1/V2 stem   β3
LAI    83 EVVLVNVTEN FNMWKNDMVE QMHEDIISLW DQSLKPCVKL TPLCV GAG SC NTSVITQACP 206
HxBc2  83 EVVLVNVTEN FNMWKNDMVE QMHEDIISLW DQSLKPCVKL TPLCV GAG SC NTSVITQACP 206

             β4         β5        ⌐A   β6    β7    β8      ⌐B     β9
LAI   207 KVSFEPIPIH YCAPAGFAIL KCNNKTFNGT GPCTNVSTVQ CTHGIRPVVS TQLLLNGSLA 266
HxBc2 207 KVSFEPIPIH YCAPAGFAIL KCNNKTFNGT GPCTNVSTVQ CTHGIRPVVS TQLLLNGSLA 266

          ⌐C    β10   ⌐D        β11        β12                      V3
LAI   267 EEEVVIRSAN FTDNAKTIIV QLNQSVEINC TRPNNNTRKS IRIQRGPGRA FVTIGKIGNM 326
HxBc2 267 EEEVVIRSVN FTDNAKTIIV QLNTSVEINC T--------- -----GAG-- ---------- 300

            β13        α2              ⌐E     β14       β15    α3   β16      β17
LAI   327 RQAHCNISRA KWNATLKQIA SKLREQFGNN KTIIFKQSSG GDPEIVTHSF NCGGEFFYCN 386
HxBc2 330 ---HCNISRA KWNNTLKQIA SKLREQFGNN KTIIFKQSSG GDPEIVTHSF NCGGEFFYCN 386

           α4  β18         V4            β19       β20    β21    ⌐F      β22
LAI   387 STQLFNSTWF NSTWSTEGSN NTEGSDTITL PCRIKQFINM WQEVGKAMYA PPISGQIRCS 446
HxBc2 387 STQLFNSTWF N--------- ---GSDTITL PCRIKQIINM WQKVGKAMYA PPISGQIRCS 446

            β23     V5       β24           α5            β25
LAI   447 SNITGLLLTR DGGNNNNGSE IFRPGGGDMR DNWRSELYKY KVVKIE 492
HxBc2 447 SNITGLLLTR DGGNSNNESE IFRPGGGDMR DNWRSELYKY KVVKIE 492
```

**Figure 1.** Primary sequence of the truncated form of LAI strain with the core structure and the V3 loop used in the current MD simulation with the numbering system according to the HXBc2 isolate. The secondary structure elements are numbered according to the X-ray core structure (PDB entry 1G9M, chain G, Kwong *et al.*, 1998) The truncated V3 loop and the disordered V4 loop that are absent in the coordinates of HXBc2 gp120 are indicated by dashes along the sequence of HXBc2. The truncated N- and C-termini are omitted for clarity. Black blocks depict differences in sequence composition of LAI versus HXBc2 that are present in the simulations.

1993; Ghiara *et al.*, 1997; Stanfield *et al.*, 1999; Tugarinov *et al.*, 1999; Tugarinov *et al.*, 2000; Sharon *et al.*, 2003), which may thus provide means for HIV-1 to evade neutralization by the immune system (Stanfield *et al.*, 1999). On one hand, biochemical studies have established the outline of the fusion process revealing its vast diversity and sophistication; on the other hand, results from the structural studies have provided static snapshots of the interaction between gp120 and its interacting partners at atomic resolution. The dynamics involved in this mechanism, however, remains elusive.

We have performed molecular dynamics (MD) simulations to investigate the conformational changes in gp120 induced upon complexation of CD4. By comparing the dynamical properties of gp120 in the free form and bound to CD4, we will show that CD4 binding substantially reduces the dynamical motions of gp120. An extensive intermolecular hydrogen bond network is formed at the interface and around the Phe43 binding cavity. We monitored motions around the CD4 binding site that are correlated to the V3 loop and trimerization interface, which likely represent the initial steps of the subsequent structural rearrangements required for co-receptor binding and gp41 mediated membrane fusion.

## Material and Methods

### *Generation of the starting structures*

The starting structure of gp120 with the truncated LAI strain sequence (SWISS-PROT accession number P03377) was obtained from SWISS-MODEL (Guex and Peitsch, 1997) using structural homologs as template (PDB entries 1G9M (Kwong *et al.*, 1998), 1G9N (Kwong *et al.*, 2000a) and 1GC1 for the gp120 core and 1CE4 for the V3 loop (Vranken *et al.*, 1995)). The particular strain was chosen because of ongoing collaborations with experimental groups (I. Braakman, Utrecht University and B. Berkhout, University of Amsterdam, personal communication). The primary sequence of gp120 of LAI differs from HxBc2 by only six amino acids in the core region (98.1% sequence identity, see Figure 1). Core gp120 is defined as the construct present in the crystal structure with the truncated V1-V3 and N- and C-termini. The V4 loop (residues 398-409), for which electron density is missing in all crystal structures, and the V3 loop, which has been proposed to undergo significant rearrangements upon CD4 binding, were modeled onto the g120 core, together with the six core mutations of the LAI strain, using the default settings of SWISS-MODEL (Guex and Peitsch, 1997). For the V3 loop, a NMR structure (residues 296-331; PDB ID: 1CE4) was taken for docking onto the core structure. This modelling and in particular the six point mutations are expected to have only minor effect on the overall structure of gp120. For example, the crystal structures of two HIV-1 variants, HxBc2 (1G9M) and YU2 (1G9N) with a sequence identity of 85.9%, show backbone (N, Cα and C') root–mean–square deviation (RMSD) of only 0.12 nm. The structural variation in core gp120 between LAI and HxBc2, which share 98.1% sequence identity, should thus be minimal and our starting model for the simulations can be considered accurate. For the simulation of the complex form (*Cplx*), the LAI gp120 model (residues 83–492) was superimposed onto the crystal structure of the complex (1G9M) to obtain the relative orientation with respect to the D1 domain of CD4 (CD4-D1, residues 1-99). Monomeric gp120 (*Free-gp120*) and CD4-D1 (*Free-CD4*) taken from the structure of the complex were used as starting structures in the free simulations for comparison.

The starting conformations of gp120 in the free and complex simulations were thus identical and correspond to the CD4–bound form.

*Molecular dynamics simulation setup and structural analysis*

The GROMACS 3.0 molecular dynamics package (Lindahl *et al.*, 2001) was used with the GROMOS 43A1 force field (Daura *et al.*, 1998). Starting structures of the free gp120 (*Free-gp120*), the free CD4 D1 domain (*Free-CD4*) and the gp120/CD4 D1 domain complex (*Cplx*) were individually solvated using the simple point charge (SPC) water model (Berendsen *et al.*, 1981) in rectangular periodic boxes with a 1.4 nm solute–wall minimum distance. After a first steepest descent energy minimisation with positional restraints on the solute, seven, three and ten chloride anions were introduced in *Free-gp120*, *Free-CD4* and *Cplx*, respectively, to obtain electro–neutralised systems. The resulting systems comprised 89608, 24054 and 95308 atoms for *Free-gp120*, *Free-CD4* and *Cplx*, respectively. A second energy minimisation was then performed, followed by five successive 20 picosecond (ps) MD runs with decreasing positional restraints force constants on the solutes ($K_{posres}$ = 1000, 1000, 100, 10 and 0 kJ mol$^{-1}$ nm$^{-2}$) prior to the 10 nanosecond (ns) production runs. A two femtosecond (fs) time step was used for the integration of the equations of motion. Solute, solvent and counter-ions were independently coupled to a reference temperature bath at 300K with a coupling constant $\tau_T$ of 0.1 ps (Berendsen *et al.*, 1984). The pressure was maintained by weakly coupling the system to an external pressure bath at one atmosphere with a coupling constant $\tau_p$ of 0.5 ps. Non–bonded interactions were calculated using twin range cutoffs of 0.8 and 1.4 nm. Long range electrostatic interactions beyond the cutoff were treated with the generalised reaction field model (Tironi *et al.*, 1995) using a dielectric constant of 54. The non–bonded pair list was updated every 10 steps. The LINCS algorithm (Hess *et al.*, 1997) was used for bond length constraining in conjunction with dummy atoms for the aromatic rings and amino group in side chains (Feenstra *et al.*, 1999). Owing to the large system sizes *ca* 2 ns were required to reach equilibrium. The analysis was performed therefore on the 2-10 ns segments for all systems.

Equilibrated trajectories (2-10 ns) of free and CD4-bound gp120 were merged for essential dynamics analysis (Amadei *et al.*, 1993). After diagonalising the backbone atom (N, C$\alpha$ and C') mass-weighted covariance matrix the primary principle component (the first eigenmode) was extracted and identified as the transition mode between the two states. From a projection of the merged trajectory along this eigenmode the structures corresponding to the two extremes were obtained.

The β–turn conformations in the V3 loop were compared with several experimental structures of this loop (PDB entries: 1CE4 (Vranken *et al.*, 1995), 1B03 (Tugarinov *et al.*, 1999), 1QNZ (Tugarinov *et al.*, 2000), 1AI1 (Ghiara *et al.*, 1997), 1F58 and 2F58 (Stanfield *et al.*, 1999)). For all types of β–turns a minimum distance between the C$\alpha$(i) and C$\alpha$(i+3) of 0.7 nm was required. The criteria for backbone torsion φ/ψ angles for type II β–turn are: φ(i+1) = –60°, ψ(i+1) = 120°, φ(i+2) = 80° and ψ(i+2) = 0° and those for type VIII β–turn were: φ(i+1) = –60°, ψ(i+1) = –30°, φ(i+2) = –120° and ψ(i+2) = 120°, each torsion angle has a ±45° tolerance.

# Results

In order to investigate the conformational changes in gp120 associated with CD4 binding, three molecular dynamics (MD) simulations of 10 ns production were performed in explicit solvent for the following systems: free gp120 (*Free-gp120*), the free D1 domain of CD4 (*Free-CD4*) and the gp120/CD4 D1 domain complex (*Cplx*). All systems showed stable trajectories with conserved secondary and tertiary structures (Table 1). Stability and equilibration of the systems was monitored by following the RMSD values as a function of time (Figure 2). The loop regions in free gp120, namely V1–V5, give rise to larger structural deviations than in the complex form (Figure 2 and Table 1). The larger RMSD values of the

**Table 1.** Structural statistics (2–10 ns)

*Average backbone RMSD[a] with respect to the starting structure (nm)*

|  | Gp120 all bb | Gp120 2° bb | CD4 all bb | CD4 2° bb | Gp120+CD4 all bb | Gp120+CD4 2° bb |
|---|---|---|---|---|---|---|
| Free | 0.42(0.02) | 0.16(0.02) | 0.10(0.02) | 0.08(0.01) |  |  |
| Cplx | 0.37(0.02) | 0.14(0.01) | 0.11(0.01) | 0.09(0.01) | 0.40(0.02) | 0.35(0.02) |

*Secondary structure elements statistics[b] (number of residues)*

|  | β sheet | α helix | Turn |
|---|---|---|---|
| Gp120 initial[c] | 102 | 52 | 17 |
| Free | 96(7) | 56(3) | 12(4) |
| Cplx | 111(7) | 50(2) | 13(3) |
| CD4 initial[c] | 41 | 0 | 9 |
| Free | 47(3) | 5(1) | 11(3) |
| Cplx | 47(4) | 6(2) | 12(3) |

**a.** The backbone (bb) positional RMSD values were calculated with respect to the starting structure (based on PDB entry 1G9M (Kwong *et al.*, 1998), see Materials and Methods) after superposition on the respective secondary structure elements (2°) as defined by DSSP (http://www.sander.ebi.ac.uk/dssp/). The high RMSD values of gp120+CD4 in the complex are due to relative motions of the two proteins while separate analysis of the individual molecule shows small deviations from the starting crystal structure. The larger deviations of the overall backbone in gp120 is due to the high mobility of the V3 loop. Standard deviations are indicated in parentheses.
**b.** The secondary structure content was calculated with DSSP excluding the modelled V3 and V4 loops.
**c.** The total numbers of residues of gp120 and CD4 in our simulations are 346 and 99, respectively (*cf*. Material and Methods).

entire gp120 backbone in the free form primarily stem from the re-positioning of the loops near the CD4 binding site as can be readily seen by comparing the snapshots of the trajectories of gp120 in the free and CD4 bound forms (Figure 3). The snapshots also illustrate that the V3 loop, which started from an identical initial configuration in the free and complex simulations, evolves into distinct configurations with well-separated spatial distributions in the two simulations. In addition, substantial reduction in the amplitude of loop motions, in particular in the V1/V2 stem, can also be observed (this will be discussed in detail in the following section). Given the compactness of CD4-D1, its RMSD values from the crystal structure are, in both runs, small and constant regardless of the presence of gp120 (see Table 1). Despite the different dynamical behaviour of gp120's flexible loops, which is reflected both in RMSD values and per
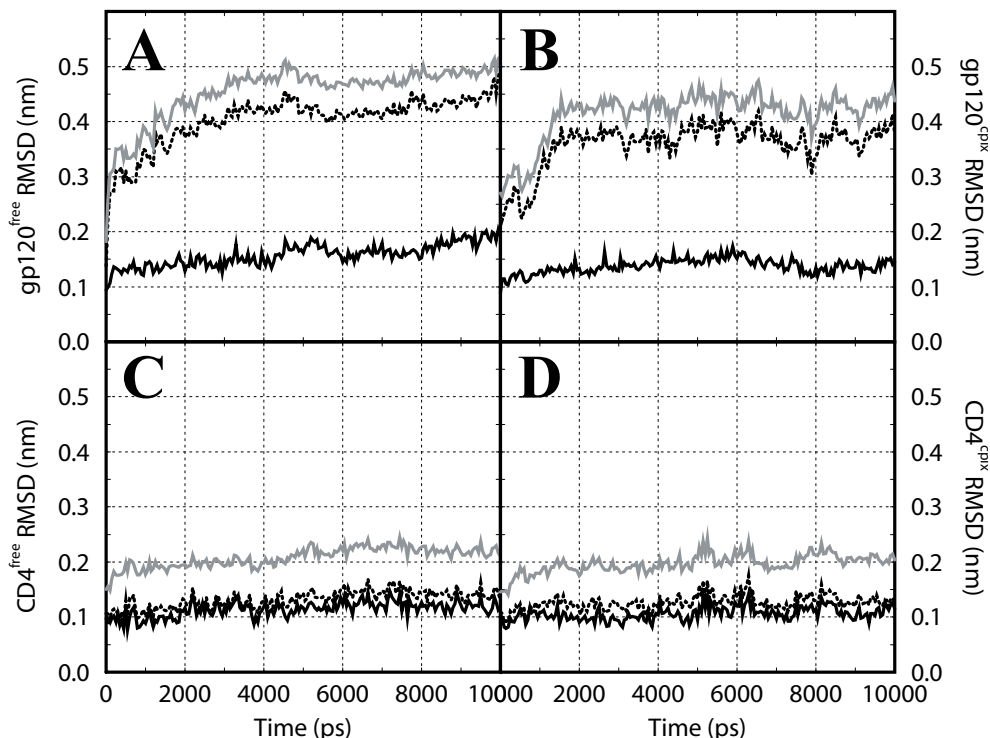
**Figure 2.** Time evolution of the RMSD from the starting structure during the 10 ns simulations of gp120 and CD4. **A.** *Free-gp120*; **B.** gp120 in *Cplx*; **C.** *Free-CD4*; **D.** CD4-D1 in *Cplx*. The backbone (N, Cα and C') RMSD values of the secondary structure elements defined in the crystal structures and the complete backbone are shown in solid and dashed lines, respectively. The all atom (including hydrogen atoms) RMSD values are shown in grey lines. The RMSD values were calculated every 2 ps after superimposition on the backbone of the secondary structure element and are plotted as a running average over a 50 ps window.

residue fluctuations, the internal energies of both gp120 and CD4 are not much perturbed upon binding (Table 2). Also worth noting is that the energy of binding ($E_{complex} - E_{free\ gp120} - E_{free\ CD4}$) amounts to $-130$ kJ/mol (considering only protein–protein and protein–solvent interactions and thus assuming that no significant changes are taking place in the bulk solvent and counter ions). This is in fair agreement with the change in enthalpy upon complexation of gp120/CD4 (-62±2 kJ/mol at 310K) reported by Doyle and co-workers (Myszka *et al.*, 2000).

### Substantial reduction of gp120 loop mobility and conformational change upon CD4 binding

A clear difference in mobility is found between the variable loops of gp120, V1–V5, and the conserved core region. Binding of CD4 substantially decreases the motions of the loops interacting with CD4, especially the V1/V2 stem and the V5 loop. It also induces a remarkable repositioning of the V3 loop that remains highly mobile in both states (Figure 3). Analysis of the backbone RMS fluctuations, expressed in terms of crystallographic temperature factor (B–factor), as a function of residue sequence for *Free-gp120*
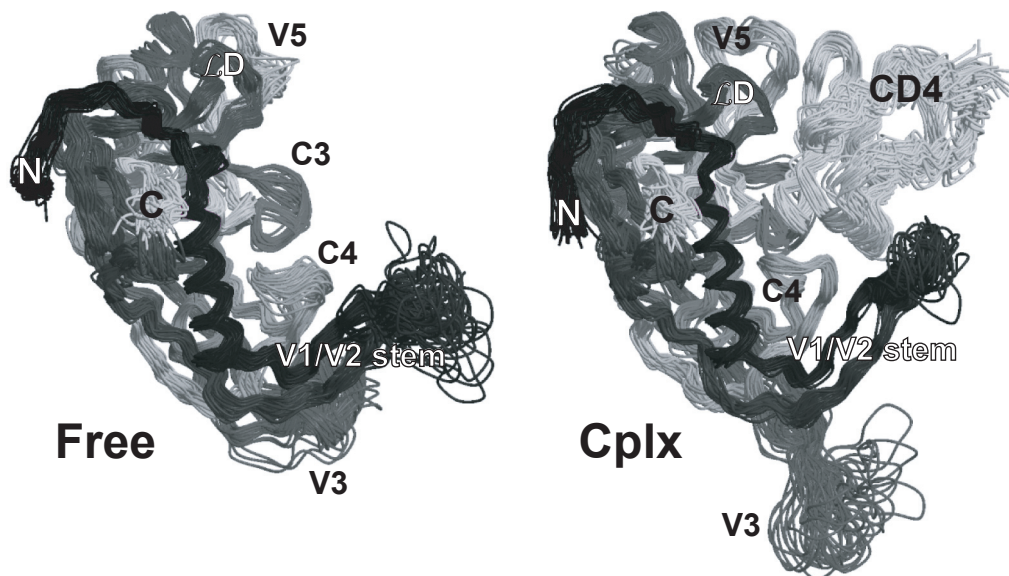
**Figure 3.** Snapshots of the MD simulations. **Left:** Free gp120; **Right:** Gp120 in complex with CD4 D1 domain. The backbone Cα traces are taken every 200 ps from the MD trajectories between 2-10 ns. The structures are superimposed on the secondary structure backbones of gp120. Gp120 is coloured from black at the N-terminus to light grey at the C-terminus. CD4-D1 structure ensemble from the *Cplx* simulation is shown in white. Loops that showed significant differences in conformation and/or flexibility are indicated (C3 consists of β15 and α3; C4 consists of β20 and β21, which form the four-stranded bridging sheet with β2 and β3 in the V1/V2 stem). Figures were generated using Molscript (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

and *Cplx* (Figure 4) reveals that most of the high-mobility sites (except V3 and V4) coincide with the CD4 binding-site. The calculated B-factors of gp120 match well with the crystallographic data with the exception of the V1/V2 stem. In CD4, however, they are significantly smaller than the experimental ones. Discrepancies in the overall B-factors amplitude are also present between the two crystal structures of free CD4, 1CDH (Ryu *et al.*, 1990) and 3CD4 (Wang *et al.*, 1990), where the latter shows similar sequential patterns but 30-40% larger overall values. Although a quantitative comparison between the simulated and experimental B-factors may not be applicable due to the difference between experimental and simulation conditions (Hünenberger *et al.*, 1995), it is apparent that the observed reduction in loop mobility upon binding near Phe43 in CD4-D1 is consistent with the experimental findings (Figure 4 inset).

The sizeable reduction of mobility upon binding is primarily determined by the intermolecular contacts in the crystal structure, *e.g.* the V1/V2 stem, ŁD and V5 regions as well as the C3 and C4 regions that contact the protruding CD4-Phe43 (Kwong *et al.*, 1998). The attenuation of the loop motions arises, to a large extent, from the formation of many intermolecular hydrogen bonds, most of which are also present in the crystal structure (Table 3, Figures 4 and 5). The large loop motions in gp120 observed in both the free and complex simulations are consistent with the fact that the crystallisation of gp120 was only possible when the variable loops, *i.e.* V1-V3, were truncated (Figure 1); the introduction of CD4 was necessary to further stabilise the ternary complex. Despite its shorter

**Table 2.** Statistics of non-bonded energies[a] (2-10 ns)

| *Coulomb's electrostatic energy (kJ/mol)* | | |
|---|---|---|
| Free gp120 | Gp120 internal | -38262 (363) |
| | Gp120–solvent | -30428 (674) |
| | Total | -68690 (766) |
| Free CD4 | CD4 internal | -12298 (314) |
| | CD4–solvent | -11761 (497) |
| | Total | -24059 (588) |
| Gp120/CD4 complex | Gp120 internal | -37923 (416) |
| | Gp120–solvent | -29930 (699) |
| | CD4 internal | -12765 (289) |
| | CD4–solvent | -11295 (503) |
| | CD4–gp120 | -1030 (177) |
| | Total | -92943 (1015) |
| *Lennard-Jones (van der Waals) energy (kJ/mol)* | | |
| Free gp120 | Gp120 internal | -12430 (115) |
| | Gp120–solvent | -2795 (144) |
| | Total | -15225 (184) |
| Free CD4 | CD4 internal | -3406 (51) |
| | CD4–solvent | -671 (78) |
| | Total | -4077 (93) |
| Gp120/CD4 complex | Gp120 internal | -12276 (103) |
| | Gp120–solvent | -2635 (137) |
| | CD4 internal | -3395 (50) |
| | CD4–solvent | -477 (25) |
| | CD4–gp120 | -452 (76) |
| | Total | -19235 (196) |

**a.** The non-bonded energies were calculated with the GROMOS96 force field (Daura *et al*., 1998) using a twin range cutoff of 0.8 and 1.4 nm with a reaction field correction (see Material and Methods). The energies are the sum of short- (SR) and long range (LR) terms; 1-4 terms were not included. Standard deviations are indicated in parentheses.

loop length, resulting in higher structural restriction, the V4 loop was still disordered and unresolved in the crystal structures (Kwong *et al*., 1998; Kwong *et al*., 2000a), probably because of its intrinsic conformational heterogeneity and remoteness from the CD4 binding interface.

### Intermolecular hydrogen bond network and Phe43 cavity

The increase in rigidity observed in the complex originates primarily from the formation of intermolecular hydrogen bonds. Although many of them are already present in the crystal structure, some are only formed during the molecular dynamics simulation. Our simulation allows, in addition, to assess the dynamical nature of these critical interactions. This can be done by analysing their occurrence throughout the trajectory. Many long-lived intermolecular hydrogen bonds bridge the tips of the mobile loops of gp120 interacting with CD4 and a hemisphere of CD4-D1 (Figure 5). This hydrogen bond network nicely encloses the docking cavity of Phe43, which is the key element for the binding mechanism. A
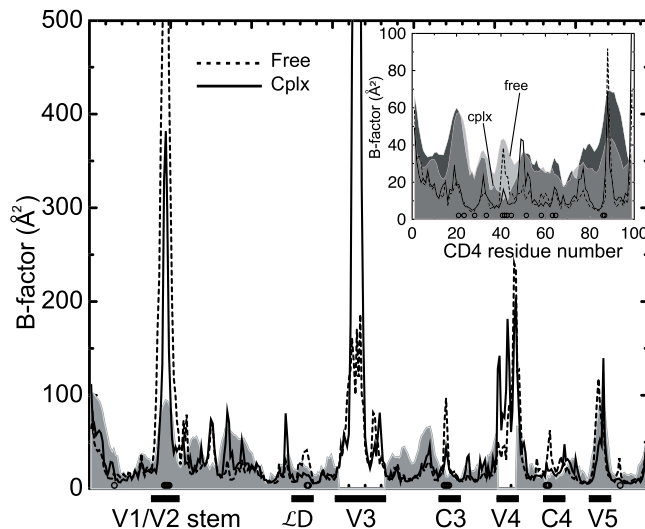
**Figure 4.** Comparison of gp120 Cα atoms temperature factors (B-factors) calculated from the RMS fluctuations in the 2-10 ns simulations of *Free-gp120* (dashed lines) and *Cplx* (solid lines). The B-factors for CD4-D1 are shown in insets at a different scale. Experimental B-factors taken from the ternary gp120/CD4/CD4i crystal structure (PDB entry 1G9M, Kwong *et al.*, 1998) are shown in grey area. The B-factors of CD4-D1 in the free form (light grey area overlaid onto the complex one) are taken from the free CD4 structure (PDB entry 3CD4)(Wang *et al.*, 1990). They are comparable to the B-factors in the complex except for the Phe43 binding site  Note that the V3 loop was truncated and the electron density of the V4 loop was missing. Residues forming stable intermolecular hydrogen bonds in the complex form are indicated with open circles along the horizontal axes (for definitions, see Table 3).

large number of side chain-side chain or side chain-backbone hydrogen bonds cover a large area of the complex interface in a non-specific manner. For instance, the side chain hydroxyl group of S42 in CD4 is able to form several hydrogen bonds by hopping amongst various backbone hydrogen bond acceptors in gp120 that are in its close proximity (Table 3). In addition, there are a few stable backbone-backbone hydrogen bonds in the vicinity of the Phe43 cavity (Table 3) involving residues of gp120 (S365, G366, G367 and D368) that are well conserved amongst various HIV-1 strains. The importance of these residues has been highlighted by structure-based mutagenic studies in which mutations of S365, G366 and D368 in gp120 diminished the affinity of the CD4-binding site specific antibody b12 (Saphire *et al.*, 2001). Overall, the residues involved in these intermolecular hydrogen bonds gave rise to an average intermolecular Coulomb's electrostatic energy of −624±124 kJ/mol, accounting for 61% of the total intermolecular Coulomb's electrostatic energy (Table 2).

These observations allow us to propose a functional role for the hydrogen bond network: the non-specific, loose, flexible and wide-spread hydrogen bonds and salt-bridges and the specific, tight, rigid and confined ones generate an affinity gradient that drives the insertion of the Phe43 phenyl ring into its binding cavity in gp120. Once inserted, it is locked in its "knob-and-socket" geometry by the specific hydrogen bond network in the proximity of the binding pocket. The hydrophobic contacts between the CD4 Phe43 side chain and the gp120 cavity then provide the short-range stabilising factor. These

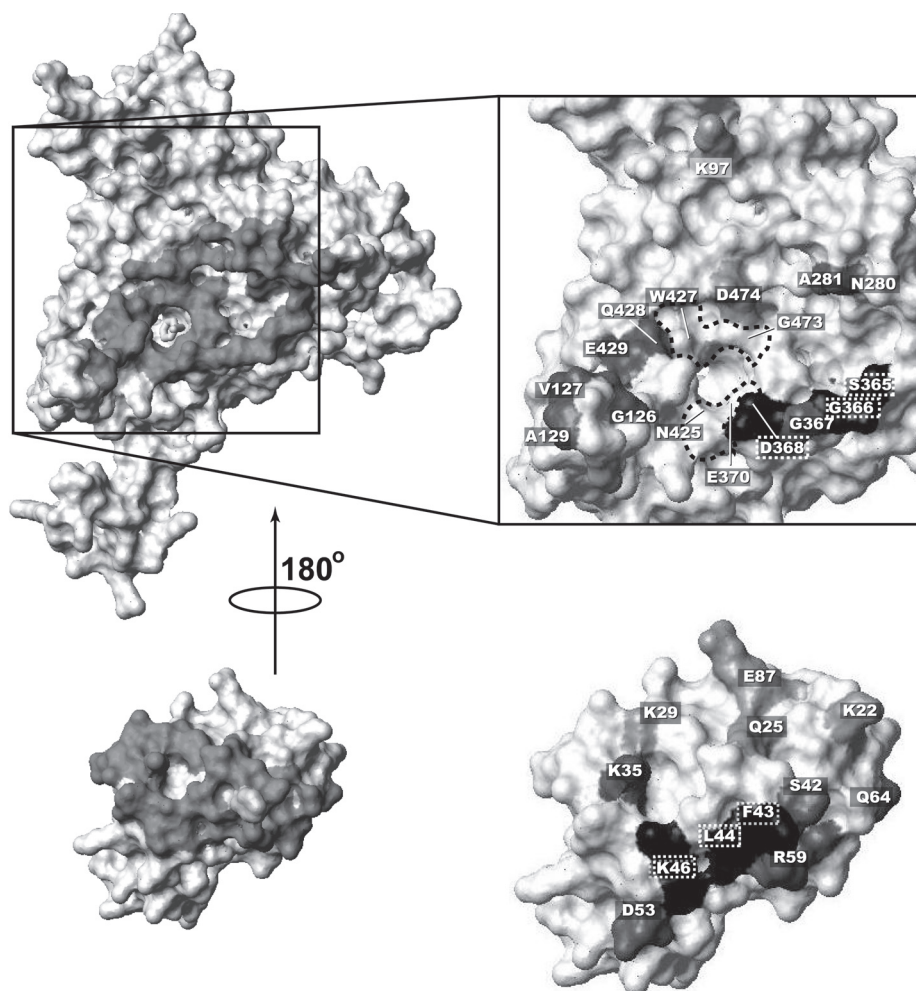**Table 3.** Intermolecular hydrogen bond and salt bridge statistics[a] (2–10 ns)

| Intermolecular hydrogen bond | | | | | |
|---|---|---|---|---|---|
| CD4 | | Gp120 | | Occurrence (%) | |
| Backbone - backbone | | | | | |
| F43 | N | G473 | O | 11.5 | |
| L44 | O | D368 | N | 67.8 | |
| K46 | N | S365 | O | 14.2 | |
| K46 | N | G366 | O | 15.7 | |
| Side chain − backbone | | | | | |
| K35 | Nζ | A281 | O | 15.9 | |
| S42 | Oγ | Q428 | O | 69.9 | |
| S42 | Oγ | E429 | O | 61.5 | |
| S42 | Oγ | E429 | N | 31.2 | |
| L44 | N | D368 | Oδ | 92.6 | |
| D53 | Oδ2 | G367 | N | 11.1 | |
| R59 | Nη1 | E429 | O | 45.9 | |
| R59 | Nη1 | V430 | O | 55.2 | |
| Q64 | Nε2 | C126 | O | 92.4 | |
| Q64 | Nε2 | V127 | O | 18.0 | |
| Q64 | Nε2 | A129[b] | O | 29.3 | |
| N66 | Nδ2 | A129[b] | O | 25.8 | |
| Side chain − side chain | | | | | |
| K22 | Nζ | E429 | Cδ | 77.8 | |
| Q25 | Nε2 | D474 | Oδ | 93.8 | |
| K29 | Nζ | N280 | Oδ | 83.6 | |
| R59 | Nε | D368 | Oδ | 80.7 | |
| Q64 | Nε2 | S131 | Oγ | 15.2 | |
| E87 | Cδ | K97 | Nζ | 61.1 | |
| D88 | Cγ | K97 | Nζ | 14.8 | |

**a.** Intermolecular hydrogen bonds and salt bridges (side chain–side chain hydrogen bonds) occurring for more than 10% during the 2–10 ns period are listed. A hydrogen bond is considered to exist when the donor-hydrogen-acceptor angle is larger than 120° and the donor-acceptor distance is smaller than 0.28 nm.

**b.** A129 is one of the linker residues GAG that were used in the crystal construct (1G9M) to replace the V1/V2 loops (128-194) (Kwong et al., 1998).

contacts are extremely well maintained during the simulation bearing very small structural variations (Table 4); translation and rotation of the phenyl ring with respect to the cavity were limited within 0.2 nm and 60°, respectively (data not shown).

## CD4 binding induced bridging sheet stabilisation

The so-called "bridging sheet" is a four-stranded β-sheet that consists of β2 and β3 in the V1/V2 stem and β20 and β21 in the C3 region. It interposes between the inner and outer domains of gp120 and is thought to be flexible and disordered in the absence of CD4. Recent thermodynamic studies on the binding of gp120 and CD4 have demonstrated unprecedented large entropy changes, which were

**Figure 5.** Mapping of the intermolecular hydrogen bond network onto the intermolecular contact interfaces. **Left panel:** The contact van der Waals surfaces of gp120 and CD4-D1. Heavy atoms of gp120 and CD4-D1 that have intermolecular distances within 0.5 nm in the starting structure are coloured dark grey. The side chain of CD4-Phe43 is shown in light grey and as sticks in gp120 to indicate the binding cavity. **Right panel:** Residues involved in intermolecular hydrogen bonds. The types of hydrogen bonding is colour-coded as: *light grey*: side chain-side chain hydrogen bond; *dark grey*: side chain-side chain and/or side chain-backbone hydrogen bond; *black*: backbone-backbone and/or side chain-backbone hydrogen bonds (for details, see Table 3). Residues that form the receptive CD4-Phe43 cavity (E370, N425, W427 and G473) are outlined with dashed lines. Note that several residues that do not have intermolecular contacts in the starting structure form intermolecular hydrogen bonds during the molecular dynamics simulations (gp120: K97, V127 and A129; CD4: K22, D53 and E87). The figures were generated using MolMol (Koradi *et al.*, 1996).

**Table 4.** Statistics of CD4 Phe43-gp120 cavity native contacts (%)[a] (2-10 ns)

| Residue | Atom | Occurrence |
|---|---|---|
| E370 | Cγ | 100.0 |
| E370 | Cδ | 99.5 |
| N425 | Cβ | 94.1 |
| N425 | C | 94.1 |
| N425 | O | 99.0 |
| W427 | N | 99.1 |
| W427 | Cα | 100.0 |
| W427 | Cβ | 83.2 |
| G473 | Cα | 99.2 |
| G473 | C | 98.9 |
| G473 | O | 100.0 |

**a.** Eleven heavy atoms of gp120 have interatomic distances with any atom of the CD4 Phe43 side chain within 0.35 nm in the crystal structure (1G9M, Kwong *et al.*, 1998) A native contact during the simulation is considered to exist if the distance between the listed atoms and the centre of the Phe43 phenyl ring is less than 0.6 nm.

proposed to be possibly accounted for by a stabilisation of the bridging sheet and a restriction of the interdomain motions upon CD4 binding. Detailed structural and dynamical information is, however, still lacking.

