

Solution structure of the X4 protein coded by the SARS related coronavirus reveals an immunoglobulin like fold and suggests a binding activity to integrin I domains

Karen Hänel^{1,2,†}, Thomas Stangler^{1,2,†}, Matthias Stoldt^{1,2} & Dieter Willbold^{1,2,*}

¹Forschungszentrum Jülich, Institut für Biologische Informationsverarbeitung (IBI-2), 52425, Jülich, Germany; ²Heinrich-Heine-Universität, Institut für Physikalische Biologie and BMFZ, 40225, Düsseldorf, Germany

Received 9 July 2005; accepted 11 October 2005
© 2005 National Science Council, Taipei

Key words: 7a, coronavirus, immunoglobulin fold, integrin, LFA-1, NMR structure determination, ORF8, SARS, U122, X4

Summary

The SARS related Coronavirus genome contains a variety of novel accessory genes. One of these, called ORF7a or ORF8, code for a protein, known as 7a, U122 or X4. We set out to determine the three-dimensional structure of the soluble ectodomain of this type-I transmembrane protein by nuclear magnetic resonance spectroscopy. The fold of the protein is the first member of a further variation of the immunoglobulin like beta-sandwich fold. Because X4 does not reveal significant sequence homologies to proteins in the data bases, we carried out a structure based similarity search for proteins with known function. High structural similarity to DI domains of ICAM-1 and ICAM-2, and common features in amino acid sequence between X4 and ICAM-1, suggest X4 to possess binding activity for the α_L integrin I domain of LFA-1. Further, based on this structure based prediction, potential functions of X4 in virus replication and pathogenesis are discussed.

Introduction

A novel coronavirus (CoV) has been shown to be the etiologic agent of the severe acute respiratory syndrome (SARS) epidemic, which affected about 30 countries in late 2002. The viral genome is almost 30 kb in length and contains at least 11 open reading frames, whereas the exact number depends on the strain and the minimal count of coded amino acid residues [1–3]. Coronaviruses are positive-strand RNA viruses that code for the characteristic proteins replicase (R), spike (S),

membrane (M), envelope (E), and nucleocapsid (N) proteins. In addition, SARS-CoV codes for subgroup-specific accessory proteins that are thought to be dispensable for viral replication in cell culture, but may be important for virus–host interactions and thus contribute to the virus’ fitness. The important roles of these so-called “accessory” proteins for viral infectivity, replication efficiency and pathogenic effects is well established and investigated e.g. for the human immunodeficiency virus (HIV) accessory proteins [4, 5], and for Rhinoviruses [6].

Most SARS-CoV accessory proteins do not reveal significant sequence homologies to proteins with known function. Thus, a possible approach to elucidate potential functions of these proteins may be to determine their three-dimensional structures

*To whom correspondence should be addressed. Fax +49-2461-612023; Tel.: +49-2461-612100; E-mail: dieter.willbold@uni-duesseldorf.de

†The authors Karen Hänel and Thomas Stangler contributed equally to this work.

and search for structural similarities to proteins with known functions.

As an example of such an approach to SARS-CoV accessory proteins, we and others [7] focused on the X4 protein [1], also called U122 [7], coded by a gene with the names ORF7a [8] and ORF8 [2]. ORF8 was suggested to encode a predicted protein of 122 amino acids that has no significant BLAST or FASTA matches to known proteins [2]. X4 is predicted to contain a transmembrane helix comprising residues 99–117. The predicted signal sequence is probably cleaved off between residues 15 and 16. Together these data indicate that X4 is likely to be a type I membrane protein, with the amino-terminal hydrophilic domain (residues 16–98) oriented inside the lumen of the ER/Golgi or on the surface of the cell membrane or virus particle, depending on the localization of the protein.

Recently, it was shown that X4 is expressed in SARS-CoV-infected cells [7, 9]. In addition, the signal peptide is cleaved in Vero E6 cells transfected with ORF 7a. The carboxy-terminus (KRKTE) represents a functional ER retrieval motif. Experiments to investigate the subcellular localization in Vero E6 cells show that X4 is indeed present in the ER compartment, probably in the ER/Golgi intermediate compartment [9], the *trans*-Golgi network and also in small amounts on the cell surface of infected cells [7].

It has been shown that an additional SARS CoV accessory protein, namely, X1 (U274, ORF3 or ORF3a), is expressed on the cell surface of SARS-CoV-infected cells [10, 11] and interacts with X4. In addition, X1 has been shown to interact with the SARS-CoV structural proteins E, S, and M [10].

Another study on X4 came to the conclusion, that X4 is able to induce apoptosis via the caspase pathway in various cell types [12]. This was speculated to be one of the underlying mechanisms for the pathogenesis of SARS-CoV infections.

These interesting findings and the lack of sequence homology to proteins with known function prompted us to determine the three-dimensional solution structure of the X4 ectodomain containing residues 16–99 by nuclear magnetic resonance (NMR) spectroscopy to derive hints for potential functions of X4.

Materials and methods

Cloning of X4e

A piece of DNA containing codons 16–99 and an additional carboxy-terminal arginine was obtained by a polymerase chain reaction (PCR) with six different synthetic oligonucleotides with sequences based on the published amino acid sequence of X4 (Swiss-Prot accession number: P59635) [1, 2]. The oligonucleotide (primer) sequences were adapted for optimal codon usage of highly expressed proteins in *E. coli*: primer A, CGGAATCA TATGCTGGAAGTTCTGTTCCAGGGGCC; primer B, GTTCTGTTCCAGGGGCCGAAC TGTACCACTACCAGGAATGCGTGCGTGG TACCACCGTGCTGCTGAAAGAACCG; primer C, CGTGCTGCTGAAAGAACCGTGCCC GAGCGGTACCTACGAAGGTAACAGCCCC TTCCACCCGCTGGCGGATAACA, primer D, GTACCATCCGCGCACGCGAACGCGAAGT GGGTGCTGGTGCAGGTGAGCGCGAATTT GTTATCCGCCAGCGGGTGG; primer E, CGG ATGAACAGTTTCGGGCTCACGCTACGCG ACCGCAGCTGGTAGGTGTGACGGGTAC CATCCGCGCACGCGA; primer F, CCGCTC GAGGGATCCTTAACGGGCTGTACAGTTCC TGCTGCACTTCTTCCTGACGGATGAACA GTTTCGGGC. Primers A and F contained *NdeI* and *BamHI* restriction sites, respectively, for insertion of the finally obtained PCR fragment into the multiple cloning site of pET15b. PCR was performed with Vent-polymerase (New England Biolabs). Three subsequent PCR-reactions were performed in 50 μ l volumes: the first with primers C and D for 10 cycles, the second with 1 μ l as a template taken from the first reaction and primers B and E for 10 cycles, and finally the third reaction with 1 μ l as a template taken from the second reaction and primers A and F for 30 cycles. Annealing temperature for the three reactions were 62, 60, and 58 °C, respectively. The PCR product obtained thereafter, was cleaved with *NdeI* and *BamHI* and inserted into the multiple cloning site of pET15b yielding plasmid pX4e that codes a poly-histidine stretch followed by thrombin and PreScission cleavage sites (MGSSHHHHHS SGLVPRGSHMLEVLFQGP), followed by residues 16–99 of X4 and a C-terminal arginine and a stop codon. The DNA sequence of the gene coding

the poly-histidine tagged X4e protein was confirmed by DNA sequencing.