The conformational changes of the bridging sheet upon removal of CD4 were monitored *in silico* by RMSD matrix analysis (Figure 6 lower right panel). The bridging sheet undergoes a series of transient structural changes with a major one at about 8 ns, after which it adopts a conformation somewhat similar to the ones in the 4.5-5.5 ns segment. The same analysis on the simulation of the complex showed little change throughout the 10 ns simulation (Figure 6 upper left panel). Several interstrand backbone hydrogen bonds between the antiparallel β2 and β21 that delimit the inner and outer domains of gp120 also show different hydrogen-acceptor distance distribution in the two simulations: While the total numbers of the interstrand backbone hydrogen bonds throughout the simulations are the same, for *Free-gp120*, the hydrogen bond length distribution peaks at 0.188 nm with a half width of 0.048 nm whereas for *Cplx* it peaks at 0.182 nm with a narrower half width of 0.035 nm. The shorter hydrogen bond distances in the complex are indicative of an increased stability of the bridging sheet upon CD4 binding while it remains flexible otherwise. This is consistent with the model proposed in the thermodynamic studies mentioned previously.

## Conformational changes and polymorphism of the V3 loop upon CD4 binding

The V3 loop has been targeted as a prominent candidate to tackle the viral fusion problem. Based on structural (Kwong *et al.*, 1998), mutagenic (Rizzuto *et al.*, 1998) and antigenic analyses (Wyatt *et al.*, 1998), co-receptor binding was proposed to require the V3 loop in concert with the bridging sheet. These form the co-receptor binding site that is created only upon CD4 binding. A large panel of monoclonal antibodies (MAbs) based on V3 loop fragments has been elicited against several primary isolates from different HIV-1 clades (subtypes) (Gorny *et al.*, 1993; Nyambi *et al.*, 1998; Gorny *et al.*, 2002). Several V3 loop–derived peptides have also been devised to inhibit viral entry into target cell in
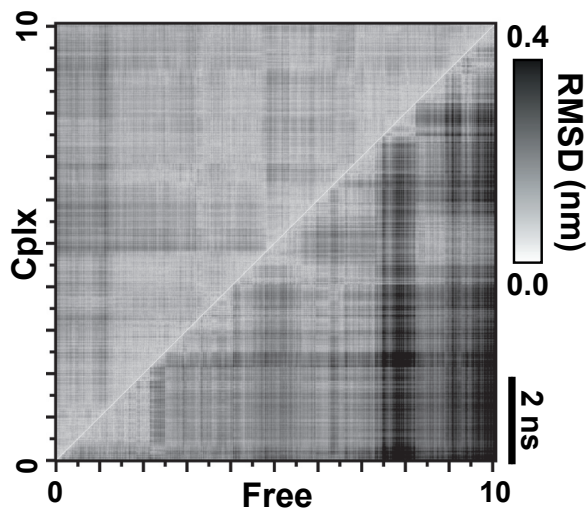
**Figure 6.** Pair-wise backbone RMSD matrix of the bridging sheet in the free (lower right panel) and the complex form (upper left panel). The four β-strands are defined as β2: 109-113; β3: 198-202; β20: 423-427; β21: 430-434. Each dot represents a positional RMSD between two conformations taken from the respective trajectories indicated on the axes and is colour-coded accordingly to the scale shown on the right. The RMSD values were calculated on backbone atoms of the bridging sheet (β2, β3, β20 and β21) after superimposition on the backbone atoms of β2 and β3. Conformations were taken every 10 ps. An equilibrated conformational sampling period is found when an off-diagonal region shows a continuous low RMSD (white to light grey).

a co-receptor-specific manner (Sakaida *et al.*, 1998; Verrier *et al.*, 1999; Basmaciogullari *et al.*, 2002). Successful vaccine design has however been hindered by the underlying structural polymorphism of the V3 loop. The well-conserved region, GPGR, can indeed adopt various type of β-turns and is found in various configurations when bound to different MAbs (Rini *et al.*, 1993; Ghiara *et al.*, 1994; Ghiara *et al.*, 1997; Stanfield *et al.*, 1999; Tugarinov *et al.*, 1999; Tugarinov *et al.*, 2000; Sharon *et al.*, 2003). Note that previous structural studies were only performed on V3 loop fragments and little is thus known about its conformation when attached to the core structure. We therefore investigated the dynamics and conformation of the V3 loop both in free gp120 and in the CD4-bound form.

   As mentioned above the V3 loop is highly mobile and occupies an ample space in the vicinity of the co-receptor binding site. Different localisations in the free and CD4-bound states can be clearly distinguished in the corresponding trajectories (Figure 3). Relocation of the V3 loop, upon CD4 binding, towards the basal part of the bridging sheet agrees with the model proposed by Sodroski and co-workers based on their mutagenesis study on the co-receptor CCR5 binding (Rizzuto *et al.*, 1998). In their comprehensive study, the removal of the V3 loop (residues 298-329) was found to abolish the co-receptor CCR5 binding completely. Many residues (K121, T123, K207, L317, E381, F383, I420, K421, Q422, P438, R440 and G441) were also identified to result in more than 90% loss in CCR5 binding when mutated. Comparing the solvent accessible surface area (SAS) of these residues in the free and complex simulations reveals an increase of SAS in the complex form only for I420 (15%) and Q422 (31%) while a small decrease is observed for R440 (8%). Other residues are

either unaffected or buried in both the free and complex forms by the V3 loop. The changes in SAS of the above three residues amount, however, to a gain of *ca* 1.10 nm², which may be an indication of the proposed exposure of the binding site. The clustering of the V3 loop and the bridging sheet can also be inferred from the lowering of the electrostatic Coulomb s energy between the V3 loop as a whole (residues 298–329) and the subset of residues that are important for CCR5 binding: -438±60 kJ/mol in the complex form versus -345±101 kJ/mol in the free form. The abundant basic residues of the V3 loop (six arginines and two lysines, see Figure 1) are thought to facilitate the recruitment of the generally acidic chemokine co-receptor subsequent to the binding of CD4 (Dragic *et al.*, 1998; Farzan *et al.*, 1998; Rabut *et al.*, 1998). They generate a positive electrostatic potential that occupies an enormous space centred at the basal region of the V3 loop. Moving from the free to the complex form relocates the positive potential toward the co-receptor binding site while neglecting the V3 loop drastically reduces this positive electrostatic potential (Figure 7)

Structurally, the V3 loop shows a high propensity of β-hairpin formation in our simulations but the position of the turn is poorly defined (Figure 8). The conserved GPGR sequence shows several transient β-turns, including type II and type VIII β-turns in the two simulations. Other tight turns were found at RGPR, QRGP and IQRG, as a result of one to three residue upstream shifts. Comparison with the available GPGR structures (see Material and Methods) gives backbone RMSD
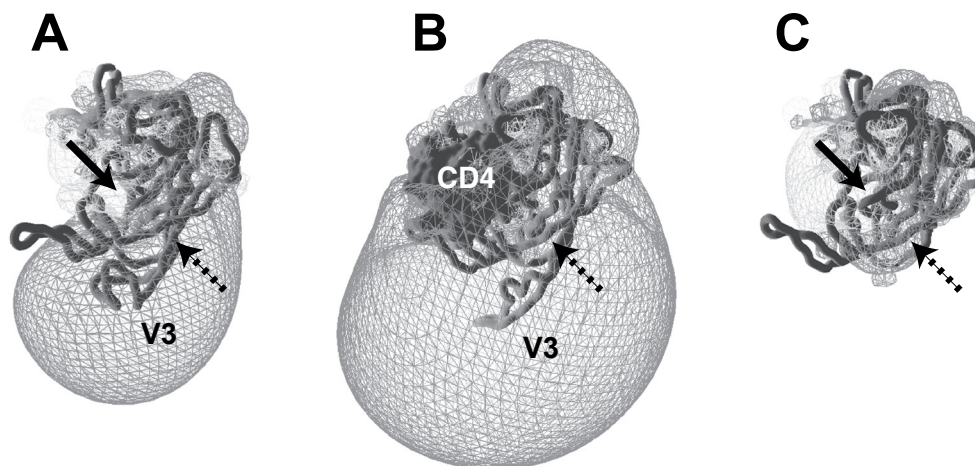


**Figure 7.** Electrostatic potential surface of gp120. **A.** Free gp120. **B** Gp120 in complex with CD4–D1. CD4–D1 is labelled and is shown in van der Waals surface. **C.** Gp120 in the complex form without CD4–D1 and the V3 loop. The core structure of gp120 is identical to that in **B** and CD4–D1 and the V3 loop were manually removed from the complex structure. The light and dark grey meshes correspond to electrostatic potentials of −2 and +2 kT/e, respectively. Negative electrostatic potentials are primarily present at the CD4 binding site of gp120. They were calculated using GRASP (Nicholls *et al.*, 1991) with protein and solvent dielectric constants of 2 and 80, respectively. The ionic strength was set to zero. The structures are taken from the snapshots at 6 ns both in the free and complex trajectories, which represent the most populated configuration of the V3 loop based on RMSD matrix analysis. The CD4 and CCR5 binding sites are indicated by solid and dashed arrows, respectively.
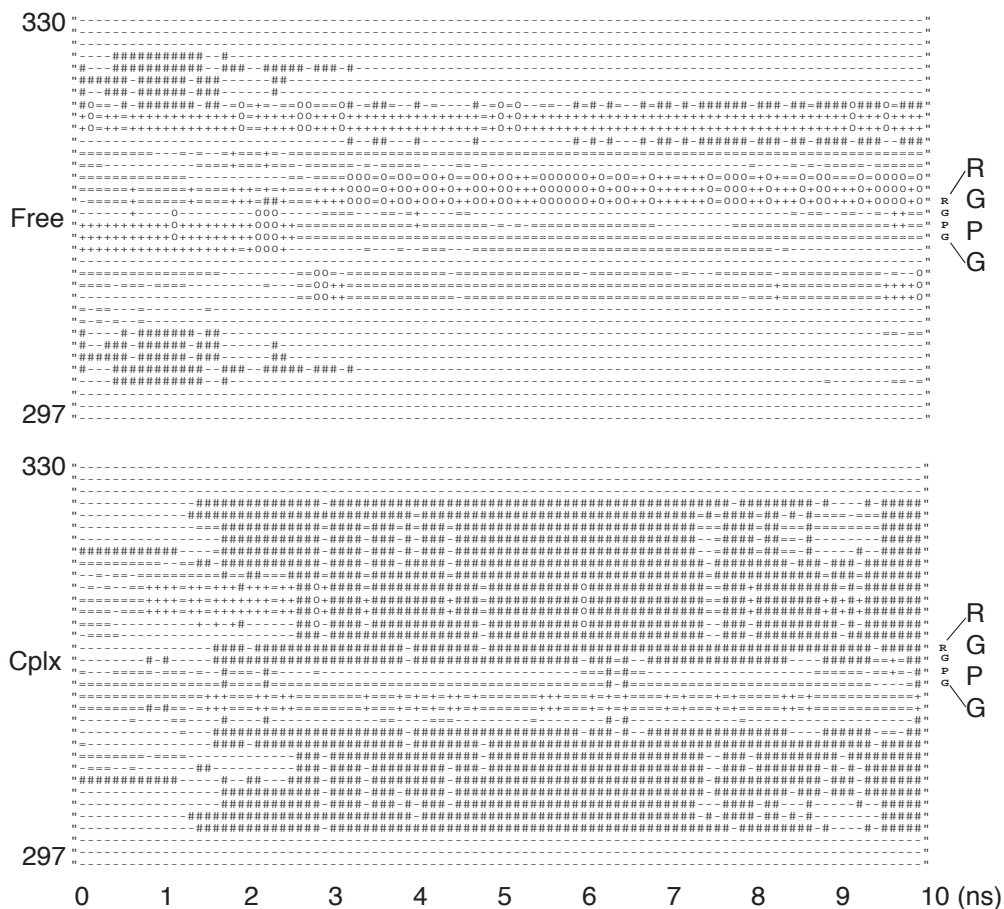
**Figure 8.** Secondary structure evolutions of the V3 loop (residues 297-330) as a function of the simulation time. Secondary structure elements are coded as following: Coil (-), β-sheet or β-bridge (#), bend (=), turn (+) and α- or $3_{10}$-helix (0). Each block corresponds to a time step of 100 ps. The position of the conserved residues, GPGR, is indicated at the right end of each diagram.

values between 0.2 and 0.6 nm with the various antibody-bound structures irrespective of the effect of CD4 complexation (data not shown).

### Essential dynamics analysis reveals concerted loop motions in gp120 upon CD4 binding

The complexity of molecular motions observed in MD simulations can be simplified by essential dynamics analysis (Amadei *et al.*, 1993), whereby sets of essential and correlated motions are extracted. Such an analysis allows distinguishing high frequency local motions, which are restrained and harmonic in nature and contain less structural information, from "essential", large amplitude global motions. In order to focus on the motions associated with CD4 binding such an analysis was performed on merged trajectories from the free and complex forms of gp120 and of CD4-D1. Two distinct distributions along the first

eigenvector, *i.e.*, with the largest eigenvalue, can be found, which correspond to the well separated free and complex states of gp120 (Figure 9A & B). Already the second eigenvector of the two trajectories shows overlap in positional fluctuation displacement suggesting similar motions between the two states. Narrow Gaussian displacement distributions appear after the first few eigenvectors suggesting restrained harmonic motions common to the two states. Projections of the two extremes of the first eigenvector along the merged trajectory on the average structure illustrate the major difference between the two states (Figure 9C). Note that the intrapolation between the two extremes does not represent the transition pathway but merely highlights the structural differences.
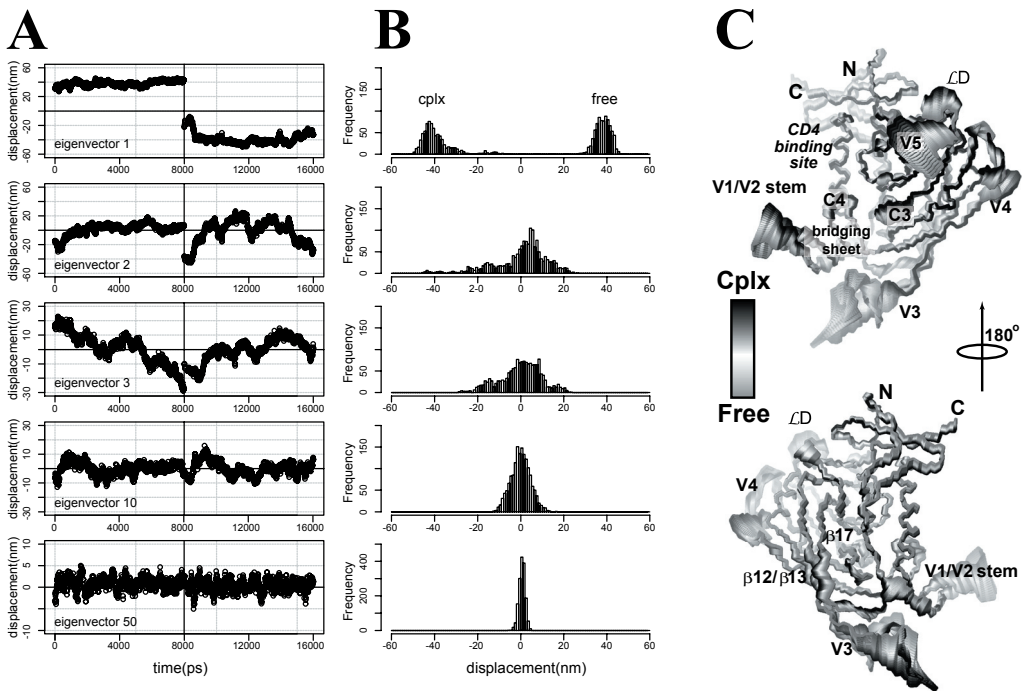


**Figure 9.** CD4 binding-induced conformational changes of gp120 extracted *via* essential dynamics analysis. **A.** Projections of the total atomic fluctuations of the merged trajectory (free: 0-8 ns and complex: 8-16 ns) along the selected eigenvectors. The eigenvectors are numbered with respect to the amplitude of displacement in descending order. **B.** Histograms of the distributions of the displacement of each eigenvector. Distinct distribution can only be found in the first eigenvector. Note that the scale is different for the 50th eigenvector (bottom row). **C.** Projections of the two extremes of the first eigenvector along the merged trajectory onto the average structure. The linear intrapolation between the two extremes is coloured from light grey (free) to black (complex) with white colour in the midst of the two to highlight the primary structural differences between the two states. The C3 and C4 regions of gp120 that are part of the Phe43 cavity show significant closure of their lid upon CD4 binding. The V3 loop shows the largest conformational change. The view from the back side of the structure after 180° rotation (*bottom*) reveals propagated structural perturbations at the interior of gp120, involving β17, which is adjacent to the Phe43 cavity as well as the two β-strands, β12 and β13 that connect the V3 loop.

In gp120, a correlated loop-contraction primarily around the CD4 binding site is revealed with, in particular, a pronounced closure of the lids of the Phe43 cavity (Figure 9C). This correlated motion increases the curvature of the CD4 binding site leading to a gain in interfacial complementarity. The concerted contraction involves primarily the V1/V2 stem, V5 and, to a less extent, the C3 and C4 regions. These finger-like structures, in particular the V1/V2 stem and V5, are in a much more open and relaxed state in the free form. The most prominent transition occurred in the V3 loop, which is driven by CD4 binding towards the basal region of the bridging sheet. In contrast to the substantial conformational changes in gp120 no major structural transition upon binding was identified for CD4-D1 suggesting that the principle modes of motion are not affected by complexation and that only their amplitudes are reduced.

## Discussion

We have shown by MD simulations that CD4 binding reduces the mobility of various loops of gp120 around the binding site and induces conformational changes that effectively lead to the wrapping up of a hemisphere of CD4-D1. While substantial changes occur in gp120, in CD4-D1, only the Phe43 containing loop showed a slight reduction of mobility upon complexation. The intermolecular interactions can be categorised into three levels: non-specific side chain-side chain or side chain-backbone hydrogen bonds (or salt-bridges), specific backbone-backbone hydrogen bonds and hydrophobic contacts between both partners resulting in the insertion of CD4 Phe43 into the receptive cavity in gp120. The large-amplitude motions of the V1/V2 stem and V3 and V4 loops may provide a shielding umbrella that masks the CD4 binding site in the *monomeric* form of gp120 as proposed by Kwong *et al.* based on their thermodynamic finding (Kwong *et al.*, 2002). The observed entropy loss upon binding of CD4 to gp120 (Myszka *et al.*, 2000; Kwong *et al.*, 2002) is a clear indication of a reduction of flexibility and is in line with our observations. When CD4 is attracted into the vicinity of the binding site, the mobility of the V1/V2 stem, V5 loop and £D reduces. The extensive intermolecular hydrogen bond formation and van der Waals contacts at the gp120/CD4 interface result in the closure of C3 and C4 regions leading to the formation of the required Phe43 binding cavity. They also reduce the overall mobility in the regions involved in the intermolecular contacts between both partners.

The formation of the extensive intermolecular hydrogen bonds as well as the stabilisation of the bridging sheet may account for the experimentally observed entropic loss (Myszka *et al.*, 2000; Kwong *et al.*, 2002), which is compensated by a gain in enthalpy through the intermolecular interactions described here. In contrast, the V3 loop remains flexible after its relocation induced by the binding of CD4 and no entropy cost is thus paid for this process. These observations are in agreement with the experimental findings that the main entropic changes occur within the core structure of gp120 (Myszka *et al.*, 2000).

One can summarise the above observations into a "binding funnel" model where dynamics and different interaction modes are coupled leading to efficient recognition and specific affinity (Tsai *et al.*, 1999). During the search of receptor CD4, gp120 is constantly subject to immune system attack. A defined binding pocket based on rigid body docking would require exhausting geometry search while antibodies neutralisation *via* structural epitope recognition is also taking place. Gp120 devised, therefore, a multi-level binding mechanism as such that the conformation with the specific binding cavity is embedded in an ensemble of less selective conformations facilitating target search: *i)* the charged residues located in the

highly mobile loops generate long range electrostatic attraction that guide CD4 to its binding interface consisting of a non-specific hydrogen bond network bearing less structural selectivity and energetic trapping; *ii)* the loose side chain-side chain hydrogen bonding that confines gp120 and CD4 into close proximity leads to the formation of specific backbone-backbone hydrogen bonds and subtle changes in the loops around CD4 lead to the closure of the lid of the receptive cavity; *iii)* while the specific hydrogen bonds are formed, the Phe43 phenyl ring plugs into its binding cavity. Only one molecule that perfectly matches with the last two steps, CD4, will be able to accomplish the structural rearrangements that weld the complex structure together, thereby completing the first stage of the fusion process: the gp120/CD4 recognition. It should be noted that we did not simulate the binding process and therefore no kinetic or free energy data could be obtained from the current study. Therefore, our "binding funnel" model cannot be interpreted in term of a free energy surface of binding; it only presents a *geometrical* and *chronological* view of the binding process.

The ability of anti-CD4-binding-site MAbs to neutralise HIV-1 has been associated with their affinity for the trimeric envelope virion structure, rather than monomeric gp120 (Poignard *et al.*, 2001). Recent thermodynamics study also suggests that the monomeric gp120 may act as decoys to the immune system (Kwong *et al.*, 2002). Most antibodies but b12 and 4KG5 are in fact elicited against the flexible monomeric gp120 with suboptimal angle of approach to the CD4 binding site such that they are sterically hindered by the interaction between the variable loops in gp120 (Zwick *et al.*, 2003). The ability to bind oligomeric gp120 is in fact the determinant to neutralise the virus (Roben *et al.*, 1994; Sattentau and Moore, 1995). On the one hand, the interpretation of our observed conformational changes is confounded by the limited knowledge about the trimeric state of gp120, the monomeric gp120 crystal structure being the only structural template available to date. On the other hand, CD4 is the unique element for the initiation of the CD4–dependent viral entry pathway. The formation of the gp120/CD4 complex is apparently the common ground for recognition, regardless of the oligomeric state. This actually demonstrates the remarkable plasticity of gp120 where the quaternary constraint imposed by the trimerisation of the envelope proteins may well be generated *via* the binding to CD4 alone, assuming that the structure of the monomeric gp120/CD4 complex resembles that in the oligomeric state. Although not presented here, we do observe correlated motions upon CD4 binding on the "back-side" of gp120, which has been proposed to be involved in the trimerisation process, either directly or more likely *via* interactions with the transmembrane trimerisation domain gp41. Under the assumption of an unique gp120/CD4 complex structure, the gp120/CD4/CD4i crystal structure has provided invaluable information about the gp120/CD4 interface, which has led to a large number of structural and biochemical studies such as structural mapping by monoclonal antibodies (Saphire *et al.*, 2001; Xiang *et al.*, 2002). Yet, it still lacks the dynamical information regarding the conformational changes occurring in gp120 upon CD4 binding, the limiting factor to the subsequent viral entry process. Some intermolecular hydrogen bonds, mostly mobile side chain-side chain hydrogen bonding (Figure 5), were only identified in our simulation. The geometry and rigidity of the well-defined CD4-Phe43 cavity, which is composed of some mobile residues in the C3 and C4 regions, is only induced upon CD4-binding (Figure 3). Furthermore, formation of the complementary gp120/CD4 interface requires correlated loop contractions as revealed by essential dynamics analysis. This process also leads to a stabilisation of the bridging sheet while the two β-hairpins forming it (the V1/V2 stem and C3) are somewhat parted

in the free form. Correlated motions at the putative trimerisation interface are also observed. We can only speculate that the perturbation at the trimerization interface, propagated from the gp120/CD4 interface, may trigger the release of tension stored within the coiled-coil structure of gp41 as proposed in the "spring-loaded" model (Eckert and Kim, 2001).

## Conclusion

Our simulations of free and CD4-bound gp120 represent two distinct states, namely the relaxed ground state and the contracted, excited or pre-fusogenic state of gp120, respectively. Large amplitude loop motions in gp120 are attenuated and show concerted contraction upon CD4 binding. The associated entropic cost is compensated by the wealth of intermolecular interactions at the CD4 interface ranging from non-specific to specific hydrogen bonds and hydrophobic contacts around and at the Phe43 binding cavity, in agreement with the large enthalpy/entropy compensation measured experimentally. This differentiated mode of interaction from long range electrostatic attraction *via* non-specific and specific short-range interactions, in combination with the dynamical nature of the system, allowed us to propose a binding funnel model for the gp120/CD4 interaction. In addition, the complexation also drives clustering of the V3 loop and the bridging sheet that generates a highly positive electrostatic attraction gradient for subsequent co-receptor binding. In line with the proposed model deduced from the mutagenesis mapping of the co-receptor binding site (Rizzuto *et al.*, 1998), our results provide a plausible explanation for the functional relationship between the V3 loop and CD4-binding and the otherwise inefficient CD4-independent entry pathway (reviewed by (Berger *et al.*, (1999)). We have here demonstrated the sophisticated plasticity of gp120, from a rigid core to a floppy exterior shielding that provides recognition specificity without compromising the capacity to evade attack from the immune system.

## Acknowledgements

# Entropy calculation of HIV-1 Env gp120, its receptor CD4 and their complex:

## an analysis of entropy changes upon complexation[#]

Shang–Te D. Hsu[1], Christine Peter[2], Wilfred F. van Gunsteren[2] and Alexandre M.J.J. Bonvin[1]

[1]*Bijvoet Center for Biomolecular Research, Utrecht University, 3584 CH Utrecht, The Netherlands*
[2]*Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH-Hönggerberg, CH-8093 Zürich, Switzerland*

## Abstract

**The HIV-1 gp120/CD4 interaction shows an unprecedented large entropy/enthalpy compensation with the capacity to fine-tune recognition over a broad range of affinity. The intermolecular interaction involves stable hydrophobic contacts with a unique protruding CD4-Phe43 structure surrounded by an intermolecular hydrogen bond network that covers the hemisphere of the CD4 D1 domain. We have applied a heuristic formula based on the covariance matrix of atom-positional fluctuations to assess the configurational entropy of the gp120/CD4 complex at different levels. The system was dissected into various subsets of atoms to evaluate the entropic contributions of different functional elements. By combining the trajectories of the free and complex forms, further insight about the conformational sampling was extracted. Despite the limited sampling time of ten nanoseconds, the theoretically derived changes in configurational entropy are in fair agreement with the experimentally determined data. The simultaneous evaluation of different interaction modes through a decomposition approach is only feasible with the knowledge of the atomic trajectory of the system. Given sufficient sampling of conformational space, the configurational entropy analysis presented here shall potentially provide accurate estimations of thermodynamic properties of biomolecules.**

## Introduction

Conformational changes are central to the viral entry of Human Immunodeficiency Virus type-1 (HIV-1), which is initiated by the mutual recognition of the HIV-1 envelope protein (Env) gp120 and the cellular receptor CD4 (Olshevsky *et al.*, 1990; Sattentau and Moore, 1991). This is followed by the binding of the chemokine co-receptor to gp120 (Trkola *et al.*, 1996; Wu *et al.*, 1996; Rizzuto *et al.*, 1998; Berger *et al.*, 1999) and the extension or stretching of the coiled coil gp41, leading to the insertion of the membrane fusion peptide into the host membrane, which induces fusion of the viral and host membranes (Chan and Kim, 1998). Specific recognition of CD4 by gp120 is evidently the linchpin of the CD4-mediated viral infection pathway. Recent extensive biochemical and biophysical studies have provided ample evidence of a substantial structural rearrangement during the gp120-CD4 recognition process (Wyatt and Sodroski, 1998). Large enthalpic and entropic changes upon binding were observed experimentally (Myszka *et al.*, 2000; Kwong et al., 2002). The free energy of binding, however, is small due to a remarkable entropy/enthalpy compensation. This means that the affinity of gp120 for CD4 and CD4-induced monoclonal antibodies (CD4MAb), is relatively small and can be fine-tuned by small variations in those two rather large contributions. The gp120/CD4 interaction is predominantly determined by the core of gp120: removal of the hypervariable loops corresponding to nearly 40% of the total sequence shows indeed little effect on the thermodynamic properties of binding (Myszka et al., 2000).

Comprehensive structural insight is, however, limited by the intrinsic flexibility of the system. The structure of the ternary complex of the core of gp120, CD4 and a CD4-induced antibody is the only available atomic information to date (Kwong *et al.*, 1998; Kwong *et al.*, 2000). This truncated form of

gp120 – core gp120 – does not compromise much the biophysical and biological properties compared to the full-length wild type (wt) (Pollard *et al.*, 1992; Wyatt et al., 1993; Myszka *et al.,* 2000; Kwong *et al.,* 2002). It provides a suitable structural template for molecular dynamics (MD) studies allowing a significant reduction in the size for the simulation system as compared to full-length gp120. Using MD simulations we have identified concerted loop motion around the CD4 binding site in gp120 upon binding (Hsu and Bonvin, 2004).

The enthalpic change upon gp120/CD4 complex formation, calculated as the difference in protein-protein plus protein-solvent potential energy between the simulations of the complex and of the separate proteins (-130 kJ mol$^{-1}$ at 300K), is of the same order of magnitude as the experimentally determined value (-62 kJ mol$^{-1}$ at 310K) (Myszka *et al.*, 2000). Encouraged by the qualitative agreement between the theoretical and experimental binding enthalpies, we aimed at extracting the binding entropy from our simulations and at gaining insight into CD4 binding-induced conformational changes in gp120 that are related to entropy changes.

The estimation of configurational entropy from molecular dynamics trajectories was first proposed by Karplus and Kushick using a quasi-harmonic method (Karplus and Kushick, 1981). The difference in configurational entropy between two molecular conformations *a* and *b* can be estimated as $\Delta S = k_B/2$ ln(det$\sigma_a$/det$\sigma_b$), where $\sigma_a$ and $\sigma_b$ are the covariance matrices of atomic positional fluctuations of the two conformers and $k_B$ is Boltzmann's constant. The method was formulated in terms of internal (non-Cartesian) coordinates, which made it less easily applicable. This approach was extended and applied to various biomolecular systems (Di Nola *et al.*, 1984; Edholm and Berendsen, 1984; Levy *et al.*, 1984). A decade later, Schlitter introduced a heuristic formula, based on Cartesian coordinates, to compute an upper bound to the *absolute entropy* of a molecule from a simulation trajectory (Schlitter, 1993). Calculation of the absolute molecular entropy would require a complete translational and rotational sampling. This is not yet reachable for proteins with the current computation time scale of nanoseconds. Instead, an assessment of the *configurational entropy* can be obtained from a MD trajectory based on the covariance matrix $\sigma$ of the Cartesian atom-positional fluctuations after elimination of translational and overall rotational motion by atom-positional least-squares fitting of molecular trajectory structures onto each other. This method was successfully tested for biomolecular simulations of peptide folding (Schäfer *et al.*, 2000; Schäfer *et al.*, 2001) and applied to simulations of protein molten globule states (Schäfer *et al.*, 2002). Recently Andricioaei and Karplus revised the quasi-harmonic approach to allow for the use of Cartesian coordinates (Andricioaei and Karplus, 2001).

One advantage of approaches based on the covariance matrix of atomic fluctuation is the possibility to compute this quantity for different subsets of atoms or degrees of freedom. One can for example resolve the entropic contributions of hydrophilic and hydrophobic residues to protein-protein binding, which is experimentally impossible. It should be noted though that entropies originating from different degrees of freedom need not be *additive* and that a decomposition will neglect correlation of motions along different degrees of freedom. Such analysis nonetheless enables us in the current study to estimate the entropic contribution of subsets of atoms or degrees of freedom of interest down to residue and atomic levels. Comparison of the configurational entropies of gp120 and the CD4 D1 domain (denoted CD4 for simplicity in the following) in the free and bound states suggests that the formation of intermolecular hydrogen bonds and hydrophobic contacts contributes most to the entropy changes. In line with previous postulates based on thermodynamic data (Myszka *et al.*, 2000; Kwong *et al.*, 2002; Xiang *et al.*, 2002), the

conformational rearrangement in the "bridging sheet" of gp120 upon CD4 binding is accompanied by a significant loss of entropy. The large-amplitude relocalisation of the V3 loop of gp120 upon binding is, however, free from any substantial entropic cost.

## Methods

Three simulations of 10 ns at 300K, 1 atm, were carried out for three different systems in explicit solvent: free HIV-1 gp120 with the truncated LAI strain sequence (SWISS-PROT accession number P03377; 346 residues and 29047 water molecules), free CD4 (99 residues of the D1 domain and 7768 water molecules) and their complex (445 residues and 30242 water molecules). Core gp120 is defined as the construct present in the crystal structure with the truncated V1-V3 loops and N- and C-termini. The V4 loop (398-409), for which electron density is missing in all crystal structures, and the V3 loop, which has been proposed to undergo significant rearrangements upon CD4 binding, were modelled onto the gp120 core, together with the six core mutations of the LAI strain (98.1% sequence identity with HxBc2, PDB entry number 1G9M) (for details see **Chapter 5**; Hsu and Bonvin 2004).

The GROMACS programme package (Lindahl *et al.*, 2001) was used for the MD simulations with the GROMOS96 43A1 force field (van Gunsteren *et al.*, 1996; Daura *et al.*, 1998) and the simple point charge (SPC) water model (Berendsen *et al.*, 1981) with rectangular periodic boxes with a 1.4 nm solute-wall minimum distance. Non-bonded interactions were calculated using twin range cutoffs of 0.8 and 1.4 nm. Long range electrostatic interactions beyond the cutoff were treated with a generalised reaction field model (Tironi *et al.*, 1995) using a dielectric constant of 54. For further simulation details we refer to (Hsu and Bonvin, 2004). Owing to the large system sizes *ca* 2 ns was required to reach equilibrium. The entropy analysis was therefore performed on the last 8 ns of each simulation, using molecular configurations that are 2 ps apart.

Schlitter's formula was used for the configurational entropy calculation, which yields an upper bound to the true entropy $S_{true}$,

$$S_{true} < S = \frac{1}{2} k_B \ln\det\left[1 + \frac{k_B T e^2}{\hbar^2} \mathbf{M}\sigma\right]$$

(1)

where $k_B$ is Boltzmann's constant, $T$ the absolute temperature, $e$ Euler's number, $\hbar$ the Plank's constant divided by $2\pi$, $\mathbf{M}$ the mass matrix that holds on the diagonal the masses belonging to the atomic Cartesian degrees of freedom, and $\sigma$ the covariance matrix of atom-positional fluctuations. The elements of $\sigma$ are given by

$$\sigma_{ij} = <(x_i - <x_i>)(x_j - <x_j>)>$$

(2)

where $x_i$ are the Cartesian coordinates after least-squares superposition of the trajectory configurations with respect to a particular subset of atoms. Although the choice of this set of atoms will determine which motions are excluded as translational and rotational ones, and so will influence the calculated internal configurational entropy, it has little effect for relative comparisons.

Two sets of atoms were used in the superposition of molecular configurations: (i) The backbone atoms (N, Cα and C') of the most stable structural elements of both proteins, the first α-helix of gp120 (residues 17-31) and part of the β-sheet of CD4 consisting of residues 26–30 and 82–86 (see Figure 1).
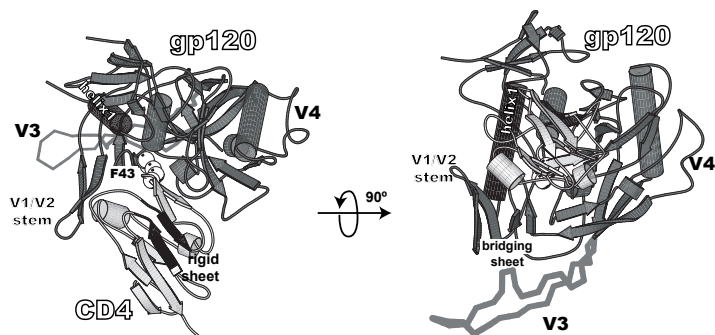
**Figure 1.** Structure of the gp120/CD4 complex. Gp120 (core) and CD4 (D1 domain) are coloured dark and light grey, respectively. CD4–F43 is shown in white spheres to indicate the position of the binding cavity. The four β-strands of the bridging sheet that link the inner and outer domains are indicated. The secondary structure elements that are used for superposition prior to the configurational entropy calculation are coloured black (for CD4 β-strands, β3: residues 26-30 and β8: residues 82-86; for gp120, α helix 1: residues 17-31). The structure corresponds to a simulated configuration taken at 6 ns, representative of the most populated V3 conformation (grey backbone Cα trace). The figure was generated using Molscript (Kraulis, 1991).

The set of atoms is indicated by the code *2nd*. (ii) The backbone (N, Cα and C') atoms of individual residues were used for superposition of trajectory configurations when calculating the entropy per residue (code: *fir*).

The covariance matrix and thus the configurational entropy was calculated for three different sets of atoms (see Table 1): (i) The gp120 molecule without the V3 loop and the CD4 molecule, both using all atoms (code: *all*) or using only backbone (N, Cα and C') atoms (code: *bb*), see Figure 4. (ii) Segments or regions that are directly involved in intermolecular interactions or that are indirectly affected upon binding; residue F43 of CD4 (code: *F43*) and the residues of gp120 forming a cavity to hold F43 (code: *F43cav*); residues of gp120 that form intermolecular hydrogen bonds (code: *interHb*) and their counterparts in CD4 (code: *interHb*), see Figure 5; the bridging sheet (code: *bridge*) and the V3 loop (code: *V3*) of gp120, see Figure 6. Detailed lists of residues are given in Table 1. (iii) The atoms of each residue (code: *res*), see Figure 7.

**Table 1.** Code definitions of atom sets used for configurational entropy calculations

| Set code | Molecule | Residue number | # of atoms |
|---|---|---|---|
| *all* | gp120 | 83–296, 331–492 | 3101 |
| | CD4 | 1–99 | 1021 |
| *bb* | gp120 | 83–296, 331–492 | 936 |
| | CD4 | 1–99 | 297 |
| *F43* | CD4 | F43 | 17 |
| *F43cav* | gp120 | E370, N425, W427, G473 | 47 |
| *interHb* | gp120 | K97, C126, V127, A129, S131, N280, A281, S365, G366, G367, D368, Q42, E429, V430, G473, D474 | 130 |
| | CD4 | K22, Q25, K29, K35, S42, F43, L44, K46, D53, R59, Q64, N66, E87, D88 | 166 |
| *bridge* | gp120 | 120–124, 198–202, 422–426, 431–435 | 198 |
| *V3* | gp120 | 297–330 | 359 |

We use the following notation for the entropy of a **set** (*all, bb*) of atoms of a **mol**ecule (*gp120, CD4, complex*) calculated from a **traj**ectory (*free-gp120, free-CD4, complex*) after superposition of the trajectory structures onto the first structure based on **fit**ting a particular set of atoms (*2nd, fir*),

$$S^{fit}_{mol,set}(traj) \tag{3}$$

The entropy difference between the bound (*b*) and free (*f*) states of CD4 is then

$$\Delta S^{2nd,b-f}_{CD4,all} \equiv S^{2nd}_{CD4,all}(complex) - S^{2nd}_{CD4,all}(CD4) \tag{4}$$

and likewise for gp120 instead of CD4.

In order to assess the degree of overlap between the configurational spaces sampled in two simulations, one may append one trajectory to the other trajectory and compute the development of the configurational entropy $S$ with time, see Figure 2. Denoting the trajectories by $I$ and $II$ we then obtain

$$S^{fit}_{mol,set}(traj\ I,\ traj\ II) \tag{5}$$

or alternatively

$$S^{fit}_{mol,set}(traj\ II,\ traj\ I) \tag{6}$$

depending on the sequence of appending.

Configurational entropy analysis of appended or combined trajectories can provide insight into the relative size and overlap of the conformational space sampled by a pair of single trajectories. For a molecule existing in two different states but with similar atom-positional fluctuations, similar configurational entropies are expected for the two states. Little information can be extracted by comparing the entropy of the two states. By combining the trajectories of these states, however, more information can be obtained. One can distinguish three cases (Figure 2):

**A.** A stepwise increase in configurational entropy after appending a second trajectory that has no or a small overlap in conformational space with the first trajectory (Figure 2A);

**B.** A rather smooth continuation of the configurational entropy build-up profile without observable change or perturbation after appending a second trajectory that has a large conformational overlap with the first one (Figure 2B);

**C.** A decrease in conformational entropy after appending a second trajectory that samples a much smaller conformational space than the first one and is contained in the conformational space of the first trajectory (Figure 2C).

These three situations, denoted cases A, B and C, will be used in the Result sections. Analysis of combined trajectories can yield information about the difference in the conformational spaces visited by the two trajectories, *e.g.,* of bound and unbound states, whereas analysis of the single trajectories only provides information about the extent of the respective conformational spaces, and not about their degree of overlap.
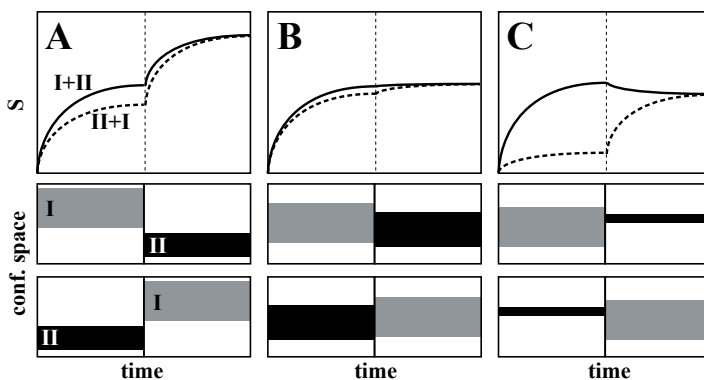
**Figure 2.** Schematic of three possible results of the configurational entropy analysis of the combined trajectory of two simulations I and II: **A** the two trajectories I and II sample different regions of conformational space, **B** the two trajectories I and II sample largely overlapping regions of conformational space and **C** the two trajectories overlap but one (II) only samples a small subspace of the conformational space of the other (I). For an explanation we refer to the Method section. *Top*: Configurational entropy build-up curves as a function of the length of the combined MD trajectory. The solid line corresponds to the configurational entropy of the I+II combined trajectory and the dashed line corresponds to the entropy obtained from the reversely combined trajectory, II+I, where I and II indicate the single trajectories. The conformational space sampling of the trajectories in the two combinations (I+II and II+I) is indicated in the lower panels.

The coverage of the conformational space of the second trajectory segment with respect to the first segment of a combined trajectory can be quantitatively measured by the difference between $S$ at the end of the second segment of the combined trajectory build-up curve and $S$ at the end of the first segment. For example, the coverage of the free trajectory with respect to the complex trajectory is

$$\Delta S_{mol,set}^{fit,c+f} \equiv S_{mol,set}^{fit}(complex,\ mol) - S_{mol,set}^{fit}(complex) \tag{7}$$

and the coverage of the complex trajectory with respect to the free trajectory is

$$\Delta S_{mol,set}^{fit,f+c} \equiv S_{mol,set}^{fit}(mol,\ complex) - S_{mol,set}^{fit}(mol). \tag{8}$$

## Results

### Configurational entropy of gp120 and CD4

The configurational entropy of various sets of atoms of the free gp120 and CD4 molecules and their complex are shown in Figures 4–6. Unlike other structural parameters such as the atom-positional root-mean-square deviations (RMSD) of the trajectory structures from the starting structure, which seem to reach equilibrium after 2ns for CD4 and 4 ns for gp120 (Figure 3), only the backbone configurational entropy starts to level off after *ca* 4ns (Figures 4A and B); the all atom values are still in the build-up phase throughout the 8 ns trajectories (Figures 4C and D). Entropy values obtained from the different

simulations are given in Table 2. Unexpectedly, the configurational entropies of both backbone and all atoms of gp120 (Figures 4A and C) in the complex form (dashed lines till 8 ns) towards the 8 ns time point are slightly higher ( $\Delta S_{gp120,bb}^{2nd,b-f}$ = 0.3 kJ K$^{-1}$mol$^{-1}$ and $\Delta S_{gp120,all}^{2nd,b-f}$ = 0.7 kJ K$^{-1}$mol$^{-1}$) than those in the free form (solid lines till 8 ns). This increase, however, only amounts to less than 2% of the respective absolute value. The difference obtained from appending the free and complex trajectories ( $\Delta S_{gp120}^{2nd,f+c}$ and $\Delta S_{gp120}^{2nd,c+f}$ , see Equations 8 and 7) are much larger (2.0 and 1.7 kJ K$^{-1}$mol$^{-1}$ for the backbone atoms and 11.1 and 10.4 kJ K$^{-1}$mol$^{-1}$ for all atoms).

**Table 2.** Comparisons of the all-atom and intra-residue configurational entropy (kJ K$^{-1}$ mol$^{-1}$) and their relative difference (%) for the molecules gp120 and CD4[a]

| Molecule (*mol*) | gp120 | | | CD4 | | |
|---|---|---|---|---|---|---|
| Fitting atom set (*fit*) | *2nd* | *fir* | | *2nd* | *fir* | |
| Entropy of atom set (*set*) | *all* [b] | *res* [c] | [d] | *all* [b] | *res* [c] | [d] |
| *Configurational entropies from single trajectories* | | | | | | |
| $S_{mol,set}^{fit}(free)$ | 49.4 | 84.3 | 71 | 22.4 | 29.8 | 33 |
| $S_{mol,set}^{fit}(complex)$ | 50.1 | 85.6 | 71 | 21.6 | 29.4 | 36 |
| $\Delta S_{mol,set}^{fit,b-f}$ | 0.7 | 1.3 | | -0.8 | -0.4 | |
| *Configurational entropies from combined trajectories* | | | | | | |
| $S_{mol,set}^{fit}(free,complex)$ | 60.5 | 92.6 | 53 | 23.9 | 31.1 | 30 |
| $\Delta S_{mol,set}^{fir,f+c}$ | 11.1 | 8.3 | | 1.5 | 1.3 | |
| $\Delta S_{mol,set}^{fit,c+f}$ | 10.4 | 7.0 | | 2.3 | 1.7 | |

**a.** The various quantities are defined in the Methods section, Equations 1–8.
**b.** All atom configurational entropies.
**c.** Sum of the intra-residue configurational entropies.
**d.** Relative differences calculated as [ $S_{mol,res}^{fir}(traj)$ – $S_{mol,all}^{2nd}(traj)$ ]/ $S_{mol,all}^{2nd}(traj)$ .

Due to the large number of atoms of gp120 the conformational sampling is not yet complete within the 8 ns simulation. The crossing over of the complex and free build-up curves at around 6 ns (Figures 4A and C) is due to a levelling off of the curve for the free form, which suggests that convergence is near for the free molecule. Appending the free and complex trajectories of CD4 gives rise to a small stepwise increase of the entropy build-up of the backbone atoms (Figure 4B at 8ns), regardless of the order of appending. This suggests that two slightly different conformational spaces are sampled in the free and complex forms of the molecule representing a situation of case A (Figure 2A). For the all atom analysis of CD4 (Figure 4D), on the other hand, a smooth continuation at 8 ns of the configurational entropy build-up is observed when the complex trajectory is appended to the free one (solid line) while a stepwise increase is visible when the trajectories are appended in the reverse order (dashed line). This illustrates case B (Figure 2B) where a subspace of the conformational space of the free molecule is sampled in the complex state.
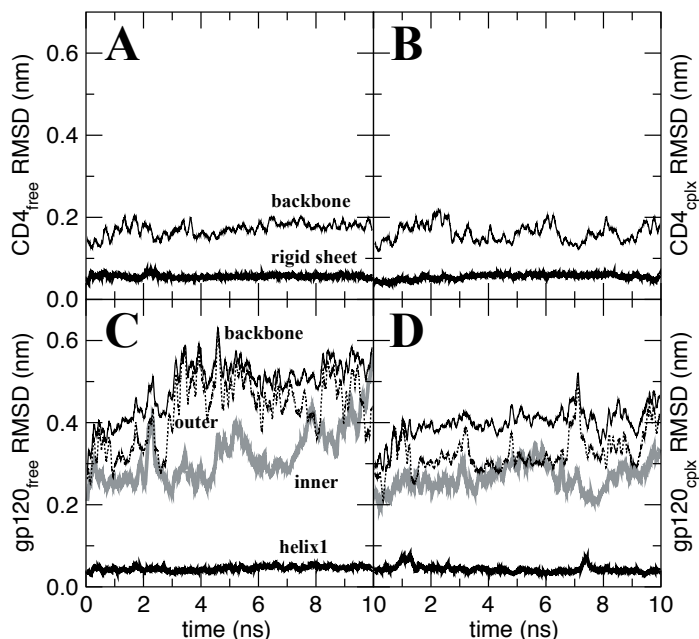
**Figure 3.** Atom–positional (backbone atoms, N, Cα and C') root–mean–square deviation (RMSD) of trajectory structures from the respective starting structure of CD4-D1 domain and gp120 (without the V3 loop) in the free and complex forms. **A & B** CD4 in the free and complex forms, respectively. **C & D** Gp120 in the free and complex forms, respectively. The RMSD values of the structural elements subjected to the least-squares superposition of backbone atoms are shown in thick solid lines (see Figure 1) and those of the backbone atoms (C', Cα and C') of all residues are shown in thin solid lines. The RMSD values of the inner and outer domains of gp120 are shown in grey and dashed lines, respectively. The RMSD values were smoothed for clarity using a 0.1 ns averaging window.

*Configurational entropy of interacting elements and CD4 binding-induced conformational changes in gp120*

The gp120/CD4 interface involves of two major interaction modes: (i) the unique "knob-and-socket" hydrophobic interaction between CD4-F43 (17 atoms) and the gp120 receptive cavity, including E370, N425, W427 and G473 (47 atoms); (ii) the stable intermolecular hydrogen bond network (166 and 130 atoms for CD4 and gp120, respectively; for details see Table 1). These intermolecular contacts restrict the mobility of the involved residues and, in turn, their configurational entropy. In the combined trajectory analysis (Figure 5), the entropy build–up curves show a substantial jump at the appending point when the trajectory of the free form is appended to that of the complex form (dashed line), indicating that the free molecule samples conformational space not visited in the complex simulation (case A; Figure 2A). Note that while the configurational entropy of the whole system is still in the early stage of equilibration (Figure 4), the entropy of most of the structural elements involved in the gp120/CD4 interface starts to level off after 4 ns as a result of their smaller number of atoms.
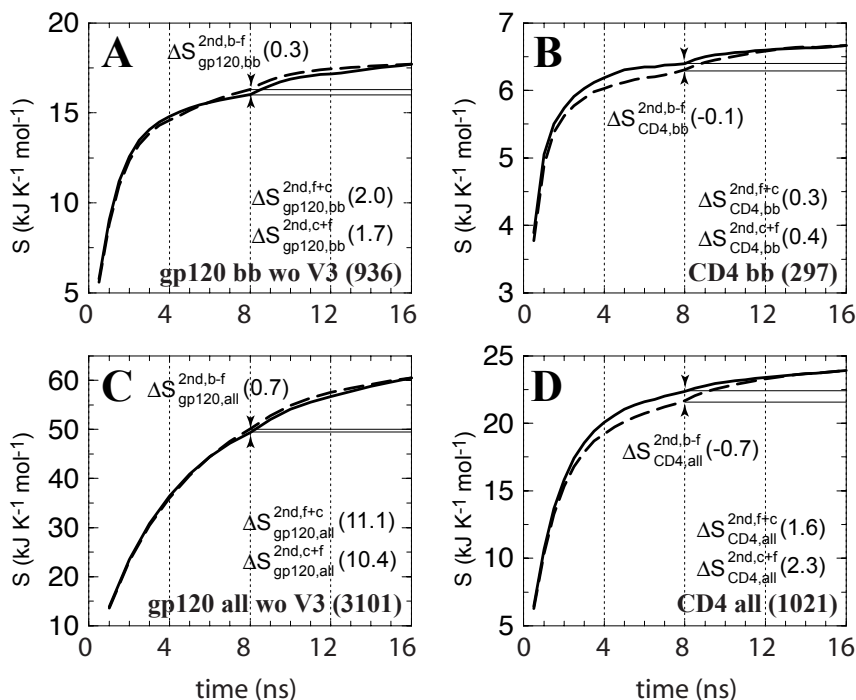
**Figure 4.** Configurational entropy of gp120 and CD4-D1 domain. **A & C** Backbone and all atoms of gp120 excluding the V3 loop. **B & D** Backbone and all atoms of CD4. The build-up curves of the free+complex combined trajectory is shown in solid lines and the ones of the complex+free combined trajectory in dashed lines. The numbers of atoms used for the calculations are indicated between parentheses at the lower right of each panel. All covariance matrices were generated after positional least-squares fitting of backbone atoms of the β sheet of CD4 and of α helix 1 of gp120, respectively (see Figure 1). The configurational entropy build-up was calculated every 0.5 ns except for the all atom analysis of gp120 (1 ns/step). The configurational entropy difference between the free and complex forms $\Delta S_{mol,set}^{2nd,b-f}$ (Equation 4) is indicated at the midpoint of the curves (8 ns). The configurational entropy gains after combining trajectories, $\Delta S_{mol,set}^{2nd,f+c}$ and $\Delta S_{mol,set}^{2nd,c+f}$, Equations 8 and 7, are calculated by subtracting the midpoint values indicated by the horizontal lines from the end point values (see Methods).

The loss of entropy stemming from the rigidification of the residues in gp120 forming stable hydrophobic contacts to CD4-F43 ( $\Delta S_{gp120,F43cav}^{2nd,b-f}$ = –0.12 kJ K$^{-1}$ mol$^{-1}$; see Figure 5A) is of the same size as the corresponding loss of entropy in CD4-F43 ( $\Delta S_{CD4,F43}^{2nd,b-f}$ = –0.13 kJ K$^{-1}$ mol$^{-1}$; see Figure 5B). Appending the complex trajectory of gp120 to the free one gives rise to a small increase ( $\Delta S_{CD4,F43}^{2nd,f+c}$ = 0.03 kJ K$^{-1}$ mol$^{-1}$) indicating that the complex form samples slightly different parts of conformational space (Figure 5A). Interestingly, for CD4–F43, the corresponding value is negative, –0.01 kJ K$^{-1}$ mol$^{-1}$ (solid line in Figure 5B). This would only occur when the averaged atom–positional fluctuations, *i.e.*, the determinant of the covariance matrix, are reduced due to the addition of the second trajectory segment (case C, Figure 2C). In other words, the second MD trajectory segment covers a small subspace of the conformational space sampled by the first segment and hence reduces the overall size of the elements of the covariance matrix σ through averaging over the combined trajectories. As an illustration of this, the root-mean-square fluctuations (RMSF) of the phenyl ring of CD4 Phe43 after least-squares fitting on
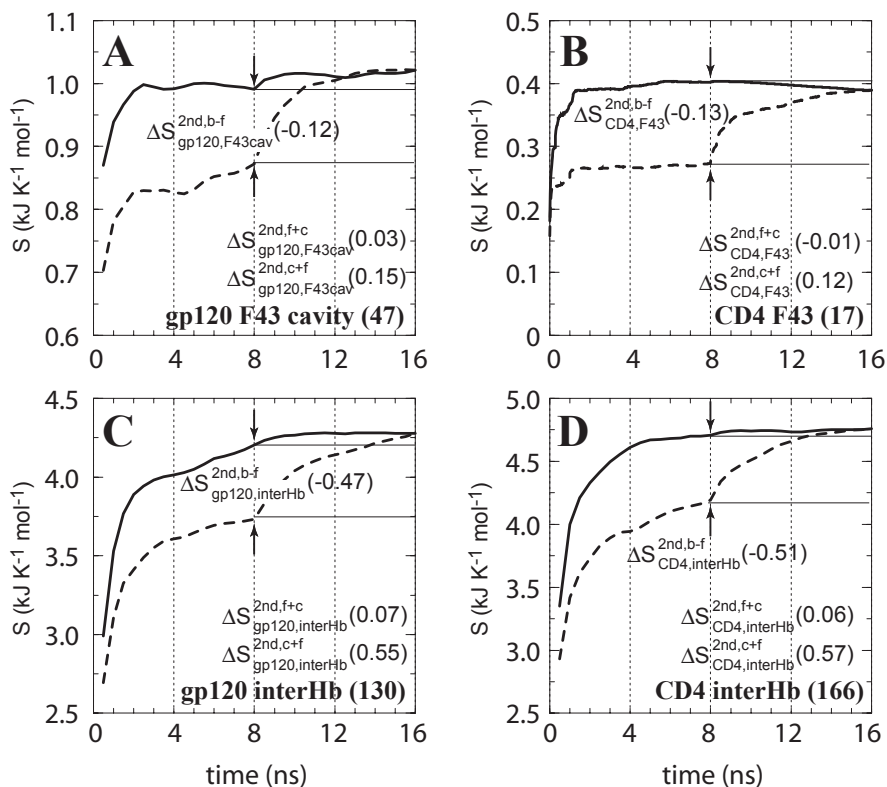
**Figure 5.** Configurational entropy of the structural elements involved in gp120/CD4 intermolecular interactions. **A** Receptive cavity of gp120 formed by residues that have hydrophobic contacts with CD4-F43. **B** CD4-F43. **C & D** Residues of gp120 and CD4 that are involved in a stable intermolecular hydrogen bond network. The configurational entropy build-up was calculated every 0.1 ns because of the fast equilibration within 1 ns. See for further explanations the legend of Figure 4.

all heavy atoms of CD4 decreases from $0.30\pm0.07$ nm in the free form to $0.08\pm0.01$ nm in the complex form. Moreover, all Phe43 side-chain conformations in the complex are contained within the ensemble of conformations of the free form.

In addition to the changes directly related to the intermolecular contact, CD4 binding also induces conformational changes away from the binding interface that are crucial for the subsequent events of viral entry into the host cell. It is generally accepted from thermodynamic and biochemical data that the β sheet of gp120 that bridges the inner and outer domain of the gp120 core, the "bridging sheet", is only fully formed and stabilised upon complexation with CD4 (Figure 1) and that this stabilisation contributes a sizeable entropic loss and/or enthalpic gain (Jardetzky, 2002; Kwong *et al.*, 2002; Xiang *et al.*, 2002). Concomitantly, the third hypervariable loop (V3) undergoes a rearrangement to a somewhat *different* conformation leading to accessibility of the epitope for co-receptor binding. Our previous MD study revealed that lid closure motions upon CD4 binding are accompanied by concerted structural changes

leading to a substantial increase of rigidity of the bridging sheet and a large-amplitude translocation of the V3 loop (**Chapter 5**; Hsu and Bonvin, 2004). The corresponding changes in entropy are shown in Figure 6. The bridging sheet shows a clear entropy difference between the free and complex forms of gp120 with a loss of entropy of -0.23 kJ K$^{-1}$ mol$^{-1}$. Not only is its flexibility reduced, but its configuration is also altered. This can be concluded from the subsequent change after appending the trajectories of the complex and of the free form to each other, which is consistent with the previously defined case A (Figure 2A). The entropy of the V3 loop in the CD4-bound form initially builds up slower than in the free form but crosses the latter curve in the last 1.5 ns resulting in a final increase of 0.24 kJ K$^{-1}$ mol$^{-1}$ (Figure 6B). In addition, the V3 loop shows an increase in entropy when the two trajectories are combined, irrespective of the order of appending. These observations suggest that the V3 loop undergoes major changes in conformation and/or localisation upon CD4 binding, while its intrinsic flexibility persists, or even possibly increases.

*Configurational entropy changes per residue*

Changes in conformational entropy upon binding can also be calculated per residue. For this, the atom-positional least-squares fitting of trajectory structures was performed using backbone atoms (N, C$\alpha$ and C') of the individual residues to exclude from the entropy contributions of collective motions of larger structural segments. The intra-residue conformational entropy obtained in this way provides, therefore, only information about side chain motions. The intra-residue configurational entropy per residue was calculated using Equation 1 and then normalised by dividing it by the number of atoms of each residue. Intra-residue entropy differences between the free and complex forms $\Delta S_{mol,all}^{fir,b-f}$ (Equation 4) are plotted as a function of residue number in Figure 7.

On average, CD4 shows an intra-residue entropy decrease of $-0.5$ kJ K$^{-1}$ mol$^{-1}$ per atom whereas gp120 shows an unexpected increase of 0.3 kJ K$^{-1}$ mol$^{-1}$. Most of the residues of CD4 and gp120 that are involved in the intermolecular hydrogen bond network (filled bars in Figure 7) show reductions in intra-residue configurational entropy upon complexation, except for K22 and D53 in CD4, and C126, S365 and E429 in gp120. The unexpected entropy increase of these five residues is probably due to their
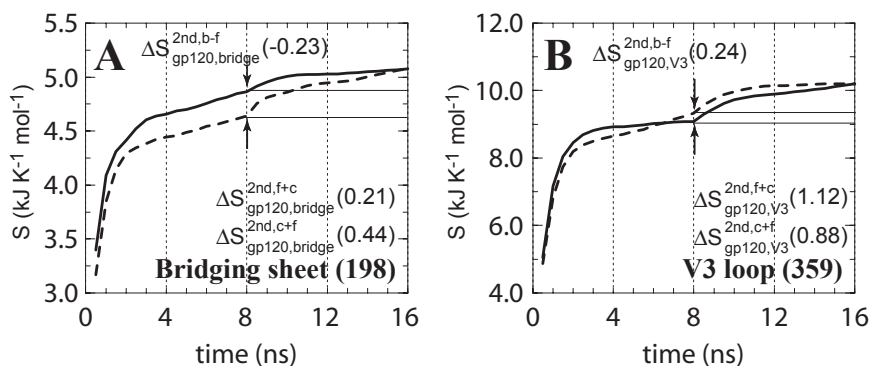


**Figure 6.** Configurational entropy of the structural elements in gp120 that experience conformational changes upon CD4 binding. **A** The bridging sheet. **B** The V3 loop (see Figure 1). The configurational entropy build-up was calculated every 0.5 ns. See for further explanations the legend of Figure 4.

intrinsic flexibility that is reflected in their large B-factors in the crystal structure (Kwong *et al.*, 1998). In addition, our previous MD analysis of the intermolecular hydrogen bonding shows that all these residues, but CD4-K22, form only marginally stable intermolecular hydrogen bonds. Despite its high occurrence (78%), the salt bridge between CD4-K22(Nζ) and gp120-E429(Cδ) located at the edge of the interface of the molecular complex might be subject to less structural limitations than those in the centre of the interface. Furthermore, the relatively long side chains of both residues might also tolerate a higher degree of flexibility despite the presence of a salt bridge.
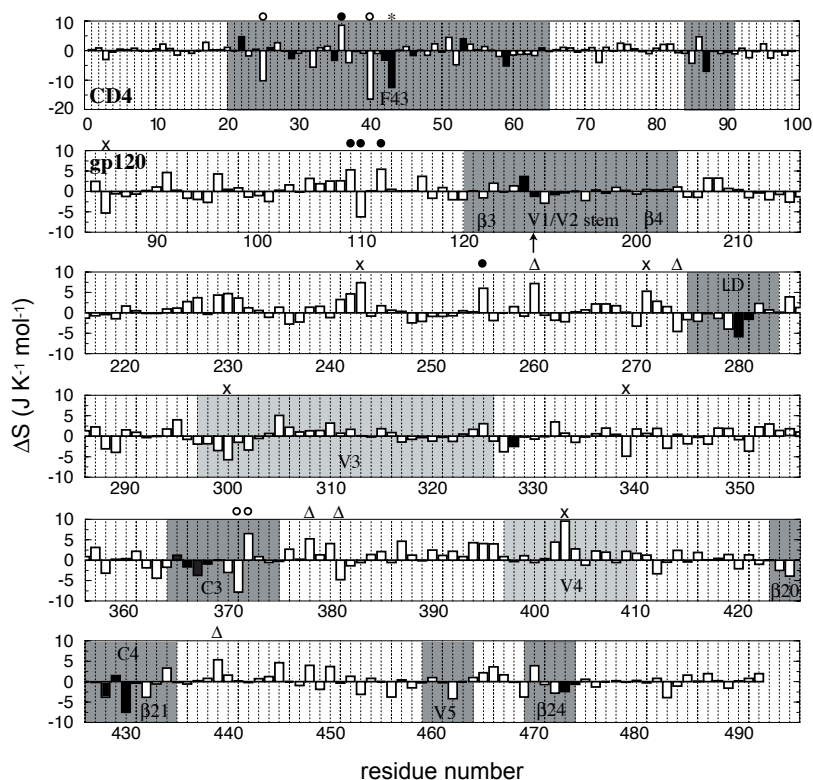


**Figure 7.** Differences in intra-residue configurational entropy in J K$^{-1}$ mol$^{-1}$ between complex and free MD simulations as a function of residue number. The intra-residue configurational entropies were calculated after positional least-squares fitting of backbone atoms of the respective residues. Top row: CD4-D1 domain. The sequence numbering of gp120 follows the HxBc2 construct with an arrow indicating the missing V1/V2 loop (residues 128-194). Residues involved in the stable intermolecular hydrogen bond network are shown in filled bars. CD4-F43 is highlighted with an asterisk. Residues not involved in stable intermolecular hydrogen bonds that exhibit entropy changes larger than two times the standard deviation from the average value are highlighted with the following symbols indicative of their minimum intermolecular distance d: open circle (d<0.5 nm), filled circle (0.5 nm < d < 1 nm), triangle (1 nm < d < 2 nm) and cross (d > 2 nm). The intermolecular distance is defined as the minimum interatomic distance from any atom of the residue of interest to any atom of the other molecule in the structure taken at 6 ns, which is well-equilibrated and is representative of the most populated V3 conformation (see Figure 1). Regions in gp120 and CD4 that are involved in intermolecular contacts are shaded in dark grey. The modelled V3 and V4 loops in gp120 are shaded in light grey.

In line with the previous analysis of the functional elements, CD4–F43 shows a large intra–residue entropy reduction upon complexation (Figure 7) due to the steric restriction imposed by the receptive cavity of gp120. The neighbouring residue Q40, which lies at the centre of the complex interface, shows the largest reduction amongst all residues  (top row in Figure 7). Although less confined than CD4–F43, it is in close contact (less than 0.5 nm inter-atomic distance) with T283, G473 and D474 of gp120, of which the latter two were found to form intermolecular hydrogen bonds with F43 and Q25 of CD4, respectively (Hsu and Bonvin, 2004). In gp120, regions that show concerted lid closure motions upon CD4 binding, namely £D, C3, C4 and V5, display substantial intra–residue entropy reductions upon complexation especially for those residues that are involved in intermolecular hydrogen bonding (Figure 7). Note that the outer half of the bridging sheet ($\beta$3 and $\beta$4) and the tip of the hairpin structure of the V3 loop (residues 297–330) do not exhibit significant entropy changes, most likely due to the lack of direct contact with CD4. This illustrates that changes in conformation do not necessarily involve a change in entropy or extent of motion.

In addition to the residues that are directly involved in intermolecular contacts or are in close proximity to the intermolecular interface, there are a few residues in gp120 that show large intra–residue entropy changes upon complexation (see Figure 7). Some residues show positive intra–residue entropy change upon complexation. Several are located in the putative trimerisation interface, S243, C378 and I439, and two of them, V255 and L260, are located in the $\beta$-turn that connects the inner and outer domains of gp120. This indicates that the intermolecular interactions between gp120 and CD4, while decreasing the intra–residue entropy of residues in direct contact, can induce intra–residue entropy changes at the inter–domain interface within the gp120 monomer and at the putative trimeric gp120 interface.

## Discussion

Sufficient conformational sampling is a prerequisite for a reliable estimation of the conformational entropy. The required simulation time is correlated to the size of the system, *i.e.*, the number of atoms taken into account in the calculation. For a moderate system size such *as* the entire CD4–D1 domain with 1021 atoms, convergence of conformational entropy of both the free and bound states is only observed in the later stage of the 8 ns trajectory segments. Nevertheless, the stabilising effect of complexation on backbone and side chains can be unambiguously identified when comparing the build–up curves of the two states (Figures 4B and D). It is, however, not obvious to explain the slight, unexpected gain in entropy upon complexation for gp120 (Figure 4A). The conformational sampling requires more time than for CD4. When considering all protein atoms, the number of atoms in gp120 becomes too large to obtain sufficient sampling in the time scale of our simulation (10ns); the build–up curves have not yet converged (Figure 4C). In contrast to the compact CD4, gp120 undergoes a number of significant loop contractions and translocations upon CD4 binding, which occur on a more extended time scale than the high frequency local fluctuations. Complete sampling of these movements would require much longer simulations.

The conformational entropy change of gp120/CD4 complexation qualitatively agrees with the experimental value despite the fact that the contribution from bulk solvent was neglected in our analysis. Assessment of the complete system was not yet feasible due to insufficient sampling, especially for gp120. Nevertheless, by considering functionally important sets of atoms, a *localised* instead of *global* analysis

provided insights into the entropic contribution of various degrees of freedom. The intermolecular-hydrogen bond network and the insertion of CD4-F43 into its receptive cavity seem to predominantly determine the large entropy loss upon complexation. The bridging sheet of gp120 also plays a key role in the entropy change, as has been proposed experimentally (Myszka *et al.*, 2000; Xiang *et al.*, 2002). While, in some cases, conventional structural parameters such as atom-positional fluctuations may fail to identify thermodynamic differences between two states, the conformational entropy analysis of combined trajectories can provide a complementary way of evaluating spatial distributions and their statistical weight. This is essentially equivalent to the clustering approach, which was previously proposed to assess equilibration and convergence of biomolecules simulations (Smith *et al.*, 2002). With the rapid advance in computing power and methodology, we are hopeful that a thorough description of the thermodynamics of such complex systems can be ultimately achieved *via* computer simulations and will meet experimental data in the near future.