Expression and purification of the X4 ectodomain

Vector pX4e was transformed in *E. coli* BL21 (DE3) Rosetta cells (Stratagene). The transformed cells were grown at 37 °C to 2 ml of Luria broth (LB) medium plus ampicillin (100 µg/ml) for 5–7 h. The cells were transferred into 50 ml LB-Medium and grown until they reached an OD₆₀₀ of about 2, added to 11 LB and grown until they reached an OD₆₀₀ of 0.7. Induction of X4e expression was induced with isopropyl β-D-galactopyranoside (IPTG) at a final concentration of 1 mM. After 7–9 h the cells were harvested and frozen at –20 °C. To obtain isotope labelled X4e protein, minimal medium [13] containing 1 g/l ¹⁵N-ammonium chloride, 2 g/l ¹³C-glucose, and a vitamin cocktail (5 mg/l thiamine, 1 mg/l biotin, 1 mg/l choline chloride, 1 mg/l folic acid, 1 mg/l niacinamide, 1 mg/l pantothenic acid, 1 mg/l pyridoxal hydrochloride, 0.1 mg/l riboflavin) was used instead of LB.

The cell pellet of 11 expression culture were resuspended in 50 ml buffer B (6 M GdmHCl, 20 mM tris-HCl, 500 mM NaCl, 5 mM imidazol, 1 mM 2-mercaptoethanol, pH 8.0). Cell lysis was carried out at room temperature for about 2 h. After centrifugation the supernatant was filtrated (0.44 µm) and added to a 5 ml nickel-loaded HiTrap chelating HP column (Amersham). The column was washed with 150 ml of buffer B, 100 ml of buffer W1 (6 M GdmCl, 20 mM Tris-HCl, 500 mM NaCl, 20 mM Imidazol, 1 mM 2-mercaptoethanol, pH 8.0), 10 ml buffer W2 (buffer W1 with additional 10 mM imidazol) and 10 ml buffer W3 (buffer W1 with additional 20 mM imidazol). X4e protein was eluted with 20 ml buffer E (6 M GdmCl, 20 mM Tris-HCl, 500 mM NaCl, 500 mM Imidazol, 1 mM 2-mercaptoethanol, pH 8.0).

Poly-histidine tagged X4e protein was dialyzed against buffer A (6 M GdmCl, 10 mM Tris-HCl, 100 mM Sodium phosphate, 1 mM EDTA, 10 mM 2-mercaptoethanol, pH 8.0), and subsequently against buffer D (10 mM Tris-HCl, 1 mM EDTA, 100 mM Sodium phosphate, pH 8). The denaturated moiety of the protein was removed by centrifugation and the supernatant was dialyzed against 10 mM sodium acetate pH 5, and

subsequently diluted 1:4 with buffer F (20 mM Tris/HCl pH 7, 1 mM EDTA). The poly-histidine tag was cleaved off with 100 µg PreScission (Amersham) per mg X4e at 4 °C for about 5 h.

The protease was removed by loading the mixture on a GST-column (2 ml resin; Amersham) equilibrated in PBS and subsequent washing with 6 ml buffer P (10 mM Tris/HCl, 100 mM sodium phosphate, pH 6). The protein was found in the flow path as well as in the fraction of buffer P. To remove the poly-histidine tag and uncleaved X4e protein, both fractions were loaded on 2 ml nickel-loaded nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen) equilibrated in buffer P. The column was washed with 20 ml buffer P. The flow path and the protein containing fraction from buffer P were combined and dialyzed against 1 mM Sodium acetate pH 5.

The pure protein was analyzed by SDS-polyacrylamide gel electrophoresis. Protein concentration was determined by absorption at 280 nm using the molar extinction coefficient of X4e of 6640 M⁻¹ cm⁻¹.

NMR spectroscopy

NMR samples contained 0.4 mM protein in 1 mM ²D₄ sodium acetate, pH 5.0, in 93% H₂O/7% D₂O. NMR spectra were recorded at 315 K on Varian ^{Unity}INOVA spectrometers equipped with triple-axis pulse-field-gradient (PFG) triple resonance probes and cryogenically cooled Z-axis PFG triple resonance probes at proton frequencies of 600 and 800 MHz. Uniformly ¹³C/¹⁵N-labelled protein was used for all experiments. The resonance assignment of X4e was obtained using the following experiments: ¹H–¹⁵N-HSQC, ¹H–¹³C-HSQCs, HNCACB, C(CO)NH, HNCO, HNHA, HCCH-COSY. Aromatic side chain resonances were assigned through (HB)CB(CGCD/CE)HD/HE and ¹³C-edited HSQC-NOESY experiments. Structural constraints were derived from ¹⁵N-edited NOESY-HSQC (120 ms mixing time) and aliphatic ¹³C-edited HSQC-NOESY (100 ms) experiments in the described buffer.

NMR Spectra for steady-state ¹H–¹⁵N NOE and T₂ relaxation time measurements [14] were recorded on ¹⁵N-labelled protein under the same conditions as the other NMR experiments. ¹H–¹⁵N NOE spectra were collected for sensitivity reasons at 800 MHz on a Varian Unity INOVA

spectrometer equipped with cryogenically cooled Z-axis PFG triple resonance probe. Spectra recorded with proton saturation utilized a 1 s recycle delay followed by a 3 s period of saturation, while spectra recorded in the absence of saturation employed a recycle delay of 4 s. T_2 experiments were recorded at 600 MHz on a Varian Unity INOVA spectrometer with a conventional probe due to concerns about heating during the CPMG period on the cryogenically cooled probe at 800 MHz. T_2 spectra were recorded with a recycle delays of 3 s. Values of the steady-state ^1H - ^{15}N NOE were obtained from the ratio of peak intensities of spectra recorded with and without proton saturation. Values of T_2 were determined by fitting the measured peak volumes to a single exponential decay curve.

Data evaluation and structure calculation

Based on the almost complete assignment of ^1H , ^{13}C , and ^{15}N resonances of X4e, a total of 1688 NOE-derived experimental constraints (including 623 long-range distance constraints) could be derived from three-dimensional NOESY spectra in an iterative procedure (Table 1). NOE analysis and assignment was performed using CARA [15] and ARIA [16]. Interproton distances were used directly to calibrate experimental peaks and to extract distance constraints. Lower and upper bounds for distance constraints were derived from the target distances empirically by estimation of the error as 12.5% of the target distance squared. Distances involving ambiguous constraints, methyl groups, aromatic ring protons and non-stereospecifically assigned methylene protons were treated as sum of separate contributions to the target function, known as “sum averaging” [17]. No hydrogen bonds or predetermined secondary structure elements were used as input.

Final structures were calculated with ARIA using CNS without automatic assignment. The 15 lowest-energy structures out of 100 calculated structures were further refined by a short simulated annealing run in an explicit water shell. Of those, the 10 lowest energy structures that did not show any distance constraint violation of more than 0.03 nm were used for further analysis. Geometry of the structures, structural parameters and secondary structure elements were analyzed and visualized using the programs MOLMOL [18],

PROCHECK [19] and WHATIF [20]. The coordinates have been deposited in the Protein Data Bank, with accession code 1YO4.

Results and discussion

Protein expression, purification and structure determination

^{15}N - ^{13}C -double-labelled X4 ectodomain (X4e) was prepared as described in the methods part. In the following, numbering of the amino acid residues of X4e is according to the residue's position in the mature protein, e.g. amino acid residue 1 of X4 is coded by codon number 16 of the ORF7a gene. To prepare a sample for NMR spectroscopy, 1.3 mg protein were used to obtain a 400 μM protein solution in 1 mM sodium acetate buffer, pH 5. Virtually all proton, ^{15}N , and ^{13}C resonances were assigned (Figure 1) and deposited in the BMRB data bank (accession code: 6824). Experimental NOE distance constraints were collected from a ^{13}C -HSQC-NOESY spectrum and in addition from a NOESY- ^{15}N -HSQC spectrum recorded with a ^{15}N -labelled sample. Altogether 1974 constraints were used for the final structure calculation using a simulated annealing protocol. The finally obtained structure family consisted often structures that fulfilled the experimental distance constraints with a maximum deviation of less than 0.03 nm (Figure 2a). The structures were deposited in the PDB data bank (accession code: 1YO4).

Structure description and fold classification

Residues 1–65 of the X4 ectodomain form a well defined beta-sandwich fold. Residues 66–84 appear to be unstructured, indicated by decreased heteronuclear ^1H - ^{15}N -NOE values (Figure 3) and the lack of experimental NOE-derived structural data for this part of the protein. The well structured part of X4e is built up from seven beta-strands so that four strands form one beta-sheet and three strands form a second sheet (Figure 2b). The sheets are closely packed or “sandwiched” against each other. Each sheet is amphipathic with the hydrophobic side facing inward. The larger four-stranded beta-sheet consists of strands A, G, F, and C, the smaller

Table 1. Constraints and structural statistics for the resulting 10 NMR structures of X4.

<i>Number of experimental restraints</i>	
Intra-residue unambiguous NOEs	703
Sequential unambiguous NOEs	417
Medium-range unambiguous NOEs	108
Long-range unambiguous NOEs	477
Total unambiguous NOEs	1705
Total ambiguous NOEs	269
<i>RMSD (nm) from the mean (Residues 1–65)</i>	
All backbone atoms	0.055 ± 0.010
All heavy atoms	0.095 ± 0.009
Secondary structure backbone atoms	0.037 ± 0.007
Secondary structure heavy atoms	0.081 ± 0.009
<i>Non-bonded energy values after water-refinement (kcal/mol)</i>	
van der Waals	-707 ± 15
electrostatic	-3286 ± 97
<i>RMSD from idealized covalent geometry</i>	
Bonds (nm)	0.00044 ± 0.00002
Angles (°)	0.52 ± 0.03
Impropers (°)	1.46 ± 0.15
<i>RMSD from experimental data</i>	
Distance (nm)	0.0021 ± 0.0002
<i>Number of restraint violations</i>	
Distance (> 0.03 nm)	0
Distance (> 0.01 nm)	26.9 ± 4.3
<i>Ramachandran analysis (Residues 1–65)</i>	
Residues in most favoured regions (%)	73.1
Residues in additional allowed regions (%)	24.5
Residues in generously allowed regions (%)	0.9
Residues in disallowed regions (%) ^a	1.5

Structural Statistics of the ensemble of X4 structures.

^aIn all structures, ser22 was found in a disallowed region. This may be due to fast local conformational exchange. This may lead to different local conformations that cannot be distinguished on the NMR time scale, which ultimately yields experimental NMR constraints that are in accordance only with a time-averaged structure rather than one of the limit conformations. Although the limit conformations may well be within either one of the allowed regions of the Ramachandran plot, the average structure does not.

three-stranded beta-sheet consists of strands B, E, and D. All beta-strands align in anti-parallel fashion, as it is well-known for most immunoglobulin-like domains, with the exception of strand A, which aligns parallel to strand G.

Two disulfide bonds link both sheets on opposite edges (Figure 2b) therefore stabilizing the beta-sandwich structure. At the top of the structure, defined by the BC, DE and FG loops [22], a disulfide bridge between Cys20 (BC loop) and Cys54 (end of the F strand) creates a compact tip in the structure. At the bottom, which is defined by the AB, CD and EF loops, a disulfide bridge links

Cys8 at the end of strand A with Cys43 at the end of strand E.

Not surprisingly, the solution structure of residues 1–65 of X4e obtained in the present study is very similar to the X-ray structure reported very recently [7] (Figure 2a). The overall backbone root mean squared (rmsd) between both structures is 0.11 nm.

The beta-sandwich domain measures approx. 3.5 × 2.7 × 2.0 nm, in which the longest distance corresponds to the top-bottom distance. The AGFC beta-sheet extends along the full height of the structure, whereas the BED beta-sheet is

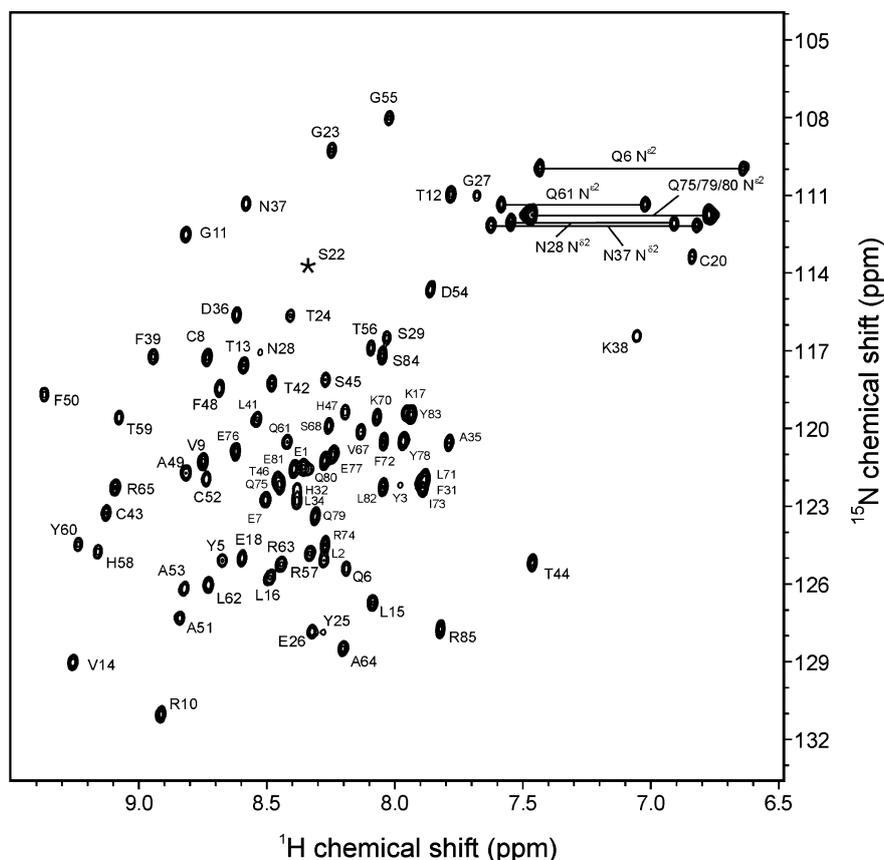


Figure 1. ^1H , ^{15}N -HSQC spectrum of SARS-CoV X4 ectodomain. Resonances are labelled according to the respective residue's position in the mature protein, e.g. residue 1 of X4 is coded by codon number 16 of the ORF7a gene. The asterisk indicates the amide resonance position of ser22, which is hardly visible, probably due to local conformational exchange.

significantly smaller and located rather to the bottom of the AGFC beta-sheet.