# A stable β-sheet fold can substitute for a disulphide bond in HIV-1 gp120[#]

Rogier W. Sanders[1‡], Shang-Te D. Hsu[2‡], Eelco van Anken[3‡], Marije Liscaljet[3],
Sonja Tillemans[3], Els Busser[1], Martijn Dankers[1], Ineke Braakman[3],
Alexandre M. J. J. Bonvin[2] and Ben Berkhout[1]

*[1]Department of Human Retrovirology, Academic Medical Center, University of Amsterdam,*
*1105 AZ Amsterdam, The Netherlands*
*[2]Department of NMR Spectroscopy and [3]Department of Bio-Organic Chemistry-1,*
*Bijvoet Center for Biomolecular Research, Utrecht University,*
*3584 CH Utrecht, The Netherlands*

## Abstract

**Protein folding is studied at several levels. First, the formation of secondary structure elements such as α-helices and β-sheets can be investigated. Second, the acquisition of tertiary structure and disulphide bonds during oxidative folding in the endoplasmic reticulum (ER) *in vivo* can be studied. Third, one can analyze the results of a completed folding process as a protein is secreted or expressed at the cell surface, the ultimate test being the analysis of function. The correlates of these different levels of protein folding are mostly unclear. We generated an HIV-1 gp120 variant through virus evolution that is functional despite the lack of the disulphide bond at the base of the V4 domain that is otherwise required for virus replication and gp120 activity. Biochemical and computational analyses indicate that virus replication is restored through the improvement of local hydrogen bonding and stabilisation of a local β-sheet fold. This study provides proof that a critically important disulphide bond can be functionally replaced by an alternative protein structure motif. It also provides evidence for the proposal that local protein stability is an important factor in escape from ER quality control during protein biosynthesis. Furthermore, our data indicate that β-sheet preference is a determinant in directing protein stability and protein folding *in vivo* and that β-sheet rules deduced from experiments with small model proteins also hold for the intricate chaperone-assisted folding of a complicated glycoprotein such as gp120.**

## Introduction

The HIV-1 envelope glycoprotein complex (Env) mediates viral attachment and entry in susceptible cells. The surface subunit (SU; gp120) sequentially binds the CD4 receptor and either CCR5 or CXCR4 as a coreceptor (Wyatt and Sodroski, 1998; Poignard *et al.*, 2001). Subsequent conformational changes result in fusion of viral and cellular membranes, mediated by the transmembrane glycoprotein (TM; gp41) (Eckert and Kim, 2001; Gallo *et al.*, 2003). Env is synthesised as a gp160 precursor protein, which is cotranslationally translocated into the endoplasmic reticulum (ER). Here, Env acquires carbohydrate chains and disulphide bonds, it folds, trimerises and looses its leader peptide (Land *et al.*, 2003). Like folding of any other glycoprotein, gp160 folding is assisted by molecular chaperones. Gp160 transiently associates with the ER resident chaperones BiP, calnexin and calreticulin (Earl *et al.*, 1991; Otteken *et al.*, 1996; Knarr *et al.*, 1999). Subsequently, gp160 is transported to the Golgi complex where it is cleaved into gp120 and gp41, which stay associated non-covalently (Stein and Engleman, 1990; Moulard and Decroly, 2000). In addition, part of the ~30 carbohydrates are modified here (Leonard *et al.*, 1990).

Most proteins have the intrinsic capacity to acquire their unique 3D structure in a spontaneous and autonomous manner, depending only on the amino acid sequence and a native environment (Jaenicke, 1999). It was assumed that folding rules based on *in vitro* folding experiments with small model proteins could be applied to the description of protein folding in living cells. However, complicated proteins need assistance from chaperones to fold properly and there is no direct evidence that these simple folding rules also apply to chaperone–assisted protein folding in the ER.

We have previously studied the role of individual disulphide bonds in oxidative Env folding. Five out of ten disulphide bonds were dispensable for folding. Surprisingly, two were also largely dispensable for Env function and viral replication. The remaining five disulphide bonds were required for proper oxidative folding of Env in the ER. The advantage of using Env over other models for glycoprotein folding is that HIV-1 provides the possibility for protein evolution when a mutant protein is tested in replication competent HIV-1. We used virus-driven protein evolution to further characterise the significance and role of specific disulphide bonds. In the current study, we describe an escape variant of an Env mutant lacking the conserved disulphide bond at the base of the V4 domain. This disulphide bond excludes the V4 loop from the gp120 core and is required for proper oxidative folding of wt gp120 and virus replication. In the evolved variant, the role of the disulphide bond during folding is replaced by increased local hydrogen bonding within a β–sheet fold, which results in an escape from ER quality control and restored virus replication.

## Materials and Methods

### Cloning

The pRS1, pcDNA3-Env-gp120 and pLAI plasmids containing the appropriate mutations in the *env* gene were generated as described previously (van Anken, 2003). PCR–generated gp120 sequences from evolved viruses (see below) were cloned into the pRS1 shuttle vector using the BsaB1 and Nhe1 sites and subsequently cloned into the pLAI infectious molecular clone (Peden *et al.*, 1991) as SalI–BamHI fragments. NotI–XhoI fragments were subcloned into the pcDNA3 expression vector for use in folding experiments. Numbering of individual amino acids is based on the sequence of HXB2 gp160.

### Cells and transfections

HeLa cells (ATCC) and HT1080 cells were cultured in MEM (Life technologies) supplemented with 10% FCS (Hybond), penicillin (100 U/ml), streptomycin (100 μg/ml). Peripheral blood mononuclear cells (PBMCs) were isolated from buffy coats from healthy individuals by Ficoll–Isopaque density centrifugation. PBMCs were cultured for three days in RPMI medium (Life Technologies Ltd., Paisley, UK) supplemented with 10% FCS, penicillin (100 U/ml), streptomycin (100 μg/ml) and phytohemagglutinin (PHA; 5 μg/ml) and subsequently cultured without PHA, but with IL-2 (100 U/ml). SupT1 cells were cultured in RPMI medium supplemented with 10% FCS, penicillin and streptomycin. LuSIV cells were cultured in RPMI medium supplemented with 10% FCS, penicillin, streptomycin and hygromycin B (Roos *et al.*, 2000). C33A cervix carcinoma cells were maintained in DMEM (Life Technologies), supplemented with 10% FCS, penicillin and streptomycin, as previously described (Sanders *et al.*, 2002a). SupT1 and C33A cells were transfected with pLAI by electroporation and $Ca_3(PO_4)_2$ precipitation, respectively, as described previously (Das *et al.*, 1999).

### Viruses and infections

Virus stocks were produced by transfecting C33A cells with the appropriate pLAI constructs. The virus containing supernatant was harvested 3 days post-transfection, filtered and stored at –80°C and the virus concentration was quantitated by capsid CA–p24 ELISA as described previously (Jeeninga *et al.*, 2000). These values were used to normalise the amount of virus in subsequent infection experiments. Infection

experiments were performed as follows. $50 \times 10^3$ SupT1 T cells or PBMCs were infected with 20 or 500 ng CA–p24, respectively, of C33A-produced HIV-$1_{LAI}$ per well in a 24-well plate, and virus spread was measured for 14 days using CA–p24 ELISA.

### Virus evolution

For evolution experiments, SupT1 cells were transfected with 10 μg pLAI constructs by electroporation, and virus cultures were inspected regularly for the emergence of revertant viruses by CA–p24 ELISA and/ or the appearance of syncytia. At regular intervals, cells and filtered supernatant were stored at –80°C and virus was quantitated by CA–p24 ELISA. When a revertant virus was identified, DNA was extracted from infected cells (Das *et al.*, 1997) and proviral gp120 sequences were PCR–amplified with primers A (5'-GCTCCATGGCTTAGGGCAACATATATCTATG-3') and B (5'-GTCTCGAGATGCTGCTCC-3') and sequenced.

### Virus entry, infectivity and neutralisation

LuSIV cells, stably transfected with an LTR–luciferase construct (Roos *et al.*, 2000), were infected with 200 ng CA–p24/$300 \times 10^3$ cells/ml in a 48 well plate. Cells were maintained in the presence of 200 nM saquinavir to prevent additional rounds of virus replication. Luciferase activity was measured after 48 hrs. Neutralisation experiments were performed similarly, but virus was preincubated for 30 min at room temperature, with the appropriate concentration of either monoclonal antibody 2G12 (Trkola *et al.*, 1996; Sanders *et al.*, 2002b; Scanlan *et al.*, 2002; Calarese *et al.*, 2003) or IgG1b12 (Burton *et al.*, 1994; Saphire *et al.*, 2001). The 50% tissue culture infectious dose (TCID$_{50}$) was determined by endpoint dilution.

### Quantitation of gp120 in cell, virion and supernatant fractions

C33A cells were transfected with 40 μg pLAI per T75 flask. Medium was refreshed at day one post-transfection. The culture supernatant was harvested at 3 days post-transfection, centrifuged and passed through a 0.45 μm filter to remove residual cells and debris. Cells were resuspended in 1.0 ml lysis buffer (50mM Tris (pH 7.4) 10mM EDTA, 100mM NaCl, 1% SDS). Virus particles were pelleted by ultracentrifugation (100.000 *g* for 45 min at 4°C) and resuspended in 0.5 ml lysis buffer. The virus free supernatant, containing gp120 shedded from the cell and virion surface, was concentrated using Amicon centrifugal filter units (Millipore, Bedford, MA) and SDS was added to a 1% end concentration.

Gp120 in cell, virion and supernatant fractions was measured as described previously (Moore and Ho, 1993; Sanders *et al.*, 2002b), with minor modifications. ELISA plates were coated overnight with sheep antibody D7324 (10 μg/ml; Aalto Bioreagents, Ratharnham, Dublin, Ireland), directed to the gp120 C5 region, in 0.1 M NaHCO$_3$. After blocking with 2% milk in Tris–buffered saline (TBS) for 30 min, gp120 was captured by incubation for 2 hr at room temperature. Recombinant HIV-$1_{LAI}$ gp120 (Progenics Pharmaceuticals, Inc. Tarrytown, NY) was used as a reference. Unbound gp120 was washed away with TBS and purified serum Ig from an HIV-1 positive individual (HIVIg) was added for 1.5 hr in 2% milk, 20% sheep serum (SS), 0.5% Tween-20. HIVIg binding was detected with alkaline phosphatase conjugated goat anti–human Fc (1:10000, Jackson Immunoresearch, West Grove, PA) in 2% milk, 20 % SS, 0.5% Tween-20. Detection of alkaline phosphatase activity was performed using AMPAK reagents (DAKO, Carpinteria, CA). The measured gp120 contents in cells, virus and supernatant were normalised for CA–p24.

## Env folding

For folding assays, mutant gp120 was expressed using a recombinant Vaccinia virus vector system (van Anken, 2003). Folding of gp120 mutants was analyzed by pulse-chase labelling and immunoprecipition with anti–Env sera as described (Land *et al.*, 2003; van Anken, 2003). Formation of disulphide bonds was assayed by SDS-PAGE mobility changes of deglycosylated, alkylated, non-reduced samples. Reduced samples were used to follow signal sequence cleavage.

## Molecular dynamics simulations

The starting structures were generated with SWISS-MODEL (Guex and Peitsch, 1997) using modelling templates as structural analogs (PDB entries 1G9M, 1GC1 for the HXB2 isolate  (Kwong *et al.*, 1998; Kwong *et al.*, 2000) and 1G9N for the YU2 isolate (Kwong *et al.*, 2000). The flexible N– and C–termini and the hypervariable V1, V2 and V3 loops, which are missing in the crystal structures, were not included to limit the sise of the simulations. The missing V4 loop was modellled by SWISS-MODEL. In addition to the wild–type sequence of HXB2 (wt), point mutations were introduced *in silico* by modifying the input primary sequences for model generations. Four variants were used in this study: the inactive mutant: C385A/C418A; the active revertant: C385A C418V T415I; the corresponding core sequence of LAI isolate; and the HXB2 core sequence with all cysteine residues replaced by alanines.

The GROMACS 3.0 molecular dynamics package (Lindahl *et al.*, 2001) was used with the GROMOS 43A1 force field (Daura *et al.*, 1998). Starting structures were individually solvated using the simple point charge (SPC) water model (Berendsen *et al.*, 1981) in periodic cubic boxes with a 1.4 nm solute–wall minimum distance. After a first steepest descent energy minimisation with positional restraints on the solute, chloride ions were introduced in all three systems to obtain an electro–neutralised system. The resulting systems are comprised of 317 amino acids and about 30,500 water molecules that give a total number of about 93,900 atoms. A second energy minimisation was then performed, followed by five successive 20 ps MD runs with decreasing positional restraint force constants on the solutes ($K_{posres}$ = 1000, 1000, 100, 10 and 0 kJ mol$^{-1}$ nm$^{-2}$) prior to the production runs.

The simulations were run for a period of 10 nanoseconds (ns) at 300K and one atm for all variants except for the core LAI gp120 (5 ns). Short 1 ns simulations at 400K and 1atm were performed starting from different time points of the 300K simulations (1, 1.5 and 2 ns) for all variants to assess their thermostability. Furthermore, 10 ns simulation at 400K and 1 atm were carried out for the wt, mutant and revertant core gp120s starting from configurations taken at 2 ns of the 300K simulations. For all simulations, solute, solvent and counterions were independently coupled to a reference temperature bath. The pressure was maintained by weakly coupling the system to an external pressure bath at one atmosphere (Berendsen *et al.*, 1984). Non-bonded interactions were calculated using twin range cutoffs of 0.8 and 1.4 nm. Long range electrostatic interactions beyond the cutoff were treated with the generalised reaction field model (Tironi *et al.*, 1995) using a dielectric constant of 54. A four femtoseconds (fs) integration time step was used for the integration of the equations of motions. The LINCS algorithm (Hess *et al.*, 1997) was used for bond length constraining in conjunction with dummy atoms for the aromatic rings and amino group in side chains (Feenstra *et al.*, 1999). The simulations required about 50 hours per nanosecond in parallel on four 1.3 GHz AMD CPUs.

# Results

*Evolution of gp120 lacking the conserved disulphide bond at the base of the V4*

In our previous work we found that five out of ten absolutely conserved disulphide bonds are essential for the oxidative folding of the HIV-1 Env protein. However, for some mutants that were folding incompetent we could still observe a minimal but reproducible infectivity when placed in a replicating virus, although not sufficient to cause a spreading infection (van Anken, 2003)†. One of these mutants was the C385A/C418A double mutant, lacking the disulphide bond at the base of the V4 variable loop (Figure 1A). This virus appeared to be a good candidate for protein evolution studies, with the aim of identifying and investigating escape routes that result in restoration of gp120 folding and virus replication. After prolonged culturing (80 days) on SupT1 T cells, we identified a replicating virus in a culture of the C385A/C418A mutant virus. Proviral *env* sequences were PCR-amplified and sequenced. Population sequencing revealed two reversions: a first-site pseudoreversion A418V and a second-site reversion at a nearby residue: T415I (Figure 1B). Thus, the wt cysteine at position 418 is not restored, which may be caused by the design of the mutant alanine codon, which requires at least two point mutation to become a cysteine codon. This combined with the fact that the original C385A substitution was still present implies that the disulphide bond at the base of the V4 is not restored. Sequencing of individual *env* clones revealed several with only the T415I reversion, implying that this mutation appeared first during the course of evolution (Figure 1). For simplicity we will hereafter refer to the respective variants as mutant (C385A/C418A), intermediate revertant (C385A/C418A T415I) and revertant (C385A C418V T415I). To establish whether the identified substitutions accounted for the revertant phenotype, the relevant *env* fragments were subcloned into a molecular clone of HIV-1$_{LAI}$. Virus stocks were produced by DNA transfection of non-susceptible C33A cells. SupT1 T cells were infected with wt, mutant and revertant viruses and virus spread was monitored by CA-p24 ELISA (Figure 2A). The C385A/C418A mutant virus did not cause a spreading infection (van Anken, 2003). The intermediate revertant (C385A/C418A T415I) replicated poorly, and revertant  (C385A C418V T415I) showed greatly improved replication, although it was still somewhat impaired compared to the wt virus. Similar results were obtained in primary cells, indicating that the revertant phenotype is not specific for the SupT1 T cell line that was used for the evolution experiment (Figure 2B). The differences in viral replication could not be attributed to differences in virus production, since the virus production, which is essentially Env-independent, was similar for wt, mutant and revertant viruses (Figure 2C). Infectivity measurements (TCID$_{50}$; fig. 2D) and single cycle viral entry experiments (Figure 2E) further corroborated the replication results and firmly established that Env-mediated entry of the revertant virus was restored. We did measure a minor increase in infectivity and no increased entry for the intermediate revertant (Figure 2D and E), but the replication advantage is obvious (Figure 2A and B). The combined results indicate that a two-step evolution process took place upon removal of the V4 base disulphide bond, and both reversions at and near residue 418 contribute to the final revertant phenotype.

---

† Van Anken, E., Sanders, R. W., Liscaljet, M., de Kok, M., Tilleman, S., Holopainen, K., Dankers, M., Dierdorp, M., Busser, E., Berkhout, B. and Braakman, I., unpublished data
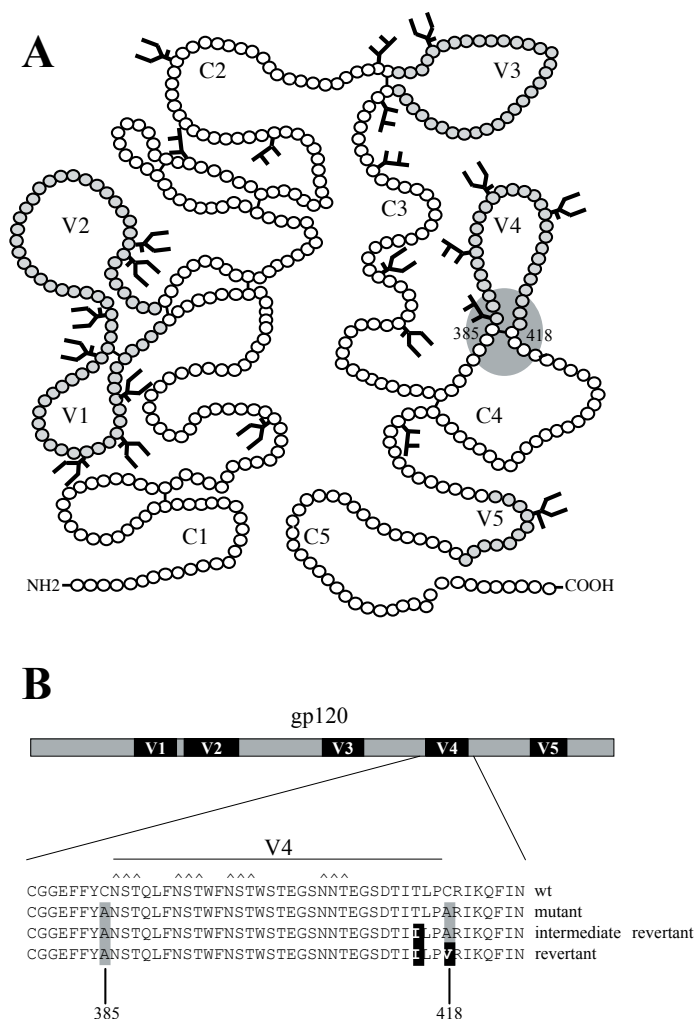
**Figure 1.** Local reversion in HIV-1 gp120. **A.** Schematic of gp120 with the five conserved domains (C1–C5 and five variable domains (V1–V5). The location of the V4 base disulphide bond is indicated (grey sphere). Figure is adapted from (Leonard *et al.*, 1990). Sites for N–linked glycosylation are given. **B.** Sequences of the V4 loop and flanking regions of wt, mutant and revertant viruses. No mutations were found outside this region. The original mutations are indicated with grey boxes, the reversion with black boxes. N–linked glycosylation sites are indicated by (^^^).

### The T415I and A418V reversions restore gp120 incorporation into virus particles

To study the effect of the various substitutions on the expression of Env in cells and on the surface of virus particles, we analyzed the gp120 content of cells and viruses using a gp120 ELISA. C33A cells were transiently transfected with the respective molecular clones. Cells were harvested after 48 hr and the virus fraction was purified from the culture supernatant by ultracentrifugation, which allowed us not only to determine the gp120 content of viruses, but also how much gp120 is shed into the supernatant from the

cellular and viral surfaces. For the wild-type construct, ~14% of the total amount of gp120 was present in the cell fraction, ~21% in the virus fraction, and ~65% in the supernatant, indicating considerable shedding of the LAI gp120 molecules from cells and/or viruses (van Anken, 2003). The corresponding
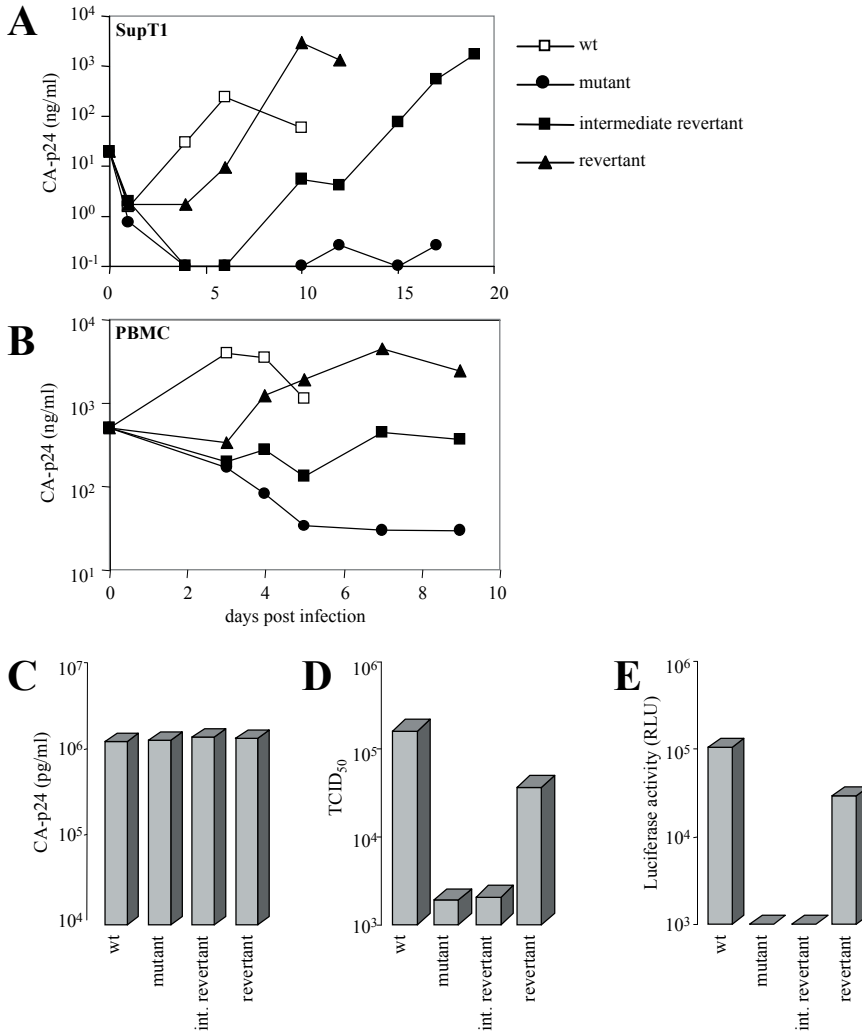


**Figure 2.** Restoration of viral infectivity and replication. **A.** $300 \times 10^3$ SupT1 T cells were infected with 20 ng CA-p24 and virus spread was measured for 20 days. **B.** $100 \times 10^3$ PBMCs were infected with 50 ng of the respective molecular clones and virus spread was monitored for 10 days. **C.** Virus stocks were produced by transfection of C33A cells and the amount of virus was quantitated using ELISA. **D.** The infectivities ($TCID_{50}$) were measured by endpoint dilution. The exact $TCID_{50}$ values of the respective mutants per µg CA-p24 are as follows: 162658 (wt); 1900 (mutant); 2083 (intermediate revertant) and 36745 (revertant). **E.** $300 \times 10^3$ LuSIV cells were infected with 200 ng CA-p24 in the presence of 200 nM Saquinavir. Viral entry into cells was quantified by measuring luciferase activity 48 hours after infection.

numbers for the viral CA-p24 protein are ~17% in cells, ~58% in the virus, and ~25% in the supernatant. We determined the gp120/CA-p24 ratio's in the respective fractions for mutant and revertant viruses and compared these to the wt gp120/CA-p24 ratio's that were arbitrarily set at 1 (Figure 3). The C385A/ C418A mutant accumulated gp120 in the cell fraction, but virtually no gp120 was incorporated into virus particles or shed into the medium (~4% gp120 incorporation into virions compared to wt). This result is consistent with the severe folding defect measured for this mutant (see below and (van Anken, 2003)). The revertant significantly increased gp120 incorporation into virions (~51% of wt). Strikingly, the revertant did hardly shed gp120 in the culture medium, suggesting that it has stabilised the gp120-gp41 interaction (~4% of wt shedding).
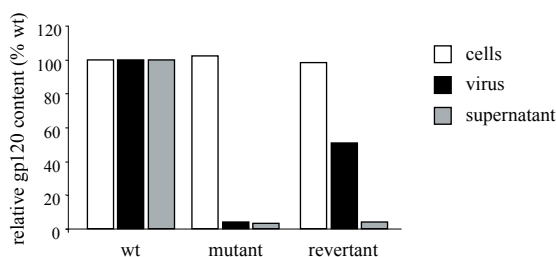


**Figure 3.** Restoration of Env incorporation into virions. gp120 content in cell, virus and supernatant fractions of virus producing cells. gp120 and CA-p24 contents were measured by ELISA. The gp120 amounts were standardised for CA-p24 input and the gp120 contents of mutants in the respective fractions are given as percentages of the wt gp120 contents (arbitrarily set at 1).

### The T415I and A418V reversions slightly improve gp120 folding

The poor replication capacity of the mutant Env could be explained by its low folding efficiency in the ER (van Anken, 2003). Apparently too few correctly folded Env molecules could leave the ER and be incorporated onto virus particles. The increased incorporation into virus particles of the revertant Env suggests that the evolutionary repair of virus replication did occur through increased folding competence of the revertant. We therefore analyzed oxidative folding of the revertant gp120 in comparison with wt and mutant gp120 by pulse-chase analysis (Figure 4). wt gp120 progressed with time via a 'smear' of partially oxidised folding intermediates (ITs) to a fully oxidised native form (NT) in nonreduced gels (Figure 4, upper left panel). Concomitantly, the signal peptide was removed from gp120 (Reduced uncleaved (Ru) vs Reduced cleaved (Rc) with time (Figure 4, lower left panel), as reported previously (van Anken, 2003). In contrast, mutant gp120 failed to display detectable amounts of NT even after 4 hrs of chase (Figure 4, upper middle panel) and the signal peptide was cleaved only from a minority of mutant gp120 molecules (Figure 4, lower middle panel), as we had observed previously (van Anken, 2003). In case of the revertant a faint NT-like band appeared at later chase periods (Figure 4, upper right panel, indicated by an arrowhead). Also, the revertant displayed slightly more signal peptide cleavage in comparison to the mutant periods (Figure 3, lower right panel, indicated by an arrowhead). Altogether, this indicates that the revertant phenotype correlates with slightly increased folding competence of gp120, although folding kinetics are far from restored to wt levels.
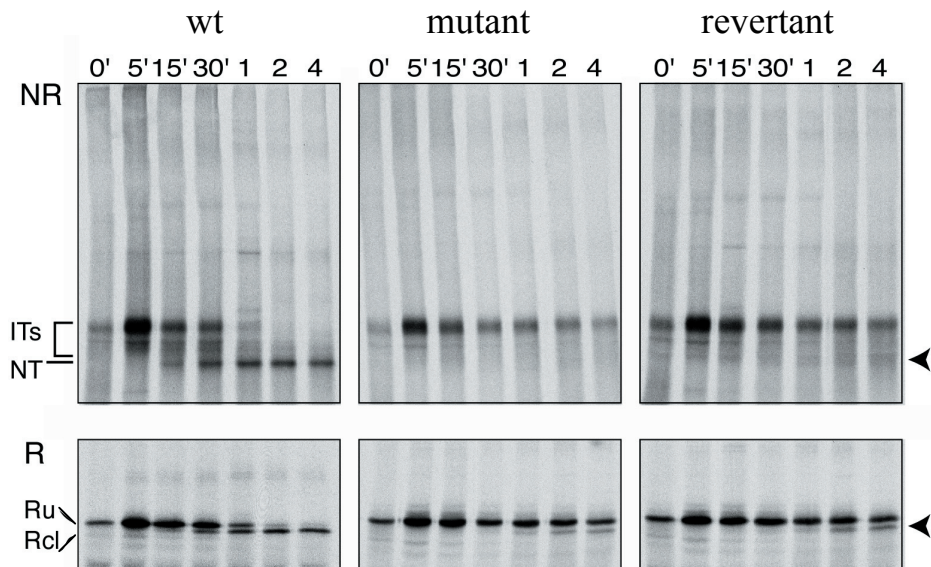
**Figure 4.** Partial restoration of gp120 folding. Hela cells were infected with VVT7 and transfected with plasmids encoding wt (left panels), mutant (middle panels) or revertant (right panels) gp120. Cells were pulse labelled for 2 min and chased for the indicated times. Cells were lysed and gp120 proteins were immunoprecipitated from lysates. Immunoprecipitates were deglycosylated and analyzed by nonreducing (upper panels) or reducing (lower panels) 7.5% SDS PAGE. Folding intermediates (ITs), the native form (NT), the reduced uncleaved (Ru) and cleaved (Rc) forms of gp120 are indicated.

### The revertant gp120 on virions is antigenically similar to wt Env

It is clear that the two local amino acid substitutions in the revertant improved gp120 folding, expression and virus incorporation compared to the original mutant. However, the question remained how these mutations could compensate for the lack of a disulphide bond that is absolutely conserved amongst natural HIV-1 isolates. To investigate whether the global fold of revertant gp120 was distinct from that of wt gp120, we performed neutralisation experiments using the 2G12 monoclonal antibody (Figure 5). 2G12 recognises a conformational carbohydrate epitope on the outer domain of gp120, and its epitope includes the carbohydrates that are attached to the asparagines at positions 332 and 392 in close proximity to the V4 base (Trkola *et al.*, 1996; Sanders *et al.*, 2002b; Scanlan *et al.*, 2002; Calarese *et al.*, 2003). In addition, the glycans attached to N295, N386 and N448 contribute directly or indirectly to the proper presentation of the 2G12 epitope. The glycan at position 386 is immediately adjacent to the 385–418 disulphide bond, and N392 is located only a few residues further downstream. Local structural alterations caused by the absence of the disulphide bond, with the additional amino acid changes causing the phenotypic reversion, could perturb the composition and/or orientation of these carbohydrates and thereby affect 2G12 binding and neutralisation. The monoclonal antibody IgG1b12, which binds to an conformational epitope that overlaps with the CD4 binding site was also included in the experiment (Burton *et al.*, 1994; Saphire *et al.*, 2001). Neutralisation experiments show that both the wt and revertant virus are inhibited by 2G12 and IgG1b12 with comparable $IC_{50}$ values (Figure 5). The mutant virus could not be analyzed in
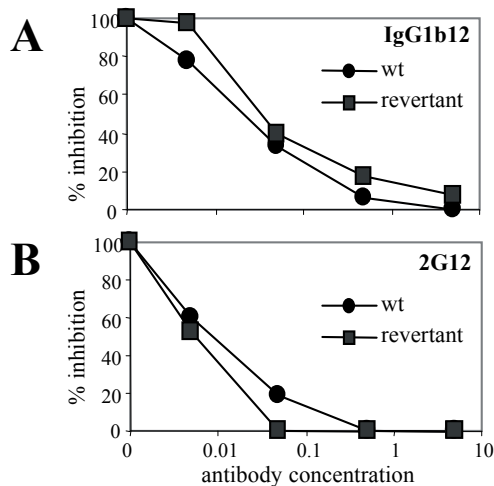
**Figure 5.** Neutralisation of wt and revertant viruses by monoclonal antibodies IgG1b12 and 2G12. Viral entry (as assayed in Figure 2D) was measured in the presence of varying concentrations of IgG1b12 and 2G12.

this assay because it does not replicate. These results indicate that wt and revertant gp120 molecules that reach the virus particle are not dramatically different in their conformation. Therefore, the restoration of replication capacity can be attributed to an increased yield of correctly folded Env species.

### Molecular dynamics simulations show no apparent effect on the global structure of gp120

To analyze in further detail the effects of the substitutions on gp120 structure and stability, we performed molecular dynamics simulations. Simulations performed at 300K revealed a remarkable stability of both wt and all variants of gp120 (results not shown). Increasing the temperature to 400K only enhanced local fluctuations, mainly in loop regions, *e.g.*, the V1/V2, V4 and V5 loops, without significant changes in energy or secondary structure. Even the removal of all disulphide bonds by alanine substitution did not lead to global unfolding or loss of secondary or tertiary structure, suggesting that the disulphide bonds do not play an important role in maintaining gp120 structure once it is folded (simulations performed at elevated temperature (400K) for 10 ns; results not shown). A complete sampling of the unfolding pathway for such a large system is beyond the reach of the current computational power. Instead, we will focus on the further description of the equilibrium state of each variant. Note that the core sequences of HXB2, present in the crystal structures, and LAI, used in our virus and folding experiments, differ by six amino acids, which are all remote from the mutation sites of interest. Comparison of the first 5 ns simulations of the HXB2 and LAI model structures revealed no significant difference (results not shown). We therefore used the HXB2 structure as wt reference and introduced point mutations *in silico* in order to minimise changes of structural variables so that the calculation of pre–equilibrating period can be reduced.

Atom positional root–mean–square deviation (RMSD) analysis of wt, mutant and revertant gp120s indicates that a 2 to 4 ns equilibration period is required for all backbone atoms to reach equilibrium. The four β–strands (β–13: 299–305, β–16/ β–17: 373–386 and β–19:412–422) around the mutation sites stabilised much faster with very little RMSD fluctuation (<0.1 nm). Therefore, to ensure the proper

sampling of the equilibrium states only the last 5 ns of all simulations were used for the analysis. A comparison of the three variants by RMSD matrix analysis also indicates that the overall backbone of the mutant core deviates slightly from wt and revertants with a maximum RMSD of 0.4 nm. The conformation of the four-stranded β-sheet that constitutes the region of interest in mutant and revertant gp120s is nevertheless very similar with only minor positional deviations from the wt (RMSD < 0.2 nm) (Figure 6). The presence of the white-to-light grey off-diagonal regions indicates sampling of similar regions of conformational space amongst these variants. Thus, no significant differences in the global structures were seen.
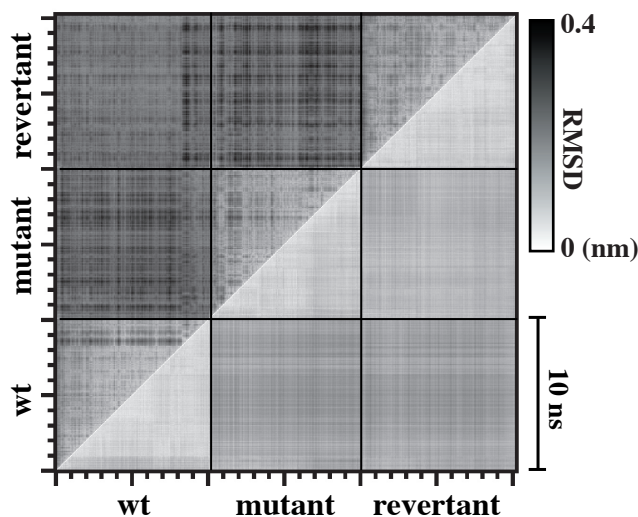


**Figure 6.** Pair-wise backbone RMSD matrix of wt, mutant (C385A/C418A) and revertant (C385A C418V T415I) gp120. Each section represents a 10 ns simulation at 300K. Each dot represents a positional root-mean-square deviation (RMSD) between two conformations taken from the respective trajectories indicated on the axes, and is coded in grey scale according to the scale shown on the right. The conformations are taken every 10 ps. The upper left panel shows the backbone RMSD fitted on the backbone atoms (N, C and Cα) of the secondary structure elements of the starting structure as identified by DSSP (Kabsch and Sander, 1983). The lower right panel shows the backbone RMSD of the β-sheet that includes β-strands β-13 (residues 299-305), β-16 and β-17 (residues 373-386), and β-19 (residues 412-422), fitted on the backbone atoms (N, C and Cα) of these residues. An equilibrated conformational sampling period is found when an off-diagonal region shows a continuous low RMSD values (white to light grey).