The overall appearance of the seven stranded beta-sandwich fold comprising residues 1–65 of X4e very much reminds one to an immunoglobulin (Ig) like fold. So far, several types of Ig like folds are described [23]. The prototype fold is called c-type with strands A, B, E, and D forming one sheet and strands C, F, and G forming the other sheet (Figure 4). A variant, called s-type has a topology with strand D being switched from the ABED sheet to the CFG sheet. This strand is then called C. X4e is a variation of the c-type with the A strand being attached to the CFG sheet. Although, it was already suggested more than 10 years ago to introduce a new subtype for such a topology within the Ig like folds [23], X4e is to our knowledge the first pure member of that subtype

of Ig-like folds. We suggest the name “p-type” in accordance with the parallel orientation of A and G strands.

A partial switch of strand A from one sheet (BED) to the other (CFG) is known from some proteins, where strand A is subdivided into A and A' with each part of the strand being attached to the BED and CFG sheets, respectively. Examples are the D1 domains of intercellular adhesion molecules ICAM-1 [24] and ICAM-2 [25]. D1 and D2 domains of the IL-1 receptor exhibit a completely switched strand A. Both domains contain, in contrast to X4e, short C' strands and were classified as “c-type” [26]. In addition, numerous “v-type” Ig-like folds are known with a nine-stranded beta-sandwich (ABED and GFCC'C'' where the A strand is completely switched from the BED sheet to the GFCC'C'' sheet (Figure 4).

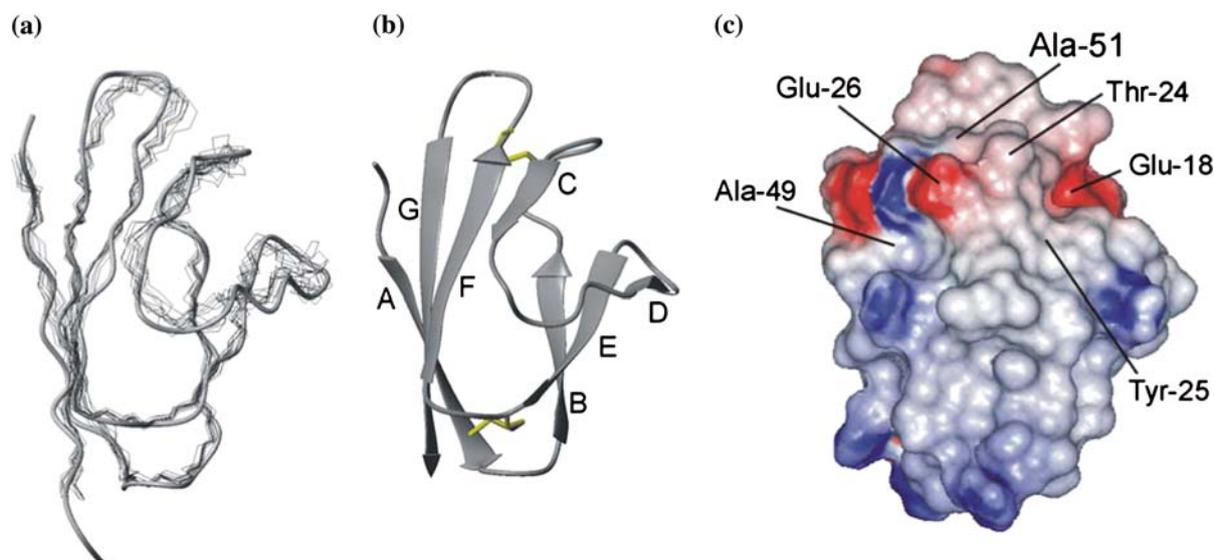


Figure 2. Solution structure of SARS-CoV X4 ectodomain after simulated annealing and refinement calculations. (a) Shown is a superposition of the protein backbones of all 10 obtained structures (fine lines) with the recently published X-ray crystal structure 1XAK (backbone worm representation, 1XAK). (b) Ribbon representation of the lowest-energy solution structure of the SARS CoV X4 ectodomain. Secondary structural elements are accentuated and labelled according to their sequential arrangement. Heavy atom sidechains of Cysteines 8, 20, 43 and 52 are shown in yellow to visualize the disulfide bridges. (c) Surface contours and charge distribution of X4 solution structure. The surface of the X4 lowest energy structure is coloured according to the electrostatic potential computed and visualized by the DelPhi module of the Accelrys Insight II molecular modelling system. Regions of basic potential are shown in blue; acidic regions are in red. Surface exposed amino acid residues of particular interest are labelled. Please note, that numbering of the amino acid residues of X4e is according to the residue's position in the mature protein, e.g. amino acid residue 1 of X4 is coded by codon number 16 of the ORF7a gene. Figures were prepared and secondary structure elements identified using MOLMOL [18, 21].

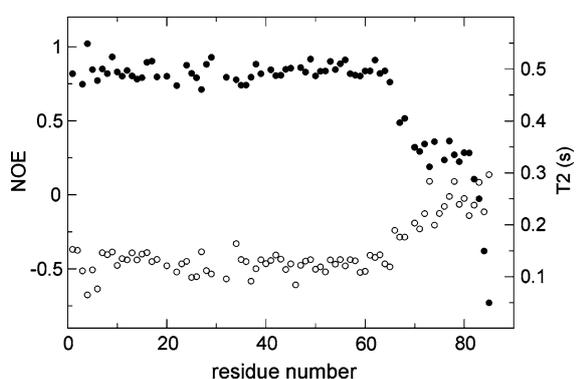


Figure 3. NMR dynamic data of SARS-CoV X4 ectodomain. Plots of ^{15}N T_2 and steady-state ^1H - ^{15}N NOE recorded at 42 °C. Steady-state ^1H - ^{15}N NOE values (black circles) were determined from data collected at 800 MHz. T_2 relaxation values (hollow circles) were determined from data collected at 600 MHz. Both ^{15}N T_2 and steady-state ^1H - ^{15}N NOE values support the hypothesis of an unstructured carboxy-terminal region encompassing residues 66–84 in contrast to the well-folded β -sandwich comprising residues 1–65. Neither ^1H - ^{15}N NOE nor ^{15}N T_2 values exhibit any special features within the β -sandwich fold.

At the top of the BED sheet, the DE loop is protruding from the structure and together with beta-strands C and D delineates a groove on the surface of X4 (Figure 2c). Central to the groove is the sidechain-carboxyl group of residue Glu 18, contributing a negative charge to the bottom of the otherwise hydrophobic groove. Such a hydrophobic patch on the molecule surface with a central negative electrostatic potential may form a potential site for ligand interaction.

The very well structured Ig-like domain of X4e comprises all residues from Glu1 to Arg65. The putative membrane-spanning segment of full-length X4 starts with residue Gln80. Residues Ser66 to Gln79 are flexible and unstructured in the solution structure, which is in accordance with the crystal structure. Although these residues might possibly form a more defined structure in the presence of a lipid bilayer, a potential role of the flexible part may be to allow the Ig-like domain to bind to membrane-distant epitopes of binding partners, or it may be a target for extracellular proteases to allow shedding of X4e.

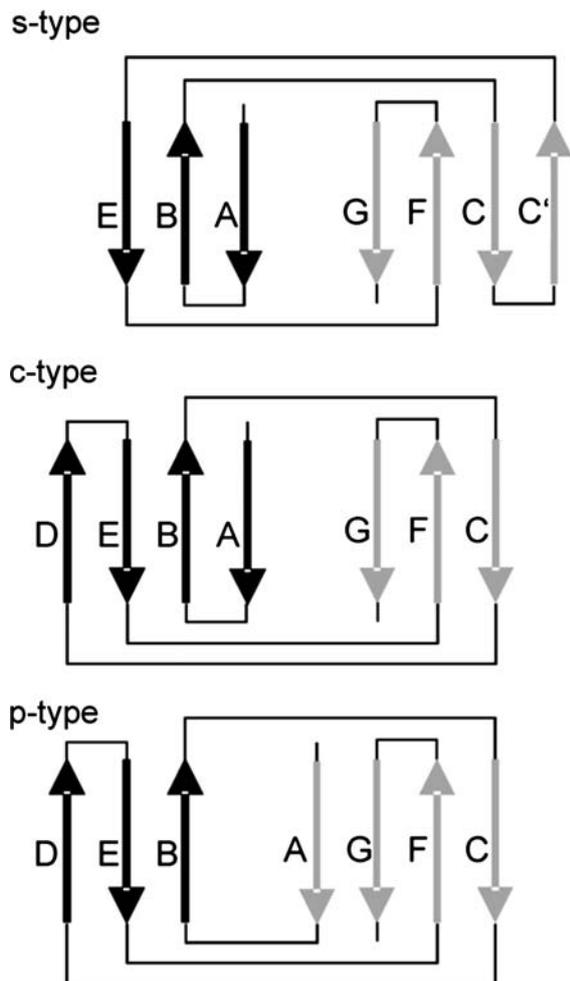


Figure 4. 2D topology diagrams of observed beta-sheet formation in various Ig-like folds based on the topology subtyping of Bork and co-workers [23]. Strands are labelled in alphabetical order from N- to C-terminus. Affiliation of each strand to either one of the two sheets is indicated by grey or black filling. Please note, that not all subtypes are shown.

Structure based prediction of X4 function

As was already found by others [2], X4 does not show any significant sequence homology to other proteins in the data bases. Therefore, no obvious function of X4 could be derived from any sequence similarity to proteins with known function. Structure based predictions of functions on the basis of similarities to proteins with known functions have been successfully used in the past [45–48], and such kind of approach is a major driving force for structural genomics projects.

In order to identify potential functions of X4 we used the solution structure to search for

proteins with similar three-dimensional structures applying the tools DALI [27] and VAST (www.ncbi.nlm.nih.gov/Structure/VAST/vast-search.html).

Employing DALI, we compared the solution structure of X4e with known structures from the Protein Data Bank (PDB). The best hit with a high reliability indicating Z-score of 4.8 is the D1 domain of ICAM-2. The structures of ICAM-2 and X4e can be aligned with each other yielding an rmsd of 0.20 nm based on 59 C α -carbon coordinate pairs (Figure 5). Another interesting hit appeared to be the D1 domain of the IL-1 receptor with a Z-score of 3.9. Both structural similarities of X4e were already found using the X4e crystal structure and DALI [7].

Using the X4e solution structure for a VAST search revealed the D1 domain of ICAM-1 to be most similar to X4e with an rmsd of 0.18 nm based on 59 C α -carbon coordinate pairs (Figure 5). ICAM-1 is very similar to ICAM-2 in respect to structure and amino acid sequence.

Comparison of X4e with ICAM-1 and ICAM-2 D1 domains

ICAM-1 and ICAM-2 are cell adhesion molecules expressed on the surface of cells, especially on endothelial cells after cytokine-mediated stimulation at inflammatory sites [28]. ICAMs belong to a subset of Ig-like superfamily proteins, which are specialized for binding to integrins. Integrins in general are non-covalently associated α/β heterodimeric transmembrane proteins, which are involved in adhesive cell–cell-interactions. ICAM-1 and ICAM-2 are known to specifically interact with lymphocyte-function-associated antigen 1 (LFA-1, CD11a/CD18, $\alpha_L\beta_2$ -integrin) that is expressed mainly on lymphocytes. ICAM LFA-1 interactions play a crucial role in lymphocyte attachment and homing to inflammation sites [29]. The ICAM-1 binding site on LFA-1 is the 180 residue containing I-domain of the α_L subunit and the binding interface is well described [30].

X4 aligns structurally very well with D1 domains of ICAM-1 and ICAM-2 (Figure 5). The most obvious differences in the structural alignment are the shortened β -strands and the corresponding BC, DE and FG loops at the top of the structure. The only topological difference between both ICAM D1 domains and X4e is

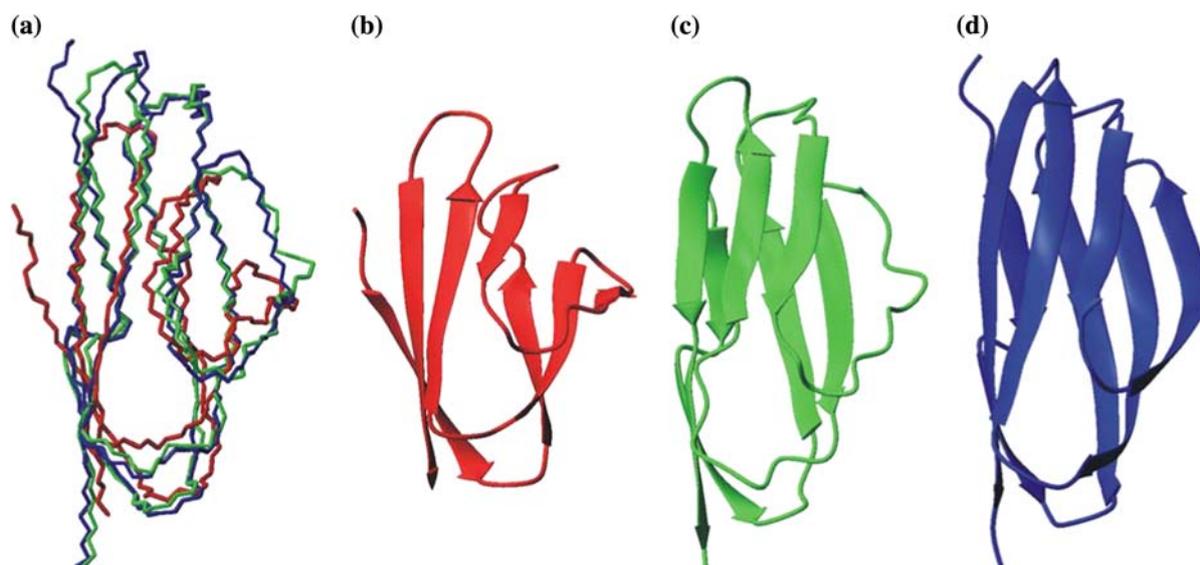


Figure 5. Structural similarity of the SARS CoV X4 ectodomain to Intercellular Adhesion Molecules (ICAMs). (a) Superposition of the lowest energy solution structure of the SARS CoV X4 ectodomain (in red) with the D1 domains of Intercellular Adhesion Molecule 1 (ICAM-1, in green) and Intercellular Adhesion Molecule 2 (ICAM-2, in blue). Shown is a backbone trace of the molecules, the structural superposition was obtained in an iterative fitting procedure with a cut-off distance of 0.5 nm and corresponds to the sequence alignment shown in Figure 4. (b–d) Ribbon representation with accentuated secondary structural elements of X4 ectodomain (b), ICAM-1 (c) and ICAM-2 (d). The orientation and colour of all molecules is the same as in panel (a). These figures were prepared using MOLMOL [18]

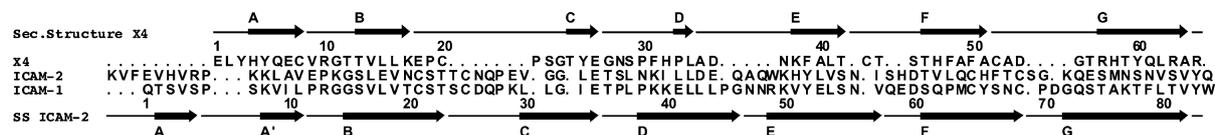


Figure 6. Structure based sequence alignment of the SARS CoV X4 ectodomain to the D1 domain of Intercellular Adhesion Molecules (ICAMs). Shown is a sequence alignment of the SARS CoV X4 ectodomain (top) with the D1 domains of Intercellular Adhesion Molecule 2 (ICAM-2, middle) and Intercellular Adhesion Molecule 1 (ICAM-1, bottom). Secondary structural elements and sequence numbering of the X4 ectodomain is shown above the sequences, secondary structural elements of ICAM-2 and sequence numbering of ICAM-1 is shown below the sequences. The sequence alignment is based on spatial proximity (cut-off distance 0.5 nm) in the structural superposition of all three structures. The alignment was prepared using UCSF Chimera [31] and ALSRIPT [32].