### The T415I and A418V reversions increase local backbone-backbone hydrogen bonding

Since no differences in the overall structures were seen, analysis at residual or atomic level may provide more insights into the effect of these mutations. The amino acids at positions 385, 415 and 418 are located in a four-stranded antiparallel β-sheet (Figure 7). Central in this β-sheet is strand β-19 with an unusual proline (P417) that causes the strand to bend. The N-terminal part of β-19 (β-19a) forms a double-stranded antiparallel β-sheet with β-13, but downstream of the unusual residue P417, β-19b interacts with β-17 to form a triple-stranded β-sheet also involving β-16. In the wt protein, C385 in β-17

and C418 in β–19 are covalently linked by the disulphide bond, conferring stability to the β–sheet that comprises the V4 base region. Inspection of the β–sheet propensities of the side chains of the revertant amino acids, revealed that both T415I and A418V increase the β–sheet propensity (Figure 7B) (Chou and Fasman, 1974; Levitt, 1978). The overall β–sheet propensity of β–19 increases from 1.11 for the mutant to 1.14 and 1.21 for the intermediate and final revertants, compared to 1.09 for the wt β–19, although it should be noted that the contribution of a disulphide bond is not taken into account for the wt β–19. Interestingly, the two reversions are on either side of the β–sheet destabilising residue P417.
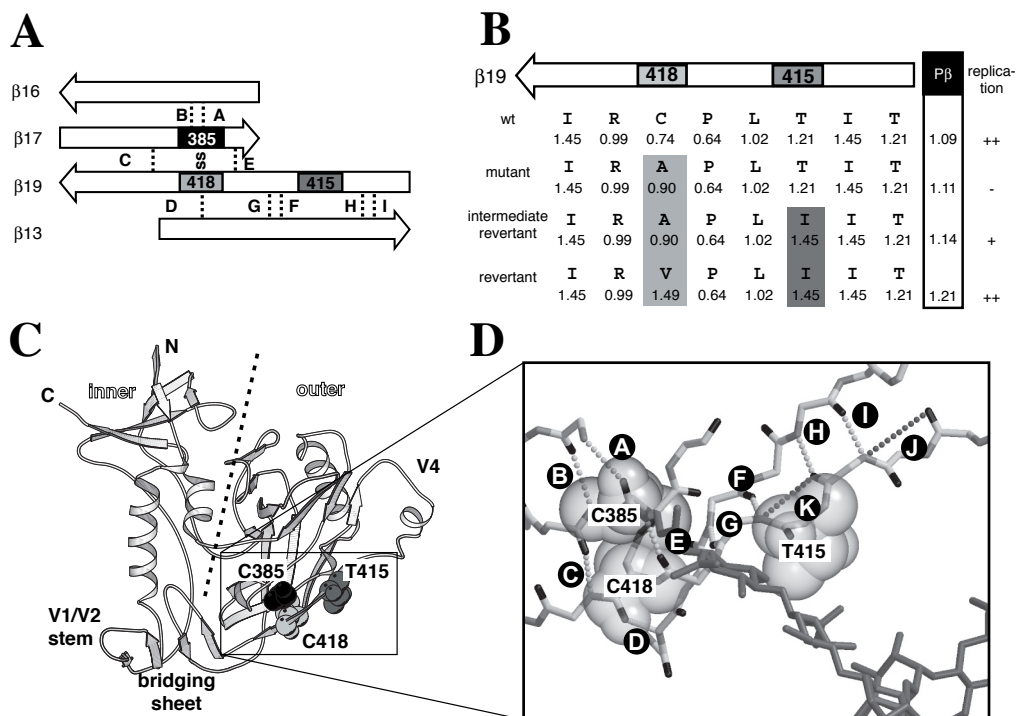


**Figure 7.** Hydrogen bond network in the vicinity of mutation sites. **A**. Schematic representation of the β–sheet fold of β–strands β–13, β–16, β–17 and β–19. The interstrand hydrogen bonds residues 385, 415 and 418 are indicated (see also Figure 7D and Table 1). **B**. β–sheet propensities of residues in β–19 including these of mutant and revertant residues, according to Levitt (1978). The composite β–sheet propensity (Pβ) of β–19 is also given Chou and Fasman (1974). **C**. Ribbon representation of the starting structure of core gp120 for simulation studies. The border between the inner and outer domains of gp120 is indicated with a dashed line. Mutation sites of C385, C418 and T415 are indicated in black, dark- and light grey spheres. The missing V4 loop is modelled by SWISS–MODEL. Detailed structure of the boxed region is shown on the right with the observed backbone-backbone hydrogen bonds around the mutation sites. Inter- and intrastrand backbone-backbone hydrogen bonds are displayed in light- and dark grey dashed-lines, respectively. Alphabetic labels correspond to the identities listed in Table 1. Glycan structure of the N-glycosylation at N386 is taken and modified from PDB entry 1H3U (Krapp *et al.*, 2003).

The formation of interresidue backbone hydrogen bonds is the determinant for secondary structure (Kabsch and Sander, 1983) and is therefore important for protein folding. A loss or a reduction in the presence of specific hydrogen bonds can lead to a deficiency in protein folding. A detailed molecular dynamics investigation into the hydrogen bond network in the vicinity of the mutation sites revealed an intriguing compensation effect. Four interstrand hydrogen bonds that are stable in wt gp120, are lost in the mutant, but are partially or fully restored in revertant (hydrogen bonds B, C, D and H in Table 1 and fig. 7C). In contrast, hydrogen bonds E and I, which are virtually absent in the wt, but present in the mutant are destabilised in the revertant. These particular hydrogen bonds may have a disadvantageous effect on the topology of the β-sheet. Some hydrogen donors and acceptors switch from interstrand bonding to intrastrand bonding (for example see hydrogen bond I and J in Table 1 and fig. 7C). Finally, in line with the experimental data, comparison of the overall average occurrence of the eleven hydrogen bonds that are in close proximity to the mutation sites indicates that wt gives the highest stability to this particular hydrogen bond network (66.2%), with partial restoration in revertant (52.5%) compared to the mutant (45.8%). The restoration of hydrogen bonding in the revertant is even more significant when one considers the absence of the disulphide bond that fixes these interactions in the wt protein.

**Table 1.** Statistics of hydrogen bond occurrence in the vicinity of the mutation and reversion sites.

| | Backbone-backbone hydrogen bond | Hydrogen bond occurrence (%) during 5-10 ns | | |
|---|---|---|---|---|
| | | wt | mutant | revertant |
| A | C385[e](O)–H374(N) | 86.8 | 90.2 | 99.6 |
| B | C385[e](N)–H374(O) | 97.4 | 0.0 | 34.1 |
| C | Y384(O)–R419(N) | 90.6 | 10.2 | 85.6 |
| D[a] | C418[e](N)–G329(O) | 94.0 | 0.2 | 34.3 |
| E[b] | P417(O)–N386(N) | 6.6 | 94.4 | 35.5 |
| F[d] | C331(O)–L416(N) | 14.4 | 0.2 | 0.2 |
| G | C331(N)–L416(N) | 92.6 | 98.0 | 86.2 |
| H[c] | I414(O)–I333(N) | 61.9 | 0.2 | 97.6 |
| I | I414(N)–I333(O) | 0.2 | 96.2 | 4.4 |
| J[d] | I414(N)–N412(O) | 96.4 | 17.0 | 0.2 |
| K[c,d] | I414(O)–L416(N) | 86.8 | 97.2 | 99.4 |
| | *Overall average* | *66.2* | *45.8* | *52.5* |

a G329 is the first linker residue for the truncated V3 loop

b N-glycosylation at N386 was not taken into account during simulation

c hydrogen bonds that are located directly proximal to residue 415 (see figure 7)

d intrastrand backbone hydrogen bonds. A hydrogen bond is considered to exist when the donor-hydrogen-acceptor angle is larger than 135° and the donor-acceptor distance is smaller than 0.25 nm.

e Cysteines in the wt protein, but not in the mutant or in revertant proteins

### The T415I and A418V reversions improve local interstrand side chain packing

In addition to the backbone hydrogen bonding that defines the secondary structure of protein folds, the side chain packing is also an important stabilising factor. The electrostatic interactions amongst side chains provide a long range attracting gradient and therefore might be crucial for protein folding during the search of the native fold. An overlay of snapshots of selected hydrophilic residues around the mutation site during molecular dynamics simulations shows a remarkable well-defined side chain packing in the wt

protein, except for R419, which exhibits a main cluster close to N386 with the remaining being poorly defined (Figure 8). Using the same criteria as for the backbone hydrogen bond network, interstrand side chain hydrogen bond are found for S375(Oγ)–Y384(Oη) (29.7% in 5-10 ns trajectory) and N386(Nδ2)–R419(Nη) (25.0%) despite the intrinsic flexibility of R419. Note that the side chain hydrogen bonding geometry of S375(Oγ)–S375(Hγ)–Y384(Oη) is surprisingly well-defined despite the intrinsic mobility of serine and tyrosine side chains. The geometry of S375 and Y384 becomes clearly disordered in the mutant and no cluster can be found in R419 that is in contact with N386. The overall geometry of the side chain organisation is slightly different and, as a consequence, both side chain hydrogen bonds are absent in the mutant. The side chain packing is substantially restored in the revertant, particularly for S375 and Y384. However, the gain in structural integrity is not sufficient to restore the hydrogen bonds that are present in the wt protein. Nevertheless, these results illustrate that the reversions contribute to an improvement of local side chain packing.
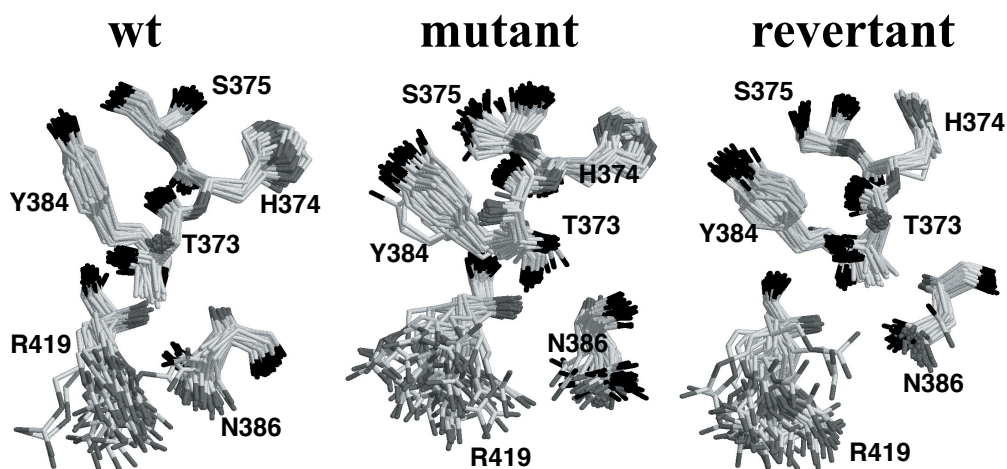


**Figure 8.** Snapshots of selected hydrophilic residues around the mutation sites, that can potentially form interstrand hydrogen bonds. The structures are sampled every 100 ps during the 5-10 ns trajectories. All structures are fitted on the backbone atoms of these residues. Oxygen, nitrogen and carbon atoms are coloured in black, dark- and light grey, respectively.

The analysis of backbone–backbone and side chain interactions illustrate that the effects of the mutations and reversions at positions 385, 415 and 418 are regional. Besides local effects on the interactions between strands β–17 and β–19 that are normally linked by the disulphide bond, there are also effects on interactions between β–17 and β–16 (*e.g.* S375-Y384; H374-C385A) and between β–19 and β–13 (*e.g.* I333-I414; L416-C331; G329-C418A). Thus, the mutations and reversions affect all four strands of the antiparallel β–sheet.

## Discussion

We describe an HIV-1 gp120 variant that emerged through virus evolution and that is functional despite the lack of the disulphide bond at the base of the V4 domain, which is otherwise required for virus replication and gp120 activity. Virological, biochemical and computational analyses suggest that virus replication is restored through improvement of a local β-sheet fold. This study provides evidence that a critically important disulphide bond can be functionally replaced by an alternative protein structure motif. It also provides evidence for the proposal that local protein stability is an important factor in escape from ER quality control during protein biosynthesis.

Our data are surprisingly well in accordance with both theoretical β-sheet propensities (Figure 7B) (Chou and Fasman, 1974; Levitt, 1978) and β-sheet preferences as established in small model proteins (Kim and Berg, 1993; Minor and Kim, 1994a; Smith *et al.*, 1994). The first reversion (T415I) slightly increases the β-sheet preference and viral replication is slightly improved. The second A418V reversion has a considerable effect on both β-sheet preference and viral replication. The generalised rules on β-sheet preference apply to β-19 since it is a central strand and not an edge strand (Minor and Kim, 1994b) (Figure 7). The counterbalancing effect of the reversions on the presence of the β-sheet disfavouring P417 is presumably only possible because β-19 is not an edge strand (Santiveri *et al.*, 2003). Thus, β-sheet preference is a major determinant in directing protein folding and protein stability and our data point out that the simple rules deduced from experiments with small model proteins also hold for the intricate folding of a complicated glycoprotein such as gp120 in living cells.

The increase of hydrogen bonding might in part be a result of the increase of van der Waals interactions where T415I is flanked by I414 and L416 with several hydrophobic residues in the opposite strand, β-13. A recent study suggested that a hydrophobic core is the stabilising factor for a β-stranded Betanova peptide (Colombo *et al.*, 2002). Kumagai and co-workers also showed that threonine-to-isoleucine mutation at position 29 does not change the structure of α-lactalbumin but has a significant positive effect on its thermostability due to the increase of hydrophobic side chain packing and hydrogen bonding (Horii *et al.*, 2001). We also did not find significant overall structural perturbation when mutations were introduced *in silico* in the native state of gp120.

The local restoration of a folding defect of a disulphide bond mutant provides new insights on the relevance of disulphide bonds for folding in the ER and on the discrimination between immature and correctly folded client proteins by ER quality control. The well-characterised calnexin/calreticulin cycle plays a role in oxidative folding in the ER and ER quality control of virtually all glycoproteins (Trombetta and Helenius, 1998; Ellgaard *et al.*, 1999; Ellgaard and Helenius, 2003). Calnexin and calreticulin are lectins that recognise monoglucosylated carbohydrate moieties on glycoproteins. They associate with Erp57, an ER resident thiol-oxidoreductase that forms transient disulphide bonds with glycoproteins bound to calnexin or calreticulin and mediates isomerisation of disulphide bonds (Molinari and Helenius, 1999). Glucosidase II hydrolyses glucose from monoglucosylated carbohydrates on folding glycoproteins, which results in glycoprotein release from calnexin and calreticulin. An other important player in the calnexin/calreticulin cycle and in ER quality control is UDP-glucose:glycoprotein glucosyltransferase (UGGT), which recognises improperly folded domains on folding glycoproteins and reglucosylates carbohydrates in the misfolded region so that it can reassociate with calnexin or calreticulin for another attempt to properly fold the domain. The determinants for recognition of improperly folded glycoprotein domains

by UGGT are not completely clear (Sousa and Parodi, 1995; Sousa and Parodi, 1996; Fernandez *et al.*, 1998; Trombetta and Helenius, 1998; Ritter and Helenius, 2000; Caramelo *et al.*, 2003). UGGT recognises only the improperly folded parts of the protein, and it does not recognise random coil. Although the carbohydrate is important for tagging the protein to be recognised by calnexin or calreticulin, it is not involved in the recognition of misfolding. It has been suggested that UGGT recognises exposed hydrophobic patches, instable domains or mobile groups.

In gp120 lacking the 385/418 disulphide bond, local strengthening of non–covalent interactions stabilises a local β–sheet fold. The reversions do not alter the local β–sheet structure, but increase the stability of the β–sheet and thereby the correct fold or a quasi–correct fold. UGGT does not recognise the protein anymore as being improperly folded and the protein is allowed to leave the ER. Although other molecular chaperones, such as BiP, play a role in the recognition of misfolding, no BiP binding sites were detected in the V4 region (Knarr *et al.*, 1999).

When UGGT senses misfolding, it tags the protein for reentering the calnexin/calreticulin pathway by reglucosylation of a local carbohydrate. The obvious candidate for marking the misfolded V4 base region is the glycan attached to N386, immediately neighboring the 385/418 disulphide bond in the wt protein. We have observed an alternative escape route in a similar evolution experiment with a C418A single mutant virus. In that particular case, the increase in viral replication and escape from ER quality control involved the elimination of the 386 glycan (results not shown). Thus, these studies may represent two different pathways for ER quality control escape: increase of local stability of protein structure or elimination of a nearby carbohydrate.

The number of revertant gp120 molecules that reached a native state was still modest compared to wt. Since only native gp120 can exit the ER, also limited amounts of revertant Env could reach the cell surface to be incorporated into virions. Nevertheless, incorporation reached levels that were close to wt. Apparently, the few Env molecules that did reach the cell surface were sufficient to produce virions with a normal Env content. This suggests that other factors are limiting in the intricate virion assembly process. This also suggests that in wt only a fraction of Env is incorporated. We note that the increase in replication capacity could be correlated directly with an increase of gp120 incorporation into virions, suggesting that the reversions cause the escape from ER quality control, but that they do not contribute to improvement of subsequent Env functions *per se* (*e.g.* (co)receptor binding, membrane fusion).

This study indicates that the 385–418 disulphide bond is not specifically required for gp120 folding, *i.e.* it can be compensated for by alternative means of protein stabilisation. There is a precedent for this phenomenon. Single chain antibody fragments (scFv) could be generated by molecular evolution, in which disulphide bonds were replaced by, for example, a salt bridge (Proba *et al.*, 1998). We have observed another escape route in a C418A single mutant, also lacking the V4 base. In addition, we have generated various escape variants from HIV–1 lacking the 54–74 disulphide bond in the C1 region of gp120 (Sanders, 2004)[*]. Restoration of the original disulphide bonds was never seen in these escape variants. Taken together with our earlier studies in which we showed that five out of ten disulphide bonds were not essential for ER folding, we can now conclude that seven out of ten disulphide bonds are not absolutely and specifically required for oxidative folding of gp120. Five

---

[*] Sanders, R.W., Busser, E., Dankers, M.M. Lu, M. and Berkhout, B., unpublished data

disulphide bonds can be replaced without penalty on folding, two can be replaced by other stabilising mechanisms. Whether the formation of specific native or non-native disulphide bonds plays a role in directing protein folding is unclear (Weissman and Kim, 1992; Cemazar *et al.*, 2003), but if such a mechanism is required for oxidative folding of Env, our work clearly indicates that most of the cysteines in Env are not involved. An exception might be formed by the cysteines in the C2-V3 region, which are absolutely required for folding. This region might act as an initiator or director for Env folding.

Of note is that the exemplary short term evolution experiments that we performed never resulted in an Env that folded as efficiently as wt and a virus that replicated at wt levels. It would be of interest to prolong the evolution of the revertant virus presented here, to see whether it can in fact reach wt levels of folding without the 385-418 disulphide bond. It is possible that, although not essential for Env folding *per se*, it increases the chances of efficient folding.

The question remains why the V4 base disulphide bond is absolutely conserved *in vivo*. One reason could be that the best alternative for this disulphide bond, such as a stabilised β-sheet, may never completely compensate for the lack of the disulphide bond and restore wt levels of replication. Prolonged evolution experiments with the revertant virus described here, might also answer this question. The loss of the disulphide bond may also be prevented because the virus would have to go through an intermediate with a free cysteine, which is usually disadvantageous because it may interfere with the formation of the correct disulphide bonds (van Anken, 2003). Alternatively, it is possible that this disulphide bond has a more distinct function *in vivo*, for example in immune evasion. It may play a role in positioning of the V4 domain so that it can optimally exert its function as antigenic shield, or it could be involved in maintaining an optimally loose gp120-gp41, resulting in shedding and the presence of immunological decoy gp120. This study underlines the evolutionary potential of HIV-1 not only because of its high mutation rate but also because of the structural plasticity of its proteins.

## Acknowledgements

# References

Altieri, A.S., Hinton, D.P. and Byrd, R.A. (1995) Association of biomolecular systems via pulsed-field gradient NMR self-diffusion measurements. *J. Am. Chem. Soc.*, **117**, 7566-7567.

Amadei, A., Linssen, A.B.M. and Berendsen, H.J.C. (1993) Essential dynamics of proteins. *Proteins*, **17**, 412-425.

Andricioaei, I. and Karplus, M. (2001) On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.*, **115**, 6289-6292.

Anglister, J., Grzesiek, S., Ren, H., Klee, C.B. and Bax, A. (1993) Isotope-edited multidimensional NMR of Calcineurin-B in the presence of the non-deuterated detergent CHAPS. *J. Biomol. NMR*, **3**, 121-126.

Bader, G.D. and Hogue, C.W.V. (2002) Analyzing yeast protein-protein interaction data obtained from different sources. *Nature Biotechnol.*, **20**, 991-997.

Bader, R., Bettio, A., Beck-Sickinger, A.G. and Zerbe, O. (2001) Structure and dynamics of micelle-bound neuropeptide Y: Comparison with unligated NPY and implications for receptor selection. *J. Mol. Biol.*, **305**, 307-329.

Baker, D. and Sali, A. (2001) Protein structure prediction and structural genomics. *Science*, **294**, 93-96.

Baltimore, D. (2002) Steering a course to an AIDS vaccine. *Science*, **296**, 2297-2297.

Barbato, G., Bianchi, E., Ingallinella, P., Hurni, W.H., Miller, M.D., Ciliberto, G., Cortese, R., Bazzo, R., Shiver, J.W. and Pessi, A. (2003) Structural analysis of the epitope of the anti-HIV antibody 2F5 sheds light into its mechanism of neutralization and HIV fusion. *J. Mol. Biol.*, **330**, 1101-1115.

Basmaciogullari, S., Babcock, G.J., Van Ryk, D., Wojtowicz, W. and Sodroski, J. (2002) Identification of conserved and variable structures in the human immunodeficiency virus gp120 glycoprotein of importance for CXCR4 binding. *J. Virol.*, **76**, 10791-10800.

Basus, V.J. (1989) Proton nuclear-magnetic-resonance assignments. *Method Enzymol.*, **177**, 132-149.

Baxter, N.J. and Williamson, M.P. (1997) Temperature dependence of $^1$H chemical shifts in proteins. *J. Biomol. NMR*, **9**, 359-369.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermans, J. (1981) Interaction models for water in relation to protein hydration. In Pullman, B. (ed.), *Intermolecular forces*. Reidel Publishing Company, Dordrecht, pp. 331-342.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Di Nola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684-3690.

Berger, E.A., Murphy, P.M. and Farber, J.M. (1999) Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.*, **17**, 657-700.

Bierbaum, G., Brötz, H., Koller, K.P. and Sahl, H.-G. (1995) Cloning, sequencing and production of the lantibiotic mersacidin. *FEMS Microbiol. Lett.*, **127**, 121-126.

Bogan, A.A. and Thorn, K.S. (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol.*, **280**, 1-9.

Bonev, B.B., Chan, W.C., Bycroft, B.W., Roberts, G.C. and Watts, A. (2000) Interaction of the lantibiotic nisin with mixed lipid bilayers: A $^{31}$P and $^2$H NMR study. *Biochemistry*, **39**, 11425-11433.

Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1991) Direct NOE refinement of biomolecular structures using 2D NMR data. *J. Biomol. NMR*, **1**, 305-309.

Bonvin, A.M.J.J. (1993) *Determination of biomolecular structures by NMR: Use of relaxation matrix calculations*. Ph.D. thesis, Utrecht University, Utrecht.

Bonvin, A.M.J.J., Rullmann, J.A., Lamerichs, R.M., Boelens, R. and Kaptein, R. (1993) "Ensemble" iterative relaxation matrix approach: A new NMR refinement protocol applied to the solution structure of crambin. *Proteins*, **15**, 385-400.

Bonvin, A.M.J.J. and Brünger, A.T. (1995) Conformational variability of solution nuclear magnetic resonance structures. *J. Mol. Biol.*, **250**, 80-93.

Bonvin, A.M.J.J. and Brünger, A.T. (1996) Do NOE distances contain enough information to assess the relative populations of multi-conformer structures? *J. Biomol. NMR*, **7**, 72-76.

Bonvin, A.M.J.J., Houben, K., Guenneugues, M., Kaptein, R. and Boelens, R. (2001) Rapid protein fold determination using secondary chemical shifts and cross-hydrogen bond $^{15}$N-$^{13}$C' scalar couplings ($^{3hb}J_{NC'}$). *J. Biomol. NMR*, **21**, 221-233.

Borders, C.L., Jr., Broadwater, J.A., Bekeny, P.A., Salmon, J.E., Lee, A.S., Eldridge, A.M. and Pett, V.B. (1994) A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Sci.*, **3**, 541-548.

Branden, C. and Tooze, J. (1998) *Introduction to protein structure*. Garland Publishing Inc., New York.

Brenner, S.E., Chothia, C. and Hubbard, T.J. (1997) Population statistics of protein structures: Lessons from structural classifications. *Curr. Opin. Struc. Biol.*, **7**, 369-376.

Breukink, E., van Kraaij, C., Demel, R.A., Siezen, R.J., Kuipers, O.P. and de Kruijff, B. (1997) The C-terminal region of nisin is responsible for the initial interaction of nisin with the target membrane. *Biochemistry*, **36**, 6968-6976.

Breukink, E., van Kraaij, C., van Dalen, A., Demel, R.A., Siezen, R.J., de Kruijff, B. and Kuipers, O.P. (1998) The orientation of nisin in membranes. *Biochemistry*, **37**, 8153-8162.

Breukink, E. and de Kruijff, B. (1999) The lantibiotic nisin, a special case or not? *Biochim. Biophys. Acta*, **1462**, 223-234.

Breukink, E., Wiedemann, I., van Kraaij, C., Kuipers, O.P., Sahl, H.-G. and de Kruijff, B. (1999) Use of the cell wall precursor lipid II by a pore-forming peptide antibiotic. *Science*, **286**, 2361-2364.

Breukink, E., van Heusden, H.E., Vollmerhaus, P.J., Swiezewska, E., Brunner, L., Walker, S., Heck, A.J. and de Kruijff, B. (2003) Lipid II is an intrinsic component of the pore induced by nisin in bacterial membranes. *J. Biol. Chem.*, **26**, 26.

Brötz, H., Bierbaum, G., Markus, A., Molitor, E. and Sahl, H.-G. (1995) Mode of action of the lantibiotic mersacidin: Inhibition of peptidoglycan biosynthesis via a novel mechanism? *Antimicrob. Agents Chemother.*, **39**, 714-719.

Brötz, H., Bierbaum, G., Reynolds, P.E. and Sahl, H.-G. (1997) The lantibiotic mersacidin inhibits peptidoglycan biosynthesis at the level of transglycosylation. *Eur. J. Biochem.*, **246**, 193-199.

Brötz, H., Bierbaum, G., Leopold, K., Reynolds, P.E. and Sahl, H.-G. (1998a) The lantibiotic mersacidin inhibits peptidoglycan synthesis by targeting lipid II. *Antimicrob. Agents Chemother.*, **42**, 154-160.

Brötz, H., Josten, M., Wiedemann, I., Schneider, U., Götz, F., Bierbaum, G. and Sahl, H.-G. (1998b) Role of lipid-bound peptidoglycan precursors in the formation of pores by nisin, epidermin and other lantibiotics. *Mol. Microbiol.*, **30**, 317-327.

Brötz, H. and Sahl, H.-G. (2000) New insights into the mechanism of action of lantibiotics-diverse biological effects by binding to the same molecular target. *J. Antimicrob. Chemother.*, **46**, 1-6.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54 (Pt 5)**, 905-921.

Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X., O'Neil, K.T. and DeGrado, W.F. (1995) Protein design: A hierarchic approach. *Science*, **270**, 935-941.

Burley, S.K., Almo, S.C., Bonanno, J.B., Capel, M., Chance, M.R., Gaasterland, T., Lin, D., Sali, A., Studier, F.W. and Swaminathan, S. (1999) Structural genomics: Beyond the human genome project. *Nat. Genet.*, **23**, 151-157.

Burton, D.R., Pyati, J., Koduri, R., Sharp, S.J., Thornton, G.B., Parren, P.W., Sawyer, L.S., Hendry, R.M., Dunlop, N., Nara, P.L. and et al. (1994) Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science*, **266**, 1024-1027.

Calarese, D.A., Scanlan, C.N., Zwick, M.B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M.R., Stanfield, R.L., Roux, K.H., Kelly, J.W., Rudd, P.M., Dwek, R.A., Katinger, H., Burton, D.R. and Wilson, I.A. (2003) Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science*, **300**, 2065-2071.

Cameron, D.W., Japour, A.J., Xu, Y., Hsu, A., Mellors, J., Farthing, C., Cohen, C., Poretz, D., Markowitz, M., Follansbee, S., Angel, J.B., McMahon, D., Ho, D., Devanarayan, V., Rode, R., Salgo, M., Kempf, D.J., Granneman, R., Leonard, J.M. and Sun, E. (1999) Ritonavir and saquinavir combination therapy for the treatment of HIV infection. *AIDS*, **13**, 213-224.

Caramelo, J.J., Castro, O.A., Alonso, L.G., De Prat-Gay, G. and Parodi, A.J. (2003) UDP-Glc:Glycoprotein glucosyltransferase recognizes structured and solvent accessible hydrophobic patches in molten globule-like folding intermediates. *Proc. Natl. Acad. Sci. USA*, **100**, 86-91.

Catasti, P., Bradbury, E.M. and Gupta, G. (1996) Structure and polymorphism of HIV-1 third variable loops. *J. Biol. Chem.*, **271**, 8236-8242.

Cavanagh, J., Fairbrother, W.J., Palmer III, A.G. and Skelton, N.J. (1996) *Protein NMR spectroscopy*. Academic Press, San Diego, CA.

Cemazar, M., Zahariev, S., Lopez, J.J., Carugo, O., Jones, J.A., Hore, P.J. and Pongor, S. (2003) Oxidative folding intermediates with nonnative disulfide bridges between adjacent cysteine residues. *Proc. Natl. Acad. Sci. USA*, **100**, 5754-5759.

Chakrabarti, P. and Janin, J. (2002) Dissecting protein–protein recognition sites. *Proteins*, **47**, 334-343.

Chan, D.C., Fass, D., Berger, J.M. and Kim, P.S. (1997) Core structure of gp41 from the HIV envelope glycoprotein. *Cell*, **89**, 263-273.

Chan, D.C. and Kim, P.S. (1998) HIV entry and its inhibition. *Cell*, **93**, 681-684.

Chan, W.C., Bycroft, B.W., Lian, L.Y. and Roberts, G.C.K. (1989) Isolation and characterisation of two degradation products derived from the peptide antibiotics nisin. *FEBS Lett.*, **252**, 29-36.

Chan, W.C., Bycroft, B.W., Leyland, M.L., Lian, L.Y., Yang, J.C. and Roberts, G.C.K. (1992) Sequence-specific resonance assignment and conformational-analysis of subtilin by 2D NMR. *FEBS Lett.*, **300**, 56-62.

Chan, W.C., Dodd, H.M., Horn, N., Maclean, K., Lian, L.Y., Bycroft, B.W., Gasson, M.J. and Roberts, G.C.K. (1996a) Structure-activity relationships in the peptide antibiotic nisin: Role of dehydroalanine 5. *Appl. Environ. Microbiol.*, **62**, 2966-2969.

Chan, W.C., Leyland, M., Clark, J., Dodd, H.M., Lian, L.Y., Gasson, M.J., Bycroft, B.W. and Roberts, G.C.K. (1996b) Structure-activity relationships in the peptide antibiotic nisin: Antibacterial activity of fragments of nisin. *FEBS Lett.*, **390**, 129-132.

Chandrasekhar, K., Profy, A.T. and Dyson, H.J. (1991) Solution conformational preferences of immunogenic peptides derived from the principal neutralizing determinant of the HIV-1 envelope glycoprotein gp120. *Biochemistry*, **30**, 9187-9194.

Chou, J.J., Kaufman, J.D., Stahl, S.J., Wingfield, P.T. and Bax, A. (2002) Micelle-induced curvature in a water-insoluble HIV-1 Env peptide revealed by NMR dipolar coupling measurement in stretched polyacrylamid gel. *J. Am. Chem. Soc.*, **124**, 2450-2451.

Chou, P.Y. and Fasman, G.D. (1974) Prediction of protein conformation. *Biochemistry*, **13**, 222-245.

Cierpicki, T. and Otlewski, J. (2001) Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR*, **21**, 249-261.

Clackson, T. and Wells, J.A. (1995) A hot-spot of binding-energy in a hormone-receptor interface. *Science*, **267**, 383-386.

Colman, P.M. and Lawrence, M.C. (2003) The structural biology of type I viral membrane fusion. *Nat. Rev. Mol. Cell Biol.*, **4**, 309-319.

Colombo, G., Roccatano, D. and Mark, A.E. (2002) Folding and stability of the three-stranded β-sheet peptide betanova: Insights from molecular dynamics simulations. *Proteins*, **46**, 380-392.

Corey, R.B. and Pauling, L. (1956) Specific hydrogen-bond formation between pyramidines and purines in deoxyribonucleic acids. *Arch. Biochem. Biophys.*, **65**, 164-181.

Crowley, P.B. and Ubbink, M. (2003) Close encounters of the transient kind: Protein interactions in the photosynthetic redox chain investigated by NMR spectroscopy. *Acc. Chem. Res.*, **36**, 723-730.

Csete, M.E. and Doyle, J.C. (2002) Reverse engineering of biological complexity. *Science*, **295**, 1664-1669.

Cudic, P., Kranz, J.K., Behenna, D.C., Kruger, R.G., Tadesse, H., Wand, A.J., Veklich, Y.I., Weisel, J.W. and McCafferty, D.G. (2002) Complexation of peptidoglycan intermediates by the lipoglycodepsipeptide antibiotic ramoplanin: Minimal structural requirements for intermolecular complexation and fibril formation. *Proc. Natl. Acad. Sci. USA*, **99**, 7384-7389.

Das, A.T.B., Klaver, B., Klasens, B.I., van Wamel, J.L. and Berkhout, B. (1997) A conserved hairpin motif in the R-U5 region of the human immunodeficiency virus type 1 RNA genome is essential for replication. *J. Virol.*, **71**, 2346-2356.