β -strand A, which is splitted between the β -sheets in the ICAMs, but is completely aligned to the CFG sheet in X4. Another striking difference is the somehow different location of the disulfide bonds in both ICAMs and X4e.

The structure-based sequence alignment of X4 with ICAM-1 and ICAM-2 D1 domains shows little sequence identities among the proteins (Figure 6). However, the key residue for LFA-1 interaction of ICAM-1 and ICAM-2, Glu34 and Glu37, respectively, is present in X4 at the homolog or analog sequence position (Glu26), which is the last residue of β -strand C. This glutamic acid

residue in the ICAM-1 D1 domain forms a direct coordination to the Mg^{2+} ion of the metal-ion dependent association site (MIDAS) in the LFA-1 I-domain [30]. A further characteristic feature of the I-domain binding site in ICAM-1 is a ring of hydrophobic residues around this glutamate residue (Pro36, Tyr66, Met64, and the aliphatic portions of Gln62 and Gln73) [30]. This feature can be also found in X4e, where Glu26 is surrounded by a ring of hydrophobic residues, too. (Figure 2c).

To our knowledge, besides the α_L subunit of LFA-1, no other cellular binding partner for the

D1 domain of ICAM-1 is known so far. I-domains of α_M integrin subunits are known to bind ICAM-1, but via its D3 domain [33]. Weak binding capabilities for ICAM-1 are further reported for α_X subunits [34], but presumably not via its D1 domain [35]. For α_D integrin subunits it was shown that they do not bind ICAM-1 [36].

The relevance of the very obvious structural similarity of X4e with ICAM-1 and ICAM-2 D1 domains is hard to estimate. But based on this structural similarity and the described common features of amino acid sequence and surface appearance of X4 with the well characterized α_L integrin I domain binding site on ICAM-1 D1, we suggest that X4 contains a binding site for the α_L integrin subunit I-domain of LFA-1. Although, experimental data will be needed to confirm the prediction, we carried out a modelling study on X4 and the α_L integrin subunit I-domain of LFA-1. Interestingly, the resulting complex did not show obvious steric problems to be formed (Figure 7).

Hypothetic consequences of the proposed LFA-1 binding activity of X4

The consequences of a predicted LFA-1 binding activity of X4 depend largely on the subcellular localization of X4 in infected cells or virus particles. In the following, we speculate on potential functions of X4 as a LFA-1 binding protein dependent from X4 subcellular localization based on examples from other proteins with known functions.

The presence of LFA-1 binding X4 molecules on the virus surface would allow the virus to use LFA-1 as a receptor for cell entry. Such an example is known from HIV-1, of which virus particles were found bearing incorporated host-encoded ICAM-1 on their surface, which leads to a 5- to 10-fold increase in infectivity, caused by an interaction between virally incorporated ICAM-1 and cell surface LFA-1 [37]. So far, however, there are no reports on positive detection of X4 in virus particles.

X4 was already described to be primarily located in the ER of infected cells and to contain an ER retention signal [9]. If X4 is able to bind LFA-1 it could prevent delivery of newly synthesized LFA-1 molecules from the ER to the cell surface. Prominent examples of viral accessory proteins with such functions are known from other

viruses, e.g. HIV-1 Vpu binds to CD4 and prevents CD4 delivery to the cell surface and induces even its degradation [38, 39]. LFA-1 is exclusively expressed on the surface of leukocytes including T cells and dendritic cells. It mediates several adhesive interactions between cells of the immune system, e.g. dendritic cells and T cells, B cells and T cells, T cells and their target cells, as well as the interactions of leukocytes with the endothelium and the transendothelial migration of leukocytes [40]. Loss of LFA-1 leads to severe defects of the immune system as can be seen in the leukocyte adhesion deficiency (LAD) syndrome [41].

One study reported small amounts of X4 on the surface of infected cells [7]. The presence of X4 with LFA-1 binding activity on the surface of infected cells could for example interfere with T cell homing, or increase the infected cells' affinity for leukocytes, and could even induce apoptosis in LFA-1 presenting T cells. Leukotoxin from *Actinobacillus actinomycetemcomitans* is an example of a protein that is expressed on the surface of infected cells, binds to LFA-1 on T cells, and induces apoptosis via caspase-3 dependent pathways in these cells [42]. Indeed, overexpression of X4 in Vero E6 cells induces apoptosis via a caspase-3 dependent pathway [12]. Whether LFA-1 is involved in the mechanism, however, is not known.

Lymphopenia is a common observation among SARS patients [43]. $CD4^+$ cells are more affected than $CD8^+$ cells. The reason for the lymphocyte depletion is not known. Interestingly, AIDS patients are characterized by a $CD4^+$ depletion. It is suggested that the disappearance of $CD4^+$ T cells in the blood is the result of increased migration of $CD4^+$ cells from the blood into tissues. Secondary signals through homing receptors received during the homing process induce many of these cells into apoptosis [44].

Comparison of X4e with IL-1 receptor D1 domain

A second site suitable for a protein-protein-interaction is suggested by the X4e similarity to the second best hit for structural similarity from VAST and DALI searches, namely the domain D1 of the IL-1 receptor. D1 and D2 domains of the IL-1 receptor are of the Ig-like fold with a switched strand A, like X4, but in contrast to X4 contain a short strand C'. The similarity between X4e and

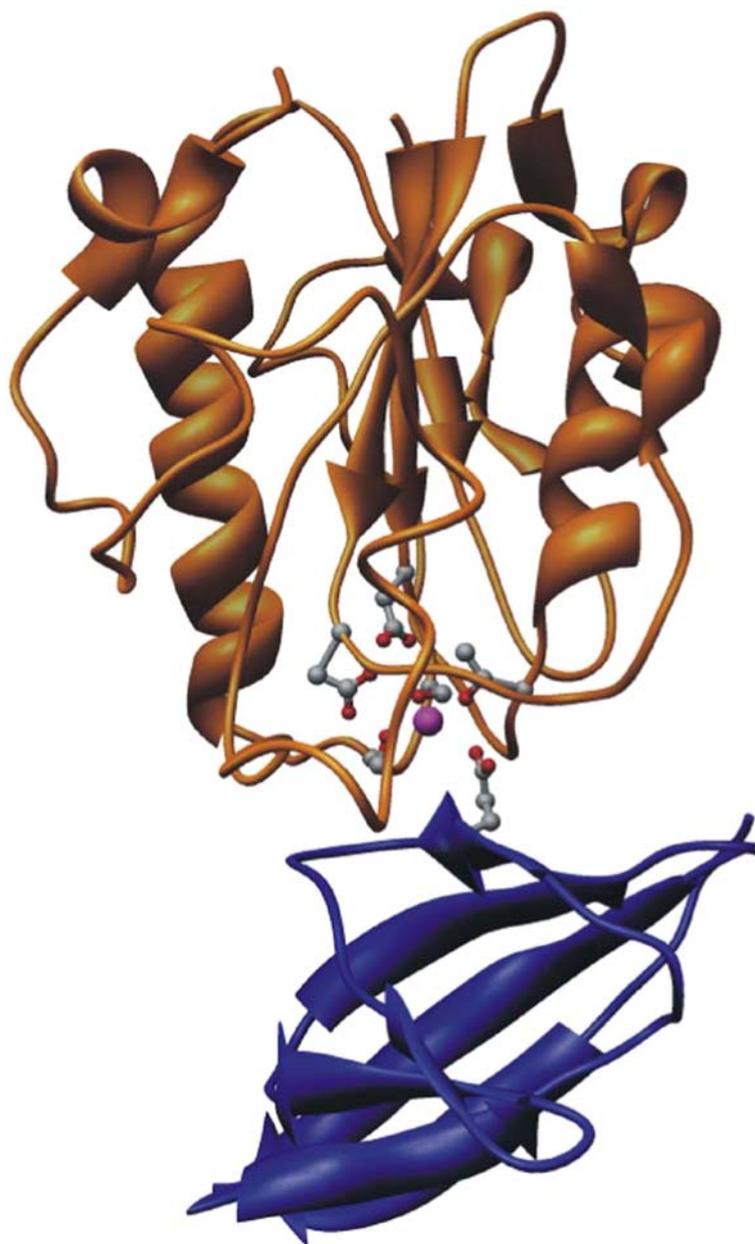


Figure 7. Modelling study of SARS CoV X4 ectodomain and the the α_L integrin subunit I-domain of LFA-1. X4 is colored in blue, the integrin I domain orange, and the Mg^{2+} ion of the metal ion-dependent adhesion site (MIDAS) is colored magenta. The I-domain MIDAS residues and Glu26 of X4 are shown as ball-and-stick models. The model was constructed based on the coordinates of the ICAM-1 complex with the intermediate affinity I-domain of $\alpha_L\beta_2$ [30] (PDB code: 1MQ8). SARS CoV X4 protein was superimposed to ICAM-1 based on protein backbone coordinates of structurally aligning residues of the potential binding interface, comprising β -strands C, D and F, and the CD loop. Sidechain torsion angles of Glu26 were adjusted for steric reasons. The model was prepared using UCSF Chimera [31].

D1 domain of IL-1 receptor is largely based on the switched strand A and the resulting low rmsd value for the majority of C_α -coordinates of both protein domains. The sequences of both domains

did not reveal the slightest similarity. An interesting feature of the IL-1 receptor D1 domain, however, is the ridge between strands A and B forming the binding site for interleukin-1. Such a

ridge can be found in X4e, as well, and was already described for the X-ray structure of X4e [7].

Taken all data together, the ectodomain of the SARS CoV coded X4 protein adopts an Ig-like fold, that shows some features resembling those of the DI domain of ICAM-1. Based thereon, we predict X4 to have a LFA-1 binding activity. Several observation from SARS patients correlate well with such an activity coded in the SARS CoV genome. In any case, an investigation of the proposed binding potential of X4 to bind LFA-1, may help to elucidate the role of X4 in replication and pathogenesis of SARS CoV.

References

- Rota P.A., Oberste M.S., Monroe S.S., Nix W.A., Campagnoli R., Icenogle J.P., Penaranda S., Bankamp B., Maher K., Chen M.H., Tong S., Tamin A., Lowe L., Frace M., DeRisi J.L., Chen Q., Wang D., Erdman D.D., Peret T.C., Burns C., Ksiazek T.G., Rollin P.E., Sanchez A., Liffick S., Holloway B., Limor J., McCaustland K., Olsen-Rasmussen M., Fouchier R., Gunther S., Osterhaus A.D., Drosten C., Pallansch M.A., Anderson L.J. and Bellini W.J., Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300: 1394–1399, 2003.
- Marra M.A., Jones S.J., Astell C.R., Holt R.A., Brooks-Wilson A., Butterfield Y.S., Khattri J., Asano J.K., Barber S.A., Chan S.Y., Cloutier A., Coughlin S.M., Freeman D., Girn N., Griffith O.L., Leach S.R., Mayo M., McDonald H., Montgomery S.B., Pandoh P.K., Petrescu A.S., Robertson A.G., Schein J.E., Siddiqui A., Smailus D.E., Stott J.M., Yang G.S., Plummer F., Andonov A., Artsob H., Bastien N., Bernard K., Booth T.F., Bowness D., Czub M., Drebot M., Fernando L., Flick R., Garbutt M., Gray M., Grolla A., Jones S., Feldmann H., Meyers A., Kabani A., Li Y., Normand S., Stroher U., Tipples G.A., Tyler S., Vogrig R., Ward D., Watson B., Brunham R.C., Krajden M., Petric M., Skowronski D.M., Upton C. and Roper R.L., The Genome sequence of the SARS-associated coronavirus. *Science* 300: 1399–1404, 2003.
- Stadler K., Masignani V., Eickmann M., Becker S., Abrignani S., Klenk H.D. and Rappuoli R., SARS – beginning to understand a new virus. *Nat. Rev. Microbiol.* 1: 209–218, 2003.
- Steffens C.M. and Hope T.J., Recent advances in the understanding of HIV accessory protein function. *Aids* 15(Suppl 5) S21–S26, 2001.
- Bour S. and Strebel K., The HIV-1 Vpu protein: a multifunctional enhancer of viral particle release. *Microbes Infect.* 5: 1029–1039, 2003.
- Stockl J., Vetr H., Majdic O., Zlabinger G., Kuechler E. and Knapp W., Human major group rhino viruses down-modulate the accessory function of monocytes by inducing IL-10. *J. Clin. Invest.* 104: 957–965, 1999.
- Nelson C.A., Pekosz A., Lee C.A., Diamond M.S. and Fremont D.H., Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. *Structure* 13: 75–85, 2005.
- Snijder E.J., Bredenbeek P.J., Dobbe J.C., Thiel V., Ziebuhr J., Poon L.L., Guan Y., Rozanov M., Spaan W.J. and Gorbalenya A.E., Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331: 991–1004, 2003.
- Fielding B.C., Tan Y.J., Shuo S., Tan T.H., Ooi E.E., Lim S.G., Hong W. and Goh P.Y., Characterization of a unique group-specific protein (U122) of the severe acute respiratory syndrome coronavirus. *J. Virol.* 78: 7311–7318, 2004.
- Tan Y.J., Teng E., Shen S., Tan T.H., Goh P.Y., Fielding B.C., Ooi E.E., Tan H.C., Lim S.G. and Hong W., A novel severe acute respiratory syndrome coronavirus protein, U274, is transported to the cell surface and undergoes endocytosis. *J. Virol.* 78: 6723–6734, 2004.
- Yu C.J., Chen Y.C., Hsiao C.H., Kuo T.C., Chang S.C., Lu C.Y., Wei W.C., Lee C.H., Huang L.M., Chang M.F., Ho H.N. and Lee F.J., Identification of a novel protein 3a from severe acute respiratory syndrome coronavirus. *FEBS Lett.* 565: 111–116, 2004.
- Tan Y.J., Fielding B.C., Goh P.Y., Shen S., Tan T.H., Lim S.G. and Hong W., Overexpression of 7a, a protein specifically encoded by the severe acute respiratory syndrome coronavirus, induces apoptosis via a caspase-dependent pathway. *J. Virol.* 78: 14043–14047, 2004.
- Maniatis T., Fritsch E.F. and Sambrook J., *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory, NY, 1982.
- Farrow N.A., Muhandiran 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., Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* 33: 5984–6003, 1994.
- Keller R., *Computer Aided Resonance Assignment Tutorial*, 1st ed., Cantina Verlag, 2004.
- Nilges M. and O'Donoghue S.I., Ambiguous NOEs and automated NOE assignment. *Progr. Nucl. Magn. Res. Spectr.* 32: 107–139, 1998.
- Nilges M., A calculation strategy for the structure determination of symmetric dimers by 1H NMR. *Proteins* 17: 297–309, 1993.
- Koradi R., Billeter M. and Wuthrich K., MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14: 51–55, 1996.
- Laskowski R.A., Rullmann J.A.C., MacArthur M.W., Kaptein R. and Thornton J.M., AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8: 477–486, 1996.
- Vriend G., What If – a molecular modeling and drug design program. *J. Mol. Graph.* 8: 52–56, 1990.
- Kabsch W. and Sander C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577–2637, 1983.
- Wang J. and Springer T.A., Structural specializations of immunoglobulin superfamily members for adhesion to integrins and viruses. *Immunol. Rev.* 163: 197–215, 1998.
- Bork P., Holm L. and Sander C., The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* 242: 309–320, 1994.
- Casasnovas J.M., Stehle T., Liu J.H., Wang J.H. and Springer T.A., A dimeric crystal structure for the N-terminal two domains of intercellular adhesion molecule-1. *Proc. Natl. Acad. Sci. USA* 95: 4134–4139, 1998.