Das, A.T.B., Klaver, B. and Berkhout, B. (1999) A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection. *J. Virol.*, **73**, 81-91.

Daura, X., Mark, A.E. and van Gunsteren, W.F. (1998) Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.*, **19**, 535-547.

Davis, S.J., Schockmel, G.A., Somoza, C., Buck, D.W., Healey, D.G., Rieber, E.P., Reiter, C. and Williams, A.F. (1992) Antibody and HIV-1 gp120 recognition of CD4 undermines the concept of mimicry between antibodies and receptors. *Nature*, **358**, 76-79.

Dayie, K.T., Wagner, G. and Lefevre, J.F. (1996) Theory and practice of nuclear spin relaxation in proteins. *Annu. Rev. Phys. Chem.*, **47**, 243-282.

De Guzman, R.N., Turner, R.B. and Summers, M.F. (1998) Protein-RNA recognition. *Biopolymers*, **48**, 181-195.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.

DeLano, W.L., Ultsch, M.H., de Vos, A.M. and Wells, J.A. (2000) Convergent solutions to binding at a protein-protein interface. *Science*, **287**, 1279-1283.

Demarest, S.J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H.J., Evans, R.M. and Wright, P.E. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*, **415**, 549-553.

Di Nola, A., Berendsen, H.J.C. and Edholm, O. (1984) Free energy determination of polypeptide conformations generated by molecular

dynamics. *Macromolecules*, **17**, 1011–1028.

Dill, K.A. and Chan, H.S. (1997) From levinthal to pathways to funnels. *Nat. Struct. Biol.*, **4**, 10–19.

Dimitrov, D.S., Hillman, K., Manischewitz, J., Blumenthal, R. and Golding, H. (1992) Kinetics of soluble CD4 binding to cells expressing human immunodeficiency virus type 1 envelope glycoprotein. *J. Virol.*, **66**, 132–138.

Dingley, A.J. and Grzesiek, S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^2J_{NN}$ couplings. *J. Am. Chem. Soc.*, **120**, 8293–8297.

Dingley, A.J., Cordier, F. and Grzesiek, S. (2001) An introduction to hydrogen bond scalar couplings. *Concepts Magn. Resonance*, **13**, 103–127.

Dinner, A.R., Sali, A., Smith, L.J., Dobson, C.M. and Karplus, M. (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.*, **25**, 331–339.

Dobson, C.M., Sali, A. and Karplus, M. (1998) Protein folding: A perspective from theory and experiment. *Angew. Chem.-Int. Edit.*, **37**, 868–893.

Dobson, C.M. (2003) Protein folding and misfolding. *Nature*, **426**, 884–890.

Dominguez, C., Boelens, R. and Bonvin, A.M.J.J. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**, 1731–1737.

Dragic, T., Trkola, A., Lin, S.W., Nagashima, K.A., Kajumo, F., Zhao, L., Olson, W.C., Wu, L.J., Mackay, C.R., Allaway, G.P., Sakmar, T.P., Moore, J.P. and Maddon, P.J. (1998) Amino-terminal substitutions in the CCR5 coreceptor impair gp120 binding and human immunodeficiency virus type 1 entry. *J. Virol.*, **72**, 279–285.

Driessen, A.J., van den Hooven, H.W., Kuiper, W., van de Kamp, M., Sahl, H.-G., Konings, R.N. and Konings, W.N. (1995) Mechanistic studies of lantibiotic-induced permeabilization of phospholipid vesicles. *Biochemistry*, **34**, 1606–1614.

Duan, Y. and Kollman, P.A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, **282**, 740–744.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C. and Obradovic, Z. (2001) Intrinsically disordered protein. *J. Mol. Graph. Model.*, **19**, 26–59.

Dykes, G.A., Hancock, R.E.W. and Hastings, J.W. (1998) Structural variations in nisin associated with different membrane mimicking and ph environments. *Biochem. Biophys. Res. Commun.*, **247**, 723–727.

Dyson, H.J. and Wright, P.E. (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struc. Biol.*, **12**, 54–60.

Earl, P.L., Moss, B. and Doms, R.W. (1991) Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein. *J. Virol.*, **65**, 2047–2055.

Eckert, D.M., Malashkevich, V.N., Hong, L.H., Carr, P.A. and Kim, P.S. (1999) Inhibiting HIV-1 entry: Discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell*, **99**, 103–115.

Eckert, D.M. and Kim, P.S. (2001a) Mechanisms of viral membrane fusion and its inhibition. *Annu. Rev. Biochem.*, **70**, 777–810.

Eckert, D.M. and Kim, P.S. (2001b) Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc. Natl. Acad. Sci. USA*, **98**, 11187–11192.

Edholm, O. and Berendsen, H.J.C. (1984) *Mol. Phys.*, **51**, 1011–1028.

El Zoeiby, A., Sanschagrin, F. and Levesque, R.C. (2003) Structure and function of the Mur enzymes: Development of novel inhibitors. *Mol. Microbiol.*, **47**, 1–12.

Ellgaard, L., Molinari, M. and Helenius, A. (1999) Setting the standards: Quality control in the secretory pathway. *Science*, **286**, 1882–1888.

Ellgaard, L. and Helenius, A. (2003) Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.*, **4**, 181–191.

Englander, S.W. and Mayne, L. (1992) Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.*, **21**, 243–265.

Englander, S.W. and Krishna, M.M. (2001) Hydrogen exchange. *Nat. Struct. Biol.*, **8**, 741–742.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed src homology-2 domain studied by $^{15}$N NMR relaxation. *Biochemistry*, **33**, 5984–6003.

Farzan, M., Choe, H., Vaca, L., Martin, K., Sun, Y., Desjardins, E., Ruffing, N., Wu, L.J., Wyatt, R., Gerard, N., Gerard, C. and Sodroski, J. (1998) A tyrosine-rich region in the n terminus of CCR5 is important for human immunodeficiency virus type 1 entry and mediates an association between gp120 and CCR5. *J. Virol.*, **72**, 1160-1164.

Feenstra, K.A., Hess, B. and Berendsen, H.J.C. (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.*, **20**, 786-798.

Feenstra, K.A., Peter, C., Scheek, R.M., van Gunsteren, W.F. and Mark, A.E. (2002) A comparison of methods for calculating NMR cross-relaxation rates (NOESY and ROESY intensities) in small peptides. *J. Biomol. NMR*, **23**, 181-194.

Feher, K., Pristovsek, P., Szilagyi, L., Ljevakovic, D. and Tomasic, J. (2003) Modified glycopeptides related to cell wall peptidoglycan: Conformational studies by NMR and molecular modelling. *Bioorg. Med. Chem.*, **11**, 3133-3140.

Fermandjian, S., Perly, B., Level, M. and Lefrancier, P. (1987) A comparative $^1$H NMR Study of MurNAc-Ala-iGln (MDP) and its analogue murabutide: Evidence for a structure involving two successive β-turns in MDP. *Carbohydr. Res.*, **162**, 23-32.

Fernandez, F., D'Alessio, C., Fanchiotti, S. and Parodi, A.J. (1998) A misfolded protein conformation is not a sufficient condition for in vivo glucosylation by the UDP-Glc:Glycoprotein glucosyltransferase. *EMBO J.*, **17**, 5877-5886.

Fersht, A.R. and Daggett, V. (2002) Protein folding and unfolding at atomic resolution. *Cell*, **108**, 573-582.

Fiaux, J., Bertelsen, E.B., Horwich, A.L. and Wüthrich, K. (2002) NMR analysis of a 900K GroEL GroES complex. *Nature*, **418**, 207-211.

Flower, D.R. (1996) The lipocalin protein family: Structure and function. *Biochem. J.*, **318 ( Pt 1)**, 1-14.

Gallo, S.A., Puri, A. and Blumenthal, R. (2001) HIV-1 gp41 six-helix bundle formation occurs rapidly after the engagement of gp120 by CXCR4 in the HIV-1 Env-mediated fusion process. *Biochemistry*, **40**, 12231-12236.

Gallo, S.A., Finnegan, C.M., Viard, M., Raviv, Y., Dimitrov, A., Rawat, S.S., Puri, A., Durell, S. and Blumenthal, R. (2003) The HIV Env-mediated fusion reaction. *Biochim. Biophys. Acta*, **1614**, 36-50.

Gardner, K. and Kay, L.E. (1998) The use of $^2$H, $^{13}$C, $^{15}$N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357-406.

Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B.H., Bhattacharya, T. and Korber, B. (2002) AIDS - diversity considerations in HIV-1 vaccine selection. *Science*, **296**, 2354-2360.

Ge, M., Chen, Z., Onishi, H.R., Kohler, J., Silver, L.L., Kerns, R., Fukuzawa, S., Thompson, C. and Kahne, D. (1999) Vancomycin derivatives that inhibit peptidoglycan biosynthesis without binding D-Ala-D-Ala. *Science*, **284**, 507-511.

Gerber, P.R., Mark, A.E. and van Gunsteren, W.F. (1993) An approximate but efficient method to calculate free energy trends by computer simulation: Application to dihydrofolate reductase-inhibitor complexes. *J. Comput. Aided Mol. Des.*, **7**, 305-323.

Ghiara, J.B., Stura, E.A., Stanfield, R.L., Profy, A.T. and Wilson, I.A. (1994) Crystal-structure of the principal neutralization site of HIV-1. *Science*, **264**, 82-85.

Ghiara, J.B., Ferguson, D.C., Satterthwait, A.C., Dyson, H.J. and Wilson, I.A. (1997) Structure-based design of a constrained peptide mimic of the HIV-1 V3 loop neutralization site. *J. Mol. Biol.*, **266**, 31-39.

Glaser, F., Steinberg, D.M., Vakser, I.A. and Ben-Tal, N. (2001) Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins*, **43**, 89-102.

Gorny, M.K., Xu, J.Y., Karwowska, S., Buchbinder, A. and Zollapazner, S. (1993) Repertoire of neutralizing human monoclonal-antibodies specific for the V3 domain of HIV-1 gp120. *J. Immunol.*, **150**, 635-643.

Gorny, M.K., Williams, C., Volsky, B., Revesz, K., Cohen, S., Polonis, V.R., Honnen, W.J., Kayman, S.C., Krachmarov, C., Pinter, A. and Zolla-Pazner, S. (2002) Human monoclonal antibodies specific for conformation-sensitive epitopes of V3 neutralize human immunodeficiency virus type 1 primary isolates from various clades. *J. Virol.*, **76**, 9035-9045.

Graham, T.A., Weaver, C., Mao, F., Kimelman, D. and Xu, W. (2000) Crystal structure of a β-catenin/Tcf complex. *Cell*, **103**, 885-896.

Graham, W.H., Carter II, E.S. and Hicks, R.P. (1992) Conformational analysis of Met-enkephalin in both aquesous solution and in the presence of sodium dodecyl sulfate micelles using multidimensional NMR and molecular modeling. *Biopolymers*, **32**, 1755-1764.

Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R.R. (1988) Clean TOCSY for $^1$H spin system-identification in macromolecules. *J. Am. Chem. Soc.*, **110**, 7870-7872.

Gross, E. and Morell, J.L. (1970) Nisin. The assignment of sulfide bridges of β-methyllanthionine to a novel bicyclic structure of identical ring size. *J. Am. Chem. Soc.*, **92**, 2919-2120.

Gruebele, M. (2002) Protein folding: The free energy surface. *Curr. Opin. Struct. Biol.*, **12**, 161-168.

Guder, A., Wiedemann, I. and Sahl, H.-G. (2000) Posttranslationally modified bacteriocins – the lantibiotics. *Biopolymers*, **55**, 62-73.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, **18**, 2714-2723.

Hansson, T., Oostenbrink, C. and van Gunsteren, W.F. (2002) Molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, **12**, 190-196.

Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T. and Kim, P.S. (1998) High-resolution protein design with backbone freedom. *Science*, **282**, 1462-1467.

Herrmann, T., Güntert, P. and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, **319**, 209-227.

Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463-1472.

Holm, L. and Sander, C. (1996) Mapping the protein universe. *Science*, **273**, 595-603.

Holmquist, M. (1998) Insights into the molecular basis for fatty acyl specificities of lipases from geotrichum candidum and candida rugosa. *Chem. Phys. Lipids*, **93**, 57-66.

Honig, B. and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science*, **268**, 1144-1149.

Horii, K., Saito, M., Yoda, T., Tsumoto, K., Matsushima, M., Kuwajima, K. and Kumagai, I. (2001) Contribution of Thr29 to the thermodynamic stability of goat $\alpha$-lactalbumin as determined by experimental and theoretical approaches. *Proteins*, **45**, 16-29.

Hsu, S.-T., Breukink, E., de Kruijff, B., Kaptein, R., Bonvin, A.M.J.J. and van Nuland, N.A.J. (2002) Mapping the targeted membrane pore formation mechanism by solution NMR: The nisin Z and lipid II interaction in SDS micelles. *Biochemistry*, **41**, 7670-7676.

Hsu, S.-T.D., Breukink, E., Bierbaum, G., Sahl, H.-G., de Kruijff, B., Kaptein, R., van Nuland, N.A.J. and Bonvin, A.M.J.J. (2003) NMR study of mersacidin and lipid II interaction in dodecylphosphocholine micelles. Conformational changes are a key to antimicrobial activity. *J. Biol. Chem.*, **278**, 13110-13117.

Hsu, S.-T.D. and Bonvin, A.M.J.J. (2004) Atomic insight into the CD4 binding-induced conformational changes in HIV-1 gp120. *Proteins*, **55**, 582-593.

Hu, Y.N., Helm, J.S., Chen, L., Ye, X.Y. and Walker, S. (2003) Ramoplanin inhibits bacterial transglycosylases by binding as a dimer to lipid II. *J. Am. Chem. Soc.*, **125**, 8736-8737.

Hughes, D. (2003) Exploiting genomics, genetics and chemistry to combat antibiotic resistance. *Nature Reviews Genetics*, **4**, 432-441.

Hünenberger, P.H., Mark, A.E. and van Gunsteren, W.F. (1995) Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular-dynamics simulations. *J. Mol. Biol.*, **252**, 492-503.

Isaacs, E.D., Schuka, A., Platzman, P.M., Hamann, D.R., Barbiellini, B. and Tulk, C.A. (1999) Covalency of the hydrogen bond in ice: A direct X-ray measurement. *Phys. Rev. Lett.*, **82**, 600-603.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA*, **98**, 4569-4574.

Jack, R.W. and Jung, G. (2000) Lantibiotics and microcins: Polypeptides with unusual chemical diversity. *Curr. Opin. Chem. Biol.*, **4**, 310-317.

Jaenicke, R. (1999) Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.*, **71**, 155-241.

Janin, J. (1995) Protein-protein recognition. *Prog. Biophys. Mol. Biol.*, **64**, 145-166.

Janin, J. and Rodier, F. (1995) Protein-protein interaction at crystal contacts. *Proteins*, **23**, 580-587.

Janin, J. (1999) Wet and dry interfaces: The role of solvent in protein-protein and protein-DNA recognition. *Struct. Fold. Des.*, **7**, R277-R279.

Janin, J., Henrick, K., Moult, J., Eyck, L.T., Sternberg, M.J., Vajda, S., Vakser, I. and Wodak, S.J. (2003) CAPRI: A critical assessment of predicted interactions. *Proteins*, **52**, 2-9.

Jardetzky, T. (2002) HIV - conformational camouflage. *Nature*, **420**, 623-624.

Jastimi, R.E., Edwards, K. and Lafleur, M. (1999) Characterization of permeability and morphological perturbations induced by Nisin on phosphatidylcholine membranes. *Biophys. J.*, **77**, 842-852.

Jeener, J., Meier, B.H., Bachman, P. and Ernst, R.R. (1982) Investigations of exchange processes by two-dimensional NMR spectroscopy.

*J. Chem. Phys.*, **71**, 4546-4553.

Jeeninga, R.E., Hoogenkamp, M., Armand-Ugon, M., de Baar, M., Verhoef, K. and Berkhout, B. (2000) Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *J. Virol.*, **74**, 3740-3751.

Johnson, B.A. and Blevins, R.A. (1994) NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603-614.

Jones, J.A., Wilkins, D.K., Smith, L.J. and Dobson, C.M. (1997) Characterisation of protein unfolding by NMR diffusion measurements. *J. Biomol. NMR*, **10**, 199-203.

Jones, S. and Thornton, J.M. (1996) Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA*, **93**, 13-20.

Jones, S., van Heyningen, P., Berman, H.M. and Thornton, J.M. (1999) Protein-DNA interactions: A structural analysis. *J. Mol. Biol.*, **287**, 877-896.

Jones, S., Daley, D.T.A., Luscombe, N.M., Berman, H.M. and Thornton, J.M. (2001) Protein-RNA interactions: A structural analysis. *Nucleic Acids Res.*, **29**, 943-954.

Jorgensen, W.L. and Tirado-Rives, J. (1988) The OPLS potential functions for proteins. Energy minimizations for crystals of cyclin peptides and crambin. *J. Am. Chem. Soc.*, **110**, 1657-1666.

Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Karplus, M. and Kushick, J. (1981) Method for estimating the configurational entropy of macromolecules. *Macromolecules*, **17**, 325-332.

Karplus, M. (2002) Molecular dynamics simulations of biomolecules. *Acc. Chem. Res.*, **35**, 321-323.

Karplus, M. and McCammon, J.A. (2003) Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.*, **9**, 646-652.

Katz, A.H. and Caufield, C.E. (2003) Structure-based design approaches to cell wall biosynthesis inhibitors. *Curr. Pharm. Des.*, **9**, 857-866.

Kay, L.E., Torchia, D.A. and Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry*, **28**, 8972-8979.

Kay, L.E., Keifer, P. and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.*, **114**, 10663-10665.

Kay, L.E. (1998) Protein dynamics from NMR. *Nat. Struct. Biol.*, **5 Suppl**, 513-517.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. and Phillips, D.C. (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, **181**, 662-666.

Kessler, H., Mierke, D.F., Saulitis, J., Seip, S., Steuernagel, S., Wein, T. and Will, M. (1992) The structure of Ro 09-0198 in different environments. *Biopolymers*, **32**, 427-433.

Kessler, H., Haessner, R. and Schuler, W. (1993) Structure of rapamycin - an NMR and molecular-dynamics investigation. *Helv. Chim. Acta*, **76**, 117-130.

Kim, C.A. and Berg, J.M. (1993) Thermodynamic beta-sheet propensities measured using a zinc-finger host peptide. *Nature*, **362**, 267-270.

Kittel, C. and Kroemer, H. (1980) *Thermal dynamics*. W. H. Freeman and Company, New York.

Knarr, G., Modrow, S., Todd, A., Gething, M.J. and Buchner, J. (1999) BiP-binding sequences in HIV gp160. Implications for the binding specificity of bip. *J. Biol. Chem.*, **274**, 29850-29857.

Koch, A.L. (2003) Bacterial wall as target for attack: Past, present, and future research. *Clin. Microbiol. Rev.*, **16**, 673-687.

Koharudin, L.M., Bonvin, A.M.J.J., Kaptein, R. and Boelens, R. (2003) Use of very long-distance NOEs in a fully deuterated protein: An approach for rapid protein fold determination. *J. Magn. Reson.*, **163**, 228-235.

Kollman, P. (1993) Free energy calculations: Application to chemical and biochemical phenomena. *Chem. Rev.*, **93**, 2395-2417.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51-55.

Krapp, S., Mimura, Y., Jefferis, R., Huber, R. and Sondermann, P. (2003) Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J. Mol. Biol.*, **325**, 979-989.

Kraulis, P.J. (1991) Molscript – a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L. and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.

Kuipers, O.P., Rollema, H.S., Yap, W.M., Boot, H.J., Siezen, R.J. and de Vos, W.M. (1992) Engineering dehydrated amino acid residues in the antimicrobial peptide nisin. *J. Biol. Chem.*, **267**, 24340–24346.

Kuipers, O.P., Rollema, H.S., de Vos, W.M. and Siezen, R.J. (1993) Biosynthesis and secretion of a precursor of nisin Z by *Lactococcus lactis*, directed by the leader peptide of the homologous lantibiotic subtilin from bacillus subtilis. *FEBS Lett.*, **330**, 23–27.

Kuipers, O.P., Bierbaum, G., Ottenwalder, B., Dodd, H.M., Horn, N., Metzger, J., Kupke, T., Gnau, V., Bongers, R., van den Bogaard, P., Kosters, H., Rollema, H.S., de Vos, W.M., Siezen, R.J., Jung, G., Gotz, F., Sahl, H.G. and Gasson, M.J. (1996) Protein engineering of lantibiotics. *Antonie Van Leeuwenhoek*, **69**, 161–169.

Kumar, S., Ma, B.Y., Tsai, C.J., Sinha, N. and Nussinov, R. (2000) Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Sci.*, **9**, 10–19.

Kutateladze, T. and Overduin, M. (2001) Structural mechanism of endosome docking by the FYVE domain. *Science*, **291**, 1793–1796.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J. and Hendrickson, W.A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, **393**, 648–659.

Kwong, P.D., Wyatt, R., Majeed, S., Robinson, J., Sweet, R.W., Sodroski, J. and Hendrickson, W.A. (2000a) Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure*, **8**, 1329–1339.

Kwong, P.D., Wyatt, R., Sattentau, Q.J., Sodroski, J. and Hendrickson, W.A. (2000b) Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus. *J. Virol.*, **74**, 1961–1972.

Kwong, P.D., Doyle, M.L., Casper, D.J., Cicala, C., Leavitt, S.A., Majeed, S., Steenbeke, T.D., Venturi, M., Chaiken, I., Fung, M., Katinger, H., Parren, P., Robinson, J., Van Ryk, D., Wang, L.P., Burton, D.R., Freire, E., Wyatt, R., Sodroski, J., Hendrickson, W.A. and Arthos, J. (2002) HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature*, **420**, 678–682.

Laberge, M. (1998) Intrinsic protein electric fields: Basic non-covalent interactions and relationship to protein-induced stark effects. *Biochim. Biophys. Acta-Protein Struct. Molec. Enzym.*, **1386**, 305–330.

Ladbury, J.E., Wright, J.G., Sturtevant, J.M. and Sigler, P.B. (1994) A thermodynamic study of the Trp repressor-operator interaction. *J. Mol. Biol.*, **238**, 669–681.

Land, A., Zonneveld, D. and Braakman, I. (2003) Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *FASEB J.*, **17**, 1058–1067.

Larsen, T.A., Olson, A.J. and Goodsell, D.S. (1998) Morphology of protein-protein interfaces. *Structure*, **6**, 421–427.

Layne, S.P., Merges, M.J., Dembo, M., Spouge, J.L., Conley, S.R., Moore, J.P., Raina, J.L., Renz, H., Gelderblom, H.R. and Nara, P.L. (1992) Factors underlying spontaneous inactivation and susceptibility to neutralization of human immunodeficiency virus. *Virology*, **189**, 695–714.

Lazar, K. and Walker, S. (2002) Substrate analogues to study cell-wall biosynthesis and its inhibition. *Curr. Opin. Chem. Biol.*, **6**, 786–793.

Leach, A.R. (2001) *Molecular modeling: Principles and applications*. Pearson Education Limited.

Leckband, D. and Israelachvili, J. (2001) Intermolecular forces in biology. *Q. Rev. Biophys.*, **34**, 105–267.

Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N. and Gregory, T.J. (1990) Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.*, **265**, 10373–10382.

Letvin, N.L., Barouch, D.H. and Montefiori, D.C. (2002) Prospects for vaccine protection against HIV-1 infection and AIDS. *Annu. Rev. Immunol.*, **20**, 73–99.

Levitt, M. (1978) Conformational preferences of amino acids in globular proteins. *Biochemistry*, **17**, 4277–4285.

Levy, R.M., Karplus, M., Kushick, J. and Perahia, D. (1984) Evaluation of the configurational entropy for proteins: Application to molecular dynamics simulations of an $\alpha$-helix. *Macromolecules*, **17**, 1370–1374.

Li, S.M., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D.J., Chesneau, A., Hao, T., Goldberg, D.S., Li, N., Martinez, M., Rual, J.F., Lamesch, P., Xu, L., Tewari, M., Wong, S.L., Zhang, L.V., Berriz, G.F., Jacotot, L., Vaglio, P., Reboul,

J., Hirozane-Kishikawa, T., Li, Q.R., Gabel, H.W., Elewa, A., Baumgartner, B., Rose, D.J., Yu, H.Y., Bosak, S., Sequerra, R., Fraser, A., Mango, S.E., Saxton, W.M., Strome, S., van den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K.C., Harper, J.W., Cusick, M.E., Roth, F.P., Hill, D.E. and Vidal, M. (2004) A map of the interactome network of the metazoan *C-elegans*. *Science*, **303**, 540-543.

Lian, L.Y., Chan, W.C., Morley, S.D., Roberts, G.C.K., Bycroft, B.W. and Jackson, D. (1992) Solution structures of nisin-A and its two major degradation productions determined by NMR. *Biochem. J.*, **283**, 312-420.

Lindahl, E., Hess, B. and van der Spoel, D. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.*, **7**, 306-317.

Linge, J.P., O'Donoghue, S.I. and Nilges, M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Method Enzymol.*, **339**, 71-90.

Linge, J.P., Williams, M.A., Spronk, C.A.E.M., Bonvin, A.M.J.J. and Nilges, M. (2003) Refinement of protein structures in explicit solvent. *Ptoreins*, **50**, 496-506.

Lins, L., Ducarme, P., Breukink, E. and Brasseur, R. (1999) Computational study of nisin interaction with model membrane. *Biochim. Biophys. Acta.*, **1420**, 111-120.

Liu, W. and Hansen, J.N. (1992) Enhancement of the chemical and antimicrobial properties of subtilin by site-directed mutagenesis. *J. Biol. Chem.*, **267**, 25078-25085.

Lo Conte, L., Chothia, C. and Janin, J. (1999) The atomic structure of protein-protein recognition sites. *J. Mol. Biol.*, **285**, 2177-2198.

Lo, M.C., Helm, J.S., Sarngadharan, G., Pelczer, I. and Walker, S. (2001) A new structure for the substrate-binding antibiotic ramoplanin. *J. Am. Chem. Soc.*, **123**, 8640-8641.

Löhr, F., Mayhew, S.G. and Rüterjans, H. (2000) Detection of scalar couplings across NH...OP and OH...OP hydrogen bonds in a flavoprotein. *J. Am. Chem. Soc.*, **122**, 9289-9295.

Long, S.B., Casey, P.J. and Beese, L.S. (2002) Reaction path of protein farnesyltransferase at atomic resolution. *Nature*, **419**, 645-650.

Ludvigsen, S., Andersen, K.V. and Poulsen, F.M. (1991) Accurate measurements of coupling-constants from 2-dimensional nuclear-magnetic-resonance spectra of proteins and determination of $\phi$ angles. *J. Mol. Biol.*, **217**, 731-736.

Luscombe, N.M., Laskowski, R.A. and Thornton, J.M. (2001) Amino acid-base interactions: A three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860-2874.

MacKerell, A.D., Brooks, B., Brooks III, C.L., Nilsson, L., Roux, B., Won, Y. and Karplus, M. (1998) CHARMM: The energy function and its parameterization with an overview of the program. In Schleyer, P.v.R. (ed.), *The encyclopedia of computational chemistry*. John Wiley & Sons, Chichester, Vol. 1, pp. 271-277.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989a) Overcoming the overlap problem in the assignment of $^1$H NMR spectra of larger proteins by use of three dimensional heteronuclear $^1$H–$^{15}$N Hartman–Hahn multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1ß. *Biochemistry*, **28**, 6150-6156.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989b) 3-dimensional heteronuclear NMR of $^{15}$N-labeled proteins. *Biochemistry*, **111**, 1515-1517.

Marrink, S.J., Tieleman, D.P. and Mark, A.E. (2000) Molecular dynamics simulation of the kinetics of spontaneous micelle formation. *J. Phys. Chem. B*, **104**, 12165-12173.

Marrink, S.J., Lindahl, E., Edholm, O. and Mark, A.E. (2001) Simulation of the spontaneous aggregation of phospholipids into bilayers. *J. Am. Chem. Soc.*, **123**, 8638-8639.

Marrink, S.J. and Mark, A.E. (2003) Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *J. Am. Chem. Soc.*, **125**, 15233-15242.

Martin, N.I., Sprules, T., Carpenter, M.R., Cotter, P.D., Hill, C., Ross, R.P. and Vederas, J.C. (2004) Structural characterization of lancticin 3147, a two-peptide lantibiotic with synergistic activity. *Biochemistry*, **43**, 4049-4056.

Matter, H., Szilagyi, L., Forgo, P., Marini, Z. and Klaic, B. (1997) Structure and dynamics of a peptidoglycan monomer in aqueous solution using NMR spectroscopy and simulated annealing calculations. *J. Am. Chem. Soc.*, **119**, 2212-2223.

Mattos, C. and Ringe, D. (2001) Proteins in organic solvents. *Curr. Opin. Struct. Biol.*, **11**, 761-764.

Mayor, U., Guydosh, N.R., Johnson, C.M., Grossmann, J.G., Sato, S., Jas, G.S., Freund, S.M., Alonso, D.O., Daggett, V. and Fersht, A.R. (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature*, **421**, 863-867.

McCafferty, D.G., Cudic, P., Frankel, B.A., Barkallah, S., Kruger, R.G. and Li, W. (2002) Chemistry and biology of the ramoplanin family of peptide antibiotics. *Biopolymers*, **66**, 261-284.

McCarrick, M.A. and Kollman, P. (1994) Use of molecular dynamics and free energy perturbation calculations in anti-human immunodeficiency virus drug design. *Methods Enzymol.*, **241**, 370-384.

McKeating, J.A., McKnight, A. and Moore, J.P. (1991) Differential loss of envelope glycoprotein gp120 from virions of human immunodeficiency virus type 1 isolates: Effects on infectivity and neutralization. *J. Virol.*, **65**, 852-860.

Melikyan, G.B., Markosyan, R.M., Hemmati, H., Delmedico, M.K., Lambert, D.M. and Cohen, F.S. (2000) Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion. *J. Cell Biol.*, **151**, 413-423.

Merritt, E.A. and Bacon, D.J. (1997) Raster3D: Photorealistic molecular graphics. *Method Enzymol.*, **277**, 505-524.

Meunier, S., Spurio, R., Czisch, M., Wechselberger, R., Guenneugues, M., Gualerzi, C.O. and Boelens, R. (2000) Structure of the fMet-tRNA$^{fMet}$-binding domain of *B. Stearothermophilus* initiation factor IF2. *EMBO J.*, **19**, 1918-1926.

Minor, D.L., Jr. and Kim, P.S. (1994a) Measurement of the β-sheet-forming propensities of amino acids. *Nature*, **367**, 660-663.

Minor, D.L., Jr. and Kim, P.S. (1994b) Context is a major determinant of β-sheet propensity. *Nature*, **371**, 264-267.

Mishima, M., Hatanaka, M., Yokoyama, S., Ikegami, T., Walchli, M., Ito, Y. and Shirakawa, M. (2000) Intermolecular $^{31}P$-$^{15}N$ and $^{31}P$-$^{1}H$ scalar couplings across hydrogen bonds formed between a protein and a nucleotide. *J. Am. Chem. Soc.*, **122**, 5883-5884.

Mitchell, J.B.O., Nandi, C.L., McDonald, I.K., Thornton, J.M. and Price, S.L. (1994) Amino/aromatic interactions in proteins – is the evidence stacked against hydrogen-bonding. *J. Mol. Biol.*, **239**, 315-331.

Molinari, M. and Helenius, A. (1999) Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature*, **402**, 90-93.

Moll, G.N., Roberts, G.C.K., Konings, W.N. and Driessen, A.J.M. (1996) Mechanism of lantibiotic-induced pore-formation. *Antonie Van Leeuwenhoek*, **69**, 185-191.

Montecalvo, M.A. (2003) Ramoplanin: A novel antimicrobial agent with the potential to prevent vancomycin-resistant enterococcal infection in high-risk patients. *J. Antimicrob. Chemother.*, **51 Suppl 3**, 31-35.

Moodie, S.L. and Thornton, J.M. (1993) A study into the effects of protein binding on nucleotide conformation. *Nucleic Acids Res.*, **21**, 1369-1380.

Moore, J.P., McKeating, J.A., Weiss, R.A. and Sattentau, Q.J. (1990) Dissociation of gp120 from HIV-1 virions induced by soluble CD4. *Science*, **250**, 1139-1142.

Moore, J.P. and Ho, D.D. (1993) Antibodies to discontinuous or conformationally sensitive epitopes on the gp120 glycoprotein of human immunodeficiency virus type 1 are highly prevalent in sera of infected humans. *J. Virol.*, **67**, 863-875.

Moore, J.P. and Sodroski, J. (1996) Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. *J. Virol.*, **70**, 1863-1872.

Moore, J.P. and Binley, J. (1998) HIV – envelope's letters boxed into shape. *Nature*, **393**, 630-631.

Morelli, X., Dolla, A., Czjzek, M., Palma, P.N., Blasco, F., Krippahl, L., Moura, J.J. and Guerlesquin, F. (2000) Heteronuclear NMR and soft docking: An experimental approach for a structural model of the cytochrome c553-ferredoxin complex. *Biochemistry*, **39**, 2530-2537.

Morris, K.F. and Johnson, C.S. (1993) Resolution of discrete and continuous molecular-size distributions by means of diffusion-ordered 2D NMR-spectroscopy. *J. Am. Chem. Soc.*, **115**, 4291-4299.

Morton, C.J. and Ladbury, J.E. (1996) Water-mediated protein-DNA interactions: The relationship of thermodynamics to structural detail. *Protein Sci.*, **5**, 2115-2118.

Moulard, M. and Decroly, E. (2000) Maturation of HIV envelope glycoprotein precursors by cellular endoproteases. *Biochim. Biophys. Acta*, **1469**, 121-132.

Moult, J., Fidelis, K., Zemla, A. and Hubbard, T. (2003) Critical assessment of methods of protein structure prediction (CASP)-round V. *Proteins*, **53 Suppl 6**, 334-339.

Mulders, J.W., Boerrigter, I.J., Rollema, H.S., Siezen, R.J. and de Vos, W.M. (1991) Identification and characterization of the lantibiotic nisin Z, a natural nisin variant. *Eur. J. Biochem.*, **201**, 581-584.

Myszka, D.G., Sweet, R.W., Hensley, P., Brigham-Burke, M., Kwong, P.D., Hendrickson, W.A., Wyatt, R., Sodroski, J. and Doyle, M.L. (2000) Energetics of the HIV gp120-CD4 binding reaction. *Proc. Natl. Acad. Sci. USA*, **97**, 9026-9031.

Nagarajan, R. (1993) Structure-activity relationships of vancomycin-type glycopeptide antibiotics. *J. Antibiot. (Tokyo)*, **46**, 1181-1195.