25. Casasnovas J.M., Springer T.A., Liu J.H., Harrison S.C. and Wang J.H., Crystal structure of ICAM-2 reveals a distinctive integrin recognition surface. *Nature* 387: 312–315, 1997.
26. Vigers G.P., Anderson L.J., Gaffes P. and Brandhuber B.J., Crystal structure of the type-I interleukin-1 receptor complexed with interleukin-1 β . *Nature* 386: 190–194, 1997.
27. Holm L. and Sander C., Touring protein fold space with Dali/FSSP. *Nucl. Acids Res.* 26: 216–319, 1998.
28. Wang C.Y., Naka Y., Liao H., Oz M.C., Springer T.A., Gutierrez-Ramos J.C. and Pinsky D.J., Cardiac graft intercellular adhesion molecule-1 (ICAM-1) and interleukin-1 expression mediate primary isograft failure and induction of ICAM-1 in organs remote from the site of transplantation. *Circ. Res.* 82: 762–772, 1998.
29. Gahmberg C.G., Tolvanen M. and Kotovuori P., Leukocyte adhesion – structure and function of human leukocyte beta2-integrins and their cellular ligands. *Eur. J. Biochem.* 245: 215–232, 1997.
30. Shimaoka M., Xiao T., Liu J.H., Yang Y., Dong Y., Jun C.D., McCormack A., Zhang R., Joachimiak A., Takagi J., Wang J.H. and Springer T.A., Structures of the alpha L I domain and its complex with ICAM-1 reveal a shape-shifting pathway for integrin regulation. *Cell* 112: 99–111, 2003.
31. Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C. and Ferrin T.E., UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25: 1605–1612, 2004.
32. Barton G.J., Alscript – a tool to format multiple sequence alignments. *Protein Engineering* 6: 37–40, 1993.
33. Diamond M.S., Staunton D.E., Marlin S.D. and Springer T.A., Binding of the integrin Mac-1 (CD11b/CD18) to the third immunoglobulin-like domain of ICAM-1 (CD54) and its regulation by glycosylation. *Cell* 65: 961–971, 1991.
34. Diamond M.S., Garcia-Aguilar J., Bickford J.K., Corbi A.L. and Springer T.A., The I domain is a major recognition site on the leukocyte integrin Mac-1 (CD11b/CD18) for four distinct adhesion ligands. *J. Cell. Biol.* 120: 1031–1043, 1993.
35. Vorup-Jensen T., Ostermeier C., Shimaoka M., Hommel U. and Springer T.A., Structure and allosteric regulation of the alpha X beta 2 integrin I domain. *Proc. Natl. Acad. Sci. USA* 100: 1873–1878, 2003.
36. Van der Vieren M., Le Trong H., Wood C.L., Moore P.F., St John T., Staunton D.E. and Gallatin W.M., A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-3. *Immunity* 3: 683–690, 1995.
36. Fortin J.F., Cantin R., Lamontagne G. and Tremblay M., Host-derived ICAM-1 glycoproteins incorporated on human immunodeficiency virus type 1 are biologically active and enhance viral infectivity. *J. Virol.* 71: 3588–3596, 1997.
38. Schubert U. and Strebel K., Differential activities of the human immunodeficiency virus type 1-encoded Vpu protein are regulated by phosphorylation and occur in different cellular compartments. *J. Virol.* 68: 2260–2271, 1994.
39. Schubert U., Antón L.C., Bacók I., Cox J.H., Bour S., Bennink J.R., Orłowski M., Strebel K. and Yewdell J.W., CD4 glycoprotein degradation induced by human immunodeficiency virus type 1 Vpu protein requires the function of proteasomes and the ubiquitin-conjugating pathway. *J. Virol.* 72: 2280–2288, 1998.
40. Bleijs D.A., Geijtenbeek T.B., Figdor C.G. and van Kooyk Y., DC-SIGN and LFA-1: a battle for ligand. *Trends Immunol.* 22: 457–463, 2001.
41. Arnaout M.A., Leukocyte adhesion molecules deficiency: its structural basis, pathophysiology and implications for modulating the inflammatory response. *Immunol. Rev.* 114: 145–180, 1990.
42. Yamaguchi N., Kubo C., Masuhiro Y., Lally E.T., Koga T. and Hanazawa S., Tumor necrosis factor alpha enhances *Actinobacillus actinomycetemcomitans* leukotoxin-induced HL-60 cell apoptosis by stimulating lymphocyte function-associated antigen 1 expression. *Infect. Immun.* 72: 269–276, 2004.
43. Cui W., Fan Y., Wu W., Zhang F., Wang J.Y. and Ni A.P., Expression of lymphocytes and lymphocyte subsets in patients with severe acute respiratory syndrome. *Clin. Infect. Dis.* 37: 857–859, 2003.
44. Cloyd M.W., Chen J.J. and Wang I., How does HIV cause AIDS? The homing theory. *Mol. Med. Today* 6: 108–111, 2000.
45. Rösch P. and Willbold D., Is EIAV Tat protein a homeodomain? *Science* 272: 1672, 1996.
46. Willbold D., Metzger A., Sticht H., Voit R., Gallert K.C., Dank N., Bayer P., Krauss G., Goody R.S. and Rösch P., Equine infectious anemia virus transactivator is a homeodomain type protein. *J. Mol. Biol.* 277: 749–755, 1998.
47. Willbold D., Hoffmann S. and Rösch P., Secondary structure and tertiary fold of the human immunodeficiency virus protein U(Vpu) cytoplasmic domain in solution. *Eur. J. Biochem.* 245: 581–588, 1997.
48. González M.E. and Carrasco L., The human immunodeficiency virus type 1 Vpu protein enhances membrane permeability. *Biochemistry* 37: 13710–13719, 1998.