Nanzer, A.P., Poulsen, F.M., van Gunsteren, W.F. and Torda, A.E. (1994) A reassessment of the structure of chymotrypsin inhibitor 2 (CI-2) using time-averaged NMR restraints. *Biochemistry*, **33**, 14503-14511.

Neidigh, J.W., Fesinmeyer, R.M., Prickett, K.S. and Andersen, N.H. (2001) Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. *Biochemistry*, **40**, 13188-13200.

Neuhaus, D. and Williamson, M. (1989) *The nuclear Overhauser effect: In structural and conformational analysis*. VCH, New York.

Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association - insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281-296.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1990) ^{1}H-NMR stereospecific assignments by conformational database searches. *Biopolymers*, **29**, 813-822.

Nilges, M., Macias, M.J., O'Donoghue, S.I. and Oschkinat, H. (1997) Automated NOESY interpretation with ambiguous distance restraints: The refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. Mol. Biol.*, **269**, 408-422.

Nilges, M. and Donoghue, S.I. (1998) Ambiguous NOEs and automated NOE assignment. *Prog. Nucl. Magn. Reson. Spectroscopy*, **32**, 107-139.

Nooren, I.M. and Thornton, J.M. (2003) Diversity of protein-protein interactions. *EMBO J.*, **22**, 3486-3492.

Northrup, S.H., Pear, M.R., Lee, C.Y., McCammon, J.A. and Karplus, M. (1982) Dynamical theory of activated processes in globular proteins. *Proc. Natl. Acad. Sci. USA*, **79**, 4035-4039.

Nyambi, P.N., Gorny, M.K., Bastiani, L., van der Groen, G., Williams, C. and Zolla-Pazner, S. (1998) Mapping of epitopes exposed on intact human immunodeficiency virus type 1 (HIV-1) virions: A new strategy for studying the immunologic relatedness of HIV-1. *J. Virol.*, **72**, 9384-9391.

Ofran, Y. and Rost, B. (2003) Analysing six types of protein-protein interfaces. *J. Mol. Biol.*, **325**, 377-387.

Olshevsky, U., Helseth, E., Furman, C., Li, J., Haseltine, W. and Sodroski, J. (1990) Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. *J. Virol.*, **64**, 5701-5707.

Opella, S.J. (1997) NMR and membrane proteins. *Nat. Struct. Biol.*, **4**, 845-848.

Opella, S.J., Ma, C. and Marassi, F.M. (2001) Nuclear magnetic resonance of membrane-associated peptides and proteins. *Method Enzymol.*, **339**, 285-313.

Otteken, A., Earl, P.L. and Moss, B. (1996) Folding, assembly, and intracellular trafficking of the human immunodeficiency virus type 1 envelope glycoprotein analyzed with monoclonal antibodies recognizing maturational intermediates. *J. Virol.*, **70**, 3407-3415.

Otting, G. and Wüthrich, K. (1989) Extended heteronuclear editing of 2D ^{1}H-NMR spectra of isotope-labeled proteins, using the X(ω1, ω2) double half filter. *J. Magn. Reson.*, **85**, 586-594.

Palma, P.N., Krippahl, L., Wampler, J.E. and Moura, J.J. (2000) BiGGER: A new (soft) docking algorithm for predicting protein interactions. *Proteins*, **39**, 372-384.

Palmer III, A.G. (1997) Probing molecular motion by NMR. *Curr. Opin. Struct. Biol.*, **7**, 732-737.

Palmer III, A.G. (2001) NMR probes of molecular dynamics: Overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 129-155.

Pauling, L. (1935) The structure and entropy of ice and of other crystals with some randomness of atomic arrangement. *J. Am. Chem. Soc.*, **57**, 2680-2684.

Pauling, L. and Delbrück, M. (1940) The nature of the intermolecular forces operative in biological processes. *Science*, **92**, 77-79.

Pauling, L. and Corey, R.B. (1953) Compound helical configurations of polypeptide chains: Structure of proteins of the α-keratin type. *Nature*, **171**, 59-61.

Pauling, L. (1967) *The nature of the chemical bond and the structure of molecules and crystals: An introduction of structural chemistry*. Ithaca, New York.

Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P. and Kollman, P.A. (1991) *AMBER 4.0*. University of California, San Francisco.

Peden, K., Emerman, M. and Montagnier, L. (1991) Changes in growth properties on passage in tissue culture of viruses derived from infectious molecular clones of HIV-1$_{LAI}$, HIV-1$_{MAL}$, and HIV-1$_{ELI}$. *Virology*, **185**, 661-672.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G. and North, A.C.T. (1960) Structure of haemoglobin. A three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature*, **185**, 416-422.

Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA*, **94**, 12366-12371.

Pervushin, K., Ono, A., Fernandez, C., Szyperski, T., Kainosho, M. and Wüthrich, K. (1998) NMR scalar couplings across Watson–Crick base pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proc. Natl. Acad. Sci. USA*, **95**, 14147-14151.

Pervushin, K. (2000) Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q. Rev. Biophys.*, **33**, 161-197.

Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661-665.

Pitera, J.W. and van Gunsteren, W.F. (2001) The importance of solute-solvent van der Waals interactions with interior atoms of biopolymers. *J. Am. Chem. Soc.*, **123**, 3163-3164.

Plotkin, S.S. and Onuchic, J.N. (2002a) Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q. Rev. Biophys.*, **35**, 111-167.

Plotkin, S.S. and Onuchic, J.N. (2002b) Understanding protein folding with energy landscape theory. Part II: Quantitative aspects. *Q. Rev. Biophys.*, **35**, 205-286.

Poignard, P., Saphire, E.O., Parren, P. and Burton, D.R. (2001) Gp120: Biologic aspects of structural features. *Annu. Rev. Immunol.*, **19**, 253-274.

Pollack, J.R. and Iyer, V.R. (2002) Characterizing the physical genome. *Nat. Genet.*, **32 Suppl**, 515-521.

Pollard, S.R., Rosa, M.D., Rosa, J.J. and Wiley, D.C. (1992) Truncated variants of gp120 bind CD4 with high-affinity and suggest a minimum CD4 binding region. *EMBO J.*, **11**, 585-591.

Prabu-Jeyabalan, M., Nalivaika, E. and Schiffer, C.A. (2002) Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure*, **10**, 369-381.

Prasch, T., Naumann, T., Markert, R.L., Sattler, M., Schubert, W., Schaal, S., Bauch, M., Kogler, H. and Griesinger, C. (1997) Constitution and solution conformation of the antibiotic mersacidin determined by NMR and molecular dynamics. *Eur. J. Biochem.*, **244**, 501-512.

Proba, K., Worn, A., Honegger, A. and Pluckthun, A. (1998) Antibody scFv fragments without disulfide bonds made by molecular evolution. *J. Mol. Biol.*, **275**, 245-253.

Rabut, G.E.E., Konner, J.A., Kajumo, F., Moore, J.P. and Dragic, T. (1998) Alanine substitutions of polar and nonpolar residues in the amino-terminal domain of CCR5 differently impair entry of macrophage- and dualtropic isolates of human immunodeficiency virus type 1. *J. Virol.*, **72**, 3464-3468.

Ramseier, H.R. (1960) Die Wirkung von Nisin auf *Clostridium butyricum* Prazm. *Arch. Mikrobiol.*, **37**, 57-94.

Reinhardt, W.P., Miller, M.A. and Amon, L.M. (2001) Why is it so difficult to simulate entropies, free energies, and their differences? *Acc. Chem. Res.*, **34**, 607-614.

Reisinger, P., Seidel, H., Tschesche, H. and Hammes, W.P. (1980) The effect of nisin on murein synthesis. *Arch. Microbiol.*, **127**, 187-193.

Riek, R., Wider, G., Pervushin, K. and Wüthrich, K. (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc. Natl. Acad. Sci. USA*, **96**, 4918-4923.

Rini, J.M., Stanfield, R.L., Stura, E.A., Salinas, P.A., Profy, A.T. and Wilson, I.A. (1993) Crystal-structure of a human-immunodeficiency-virus type-1 neutralizing antibody, 50.1, in complex with its V3 loop peptide antigen. *Proc. Natl. Acad. Sci. USA*, **90**, 6325-6329.

Ritter, C. and Helenius, A. (2000) Recognition of local glycoprotein misfolding by the ER folding sensor UDP-glucose:Glycoprotein glucosyltransferase. *Nat. Struc. Biol.*, **7**, 278-280.

Rizzuto, C.D., Wyatt, R., Hernandez-Ramos, N., Sun, Y., Kwong, P.D., Hendrickson, W.A. and Sodroski, J. (1998) A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science*, **280**, 1949-1953.

Roben, P., Moore, J.P., Thali, M., Sodroski, J., Barbas, C.F. and Burton, D.R. (1994) Recognition properties of a panel of human recombinant Fab fragments to the cd4 binding-site of gp120 that show differing abilities to neutralize human-immunodeficiency-virus type-1. *J. Virol.*, **68**, 4821-4828.

Rogers, H.J. and Perkins, H.R. (1968) *Cell walls and membranes*. E. & F.N. Spon Ltd., London.

Roos, J.W., Maughan, M.F., Liao, Z., Hildreth, J.E. and Clements, J.E. (2000) LuSIV cells: A reporter cell line for the detection and quantitation of a single cycle of HIV and SIV replication. *Virology*, **273**, 307-315.

Root, M.J., Kay, M.S. and Kim, P.S. (2001) Protein design of an HIV-1 entry inhibitor. *Science*, **291**, 884-888.

Ruhr, E. and Sahl, H.-G. (1985) Mode of action of the peptide antibiotic nisin and influence on the membrane potential of whole cells and on cytoplasmic and artificial membrane vesicles. *Antimicrob. Agents Chemother.*, **27**, 841-845.

Rusche, J.R., Javaherian, K., McDanal, C., Petro, J., Lynn, D.L., Grimaila, R., Langlois, A., Gallo, R.C., Arthur, L.O., Fischinger, P.J., Bolognesi, D.P., Putney, S.D. and Matthews, T.J. (1988) Antibodies that inhibit fusion of human immunodeficiency virus- infected cells bind a 24-amino acid sequence of the viral envelope, gp120. *Proc. Natl. Acad. Sci. USA*, **85**, 3198-3202.

Ryu, S.E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X.P., Xuong, N.H., Axel, R., Sweet, R.W. and et al. (1990) Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature*, **348**, 419-426.

Sahl, H.-G. and Bierbaum, G. (1998) Lantibiotics: Biosynthesis and biological activities of uniquely modified peptides from gram-positive bacteria. *Annu. Rev. Microbiol.*, **52**, 41-79.

Sailer, M., Helms, G.L., Henkel, T., Niemczura, W.P., Stiles, M.E. and Vederas, J.C. (1993) $^{15}$N- and $^{13}$C-labeled media from *Anabaena sp.* For universal isotopic labeling of bacteriocins: NMR resonance assignments of leucocin A from *leuconstoc gelidum* and nisin A from *Lactococcus lactis*. *Biochemistry*, **32**, 310-318.

Sakaida, H., Hori, T., Yonezawa, A., Sato, A., Isaka, Y., Yoshie, O., Hattori, T. and Uchiyama, T. (1998) T-tropic human immunodeficiency virus type 1 (HIV-1) derived V3 loop peptides directly bind to CXCR-4 and inhibit T-tropic HIV- 1 infection. *J. Virol.*, **72**, 9763-9770.

Sali, A., Glaeser, R., Earnest, T. and Baumeister, W. (2003) From words to literature in structural proteomics. *Nature*, **422**, 216-225.

Sanders, R.W., Schiffner, L., Master, A., Kajumo, F., Guo, Y., Dragic, T., Moore, J.P. and Binley, J.M. (2000) Variable-loop-deleted variants of the human immunodeficiency virus type 1 envelope glycoprotein can be stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits. *J. Virol.*, **74**, 5091-5100.

Sanders, R.W., de Jong, E.C., Baldwin, C.E., Schuitemaker, J.H., Kapsenberg, M.L. and Berkhout, B. (2002a) Differential transmission of human immunodeficiency virus type 1 by distinct subsets of effector dendritic cells. *J. Virol.*, **76**, 7812-7821.

Sanders, R.W., Venturi, M., Schiffner, L., Kalyanaraman, R., Katinger, H., Lloyd, K.O., Kwong, P.D. and Moore, J.P. (2002b) The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120. *J. Virol.*, **76**, 7293-7305.

Sanders, R.W. (2004) *The HIV-1 envelope glycoproteins: Folding, function and vaccine design*. Ph.D. thesis, Universiteit van Amsterdam, Amsterdam.

Santiveri, C.M., Rico, M., Jimenez, M.A., Pastor, M.T. and Perez-Paya, E. (2003) Insights into the determinants of beta-sheet stability: $^{1}$H and $^{13}$C NMR conformational investigation of three-stranded antiparallel β-sheet-forming peptides. *J. Pept. Res.*, **61**, 177-188.

Saphire, E.O., Parren, P., Pantophlet, R., Zwick, M.B., Morris, G.M., Rudd, P.M., Dwek, R.A., Stanfield, R.L., Burton, D.R. and Wilson, I.A. (2001) Crystal structure of a neutralizing human IgG against HIV-1: A template for vaccine design. *Science*, **293**, 1155-1159.

Sarkhel, S. and Desiraju, G.R. (2004) N-H...O, O-H...O, and C-H...O hydrogen bonds in protein-ligand complexes: Strong and weak interactions in molecular recognition. *Proteins*, **54**, 247-259.

Sattentau, Q.J. and Moore, J.P. (1991) Conformational-changes induced in the human-immunodeficiency- virus envelope glycoprotein by soluble CD4 binding. *J. Exp. Med.*, **174**, 407-415.

Sattentau, Q.J. and Moore, J.P. (1995) Human-immunodeficiency-virus type-1 neutralization is determined by epitope exposure on the gp120 oligomer. *J. Exp. Med.*, **182**, 185-196.

Sattler, M., Schleucher, J. and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.*, **34**, 93-158.

Scanlan, C.N., Pantophlet, R., Wormald, M.R., Ollmann Saphire, E., Stanfield, R., Wilson, I.A., Katinger, H., Dwek, R.A., Rudd, P.M. and Burton, D.R. (2002) The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of α1-2 mannose residues on the outer face of gp120. *J. Virol.*, **76**, 7306-7321.

Schäfer, H., Mark, A.E. and van Gunsteren, W.F. (2000) Absolute entropies from molecular dynamics simulation trajectories. *J. Chem. Phys.*, **113**, 7809-7817.

Schäfer, H., Daura, X., Mark, A.E. and van Gunsteren, W.F. (2001) Entropy calculations on a reversibly folding peptide: Changes in solute free energy cannot explain folding behavior. *Proteins*, **43**, 45-56.

Schäfer, H., Smith, L.J., Mark, A.E. and van Gunsteren, W.F. (2002) Entropy calculations on the molten globule state of a protein: Side-chain entropies of α–lactalbumin. *Proteins*, **46**, 215–224.

Schlitter, J. (1993) Estimation of absolute and relative entropies of macromolecules using the covariance-matrix. *Chem. Phys. Lett.*, **215**, 617–621.

Schneider, T.R., Karcher, J., Pohl, E., Lubini, P. and Sheldrick, G.M. (2000) *Ab initio* structure determination of the lantibiotic mersacidin. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 705–713.

Schreiber, G. and Fersht, A.R. (1996) Rapid, electrostatically assisted association of proteins. *Nat. Struc. Biol.*, **3**, 427–431.

Schwabe, J.W. (1997) The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.*, **7**, 126–134.

Schwartz, B., Markwalder, J.A. and Wang, Y. (2001) Lipid II: Total synthesis of the bacterial cell wall precursor and utilization as a substrate for glycosyltransfer and transpeptidation by penicillin binding protein (PBP) 1b of *Escherichia coli*. *J. Am. Chem. Soc.*, **123**, 11638–11643.

Selzer, T., Albeck, S. and Schreiber, G. (2000) Rational design of faster associating and tighter binding protein complexes. *Nat. Struct. Biol.*, **7**, 537–541.

Selzer, T. and Schreiber, G. (2001) New insights into the mechanism of protein-protein association. *Proteins*, **45**, 190–198.

Sharon, M., Kessler, N., Levy, R., Zolla-Pazner, S., Gorlach, M. and Anglister, J. (2003) Alternative conformations of HIV-1 V3 loops mimic β hairpins in chemokines, suggesting a mechanism for coreceptor selectivity. *Structure*, **11**, 225–236.

Sheinerman, F.B., Norel, R. and Honig, B. (2000) Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.*, **10**, 153–159.

Sheldrick, G.M., Jones, P.G., Kennard, O., Williams, D.H. and Smith, G.A. (1978) Structure of vancomycin and its complex with acetyl-D-alanyl-D-alanine. *Nature*, **271**, 223–225.

Shimoni, L. and Glusker, J.P. (1995) Hydrogen bonding motifs of protein side chains: Descriptions of binding of arginine and amide groups. *Protein Sci.*, **4**, 65–74.

Shoup, D., Lipari, G. and Szabo, A. (1981) Diffusion-controlled bimolecular reaction rates. The effect of rotational diffusion and orientation constraints. *Biophys. J.*, **36**, 697–714.

Shoup, D. and Szabo, A. (1982) Role of diffusion in ligand binding to macromolecules and cell-bound receptors. *Biophys. J.*, **40**, 33–39.

Shuker, S.B., Hajduk, P.J., Meadows, R.P. and Fesik, S.W. (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science*, **274**, 1531–1534.

Simonson, T., Archontis, G. and Karplus, M. (2002) Free energy simulations come of age: Protein-ligand recognition. *Acc. Chem. Res.*, **35**, 430–437.

Sinensky, M. (2000) Functional aspects of polyisoprenoid protein substituents: Roles in protein-protein interaction and trafficking. *Biochim. Biophys. Acta.*, **1529**, 203–209.

Sitia, R. and Braakman, I. (2003) Quality control in the endoplasmic reticulum protein factory. *Nature*, **426**, 891–894.

Slijper, M., Kaptein, R. and Boelens, R. (1996) Simultaneous $^{13}$C and $^{15}$N isotope editing of biomolecular complexes. Application to a mutant lac repressor headpiece DNA complex. *J. Magn. Reson. B*, **111**, 199–203.

Smith, C.K., Withka, J.M. and Regan, L. (1994) A thermodynamic scale for the β-sheet forming tendencies of the amino acids. *Biochemistry*, **33**, 5510–5517.

Smith, L., Novak, J., Rocca, J., McClung, S., Hillman, J.D. and Edison, A.S. (2000) Covalent structure of mutacin 1140 and a novel method for the rapid identification of lantibiotics. *Eur. J. Biochem.*, **267**, 6810–6816.

Smith, L.J., Daura, X. and van Gunsteren, W.F. (2002) Assessing equilibration and convergence in biomolecular simulations. *Proteins*, **48**, 487–496.

Sousa, M. and Parodi, A.J. (1995) The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:Glycoprotein glucosyltransferase. *EMBO J.*, **14**, 4196–4203.

Sousa, M.C. and Parodi, A.J. (1996) The interaction of the UDP-Glc:Glycoprotein glucosyltransferase with the acceptor glycoprotein. *Cell Mol. Biol.*, **42**, 609–616.

Sprinzak, E., Sattath, S. and Margalit, H. (2003) How reliable are experimental protein-protein interaction data? *J. Mol. Biol.*, **327**, 919–923.

Stanfield, R.L., Cabezas, E., Satterthwait, A.C., Stura, E.A., Profy, A.T. and Wilson, I.A. (1999) Dual conformations for the HIV-1 gp120 V3 loop in complexes with different neutralizing Fabs. *Struct. Fold. Des.*, **7**, 131–142.

Stanfield, R.L., Gorny, M.K., Williams, C., Zolla-Pazner, S. and Wilson, I.A. (2004) Structural rationale for the broad neutralization of HIV-1 by human monoclonal antibody 447-52d. *Structure*, **12**, 193-204.

Stein, B.S. and Engleman, E.G. (1990) Intracellular processing of the gp160 HIV-1 envelope precursor. Endoproteolytic cleavage occurs in a cis or medial compartment of the Golgi complex. *J. Biol. Chem.*, **265**, 2640-2649.

Szekat, C., Jack, R.W., Skutlarek, D., Farber, H. and Bierbaum, G. (2003) Construction of an expression system for site-directed mutagenesis of the lantibiotic mersacidin. *Appl. Environ. Microbiol.*, **69**, 3777-3783.

Takahashi, H., Nakanishi, T., Keiichiro, K., Arata, T. and Shimada, I. (2000) A novel NMR method for determining the interfaces of large protein–protein complexes. *Nat. Struc. Biol.*, **7**, 220-223.

Tame, J.R., Sleigh, S.H., Wilkinson, A.J. and Ladbury, J.E. (1996) The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat. Struct. Biol.*, **3**, 998-1001.

Taroni, C., Jones, S. and Thornton, J.M. (2000) Analysis and prediction of carbohydrate binding sites. *Protein Eng.*, **13**, 89-98.

Thali, M., Moore, J.P., Furman, C., Charles, M., Ho, D.D., Robinson, J. and Sodroski, J. (1993) Characterization of conserved human-immunodeficiency-virus type-1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J. Virol.*, **67**, 3978-3988.

Thornton, J.M., Macarthur, M.W., McDonald, I.K., Jones, D.T., Mitchell, J.B.O., Nandi, C.L., Price, S.L. and Zvelebil, M. (1993) Protein structures and complexes – what they reveal about the interactions that stabilize them. *Phiols. T. Roy. Soc. A*, **345**, 113-129.

Tironi, I.G., Sperb, R., Smith, P.E. and van Gunsteren, W.F. (1995) A generalized reaction field method for molecular-dynamics simulations. *J. Chem. Phys.*, **102**, 5451-5459.

Tjandra, N. and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, **278**, 1111-1114.

Tjandra, N., Garrett, D.S., Gronenborn, A.M., Bax, A. and Clore, G.M. (1997) Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat. Struct. Biol.*, **4**, 443-449.

Tolman, R.L., Bednarek, M.A., Johnson, B.A., Leanza, W.J., Marburg, S., Underwood, D.J., Emini, E.A. and Conley, A.J. (1993) Cyclic V3-loop-related HIV-1 conjugate vaccines. Synthesis, conformation and immunological properties. *Int. J. Pept. Protein Res.*, **41**, 455-466.

Tolstorukov, M.Y., Jernigan, R.L. and Zhurkin, V.B. (2004) Protein-DNA hydrophobic recognition in the minor groove is facilitated by sugar switching. *J. Mol. Biol.*, **337**, 65-76.

Torda, A.E., Scheek, R.M. and van Gunsteren, W.F. (1990) Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. *J. Mol. Biol.*, **214**, 223-235.

Trkola, A., Dragic, T., Arthos, J., Binley, J.M., Olson, W.C., Allaway, G.P., Cheng-Mayer, C., Robinson, J., Maddon, P.J. and Moore, J.P. (1996a) CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature*, **384**, 184-187.

Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., Sullivan, N., Srinivasan, K., Sodroski, J., Moore, J.P. and Katinger, H. (1996b) Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.*, **70**, 1100-1108.

Trombetta, E.S. and Helenius, A. (1998) Lectins as chaperones in glycoprotein folding. *Curr. Opin. Struct. Biol.*, **8**, 587-592.

Tsai, C.J., Xu, D. and Nussinov, R. (1998) Protein folding via binding and vice versa. *Fold. Des.*, **3**, R71-R80.

Tsai, C.J., Kumar, S., Ma, B.Y. and Nussinov, R. (1999) Folding funnels, binding funnels, and protein function. *Protein Sci.*, **8**, 1181-1190.

Tugarinov, V., Zvi, A., Levy, R. and Anglister, J. (1999) A cis proline turn linking two beta-hairpin strands in the solution structure of an antibody-bound HIV-1(IIIB) V3 peptide. *Nat. Struct. Biol.*, **6**, 331-335.

Tugarinov, V., Zvi, A., Levy, R., Hayek, Y., Matsushita, S. and Anglister, J. (2000) NMR structure of an anti-gp120 antibody complex with a V3 peptide reveals a surface important for co-receptor binding. *Struct. Fold. Des.*, **8**, 385-395.

Tugarinov, V., Muhandiram, R., Ayed, A. and Kay, L.E. (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: Chemical shift assignments and secondary structure of malate synthase G. *J. Am. Chem. Soc.*, **124**, 10025-10035.

Turner, B.G. and Summers, M.F. (1999) Structural biology of HIV. *J. Mol. Biol.*, **285**, 1-32.

Tyers, M. and Mann, M. (2003) From genomics to proteomics. *Nature*, **422**, 193-197.

Valencia, A. and Pazos, F. (2002) Computational methods for the prediction of protein interactions. *Curr. Opin. Struct. Biol.*, **12**, 368-373.

van Anken, E. (2003) *Versatility of the ER folding factory and the making and breaking of disulfide bond*. Ph.D. thesis, Utrecht University, Utrecht.

van Asselt, E.J., Kalk, K.H. and Dijkstra, B.W. (2000) Crystallographic studies of the interactions of *Escherichia coli* lytic transglycosylase Slt35 with peptidoglycan. *Biochemistry*, **39**, 1924-1934.

van de Ven, F.J., van den Hooven, H.W., Konings, R.N. and Hilbers, C.W. (1991) NMR studies of lantibiotics. The structure of nisin in aqueous solution. *Eur. J. Biochem.*, **202**, 1181-1188.

van den Hooven, H.W. (1995) *Structure elucidation of the lantibiotic nisin in aqueous solution and in membrane-like environments*. Ph.D. thesis, Katholieke Universteit Nijmegen, Nijmegen.

van den Hooven, H.W., Doeland, C.C., van de Kamp, M., Konings, R.N., Hilbers, C.W. and van de Ven, F.J. (1996a) Three-dimensional structure of the lantibiotic nisin in the presence of membrane-mimetic micelles of dodecylphosphocholine and of sodium dodecylsulphate. *Eur. J. Biochem.*, **235**, 382-393.

van den Hooven, H.W., Spronk, C.A., van de Kamp, M., Konings, R.N., Hilbers, C.W. and van de Van, F.J. (1996b) Surface location and orientation of the lantibiotic nisin bound to membrane-mimicking micelles of dodecylphosphocholine and of sodium dodecylsulphate. *Eur. J. Biochem.*, **235**, 394-403.

van Gunsteren, W., Bakowies, D., Burgi, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Glattli, A., Hansson, T., Oostenbrink, C., Peter, C., Pitera, J., Schuler, L., Soares, T. and Yu, H.B. (2001) Molecular dynamics simulation of biomolecular systems. *Chimia*, **55**, 856-860.

van Gunsteren, W.F. and Berendsen, H.J. (1987) Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *J. Comput. Aided Mol. Des.*, **1**, 171-176.

van Gunsteren, W.F. and Berendsen, H.J.C. (1990) Computer-simulation of molecular-dynamics – methodology, applications, and perspectives in chemistry. *Angew. Chem.-Int. Edit.*, **29**, 992-1023.

van Gunsteren, W.F., Beutler, T.C., Fraternali, F., King, P.M., Mark, A.E. and Smith, P.E. (1993) Computation of free energy in practice: Choice of approximations and accuracy limiting factors. In van Gunsteren, W.F., Weiner, P.K. and Wilkinson, A.J. (eds.), *Computer simulation of biomolecular systems, theoretical and experimental applications*. Escom Science Publishers, Leiden, Vol. 2, pp. 315-348.

van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. and Tironi, I.G. (1996) *Biomolecular simulation: The GROMOS96 manual and user guide*. Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland.

van Gunsteren, W.F., Daura, X. and Mark, A.E. (2002) Computation of free energy. *Helv. Chim. Acta*, **85**, 3113-3129.

van Heijenoort, J. (1994) Biosynthesis of the bacterial peptidoglycan unit. In Ghuysen, J.-M. and Hakenbeek, R. (eds.), *Bacterial cell wall*. Elsevier Science B.V., Amsterdam, Vol. 27, pp. 39-54.

van Heijenoort, J. (2001) Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology*, **11**, 25R-36R.

van Heusden, H., de Kruijff, B. and Breukink, E. (2002) Lipid II induces an overall transmembrane orientation of the pore-forming peptide lantibiotic nisin. *Biochemistry*, **41**, 12171-12178.

van Holde, K.E., Johnson, W.C. and Ho, P.S. (1998) *Principles of physical biochemistry*. Prentice-Hall, Inc., London.

van Kraaij, C., Breukink, E., Noordermeer, M.A., Demel, R.A., Siezen, R.J., Kuipers, O.P. and de Kruijff, B. (1998) Pore formation by nisin involves translocation of its C-terminal part across the membrane. *Biochemistry*, **37**, 16033-16040.

VanNieuwenhze, M.S., Mauldin, S.C., Zia-Ebrahimi, M., Aikins, J.A. and Blaszczak, L.C. (2001) The total synthesis of lipid I. *J. Am. Chem. Soc.*, **123**, 6983-6988.

VanNieuwenhze, M.S., Mauldin, S.C., Zia-Ebrahimi, M., Winger, B.E., Hornback, W.J., Saha, S.L., Aikins, J.A. and Blaszczak, L.C. (2002) The first total synthesis of lipid II: The final monomeric intermediate in bacterial cell wall biosynthesis. *J. Am. Chem. Soc.*, **124**, 3656-3660.

Vendruscolo, M., Paci, E., Dobson, C.M. and Karplus, M. (2001) Three key residues form a critical contact network in a protein folding transition state. *Nature*, **409**, 641-645.

Verrier, F., Borman, A.M., Brand, D. and Girard, M. (1999) Role of the HIV type 1 glycoprotein 120 V3 loop in determining coreceptor usage. *Aids Res. Hum. Retrovir.*, **15**, 731-743.

Vertesy, L., Aretz, W., Bonnefoy, A., Ehlers, E., Kurz, M., Markus, A., Shiell, M., Vogel, M., Wink, J. and Kogler, H. (1999) Ala(0)-actagardine, a new lantibiotic from cultures of atagardines liguriae ATCC 31048. *J. Antibiot. (Tokyo)*, **52**, 730-741.

Vranken, W.F., Budesinsky, M., Fant, F., Boulez, K. and Borremans, F.A.M. (1995) The complete consensus V3 loop peptide of the envelope protein gp120 of HIV-1 shows pronounced helical character in solution. *FEBS Lett.*, **374**, 117-121.

Wallace, A.C., Laskowski, R.A. and Thornton, J.M. (1995) LIGPLOT – a program to generate schematic diagrams of protein ligand interactions. *Protein Eng.*, **8**, 127-134.

Wand, A.J., Ehrhardt, M.R. and Flynn, P.F. (1998) High-resolution NMR of encapsulated proteins dissolved in low- viscosity fluids. *Proc. Natl. Acad. Sci. USA*, **95**, 15299-15302.

Wang, J. (2002) Protein recognition by cell surface receptors: Physiological receptors versus virus interactions. *Trends Biochem. Sci.*, **27**, 122-126.

Wang, J.H., Yan, Y.W., Garrett, T.P.J., Liu, J.H., Rodgers, D.W., Garlick, R.L., Tarr, G.E., Husain, Y., Reinherz, E.L. and Harrison, S.C. (1990) Atomic-structure of a fragment of human CD4 containing 2 immunoglobulin-like domains. *Nature*, **348**, 411-418.

Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.

Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J. and Wiley, D.C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature*, **387**, 426-430.

Weissman, J.S. and Kim, P.S. (1992) Kinetic role of nonnative species in the folding of bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*, **89**, 9900-9904.

Wiedemann, I., Breukink, E., van Kraaij, C., Kuipers, O.P., Bierbaum, G., de Kruijff, B. and Sahl, H.-G. (2001) Specific binding of nisin to the peptidoglycan precursor lipid II combines pore formation and inhibition of cell wall biosynthesis for potent antibiotic activity. *J. Biol. Chem.*, **276**, 1772-1779.

Wiedemann, I., Benz, R. and Sahl, H.-G. (2004) Lipid II-mediated pore formation by the antibiotic nisin - A black lipid membrane study. *J. Bacteriol.*, in press.

Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A. and Smith, L.J. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, **38**, 16424-16431.

Williams, K.A., Farrow, N.A., Deber, C.M. and Kay, L.E. (1996) Structure and dynamics of bacteriophage like major coat protein in MPG micelles by solution NMR. *Biochemistry*, **35**, 5145-5157.

Williamson, J.R. (2000) Induced fit in RNA-protein recognition. *Nat. Struct. Biol.*, **7**, 834-837.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J. and Kent, S.B. (1989) Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. *Science*, **245**, 616-621.

Wodak, S.J. and Janin, J. (2003) Structural basis of macromolecular recognition. In *Protein modules and protein-protein interactions*, Vol. 61, pp. 9-73.

Wolynes, P.G., Onuchic, J.N. and Thirumalai, D. (1995) Navigating the folding routes. *Science*, **267**, 1619-1620.

Wu, L.J., Gerard, N.P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A.A., Desjardin, E., Newman, W., Gerard, C. and Sodroski, J. (1996) CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature*, **384**, 179-183.

Wyatt, R., Sullivan, N., Thali, M., Repke, H., Ho, D., Robinson, J., Posner, M. and Sodroski, J. (1993) Functional and immunological characterization of human- immunodeficiency-virus type-1 envelope glycoproteins containing deletions of the major variable regions. *J. Virol.*, **67**, 4557-4565.

Wyatt, R., Kwong, P.D., Desjardins, E., Sweet, R.W., Robinson, J., Hendrickson, W.A. and Sodroski, J.G. (1998) The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature*, **393**, 705-711.

Wyatt, R. and Sodroski, J. (1998) The HIV-1 envelope glycoproteins: Fusogens, antigens, and immunogens. *Science*, **280**, 1884-1888.

Wymore, T., Gao, X.F. and Wong, T.C. (1999) Molecular dynamics simulation of the structure and dynamics of a dodecylphosphocholine micelle in aqueous solution. *J. Mol. Struct.*, **486**, 195-210.

Xiang, S.H., Kwong, P.D., Gupta, R., Rizzuto, C.D., Casper, D.J., Wyatt, R., Wang, L.P., Hendrickson, W.A., Doyle, M.L. and Sodroski, J. (2002) Mutagenic stabilization and/or disruption of a CD4-bound state reveals distinct conformations of the human immunodeficiency virus type 1 gp120 envelope glycoprotein. *J. Virol.*, **76**, 9888-9899.

Xu, D., Lin, S.L. and Nussinov, R. (1997a) Protein binding versus protein folding: The role of hydrophilic bridges in protein associations. *J. Mol. Biol.*, **265**, 68-84.

Xu, D., Tsai, C.J. and Nussinov, R. (1997b) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.*, **10**, 999-1012.

Xu, R.X., Olejniczak, E.T. and Fesik, S.W. (1992) Stereospecific assignments and $\chi 1$ rotamers for FKBP when bound to ascomycin from $^3J_{H\alpha,H\beta}$ and $^3J_{N,H\beta}$ coupling-constants. *FEBS Lett.*, **305**, 137-143.

Ye, X.Y., Lo, M.C., Brunner, L., Walker, D., Kahne, D. and Walker, S. (2001) Better substrates for bacterial transglycosylases. *J. Am. Chem. Soc.*, **123**, 3155-3156.

Young, L., Jernigan, R.L. and Covell, D.G. (1994) A role for surface hydrophobicity in protein-protein recognition. *Protein Sci.*, **3**, 717-729.

Zasloff, M. (2002) Antimicrobial peptides of multicellular organisms. *Nature*, **415**, 389-395.

Zimmermann, N. and Jung, G. (1997) The three-dimensional solution structure of the lantibiotic murein-biosynthesis-inhibitor actagardine determined by NMR. *Eur. J. Biochem.*, **246**, 809-819.

Zuiderweg, E.R. (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry*, **41**, 1-7.

Zvi, A., Kustanovich, I., Hayek, Y., Matsushita, S. and Anglister, J. (1995) The principal neutralizing determinant of HIV-1 located in V3 of gp120 forms a 12-residue loop by internal hydrophobic interactions. *FEBS Lett.*, **368**, 267-270.

Zwick, M.B., Kelleher, R., Jensen, R., Labrijn, A.F., Wang, M., Quinnan, G.V., Parren, P. and Burton, D.R. (2003) A novel human antibody against human immunodeficiency virus type 1 gp120 is V1, V2, and V3 loop dependent and helps delimit the epitope of the broadly neutralizing antibody immunoglobulin G1 b12. *J. Virol.*, **77**, 6965-6978.

# Summary

Antibiotic-resistant super-bugs and HIV-caused AIDS have imposed major challenges to modern medicine. Choices for infectious disease treatments are gradually depleting because antibiotics are often mis- or overused. A vaccine that can effectively inhibit HIV-1 infection, which causes tens of millions of AIDS-related casualties annually, is thus far unavailable. Scientific efforts are devoted to tackle these life-threatening diseases from various angles. In this thesis, we choose to approach the problems from a molecular point of view using NMR spectroscopy and MD simulations.

Most vital cellular functions are achieved by a structured ensemble of biomolecules, *i.e.*, a biomolecular complex. The assembly process of such complexes requires specific recognition, conferred by the physical and chemical complementarity of the counterparts, which is dictated by the 3D structures. In the first part of **Chapter 1**, current progress in structural studies on biomolecules is briefly reviewed and potential pharmaceutical and biomedical applications are discussed. The growing number of 3D structures of biomolecular complexes enables systematic analysis of the different interaction modes and of the structural motifs that are utilised for specific recognition. The second and third parts then focus on the two main themes of this thesis: lantibiotics–lipid II and HIV-1 gp120–CD4 interactions, respectively. The common feature of these two systems is that both lipid II and CD4 are membrane associated receptor molecules.

Nisin is a peptide antibiotic, which kills bacteria by membrane pore formation facilitated by binding to the bacterial cell wall precursor lipid II. **Chapter 2** describes the mapping of the lipid II binding site on nisin by means of NMR titration experiments using lipid II-containing SDS micelles. Under these NMR conditions nisin binds to lipid II in a 1:1 stoichiometry. Furthermore, the strong binding affinity between nisin and lipid II is confirmed by the slow exchange NMR titration profile. Formation of the nisin/lipid II complex induces large chemical shift perturbations localised in the N-terminal part of nisin. In conjunction with the analysis of the chemical shift temperature dependency, a detailed picture emerges: only the N-terminal part of nisin, primarily the first two rings A and B, is involved in lipid II recognition. The hinge region and the C-terminal part of nisin are responsible for the subsequent membrane insertion and the assembly process leading to pore formation, as derived from fluorescence spectroscopy data and other biophysical methods.

Structure elucidation of this complex, however, was prohibited by formation of large aggregates of the interacting molecules within several days. The increase in molecular size and structural heterogeneity caused strong line broadening in the NMR spectra. This obstacle was alleviated by using a short prenyl chain variant of lipid II (3LII) and a stable and soluble nisin/3LII complex could be obtained (**Chapter 3**). Despite differences in sample conditions, the chemical shift perturbations observed in nisin upon lipid II binding are qualitatively the same in DMSO as those observed in SDS micelles, indicating a specific and unique mode of interaction between nisin and lipid II. Moreover, the improved sample condition enabled the detection of intermolecular hydrogen bonds from the measurement of cross hydrogen bond amide to phosphate scalar coupling. Based on the solution structure of the nisin/3LII complex and 3D homology modelling of several other lipid II-binding lantibiotics a *pyrophosphate cage* model is proposed. This model explains the residue conservation amongst lipid II-binding lantibiotics. This first high-resolution structure of a lantibiotic in complex with lipid II enables us to rationalise the different binding modes amongst various lipid II binding antibiotics, including vancomycin and ramoplanin.

Mersacidin is an example of type B lantibiotics and it interferes with the bacterial cell wall synthesis by binding to lipid II. **Chapter 4** describes conformational changes of mersacidin monitored by NMR in relation with non-specific DPC micelles binding and specific lipid II binding. Because of the weak binding between mersacidin and lipid II, as observed by the fast exchange profile in the NMR titration experiment, the structural characterisation was limited to describing the effect of the environment on the 3D structure of mersacidin. An important result is that the observed chemical shift perturbations are predominantly governed by conformational changes in the hinge region of mersacidin, which in turn modulate the distribution of surface charges. The observed modulation of charge accessibility has led to the conclusion that electrostatic interactions play a key role in mersacidin–lipid II recognition. This hypothesis has been confirmed later by mutagenesis studies.

Dynamics is essential for the viral entry of HIV-1 into a host cell. Recognition of the T-cell receptor CD4 by the HIV-1 envelope glycoprotein gp120 triggers a series of conformational rearrangements, which are required for the fusion of the viral and host cell membranes. **Chapter 5** describes MD simulations of HIV-1 gp120, CD4 and their complex in order to assess the structural and dynamical differences between free and bound states. Concerted motions upon CD4 binding are identified in gp120 not only at the binding interface but also at the putative gp120 trimerisation interface. Analysis of the modes of interactions between gp120 and CD4 allow us to introduce the concept of a *geometrical binding funnel* in order to describe the observed spatial distribution of interactions. The specificity of the monitored interactions is anti-correlated with the dynamics of the residues involved. The translocation of the V3 loop upon CD4 binding generates an electrostatic attraction gradient, which may contribute to the subsequent co-receptor binding.

The analysis of the MD simulations continues in **Chapter 6** where calculations of configurational entropy are used in order to assess changes upon complexation in various entropy components of gp120 and CD4. We exploit the possibility to compute the configurational entropy of subsets of atoms in the system by dissecting contributions originating from different modes of interactions and various segments of the molecules. In addition, a *combined* trajectory analysis is introduced from which new information can be extracted by appending two separate simulation trajectories. Changes in entropy after combination can provide new insight regarding the extent of overlap of the separate trajectories.

Finally, **Chapter 7** describes results of MD simulations that were performed to complement various biochemical assays and rationalise the observed restoration of folding efficiency and thus viral infectivity of a *revertant* gp120. It has been demonstrated that removal of a disulphide bond at the base of the V4 domain of gp120 by cysteine to alanine mutations can disrupt HIV-1's viral replication and gp120 activity. Through viral evolution, however, a revived revertant gp120 was identified in which two β-branched amino acids were introduced by spontaneous mutation at and near the mutated disulphide bond site. Comparison of MD simulations of the functional wild-type gp120, the dysfunctional mutant and the partially functional revertant suggests that a localised increase in stability of the interstrand hydrogen bond network observed in the revertant with respect to the mutant is very likely to be the cause of the restoration of function. An increase in local β-sheet stability/propensity might facilitate gp120 to overcome a currently unknown threshold in the ER folding quality control machinery. The experimental data obtained *via* cell biological folding assays and virological infectivity assays are in qualitative agreement with the computational data, in spite of the very different time scales that are probed by each technique,

from picoseconds to nanoseconds for MD simulations, from minutes to hours for folding assays and from days to weeks for infectivity and replication assays.

# Samenvatting

Resistente bacteriestammen en AIDS, veroorzaakt door HIV besmetting, leggen een enorme druk op de moderne geneeskunde. De mogelijkheden om infecties te bestrijden worden steeds kleiner omdat antibiotica te vaak of verkeerd worden gebruikt. Een vaccin dat effectief een infectie met HIV-1 zou kunnen voorkomen is tot op heden niet beschikbaar. Hierdoor sterven tientallen miljoenen mensen wereldwijd aan AIDS-gerelateerde aandoeningen. Wetenschappers werken vanuit verschillende invalshoeken om deze levensbedreigende verschijnselen een halt toe te roepen. In dit proefschrift hebben we gekozen om naar bovenstaande problemen te kijken vanuit een moleculaire oogpunt met behulp van NMR-spectroscopie en MD-simulaties.

De meeste vitale celfuncties worden gecontroleerd door een gestructureerd ensemble van biomoleculen, het zogenoemde biomoleculaire complex. Het proces van assemblage berust op specifieke herkenning, beschreven door de zowel fysische als chemische complementariteit, welke opgesloten ligt in de driedimensionale (3D) structuur van de partners. In het eerste gedeelte van **Hoofdstuk 1** wordt een kort overzicht gegeven van de huidige stand van zaken in structuurstudies van biomoleculen en worden hun potentiële farmaceutische en biomedische toepassingen besproken. Het groeiend aantal 3D structuren van biomoleculaire complexen maakt een systematische analyse mogelijk van de verschillende interactie vormen en van de structurele motieven die gebruikt worden voor specifieke herkenning. Het tweede en derde gedeelte van Hoofdstuk 1 richt zich op de twee voornaamste thema's van dit proefschrift: lantibiotica-lipide II en HIV-1 gp120-CD4 interacties. Het gemeenschappelijke kenmerk van deze twee systemen is dat zowel lipide II als CD4 membraangebonden receptormoleculen zijn.

Nisine is een antibiotisch peptide, dat bacteriën doodt door poriën te vormen in de membraan. Het wordt daarbij geholpen door een precursor van de bacteriële celwandsynthese, lipide II. Hoofdstuk 2 brengt die gedeelten van nisine in kaart waar lipide II aan bindt. Deze informatie is verkregen door middel van NMR-titratie-experimenten met SDS-micellen die lipide II bevatten. Nisine bindt lipide II in een 1:1 stoichiometrie onder deze NMR-condities. De sterke bindingsaffiniteit tussen nisine en lipide II wordt bevestigd door het langzame uitwisselingspatroon in de NMR-titratie-experimenten . De vorming van het nisine-lipide II complex zorgt voor grote veranderingen in de chemishe verschuivingen in het N-terminale gedeelte van nisine, in overeenkomst met de analyse van de temperatuursafhankelijkheid van de chemishe verschuivingen. Dit leidt tot het volgende gedetailleerd beeld: slechts het N-terminale gedeelte van nisine, voornamelijk de eerste twee ringen A en B, is betrokken bij de herkenning van lipide II. Fluorescentie-spectroscopie en andere biofysische methoden hebben uitgewezen dat het schanierpunt en het C-terminale gedeelte van nisine belangrijk zijn voor de daaropvolgende membraaninsertie en complexvorming die leiden tot het ontstaan van gaten in de membraan.

Structuuropheldering van dit complex bleek echter niet mogelijk omdat binnen een aantal dagen de moleculen samenklonterden tot aggregaten. De toenemende grootte van het complex en de verscheidenheid aan structuren in het monster veroorzaakten sterke lijnverbreding in de NMR-spectra. Dit probleem werd opgelost door een kortere variant van lipide II (een prenyl keten; 3LII) te gebruiken waardoor een stabiel en oplosbaar nisine-3LII complex verkregen werd in DMSO (**Hoofdstuk 3**). Ondanks de verschillende samenstellingen van de monsters werd ook bij de titratie van dit complex in DMSO kwalitatief dezelfde veranderingen in de chemishe verschuivingen waargenomen als in de SDS micellen, hetgeen wijst op een unieke en specifieke interactie tussen nisine en lipide II. Bovendien

konden door de verbeterde condities, intermoleculaire waterstofbruggen gedetecteerd worden door de scalaire waterstofbrug-kruiskoppeling van de amide naar de fosfaatgroep te meten. Op basis van de 3D structuur van het nisine-3LII complex en 3D homologie-modellering van verschillende andere lipide II-bindende lantibiotica stellen wij een *pyrofosfaatkooimodel* voor. Dit model verklaart het evolutionaire behoud van verscheidene residuen in lipide II-bindende lantibiotica. Deze eerste hogeresolutiestructuur van een lantibioticum-lipide II complex maakt het mogelijk het bindingspatroon van verschillende andere lipid-II bindende antibiotica, waaronder vancomycine en ramoplanine, te rationaliseren.

Mersacidine is een voorbeeld van een type B lantibioticum. Het verstoort de bacteriële celwandsynthese door aan lipide II te binden. **Hoofdstuk 4** beschrijft de conformationele veranderingen van mersacidine door a-specifieke binding aan DPC micellen en specifieke lipide II binding, bestudeerd met behulp van NMR-spectroscopie. Door de zwakke binding van mersacidine aan lipide II, zoals waargenomen door het snelle uitwisselingspatroon in de NMR-titratie-experimenten, blijft de karakterisering van de structuur van het complex beperkt tot het beschrijven van het effect van de directe omgeving op de 3D structuur van mersacidine. Een belangrijk resultaat is dat de veranderingen in de chemishe verschuivingen voornamelijk het gevolg zijn van conformationele veranderingen in een onverwacht scharnierpunt van mersacidine, en daarbij de ladingsverdeling aan het oppervlak moduleert. Deze modulatie in beschikbare lading heeft tot de conclusie geleid dat elektrostatische interacties een sleutelrol vervullen voor de herkenning van mersacidine en lipide II. Deze hypothese is later bevestigd door mutagenese-onderzoek.

Dynamica speelt een essentiële rol bij het viraal binnen dringen van HIV-1 in de gastcel. De herkenning van een T-cel receptor CD4 door het HIV-1 envelop glyco-eiwit gp120 zet een cascade van conformationele veranderingen in gang, welke nodig zijn voor de samensmelting van de virale- en gastcelmembranen. **Hoofdstuk 5** beschrijft MD simulaties van HIV-1 gp120, CD4 en hun complex zodat de structurele en dynamische verschillen tussen de vrije en gebonden vormen bestudeerd kunnen worden. Samenhangende bewegingen door binding van CD4 aan gp120 zijn niet alleen waargenomen aan het bindingsoppervlak maar ook aan die kant van gp120 waarvan aangenomen wordt dat deze betrokken is bij trimerisatie. De analyse van de interacties tussen gp120 en CD4 bracht ons tot het concept van de *geometrische bindingstrechter* om de waargenomen ruimtelijke, structurele verdeling van interacties te kunnen beschrijven. De specificiteit van de waargenomen interacties is niet gecorreleerd aan de dynamiek van de betrokken residuen. De verplaatsing van de V3-lus als gevolg van de binding aan CD4 zorgt voor een elektrostatische aantrekkingsgradiënt, welke zou kunnen bijdragen aan de subsequente binding van de co-receptor.

De analyse van de MD-simulaties wordt voortgezet in **Hoofdstuk 6**. Hier worden berekeningen aan de configuratie-entropie gebruikt om veranderingen in de verschillende entropiecomponenten van gp120 en CD4 vast te leggen wanneer zij een complex vormen. Er is gebruik gemaakt van de mogelijkheid om de configuratie-entropie van atoom-subgroepen in het systeem te berekenen door de bijdragen te ontleden die afkomstig zijn van verschillende interactievormen en van verschillende segmenten van de moleculen. Bovendien hebben we een gecombineerde trajectorie-analyse geïntroduceerd waaruit nieuwe informatie kan worden onttrokken door twee aparte trajectoriën samen te voegen. Verschillen in de entropie na de samenvoeging kunnen nieuwe inzichten geven in de mate van overeenstemming van de afzonderlijke trajectoriën.

Tenslotte worden in **Hoofdstuk 7** MD-simulaties beschreven die uitgevoerd werden om verscheidene biochemische studies aan te vullen en om het waargenomen herstel in de vouwingseffectiviteit en dus

de virale besmettelijkheid van een functionele variant van gp120 te verklaren. Er is aangetoond dat verwijdering van de disulfide binding aan de basis van het V4-domein van gp120, door een cysteïne te muteren naar een alanine, leidt tot verhindering van de virale replicatie van HIV-1 en tot afname van de activiteit van gp120. Echter, door virale evolutie is een functionele variant van gp120 ontdekt met twee beta-vertakte aminozuren op en dichtbij de plaats van de gemuteerde disulfide-binding. Vergelijking van MD-simulaties van het functionele wildtype gp120, een niet-functionele mutant en de gedeeltelijk functionele variant suggereert dat een gelocaliseerde toename in stabiliteit van het waterstofbrugnetwerk tussen de beta-strengen in de functionele variant ten opzichte van de mutant hoogstwaarschijnlijk de oorzaak van het herstel in activiteit is. Een toename in stabiliteit of vormingswaarschijnlijkheid van een lokale beta-sheet zou gp120 in staat kunnen stellen een tot op heden onbekende drempel in het ER-vouwingscontroleapparaat te overkomen. De experimentele data die verkregen zijn op basis van biologische vouwingsstudies aan de cel en virale infectiestudies geven kwalitatief hetzelfde beeld als de *in silico*-data, ondanks de zeer verschillende tijdschalen die bestudeerd kunnen worden met de verschillende methoden; van pico- tot nanoseconden bij MD-simulaties, tot minuten/uren bij vouwingsstudies en zelfs dagen tot weken bij infectie- en replicatiestudies.

# 中文摘要

現今醫學所面臨的兩大重要課題，一是細菌的多重抗藥性危機，二是後天免疫不全症候群（AIDS，簡稱為愛滋病）疫情的漫延。由於抗生素的不當使用及濫用，現階段台灣多重抗藥性比例已遠之高於世界各國的抗藥性表現。根據三軍總醫院針對全台灣四十三家醫院之病童在院內感染侵襲性肺炎鏈球菌的病例調查，有高達76%的院內致病細菌對盤尼西林產生抗藥性，對紅黴素的抗藥性更高於90%，與美國兒童25-30%的抗藥性呈強烈對比！許多抗生素基本上已經喪失其臨床醫療的價值。根據聯合國愛滋病組織（UNAIDS）的資料統計，愛滋病自首例被證實至今的短短二十年間已奪走近兩千兩百萬人的性命，因此愛滋病被喻為二十世紀的黑死病。愛滋病源自於感染人類免疫缺乏病毒（HIV）所導致的免疫系統衰敗。由於HIV病毒突變頻繁，現今醫療對策除了試圖控制HIV帶原者的病情之外，仍舊缺乏有效疫苗以作徹底防治。

本論文應用核磁共振光譜學（nuclear magnetic resonance spectroscopy）與分子動態電腦模擬（molecular dynamics simulations）試圖經由探索這些生物巨分子的基本辨識機制以提供細菌多重抗藥性與愛滋病新的思考方向。生物分子的辨識主要源自於彼此之間的相容與互補性。為了能在細胞內與細胞之間數以萬計的各類分子之間有效地尋找其特定的結合標的物，生物分子往往具備獨特的辨識結構與機制。這些辨識結構的互補性包涵了互動分子之間物理上幾何結構的契合性以及化學上組成官能基之非鍵結性的吸引力。為了對這些生物分子取得更深入的了解，我們藉由結構生物學建立各種生物巨分子的三度空間模型。高解析度的分子模型提供了該分子內各個組成原子的三度空間座標。藉由這些分子結構，結構生物學的豐富資訊可為相關的生物化學、藥學及醫學提供深度的探索與不同的研究方向。

核磁共振光譜學是現今少數具備決定分子結構能力的技術之一。同時它也被廣泛地運用在分子結合點的定位及分析。尤其在製藥的藥物設計與篩選過程中，它扮演了不可或缺的角色。本論文的前半部描述核磁共振光譜學之應用在研究單硫環抗生素（lanthionine-containing antibiotics）與細菌細胞壁前驅物（Lipid II）的辨識與結合機制。單硫環抗生素因為其獨特的基因轉譯後修飾（post-translation modification）之單硫環狀結構（lanthionine rings，見第一章圖一）而得其名。許多單硫環抗生素的殺菌功效需要經由Lipid II的媒

*157*

介。Lipid II是細胞壁多層網狀結構聚合反應前的最終前驅物。藉由獨特的非飽合性脂肪雙磷酸（undecaprenyl phosphate），細胞壁的基本組成單位得以由細胞質被輸送穿過細胞膜至細胞體表準備接受聚合反應（見第一章圖三及圖四）。在多種不同單硫環抗生素中，第一類單硫環抗生素利用Lipid II 作為細胞膜結合定點以有效地造成細胞膜穿孔而達成殺菌功效。其中nisin 是最早被發現的第一類單硫環抗生素。同是它也是現今唯一具有商業價值的個例。Nisin在食品加工上被廣泛地運用為乳製品及啤酒添加劑（又稱為E234）以抑制細菌滋生（見第一章圖二）。本論文所探討的mersa-cidin（見第一章圖四）屬於第二類單硫環抗生素。這類抗生素藉由與Lipid II結合抑制細胞壁合成並導致細胞死亡。

我們藉由核磁共振光譜學決定nisin與Lipid II複合物的三維結構，並詳細探討這兩者之間的結合機制。Nisin利用兩個單硫環及其胺基末稍籠罩住 Lipid II 的雙磷酸基。同時兩分子間的界面形成了數個氫鍵使其建立穩定的複合物結構。在演化的過程中，許多其他單硫環抗生素也保存相似的單硫環結構。這些家族成員因為其演化上的相似性，很可能也具有與nisin相同的Lipid II辨識機制。由於Lipid II 的雙磷酸官能基在細胞壁合成及反應物輸送的過程扮演著不可或缺的角色。我們相信未來的抗生素研發可套用nisin辨識Lipid II的基本架構作為出發點以解決日漸增長的抗藥性問題（見第二及第三章）。

我們對於第二類單硫環抗生素mersacidin進行了類似的結構研究。但由於mersacidin 與Lipid II之間較弱的親和性使得穩定的複合結構不易獲取，因此分析分子三維結構僅侷限於mersacidin在不同的實驗條件下對於其週遭環境之反應。在分析一系列結構之後，我們歸納出靜電作用力是驅使mersacidin與Lipid II結合的主要關鍵。Mersacidin的三維結構可以因應週遭的親水性或疏水性環境作大幅度調整（見第四章）。

本論文的後半部著重於HIV病毒顆粒封套醣蛋白（virion envelope glyco-protein）gp120的分子動態模擬研究。HIV病毒屬於第一類反轉錄病毒。這一類病毒之增生繁殖仰賴其宿主的基因複製機制。換言之，HIV 病毒需要將其基因資訊轉殖入宿主的基因體；對人類而驗便是病毒感染。HIV病毒侵襲的主要目標為人體免疫系統的輔助性Ｔ淋巴球。這一類型的淋巴球體表帶有特定的抗原/抗體受器蛋白（CD4）。經由HIV病毒的封套蛋白與免疫系統的受器蛋白的結合，病毒細胞膜得以與宿主細胞膜融合，最終導致病毒與宿主基因資

訊的融合。HIV病毒藉由宿主增生複製大量的病毒蛋白，使病毒個體的數量呈幾何性增長（HIV病毒的生長週期見第一章圖七）。

　　仔細檢視病毒感染的機制，我們不難發現：有效的抑制病毒細胞膜的融合也就等於抑制病毒的感染。既然所有的病毒宿主融合機制都得經由HIV病毒的封套蛋白與免疫系統的受器蛋白的結合引發，很明顯地，科學家們面臨的首要問題便是：HIV病毒的封套蛋白是如何辨識免疫系統的受器蛋白？欲有效地發展對抗HIV病毒之疫苗，我們需要取得在尚未與其受器蛋白結合前之HIV病毒的封套蛋白的結構資訊。然而受限於實驗上的種種因素，現今蛋白質晶體結構學家們僅能提供封套蛋白在與受器蛋白複合的三維晶體結構。後續熱力學及分子生物學實驗顯示HIV病毒封套蛋白在與免疫系統之受器蛋白結合前後有相當不同的三維結構。但更進一步的結構資訊仍舊相當缺乏。因此我們利用電腦模擬程式分析外套蛋白在與受器蛋白結合前後之結構與動態轉變。熱力學及分子生物學實驗上所觀察到的現象可以經由特殊的模擬軌跡分析方法得到證實（見第五及第六章）。

　　電腦模擬最大的優勢在於模擬的分子及原子動態能被精確地在特定的時間點截取下來。相較於其他研究方法，電腦模擬提供大量的分子動態資訊。也正因其大量的數學計算程序，現今的高速電腦僅能提供有限的有效計算時間範圍。生物巨分子的電腦模擬通常侷限在奈秒（十億分之一秒）與微秒（百萬分之一秒）之間。因此電腦模擬所觀察到的現象尚需實驗之輔佐與證實。本論文的最後一章結合了病毒演化實驗、細胞生物學的蛋白摺疊（Protein folding）實驗與生物巨分子動態模擬得以涵蓋廣泛的觀察時間範圍 （由奈秒至數週）及解析度（由蛋白的組成成原子至病毒與細胞的互動）。理論模擬的結果顯示外套蛋白內特定共價雙硫鍵的結構穩定性可以被局部的穩定性氫鍵網所取代。運用該模型可以解釋實驗上所觀察到的HIV病毒封套蛋白的突變與功能得失（見第七章）。經由這一系列的研究，我們了解到HIV病毒的封套蛋白結構具有高度多樣性。也正因其結構上及序列上的多變使得許多愛滋病的藥物與單株抗體（monoclonal antibody）無法持續有效地抑制病毒的入侵。未來的疫苗設計勢必將仰賴更多的生物結構資訊。雖然本論文的研究成果尚未提供完整的解決方法，但經由更多整合性的實驗方法，我們相信現階段對於其分子動態的理解將有助於日後的藥物發展。

# Curriculum Vitae

Shang-Te Danny Hsu was born in Taipei city, Taiwan, November 23, 1975. After junior high school he moved to Hsinchu city and completed his senior high school diploma from the National Experimental High School at the Science-Based Industrial Park in 1994. Thereafter he started his university education in the Department of Physics, National Tsing Hua University, Hsinchu. Under the supervision of Prof. Jya-Wei Cheng in the Department of Life Sciences, his undergraduate thesis characterised the hydration properties of a chimeric DNA/RNA hybrid duplex using nuclear magnetic resonance (NMR) spectroscopy. In 1998, he completed his bachelor degree with a double major in Physics and Life Sciences. The structural study of the DNA/RNA hybrid duplex continued. In June 2000, he received his master degree from the Department of Life Sciences with an honour membership of the Phi Tau Phi Scholastic Honour Society of the Republic of China (Taiwan). During the six years in the National Tsing Hua University, he was also in charge of photography and visual design of the Arts Centre of the university. In September 2000, he set off to Utrecht, the Netherlands, to pursue his Ph.D. degree under the guidance of Prof. Robert Kaptein, Dr. Alexandre Bonvin and Dr. Nico van Nuland in the Bijvoet Center, Utrecht University. Specialised in interactions of biomolecular complexes using solution NMR spectroscopy and molecular dynamics (MD) simulations, he will be defending his Ph.D. thesis in June 2004.

# Publications

M. Ekkelenkamp, M.G.M. Hanssen, **S.-T.D. Hsu**, A. de Jong, J. Verhoef, and N.A.J. van Nuland. "Isolation and characterization of a novel lantibiotic Epilancin 15X" *manuscript in preparation*

**S.-T.D. Hsu**, E Breukink, E. Tischenko, B. de Kruijff, M.A.G. Lutters, R. Kaptein, A.M.J.J. Bonvin and N.A.J. van Nuland "The pyrophosphate cage: the structure of the nisin/3LII complex provides a blueprint for novel antibiotics" *submitted*

Y.N. Liu, K.F. Lin, **S.-T.D. Hsu** and P.C. Lyu "Solution structure of non-specific lipid transfer protein 1 from mung bean (*Vigna radiate var. radiata*)" *manuscript in preparation*

R.W. Sanders[#], **S.-T.D. Hsu**[#], E. van Anken[#], M. Liscaljet, S. Tillemans, E. Busser, M. Dankers, I. Braakman, A.M.J.J. Bonvin, B. Berkhout "A stable beta-sheet fold can substitute for a disulfide bond in HIV-1 gp120" *manuscript in preparation*

**S.-T.D. Hsu**, C. Peter, W.F. van Gunsteren and A.M.J.J. Bonvin, "Entropy calculation of HIV-1 Env gp120, its receptor CD4 and their complex: an analysis of entropy changes upon complexation" *submitted*

**S.-T.D. Hsu** and A.M.J.J. Bonvin, "Dynamics of CD4 binding-induced conformational changes in HIV-1 gp120" *Proteins*, (2004) **55**, 582-593

**S.-T.D. Hsu**, E. Breukink, G. Bierbaum, H.-G. Sahl, B. de Kruijff, R. Kaptein, N.A.J. van Nuland and A.M.J.J. Bonvin, "NMR study of mersacidin and lipid II interaction in DPC micelles: from conformational change to antimicrobial activity" *J. Biol. Chem*. (2003) **278**, 13110-14117

**S.-T. Hsu**, E. Breukink, B. de Kruijff, R. Kaptein, A.M.J.J. Bonvin, and N.A.J. van Nuland, "Mapping the targeted membrane pore formation mechanism by solution NMR: the nisin Z and lipid II interaction in SDS micelles" *Biochemistry*, (2002) **41**, 7670-7676.

Y.-P. Tsao, L.-Y. Wang, **S.-T. Hsu**, M.L. Jain, S.-H. Chou, W.-C. Huang, and J.-W. Cheng, "The solution structure of [d(CGC)r($a_m a_m a_m$)d(TTTGCG)]$_2$." *J. Biomol. NMR*, (2001) **21**, 209-220.

**S.-T. Hsu**, M.-T. Chou, and J.-W. Cheng, "The solution structure of [d(CGC)r(aaa)d(TTTGCG)]$_2$: Hybrid junctions flanked by DNA duplexes." *Nucleic Acids Research*, (2000) **28**, 1322-1331.

**S.-T. Hsu**, M.-T. Chou, S.-H. Chou, W.-C. Huang, and J.-W. Cheng, "Hydration of [d(CGC)r(aaa)d(TTTGCG)]$_2$." *J. Mol. Biol.*, (2000) **295**, 1129-1137.

[#] These authors contributed equally

# Acknowledgement

I am deeply indebted to Rob, Rolf, Alexandre and Nico for your courage of making this risky and certainly costly decision to offer me the Ph.D. position with very little information about me. I am glad that I made it and I hope you find it worthwhile. Over the years, the daily supervision of Alexandre and Nico is the driving force that pushes me to the limit but protects me from breaking down!

This work is a collection of successful collaborations with very different disciplines and I am grateful to those who have been involved in these activities. The NMR study on nisin/lipid II interactions was initiated and powered by Eefjan (CBLE) and, despite the fact that it is meant to be a "temporary" project, my Ph.D. begins and ends with it. Hester's parallel research consolidates my part and together we are building a bigger picture. Discussions with Ben's perception make this collaboration productive and enjoyable. Before the end of 2001, I switched my focus on mersacidin and the participation of Hans-Georg and Gabriele (Institute of Medical Microbiology and Immunity, University of Bonn) formed an interesting and successful combination of microbiologists, biochemists and structural biologists. When available, Rainer, Carine, Eugene, Tammo and Rolf have helped me a great deal to implement ideas into workable NMR experiments. Patrick and Eiso are my ultimate helpers whenever I am puzzled by programme syntax. In the spring of 2002, Eelco and Ineke (BOC-1, Bijvoet Center) initiated the simulation of the HIV-1 envelope protein gp120. Alexandre introduced me to the world of molecular dynamics and Wilfred's passionate and inspiring lecture series in Leiden broadened my view. Once in a meeting, Rogier and Ben (Academic Medical Center, UvA) noticed the conformational changes at the CD4 binding site of gp120 from my molecular movie. This remark soon evolved into an independent project itself, which then lead to the collaboration with Christine and Wilfred (ETH, Zürich) on the configurational entropy analysis of gp120/CD4 complexation. During a short stay in ETH in November 2002, Roberto Lins, Chris and Indira generously provided their latest carbohydrate and lipid force fields. About Zürich and dynamics, Jolande and Wytze's "Dynamic Presentation" coaching is certainly one thing that I appreciate very much. After many discussions and refinements, the challenging and exciting team work on gp120 revertant with Eelco and Roiger is at last to be finalised and I am looking forward to the outcome of this remarkable combination of virology, cell biology and computer simulations. Finally, I would like to thank you all for your careful reading of the manuscripts and your critical comments and remarks.

Amongst many close colleagues I would like to personally thank a few: Roberto, no matter how different our paces may be, it has always been a pleasure to share room 1.10 with you. Klaartje, you are spontaneous and sensitive, and your positive attitude is infectious. I am grateful to your support and sharing that alleviated my depression. They meant a lot to me. A cheese plate in Marseilles in exchange with a chilly fondue in Beijing, Cyril, I enjoyed the discussions and trips that we have made together. Through your eyes I discover a world of a different kind, which is simple and open. From Cargese to Sanceyroux, from Murakami to Canetti, Nocky, you are a precious companion with whom I share wonderful memories and thoughtful ideas. Thinking of you always reminds me of your first sentence when we met in Florence – *"Are you going to learn Dutch?"* you asked. I find no simple term to express my appreciation to our friendship, so *thank you* for being a dear friend.

Outside the magnetic field, there are a few more friends to whom I wish to thank. Kathrin and Sascha in Wuppertal, your hospitality has always been the source of my relaxation. Isn't it amazing that we realised the trip to China with Cyril and Sandrine? Suat and Babette, our discussions about arts, societies, philosophies and religions – medieval or modern age – are fascinating. Through our conversations I somehow rediscover my identity. Jurjen, Robert and Willem, we met because of Nocky and I am really glad that we did because of your sincerity. Maggie and Qa, your warmth and generosity are your treasured characters and our common passion about travel, delicacy and design make those days memorable. How wonderful to have friends like you who always bring me laughter! Frances, you are my trustful consolation and advisor, and you are a comforting depressurisor. Yenmin, ZnMg, Yenju, Hui-Yin, Hungjen, MMJ, Kuen-Phon and Lou, my *old* NTHU friends, our directions parted since we all left Hsinchu but my affection for you does not lose in time or space. Thank you for your distant support over the years. Last but not least, Wies and Yu-Hsin, *my paranimfen*, a medical doctor-to-be and an acupuncturist. Wies, I admire your tenderness and understanding towards people. These are your treasured assets. However busy we manage to squeeze some time together. So every occasional attendance to *het Concertgebouw* and dinner session with you is particularly delightful. Yu-Hsin, thank you for those frequent conversations and for your careful treatments – remedies that work psychologically and physically. Although my stay in Utrecht will not be forever, remember, an enduring friendship shall not be limited by distance.

感謝家人們長年來的支持與信任讓我選擇走完這段另類的成長過程。宏觀，遠見是我從你們身上所學得的珍貴寶藏。希望你們能與我分享學成的喜悅。

志弘